Copernicus Global Land Operations
"Vegetation and Energy"
"CGLOPS-1"
Framework Service Contract N° 948120 – IPR - 2023

QUALITY ASSESSMENT REPORT

NORMALIZED DIFFERENCE VEGETATION INDEX (NDVI) 300M

VERSION 3

Issue 1.10

Organization name of lead contractor for this deliverable: VITO

Book Captain: Sarah Gebruers
Contributing Authors: Carolien Toté
Else Swinnen

Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Document Release Sheet

Book captain:	Sarah Gebruers (VITO)	Sign Sarah Gebruers	Date 27.02.2026
Approval:	Roselyne Lacaze (HYGEOS)	Sign Roselyne Lacaze	Date 13.03.2026
Endorsement:	Nadine Gobron (JRC)	Sign	Date
Distribution:	Public		

Change Record

Date	Page(s)	Description of Change	Release
07.11.2025	All	First version	I1.00
27.02.2026	140 - 155	Updated Annex 2: analysis extended to December 2025.	I1.10

TABLE OF CONTENTS

List of Figures

List of Tables

List of Acronyms

ACT	ACTual data
AD	Applicable Document
AOT	Aerosol Optical Thickness
ATBD	Algorithm Theoretical Basis Document
BA	Bare areas
BRDF	Bidirectional Reflectance Distribution Function
C	Collection
CAL/VAL	Calibration/Validation
CAMS	Copernicus Atmosphere Monitoring Service
CCD	Charge Coupled Device
CEOS	Committee on Earth Observing Satellites
CGLOPS	Copernicus Global Land Operations
COG	Cloud Optimized Geotiff
CRO	Cropland
DBF	Deciduous Broadleaf Forest
EBF	Evergreen Broadleaf Forest
EOS	Earth Observing System
EPSG	European Petroleum Survey Group
FOV	Field-of-View
FR	Full resolution
GBOV	Ground-Based Observations for Validation
GCP	Ground Control Points
GIO	GMES Initial Operations
GM	Geometric Mean
GMR	Geometric Mean Regression
GSD	Ground Sampling Distance
HER	Herbaceous cover
ICP	Instrument Calibration Parameter
INSPIRE	Infrastructure for Spatial Information in Europe
JRC	Joint Research Centre
L2A	Level2A
LANDVAL	LAND VALidation network
LC100	Land Cover at 100m resolution
LPV	Land Product Validation

LTAN	Local Time at Ascending Node
LTDN	Local Time at Descending Node
MBE	Mean Bias Error
MCD43A4	MODIS/Terra and Aqua Nadir BRDF-Adjusted Reflectance Daily L3
MERRA-2	Modern-Era Retrospective analysis for Research and Applications, Version 2
MODIS	MODerate resolution Imaging Spectroradiometer
MXF	Mixed Forest
NDVI	Normalized Difference Vegetation Index
NDVI_unc	NDVI uncertainty
NetCDF	Network Common Data Format
NIR	Near Infra-Red
NLF	Needleleaved Forest
NOBS	Number of observations
NRT	Near-real time
Oa	OLCI band a
OLCI	Ocean and Land Colour Imager
OSCAR	Optical Sensor Calibration with simulated Radiance
PROBA-V	Project for on-board autonomy – Vegetation instrument
PSF	Point Spread Function
PUM	Product User Manual
QA	Quality Assessment
QAR	Quality Assessment Report
QFLAG	Quality Flag
QWG	Quality Working Group
R^2	Coefficient of determination
RMSD	Root Mean Squared Deviation
RSRF	Relative Spectral Response Function
S10	10-daily Synthesis product
S3	Sentinel-3
SAIL	Scattering by Arbitrary Inclined Leaves
SHR	Shrubland
SLSTR	Sea and Land Surface Temperature Radiometer
SMAC	Simplified Method for Atmospheric Correction
SPOT-VGT	Satellite Pour l'Observation de la Terre – Vegetation
SQM	Scientific Quality Monitoring
SWIR	Short Wave Infrared

SZA	Solar Zenith Angle
TDS	Test Dataset
TOA	Top of Atmosphere
TOC	Top of Canopy
UTM	Universal Transverse Mercator
V	Version
VIIRS	Visible Infrared Imaging Radiometer Suite
VNIR	Visible and Near Infrared
VR	Validation Report
VZA	Viewing Zenith Angle
WGS	World Geodetic System

EXECUTIVE SUMMARY

The Copernicus Land Monitoring Service (CLMS) produces a series of qualified bio-geophysical products on the status and evolution of the land surface. The products are used to monitor vegetation, crops, water cycle, energy budget and terrestrial cryosphere. Production and delivery of the parameters take place in a timely manner and are complemented by the constitution of long-term time series.

This document presents the results of the validation of the 300m Normalized Difference Vegetation Index (NDVI) Version 3 (V3).

The time series of NDVI 300m V3 is based on Top-Of-Canopy (TOC) reflectances, normalized from Bidirectional Reflectance Distribution Function (BRDF) effects, acquired by PROBA-V and by the Ocean and Land Colour Imager (OLCI) onboard Sentinel-3 (S3) platform.

The NDVI 300m V3 differs from NDVI 300 V2 and NDVI 300m V1 in three major aspects (see [CGLOPS1_ATBD_NDVI300m-V3]):

(1) Input data: NDVI 300m V3 is derived from PROBA-V Collection 2 daily (S1) TOC data and from reprocessed Sentinel-3/OLCI TOC V2.3 data, whereas NDVI 300m V2 was derived from Sentinel-3/OLCI TOC V1 data, and NDVI 300m V1 from PROBA-V Collection 0 and Collection 1 10 daily (S10) products. NDVI 300m V2 and NDVI 300m V1 are only available for the Sentinel-3 resp. PROBA-V period.
(2) Compositing strategy: NDVI 300m V3 is corrected for BRDF effects using ReBeLS v1.6 which provides improved outlier detection and filtering of extreme geometries, as opposed to ReBeLS v1.3 that was used for NDVI 300m V2, and also different from NDVI 300m V1 which is a maximum value NDVI composite based on directional reflectances, updated every 10 days using a sliding window.
(3) Output quality information: In NDVI 300m V3, the number of snow observation, contained in the BRDF output layer, is propagated to the NDVI quality flags. Moreover, a quality flag is raised when the prior information on BRDF is based on interpolated values.

The quality assessment (QA) of the NDVI 300m V3 product is based on the periods 2019 (for comparison with NDVI 300m V1, and consistency between V3 derived from PROBA-V and from Sentinel-3/OLCI) and 2023 (for comparison with NDVI 300m V2) for consistency analysis, and on the full time series 2014 – 2024 for analysis of temporal stability and consistency.

The NDVI 300m V3 product is based on 10-daily BRDF corrected TOC reflectances, normalized to nadir viewing and sun zenith angle at local noon. The product contains an NDVI layer, its associated uncertainty (NDVI_unc), a layer containing the number of clear observations in the compositing

window (NOBS), and a quality flag (QFLAG) describing the BRDF inversion quality in the red and NIR bands, indicating whether (snow) observations are present, and whether the BRDF priors are interpolated.

The occurrences in the QFLAG, NDVI_unc, and NOBS layer of NDVI 300m V3 were investigated. The 'NOBS=0' quality flag and high NDVI uncertainties unsurprisingly mainly occurs during the Northern hemisphere winter, at high latitudes, and in regions with persistent cloud cover. There are no clear spatial or temporal patterns for the 'Warning' and 'Extreme warning' quality flags. The analysis indicates no relation between the NDVI values and their uncertainties, but the NDVI_unc exponentially decreases with increasing NOBS. Users are recommended to exclude NDVI values when the 'Extreme warning', 'Interpolated prior', or 'snow' flags are raised.

Intercomparison with NDVI 300m V1, NDVI 300m V2, NDVI 1km V3, and MODIS MCD43A4 V6.1 were performed with a selection of full resolution tiles, global sub-sampled images, and extractions over the LANDVAL V2 sites (Martínez-Sánchez et al., 2024).

It is concluded that the NDVI 300m V3 dataset shows high product completeness, good stability over time, and large statistical consistency with the reference datasets. Some inconsistencies are expected between product versions, due to differences in sensors and processing approaches, most importantly BRDF normalisation, but overall biases are minimal and mostly unsystematic. The agreement between PROBA-V and Sentinel-3/OLCI in NDVI 300m V3 is very good, resulting in a consistent 300m NDVI product from 2014 up until present.

1 BACKGROUND OF THE DOCUMENT

1.1 SCOPE AND OBJECTIVES

This document presents the results of the quality assessment of the 300m Normalized Difference Vegetation Index (NDVI) Version 3 (V3) product, available to users from 2014 onwards. NDVI 300m V3 was designed to provide a continuous time series integrating input data from both PROBA-V (2014 - 2018) and Sentinel-3/OLCI (2019 – onwards). NDVI 300m V3 is now for the whole time series based on bi-directional reflectance distribution function (BRDF) corrected surface reflectances (unlike previous versions, see §3.1).

The validation encompasses a robust and scientific analysis of the new version of the NDVI 300m product, with the objective to evaluate the product stability and performance. The NDVI 300m V3 is compared to the NDVI 300m V1 and NDVI 300m V2 products, and to BRDF-corrected NDVI from MODIS.

A tool is available to resample the NDVI 300m V3 to 1 km to extend the existing 1km NDVI archive to near-real time[1]. Therefore, the NDVI 300m V3 is also compared with the currently available NDVI 1km V3 dataset, which is based on BRDF-normalized surface reflectance from SPOT/VGT and PROBA-V for the time series 1999 – 2020.

1.2 CONTENT OF THE DOCUMENT

The document is structured as follows:

- Chapter 2 recalls the requirements and the expected performance.
- Chapter 3 describes the scientific questions to be answered in the quality assessment.
- Chapter 4 describes the methodology for the quality assessment, the metrics, and the evaluation criteria.
- Chapter 5 presents the results of the analyses for the NDVI 300m V3.
- Chapter 6 summarises the main conclusions, perspectives, and recommendations of the study.

[1] https://github.com/cgls/ResampleTool_notebook_Python

1.3 RELATED DOCUMENTS

1.3.1 Applicable documents

AD1: Part 2: Technical specifications of Framework Service Contract – Operation of the bio-geophysical variables systematic monitoring of the Global Land Component of the Copernicus Land Service 'CGLOPS' JRC/2023/OP/0273, 19th April 2023.

Available at https://etendering.ted.europa.eu/cft/cft-display.html?cftId=13795

1.3.2 Input

Document ID	Descriptor
CGLOPS1_ATBD_NDVI300m-V3	Algorithm Theoretical Basis Document of the NDVI 300m Version 3 product
CGLOPS1_QAR_S3-CloudMask	Evaluation report of OLCI and SLSTR cloud, cloud shadow and snow detection
CGLOPS1_ATBD_S3-AC-V1.1	Algorithm Theoretical Basis Document of the atmospheric correction for Sentinel-3/OLCI and SLSTR data
CGLOPS1_ATBD_BRDFCorrection300m-V1.6	Algorithm Theoretical Basis Document of the BRDF model retrieval Version 1.6
CGLOPS1_ATBD_NDVI300m-V2	Algorithm Theoretical Basis Document of the NDVI 300m Version 2 product
CGLOPS1_QAR_NDVI300m-V2	Quality Assessment Report of the NDVI 300m Version 2 product
GIOGL1_ATBD_NDVI300m-V1	Algorithm Theoretical Basis Document of the NDVI 300m Version 1 product
GIOGL1_QAR_NDVI300m-V1	Quality assessment report of the NDVI 300m Version 1 product
CGLOPS1_ATBD_NDVI1km-V3	Algorithm Theoretical Basis Document of the NDVI 1km Version 3 product
CGLOPS1_QAR_NDVI1km-V3	Quality assessment report of the NDVI 1km Version 3 product

CGLOPS1_ATBD_LC100mV3	Algorithm Theoretical Basis Document of the global Land Cover 100m map

Available at https://land.copernicus.eu/en/technical-library

1.3.3 Output

Document ID	Descriptor
CGLOPS1_PUM_NDVI300m-V3	Product User Manual summarizing all information about the NDVI 300m Version 3 product

Available at https://land.copernicus.eu/en/technical-library

1.3.4 External documents

Document ID	Descriptor
GCOS#245	The 2022 GCOS ECVs requirements, World Meteorological Organization, Geneva, Switzerland (2022)

Available at: https://library.wmo.int/idurl/4/58111

Sentinel-3/OLCI User Guide

Available at https://sentinel.esa.int/web/sentinel/user-guides/sentinel-3-olci

Sentinel-3 OLCI Level-1B Reprocessed BC004 Product User Guide

Available at https://user.eumetsat.int/s3/eup-strapi-media/S3_OLCI_Level_1_B_Reprocessed_BC_004_Product_User_Guide_3fd1fa3787.pdf

OPT-MPC Sentinel-3 OLCI and SYN Annual Performance Report – Year 2024

Available at https://sentiwiki.copernicus.eu/__attachments/1681931/OMPC.ACR.APR.006%20-%20S3%20OLCI-SYN%20Annual%20Performance%20Report%202024%20-%201.1.pdf

PROBA-V C2 PUM	User Manual of PROBA-V C2 data

Available at https://proba-v.vgt.vito.be/sites/probavvgt/files/downloads/PROBA-V_C2_Products_User_Manual.pdf

PROBA-V: https://proba-v.vgt.vito.be

2 PRODUCT REQUIREMENTS

According to the applicable documents [AD1], the requirements relevant for CLMS global NDVI product are described below.

2.1 SPECIFIC TECHNICAL DETAILS AND REQUIREMENTS

PRODUCT SPECIFICATION	
Geometric Properties:	
Spatial resolution	300m,
Geolocation precision	Better than 0.5 pixels
Coordinate position	Centre of the pixel
Geodetic datum	WGS84
Geographic projection	regular latitude/longitude grid
Geographic Coverage	Global
Temporal resolution	10-day period (dekad: days 1-10, 11-20, 21 end of month)
Timeliness	Within 2 days (optimally 1 day) after the end of each dekad

NDVI – NORMALIZED DIFFERENCE VEGETATION INDEX	
The Normalized Difference Vegetation Index (NDVI) is a dimensionless parameter that is chlorophyll sensitive and indicative for the greenness of the biomes and is expressed as the difference between the near-infrared and red bands, normalized to the summation.	
Uncertainty (2-sigma)	Threshold: 5% Goal: 2.5%
Stability (per decade, i.e. 10 years)	Threshold: 3% Goal: 1.5%

2.2 FURTHER REQUIREMENTS

2.2.1 Output product composition

Products may contain various information layers and ancillary information – the base reference for product packages are the operational products as on 01.03.2023.

2.2.2 Data Structure

Data coding[2] shall be compatible with the Global Land products as on 01.03.2023 and/or follow the INSPIRE specifications, where applicable.

Ancillary information shall be as currently used and include at least the following:

- The number of measurements per pixel used to generate any synthesis product
- The per-pixel date of the individual measurement or the start-end dates of the period covered
- Quality indicators, with explicit per-pixel identification of the cause of anomalous parameter result.

The product naming and filename conventions that are used in the Copernicus Global Land component production as on 01.03.2023 shall be followed. This may be adapted for complete product collections upon agreement with the contracting authority during the Framework.

