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

Richard Fernandes[1], Luke Brown[2], Francis Canisius[1], Jadu Dash[3], Liming He[1], Gang Hong[1], Lucy Huang[1], Nhu Quynh Le[1], Camryn MacDougall[1], Courtney Meier[4], Patrick Osei Darko[5], Hemit Shah[1], Lynsay Spafford[6], Lixin Sun[1]

[1]Canada Centre for Remote Sensing, Natural Resources Canada, 580 Booth Street, Ottawa, Canada, K1A 0Y7

[2]School of Science, Engineering & Environment, University of Salford, Manchester M5 4WT, United Kingdom

[3]School of Geography and Environmental Science, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom

[4]National Ecological Observatory Network, Boulder, CO 80301, United States

[5]Department of Geography, McGill University, 805 Rue Sherbrooke O, Montréal, QC H3A 0B9

[6]Department of Environmental Sciences, Faculty of Sciences, , Memorial University of Newfoundland, Antigonish, Canada

Highlights

- CEOS Stage 3 validation of SL2P over North American forests.
- SNAP implementation of SL2P exhibits artifacts due to coding errors.
- SL2P underestimates LAI; lower bias for fAPAR and fCOVER.
- Bias correction reduces LAI and fAPAR error for temperate and boreal forests.

Abstract

Canopy biophysical variables such as the fraction of canopy cover (fCOVER), fraction of absorbed photosynthetically active radiation (fAPAR), and leaf area index (LAI) are widely used for ecosystem modelling and monitoring. The Sentinel-2 mission was designed for systematic global mapping of these variables at 20m resolution using imagery from the MultiSpectral Instrument. The Simplified Level 2 Prototype Processor (SL2P) is available as a baseline mapping solution. Previous validation over limited sites indicates that SL2P generally satisfies user requirements for all three variables over crops but underestimates LAI over forests. In this study, Sentinel-2 fAPAR, fCOVER and LAI products from SL2P were validated over 281 sites representative of most North American forest ecozones and also compared to Moderate Resolution Imaging Spectrometer (MODIS) and Copernicus Global Land Service (CGLS) products. In addition to meeting the Committee on Earth Observation Satellites Stage 3 validation requirements for these areas, our study also explores the relationship between bias in SL2P products and canopy clumping and provides empirical bias correction functions for each variable.

SL2P was implemented within the Landscape Evolution and Forecasting Toolbox in Google Earth Engine both for efficiency and due to bugs in the Sentinel Application Platform (SNAP) implementation. SL2P was found to underestimate LAI by 20% to 50% over forests with LAI>2; in agreement with other studies and with comparisons to MODIS and CGLS products. SL2P bias for fCOVER and fAPAR transitions from ~0.1 at low values to ~-0.1 at high values. Precision error, at one standard deviation, was ~0.5 for LAI and slightly less than ~0.1 for fCOVER and fAPAR. Total uncertainty was dominated by bias for LAI and slightly greater than precision error for fCOVER and fAPAR. Target user requirements were satisfied for 51% of LAI, 37% of fCOVER and 31% of fAPAR comparisons. For all variables, accuracy exhibited weak to moderate linear relationships to clumping (r^2<=0.52) but scatter plots indicated larger negative LAI biases over northern latitude sites where canopies exhibited greater clumping. With the exception of evergreen broadleaf forests, empirical bias correction using in-situ data reduced LAI accuracy error by up to 96% and increased the agreement rate with uncertainty requirements by up to 8%. Users of SL2P LAI over forests are recommended to apply bias correction or consider recalibrating SL2P with spatially heterogenous radiative transfer model simulations.

1. Introduction

A primary goal of the Sentinel-2 (S2) mission is the systematic mapping of canopy variables, including the fraction of canopy cover (fCOVER), fractional of absorbed photosynthetically active radiation (fAPAR), and leaf area index (LAI), using measurements from the MultiSpectral Instrument (MSI) on a constellation of polar orbiting satellites (ESA Sentinel-2 Team, 2007). The European Space Agency (ESA) sponsored the development of the Simplified Level 2 Prototype Processor (SL2P) for mapping these variables using Level 2A bottom of atmosphere reflectance (L2A) products derived from MSI data (Weiss and Baret, 2016). SL2P versions are implemented in ESA's Sentinel Application Platform (SNAP) 9.0.0 (http://step.esa.int), used by the European Union SEN4CAP agricultural sustainability project ((http://esa-sen4cap.org/) and the open source LEAF-Toolbox (Fernandes et al., 2021) implemented in Google Earth Engine (GEE), and used by the Government of Canada Earth Observation for Cumulative Effects Monitoring Programme (Janzen et al., 2020).

S2 mission requirements correspond to Global Climate Observing System requirements (GCOS, 2022; "Goal" in Table 1). The Copernicus Global Land Service (CGLS) has also identified less stringent requirements considered acceptable by downstream services such as crop monitoring, hydrological and ecosystem models and habitat mapping (Sanchez-Sapero and Martinez-Sanchez, 2022; "Target" in Table 1). Good practice for validation requires quantification of uncertainty (U), accuracy (A), precision (P) and uncertainty agreement ratio (UAR), defined in Table 2, as a function of the reference product value. (Fernandes et. al., 2014). The Committee on Earth Observation Satellites (CEOS, https://lpvs.gsfc.nasa.gov/) describes four validation levels corresponding to increasing spatial and temporal representativeness of validation statistics: Level 1, validation over <30 sites and time periods; Level 2, validation of >30 sites and time periods together with intercomparisons with similar products; Level 3, validation over >30 sites with validation statistics quantified as a function of reference conditions and with globally representative intercomparison; and Level 4 corresponding to ongoing Level 3 validation with fiducial reference networks.

Table 1. Product definitions (Sanchez-Sapero, J. and Martinez-Sanchez, 2022 and GCOS, 2022) and user requirements for target (CGLS) goal (GCOS) thematic uncertainty levels.

Variable	Definition	Goal	Target
fAPAR	Fraction of absorbed photosynthetically active radiation by green vegetation for a given solar illumination condition.	max(10%,0.05)	max(10%,0.05)
fCOVER	Fraction of ground covered by green vegetation.	5%	max(10%,0.05)
LAI	Half the total foliage area per unit horizontal ground area.	15%	max(15%,0.5)

Table 2. Thematic error metrics used for product validation.

Metric	Acronym	Definition
Uncertainty	U	Expected value of the absolute difference of estimated and product values.
Accuracy	A	Expected value of the estimated value minus the product value.
Precision	P	Square root of the expected value of the square of the total of the estimated value minus both the product value and the accuracy metric.
Uncertainty agreement ratio	UAR	The fraction of validated samples that meet a given uncertainty requirement, in this case, GCOS.

Level 2 and Level 3 validation has been performed for low resolution (>250m) LAI, fAPAR, and fCOVER products (Weiss et al., 2007; Garrigues et al., 2008; Weiss et al., 2014; Camacho et al 2013; Yan et al., 2016; Brown et al., 2020; Fuster et al., 2020). Some of these studies predate community good practices for quantifying errors as a function of retrieved or reference value (Fernandes et al., 2014) and none of these studies quantify the uncertainty of validation metrics due to limited reference measurements. There are few Level 2 studies validating medium resolution products due to the limited number of systematic products, the difficulty of product generation for large areas, and the complexity of matching reference and higher spatial resolution product values. Brown et al. (2021a) performed a level 2 validation of SL2P LAI that included detailed error analysis but did not perform intercomparison with other products. In fact, to date, SL2P has only been validated at Level 2 partly because co-incident in-situ data is limited and partly due to the challenges in processing sufficient products for intercomparison using the SNAP toolbox. Nevertheless, SL2P L2 validation indicates it generally satisfies target requirements for crops (Djamai et al., 2019; Hu et al., 2020; Brown et al., 2021a) but underestimates LAI and fAPAR over dense forests (Putzenlechner et al., 2019; Brown et al., 2021a) (Supplementary material

Table S1). Brown et al. (2021a) hypothesised the LAI underestimation was due to spatial clumping of foliage not accounted for within SL2P but did not test this with data. This also raises the concern that error of SL2P fAPAR and fCOVER estimates will also increase as spatial clumping increases. However, SL2P has not been simultaneously validated for LAI, fAPAR and fCOVER over high latitude forests that exhibit substantial clumping (He et al., 2012).

A Level 3 validation of SL2P estimates of LAI, fCOVER and fAPAR over a broad geographical range of forests would allow product users to integrate validation statistics in down stream applications, improve our understanding of the limitations of SL2P, and serve to prioritise improvements to SL2P or similar algorithms for mapping vegetation parameters using the S2 MSI. Additionally, Level 3 validation may facilitate bias correction using empirical statistical models between the SL2P response and in-situ reference measurements. For example, Brown et al. (2020) demonstrated that linear regression models between SL2P and reference measurements can produce unbiased estimates of LAI and fAPAR with a UAR exceeding 74%. The extent to which this strategy can be applied at continental scale has yet to be determined.

Here, SL2P is validated using in-situ reference measurements (RM) for 14 of 17 forest ecological zones of North America (Commission for Environmental Cooperation, 2022; Figure 1), not including tropical rainforests, temperate steppe and subtropical dry forests, and compared to previously validated coarse resolution Moderate Resolution Imaging Spectrometer (MODIS) and CGLS products. A new Canadian field campaign was conducted to address the issue of limited in-situ sampling. To facilitate product intercomparison, SL2P was implemented in GEE to produce products over a replicate sample of 100km x 100km regions of all North America forest ecological zones. Even so, this study does not evaluate the temporal stability of SL2P products as, prior to 2019, imagery over North America was not systematically processed to Level 2A surface reflectance products by ESA.

The goals of our study are to:

I) validate SL2P LAI, fAPAR and fCOVER estimates over typical North American Forests by quantifying their accuracy, precision, uncertainty and uncertainty agreement ratio,

II) determine if indeed SL2P accuracy for each variable is related to canopy clumping,

III) assess the applicability of empirical bias correction to reduce retrieval error, and

IV) provide good practices for achieving a CEOS Stage 3 validation of medium resolution vegetation biophysical variable products.

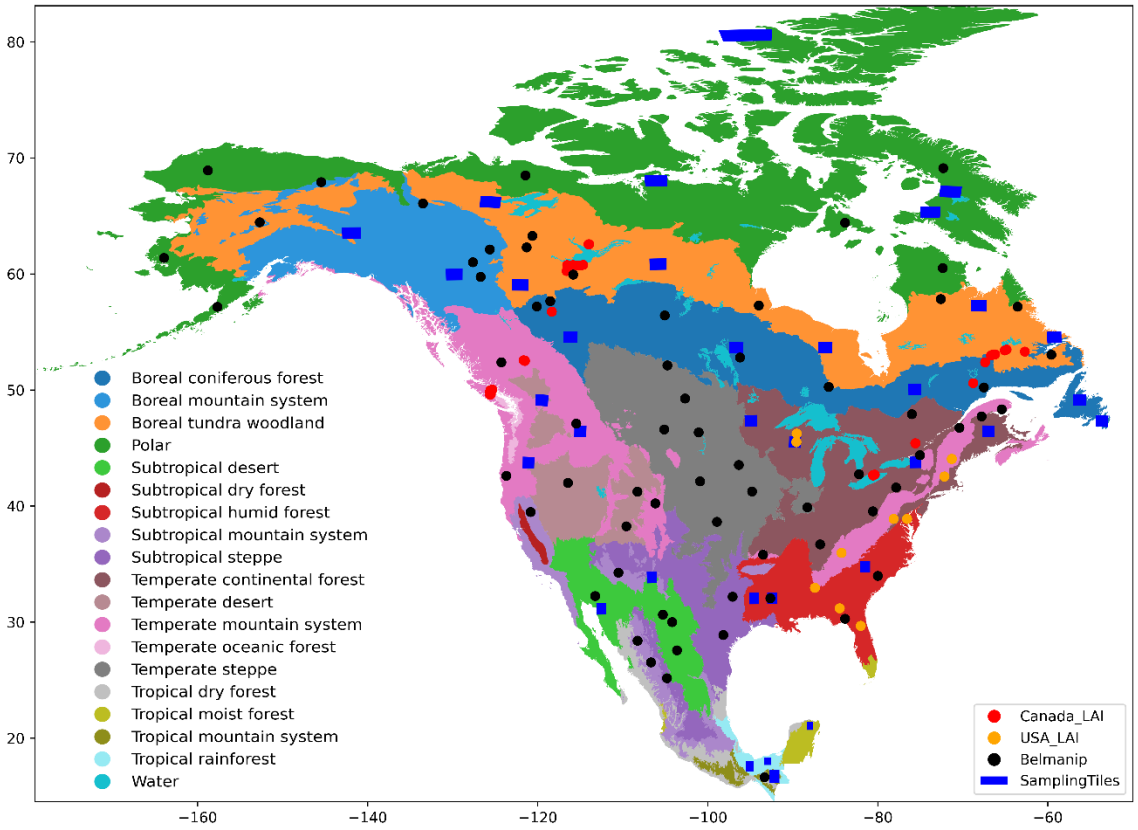


Figure 1. Location of intercomparison regions (blue rectangles) and in-situ reference measurement sites (red and yellow circles) within North American Forest Ecological Regions. Global BELMANIP intercomparison sites are indicated as well with black circles

2. Data Sets

2.1 Geospatial Datasets

Geospatial datasets described in the following subsections were used for SL2P product generation, stratification of intercomparison, and product intercomparison (Table 3).

Table 3. Geospatial data sets.

Dataset	Description	Access	Reference
CGLSV1	333m 10d fAPAR (at 10.15 Local Time), fCOVER and LAI	http://land.copernicus.eu/global/products/	Verger and Descals, 2022.
MERIT DEM	3 arc second Multi Error Removed Improved DEM	https://developers.google.com/earth-engine/datasets/	Yamazaki, et al. 2017
MCD15	500m 4d MODIS Collection 6 Leaf Area Index/FPAR	https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MCD15A3H	Myneni et al., 2015
MSI L2A	MSI Level 2A Bottom of Atmosphere reflectance	https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR	Copernicus Sentinel data 2019,2020
Forest Ecozones	North American Forests Primary Ecological Zones	https://www.cec.org/north-american-environmental-atlas/north-american-forests-2022/	Commission for Environmental Cooperation, 2022
NALC2015	30m North American Land Cover 2015	https://www.cec.org/north-american-environmental-atlas/land-cover-30m-2015-landsat-and-rapideye/	Commission for Environmental Cooperation, 2020

2.1.1 MERIT-DEM

MERIT-DEM is a 3 arc second resolution digital elevation model (DEM) produced by combining a number of available DEMs. The vertical uncertainty is <±9m over forested areas with slope <10% (Yamazaki, et al. 2017).

2.1.2 CEC Forest Ecozones

CEC Forest Ecozones is a polygon coverage of forest ecological zones for North America produced jointly by the Governments of Canada, United States of America and Mexico and published by the Commission for Environmental Cooperation (2022) (Figure 1). The map indicates 17 different primary forest ecological zones based on a combination of climate and potential vegetation classifications. The thematic error of this map is not given as it is a potential rather than actual geophysical dataset.

2.1.3 NALC2015

NALC2015 is a 30m resolution land cover map for North America circa 2015 with an 18-class legend (Table 4) (Commission for Environmental Cooperation, 2020). The land cover is based on peak growing season satellite imagery from Landsat 5 and 8, as well as RapidEye over the U.S.A., from 2015; with missing pixels replaced using the peak season estimate in the most recent valid year. The thematic error of products using the same monitoring system has been assessed over Canada with 79.9% correct labelling for all 18 classes and 83% correct labelling of forest classes (Latifovic et al., 2012). The spatial uncertainty of NALC2015 is less than 5m, 67.5% circular error probable. The NALC2015 legend was translated to International Geosphere Biosphere Programme (IGBP) classes (Lambin and Geist, 2006) used to label the RM sites.

