globdiversity

rs-enabled ebvs for terrestrial ecosystems

Algorithm Theoretical Basis Document

Canopy Chlorophyll Content

Document Ref:	ATBD-CCC
Version:	3.4
Contract No:	4000120011/17/I-NB
Creation Date:	2018-05-03
Last Modified:	2020-09-30

Submitted by

Distribution List

This GlobDiversity document is currently stored at the Dropbox are under the folder Dropbox\03_Tasks\03_Task3\01_ATBD

This document is accessible to all participants of the GlobDiversity project.

Summary

This document refers to the activities of task 2 and 3, including sub-packages. This document is a draft and the version 2.0 of the Algorithm Theoretical Baseline Document (ATBD).

Document Release Sheet

Authors:	Andrew Skidmore (UT) Abebe Ali (UT) Roshanal Darvishzadeh (UT) David Marshall (DLR) Valentin Ziel (DLR) Ruben van de Kerchove (VITO)	
Approval (internal)	Claudia Röösli (UZH) Vladimir Wingate (UZH)	
Approval (ESA)	Marc Paganini (ESA)	
Distribution:		

Change Record

#	*Date*	*Page(s)*	*Change record*	*Release*
0	2018-05-04		First document draft	0.1
1	2018-11-29			0.1
2	2020-01-31		Finalization of current implementation status	2.0
3	2020-03-20		Updates in Chapter 4, 5 and 6	3.0
3	2020-03-26		Update for CCC LUT approach and new formatting	3.1
3	2020-05-20		Final edits	3.2
4	2020-08-28		Added upscaling	3.3
5	2020-09-30		Finalisation	3.4

Applicable and reference documents

The following documents apply to the extent specified herein.

Document ID	Document Title	Issue
AD-1	Project study report	Latest
AD-2		
...		

Table of Contents

1. Introduction 1

2. Scientific background 3

3. Input Data 6

3.1. Input Satellite and image types 6

3.2. Image Preprocessing 6

3.3. Sentinel 2 product 7

3.4. Land cover product 7

3.5. Sentinel-2 SCL Scene Classification Layer 8

3.6. RTM input parameters 8

3.7. Final input product 9

4. Algorithm description 11

4.1. Theoretical description 11

4.2. Algorithms overview 12

4.2.1. RTMs inversion 12

4.2.2. The simple ratio vegetation indices 15

4.3. Invalid CCC values 17

4.4. Source code 17

4.4.1. General 17

4.4.2. SRVI approach 17

4.4.3. LUT approach 18

5. Product 19

5.1. Product description 19

5.2. Accuracy analysis (expected accuracy) 19

5.3. Error budget 20

5.4. Validation (verification) summary 21

6. Practical considerations for implementation 23

6.1. Memory requirements 23

6.2. System requirements 23

6.3. Error handling 23

6.4. External databases 23

6.5. Manual interaction 23
6.6. Algorithm validation 24
7. Upscaling results 25
7.1. Introduction 25
7.2. Results 26
8. References 30

globdiversity rs-enabled ebvs

esa

List of Figures

Figure 1: Scene Classification Layer 8

Figure 2: Overview of the input parameters of RTMs (i.e., Leaf and canopy parameters, sensor configurations and background reflectance) to simulate top of canopy reflectance (R(λ))........ 9

Figure 3: Virtual Layer Stack (VRT) for the CCC algorithm........ 10

Figure 4: Canopy chlorophyll content generation workflow using RTMs inversion and SRVIs on Sentinel-2 surface reflectance data. The dark arrows show the process of the main algorithm (RTM inversion) and the red lines the process of SRVIs (optional algorithm)........ 12

Figure 5: SRVI algorithm implementation Work flow........ 18

Figure 6: Canopy chlorophyll content (g/m^2) derived from Sentinel-2 data using INFORM inversion (a) and SRVI (b) in Bavarian forest national park, Germany. 19

Figure 7: Scatter plot of the measured (in situ) CCC and predictions made by candidate algorithms: validation results of INFORM inversion by using LUT (a) and the Simple Ratio Vegetation Index (SRVI) (b). The line in Black colours show the 1:1 relationship, whilst the line in red indicates the relationship between the fields measured and predicted values of CCC. 21

Figure 8: 2018 Corine land cover product for Finland at 20 m resolution available from https://www.syke.fi/en-US/Open_information/Spatial_datasets/Downloadable_spatial_dataset#C...... 25

Figure 9: CCI 2016 20 m land cover prototype over Senegal available from http://2016africalandcover20m.esrin.esa.int/........ 26

Figure 10: mosaic of several CCC images taken on December 26th 2019 over Senegal. The Sentinel-2 orbit (290 km width) is clearly visible. White areas are either not covered on that day by the S2 orbit or are masked due to either the SCL flags or LC flags........ 27

Figure 11: per pixel CCC time series percentiles over Finland: 10th percentile (P10), 50th percentile (P50), 90th percentile (P90) and an RGB composite with R= P10, G= P50 and B=P90) 27

Figure 12: per pixel CCC time series percentiles over Senegal: 10th percentile (P10), 50th percentile (P50), 90th percentile (P90) and an RGB composite with R= P10, G= P50 and B=P90) 28

Figure 13: per pixel CCC time series percentiles over a region (Karkkila, 60°31'57.9"N 24°12'42.9"E) in Finland (see Figure 11 for the legend): 10th percentile (P10), 50th percentile (P50), 90th percentile (P90) and an RGB composite with R= P10, G= P50 and B=P90). Additionally, also the Corine Land Cover (see Figure 8 for the legend) and a Bing maps image is given as a reference. For the legend, the reader is referred to Figure 11 and Figure 8. 29

List of Tables

Table 1: Spectral features and their relationship with vegetation biochemical content (Curran et al., 2018). 3

Table 2: Classes and codes of the global land cover (GLC) product as adapted for the generic land cover classes required for the proposed algorithms........ 7

Table 3: INFORM input parameters used to generate the LUT as defined based on literature review and sensor configuration (sentinel-2 MSI)........ 13

Table 4: The PROSAIL radiative transfer model input parameters used to generate the lookup table. 14
Table 5: summary of the CCC upscaling processing results.. 26

Terms, Definitions and Abbreviations

Acronym	Definition
ALA	Average Leaf Angle
ATBD	Algorithm Theoretical Basis Document
BOA	Bottom of Atmosphere
BRF	Bidirectional Reflectance Function
Cab	Leaf chlorophyll content
CC	Canopy Closure
CCC	Canopy Chlorophyll Content
CD	Crown Diameter
C_m	Leaf dry mass per unit leaf area
C_w	Leaf water mass per unit leaf area
EBV	Essential Biodiversity Variables
EMS	Electromagnetic Spectrum
INFORM	Invertible Forest Reflectance Model
LAI	Leaf Area Index
LAI_s	Single tree LAI
N	Leaf (mesophyll) internal structure
NIR	Near-Infrared
PROSAIL	The PROSPECT leaf model and Scattering by Arbitrarily Inclined Leaves
PSR	Project Study Report
RS-enabled EBV	Remote sensing enabled Essential Biodiversity Variable
SD	Stem density (number of trees per hectare)
SH	Stand height
SRVI	Simple Ratio Vegetation Index
TOA	Top of Atmosphere
TOC	Top of Canopy
VI	Vegetation Index

1. Introduction

This Algorithm Theoretical Baseline Document (ATBD) describes all technical issues from the prototyping of the Canopy Chlorophyll Concentration (CCC) in the context of the Remotely Sensed Essential Biodiversity Variables (RS-enabled EBVs) product of the ESA funded GlobDiversity Project. This document shall specify the process flow of the prototyped algorithm and the associated program in more detail.

GlobDiversity is the first large-scale project explicitly designed to develop and engineer RS-enabled EBVs. This project initiated and funded by the European Space Agency (ESA) supports the efforts of the Convention on Biological Diversity (CBD) and Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), among others, and is adopted under the umbrella of the Group on Earth Observations Biodiversity Observation Network (GEO BON). The GlobDiversity project shall support the initiative to build a global knowledge of biological diversity of terrestrial ecosystems (= on land) and of relevance for society.

There are three RS-enabled EBVs designed as part of the GlobDiversity project with each algorithm documented by such an ATBD:

- Fragmentation (Wageningen Environmental Research WEnR, Wageningen University & Research)
- Canopy chlorophyll concentration (Faculty of Geo-Information Science and Earth Observation ITC, University of Twente)
- Land surface phenology (Dept. of Geography, University of Zurich (UZH), the hereby documented algorithm)

Within the project, these three variables were investigated in detail to contribute to an observation system to assess the variable in an efficient and effective way, covering extensive areas at a fine spatial and temporal resolution. The definition and selection, name and definition of the three RS-enabled EBVs was based on the expertises existing within the project consortium and independent from any efforts of defining and prioritising possible candidate EBVs and RS-enabled EBVs that might have existed at the time of the project's start in 2018.

In the following, the algorithm of the processing chain to derive Canopy Chlorophyll Concentration (CCC) is described in detail. The algorithm was chosen and developed by the University of Twente (ITC) and then transmitted to the German Aerospace Center (DLR) to be translated into a code suitable for cloud computing of larger areas of interest. The algorithm has been chosen and developed with the goal of a potential future global application based on high-resolution satellite data (10-30m) and with a computational efficient implementation. The ATBD includes a description of the necessary preprocessing steps and the processing step of the core algorithm. In addition, results from the project performed on a few test sites globally distributed are presented with a chosen validation approach. In addition, the last chapter presents restrictions of the current implementation and modifications that might be necessary for a potential global processing.

The organization of this document is structured in 8 chapters as shown in the table below.

	Explanation
Chapter 1	Provides an introduction
Chapter 2	Describes the scientific background, and addresses the current standard processing schemes
Chapter 3	Provides information about the input data
Chapter 4	Includes the algorithms of the proposed processing
Chapter 5	Provides information about the product and the error budget estimates
Chapter 6	Outlines the implementations
Chapter 7	Shows the results of the upscaling experiments for Senegal and Finland
Chapter 8	Includes the references

2. Scientific background

It is argued that the chlorophylls are Earth's most important organic molecules as they are necessary for photosynthesis (Blackburn, 2007). Plants chlorophylls are of two types, chlorophyll-*a,* and chlorophyll-*b*. Canopy chlorophyll Content (CCC) is defined as "the total amount of chlorophyll *a* and *b* pigments in a contiguous group of plants per unit ground area often expressed in mg/m^2" (Gitelson et al., 2005). CCC is a compound variable. It is a product of chlorophyll content of a fresh green leaf per unit leaf area, and leaf area index (LAI). CCC is a terrestrial ecosystem functional EBV that describes chlorophyll pigments (distribution within the 3D canopy surface. Thus, CCC determines the total photosynthetically active radiation absorbed by vegetation (Gitelson et al., 2015, and 2005). Quantification of CCC has been long used for a wide range of ecological applications from being an important input variable of terrestrial biosphere models to quantify carbon and water fluxes (Luo et al., 2018) to primary productivity prediction (Houborg et al., 2013, Peng and Gitelson, 2011), and light use efficiency assessment (Wu et al., 2012). Changes in CCC are an indicator of vegetative growth, disease, nutritional and environmental stresses (Korus, 2013, Zhao et al., 2011, Inoue et al., 2012). It is a plant pigment that provides valuable information about plant physiology and ecosystem processes (functions) at different scales and enables the ecologists, farmers, and decision-makers to assess the influence of climate change, and other anthropogenic and natural factors on plant functions. Monitoring the dynamics of CCC helps to understand the adaptation of forest, crops, and other vegetation canopies to such factors (Féret et al., 2017).

