



Land Carbon Constellation

Satellite Data Uncertainty analysis

Scientific Report D-7

Version 1.0
19 October 2021

Prepared by



Land Carbon Constellation

Satellite Data Uncertainty analysis

Scientific Report D-7

ESA Land Carbon Constellation Project
Technical Officer: Matthias Drusch

CESBIO-SMOS report **SO-TN-CB-GS-0102**

N.J. Rodríguez-Fernández¹, J. Verrelst², H. Lindqvist³, M. Barbier¹, E. Bueechi⁴,
P. Reyez Muñoz², A. Mialon¹, W. Dorigo⁴, A. Bouvet¹, T. Kaminski⁵, M. Scholze⁶

¹ **CESBIO**

UMR 5126 (CNRS, CNES, UPS, IRD, INRAE),
18 av. Edouard Belin, bpi 2801, 31401 Toulouse cedex 9, France

² **Universitat de Valencia**

Valencia, Spain

³ **Finnish Meteorological Institute**

Erik Palmenin aukio 1, 00560 Helsinki, Finland

⁴ **TU Wien**

Vienna, Austria

⁵ **The Inversion Lab**

Hambourg, Germany

⁶ **University of Lund**

Lund, Sweden

19th October, 2021

Number of pages: 107

Authors:	N.J. Rodríguez-Fernández, J. Verrelst, H. Lindqvist, W. Dorigo, M. Barbier E. Bueechi, P. Reytez Muñoz, Y. Kerr, A. Mialon, A. Bouvet, T. Kaminski, M. Scholze		
Circulation:	ESA document		
Release	Date	Details	Editors
0.1	3 August 2021	Template and first figures	NRF
0.2	27 August 2021	Adding contributions from individual data sets	MB, EB, HL, PRM
0.3	17 September 2021	Adding microwave instruments comparison Adding optical data comparison Adding VOD-SIF comparison	MB HL, PRM MB
0.4	5 October 2021	Modifications after review by NRF	HL, PRM, MB, EB
0.5	6 October 2021	Adding introduction, abstract and conclusions	NRF
0.6	15 October 2021	Replacing temporal moving average for SMOS from 7 to 11 days	MB MB
0.7	18 October 2021	Review by MS and NRF	MS, NRF
1.0	19 October 2021	Review of Sect. 9 and new Appendix Delivery to LCC PI	NRF NRF

For any clarifications please contact N. Rodríguez-Fernández (nemesio.rodriguez@cesbio.cnes.fr).

Work Package 1 project partners

CESBIO, Centre d'Etudes Spatiales de la BIOsphère,
Toulouse, (France).

Universitat de Valencia, Valencia (Spain)

FMI, Finnish Meteorological Institute,
Helsinki (Finland)

TU Wien, Technical University of Vienna,
Vienna (Austria)

Contents

List of Figures	8
List of Tables	10
Reference documents and data sets	11
1 Executive summary	12
2 Introduction	12
3 SMOS data	13
3.1 Data (SM, L-VOD) and associated uncertainties	13
3.2 Data analysis: spatial and temporal variability analysis	18
3.3 Comparison of temporal and spatial average to VOD DQX	23
3.4 Conclusions: new estimation of uncertainties	25
4 ASCAT data	26
4.1 Data and associated uncertainties	26
4.2 Data analysis	26
4.3 Conclusions: new estimation of uncertainties	29
5 AMSR-2 data	30
5.1 Data and associated uncertainties	30
5.2 Data analysis	30
5.3 Conclusions	30
6 Optical data: Sentinel-3 OLCI	33
6.1 Data and associated uncertainties	33
6.2 Gaussian processes regression (GPR)	33
6.3 Data analysis: spatial and temporal variability analysis	34
6.4 Data analysis: temporal variability analysis	45
6.5 Conclusions: new estimation of uncertainties	47
7 SIF data	48
7.1 Data and associated uncertainties	48
7.2 Data analysis: spatial and temporal variability analysis	48
7.3 Conclusions: new estimation of uncertainties	54
8 AGB data	57

9	Vegetation-related variables and their uncertainty. Cross-comparison	59
9.1	Microwave optical depth	59
9.2	Optical indices and SIF	63
9.3	Microwaves vs optical data	67
10	Conclusions	71
10.1	Individual data sets	71
10.2	Data sets comparison	72
	Bibliography	74
A	Data base accompanying documentation	74

List of Figures

1	SMOS-L2 VOD (tau) average from July to September 2015 over the three regions selected for the study (top panel). The red circles show the location of Las Majadas, Reusel and Sodankyla.	14
2	SMOS L2 SM v700 vs v650	16
3	SMOS L2 tau v700 vs v650	17
4	SMOS L2 tau v700 FNO and FFO	18
5	Difference of SMOS L2 VOD and its neighbours between two days	19
6	Time serie of VOD	20
7	Time series VOD with obs, filter and 11d moving average	20
8	Time series of VOD and polynomial fitting	22
9	From top to bottom : Standard deviation of δ_{VODDQX} and $\delta_{TemporalVODDQX}$	24
10	biasDQX-VOD-VODAverage	25
11	ASCAT VOD	27
12	ASCAT VODts	27
13	ASCAT VOD	28
14	ASCAT VODts	28
15	AMSR-2 VOD	31
16	AMSR-2 VODts	32
17	FAPAR estimates (top) and associated uncertainties (bottom) over Las Majadas for four dates in 2009.	36
18	LAI estimates (top) and associated uncertainties (bottom) over Las Majadas for four dates in 2009.	36
19	LCC estimates (top) and associated uncertainties (bottom) over Las Majadas for four dates in 2009.	37
20	FVC estimates (top) and associated uncertainties (bottom) over Las Majadas for four dates in 2009.	37
21	FAPAR estimates (top) and associated uncertainties (bottom) over Reusel for four dates in 2009.	38
22	LAI estimates (top) and associated uncertainties (bottom) over Reusel for four dates in 2009.	39
23	LCC estimates (top) and associated uncertainties (bottom) over Reusel for four dates in 2009.	39
24	FVC estimates (top) and associated uncertainties (bottom) over Reusel for four dates in 2009.	40
25	FAPAR estimates (top) and associated uncertainties (bottom) over Sodankyla for four dates in 2009.	41
26	LAI estimates (top) and associated uncertainties (bottom) over Sodankyla for four dates in 2009.	42
27	LCC estimates (top) and associated uncertainties (bottom) over Sodankyla for four dates in 2009.	43

28	FVC estimates (top) and associated uncertainties (bottom) over Sodankyla for four dates in 2009.	44
29	Averaged temporal profiles (green) over Las Majadas area together with associated averaged uncertainties (gray) for LAI, FAPAR, LCC & FVC.	45
30	Averaged temporal profiles (green) over Reusel area together with associated averaged uncertainties (gray) for LAI, FAPAR, LCC & FVC.	46
31	Averaged temporal profiles (green) over Sodankyla area together with associated averaged uncertainties (gray) for LAI, FAPAR, LCC & FVC.	46
32	TROPOMI SIF estimates over Sodankylä for three days in 2019.	49
33	TROPOMI SIF estimates over Reusel for three days in 2019.	50
34	TROPOMI SIF estimates over Las Majadas for four days in 2019.	50
35	OCO-2 SIF estimates over Sodankylä for three months in 2019.	51
36	OCO-2 SIF estimates over Reusel for three months in 2019.	52
37	OCO-2 SIF estimates over Majadas for four months in 2019.	52
38	One-to-one plots of two TROPOMI SIF products: the all-sky product (743 nm) and the clear-sky product (735 nm).	53
39	Time series of the two TROPOMI SIF products: the all-sky product (743 nm) and the clear-sky product (735 nm) (left) and their difference (right).	54
40	Time series of the OCO-2 SIF retrieval products at 757 nm and 771 nm, and the computational estimate at 740 nm.	55
41	Time series of TROPOMI SIF presented for all of the three regions. Both individual SIF observations and their hourly averages are shown. The error bars correspond to the standard deviation from the hourly average.	56
42	AGB CCI 2018 histograms	57
43	AGB CCI VOD SMOS Scatter	57
44	AGB CCI VOD SMOS RMSD	58
45	FFO SMOS	58
46	VOD ASCAT vs VOD SMOS	60
47	MODIS LCType	61
48	Correlation VOD ASCAT vs VOD SMOS	62
49	Correlation VOD-Noise ASCAT vs VOD DQX SMOS	62
50	VOD AMSR-2 vs VOD SMOS	64
51	Correlation coefficients VOD AMSR-2 vs VOD SMOS	65
52	Time series (left panel) and the correlation (right panel) of the TROPOMI SIF and OLCI FAPAR estimates over the three focus areas. The co-located data are based on daily averages of the SIF and FAPAR over the entire focus areas. The dashed lines in the right panel figures are the best-fit linear estimates of the data.	66
53	SIF - FAPAR scatterplots: a) Las Majadas 16-07-2018. R2: 0.30 and b) Reusel 29-03-2019. R2: 0.16	67
54	Mean of SIF Tropomi vs VOD SMOS	68
55	Correlation coefficient map SIF Tropomi vs VOD SMOS	69

56	Time series of SIF and VOD	70
----	--------------------------------------	----

List of Tables

1	Summary of the data sets in the Land Surface Constellation project data base. . . .	14
2	Threshold values of RFI-Prob, VOD DQX and Chi2 when removing the 5 % highest values of those parameters in each region.	59

Reference documents

Doc/data name	Reference	Label
Satellite observation documentation including justification for selection	LCC D-4	RD1
Satellite observation data base including uncertainty estimates as provided with the data sets	LCC D-5	RD2
Satellite observation data base, quality controlled with revised uncertainty estimates	LCC D-6	RD3

1 Executive summary

The quality assessment and uncertainties evaluation of the Earth Observation (EO) data in the Land Carbon Constellation project data base [RD2] is presented. In addition, a first product cross-comparison is discussed. A new release of the project data base [RD3] has been done containing the quality assessed data with new uncertainties estimations when needed.

The focus of the analysis was put on the data sets that will be used for data assimilation later during the project: SMOS SM and VOD, ASCAT VOD, Sentinel 5-P SIF and Sentinel 3 FAPAR, but other data sets such as AMSR-2 VOD, OCO-2 SIF and Sentinel-3 LAI, FVC, LCC have also been analyzed.

The SMOS Level 2 v720 data were compared to v650. As expected, the main differences arise in regions with mixed land cover, where the new v720 shows a smoother behavior in forest/low vegetation transition zones. The temporal and spatial variability of VOD were compared to the VOD uncertainty provided with the data and a new estimation of the uncertainty was provided. A similar approach was used for AMSR-2 C1, C2 and X-bands VOD. ASCAT VOD time series are smoothed with a kernel window of 42 days. The VOD uncertainty increases with increasing VOD values but they are in general very low ($10^{-2} - 10^{-3}$). No alternatives were found to provide more realistic uncertainties but a filtering recommendation was provided.

Regarding the S3-OLCI vegetation products (LAI, FAPAR, FVC, LCC), they are retrieved using a Bayesian framework in which uncertainties are implicitly calculated together with the mean estimations. Within this study, the training data set was generalized by mixing simulated and real spectra data taken from original S3 scenes providing a new more robust uncertainty estimation.

SIF data from OCO-2 and Sentinel 5p were analyzed. There is a large variability in the single SIF observations and it is recommended to use hourly averages with the standard deviation of all SIF values used to compute the hourly average as associated uncertainty.

The VOD data from different sensors were compared. Bias maps of ASCAT and AMSR-2 with respect to SMOS show both regions with positive and regions with with negative bias. There are also regions with negative temporal correlations. The maps were compared to land cover maps but no clear relationship was found so far.

The temporal correlation of SIF and FAPAR was found to vary between the focus areas with values from $R=0.59$ in Sodankyla, to 0.71 in Las Majadas and 0.81 in Reusel. FAPAR and SIF seasonal cycles are similar but FAPAR increases in early spring and only in late May or June for SIF.

Finally, a first comparison between microwave and visible data was performed by comparing SMOS VOD and Sentinel 5P SIF. Temporal correlation maps show regions with negative values, in particular in the Las Majadas regions.

2 Introduction

This document is the Scientific Report associated to the Work Package 1.3 of the Land Carbon Constellation (LCC) project. Within this work package a comprehensive data base was built containing vegetation-related data collected from different instruments operating at wavelengths from

the near-infrared to microwaves. These different data sets were presented in document **RD1**, which includes a justification for their selection. Table 1 shows a summary of the data available in the LCC data base and their temporal coverage. In addition, there are three ESA Climate Change Initiative above ground biomass (AGB) maps representative of years 2010, 2017 and 2018.

Figure 1 shows the three regions selected for the LCC study. In the following, an analysis of the data that will be assimilated (SMOS Soil Moisture and Vegetation Optical Depth -VOD-, ASCAT VOD, Sentinel 3 Fraction of Absorbed Photosynthetically Active Radiation, Sentinel 5p Sun Induced Fluorescence) in addition to some of the auxiliary data sets (AMSR-2 VOD, OCO-2 SIF, Sentinel 3 Leaf Area Index, Leaf Chlorophyll Content, Fraction Vegetation Cover, AGB) is presented for the three regions shown in Fig. 1.

The goals of the analysis are:

- to perform a quality assessment of the individual data sets
- to provide new uncertainty estimations if needed
- to look for possible uncertainties correlation
- to perform a first cross-comparison of the different vegetation-related parameters obtained from different sensors at different wavelengths.

The structure of this document is as follows: Section 3 discusses the SMOS VOD data. Sections 4 and 5 present the ASCAT and AMSR-2 VOD, respectively. Section 6 discusses optical data from Sentinel 3 and Section 7 presents the OCO-2 and Sentinel-5 SIF data. The auxiliary above ground biomass data set is briefly discussed in Sect. 8. Different data sets are compared in Sect. 9, in particular the active and passive microwave VOD data obtained at different frequencies are compared in Sect. 9.1, SIF and FAPAR are compared in Sect. 9.2 and SMOS VOD and Sentinel 5 SIF are compared in Sect. 9.3. Finally, the conclusions are summarized in Sect. 10. A technical note describing the data formats in the data base is provided in Appendix A.

3 SMOS data

3.1 Data (SM, L-VOD) and associated uncertainties

The SMOS data version available when the Land Carbon Constellation (LCC) project started was v650. The new version v700 became available in the first quarter of 2021. As stated in the proposal, it was decided to move to the new version of SMOS data for the LCC project. Therefore, the first step in the data analysis consisted in comparing the two versions for the three regions of interest. Figure 2 shows the mean of SM for v700 from 2011 to March 2021, it also shows the bias and the correlation with respect to v650. Figure 3 is equivalent to Fig. 2 but for L-VOD (τ) instead of SM.

According to Figs. 2 and 3, the first thing to be noticed in the differences between v650 and v700 for SM and VOD is that they follow a geographical pattern. It means that one version provides higher or lower values than the other for a particular area. By cross-checking these differences with

Table 1: Summary of the data sets in the Land Surface Constellation project data base.

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Exact period
SMOS L-VOD												01/2011-12/2021
ASCAT C-VOD												01/2011-12/2021
OCO-2 SIF												09/2014-12/2021
Sentinel 5P SIF												12/2017-12/2021
S3 FAPAR/LAI												04/2016-12/2021
ASCAT backscattering												01/2011-12/2021
SMOS soil moisture												01/2011-12/2021
SMOS TB												01/2011-12/2021
AMSR-2 C-VOD												11/2012-12/2021
AMSR-2 X-VOD												11/2012-12/2021
MODIS LST												01/2011-12/2021
MODIS PRI												01/2011-12/2021
S3 LCC												04/2016-12/2021
S3 FVC												04/2016-12/2021

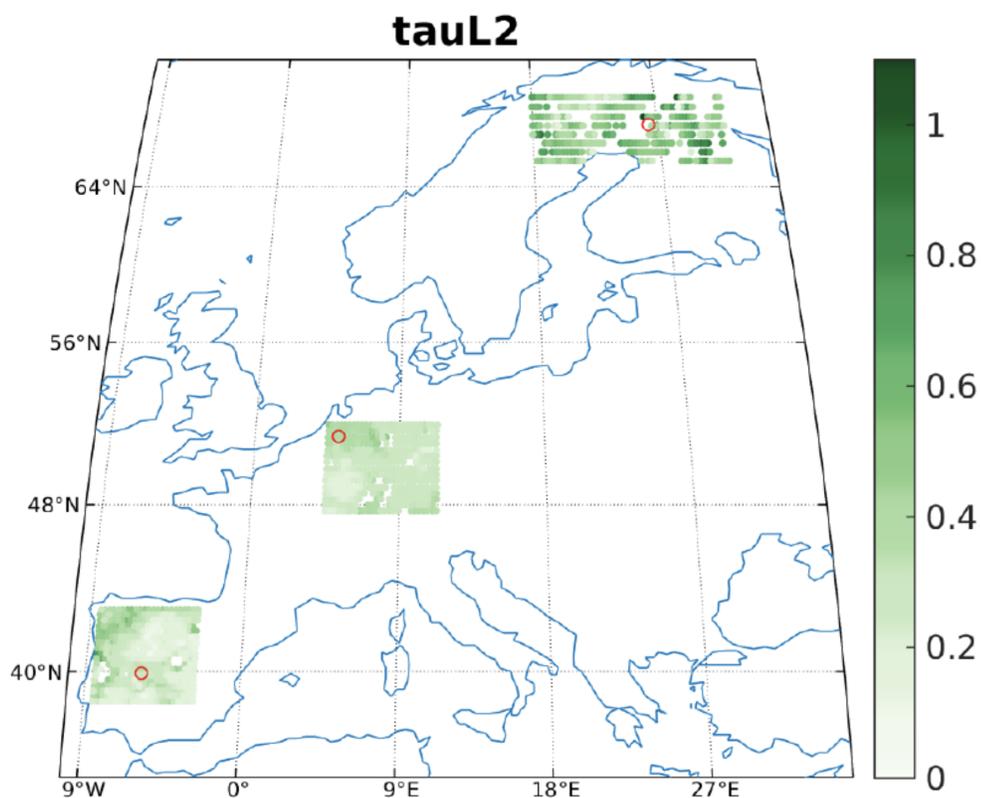


Figure 1: SMOS-L2 VOD (tau) average from July to September 2015 over the three regions selected for the study (top panel). The red circles show the location of Las Majadas, Reusel and Sodankyla.

forest and low vegetation maps (fig. 4), it appears that the main difference between v650 and v700 is the way they are processing areas without a predominant vegetation type. Areas with the most significant difference are those with a transition between nominal soil and forest. Comparing Fig. 4 with bias maps of v650 and v700 (2nd line maps of Figs. 2 and 3) shows clearly a correlation between areas where v650 gives higher results than v700 for soil moisture and vegetation optical depth and areas with a nominal and forest cover between 40% and 60%.

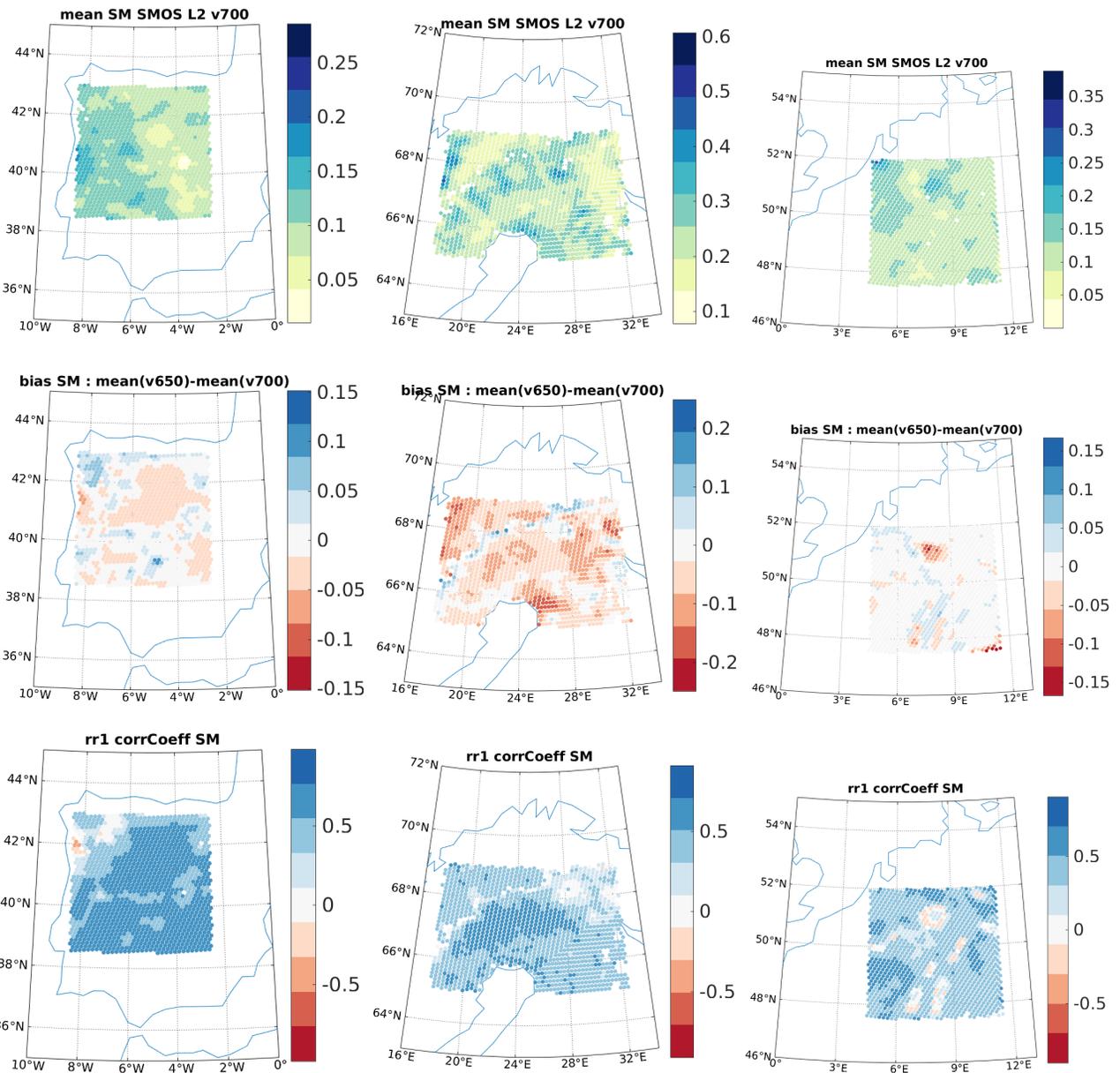


Figure 2: From top to bottom: (i) Mean of SMOS L2 v700 SM. (ii) Bias of SMOS L2 v650 SM versus v700 SM. (iii) Correlation of SMOS L2 SM v650 versus v700. The results are presented for the three regions, from left to right: Las Majadas, Sodankyla, Reusel

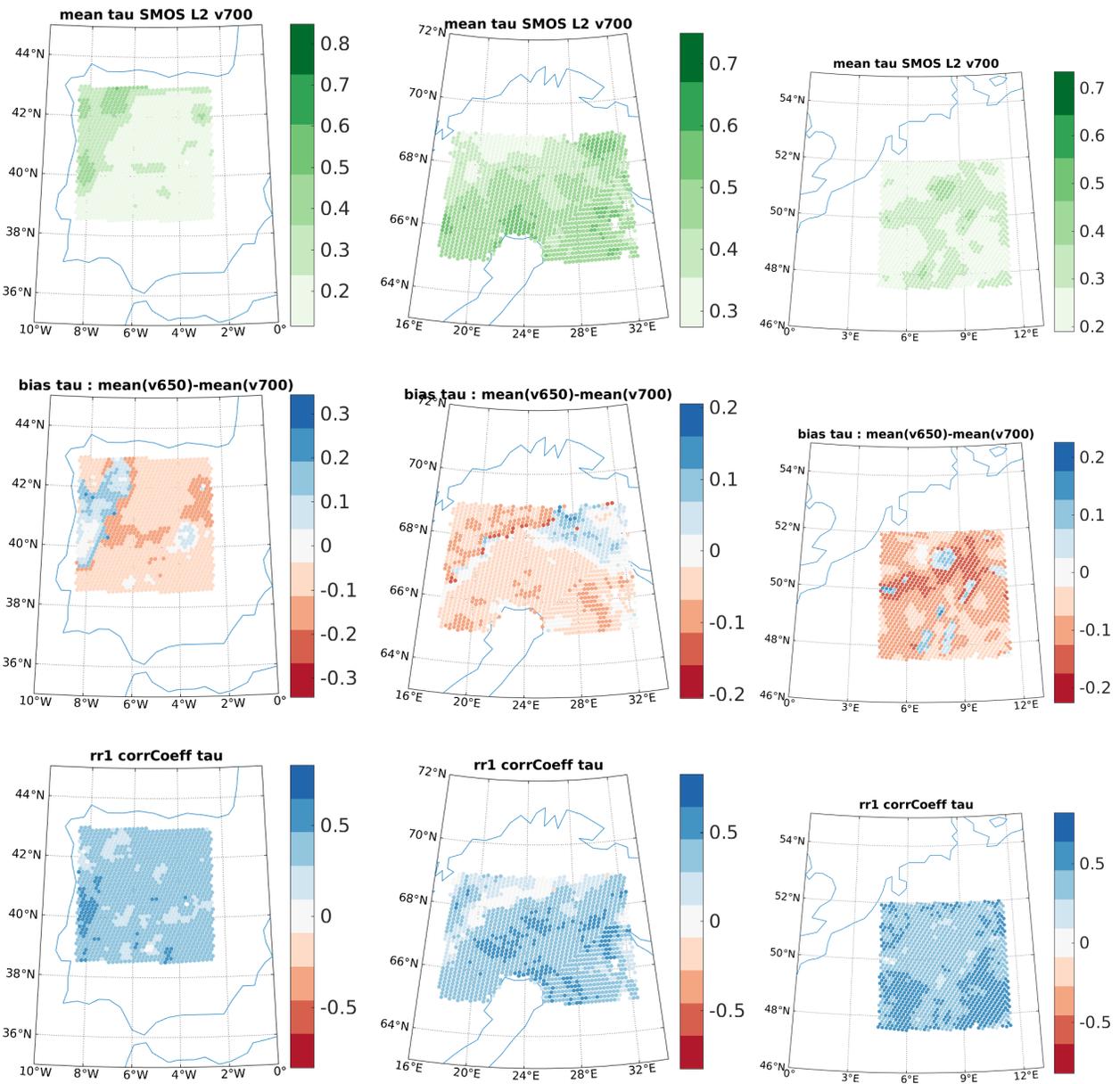


Figure 3: From top to bottom: (i) Mean of SMOS L2 v700 tau (L-VOD). (ii) Bias of SMOS L2 v650 tau versus v700 tau. (iii) Correlation of SMOS L2 tau v650 versus v700. The results are presented for the three regions, from left to right: Las Majadas, Sodankyla, Reusel

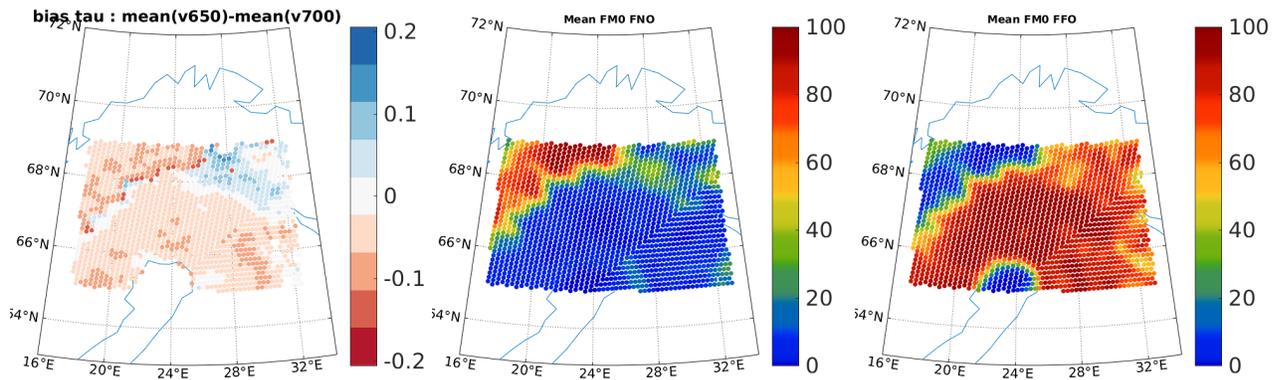


Figure 4: From left to right: (i) Bias of SMOS L2 v650 tau versus v700 tau. (ii) Mean value of nominal soil fraction. (iii) Mean value of forest soil fraction. The results are presented here for the region of Sodankylä

3.2 Data analysis: spatial and temporal variability analysis

The instrument resolution of SMOS is on average 43 km. This corresponds to the projection on the Earth surface of the full width at half maximum of the main lobe of the synthesized beam. For instance, to take into account all the surface contributing to the signal, the Level 2 algorithm uses a "working area" of 123 km \times 123 km. The SMOS v700 data are provided in an ISEA 15 km grid, which implies that adjacent grid points are not fully independent. The difference of SM and L-VOD from one point to the closest neighbors is a combination of geophysical signal variations and observational uncertainties.

On the other hand, SM can vary fast with time at scales of hours. In contrast, L-VOD is not expected to vary strongly in intervals of several days. Thus, analyzing the time series variations in short time periods could provide a new estimation of uncertainties in the data-set, that can be compared to the already existing uncertainty indices (dqx) and other quality index provided with the data (RFI , $Chi2$).

The data provided in SMOS-L2 present high space and temporal variations. To provide the most complete and clean data-set it is necessary to make a spatial and temporal analysis of the data, and compare these variations to the known quality indices in order to decide which data should be kept or not. To complete these quality indices and be able to observe a clear trend on the data, a moving average of these data will be used as a reference to compare the difference between this reference and the data and then comparing it with the quality indices.

3.2.1 Spatial analysis

Figure 5 shows the difference of SMOS L2 VOD with respect to the spatial average for the 8 closest grid points for two dates over the Las Majadas region. To do so, the distance from every observed points to the others has been calculated. The spatial average was then calculated using the VOD value of a point, and the values of the 8 closest points, using therefore the value of 9 different points.

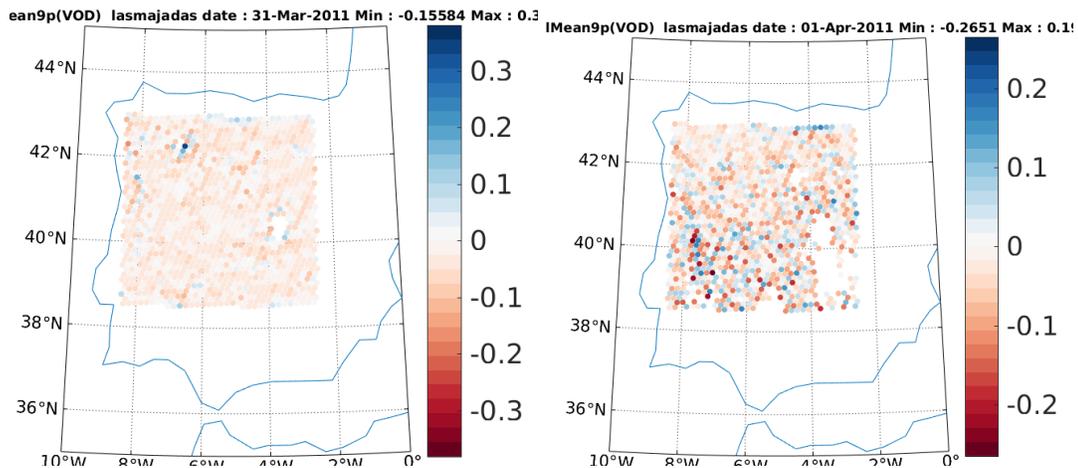


Figure 5: Difference between SMOS L2 VOD and the mean of VOD of its 8 closest neighbours. Comparison between two days in the region of Las Majadas

There is no spatial or temporal pattern in the evolution of these differences that seems to evolve randomly day after day. Confirming that they are driven by noise. As expected, the noise is reduced by the spatial averaging. Figure 6 shows the spatial average a VOD times series for a given point compared to the original VOD time series at that point. As expected, the spatial average reduces the noise of the original time series but there is still a significant variability and the spatial average time series does show some spikes that are probably noise. The VOD difference with respect the 8 neighbours spatial average is compared to VOD DQX in Sect. 3.3.

3.2.2 Temporal variability analysis. Moving averages

Taking into account the previous considerations on the temporal variability of VOD, a centered moving average on 11 days has been used to compare with the original VOD values (the same window size was also used for the AMSR-2 analysis, Sect. 5).

These quantities are compared to VOD DQX in Sect. 3.3.

