|  |  |
| --- | --- |
| SL2P- CCRS Simplified Level 2 Prototype Processor - CCRS |  |
|  | Algorithm Theoretical Basis Document |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Name | Company | Date | Signature |
| Prepared by : | Richard Fernandes | CCRS | July 14 2023 |  |
| Checked by : | Gang Hong | CCRS |  |  |
| Approved by : | Darren Janzen | CCRS |  |  |

|  |  |
| --- | --- |
| Document reference : | **Error! Unknown document property name.** |
| Issue.Revision : | 1.0 |
| Date : | July 14, 2023 |
| Client :  EN-MAG-MOD.DT-GEN-034-v1.3 | NRCan |
| Ref., Tender : | Cumulative Effects Phase 2 |

Document Change Record

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Iss. | Rev. | Date | Reason | Comments |
| 1 | 0 | July 14 ,2023 | Creation of the document | First draft |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

# Executive Summary

There is a consensus requirement to globally monitor vegetation canopy biophysical variables at medium resolution (<1ha) and a frequency of <=10days (World Meteorological Organization , 2022a) . Multispectral satellite based imagers designed to satisfy measurement requirements for this task are and will continue to be available (World Meteorological Organization , 2022b**Error! Hyperlink reference not valid.**The Simplified Level 2 Processor – CCRS (SL2P-CCRS) produces estimates of canopy biophysical variables given inputs of either a top-of-atmosphere (TOA) or top-of-canopy (TOC) bi-directional multispectral reflectance together with the illumination, view and relative azimuth angles. Separate non-linear regression models, calibrated for land cover conditions typical of North America, are used to estimate the expected value and the expected root mean square error of each variable. The regression estimators are optimized for multi-spectral (i.e. <10 bands with >10nm bandwidth) reflectance inputs but can be applied to arbitrary spectra as long as a radiative transfer model with sufficient accuracy to simulate such spectra is included in the processor. The parameterization, algorithm and sample results of SL2P-CCRS are presented and compared with its predecessor, SL2P (Weiss and Baret, 2016, WB2016).

Contents

[Simplified Level 2 Prototype Processor - CCRS 1](#_Toc140248568)

[Executive Summary 3](#_Toc140248569)

[1.0 Introduction 15](#_Toc140248570)

[1.1 Background 15](#_Toc140248571)

[1.2 Scope and Objectives 16](#_Toc140248572)

[1.3 Content of the Document 17](#_Toc140248573)

[2.0 Symbols and Acronyms 18](#_Toc140248574)

[AOD550 18](#_Toc140248575)

[AOD 550nm 18](#_Toc140248576)

[DIM 18](#_Toc140248577)

[B 18](#_Toc140248578)

[% 18](#_Toc140248579)

[Cab 18](#_Toc140248580)

[µg.cm-2 18](#_Toc140248581)

[Cbd 18](#_Toc140248582)

[g.cm-2 18](#_Toc140248583)

[CC 18](#_Toc140248584)

[DIM (0-1) 18](#_Toc140248585)

[Ccar 18](#_Toc140248586)

[g.cm-2 18](#_Toc140248587)

[CCC 18](#_Toc140248588)

[Canopy chlorophyll content 18](#_Toc140248589)

[µg.cm-2 18](#_Toc140248590)

[Cdm 18](#_Toc140248591)

[g.cm-2 18](#_Toc140248592)

[CH20 18](#_Toc140248593)

[Water vapour concentration 18](#_Toc140248594)

[g.cm-2 18](#_Toc140248595)

[CO 18](#_Toc140248596)

[mol 18](#_Toc140248597)

[Cw 18](#_Toc140248598)

[µg.cm-2 18](#_Toc140248599)

[CWC 18](#_Toc140248600)

[Campy water content 18](#_Toc140248601)

[g.cm-2 18](#_Toc140248602)

[Cx 18](#_Toc140248603)

[µg.cm-2 18](#_Toc140248604)

[D 18](#_Toc140248605)

[Canopy directional scattering factor 18](#_Toc140248606)

[DIM 18](#_Toc140248607)

[DOY 18](#_Toc140248608)

[Day number 18](#_Toc140248609)

[fAPAR 18](#_Toc140248610)

[Fraction absorbed photosynthetically active radiation 18](#_Toc140248611)

[DIM (0-1) 18](#_Toc140248612)

[fCOVER 18](#_Toc140248613)

[Fraction cover 18](#_Toc140248614)

[DIM (0-1) 18](#_Toc140248615)

[LAD 19](#_Toc140248616)

[Leaf angle distribution 19](#_Toc140248617)

[probability histogram 19](#_Toc140248618)

[LAI 19](#_Toc140248619)

[Leaf area index 19](#_Toc140248620)

[DIM 19](#_Toc140248621)

[Lat 19](#_Toc140248622)

[Latitude 19](#_Toc140248623)

[°N 19](#_Toc140248624)

[Long 19](#_Toc140248625)

[Longitude 19](#_Toc140248626)

[°E 19](#_Toc140248627)

[N 19](#_Toc140248628)

[Leaf layers 19](#_Toc140248629)

[Fractional count 19](#_Toc140248630)

[Patm 19](#_Toc140248631)

[KPa 19](#_Toc140248632)

[19](#_Toc140248633)

[Black sky albedo 19](#_Toc140248634)

[DIM 19](#_Toc140248635)

[19](#_Toc140248636)

[Leaf single scattering albedo 19](#_Toc140248637)

[DIM 19](#_Toc140248638)

[19](#_Toc140248639)

[Shoot single scattering albedo 19](#_Toc140248640)

[DIM 19](#_Toc140248641)

[19](#_Toc140248642)

[Needle-to-shoot area ratio 19](#_Toc140248643)

[DIM 19](#_Toc140248644)

[3.0 Overview of SL2P-CCRS 20](#_Toc140248645)

[4.0 Parameter File 21](#_Toc140248646)

[Start 22](#_Toc140248647)

[Learning Data 22](#_Toc140248648)

[Bandnames 24](#_Toc140248649)

[Sensor 25](#_Toc140248650)

[Configuration 26](#_Toc140248651)

[5.0 Inversion Algorithm Definition 27](#_Toc140248652)

[NNET Inversion Algorithm Definition 27](#_Toc140248653)

[Canopy and Atmosphere Parameters 29](#_Toc140248654)

[6.0 Simulation 34](#_Toc140248655)

[6.1 Radiative Transfer Model 34](#_Toc140248656)

[Shoot Clumping 34](#_Toc140248657)

[Crown Clumping 34](#_Toc140248658)

[6.2 Simulation Laws 35](#_Toc140248659)

[Acquisition Geometry 35](#_Toc140248660)

[Canopy Laws 36](#_Toc140248661)

[Soil Laws 38](#_Toc140248662)

[Simulated Variables 40](#_Toc140248663)

[Performance 42](#_Toc140248664)

[Single Retrieval Cross-validation 43](#_Toc140248665)

[Albedo 43](#_Toc140248666)

[FAPAR 44](#_Toc140248667)

[FCOVER 45](#_Toc140248668)

[LAI 46](#_Toc140248669)

[CWC 47](#_Toc140248670)

[CAB 48](#_Toc140248671)

[Multiple Retrieval Independent Validation 48](#_Toc140248672)

[Quality Indicators 55](#_Toc140248673)

[Name 56](#_Toc140248674)

[Bit 56](#_Toc140248675)

[Value=0 56](#_Toc140248676)

[Value=1 56](#_Toc140248677)

[Terrain 56](#_Toc140248678)

[1 56](#_Toc140248679)

[Simple Terrain 56](#_Toc140248680)

[Complex Terrain 56](#_Toc140248681)

[LandCover 56](#_Toc140248682)

[2 56](#_Toc140248683)

[Vegetated 56](#_Toc140248684)

[Specific 56](#_Toc140248685)

[Input range 56](#_Toc140248686)

[3 56](#_Toc140248687)

[In range 56](#_Toc140248688)

[Out of range 56](#_Toc140248689)

[Output range 56](#_Toc140248690)

[4 56](#_Toc140248691)

[In range 56](#_Toc140248692)

[Out of range 56](#_Toc140248693)

[Unused 56](#_Toc140248694)

[5 56](#_Toc140248695)

[Unused 56](#_Toc140248696)

[Unused 56](#_Toc140248697)

[Unused 56](#_Toc140248698)

[6 56](#_Toc140248699)

[Unused 56](#_Toc140248700)

[Unused 56](#_Toc140248701)

[Unused 56](#_Toc140248702)

[7 56](#_Toc140248703)

[Unused 56](#_Toc140248704)

[Unused 56](#_Toc140248705)

[Unused 56](#_Toc140248706)

[8 56](#_Toc140248707)

[Unused 56](#_Toc140248708)

[Unused 56](#_Toc140248709)

[Algorithm Implementation 56](#_Toc140248710)

[Conclusions 57](#_Toc140248711)

[References 58](#_Toc140248712)

[Appendix I – Physical variables 61](#_Toc140248713)

[Fraction Canopy Cover 61](#_Toc140248714)

[FCOVER 61](#_Toc140248715)

[% 61](#_Toc140248716)

[Fraction of horizontal surface area covered by vegetation 61](#_Toc140248717)

[0-1 61](#_Toc140248718)

[Leaf Area Index 61](#_Toc140248719)

[LAI 61](#_Toc140248720)

[DIM 61](#_Toc140248721)

[Half the total live foliage area per horizontal ground area 61](#_Toc140248722)

[0-15 61](#_Toc140248723)

[Canopy Chlorophyll Content 61](#_Toc140248724)

[CCC 61](#_Toc140248725)

[g.m-2 61](#_Toc140248726)

[Mass of chlorophyll a+b per unit ground area 61](#_Toc140248727)

[0-100 61](#_Toc140248728)

[Canopy Water Content 61](#_Toc140248729)

[CWC 61](#_Toc140248730)

[g.m-2 61](#_Toc140248731)

[Mass of water per unit ground area 61](#_Toc140248732)

[0-100 61](#_Toc140248733)

[Albedo, black sky 61](#_Toc140248734)

[A 61](#_Toc140248735)

[0-1 61](#_Toc140248736)

[Ratio of top of canopy upper hemispherical upwelling radiance to top of canopy incident direct irradiance 61](#_Toc140248737)

[0-1 61](#_Toc140248738)

[D 61](#_Toc140248739)

[D 61](#_Toc140248740)

[0-1 61](#_Toc140248741)

[canopy scattering coefficient for a foliage single scattering albedo of one under direct irradiance 61](#_Toc140248742)

[0-1 61](#_Toc140248743)

[Heterogenity 61](#_Toc140248744)

[None 61](#_Toc140248745)

[Nominal 61](#_Toc140248746)

[Canopy heterogeneity 61](#_Toc140248747)

[Homogenous 61](#_Toc140248748)

[Het: heterogenous 61](#_Toc140248749)

[- 61](#_Toc140248750)

[Fraction of matter that is senescent 61](#_Toc140248751)

[0-1 61](#_Toc140248752)

[- 61](#_Toc140248753)

[Fraction of matter that is woody 61](#_Toc140248754)

[0-1 61](#_Toc140248755)

[- 61](#_Toc140248756)