2.1.4 MCD15

MCD15 corresponds to the global 4-day composites of fAPAR and LAI gridded at 500m resolution. Retrievals are produced using the MODIS Version 6.1 fAPAR and LAI algorithm (Myneni et al., 2015) applied to MODIS imagery from Terra and Aqua satellites. Here, only primary algorithm retrievals corresponding to the maximum fAPAR in the 4-day interval are selected. The thematic error of MCD15 over forests is reported in a number of studies (Supplementary Materials, Table S2). Brown et al. 2020 report a root mean square difference (RMSD) between 0.48 to 1.05 for LAI and 0.09 to 0.14 for FAPAR for 547 samples at 18 sites across North America. Yan et al. (2016) report a maximum LAI (fAPAR) residual of +2/-1 LAI and ±0.2 fAPAR for 50 samples with reference LAI ranging from 1.2 to 6 and fAPAR from 0.25 to 0.9. MCD15 underestimated LAI by between -0.14 (Brown et al., 2020) and -0.41 (Fuster et al., 2020) over ENF and by -1.47 for MF (Lin et al., 2017). The geolocation uncertainty of MCD15 is

204 better than 53m root mean square error (RMSE; Lin et al., 2019) although the projected instantaneous
205 field of view (PIFOV) of measurements can vary by a factor of 4.83 across-track (2.01 along track) for
206 extreme view angles over flat terrain due to the MODIS 110˚ field of view (Wolfe et al., 1998).

207 *Table 4. NALC2015 land cover classes, IGBP Class acronym and forest land class designation. IGBP Classes: mixed forest (MF),*
208 *deciduous broadleaf forest (DBF), evergreen needleleaf forest (ENF), grassland (GR), shrub (SH). evergreen broadleaf forest*
209 *(EBF), barren land (BL), cropland (CR), wetland (WL), urban (UB), water (WA), snow or ice (SI).*

NALC2015 Class	IGBP Class	Forestland Class
Temperate or sub-polar needleleaf forest	ENF	Yes
Sub-polar taiga needleleaf forest	ENF	Yes
Tropical or sub-tropical broadleaf evergreen forest	EBF	Yes
Tropical or sub-tropical broadleaf deciduous forest	DBF	Yes
Temperate or sub-polar broadleaf deciduous forest	DBF	Yes
Mixed forest	MF	Yes
Tropical or sub-tropical shrubland	SH	Yes
Temperate of sub-polar shrubland	SH	Yes
Tropical or sub-tropical grassland	GR	No
Tropical or sub-polar grassland	GR	No
Sub-polar or polar shrubland-lichen-moss	SH	No
Sub-polar or polar grassland-lichen-moss	GR	No
Sub-polar or polar barren-lichen-moss	GR	No
Wetland	WL	Yes
Cropland	CR	No
Barren lands	BL	No
Urban and built-up	UB	No
Water	WA	No
Snow/Ice	SI	No

210
211

212 2.1.5 CGLSV1

213

CGLSV1 corresponds to CGLS Ocean Land Colour Instrument (OLCI) Version 1.1 fAPAR, fCOVER and LAI products gridded at 300m resolution for non-overlapping 10-day periods (Verger and Descals, 2022). CGLSV1 is a Level 4 product in that temporal interpolation using both current and historical retrievals for a mapped pixel are used to filling in empty periods. Only non-interpolated pixels were used in this study. The input Level 2 fAPAR and LAI products are derived by separate neural networks that relate OLCI surface reflectance to CGLS ProbaV Products (PBV 300m V1 v1.0; Baret et al., 2016). The ProbaV products are in turn estimated using a weighting of corresponding MCD15 Version 5 (Yang et al., 2006) and CYCLOPES V3.1 (Baret et al., 2007) products; with the CYCLOPES V3.1 weighing transitioning from >80% for LAI<1 to <20% for LAI>2. This weighting reduces the independence of reference MCD15 and CGLS products since the latter are essentially scaled versions of the former for LAI>2. Level 2 fCOVER is derived using a fixed logarithmic relationship between fCOVER and LAI.

CGLS OLCI products have not been extensively validated due to geolocation issues as only tie-points rather than image chip control points were used when generating CGLS L4 products from Level 2 inputs (Prikaziuk et al., 2021) These geolocation error is generally less than 333, and its impact on intercomparisons is reduced in our study by using 5x5 pixel CGLS product averages during intercomparison. However, the error has inhibited CGLSv1 OLCI product validation against in-situ measurements. Nevertheless, CGLS ProbaV products using a virtually identical algorithm adapted for ProbaV spectral bands has been extensively validated (Supplementary Material Table S3) . Brown et al., 2020 reported ProbaV uncertainty of 0.25 to 0.91 for LAI and 0.05 to 0.09 for FAPAR for 538 samples at 18 sites across North America. Over the same sites, Fuster et al. (2020) reported much larger residuals for ProbaV products (often exceeding ±1 LAI and ±0.1 fAPAR) but did not use spatial weighting with ancillary layers when upscaling in-situ RM. Information regarding the OLCI PIFOV variation is not available but it is likely less than MODIS given the smaller 68.5˚ field of view of OLCI (Bourg et al., 2021).

2.1.6 MSI L2A

MSI L2A data corresponds to bottom-of-atmosphere reflectance (ρ) processed from MSI top-of-atmosphere L1B products by ESA from the MSI on S2A or S2B satellites, using version 2.10 or higher of the Sen2Cor algorithm (Müller-Wilm, 2018). Clear sky pixels over land and water are mapped with 98% accuracy and radiometric uncertainty is better than 0.005 + 0.05ρ for flat surfaces (Doxani et al., 2018); although the latter could increase substantially over terrain with adjacent slopes exceeding 10° (Djamai and Fernandes, 2019). The geolocation uncertainty is less than 12.5m (95 percentile of the circular error) (Gascon et al., 2017). The full width half maximum point spread function ranges from 22.0 m for the 10m Band 4, to 33.4 m and 39.1m for Band 5 and Band 11 respectively (Radoux et al., 2016). PIFOV variation is far less than the pixel size as the field of view is only 20.6˚ (Gascon et al., 2017).

2.2 In-Situ Reference Measurements

In-situ RM spanning 2019 and 2020 were acquired from the Ground-Based Observations for Validation (GBOV) component of the CGLS (Brown et al., 2021b) and by the authors at Canada Centre for Remote Sensing (CCRS) (Table 5). Elementary Sampling Units (ESUs), corresponding to the spatial footprint of a RM, were located within long term monitoring sites (GBOV) or regional transects (CCRS). ESUs were centred within patches of ~100m x 100m that were qualitatively assessed as having similar canopy characteristics based on high resolution imagery and in-situ survey.

2.2.1 GBOV

The fraction of intercepted PAR (fIPAR), fCOVER and LAI RM were derived from Digital Hemispherical Photographs (DHPs) for 142 Elementary Sampling Units (ESUs) at 14 forest or shrubland sites within the National Ecological Observatory Network (Barnett et al., 2019) in North America (Table 5, Figure 1) using the method described in Brown et al. (2021a). GBOV fIPAR is defined as the black-sky PAR intercepted by overstory and understory vegetation at 10:00 local solar time and LAI as the one-sided leaf area per unit ground surface area. In fact, GBOV LAI actually corresponds to half the total plant area per unit horizontal ground area (PAI).

At each site, three 20m by 20m square ESUs, located within 1km of NEON tower locations, were sampled bi-weekly from leaf-out to senescence. In each ESU, 12 co-located upward and downward

looking DHP images were acquired with 4m spacing in North-South and East-West transects through the plot centre (Figure 2)using 36.3MPixel Nikon D810 or D800 cameras (https://en.nikon.ca/nikon-products/product-archive/dslr-cameras/d810.html) with a Nikon 16mm Fisheye lens(https://en.nikon.ca/nikon-products/product/camera-lenses/af-fisheye-nikkor-16mm-f%252f2.8d.html) giving a 180° diagonal field of view (FOV). The ESU centre location was determined within 1m 90% circular error probable. Five sites included additional intensive sampling dates at up to 20 additional ESUs distributed throughout the dominant vegetation types across each site (land cover types < 5% of total site area were not sampled). Details regarding the estimation of gap fraction from DHP images, and subsequently fIPAR and PAI, are given in Appendix B. of Brown et al. (2020). Briefly, upwards and downward images were processed separately. fIPAR (fCOVER) was estimated as one minus the mean gap fraction within ±5° of the solar zenith angle at 10:00 local solar time (±5° of nadir). PAI was estimated as the average of effective PAI estimates for 10° azimuthal intervals within ±5° of 57.5° zenith angle to minimize sensitivity to leaf angle. The effective PAI for each azimuth interval was estimated as the negative logarithm of the gap fraction divided by the cosine of the zenith angle. The effective PAI for the entire image was also quantified using the mean gap fraction for all azimuths at 57.5° zenith angle. ESU mean PAI or PAIe was determined by the sum of the mean PAI or PAIe for upward and downward DHPs. The 1σ uncertainty of each RM was computed following Fiducial Reference Measurements for Vegetation guidelines, by propagating the variability in gap fraction at the ESU level through each measurement equation (Brown et al., 2021b), whilst uncertainty due to instrument levelling was based on Origo et al. (2017).

2.2.2 CCRS

fIPAR, fCOVER and PAI RM were derived from DHPs acquired for 133 ESUs in Natural Resources Canada's Cumulative Effects study sites across Canada (Table 5, Figure 1). ESUs were located within the dominant land cover types at each site with replication where logistics permitted. For each ESU, seven co-located upward and downward DHP images were acquired every 5m along two parallel transects spaced 15m apart (Figure 2) using 45.7Mpixel Nikon D850 cameras (https://en.nikon.ca/nikon-products/product/dslr-cameras/d850.html#tab-ProductDetail-ProductTabs-Overview) with a Nikon 8mm Fisheye lens (https://en.nikon.ca/nikon-products/product/camera-lenses/af-s-fisheye-nikkor-8-15mm-f%252f3.5-4.5e-ed.html) giving a 180° FOV in all directions . The ESU centre was located to

within 5m 90% circular error probable. In addition, the IGBP land cover class and the approximate surface cover fraction of bryophyte, lichen, mineral soil or litter was noted.

DHPs for each ESU sampling date were quality controlled visually, contrast enhanced using ViewNXi software (https://en.nikon.ca/nikon-products/product/imaging-software/viewnx-i.html) and masked to remove the field operator. CANEYE V6.45 (https://www6.paca.inrae.fr/can-eye/Download/) was used to derive RM estimates as well as PAIe and associated 1σ uncertainties for either the upward or downward looking DHPs acquired during a ESU visit. CANEYE uses the same approach as GBOV to derive the RM but also provides an alternate PAIe and PAI estimate that minimises the difference of the observed gap fraction for each position in the hemisphere and the modelled gap fraction with a penalty function proportional to the deviation from the PAIe estimate within ±5° of 57.5° zenith. Agreement of the GBOV and CANEYE algorithms is a necessary condition if indeed the canopy was sufficiently sampled and DHPs were adequately processed. As part of the CCRS protocol, ESU measurements where these two approaches differed in excess of their average standard error were processed after further quality control and enhancement. This approach also ensured a level of consistency between GBOV and CCRS RM.

The standard error of RM estimates for each DHP sampled during an ESU visit was used to quantify RM measurement error using the same approach applied with GBOV. However, as CANEYE does not directly provide the fIPAR for each DHP, the fCOVER within plot relative standard error was used to approximate the fAPAR standard error, recognizing that this will be pessimistic as the former is based on gap fraction within ±10° of nadir and the latter at 10:00 local solar time.

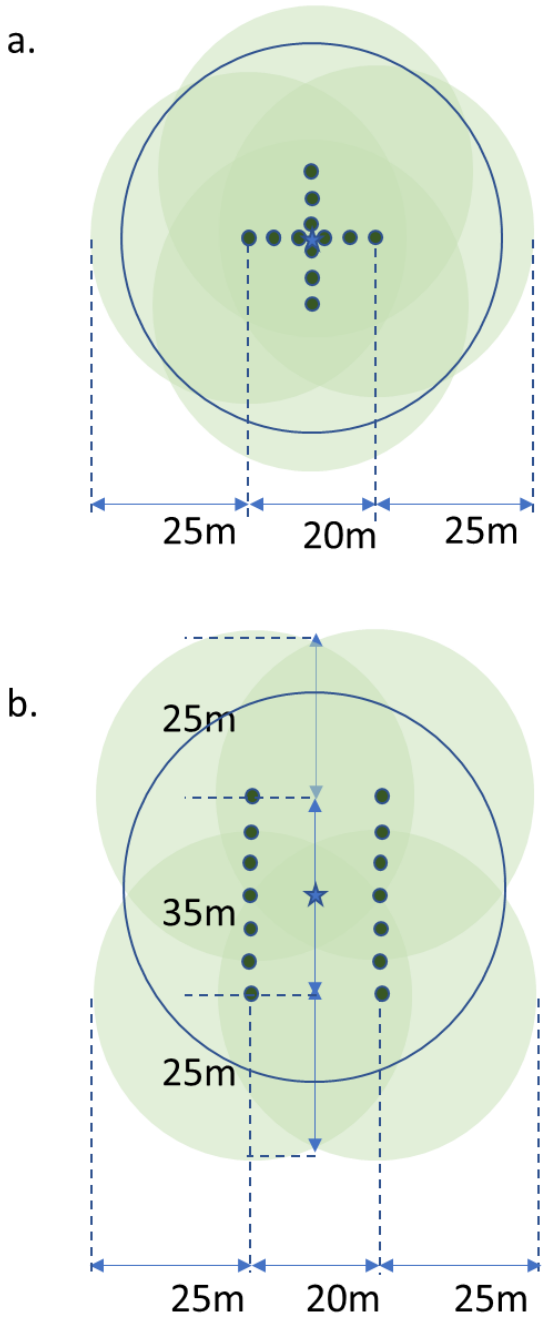


Figure 2. Schematic of a. NEON and b. CCRS Elementary Sampling Unit (ESU) design. Star: ESU centre; dark green shaded circles: understory DHP 60° FOV for 50cm canopy; light green shaded circles: overstory DHP 60° FOV for 20m tall canopy; blue open circle: 30m radius SL2P product sampling buffer.