2.2.3 Data format

To ensure interoperability with the current Global Land component (operational product data formats and archive data formats) and other Copernicus services, all datasets will be available in NetCDF. Additional format such as Cloud Optimized Geotiff (COG) or ZAR format can be proposed for production or could be requested by the Contracting Authority during the Framework Contract.

2.2.4 Uncertainties and Validation

Uncertainties indicated in the product specifications above follow the threshold and goal proposed by GCOS#245.

Uncertainty estimates should account for the error propagation uncertainty coming from input data though the retrieval algorithms as during the contract period, ESA plans to imbed uncertainties in the Sentinel-3 ground segment products and, for Sentinel-2, there is an offline tool to determine Level 1 uncertainties; these can be used for propagation in the production chain.

Validation of the products shall conform to at least the CEOS LPV standards. Wherever appropriate the bio-geophysical variables shall be validated and compared to CEOS CAL/VAL data sets and/or Ground-Based Observations for Validation (GBOV) of Copernicus Global Land component when biophysical parameters are available.

[2] Data coding is the provision for the number of bits, range of values, usage of reserved values, content of status map, conversion to physical values formulas.

2.2.5 Input data

Copernicus Sentinel data are available from https://dataspace.copernicus.eu

Bio-geophysical variable products should be based on common base reflectance data:

- Sentinel 3: Derived Top-of-Canopy reflectance may be brokered or produced as it is the case on 01.03.2023 under the CGLOPS contracts;
- Sentinel 2: the Global Land Sentinel-2 Global Mosaic (S2GM)[3] component provides temporal mosaic of surface spectral bands that can be brokered and/or directly used.

Up to and including 2019, products archives of the Global Land component have been based on SPOT VGT, PROBA-V, ENVISAT, MODIS, TOPEX/Poseidon, Jason-1, Jason-2, Jason-3, datasets, which are available through the https://dataspace.copernicus.eu.

Ancillary satellite data that is purchased through Copernicus and put at disposal of the Services is available through the Data Warehouse and will become available on the Copernicus Dataspace Ecosystem.

Ancillary data sets, other than satellite imagery described above, that might be required shall be the responsibility of the contractor.

2.2.6 Product delivery

Products shall be delivered to the Copernicus Land Component dissemination.

The Copernicus Data Space Ecosystem infrastructure will be used to provide access to the final map products.

[3] https://land.copernicus.eu/en/products/global-image-mosaic

3 SCIENTIFIC QUESTIONS

3.1 DIFFERENCES BETWEEN PRODUCT VERSIONS

The product under evaluation in the quality assessment report is the NDVI 300m V3 dataset. This dataset is created based on input data from two different sensors. From 2014 to 2018, NDVI 300m V3 is based on TOC reflectances from PROBA-V Collection 2 (C2) daily (S1) synthesis 300m data. From 2019 onwards, the product is based on TOC reflectances V2.3 from the OLCI sensor onboard Sentinel-3A and -3B (with future continuity ensured by the planned Sentinel-3C and -3D missions). For validation purposes, a test dataset of NDVI 300m V3 based on PROBA-V TOC reflectances covering the year 2019 is available. This period is used for quality assessment and comparison with the Sentinel-3/OLCI NDVI 300m V3 dataset. However, this test dataset is not distributed to users.

The PROBA-V and Sentinel-3/OLCI sensors have some differences (as explained in more detail in §4.2.5). To make the data from both sensors as consistent as possible, a few steps have been taken. Firstly, to minimise the influence of differences and shifts in overpass time (and thus illumination conditions) and observation geometry, related to the orbital drift of PROBA-V (see §4.2.5.1) and the switch from PROBA-V to Sentinel-3/OLCI (January/2019), a BRDF normalisation is applied on the TOC reflectances used as input to NDVI 300m V3 with ReBeLS v1.6 [CGLOPS1_ATBD_NDVI300m-V3]. This version of the ReBeLS algorithm contains an improved outlier detection, filtering of extreme geometries, and raises a quality flag when the MODIS MCD43P climatology prior is based on interpolated values (see CGLOPS1_ATBD_BRDFCorrection300m-V1). Secondly, to reach consistency between PROBA-V and OLCI, which have different spectral bands (see §4.2.5.5), a selection of narrow OLCI bands is combined into a single red and a single NIR band, comparable to the wide PROBA-V spectral bands. The optimal band selection resulted in bands Oa07 and Oa08 that are combined into one red band, and bands Oa16 and Oa18 that are combined into one NIR band (see Section 5 in CGLOPS1_ATBD_NDVI300m-V3). Additionally, the PROBA-V NDVI values are multiplied by a correction factor of 4.5% to minimise the bias between NDVI from both sensors.

To assess the quality of the NDVI 300m V3 dataset, it is compared to previous NDVI products. An overview of the CGLOPS 300m and 1km NDVI product versions that are currently available to the user is provided in Table 1. Inconsistencies between the NDVI 300m V1, V2, and V3 datasets are expected because of differences in processing approaches, especially the BRDF-normalisation, but also related to differences in surface reflectance input collections. These differences are discussed in Section 4.2.5.

Table 1: Algorithm differences between successive versions of the NDVI 300m and 1km products.

Spatial resolution	NDVI version	Algorithm summary	Time series available
300m	V1	Derived from PROBA-V C1 S10 maximum NDVI composite (i.e. directional reflectance) products, with incorporation of the Status Map and land mask.	Jan/2014 - Jun/2020 (only PROBA-V)
	V2	Angular corrected composite over land pixels, using ReBeLS v1.3, derived from Sentinel-3/OLCI data, with a 30-days composite window, updated every 10 days using a sliding window. The normalized TOC V1 reflectances from OLCI bands Oa7, Oa8, Oa9, and Oa10 are averaged to a mean(RED) reflectance and the OLCI bands Oa16, Oa17, and Oa18 are averaged to a mean(NIR) reflectance prior to NDVI computation.	From Jul/2020 onwards (only Sentinel-3/OLCI)
	V3	BRDF normalization is done on PROBA-V C2 S1 TOC data (up to December/2018) and Sentinel-3 TOC V2.3 (from January/2019 onwards), using ReBeLS v1.6. The number of snow observations in the BRDF compositing window ($Nobs_{snow}$) is propagated to the QFLAG. A quality flag is raised when the BRDF priors are based on interpolated values. The normalized TOC V2.3 reflectances from OLCI bands Oa7 and Oa8 are averaged to a mean(RED) reflectance and the OLCI bands Oa16 and Oa18 are averaged to a mean(NIR) reflectance prior to NDVI computation.	From Jan/2014 onwards (PROBA-V and Sentinel-3/OLCI)
1km	V3	Angular corrected composite, derived from SPOT/VGT C3 TOA segments and PROBA-V C1 TOA data; BRDF normalized surface reflectance, 10-days composite window (extendable to 16 days), updated every 10 days using a sliding window. A spectral harmonization between VGT1 and VGT2 is performed.	Jan/99 – Jun/2020 (VGT and PROBA-V)

It is important to note that the NDVI 300m V3 dataset on which this QAR is based (2014 – 2024), is generated with a different configuration than the approach that is used in the near-real time (NRT) processing: in case of off-line processing (i.e. input data are available both before and after the target product date) the accumulation window for BRDF correction comprises an entire year of TOC input data, extended with one month before and one month after the target period (resulting in a total span of 14 months), while the accumulation window in the NRT processing spans only 30 days. This

impacts the performance of the ReBeLS outlier detection, enhancing its ability to filter spurious observations as the length of the available time series increases. The effect of these different time windows was investigated in the QAR of NDVI 300m V2 (see CGLOPS1_QAR_NDVI300m-V2), and is repeated here for NDVI 300m V3 (see Annex 2: Back-processing vs NRT processing). From these analyses we conclude that there is no systematic bias between both processing modes for the NDVI, although NDVI uncertainties are much lower when the accumulation window is longer. The conclusions of this QAR for the back-processed NDVI product are thus valid for the NRT NDVI 300m V3 product.

3.2 RESEARCH QUESTIONS

From a user's perspective, one of the most relevant questions is the temporal stability over the product time series and consistency between PROBA-V and Sentinel-3/OLCI for NDVI 300m V3. As stated above, the latter is evaluated by consistency analysis of a test dataset of NDVI 300m V3 based on PROBA-V with the Sentinel-3/OLCI based NDVI 300m V3 over the year 2019. Temporal stability is analysed through evaluation of temporal smoothness, temporal pixel profiles and inter-annual precision, and by comparison with an external dataset.

Secondly, it is important to quantify and explain possible inconsistencies between the existing NDVI 300m V1 archive on PROBA-V Collection 1, the NDVI 300m V2 archive on Sentinel-3/OLCI, and NDVI 300m V3 based on the combination of PROBA-V Collection 2 and reprocessed Collection 4 Sentinel-3/OLCI.

In case users require a longer time series, the NDVI 300m V3 dataset can be extended in the past by including the time series of SPOT/VGT1, SPOT/VGT2 and PROBA-V (1999-2020) at 1km resolution. We therefore also investigate the consistency between the NDVI 300m V3 and the NDVI 1km V3 time series.

Furthermore, the NDVI 300m V3 is compared to an external reference, namely the MCD43A4 V6.1 product derived from MODIS on EOS/Terra and Aqua platforms. This allows evaluating the quality of the NDVI 300m V3 with respect to the user requirements in terms of acceptable differences with existing satellite-derived products.

The quality assessment of the BRDF-corrected NDVI 300m V3 product therefore focuses on the following scientific questions:

Q1. What are the temporal and spatial distributions of the quality flags? Which flags are used in the following analyses and are recommended for users?

Q2. What are the temporal variation and spatial distribution of the product completeness of NDVI 300m V3? How do the results compare to those of NDVI 300m V1, NDVI 300m V2, NDVI 1km V3, and MCD43A4 V6.1?
Q3. How do the NDVI 300m V3 Sentinel-3/OLCI and PROBA-V datasets compare visually to the NDVI 300m V1, NDVI 300m V2, NDVI 1km V3, and MODIS MCD43A4 V6.1 datasets?
Q4. What is the statistical and spatial consistency between PROBA-V and Sentinel-3/OLCI for NDVI 300m V3?
Q5. How does NDVI 300m V3 compare with the versions NDVI 300m V1 and NDVI 300m V2, spatially and statistically?
Q6. What is the statistical and spatial consistency with NDVI 1km V3 (including 1km data of SPOT/VGT1, SPOT/VGT2, and PROBA-V)?
Q7. What is the statistical and spatial consistency of NDVI 300m V3 in comparison with external reference MODIS data (MCD43A4 V6.1)?
Q8. What is the temporal consistency of NDVI 300m V3 in comparison to MCD43A4 V6.1? How do the results of this analysis compare with those of NDVI 300m V1, NDVI 300m V2, and NDVI 1km V3? What is the stability of the NDVI 300m V3 product?

4 QUALITY ASSESSMENT METHODS

4.1 OVERALL PROCEDURE

4.1.1 Introduction

The NDVI 300m V3 product is based on BRDF corrected TOC reflectances (TOC-r), normalised to nadir viewing and sun zenith angle at local noon. The product contains four output layers:

- the NDVI values (NDVI),
- their associated uncertainties (NDVI_unc),
- a quality flag (QFLAG) describing the BRDF inversion quality in the red and NIR bands (see Table 2 for bitwise interpretation), indicating the possible presence of (snow) observations, and whether the MCD43P BRDF priors are interpolated,
- and a layer containing the number of observations used in the BRDF inversion procedure, propagated from the quality information layer of the TOC-r (NOBS).

Details are given in the PUM [CGLOPS1_PUM_NDVI300m-V3]. In the first scientific research question (see Chapter 3), we investigate the occurrences in the QFLAG, NDVI_unc, and NOBS layers in NDVI 300m V3 and define which flags and/or uncertainty thresholds are used in the subsequent analyses.

Table 2: Bit number, description, and value of the NDVI QFLAG layer.

Bit	Description	Value
0	No observations found in the compositing window in, at least, one of the red or NIR bands	1
1	At least one of the observations in the compositing window is flagged as 'snow'	2
2	At least one TOC-r input band for red has quality 'warning'	4
3	At least one TOC-r input band for red has quality 'extreme warning'	8
4	At least one TOC-r input band for NIR has quality 'warning'	16
5	At least one TOC-r input band for NIR has quality 'extreme warning'	32
6	TOC-r out of range (TOC-r < 0 | TOC-r > 1) for at least one band	64
7	The BRDF MCD43P priors are gap filled	128

The NDVI 300m V3 datasets used for this QAR are global sub-sampled images (see §4.3.2), a selection of 10° x 10° tiles and 1° x 1° mini tiles at full resolution (FR) (see §4.3.3), and extractions

over the LANDVAL V2 (Martínez-Sánchez et al., 2024) sites (see §4.3.4). For the spatial and statistical consistency analyses, the years 2019 and 2023 are selected to have overlap with the reference products (see §4.2). The temporal analysis is done on the full time series from 2014 up to the end of 2024.

4.1.2 Criteria

Table 3 provides a description of the criteria under evaluation.

Table 3: Description of the criteria under evaluation.

Criterium	Description
Product completeness	Product completeness is linked with the absence of gaps and the occurrences of quality flags in the product, both in space and in time. Gaps in the product inputs are caused by cloud or snow contamination, bad illumination conditions (e.g. in winter), poor atmospheric conditions, or technical problems during image acquisition, and – if propagated to the final product – are generally considered as a severe limitation. We focus on the temporal evolution and spatial distribution of product completeness, the occurrences of quality flags, and on the frequency distribution of gap length. A user recommendation on the use of quality flags is formulated and applied in the subsequent analyses.
Spatial consistency	Spatial consistency refers to the realism and repeatability of the spatial distribution of retrievals over the globe, including the absence of artefacts (e.g., missing data, stripes, unrealistic values, etc.), based on expert knowledge. The analysis is based on qualitative visual checks at full resolution, and spatial distribution of validation metrics at global level (on a systematic subsample).
Statistical consistency	The evaluation of statistical consistency is based on (i) histograms of the overall bias between datasets, (ii) evaluation of the proportion of residuals between datasets below a goal (2.5%) and threshold (5%) uncertainty level[4], and (iii) the geometric mean regression between datasets and associated statistics.
Temporal consistency	The realism of temporal variations of the product is qualitatively assessed for point locations, well distributed over the globe. Spatio-temporal evolution of validation metrics is assessed through Hovmöller plots. Frequency histograms of temporal smoothness are compared between datasets. Product stability is assessed based on the inter-annual precision.

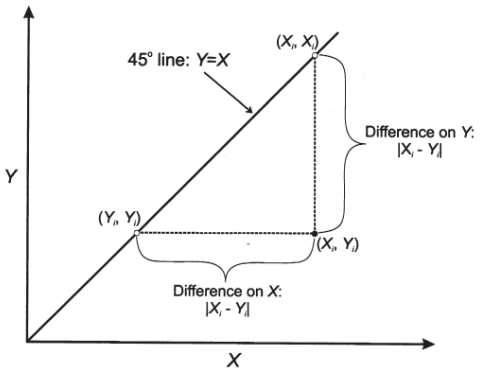
[4] according to the user requirement for NDVI (see Chapter 2 and [AD1])

4.1.3 Metrics

The validation metrics that are used for the assessment of statistical and temporal consistency are listed in Table 4. A selection of these validation metrics can be calculated per pixel over time (to evaluate spatial consistency) or per time step over grouped pixels (to evaluate temporal consistency). The dataset to be validated is X, and the reference datasets are Y.