Remote sensing data provide a unique way to measure vegetation properties. The spectral reflectance of vegetation is characterized by absorption features due to a chemical constituent of leaves such as chlorophyll, water, nitrogen, and carbon-containing compounds, comprising primarily protein, lignin, and cellulose. When incoming radiation interacts with vegetation, some part of it is reflected, some absorbed and rest is transmitted. A typical reflectance spectrum of a vegetation canopy can be subdivided into three parts, visible (400- 700 nm), near-infrared (NIR) (701 – 1300 nm) and middle-infrared (1301- 2500 nm) (Curran et al., 2018). The visible region mainly contains information about leaf pigments (i.e., chlorophyll *a*&*b*, carotene, and xanthophyll), the NIR domain about leaf structure, and the middle-infrared region contains information about the absorption of radiation by water, cellulose, and lignin (Table 1).

Table 1: Spectral features and their relationship with vegetation biochemical content (Curran et al., 2018).

Waveband	Waveband width (nm)	Characteristics	Relation to vegetation amount
Ultraviolet/blue	350-500	Strong chlorophyll and carotene absorption	Strong negative
Green	500-600	Reduced level of pigment absorption	Weak positive
Red	600-700	Strong chlorophyll absorption	Strong negative
Red edge	700-740	Transition between strong	Weak negative

		absorption and strong reflectance	
Near-infrared	740-1300	High vegetation reflectance	Strong positive
Middle-infrared	1300-2500	Water, cellulose, and lignin absorption	Not specific

Chlorophyll is the major absorber of radiation in the visible region, and the level of absorption can be used for retrieving CCC from remote sensing data for spatially extensive areas from the landscape to the global scale, using the strong and distinctive chlorophyll absorption bands in the visible spectral region. Numerous algorithms combining reflectance at discrete spectral wavelengths have been constructed to amplify the sensitivity of the spectral reflectance to chlorophyll content (Main et al. 2011). Theoretically, the "ideal" CCC retrieval algorithm from RS data should be sensitive only to chlorophyll content, but insensitive or little affected by any other factors. However, it is impossible to design an algorithm which is sensitive only to the desired variable and entirely insensitive to all other parameters (Hunt et al., 2012). Consequently, CCC retrieval algorithms remain always sensitive to the artifacts caused by canopy structure, radiometric distortions due to topography, atmosphere, solar illumination geometry, and sensor viewing conditions, and soil optical properties particularly in sparse vegetation (Darvishzadeh et al., 2008a, Ollinger, 2011, Gitelson et al., 2005).

Canopy structure controls how the photons scatter within the canopy before escaping or being absorbed. One of the most common canopy structure descriptors is the leaf area index (LAI) that quantifies the number of leaf layers with which a photon can potentially interact (Ollinger, 2011). Another canopy structural variable that affects the relationship between CCC and reflectance is leaf angle distribution (LAD). The more vertically aligned the leaves, the deeper the light penetrates within the canopy (Ellsworth and Reich, 1993). Therefore, canopy structure quantified by LAI and LAD affects the accuracy of CCC retrieval from remote sensing data.

In addition, bare soils have different spectral properties depending on mineral composition, color, moisture, organic matter content, salt and Sodium content, roughness, and texture. These soil property variations affect the spectral response of soil and canopies and induce noise to the relationship between canopy reflectance and vegetation parameters, such as CCC (Darvishzadeh et al., 2008a, Ollinger, 2011, Gitelson et al., 2005). Many studies have evaluated and compared the sensitivity of CCC retrieval algorithms to these structural, atmospheric, and soil optical property conditions (e.g., Vincini et al., 2016, Wu et al., 2008, Niemann et al., 2012).

Nevertheless, several methods ranging from empirical to 3D radiative transfer model (RTM) inversion can be used to estimate CCC from remote sensing data. Algorithms that rely on the red-edge and near-infrared region reflectance spectra are proved to be less sensitive to external factors and capable of accurately estimating CCC. The provision of reflectance spectra in narrow red and NIR bands with a high spatial resolution by Sentinel-2 spacecraft instrument enables to perform vegetation monitoring via CCC estimated from remote sensing data. A plethora of algorithms exist in the literature, and it is imperative to identify the one operationally feasible for CCC. Therefore, through literature review, experimental analysis, and robustness verification across biomes it is proposed to use physically based models (i.e., INFORM (Invertible Forest Reflectance Model) for forest ecosystem and PROSAIL (the PROSPECT leaf model and Scattering by Arbitrarily Inclined Leaves) for short plant ecosystems) as the standardized and harmonized approach to retrieve CCC products from Sentinel-2 MSI. The simple ratio vegetation index (SRVI) based on the red band and near-infrared are also proposed as a backup algorithm.

3. Input Data

3.1. Input Satellite and image types

The RS-enabled EBV-CCC products development focuses here on passive Earth observation satellites that are specifically designed for environmental monitoring. Thus, Earth observation satellites, which carry instruments that operate in the visible and near-infrared region of the electromagnetic spectrum (EMS) with high spatial and high spectral resolution, are the preferred satellite types. A sun-synchronous polar or near-polar orbit satellite generally suited for generating CCC global products for monitoring change in the state of biodiversity over time.

Prediction of CCC from remote sensing data using RTM inversion demands spectral information in a number of chlorophyll sensitive bands (wavelengths) in the visible and NIR region of the EMS. Thus, panchromatic image types with only one band do not fit the purpose. Hyperspectral images with hundreds of very narrow discrete/contiguous bands are the most superior and Input Satellite and image types preferable but rarely available. Satellite imageries with 10 - 15 spectral bands with several bands in the visible and NIR regions together with 10 - 30m ground resolution are the good alternative imagery for accurate mapping of global CCC products. Therefore, the potential input satellite data focus on the new data streams such as those offered by the Sentinel-2 MSI mission. Sentinel-2 provides multispectral images with bands in the visible, NIR and MIR spectrum. Data from this satellite are available at 10, 20 and 60 m spatial resolution that can significantly contribute to CCC mapping. In this approach, we only use data with a resolution of 10 m and 20 m. For a comparable data set and image cube, the images were resampled to 10 m (GDAL build VRT (Virtual Dataset) options were set to: resolution = highest and separate = true. The last parameter creates a new layer for every stacked band).

3.2. Image Preprocessing

Predictions of global CCC products are based on top of the canopy (TOC) or surface reflectance data. Satellite imagery radiance records have to be converted into TOC reflectance through radiometric and atmospheric correction. The remote sensing data must be orthorectified and georeferenced using standard sensor in-flight information. The images have to be resampled (if necessary) and mosaicked by removing the view angle effects via view-angle correction technique, and applying filtering techniques to correct for random and systematic noise. Finally, non-vegetated areas and clouds must be masked out.

In this study, the experimental analysis focused on the evaluation of algorithms for CCC retrieval from Level-2A products of Sentinel-2 data. The preprocessing includes an atmospheric correction applied to Top-Of-Atmosphere (TOA) Level-1C ortho-image Sentinel-2 products, which resulted in an ortho-image Bottom-Of-Atmosphere (BOA) corrected reflectance product or TOC. The Level-2A generation was performed through Sen2Cor Version 2.5 using the Sentinel-2 Level-1C product as input.

3.3. Sentinel 2 product

As will be described later, not all scenes and bands from the given Sentinel-2 images are necessary. Only the following bands, which are required for the calculations of CCC, are used here: 04, 05, 06, 08 and 8A. Bands 04 and 08 were available in the spatial resolution of 10 meters, and bands 05, 06, 8A and SCL (see later chapter SCL) were available in the spatial resolution of 20 meters. The highest spatial resolution was always chosen if scenes were available in different resolutions.

3.4. Land cover product

The proposed algorithms depend on the land cover type. On the one hand, a land cover map is used to segment vegetation and non-vegetated parts of the terrestrial ecosystems. On the other hand, it is used to group the vegetated areas into 'short vegetation,' and forest so that selected algorithms for each group applied separately. Vegetation types such as grasses, crops, shrubs, and bushes are considered 'short vegetation' and vegetation cover with tree treated as forest. Preferably, a land cover product with a higher spatial resolution ($\leqslant$ 30 meters) should be used for this purpose. Gong et al. (2019) have recently generated a high-resolution (10m) global land cover (GLC) product for the year 2017, which is freely and openly available at http://data.ess.tsinghua.edu.cn/fromglc10_2017v01.html. The ten GLC classes of this product are grouped into three generic classes that can be utilized for both masking and stratifying into short vegetation, and forest (Table 2).

A global land cover dataset was downloaded to the processing servers. In the next step, a virtual, global scene was created using GDAL VRT, which serves as the basis for the following steps.

Table 2: Classes and codes of the global land cover (GLC) product as adapted for the generic land cover classes required for the proposed algorithms

GLC class name	GLC class Code	Generic class	Generic class code
Cropland	10	Short vegetation	1
Grassland	30		
Shrubland	40		
Wetland	50		
Tundra	70		
Forest	20	Forest	2
Water	60	Non vegetation	0
Impervious surface	80		
Bareland	90		

Snow/Ice	100		

3.5. Sentinel-2 SCL Scene Classification Layer

Scene Classification (SC) aims at providing a pixel classification map (cloud, cloud shadows, vegetation, soils/deserts, water, snow, etc.). This classification map is used to exclude pixels which are not suitable for further processing and would falsify the result. For full information about the process and the algorithms visit the ESA webpage. The classification mask is generated along with the process of generating the cloud mask quality indicator and by merging the information obtained from cirrus cloud detection and cloud shadow detection. The classification map is produced for each SENTINEL-2 Level-1C product at 20 m resolution, and byte values of the classification map are organized as shown in Figure 1.

Label	Classification
0	NO_DATA
1	SATURATED_OR_DEFECTIVE
2	DARK_AREA_PIXELS
3	CLOUD_SHADOWS
4	VEGETATION
5	NOT_VEGETATED
6	WATER
7	UNCLASSIFIED
8	CLOUD_MEDIUM_PROBABILITY
9	CLOUD_HIGH_PROBABILITY
10	THIN_CIRRUS
11	SNOW

Figure 1: Scene Classification Layer

For the product generation, only label number four (Vegetation) was used for further processing. Although also other classes contain information about vegetation, a conservative strategy was chosen for this prototype. Moreover, there the quality of this mask needs further evaluation. Therefore, results should be treated with caution. For example, the dark areas can also belong to dark vegetation and the unclassified layer (number seven) can also contain pixels that are necessary for processing. In this approach, it was more important to classify as well as possible and then to continue working with reliable masks. The change for other or many classes can be easily modified and adapted later in the source code.

3.6. RTM input parameters

The proposed RTM inversion methods require generating a large spectral lookup table (LUT) by varying leaf and canopy parameters as well as sensor geometry. Leaf reflectance and transmittance properties are simulated using the leaf model-PROSPECT, which is integrated with the two canopy RTMs. The main leaf input parameters required for PROSPECT includes chlorophyll content (Cab) in µg/cm^2, leaf dry mass per unit area (C_m) in mg/cm^2, leaf water mass per unit area (C_w) in mg/cm^2, and effective number of Leaf (mesophyll) internal structure (N) among others. The leaf level spectral property is up-scaled to canopy using the PROSAIL and INFORM canopy RTMs by using the leaf level simulated spectra, canopy structural variables, and sensor geometry (Figure 2). The type and number of canopy structural input parameters vary depending on the canopy RTM. Type and value range of each model input parameter are described in section 5.