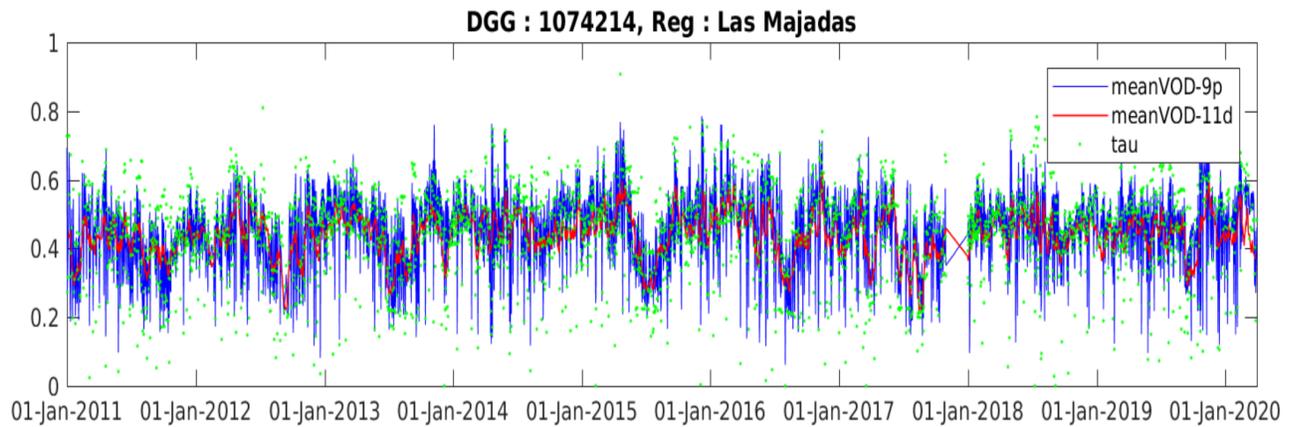


Figure 6: Time serie of VOD with spatial and temporal moving average for the point 1074214 of Las Majadas region

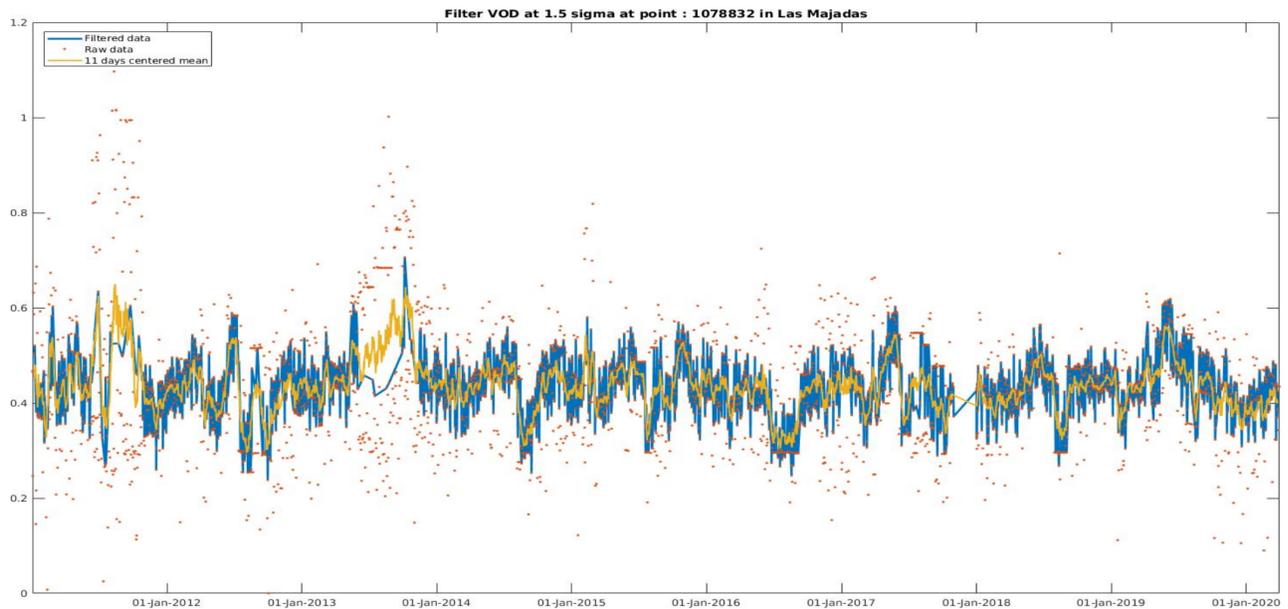


Figure 7: Time series of VOD, containing observations, filtered data and 11 days moving average for the point 1078832 in the region of Las Majadas

3.2.3 Temporal variability analysis and time series fitting

Instead of using a moving average, another way to analyse these data is the use of a mathematical fitting model. As for both SM and VOD are subject to seasonal and periodic effects they can be fitted to a polynomial function. Figure 8 shows a time-series with a strong seasonality (grid point n°2002894), and an other with a smoother one (grid point n°1082426). Four major cycles per year were identified (corresponding to the seasonal cycle) for the VOD and as the observation duration

occurred on nine years it was decided to use a 36 degree polynomial. The upper and lower envelope (purple and yellow on fig. 8) of the polynomial are the initial polynomial with a translation of 1.5σ . σ corresponds to the standard deviation of detrended data. As the 11 days moving average is smoothing the data, in order to be the most conservative as possible, upper and lower envelopes have to be respectively above and beneath this moving average. However, this method is not easily applicable to time series with a less significant seasonality and therefore it will not be discussed further.

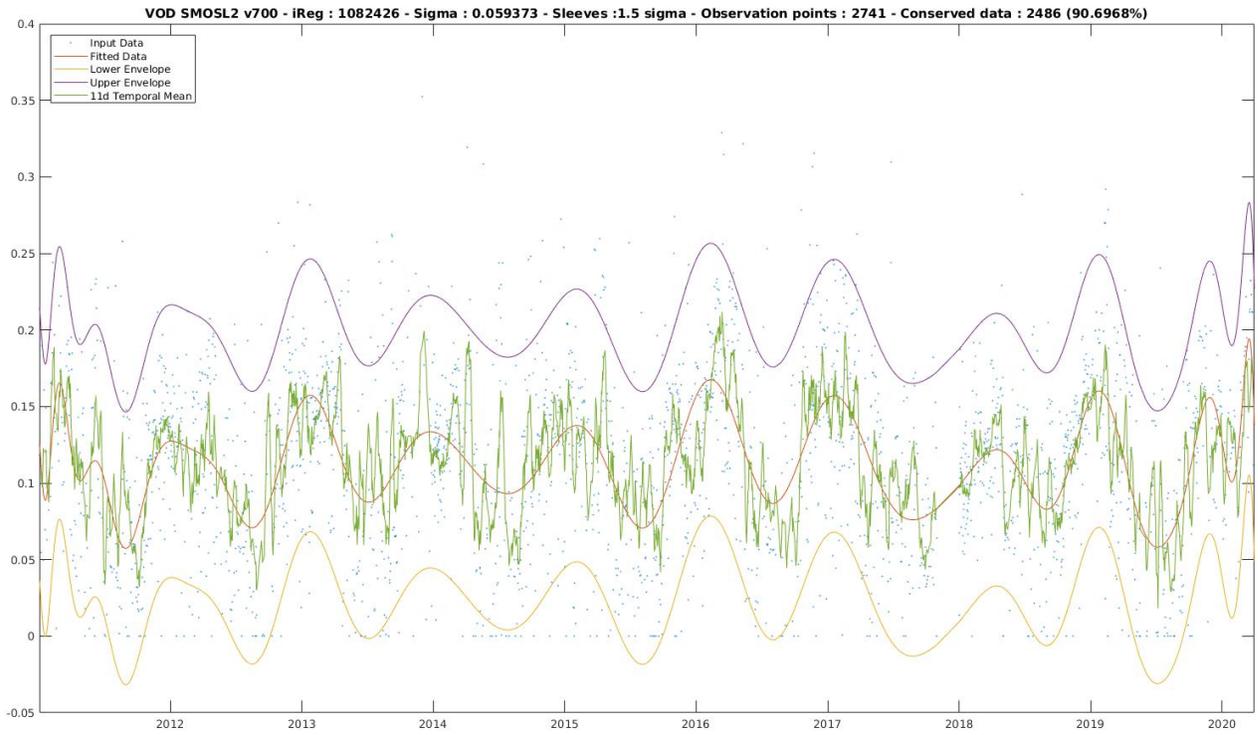
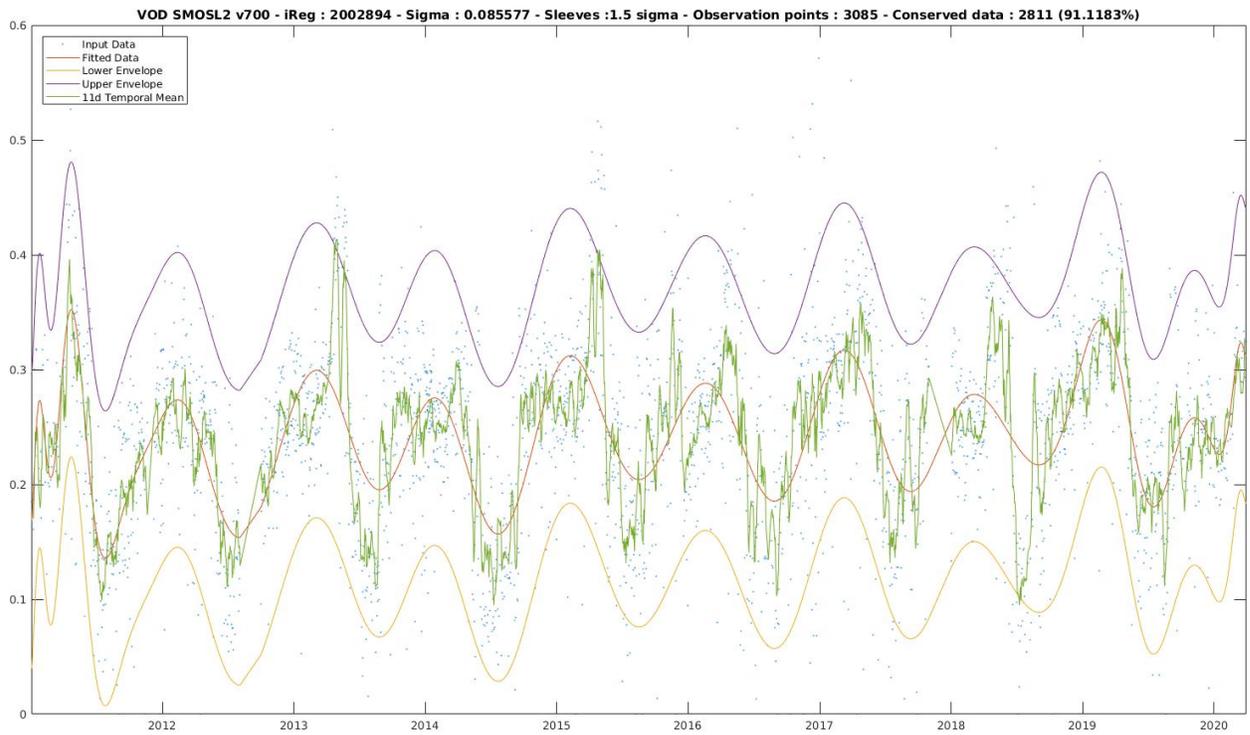


Figure 8: Time series of VOD with their spatial moving average and polynomial fitting. From top to bottom are points 2002894 and 1082426 of Las Majadas region

3.3 Comparison of temporal and spatial average to VOD DQX

The spatial and temporal VOD variations computed in the sections above were compared to the VOD DQX uncertainty associated to the VOD values. VOD DQX is defined as follows (Kerr et al., 2011):

$$DQX = RSTD \times C(PRFI) \quad (1)$$

With

$$C(PRFI) = 4 \times PRFI + 1, \text{ for } PRFI \in [0; 1] \quad (2)$$

where PRFI is the Radio Frequency Interference probability.

$$RSTD(SM^*, VOD^*) = \sqrt{([J(SM^*, VOD^*)]^t \times COV_Z^{-1} \times [J(SM^*, VOD^*)])^{-1}} \quad (3)$$

Where RSTD is the Retrieval Standard Deviation, $[J(SM^*, VOD^*)]$ is the Jacobian matrix of retrieved SM and VOD, and COV_Z a diagonal matrix containing antenna radiometric uncertainties and prior variances.

In order to compare the temporal and spatial variation of VOD with respect to VOD DQX, the differences :

$$\delta_{SpatialVODDQX} = DQX_VOD - |VOD_{Spatial-Mean} - VOD| \quad (4)$$

and

$$\delta_{TemporalVODDQX} = DQX_VOD - |VOD_{temporal-Mean} - VOD|, \quad (5)$$

were computed. The standard deviation of $\delta_{SpatialVODDQX}$ and $\delta_{TemporalVODDQX}$ are shown in the right and left panels of Figure 9, respectively. Figure 10 shows the mean of $\delta_{TemporalVODDQX}$ and $\delta_{SpatialVODDQX}$, in other words, the bias of VOD DQX with respect to $|VOD_{Spatial-Mean} - VOD|$ and $|VOD_{temporal-Mean} - VOD|$, respectively. It appears that $|VOD_{Spatial-Mean} - VOD|$ is lower than VOD DQX on average. For the temporal average the results are more variable. In Reusel VOD DQX is higher than $|VOD_{temporal-Mean} - VOD|$ over most of the region while in Sodankyla $|VOD_{temporal-Mean} - VOD|$ is higher. In Las Majadas region, the bias shows both regions with higher VOD DQX and other regions with higher $|VOD_{temporal-Mean} - VOD|$.

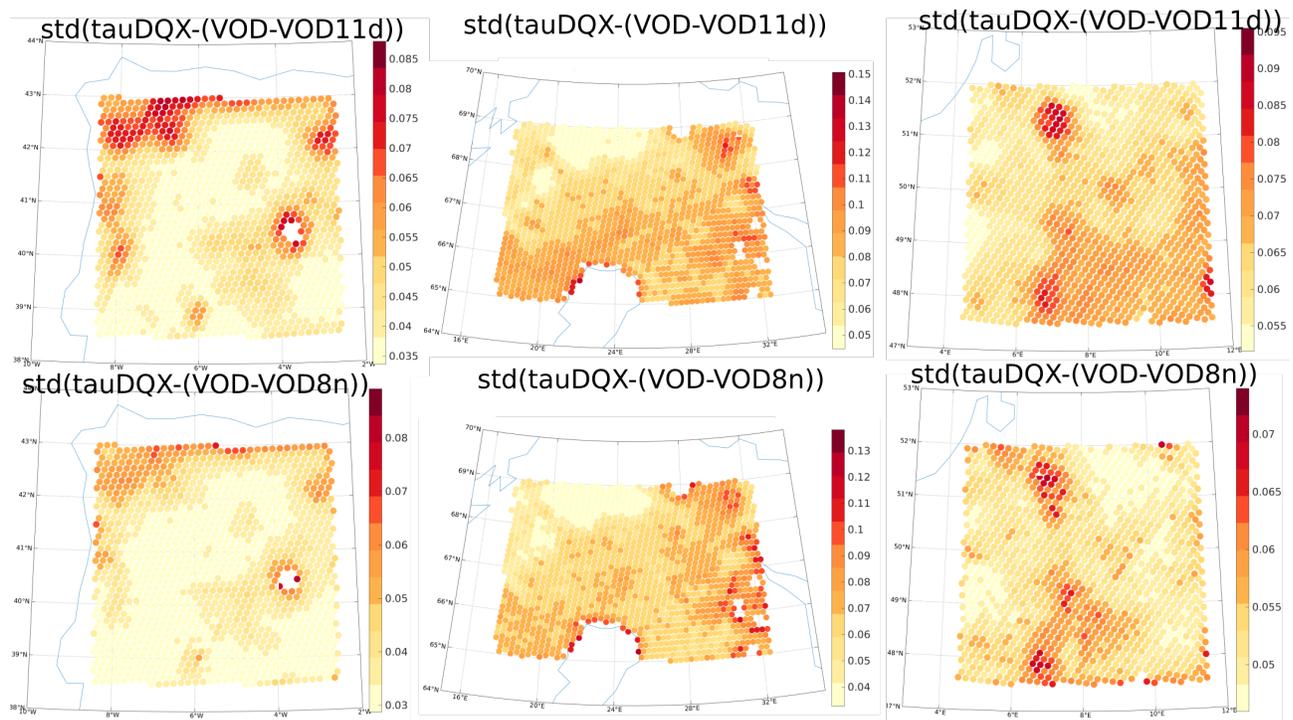


Figure 9: From the left to the right : (i) Standard deviation of the $\delta_{TemporalVODDQX}$. (ii) Standard deviation of $\delta_{SpatialVODDQX}$. In the region of Las Majadas, Sodankyla and Reusel.

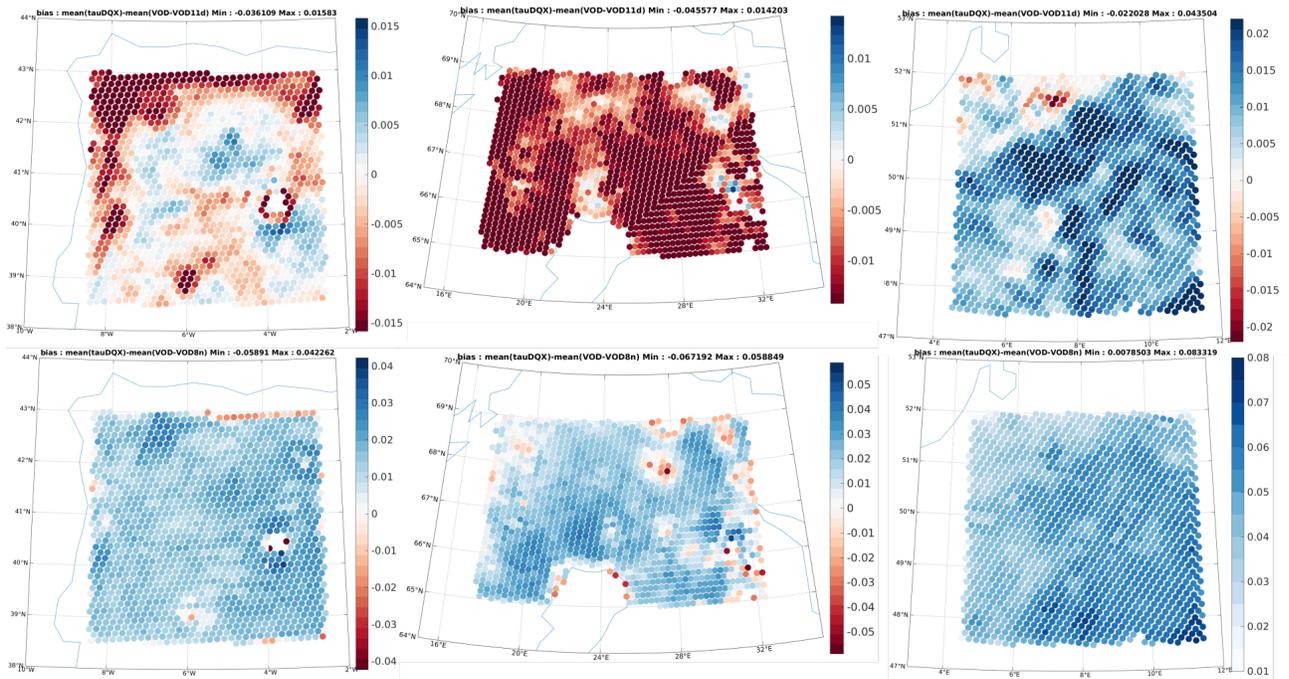


Figure 10: From top to bottom mean of $\delta_{TemporalVODDQX}$ and mean of $\delta_{SpatialVODDQX}$; From the left to the right : Las Majadas, Sodankyla, Reusel

3.4 Conclusions: new estimation of uncertainties

The evaluation confirms the quality improvement of SMOS Level 2 v700 with respect to v650. Observations are provided with an uncertainty index, DQX, that combines the model sensitivity to uncertainties in the input data and uncertainties arising from the possible presence of RFI. Taking into account that the ESA L2 grid oversamples the instrument resolution and that strong VOD variations are not expected in 1-2 weeks time, a spatial and temporal analysis of VOD was done. VOD averages of nine adjacent grid points were computed. In addition, smoothed VOD time series (keeping the original number of samples) were obtained computing averages in 11 days moving windows. The difference of local VOD values were compared to the spatial averages and the difference of VOD at a give time was compared to the 11 days temporal average. Assuming that at those spatial and temporal scales, VOD variations are mainly due to noise, those quantities can be compared to DQX. The analysis showed that they have a similar magnitude, some times being higher and some times lower than DQX. Therefore, a new way to define conservative uncertainty estimations of SM and VOD could be to use the highest of those values as the uncertainty of the observations.

4 ASCAT data

4.1 Data and associated uncertainties

Vreugdenhil et al. (2016) developed a method to retrieve VOD from ASCAT backscatter observations using the TUW soil moisture retrieval algorithm and a water-cloud model. Using the multi-incidence angle observation capabilities of the ASCAT instrument, the TUW soil moisture retrieval algorithm corrects for vegetation based on the relationship between backscatter and incidence angle. It is assumed that an increase in vegetation density increases volume scattering and subsequently increases the slope of the backscatter – incidence angle relationship. The slope is then used to correct for vegetation, where an increase in vegetation density decreases the sensitivity to soil moisture. Vreugdenhil et al. (2016) used the water-cloud model to model the sensitivity to soil moisture obtained through the TUW algorithm and subsequently obtain VOD.

One of the disadvantages of the VOD calculation based on the TUW algorithm is its dependence on parameters within the TUW algorithm. As the TUW algorithm is a change detection method it scales backscatter between the backscatter related to the wettest and driest observed conditions, the so-called wet and dry reference. Over very dry climates (Koeppen Geiger climate classes B) it is assumed that these wettest conditions are not met, and thus is the wet reference set to a static backscatter level. Within the VOD retrieval this can lead to spatial jumps in the VOD, which can be observed over Spain (Figure 11). Alternatively, the slope can be directly used to observe vegetation dynamics as the slope of the backscatter incidence angle relationship is related to vegetation dynamics and is independent of parameter adjustments in the wet reference.

The VOD data shows a clear seasonality in all three regions. The lowest values are reached around the end of the year, and the highest values between May and August. Las Majadas has its peaks generally a bit earlier in the year than the other two regions (Figure 12). The spatial distribution shows a high variability in Las Majadas and generally low values in Reusel and high values in Sodankylaa. The noise shows a similar pattern as VOD: it is higher where VOD is high (Figure 11). Overall, the noise is low. The backscatter signal shows a high spatial variability in Las Majadas too and rather high values in Reusel. The noise of the backscatter is low except one cluster in Las Majadas with higher noise (see Figure 13). The rather low noise in the backscatter data can be seen in the timeseries too (Figure 14). However, as a not smoothed data set, there is a high temporal variability. This is likely due to a combination of a changing soil moisture signal and noise.

4.2 Data analysis

A first potential source of uncertainties of the ASCAT VOD and sigma40 data set arise from the measurement uncertainties which concern both the backscatter signal and its slope. The parameter VOD noise, sigma40 noise, respectively cover the measurement uncertainties. The VOD noise was calculated via error propagation. A second source of uncertainties, from the retrieval model, concerns the VOD data only. These uncertainties are harder to evaluate individually. However, the combined effect of the uncertainties of the measurement and the retrieval algorithm are tried to be reduced by applying a kernel smoothing window of 42 days. This reduces the not realistic daily variations

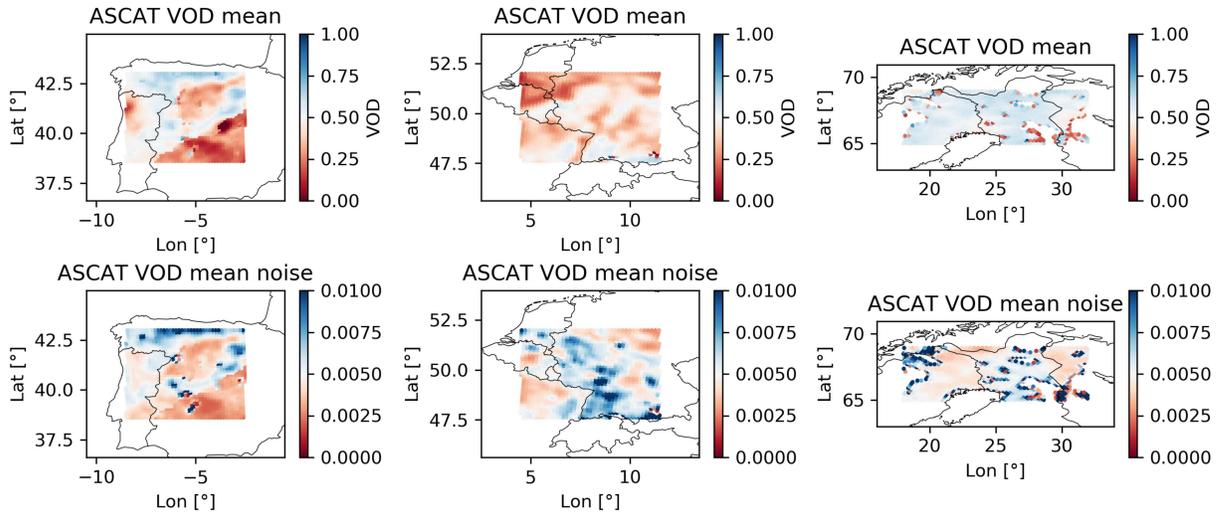


Figure 11: Mean ASCAT VOD values from 2011 to 2020 above and mean noise below

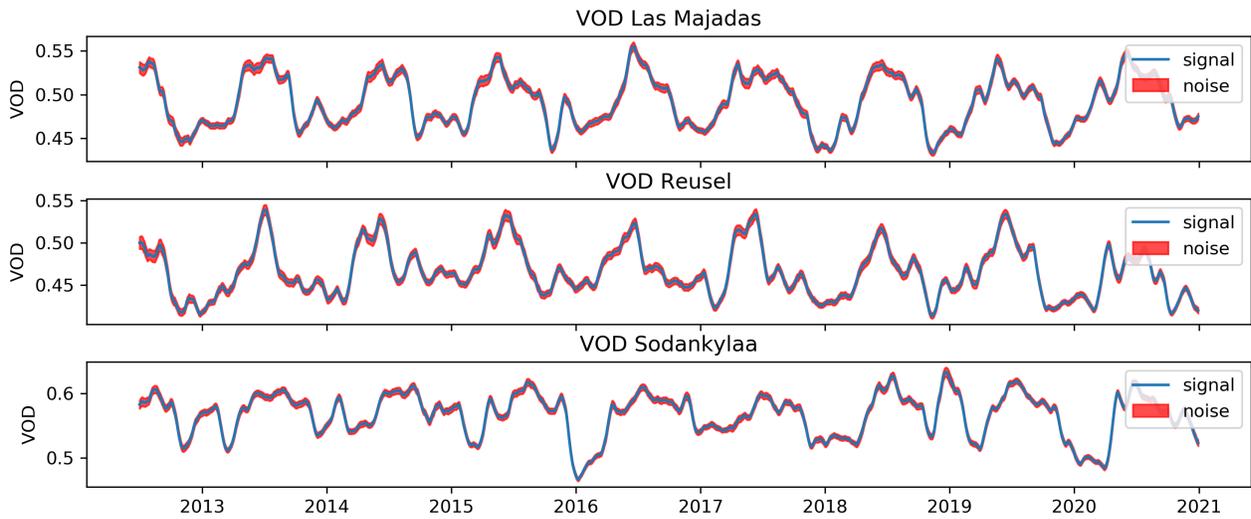


Figure 12: ASCAT VOD timeseries of one location per region

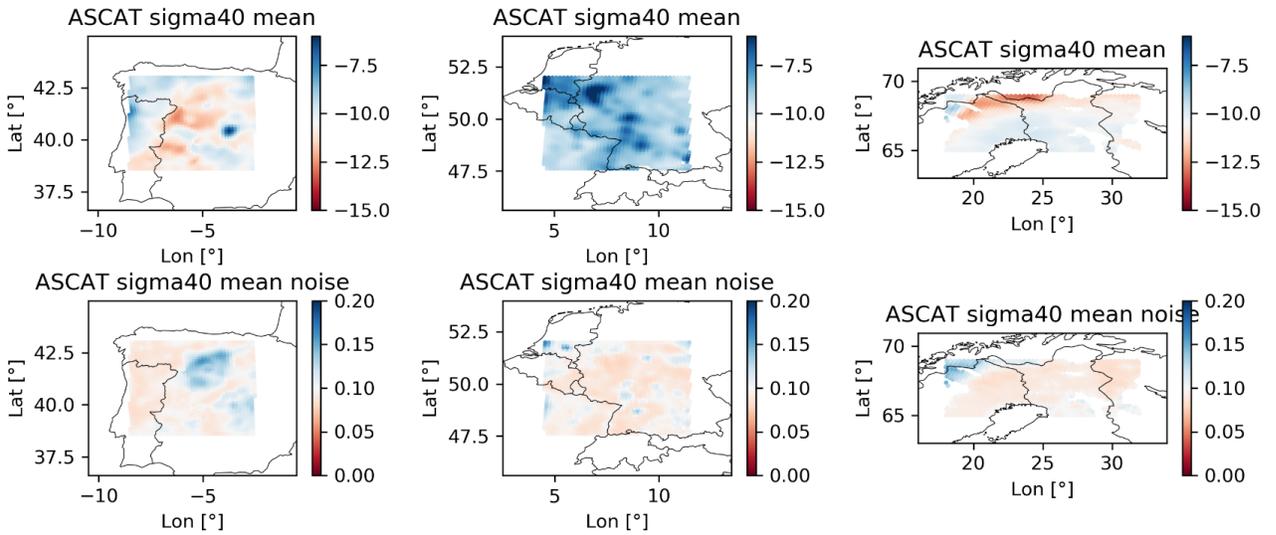


Figure 13: Mean ASCAT backscatter values from 2011 to 2020 above and mean noise below

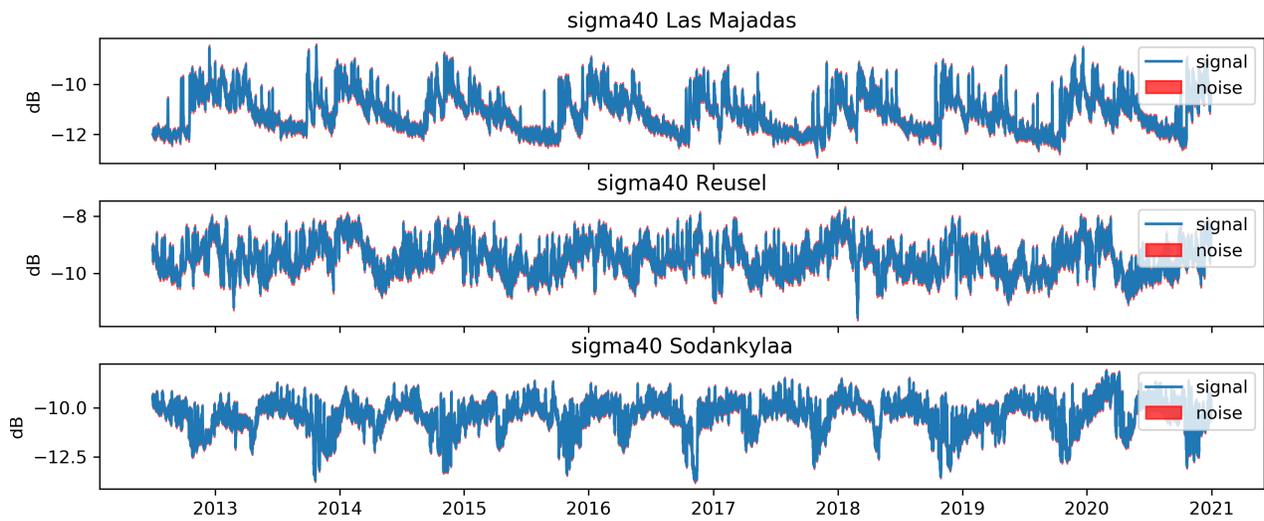


Figure 14: ASCAT backscatter timeseries of one location per region

of VOD (cf. 3.2). Additionally, the noise is reduced by that as well. It is not expected that the variable *noise* in the data sets cover all sources of uncertainties. Still, it can be used as reasonable proxy for the uncertainties. Finally, it needs to be stated, that these uncertainty estimations are only for conditions when the soils are neither frozen nor covered with snow. Frozen soils and snow cover lead to a systematic error which is not considered in the noise, and should, therefore, be masked out by using the respective flags, for both backscatter and VOD.

4.3 Conclusions: new estimation of uncertainties

The main source of measurement uncertainties are included in both data sets as the variable *VOD noise* and *sigma40 noise*. The noise is overall rather low and varying a bit with the location. For further improvement of the use of the data it is recommended to apply the provided flagging to get rid of measurements with systematic errors. First of all, locations flagged in the variable *Proc-flag* with a 2 should be masked out. Additional flagging is required for observations on frozen and / or snow covered soils. Ideally, this is done with an additional external dataset like ERA5-Land or similar. If this is not desired, the variables *probability of frozen soils* and *snow probability* can be used to mask observations with a high probability (eg. >60) of being frozen and / or covered with snow. Above that, the noise of the VOD data was reduced by the smoothing process. Overall, it can be stated that the backscatter and its slope are likely more robust and reliable than the VOD, as there are no additional uncertainties introduced by the modelling to retrieve the VOD signal. This brings us to the conclusion that it may be a better choice to use the slope directly for the assimilation instead of the VOD data set. As a next step, there will be some further tests done by TUW to check if it is feasible to use backscatter slope instead of VOD for the assimilation.