Table 4: Validation metrics used for quality assessment of the NDVI.

Metric	Description
R^2	The coefficient of determination (R^2) indicates agreement or covariation between two datasets with respect to a linear regression model. It summarises the total data variation explained by this linear regression model. The result varies between 0 and 1, and higher R^2 values indicate higher covariation between the datasets: $R^2 = \left(\frac{\sigma(X,Y)}{\sigma(X)\cdot\sigma(Y)}\right)^2,$ with σ(X) and σ(Y) the standard deviation of X and Y, and σ(X, Y) the co-variation of X and Y. R^2 cannot be used for the comparison of the time series per pixel, due to temporal autocorrelation.
GM regression	The geometric mean (GM) regression (GMR) model is used to identify the relationship between two datasets of remote sensing measurements, with both datasets subject to noise. The GM regression model minimizes the sum of the products of the vertical and horizontal distances (errors on Y and X): $Y = a + b \cdot X$ By applying an eigen decomposition to the covariance metrics of X and Y, two eigenvectors v1 and v2 are obtained that describe the principal axes of the point cloud (Duveiller et al., 2016): $b = \frac{\lambda_1 - {\sigma_X}^2}{\sigma_{XY}}$ (GMR slope) $a = \bar{Y} - b \cdot \bar{X}$ (GMR intercept) with λ_1: the eigenvalue associated with the first eigenvector defining the principal axis σ_X: the standard deviation of X σ_{XY}: the covariance between X and Y $\bar{X}$: the mean value of X $\bar{Y}$: the mean value of Y

Metric	Description
Bias	The bias is defined as the actual difference between two data points: $$bias = X - Y$$
MBE or Accuracy (A)	The Mean Bias Error (MBE) or Accuracy (A) is defined as the actual difference or bias between two datasets, including positive and negative differences between observations: $$MBE = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i) = \bar{X} - \bar{Y}$$
STD or Precision (P)	The standard deviation (STD) of the bias or Precision (P) is a measure for repeatability: $$STD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - Y_i - A)^2}$$
RMSD or Uncertainty (U)	The Root Mean Squared Deviation (RMSD) or Uncertainty (U) measures how far the difference between the two datasets deviates from zero (Figure 1). $$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2}$$  **Figure 1: The difference between X and Y is calculated based on the distance between the point (X_i, Y_i) and the 1:1 line** (Ji and Gallo, 2006)**.**
MD	The median bias or median deviation is defined as the median of the difference between two datasets: $$MD = median(X - Y)$$

Metric	Description
MAD	The median absolute deviation is the median of the absolute difference between two datasets: $$MAD = median(\|X - Y\|)$$
Temporal smoothness	The temporal smoothness δ (Weiss et al., 2007)(Weiss et al., 2007) is evaluated by taking three consecutive observations and computing the absolute value of the difference between the centre $P(d_{n+1})$ and the corresponding linear interpolation between the two extremes $P(d_n)$ and $P(d_{n+2})$ as follows: $$\delta(d_n) = \left\| P(d_{n+1}) - P(d_n) - \frac{P(d_n) - P(d_{n+2})}{d_n - d_{n+2}}(d_{n+1} - d_n) \right\|$$
Time series noise	The time series noise can be estimated by averaging δ over the time series (Vermote et al., 2009): $$Noise = \sqrt{\frac{\sum_{n=1}^{N-2} \delta(d_n)^2}{N-2}}$$ Relative noise is computed by dividing the noise by the mean of the time series (Claverie et al., 2013): $$Relative\ noise = 100 \times \frac{Noise}{\bar{X}}$$
Inter-annual precision	The inter-annual precision is determined as the median absolute deviation of the 5th and 95th percentiles between consecutive years.
Stability	Stability is estimated as the slope of inter-annual precision over time.

The MBE, STD, and RMSD statistics are commonly referred to as APU (Accuracy, Precision, and Uncertainty) (Vermote and Kotchenova, 2008).

One of the tasks of this QAR is to determine if the NDVI 300m V3 data is conform with the uncertainty requirement (see §2.1). The uncertainty requirement is defined as a relative value and is here computed as the percentage of LANDVAL V2 pixels (see §4.3.4) for which the NDVI value lies within the goal or threshold distance from the GM regression line with respect to a certain reference dataset. The uncertainty for a certain site is thus defined as:

$$\boldsymbol{uncertainty} = \frac{\sqrt{[(a+b\cdot X)-Y]^2}}{|a+b\cdot X|}, \qquad \textbf{(Eq. 1)}$$

with *X* the NDVI 300m V3 NDVI value and *Y* the corresponding NDVI value of a reference dataset, and a and b are the intercept and slope of the GM regression between NDVI 300m V3 and the reference dataset.

4.1.4 List of procedures

The overall validation procedures are summarized in Table 5. The procedures are designed to answer the scientific questions (Q1 to Q8) as described in Chapter 3. Depending on the scientific question, emphasis of the procedures is on statistical, spatial and/or temporal consistency. Statistical consistency focuses on the distribution of values and the statistical relationship between two datasets. Spatial consistency investigates spatial artefacts, spatial distribution of validation metrics, and distribution of values per biome. Temporal consistency focuses on pixel profiles and temporal smoothness and stability.

Table 5: Procedures for the validation of NDVI 300m V3.

	Criterium	Method and/or Validation metric	Reference products	Spatial coverage	Temporal coverage
Q1	/	• Quantification (in %) of flag occurrences over land: temporal and spatial distribution	/	Global subsample	2019
Q2	Product completeness	• Temporal distribution of % of missing values or pixels classified as 'invalid' over land • Spatial distribution of % of missing values or pixels classified as 'invalid' over land	NDVI 300m V1/V2 NDVI 1km V3 MCD43A4 V6.1	Global subsample	2014 – 2024
		• Frequency distribution of the gap length		LANDVAL V2	2019/2023
Q3	Spatial consistency	• Visual comparison	NDVI 300m V1/V2 MCD43A4 V6.1	1°x1° tiles (FR) 10°x10° tiles (FR)	2019/2023
Q4	Statistical consistency	• Histogram of overall bias • GM regression analysis, overall and per biome	/	LANDVAL V2	2019
	Spatial consistency	• Histograms per biome • Spatial distribution of the validation metrics		Global subsample	
Q5	Statistical consistency	• Histogram of overall bias • GM regression analysis, overall and per biome • Uncertainty requirement analysis	NDVI 300m V1 NDVI 300m V2	LANDVAL V2	2019/2023
	Spatial consistency	• Histograms per biome • Spatial distribution of the validation metrics		Global subsample	
Q6	Statistical consistency	• Histogram of overall bias • GM regression analysis • Uncertainty requirement analysis	NDVI 1km V3	LANDVAL V2	2019
	Spatial consistency	• Histograms per biome • Spatial distribution of the validation metrics		Global subsample	

Q7	Statistical consistency	• Histogram of overall bias • GM regression analysis • Uncertainty requirement analysis	MCD43A4 V6.1	LANDVAL V2	2019
	Spatial consistency	• Histograms per biome • Spatial distribution of the validation metrics		Global subsample	
Q8	Temporal consistency	• Temporal pixel profiles • Frequency distribution of the temporal smoothness • Stability requirement analysis	MCD43A4 V6.1 NDVI 300m V1/V2 NDVI 1km V3	LANDVAL V2	2014 – 2024
		• Spatio-temporal evolution of the validation metrics (Hovmöller plots)		Global subsample	

4.2 SATELLITE REFERENCE PRODUCTS

This section lists the reference NDVI products that are used in the quality assessment, i.e. previous CGLOPS product versions NDVI 300m V1 (§4.2.1), NDVI 300m V2 (§4.2.2), and NDVI 1km V3 (§4.2.3), and an external reference dataset derived from MODIS MCD43A4 V6.1 (§4.2.4). Next, a paragraph is dedicated to the factors related to satellite/sensor and processing that affect NDVI agreement between different datasets (§4.2.5).

4.2.1 NDVI 300m V1

The NDVI 300m V1 is a standard 10-day directional NDVI synthesis product, based on maximum value compositing over a 10-day period. The time series of NDVI 300m V1 is derived from Collection 1 PROBA-V (Toté et al., 2018)(Toté et al., 2018)(Toté et al., 2018) 10-daily TOC composites (January 2014 – June 2020). More details on the NDVI 300m V1 are available in [GIOGL1_ATBD_NDVI300m-V1].

The PROBA-V based NDVI 300m V1 datasets used in the validation are global sub-sampled images, a selection of 1° x 1° and 10° x 10° tiles at full resolution, and the LANDVAL V2 sites over the year 2019 for statistical and spatial analyses, and over the full time series (January 2014 – June 2020) for the temporal analysis.

The NDVI 300m V1 dataset has no separate QFLAG layer and is therefore only filtered based on the valid range in the NDVI data itself. This excludes missing data, clouds, snow, sea, and background pixels with unscaled Digital Number (DN) values from 251 to 255, respectively [GIOGL1_ATBD_NDVI300m-V1].

4.2.2 NDVI 300m V2

The NDVI 300m V2 products are BRDF-normalized NDVI products derived from Sentinel-3/OLCI TOC reflectance [CGLOPS1_ATBD_S3-AC-V1, CGLOPS1_ATBD_BRDFCorrection300m-V1]. The time series of NDVI 300m V2 spans from July 2020 to the present. The dataset is described in more detail in [CGLOPS1_ATBD_NDVI300m-V2].

The Sentinel-3/OLCI based NDVI 300m V2 datasets used in the validation are global sub-sampled images, a selection of 1° x 1° and 10° x 10° tiles at full resolution, and the LANDVAL V2 sites over the year 2023 for statistical and spatial analyses, and over the period July 2020 – December 2024 for the temporal analysis.

The NDVI 300m V2 data is filtered based on the QFLAG layer. Pixels with missing TOC-r, red or NIR ‘extreme warning’, TOC-r out of range, and/or sea pixels are excluded from the analyses (i.e.

pixels for which bit 1, 3, 5, 6, and/or 7 are set), in line with user recommendations [CGLOPS1_QAR_NDVI300m-V2].

4.2.3 NDVI 1km V3

The NDVI 1km V3 products are BRDF-normalised NDVI products generated from the SPOT/VEGETATION C3 VGT-P and PROBA-V C1 L2A Top-of-Atmosphere (TOA) retrievals. The time series of NDVI 1km V3 includes data from different sensors, i.e. SPOT4/VGT1 (May 1998 – January 2003), SPOT5/VGT2 (February 2003 – December 2013), and PROBA-V (January 2014 – June 2020). The SPOT4 and SPOT5 satellites have a local equator-crossing time of around 10:30 AM (see §4.2.5.1). The overpass time for PROBA-V at launch was 10:45 AM. Since PROBA-V has no on-board propulsion, natural drift has led to consequent earlier overpass times, till around 9:23 AM by the end of the mission in July 2020. Data of SPOT/VEGETATION Collection 3 (Toté et al., 2017) and PROBA-V Collection 1 (Toté et al., 2018) TOA orbital segments are used in the NDVI 1km V3 processing line. A spectral harmonization is applied on the NDVI derived from SPOT/VGT1. This is explained in more detail in [CGLOPS1_ATBD_NDVI1km-V3].

The PROBA-V based NDVI 1km V3 datasets used in the validation are global sub-sampled images and the LANDVAL V2 sites over the year 2019 for statistical and spatial analyses. For the temporal consistency analysis, LANDVAL extractions (temporal profiles) and global subsamples (Hovmöller plots) for the period 2014 – 2024 are used.

NDVI 1km V3 pixels with no clear observations, failed BRDF inversion, or sea pixels are excluded from further analyses. Only pixels with QFLAG equal to 3 or 8 are kept, in line with user recommendations [CGLOPS1_QAR_NDVI1km-V3].

4.2.4 MODIS MCD43A4 V6.1 NDVI

The MODerate Resolution Imaging Spectroradiometer (MODIS) is mounted on two platforms, Earth Observing System (EOS)-Terra and Aqua. The MODIS instruments on Terra and Aqua capture data in 36 bands with a spatial resolution of 250m (2 bands), 500m (5 bands) and 1km (29 bands). Both instruments image the same area on Earth approximately three hours apart: Terra has a morning equator overpass (descending node) at around 10:30 local time, while Aqua is timed to cross the equator (ascending node) at approximately 13:30 local time (see §4.2.5.1).

The MCD43A4 Version 6.1 dataset is a nadir BRDF-adjusted reflectance at 500m resolution, adjusted to the solar zenith angle at local solar noon (Schaaf et al., 2011, 2002). The MCD43A4 reflectance represents the best characterization of the surface possible from the inputs available over a 16-day

period from both Terra and Aqua data. The NDVI is calculated by combining the reflectances of band 1 (red) and 2 (NIR).

MCD43A4 is provided in Sinusoidal projection, hence the derived NDVI was resampled to the CGLOPS products 1km grid in geographical projection system, based on a cubic resampling method. Since the MCD43A4 dataset is produced every day (with 16 days of acquisition, i.e. product date +/- 8 days), the temporal resolution was adapted to the temporal resolution of the CGLOPS NDVI products (10-daily) by considering three products per month, each with maximum overlap with the compositing window of the CGLOPS NDVI products. Quality flags (missing/invalid, snow, water) and uncertainties were derived from the related MCD43A2 dataset describing the overall condition of the BRDF product. Only data over land based on, at minimum, 2 observations for BRDF inversion were included. Also snow pixels were removed from further analysis.

The MODIS MCD43A4 NDVI datasets used in this study are global sub-sampled images, a selection of 10° x 10° tiles at full resolution (i.e. 1km resampled) and the LANDVAL V2 sites over the years 2019 and 2023 for statistical and spatial analyses. For the temporal consistency analysis, LANDVAL extractions (temporal profiles) and global subsamples (Hovmöller plots) for the period 2014 – 2024 are used.

4.2.5 Expected differences based on satellite/sensor and processing definition

The following paragraphs summarize how the agreement between different NDVI datasets under consideration is affected by several factors related to satellite/sensor and its processing.

4.2.5.1 Orbital characteristics and overpass time

The overpass time determines surface illumination conditions and hence the observed radiance by the satellite. Orbital drift is the process in which the satellite's overpass time is gradually shifted to a later or earlier time. While Sentinel-3, Terra, and Aqua are maintained at a constant overpass time, SPOT5 was not controlled in orbit since the beginning of 2013. For this sensor, the imagery is therefore acquired progressively earlier in the day, causing a systematic change in illumination conditions and hence relates to BRDF effects (see also §0). The gradual increase in sun zenith angle (SZA) leads to increased NDVI (Galvao et al., 2004; Sellers, 1985), as was shown in the inter-comparison between SPOT/VGT and MODIS (Toté et al., 2017) and in simulations of red and NIR reflectance and its corresponding NDVI with varying viewing zenith angle (VZA) and SZA (Swinnen and Dierckx, 2015) using the SAIL model (Verhoef, 1984). Since PROBA-V has no onboard propulsion, the satellite overpass time is variable throughout its operational life. Figure 2 compares the overpass time of SPOT4, SPOT5, PROBA-V, Sentinel-3, Terra, and Aqua. Table 6 summarizes satellite flight characteristics.