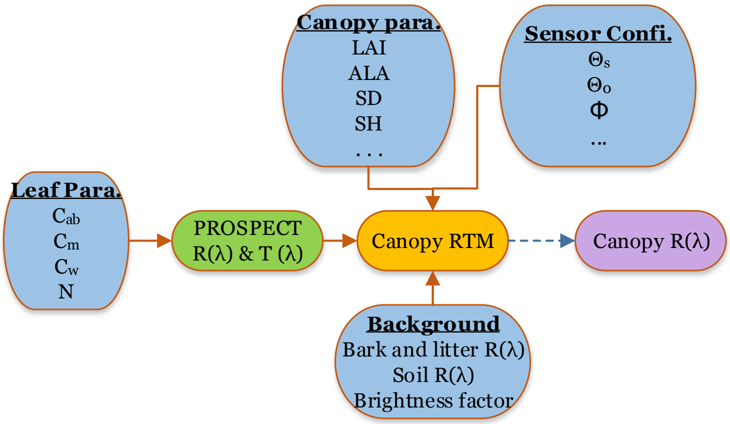

Figure 2: Overview of the input parameters of RTMs (i.e., Leaf and canopy parameters, sensor configurations and background reflectance) to simulate top of canopy reflectance (R(λ))

3.7. Final input product

The final input product is a data stack of all required classes, scenes and images (Figure 3). This product is used by both, the SRVI and the LUT prototype. This product is a layer stack built with the GDAL open source tools and is called VRT (Virtual Dataset). The VRT driver is a format driver for GDAL that allows a virtual GDAL dataset to be composed from other GDAL datasets with repositioning and algorithms potentially applied, as well as altering or adding to the metadata. VRT descriptions of datasets can be saved in an XML format and are normally given the extension .vrt. The program builds a VRT that is a mosaic of the list of input GDAL datasets. The list of input GDAL datasets can be specified. With the option “separate”, each file goes into a separate band in the VRT dataset. Otherwise, the files are considered as tiles of a larger mosaic and the VRT file has as many bands as one of the input files. The use of virtual layers leads to a minimal copying effort on the processing systems, since no data has to be copied, but only virtual work is done on the scenes. The command *gdalbuildvrt* is used with the “separate” option to build the input data stack. The data stack includes the required Sentinel 2 bands for the SRVI and LUT calculations (Band 4, 5, 6, 8 and 8A); the land cover classification; and the cloud cover (Figure 5).

All images were spatially resampled to 10 m (close to sample plots size) by using GDAL buildvrt option “resolution” (see: https://gdal.org/programs/gdalbuildvrt.html, visited 23.03.2020) with the parameter “highest” (in this case: 10 meters, because of band 04 and 08 and global landcover scene). The global landcover VRT scene was cropped to the specific extension of the Sentinel 2 scene. The fourth band of the Sentinel 2 scene always serves as the basis for this. The cutting was done with a Python script called Rasterio. Afterwards, the cropped Landcover scene could be inserted into the layer stack without problems. Cloud and its shadow were identified by the integrated Sentinel-2 scene classification layer (SCL) and masked for this specific study by the automated processing. Non-vegetative areas masked out using available land cover products by using the generic class code zero.

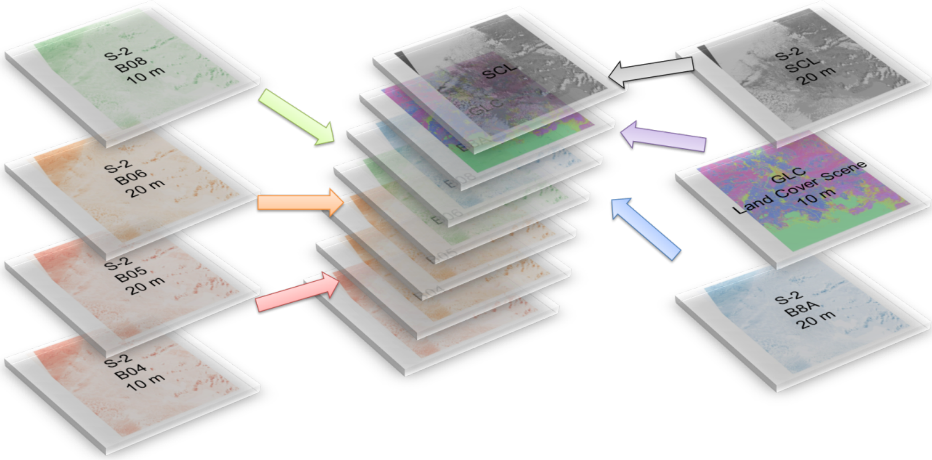

Figure 3: Virtual Layer Stack (VRT) for the CCC algorithm

4. Algorithm description

4.1. Theoretical description

This part of the Algorithmic Theoretical Basis Document (ATBD) describes the algorithms proposed to produce Canopy chlorophyll content (CCC) over terrestrial ecosystems from atmospherically corrected surface reflectance. After a thorough literature review, experimental analysis, robustness and spatiotemporal consistency check, we proposed RTMs inversion and the simple ratio vegetation index (SRVI) for canopy chlorophyll content retrieval. This section outlines the physical principles and mathematical background of these selected algorithms. The proposed algorithms follow the recommendations issued in the Project Study Report (PSR) that was reviewing the state-of-the-art algorithms available in the literature.

A review of the literature indicated that algorithms based on spectral information in the red band, the red-edge (which is a region within the red-NIR transition zone), and NIR are generally robust and could be applied for producing chlorophyll products. This is because there is a strong correlation between an increase in chlorophyll and an increase in the absorption of radiation energy in the red edge (Curran et al., 1990). The algorithms that can be applied to quantify chlorophyll using spectral information in the stated spectral region are plenty. Verrelst et al. (2015) grouped all algorisms that can be used to quantify vegetation biophysical and biochemical variables from remote sensing data into four methodological categories: Parametric regression, non-parametric regression, Physically-based, and Hybrid methods.

All approaches have their advantage and drawbacks. Some are superior to all other methods in the overall predictive ability of CCC, but not feasible for global application. Therefore, the selection of the algorithm should be a trade-off. This urges a hierarchical selection procedure. First, all algorithms with top predictive ability were identified from the vast spectrum literature. Second, selection requirement criteria were set and the methods evaluated by considering: i) volume of spectral information need ii) accuracy, and iii) computational efficiency. Subsets of algorithms are shortlisted based on those criteria, and finally, an experimental analysis performed to benchmark the algorithm that is applicable for global CCC product retrieval.

During the experimental analysis, benchmarking was determined by computing the coefficients of determination (R^2) between canopy chlorophyll content and the selected algorithms. For this purpose, canopy chlorophyll content collected from representative sample plots randomly distributed in the mixed mountain forest of the Bavarian Forest National Park (BFNP) together with TOC reflectance spectra extracted from Sentinel-2 L2A data in July 2017 were used. Algorithms with higher R^2 and lower RMSE (high accuracy) were further investigated for their operational use to retrieve CCC consistently from Sentinel-2 imagery irrespective of the difference in vegetation structure and composition across a wide range of vegetation types. Thus, through robustness verification, LUT based inversion of INFORM on forest ecosystems (Ali et al., 2020) and PROSAIL over non-forest (short plant) ecosystems as well as the simple ratio vegetation index with band 8a and band 4 of Sentinel-2 proved to be more robust than all other tested algorithms (AD-1).

Consequently, the INFORM and PROSAIL inversions using LUT are found to be a plausible algorithm and recommended for retrieval of CCC products in terrestrial ecosystems from remote sensing data. See AD-1 on the algorithm selection process.

4.2. Algorithms overview

The analytical workflow in generating CCC products using the benchmarked algorithms is illustrated in Figure 4. To apply the proposed algorithms, the preprocessed image data has to be segmented into the forest and non-forest vegetation followed by applying the corresponding RTM inversion or SRVIs and recombining the results to generate the required CCC product (s).

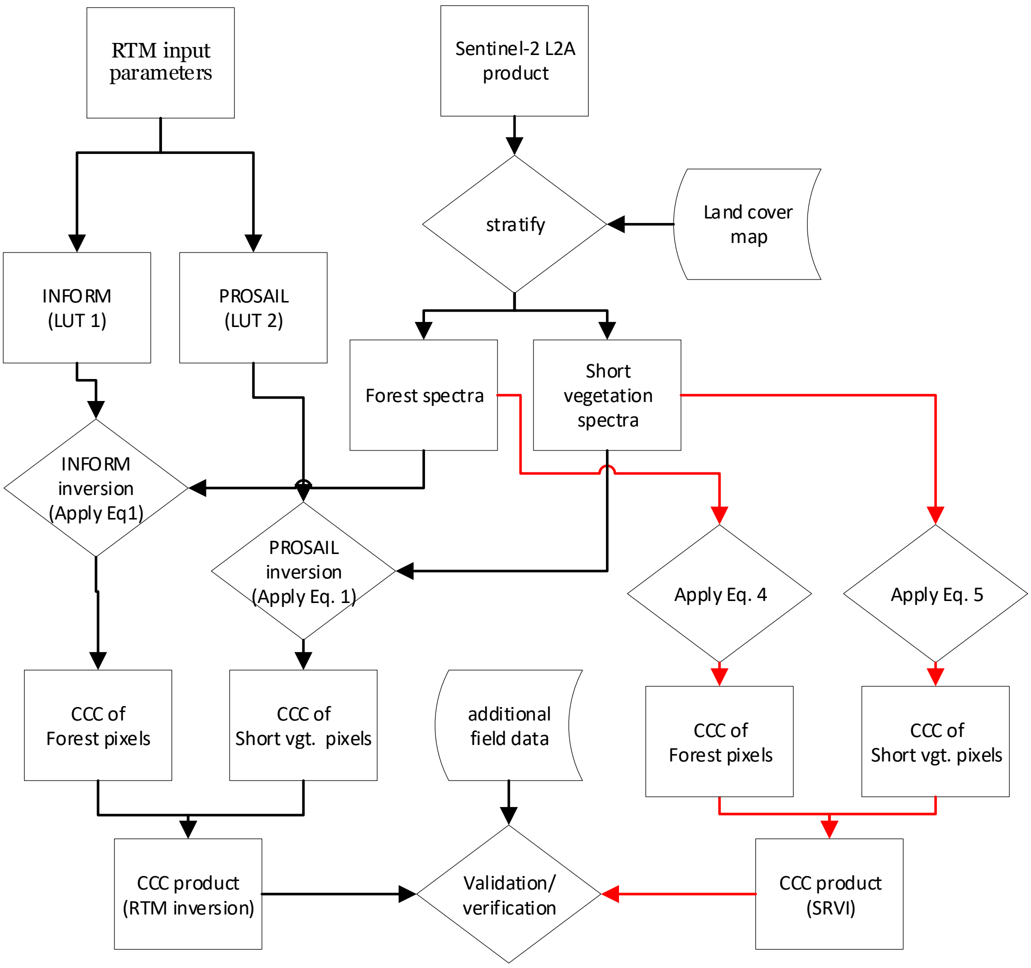

Figure 4: Canopy chlorophyll content generation workflow using RTMs inversion and SRVIs on Sentinel-2 surface reflectance data. The dark arrows show the process of the main algorithm (RTM inversion) and the red lines the process of SRVIs (optional algorithm).