[Age since first growth 61](#_Toc140248757)

[0-100 61](#_Toc140248758)

[Leaf angle distribution 61](#_Toc140248759)

[LAD 61](#_Toc140248760)

[Discrete probability histogram 61](#_Toc140248761)

[Proportion of leaves for angles: [xx),[xx),[xx),[xx),[xx),[xx] 61](#_Toc140248762)

[0-1; sum 1 61](#_Toc140248763)

[Leaf diameter 61](#_Toc140248764)

[d 61](#_Toc140248765)

[cm 61](#_Toc140248766)

[Diameter of leaf 61](#_Toc140248767)

[0-100 61](#_Toc140248768)

[Leaf layers 61](#_Toc140248769)

[N 61](#_Toc140248770)

[Fractional count 61](#_Toc140248771)

[Number of effective scattering layers in a leaf 61](#_Toc140248772)

[0-4 61](#_Toc140248773)

[Cab 61](#_Toc140248774)

[µg.cm-2 61](#_Toc140248775)

[Foliage chlorophyll a+b content 61](#_Toc140248776)

[0-100 61](#_Toc140248777)

[Cw 61](#_Toc140248778)

[g.cm-2 61](#_Toc140248779)

[0-100 61](#_Toc140248780)

[Cdm 61](#_Toc140248781)

[g.cm-2 61](#_Toc140248782)

[0-100 61](#_Toc140248783)

[Cbd 61](#_Toc140248784)

[g.cm-2 61](#_Toc140248785)

[0-100 61](#_Toc140248786)

[Ccar 61](#_Toc140248787)

[µg.cm-2 61](#_Toc140248788)

[0-100 61](#_Toc140248789)

[Cx 61](#_Toc140248790)

[µg.cm-2 61](#_Toc140248791)

[0-100 61](#_Toc140248792)

[Appendix II SL2P-C Configurations for Canadian Ecozones 62](#_Toc140248793)

[Region 63](#_Toc140248794)

[Area 63](#_Toc140248795)

[Lat 63](#_Toc140248796)

[Min 63](#_Toc140248797)

[Lat 63](#_Toc140248798)

[Max 63](#_Toc140248799)

[1000km](#_Toc140248800)[2](#_Toc140248800) [63](#_Toc140248800)

[°N 63](#_Toc140248801)

[°N 63](#_Toc140248802)

[DOY 63](#_Toc140248803)

[Global 63](#_Toc140248804)

[100000 63](#_Toc140248805)

[41 63](#_Toc140248806)

[83 63](#_Toc140248807)

[365 63](#_Toc140248808)

[Canada 63](#_Toc140248809)

[10000 63](#_Toc140248810)

[41 63](#_Toc140248811)

[83 63](#_Toc140248812)

[365 63](#_Toc140248813)

[Northern Arctic 63](#_Toc140248814)

[2886 63](#_Toc140248815)

[58 63](#_Toc140248816)

[83 63](#_Toc140248817)

[294 63](#_Toc140248818)

[1673 63](#_Toc140248819)

[53 63](#_Toc140248820)

[72 63](#_Toc140248821)

[308 63](#_Toc140248822)

[338 63](#_Toc140248823)

[57 63](#_Toc140248824)

[83 63](#_Toc140248825)

[308 63](#_Toc140248826)

[265 63](#_Toc140248827)

[60 63](#_Toc140248828)

[70 63](#_Toc140248829)

[308 63](#_Toc140248830)

[468 63](#_Toc140248831)

[56 63](#_Toc140248832)

[65 63](#_Toc140248833)

[322 63](#_Toc140248834)

[816 63](#_Toc140248835)

[51 63](#_Toc140248836)

[60 63](#_Toc140248837)

[322 63](#_Toc140248838)

[2072 63](#_Toc140248839)

[44 63](#_Toc140248840)

[60 63](#_Toc140248841)

[350 63](#_Toc140248842)

[654 63](#_Toc140248843)

[57 63](#_Toc140248844)

[69 63](#_Toc140248845)

[322 63](#_Toc140248846)

[447 63](#_Toc140248847)

[50 63](#_Toc140248848)

[59 63](#_Toc140248849)

[336 63](#_Toc140248850)

[169 63](#_Toc140248851)

[41 63](#_Toc140248852)

[48 63](#_Toc140248853)

[365 63](#_Toc140248854)

[738 63](#_Toc140248855)

[49 63](#_Toc140248856)

[62 63](#_Toc140248857)

[336 63](#_Toc140248858)

[281 63](#_Toc140248859)

[43 63](#_Toc140248860)

[50 63](#_Toc140248861)

[365 63](#_Toc140248862)

[466 63](#_Toc140248863)

[49 63](#_Toc140248864)

[55 63](#_Toc140248865)

[365 63](#_Toc140248866)

[448 63](#_Toc140248867)

[49 63](#_Toc140248868)

[58 63](#_Toc140248869)

[365 63](#_Toc140248870)

[245 63](#_Toc140248871)

[48 63](#_Toc140248872)

[60 63](#_Toc140248873)

[365 63](#_Toc140248874)

[Appendix III SL2P-C Canopy biophysical variable parameterizations 64](#_Toc140248875)

[LAI 64](#_Toc140248876)

[Leaf Angle Distribution 64](#_Toc140248877)

[Crown Cover 65](#_Toc140248878)

[Hot Spot Dimension 65](#_Toc140248879)

[Figure 1.](C:\\Users\\rfernand\\Downloads\\LEAF Service System Design.docx" \l "_Toc140248880) *[SL2P-CCRS architecture. Solid lines indicate data flow during calibration; dashed lines indicate data flow during application.](C:\\Users\\rfernand\\Downloads\\LEAF Service System Design.docx" \l "_Toc140248880)* [21](C:\\Users\\rfernand\\Downloads\\LEAF Service System Design.docx" \l "_Toc140248880)

[Figure 2. Comparison of sampling designs for the nominal SL2P canopy architecture distribution using only 1296 samples (in contrast to the nominal 43782 samples) for demonstration purposes. 32](#_Toc140248881)

[Figure 3. Density and scatter plots of cosine of indicated geometric quantity for 40000 samples drawn from Geometry Laws for all SL2P-CCRS databases for S2A. Contours correspond to 0.2 probability intervals from 0 to 1. 35](#_Toc140248882)

[Figure 4. Fitted minimum and maximum conditional range of canopy cover given LAI for DBF (left) and ENF (right). 37](#_Toc140248883)

[Figure 5. Kernel density fits to 40000 samples from each canopy parameter Law as a function of Database Class. Gamma is constant at 1 for SL2P and DBF. 38](#_Toc140248884)

[Figure 6. Soil spectra shapes for SL2P and DBF based on WB2016. 39](#_Toc140248885)

[Figure 8. Soil spectral shapes for ENF based on both WB2016 and Miler et al., 1997. 39](#_Toc140248886)

[Figure 8. Comparison of histogram bar heights (blue circles) with fitted lognormal pdf (red) and WB2016 normal pdf prior to truncation. 40](#_Toc140248887)

[Figure 9. Kernel density plots of univariate (diagonal) and bivariate (lower) distributions of simulated response variables for each Class. Contours correspond to 0.2 probability intervals from 0 to 1. 41](#_Toc140248888)

[Figure 10. Kernel density plots of univariate (diagonal) and bivariate (lower) distributions of simulated reflectances for each Class. Contours correspond to 0.05,0.2,0.4,0.6,0.8 and 1.0 probability levels. 42](#_Toc140248889)

[Figure 12. Kernel density plots of SL2P cross-validation residuals for single sample versus smoothed retrievals based on 10 (blue) or 100 (orange) nearest neighbours in input space. 50](#_Toc140248890)

[Figure 13. Kernel density plots of DBF cross-validation residuals for single sample versus smoothed retrievals based on 10 (blue) or 100 (orange) nearest neighbours in input space. 51](#_Toc140248891)

[Figure 14. Calibrated distributions (red) and sampled histograms (black) for two forest classes. 64](#_Toc140248892)

# 1.0 Introduction

## 1.1 Background

The Simplified Level 2 Processor – CCRS (SL2P-CCRS) produces estimates of canopy biophysical variables (outputs, Table **1**) given inputs of either a top-of-atmosphere (TOA) or top-of-canopy (TOC) bi-directional reflectance spectrum together with the illumination, view and relative azimuth angles. Separate non-linear regression models are used to estimate the expected value and the expected root mean square error (RMSE) of each output. The regression estimators are optimized for multi-spectral (i.e. <10 bands with >10nm bandwidth) reflectance inputs but can be applied to arbitrary spectra as long as a radiative transfer model with sufficient accuracy to simulate such spectra is included in the processor.

Table 1. SL2P-D output variables. Uncertainty requirements provided by GCOS (WMO, 2022a) or the Copernicus Global Land Service (Lang and Tychon, 2015).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Acronym | Units | Description | Requirement | Range |
| Fraction Canopy Cover | fCOVER | fraction | Fraction of horizontal surface area covered by vegetation | Max(0.5,10%) | 0-1 |
| Fraction Absorbed PAR | fAPAR | fraction | Fraction of absorbed PAR by vegetation | Max(0.5,10%) | 0-1 |
| Leaf Area Index | LAI | m2.m-2 | Half the total live foliage area per horizontal ground area | Max(0.5,15%) | 0-20 |
| Canopy Chlorophyll Content | CCC | ug.m-2 | Mass of chlorophyll A+B per horizontal ground area | Max(20,20%) | 0-1000 |
| Canopy Water Content | CWC | g.m-2 | Mass of H20 per horizontal ground area | Max(20,20%) | 0-1000 |
| Albedo, black sky | α | fraction | Ratio of top of canopy upper hemisphere upwelling radiance to top of canopy incident direct irradiance | Max(0.025,5%) | 0-1 |
| Directional area scattering factor, black sky | D | fraction | Scattering coefficient for foliage single scattering albedo of one, zero boundary reflectance and direct irradiance | None | 0-1 |

SL2P-CCRS is an implementation of a subset of the SL2P-D processor (Fernandes and Djamai, <https://github.com/rfernand387/SL2PD/blob/master/Reports/sl2p-d.docx>; see also Appendix A of Brown et al., 2021) that in turn is a modification of the SL2P processor (Weiss and Baret, 2016). Differences between SL2P-CCRS and these two processors are summarized in Table **2**. SL2P-CCRS is an extension of SL2P to account for canopy clumping for forests and shrubs and reverts to SL2P, with a slightly more robust regression strategy, otherwise. SL2P-CCRS has three phases: calibration of the regression models, validation of the regression models with simulations and application of the regression models to measurements. Users are encouraged to include a prior phase corresponding to benchmarking of the simulation model and quantification of input measurement error and a posterior phase corresponding to validation of estimates using internationally accepted protocols (Widlowski et al. 2007, Fernandes et al., 2014).

2

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

## 1.2 Scope and Objectives

The objective of this document is to provide a theoretical basis for the implementation of SL2P-CCRS available at (https://github.com/rfernand387/SL2P-CCRS).**Error! Hyperlink reference not valid.**To satisfy this objective, the inputs to the SL2P-CCRS are defined and the algorithm is discussed on a functional level. At each stage a theoretical justification is provided. In many instances, the justification simply follows from SL2P and is noted as such. Although, especially with respect to additional features such as sampling , new RTMs , and new regression approaches, a detailed theoretical basis is provided.