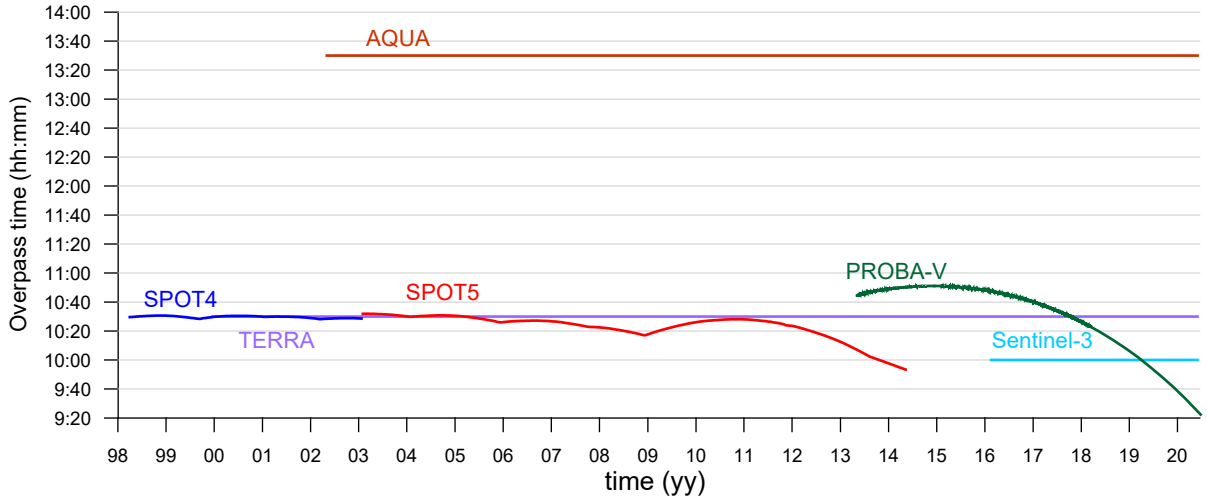


Figure 2: Comparison of equator crossing time. Local Time at Descending Node (LTDN) for SPOT4, SPOT5, PROBA-V, Sentinel-3, and Terra; Local Time at Ascending Node (LTAN) for Aqua.

Table 6: SPOT4, SPOT5, PROBA-V, Sentinel-3, Terra, and Aqua flight characteristics and equatorial overpass times.

	SPOT4	SPOT5	PROBA-V	Sentinel-3	Terra	Aqua
Altitude [km]	832	832	820	815	705	705
Inclination angle [°]	98.7	98.7	98.7	98.6	98.2	98.2
Swath [km]	2200	2200	2250	1270 [5]	2330	2330
Equator crossing time	~10:30	10:32 – 9:53 (02/2003 – 05/2014)	10:47 – 9:24 (10/2013 – 06/2020)	10:00	10:30	13:30
Stability of the platform	Minimal orbital drift, except from mid 2007 onwards	Minimal orbital drift, except from 2013 onwards	Orbital drift, no on-board propellant	No orbital drift	No orbital drift	No orbital drift

4.2.5.2 BRDF effects

Surface reflectance varies with illumination and viewing geometry for anisotropic surfaces, like most of the natural surfaces. A Bi-directional Reflectance Distribution Function (BRDF) describes this

[5] For OLCI onboard Sentinel-3

dependency for each surface type corresponding with a pixel. As a ratio, the NDVI implicitly contains a directional correction, because the effects are similar in the two bands (Vermote et al., 2009). Nevertheless, correction for BRDF effects is considered as one of the most important requirements for long-term and multi sensor time series analysis with satellite imagery (Cihlar et al., 2004, 1998).

The CGLOPS NDVI 300m V3, NDVI 300m V2, NDVI 1km V3, and the MCD43A4 NDVI time series are based on BRDF-normalized surface reflectances adjusted to nadir viewing. All of them are normalised to solar zenith angle at local solar noon. In contrast, CGLOPS NDVI 300m V1 is based on directional reflectances.

High frequency variations that are present in a time series of directional NDVI are likely to be reduced by BRDF correction, especially at mid-latitudes (Vermote et al., 2009). The effect of viewing geometry is larger for grasslands, and almost negligible for dense forests (Meyer et al., 1995). In general, NDVI tends to increase with higher solar zenith angles (Galvao et al., 2004; Sellers, 1985), but the effect is influenced by the canopy architecture. Tucker et al. (2005) reported the effect of varying solar zenith angles on the NDVI to be highest in the tropics for tropical forests, more moderate for less densely vegetated areas, and lowest at high latitudes. In general, NDVI from forward scattering direction are usually higher than those derived from the backscattering view (Breunig et al., 2011).

4.2.5.3 Scanning System

The differences in scanning system and optics between the VGT, PROBA-V, Sentinel-3/OLCI, and MODIS sensors affect the off-nadir spatial resolution of the imagery. The field-of-view (FOV) of Sentinel-3/OLCI is divided between five cameras. The instrument scans the Earth's surface using the push-broom method. Charge Coupled Device (CCD) arrays provide spatial sampling in the across track direction, while the satellite's motion provides scanning in the along-track direction. The OLCI swath is tilted 12.6° westwards to mitigate the negative impact of sun glint contamination over water surfaces.

Focusing on the red and NIR bands, the nominal (sub-nadir) resolution of the SPOT/VGT sensors is approximately 1x1 km, whereas this is 100m at nadir and up to 350m at the extremity of the swath for PROBA-V VNIR channels (Francois et al., 2014), 300m at nadir for Sentinel-3/OLCI and 250m for MODIS. For sensors with a swath width of more than 2000 km, there is global coverage with a near daily frequency at the equator, to a sub-daily frequency at higher latitudes. Since the swath width of Sentinel-3/OLCI is smaller, the combination of two Sentinel-3 OLCI sensing systems is needed to obtain similar coverage.

The VGT scanning system operates with an array of 1728 detectors positioned in one row perpendicular to the satellite orbital track simultaneously scanning one full scan line (i.e. push broom scanner). There is a fixed geometrical relationship between the detector elements in each scan line and the ground resolution measured by the sensor, accounting for a large part of off-nadir pixel deformation. Each individual detector measures the energy for a single ground resolution cell and thus the size and instantaneous FOV of the detectors determines the spatial resolution of the system. The VGT data are acquired essentially distortion free up to a 50° off-nadir angle if the Earth's curvature is not considered (Wolters et al., 2016).

The optical design of PROBA-V consists of three push broom cameras (Wolters et al., 2018). Each camera has two focal planes, one for the short-wave infrared (SWIR) and one for the visible and near-infrared (VNIR) bands. The VNIR detector consists of four lines of 5200 pixels. The SWIR detector is a linear array composed of three staggered detectors of 1024 pixels.

The MODIS sensors are across-track scanners that scan the Earth in a series of lines, perpendicular oriented to the direction of the orbit. Each line is scanned from one side to the other, using a rotating mirror placed in front of a sensor. The mirror sweeps with a constant angular velocity, resulting in the same angular resolution for every measurement. The sensors instantaneous FOV remains the same, and when sweeping away from the nadir position, the distance to the Earth increases and so does the ground surface resolved by the satellite. Because of the large swath width, the Earth's curvature adds an additional panoramic distortion to the off-nadir pixels. This leads to large off-nadir spatial deformations and the bow tie effect (Meyer, 1996).

4.2.5.4 Geometric Accuracy and Spatial Resolution

Mis-registration, together with the point spread function (PSF), induces image blur in an image time series (Meyer, 1996). Geometric accuracy is therefore another important issue to be considered when comparing time series of remote sensing data.

Geometric stability has been revised in the Sentinel-3/OLCI-A and OLCI-B Collection-4 reprocessing used as input for CGLOPS Sentinel-3 TOC reflectance V2.3. The geometric performance of both instruments is within specifications (i.e. below 0.5 pixel) and generally shows an improvement over operational data [Sentinel-3 OLCI Level-1B Reprocessed BC004 Product User Guide].

PROBA-V geolocation accuracy is maximized through a continuous and highly autonomous in-flight geometrical calibration/validation software, which estimates and monitors the Exterior Orientation parameters (boresight angles) and Interior Orientation deformations (CCD line of sight vectors) of each camera of PROBA-V (Mica et al., 2012). The geometrical parameters are estimated using a large

number of globally distributed ground control points (GCPs). The estimated geometrical calibration parameters are then used to update the Instrument Calibration Parameter (ICP) file. After a period of relatively high Absolute Location Error (ALE) with peak values up to ~150 – 250 m in the first months of the PROBA-V operations due to a star tracker issue, the absolute location error became much lower and was generally well below 100 m [6].

The geolocation accuracy of MODIS is reported to be approximately 50 m through the combined use of the onboard exterior orientation measurement system and the usage of a large set of GCPs (Wolfe et al., 2002).

For VGT1, high absolute, multi-temporal and multi-band registration accuracy is obtained using an elaborate database of GCPs. The absolute location accuracy from March 2000 to March 2003 was estimated to 190 m as root mean squared deviation (RMSD) (Sylvander et al., 2004). For VGT2, the stars tracker onboard SPOT5 allows for accurate geometric modelling without any need for GCPs, leading to a global performance of 170 m.

The native resolution of Sentinel-3/OLCI is 300 m at nadir. The native resolution of PROBA-V is 100 m at nadir, and 360 m at the edge of the swath. In contrast, SPOT/VGT reaches a ground sampling distance (GSD) of 1165 m at nadir, and 1700 m at the edge of the swath. MODIS reaches a spatial resolution of 250 m for the red and NIR bands. This means that, in PROBA-V derived NDVI, slightly more details are visible compared to Sentinel-3 NDVI and MODIS NDVI, and much more compared to SPOT/VGT.

4.2.5.5 Spectral Response Functions

Figure 3 shows the relative spectral response functions (RSRF) of the red and NIR spectral bands of VGT1, VGT2, PROBA-V, Sentinel-3/OLCI, and Terra/MODIS. Although the RSRFs overlap, the curves show differences in shape, central wavelength locations, bandwidths and degrees of overlap. The difference in RSRF-definition of the red and NIR channels is probably the largest source of difference between data of different sensors. In general, smaller bandwidths, such as for the Sentinel-3/OLCI and MODIS sensors, may lead to higher NDVI values. When red and/or NIR bands are positioned closer to the so-called ‘red edge’, where vegetation spectra show an abrupt change in reflectance (Dawson and Curran, 1998), this may lead to lower NDVI values. This is for example the case for SPOT4/VGT1 (with red band slightly overlapping the red edge) and Sentinel-3/OLCI band Oa16 (especially in comparison to the MODIS NIR band).Of particular interest is the spectral

[6] http://proba-v.vgt.vito.be/en/quality/platform-status-information/geolocation-accuracy

response in the transition range from the chlorophyll absorption band to the foliage reflection band (0.68 - 0.72µm) (Trishchenko et al., 2002), i.e. the so-called 'red edge'. This leads to differences in NDVI among the sensors for the same surface. Both the VGT1 and VGT2 red channels extend beyond the 0.7µm limit. This overlap with the red edge has a significant impact on the red reflectance, and results in lower NDVI values. The effect is largest for VGT1.

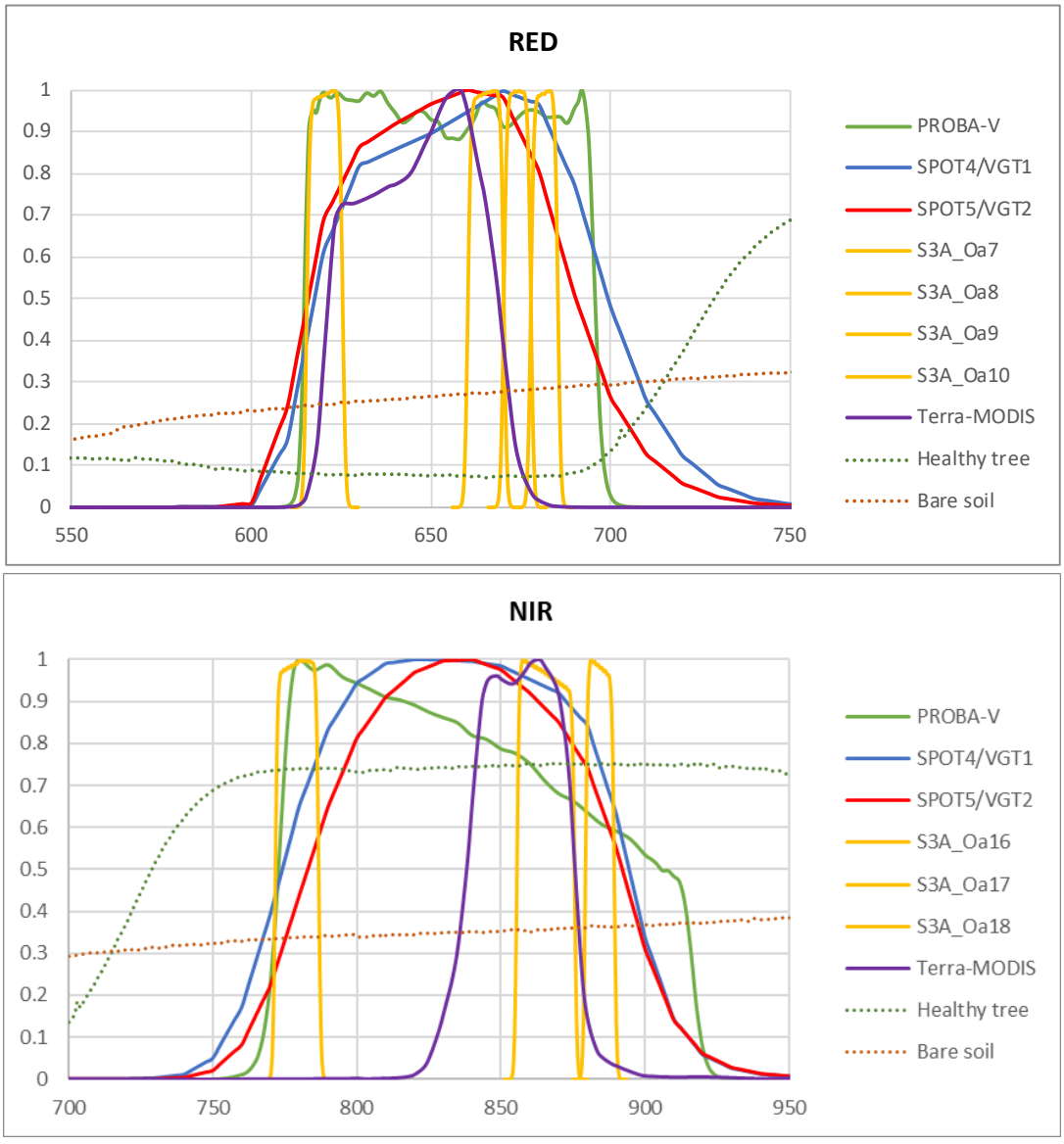


Figure 3: Relative spectral response functions (RSRF) for red (top) and NIR (bottom) spectral bands of PROBA-V, SPOT4/VGT1, SPOT5/VGT2, Sentinel-3/OLCI, and Terra/MODIS.