337 *Table 5. In-situ sites. ESU: Elementary Sampling Unit. IGBP Class defined in Table 4.*

Site	Ecoregion	IGBP Class	Lat. °N	Long. °E	Elev. (m a.s.l.)	Date Start	Date End	#ESU	Network
Bartlett Exp. Forest	Temperate continental forest	MF	44.06	-71.28	232	2019-05-13	2020-10-12	3	GBOV
Blandy Exp. Farm	Temperate continental forest	DBF	39.08	-77.95	183	2019-03-26	2020-10-14	22	GBOV
Dead Lake	Subtropical humid forest	DBF	32.53	-87.8	22	2019-03-13	2020-08-07	22	GBOV
Disney Wilderness Preserve	Tropical moist Forests	MF	28.12	-81.44	15	2019-01-03	2020-12-15	24	GBOV
Geraldton	Boreal coniferous forest	ENF	49.85	-86.88	348	2020-07-15	2020-07-21	56	CCRS
Guanica Forest	Tropical Dry Forests	ENF	17.97	-66.87	143	2019-01-07	2020-12-16	3	GBOV
Harvard Forest	Temperate continental forest	MF	42.54	-72.17	351	2019-05-08	2020-10-06	3	GBOV
Hay River	Boreal coniferous forest	ENF	60.57	-116.12	165	2019-09-05	2019-09-07	28	CCRS
Jones Ecological Research	Temperate continental forest	ENF	31.20	-84.47	44	2019-01-10	2020-12-08	24	GBOV
Joronarda	Subtropical steppe	SH	32.59	-106.84	36	2019-01-10	2020-12-08	3	GBOV
Labrador	Boreal tundra woodland	DBF	52.65	-66.13	12	2019-07-24	2019-07-31	7	CCRS
Mer Bleue	Temperate continental forest	ENF	45.40	-75.57	86	2019-09-18	2019-09-18	3	CCRS
Moab	Temperate desert	SH	38.25	-109.39	1799	2019-01-10	2020-12-08	3	GBOV
MtPolley	Temperate mountain system	SH	52.52	-121.55	917	2019-08-12	2019-08-15	6	CCRS
Nova Scotia	Temperate oceanic forest	SH	45.18	-63.03	34	2021-06-09	2021-08-27	7	CCRS
Oak Ridge	Subtropical mountain system	SH	35.96	-84.28	334	2019-04-16	2020-10-24	3	GBOV
Onaqui Ault	Temperate desert	SH	40.18	-112.45	1685	2019-03-20	2020-08-26	3	GBOV
Ordway	Subtropical humid	EBF	29.70	-81.99	45	2019-	2020-	3	GBOV

Swisher Biological Stn.	forest					01-28	11-13		
Peace River	Boreal Mountain System	ENF	56.74	-118.35	330	2019-08-12	2019-08-12	9	CCRS
Santa Rita	Subtropical desert	MF	31.91	-110.84	983	2019-03-04	2020-10-06	3	GBOV
Smithsonian Conservation Biology Inst.	Subtropical humid forest	ENF	38.89	-76.56	361	2019-05-21	2020-10-15	3	GBOV
Talladega National Forest	Subtropical humid forest	ENF	32.95	-87.39	135	2019-03-19	2020-07-01	3	GBOV
Turkey Point	Temperate continental forest	DBF	42.68	-80.46	222	2019-06-25	2019-06-27	4	CCRS
University Notre Dame Conservation	Temperate continental forest	DBF	46.23	-89.54	518	2019-05-08	2020-09-30	23	GBOV
Vancouver Island	Temperate oceanic forest	MF	49.84	-125.54	50	2019-08-09	2019-08-10	7	CCRS
Yellowknife	Boreal Tundra Woodland	ENF	62.56	-113.99	206	2019-08-09	2019-08-12	6	CCRS

3. Methods

3.1 Reference Measurements

RM were uploaded to GEE as feature collections (Fernandes et al., 2023). PAIe, PAI, fIPAR and fCOVER were converted to LAIe, LAI, fAPAR and green fCOVER, respectively, by multiplying the measured value by woody area to total area ratios given in Brown et al. (2021a) for the overstory and 0.05 (±0.025 1σ) for the understory. The understory value was selected assuming herbaceous and shrub understory cover have non-zero woody to total area ratio that is typically less than trees due to absence of trunks. Total LAIe, LAI, fAPAR and fCOVER were estimated by combining understory and overstory values as in Brown et al. (2021a). Clumping was calculated as the ratio of total LAIe to total LAI (Chen and Cihlar, 1996). A

clumping of one corresponds to a canopy with random foliage locations while lower values indicate foliage that has increasing spatial clumping.

The 1σ uncertainty of overstory or understory components of each RM was estimated as the Euclidean sum of the 1σ uncertainties due to levelling error, sampling variability, the applied woody to total area ratio, and for LAI, a 0.025 1σ uncertainty due to clumping. The clumping uncertainty is the Euclidean sum of the half the reported 0.03 root mean square error difference between the Liang and Yuequin (1986) and Chen and Cihlar (1995) clumping estimates for a range of sites and the ~0.02 change in estimated clumping when using 10 versus 20 sampling points in a plot reported in Ryu et al. (2010). This approach assumes that the two methods for estimating clumping are equally uncertain although their accuracy varies with canopy type (Leblanc et al., 2005; Woodgate et al., 2017). The 1σ uncertainty of the corresponding total RM value was estimated as the Euclidean sum of the constituent understory and overstory 1σ uncertainties weighted by their proportion of the total RM.

3.2 Validation

SL2P uses separate non-linear regression predictors to estimate the expected value (retrieved estimate) and the standard error (theoretical precision) of LAI, fAPAR and fCOVER given L2A ρ and associated acquisition geometry (Weiss and Baret, 2016). Each non-linear regression is calibrated using a database of 42,378 simulated L2A ρ with associated canopy variables produced by applying the turbid spatially homogeneous (i.e. clumping=1) PROSAILH radiative transfer model (RTM) (Verhoef et al., 2007) to a database of canopy parameters sampled from globally representative priors. Additionally, retrievals are flagged as out of domain (out of range) if the inputs (outputs) fall outside of the domain (range) of the calibration database simulations with a 10% tolerance as described in Djamai et al. (2019). The LEAF-Toolbox implementation of SL2P was used to map LAI, fAPAR and fCOVER and associated quality control and theoretical precision layers at 20m resolution for clear sky land pixels whose centroid fell within a 30m radius of the centre of each ESU and ±7d interval of each RM.

The SL2P implementation in SNAP was not used as two bugs were identified in the MATLAB libraries provided to SNAP and subsequently verified by their author (M. Weiss , personal communication): i) incorrect truncation of the prior probability distributions used to specify realizations of simulated

canopies for calibration of the regressions used to estimate product values, and ii) a coding error in the algorithm used to flag out of domain retrievals that significantly overestimated the frequency of samples falling outside the domain. To demonstrate the impact of these bugs for current users of SNAP, SNAP and LEAF-Toolbox products were compared over the validation site used in Djamai et al. (2019).

The ±7d interval for matches ensured at least three MSI acquisitions for each sample date given the <5d revisit of Sentinel-2A and 2b MSI imagers. Both temporal differences in SL2P matches during the start and end of season or during rapid disturbance or drought and SL2P measurement error due to geolocation or atmospheric correction uncertainty could result in some matched dates having much larger residuals than others. In circumstances where multiple dates were matched to a single RM, these errors were reduced by discarding dates where at least half the samples exceeded the 50%ile absolute residual of all matched pixels for the 15d period at the ESU.

Ideally, the spatial footprint of matching product pixels should correspond to the spatial footprint of in-situ measurements (Fernandes et al., 2014). This was not possible to implement here as spatially explicit canopy height information was not available for CCRS ESUs. Also, product pixel footprints will vary with the L2A pixel PIFOV that in turn depends on actual geolocation, view zenith angle and spectral band (Gascon et al., 2017). To avoid these complexities a fixed 30m radius was used corresponding to the approximate footprint of the ESUs with the tallest canopies measured. This choice assumes that, for shorter canopies, the ESU RM also applies within the spatial footprint of matching product pixels.

The 1σ uncertainty of the mean SL2P estimate for a given match-up was modelled as the Euclidean sum of the mean SL2P theoretical and the standard error of the matching SL2P samples. For each variable, statistics for A, U, P and uncertainty agreement ratio (UAR) and coefficient of determination (r^2) were computed as in Brown et al. (2021a) both for the entire population of RM and by IGBP cover class.

Good practice requires reporting thematic error metrics as a function of the RM value to allow for comparisons of validation results across studies with different sampling distributions (Fernandes et al., 2014; Doxani et al., 2018; Brown et al., 2021a). To do so, A, P and U were modelled using third order polynomial weighted least squares regressions fitted to quantities based on residuals between the mean of matching SL2P product pixels and RM values for each sampled ESU. For A (U) the residual (absolute residual) between SL2P estimates and RM was regressed against the RM. For P, the SL2P estimates

were first corrected by adding the modelled A followed by regressing the absolute residuals between the corrected SL2P estimates and RM against the RM. Weighting was inversely proportional to the sum of squares of the 1σ uncertainty of the mean SL2P estimate and the 1σ uncertainty of the RM. The expected conditional value and ±95%ile confidence intervals of each regression were computed using statsmodel version 01.3 (https://www.statsmodels.org/stable/index.html). Population level statistics of A,P,U and UAR were also computed as:

$$A = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i) \quad (1)$$

$$U = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \quad (2)$$

$$P = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i - A)^2} \quad (3)$$

$$UAR = \frac{1}{N}\sum_{i=1}^{N} I\left(\left|\frac{\hat{y}_i - y_i}{y_i}\right| \leq \varepsilon_{rel} \cup |\hat{y}_i - y_i| \leq \varepsilon_{max}\right) \quad (4)$$

Where $\hat{y}_i, y_i$ are, respectively, the SL2P estimate and RM for the i^{th} of N comparisons; $\varepsilon_{rel}, \varepsilon_{max}$ are, respectively, the relative and maximum target uncertainty requirement (Table 1) and I is the indicator function.

Brown et al. (2020) demonstrated that accuracy of local SL2P LAI and fAPAR maps can be improved by bias correction using empirical relationships between SL2P retrievals and local RM. Here, for each variable, this approach was extended to all sites by fitting a third order polynomial weighted least squares regression to predict the RM given the corresponding SL2P estimate using all match-ups. The weights were inversely proportional to the Euclidean sum of the 1σ uncertainty of the mean SL2P estimate and the 1σ uncertainty of the of the RM. The fitted polynomials were validated by fitting and applying similar hold out site specific calibrations to the SL2P estimates at each site and then applying the same validation protocol used for uncorrected SL2P estimates. The hold out approach was used to ensure statistical independence between validation RM and the bias correction.

3.3 Intercomparison

Product intercomparison should be performed over replicate regions sampled within unique land surface conditions (Fernandes et al., 2014). The BELMANIP2 sampling design was previously developed to sample regions based on strata representative of global biomes, land cover and phenology together with the constraint that sample regions minimize conditions not relevant to the stratum and have relatively flat terrain (Weiss et al., 2014). The last condition reflects the fact that complex terrain can result in both radiometric and geolocation uncertainty that can mask differences in the retrieval algorithms. BELMANIP2 was not used here for three reasons: i) it misses two North American forest ecozones, ii) even within a forest ecozone, BELMANIP2 regions are located to match the expected distribution of all land cover rather than only forest cover, and iii) BELMANIP2 does not consider the number of valid product intercomparisons available within forested regions. Instead, since our study only considers North American forests, we relied on the North American forest ecozone map to stratify by geographic location, forest type, and phenology. Sample regions were restricted to 100km x 100km Military Grid Reference System (MGRS) tiles (Defence Mapping Agency, 1990) since they offer a global equal area grid and because L2A products are formatted using these tiles.

MGRS tiles were scored using the product of four relative criteria scores to select tiles that maximise potential valid intercomparisons (Table 6). Each relative criteria score corresponds to a raw criteria score divided by the sum of the same raw score of all tiles overlapping the ecozone. The product of relative scores ensured that the selected MGRS tile would not rank low on any one of the relative scores. For each forest ecozone, all MGRS tiles with at least 50% overlap with the ecozone were scored and the two tiles with the highest score selected.

Table 6. Raw criteria scores for selecting MGRS tiles.

Name	Definition	Inputs
Forest Area	Average forest area within 3x3 MCD15 pixel footprints.	NALC2015, MCD15
Land Area	Average land area within 3x3 MCD15 pixel footprints.	NALC2015, MCD15
Elevation Deviation	1σ of MERIT elevation within 3x3 MCD15 pixel footprints.	MERIT DEM, MCD15
Vegetation	Growing season average of 1σ SL2P NDVI within a 3x3	MSI L2A Products April-September,

Homogeneity	MCD15 pixel footprints	2019, MCD15
Clear Sky Count	Number of dates with >90% coverage of clear sky MSI L2A retrievals in each 3x3 MCD15 pixel footprint.	MSI L2A Products April-September, 2019, MCD15

Match-ups were extracted for each L2A product acquired during 2019 over the selected MGRS tile using aggregated valid coarse resolution product pixels to reduce uncertainty due to geolocation or spatial footprint variability. For each match-up sample of non-overlapping 3x3 (for MODIS) or 5x5 (for CGLS) coarse resolution product pixels, areas labelled as either water in the L2A scene classification mask, water or built-up in the NALC2015 land cover, or L2A normalised difference vegetation index less than 0.1, were assigned biophysical parameter estimates of zero. SL2P was applied to other areas in the footprint, using the corresponding 20m L2A product, to estimate canopy biophysical parameters and associated theoretical precision. Intercomparisons were considered valid if at least 90% of the match-up footprint area was mapped with valid SL2P retrievals or zero values and with 100% coverage of valid highest quality coarse resolution product retrievals. For each match-up, the expected value of the SL2P product estimate was estimated using area-weighted binning followed by dividing by the mapped match-up footprint area. Additionally, the standard deviation of valid retrievals and the proportion of each MCD15 biome type was determined for each MODIS or CGLS pixel using SL2P output and a NALC2015 land cover respectively.

4. Results

4.1 Verification

Comparisons of SNAP and LEAF-Toolbox implementations of SL2P indicated differences in both LAI and quality masks (Figure 3). SNAP overestimates (underestimates) LEAF-Toolbox for LAI>4 (LAI< 2) (Figure 3e). SNAP also typically designates >50% of retrievals as invalid due to being 'Out of Domain' of the calibration dataset in comparison to only ~10% for LEAF-Toolbox (compare Figure 3c. to Figure 3d). Further, SNAP maps all L2A reflectance measurements including clouds while LEAF-Toolbox only maps clear sky land using the L2A cloud mask.

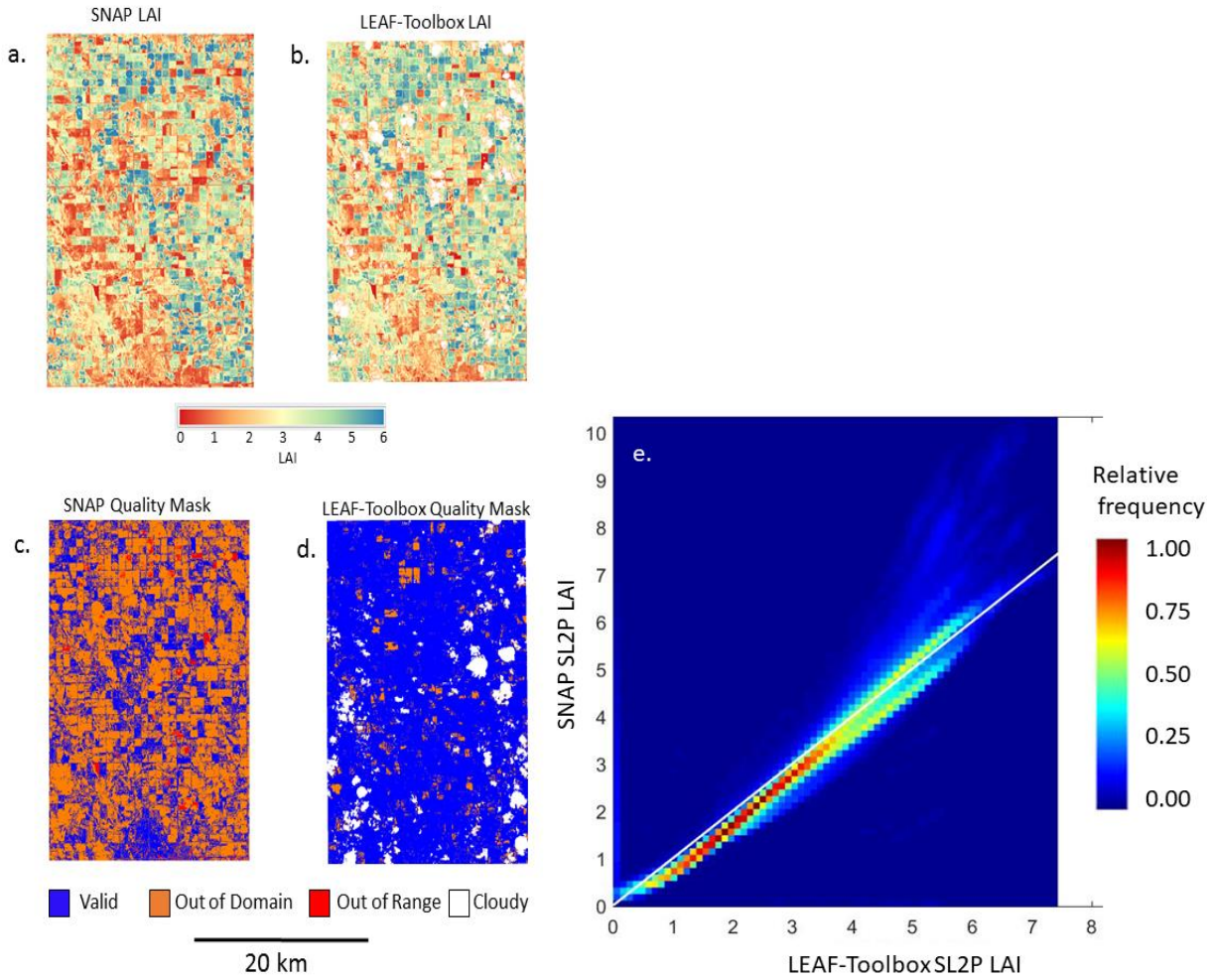


Figure 3. Comparison of SL2P LAI and quality mask implemented in (a. and c.) SNAP and (b. and d) LEAF-Toolbox. as well as a (e.) scatterplot of SNAP versus LEAF-Toolbox LAI for the agricultural region reported in Djamai et al. (2019).