After reading this document one should be able to produce a new calibration of SL2P-CCRS, develop regression estimators for outputs and their RMSE, and assess the performance by cross-validation. However, this document is not an engineering manual for the SL2P-CCRS code or for modification of the code. Moreover, this release of SL2P-CCRS is not intended for high-performance application to arbitrary input images; in part because of the complexity of image formats and in part because of varying user production requirements. The regression algorithms produced by SL2P-CCRS have been implemented in Google Earth Engine (https://github.com/rfernand387/LEAF-Toolbox) for this purpose. Readers interested in modifying the code are encouraged to do so and to contact either the current authors or the SL2P authors for advice but should be aware they are entering terra incognito and walking into the realm of science (as Fred and Marie would say ‘bonne chance!’).

## 1.3 Content of the Document

The document begins with a list of symbols and acronyms. An overview of SL2P-CCRS is provided and each component is then described in detail in separate sections.

# 2.0 Symbols and Acronyms

|  |  |  |
| --- | --- | --- |
| Symbol/Acronym | Name | Units |
| AOD550 | AOD 550nm | DIM |
| B | Soil brightness | % |
| Cab | Leaf chlorophyll content | µg.cm-2 |
| Cbd | Leaf brown pigment content | g.cm-2 |
| CC | Canopy cover | DIM (0-1) |
| Ccar | Leaf carotenid content | g.cm-2 |
| CCC | Canopy chlorophyll content | µg.cm-2 |
| Cdm | Leaf dry matter content | g.cm-2 |
| CH20 | Water vapour concentration | g.cm-2 |
| CO | ozone concentration | mol |
| Crop | Cropland | Nominal |
| Cw | Leaf water content | µg.cm-2 |
| CWC | Campy water content | g.cm-2 |
| Cx | Leaf Xanthophyll content | µg.cm-2 |
| D | Canopy directional scattering factor | DIM |
| DBF | Deciduous Broadleaf Closed Forest | Nominal |
| DNF | Deciduous Needleleaf forest | Nominal |
| DOY | Day of year | Day number |
| EBF | Evergreen Broadleaf Forest | Nominal |
| ENF | Evergreen Needleleaf forest | Nominal |
| fAPAR | Fraction absorbed photosynthetically active radiation | DIM (0-1) |
| fCOVER | Fraction cover | DIM (0-1) |
| Gr | Grassland or pasture | Nominal |
| GrPolar | Polar grassland | Nominal |
| LAD | Leaf angle distribution | probability histogram |
| LAI | Leaf area index | DIM |
| Lat | Latitude | °N |
| Long | Longitude | °E |
| MF | Mixed forest | Nominal |
| MSI | Multispectral Imager | Nominal |
| N | Leaf layers | Fractional count |
| Patm | Atmospheric pressure | KPa |
| R | Bi-directional reflectance | 0-1 |
| RTM | Radiative transfer model | Nominal |
| Sh | Shrubland | Nominal |
| ShPolar | Polar shrubland | Nominal |
| Toa | Top of atmosphere | Nominal |
| Toc | Top of canopy | Nominal |
| Veg | Vegetated | Nominal |
| Wet | Herbaceous wetland | Nominal |
|  | Black sky albedo | DIM |
|  | Leaf single scattering albedo | DIM |
|  | Shoot single scattering albedo | DIM |
|  | Needle-to-shoot area ratio | DIM |

# 3.0 Overview of SL2P-CCRS

This section provides a description of the process flow of SL2P-CCRS to produce regression algorithms that can be used to predict vegetation parameters from optical satellite imagery.

SL2P-CCRS consist of Databases of Classes, each of which correspond to a separate surface land cover class. Each Class corresponds to a different set of regression algorithms for predicting desired canopy variables (Table **1**) together with their associated calibration and validation information. SL2P-CCRS differs from SL2P in that it contains multiple Classes, one of which is identical to SL2P. Each Class is housed in separate Databases (Table **3**) for maximum modularity.

Table 3. SL2P-CCRS Databases.

|  |  |
| --- | --- |
| Name | Description |
| SL2P | Original SL2P database as defined in WB2016.  Applies to all non-woody land cover. |
| ENF | Evergreen needleleaf forest database for North America.  Also applied to deciduous needleaf forests. |
| DBF | Deciduous broadleaf forests database for North America.  Also applies to evergreen broadleaf forests and all shrubs. |

Each SL2P-CCRS Database uses an input parameter file to perform three sub-processes: simulation, calibration and cross-validation, with corresponding outputs (Figure 1). The regression predictors are also applied to independent simulation databases to perform independent validation. Application of these predictors to map satellite imagery is beyond the scope of this document.

Figure 1. *SL2P-CCRS architecture. Solid lines indicate data flow during calibration; dashed lines indicate data flow during application.*

Regression Calibration

Simulation

Cross-Validation

Parameter File

Simulation

Database

Performance Visualization

Regression Predictors

# 4.0 Parameter File

There is a separate parameter file for each SL2P-CCRS database . Each parameter file is a Microsoft Excel .xls that defines algorithm calibration, including specification of the input measurements and outputs, the regression model for estimating outputs, and the nominal validation database. The parameter file has a number of worksheets that are described below. Worksheet formatting must be preserved during calibration. In most instances the MATLAB implementation will indicate an error if formatting or errors in data entry result in inconsistent parameter values or naming of resources (database or classes) that do not exists.

## Start

This sheet contains the title of the algorithm, a database, the name of the validation database and a flag (CopyFlag) indicating if simulations can be reused between land cover class (**Table 4**).

Table 4. Description of Start worksheet entries.

|  |  |
| --- | --- |
| Database directory name | Name of the directory that will contain the report |
| Comments | Comments |
| Inversion Algorithm Name | Inversion Algorithm Name |
| Name of Validation Database Directory | Name of Validation Database Directory |
| CopyFlag | CopyFlag |

By convention, uses should tile databases as:

{Sensor input}\_{geographiczone}\_{soildatabase}\_{terraincomplexity}\_{size}\_{sampling scheme}.xls

Phrases in brace brackets are selected from valid values from tables described below. The title is assigned to the database andused to determine if this database can be used to validate regression models from another database.

## Learning Data

The Learning Data worksheet contains global parameters for the algorithm as defined in Table **5**.

Table 5. Learning Data worksheet description.

|  |  |  |
| --- | --- | --- |
| Name | Description | Range |
| Toc\_Toa | Switch between TOA and TOC reflectance | ‘Toc’ or ‘Toa’ |
| Terrain | Terrain complexity for input noise | ‘Simple’ or ‘Complex’ |
| Classification | Land cover classification for documentation | ‘CCRS’ , ‘SL2P’ |
| Nb\_Classes | Number of classes used; each requires a canopy/atmosphere description worksheet | <= #classes in Classification |
| FAPAR\_Time | Time used for fAPAR computation | Local Time hh:mm |
| RTM | Radiative transfer model used for all simulations | ‘sail3’; ‘4SAIL2’; ‘Flight’ |
| Max\_Sims | Maximum # simulations per class; used to simulate geometry. | >=Nb\_Sims of all of the classes |

‘Toc\_Toa’ specifies the input reflectance measurements as either top of canopy or top of atmosphere. If top of atmosphere is specified the SMAC radiative transfer model is used with a sampled distribution of atmosphere parameters specified under the Canopy\_Atmosphere\_Class1 worksheet to convert produce a calibration database of top of atmosphere reflectance; else the output of the selected canopy RTM is used.

‘Terrain’ specifies the terrain complexity in the sense defined by the SEN2COR atmospheric correction algorithm. Complex terrain implies the contribution from surrounding land surfaces and the local slope cannot be neglected when determining TOA reflectance. Since only 1D RTMs are used to model reflectance this contribution is considered noise and specified as such in the ‘Sensor’ worksheet. Currently the results of Djamai and Fernandes (2018) are used to quantify the noise levels.

‘Classification’ specifies the land cover classification associated with each Class. This is only for documentation purposes since the parameters for each Class are specified in individual worksheets.

‘Nb\_Classes’ specifies the number of classes in this database. Each class must be specified in a separate Canopy\_atmosphere\_Class# worksheet.

‘FAPAR\_Time’ species the local time at with black sky fAPAR is computed in the simulation database.

‘RTM’ specifies the canopy radiative transfer model used for all simulations. The same model is used for all classes to allow for copying of simulations between classes. SL2P-D uses 4SAIL2.

‘Max\_Sims’ specifes the maximum allowable simulations for a class. It is used to produce a database of randomly sampled geometric acquisition variables based on the spatiotemporal extents specified in the ‘Configuration’ worksheet. The database is randomized before use.

‘Validation\_Name’ specifies the name of the database containing validation data. The database may be i. the same as the current database in which case all calibration data is used for validation ii. A different complete database or iii. A partially specified database containing only input and output values. In all cases, the database must include a definition (Def\_Base) specifying the number of classes and corresponding input and output data for each class.

‘Copy\_Flag’ is set to ‘yes’ if simulated input-output pairs for a class should be first copied from existing simulated classes in the same database. This avoids having to redo similar simulations but slightly modifies the initial sampling scheme for the class since a desired simulation law may not have an exact match in any other class. The maximum tolerated mismatch is determined by the most constraining Nb\_levels parameter specified in the Canopy\_Atmosphere\_Class# worksheet for the class being populated.

## Bandnames

The ‘Bandnames’ worksheet specifies the allowable band names that can be selected in the ‘Sensors’ worksheet as inputs to the calibration database. Each new sensor (or newly calibrated sensor) is specified in a new column in this worksheet. The name of the sensor at the top of the column and the band names in lower rows must match the name of a structure and associated structure members in the ‘Orbito\_Sensor.m’ and ‘Filters\_Smac.mat’ file. Following Djamai et al. (2018) the following bands used for Landsat OLI and Sentinel 2 MSI are indicated in Table **6** and Table **7**.

Table 6. Landsat OLI bands used within SL2P-CCRS.

|  |  |  |  |
| --- | --- | --- | --- |
| OLI Band | Central wavlength | Nominal Width | Spatial resolution |
| name | nm | nm | m |
| B2 | 482.5 | 66 | 30 |
| B3 | 562.5 | 76 | 30 |
| B4 | 655 | 51 | 30 |
| B5 | 865 | 41 | 30 |
| B6 | 1610 | 101 | 30 |
| B7 | 2200 | 201 | 30 |

Table 7. Sentinel 2 MSI bands used within SL2P-D.

|  |  |  |  |
| --- | --- | --- | --- |
| MSI Band | Central wavlength | Nominal Width | Spatial resolution |
| name | nm | nm | m |
| B3 | 560 | 35 | 10 |
| B4 | 665 | 30 | 10 |
| B5 | 705 | 15 | 20 |
| B6 | 740 | 15 | 20 |
| B7 | 783 | 20 | 20 |
| B8a | 865 | 20 | 20 |
| B11 | 1610 | 90 | 20 |
| B12 | 2190 | 180 | 20 |

## Sensor

The ‘Sensor’ worksheet specifies the Sensor, input bands and level of additive and multiplicative noise for each band and acquisition geometry parameter under simple and complex terrain conditions. SL2P-CCRS currently assumes simple terrain. The Sensor must be selected from a pre-defined list indicated in the worksheet for which orbit and SMAC parameters are pre-tabulated in files ’Orbito\_Sensor.m’ and ‘Flitres\_Smac.mat’ respectively. The user should update these files and the ‘Bandnames’ worksheet with new sensors or if sensor specifications change; ideally using a calibration update date for the latter as a suffix for the sensor name. Input noise specifications for Landsat 8/9 and Sentinel 2A/2B L2Aproducts are provided in Table **8** and Table **9** respectively according to Djamai and Fernandes (2018).