Several authors (Gonsamo and Chen, 2013; Trishchenko, 2009) analysed these differences and provide adjustment methods to reduce them. Based on the approach of Trishchenko (2009), in CGLOPS NDVI 1km V3 a spectral correction is applied to harmonize VGT1 red reflectance to VGT2, before NDVI calculation. No spectral correction is applied on PROBA-V (see [CGLOPS1_ATBD_NDVI1km-V3]), since, in this case, the differences in absolute calibration are larger than adjustments possible based on the RSRFs (see also §4.2.5.6).

For the Sentinel-3 based CGLOPS NDVI 300m V3 and CGLOPS NDVI 300m V2, the narrow Sentinel-3 bands are combined into single broad red and NIR bands to reach consistency. The band combination algorithm was revised and is therefore not the same in V3 and V2. In V2, all available red and NIR bands are used, i.e. Oa07, Oa08, Oa09, and Oa10 are combined into a broad RED band, and Oa16, Oa17, and Oa18 are averaged into a broad NIR band. In V3, all possible band combinations were investigated to find the best consistency with PROBA-V, leading to the selection of two red bands, Oa07 and Oa08, for the broad red band, and two NIR bands, Oa16 and Oa18, for the broad NIR band. More details on this analysis are available in [CGLOPS1_ATBD_NDVI300m-V2, CGLOPS1_ATBD_NDVI300m-V3].

4.2.5.6 Radiometric Calibration Accuracy

Radiometric calibration is necessary to allow the conversion of the signals measured by a satellite sensor to absolute radiance units. The radiometric performance of a sensor in the visible and near-infrared region usually degrades in orbit (Gutman et al., 1999).

The VGT, Sentinel-3/OLCI, and MODIS sensors, unlike PROBA-V, have an onboard calibration device, allowing accurate radiometric calibration (Fougnie et al., 2000; Henry and Meygret, 2000; Xiong and Barnes, 2006). Sentinel-3/OLCI-A has been reported to show an excess of brightness of around 2-3% in the red and NIR bands, while radiometric performance for OLCI-B is within the 2% requirement [OPT-MPC Sentinel-3 OLCI and SYN Annual Performance Report – Year 2024]. Although in Collection-4 reprocessing the radiometric stability has been revised, this OLCI-A bias was not corrected for. An OLCI-A bias correction is therefore applied in the CGLOPS TOC reflectance V2.3 pre-processing chain. Lachérade et al. (2013) reported SPOT5/VGT2 and MODIS calibrations to show biases of 3% for the red and NIR bands (MODIS being brighter). Similarly, VGT was reported to show TOA reflectances below the simulations over calibration target Libya4 (-3.1% for red, -2.5% for NIR), while the opposite is true for Aqua/MODIS (+0.6% for red, +1.6% for NIR) (Govaerts et al., 2013). Updated absolute calibration parameters were generated for the reprocessing of the entire VGT archive (resulting in Collection 3), aiming to improve the stability of the instrument

responses (Toté et al., 2017). The last reprocessing of MODIS (resulting in Collection 6.1) (Doelling et al., 2015) includes enhanced radiometric calibration and adjustments for sensor degradation[7].

Due to the absence of on-board calibration devices, the radiometric calibration and stability monitoring of the PROBA-V instrument relies solely on vicarious calibration. Therefore, the Optical Sensor CAlibration with simulated Radiance (OSCAR) Calibration/Validation facility (Sterckx et al., 2013) was developed for the PROBA-V mission. The PROBA-V archive was reprocessed (resulting in Collection 1) to improve and harmonize radiometric calibration (Toté et al., 2018). In the latest archive reprocessing, that was initiated after the operational phase of PROBA-V and resulted in Collection 2, used as input for NDVI 300m V3, absolute calibration and equalization coefficients were updated, with small impacts on TOA reflectances (< 0.2%) (Toté et al., 2023).

It is to be noted that SPOT/VGT, PROBA-V, and Sentinel-3 are not radiometrically intercalibrated. Preliminary analysis has shown that, compared to Sentinel-3/OLCI-B, PROBA-V Collection 1 is slightly less bright in red (~0.8%) and NIR (~0.9%), but that differences are larger for OLCI-A (~1.4% for red and ~2.1% for NIR), related to the excess in brightness of OLCI-A (Sterckx and Adriaensen, 2019). Compared to SPOT/VGT Collection 3, TOA reflectances were estimated to be slightly higher for PROBA-V Collection 1, with differences between 1.6 % and 2.2 % for the red band, and between 0.8% and 3.6 % for the NIR band, depending on the PROBA-V camera (Sterckx et al., 2016). Since there were no major changes to the absolute radiometric calibration of PROBA-V in the latest reprocessing (Toté et al., 2023), similar differences are expected for PROBA-V Collection 2.

4.2.5.7 Atmospheric Correction

Due to the different spectral band definitions, the influence of atmospheric perturbation will be different for data from OLCI, PROBA-V, VGT, and MODIS. The NIR spectral bands from OLCI, PROBA-V, VGT, and MODIS were specifically designed to avoid the 0.935µm water vapor absorption band. Van Leeuwen et al. (2006) investigated the influence of Rayleigh scattering, water absorption, ozone absorption, and aerosol optical thickness (AOT) on multi-sensor NDVI (MODIS and VIIRS) and concluded that NDVI data continuity across sensors would be largely enhanced if adequate atmospheric corrections are applied, because different sensors are differently affected by the atmosphere. Accurate atmospheric correction largely depends on the knowledge about the atmospheric conditions at the time of image acquisition.

[7] https://landweb.modaps.eosdis.nasa.gov/data/userguide/MODIS_Land_C61_Changes.pdf

The VGT, PROBA-V (for NDVI 300m V1, NDVI 300m V3, and NDVI 1km V3) and OLCI (for NDVI 300m V2 and NDVI 300m V3) imagery are processed with the same method for atmospheric correction (Simplified Method for Atmospheric Correction, SMAC (Rahman and Dedieu, 1994)). For the VGT and PROBA-V derived products, identical atmospheric inputs are used, except for AOT. For atmospheric correction of Sentinel-3/OLCI, in addition to different AOT, also other atmospheric parameters are different: total column of water vapor and sea-level surface pressure come from the OLCI Level1 input data, and surface temperature comes from a monthly mean climatology of MERRA-2 [CGLOPS1_ATBD_S3-AC-V1].

Over densely vegetated surfaces (i.e. low red reflectances), higher AOT values will lead to lower TOC red reflectances and slightly higher TOC NIR reflectances (although the effect of AOT is larger on red compared to NIR), resulting in higher NDVI values.

CGLOPS NDVI 300m V2 and CGLOPS NDVI 300m V3 use SMAC with near-real time AOT information from Copernicus Atmosphere Monitoring Service (CAMS) and the MERRA-2 (Gelaro et al., 2017) monthly climatology as back-up. CGLOPS NDVI 1km V3 uses SMAC with a monthly AOT climatology derived from CAMS data as input, with a latitudinal function of AOT as fall back when the default AOT input results in negative TOC reflectances. Atmospheric correction in PROBA-V Collection 1 (input for NDVI 300m V1) is done based on image-retrieved AOT. Analysis within the PROBA-V Quality Working Group (QWG) has shown that image based AOT often leads to high AOT estimations, thus resulting in higher NDVI values. In PROBA-V Collection 2 (input for NDVI 300m V3), atmospheric correction is done using AOT inputs from MERRA-2.

The MODIS data is corrected for the effects of gaseous and aerosol scattering and absorption as well as adjacency effects caused by variation of land cover, atmosphere coupling effects, and contamination by thin cirrus. The inputs for the atmospheric correction are derived from the MODIS data itself (Vermote et al., 2002).

It is worth mentioning that in the CGLOPS Sentinel-3 TOC reflectance processing, input per-pixel TOA uncertainties are propagated through the atmospheric correction, and uncertainties related to the selection of the aerosol model are included in the output TOC products [CGLOPS1_ATBD_S3-AC-V1]. This is not the case for the standard PROBA-V products nor the CGLOPS processing chains based on SPOT/VGT or PROBA-V.

4.2.5.8 Cloud Masking

The cloud masking is done with different algorithms for the various sensors, because the method depends on the spectral bands of the sensor. For VGT and PROBA-V, only optical channels are used in the cloud detection (Wolters et al., 2018, 2016), which leads in general to a less accurate cloud

mask. It is known that the PROBA-V Collection 1 (for NDVI 300m V1) cloud detection shows large commission errors over bright surfaces, and high occurrences of omission errors at higher latitudes in the winter months. A better performing cloud detection algorithm (Gómez-Chova et al., 2017; Toté et al., 2021) was introduced in Collection 2 (for NDVI 300m V3).

For Sentinel-3/OLCI, the IdePix cloud masking algorithm is done based on the 21 OLCI bands (Wevers et al., 2019). The IdePix cloud mask generally performs well, although some omission errors over semi-transparent clouds and cloud edges and commission errors over bright surfaces (including clear-sky snow) are observed [CGLOPS1_QAR_S3-CloudMask].

MODIS has thermal infrared bands, which facilitate cloud, snow and ice detection. For MODIS, the cloud detection algorithm is described by Ackerman et al. (2010).

4.2.5.9 Temporal Compositing

CGLOPS NDVI 300m V1 is a 10-daily maximum NDVI value composite. Observations in days 1-10, 11-20, and 21-end of the month are considered to create three products per month. While minimizing atmospheric effects, maximum value compositing might result in the selection of observations away from nadir, where a relatively larger proportion of the surface appears to be vegetated (Meyer et al., 1995) (see also §0). In case of CGLOPS NDVI 300m V2, CGLOPS NDVI 300m V3, and NDVI 1km V3, BRDF inversion is applied on a 30-days (NDVI 300m V2/V3 in NRT processing), 365-days (NDVI 300m V2 in back processing), 14 months (NDVI 300m V3 back processing), or 10-days (NDVI 1km V3) accumulation period, preceding the last day of each dekad (i.e. day 10, day 20 and the last day of the month), whereby the more recent observations get a higher weight in the inversion. In contrast, MCD43A4 NDVI is based on BRDF inversion over an accumulation period fixed to 16 days, which, in principle, will result in a smoother profile.

4.3 REGIONAL/BIOME ASSESSMENT AND SPATIAL SUBSAMPLING

4.3.1 Stratification per biome

An aggregated version of the CGLOPS Global Land Cover at 100m, epoch 2018 (Buchhorn et al., 2019) was used to distinguish between major land cover classes at the global scale (Figure 4). The classes were aggregated according to the scheme in Table 7. The map at 300m spatial resolution is generated from the CGLOPS-LC100 discrete classification map. For each of the 60 global UTM zones, a downscaling from 100m to 300m was performed by aggregating following the mode of the discrete classes. In case of equal occurrence of discrete classes, a set of expert rules based on cover fractions which can be retrieved from the CGLOPS-LC100 documentation is applied [CGLOPS1_ATBD_LC100mV3]. A new global tiling grid at 300m resolution in EPSG:4326

(WGS84) is created to which each of the UTM zones is transformed following the best-available-pixel approach.

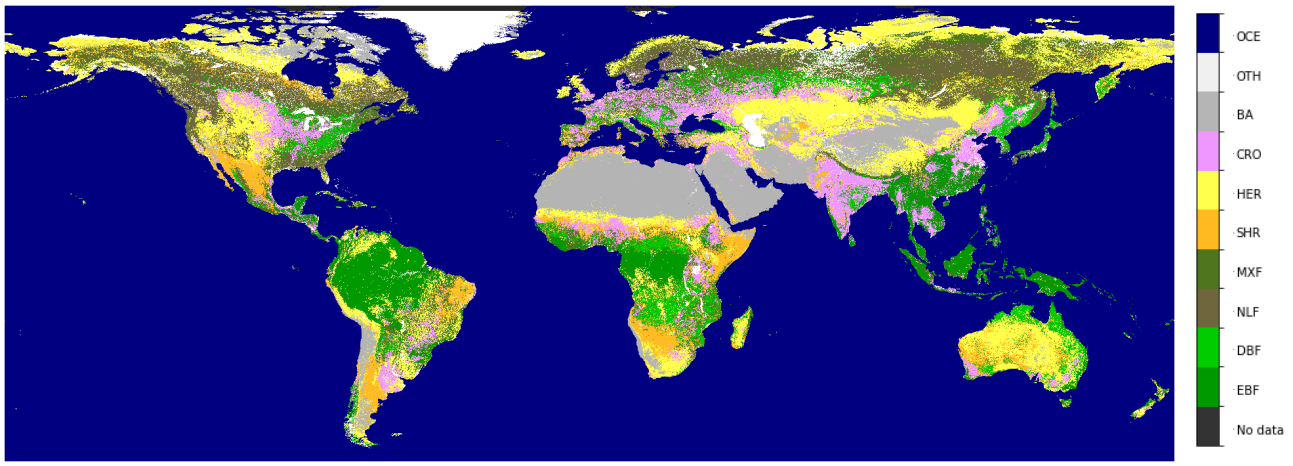


Figure 4: The CGLOPS LC100 epoch 2018 classification aggregated into eight classes.

Table 7: Aggregation scheme for CGLOPS Land Cover 100m classes into eight major biomes and proportion of each biome at global scale.

Abbreviation	Name	CGLOPS-LC100 classes	Proportion at global scale (%)
EBF	Evergreen broadleaf forest	112, 122	8.4
DBF	Deciduous broadleaf forest	114, 124	5.7
NLF	Needleleaf forest	111, 113, 121, 123	9.9
MXF	Mixed forest	115, 125	12.7
SHR	Shrubland	20	6.9
HER	Herbaceous	30	22.4
CRO	Crop	40	9.8
BA	Bare/sparse vegetation	60, 100	14.7
	Other (not considered in the analyses)	0, 50, 70, 80, 90, 116, 126	9.5

4.3.2 Global systematic subsample

Global images are systematically subsampled over the whole globe, taking the central pixel in each (arbitrary) window of 51 by 51 pixels for 300m products, and 17 by 17 pixels for 1 km products. This subsample is representative for the global patterns of vegetation and considerably reduces the processing time. No averaging is done to perform the subsampling.

To be able to perform pixelwise comparison at global scale between the NDVI 300m V3 product on the one hand and the NDVI 1km V3 and MODIS MCD43A4 V6.1 products on the other hand, an additional global subsample is extracted for the 300m data. This subsample is also created for every 51st pixel, but instead of using just one 300m x 300m pixel, the central 3 by 3 pixels are averaged to mimic a 1km pixel. The average is only computed when at least seven of the nine pixels have a valid NDVI value, otherwise the value is discarded.

4.3.3 10° x 10° tiles and 1° x 1° mini tiles at full resolution

To evaluate spatial patterns, a selection of the inter-comparison procedures can be applied on 10° x 10° tiles at full resolution (Figure 5). The CGLOPS Sentinel-3 processing tiling grid is used as reference.