4.2 Sampled SL2P and RM Estimates

On average, 8.07 (range [6.5,28]) matching SL2P pixels were found over each of 1107 RM samples of LAIe, LAI, fCOVER and fAPAR (Supplementary Material Table S4). RM values spanned 0.01 to 7.41 for LAI and 0.002 to 0.95 for both fAPAR and fCOVER (Figure 4). The RM measurement uncertainty was typically on the order of ±1 unit for LAIe and LAI and ±0.05 units for fAPAR and fCOVER (Supplementary Material Figure S1)

The modal LAI value for MF and DBF was ~4.5 although the DBF sites had second mode of ~1.5 due to the inclusion of early season NEON sampling. ENF sites had a lower modal LAI than other forests (~1.5) due to the northern latitude CCRS sites. Closed SH (CSH) sites showed a relatively uniform distribution of LAI between ~2-3 while open SH (OSH) sites had a narrow mode at ~0.5 LAI.

As expected, fAPAR and fCOVER RM were almost linearly related (r^2=0.98) since both are weighted gap fraction estimates from the same DHP samples. Both quantities were logarithmically related to LAI (r^2>0.91) as expected from the relationship between gap fraction and LAI (Monteith and Unsworth, 2014). Except for shrublands, where clumping was always near 1, there was no clear relationship between clumping and land cover; although clumping was generally lower for CCRS sites versus NEON sites even though both use the same method to estimate clumping (Figure 4).

Univariate and bi-variate distributions of SL2P estimates (Figure 5) were similar to their RM counterparts but with somewhat greater range and stronger linear relationships between fAPAR and fCOVER (r^2 =0.99) and logarithmic relationships between these variables and LAI (r^2>0.95). The stronger relationships reflect the fact that SL2P assumes homogenous turbid canopies while the RM include variations in spatial clumping and crown shape that also impact gap fraction (Stenberg et al., 2014). In contrast to the RM distributions, there was no visible distinction in bivariate distributions of SL2P products between CCRS and NEON sites since all products were derived using the same radiative transfer model and therefore have same biases due to clumping and crown shape.

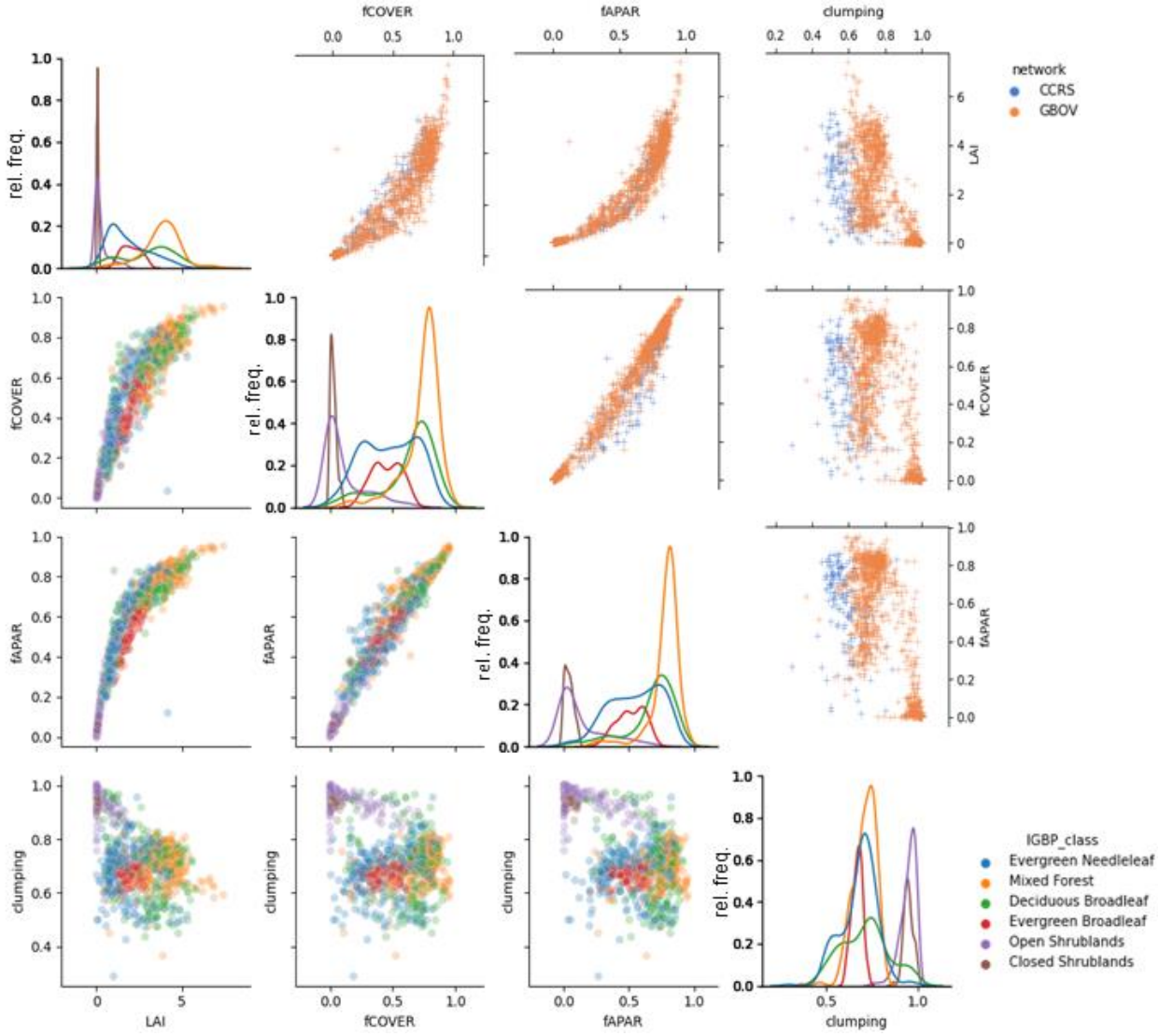
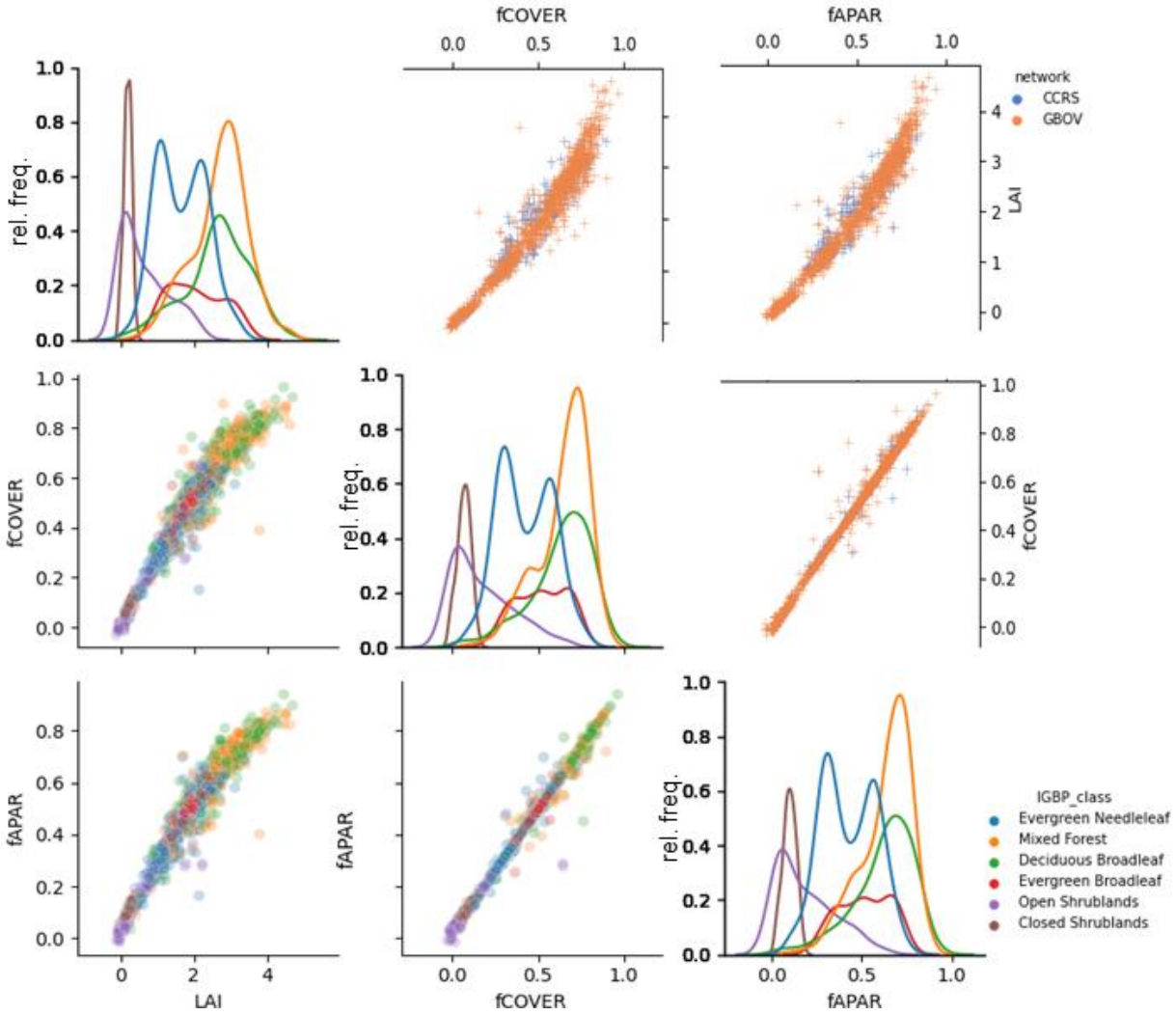


Figure 4. RM scatter plots by measurement network (upper diagonal figures) and IGBP land cover class (lower diagonal figures) together with relative frequency kernel density histograms for each variable by land cover (diagonal)

536 .



537
538 *Figure 5. Scatter plots of SL2P retrievals by measurement network (upper diagonal figures) and IGBP land cover class (lower*
539 *diagonal figures) together with relative frequency kernel density histograms for each variable by land cover (diagonal).*

540

541 4.3 Comparisons of RM and SL2P

542

543 4.3.1 Population Level Statistics

544

545 While not a GCOS requirement, it is useful to compare SL2P LAI to both RM LAIe and RM LAI since lower
546 error for LAIe versus LAI would supports the hypothesis that SL2P is not properly accounting for
547 clumping. Scatter plots of SL2P LAI versus RM LAIe (Figure 6a) and LAI (**Error! Reference source not**

found.b) indicate relatively linear relationships (r^2 =0.65 for LAIe and r^2 =0.64 for LAI). A U of 0.68 and UAR of 58% was observed for LAIe and 0.99 and 48%, respectively, for LAI. LAIe was consistently overestimated but with only a modest bias of 0.33. LAI was relatively unbiased for LAI<2 and underestimated for LAI>2 so the population level LAI bias of -0.38 was not representative of local performance. Both fCOVER and fAPAR exhibited relatively linear relationships (r^2≥0.7) but fAPAR had lower slightly lower A than fCOVER (-0.07 versus -0.02) that translated into a lower UAR (31% versus 37%) even though U for both was ~0.15.

Qualitatively, there was no evidence of systematic differences in residuals when comparing CCRS and GBOV networks for the same land cover and RM value (**Error! Reference source not found.**). Quantitative tests were not performed due to the imbalance of sample sizes between networks. In terms of land cover classes, for all variables the closed SH were typically estimated with low uncertainty (e.g. <0.5 LAI U, <0.05 fAPAR and fCOVER U) while open SH tended to be overestimated by between 0.5 LAI to 1 LAI and between 0.05 and 0.2 fAPAR and fCOVER) (**Error! Reference source not found.** and Table 7). Error metrics were similar across the forest classes with the exception of EBF where both fAPAR and LAI were unbiased (Table 7).

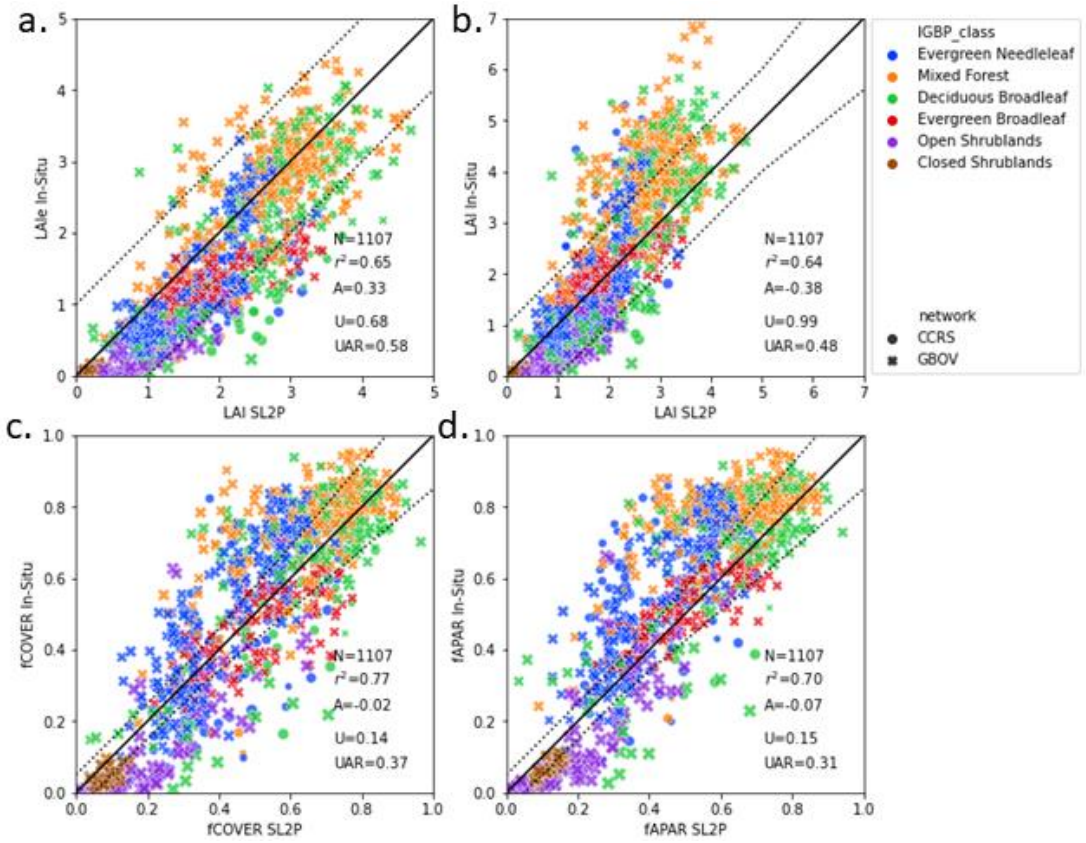


Figure 6. Scatter plots of ESU SL2P estimates versus matching RM for each variable together with population validation metrics. Symbol shape indicates network and symbol colour indicates IGBP class. Dashed lines bound target user requirement around solid 1:1 line.