Table 8. Input noise specification for simple and complex terrain for Landsat 8 and 9 OLI.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Input | Simple Terrain | | Complex Terrain | |
|  | Additive Noise | Multiplicative Noise | Additive Noise | Multiplicative Noise |
| %refl or ° | % | %refl or ° | % |
| B2 | 2 | 2 | 3 | 2 |
| B3 | 2 | 2 | 3 | 2 |
| B4 | 2 | 2 | 3 | 2 |
| B5 | 2 | 2 | 4 | 4 |
| B6 | 2 | 2 | 4 | 4 |
| B7 | 2 | 2 | 4 | 4 |
| Solar Zenith Angle | 0 | 0 | 10 | 0 |
| View Zenith Angle | 0 | 0 | 10 | 0 |
| Relative Azimuth Angle | 0 | 0 | 0 | 0 |

Table 9. Input noise specification for simple and complex terrain for Sentinel 2A and 2B

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Input | Simple Terrain | | Complex Terrain | |
|  | Additive Noise | Multiplicative Noise | Additive Noise | Multiplicative Noise |
| %refl or ° | % | %refl or ° | % |
| B03 | 2 | 2 | 3 | 2 |
| B04 | 2 | 2 | 3 | 2 |
| B05 | 2 | 2 | 3 | 2 |
| B06 | 2 | 2 | 4 | 4 |
| B07 | 2 | 2 | 4 | 4 |
| B8A | 2 | 2 | 4 | 4 |
| B11 | 2 | 2 | 4 | 4 |
| B12 | 2 | 2 | 4 | 4 |
| Solar Zenith Angle | 0 | 0 | 10 | 0 |
| View Zenith Angle | 0 | 0 | 10 | 0 |
| Relative Azimuth Angle | 0 | 0 | 0 | 0 |

## Configuration

The ‘Configuration’ worksheet specifies the geometric configuration, spatial and temporal extents for sampling the acquisition geometry for all simulations in the database. One of a number of possible extent specifications can be specified (Table **10**). Sensitivity analysis indicates that global coverage of vegetated land surfaces is sufficient (Table **11**).

Table 10. Cases considered for sensor configuration, spatial and temporal extents.

|  |  |  |
| --- | --- | --- |
| Case | Definition | Example |
| Case Single Date, Location and Configuration | The sun position can be specified either or from time and location. | One experiment at one date, nadir viewing. |
| Case Single Date, Location and Multiple Configuration | The view directions are specified in a file. The sun position can be specified either directly or from time and location. | Acquisition at a given date and location, but over a large extent (a SPOT or TM image) or with several view directions (CHRIS/POLDER) |
| Case Multiple Dates, Locations and Single Configuration | The locations can be specified by a range of lat/lon of by a list of coordinates. | Processing of global data from directionally normalized data (CYCLOPES) |
| Case Multiple Dates, Locations and Configurations | The locations can be specified by a range of lat/lon of by a list of coordinates. | Processing of global data from instantaneous observations (MERIS, Landsat OLI, S2MSI) |
| Case Multiple Dates, Single Location and Single Configuration | The locations can be specified by a range of lat/lon of by a list of coordinates. | Processing of images acquired with the same view angle, over the same target, at multiple dates (formosat) |

Table 11. Template datasheet for specification of a selected configuration, spatial and temporal extents. Allowable entries, identified in yellow, vary with configuration indicated in Table 2.

|  |  |  |  |
| --- | --- | --- | --- |
| Selected | **Min** | **Max** | **File** |
| **Sun zenith (°)** | 999 | 999 |  |
| **Sun Azimuth (°)** |  |  |  |
| **View zenith (°)** |  |  |  |
| **View Azimuth (°)** |  |  |  |
| **Year** |  |  |  |
| **Day of year** | 1 | 365 |  |
| **Hour (UT)** |  |  |  |
| **Minute** |  |  |  |
| **Lat (°)** | -56 | 83 |  |
| **Lon (°)** | 0 | 360 |  |

## 5.0 Inversion Algorithm Definition

The inversion algorithm corresponds to a list of outputs and a specification for an estimation algorithm. Multiple inversion algorithms can be applied to a database by repeatedly executing the calibration process as long as the list of desired outputs does not change. This restriction is applied to minimize the size and data processing demands on producing a calibration database since otherwise all possible outputs would always have to be included in the calibration database. Table **12** lists the supported inversion algorithms. Each inversion algorithm requires parameters specified in sub-sections below. Currently the NNET algorithm, corresponding to a separate single neural network regression per output variable, is used for SL2P.

Table 12. Supported inversion algorithms.

|  |  |  |
| --- | --- | --- |
| Name | Description | Implementation |
| NNET | backpropagation network | MATLAB net |
| NNETP | partitioned backpropagation network | MATLAB net applied first to estimate one output variable followed by multiple MATLAB net networks each for small range of this variable |
| LUT | Look up table | Not implemented |
| GPR | Gaussian process regression | Not implemented |
| SVR | Support vector regression | Not implemented |

### NNET Inversion Algorithm Definition

Following the theoretical basis provided in SL2P, a single hidden layer multilayer perceptron neural network regression (implemented in MATLAB using the ‘train’ function) is defined to estimate a specified output given input reflectance measurements, cosine of view zenith angle, cosine of solar zenith angle and cosine of relative azimuth angle using fixed and used defined parameters provided in Table **13**.

For SL2P-CCRS, the calibration database is partitioned into clusters of approximately 100 samples using the k-means algorithm (<https://www.mathworks.com/help/stats/kmeans.html>) applied to input distributions for ALA, Cab, N, Cdm, Cw\_rel and Bs with uniform initialization. Canopy structure variables were not included since they cannot be assumed to be locally correlated. Ideally the network would be trained using a batch of different clusters held out for validation at each epoch. As this feature is not implemented in MATLAB, replicate (user specified as Num\_networks) network is trained using a batch corresponding to a random sample of 1/3 of the clusters and initial weights from the previous batch (or randomly initialized for the first batch). The performance of each network is quantified using a regularized metric corresponding to the sum of the mean square prediction error over the cross-validation samples and a faction of the mean square weights.

Table 13. Regression network parameters.

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter | Description | Nominal Value | Range |
| Number of networks | Number of replicate networks | 5 | >=1 |
| Number of hidden layers | Number of hidden layers per network | 1 | 1,2 |
| Transfer function L1 | Node transfer function for layer 1 | tansig | ‘poslin’ for positive linear; ‘tansig’ for hyperbolic tangent sigmoid |
| Number of neurones L1 | Number of neurons for layer 1 | 10 | >=1 |
| Transfer function L2 | Node transfer function for layer 2 | 0 | ‘poslin’ for positive linear; ‘tansig’ for hyperbolic tangent sigmoid |
| Number of neurones L2 | Number of neurons for layer 2 | 0 | >=1 |
| Tolerance | Limits for considering out of range retrievals still possible | 0.1 | varies with output variable |
| Time of instantaneous fAPAR (HH.MM) | Used to document time of the fAPAR used during calibration | 10:00 | Local time hh:mm |
| Performance regularization | Additional error contribution from mean square magnitudes of weights. | 0.1 | 00.1 |
| Epochs | Maximum number of sweeps through training database | 250 | 250 |
| Performance function | Objective fuction metric used | mse | ‘mse’ for mean square error |
| goal | Stopping error level. | 1e-3 | 1e-3 |
| Update | Algorithm used to update network | Levenberg-Marquardt | Levenberg-Marquardt |

## Canopy and Atmosphere Parameters

Canopy and atmosphere parameter probability distribution functions (pdf) required to produce the calibration database are specified for each Class in separate worksheets named ‘Canopy\_Atmosphere\_Class#’ (Table **14**, Table **15**,Table **16**). For simplicity sheets must be follow increasing natural numbers corresponding to the class names identified in the ‘Learning’ worksheet. To facilitate independent validation, SL2P-CCRS uses single class worksheets only so there are separate databases for each land cover type. Like SL2P, SL2P-CCRS can simulate both to of canopy and top of atmosphere calibration datasets. Here, only top of canopy calibration is described. The atmosphere calibration is identical to that in WB2016.

There are currently 11 canopy parameters. Each parameter is specified using a truncated two parameter pdf subsequently scaled to reflect their covariation with LAI. Refer to <https://www.mathworks.com/help/stats/continuous-distributions.html> valid distributions and definition of the two parameters (P1 and P2). The exception being the use of three parameter extreme value distributions used for LAI pdf , where the third parameter is passed directly as a function argument to SL2P to avoid restructuring the ‘Canopy\_Atmosphere\_Class#’ worksheet. The MATLAB ‘trunc’ option (<https://www.mathworks.com/help/stats/prob.normaldistribution.truncate.html>) is used with the provided lower and upper bound parameters. This differs from SL2P that performed truncation after generating distributions that can result in oversampling at the truncation bounds.

Table 14. Canopy parameters for SL2P class .