4.2.1. RTMs inversion

a) Parameterization and generation of LUT using INFORM

The Invertible Forest Reflectance model "INFORM" (Schlerf and Atzberger, 2006, Atzberger, 2001) is a combination of the forest light interaction model (Rosema et al., 1992) and SAIL

(Verhoef, 1984) canopy RTMs with the PROSPECT (Jacquemoud and Baret, 1990) leaf RTM. INFORM is parameterized by leaf parameters such as C_m, C_w, C_{ab} and leaf Structure parameter (N), canopy parameters such as stem density (SD), single tree leaf area index (LAI_s), stand height (SH), crown diameter (CD) and average leaf angle (ALA), as well as external parameters such as, view zenith (θ_o), sun zenith (θ_s) and relative azimuth angle (Φ), and simulates canopy spectral reflectance of forest stands between the 400 and 2500 nm wavelengths'. The model has been successfully used to retrieve forest biophysical and biochemical parameters (Ali et al. 2020, 2017; Darvishzadeh et al. 2019; Schlerf et al. 2012; Wang et al. 2018). As demonstrated in Figure 4, INFORM (Canopy RTM) provides spectral reflectance on top of a canopy under specified conditions.

To simulate the spectral property of forest ecosystems, INFORM is parameterized based on the range of input parameters determined through literature review and sensor configurations (A LUT of 100,000 spectra is generated by varying the free variables randomly within their range. This size of PROSAIL LUT (though is smaller than LUT of INFORM due to less model input parameters), is confirmed to be large enough for retrieval of vegetation parameters in different ecosystem (e.g., Atzberger et al., 2015, Darvishzadeh et al., 2008b). A random Gaussian noise value of 0.3% was also added to each simulated spectrum to account for model uncertainties and reduce auto-correlation between the spectrum and input variables.

Table 3). In INFORM, LAI is represented by the leaf area indices of single trees. Hence, the ground truth values for LAIs were computed from Leaf area index (LAI) and canopy closure (CC) (i.e., LAI_s = LAI/CC).

A LUT of 100,000 spectra is generated by varying the free variables randomly within their range. This size of PROSAIL LUT (though is smaller than LUT of INFORM due to less model input parameters), is confirmed to be large enough for retrieval of vegetation parameters in different ecosystem (e.g., Atzberger et al., 2015, Darvishzadeh et al., 2008b). A random Gaussian noise value of 0.3% was also added to each simulated spectrum to account for model uncertainties and reduce auto-correlation between the spectrum and input variables.

Table 3: INFORM input parameters used to generate the LUT as defined based on literature review and sensor configuration (sentinel-2 MSI).

Parameter	Symbol	Unit	range or fixed values	
			Min	**Max**
Leaf dry mass per area	C_m	g/cm^2	0.005	0.03
Equivalent water thickness	C_w	g/cm^2	0.006	0.035
Leaf structural parameter	N	NA	1	2.5
Leaf chlorophyll content	C_{ab}	µg/cm^2	5	65
Single tree LAI	LAI_s	NA	2	10
Understory LAI	LAI_u	NA	0.2	1
Stem density	SD	n/hr	200	2000
Stand height	SH	m	5	40
Crown diameter	CD	m	3	10

Average leaf angle	ALA	degree	40	60
Sun zenith angle	θ_s	degree	25	35
Observation zenith angle	θ_0	degree	0	15
Azimuth angle	Φ	degree	50	210
Scale		NA	0.5	1.5
Fraction of diffused radiation	Sky1	fraction	0.1	

b) Parameterization and generation of LUT using PROSAIL

PROSAIL, which is a one-dimensional bidirectional turbid medium radiative transfer model, is used for the simulation of the canopy bidirectional reflectance data in 'short vegetation' such as wetlands, taiga, and tundra. There are different versions of PROSAIL. Here we used the latest version (version: PROSAIL-D), but other versions of the model (e.g., PROSAIL-H) can be used instead. PROSAIL requires leaf input parameters such as Cab, N, C_m, and C_w, and canopy and sensor configuration parameters. These canopy and sensor parameters are the sun zenith angle, observer zenith angle, relative azimuth angle, soil factor, LAI, hot spot size parameter, and two leaf inclination distribution function (LIDF) parameters. The solar and observation angle parameters are obtained from the metadata file of the remote sensing data (Sentinel-2). The hot spot is parameterized as a function of LAI based on previous studies (Yin et al., 2016). The parameters soil factor, Cab, N, C_m, C_w, and LAI are taken as free variables (Table 4). The model default values are used for other parameters. The variation ranges of the free variables are based on prior knowledge in the literature.

A LUT of 100,000 spectra is generated by varying the free variables randomly within their range. This size of LUT is confirmed large enough for retrieval of vegetation parameters in different vegetation (e.g., Atzberger et al., 2015, Darvishzadeh et al., 2008b). A random Gaussian noise value of 0.3% was also added to each simulated spectrum to account for model uncertainties and reduce auto-correlation between the spectrum and input variables.

Table 4: The PROSAIL radiative transfer model input parameters used to generate the lookup table

Parameter	Symbol	Unit	range or fixed values	
			Min	Max
Leaf dry mass per area	C_m	g/cm^2	0.005	0.025
Equivalent water thickness	C_w	cm	0.05	0.03
Leaf structural parameter	N		1.2	2.2
Chlorophyll content	C_{ab}	μg/cm^2	5	70
Carotenoid content	Car	μg/cm^2	8	
Anthocyanin content	Ant	μg/cm^2	0	
brown pigment content	C_{brown}		0	
Leaf area index	LAI	m^2/m^2	0.2	8

Leaf inclination distribution function type	TypeLidf		2	
Leaf inclination distribution function a	LIDFa	degree	20	70
Leaf inclination distribution function b	LIDFb		0	
Hot spot factor	H_{spot}		0.5/LAI	
Soil reflectance factor	p_{soil}		0.3	0.6
Sun zenith angle	t_s	degree	25	35
Observation zenith angle	t_0	degree	0	15
Azimuth angle	p_{si}	degree	50	210

c) LUTs Inversion

For both LUT generated by INFORM and PROSAIL, LUT inversion involved matching the similarity between measured spectra (Sentinel-2) and simulated spectra (INFORM or PROSAIL). Spectrum matching was performed using the least root mean square error (RMSE) comparison of the measured and simulated spectra according to Eq.1.

$$RMSE = \sqrt{\frac{\sum(R_{measured} - R_{modelled})^2}{n}} \qquad \textit{Eq. 1}$$

where $R_{measured}$ is a Sentinel-2 reflectance at wavelength λ and $R_{modelled}$ is a simulated reflectance at wavelength λ in the LUT, and n is the number of wavelengths.

LUT inversion is traditionally very slow. To speed up the processor, we use a kd-tree for quick nearest-neighbour lookup (Maneewongvatana and Mount, 1999). The increased efficiency of the processor comes from the fact that not every entry in the LUT is compared to the input spectral data. Instead, the binary tree splits the LUT into nodes, subspaces of the LUT, and the comparison is made against the data in these nodes. The sliding midpoint method ensures that the nodes are appropriately sized in every dimension. Here, the dimensions are the 3 Sentinel bands. It is especially important that a node is not too small in any one dimension. The binary tree can find the N closest neighbours of given spectra (where N, in this case, equals 100), and the algorithm can be made more efficient by choosing the approximate closest neighbours.

The inversion was carried out using those Sentinel-2 bands found between 650 and 750 nm (three bands: Band 4, 5 &6), which are found in the red and NIR transition zone wavelength range where an increase in chlorophyll content increases absorption. The solution to the inverse problem is the set of input parameters corresponding to the reflectance in the database that provided the smallest RMSE. Because of the potential insufficiency in model formulation and parameterization, and noise related to calibration and preprocessing errors in the observed reflectance, the least RMSE solution might not necessarily provide the best estimates. For this reason, for each measured spectrum, the 100 closest matching spectra are selected from the LUT. From the multiple available solutions (q), we chose the median CCC value of the multiple solutions as a final solution after experimenting the performance of other statistical measures of central tendency such as mean and mode for several closest matching spectral subsets.

4.2.2. The simple ratio vegetation indices

a) Mathematical description

Two simple ratio vegetation indices optimized for forests and non-forest vegetation were proposed as optional algorithms, which can be used for the comparative purpose. The SRVIs use spectral information in two spectral bands to compute CCC. Optimization of the benchmarked algorithms during experimental analysis revealed that SRVI based on spectral band 4 and band 8a of Sentinel-2 MSI outperformed other indices for retrieving CCC in forest ecosystems.

The simple ratio vegetation index 1 (SRVI_1) proposed for forest ecosystems is a ratio of reflectance of a band centred at 865 nm and 665nm. In terms of Sentinel-2 MSI band setting SRVI is computed using Eq. 2. And the simple ratio vegetation index 2 (SRVI_2) proposed for non-forest vegetation is the one proposed by Inoue et al. (2016), which is calibrated and validated for retrieving CCC from remote sensing data for a wide range of crops and natural grasses (Eq. 3).

$$\text{SRVI-1} = \frac{B8a}{B4} = \frac{R_{865}}{R_{665}} \qquad \text{(Eq. 2)}$$

$$\text{SRVI-2} = \frac{B8}{B5} = \frac{R_{835}}{R_{704}} \qquad \text{(Eq. 3)}$$

where R865, R835, R704, and R665 are reflectances in the centre wavelengths of the Sentinel-2 band setting.

As discussed in the previous sections, the computation of SRVI requires radiometrically and atmospherically corrected reflectance (Top-of-canopy reflectance). Besides, the selected algorithms have to be applied only on vegetative land covers. Therefore, vegetated and non-vegetated pixels have to be identified first. Globally available land cover products can be used for this purpose. Alternatively, a threshold value can be applied to the reflectance value in the red edge region to distinguish between the vegetated and bare land pixels. Nonvegetated land pixels, i.e., pixels which are barren have a high reflectance in this region. Therefore, pixels with high reflectance in the red edge region(e.g., 30% or more reflectance) can be considered non-vegetated/bare lands (Curran et al., 2018).

Likewise, pixels with cloud cover can be removed by observing the difference in reflectance between bands in the red edge and NIR. Particularly, the reflectance of pixels with low cloud cover is a mixture of reflectance from the ground and top of the cloud. Since such pixels do not indicate the actual ground vegetation condition, they have to be removed. Pixels with low cloud cover have a very similar reflectance in the transition zone between the visible and NIR region (Curran et al., 2018). Therefore, pixels with minimal reflectance difference in two or more bands of this region can be flagged as cloud mixed-pixels and avoided.

b) Fitting linear equations

The canopy chlorophyll content is retrieved by investigating the relationship between SRVIs and measured CCC values. This has been done by fitting an equation on the relationship. Equation 4 and equation 5 indicates the linear relationship between SRVIs extracted from Sentinel-2 level 2A product and ground measured CCC for forest and non-forest vegetation, respectively.

CCC (g/m^2) in forest = 0.071*SRVI-1 + 0.217 (Eq. 4)

CCC (g/m^2) in 'short vegetation' = 0.325 *SRVI-2 – 0.358 (Eq. 5)

4.3. Invalid CCC values

The CCC values have to be within the global CCC range (0-10 g/m^2) (Singh, 2018). Because of uncertainties associated with models and remote sensing data, predicted values may go beyond the expected range. In the case where any of the selected algorithms provide estimates outside this limit, it is removed during the processing as an invalid value. Values greater than the threshold value were labelled as NA, after consultation with the ITC developers of the algorithms.

4.4. Source code

4.4.1. General

The original source code developed by ITC was delivered as Matlab code (LUT) and as theoretical pseudo code in the ATBD (SRVI). Both prototypes are written in Python 3. Python is characterized by its simplicity and speed of prototyping but is an order of magnitude slower than, for example, C++. After various tests and considerations of the subsequent target platform in Task 4 at VITO, it was decided to write everything in modern Python, since the program with the appropriate functions is fast enough for the test sites. The program can be called for SRVI and LUT approach separately, but also for both approaches at the same time. This is implemented with a flag when the function is called. The program expects the previously mentioned Sentinel-2 L2A and landcover files for the VRT data cube and an output folder. For the LUT approach, the two pre-calculated INFORM and PROSAIL LUTs are needed. The program dependencies consist in the functionalities or packages GDAL, Rasterio, Numpy, Scipy and system libraries for folder creation and file searching.