5 AMSR-2 data

5.1 Data and associated uncertainties

AMSR-2 VOD data is based on the Land Parameter Retrieval Model v6 using horizontal and vertical polarized brightness temperature. The VOD that were calculated for the bands C1, C2, and X are used for this study validation purposes. There are no uncertainties calculated from the model. However, observations which are likely to have high uncertainties are flagged. For the purpose of this study an uncertainty estimate was calculated similarly as in 3.2.2. The data shows overall the same patterns as ASCAT VOD: in Las Majadas the highest mean VOD values are reached in the Northwest, and the lowest variability over the study areas is in Sodankylaa (Figure 15). Above that, there is at all locations a clear seasonality and the peaks are between May and August as well. In Sodankylaa, there are some increasing values towards the end and the beginning of the year. There, the uncertainties are large, hence, it could be due to a systematic error from snow cover or frozen soils which are not flagged yet (cf. Figure 16).

5.2 Data analysis

The AMSR-2 VOD for all bands is not smoothed. The high daily variabilities are expected to be due to noise (cf. 3.2). Therefore, the temporal variability analysis was performed too. As there is only one daily value, it was decided to make the moving average window slightly longer and set to 11 days. This 11-day mean was computed using the value of the current day plus the values of the previous and the subsequent five days. The 11-day mean was only calculated for time periods with more than six available measurements. The noise was then computed as the difference from the current day to the 11-day mean. The noise estimation obtained like this should be a reasonable proxy for the uncertainties within these data sets. The noise for both C1 and X band is generally low (<0.025). With C1 band having a higher variability. Amongst other, urban areas (close to Munich, Bruxelles and Madrid) seem to lead to higher uncertainties for C1 band (>0.1). This is most likely due to radio frequency interference. X Band VOD has overall lower uncertainties, and more evenly distributed over the study areas (cf. Figure 15).

5.3 Conclusions

Measurements with likely high uncertainties are masked out in the data. Additionally, flags are provided in the variable *FLAGS* to give information why certain measurements are / should not be used. In the frame of this project an additional uncertainty measure was calculated: the temporal variability analysis. This is an estimation for the noise obtained in the data sets. It showed that noise values computed in this way are generally low. For C1 Band, it showed that the noise has a high spatial variability. Overall, the conclusion of 3.4 is valid for AMSR-2 VOD too: the moving window is an effective way to estimate the uncertainties. For the use of the data it can be stated that most unreliable measurements are already flagged. Therefore, it can be used as it is provided. If additional flagging for uncertain measurements is desired, then areas with a low signal to noise ratio (eg. <0.8) can be flagged.

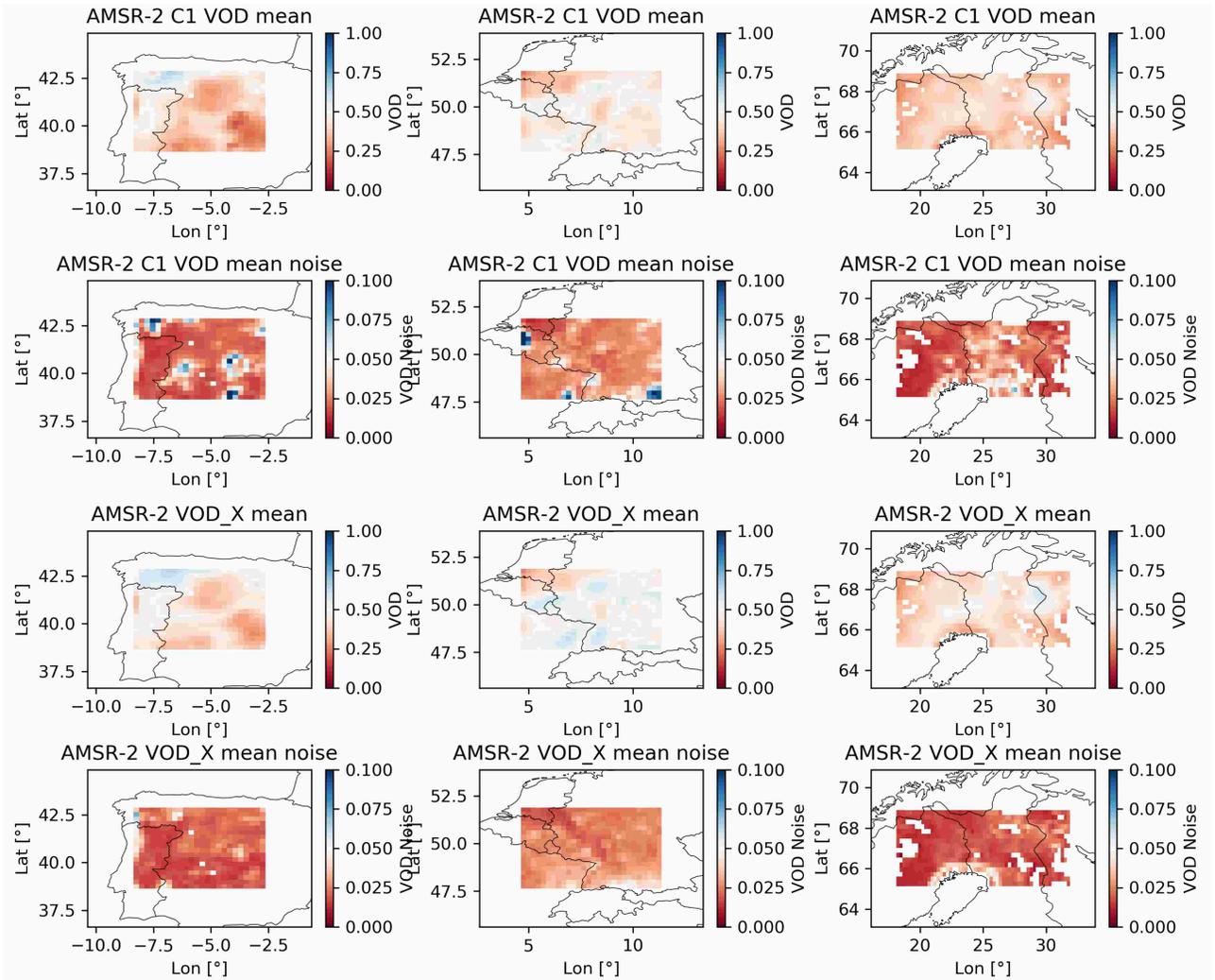


Figure 15: mean of AMSR-2 VOD of C1 Band from 2012 to 2020

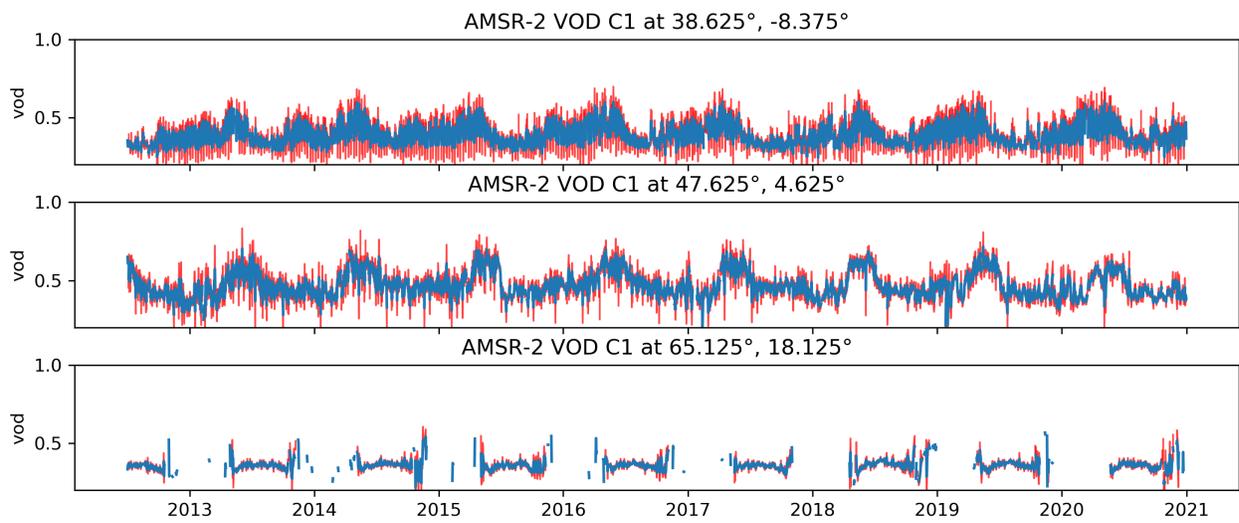


Figure 16: AMSR-2 VOD timeseries of C1 Band of one location per region. Noise estimation is displayed in red

6 Optical data: Sentinel-3 OLCI

6.1 Data and associated uncertainties

The S3-OLCI vegetation products (LAI, FAPAR, FVC, LCC) are retrieved using a hybrid machine learning approach based on Gaussian Processes Regression (GPR) models. The approach and a first model version has been described in De Grave et al. (2020).

The advantage of using GPR is that it is developed in a Bayesian framework, i.e. uncertainties are implicitly calculated together with the (mean) estimations. The GPR algorithm is based on (Bayesian) probability distribution functions for the target variables, constrained by the incoming OLCI spectral data. Uncertainties are based on deviations in estimations between an initial state of the model and last obtained results through an automatic learning process from the new incoming data. The deviations give us information on how well behave our model when facing real data.

The accuracy of the model estimations depends on the performance of the training dataset when covering all the possible conditions in which the model will be applied. Originally the GPR models were trained with exclusively simulations coming from the radiative transfer model (RTM) SCOPE, along with spectra of non-vegetated surfaces to account for the variability present in OLCI images. As opposed to De Grave et al. (2020), the training dataset was subsequently upscaled with the 6SV atmospheric RTM in order to be able running at top-of-atmosphere (TOA) scale, i.e. the OLCI product L1C. The reason to do so, was to ensure processing the full OLCI data stream from its very beginning.

During the course of this project, it was studied on how to optimize further the used training dataset. For this reason, at the writing of this report, an improvement of the original first models version (V1.0) have been applied, becoming a new version of model: V1.1. In this version, our models were trained by mixing SCOPE simulations together with real spectra data taken from original S3 scenes.

With regards to the MODIS ancillary data LST (Land Surface Temperature) and PRI (Photochemical Reflectance Index) collected by the same group, uncertainties were not provided by the original source, and therefore are not present in this delivery.

6.2 Gaussian processes regression (GPR)

This section provides the theoretical framework how estimates and associated uncertainties are calculated. The GPR model establishes a relation between the input (B -bands spectra) $\mathbf{x} \in \mathbb{R}^B$ and the output variable (canopy parameter) $y \in \mathbb{R}$. Assuming that the observed variable y is formed by noisy observations of the true underlying function $f(\mathbf{x})$, i.e. $y = f(\mathbf{x}) + \epsilon$, and that the noise ϵ is additive band-independent Gaussian distributed with zero mean and variance σ_n , the estimation of y provided by GPR can be expressed as :

$$\hat{y} = \sum_{i=1}^N \alpha_i K(\mathbf{x}_i, \mathbf{x}), \quad (6)$$

where $\{\mathbf{x}_i\}_{i=1}^N$ are the training spectra, $\alpha_i \in \mathbb{R}$ is the weight assigned to each one of them, and K is the Kernel function evaluating the similarity between the test spectrum \mathbf{x} and the generic training spectra \mathbf{x}_i , $i = 1, \dots, N$. The kernel selected for this work is the standard scaled radial basis function (RBF), defined as

$$K(\mathbf{x}_i, \mathbf{x}_j) = \nu \exp \left(- \sum_{b=1}^B \frac{(\mathbf{x}_i^{(b)} - \mathbf{x}_j^{(b)})^2}{2\sigma_b^2} \right) + \delta_{ij}\sigma_n^2, \quad (7)$$

where ν is a scaling factor, σ_b is a dedicated parameter controlling the spread of the training information for each particular spectral band b , σ_n is the noise standard deviation and δ_{ij} is the Kronecker's symbol. The kernel is thus parametrized by signal (ν , σ_b) and noise (σ_n) hyperparameters, collectively denoted as $\boldsymbol{\theta} = \{\nu, \boldsymbol{\sigma}, \sigma_n\}$, where $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_B)$.

Let us define the stacked output training samples $\mathbf{y} = (y_1, \dots, y_N)^\top$, the covariance terms of the test point $\mathbf{k}_* = (k(\mathbf{x}_*, \mathbf{x}_1), \dots, k(\mathbf{x}_*, \mathbf{x}_N))^\top$, and $k_{**} = k(\mathbf{x}_*, \mathbf{x}_*)$. For prediction purposes, the GPR is obtained by computing the posterior distribution over the unknown output y_* , $p(y_* | \mathbf{x}_*, \mathcal{D})$, where $\mathcal{D} \equiv \{\mathbf{x}_n, y_n | n = 1, \dots, N\}$ is the training dataset. This posterior can be shown to be a Gaussian distribution, for which one can estimate the *predictive mean* (point-wise predictions):

$$\mu_{y_*} = \mathbf{k}_*^\top (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{y}, \quad (8)$$

and the *predictive variance* (confidence intervals, or uncertainty estimates):

$$\sigma_{y_*}^2 = k_{**} - \mathbf{k}_*^\top (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{k}_*. \quad (9)$$

The corresponding hyperparameters $\boldsymbol{\theta}$ are typically obtained by maximizing the marginal likelihood (also called *evidence*) of the observations. See Rasmussen and Williams (2006) for further details. A more detailed survey on GPR properties and alternative kernel versions is also provided in Camps-Valls et al. (2016).

On the application side, the following important properties of the method are worth stressing. First, the obtained weights α_i after hyperparameter optimization gives the relevance of each training spectrum \mathbf{x}_i . The predictive mean is essentially a weighted average of the canopy parameter values associated to the training samples: the closer to the test sample, the higher the weight. Second, and probably most interesting for mapping applications, the GPR model provides not only a per-pixel prediction, but also an uncertainty (or confidence) level for the prediction, which enables to evaluate the model transferability in space and time.

6.3 Data analysis: spatial and temporal variability analysis

Regarding the S3 OLCI instrument, all the 21 bands were used for the retrieval of the biophysical variables. The GPR models were upscaled to TOA radiance (L1C) and subsequently implemented into Google Earth Engine (GEE) where the processing has been done. That step allowed us to process the vegetation products directly and only over the selected test areas. For the three test areas, the products were extracted at the nominal temporal (approx. 2 days) and spatial resolution

(300 m). Data data is processed from April 2016 until 20th of November 2020. For each test site an area of 500x500 km² were processed.

Because the probabilistic GPR model provides directly the mean estimate along with the associated uncertainty, in the following some output maps of both products are provided. As a demonstration case, maps are provided for 4 particular days within the year of 2019 and the months of March, July, September and December, obtaining daily results belonging to the 4 different seasons. The selected day depended on the cloud coverage; for each site it was attempted to show maps with a minimum amount of cloud coverage. The clouds are masked out the maps. First fAPAR is given, being a robust variable. Following LAI is given, presenting biggest uncertainties compared to fAPAR. For reasons of consistency, third also LCC and finally FVC are given. Each time first the mean estimates are provided, with the associated uncertainties underneath.

6.3.1 Las Majadas

Starting with Las Majadas, Spain, this site led to the best quality maps because of the following reasons: (1) lack of clouds, and (2) bright illumination. Estimates are on top in the figure, uncertainties underneath. It must be remarked while for the estimates the same color range was preserved over the 4 maps, for the uncertainties this was not always the case due to substantial outliers. Particularly LAI still gives high uncertainties (Figure 18). The retrieval of this variable appeared to be the most challenging: although the estimates throughout the year greatly improved as opposed to V1.0, the uncertainties still suggest the GPR model faces difficulties, specially for particular land covers. The maximum LAI uncertainties were obtained for low vegetated areas in the eastern part of this site, as observed in the images on different dates.

Also, LCC was characterized by overly high uncertainties for July. However, because these high uncertainties are not systematic, they are rather related to atmospheric effects present in some of the bands.

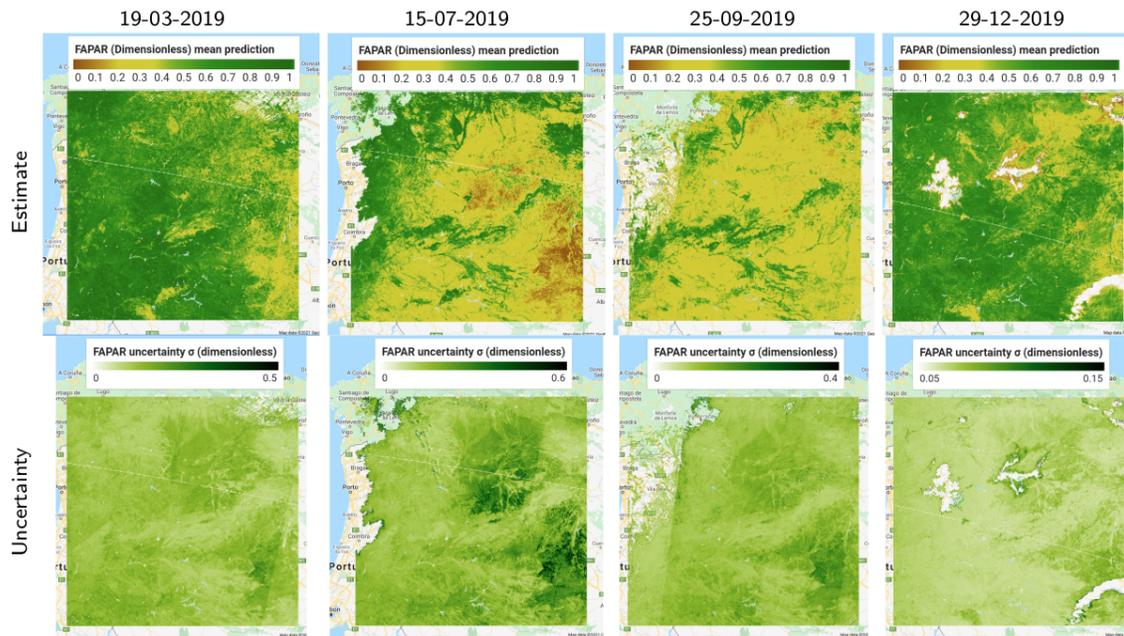


Figure 17: FAPAR estimates (top) and associated uncertainties (bottom) over Las Majadas for four dates in 2009.

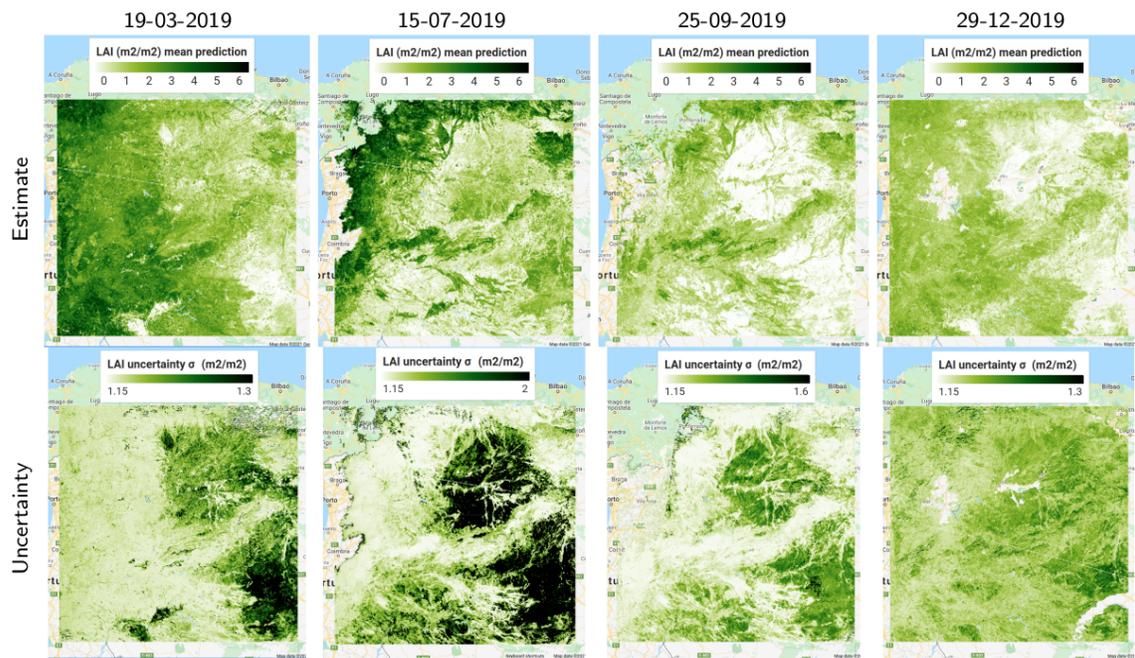


Figure 18: LAI estimates (top) and associated uncertainties (bottom) over Las Majadas for four dates in 2009.

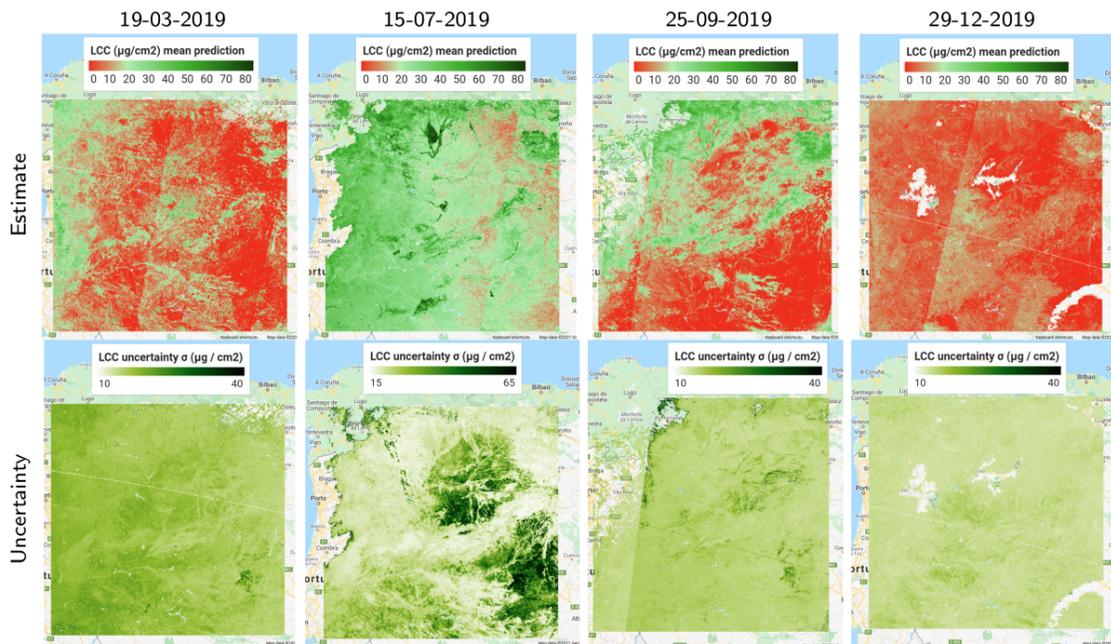


Figure 19: LCC estimates (top) and associated uncertainties (bottom) over Las Majadas for four dates in 2009.

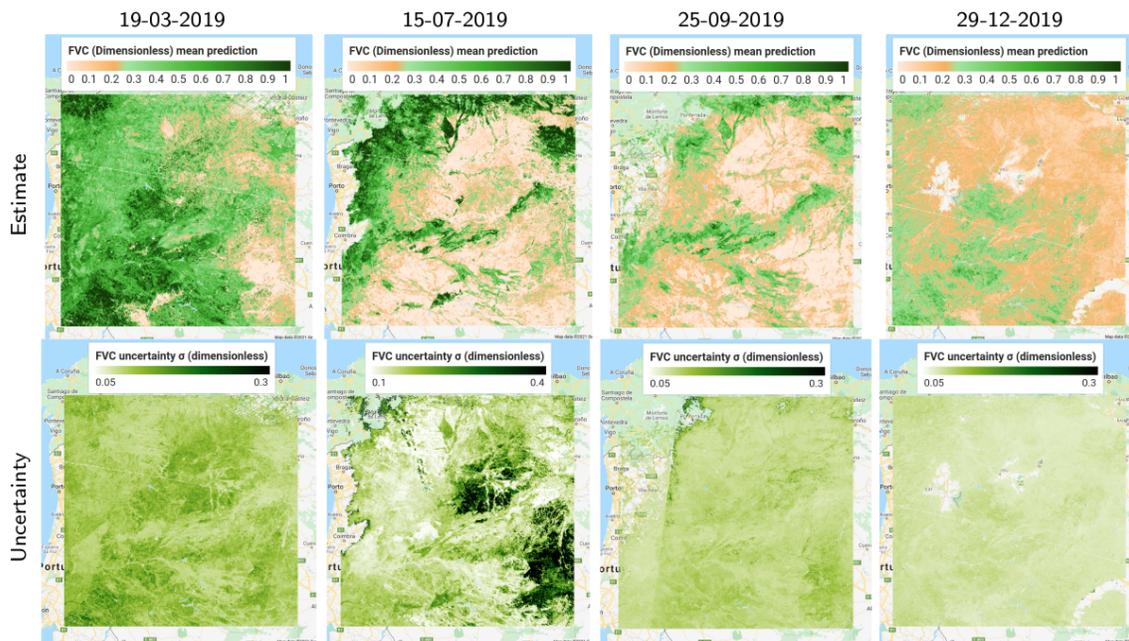


Figure 20: FVC estimates (top) and associated uncertainties (bottom) over Las Majadas for four dates in 2009.

6.3.2 Reusel

Following, Reusel, Netherlands, here already more clouds are present, leading to gaps. LAI uncertainties are still irregular over time, confirming the difficulties this model faces. The uncertainties of the other variables are more stable and within expected range.

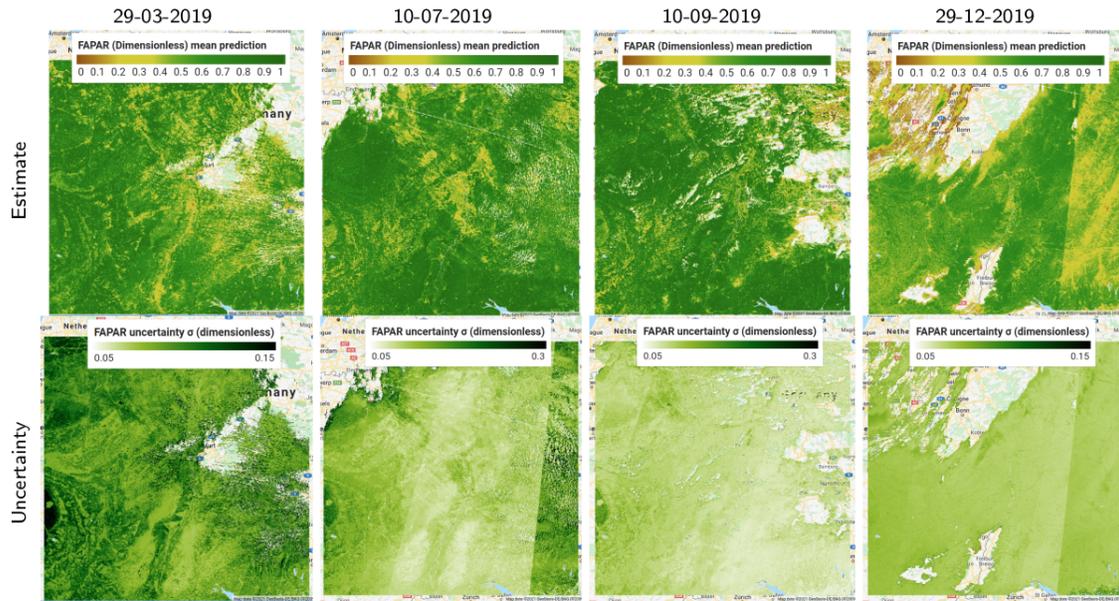


Figure 21: FAPAR estimates (top) and associated uncertainties (bottom) over Reusel for four dates in 2009.



Figure 22: LAI estimates (top) and associated uncertainties (bottom) over Reusel for four dates in 2009.

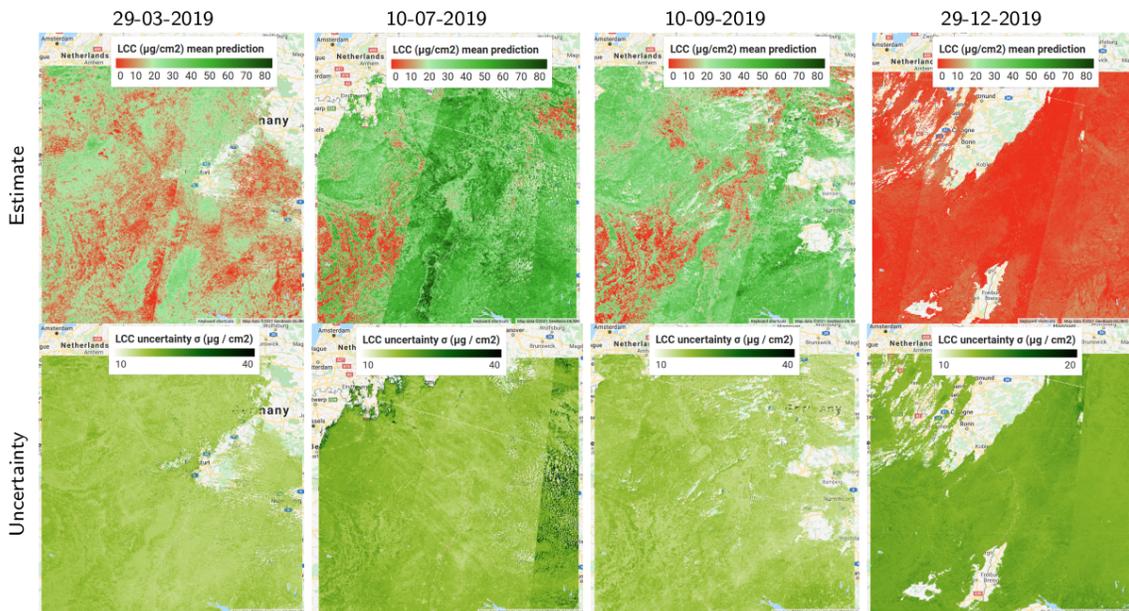


Figure 23: LCC estimates (top) and associated uncertainties (bottom) over Reusel for four dates in 2009.

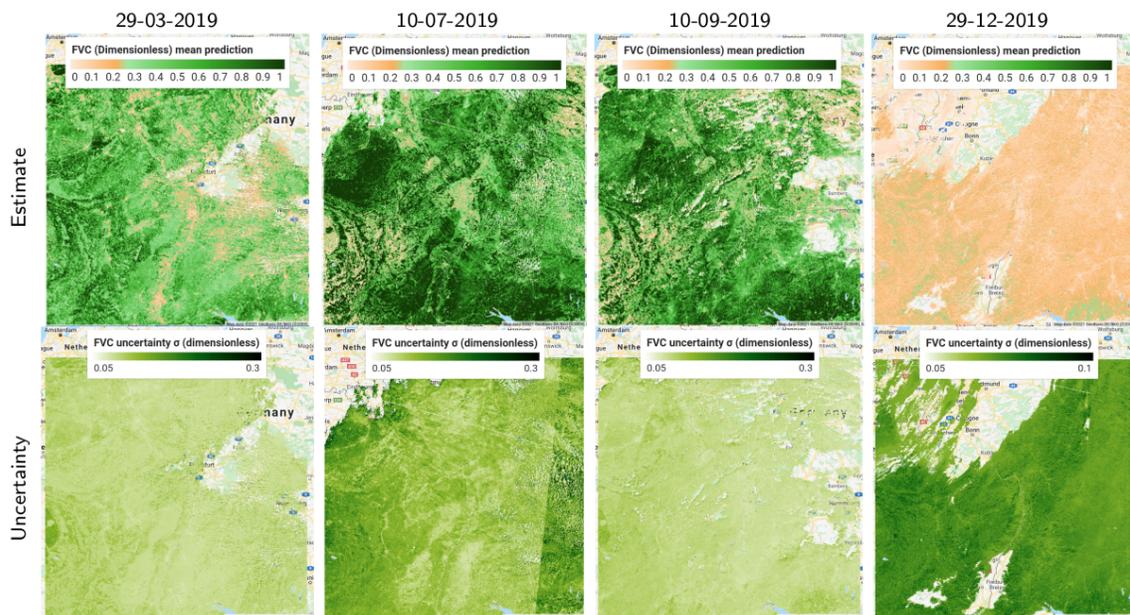


Figure 24: FVC estimates (top) and associated uncertainties (bottom) over Reusel for four dates in 2009.