To cover different climatic zones and vegetation conditions, the following tiles are considered:

- Northern Europe, tile X20Y01, coordinate ranges [80° - 70°N, 20° - 30° E]
- Western Europe, tile X18Y03, coordinate ranges [55° - 45° N, 0° - 10° E]
- South America, tile X10Y08, coordinate ranges [5° N - 5° S, 80°W – 70° W]
- Central Africa, tile X20Y07, coordinate ranges [15° - 5°N, 20° - 30° E]

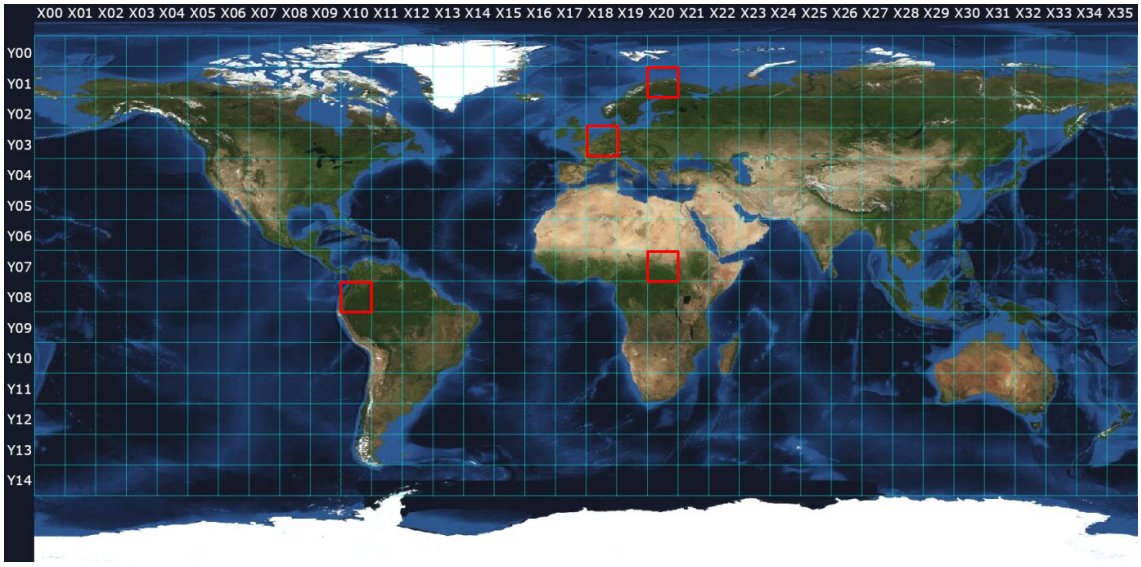


Figure 5: Grid tiles on world map[8]. Red squares indicate tiles under consideration.

For the Western Europe and Central Africa tiles, additionally one 1° x 1° mini tile is depicted for in detail visual comparison between 300m products. The following mini tiles are considered:

- Western Europe, region around Brussels, coordinate ranges [51° - 50° N, 4° - 5° E]
- Eastern Africa, South Sudan, coordinate ranges [9° - 8°N, 27° - 28° E]

4.3.4 LANDVAL V2 network

The LANDVAL V2 network of 2000 sites (Figure 6) is used to evaluate the statistical consistency among and temporal realism of the NDVI time series (Martinez-Sánchez et al., 2024). The sites are selected based on three criteria: i) same land cover and ecoregion (> 90% homogeneity) over an area of 3km x 3km; ii) distance between samples at least of 2 degrees on the hypotenuse; and iii) slope within the 3km x 3km area lower than 100m.

The LANDVAL V2 sites are used for the statistical analyses and for the generation of temporal plots. The 300m data extraction is based on 3x3 pixels covering the exact same footprint as the 1km pixel, calculating the average NDVI and NDVI_unc only when at least seven (out of nine) values are valid. Stratification per biome is based on the aggregated version of the CGLOPS Global Land Cover at 100m (see §4.3.1).

[8]Background image source: http://server.arcgisonline.com/arcgis/rest/services/ESRI_Imagery_World_2D/MapServer

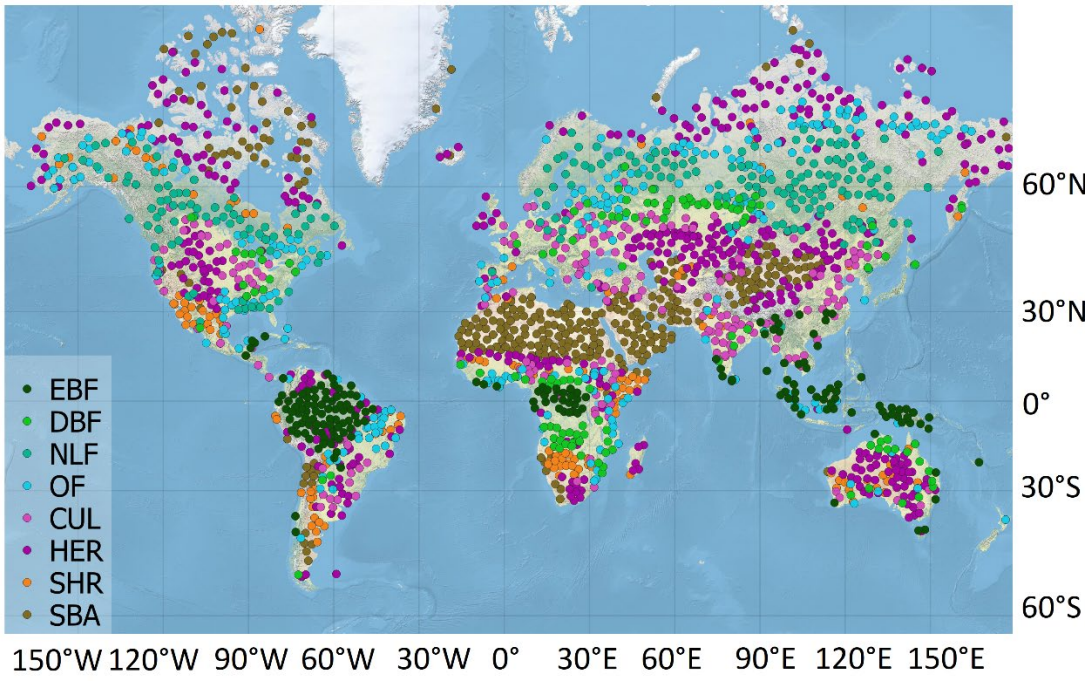


Figure 6: Global distribution of the selected LANDVAL V2 sites (Martínez-Sánchez et al., 2024)

5 RESULTS

5.1 Q1: QUALITY FLAGS – OCCURRENCES AND RECOMMENDATIONS?

5.1.1 Introduction

The scientific questions to be answered in this chapter are the following: what is the occurrence of quality flags (and NDVI uncertainties) in the NDVI 300m V3 product in terms of temporal and spatial distribution? Which flags are recommended for users and are applied in the subsequent analyses?

The analyses are performed over a one-year period (2019) for both PROBA-V and Sentinel-3/OLCI and are based on a global systematic subsample (see §4.3.2). Although the NDVI 300m V3 dataset switches from PROBA-V to Sentinel-3/OLCI at the start of 2019, the year 2019 was processed for both sensors and this overlapping year is used for intercomparison.

As described before, it is important to note that the NDVI 300m V3 dataset on which this QAR is based, is generated with a different configuration than the approach that is used in the NRT processing: in case of back processing (i.e. input data are available both before and after the target product date) the accumulation window for BRDF correction spans an entire year plus one month before and one month after (i.e. 14 months), while it spans only 30 days in NRT processing. We have investigated the differences between both processing modes in [Annex 2: Back-processing vs NRT processing] and concluded that there is no systematic bias between both processing modes for the NDVI, although NDVI uncertainties are much lower and the number of NOBS=0, warning, and extreme warning occurrences are higher when the accumulation window is longer. The results presented in the sections below are based on the back processed time series.

5.1.2 Quality flags

The NDVI 300m V3 product has an ancillary QFLAG layer which contains quality information for all land pixels. This layer can be used to filter out bad data on top of the values that are already flagged in the NDVI layer (i.e. DN values above 250). The NDVI and NDVI_unc are not computed, and flagged, in four cases:

- when the pixel is not above land,
- if the TOC-r values are out of range (this coincides with QFLAG bit 6 being set),
- when more than half of the observations in the compositing window are classified as snow (this is related to, but not the same as QFLAG bit 1),

- and when the following three conditions are met: the MODIS MCD43P prior is based on linearly interpolated values (QFLAG bit 7 is set), NOBS is equal to zero (QFLAG bit 1 is set), and the latitude is above 55°N.

Since NDVI will never be computed when QFLAG bit 6 is set, this bit is omitted in the following analysis and only the other bits are investigated.

The following results show the temporal and spatial distribution of the following occurrences:

- QFLAG bit 0 = 1: NDVI warning, NOBS equals zero, i.e. no observations are found in the compositing window in at least one of the red or NIR bands,
- QFLAG bit 1 = 1: presence of snow, at least one of the observations in the compositing window is flagged as snow,
- QFLAG bit 2 = 1 and/or QFLAG bit 4 = 1: 'Warning', i.e. at least one of the normalized TOC reflectance input bands for red or NIR has quality 'warning',
- QFLAG bit 3 = 1 and/or QFLAG bit 5 = 1: 'Extreme warning', i.e. at least one of the normalized TOC reflectance input bands for red or NIR has quality 'extreme warning'.
- QFLAG bit 7 = 1: the MCD43P climatology (BRDF priors) are gap filled by means of a linear interpolation.

It is noteworthy to mention that in NRT production, not all Sentinel-3 tiles are being processed. Whenever there is no data available for a certain tile over the whole accumulation window of 30 days, the BRDF processing is not triggered. This, in practice, only impacts tiles at high latitudes during winter months. Since these tiles don't contain valid NDVI data, it does not have an influence on the completeness of valid NDVI data, but when visualising the quality flags on a global scale these tiles do pop out as the flag values are not properly set for these data.

The temporal distribution of flag occurrences is shown in Figure 7. For both PROBA-V and Sentinel-3/OLCI, the occurrence of NOBS=0 is larger in the Northern hemisphere winter period than in the summer period. This is related to snow, cloud and/or shadow cover, and less favourable illumination conditions, resulting in a lower number of valid observations in the input data. The percentage of NOBS=0 is in general higher for Sentinel-3/OLCI than for PROBA-V. This is linked to the difference in uncertainties between both sensors. The NDVI uncertainties of Sentinel-3/OLCI are higher than those of PROBA-V. This impacts the number of observations that are used by ReBeLS, which has a selection criterion based on the uncertainty (see CGLOPS1_ATBD_BRDFCorrection300m-V1). The Sentinel-3/OLCI based products are therefore more likely to have the NOBS=0 quality flag raised. Additionally, more pixels are flagged as snow in the NDVI layer of NDVI 300m V3 based on PROBA-V and are therefore not considered in the quality flag analysis of the QFLAG layer. This is

also related to the commission of clear sky snow observations in the IdePix cloud masking (see §4.2.5.8 and [CGLOPS1_QAR_S3-CloudMask]).

The occurrences of warning and extreme warning remain roughly stable over time, with a slight increase during the period of March – June. The warning and extreme warning quality flags depend on the quality of the BRDF inversion which in turn depends on the comparison of predicted and observed TOC reflectances and their uncertainties by means of a mean z-score [CGLOPS1_ATBD_BRDFCorrection300m-V1]:

$$z = \frac{TOC_{obs} - TOC_{pred}}{\sqrt{(\sigma^2_{TOC_{obs}} + \sigma TOC^2_{pred})}}. \quad \textbf{(Eq. 2)}$$

Since the Sentinel-3/OLCI uncertainties are higher, the NDVI values will in general have a lower z-score and a better goodness of fit. Therefore, there are less occurrences of warning and extreme warning for this sensor. Overall, average values of warning and extreme warning for both Sentinel-3/OLCI and PROBA-V are however well below 5%.

The number of snow observations is also more or less stable over time. It decreases slightly during the Northern hemisphere summer, when less snow is expected. The number of snow pixels is higher in PROBA-V than for Sentinel-3/OLCI due to different cloud/snow masking algorithms. The IdePix algorithm, used for Sentinel-3/OLCI, is known to underperform causing clear-sky snow to be often misclassified as cloud (see §4.2.5.8 and [CGLOPS1_QAR_S3-CloudMask]).

The interpolated prior quality flag behaves similarly for PROBA-V and Sentinel-3/OLCI, as it is related to the BRDF prior and is independent of the sensor. Differences that appear in this analysis are related to the fact that only valid NDVI pixels are included and the filtering of valid/invalid pixels in the NDVI layer is different for PROBA-V and Sentinel-3/OLCI (see above). This also explains why the number of occurrences of the interpolated prior quality flag is not highest – as one would expect – during the Northern hemisphere wintertime, when at high latitudes most NDVI values are based on interpolated BRDF priors, because for those pixels NDVI is flagged in the NDVI layer itself (some based on this bit7 quality flag) and thus they are not included in the analysis shown in Figure 7. The increase of the interpolated prior quality flag during the period July – September is related to gaps in the tropical regions caused by high AOT values and clouds.

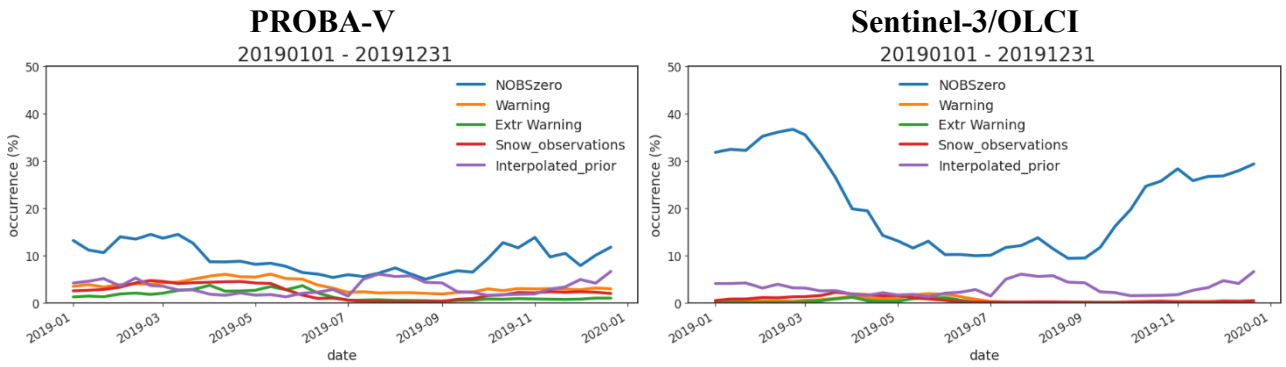


Figure 7: Temporal distribution of flag occurrences over land pixels with valid NDVI during the year 2019. NOBS=0 (blue), warning (orange), extreme warning (green), snow observations (red), and interpolated prior (purple) are shown for PROBA-V (left) and Sentinel-3/OLCI (right).

Figure 8 shows larger occurrences of NOBS=0 at higher latitudes (in agreement with the previous result) and in areas with persistent cloud cover (e.g. the wet tropics). Here as well, a large difference in NOBS=0 exists between PROBA-V and Sentinel-3/OLCI, related to the difference in uncertainties (see above). No clear patterns are discernible in the spatial distribution of warning and extreme warning, which seem more random except for some higher occurrences in mountainous areas (e.g. Tibetan plateau) and in the wet tropics. The number of snow observations is higher at higher latitudes (e.g. Greenland) and in mountainous regions, which is to be expected. PROBA-V clearly has more snow pixels related to the better cloud/snow masking algorithm. The spatial pattern of the interpolated prior quality flag is quasi-identical for PROBA-V and for Sentinel-3/OLCI, differences are due to pre-filtering of valid NDVI values based on the flags in the NDVI layer.