After calling the program, the VRT data cube is created first. This internal object is then passed to the selected LUT and/or SRVI functions and the results are written to the specified output folder with the respective extension lut.tif or srvi.tif. The program processes the data cube pixel by pixel and does not use internal subtitling. All temporary data is kept in the main memory (RAM), and the results are calculated directly, which is possible on any modern computer. The final result is available as GeoTiff and can be viewed and analyzed with any GIS program.

4.4.2. SRVI approach

The calculation is done step by step (Figure 8), the logic is conceivable as a binary tree and as specified in Figure 5. The first step is to inspect the Sentinel-2 SCL file and check if the pixel belongs to the vegetation class. If it is not such a class, the result pixel value is written with the NA value. Otherwise, the Global Landcover File is inspected. Here a distinction is made between the two classes, forest and short vegetation. If it is not such a class, the result pixel value is written with the NA value again. Depending on the class, the corresponding algorithm is now applied to the pixel. The program automatically recognizes the correct bands (layers) and calculates the CCC value. All values above 10 are assigned the NA value, as already written in (Singh, 2018).

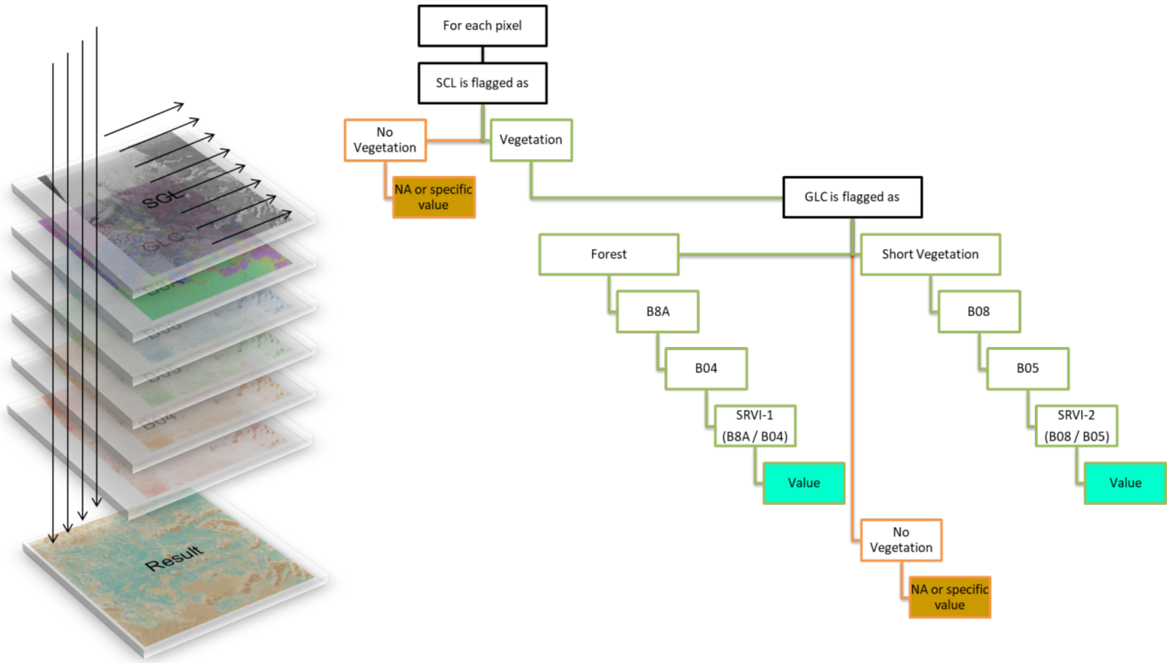

Figure 5: SRVI algorithm implementation Work flow.

4.4.3. LUT approach

The calibration and implementation of the benchmarked algorithm can be performed on different platforms. In this specific case, the predictions of the CCC values using the selected algorithms were carried out originally by using the Matlab R2017b software. Each RTM has sets of source codes for the generation of LUT and inversion. The PROSAIL source code for different platforms such as Matlab, Python and Fortran can be downloaded from http://teledetection.ipgp.jussieu.fr/prosail/. The INFORM code for Matlab platform was obtained by personal communication of the authors. Its GUI version can be found http://ipl.uv.es/artmo/index.php/download.

After image preprocessing (section 4.2), the spectral information of the level 2A product of Sentinel-2 is stored in band-sequential (BSQ) ENVI standard file format. The implementation of the proposed algorithm starts and ends with the reading and writing of ENVI files. The LUT approach takes two LUTs as input: one from INFORM and one from PROSAIL, as described in Section 5.2.1. The LUTs contain 200,000 spectra records with their corresponding model input parameters. Another code file that reads the Sentinel-2 images and searching the best match of each pixel spectral from the LUT using Eq. 1 runs in the backward mode to find the equivalent CCC for each pixel spectrum.

5. Product

5.1. Product description

The output products from any of the selected algorithms contain the pixel value of CCC in g/m^2 retrieved from Sentinel-2 data. There is a significant variation in CCC within and between individual plants of the same species depending on the availability of nutrients and environmental factors (Hamblin et al., 2014). Hence, both high spectral information and high spatial resolution are required for quantification of vegetation biochemical variables. There is a common consensus that higher spatial resolution satellite sensors can resolve finer details of canopy variables while the lower spatial resolution sensors lose, the finer details (Knyazikhin et al., 1999). Thus, the products presented here are obtained by resampling all the Sentinel-2 data into 20m resolution. Figure 6 shows the spatial distribution of CCC predicted by INFORM inversion and the SRVI in the temperate forest ecosystem.

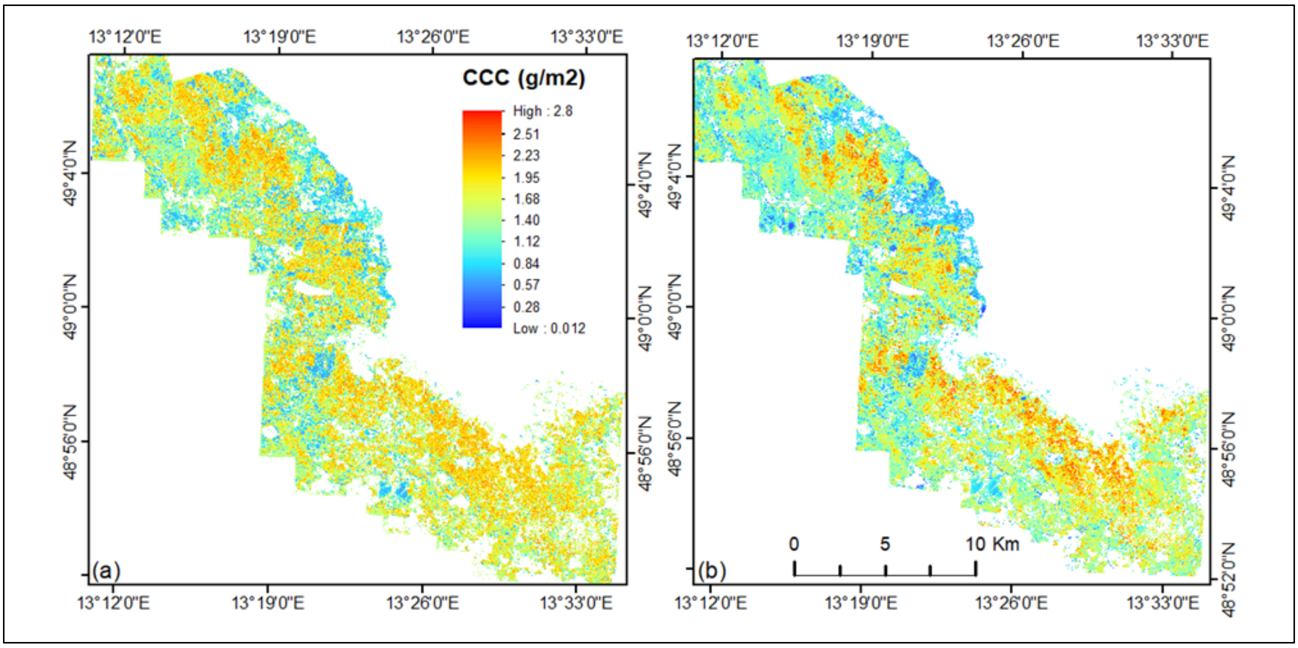

Figure 6: Canopy chlorophyll content (g/m^2) derived from Sentinel-2 data using INFORM inversion (a) and SRVI (b) in Bavarian forest national park, Germany.

5.2. Accuracy analysis (expected accuracy)

CCC is not a directly observable variable in the field. Errors associated with measurement of leaf chlorophyll content and LAI adds up errors to what is assumed to be measured CCC. Therefore, there is no as such "true" value of CCC. On the other hand, the prediction of CCC from remote sensing data has several error sources such as radiometric and atmospheric correction (see section 6.3). In other words, 'true' CCC is difficult to validate unless undertaken with careful field sampling and analysis in a chemistry laboratory (Hamblin et al., 2014). On the other hand, CCC products validation is not a onetime activity, and at least seasonal (4 times per year) in-situ assessments are required (Knyazikhin et al., 1999). Acquiring such dataset from a range of biomes representing a logical subset of the whole terrestrial ecosystems is very expensive and time-consuming. Since there is a lack of field data for most of the pilot sites (which is generally the case for upscaling of chlorophyll and many other RS-EBVs), the following verification strategies are proposed:

1. Geographical consistency: through observing spatial distribution differences between pairs of CCC products generated by the LUT and SRVI approaches across biomes.
2. Compute statistics such as R^2, RMSE, and Bias for pairs of products of LUT and SRVI as a measure of robustness.
3. Temporal consistency analysis: check the consistency of the robustness between pairs of products through time.
4. Comparison of products: compare the products generated by the proposed algorithms with existing small and large scale CCC products in the literature.

R^2, RMSE, and Bias are computed by applying Eq. 6, 7 and 8, and were used to measure the closeness/agreement of CCC retrieved from remote sensing data to field-based measurements, or the robustness of products generated by the two candidate algorithms.

$$R^2 = 1 - \frac{\sum(y_i - y_i')^2}{\sum(y_i - \bar{y_i})^2}$$ *Eq. 6*

$$RMSE\ (\%) = \sqrt{\frac{\sum(y_i - y_i')^2}{n}} / \bar{y_i} * 100$$ *Eq. 7*

$$Bias = \frac{\sum y_i - y_i'}{n}$$ *Eq. 8*

where y_i and y_i' are the actual and predicted values for sample i, and n is the number of samples considered.

5.3. Error budget

Like other variables, the accuracy of retrieving CCC from remote sensing data is affected by a number of error sources emanating from the remote sensing data itself such as satellite data calibration, geo-registration, cloud screening, and atmospheric correction, or uncertainty and discrepancies between measured and simulated spectral datasets due to simplification of RTM input parameters. Depending on the method applied and ancillary data used for correcting those effects, the accuracy of CCC product varies. Soil brightness and sensor view angle also alter reflectance in different bands. Although RTM inversion takes into account the effect of soil background and view angles, care must be taken while applying SRVI for different soil background and view angles. Generally, lower view zenith angles (-30 to + 30 degrees) have minimal effects (Dash and Curran, 2004).