6.3.3 Sodankyla

Finally, for Sodankyla it was impossible to retrieve vegetation variables in winter time due to permanent darkness. Here again, LAI can lead to large uncertainties, likely due to atmospheric effects as they appear around cloud cover. The same high uncertainties are also observed for the other variables.

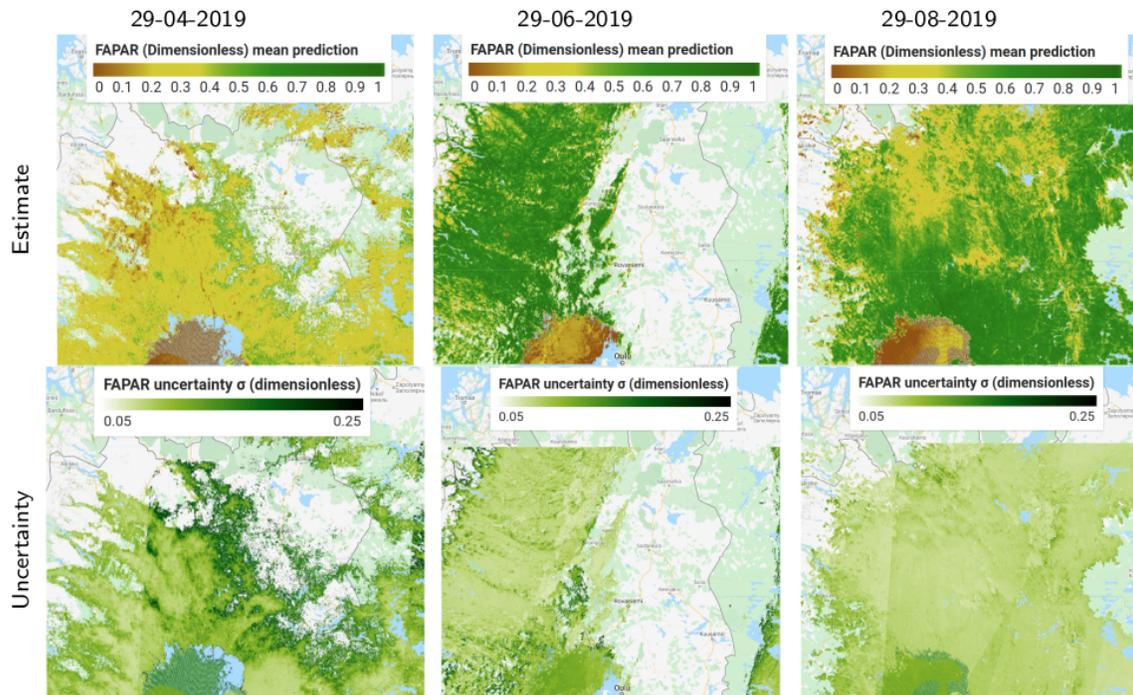


Figure 25: FAPAR estimates (top) and associated uncertainties (bottom) over Sodankyla for four dates in 2009.

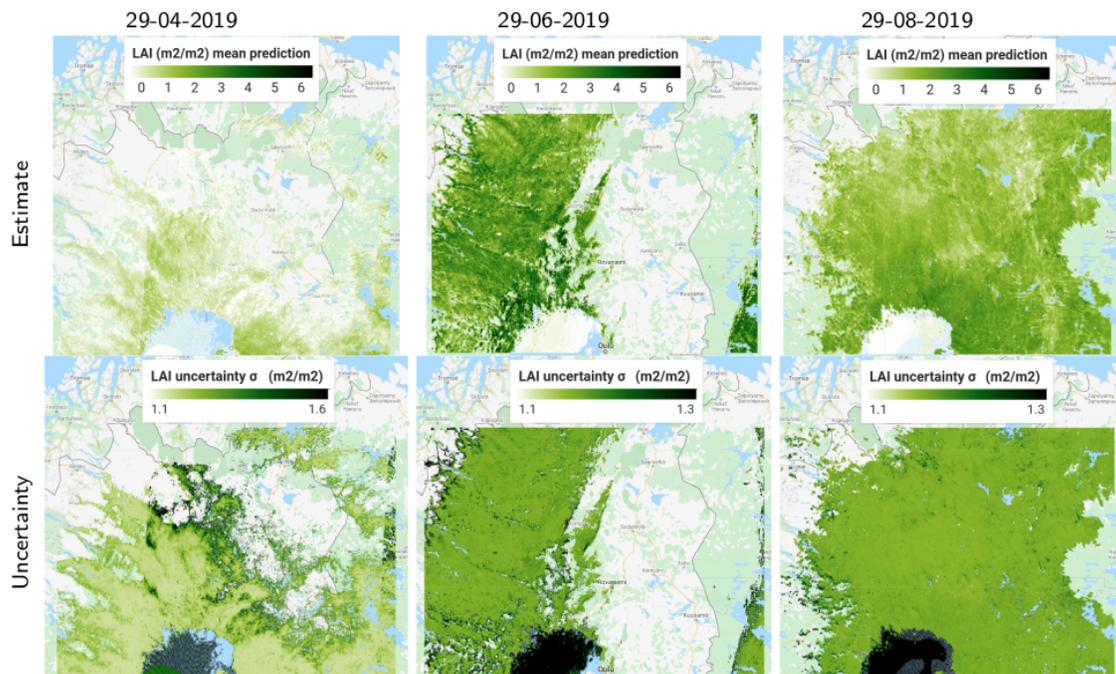


Figure 26: LAI estimates (top) and associated uncertainties (bottom) over Sodankyla for four dates in 2009.

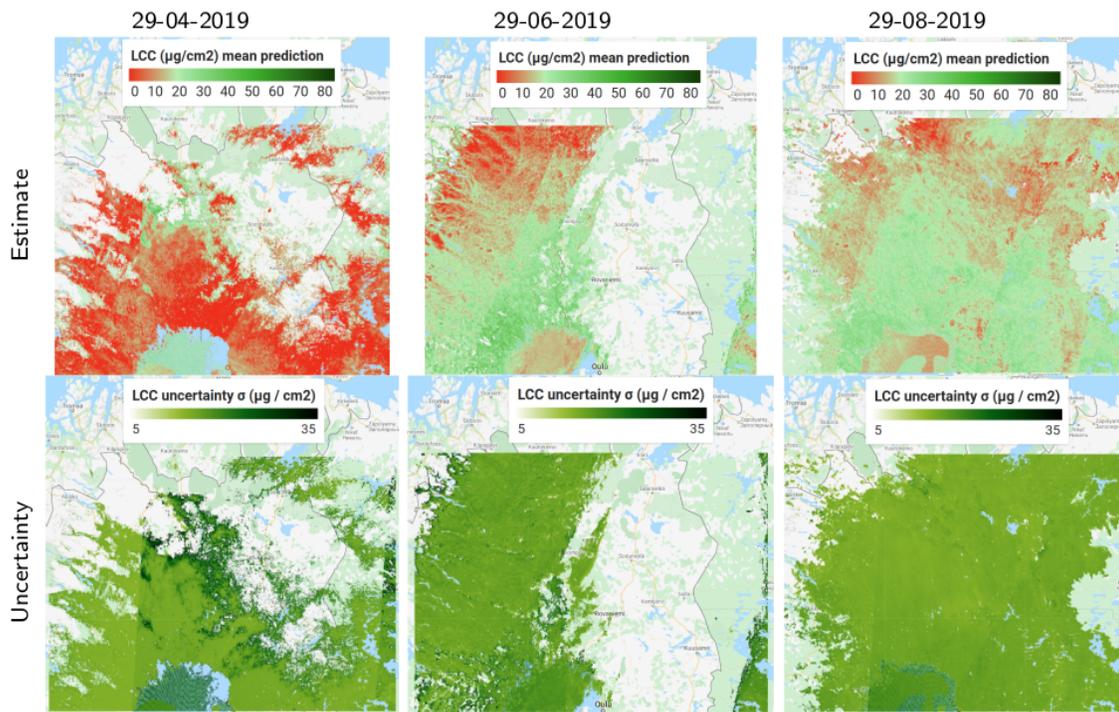


Figure 27: LCC estimates (top) and associated uncertainties (bottom) over Sodankyla for four dates in 2009.

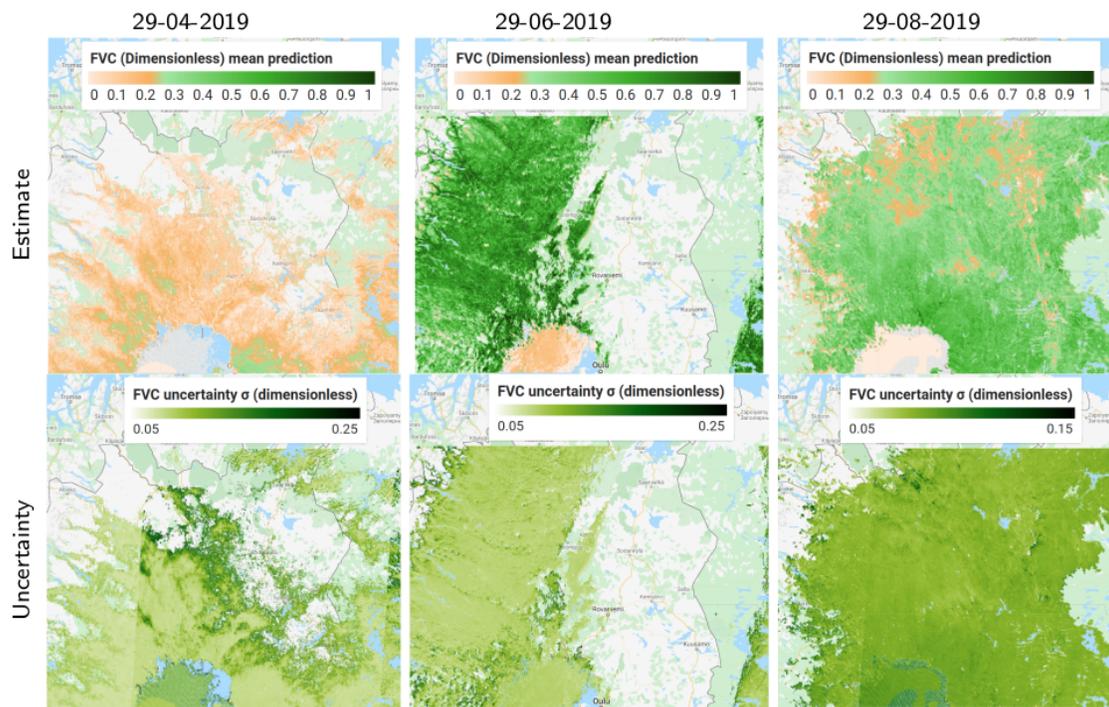


Figure 28: FVC estimates (top) and associated uncertainties (bottom) over Sodankyla for four dates in 2009.

6.4 Data analysis: temporal variability analysis

As a next step, the temporal profiles were plotted for the three test areas. the pursued approach was to plot the average estimate of the 500x500 km along with the average of the associated uncertainties. For all the three sites, it shows the seasonal cycles. The averaged estimates seems to fall within expected range.

For all the three sites, LAI goes along with largest uncertainties, followed by LCC. Hence, the retrievals of these models had the least confidence, suggesting that these models would benefit from improvements in training. The uncertainties of FAPAR and FVC are generally within expected range, yet it is also characterized by substantial outliers.

It must also be noted that for Sodyankyla no meaningful retrievals were extracted during winter time due to darkness. Hence, these time period is gapped.

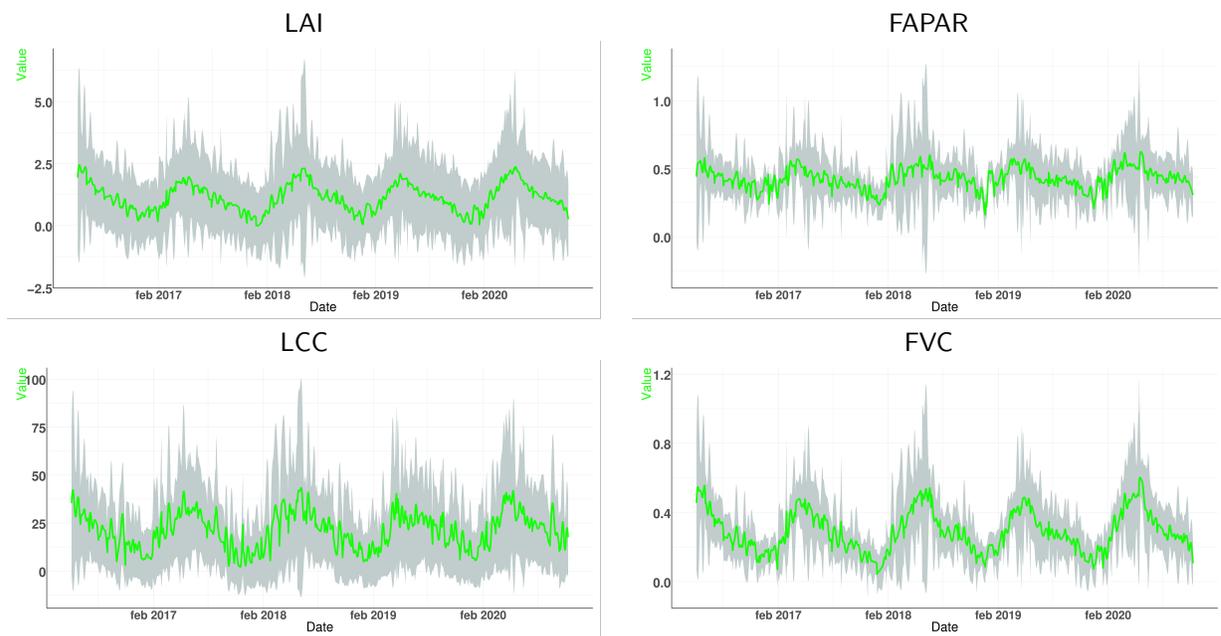


Figure 29: Averaged temporal profiles (green) over Las Majadas area together with associated averaged uncertainties (gray) for LAI, FAPAR, LCC & FVC.

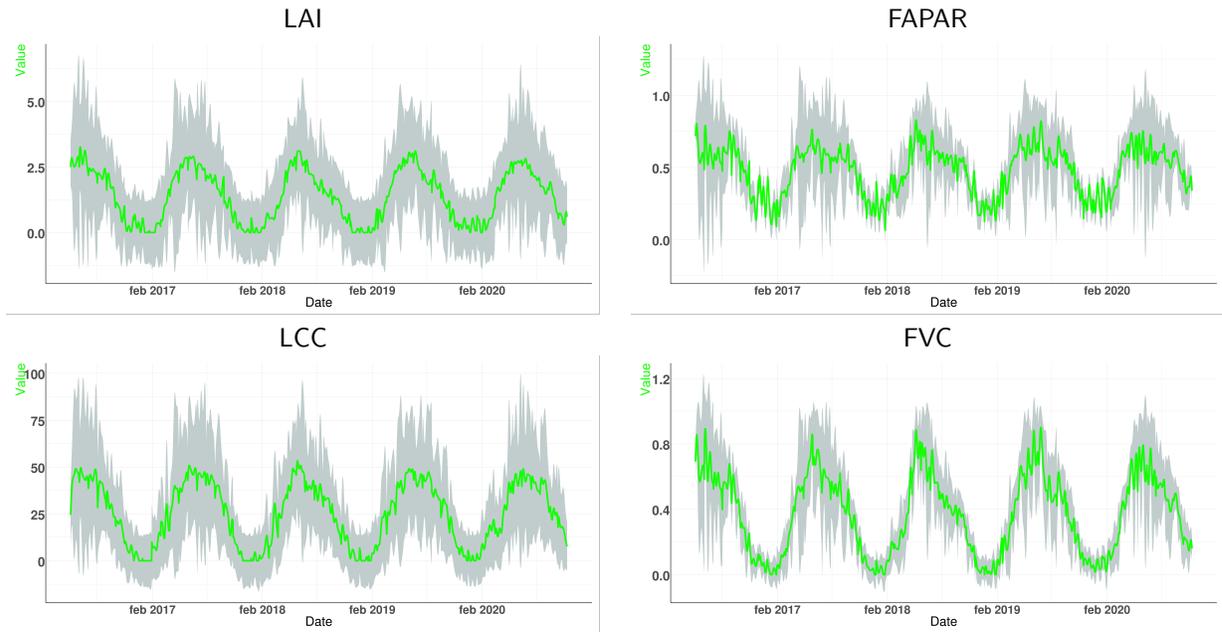


Figure 30: Averaged temporal profiles (green) over Reusel area together with associated averaged uncertainties (gray) for LAI, FAPAR, LCC & FVC.

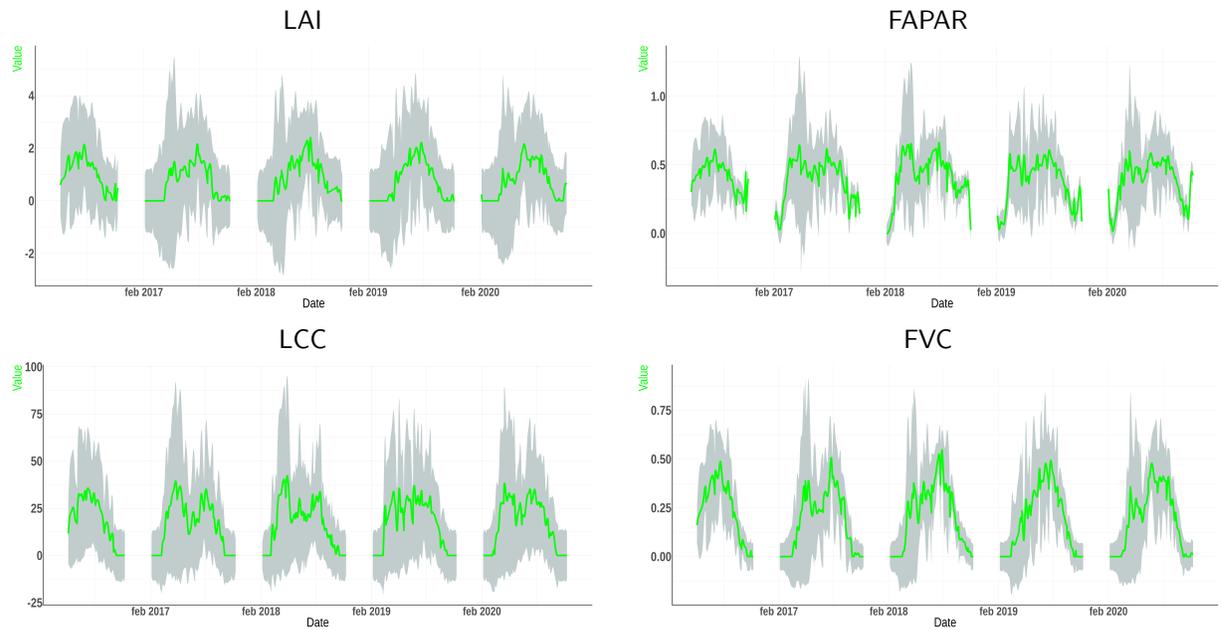


Figure 31: Averaged temporal profiles (green) over Sodankyla area together with associated averaged uncertainties (gray) for LAI, FAPAR, LCC & FVC.

6.5 Conclusions: new estimation of uncertainties

The shown spatial maps and temporal profiles provides insight into the quality of the latest GPR models version, i.e. V1.1. The development of V1.1 was required because with V1.0 we experienced the following artifacts: (1) unrealistic estimates during winter time, (2) poor LAI retrievals, and (3) high uncertainties. Besides, the V1.1 associated uncertainties are more within the expected range. Effectively, thanks to the uncertainties some idea about the quality of the retrievals can be obtained, although it must also be remarked that the delivered uncertainties depend on the interactions between both: (1) the trained model, and (2) the provided observations. If, for some reason these two factors do not match then uncertainties increase. In practice, the uncertainties associated to the vegetation products vary according to different factors: The quality of the model, the quality of images, the land cover, the season of the year, etc. Specifically, because the retrievals are done at the TOA scale, large uncertainties can appear due to atmospheric effects present in any of the bands. Because of the many factors involved, the uncertainties have to be used with care. Commonly, rather relative uncertainties are used, i.e. the SD divided by the prediction and expressed as a percentage (also called: coefficient of variation).

7 SIF data

7.1 Data and associated uncertainties

The solar-induced fluorescence (SIF) satellite data for the project have been collected from the observations of two missions: Nasa Orbiting Carbon Observatory -2 (OCO-2) and the Sentinel 5P Tropospheric Monitoring Instrument (TROPOMI). It should be noted that SIF from both TROPOMI and OCO-2 is not the primary product of the missions, i.e., the missions were not planned to be optimal for retrieving SIF. SIF is retrieved from the near-infrared wavelengths coinciding with the solar Fraunhofer spectral lines. The OCO-2 SIF product analysed is the Level 2 SIF, v10r. The TROPOMI SIF product is the Level 2B TROPOSIF v2.0 product that was received for the project prior to its original publication. The uncertainty estimates for each retrieval are the statistical retrieval uncertainties reported in the official products. The product quality flags or other filtering recommendations (quality flags QF=0 and QF=1 for OCO-2 and cloud fraction < 0.2 for TROPOMI) have been applied while collecting the data to this project. Additional investigations of the uncertainties in the data have been carried out within the LCC project and are described more in Sect.7.2.

7.2 Data analysis: spatial and temporal variability analysis

We present here example results for all of the three research sites and both satellite instruments. TROPOMI should be considered as the primary SIF data source for this project purely due to the better spatiotemporal coverage of the product as compared to OCO-2.

7.2.1 SIF from TROPOMI

The spatial resolution of the TROPOMI SIF pixels is on average $5.5 \text{ km} \times 7 \text{ km}$, and the swath width of the instrument is 2600 km. This allows for a daily global coverage with medium resolution observations, and at higher latitudes (such as in Sodankylä), several overpasses occur during one day.

TROPOMI SIF over Sodankylä is presented in Fig. 32 for three separate days in 2019. It is evident that in late April, SIF signal is very small and close to zero. The product includes either some noise or faint signatures of SIF over Finland, south from Sodankylä. The uncertainties are large as compared to the SIF values, thus realistically reflecting the uncertainty of the product at this time of year, considering large solar zenith angles and other retrieval challenges. Later, towards the end of June, the SIF signal is higher in the vicinity of the site. One TROPOMI orbit from this day has been evaluated as cloudy and thus results in the loss of some data. The uncertainties are comparable to April values, and are thus relatively smaller now that SIF signal is higher, also realistically reflecting the conditions for the retrieval around midsummer. In late August, SIF signal is again weak and appears noisy. Uncertainties are smaller than for the April and June cases, which may be too optimistic and should be re-evaluated. Overall, the TROPOMI product provides good coverage over the Sodankylä site, also seasonally. The seasonal variability is realistic and highlights the short period of photosynthesis in the vicinity of this high-latitude site.

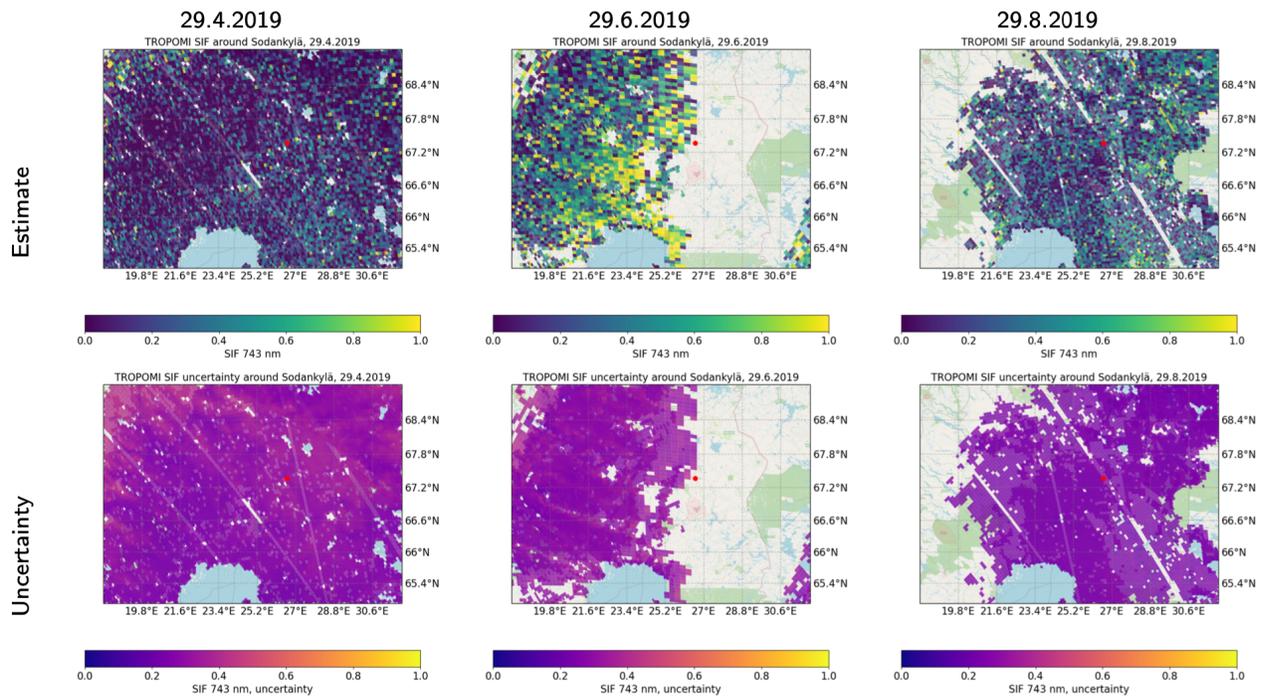


Figure 32: TROPOMI SIF estimates over Sodankylä for three days in 2019.

TROPOMI SIF over Reusel is presented in Fig. 33 for three separate days in 2019. In late March, it appears that photosynthesis is ongoing in the vicinity of the site. SIF signal shows some spatial variability but the uncertainties do not vary equally. Some cloud contamination can be seen. In early July, SIF shows expected high values, and spatial features in the signal can be seen. Again, uncertainties do not show similar variability and are relatively small. In the September case, cloud contamination has resulted in the filtering of most retrievals; yet, high SIF values can be distinguished between the patchy cloud fields, surprisingly even higher than in July (see also the TROPOMI SIF time series which shows two seasonal maxima for Reusel in Fig. 39). The uncertainties do not fully reflect this variability and should be evaluated using additional methods.

TROPOMI SIF over Las Majadas is presented in Fig. 34 for four separate days in 2019. Each day is affected by fractional cloud coverage, causing partial lack of data. In March, the site and its neighboring areas show high SIF signatures, and in the immediate vicinity of the site, this is the highest signature of all seasons depicted in Fig. 34. In July, SIF signal is surprisingly lower than in March near the site but shows high values north from the site, which are likely to affect the time series from the entire area of interest where the SIF maximum is identified around May (see, e.g., Fig. 39). Uncertainties are somewhat higher in July than in March and more uniform than the spatial distribution of SIF signal. September and December SIF resemble each other spatially, except that in the north, the values are lower in December whereas south from the site, December SIF values are higher than in September. The uncertainties are overall higher in September than in

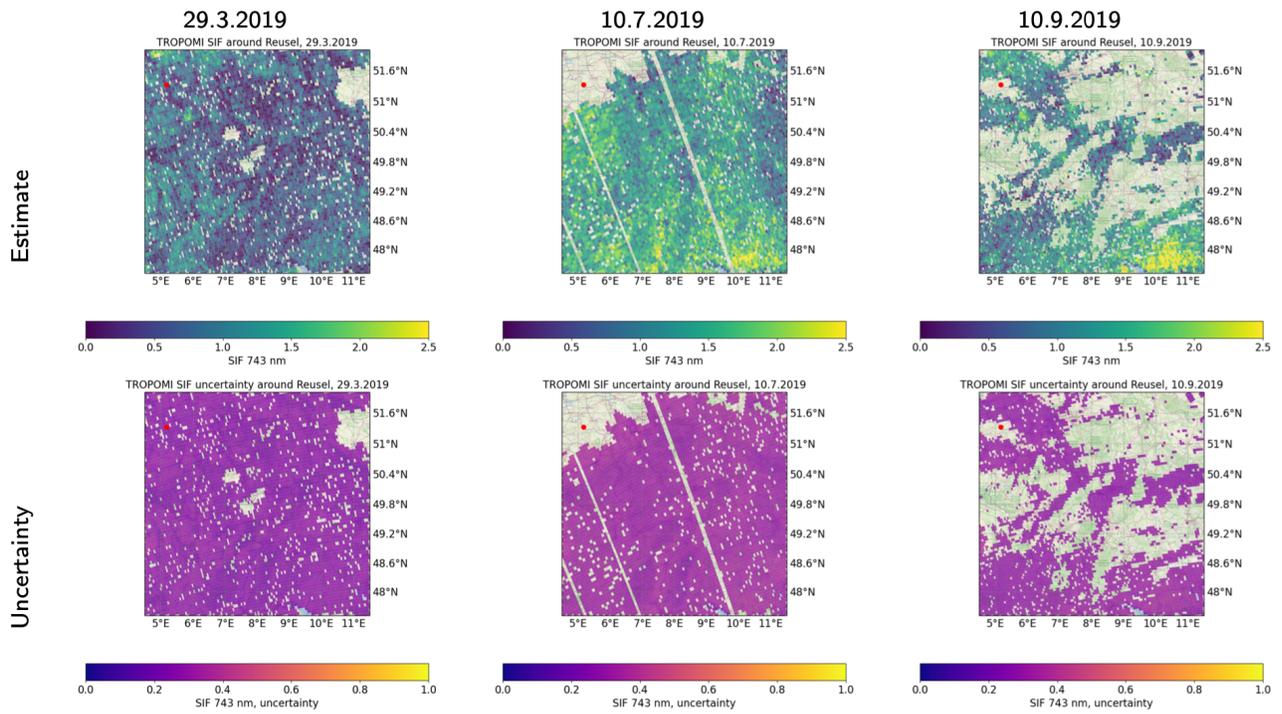


Figure 33: TROPOMI SIF estimates over Reusel for three days in 2019.

December.

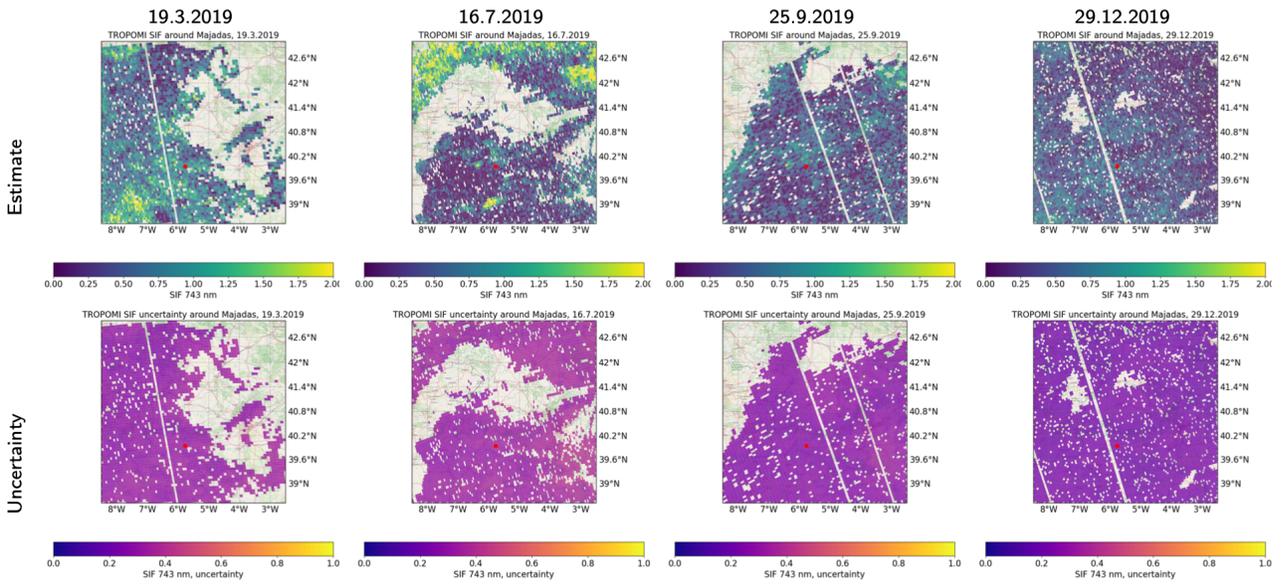


Figure 34: TROPOMI SIF estimates over Las Majadas for four days in 2019.

7.2.2 SIF from OCO-2

The resolution of the OCO-2 SIF pixels is on average $1.3 \text{ km} \times 2.3 \text{ km}$, and the swath width of the instrument is about 10 km. The narrow swath combined with the kilometer-scale pixel size lead to a revisit time of approximately 16 days. In a spatial analysis, the narrow swath can only be used to support observations with a better spatial coverage, in this case the TROPOMI SIF product. For OCO-2, SIF is retrieved from the wavelength bands at 757 nm and 771 nm, and based on these, also estimated computationally at 740 nm. Time series of these three SIF estimates are shown later in Fig. 40. The maps shown in this section are based on OCO-2 SIF estimate at 740 nm, which is comparable to the TROPOMI SIF.