4.3.2 Accuracy, Precision and Uncertainty as a Function of RM Value

SL2P LAI estimated RM LAIe with bias decreasing from 0.5 LAI at LAI~0 to -0.5 LAI at LAI~4 (**Error! Reference source not found.**a.). However, when compared to RM LAI, SL2P LAI was nearly unbiased (for LAI<2) but increasingly underestimated larger RM LAI; reaching an underestimate of -3 LAI units at LAI 6 (**Error! Reference source not found.**b). In contrast the precision of SL2P LAIe and LAI was approximately ~0.5 LAI units for all levels of LAI. As a result, uncertainty of SL2P LAI as an estimate of RM LAIe was within target requirements, assuming LAI could be replaced by LAIe, but SL2P LAI fell outside LAI target requirements for LAI>3.

SL2P estimated RM fCOVER and fAPAR with a bias of ~0.1 for the lowest RM values, to ~0 for mid-range RM values (~0.3 fCOVER and ~0.5fAPAR) and ~-0.15 for high (>0.8) RM values (**Error! Reference source not found.**c. and **Error! Reference source not found.**d). SL2P fCOVER and fAPAR precision was relatively

587 constant ranging from ~0.05 at extreme RM values to ~0.1 for mid range RM values. For both fCOVER
588 and fAPAR, the combination of relatively constant precision and changing accuracy resulted in greater
589 uncertainty at low and high RM values; although uncertainty was typically between 0.08 and 0.15 for all
590 retrievals.
591
592 The 95% confidence intervals of the A, P and U model fits were narrow for fAPAR and fCOVER (<±0.05)
593 and for LAI<5 (<0.2) but was wider for high (>3) LAIe and high (>5) LAI due to increased non-linearity and
594 decreased sample density. However, in these latter cases the magnitude of both A and U also increased
595 so the A, P and U model confidence interval widths were approximately constant relative to
596 corresponding regression line magnitudes.
597
598 Bias correction maintained or reduced accuracy error and increased UAR for LAI and fAPAR with the
599 exception of EBF LAI, that was unbiased prior to correction (Table 7). Bias correction also preserved or
600 increased UAR for land cover other than EBF; with substantial increases in UAR for fAPAR. The MF class
601 showed the largest improvement in accuracy and UAR due to bias correction. In contrast, bias
602 correction results in little change in error statistics for open shrubs. Bias correction equations without
603 hold-out were also produced (Supplementary Material Table S5).
604
605 For all variables, the relationship between clumping and A was weak and non-monotonic (Figure 8). The
606 weak relationship spanning our entire sample follows since, for all parameters, A was correlated to the
607 RM magnitude (**Error! Reference source not found.**) which in turn was not correlated to clumping but
608 rather the ground reference network (Figure 4). However, both P and U error decreased as clumping
609 index increased in a relatively linear manner, as evidenced by r^2<-0.4 in all but one case; reflecting the
610 increase in scatter of residuals as clumping index decreases.
611
612
613
614
615
616

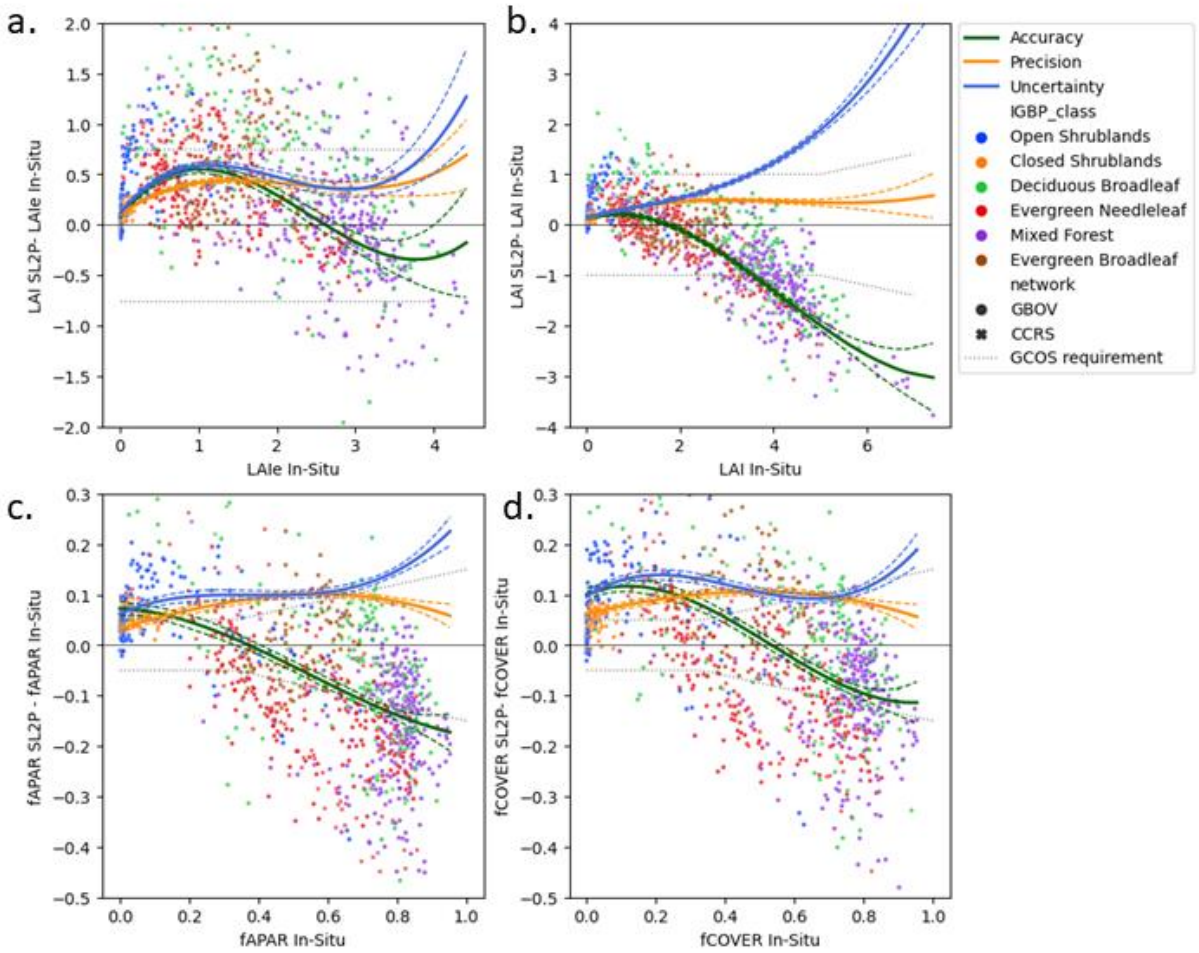


Figure 7. Residual plots of SL2P estimates versus matching RM as a function of RM value together with fitted models of Accuracy, Precision and Uncertainty (solid colours) and their 95%ile confidence intervals (dashed colours) Symbol shape indicates network, symbol colour indicates IGBP land cover and symbol size indicates clumping. Dashed lines bound target user requirement around solid 1:1 line.

Table 7. Thematic error statistics for validation of original and bias corrected SL2P retrievals with in-situ reference measurements for all data and by IGBP biome. For A,P,U the grey shading indicates absolute change less than the greater of 10% of mean estimate or 0.02 For UAR, the grey shading indicates a change less than 10% UAR. Green (gold) shading indicates improved (worsened) metric due to bias correction.

Variable	Biome	Number of comparisons	Mean estimate	Original				Bias Corrected			
				A	P	U	UAR	A	P	U	UAR
LAI	ALL	1107	2.31	-0.38	0.71	0.91	0.51	-0.03	0.74	0.82	0.54
	CSH	66	0.08	0.11	0.08	0.14	1.00	0.03	0.04	0.05	1.00
	DBF	203	3.05	-0.42	1.07	1.15	0.36	0.16	1.06	1.08	0.40
	EBF	104	2.04	0.01	0.39	0.39	0.85	0.42	0.72	0.84	0.46
	ENF	297	1.86	-0.20	0.76	0.78	0.63	0.02	0.71	0.71	0.58
	MF	297	3.82	-1.12	0.82	1.39	0.20	-0.56	0.84	1.01	0.48
	OSH	140	0.27	0.37	0.39	0.54	0.68	0.38	0.49	0.62	0.65
fAPAR	ALL	1107	0.55	-0.07	0.11	0.15	0.31	-0.03	0.11	0.13	0.43
	CSH	66	0.04	0.06	0.02	0.06	0.32	0.04	0.02	0.05	0.64
	DBF	203	0.67	-0.05	0.14	0.15	0.34	0.00	0.14	0.14	0.49
	EBF	104	0.52	0.00	0.07	0.07	0.55	0.08	0.08	0.12	0.31
	ENF	297	0.55	-0.12	0.13	0.18	0.21	-0.07	0.13	0.15	0.34
	MF	297	0.78	-0.15	0.11	0.18	0.26	-0.10	0.10	0.14	0.47
	OSH	140	0.14	0.05	0.09	0.10	0.42	0.06	0.11	0.12	0.44
fCOVER	ALL	1107	0.50	-0.02	0.12	0.13	0.37	-0.03	0.12	0.14	0.42
	CSH	66	0.03	0.04	0.03	0.05	0.58	-0.01	0.02	0.02	0.95
	DBF	203	0.63	0.00	0.15	0.15	0.39	0.00	0.15	0.15	0.45
	EBF	104	0.45	0.07	0.10	0.12	0.38	0.08	0.13	0.15	0.25
	ENF	297	0.47	-0.05	0.13	0.14	0.25	-0.07	0.14	0.15	0.26
	MF	297	0.73	-0.09	0.12	0.15	0.42	-0.09	0.12	0.15	0.43
	OSH	140	0.11	0.07	0.10	0.12	0.41	0.05	0.10	0.12	0.58

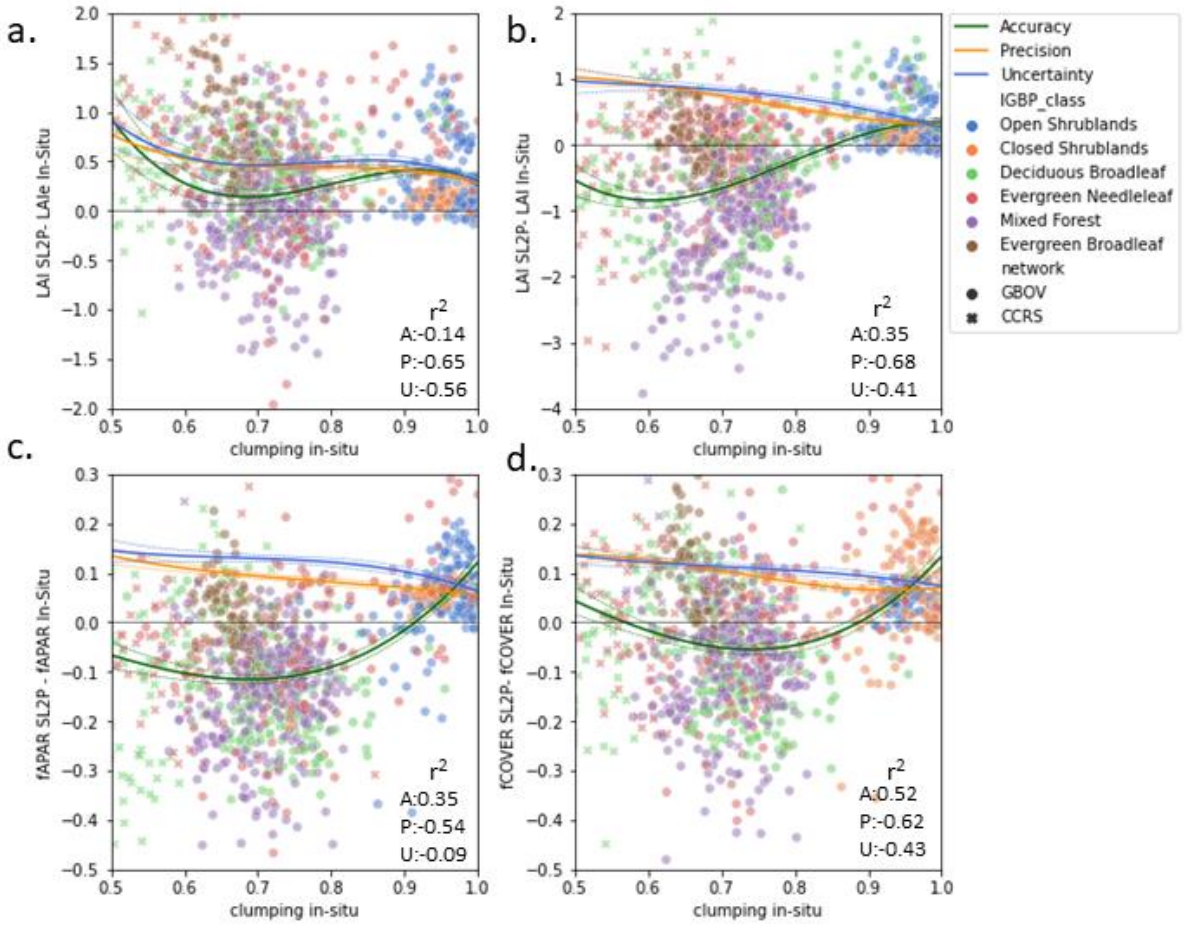


Figure 8. Residual plots of SL2P estimates versus matching RM as a function of clumping together with fitted models of Accuracy, Precision and Uncertainty (solid colours) and their 95%ile confidence intervals (dashed colours). Symbol shape indicates network, symbol colour indicates IGBP land cover class. Pearson correlation coefficient (significant at p<=0.002.) between modelled A, P and U and in-situ clumping index indicated in each panel.

4.3.3 Validation of SL2P Theoretical Precision

For all variables, SL2P theoretical precision was poorly related to the P estimated from observed residuals (Figure 9). For LAIe and LAI, SL2P generally overestimates P with an almost uniform distribution below the 1:1 line. In contrast, SL2P theoretical precision was almost constant at ~0.04 for

fAPAR and fCOVER over forests and uncorrelated (r^2<0.1) with the modelled precision. It was only for fCOVER over shrublands that SL2P theoretical precision was relatively unbiased and within ±0.05 of P.