Another source of error is impracticality and expense of collecting field data over a large number of different ground/atmosphere combinations over sufficiently long timescales. A lack of agreement can depend amongst others on uncertainties from field data measurement tools and enumerators, and sample systematic and random errors. It is more challenging to measure chlorophyll content and LAI reliably in the field. There are currently a wide variety of approaches in use for the field measurement of LAI and chlorophyll content. There can be Bias associated with these different measurement schemes. Bias could stem from sampling error due to the number and location of the measurements, and/or the quality of the theory that relates these measurements to the actual canopy variables (Knyazikhin et al., 1999).

5.4. Validation (verification) summary

The prototyped products presented in section 6.1 are validated with CCC collected from the field in BFNP. Validation of the CCC product predicted by the SRVI (865, 665 nm) resulted in R^2 = 0.75 and RMSE = 13.52%. The CCC product obtained by INFORM inversion provided R^2 = 0.66 and RMSE = 21.72%. The SRVI prediction shows the over-estimation of lower CCC values and under-estimation of higher CCC values. Whereas the RTM inversion does not demonstrate such a trend (Figure 7).

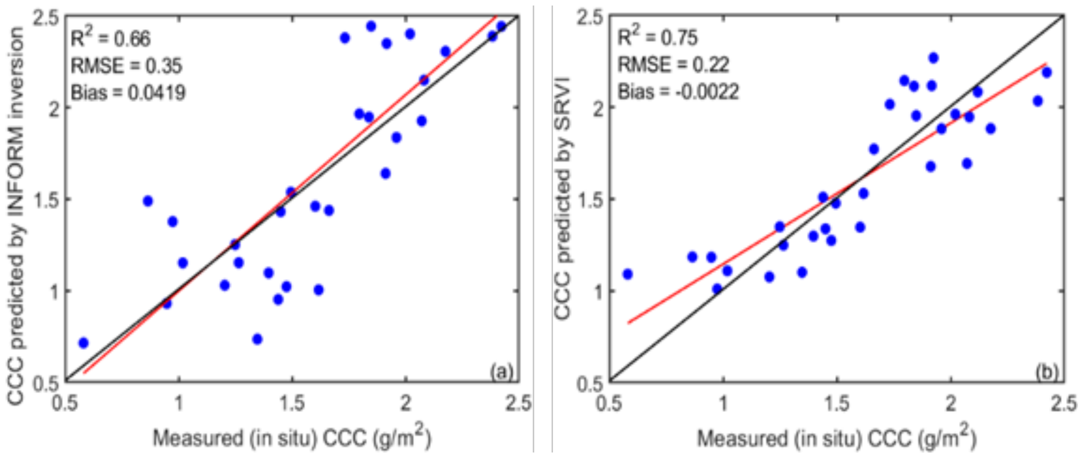

Figure 7: Scatter plot of the measured (in situ) CCC and predictions made by candidate algorithms: validation results of INFORM inversion by using LUT (a) and the Simple Ratio Vegetation Index (SRVI) (b). The line in Black colours show the 1:1 relationship, whilst the line in red indicates the relationship between the fields measured and predicted values of CCC.

The robustness of the proposed algorithms across biomes is verified by computing R^2, RMSE, and Bias between the CCC products predicted by the two algorithms. Higher R^2 with lower RMSE and Bias close to zero combination is considered as an indicator of robustness. The linearity and robustness measures between RTMs based and the SRVI based products in five biomes were reported in the project study report (PSR). The products from the two approaches are in good agreement. More discrepancy observed in the tundra biome. This discrepancy could be partly attributed to the low level of CCC in the Tundra Biome (0 to 0.5 g/m2). The SRVI approach generally has a limitation of overestimating lower CCC values and underestimating higher CCC values. The low level of chlorophyll in this biome (tundra) may exacerbate the weakness of the CCC product derived by SRVI. Validation using in-situ data can help to understand the underlining cause better.

Besides, the temporal consistency of the robustness was investigated to check if the relationship changes through time. For this purpose, the proposed algorithms were applied on cloud-free time-series Sentinel-2 MSI data available for the period June 2017 to September 2018. The relationship between the pairs of methods does not show a significant difference through time (details can be found in the verification section of the PSR document (AD-1)). The time series scatter plots have a similar pattern to the result obtained by applying the methods in single time Sentinel-2 data, which confirms that the proposed algorithms temporally consistent, and can be used to generate CCC products from Sentinel-2 data acquired any time.

Overall, the geographical and temporal consistency, as well as computing measure of robustness between CCC products of the LUT inversion and SRVI in ten representative pilot sites in five biomes of the world, have shown that the two approaches retained spatiotemporal consistency and robustness across biomes. The two approaches were more robust and exhibited lower spatiotemporal discrepancies in wetlands, shrubs, savannah and grasslands than other biomes. This study has found that generally, the SRVI methods overestimate CCC of tundra biomes, which have very low CCC range (< 0.5 g/m2). Whereas in heterogeneous biomes with a wide range of CCC, such as forest ecosystems, systematic over/underestimation was predominant in the relationship of the two approaches. Consequently, the statistical measures such as R^2, RMSE and Bias between CCC products of the two approaches were less vigorous in Tundra and Forest biomes compared to wetlands and grasslands. Therefore, investigation in all pilot sites revealed that the systematic errors stem from the drawbacks of statistical approaches, and reaffirmed the fact that CCC product at a global scale should be based on RTMs inversion using LUT approaches, which account for spatiotemporal variations.

6. Practical considerations for implementation

6.1. Memory requirements

Commonly, RTM inversion requires computing RMSE between the individual pixels spectra and all records of the LUT. It demands a lot of memory and takes a considerably long time. This looping requirement of the merit function makes RTM inversion using LUT computationally the most expensive method. For this reason, we have implemented a different approach using the tree algorithm (see 5.4.3), which has significantly reduced the computation time. The tree algorithm was found to reduce the time for a single pixel calculation by a factor of 10,000. The maximum amount of RAM was 4GB.

The SRVI approach is already very performant due to the simple nature of the calculation and the vectorised nature of matrix calculations in Python. During prototyping, a single Sentinel tile could be processed in less than a minute (often only seconds), using a maximum amount of 4 GB RAM.

The SRVI and LUT approaches can be calculated subsequently in the prototype. The LUT approach takes more time than the SRVI approach to calculate the CCC, so this is the limiting factor. One tile (or granule) was therefore computed within 30 minutes with a maximum amount of 8 GB RAM. For the processor, a normal current processor with about 3 Ghz is sufficient for the calculations. Hyper-threading is not necessary because the application is not optimized for multiprocessing. Of course, this depends on the computing system, but also common notebooks are capable of running this prototype.

6.2. System requirements

Modern Intel or AMD processor with 3 Ghz, 8 GB RAM and enough space for the input and output data

6.3. Error handling

Running the inversion with smaller area size helps reducing memory problems. Therefore, inversion in several blocks of an image is recommended whenever a memory problem occurs. The size of a sentinel granule should not be exceeded and has not been tested. With appropriate memory, you can still improve the speed, but the tiles are already well chosen and it is more advisable to create small sub tiles and calculate them with several processor cores at once.

6.4. External databases

Land cover product to distinguish vegetated and non-vegetated areas and further forest from non-forest vegetated areas. Land cover products of any type, preferably fine resolutions can be used for this purpose.

6.5. Manual interaction

No manual user intervention is necessary during processing.

6.6. Algorithm validation

Comparison of products can be used as algorithms robustness verification. R^2, RMSE and Bias can be computed between the RTM inversion and SRVI CCC sample products to verify the agreement of algorithms in predicting CCC.

7. Upscaling results

7.1. Introduction

Time series of CCC were retrieved for two pre-defined areas of interests, i.e. Finland and Senegal, for the year 2019. Only the SRVI approach was applied due to the computational cost of the RTM inversion model.

For Finland, the 2018 20 m Corine Land Cover product was used to distinguish vegetated and non-vegetated areas and forest vs low vegetation (Figure 8).

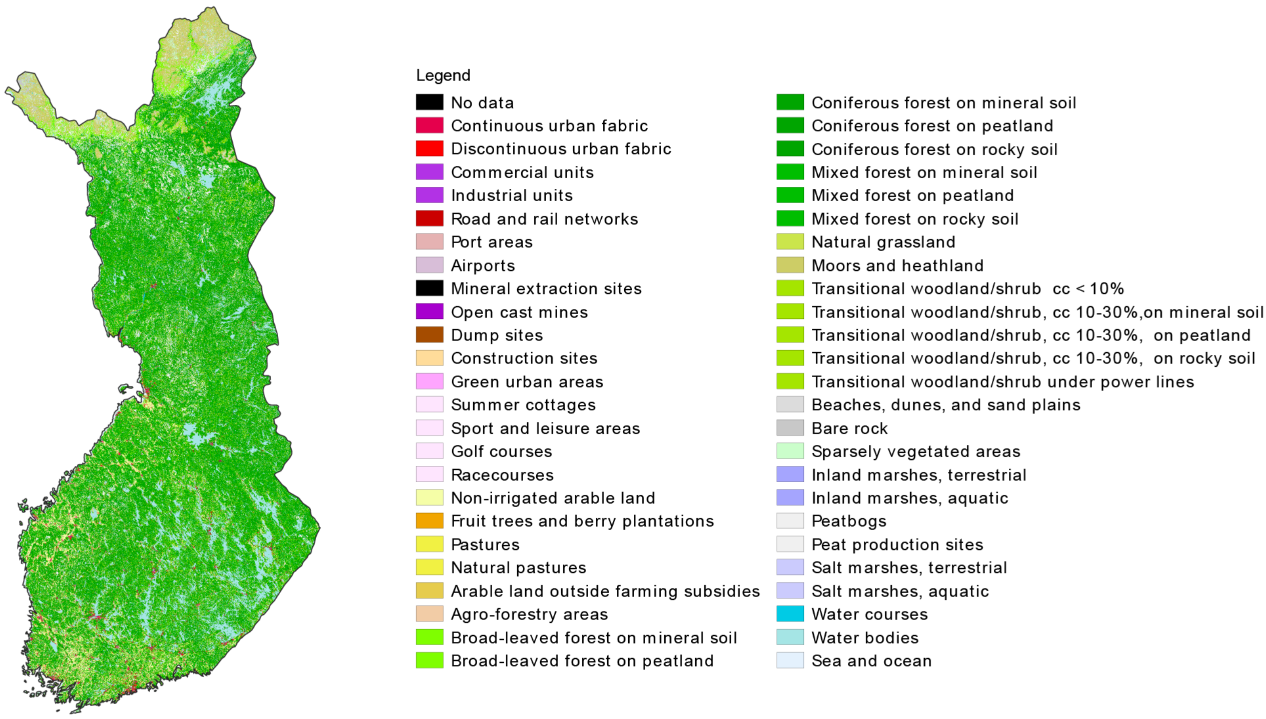

Figure 8: 2018 Corine land cover product for Finland at 20 m resolution available from https://www.syke.fi/en-US/Open_information/Spatial_datasets/Downloadable_spatial_dataset#C

For Senegal, the 2016 20 m ESA CCI land cover product for Africa was used (see Figure 9)