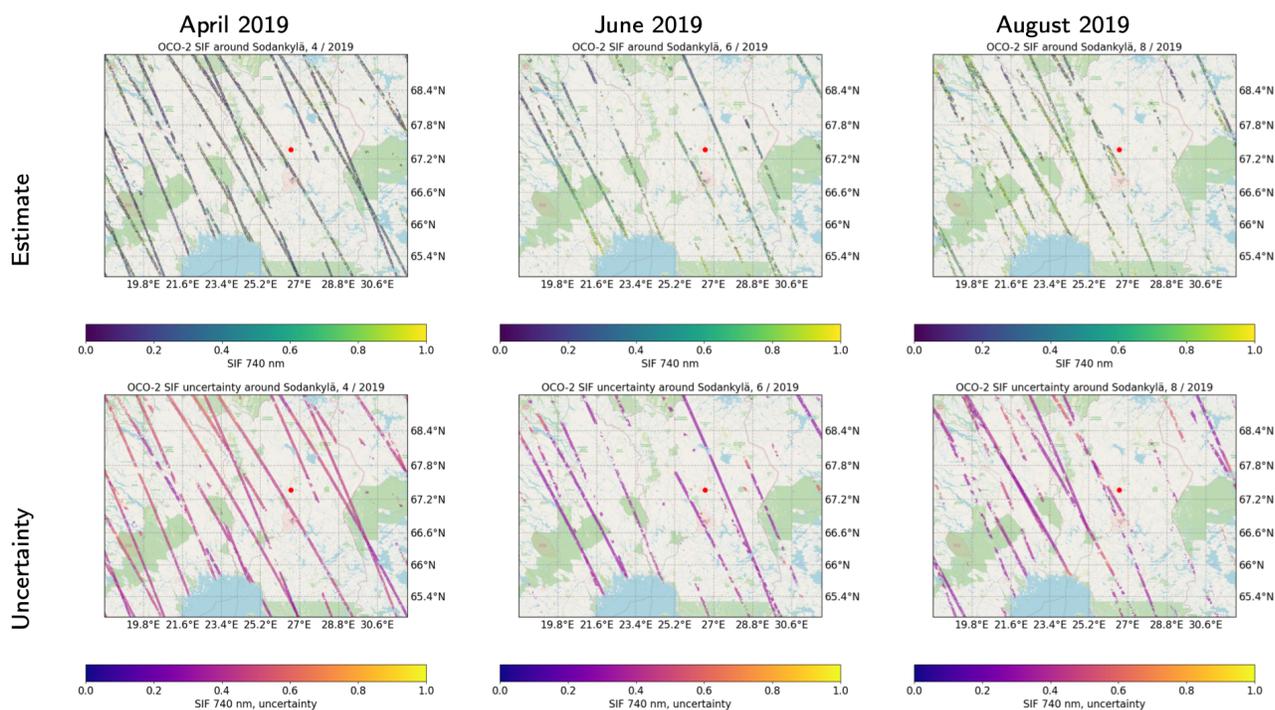


Figure 35: OCO-2 SIF estimates over Sodankylä for three months in 2019.

7.2.3 Comparison of the TROPOMI SIF products at 743 nm and 735 nm

Because the TROPOMI SIF product was only recently released, initial product evaluation is important. We carried out a comparison of TROPOMI all-sky (based on 743 nm – 758 nm retrieval window) and clear-sky (based on 735 nm – 758 nm retrieval window; more experimental) products at the sites of interest. The figures 38 and 39 show that the 735 nm product is mostly systematically lower in SIF estimate than the 743 nm product. The difference is not a constant bias but shows seasonal variability, especially at Reusel. The reason for the seasonally-varying difference is currently not known. These results have been reported to the TROPOMI team and their recommendation

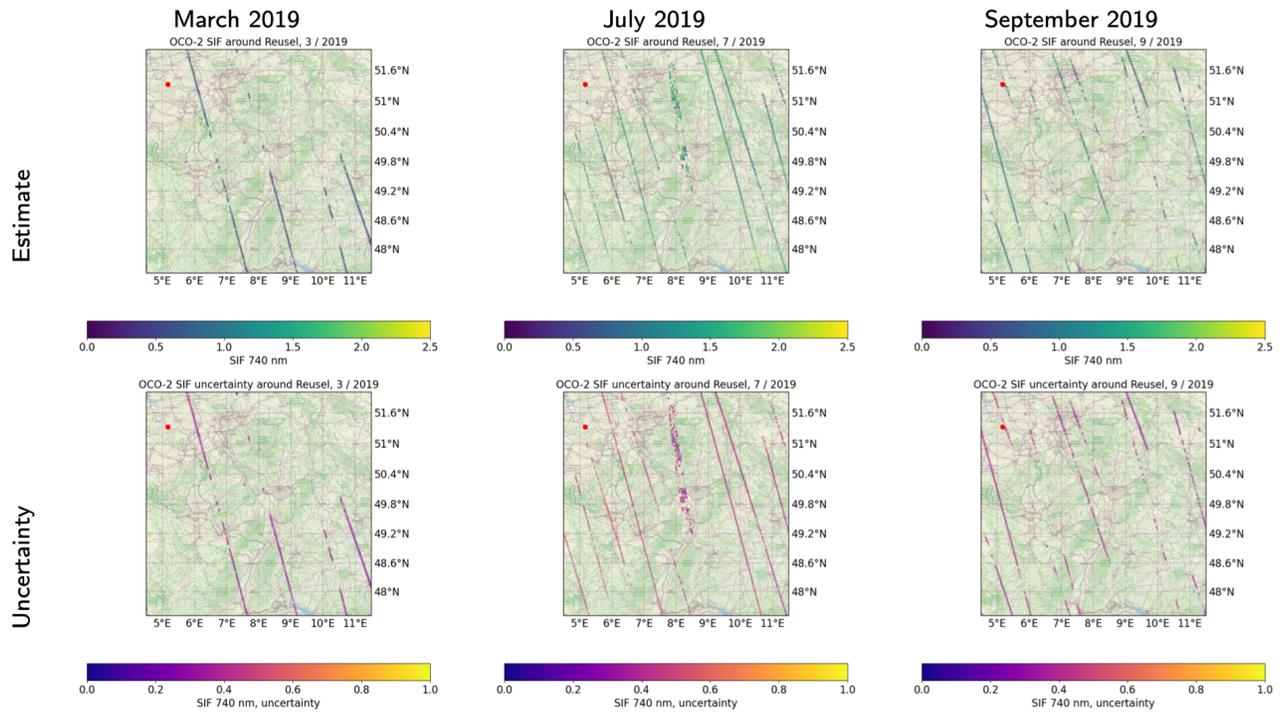


Figure 36: OCO-2 SIF estimates over Reusel for three months in 2019.

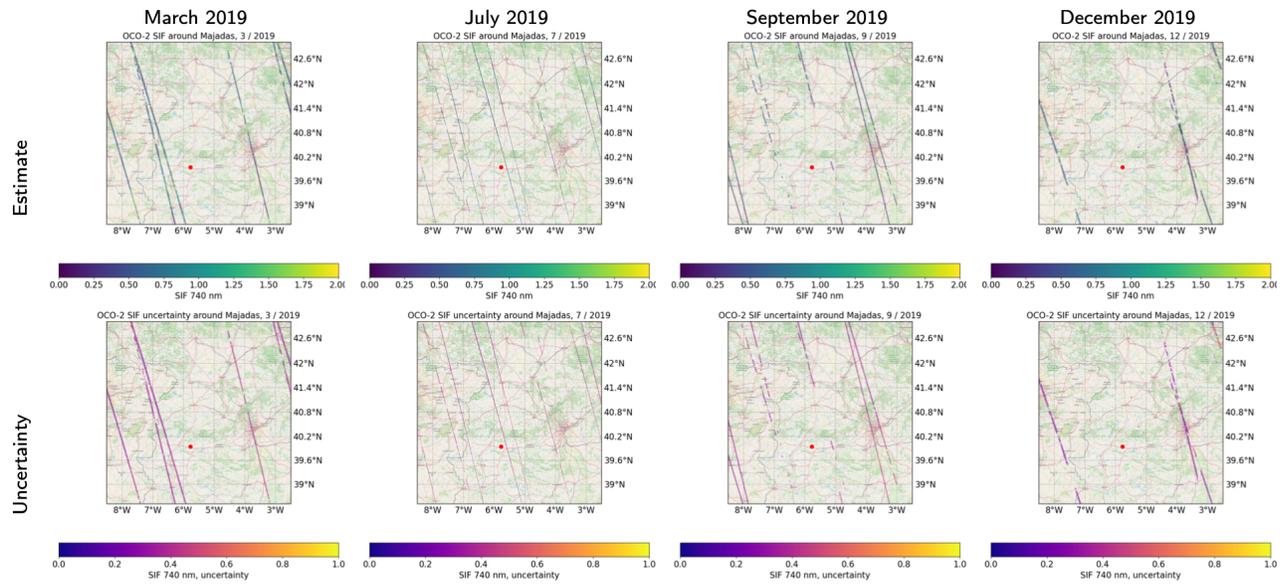


Figure 37: OCO-2 SIF estimates over Majadas for four months in 2019.

(also given in Guanter et al. (2021)) is that the 743 nm product should be preferred until further studies have improved and confirmed the usability of the 735 nm product.

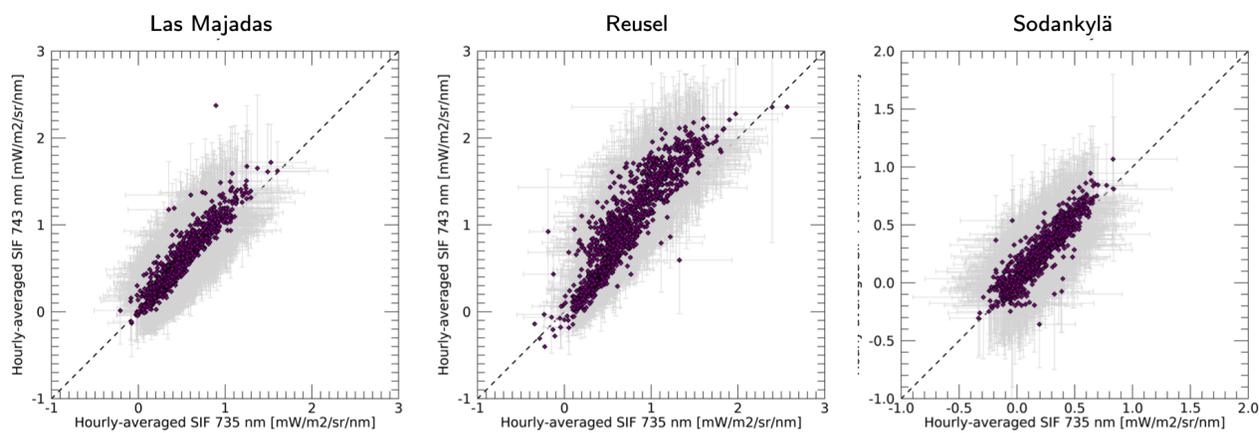


Figure 38: One-to-one plots of two TROPOMI SIF products: the all-sky product (743 nm) and the clear-sky product (735 nm).

7.2.4 Comparison of the OCO-2 SIF products at 740 nm, 757 nm, and 771 nm

The spectrum of solar-induced fluorescence is wavelength-dependent, and therefore the retrieval window affects the retrieved SIF value. OCO-2 data can be used for SIF retrievals at the wavelengths of 757 nm and 771 nm. Based on these two retrieved values, a computational value at 740 nm is estimated, which can then be used in comparisons to, e.g., TROPOMI SIF retrievals. Figure 40 shows the time series of these OCO-2 SIF values. Based on the SIF spectral shape, it is expected that the 771 nm band produces the lowest SIF values, as can be seen from the figure.

7.2.5 Analysis of the temporal variability and uncertainties

The spatial distribution of SIF observations shows the significant amount of noise present in the individual observations of each satellite overpass, and also that this noise is not present in the individual SIF uncertainty estimates. Therefore, we explore the temporal variability in all three focus regions. Figure 41 shows the large variability in the single observations, and how this noise is brought down to a reasonable level with hourly averaging. The figure shows as error bars the standard deviations computed as the one-sigma deviation of all SIF values within a given hour from the one-hour average SIF. In this format, the interpretation of the seasonal cycle of photosynthesis is easier: the start and end timing as well as the annual maximum and its timing can be determined from the time series. This also highlights the importance of not filtering the seemingly unrealistic, negative SIF retrievals, which in a statistical sense are reasonable retrieval results.

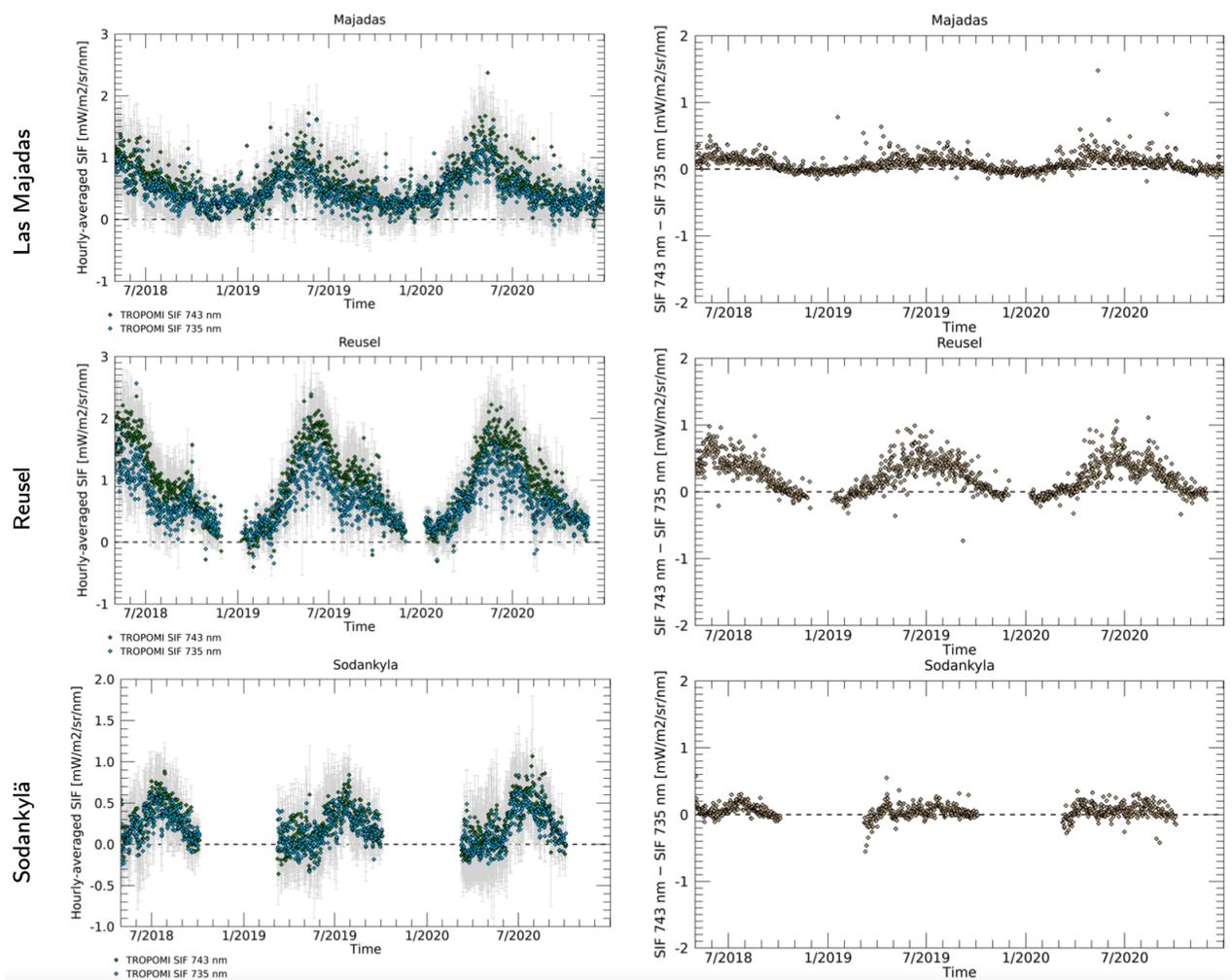


Figure 39: Time series of the two TROPOMI SIF products: the all-sky product (743 nm) and the clear-sky product (735 nm) (left) and their difference (right).

7.3 Conclusions: new estimation of uncertainties

The source of uncertainty from the observations is included in the SIF datasets as the statistical retrieval uncertainty. The uncertainty is overall rather low and varying only modestly with the location or season. The uncertainty estimate was found to not reflect the true spatial or temporal variability of the products. The noise in the observations was significantly reduced by carrying out hourly averaging. From these hourly-averaged data, the interpretation of the start of photosynthesis, the photosynthesis maximum as well as its timing, and the end of photosynthesis is more reliable than from the instantaneous SIF data. As the uncertainty, it is recommended to use the standard deviation of the hourly-averaged SIF.

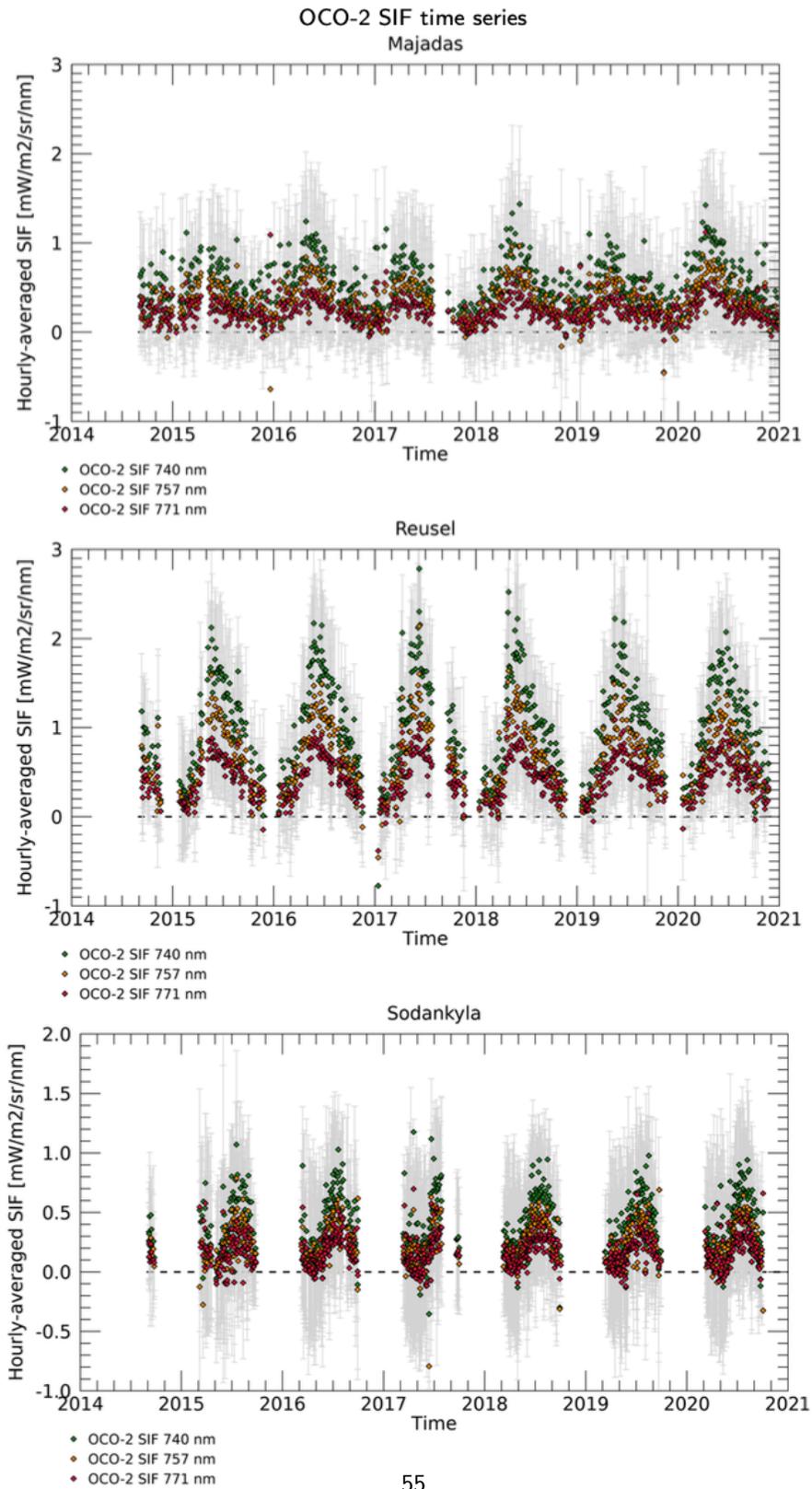


Figure 40: Time series of the OCO-2 SIF retrieval products at 757 nm and 771 nm, and the computational estimate at 740 nm.

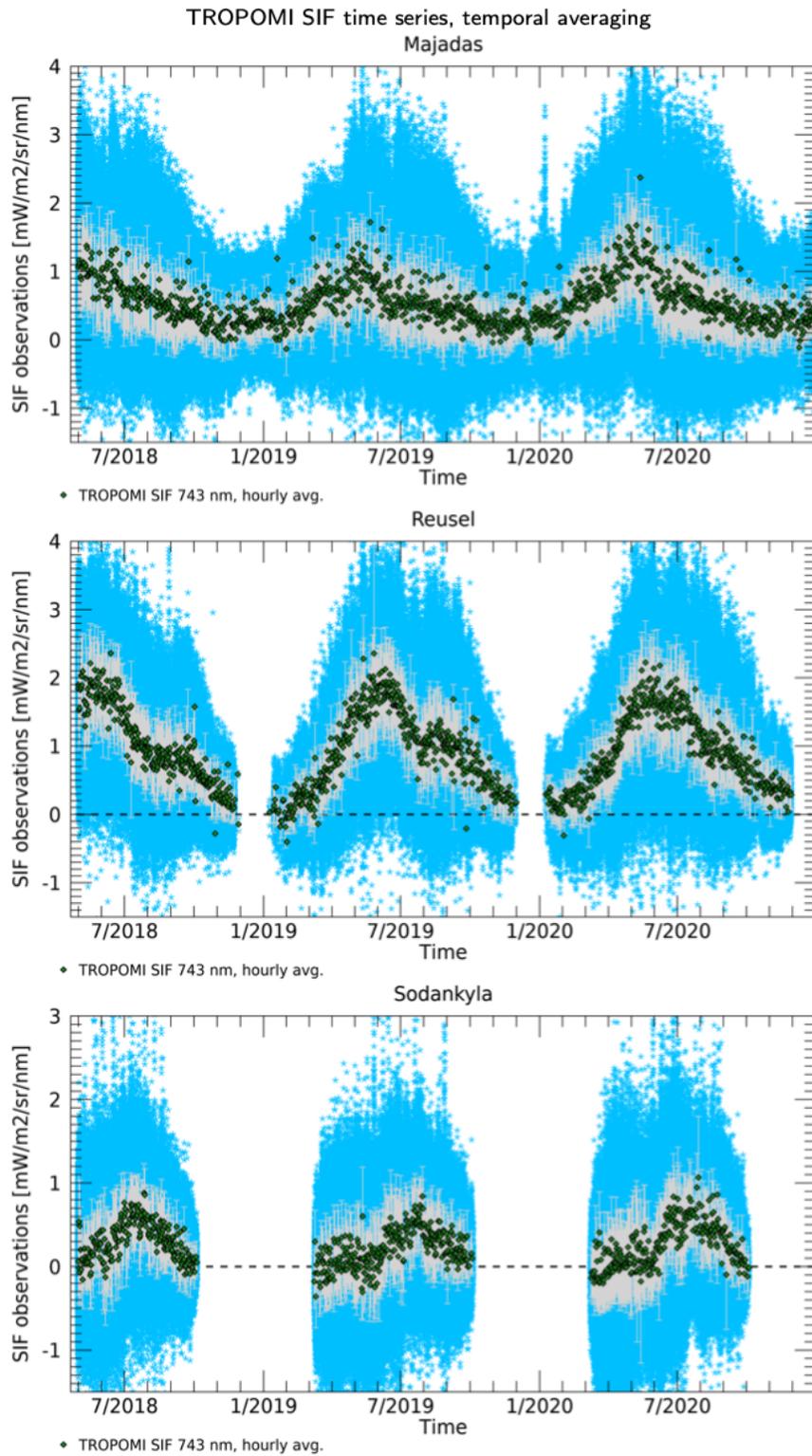


Figure 41: Time series of TROPOMI SIF presented for all of the three regions. Both individual SIF observations and their hourly averages are shown. The error bars correspond to the standard deviation from the hourly average.

8 AGB data

AGB data were provided by the CCI Biomass project using several sensors. Three dates are provided: 2010, 2017 and 2018. The 2010 maps were produced using Envisat SAR and ALOS PALSAR. The 2017 and 2018 maps were produced using a similar approach but Sentinel-1 is used instead of Envisat ASAR and ALOS-2 PALSAR-2 is used instead of ALOS PALSAR. AGB data resolution is 100m and their unit is Mg.ha⁻¹. Fig. 42 shows that AGB is distributed in vary various ways depending on the region. It also shows that the data-set also contain probable outliers. Therefore, it will be recommended to filter out data with AGB value over 200 Mg.ha⁻¹ in Las Majadas, 125 Mg ha⁻¹ in Sodankyla and 350 Mg ha⁻¹ in Reusel.

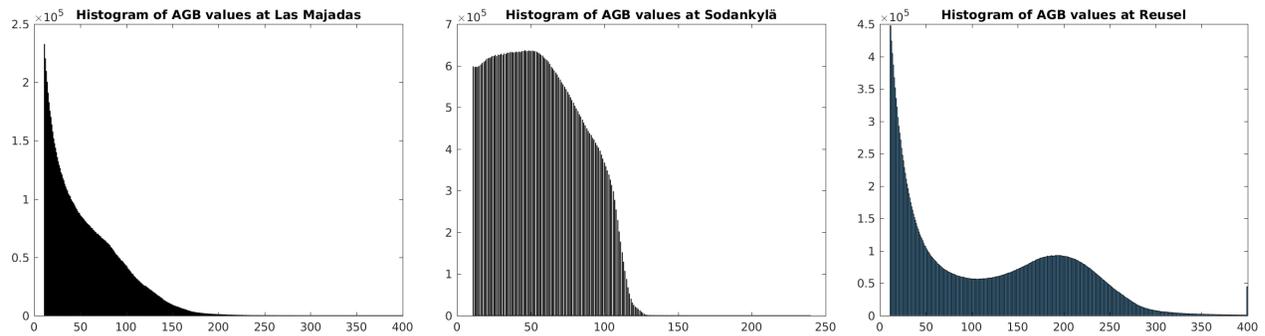


Figure 42: Histograms of AGB CCI 2018 values. For a better scale visibility values between 0 to 10 have been removed. From left to right : Las Majadas, Sodankyla, Reusel

Fig. 43 shows AGB value on the vertical axis and VOD value on horizontal axis for every grid points of a region. According to this figure, AGB data seems to be very well correlated to SMOS VOD. Fig. 44 shows the root-mean-square deviation of the difference between AGB and VOD data. Comparing this figure to Fig. 45 it appears that the area where we measure the highest difference between AGB and VOD are those with the denser vegetation.

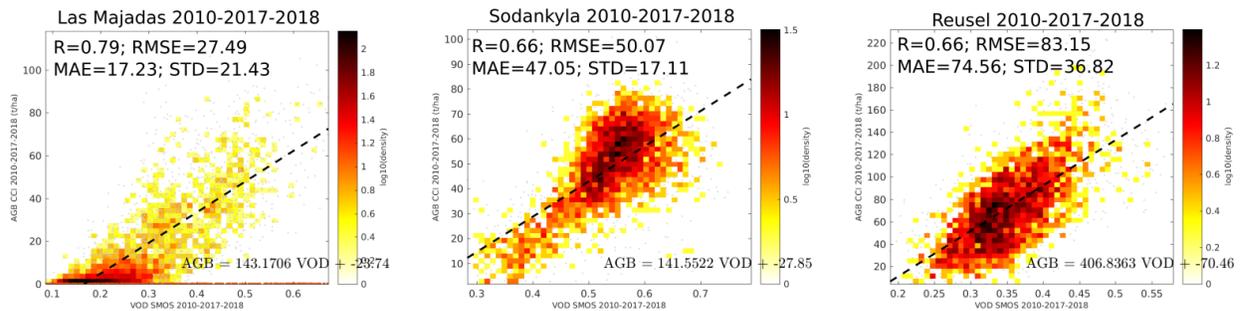


Figure 43: Scatter-plot of AGB CCI 2010-2017-2018 values compared to VOD SMOS values (2011-2017-2018). From left to right : Las Majadas(R=0.79), Sodankyla(R=0.66), Reusel(R=0.66). R : Correlation Coefficient

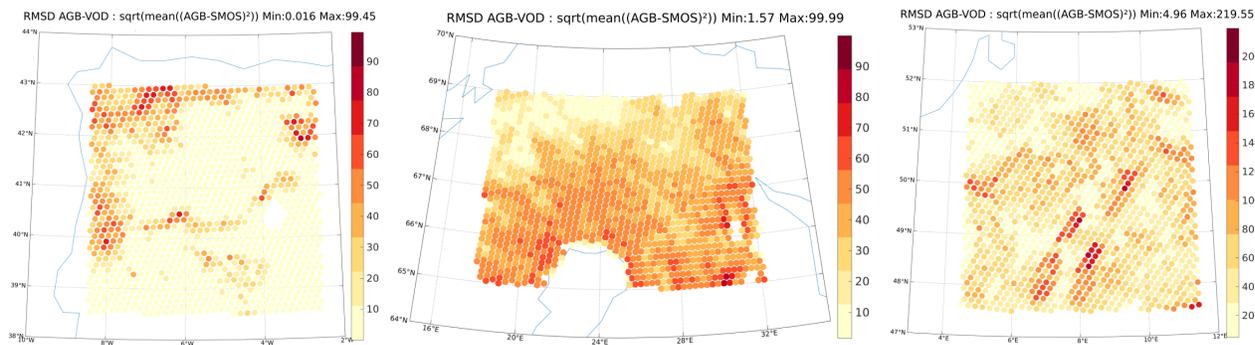


Figure 44: Root-mean-square deviation of AGB CCI 2010-2017-2018 values compared to VOD SMOS values (2011-2017-2018). From left to right : Las Majadas, Sodankyla, Reusel.

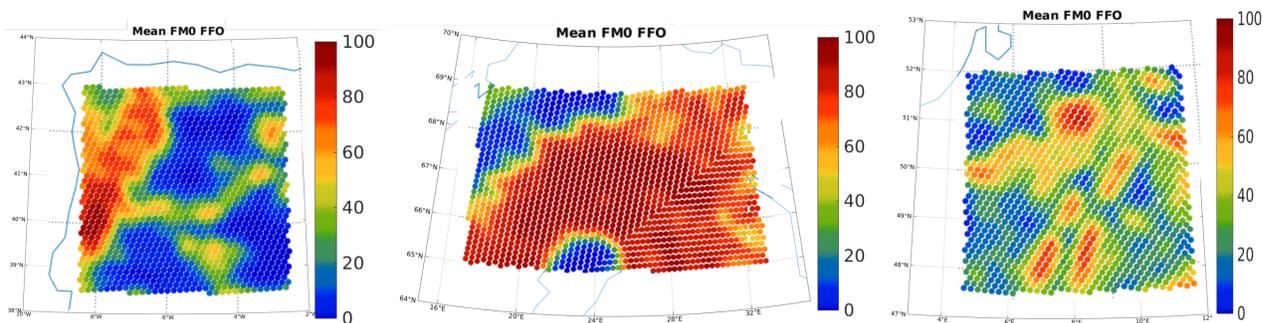


Figure 45: Forest fraction of land cover from SMOS auxiliary data. From left to right : Las Majadas, Sodankyla, Reusel.

	RFI-Prob	VOD DQX	Chi2
Las Majadas	0.23	0.15	2.4
Sodankyla	0.24	0.31	2.2
Reusel	0.39	0.20	3.1

Table 2: Threshold values of RFI-Prob, VOD DQX and Chi2 when removing the 5 % highest values of those parameters in each region.

9 Vegetation-related variables and their uncertainty. Cross-comparison

9.1 Microwave optical depth

In order to compare the data properly, a set of filtering criteria was defined and applied to SMOS, ASCAT and AMSR-2 VOD data sets. The data meeting one of the next conditions were removed for the following comparison:

ASCAT

$$\frac{VOD_{Noise}}{VOD} \geq 0.02$$

$$Proc.Flag > 0$$

AMSR-2 One of the following flag is up : No data, Snow/Frozen, RFI, No band retrieval, No model convergence, Low quality.