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Theme** | **Variable** | **Lower**  **Bound** | **Upper**  **Bound** | **P1** | **P2** | **Nb\_Class** | **Law** | **LAI\_**  **Conv** | **Var\_min (0)** | **Var\_max (0)** | **Var\_min (LAI\_Conv)** | **Var\_max (LAI\_Conv)** |
| **Canopy** | **LAI** | 0.0 | 15.0 | 2.0 | 3.0 | 6 | Normal | 1000 | 0 | 20 | 0 | 20 |
| **ALA (°)** | 30 | 80 | 60 | 30 | 3 | Normal | 10 | 30 | 80 | 55 | 65 |
| **Crown\_**  **Cover** | 1.0 | 1.0 | 1.0 | 1.0 | 1 | Uniform | 10 | 1 | 1 | 1 | 1 |
| **HsD** | 0.1 | 0.5 | 0.2 | 0.5 | 1 | Normal | 1000 | 0.1 | 0.5 | 0.1 | 0.5 |
| **Leaf** | **N** | 1.20 | 2.20 | 1.50 | 0.30 | 3 | Normal | 10 | 1.20 | 2.20 | 1.30 | 1.80 |
| **Cab (µg.m-2)** | 20 | 90 | 45 | 30 | 4 | Normal | 10 | 20 | 90 | 45 | 90 |
| **Cdm (g.m-2)** | 0.0030 | 0.0110 | 0.0050 | 0.0050 | 4 | Normal | 10 | 0.0030 | 0.0110 | 0.0050 | 0.0110 |
| **Cw\_Rel** | 0.60 | 0.85 | 0.60 | 0.85 | 4 | Uniform | 10 | 0.60 | 0.85 | 0.70 | 0.80 |
| **Cbp** | 0.00 | 0.20 | 0.00 | 0.30 | 3 | Normal | 10 | 0.00 | 0.20 | 0.00 | 0.20 |
| **Soil** | **Bs** | 0.50 | 3.50 | 1.20 | 2.00 | 4 | Normal | 10 | 0.50 | 3.50 | 0.50 | 1.20 |
| **RTM** | **Gamma** | 1 | 1 | 1 | 1 | 1 | Normal | 10 | 1 | 1 | 1 | 1 |

Table 15 Canopy and atmosphere parameters for DBF class.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Theme** | **Variable** | **Lower Bound** | **Upper Bound** | **P1** | **P2** | **Nb\_Class** | **Law** | **LAI\_Conv** | **Var\_min (0)** | **Var\_max (0)** | **Var\_min (LAI\_Conv)** | **Var\_max (LAI\_Conv)** |
| **Canopy** | **LAI** | 0.0 | 10.0 | 0.0 | 2.0 | 6 | GenEx | 1000 | 0.0 | 10.0 | 0.0 | 10.0 |
| **ALA (°)** | 20 | 60 | 37 | 20 | 3 | Normal | 10 | 20 | 60 | 20 | 60 |
| **Crown\_Cover** | 0.05 | 1 | 0.05 | 1 | 6 | Uniform | 7 | 0 | 1 | 0.9 | 1 |
| **HsD** | 0.1 | 0.5 | 0.2 | 0.5 | 1 | Normal | 1000 | 0.1 | 0.5 | 0.1 | 0.5 |
| **Leaf** | **N** | 1.1 | 2.3 | 1.1 | 2.3 | 3 | Uniform | 10 | 1.1 | 2.3 | 1.1 | 2.3 |
| **Cab (µg.m-2)** | 20 | 60 | 20 | 60 | 4 | Uniform | 10 | 20 | 60 | 45 | 60 |
| **Cdm (g.m-2)** | 0.005 | 0.01 | 0.005 | 0.01 | 4 | Uniform | 10 | 0.005 | 0.01 | 0.005 | 0.01 |
| **Cw\_Rel** | 0.7 | 0.9 | 0.8 | 0.08 | 4 | Normal | 10 | 0.65 | 0.9 | 0.75 | 0.9 |
| **Cbp** | 0.00 | 0.20 | 0.00 | 0.30 | 3 | Normal | 10 | 0.00 | 0.20 | 0.00 | 0.20 |
| **Soil** | **Bs** | 0.50 | 3.50 | 0.00 | 0.60 | 4 | Lognormal | 10 | 0.50 | 3.50 | 0.50 | 1.20 |
| **RTM** | **Gamma** | 1 | 1 | 1 | 1 | 1 | Uniform | 10 | 1 | 1 | 1 | 1 |

Table 16 Canopy and atmosphere parameters for ENF class.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Theme** | **Variable** | **Lower Bound** | **Upper Bound** | **P1** | **P2** | **Nb\_Class** | **Law** | **LAI\_Conv** | **Var\_min (0)** | **Var\_max (0)** | **Var\_min (LAI\_Conv)** | **Var\_max (LAI\_Conv)** |
| **Canopy** | **LAI** | 0.0 | 10.0 | 0.0 | 2.0 | 6 | GenEx | 1000 | 0.0 | 10.0 | 0.0 | 10.0 |
| **ALA (°)** | 20 | 70 | 45 | 22 | 3 | Normal | 10 | 20 | 70 | 20 | 70 |
| **Crown\_Cover** | 0.05 | 1 | 0.05 | 1 | 6 | Uniform | 7 | 0 | 0.2 | 0.9 | 8.75 |
| **HsD** | 0.1 | 0.5 | 0.2 | 0.5 | 1 | Normal | 1000 | 0.1 | 0.5 | 0.1 | 0.5 |
| **Leaf** | **N** | 1.1 | 2.3 | 1.1 | 2.3 | 3 | Uniform | 10 | 1.1 | 2.3 | 1.1 | 2.3 |
| **Cab (µg.m-2)** | 20 | 60 | 20 | 60 | 4 | Uniform | 10 | 20 | 60 | 45 | 60 |
| **Cdm (g.m-2)** | 0.005 | 0.01 | 0.005 | 0.01 | 4 | Uniform | 10 | 0.005 | 0.01 | 0.005 | 0.01 |
| **Cw\_Rel** | 0.7 | 0.9 | 0.8 | 0.08 | 4 | Normal | 10 | 0.65 | 0.9 | 0.75 | 0.9 |
| **Cbp** | 0.00 | 0.20 | 0.00 | 0.30 | 3 | Normal | 10 | 0.00 | 0.20 | 0.00 | 0.20 |
| **Soil** | **Bs** | 0.50 | 3.50 | 0.00 | 0.60 | 4 | Lognormal | 10 | 0.50 | 3.50 | 0.50 | 1.20 |
| **RTM** | **Gamma** | 1 | 2 | 1 | 2 | 1 | Uniform | 10 | 1.00 | 2.00 | 1.00 | 2.00 |



One of five sampling schemes (Table **17**) are provided to sample the joint distribution of either canopy or atmosphere parameters. All of these schemes are unbiased so the calibrated regression will be unbiased if tested with samples from the same distribution. However, the precision of the calibrated network is a a function of training sample size and the effective dimensionality of the problem (which will lie somewhere between 1 and 11 for SL2P-CCRS). Full orthogonal sampling guarantees a reduction rate better than Monte Carlo and potentially as good as most of the other sampling schemes. Full orthogonal has been used for single class applications by SL2P but is problematic if the sample size must vary since increases in sample size must be performed in steps of ~ and resampling, for example when reusing simulations for another class, must be done carefully to avoid large gaps between samples. Scrambled Halton sampling offers better precision than Monte Carlo sampling when the effective dimensionality of the problem space, , is less than 3 but does not always guarantee better performance than the best case Full Orthogonal. However, scrambled Sobol sampling gives better precision the Full Orthog sampling for and better precision than Monte Carlo sampling for . Scrambled Sobol sampling is used as default for calibration. However, estimation of the regression performance is performed using Monte Carlo sampling to ensure there is no correlation between residuals due to the validation sample design.

Table 17. Sampling designs for calibration and validation.

|  |  |  |  |
| --- | --- | --- | --- |
| Sampling Design | Description | Theoretical Precision | MATLAB Implementation |
| Monte Carlo | Random sampling |  | rand |
| Full Orthogonal | Random sampling within full factorial sampling |  | fullfact |
| Latin Hypercube | Random sampling within Latin hypercube |  | lhsdesign |
| Quasi-randomized Halton Set | Halton set sampling followed by reverse-radix scrambling |  | haltonset followed by scramble with ‘rr2’ option |
| Quasi-randomized Sobol Set | Sobol set sampling followed by affine scrambling | < | Sobolset followed by scramble with 'MatousekAffineOwen' |

As in SL2P, for efficiency, sampling is initially performed assuming parameters are independent. In reality, canopy parameters exhibit covariation. Following SL2P only the covariation of the truncation bounds of each parameter with respect to LAI is specified. This is performed by using linear bounds as a function of LAI. Out of bounds samples must be resampled as discarding a sample will adjust the marginal distribution of all canopy parameters. SL2P resampled by reflecting out of bounds values into valid bounds (e.g. if Cab was out of bounds by 10gcm-2 it would be adjusted to the minimum of 30gcm-2 or the upper Cab bound for the current LAI). This heuristic tends to increase sampling near boundaries and can result in a pdf that deviates from the shape of the user specified pdf for the out of bound parameter (typically resulting in heavier tails). SL2P-CCRS resamples using a pdf specified with the new bounds. This ensures that the conditional distribution of parameters does not change for a given LAI although it does potentially increase the uncertainty of the sampling design somewhat since the final sampling design will now deviate from a hypothetically perfect design.

The combination of constraints on bounds and the sampling method results in sampled canopy parameters that exhibit trends in terms of range with respect to LAI and clumps (Figure **2**). Typically, the sample size is set at a level where sensitivity of the accuracy of estimated canopy variables to the sampling design is small. However, both Sobol and Halton sampling have fewer clumps suggesting that they may provide more uniform estimation performance for predicting algorithm uncertainty over the range of canopy parameters.

A group of blue dots

Description automatically generated

Figure 2. Comparison of sampling designs for the nominal SL2P canopy architecture distribution using only 1296 samples (in contrast to the nominal 43782 samples) for demonstration purposes.

The same canopy and atmosphere parameters used in SL2P are used here but with different distributions based on land cover class (Table **18**).

Table 18. Radiative transfer model and soil reflectance database for each land cover class.

|  |  |  |
| --- | --- | --- |
| Land Cover Class | Canopy | Soils |
| Vegetated | SL2P | Global |
| Evergreen Needleleaf forest | ENF | NA ENF |
| Deciduous Needleleaf forest | ENF | NA ENF |
| Evergreen Broadleaf Forest | DBF | Global |
| Deciduous Broadleaf Forest | DBF | Global |
| Mixed forest1 | DBF+ENF | None |
| Shrubland | DBF | Global |
| Herbaceous wetland | SL2P | Global |
| Moss&Lichen | SL2P | Global |
| Bare/sparse vegetation | SL2P | Global |
| Cropland | SL2P | Global |
| Polar shrubland | SL2P | Global |
| Grassland or pasture | SL2P | Global |
| Polar grassland | SL2P | Global |

1Mixed forest is estimated using a weighting of Decidual Broadleaf Forest and Evergreen Needleaf Forest

Atmosphere distribution parameters are fixed for all simulations following SL2P (Table **19**) although only top of canopy calibration is currently used and tested.

Table 19. Atmosphere parameters.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Lower Bound** | **Upper Bound** | **P1** | **P2** | **Nb\_Class** | **Variable** |
| **P (mbar)** | 0.5 | 3.5 | 0.5 | 3.5 | 1 | Uniform |
| **t550** | 950 | 1080 | 950 | 1080 | 1 | Uniform |
| **H2O (cm)** | 0 | 0.8 | 0 | 0.8 | 1 | Uniform |
| **O3 (dbs)** | 1 | 6.5 | 1 | 6.5 | 1 | Uniform |



# 6.0 Simulation

The Simulation process produces input-output samples for each Database Class for calibration, cross-validation and independent validation of regression estimators. Each sample input corresponds to satellite bi-directional reflectance (R) and acquisition geometry and the output corresponds to canopy biophysical variables ( Table **1**). Each calibration sample is generated by applying a sample of simulation Laws to the RTM. As such the RTM input requirements govern the quantities defined by the simulation Laws.

## 6.1 Radiative Transfer Model

The SL2P-CCRS RTM corresponds to the combination of a leaf optical model (PROSPECTD, Feret et al., 2017) modified to account for shoot clumping and the 4SAIL2 canopy RTM (Verehoef and Bach, 2007). This model is a generalization of the PROSAIL model (Jaquemoud et al., 2007) used in SL2P in terms of shoot clumping and crown clumping.