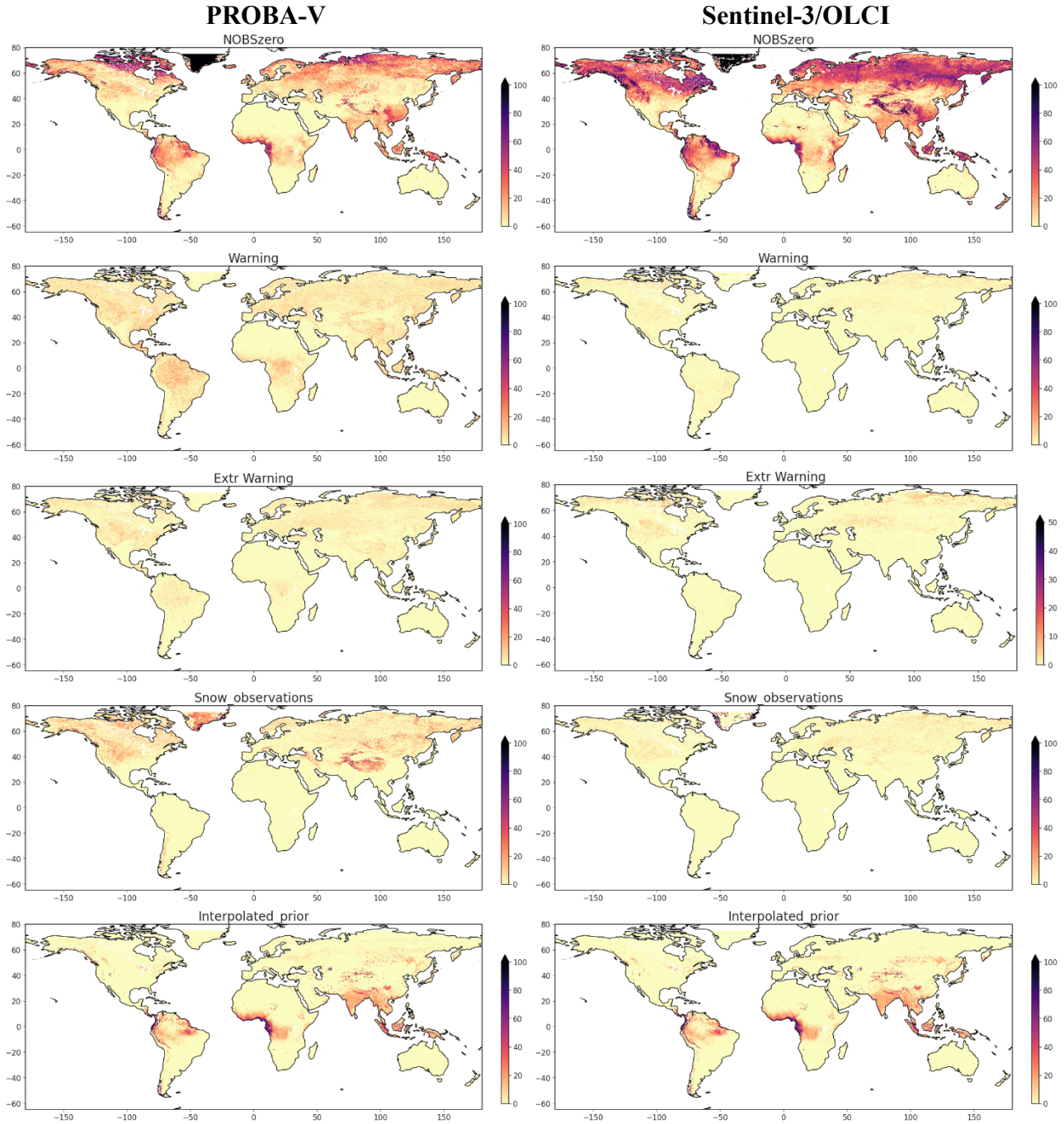


Figure 8: Spatial distribution of flag occurrences over land pixels with valid NDVI. From top to bottom the percentage of pixels with NOBS=0, warning, extreme warning, snow observations, and interpolated priors are shown for NDVI 300m V3 PROBA-V (left) and for NDVI 300m V3 Sentinel-3/OLCI (right) for the year 2019.

The distribution per biome type (Figure 9) shows no clear pattern for the occurrences of warning and extreme warning. They are lowest for BA and SHR and highest for EBF and the other forest biome types. NOBS=0 occurs more often in biomes located at higher latitudes such as NLF, MXF, and HER, and in the wet tropics EBF. The interpolated prior flag occurs mostly for EBF.

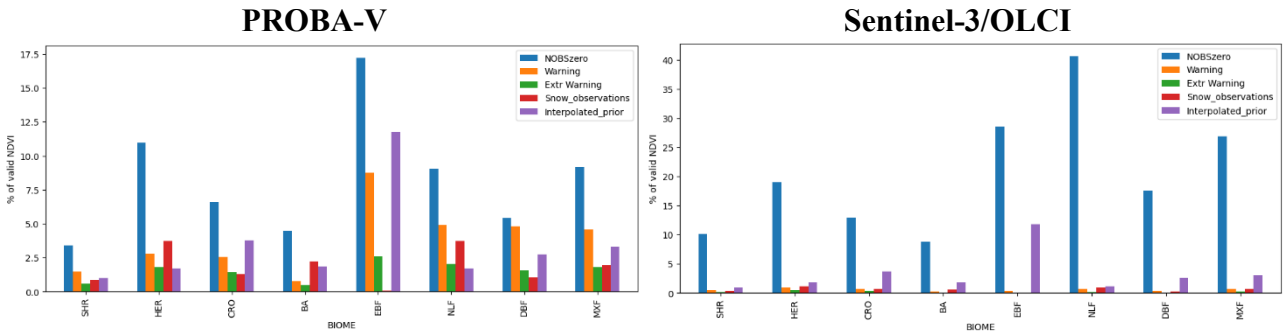


Figure 9: Distribution of flag occurrences over land pixels with valid NDVI per biome type for the year 2019. NOBS=0 (blue), warning (orange), extreme warning (green), snow observations (red), and interpolated prior (purple) are shown for PROBA-V (left) and Sentinel-3/OLCI (right).

5.1.3 Uncertainties

The temporal distribution of NDVI 300m V3 PROBA-V and Sentinel-3/OLCI and associated uncertainties are shown in Figure 10 by plotting the percentiles P10 to P90, calculated over a global subsample of all land pixels. The NDVI percentiles show the expected pattern, with lower NDVI values in the Northern hemisphere wintertime and slightly higher NDVI uncertainties during the same period. The plots for PROBA-V and Sentinel-3/OLCI based NDVI are very similar. For both datasets, the NDVI uncertainties peak around November, when there are still a lot of observations present in the Northern hemisphere, unlike during the winter period when many pixels are masked by the interpolated prior quality flag (QFLAG bit 7). This figure also confirms that Sentinel-3/OLCI uncertainties are higher than those of PROBA-V.

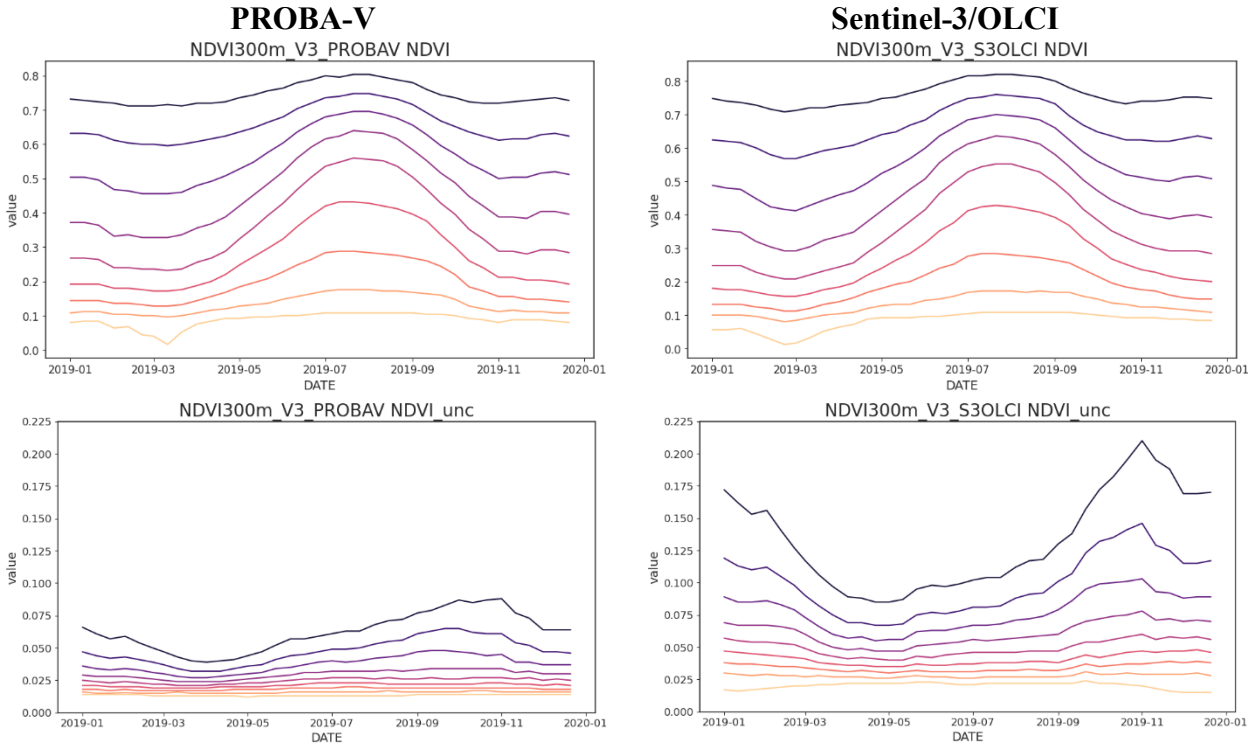


Figure 10: Temporal pattern of NDVI 300m V3 PROBA-V (left) and NDVI300m V3 Sentinel-3/OLCI (right) NDVI percentiles (top) and associated NDVI uncertainty percentiles (bottom). Percentiles P10 (yellow) up to P90 (purple) are shown for the year 2019.

The spatial distribution of mean NDVI and mean NDVI uncertainty show the expected spatial patterns (see Figure 11). The NDVI is highest in tropical areas and lowest in deserts, while the uncertainties are highest at high latitudes and in regions with persistent cloud cover (e.g. the wet tropics). The spatial patterns are nearly identical for PROBA-V and Sentinel-3/OLCI based NDVI 300m V3, and very comparable to those of NDVI 300m V2 (see Fig. 12 of CGLOPS1_ATBD_NDVI300m-V2, not shown here), but the uncertainties in V3 are higher compared to uncertainties in V2. A difference between the uncertainties in PROBA-V and Sentinel-3/OLCI is clearly visible: while the spatial patterns are similar, the uncertainties of Sentinel-3/OLCI are around a factor two higher (see Figure 35).

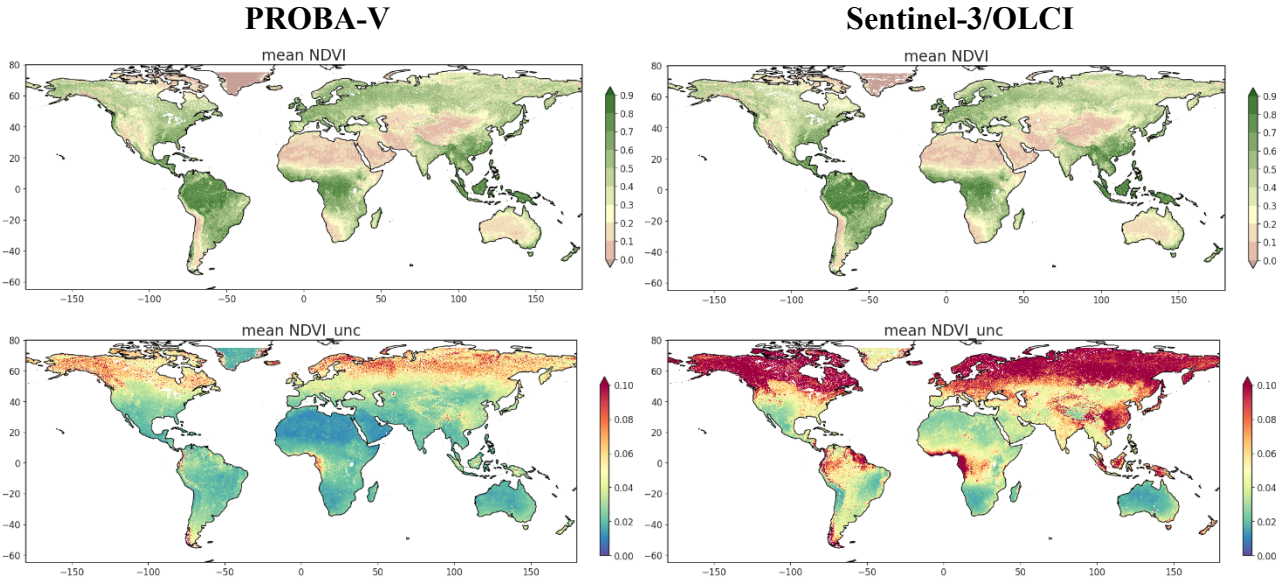


Figure 11: Spatial pattern of mean NDVI 300m V3 (top) and associated NDVI uncertainties (bottom) for PROBA-V (left) and Sentinel-3/OLCI (right) for the year 2019.

Figure 12 shows the relation between NDVI and NDVI_unc at the top and between NDVI_unc and NOBS at the bottom, for NDVI 300m V3 PROBA-V and Sentinel-3/OLCI. There is no clear relation between NDVI and NDVI_unc, indicating that users can apply a fixed threshold on NDVI uncertainty when required. On the other hand, the NDVI uncertainty clearly depends on NOBS, with higher uncertainties for lower NOBS values and an exponentially decreasing relation between both: including more observations will increase the precision of the determined values and hence decrease the uncertainty. The maximum value of NOBS, i.e. the number of clear input observations in the dekad time window, is 11 for PROBA-V (S1 TOC reflectance is used as input to the BRDF algorithm) and can be higher for Sentinel-3/OLCI (sub-daily TOC reflectance from Sentinel-3A and Sentinel-3B are used in the BRDF inversion).