SMOS The distributions of RFI-Prob, VOD DQX and Chi2 was analyzed and the 5% highest values were removed. The actual threshold values are given in Table 2.

ASCAT, AMSR-2, SMOS

$$VOD \leq 0$$

For the comparison between the different data sets, the data have been projected in the grid with the lowest resolution. Therefore, ASCAT data were projected into the SMOS grid for the ASCAT-SMOS comparison, and SMOS data were projected into the AMSR-2 grid for the AMSR2-SMOS comparison.

9.1.1 VOD comparison ASCAT - SMOS

Figure 46 shows the SMOS VOD and ASCAT VOD temporal averages for the three regions. It also shows their difference or bias (Fig. 46 a, d, g,) computed as the mean of ASCAT VOD minus the mean of SMOS VOD. The bias is not randomly distributed over the map, but follows distinctive geographical patterns. Regions with negative and regions with positive bias were found for the three regions. A preliminar comparison of the bias maps with MODIS Land Cover maps (Fig. 47), shows that there might be a correlation between some classes of land cover and the sign of the bias. In

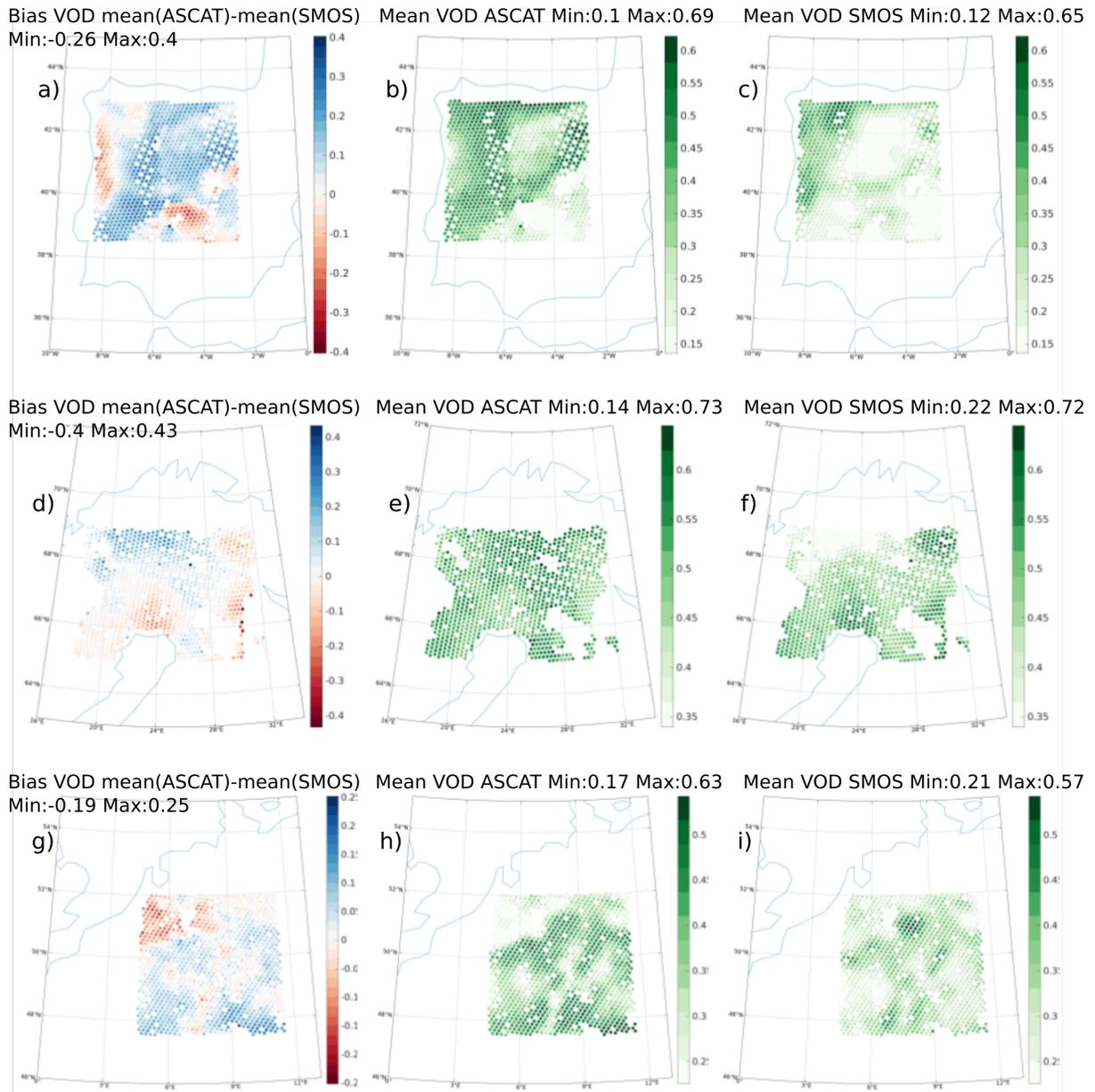


Figure 46: From left to right: (i) VOD bias between ASCAT and SMOS. (ii) Mean of ASCAT VOD from 2011 to 2020. (iii) Mean of SMOS VOD from 2011 to 2020. The results are presented for the three regions, from top to bottom: Las Majadas, Sodankyla, Reusel. Data used for these results have been previously filtered.

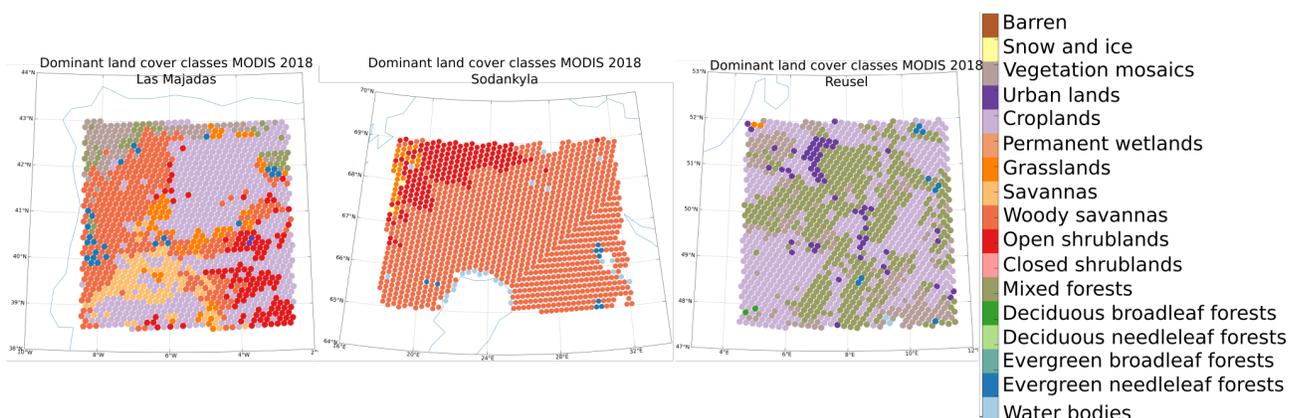


Figure 47: MODIS 2014 Land cover IGBP classification. From left to right: Las Majadas, Sodankyla, Reusel.

Sodankyla, ASCAT VOD is higher on average than SMOS VOD in the Northwest when the land cover goes from woody savannas to open shrublands. In Las Majadas, ASCAT VOD is higher than SMOS VOD for most of the region but a negative bias is found in the West and the South-East, in regions mostly classified as woody savannas, in agreement with Sodankyla. However, there are also regions classified as woody savannas where ASCAT VOD is higher than SMOS VOD. In Reusel, lower ASCAT VOD than SMOS VOD are found in the North-West, where the land cover is a mix of croplands and vegetation mosaics. In contrast, ASCAT VOD is in general higher than SMOS VOD in other regions with croplands.

The temporal correlation between VOD ASCAT and VOD SMOS data was studied by computing Pearson correlations (Fig. 48). The absolute values of the correlation are lower than 0.4 for most of the regions. The correlation is mostly positive at Reusel and negative at Las Majadas, while there are both negative and positive correlation regions in Sodankyla. A more quantitative analysis as a function of the land cover should be done to get more insight into the bias and correlations maps to fully understand SMOS and ASCAT VOD.

The possible correlations of the uncertainties of both data sets were investigated by computing the Pearson correlation of SMOS VOD DQX and ASCAT VOD-Noise parameters (Fig. 49). No significant correlation was found, as expected from independent observations with different instruments and retrievals performed with different algorithms.

9.1.2 Comparison AMSR-2 - SMOS

SMOS VOD was also compared to AMSR-2 VOD. Figure 50 shows the bias for AMSR-2 VOD estimations at bands C1, C2 and X with respect to SMOS L-band VOD. Similar results were obtained for the three AMSR-2 bands. In contrast, AMSR-2 VOD is higher than SMOS VOD in Reusel, lower in Sodankyla, while in Las Majadas it is higher except in the north western region. Fig. 51 shows that the correlation coefficient between SMOS and AMSR-2 VOD is negative and low in the north of Las Majadas while it is positive and high in the south, like for the bias from the previous figure. For Sodankyla it is the opposite, we have a low positive correlation in the North and a low negative

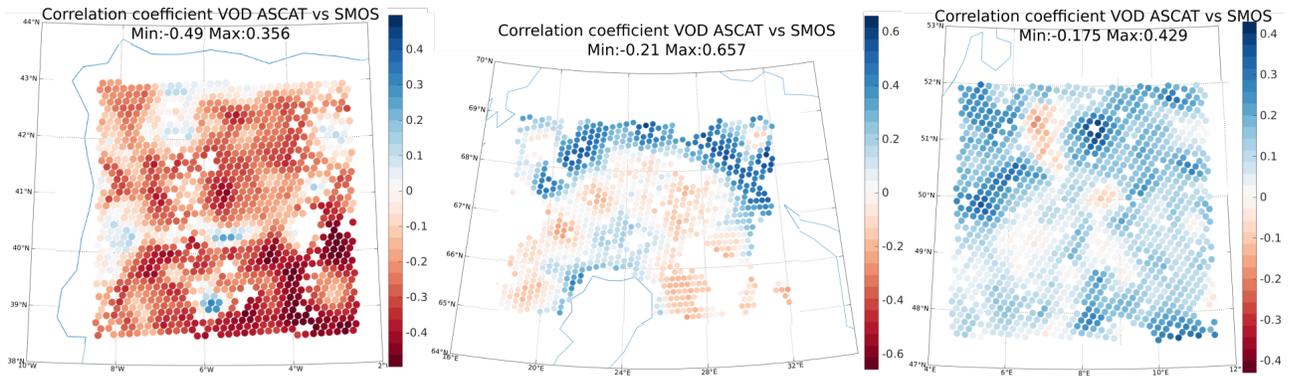


Figure 48: Maps of correlation coefficient between VOD values of ASCAT and SMOS, from left to right: (i) Las Majadas. (ii) Sodankyla. (iii) Reusel. The results are using data from 2011 to 2020 of ASCAT and SMOS

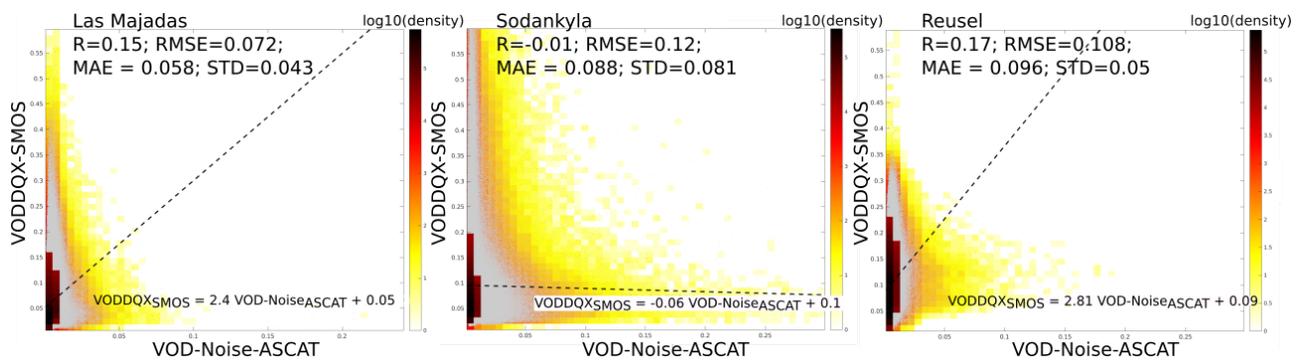


Figure 49: Scatter plot of SMOS VOD DQX in function of ASCAT VOD-Noise. Correlation coefficient and linear regression. From left to right : Las Majadas, Sodankyla, Reusel

in the South. For Reusel, the correlation is always low either when it is positive or negative.

9.2 Optical indices and SIF

9.2.1 Temporal correlation SIF – FAPAR

The temporal relation of solar-induced chlorophyll fluorescence (SIF) and the fraction of absorbed photosynthetically active radiation (FAPAR) was explored for the three focus areas. First, we aggregated the observations daily for all pixels within the focus regions. Then, we co-located both aggregated data sets by using same-day observations as the co-location criterion. Figure 52 shows the time series of the co-located observations at all three sites. Co-location is temporally limited to the years 2018, 2019 and 2020 due to the availability of TROPOMI data.

In Sodankylä, the correlation of SIF and FAPAR is 0.59, the lowest of all three sites. The time series shows an interesting detail in spring: FAPAR values peak already in March while SIF values are still close to zero. According to SIF, the start of photosynthesis occurs in late May or June. Despite their different temporal changes in spring, both SIF and FAPAR start to decrease in August and appear coupled during the autumn. The reasons for these discrepancies can be speculated. First of all, evergreen boreal forests do not change reflectance according to seasons contrary to deciduous forests. In spring, the pine tree needles appear green although photosynthesis is not yet possible due to the frozen ground. The time series also includes the early and late winter months when SIF retrievals were not possible due to the limited amount of solar radiation and large solar zenith angles. The FAPAR estimates during these time periods may also be affected by these challenges and therefore appear unrealistic (e.g. the increase in FAPAR towards November may not be realistic). In spring, it may also be possible that snow-covered surface induces bias in the FAPAR estimate, although bright pixels have been filtered in earlier data processing.

In Reusel, the correlation of SIF and FAPAR is 0.81, which is the highest of the three sites. The time series shows coupled variability at all seasons. An interesting detail is the double peak in SIF: the first and most significant maximum in June and another local maximum, although not as pronounced, in August-September. This appears also in FAPAR, although not as clearly. Possible reasons could be the land and crops management: e.g., harvest and irrigation.

In Las Majadas, the correlation of SIF and FAPAR is 0.71. The time series shows coupled variability at most seasons, except during late summer, FAPAR remains high although SIF decreases. This may be caused by limited water availability for photosynthesis which is seen in the decrease of SIF but, for the dominant vegetation type in the region, FAPAR may be a less sensitive indicator of drought.

To conclude, the temporal correlation of SIF and FAPAR was found to vary between the focus areas. The needleleaf forest dominated regions were found to have the lowest correlation. The coupling between SIF and FAPAR could be studied further in more detail and over smaller areas around the measurement sites to better link and scale from the satellite-scale to the campaign measurements data and to the underlying processes driving the coupling or decoupling.

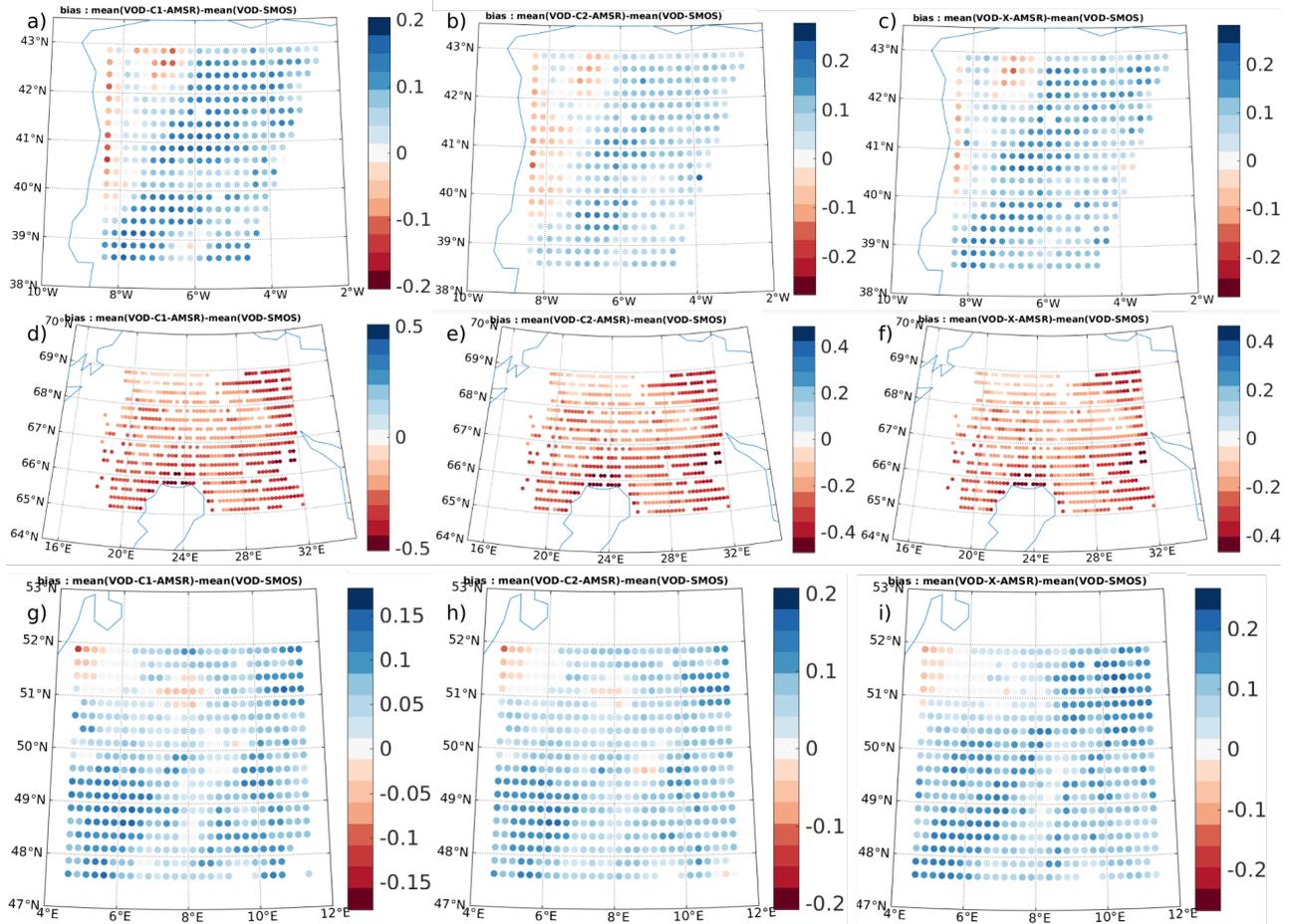


Figure 50: From left to right: (i) VOD bias between AMSR-2 C1 band and SMOS. (ii) VOD bias between AMSR-2 C2 band and SMOS. (iii) VOD bias between AMSR-2 X band and SMOS. The results are presented for the three regions, from top to bottom: Las Majadas, Sodankyla, Reusel. Data used for these results have been previously filtered. Bias are the difference between the VOD mean value from 2012 to 2020 for AMSR-2 and from 2011 to 2020 for SMOS.

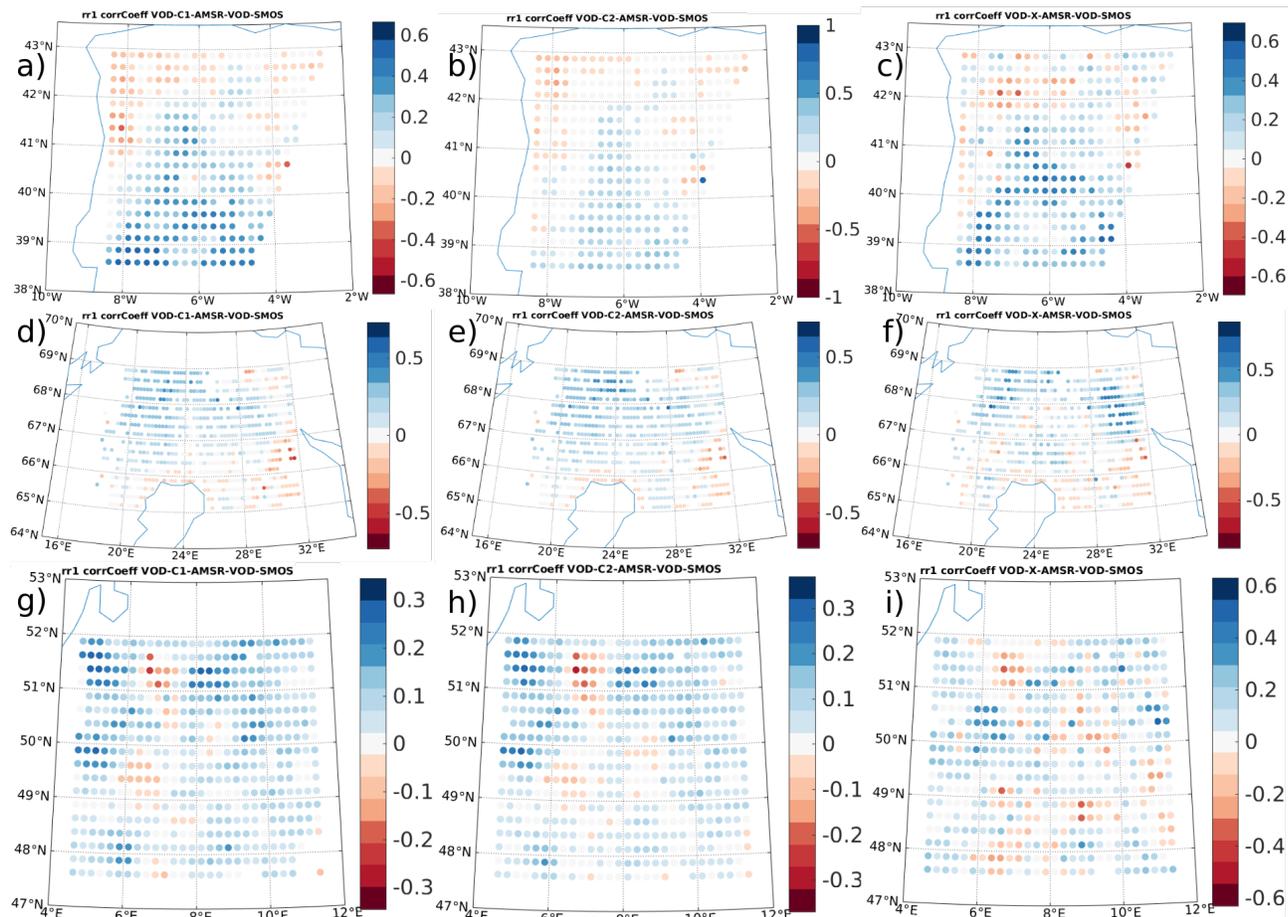


Figure 51: From left to right: (i) Correlation coefficients of VOD between AMSR-2 C1 band and SMOS. (ii) Correlation coefficients of VOD between AMSR-2 C2 band and SMOS. (iii) Correlation coefficients of VOD between AMSR-2 X band and SMOS. The results are presented for the three regions, from top to bottom: Las Majadas, Sodankyla, Reusel. Data used for these results have been previously filtered. Coefficients have been calculated using VOD value from 2012 to 2020 for AMSR-2 and from 2011 to 2020 for SMOS.

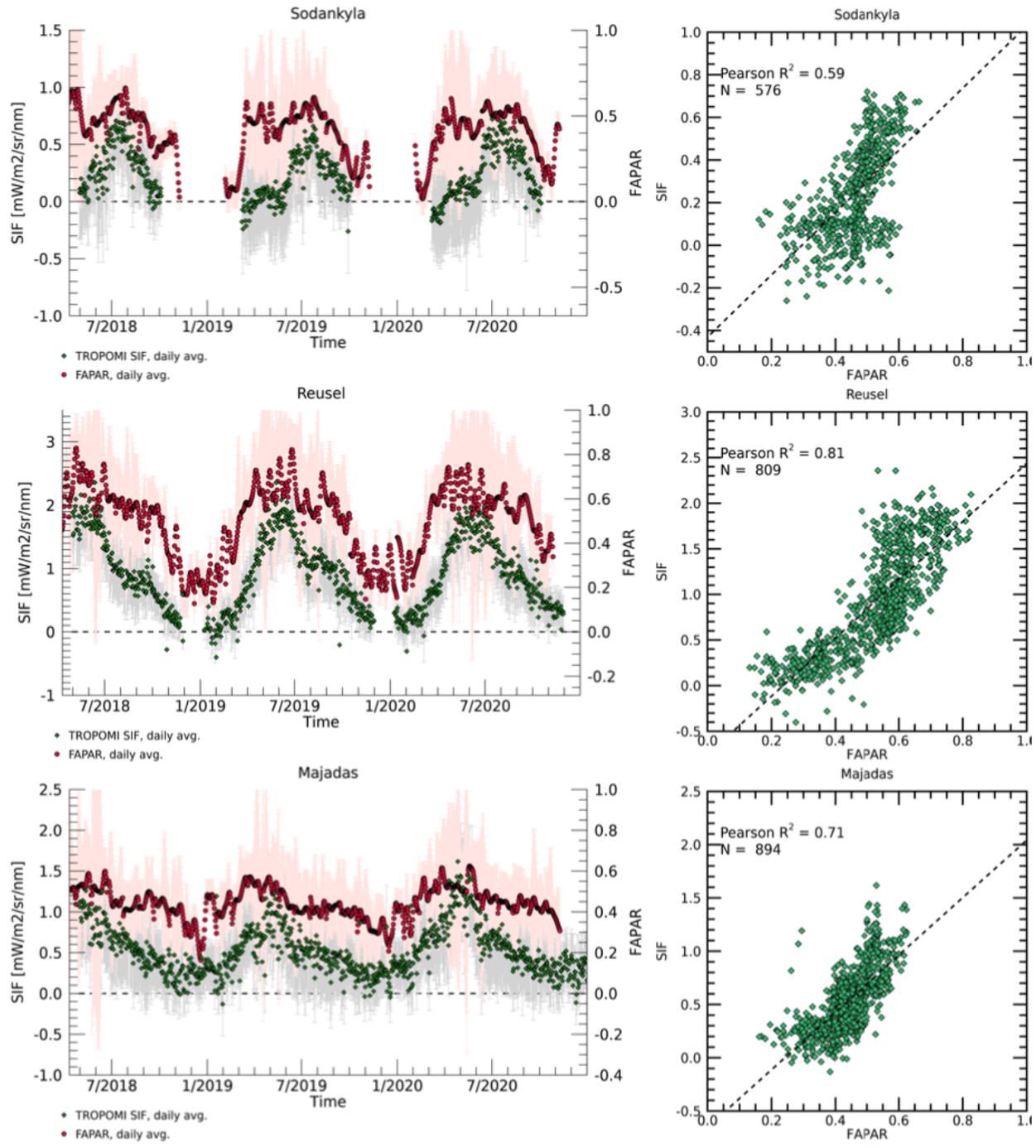


Figure 52: Time series (left panel) and the correlation (right panel) of the TROPOMI SIF and OLCI FAPAR estimates over the three focus areas. The co-located data are based on daily averages of the SIF and FAPAR over the entire focus areas. The dashed lines in the right panel figures are the best-fit linear estimates of the data.

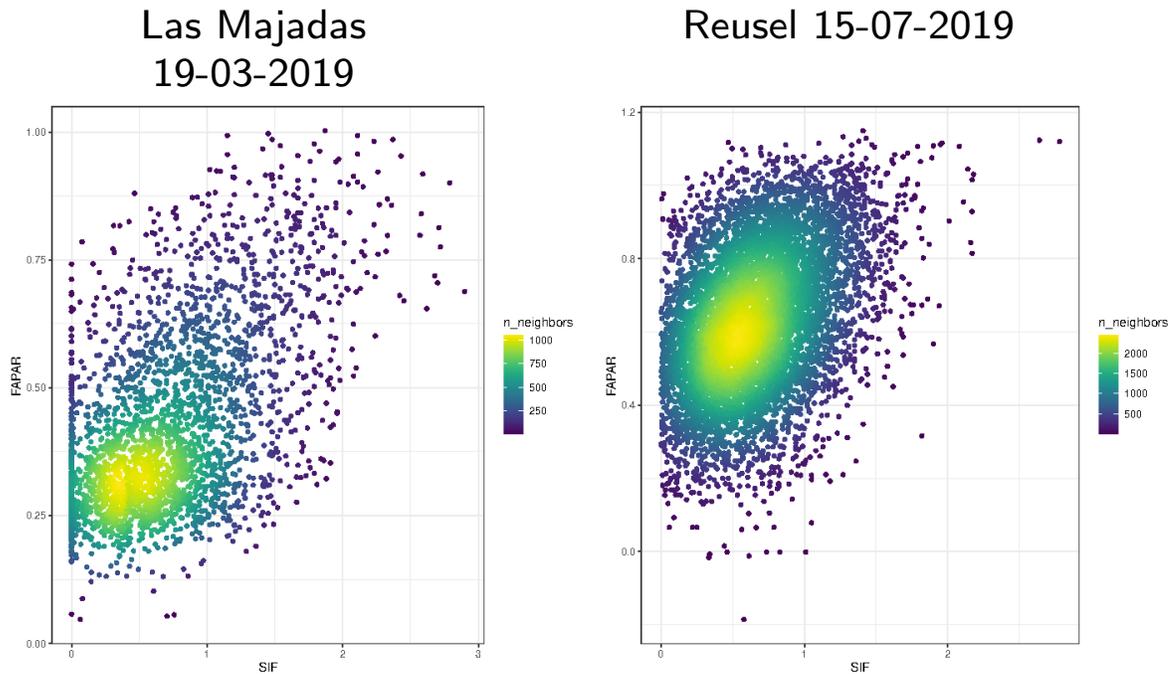


Figure 53: SIF - FAPAR scatterplots: a) Las Majadas 16-07-2018. R2: 0.30 and b) Reusel 29-03-2019. R2: 0.16

9.2.2 Spatial correlation SIF - FAPAR

The relation of both variables over space was analyzed by comparing values obtained at the same locations for specific days. The scatterplots in Fig. 53 show only small correlation between SIF and FAPAR in the case of Las Majadas and Reusel, being non significant in the case of Sodankyla. Although a stronger correlation could be expected between this two variables, being both related with vegetation productivity, many factors could lead to the miss of observed spatial correlation, as for instance differences derived of sensor resolutions (300 m OLCI Vs 7 km TROPOMI), acquisition times and model behaviors. Although reprojection of FAPAR retrieved from OLCI were carried out to fit the 7 km resolution of TROPOMI, correlations remained low, presuming that also the reprojection process generate an additional uncertainty. In the case of the time correlation, this factor would be compensate as working with averages over whole areas.

9.3 Microwaves vs optical data

9.3.1 Comparison SIF Tropomi and VOD SMOS

The relationship between Tropomi SIF data-set and SMOS VOD was studied for the three regions from 2018 to 2020. The following results concern 743nm SIF values. The same analysis made for 735nm lead to the same conclusions. In order to compare SIF with SMOS data, the mean value of SIF data from all the points associated to SMOS-grid points has been calculated using an

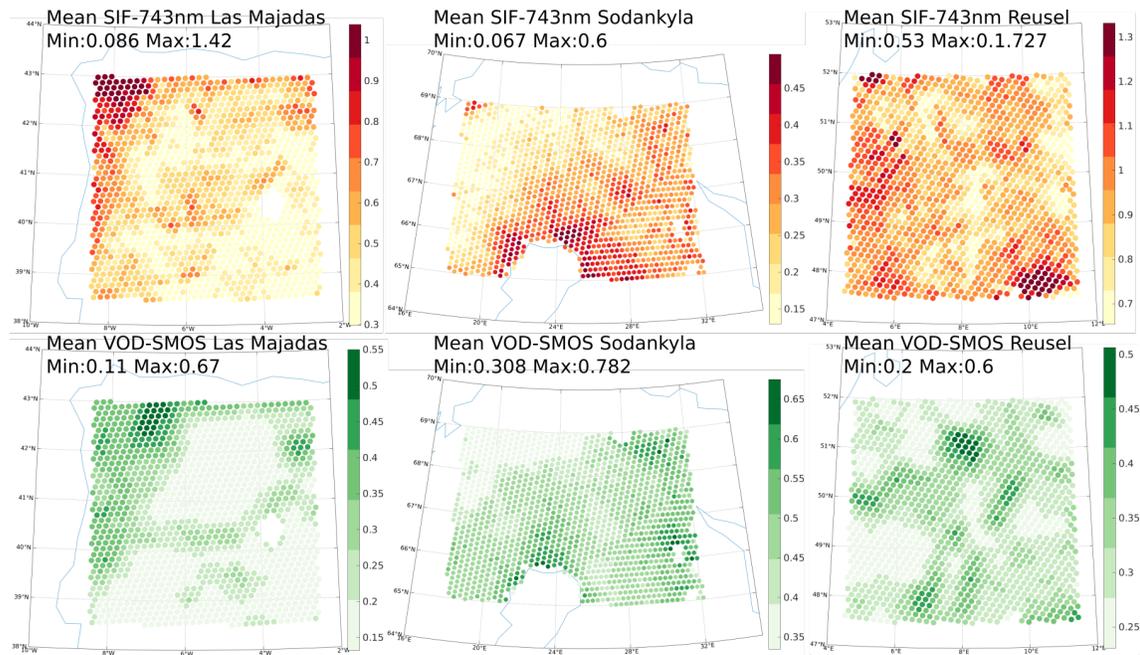


Figure 54: From top to bottom: (i) Mean of SIF-743nm Tropomi values from 2018 to 2020. (ii) Mean of VOD SMOS values from 2018 to 2020. The results are presented for the three regions, from left to right: Las Majadas, Sodankyla, Reusel.

aggregation of the data in area of $43 \times 43 \text{ km}^2$. As Fig. 54 shows, there is an overall good similarity of the high and low values distribution for Las Majadas and Sodankyla. In contrast, there is no clear relationship for Reusel.