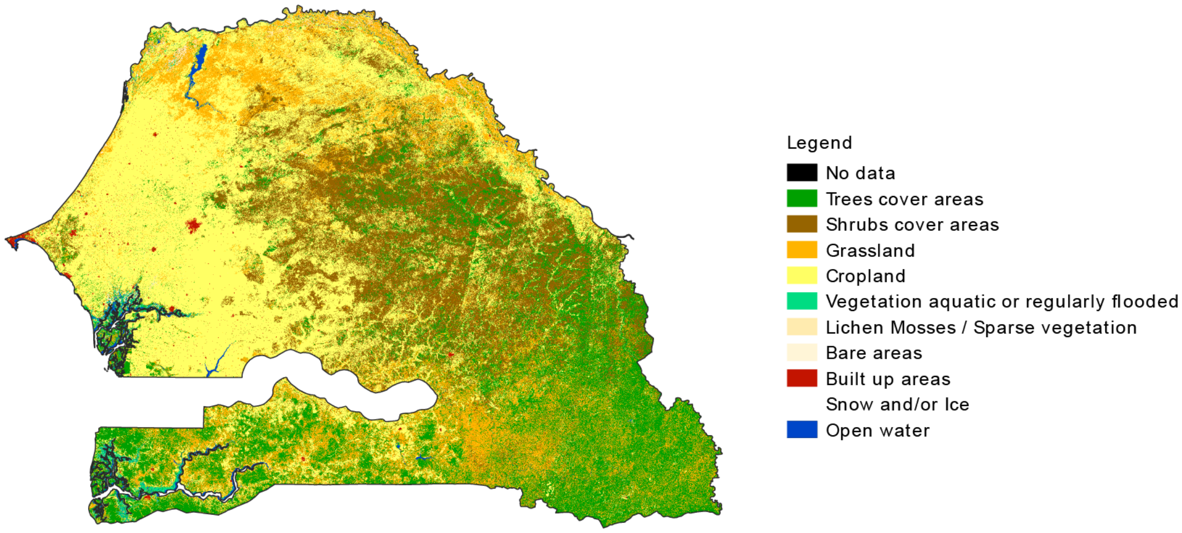

Figure 9: CCI 2016 20 m land cover prototype over Senegal available from http://2016africalandcover20m.esrin.esa.int/

As a preprocessing step, both land cover products were first resampled to the Sentinel-2 UTM tiling grid at 20 m resolution using nearest neighbor resampling and reclassified into vegetated and non-vegetated areas and forest vs low vegetation as required by the SRVI algorithm.

7.2. Results

Table 5, summarizes the total number of CCC products and processing parameters, derived from single Sentinel-2 images for both Senegal and Finland.

Table 5: summary of the CCC upscaling processing results

Parameter	**Senegal**	**Finland**
Number of processed S2 tiles	40	65
Number of processed CCC products	4394	15546

Figure 10 illustrates a mosaic (i.e. covering a Sentinel-2 orbit) of several CCC products, for one specific day for Senegal.

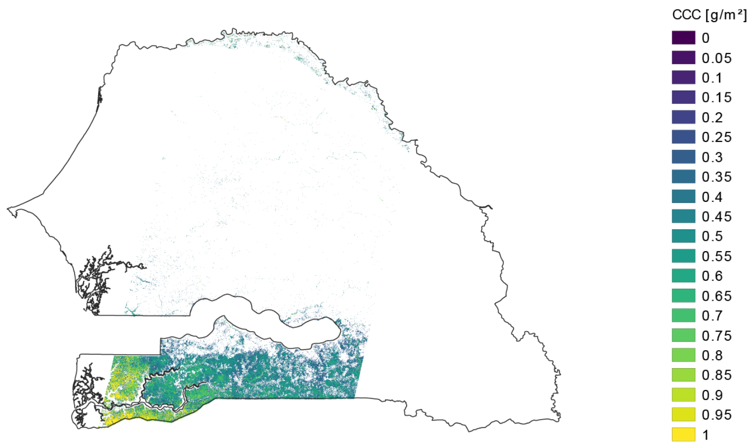

Figure 10: mosaic of several CCC images taken on December 26th 2019 over Senegal. The Sentinel-2 orbit (290 km width) is clearly visible. White areas are either not covered on that day by the S2 orbit or are masked due to either the SCL flags or LC flags.

Based on those time series of CCC, temporal statistics were calculated and visualized showing the spatio-temporal variation in CCC. Figure 11 and Figure 12 demonstrate the 10th, 50th (i.e. median) and 90th percentile for whole Finland and Senegal respectively, as well as an RGB composite of the 3 layers. Figure 13 finally, illustrates these layers zoomed in on a particular region in Finland (village of Karkkila).

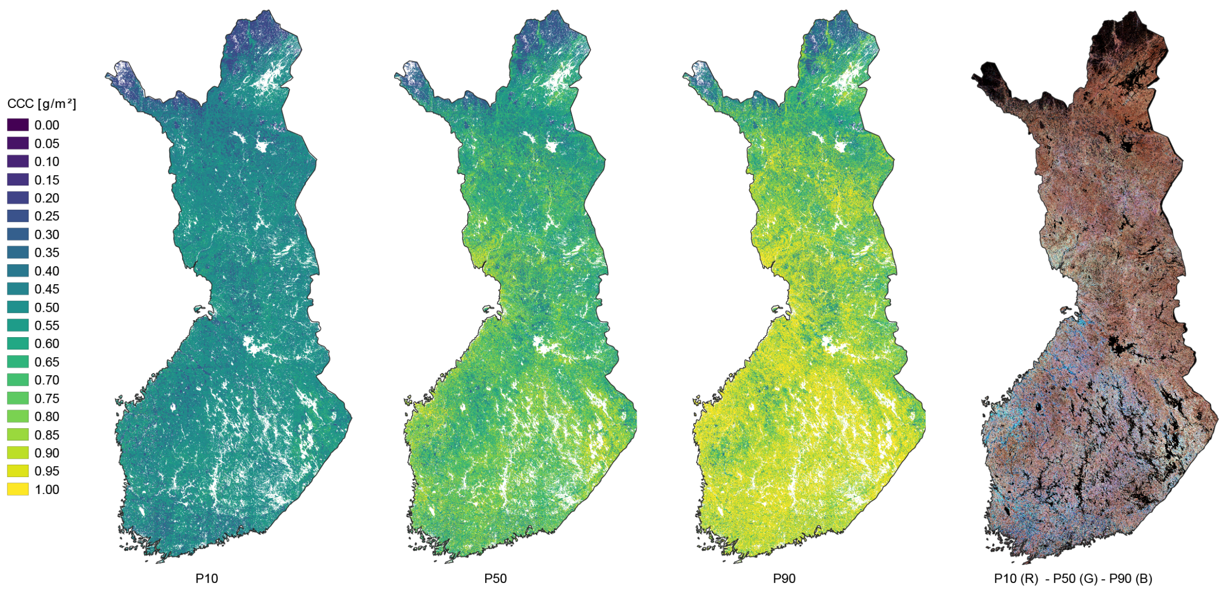

Figure 11: per pixel CCC time series percentiles over Finland: 10th percentile (P10), 50th percentile (P50), 90th percentile (P90) and an RGB composite with R= P10, G= P50 and B=P90)

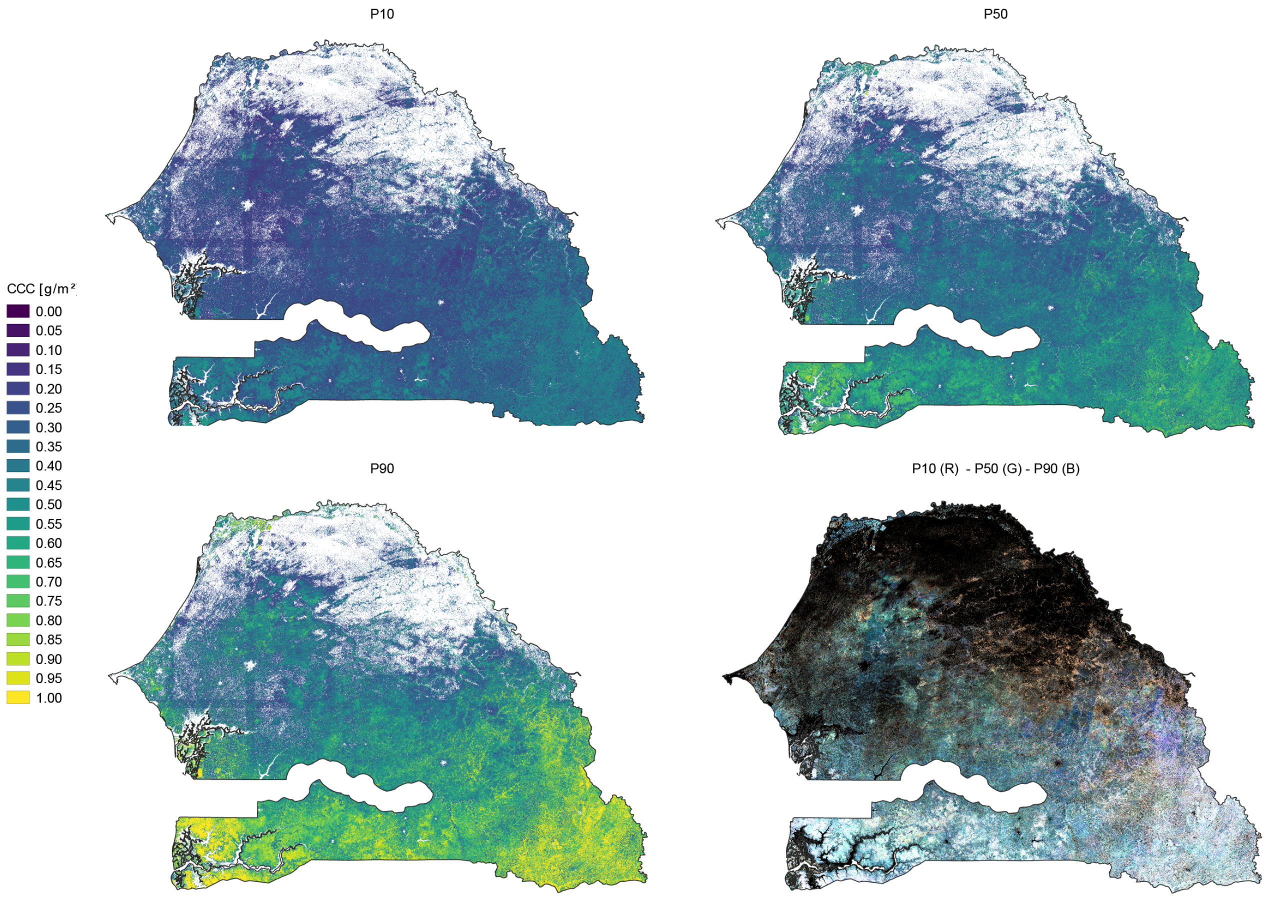

Figure 12: per pixel CCC time series percentiles over Senegal: 10th percentile (P10), 50th percentile (P50), 90th percentile (P90) and an RGB composite with R= P10, G= P50 and B=P90)