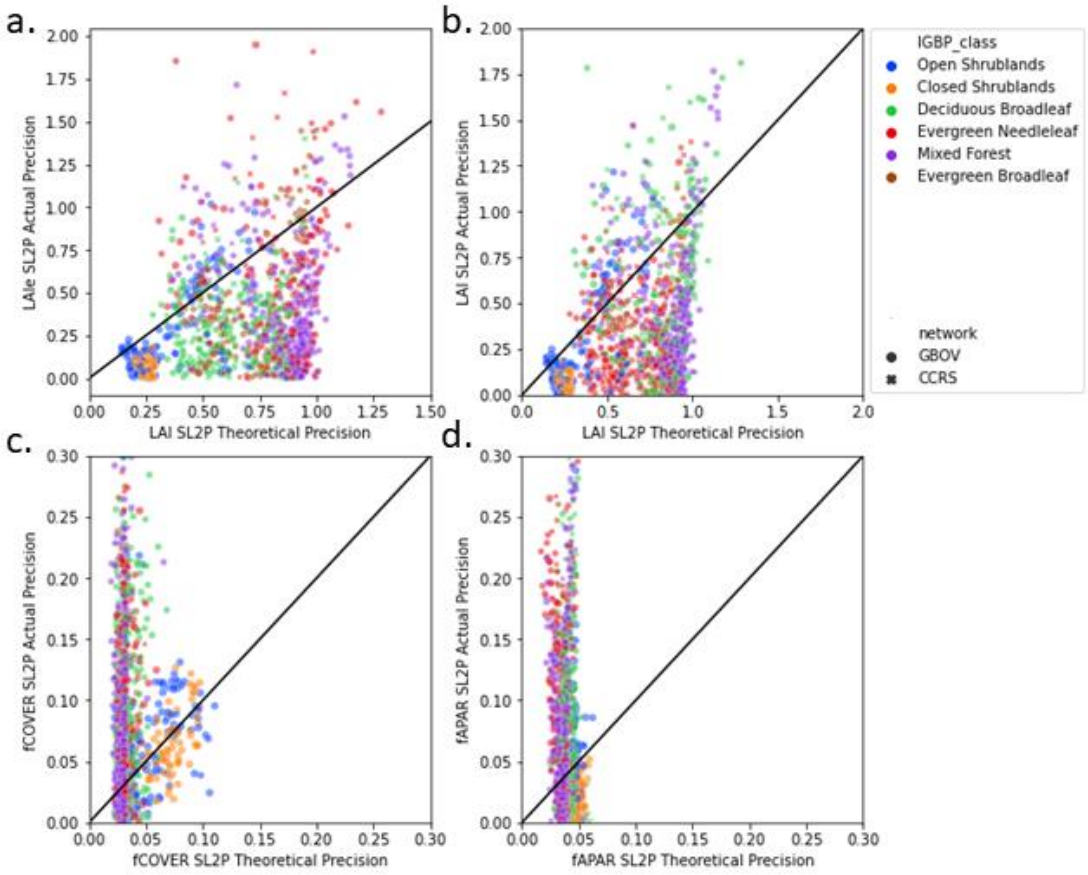


Figure 9. Scatter plots of SL2P modelled precision based on comparisons to RM versus theoretical precision based on the retrieval algorithm for in-situ sites. 1:1 line indicated in black

4.4 Intercomparison

The number of intercomparisons for a given aggregated SL2P product estimate value ranged from 500 to 10000 for CGLS versus SL2P and from 1000 to 15000 for MODIS versus SL2P (Supplementary Material Figure S2). The larger number of MODIS comparisons was due to the higher temporal revisit frequency of MODIS on Aqua and Terra in comparison to the OLCI imagers on Sentinel 3A and 3B. Sub-pixel variability in LAI was small for both MODIS and CGLS pixels used for intercomparison; with a coefficient of variation <0.2 for LAI>1 (Figure 10a. and Figure 10c.). Sub-pixel biome type variation was also low between shrubs and forest, with shrub dominated pixels containing at least 80% shrub cover (Figure

10b. and Figure 10d.). The majority of forested pixels had at least 75% cover of either broadleaf or needleleaf forests but there was also a lesser number of mixed forest pixels.

For brevity, we summarise intercomparison results across all ecozones (Figure 11) since between ecozone variations were correlated to the magnitude of aggregated SL2P estimates (not shown) . MODIS and CGLS comparisons to SL2P were remarkably similar: SL2P underestimates both by ~50% LAI for SL2P LAI between 0.5 and 3.5 and by ~20% for SL2P fAPAR between 02 and 0.8. However, MODIS and CGLS LAI and fAPAR saturated with respect to SL2P LAI for SL2P LAI>4. This agrees with results indicating MODIS and CGLS products saturate when in-situ LAI >4 (Brown et al., 2020). For low (<0.5) LAI and (<0.25) fAPAR, both MODIS and CGLS showed good agreement (within ±0.25 LAI) with SL2P but were still overestimating SL2P fAPAR by between 0.05 and 0.1 units. Results for CGLS and SL2P fAPAR comparisons were similar to those for fCOVER as expected given the similarity of SL2P fCOVER and fAPAR (Figure 5) and the fact that CGLS fCOVER is a deterministic function of CGLS fAPAR.

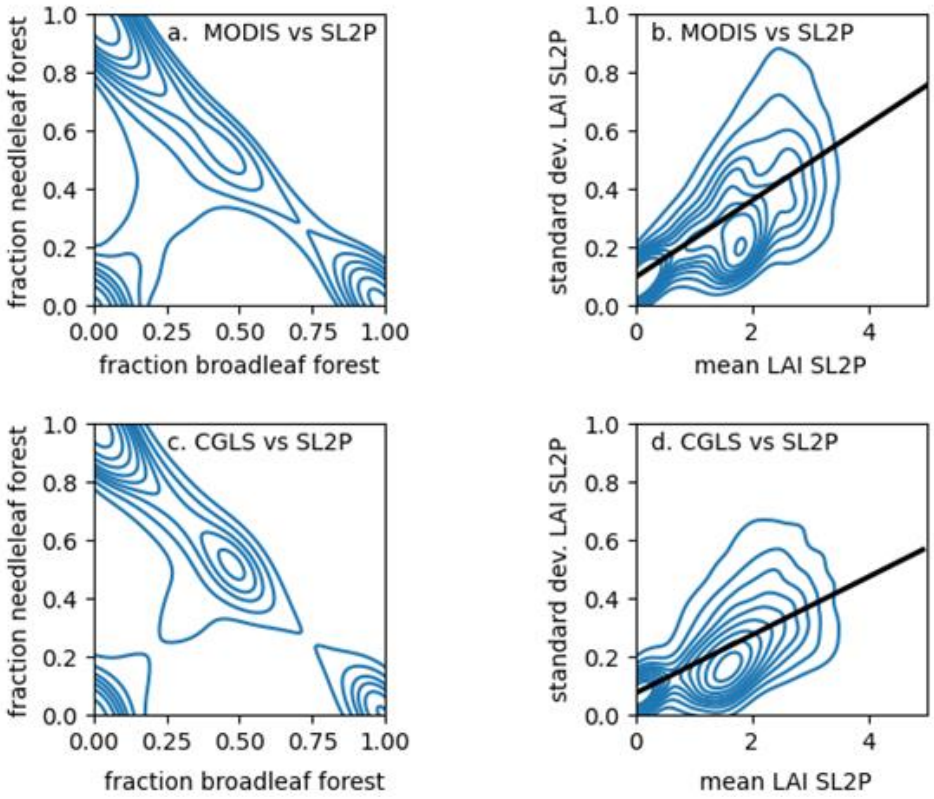


Figure 10. Density contour plots of sub-pixel fraction forest biome type for intercompared (a) MODIS (a) and (b) CGLS products and sub-pixel LAI standard deviation versus LAI for intercompared (a) MODIS (c) and (d) CGLS products together with linear regression fits (black lines). Contour intervals range from quantiles of 0.1 to 1.0 in steps of 0.1.

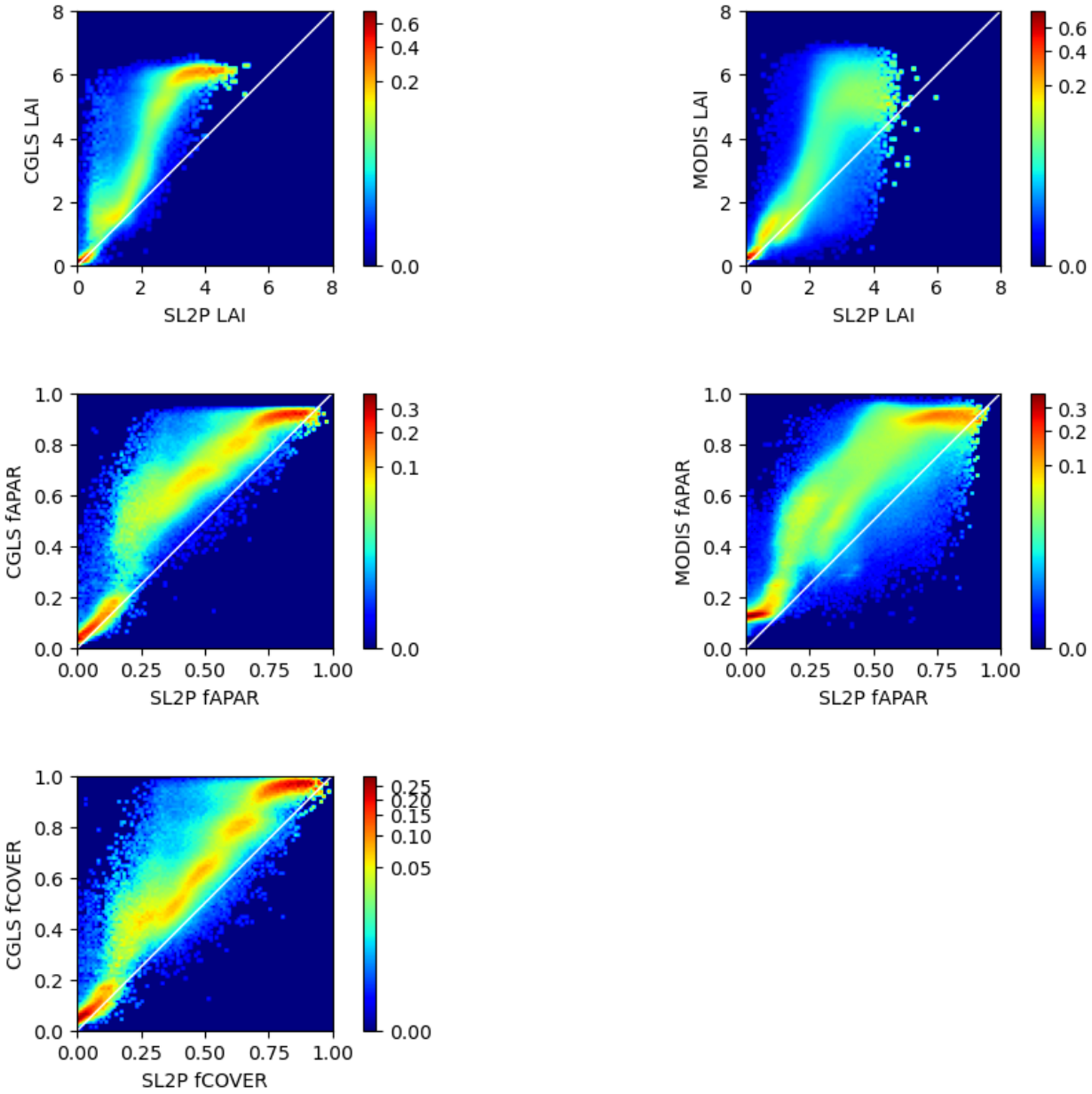


Figure 11. Density plots of the conditional probability of CGLS (left column) or MODIS (right column) comparisons given a matching aggregated SL2P estimate for 1.5km x 1.5km grid cells. Colours correspond to conditional probability of reference product variable given the aggregated SL2P match-up value.

5.Discussion

This study focussed on SL2P both because it is being used for science and applications and because it serves as a free and open baseline for benchmarking new approaches. We further considered forests because previous studies have shown SL2P typically meets user requirements over croplands but not over forests (Supplementary Material Table S1). Brown et al. (2021a) showed that SL2P underestimated forest LAI even with uniform calibration priors; suggesting a systematic limitation in either L2A inputs or PROSAILH. We hypothesized the lack of clumping within the PROSAILH was the cause of the bias as empirical algorithms using similar input imagery are unbiased for LAI over North American forests (Fernandes et al., 2003). We also wanted to determine if the bias seen for LAI also occurred for fAPAR and fCOVER, and if the bias could be corrected using empirical models of bias as a function of SL2P retrievals. To address these questions, we expanded the spatially limited sample used in Brown et al. (2021a) to include 133 new northern latitude forest ESUs and refined the validation approach of Brown et al. (2021a) to quantify thematic error as a function of RM and a function of clumping.

This study began with a code verification that included a review of the SNAP and LEAF Toolbox implementations of SL2P by a software engineering team at CCRS not involved in authoring either implementations. Two bugs were found in the MATLAB code provided to the SNAP developers: i. incorrect truncation of priors used to specify radiative transfer model simulations and ii. incorrect logic when checking if inputs fell within the domain of the calibrated regression predictors and associated quality control masks. These bugs result in systematic errors in high and low product estimates and reduced the consistency and replicability of products and validation studies (Figure 3). Code verification is not currently part of CEOS good practices but should be applied prior to product validation.

The RM dataset sampled all North American forest ecozones except tropical forest and steppe ecozones (Figure 1,Table 5). The tropical forest ecozones are of global importance and should be incorporated in future validation but are of limited extent over North America. The RM samples covered the typical range of LAI (0 to 7.5) (Scurlock et al., 2001; Fernandes et al., 2003), except for Douglas Fir forests on the Pacific coast that approach values of 10. On the other hand, this is the only study that validates SL2P over forests using a representative range of fAPAR and fCOVER. Simultaneous validation of all three variables is important since they are closely related in-situ (Figure 4) and for SL2P (Figure 5), with r^2>0.9

738 between variables. Indeed, the close SL2P relationships suggest that there may be a benefit to using
739 variables that are estimated with low bias (e.g. fCOVER) to constrain other variables such as LAI.
740
741 The CCRS network sites provided increased sampling of lower clumping index and a broader range of
742 ecozones that used in Brown et al. (2021). This, together with intercomparison, allowed us to achieve a
743 CEOS Stage 3 validation. The CCRS sites provided broad spatial coverage with a single date sampling
744 while the subset of NEON sites selected for this study provided the obverse. Current in-situ survey
745 networks do not provide dense sampling in space and time, and it seems that most networks prefer the
746 latter since they are typically related to long term measurement sites. However, we suggest that a
747 broad spatial sample is desirable, as it will eventually include temporal variability if sampling dates are
748 allowed to vary, while a fixed network will never improve spatial coverage.
749
750 The combination of CCRS and GBOV RM was possible due to their similar acquisition and processing
751 protocols and a shared theoretical basis for estimating RM from gap fraction. Validation of the RM
752 methods is beyond the scope of this study but consistency between networks was supported by their
753 similar bi-variate distributions (Figure 4). Ad hoc intercomparisons over 6 plots found random
754 differences less than 5% for GBOV versus CCRS data processing (not shown) suggesting that the
755 uncertainty between networks is far less than the total uncertainty of each RM (e.g. Supplementary
756 Material Figure S1).
757
758 Both the NEON and CCRS ESUs were situated in the centre of large (≥100m x 100m) patches (Figure 2)
759 that were qualitatively assessed as spatially homogenous both to minimize error in DHPs whose FOV
760 extend past the nominal ESU footprint and to reduce the impact of geolocation errors on match-ups.
761 This limited our ability to quantify SL2P error in heterogenous landscapes, including disturbed and
762 developed areas where forested pixels may be adjacent to non-forested areas. It is likely our error
763 estimates are optimistic for these circumstances since SL2P does not consider heterogenous mixed
764 pixels (clumping≠1) since the RTM used for SL2P calibration does not account for spatial heterogeneity
765 of vegetation. Even pure forested pixels may have increased error if they are adjacent to different
766 vegetation since the input L2A BRF to SL2P does not correct for lateral fluxes between pixels below the
767 top of the canopy. Simulations indicates that one requires ~100m pixel resolution to ensure the ratio of
768 later scattering to BRF is below 0.1 in the worst case where a forest is surrounded by absorbing
769 boundaries (Widlowsk et al., 2008). In reality, forests usually only have one pixel edge adjacent to a low

vegetated pixel so a ~25m pixel size should correspond to the same error. Nevertheless, this error is not accounted for either in the input L2A products or the SL2P algorithm. Ideally, new technologies to permit automated regional estimates of RM using airborne surveys should be developed; especially those that include measurement approaches such as LIDAR or structure from motion that can map three dimensional canopy structure over heterogenous landscapes (Fang et al., 2019).