### Shoot Clumping

Needleleaf canopies often have foliage clumped in shoots. The single scattering albedo of shoots is generally less than that of needles due to multiple scattering within a shoot Smolander and Stenberg (xx). Here, the needle-to-shoot area ratio ), relates PROSPECTD leaf single scattering albedo, , to shoot single scattering albedo following Smolander and Stenberg (2001):

(1)

Leaf reflectance and transmittance are then scaled by and LAI by prior to use in 4SAIL2.

### Crown Clumping

Forest and shrub canopy foliage is organized in crowns. Crown clumping will typically decrease shadows and increase the likelihood of illuminating and viewing the understory in a manner that results in a non-trivial impact on R (Chen and Leblanc, 1997). PROSAIL does not account for crown clumping. There are many RTMs that account for crown clumping. The 4SAIL2 RTM model, with a single canopy layer, is used because i) it have been validated using international standard (Widlowski et al., 2007), ii) it relies on only one additional parameter compared to PROSAIL, the crown cover (CC) and iii) it is equivalent to the RTM used in PROSAIL when CC=1.

## 6.2 Simulation Laws

### Acquisition Geometry

The Acquisition Geometry Laws are identical for each Class since, in principle, land cover change due to disturbance or plantation can result in a Class appearing over most of the global domain considered by SL2P. These Laws are defined using either the Landsat 8 or Sentinel 2A orbital geometry over a domain defined by global vegetated land (Configuration). The Laws are uniformly sampled for each Database (Figure **3**).

A group of blue lines

Description automatically generated

Figure 3. Density and scatter plots of cosine of indicated geometric quantity for 40000 samples drawn from Geometry Laws for all SL2P-CCRS databases for S2A. Contours correspond to 0.2 probability intervals from 0 to 1.

### Canopy Laws

Canopy laws vary by Class since the prior joint distributions of canopy parameters are land cover specific. SL2P uses global in-situ datasets acquired circa 2016 to calibrate the prior distribution for each parameter independent on constraints due to covariation with other parameters. This decision will result in biased regression estimates over local conditions where these priors are not representative. Validation studies indicate the SL2P priors are generally unbiased over crops, grasslands, and herbaceous and cover but biased for forests and tall shrubs (Djamai et al., 2018; Brown et al. 202; Fernandes et al., 2023). As such, the SL2P priors are used for the SL2P Database/Class applied to non-woody land cover and new priors are defined for the ENF and DBF classes.

The new ENF and DBF priors are defined using measurements from North America and Europe and are therefore limited to biomes in these regions. Furthermore, in contrast to the SL2P philosophy of fitting priors closely to data, the prior distributions are fit using one of three approaches depending on a qualitative assessment of the representativeness of in-situ data according to CEOS Validation Sampling levels (<https://lpvs.gsfc.nasa.gov/>) (Table **20**). Only uniform or normal distributions were fit for variables with less than 30 samples or where sampling was possibly biased since these maximize the conditional entropy of the fit, and hence make the least statistical assumptions, constrained to the range (for uniform) or mean and standard deviation (for normal). For biased sampling with >30 measurements , a uniform distribution calibrated to the range of measurements was fit. For representative sampling for <30 sites a truncated normal distribution was fit but the fitted standard deviation was doubled to reflect the additional uncertainty due to the small sample size. For representative sampling with >30 sites a distribution maximizing the Aikike Information Criteria was fit using FITTER (<https://fitter.readthedocs.io/en/latest/>) truncated to the 2.5&ile to 97.5%ile range of physically valid measurements.

Table 20. Canopy LAW fitting approaches.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Level | Description | Distribution | Mean | Stdev | Range | Variables |
| 0 | Opportunistic Sampling <30 sites | Uniform | Mid point of range | Defined by range | Physically valid values | None |
| 1 | Biased sampling | Uniform | Mid point of range | Defined by range | Extreme values | CC,HsD, Cab, Cdm, Gamma |
| 2 | Representative sampling <30 sites | Truncated Normal | Mean value of samples | Twice stdev of samples | Extreme values | ALA, CbP, Bs |
| 3 | Representative sampling, >30 sites | Truncated Fitted | Fitted | Fitted | 2.5%ile to 97.5%ile | LAI, CW\_rel |

SL2P relies on the co-variation of canopy parameters with LAI to further constrain these priors. Here, the SL2P covariates were used for ENF and DBF as well due to lack of additional information. Covariation of the new gamma parameter was not enforced due to lack of data. Covariation for the new CC parameter was modelled using measurements from Fernandes et a al. (2023) and plausible ranges of the canopy foliage interaction cross-section (k) (Figure **4**). In this case, mixed forest data was used for fitting both DBF and ENF Classes. Further, the DBF class upped bound was not reduced at low LAI since senescent trees could have very low LAI for a given CC.

A comparison of a graph

Description automatically generated

Figure 4. Fitted minimum and maximum conditional range of canopy cover given LAI for DBF (left) and ENF (right).

The resulting Laws were used to generate 41472 samples for SL2P based on WD2016 and 248832 samples for ENF and DBF to account for the addition of CC and gamma parameters (Figure **5**)

A group of graphs showing different colored lines

Description automatically generated

Figure 5. Kernel density fits to 40000 samples from each canopy parameter Law as a function of Database Class. Gamma is constant at 1 for SL2P and DBF.

### Soil Laws

Soil spectra shapes were randomly sampled from Class specific databases corresponding to mineral and litter covered soils for SL2P and DBF (Figure **6**) and with the addition of moss and lichen for ENF from Miller et al. 1997 (Figure **7**). Spectra were scaled using a random value samples from a truncated log normal distribution with parameter 0.6 fitted to a histogram of scaling factors reported in WB2016. Note that WB2016 fit this histogram using a normal distribution with mode 1.2 and standard deviation 3 truncated between 0 and 3.5. Comparison of the shapes of (non-truncated) log normal and normal distributions with these parameters to the data from WB2016 suggest the log normal distribution, userd here, is more accurate (Figure **8**).

A graph of different colored lines

Description automatically generated

Figure 6. Soil spectra shapes for SL2P and DBF based on WB2016.

A graph of a spectrum

Description automatically generated

Figure 7. Soil spectral shapes for ENF based on both WB2016 and Miler et al., 1997.

A graph of a normal distribution

Description automatically generated

Figure 8. Comparison of histogram bar heights (blue circles) with fitted lognormal pdf (red) and WB2016 normal pdf prior to truncation.

### Simulated Variables

Response and regressor variables were simulated using the coupled PROSPECTD and 4SAIL2. The marginal distributions of structure and biochemistry variables were similar between ENF and DBF while the marginal distributions of albedo and D were similar between DBF and SL2P (Figure **9**). ENF differed from the broadleaf canopies in terms of albedo and D due to shoot clumping. Bivariate distributions of LAI, fAPAR and fCOVER showed less dispersion for SL2P compared to ENF and DBF due to the lack of clumping. Even so, fAPAR and fCOVER were almost linearly related as expected considering both are essentially canopy gap fraction measures.

A graph of a diagram

Description automatically generated

Figure 9. Kernel density plots of univariate (diagonal) and bivariate (lower) distributions of simulated response variables for each Class. Contours correspond to 0.2 probability intervals from 0 to 1.

The combination of darker soils and shoot clumping resulted in lower reflectance for ENF versus DBF although the shape of their distributions was very similar (Figure **10**). In contrast had lower visible and higher NIR reflectance than both ENF and DBF due to the uniform canopy cover masking the understory. The difference was negligible for SWIR as the SWIR reflectance of the understory and foliage was much more similar. Bivariate distributions indicate that only B07 and B08A within the red-edge were strongly correlated. This suggests that most bands provide complimentary information.

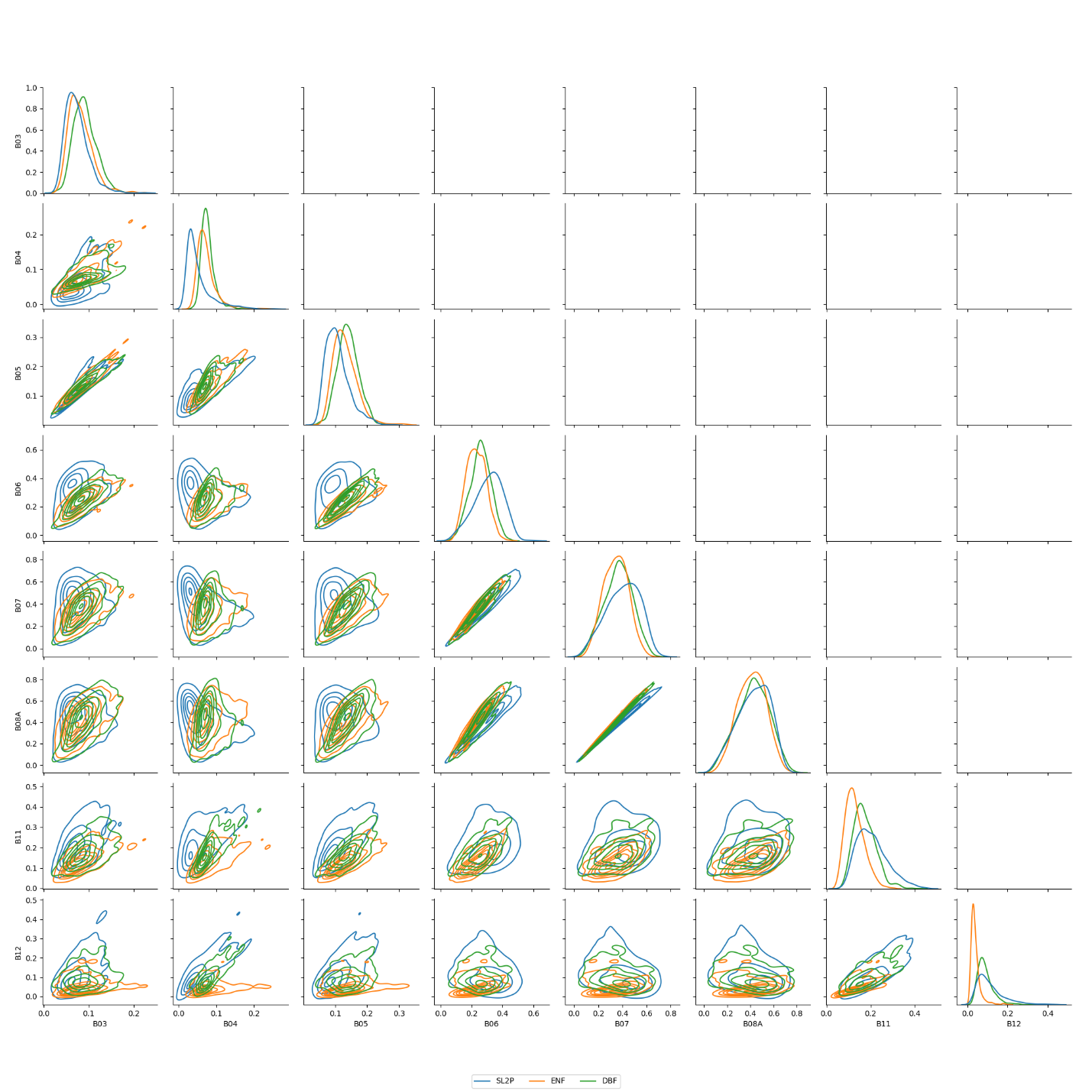


Figure 10. Kernel density plots of univariate (diagonal) and bivariate (lower) distributions of simulated reflectances for each Class. Contours correspond to 0.05,0.2,0.4,0.6,0.8 and 1.0 probability levels.