In order to get further insight of the VOD and SIF relationship, maps of Pearson correlation were computed (Fig. 55). Correlations are positive for most of the grid points in Sodankyla and Reusel, while there are negative values in large region in Las Majadas. In any case, independently of the sign, the absolute values are low. However, one must have in mind that this is a very preliminary comparison and that data filtering before comparison might not be optimal.

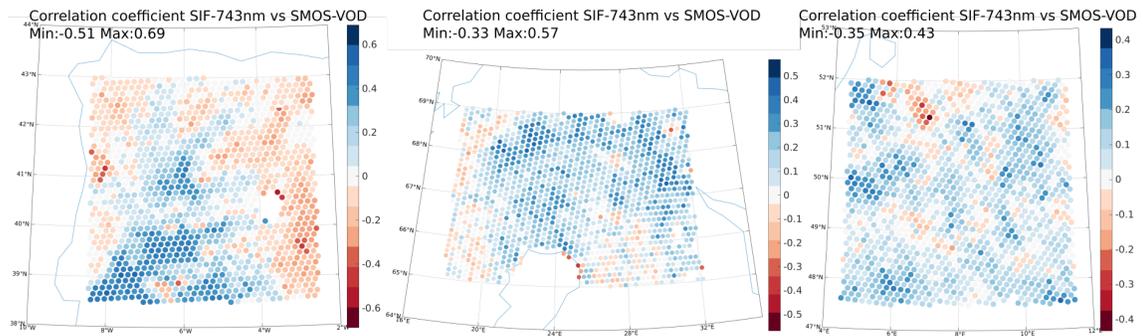


Figure 55: Correlation coefficients maps between SIF (743nm) Tropomi and VOD SMOS. From left to right : Las Majadas, Sodankyla, Reusel.

The top panel of Fig. 56 shows SIF values for both wavelengths, SMOS VOD value and its 11 days moving average for the closest grid point to the actual Las Majadas instrumented site. The bottom panel shows SIF uncertainty for both wavelength and SMOS VOD DQX. Interestingly, uncertainties show the same seasonal cycle, while SIF and VOD values are not showing a clear correlation in their behaviour.

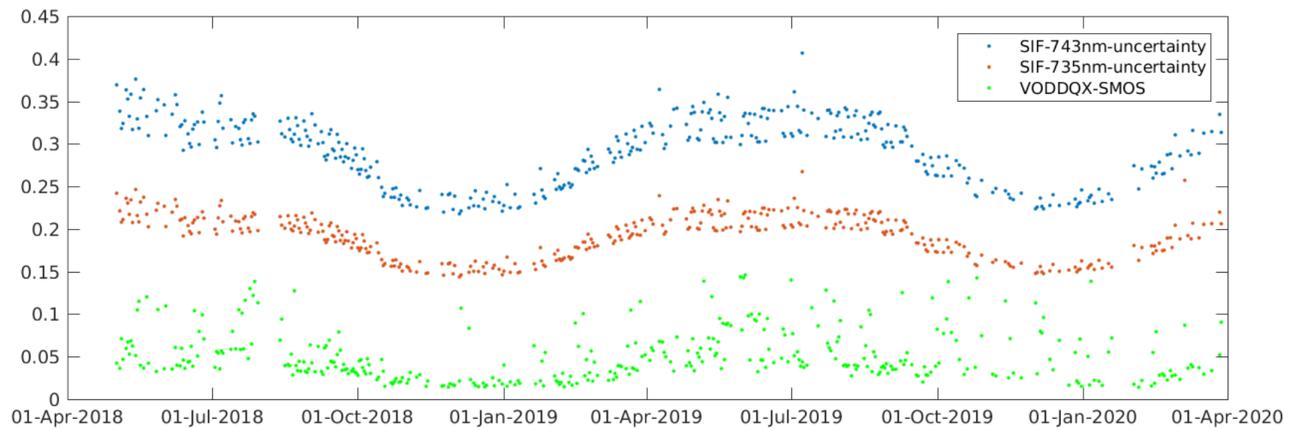
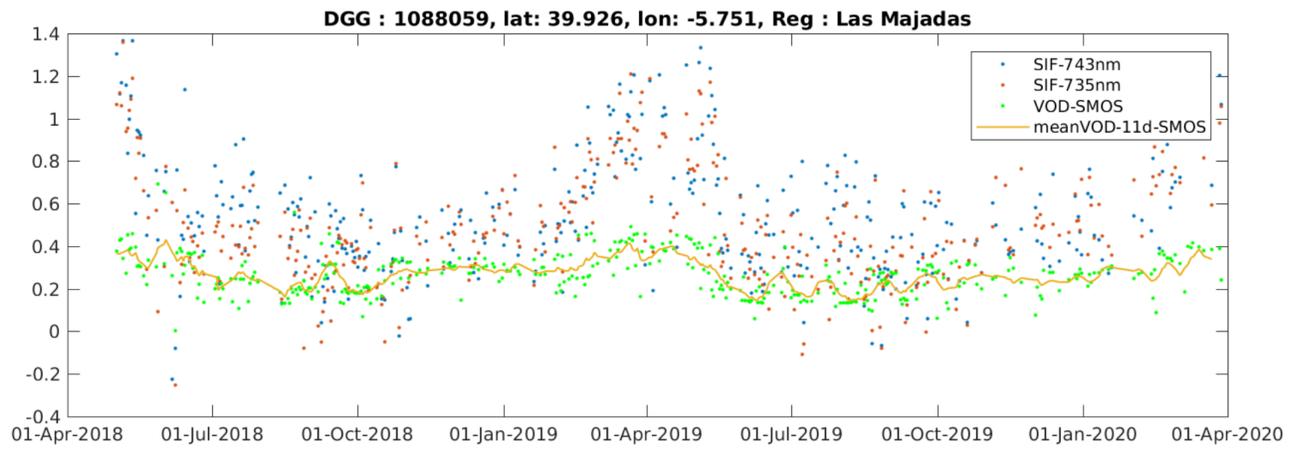


Figure 56: Time series of SIF Tropomi and VOD SMOS with their uncertainty and temporal moving average for the point 1088059 of Las Majadas region

10 Conclusions

The data base [RD2] was analyzed with the following objectives:

- Performing a quality assessment of the data sets
- Providing appropriate filtering strategies and new uncertainties estimations when needed
- Performing a first product cross-comparison

The emphasis of the analysis was put into the data sets that will be used for data assimilation: SMOS SM and VOD, ASCAT VOD, Sentinel 5-P SIF and Sentinel 3 FAPAR. In addition, ASMR-2 VOD, OCO-2 SIF and Sentinel-3 LAI, FVC, LCC have also been analyzed.

A new release of the project data base [RD3] has been done containing the quality assessed data with new uncertainties estimations when needed.

10.1 Individual data sets

The SMOS Level 2 v720 data that became available short after the kick-off of the LCC project were compared to v650. As expected, the main differences arise in regions with mixed land cover. The new v720 shows a smoother behavior in forest/low vegetation transition zones, as expected from the algorithmic changes implemented in v720. This is the SMOS data set that should be used for the rest of the LCC study. SMOS L2 V720 VOD data were further analyzed in the three regions of the study. Taking into account that SMOS observations for adjacent points of the ESA SMOS L2 grid are not fully independent, and that strong variations of VOD in short time scales are not expected, a temporal and spatial variability analysis of VOD was performed. The temporal and spatial variability of VOD were compared to the VOD uncertainty parameter DQX_{tau} . The difference of the daily (local) values and the temporal (spatial) averages was computed and compared to DQX_{tau} . These new uncertainty estimate is sometimes higher and sometimes lower than DQX_{tau} . A conservative approach could be to use the higher estimation of the uncertainty.

ASCAT VOD retrieval is based in the analysis of the slope of the backscatter versus incidence angle relationship. The ASCAT VOD time series are smoothed with a kernel window of 42 days. The uncertainty of the VOD provided with the data increases with increasing VOD values but it remains very low ($10^{-2} - 10^{-3}$). However, no alternatives were found to provide more realistic uncertainties. A filtering recommendation was provided.

Regarding, AMSR-2, a similar approach to that explained above for SMOS VOD was used. The VOD values for a given local time series were compared to the 11-days average. The difference is assumed to be noise as VOD is not expected to change significantly in this time scale.

The S3-OLCI vegetation products (LAI, FAPAR, FVC, LCC) are retrieved using a Bayesian framework in which uncertainties are implicitly calculated together with the (mean) estimations. The accuracy of the model estimations depends on the performance of the training dataset when covering all the possible conditions in which the model will be applied. Originally the GPR models were trained with exclusively simulations coming from the radiative transfer model (RTM) SCOPE, along with spectra of non-vegetated surfaces to account for the variability present in OLCI images. During

LCC WP1 studies, the training data set was generalized by mixing SCOPE simulations together with real spectra data taken from original S3 scenes providing a new more robust uncertainty estimation.

SIF data from OCO-2 and Sentinel 5p were analyzed. The two S5p SIF indices (743 and 735 nm) were compared. The difference is not a constant bias but it shows a seasonal variability, whose origin is unknown. OCO-2 SIF retrievals at 757 nm and 771 nm were used to estimate a 740 nm SIF, which can then be used in comparisons to S5P SIF. There is a large variability in the single SIF observations and it is recommended to use hourly averages. It is recommended to use as associated noise the standard deviation of all SIF values used to compute the hourly average. In this way the interpretation of the seasonal cycle of photosynthesis is easier.

AGB data by the CCI Biomass are provided without uncertainties and detailed study of AGB uncertainties is out of the scope of this project. Nevertheless, the data were analyzed and unrealistic high values were found for some pixels. Threshold values are proposed for data filtering.

10.2 Data sets comparison

The VOD data from different sensors were compared. The bias of ASCAT and SMOS VOD was analyzed for the three regions. The bias (ASCAT-SMOS) maps show regions where it is negative and regions where it is positive. The spatial distribution of the bias maps were compared to land cover maps but no clear relationship has been found so far. Correlation maps show a low but positive correlation in Reusel. In Sodankyla there are both regions where the correlation is positive and regions where it is negative. Nevertheless, the correlation is negative for most of the Las Majadas region. No correlation was found in between the SMOS and ASCAT VOD uncertainties.

Bias maps with of AMSR-2 VOD were also computed with respect to SMOS VOD for the three lowest frequency bands of AMSR-2. The bias (AMSR-2 - SMOS) is mainly positive in Las Majadas except in the North West and also in Reusel. In contrast, it is negative in Sodankyla. Similar maps were obtained for the C1, C2 and X frequencies with respect to the L-band. Pearson correlations of AMSR-2 and SMOS VOD also show negative values in the North West of Las Majadas, in the South East of Sodankyla and in some regions of Reusel.

Different visible data were also compared. In particular, SIF data was compared to FAPAR data. The temporal correlation of SIF and FAPAR was found to vary between the focus areas. The needleleaf forest dominated regions were found to have the lowest correlation with $R=0.59$ in Sodankyla, 0.71 in Las Majadas and 0.81 in Reusel. The time series in Sodankyla show an interesting behaviour in spring: FAPAR values peak already in March while SIF values are still close to zero. According to SIF, the start of photosynthesis occurs in late May or June. Despite their different temporal changes in spring, both SIF and FAPAR start to decrease in August and appear coupled during the autumn. A tentative explanation is that in spring, the pine tree needles appear green although photosynthesis is not yet possible due to the frozen ground. Other difference is that FAPAR annual time series are longer than those of SIF as they also include the early and late winter months when SIF retrievals were not possible due to the limited amount of solar radiation and large solar zenith angles. However, the FAPAR estimates during these time periods may also be affected by these challenges and therefore appear unrealistic.

Finally, a first comparison microwaves and visible data was performed comparing SMOS VOD and

Sentinel 5P SIF. Temporal correlation maps show again regions with negative values, in particular in the Las Majadas regions.

Bibliography

- Camps-Valls, G., J. Verrelst, J. Muñoz-Marí, V. Laparra, F. Mateo-Jiménez, and J. Gómez-Dans, 2016: A survey on Gaussian processes for earth observation data analysis. *IEEE Geoscience and Remote Sensing Magazine*, 4 (2).
- De Grave, C., J. Verrelst, P. Morcillo-Pallarés, L. Pipia, J. Rivera-Caicedo, E. Amin, S. Belda, and J. Moreno, 2020: Quantifying vegetation biophysical variables from the Sentinel-3/FLEX tandem mission: Evaluation of the synergy of OLCI and FLORIS data sources. *Remote Sensing of Environment*, 251, 112 101.
- Guanter, L., et al., 2021: The tropisif global sun-induced fluorescence dataset from the sentinel-5p tropomi mission. *Earth System Science Data Discussions*, 1–27.
- Kerr, Y., P. Waldteufel, P. Richaume, I. Davenport, P. Ferrazzoli, and J. Wigneron, 2011: SMOS level 2 processor soil moisture algorithm theoretical basis document (ATBD). *SM-ESL (CBSA), CESBIO, Toulouse, SO-TN-ESL-SM-GS-0001-3.g, SO-TN-ARR-L2PP-0037-3.6*.
- Rasmussen, C. E. and C. K. I. Williams, 2006: *Gaussian Processes for Machine Learning*. The MIT Press, New York.

A Data base accompanying documentation

Land Surface Carbon Constellation Study

Technical Note (TN)

Satellite observation data base description

**Jochem Verrelst^a, Nemesio Rodríguez-Fernández^b,
Hannakaisa Lindqvist^c, Santiago Belda^a, Pablo Reyes^a,
Wouter Dorigo^d, Emanuel Bueechi^d, Arnaud Mialon^b,
Alexandre Bouvet^b, Martin Barbier^b**

^a Universidad de Valencia, Valencia, Spain

^b Centre d'Etudes Spatiales de la Biosphère, Toulouse, France

^c Finnish Meteorological Institute, Finland

^d Technische Universität Wien, Austria

Background

The present document describes the required processing steps for obtaining the various satellite products that are selected for the study. A description of the selected products is provided in **D4**. Regarding the processing, some products are already readily available and the required steps are merely intersecting over the regions and writing it away. Other products required additional retrieval processing steps.

All the products are available in regions of 500 x 500 km² around the three study sites (1: Sodankylä; 2: Majadas de Tietar; 3: Reusel), and they span several years in the study time period from 2011 to 2021. The datasets selected are not space or time averages and are Level 2 datasets, and some of them with well-defined uncertainties.

Table 1. Summary of the remote sensing datasets used in this study and time period covered by each of them.

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Exact period
SMOS L-VOD												01/2011-12/2021
ASCAT C-VOD												01/2011-12/2021
OCO-2 SIF												09/2014-12/2021
Sentinel 5P SIF												12/2017-12/2021
S3 FAPAR/LAI												04/2016-12/2021
ASCAT backscattering												01/2011-12/2021
SMOS soil moisture												01/2011-12/2021
SMOS TB												01/2011-12/2021
AMSR-2 C-VOD												11/2012-12/2021
AMSR-2 X-VOD												11/2012-12/2021
MODIS L ST												01/2011-12/2021
MODIS PRI												01/2011-12/2021
S3 LCC												04/2016-12/2021
S3 FVC												04/2016-12/2021

In the following section, for each product a concise overview is given of the key steps necessary to obtain the satellite products in a workable format (NetCDF) for the continuation of the project.

SMOS Soil Moisture and L-VOD

SMOS Level 2 v700

SMOS measures the thermal emission from the Earth in the 1.4 GHz protected frequency range in full-polarization and for incidence angles from 0° to ~60°. The footprint size and shape depend on the position within the field of view but the full width at half maximum size of the synthesized beam is ~43 km on average. The equator overpass time is 6:00 AM/PM for ascending/descending orbits.

The soil temperature profile is taken from European Centre for Medium Range Weather Forecast (ECMWF) Integrated Forecast System (IFS) data. The forward model contributions are computed at 4 km resolution pixels and aggregated to the sensor resolution using the mean synthetic antenna pattern. For footprints with mixed land cover, the SM content of the minor land cover is estimated from the ECMWF IFS predictions, and its contribution to the BT is fixed. In this case, the SMOS retrieval is only performed for the dominant land cover class within the footprint. In order to initialize the minimization over forest, a first guess for L-VOD is estimated from the maximum LAI (Rahmoune et al. 2013).

At the date of the kick-off meeting of the project, the operational version of SMOS Level 2 data was v650. During the first quarter of 2021, v700 became available, and the data in the project data base was updated. The current version of SMOS Level 2 data in the data base is v700.

The data is organized in three files, one per region, with names

“smosL2_1D_v700_REGION_trans.nc”

where “REGION” is “sodankyla”, “lasmajadas” or “reusel”.

Data are in netcdf4 format with deflate level 4.

Example of netcdf file generated for smosL2 data :

```
dimensions:
  nobs = 4752455;
  ntc = 6;
variables:
  float DGG(nobs=4752455);
    :long_name = "DGG";
    :units = "-";
    :_ChunkSizes = 573329U; // uint

  float lat(nobs=4752455);
    :long_name = "Latitude";
    :units = "degrees_north";
    :_FillValue = -99999.0f; // float
    :_ChunkSizes = 573329U; // uint

  float lon(nobs=4752455);
```

```
:long_name = "Longitude";
:units = "degrees_east";
:_FillValue = -99999.0f; // float
:_ChunkSizes = 573329U; // uint

int yymmddHHMMSS(nobs=4752455, ntc=6);
:long_name = "Aquisition time (UTC)";
:units = "year month day hour min sec";
:_FillValue = -99999; // int
:_ChunkSizes = 174762U, 6U; // uint

float soil_moisture(nobs=4752455);
:_FillValue = -99999.0f; // float
:long_name = "SM";
:units = "m3/m3";
:footprint_diameter = "43";
:footprint_diameter_unit = "km";
:_ChunkSizes = 573329U; // uint

float soil_moisture_dqx(nobs=4752455);
:_FillValue = -99999.0f; // float
:long_name = "SM Dqx";
:units = "m3/m3";
:_ChunkSizes = 573329U; // uint

float tau(nobs=4752455);
:_FillValue = -99999.0f; // float
:long_name = "VOD";
:units = "-";
:footprint_diameter = "43";
:footprint_diameter_unit = "km";
:_ChunkSizes = 573329U; // uint

float tau_dqx(nobs=4752455);
:long_name = "VOD Dqx";
:units = "-";
:_FillValue = -99999.0f; // float
:_ChunkSizes = 573329U; // uint

float Chi2(nobs=4752455);
:long_name = "Goodness of fit";
:units = "-";
:_FillValue = -99999.0f; // float
:_ChunkSizes = 573329U; // uint

float Chi2P(nobs=4752455);
:long_name = "Goodness of fit probability";
:units = "-";
:_FillValue = -99999.0f; // float
:_ChunkSizes = 573329U; // uint

float FL_forest(nobs=4752455);
:long_name = "Flag forest";
:units = "-";
```

```

:_FillValue = -99999.0f; // float
:_ChunkSizes = 573329U; // uint

float FL_nominal(nobs=4752455);
:long_name = "Flag nominal soil";
:units = "-";
:_FillValue = -99999.0f; // float
:_ChunkSizes = 573329U; // uint

float FL_frost(nobs=4752455);
:long_name = "Flag frost";
:units = "-";
:_FillValue = -99999.0f; // float
:_ChunkSizes = 573329U; // uint

float FL_ice(nobs=4752455);
:long_name = "Flag ice";
:units = "-";
:_FillValue = -99999.0f; // float
:_ChunkSizes = 573329U; // uint

float RFI_Prob(nobs=4752455);
:long_name = "RFI Probability";
:units = "--";
:_FillValue = -99999.0f; // float
:_ChunkSizes = 573329U; // uint

float X_swath(nobs=4752455);
:long_name = "Distance to the center of the swath";
:units = "kilometers";
:_FillValue = -99999.0f; // float
:_ChunkSizes = 573329U; // uint

float soil_temperature(nobs=4752455);
:_FillValue = -99999.0f; // float
:long_name = "Soil temperature first layer ECMWF from AUXEC files";
:units = "K";
:_ChunkSizes = 573329U; // uint

float snow_depth(nobs=4752455);
:long_name = "Snow depth ECMWF from AUXEC files";
:units = "meters";
:_FillValue = -99999.0f; // float
:_ChunkSizes = 573329U; // uint

tau_11d_average
  Size: 4493327x1
  Dimensions: npt
  Datatype: double
  Attributes:
    _FillValue = -99999
    long_name = 'VOD 11d Moving Average'
    units = '-'

tau_9_neighbours_average

```

```

Size: 4493327x1
Dimensions: npt
Datatype: double
Attributes:
    _FillValue = -99999
    long_name = 'VOD 9 neighbours Moving Average'
    units = '.'
Diff_tau_tau_11d_average
Size: 4493327x1
Dimensions: npt
Datatype: double
Attributes:
    _FillValue = -99999
    long_name = 'Difference between VOD and VOD temporal average'
    units = '.'
Diff_tau_tau_9_neighbours_average
Size: 4493327x1
Dimensions: npt
Datatype: double
Attributes:
    _FillValue = -99999
    long_name = 'Difference between VOD and VOD spatial average'
    units = '.'
std_tau_tau_11d_average
Size: 4493327x1
Dimensions: npt
Datatype: double
Attributes:
    _FillValue = -99999
    long_name = 'Standard deviation of VOD values use for the temporal average'
    units = '.'
std_tau_tau_9_neighbours_average
Size: 4493327x1
Dimensions: npt
Datatype: double
Attributes:
    _FillValue = -99999
    long_name = 'Standard deviation of VOD values use for the spatial average'
    units = '.'

```

```

// global attributes:
:creation_time = "28-May-2021 08:48:25";
:institution = "CESBIO";
:title = "SMOS Level 2 v700";
:contact = "nemesio.rodriguez@cesbio.cnes.fr";

```

The files contain soil moisture and VOD (tau) with their corresponding uncertainties (“_dqx”). Other variables are also contained in the files to help to the data analysis to be performed in WP1.3.

SMOS INRA-CESBIO (IC)

SMOS IC version 105 (the same dataset used for the SMOS+Veg project) has also been added to the data base for comparison with the new Level 2 SMOS data set v700 if needed. SMOS-IC v105 is only available from 2011 to 2016.

SMOS IC L-VOD is provided as netcdf file, one per site, i.e.

SMOSIC_SMVOD__lasmajadas.nc

SMOSIC_SMVOD__reusel.nc

SMOSIC_SMVOD__sodankyla.nc

The file contains the latitude and the longitude coordinates of each SMOS nodes as a vector. The data are stored in a matrix of dimensions: npt x nvar x ntime

Npt is the number of SMOS nodes peer zones

Nvar is 8 such as :

- 1) Days : number of days since 01/01/2000
- 2) Seconds (seconds) : number of seconds, UTC time
- 3) Microseconds: number of seconds, UTC time
- 4) VOD : Vegetation Optical Depth (-)
- 5) VOD_stdError : Vegetation Optical Depth (-), standard error, is the range of VOD that is due to radiometric accuracy of SMOS measurements
- 6) SM (m³/m³) is Surface Soil Moisture
- 7) SM_stdError (m³/m³) is Surface Soil Moisture error. It is the range of SM due to the radiometric accuracy of SMOS measurements
- 8) RMSE (K) Root Mean Square error, between SMOS acquired Brightness Temperatures and modeled ones, a measure of good fit.

Fill value is -999

Good Practice :

i) SMOS is subject to RFI sources. To filter out suspicious data, one can use the RMSE and a threshold of 6K.

ii) The VOD can be < 0. If mathematically possible, it is not acceptable. It is then recommended to keep VOD values that are within the range [0 1.8] .

The data are stored such as :

variables:

```
float npt(npt) ;
```

```
float nvar(nvar) ;
```

```
float ntime(ntime) ;
```

```
float data(ntime, nvar, npt) ;
```

```
data :long_name = "" ;
```

```

data :units = "- , s, s, -, -, m3/m3, m3/m3, K" ;

data :_FillValue = -99999. ;

data :footprint_diameter = "43" ;

data :footprint_diameter_unit = "km" ;

float lat(npt) ;

lat:long_name = "Latitude" ;

lat:units = "degree" ;

lat:_FillValue = -99999. ;

float lon(npt) ;

lon:long_name = "Longitude" ;

lon:units = "Degree" ;

lon:_FillValue = -99999. ;

```

SMOS Level 3 Brightness temperatures

SMOS brightness temperatures have been provided to the data base. The brightness temperatures chosen are Level 3 TBs because there are already expressed in vertical and horizontal polarization (and not in the antenna reference frame as L1 TBs) and they are angle-binned. These L3TBs are processed with the latest version of the Level 3 processor, version 330, done for reprocessing campaign number 7, RE07. The data cover the period April 2010 to Mars 2020. A catch up campaign is planned in 2021 to process the data from April 2020 onwards. The files used for this ITT will be updated accordingly.

For each node and overpass, SMOS TB are acquired at various incidence angle at the antenna frame (X and Y polarisations). The CATDS developed the L3 TB product which provides data : i) TB in H and V polarisations ii) at regular incidence angle, i.e. the TB are averaged over ranges of incidence angles.

For the Carbon Constellation project, 3 files are created, one per regions of interests. The data are stored as netcdf file, with the following content :

- *) lat : latitude of each nodes, vector of N_nodes
- *) lon : Longitude of each nodes, vector of N_nodes
- inc: incidence angle, middle angle per bin, 1x15 vector (bin # 15 is not used for the time being)
- *) dinc : range of angle for each bins, in degrees. For instance, a bin is 5 degree wide, so it is inc +/- 2.5 deg. So for the bin 8, the "inc" angle is 37.5 deg., its dinc is 5 deg, which means

the bin covers the smos measurements acquired between 35 and 40 deg.

- *) BT_H : Brightness Temperatures in H polarisation ; matrix of N_nodes x N_bin x N_time
- *) BT_V: Brightness Temperatures in V polarisation ; matrix of N_nodes x N_bin x N_time
- *) Pixel_Radiometric_Accuracy_V: Radiometric Accuracy of TB V per bin, is the average of radiometric accuracy of TBs acquired at an incidence angle within [inc+/- 0.5dinc]
- *) Pixel_Radiometric_Accuracy_H: Radiometric Accuracy of TB H per bin, is the average of radiometric accuracy of TBs acquired at an incidence angle within [inc+/- 0.5dinc]
- *) Footprint_Axis1: Dimension (in km) of major axes of the ellipse of the footprint ; matrix of N_nodes x N_bin x N_time
- *) Footprint_Axis2 : Dimension (in km) of minor axes of the ellipse of the footprint ; matrix of N_nodes x N_bin x N_time
- *) Nb_RFI_Flags: Number of TB with potential RFI detection ; matrix of N_nodes x N_bin x N_time
- *) Nviews : Number of Brightness Temperatures per bin ; matrix of N_nodes x N_bin x N_time
- *) yymmddHHMMSS : Time of acquisition, in UTC time. Is a matrix N_time x 6, such as year, month, days of the month, hour, minute, secondes.

Good practice and recommendations : Unfortunately, SMOS TB are affected by RFI sources. Our best option is to filter out data that seem to be wrong because of RFI contributions. It is recommended to use the ratio N_RFI_flags / Nviews with a threshold of 0.4/0.5. This is still under investigations. For the time being, please remove TB>320K, and TB with a N_NRFIFlag/Nview > 0.5.

For more informations regarding RFI filtering, please contact Cesbio.

References :

- CATDS (2016). CATDS-PDC L3TB - Global polarised brightness temperature product from SMOS satellite. CATDS (CNES, IFREMER, CESBIO)
<http://dx.doi.org/10.12770/6294e08c-baec-4282-a251-33fee22ec67f>
- Al Bitar Ahmad, Mialon Arnaud, Kerr Yann H., Cabot Francois, Richaume Philippe, Jacqueline Elsa, Quesney Arnaud, Mahmoodi Ali, Tarot Stephane, Parrens Marie, Al-Yaari Amen, Pellarin Thierry, Rodriguez-Fernandez Nemesio, Wigneron Jean-Pierre (2017). The global SMOS Level 3 daily soil moisture and brightness temperature maps. Earth System Science Data, 9(1), 293-315. Publisher's official version : <http://doi.org/10.5194/essd-9-293-2017>, Open Access version : <http://archimer.ifremer.fr/doc/00388/49941/>

ASCAT VOD

ASCAT VOD is derived from ASCAT sigma0 from Metop-A (2007/01 - 2020/12), Metop-B (2012/01 - 2020/12), and Metop-C (2019/04 - 2020/12). The computation of the dry and wet backscatter reference is based on estimating the slope and curvature parameters using a Kernel Smoother. The VOD computation is based on TU-Wien retrieval algorithm and a water-cloud model (for further info see Vreugdenhil et al. 2016).

Per region, one netcdf file is provided: ASCAT_vod_so.nc, ASCAT_vod_re.nc, and ASCAT_vod_lm.nc

The final product contains the information of VOD, its noise, flag and the correspondent *yymmddHH*, *lat* and *lon* vectors. The data is stored on the WARP 5 grid representing a radius of approximately 12.5 km. Hence, the data is structured as followed: *lat lon* vectors represent the centres of the considered datapoint. The vector *yymmddHH* then represents how VOD data is stored. All *lat/lon* combinations are repeated for each individual timestep. The vectors *VOD*, *VOD_Noise*, *Proc_Flag*, thus, represent the corresponding values starting with the whole time series of the first *lat/lon* combination. The last two variables in the datasets are *probability_of_frozen_soils* and *snow_probability* which provide probabilities (in %) if the soils are frozen and covered by snow at the moment and location of the observation.

The individual variables are described in the nc file itself. Here, only few additional information is provided:

Lat-, and *lon_vertices* are not provided, as the ASCAT data has no regular grid with rectangular pixels.

VOD and VOD_Noise are provided as obtained from the retrieval algorithm. Hence, no detrending or similar has been applied.

Fill value is -999999. The additional information in the nc file of *_FillValue* can be neglected!

Flagging of the data is only basically provided and is either 0,1,or 2. 0 means that data is ok, 1=no data and 2 stands for unreliable results. At some locations there are many missing values (>5% of all times). At such locations it happens frequently that the remaining data is of poor quality. Additional flagging is required for observations on frozen and / or snow covered soils. Ideally, this is done with an additional external dataset like ERA5-Land or similar. If this is not desired, the variables *probability_of_frozen_soils* and *snow_probability* can be used to mask observations with a high probability (eg. >60%) of being frozen and / or covered with snow.

The data are stored such as :

Source:

las_majadas.nc

Dimensions:

nobs = 5895942;

Variables:

```
    long nobs(nobs=5895942);

float VOD_Noise(nobs=5895942);
  :_FillValue = NaNf; // float
  :FillValue = -999999; // int
  :units = "-";
  :long_name = "Noise of VOD data";
  :approximate_footprint_diameter = 0.125; // double
  :footprint_diameter_unit = "degrees";
  :comment = "the footprint diameter covers an area of: 0.113x0.144, latitude x longitude";
  :_ChunkSizes = 982657U; // uint

float VOD(nobs=5895942);
  :_FillValue = NaNf; // float
  :FillValue = -999999; // int
  :units = "-";
  :long_name = "Vegetation Optical Depth from ASCAT data";
  :approximate_footprint_diameter = 0.125; // double
  :footprint_diameter_unit = "degrees";
  :comment = "the footprint diameter covers an area of: 0.113x0.144, latitude x longitude";
  :_ChunkSizes = 982657U; // uint

float lon(nobs=5895942);
  :_FillValue = NaNf; // float
  :units = "degrees east";
  :long_name = "longitude";
  :comment = "vector point representing center of footprint";
  :_ChunkSizes = 982657U; // uint

long yymmddHHMMSS(nobs=5895942);
  :units = "yymmddHHMMSS";
  :long_name = "time point of observation";
  :comment = "time coordinate given as UTC";
  :_ChunkSizes = 491329U; // uint

int Proc_Flag(nobs=5895942);
  :units = "-";
  :long_name = "Flagging of vod data";
  :comment = "Contains the following flags: 0=Data is ok; 1=No Data; 2=Data quality
uncertain. Comes from location with many no data points (>5%)";
  :_ChunkSizes = 982657U; // uint

float lat(nobs=5895942);
  :_FillValue = NaNf; // float
  :units = "degrees north";
  :long_name = "latitude";
  :comment = "vector point representing center of footprint";
  :_ChunkSizes = 982657U; // uint

int snow_probability(nobs=5895942);
  :comment = "Should only be used if no better information is available about the
occurrence of snow. If that is the case at least all observations with values >60 should
be flagged";
  :long_name = "Probability of snow cover in %";
```

```
:units = "-";  
:_ChunkSizes = 982657U; // uint  
  
int probability_of_frozen_soils(nobs=5895942);  
:comment = "Should only be used if no better information is available about the soil  
conditions. If that is the case at least all observations with values >60 should be  
Flagged";  
:long_name = "Probability of frozen soils in %";  
:units = "-";  
:_ChunkSizes = 982657U; // uint
```

Literature:

M. Vreugdenhil, W. A. Dorigo, W. Wagner, R. A. M. de Jeu, S. Hahn and M. J. E. van Marle, "Analyzing the Vegetation Parameterization in the TU-Wien ASCAT Soil Moisture Retrieval," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3513-3531, June 2016, doi: 10.1109/TGRS.2016.2519842..