LAI validation indicated similar precision and bias as reported in Brown et al. (2021a) but here included forests up to LAI 7.5. In contrast to LAI, we noticed a curious almost sinusoidal pattern to LAIe, fAPAR and fCOVER bias that shifted from positive to negative with increasing RM value. This result may explain why previous studies found SL2P to be negatively biased for high fAPAR forests (Putzenlechner et al., 2019) but relatively unbiased for moderate fAPAR over crops (Djamai et al., 2019). Our analysis also indicates that statistics based on the entire population of RM can be misleading for bias and uncertainty and should be accompanied with breakdowns into sub-populations of land cover and reference value in future validation studies. All variables were estimated with relatively consistent precision that, in the absence of bias, would result in uncertainty falling within user requirements for most levels of each variable. This is encouraging since it indicates that the propagation of measurement error from inputs to product is modest. In fact, precision errors may even decrease further with temporal or spatial smoothing although this was beyond the scope of our study. The main challenge, especially for LAI, remains reducing bias.

Our novel approach of using regression to model the expected value of A,P and U as a function of reference value differs from previous good practice of using binned or population statistics (Fernandes et al., 2014). Regression naturally provides an unbiased estimator with prediction confidence intervals that can be used to identify where further RM sampling is required. However, it is good practice to examine patterns in regression residuals due to factors such as clumping or biome type. While we did not observe effects due to clumping it is clear that SL2P performance varies with biome. Specifically, the uncorrected SL2P resulted is almost no bias for EBF. EBF correspond to relatively high fCOVER that may satisfy the turbid medium assumptions within the SL2P radiative transfer model better than other validated biomes. Ideally, our experiment should be replicated using large samples over each biome to verify the observed structure in residuals.

Bias correction using empirical regression models generally improved LAI and fAPAR accuracy and UAR over forests other than EBF (Table 7). The observed improvements are conservative since i) we completely held out each site when fitting and applying the bias correction and, ii) our RM sample were not optimized for bias correction in that they did uniformly sample the range of possible SL2P retrievals for a given variable.

Our attempt to test the hypothesis that SL2P LAI bias was due to clumping had mixed results. There was weak to moderate quantitative evidence (r^2<0.52) that errors in fAPAR, fCOVER or LAI were related to clumping (Figure 8). The covariation of clumping with land cover and LAI (e.g. shrub clumping close to 1 in Figure 4) meant that we could not separate these two effects from clumping. Indeed, it may be impossible to control for these effects when using undisturbed sites and may require real or numerical canopy manipulation experiments (e.g. Stenberg et al., 2014; Widlowski et al., 2015). However, there was qualitative evidence that, for forests other than EBF, LAI bias increases as the clumping index decreases. For example, for similar RM LAI underestimation was worse for the CCRS network versus the GBOV network (Figure 7). Furthermore, the relatively unbiased estimation of LAIe (Figure 7) suggests that SL2P LAI is actually mapping the LAI for a spatially homogenous canopy. This may be expected given that SL2P is calibrated with a homogenous radiative transfer model. An alternative test of the hypothesis that clumping causes SL2P LAI bias would be to determine if replacing PROSAILH in SL2P with an accurate heterogeneous radiative transfer model, that consider the realistic three dimensional canopies, reduces LAI bias while increasing LAIe bias.

Estimates of theoretical precision provided by SL2P are not sufficiently accurate for fAPAR and fCOVER or precise for LAIe and LAI to be of practical use (Figure 9). Indeed, it is contradictory to assume that precision could be predicted for each retrieval with low uncertainty since if that were the case one could use this prediction to improve retrievals. Nevertheless, one could expect an unbiased estimate of precision if cross validation was sufficiently robust during SL2P calibration. The current cross validation approach is to test retrievals for PROSAILH simulations drawn from the same priors used for calibration. At a minimum, precision should be modelled using different priors and ideally using an ensemble of model simulations. One alternative may be to use perturbations of paired RM and satellite measurements, based on the associated uncertainties, to quantify precision using the validation approach presented here.

Intercomparison of medium and coarse resolution products needs to control for differences due to the spatial scale of their respective inputs. Coarse resolution reference products will typically have additional uncertainty due to sub-pixel variation in land cover or vegetation density (Yan et al., 2016; Dong et al., 2022). Our intercomparison involved objective criteria to sample MGRS tiles and to reject samples without sufficient forest cover in each coarse resolution product pixel. Both of these criteria resulted in low sub-pixel variability for sampled MCD15 and CGLS pixels (Figure 10). Uncertainty in intercomparisons due to sub-pixel variability was further reduced during intercomparison due to the use of 1500m x 1500m averages of reference and SL2P products. Hence, the error in the aggregated coarse resolution reference products should be approximately their reported accuracy error of <0.5 LAI and <0.1 fAPAR (Supplementary Material Tables S2 and S3). This indicates that the SL2P underestimation of both MODIS LAI (Figure 11a) and CGLS LAI (Figure 11b.) cannot be attributed to uncertainty in the reference data. The SL2P LAI underestimation also agrees with observations over in-situ sites. SL2P also consistently underestimates MODIS fAPAR (Figure 11c.) and CGLS fAPAR (Figure 11d.) with magnitude similar to underestimates observed when compared to RM data (Table 7).

SL2P fCOVER comparisons to CGLS indicate a similar pattern as seen with fAPAR. However, we are hesitant to draw further insights for comparisons with for two reasons: i) CGLS retrievals of LAI and fAPAR are either heavily weighted to the MCD15 values for LAI>2 and essentially based on the SL2P algorithm for LAI<1 since SL2P shares a similar calibration database and inversion approach as the GEOV1 products used by CGLS and, ii) CGLS fCOVER is a functional transformation of CGLS fAPAR so fCOVER intercomparisons provide no new insights compared to fAPAR intercomparisons.

The intercomparison was limited to sample regions due to our initial concern regarding computation demands of having to generate SL2P products over larger areas. Ongoing work with GEE is generating Canada wide monthly SL2P products that could be extended globally with modest costs. This suggests that future intercomparison could be done using exhaustive sampling. Additional sampling may be useful to understand product differences at high LAI where precision is low and for specific land cover conditions such as regeneration or disturbance. Areas with poor agreement could also be useful to focus new in-situ measurements. An objective approach is required to design a sampling design for these cases.

This study built on the large number of previous coarse resolution product validation studies as well as recent medium resolution validation studies (Supplementary Materials Tables S1, S2 and S3) by

developing new good practices for survey, code verification, validation, and intercomparison. However, there remain three elephants in the room when it comes to the practice of land parameter validation:

i. Agencies invest significant funding into observing systems, algorithm development, and services but limited amounts to the acquisition of reference measurements. Even GBOV, the Copernicus validation service, does not itself operate systematic survey systems other than for benchmarking. Satellite validation has relied primarily on long term networks. These may be sufficient for publishing a new algorithm but do not have enough spatial coverage to meet CEOS requirements. CEOS and designers of new systems and services need to require proof that quality can be assured within the design of new satellite observing systems.

ii. There is no accepted approach to track the quality assurance of algorithms and mapping systems. ISO and IEEE standards are widely used for analogous systems across other industries but we have only seen such standards for radiative transfer models in terms of land surface mapping with satellite earth observation (Widlowski et al., 2015). The lack of standards can influence downstream services. For example, the SNAP implementation of SL2P has yet to be updated to address the bugs identified in our study. CEOS is in a position to require algorithm developers to adopt standards and to publish the level of standards met by various systems. This is especially important as future systems may not be in the public domain.

iii. Both product generation and validation rely on computer code that is often not public. This prevents replication of results in an efficient and accurate manner and understanding if product differences are due to bugs in code. We were fortunate to have access to the original SL2P code. The MCD15 product had six revisions that vastly improved efficiency and performance. However, the MCD15 algorithm and its associated RT models are not published in a free and open manner so future science may have limited access to the progress of this work. We feel it is critical for the scientific community and funding agencies to embrace the free and open publication of code (our code is pubished at Fernandes et al. , 2021)

6. Conclusions

This study validated SL2P for mapping LAI, fAPAR and fCOVER over North American forests using Sentinel-2 MSI data. It enhanced previous studies by increasing the spatial coverage of in-situ reference measurements so they represented most forest ecozones, by quantifying thematic error metrics for all three variables simultaneously and by conducting product intercomparison. These steps and the publication of this work meet necessary conditions for CEOS Stage 3 validation. The study had four goals: report on thematic error of SL2P, determine if biases are due to clumping, evaluate empirical bias correction, and provide new good practices.

Code verification of SL2P identified bugs in the SNAP implementation that affect both product retrieval and quality control flags. The SNAP implementation should not be used until revision. The verified SL2P implementation within the LEAF-Toolbox underestimated in-situ LAI by between 20% and 50% and MODIS LAI by ~50% for LAI>2. Compared to in-situ measurements, SL2P fAPAR and fCOVER bias trended from approximately +0.1 to -0.1 as both variables increased. The fAPAR underestimation was also observed when comparing to MODIS for fAPAR>0.5. Precision was relatively constant at <0.1 units for fAPAR and fCOVER and <0.5 units for LAI. SL2P satisfied target uncertainty requirements for 48% of LAI, 37% of fCOVER and 31% of fAPAR comparisons to in-situ data. The low level of agreement reflects both biases in products and stringent requirements for fCOVER and fAPAR. Reducing product bias is fundamental to reducing thematic error.

Clumping showed only weak to moderate linear relationships ($|r^2|<0.5$) to bias for all variables and even these were at times counterintuitive, with bias increasing as canopies were less clumped. Our sample was not able to control for the covariation of clumping and the SL2P estimates to quantify the impact of clumping on SL2P bias. Further studies are required in disturbed landscapes or with simulations. Nevertheless, as in Brown et al. (2021a), the fact that SL2P estimates LAIe with little or no bias but underestimates LAI suggests the absence of clumping in PROSAILH is to blame for LAI bias. Scatterplots of residuals also indicated the CCRS network sites tended to show greater negative LAI bias than the less clumped GBOV network sites at comparable LAI.

Empirical bias correction preserved or improved SL2P validation metrics for all land cover classes except evergreen broadleaf forests. Ideally, SL2P should be updated to include calibration using simulations from heterogenous radiative transfer model simulations when mapping forests. Until that time, with the exception of EBF, empirical bias correction may be useful for SL2P LAI and fAPAR over forests similar to those in our in-situ sample.

This study promoted new good practices for validation of canopy biophysical variables that may benefit future studies. We acquired a representative sample or RM from networks that followed consistent acquisition and processing standards and that included uncertainty propagation. We used regression to model thematic error metrics as a function of the mapped value. We applied strict sampling criteria to minimize sub-pixel heterogeneity when using coarse resolution products. Finally, we provided all code, including SL2P and the validation methods and data, in a free and open manner.

We conclude by noting that the publishing of new algorithms and products is becoming both easier and more frequent with on-line journals and free and open computing platforms. These algorithms and products represent complex hypotheses about the physics and structure of our environment. The value of these hypotheses should be measured in our ability to test and potentially defeat them. Validation data, methods and the human activity of publishing the validation results is fundamental for users to understand the limitations of algorithms and products and for developers to make improvements.

7.Acknowledgements

Dr. Fernandes led the CCRS field campaigns, performed data analysis and drafted the manuscript. We acknowledge Dr. Wenjun Chen, P. Osei Darko, Dr. Robert Fraser, Lucy Huang, Nhu Quynh Le, Julie Lovitt, Lynsay Spafford, Dr. H. Peter White and Dr. Yu Zhang for data collection; Dr. Fred Baret and Dr. Marie Weiss for providing the original SL2P code; Hemit Shah and Camryn MacDougall for software verification and the rest of the co-authors for assisting in revising the manuscript. We acknowledge the use of modified Sentinel 2 and Sentinel 3 data and derived products. This study has been undertaken using data from GBOV "Ground Based Observation for Validation" (https://land.copernicus.eu/global/gbov) funded by European Commission Joint Research Centre FWC932059, part of the Global Component of the European Union's Copernicus Land Monitoring Service. GBOV products are developed and managed by ACRI-ST with the support from University College London, University of Leicester, University of Southampton, University of Valencia and Informus GmbH. We thank the NEON network for the measurements collected in the field and used to generate GBOV products. We thank four excellent reviewers and the editors of this paper. The work was funded by Natural Resources Canada's Earth Observation for Cumulative Effects Project.

8.References

Baret,F., Hagolle, O., Geiger,B., Bicheron, P., Miras,B., Huc, M. , Berthelot,B. , Niño,F. , WeissM., Samain, O., Roujean,J.L. , Leroy, M. , 2007. LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION: part 1: principles of the algorithm. Remote Sens. Environ., 110, pp. 275-286

Baret , F., Weiss, M., Verger, A., and Smets, B., 2016. ATBD FOR LAI, FAPAR AND FCOVER FROM PROBA-V PRODUCTS AT 300M RESOLUTION (GEOV3), ImagineS, FP7-Space-2012-1, I1.73, https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/ImagineS_RP2.1_ATBD-LAI300m_I1.73.pdf

Barnett, D. T., Duffy, P. A., Schimel, D. S., Krauss, R. E., Irvine, K. M., Davis, F. W., Gross, J. E., Azuaje, E. I., Thorpe, A. S., Gudex-Cross, D., Patterson, M., McKay, J. M., McCorkel, J. T., and Meier, C. L., 2019. The terrestrial organism and biogeochemistry spatial sampling design for the National Ecological Observatory Network. Ecosphere, 10:e02540. 10.1002/ecs2.2540.

Bourg, L., Bruniquel, J., Morris, H., Dash, J., Preusker, R., Dransfeld, S., 2021. Copernicus Sentinel-3 OLCI Land User Handbook. European Space Agency, version 1, S3MPC.ACR.HBK.001.

Brown, L.A., Meier, C., Morris, H., Pastor-Guzman, J., Bai, G., Lerebourg, C., Gobron, N., Lanconelli, C., Clerici, M., Dash, J., 2020. Evaluation of global leaf area index and fraction of absorbed photosynthetically active radiation products over North America using Copernicus Ground Based Observations for Validation data, Remote Sensing of Environment, Volume 247, 111935, ISSN 0034-4257, https://doi.org/10.1016/j.rse.2020.111935.