# 7.0 Performance

Following good practices (Fernandes et al., 2014) the performance of the regression models were evaluated using hold out cross-validation with the calibration database for each Class and output variable in terms of in terms of the overall Uncertainty Agreement Ratio (UAR) with respect to product user requirements (Table **1**) and the Accuracy (A) and Uncertainty (U conditional on the estimated value or on the reference validation value. The thematic performance should be considered optimistic as it assumes no RTM model error and correct priors.

## 7.1 Retrieval Cross-validation

### Albedo

The distribution of estimated versus validation albedo was similar for SL2P and DBF, with symmetry along the 1:1 line except for very low values. The residual plots also show that Sl2P and DBF have similar bias trend: slightly overestimating very low albedo, slightly underestimating moderate albedo and then showing a relatively large overestimate (>0.025) for albedo approaching 0.2. The bias trend for ENF is even more complicated, with an similar overestimate as the other classes at 0.02 albedo raplidly transitioning to an large underestimate at 0.3 albedo and then dropping again to an overestimate. The biases seen for high (>0.2) albedo are most likely due to the regression extrapolation as the prior probability of high albedo is low in the calibration database. This suggests that all three approaches should not be used for quantifying albedo, and possibly other variables, for very sparse vegetation over bright soils (e.g. arid regions) and for canopies with substantial senescent vegetation (advanced drought, curing of grasslands, or fall senesence of forests).

A collage of data

Description automatically generatedA collage of graphs and diagrams

Description automatically generatedA collage of graphs and diagrams

Description automatically generated

### FAPAR

A screenshot of several graphs

Description automatically generatedA collage of graphs and diagrams

Description automatically generatedA collage of graphs and diagrams

Description automatically generated

### FCOVER

A collage of graphs and diagrams

Description automatically generatedA collage of graphs and diagrams

Description automatically generatedA collage of graphs and diagrams

Description automatically generated

### LAI

A collage of graphs and diagrams

Description automatically generatedA group of images of a diagram

Description automatically generatedA group of images of different colored lines

Description automatically generated

### CWC

A collage of graphs and diagrams

Description automatically generatedA collage of graphs and diagrams

Description automatically generatedA collage of graphs and diagrams

Description automatically generated

### CAB

A collage of graphs and diagrams

Description automatically generatedA collage of graphs and diagrams

Description automatically generatedA group of graphs showing different types of data

Description automatically generated

## 7.2 Multiple Retrieval Independent Validation

Independent validation was performed using a hold out database of simulations with the same priors as the calibration data. Overall uncertainty is quantified using the UAR and the histogram of residuals rather than conventional metrics such as root mean square error or coefficient of determination as the latter depend on the magnitude of the retrievals. Additionally, overall uncertainty was quantified using averages of retrievals with similar inputs, defined as the k-nearest neighbours in input feature space, to approximate the effect of local spatial or temporal averaging. Both k=10 and k=100 were considered.

The UAR without averaging ranged from 0.72 to 0.97 across all Classes for variables not involving LAI and from 0.42 to 0.56 for variables involving LAI. SL2P UAR for single retrievals was between 0.05 and 0.15 higher than either ENF or DBF. These results indicate that LAI remains an ill-posed solution and that having to account for canopy clumping further increases uncertainty. These results may seem disappointing but spatial averaging of 20m pixels may be appropriate considering GCOS spatial requirements are 50m (~9 pixels) for adaptation studies and 250m (~150pixels) for climate modelling. The averaging treatments resulted in substantial improvements in UAR after only 10 nearest spectral neighbour pixels; with UAR exceeding 0.9 for all SL2P variables and 0.8 for DBF and ENF. This effect suggests that the single pixel retrievals are actually already close to meeting thematic uncertainty requirements and that even a little averaging brings many retrievals within requirements. It also suggests that retrievals are for the most part unbiased and errors are highly uncorrelated. Comparison of the distribution of residuals before and after averaging support the latter hypothesis; with residuals decreasing by a factor of ~3 for 10 pixel averaging and a further factor of ~2 for 100 pixel averaging for all variables but Albedo. The lower than expected decrease with 100 pixel averaging indicates that the residuals are not perfectly uncorrelated. The lack of improvement for Albedo when averaging errors suggests that random errors in input reflectance are not the major source of uncertainty in estimates for any of the classes. Rather, it seems that averaging regularize ill-posed solutions, especially for dense canopies where many solutions can closely match an observation.

Table 21. Uncertainty Agreement Ratio statistics.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | SL2P |  |  | DBF |  |  | ENF |  |
| Averaging | 1 | 10 | 100 | 1 | 10 | 100 | 1 | 10 | 100 |
| Albedo | 0.76 | 0.91 | 0.95 | 0.75 | 0.93 | 0.96 | 0.78 | 0.95 | 0.96 |
| fCOVER | 0.87 | 1.00 | 1.00 | 0.72 | 0.99 | 1.00 | 0.56 | 0.94 | 0.99 |
| fAPAR | 0.85 | 1.00 | 1.00 | 0.80 | 1.00 | 1.00 | 0.52 | 0.93 | 1.00 |
| LAI | 0.56 | 0.95 | 0.99 | 0.47 | 0.94 | 1.00 | 0.45 | 0.91 | 0.99 |
| LAI.Cab | 0.51 | 0.92 | 1.00 | 0.43 | 0.89 | 0.99 | 0.40 | 0.85 | 0.98 |
| LAI.Cwc | 0.61 | 0.89 | 0.96 | 0.40 | 0.81 | 0.95 | 0.37 | 0.76 | 0.92 |

A group of images of a diagram

Description automatically generated

Figure 11. Kernel density plots of SL2P cross-validation residuals for single sample versus smoothed retrievals based on 10 (blue) or 100 (orange) nearest neighbours in input space.

A group of graphs with blue and orange lines

Description automatically generated

Figure 12. Kernel density plots of DBF cross-validation residuals for single sample versus smoothed retrievals based on 10 (blue) or 100 (orange) nearest neighbours in input space.

A group of graphs with lines

Description automatically generated

## Quality Indicators

Quality indicators are provided for each retrieval using a single byte to code 8 logical indictors (Table 2223).

The terrain indicator identifies the designated terrain complexity. Complex terrain corresponds to approximately twice the expected noise of input measurements than simple terrain resulting in fewer out of range inputs but a larger retrieval uncertainty.

The LandCover indicator is set to Specific if a non-generic land cover is applied. By convention Class 1 is the generic vegetated land cover. Generic land cover will result in fewer out of range inputs and outputs but with a larger retrieval uncertainty if the correct land cover is known and specified.

The Input range indicator is set to Out of Range if the input measurements fall outside the convex hull of the noisy input training dataset.

The Output range indicator is set to Out of Range if the output estimates fall outside the convex hull of the output training dataset and the user specified range of LAI.

Table 22. Quality indictor coding.

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Bit | Value=0 | Value=1 |
| Terrain | 1 | Simple Terrain | Complex Terrain |
| LandCover | 2 | Vegetated | Specific |
| Input range | 3 | In range | Out of range |
| Output range | 4 | In range | Out of range |
| Unused | 5 | Unused | Unused |
| Unused | 6 | Unused | Unused |
| Unused | 7 | Unused | Unused |
| Unused | 8 | Unused | Unused |

# Algorithm Implementation

To implement SL2P-D one requires MATLABR2019 or later with the Statistics and Machine Learning package and the Parallel Processing toolbox. One could disable parallel processing by searching and replacing all ‘parfor’ commands with ‘for’ commands.

SL2P-X executes for the first time for a provided parameter file by invoking at the Matlab command line:

status = SL2PX(‘parameter\_file\_name,1,1,p3)

Here p3 is an optional parameter for extreme value LAI prior distributions. Successful execution will return a status of ‘1’ and produce a report directory, containing the calibration databases, as named in the parameter file. Validation is not essential.

SL2P-C can be executed multiple times to either calibrate additional regression algorithms or to apply a selected algorithm to a validation database. Both actions will result in additions to the Results.mat file for each class – either in terms of new networks or in terms of new validation results. If a new network is calibrated a new .xls worksheet will be added to the parameter file for the network details. If new validation is performed, will result in new visualizations for each class. One can make use of the .xls worksheet for a network to apply the regression algorithm to other datasets.

To apply SL2P-c in Google Earth Engine, the matlab function SL2P2GEE must be executed with an input database class directory specified. This function will produce one (using no partitioning) or more (using many partitions) CSV files corresponding to each network calibrated for the class. The user can then upload these CSV files as GEE assets. The LEAF Toolbox within GEE can then be configured to apply these networks based on a user specified index layer.

# Conclusions

The Simplified Level 2 Processor – CCRS for deriving vegetation biophysical variables from multispectral reflectance data is described. SL2P-C is generalization of SL2P to allow for canopy clumping at crown and shoot scales. The current CCRS implementation of SL2P-C relies on three cover classes: SL2P for non woody land cover, DBF for broadleaf woody land cover, and ENF for needleleaf woody land cover. The ENF class also includes moss and lichen soils. Cross-validation indicates SL2P-C is generally unbiased for LAI, fAPAR and fCOVER increasingly biased for Albedo >0.2 and shows inconsistent biases for canopy water content and canopy chlorophyll content.

Validation using independent simulations with the same priors as the calibration datasets indicates single SL2P-C retrievals meet GCOS threshold requirements over 80% of the time for Albedo, fCOVER and fAPAR and 45% of the time for LAI and canopy water content and canopy chlorophyll content . However, averaging 10 spectrally similar pixels reduces uncertainty by a factor of ~3 and results in meeting GCOS requirements between 80% and 100% of the time depending on variable and cover class. The averaging results indicate that SL2P-C retrieval errors will likely be spatially uncorrelated assuming zero mean (unbiased) input errors and could also be temporally uncorrelated over short intervals where canopy reflectance remains relatively constant. Validation against in-situ measurements should be performed with new protocols required to quantify the short length scale spatial and short extent temporal pattern of residuals.

SL2P-C and sample parameters files and output are available at <https://github.com/rfernand387/SL2PC/upload/master/Reports>. This resource can be used with Matlab Version 2019 and the Statistics and Machine Learning toolbox to apply SL2P to user databases formatted to meet the current simulation databases. SL2P-C can be applied to imagery using the LEAF-Toolbox (<https://github.com/rfernand387/LEAF-Toolbox> ). Further testing with the FLIGHT RTM (for forests) and using validation with in-situ measurements is underway.

# References

[Canada's National Forest Inventory](https://nfi.nfis.org/en/standardreports), 2013, Table 5.1 Area (1000 ha) of forest land by forest type, age class in Canada, 2006 revised baseline, Version 3, December 2013 (accessed April 7, 2017).