For additional information please contact TU Wien: emanuel.bueechi@geo.tuwien.ac.at

OCO-2 SIF

1. Daily netCDF files of OCO-2 version 10r SIF product were downloaded from Nasa EarthData GES DISC at https://disc.gsfc.nasa.gov/datasets/OCO2_L2_Lite_SIF_10r/summary?keywords=SIF
2. The data were selected based on the central latitude and longitude of each sounding.
3. The data were quality screened based on the product quality flag (QF). Retrievals with QF = 0 ('best') and QF = 1 ('good') were accepted.
4. OCO-2 provides data for SIF retrievals at 757 nm and 771 nm. Based on these, SIF at 740 nm is also estimated.
5. Finally, a netCDF file was created per each region using SIF retrievals from 2014 to the end of 2020.
6. NetCDF files have been compressed using nccopy -d9.
7. Data description:
 - Approximately 1.2 km x 2.3 km resolution per sounding at nadir geometry.
 - The pixel geometry varies according to the satellite viewing mode (nadir, glint, target). Therefore, all corner coordinates are given.
 - Data provided as single soundings, no averaging or aggregation.
 - SIF at 740 nm corresponds to SIF around the far-red peak in the SIF spectrum. This is comparable e.g. to the TROPOMI SIF.
 - Retrievals at other wavelengths (757 nm and 771 nm) are also given as well as the illumination geometry (zenith and azimuth angles) in case modelling can benefit from these.
8. File Structure of netcdf file (see below). The variables are:
 - latitude_corners
 - longitude_corners
 - yymmddHHMMSS (time)
 - SIF_740nm
 - SIF_740nm_uncertainty
 - SIF_757nm
 - SIF_757nm_uncertainty
 - SIF_771nm
 - SIF_771nm_uncertainty
 - SZA (Solar Zenith Angle)
 - VZA (Viewing Zenith Angle)
 - SAz (Solar Azimuth Angle)
 - VAz (Viewing Azimuth Angle)
9. Status:
 - Majadas OCO-2 SIF is **ready from 2014 to 2020 (file size: 22.2 MB)**
 - Sodankyla OCO-2 SIF is **ready from 2014 to 2020 (file size: 17.6 MB)**
 - Reusel OCO-2 SIF is **ready from 2014 to 2020 (file size: 16.7 MB)**
10. The datasets are located here:
https://fmi100-my.sharepoint.com/:f/g/personal/hannakaisa_lindqvist_fmi_fi/EkjY3MIHxtRPrIR0n5UqZhcBqSRQILIOvV-_Y0O4vzPPRg?e=THnZwM

GENERATED NETCDF FILE

Source:

Majadas_SIF_OCO2_final.nc4

Reusel_SIF_OCO2_final.nc4

Sodankyla_SIF_OCO2_final.nc4

Format:

netCDF-4

Dimensions (from Majadas file):

sounding_dim = 526907

vertex_dim = 4

ntc = 6

Variables:

latitude_corners(sounding_dim, vertex_dim)

Data type: float

long_name = "Corner latitude of the measurement"

units = "degrees_north"

longitude_corners(sounding_dim, vertex_dim)

Data type: float

long_name = "Corner longitude of the measurement"

units = "degrees_east"

yymmddHHMMSS(sounding_dim, ntc)

Data type: int

long_name = "Time point of observation"

units = "Time given as UTC"

SZA(sounding_dim)

Data type: float

long_name = "Solar zenith angle"

units = "degrees"

VZA(sounding_dim)

Data type: float

long_name = "Viewing zenith angle"

units = "degrees"

SAz(sounding_dim)

Data type: float

long_name = "Solar azimuth angle"

units = "degrees"

VAz(sounding_dim)

Data type: float

long_name = "Viewing azimuth angle"

units = "degrees"

SIF_740nm(sounding_dim)

Data type: float

long_name = "Estimated Solar Induced Fluorescence at 740 nm"

units = "W/m^2/sr/μm"

SIF_740nm_uncertainty(sounding_dim)

Data type: float

long_name = "Estimated 1-sigma uncertainty of Solar Induced Fluorescence at 740 nm"

units = "W/m²/sr/μm"

SIF_757nm(sounding_dim)

Data type: float

long_name = "Solar Induced Fluorescence at 757 nm"

units = "W/m²/sr/μm"

SIF_757nm_uncertainty(sounding_dim)

Data type: float

long_name = "1-sigma uncertainty of retrieved Solar Induced Fluorescence at 757 nm"

units = "W/m²/sr/μm"

SIF_771nm(sounding_dim)

Data type: float

long_name = "Solar Induced Fluorescence at 771 nm"

units = "W/m²/sr/μm"

SIF_771nm_uncertainty(sounding_dim)

Data type: float

long_name = "1-sigma uncertainty of retrieved Solar Induced Fluorescence at 771 nm"

units = "W/m²/sr/μm"

For additional information, please contact Finnish Meteorological Institute
(hannakaisa.lindqvist@fmi.fi).

Sentinel-5P TROPOMI SIF

1. Daily netCDF files of TROPOSIF version 2.0 Level 2B SIF product were downloaded from <ftp://ftp.sron.nl/open-access-data-2/TROPOMI/tropomi/sif/v2.0/> and are courtesy of Luis Guanter's team.
2. The data were selected based on the central latitude and longitude of each sounding.
3. The data were quality screened based on the cloud fraction. Retrievals with cloud fraction < 0.2 were accepted, as recommended by the product ATBD.
4. A rough outlier filtering has been applied.
5. Please note that negative SIF values may occur due to the retrieval process. These should not be cut from the data because that would bias the mean SIF values.
6. A comparison was made between the all-sky (SIF_743nm) and clear-sky (SIF_735nm) SIF data sets. The differences were found to be small.
7. Finally, a netCDF file was created per each region using SIF retrievals from 5/2018 to the end of 2020.
8. NetCDF files have been compressed using nccopy -d9.
9. Data description:
 - Approximately 3.6 km x 5.6 km resolution per sounding at nadir geometry.
 - The pixel size varies according to the pixel placement in the instrument swath.
 - Data provided as single soundings, no averaging or aggregation.
 - SIF at 740 nm corresponds to SIF around the far-red peak in the SIF spectrum.
10. File Structure of netcdf file (see below). The variables are:
 - latitude
 - longitude
 - latitude_corners
 - longitude_corners
 - yymmddHHMMSS (time)
 - SIF_743nm
 - SIF_743nm_uncertainty
 - SIF_735nm
 - SIF_735nm_uncertainty
11. Status:
 - MAJADAS TROPOMI SIF is ready from 2018 to 2020 (file size: 129 MB)
 - Sodankyla TROPOMI SIF is ready from 2018 to 2020 (file size: 97.6 MB)
 - Reusel TROPOMI SIF is ready from 2018 to 2020 (file size: 86.8 MB)
12. The datasets are located here:
https://fmi100-my.sharepoint.com/:f/g/personal/hannakaisa_lindqvist_fmi_fi/EkjY3MIHxtRPrIR0n5UqZhcBqSRQILIOvV-_Y0O4vzPPRg?e=THnZwM

GENERATED NETCDF FILE

Source:

Majadas_SIF_TROPOMI_final.nc4
Reusel_SIF_TROPOMI_final.nc4
Sodankyla_SIF_TROPOMI_final.nc4

Format:

netCDF-4

Dimensions (from Majadas file):

sounding_dim = 3765534
vertex_dim = 4
ntc = 6

Variables:

latitude_corners(sounding_dim, vertex_dim)
Data type: float
long_name = "Corner latitude of the measurement"
units = "degrees_north"

longitude_corners(sounding_dim, vertex_dim)
Data type: float
long_name = "Corner longitude of the measurement"
units = "degrees_east"

latitude(sounding_dim)
Data type: float
long_name = "Center latitude of the measurement"
units = "degrees_north"

longitude(sounding_dim)
Data type: float
long_name = "Center longitude of the measurement"
units = "degrees_east"

yymmddHHMMSS(sounding_dim, ntc)
Data type: int
long_name = "Time point of observation"
units = "Time given as UTC"

SIF_743nm(sounding_dim)
Data type: float
long_name = "Retrieved Solar Induced Fluorescence at 743 nm"
units = "mW/m^2/sr/nm"

SIF_743nm_uncertainty(sounding_dim)
Data type: float
long_name = "Uncertainty of Solar Induced Fluorescence at 743 nm"
units = "mW/m^2/sr/nm"

SIF_735nm(sounding_dim)
Data type: float
long_name = "Retrieved Solar Induced Fluorescence at 735 nm"
units = "mW/m^2/sr/nm"

SIF_735nm_uncertainty(sounding_dim)
Data type: float
long_name = "Uncertainty of Solar Induced Fluorescence at 735 nm"

units = "mW/m^2/sr/nm"

For additional information, please contact Finnish Meteorological Institute
(hannakaisa.lindqvist@fmi.fi).

ASCAT Backscatter

The ASCAT Backscatter is derived from ASCAT sigma40 from Metop-A (2007/01 - 2020/12), Metop-B (2012/01 - 2020/12), and Metop-C (2019/04 - 2020/12). The computation of the dry and wet backscatter reference is based on estimating the slope and curvature parameters using a Kernel Smoother. This dataset provides the information of the backscattering and the slope signal at 40° from ASCAT 12.5 km sampling. This dataset was established in the frame of the H SAF project at TU Vienna.

The delivered datasets are structured similarly as ASCAT vod data. For each of the three regions there is one dataset:

bs40_lm.nc; bs40_re.nc; bs40_so.nc, for Las Majadas, Reusel, and Sodankylaa respectively.

The final product contains the information of the backscatter and the slope at 40°, their noises, the correspondent time, lat and lon vectors, as well as confidence flag and surface state flag. The data is stored on the WARP 5 grid representing a radius of approximately 12.5 km. Hence, the data is structured as followed: lat lon vectors represent the centres of the considered datapoint. The vector time then represents how VOD data is stored. All lat/lon combinations are repeated for each individual timestep. The vectors sigma40 and sigma40_noise represent the corresponding values starting with the whole time series of the first lat/lon combination. Flagging information for potentially unreliable observations are stored in the two variables *conf_flag*, and *surface_state_flag*. At least observations on frozen or soil covered soils should be masked by using these flags. This information is provided in the variable *surface_state_flag*.

The individual variables are described in the nc file itself. Here, only few additional information is provided:

Lat-, and lon_vertices are not provided, as the ASCAT data has no regular grid with rectangular pixels.

Fill value is -9999. The additional information in the nc file of *_FillValue* can be neglected!

Example NetCDF file:

Dimensions:

```
nobs = 11173095;  
ntc = 4;
```

Variables:

```
float sigma40(nobs=11173095);  
: _FillValue = NaNf; // float  
: FillValue = -9999; // int  
: units = "dB";  
: long_name = "Backscatter at 40 degrees from ASCAT data";  
: geospatial_sampling = 0.125; // double  
: geospatial_resolution = "25-34 km";  
: footprint_diameter_unit = "degrees";  
: comment = "Resolution is dependent on location. Therefore, range of resolution is
```

```

    Given";
    :_ChunkSizes = 1015736U; // uint

float sigma40_noise(nobs=11173095);
:_FillValue = NaNf; // float
:FillValue = -9999; // int
:units = "dB";
:long_name = "Noise of Backscatter at 40 degrees from ASCAT data";
:geospatial_sampling = 0.125; // double
:geospatial_resolution = "25-34 km";
:footprint_diameter_unit = "degrees";
:comment = "Resolution is dependent on location. Therefore, range of resolution is
    Given";
:_ChunkSizes = 1015736U; // uint

float slope40(nobs=11173095);
:_FillValue = NaNf; // float
:FillValue = -9999; // int
:units = "dB";
:long_name = "Slope at 40 degrees from ASCAT data";
:geospatial_sampling = 0.125; // double
:geospatial_resolution = "25-34 km";
:footprint_diameter_unit = "degrees";
:comment = "Resolution is dependent on location. Therefore, range of resolution is
    Given";
:_ChunkSizes = 1015736U; // uint

float slope40_noise(nobs=11173095);
:_FillValue = NaNf; // float
:FillValue = -9999; // int
:units = "dB";
:long_name = "Noise of slope at 40 degrees from ASCAT data";
:geospatial_sampling = 0.125; // double
:geospatial_resolution = "25-34 km";
:footprint_diameter_unit = "degrees";
:comment = "Resolution is dependent on location. Therefore, range of resolution is
    Given";
:_ChunkSizes = 1015736U; // uint

float conf_flag(nobs=11173095);
:flag_meanings = "bad_surface_state_flag topographic_complexity_above_50perc
    wetland_above_50perc soil_moisture_noise_above_50perc
    sensitivity_to_soil_moisture_below_1dB reserved_for_future_use
reserved_for_future_use";
:valid_range = 0B, 127B; // byte
:_FillValue = NaNf; // float
:name = "conf_flag";
:long_name = "confidence flag";
:flag_masks = 1B, 2B, 4B, 8B, 16B, 32B, 64B; // byte
:_ChunkSizes = 1015736U; // uint

float surface_state_flag(nobs=11173095);
:_FillValue = NaNf; // float
:name = "ssf";
:long_name = "surface state flag";

```

```
:flag_values = 0B, 1B, 2B, 3B, 4B; // byte
:flag_meanings = "unknown unfrozen frozen_temporary melting_water_on_the_surface
  Permanent_ice";
:valid_range = 0B, 4B; // byte
:_ChunkSizes = 1015736U; // uint

float lat(nobs=11173095);
:comment = "vector point representing center of footprint";
:long_name = "latitude";
:_FillValue = NaNf; // float
:units = "degrees north";
:_ChunkSizes = 1015736U; // uint

float lon(nobs=11173095);
:_FillValue = NaNf; // float
:long_name = "longitude";
:comment = "vector point representing center of footprint";
:units = "degrees east";
:_ChunkSizes = 1015736U; // uint

int yymmddHH(nobs=11173095, ntc=4);
:long_name = "time point of observation";
:comment = "time coordinate given as UTC";
:units = "yymmddHH";
:_ChunkSizes = 1596157U, 1U; // uint
```

For additional information please contact TU Wien: emanuel.bueechi@geo.tuwien.ac.at

AMSR-2 VOD

AMSR-2 VOD is obtained applying the Land Parameter Retrieval Model (LPRM) v6 using horizontal and vertical polarized brightness temperature from AMSR-2 data (for further information on LPRM: van der Schalie et al., 2016).

Per region, one netcdf file is provided: AMSR2_lm.nc, AMSR2_re.nc, AMSR2_so.nc

The final contains all information obtained from LPRM, including VOD and surface soil moisture for the different bands (C1, C2, X, Ku), Scantime, Flags, lat, lon, Surface temperature, and time. As an additional information, a noise estimation was calculated for VOD of C1, C2 and X Band. They are stored as *VOD_C1_noise*, *VOD_C2_noise*, and *VOD_X_noise*.

The data is stored on a regular 0.25° grid, where the lat and lon vectors represent the centre of the cell. The variables *lat_vertices* and *lon_vertices* show the respective corners of the pixels.

The dimension time provides the date when the data was taken. The variable SCANTIME shows for each pixel the exact moment (in seconds) when on the considered day the observation was measured. Hence to get the precise moment SCANTIME needs to be added to time. This is what the variable *yymmddHH* does. The mean of SCANTIME per day is taken and added to time. The values are rounded to hours.

Fill value is -9999. The additional information in the nc file of *_FillValue* can be neglected!

The individual variables are described in the nc file itself. Here, only few additional information is provided:

FLAGS: shows if the data can be used. The flags are described in the dataset and range from no data, over rain/desert to too dense vegetation. Can include combinations of multiple flags.

The data are stored such as :

dimensions:

```
lat = 16;  
lon = 56;  
time = 3106;  
nv = 4;
```

variables:

```
float lat(lat=16);  
  :_FillValue = NaNf; // float  
  :units = "degrees_north";  
  :long_name = "latitude";
```

```
float lon(lon=56);  
  :_FillValue = NaNf; // float  
  :units = "degrees_east";  
  :long_name = "longitude";
```

```
[...]
float lat_vertices(lat=16, nv=4);
:units = "degrees_north";
:comment = "ordering SW/SE/NE/NW";
:_ChunkSizes = 16U, 4U; // uint
:_FillValue = NaNf; // float
:long_name = "latitude_of_pixel_corners";

int yymmddHH(time=3106, ntc=4);
:units = "yymmddHH";
:long_name = "time point of observation";
:comment = "time coordinate given as UTC";
:_ChunkSizes = 3106U, 4U; // uint

float SCANTIME(lat=16, lon=56, time=3106);
:_FillValue = NaNf; // float
:FillValue = -9999; // int
:units = "seconds";
:long_name = "Observation time on UTC day of overpass";
:_ChunkSizes = 8U, 28U, 1553U; // uint
```

```
[...]

float VOD_C1(lat=16, lon=56, time=3106);
:long_name = "Vegetation Optical Depth from AMSR2 C1-band";
:_ChunkSizes = 8U, 28U, 1553U; // uint
:_FillValue = NaNf; // float
:FillValue = -9999; // int
:units = "-";
```

```
float VOD_C1_noise(lat=16, lon=56, time=3106);
:FillValue = -9999; // int
:long_name = "Noise of Vegetation Optical Depth from AMSR2 C1-band";
:_ChunkSizes = 8U, 28U, 1553U; // uint
:_FillValue = NaNf; // float
:units = "-";
```

```
[...]

// global attributes:
}
```

Literature:

R.van der Schalie, Y.H. Kerr, J.P. Wigneron, N.J. Rodríguez-Fernández, A. Al-Yaari, R.A.M.de Jeu, Global SMOS Soil Moisture Retrievals from The Land Parameter Retrieval Model, International Journal of Applied Earth Observation and Geoinformation, Volume 45, Part B, 2016, Pages 125-134, <https://doi.org/10.1016/j.jag.2015.08.005>.

For additional information please contact TU Wien: emanuel.bueechi@geo.tuwien.ac.at

MODIS LST and PRI

- Daily HDF files were downloaded from <https://e4ftl01.cr.usgs.gov/>
- LST was taken from MYD11A1 product (AQUA). It was converted to Geotiff format
- PRI was estimated using MYDOCGA product (AQUA) and making use of band 11 (526-536 nm) and band 12 (546–556 nm). It was converted to Geotiff format.

$$PRI = \frac{\rho_{11} - \rho_{12}}{\rho_{11} + \rho_{12}}$$

- Finally, a big netcdf file was created per each variable and region using the Geotiff images from 2011 to 2020.
- A quality control was applied, i.e excluding pixels with poor quality [clouds or other reasons (NaN values)].
- Data description:
 - Daily Resolution
 - Spatial Resolution: 1 KM
 - **Option 1:** Sinusoidal projection (this is the original Coordinate Reference System in MODIS). Coordinates expressed in meters (X,Y). Reprojection technique was not applied to keep the maximum level of accuracy and consistency w.r.t. the original data.
 - **Option2:** Reprojection to WGS84 coordinate system. Coordinates expressed in latitude and longitude.

File Structure of netcdf file (see below). The variables are:

- **crs** (Coordinate Reference System information): WGS84 or SINUSOIDAL
- **latitude** (or X coordinate)
- **longitude** (or Y coordinate)
- **yymmddHHMMSS**(Time vector)
- **LST** (Land Surface temperature) or **PRI** (Photochemical Reflectance Index)

Status:

- MAJADAS LST is ready from 2011 to 2020
- Sodankyla LST is ready from 2011 to 2020
- Reusel LST is ready from 2011 to 2020
- MAJADAS PRI is ready from 2011 to 2020
- Reusel PRI is ready from 2011 to 2020

- Sodankyla PRI is ready from 2011 to 2020

EXAMPLE: GENERATED NETCDF FILE

Source:

LST_ESTIMATE_MAJADAS.nc

Format:

netcdf4

Dimensions:

time = 3651 (UNLIMITED)
rows = 438
columns = 658
lon = 658
lat = 438
t1 = 6

Variables:

crs

Size: 1x1

Dimensions:

Datatype: char

Attributes:

grid_mapping_name = 'latitude_longitude'
long_name = 'CRS definition'
longitude_of_prime_meridian = 0
semi_major_axis = 6378137
inverse_flattening = 298.2572
spatial_ref= 'GEOGCS["WGS 84",DATUM["WGS_1984",SPHEROID["WGS 84",6378137,298.257223563,AUTHORITY["EPSG","7030"]],AUTHORITY["EP SG","6326"]],PRIMEM["Greenwich",0],UNIT["degree",0.0174532925199433], AUTHORITY["EPSG","4326"]]'
GeoTransform = '-9.092727616310089 0.01025883285760756 0 43.00000000000036 0 -0.01025883285760756 '

lon

Size: 658x1

Dimensions: lon

Datatype: double

Attributes:

_FillValue = -99999.0
standard_name = 'longitude'
long_name = 'longitude'
units = 'degrees_east'
Comments = 'First coordinate corresponds to the SW corner of the LST product'

lat

Size: 438x1

Dimensions: lat

Datatype: double

Attributes:

_FillValue = -99999.0
standard_name = 'latitude'
long_name = 'latitude'

units = 'degrees_north'
Comments = 'First coordinate corresponds to the SW corner of the LST product'

yymmddHHMMSS

Size: 6x3651
Dimensions: t1,time
Datatype: double
Attributes:
_FillValue = -99999.0

LST

Size: 438x658x3651
Dimensions: rows,columns,time
Datatype: single
Attributes:
_FillValue = -99999.0
Units = 'Kelvin'
Pixel Size = '926.625 meters';
Region = 'MAJADAS'
Orientation = 'North'
Corner Coordinates (degrees) = 'Upper Left (-9.0927276, 43.0000000)
Lower Left (-9.0927276, 38.5066312) Upper Right (-2.3424156,
43.0000000) Lower Right (-2.3424156, 38.5066312)'

Path Local server:

- LST: Z:\LCCstudy\MODIS\netCDF
- PRI: Z:\LCCstudy\MODIS\netCDF

S3 LAI, LCC, FAPAR, FVC

Sentinel-3 Vegetation traits: LAI, LCC, FAPAR, FVC based on GPR (Gaussian Process Regression) Machine Learning methods

Status: New V1.1 delivered (15th October 2021).

Previous: V1.0 (15th April 2021)

- Sodankyla already processed (2016-2020)
- Reusel already processed (2016-2020)
- Las Majadas already processed (2016-2020).

Error found: After 24th November 2020, data of two OLCI bands (OA01 and OA10) are provided with anomalous values on our study region, involving that our masking function filters these pixels and no maps are provided after this date.

Proposed format changes:

- NetCDF compression applied by nccopy tool, deflation level 5 (from 0 to 9)
- Fill value: Changed from NaN to -9999
- Datetime: This information is encapsulated in a 1D monotonic coordinate variable (Time). It is provided in days resolution. The daily acquisition time is set to 10:00:00, as the Mean Local Time of Sentinel 3 crossing. More info in: <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-3/satellite-description/orbit> Additionally, a yymmddHHMMSS 2D variable is added as proposed.

A daily NetCDF is converted from a daily GeoTiff and next merged from timesteps into one single NetCDF with a Time coordinate variable.

- 1D latitude and 1D longitude variables are provided together with footprint corner coordinates. 2D corners coordinates have been added additionally by our workflow.

Format: both .Geotif and .NetCDF Common Data Format.

Vegetation variables (VV): Total of 12 daily variables: 4 predictions + 4 uncertainties + 4 quality flags

NetCDF file variables: 1D CRS definition, Geo2D VV, 1D latitude, 1D longitude, 1D time (number of days since 2000-01-01), 2D latitude corners, 2D longitude corners, 2D timedata yymmddHHMMSS

Format of acquisition datetime: the products are daily generated at 10:00 am Mean Local Solar Time. One NetCDF file per year and variable (So also a separate file for uncertainties and quality flag)

Size of files: 2228x, 1670y pixels. Around 5-6Gb per year and variable

Image resolution: 300 meters

CRS: WGS84

Quality flag:

0: data are in range

1: data are out of range

Ranges:

LAI: 0 - 9.99

LCC: 0 - 100

FAPAR: 0 - 1

FVC: 0 - 1

We followed the next steps in order to obtain our daily GPR-based maps:

1. Retrieving GPR parameters from a S3-TOA and vegetation traits LUT by using ARTMO software
2. Implementing algorithms into GEE to obtain our products from S3 images.
3. Exporting 12 multibands maps in .tif format and stored in GEE assets
4. Downloading daily single bands to local repository
5. Conversion from .tif daily files into single .NetCDF per year, variable and site

NetCDF Metadata:

Driver: netCDF/Network Common Data Format

Files: FAPAR_ESTIMATE_SITE.nc, FAPAR_UNCERTAINTY_SITE.nc,
FAPAR_QUALITY_FLAG_SITE.nc, FVC_ESTIMATE_SITE.nc,
FVC_UNCERTAINTY_SITE.nc, FVC_QUALITY_FLAG_SITE.nc, LAI_ESTIMATE_SITE.nc,
LAI_UNCERTAINTY_SITE.nc, LAI_QUALITY_FLAG_SITE.nc, LCC_ESTIMATE_SITE.nc,
LCC_UNCERTAINTY_SITE.nc, LCC_QUALITY_FLAG_SITE.nc

Size is 2228, 1670

Coordinate System is:

GEOGCRS["WGS 84",

DATUM["World Geodetic System 1984",

ELLIPSOID["WGS 84",6378137,298.257223563,

LENGTHUNIT["metre",1]],

PRIMEM["Greenwich",0,

ANGLEUNIT["degree",0.0174532925199433]],

```
CS[ellipsoidal,2],
  AXIS["geodetic latitude (Lat)",north,
  ORDER[1],
  ANGLEUNIT["degree",0.0174532925199433]],
  AXIS["geodetic longitude (Lon)",east,
  ORDER[2],
  ANGLEUNIT["degree",0.0174532925199433]],
  ID["EPSG",4326]]
```

Data axis to CRS axis mapping: 2,1

Origin = (-8.502554164191272,43.000556020233255)

Pixel Size = (0.002694945852359,-0.002694945852359)

Metadata:

dimensions:

```
time = UNLIMITED; // (1531 currently)
```

```
lon = 2228;
```

```
lat = 1670;
```

```
nv = 4;
```

```
nobslat = 1670;
```

```
nobslon = 2228;
```

```
ntc = 6;
```

```
nobsdate = 1531;
```

variables:

```
int time(time=1531);
```

```
:standard_name = "time";
```

```
:long_name = "Time";
```

```
:units = "days since 2000-01-01 10:00:00";
```

```
:calendar = "standard";
```

```
:axis = "T";
```

```

        :_ChunkSizes = 512U; // uint
double lon(lon=2228);

        :standard_name = "projection_x_coordinate";

        :long_name = "longitude";

        :units = "degrees_east";

        :axis = "X";

        :comment = "centre of pixel";

        :_ChunkSizes = 2228U; // uint
double lat(lat=1670);

        :standard_name = "projection_y_coordinate";

        :long_name = "latitude";

        :units = "degrees_north";

        :axis = "Y";

        :comment = "centre of pixel";

        :_ChunkSizes = 1670U; // uint
char crs;

        :grid_mapping_name = "latitude_longitude";

        :long_name = "CRS definition";

        :longitude_of_prime_meridian = 0.0; // double

        :semi_major_axis = 6378137.0; // double

        :inverse_flattening = 298.257223563; // double

        :spatial_ref = "GEOGCS[\"WGS 84\",DATUM[\"WGS_1984\",SPHEROID[\"WGS
84\",6378137,298.257223563,AUTHORITY[\"EPSG\",\"7030\"]],AUTHORITY[\"EPSG\",
\"6326\"],PRIMEM[\"Greenwich\",0],UNIT[\"degree\",0.0174532925199433,AUTHORITY[\"EPSG\",
\",\"9122\"],AXIS[\"Latitude\",NORTH],AXIS[\"Longitude\",EAST],AUTHORITY[\"EPSG\",
\",\"4326\"]]]";

        :GeoTransform = "-8.502554164191272 0.002694945852358565 0
43.00055602023325 0 -0.002694945852358565 ";

float FAPAR_GREEN(time=1531, lat=1670, lon=2228);

```

```

:long_name = "fraction_absorbed_photosynthetically_active_radiation_prediction,";

:grid_mapping = "crs";

:_FillValue = -9999.0f; // float

:missing_value = -9999.0f; // float

:_ChunkSizes = 1U, 1U, 2228U; // uint

double lat_vertices(nobslat=1670, nv=4);

:units = "degrees_north";

:comment = "ordering SW/SE/NE/NW";

:long_name = "latitude_of_pixel_corners";

double lon_vertices(nobslon=2228, nv=4);

:comment = "ordering SW/SE/NE/NW";

:long_name = "longitude_of_pixel_corners";

:units = "degrees_east";

int yymmddHHMMSS(nobsdate=1531, ntc=6);

:units = "Year Month Day Hour Minute Second";

:comment = "time coordinates given as UTC";

:long_name = "time point of observation";

// global attributes:

:CDI = "Climate Data Interface version 1.9.9rc1 (https://mpimet.mpg.de/cdi)";

:Conventions = "CF-1.5";

:GDAL_AREA_OR_POINT = "Area";

:GDAL = "GDAL 3.0.4, released 2020/01/28";

:NCO = "netCDF Operators version 4.9.1 (Homepage = http://nco.sf.net, Code = http://github.com/nco/nco)";

:CDO = "Climate Data Operators version 1.9.9rc1 (https://mpimet.mpg.de/cdo)";

:PIXEL_SIZE_meters = "300 m";}

```

Above Ground Biomass

AGB data were provided by the CCI Biomass project using several sensors. Three dates are provided: 2010, 2017 and 2018. The 2010 maps was produced using Envisat SAR and ALOS PALSAR. The 2017 and 2018 maps were produced using a similar approach but Sentinel-1 is used instead of Envisat ASAR and ALOS-2 PALSAR-2 is used instead of ALOS PALSAR.

Per region we provide one netcdf file for every requested year :
'AGB_CCI_YEAR_REGION.nc'.

For *YEAR* we have 2010, 2017 and 2018 and for *REGION* Lasmajadas, Reusel and Sodankyla.

AGB data contain an array with a dimension of longitude X latitude with a resolution of 100 m. Data units are in tons/hectare (or Mg.ha-1).

Example of netcdf file generated for AGB data :

variables:

```
float agb(lat=5063, lon=6750);
    :long_name = "Above Ground Biomass";
    :grid_mapping = "crs";
    :units = "t/ha";
    :_FillValue = -9999.0f; // float
    :_ChunkSizes = 844U, 1125U; // uint

char crs;
    :grid_mapping_name = "latitude_longitude";
    :long_name = "CRS definition";
    :longitude_of_prime_meridian = 0.0; // double
    :semi_major_axis = 6378137.0; // double
    :inverse_flattening = 298.257223563; // double
    :spatial_ref = "GEOGCS[\"WGS 84\", DATUM[\"WGS 1984\", SPHEROID[\"WGS 84\", 6378137, 298.257223563, AUTHORITY[\"EPSG\", 7030]], PRIMEM[\"Greenwich\", 0], UNIT[\"degree\", 0.0174532925199433], AUTHORITY[\"EPSG\", 4326]]";
    :GeoTransform = "-8.5 0.0008888888888888889 0 43 0 -0.0008888011060635987 ";

double lat(lat=5063);
    :standard_name = "latitude";
    :long_name = "latitude";
    :units = "degrees_north";
    :_ChunkSizes = 5063U; // uint
```

```
double lon(lon=6750);
:standard_name = "longitude";
:long_name = "longitude";
:units = "degrees_east";
:_ChunkSizes = 6750U; // uint

// global attributes:
:Conventions = "CF-1.5";
:GDAL = "GDAL 2.3.0, released 2018/05/04";
:history = "Wed Feb 17 10:04:42 2021: GDAL Create( AGB_CCI_2010_Lasmajadas.nc, ... )";
:Resolution = "100 m";
```