Brown, L., Fernandes, R., Djamai, N., Meier, C., Gobron, N., Morris, H., Canisius, C., Bai, G., Lerebourg, C., Lanconelli, C., Clerici, M., Dash, J., 2021a. Validation of baseline and modified Sentinel-2 Level 2 Prototype Processor leaf area index retrievals over the United States, IISPRS Journal of Photogrammetry and Remote Sensing, 175, 71-87, https://doi.org/10.1016/j.isprsjprs.2021.02.020.

Brown, L., Camacho, F., García-Santos, V., Origo, N., Fuster, B., Morris, H., Pastor-Guzman, J., Sánchez-Zapero, J., Morrone, R., Ryder, J., Nightingale, J., Boccia, V., Dash, J., 2021b. Fiducial Reference Measurements for Vegetation Bio-Geophysical Variables: An End-to-End Uncertainty Evaluation Framework. Remote Sensing, 13, 3194. https://www.mdpi.com/2072-4292/13/16/3194.

Camacho, F., Cernicharo, J., Lacaze, R., Baret, F., and Weiss, M, 2013. GEOV1: LAI, FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part 2: Validation and intercomparison with reference products. Remote Sensing of Environment, 137, 310-329.

Chen, J.M., Cihlar, J., 1995. Plant canopy gap-size analysis theory for improving optical measurements of leaf-area index. Applied Optics, 34, 6211–6222.

Commission for Environmental Cooperation, 2020. 2015 Land Cover of North America at 30 Meters, North American Land Change Monitoring System, Ed. 2.0. Accessed at http://www.cec.org/north-american-environmental-atlas/land-cover-30m-2015-landsat-and-rapideye/ on March 15, 2023.

Commission for Environmental Cooperation, 2022. CEC Terrestrial Ecological Regions of North America, Montreal, Commission for Environmental Cooperation, accessed at http://www.cec.org/north-american-environmental-atlas/north-american-forests-2022/ on March 15, 2023.

Defence Mapping Agency, 1990. DEFENSE MAPPING AGENCY TECHNICAL MANUAL 8358.1 DATUMS, ELLIPSOIDS, GRIDS, AND GRID REFERENCE SYSTEMS. DM ™ 8358.1, Defence Mapping Agency, U.S.A., http://everyspec.com/DoD/DOD-General/download.php?spec=DMA_TM-8358.1.006300.PDF.

Djamai, N and Fernandes, R., 2018. Comparison of SNAP-Derived Sentinel-2A L2A Product to ESA Product over Europe. Remote Sensing. 10:926. https://doi.org/10.3390/rs10060926.

Djamai, N., Fernandes, R., Weiss, M., McNairn, H., Goïta, K., 2019. 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., 225, 416-430, 10.1016/j.rse.2019.03.020.

Doxani, G., Vermote, E.,Roger, J.-C., Gascon, F., Adriaensen, S., Frantz, D., Hagolle, O., Hollstein, A., Kirches, G, ; Li, F., Louis, J., Mangin, A., Pahlevan, N., Pflug, B., and Vanhellemont, Q., 2018. Atmospheric Correction Inter-Comparison Exercise. Remote Sens., 10, 352.

ESA Sentinel-2 Team, 2007. GMES Sentinel-2 Mission Requirements Document. EOP-SM/1163/MR-dr. European Space Agency. accessed at https://earth.esa.int/pub/ESA_DOC/GMES_Sentinel2_MRD_issue_2.0_update.pdf on January 20, 2021.

Fang, H., Baret, F., Plummer, S., and Schaepman-Strub, G., 2019. An overview of global leaf area index (LAI): Methods, products, validation, and applications. Reviews of Geophysics. 57, 739– 799. https://doi.org/10.1029/2018RG000608.

Fernandes, R., Plummer, S., Nightingale, J., Baret, F., Camacho, F., Fang, H., Garrigues, S., Gobron, N., Lang, M., Lacaze, R., Leblanc, S., Meroni, M., Martinez, B., Nilson, T., Pinty, B., Pisek, J., Sonnentag, O., Verger, A., Welles, J., Weiss, M., Widlowski, J.-L., Schaepman-Strub, G., Roman, M., and Nickeson, J., 2014. Global Leaf Area Index Product Validation Good Practices, in: Fernandes, R., Plummer, S., Nightingale, J. (Eds.), Best Practice for Satellite-Derived Land Product Validation. Land Product Validation Subgroup (Committee on Earth Observation Satellites Working Group on Calibration and Validation), https://doi.org/10.5067/doc/ceoswgcv/lpv/lai.002.

Fernandes, R. et al., 2021. LEAF Toolbox, Canada Centre for Remote Sensing, accessed at https://github.com/rfernand387/LEAF-Toolbox/wiki on March 15, 2023, DOI: 10.5281/zenodo.4321298.

Fernandes, R., Canisius, F. et al., 2022. In-Situ Leaf Area Index, fAPAR and Canopy Cover Measurements over Canadian Forests in Support of Cumulative Effects Assessments, Geomatcis Canada Open File 64, accessed at https://publications.gc.ca/collections/collection_2022/rncan-nrcan/m103-3/M103-3-64-2023-eng.pdf.

Fernandes, R., Butson, C., Leblanc, S. and Latifovic, R., 2003.Landsat-5 TM and Landsat-7 ETM+ based accuracy assessment of leaf area index products for Canada derived from SPOT-4 VEGETATION data, Canadian Journal of Remote Sensing, 29:2, 241-258, DOI: 10.5589/m02-092.

Fuster B, Sánchez-Zapero J, Camacho F, García-Santos V, Verger A, Lacaze R, Weiss M, Baret F, Smets B, 2020. Quality Assessment of PROBA-V LAI, fAPAR and fCOVER Collection 300 m Products of Copernicus Global Land Service. Remote Sensing. 12, 1017. https://doi.org/10.3390/rs12061017.

Garrigues, S., Lacaze, R., Baret, F. J. T. M., Morisette, J. T., Weiss, M., Nickeson, J. E., Fernandes, R., Plummer, S., Shabanov, N.V., Myneni, R., Knyazhikin ,Y., and Yang, W., 2008. Validation and intercomparison of global Leaf Area Index products derived from remote sensing data. Journal of Geophysical Research: Biogeosciences, 113(G2).

Gascon F, Bouzinac C, Thépaut O, Jung M, Francesconi B, Louis J, Lonjou V, Lafrance B, Massera S, Gaudel-Vacaresse A, Languille F, Alhammoud B, Viallefont F, Pflug B, Bieniarz J, Clerc S, Pessiot L, Trémas T, Cadau E, De Bonis R, Isola C, Martimort P, Fernandez V., 2017. Copernicus Sentinel-2A Calibration and Products Validation Status. Remote Sensing. 2, 9, :584. https://doi.org/10.3390/rs9060584.

GCOS, 2022.The 2022 GCOS Implementation Plan. GCOS No. 244. World Meteorological Organization, © 2022.

He, Li. , Chen, Ji., Pisek, J., Schaaf, C.B, and Strahler, A., 2012. Global clumping index map derived from the MODIS BRDF product. Remote Sensing of Environment. 119. 118-130. 10.1109/IGARSS.2011.6049427.

Hu, Q., Yang, Y., Xu, B., Huang, J., Memon, M.S., Yin, G., Zeng, Y., Zhao, J., Liu, K., 2020. Evaluation of global decametric-resolution LAI, FAPAR and FVC estimates derived from Sentinel-2 imagery. Remote Sensing, 12, 912, 10.3390/rs12060912.

Janzen, D., et al., 2020. EO Baseline Data for Cumulative Effects Year End Report (FY 2019/20), Geomatics Canada Open File 60, accessed at https://publications.gc.ca/collections/collection_2020/rncan-nrcan/m103-3/M103-3-60-2020-eng.pdf on March 15, 2023.

Lambin, E.F. and Geist, H.J., 2006. Land-Use and Land-Cover Change. Local processes and Global Impacts. Lambin, E.F. and H.J. Geist (Eds). The IGBP Series, Springer-Verlag, Berlin, 222 pp.

Lang, A.R.G., and Yueqin, X., 1986. Estimation of leaf area index from transmission of direct sunlight in discontinuous canopies Agric. For. Meteorol., 37,229-243, 10.1016/0168-1923(86)90033-X.

Latifovic, R., Homer, C., Ressl, R., Pouliot, D.A., Hossian, S., Colditz, R., Olthof, I., Chandra, G., Victoria, A., 2012. North American Land Change Monitoring System. Remote Sensing of Land Use and Land Cover: Principles and Applications. 303-324. 10.1201/b11964-24.

Leblanc, S.G., Chen, J.M., Fernandes, R., Deering, D.W., Conley, A., 2005. Methodology comparison for canopy structure parameters extraction from digital hemispherical photography in boreal forests. Agricultural and Forest Meteorology, 129, 187–207.

Lin, G., Wolfe, R., Zhang, P, Tilton, J., Dellomo, J., Tan, B., 2019. Thirty-six combined years of MODIS geolocation trending. Proc. SPIE 11127, Earth Observing Systems XXIV, 1112715 (9 September 2019); doi: 10.1117/12.2529447.

Monteith, J.L. and Unsworth, M.H., 2014. Principles of Environmental Physics, 4th ed. Academic Press, https://doi.org/10.1016/C2010-0-66393-0.

Müller-Wilm, U., 2018. Sen2Cor Configuration and User Manual (2nd ed.), CS, Toulouse, France.

Myneni, R., Knyazikhin, Y., Park, T., 2015. MCD15A3H MODIS/Terra+Aqua Leaf Area Index/FPAR 4-day L4 Global 500m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC. Accessed 2021-04-19 from https://doi.org/10.5067/MODIS/MCD15A3H.006.

Origo, N., Calders, K., Nightingale, J., & Disney, M., 2017. Influence of levelling technique on the retrieval of canopy structural parameters from digital hemispherical photography. AGRICULTURAL AND FOREST METEOROLOGY, 237–238, 143–149. https://doi.org/10.1016/j.agrformet.2017.02.004

Prikaziuk E, Yang P, and van der Tol C., 2021. Google Earth Engine Sentinel-3 OLCI Level-1 Dataset Deviates from the Original Data: Causes and Consequences. Remote Sensing,13(6):1098. https://doi.org/10.3390/rs13061098.

Putzenlechner, B., Castro, S., Kiese, R., Ludwig, R.;, Marzahn, P., Sharp, I., Sanchez-Azofeifa, A., 2019. Validation of Sentinel-2 fAPAR products using ground observations across three forest ecosystems. Remote Sens. Environ,, 232, 111310.

Radoux J, Chomé G, Jacques DC, Waldner F, Bellemans N, Matton N, Lamarche C, D'Andrimont R, Defourny P., 2016. Sentinel-2's Potential for Sub-Pixel Landscape Feature Detection. Remote Sensing., 8, :488. https://doi.org/10.3390/rs8060488.

Ryu, Y., Nilson, T., Kobayashi, H., Sonnentag, O., Law, B.E., Baldocchi, D.D., 2010.

On the correct estimation of effective leaf area index: Does it reveal information on clumping effects?,

Agricultural and Forest Meteorology,150, 463-472, https://doi.org/10.1016/j.agrformet.2010.01.009.

Sanchez-Sapero, J. and Martinez-Sanchez, E., 2022. QUALITY ASSESSMENT REPORT LAI, FAPAR, FCOVER FROM SENTINEL-3/OLCI COLLECTION 300M VERSION 1.1, Copernicus Global Land Operations – Lot 1, Issue: I1.20, accessed at

https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS1_QAR_LAI300m-V1.1_I1.20.pdf on March 15, 2023.

Scurlock, J., Asner, G.,Gower, S., 2001. Worldwide Historical Estimates of Leaf Area Index, 1932-2000. ORNL/TM-2001/268, Oak Ridge National Laboratory, U.S.A. bhttps://info.ornl.gov/sites/publications/Files/Pub57077.pdf.

Stenberg P., Mõttus M., Rautiainen, M., Sievänen, R., 2014. Quantitative characterization of clumping in Scots pine crowns. Ann Bot., 114(4), 689-694, doi: 10.1093/aob/mct310.

Verger, A. and Descals, A., 2022. ALGORITHM THEORETICAL BASIS DOCUMENT Leaf Area Index (LAI) Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) Fraction of green Vegetation Cover (FCover) Collection 300m Version 1.1, CGLOPS1_ATBD_LAI300m-V1.1 © CGLOPS Lot1 consortium, https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS1_ATBD_LAI300m-V1.1_I1.10.pdf.

Verhoef,W., Jia, L., Xiao, Q. Su, Z., 2007. Unified optical-thermal four-stream radiative transfer theory for homogeneous vegetation canopies, IEEE Trans. Geosci. Remote Sens., 45 1808-1822.

Weiss, M., Baret, F., Garrigues, S., and Lacaze, R., 2007. LAI and fAPAR CYCLOPES global products derived from VEGETATION. Part 2: Validation and comparison with MODIS collection 4 products. Remote sensing of Environment, 110(3), 317-331.

Weiss, M., Baret,F. Block, T.,Koetz, B., Burini, A., Scholze, B., Lecharpentier, P., Brockmann, C., Fernandes, R., Plummer, S., Myneni, R., Gobron, N., Nightingale, J., Schaepman-Strub, G., Camacho de Coca, F., Sanchez-Azofeifa, G.A., 2014. On Line Validation Exercise (OLIVE): A Web Based Service for the Validation of Medium Resolution Land Products. Application to FAPAR Products. Remote Sensing. 6. 4190-4216. 10.3390/rs6054190.

Weiss, M., Baret, F., 2016. S2ToolBox Level 2 Products: LAI, FAPAR, FCOVER, 1.1. ed. Institut National de la Recherche Agronomique, Avignon, France. Accessed at https://step.esa.int/docs/extra/ATBD_S2ToolBox_L2B_V1.1.pdf , on March 15, 2023.

Widlowski, J.L., Lavergne, T., Pinty, B. Gobron, N. Verstraete, M.M., 2008. Towards a high spatial resolution limit for pixel-based interpretations of optical remote sensing data, Adv. Space Res., 41, pp. 1724-1732.

Widlowski, J.L. Mio, C. Disney, M. Adams, J. Andredakis, I. Atzberger, C. Brennan, J. Busetto, I. Chelle, M. Ceccherini, G. et al., 2015. The fourth phase of the radiative transfer model intercomparison (RAMI) exercise: actual canopy scenarios and conformity testing. Remote Sens. Environ., 169, pp. 418-437.

Wolfe, R., Roy, D. and Vermote, E., 1998. MODIS land data storage, gridding and compositing methodology: Level 2 Grid. IEEE Trans Geosci Remote Sens, 36, 1324-1338.

Woodgate, W., Armston, J.D., Disney, M., Suarez, L. Jones, S.D., Hill, M.J., Wilkes, P., Soto-Berelov, M., 2017. Validating canopy clumping retrieval methods using hemispherical photography in a simulated Eucalypt forest. Agric. For. Meteorol., 247, 181-193, 10.1016/j.agrformet.2017.07.027.

Yamazaki D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J.C., Sampson, C.C., Kanae, S. and Bate, P.D., 2017. A high accuracy map of global terrain elevations. Geophysical Research Letters, vol.44, pp.5844-5853, 2017. doi:10.1002/2017GL072874.

Yan, K., Park, T., Yan, G., Liu, Z., Yang, B., Chen, C., Nemani, R.R., Knyazikhin, Y., Myneni, R.B., 2016. Evaluation of MODIS LAI/FPAR Product Collection 6. Part 2: Validation and Intercomparison. Remote Sensing, 8, 460. https://doi.org/10.3390/rs8060460.

1214

1215

1216

1217

1218