Brown, L., Fernandes, R.A., 2021. Validation of baseline and modified Sentinel-2 Level 2 Prototype Processor leaf area index retrievals over the United States. IISPRS J. Photogramm. Remote Sens., 175, pp. 71-87, [10.1016/j.isprsjprs.2021.02.020](https://doi.org/10.1016/j.isprsjprs.2021.02.020" \t "_blank).

[Djamai and Fernandes, 2018](https://www.sciencedirect.com/science/article/pii/S0034425719301117" \l "bbb0180), **Comparison of SNAP-derived sentinel-2A L2A product to ESA product over Europe**, Remote Sens., 10 (2018), p. 926.

Djamai, N.; Fernandes, R.; Weiss, M.; McNairn, H.; Goïta, K. Validation of the Sentinel Simplified Level 2 Product Prototype Processor (SL2P) for mapping cropland biophysical variables using Sentinel-2/MSI and Landsat-8/OLI data. Remote Sens. Environ. 2019, 225, 416–430.

Fernandes, R., C. Butson, S. Leblanc, R. Latifovic, 2003.

Landsat-5 TM and Landsat-7 ETM+ based accuracy assessment of leaf area index products for Canada derived from SPOT-4 VEGETATION data.

Can. J. Remote. Sens., 29 ,[10.5589/m02-092](https://doi.org/10.5589/m02-092" \t "_blank).

[Fernandes et al., 2014](https://www.sciencedirect.com/science/article/pii/S0034425719301117" \l "bbb0225), **Global Leaf Area Index Product Validation Good Practices**, Committee of Earth Observing Systems Working Group on Calibration and Validation (2014), p. 75.

Fernandes, R., Maloley, M., Canisisu, F., 2018. Relationship between leaf area index and Landsat Operational Land Imager equivalent reduced simple ratio vegetation index for the Athabasca oil sands region, northern Alberta. Geomatics Canada, Open File 39, <https://doi.org/10.4095/308333>

Fernandes, R.A. and Djamai. N., 2019. Well posed retrieval of canopy biophysical variables from multi-spectral imagery using the canopy directional scattering factor. In preparation.

Fernandes, R., Luke Brown, Francis Canisius, Jadu Dash, Liming He, Gang Hong, Lucy Huang, Nhu Quynh Le, Camryn MacDougall, Courtney Meier, Patrick Osei Darko, Hemit Shah, Lynsay Spafford, Lixin Sun, 2023. Validation of Simplified Level 2 Prototype Processor Sentinel-2 fraction of canopy cover, fraction of absorbed photosynthetically active radiation and leaf area index products over North American forests, Remote Sensing of Environment, 293, https://doi.org/10.1016/j.rse.2023.113600.

Féret J-B, Gitelson AA, Noble SD & Jacquemoud S, 2017. PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle. Remote Sensing of Environment, 193, 204–215. <https://doi.org/10.1016/j.rse.2017.03.004>

Jacquemoud S, Verhoef W, Baret F, Bacour C, Zarco-Tejada PJ, Asner GP, François C & Ustin SL, 2009. PROSPECT+ SAIL models: A review of use for vegetation characterization. Remote Sensing of Environment, 113:S56–S66. <https://doi.org/doi:10.1016/j.rse.2008.01.026>

Lang, M. and Tychon, B. 2015. GIO Copernicus Global Land Technical User Group Service Specification and Product Requirements Proposal, SPB-GIO-307-TUG-SS-004.

Miller, J. R., et al. , 1997. , Seasonal change in understory reflectance of boreal forests and influence on canopy vegetation indices, J. Geophys. Res., 102( D24), 29475– 29482, doi:[10.1029/97JD02558](https://doi.org/10.1029/97JD02558" \o "Link to external resource: 10.1029/97JD02558" \t "_blank).

Verhoef W & Bach H, 2007. Coupled soil–leaf-canopy and atmosphere radiative transfer modeling to simulate hyperspectral multi-angular surface reflectance and TOA radiance data. Remote Sensing of Environment, 109:166-182. <https://doi.org/10.1016/j.rse.2006.12.013>

Weiss and Baret, 2016, **S2ToolBox level 2 products. Version 1.1** , Available online [step.esa.int/docs/extra/ATBD\_S2ToolBox\_L2B\_V1.1.pdf](http://step.esa.int/docs/extra/ATBD_S2ToolBox_L2B_V1.1.pdf" \t "_blank)

Widlowski, J-L., M. Taberner, B. Pinty, V. Bruniquel-Pinel, M. Disney, R. Fernandes, J.-P. Gastellu-Etchegorry, N. Gobron, A. Kuusk, T. Lavergne, S. Leblanc, P. Lewis, E. Martin, M. Mottus, P. J. R. North, W. Qin, M.Robustelli, N. Rochdi, R.Ruiloba, C.Soler, R.Thompson, W. Verhoef, M. M.Verstraete, and D. Xie (2007), 'The third RAdiation transfer Model Intercomparison (RAMI) exercise: Documenting progress in canopy reflectance modelling', Journal of Geophysical Research-D.

World Meteorological Organization, 2022a, The Global Observing System for Climate: ECV Requirements 2022, WMO Pub No. GCOS – 245.

World Meteorological Organization, 2022b, The Global Observing System for Climate: Space Agencies Supplement 2022, WMO Pub No. GCOS – 246.

World Meteorological Organization, 2022c, The Global Observing System for Climate: Implementation Plan 2022, WMO Pub No. GCOS – 244.

# Appendix I – Physical variables

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fraction Canopy Cover | FCOVER | % | Fraction of horizontal surface area covered by vegetation | 0-1 |
| Leaf Area Index | LAI | DIM | Half the total live foliage area per horizontal ground area | 0-15 |
| Canopy Chlorophyll Content | CCC | g.m-2 | Mass of chlorophyll a+b per unit ground area | 0-100 |
| Canopy Water Content | CWC | g.m-2 | Mass of water per unit ground area | 0-100 |
| Albedo, black sky | A | 0-1 | Ratio of top of canopy upper hemispherical upwelling radiance to top of canopy incident direct irradiance | 0-1 |
| D | D | 0-1 | canopy scattering coefficient for a foliage single scattering albedo of one under direct irradiance | 0-1 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Heterogenity | None | Nominal | Canopy heterogeneity | HomogenousHet: heterogenous |
| Senes. Frac | - | 0-1 | Fraction of matter that is senescent | 0-1 |
| Woody Frac | - | 0-1 | Fraction of matter that is woody | 0-1 |
| Age | - | years | Age since first growth | 0-100 |
| Leaf angle distribution | LAD | Discrete probability histogram | Proportion of leaves for angles: [xx),[xx),[xx),[xx),[xx),[xx] | 0-1; sum 1 |
| Leaf diameter | d | cm | Diameter of leaf | 0-100 |
| Leaf layers | N | Fractional count | Number of effective scattering layers in a leaf | 0-4 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Chlorophyll a+b content | Cab | µg.cm-2 | Foliage chlorophyll a+b content | 0-100 |
| Water content | Cw | g.cm-2 | Foliage water content | 0-100 |
| Dry matter content | Cdm | g.cm-2 | Foliage dry matter content | 0-100 |
| Brown pigment content | Cbd | g.cm-2 | Foliage brown pigment content | 0-100 |
| Carotenid content | Ccar | µg.cm-2 | Foliage carotenid content | 0-100 |
| Xanthophyll content | Cx | µg.cm-2 | Foliagexanthophyll content | 0-100 |

# Appendix II SL2P-C Configurations for Canadian Ecozones

Table 23. Spatial partitions for calibration datasets. Regions correspond to Ecozones of Canada. Snow End (Onset) dates taken from earliest (latest) dates observed over CCRS, NOAA and MODIS snow cover poducts betweenb 2006 and 2010..

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Region | Area | LatMin | LatMax | Long Min | Long Max | Snow End | Snow Onset |
| 1000km2 | °N | °N | °E | °E | DOY | DOY |
| Global | 100000 | 41 | 83 | -141 | -47 | 1 | 365 |
| Canada | 10000 | 41 | 83 | -141 | -47 | 1 | 365 |
| Northern Arctic | 2886 | 58 | 83 | -128 | -62 | 122 | 294 |
| Southern Arctic | 1673 | 53 | 72 | -141 | -68 | 104 | 308 |
| Arctic Cordillera | 338 | 57 | 83 | -95 | -61 | 140 | 308 |
| Taiga Cordillera | 265 | 60 | 70 | -141 | -125 | 104 | 308 |
| Boreal Cordillera | 468 | 56 | 65 | -141 | -122 | 86 | 322 |
| Taiga Shield | 816 | 51 | 60 | -80 | -55 | 86 | 322 |
| Boreal Shield | 2072 | 44 | 60 | -112 | -47 | 86 | 350 |
| Taiga Plain | 654 | 57 | 69 | -137 | -113 | 86 | 322 |
| Hudson Plains | 447 | 50 | 59 | -96 | -75 | 86 | 336 |
| Mixedwood Plains | 169 | 41 | 48 | -85 | -69 | 1 | 365 |
| Boreal Plains | 738 | 49 | 62 | -123 | -96 | 32 | 336 |
| Atlantic Maritime | 281 | 43 | 50 | -73 | -59 | 1 | 365 |
| Prairies | 466 | 49 | 55 | -114 | -95 | 1 | 365 |
| Montane Cordillera | 448 | 49 | 58 | -131 | -112 | 1 | 365 |
| Pacific Maritime | 245 | 48 | 60 | -138 | -121 | 1 | 365 |

Day of Year, min and max correspond to Snow End and Snow Onset dates respectively

# Appendix III SL2P-C Canopy biophysical variable parameterizations

This section describes the calibration of prior probability distribution functions for RTM model parameters.

## LAI

LAI representing North American Forests ( Fernandes et al., (2003), Fernandes et al., 2018, and Fernandes et al. (2023)) were used to calibrate pdf fits qualitatively. Qualitative calibration was performed since standard fitters cannot handle truncated distributions and to ensure the calibrated distribution dispersion was larger than the sampling distribution to account for the limited sample size. Mixed forests were included in both broadleaf and needleleaf class calibrations since monocultures are relatively infrequent in North American forests. A total of 359 plots were used for calibrating the broadleaf forest prior and 516 plots for the needleleaf forest prior. The resulting distribution was identical for both cover classes: generalized extreme value, mean 3.2, standard deviation 2, kurtosis parameter -0.4.

A comparison of graphs with red lines

Description automatically generated

Figure 13. Calibrated distributions (red) and sampled histograms (black) for two forest classes.

## Leaf Angle Distribution

Leaf angle distribution priors were defined using uniform distributions due to limited samples (xx, xx).

## Crown Cover

Crown cover was defaulted to a uniform distribution with range specified by the 2.5%ile to 97.5%ile interval of data from Fernandes et al. (2023). A uniform distribution was selected both because the sample size was limited in location and because crown cover is frequently manipulated by management practices not sampled in Fernandes et al. (2023).

## Hot Spot Dimension

The hotspot dimension parameter corresponds to the ratio of vertical layer separation to foliage mean width. In the absence of additional data the SL2P prior was used. Note that the 4SAIL2 model provides a stronger hotspot effect due to crown structure that is parameterized by crown cover and crown shape ratio. Here, the crown shape ratio ranged with a uniform distribution between 1 and 4 based on literature data (xx).