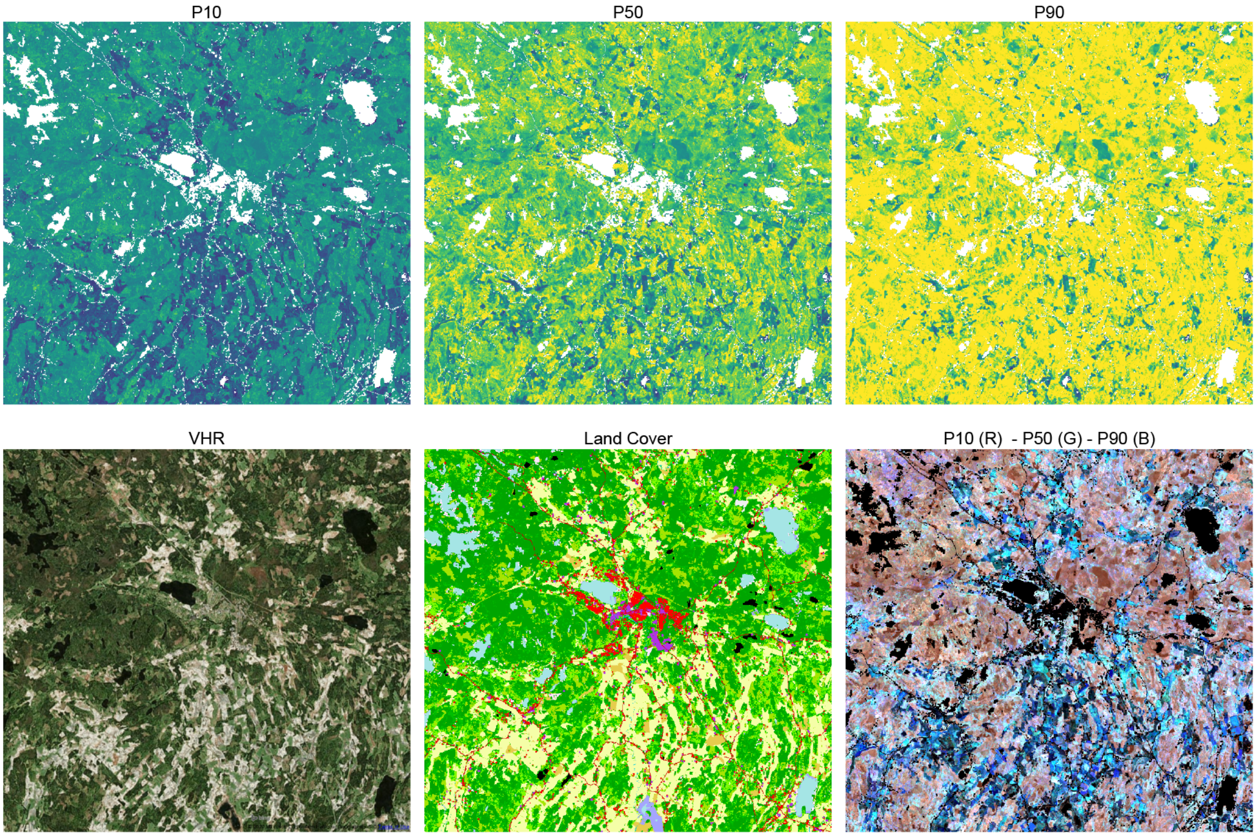

Figure 13: per pixel CCC time series percentiles over a region (Karkkila, 60°31'57.9"N 24°12'42.9"E) in Finland (see Figure 11 for the legend): 10th percentile (P10), 50th percentile (P50), 90th percentile (P90) and an RGB composite with R= P10, G= P50 and B=P90). Additionally, also the Corine Land Cover (see Figure 8 for the legend) and a Bing maps image is given as a reference. For the legend, the reader is referred to Figure 11 and Figure 8.

8. References

Ali, A. M., Darvishzadeh, R., Skidmore, A., Gara, T. W., O'Connor, B., Roeoesli, C., . . . Paganini, M. 2020. Comparing methods for mapping canopy chlorophyll content in a mixed mountain forest using Sentinel-2 data. *International Journal of Applied Earth Observation and Geoinformation,* 87**,** 102037.

Atzberger, C. 2001. Development of an invertible forest reflectance model: The INFOR-model. *Decade of Trans-European Remote Sensing Cooperation***,** 39-44.

Atzberger, C., Darvishzadeh, R., Immitzer, M., Schlerf, M., Skidmore, A. & le Maire, G. 2015. Comparative analysis of different retrieval methods for mapping grassland leaf area index using airborne imaging spectroscopy. *International Journal of Applied Earth Observation and Geoinformation,* 43**,** 19-31.

Blackburn, G. A. 2007. Hyperspectral remote sensing of plant pigments. *Journal of Experimental Botany,* 58**,** 855-867.

Curran, P., Dash, J. & Esl, M. 2018. *ALGORITHM THEORETICAL BASIS DOCUMENT ATBD 2.22 CHLOROPHYLL INDEX Version-2.2*.

Curran, P. J., Dungan, J. L. & Gholz, H. L. 1990. Exploring the relationship between reflectance red edge and chlorophyll content in slash pine. *Tree Physiology,* 7**,** 33-48.

Darvishzadeh, R., Skidmore, A., Atzberger, C. & van Wieren, S. 2008a. Estimation of vegetation LAI from hyperspectral reflectance data: Effects of soil type and plant architecture. *International Journal of Applied Earth Observation and Geoinformation,* 10**,** 358-373.

Darvishzadeh, R., Skidmore, A., Schlerf, M. & Atzberger, C. 2008b. Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. *Remote Sensing of Environment,* 112**,** 2592-2604.

Darvishzadeh, R., Skidmore, A.K., Abdullah, H., Cherenet, E., Ali, A.M., Wang, T., Nieuwenhuis, W., Heurich, M., Vrieling, A., O'Connor, B., Paganini, M., (2019). 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 (JAG), 79, 58-70.

Dash, J. & Curran, P. J. 2004. The MERIS terrestrial chlorophyll index. *International Journal of Remote Sensing,* 25**,** 5403-5413.

Ellsworth, D. S. & Reich, P. B. 1993. Canopy structure and vertical patterns of photosynthesis and related leaf traits in a deciduous forest. *Oecologia,* 96**,** 169-178.

Féret, J. B., Gitelson, A. A., Noble, S. D. & Jacquemoud, S. 2017. PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle. *Remote Sensing of Environment,* 193**,** 204-215.

Gitelson, A. A., Viña, A., Ciganda, V., Rundquist, D. C. & Arkebauer, T. J. 2005. Remote estimation of canopy chlorophyll content in crops. *Geophysical Research Letters,* 32**,** n/a-n/a.

Gitelson, A. A., Peng, Y., Arkebauer, T. J. & Suyker, A. E. 2015. Productivity, absorbed photosynthetically active radiation, and light use efficiency in crops: Implications for remote sensing of crop primary production. *Journal of Plant Physiology,* 177**,** 100-109.

Hamblin, J., Stefanova, K. & Angessa, T. T. 2014. Variation in Chlorophyll Content per Unit Leaf Area in Spring Wheat and Implications for Selection in Segregating Material. *PLOS ONE,* 9**,** e92529.

Houborg, R., Cescatti, A., Migliavacca, M. & Kustas, W. P. 2013. Satellite retrievals of leaf chlorophyll and photosynthetic capacity for improved modeling of GPP. *Agricultural and Forest Meteorology,* 117**,** 10-23.

Hunt, E. R., Doraiswamy, P. C., McMurtrey, J. E., Daughtry, C. S., Perry, E. M. & Akhmedov, B. 2012. A visible band index for remote sensing leaf chlorophyll content at the Canopy scale. *International Journal of Applied Earth Observation and Geoinformation,* 21**,** 103-112.

Inoue, Y., Sakaiya, E., Zhu, Y. & Takahashi, W. 2012. Diagnostic mapping of canopy nitrogen content in rice based on hyperspectral measurements. *Remote Sensing of Environment,* 126**,** 210-221.

Inoue, Y., Guérif, M., Baret, F., Skidmore, A., Gitelson, A., Schlerf, M., . . . Olioso, A. 2016. Simple and robust methods for remote sensing of canopy chlorophyll content: a comparative analysis of hyperspectral data for different types of vegetation. *Plant Cell and Environment,* 39**,** 2609-2623.

Jacquemoud, S. & Baret, F. 1990. Prospect - a Model of Leaf Optical-Properties Spectra. *Remote Sensing of Environment,* 34**,** 75-91.

Knyazikhin, Y., Glassy, J., Privette, J., Tian, Y., Lotsch, A., Zhang, Y., . . . Running, S. 1999. *MODIS Leaf Area Index (LAI) and Fraction of Photosynthetically Active Radiation Absorbed by Vegetation (FPAR) Product (MOD15) Algorithm Theoretical Basis Document*.

Korus, A. 2013. EFFECT OF PRELIMINARY AND TECHNOLOGICAL TREATMENTS ON THE CONTENT OF CHLOROPHYLLS AND CAROTENOIDS IN KALE (BRASSICA OLERACEA L. VAR. ACEPHALA). *Journal of Food Processing and Preservation,* 37**,** 335-344.

Luo, X., Croft, H., Chen, J. M., Bartlett, P., Staebler, R. & Froelich, N. 2018. Incorporating leaf chlorophyll content into a two-leaf terrestrial biosphere model for estimating carbon and water fluxes at a forest site. *Agricultural and Forest Meteorology,* 248**,** 156-168.

Niemann, K. O., Quinn, G., Goodenough, D. G., Visintini, F. & Loos, R. 2012. Addressing the Effects of Canopy Structure on the Remote Sensing of Foliar Chemistry of a 3-Dimensional, Radiometrically Porous Surface. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 5**,** 584-593.

Ollinger, S. V. 2011. Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytologist,* 189**,** 375-394.

Peng, Y. & Gitelson, A. A. 2011. Application of chlorophyll-related vegetation indices for remote estimation of maize productivity. *Agricultural and Forest Meteorology,* 151**,** 1267-1276.

Rosema, A., Verhoef, W., Noorbergen, H. & Borgesius, J. J. 1992. A New Forest Light Interaction-Model in Support of Forest Monitoring. *Remote Sensing of Environment,* 42**,** 23-41.

Schlerf, M. & Atzberger, C. 2006. Inversion of a forest reflectance model to estimate structural canopy variables from hyperspectral remote sensing data. *Remote Sensing of Environment,* 100**,** 281-294.

Singh, D. S., Sarnam 2018. Geospatial Modeling of Canopy Chlorophyll Content Using High Spectral Resolution Satellite Data in Himalayan Forests. *Climate Change and Environmental Sustainability,* 6.

Verhoef, W. 1984. Light-Scattering by Leaf Layers with Application to Canopy Reflectance Modeling - the Sail Model. *Remote Sensing of Environment,* 16**,** 125-141.

Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J. P., Veroustraete, F., Clevers, J. G. P. W. & Moreno, J. 2015. Optical remote sensing and the retrieval of terrestrial vegetation biogeophysical properties – A review. *ISPRS Journal of Photogrammetry and Remote Sensing,* 108**,** 273-290.

Vincini, M., Calegari, F. & Casa, R. 2016. Sensitivity of leaf chlorophyll empirical estimators obtained at Sentinel-2 spectral resolution for different canopy structures. *Precision Agriculture,* 17**,** 313-331.

Wang, Z., Skidmore, A. K., Wang, T., Darvishzadeh, R., Heiden, U., Heurich, M., . . . Hearne, J. 2017. Canopy foliar nitrogen retrieved from airborne hyperspectral imagery by correcting for canopy structure effects. *International Journal of Applied Earth Observation and Geoinformation,* 54**,** 84-94.

Wu, C., Niu, Z. & Gao, S. 2012. The potential of the satellite derived green chlorophyll index for estimating midday light use efficiency in maize, coniferous forest and grassland. *Ecological Indicators,* 14**,** 66-73.

Wu, C. Y., Niu, Z., Tang, Q. & Huang, W. J. 2008. Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agricultural and Forest Meteorology,* 148**,** 1230-1241.

Yin, C., He, B., Quan, X. & Liao, Z. 2016. Chlorophyll content estimation in arid grasslands from Landsat-8 OLI data. *International Journal of Remote Sensing,* 37**,** 615-632.

Zhao, C., Wang, Z., Wang, J., Huang, W. & Guo, T. 2011. Early detection of canopy nitrogen deficiency in winter wheat (Triticum aestivum L.) based on hyperspectral measurement of canopy chlorophyll status. *New Zealand Journal of Crop and Horticultural Science,* 39**,** 251-262.