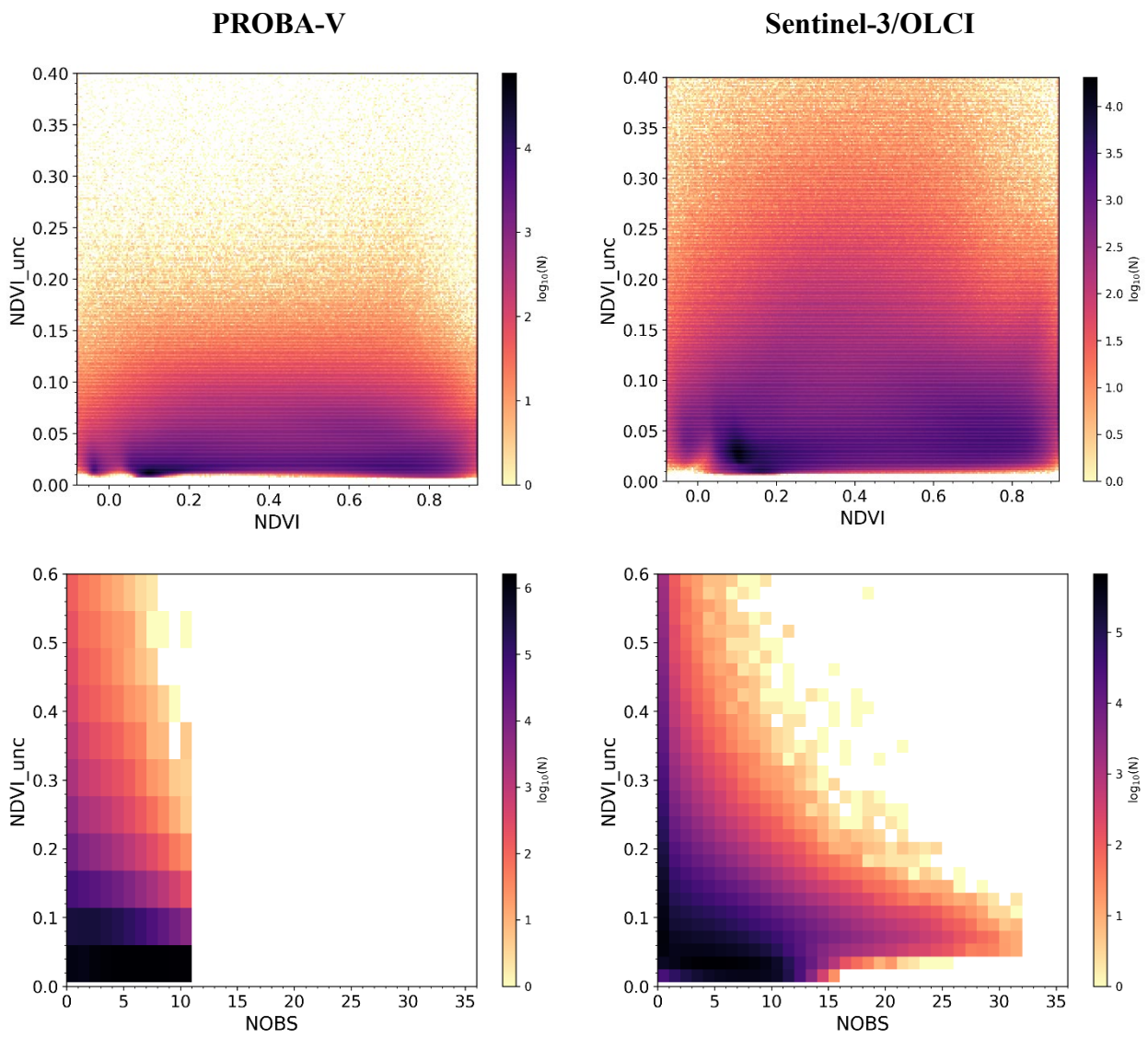


Figure 12: Scatter density plots showing the relation between NDVI and NDVI_unc (top), and the relation between NOBS and NDVI_unc (bottom) for NDVI 300m V3 PROBA-V (left) and NDVI 300m V3 Sentinel-3/OLCI (right) for the year 2019.

5.1.4 Conclusion

What is the occurrence of quality flags (and NDVI uncertainties) in the NDVI 300m V3 product in terms of temporal and spatial distribution?

The temporal and spatial distribution of quality flags 'Warning' and 'Extreme warning' have no clear temporal nor spatial pattern. In contrast, NOBS=0 occurs mainly during the Northern hemisphere winter, at higher latitudes, and in the wet tropics. This is due to persistent snow, cloud and/or cloud

shadow cover, and less favourable illumination conditions. Related to this, NOBS=0 occurs more often in biomes located at high latitudes (NLF and MXF) and in the wet tropics (EBF). The frequency of NOBS=0 is much higher for Sentinel-3/OLCI than for PROBA-V, while this is the other way around for the warning and extreme warning quality flags. This is related to the higher NDVI uncertainties of Sentinel-3/OLCI and how they affect the BRDF algorithm. The difference in number of snow observations is caused by different cloud masking methods applied to data from the two sensors, with IdePix, the algorithm used for Sentinel-3/OLCI, known to underperform for snow classification.

High NDVI uncertainties occur mainly during the Northern hemisphere wintertime, at high latitudes, and in areas with persistent cloud cover. The analysis indicates no relation between NDVI and NDVI uncertainty, but there is an exponentially decreasing relation between NDVI uncertainty and NOBS.

Which flags are recommended for users and are applied in the subsequent analyses?

It is recommended to exclude NDVI values with BRDF inversion quality flag 'Extreme warning' (QFLAG bit 3 or 5 is set) and values for which the MCD43P BRDF priors are based on linearly interpolated values (QFLAG bit 7 is set). Additionally, it is also recommended to exclude NDVI values that are classified as snow (QFLAG bit 1 is set) since they have no useful vegetation information, there is possible confusion with cloud classification, and only a small fraction of pixels is affected.

5.2 Q2: PRODUCT COMPLETENESS OF NDVI 300M V3?

5.2.1 Introduction

The scientific questions to be answered are the following: what are the temporal variation and spatial distribution of the product completeness of NDVI 300m V3 in comparison to MCD43A4 V6.1? How do the results based on NDVI 300m V3 compare to those of NDVI 300m V1, NDVI 300m V2, and NDVI 1km V3?

The analyses are performed over two one-year periods (2019 and 2023), the latter for comparison with NDVI 300m V2, and on the full period (2014 – 2024) for the temporal analysis. They are based on global systematic subsamples for the spatial and temporal distribution of gaps (see §4.3.2) and on extractions of valid NDVI values for LANDVAL V2 sites for the evaluation of gap lengths (see §4.3.4).

The analyses include:

- Spatial distribution of gaps
- Temporal distribution of gaps
- Gap length frequency

5.2.2 Spatial distribution of gaps

Figure 13 shows the spatial patterns of gaps in NDVI 300m V3 compared to the reference datasets for the year 2019, while Figure 14 shows the spatial gap patterns for the year 2023. The large-scale spatial patterns are similar for all datasets with gaps being mostly located at high latitudes and in areas with persistent cloud cover (e.g. the wet tropics). NDVI 300m V3 shows a somewhat higher completeness than the reference datasets, related to the longer accumulation period (see Annex 2: Back-processing vs NRT processing) and the use of a climatology in the BRDF inversion. When a climatology is used, NDVI values are produced even when there is no input data available, therefore leading to higher completeness. The effect is clearly visible in the completeness figures, when comparing areas with persistent cloud cover (e.g. the Amazone) for datasets that use a climatology (NDVI 300m V3, NDVI 300m V2, and NDVI 1km V3) with datasets that don't (NDVI 300m V1 and MCD43A4 V6.1).

There are more gaps present in the NDVI 300m V3 PROBA-V dataset than in the NDVI 300m V3 Sentinel-3/OLCI dataset. More data in the PROBA-V dataset are missing due to more NDVI values being filtered by the snow quality flag (bit 1). The difference in completeness is hence also caused by the difference in cloud and snow masking methodology.

The difference in completeness between NDVI 300m V3 and NDVI 300m V2 is related to the difference in accumulation window. The spatial completeness patterns are much more similar when comparing NDVI 300m V2 from Figure 14 with the spatial pattern of NDVI 300m V3 for NRT processing of 2025 (see Figure 70).

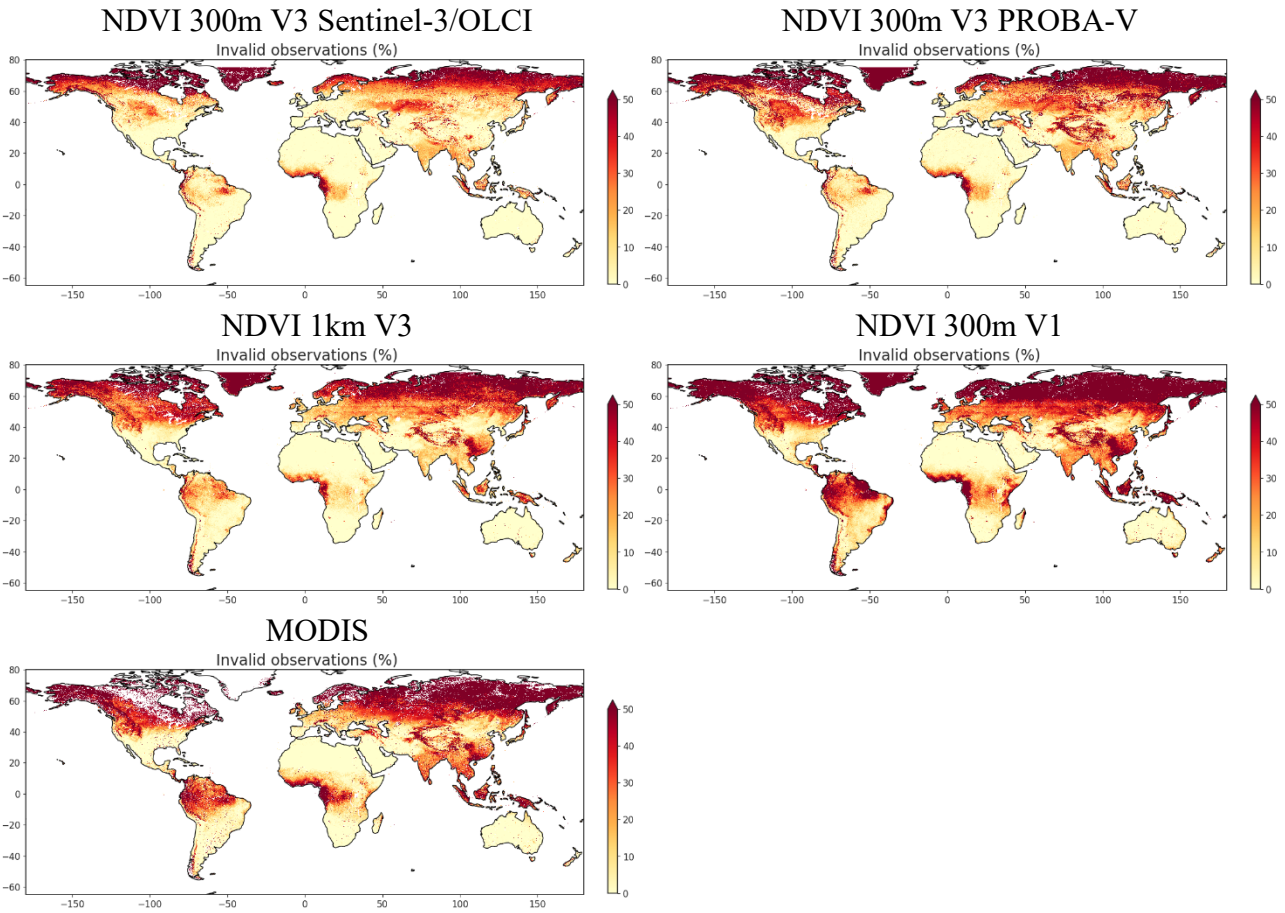


Figure 13: Spatial distribution of the number of missing values (%) of NDVI 300m V3 Sentinel-3/OLCI, NDVI 300m V3 PROBA-V, NDVI 1km V3, NDVI 300m V1, and MODIS MCD43A4 V6.1 for the year 2019.

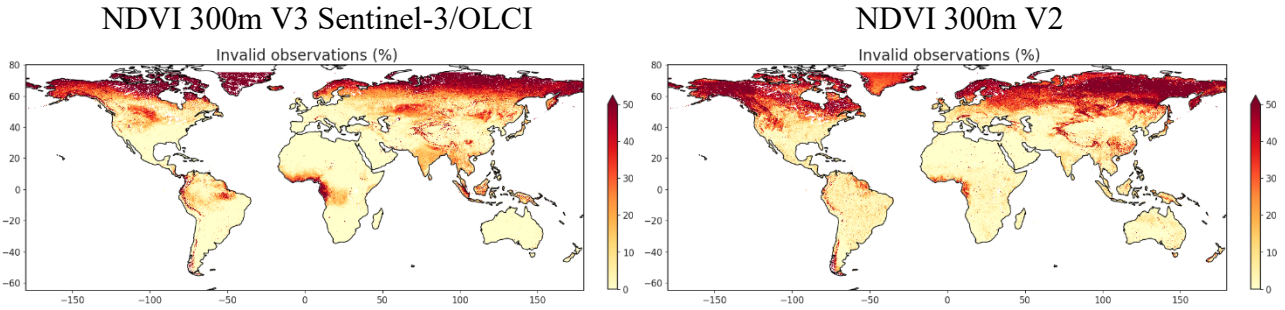


Figure 14: Spatial distribution of the number of missing values (%) of NDVI 300m V3 Sentinel-3/OLCI and NDVI 300m V2 for the year 2023.

5.2.3 Temporal distribution of gaps

Similar conclusions can be drawn from the analysis of the temporal distribution of the number of valid observations over land (top panel of Figure 15). The completeness of NDVI 300m V3 is much larger than that of NDVI 300m V1 and MODIS, and is comparable to the completeness of NDVI 300m V2 and NDVI 1km V3. This is again related to the longer accumulation period in NDVI 300m V3 and the use of a climatology in the BRDF inversion of NDVI 300m V3, NDVI 300m V2, and NDVI 1km V3. For all datasets, the completeness is larger during the Northern hemisphere summer months, when the northern regions are not impacted by bad illumination conditions.

There is quite some difference between the temporal patterns of product completeness of PROBA-V and Sentinel-3/OLCI for NDVI 300m V3. The completeness of Sentinel-3/OLCI is higher at the start of each year, increases fast, and then shows a decrease during March/April. The rapid increase in Spring is caused by pixels with very low NVDI values located at high latitudes. These are probably undetected snow pixels. The flagging of snow pixels is different for PROBA-V and Sentinel-3/OLCI as can be seen in the bottom panel of Figure 15. In this figure, the occurrences of three of the four quality flags present in the NDVI layer itself are plotted over time. The ‘snow’ flag is applied when more than half of the recent observations are identified as snow, ‘missing’ pixels have TOC-r out of range or not available for at least one of the bands, and pixels identified as ‘unknown’ have NOBS=0 and interpolated BRDF priors. There are large peaks of snow occurrences around March of each year and smaller peaks around November, but the amplitude of the peaks is lower for the Sentinel-3/OLCI period. This pattern coincides with the patterns in the completeness of NDVI 300m V3, indicating that the higher completeness of Sentinel-3/OLCI during Winter/Spring compared to PROBA-V is caused by less snow identification for Sentinel-3/OLCI and thus more valid (low) NDVI values.

The under detection of snow pixels also explains the difference between NDVI 300m V3 and NDVI 300m V2. In the V2 version, there is no snow filtering or interpolated-prior masking applied, but

many of these 'snow' pixels at high latitudes are removed by applying the 'TOC-r out of range' filter. In NRT processing, more pixels are assigned this flag value and hence the temporal completeness pattern over NRT NDVI 300m V3 will be closer to that of NDVI 300m V2 (see Annex 2: Back-processing vs NRT processing).

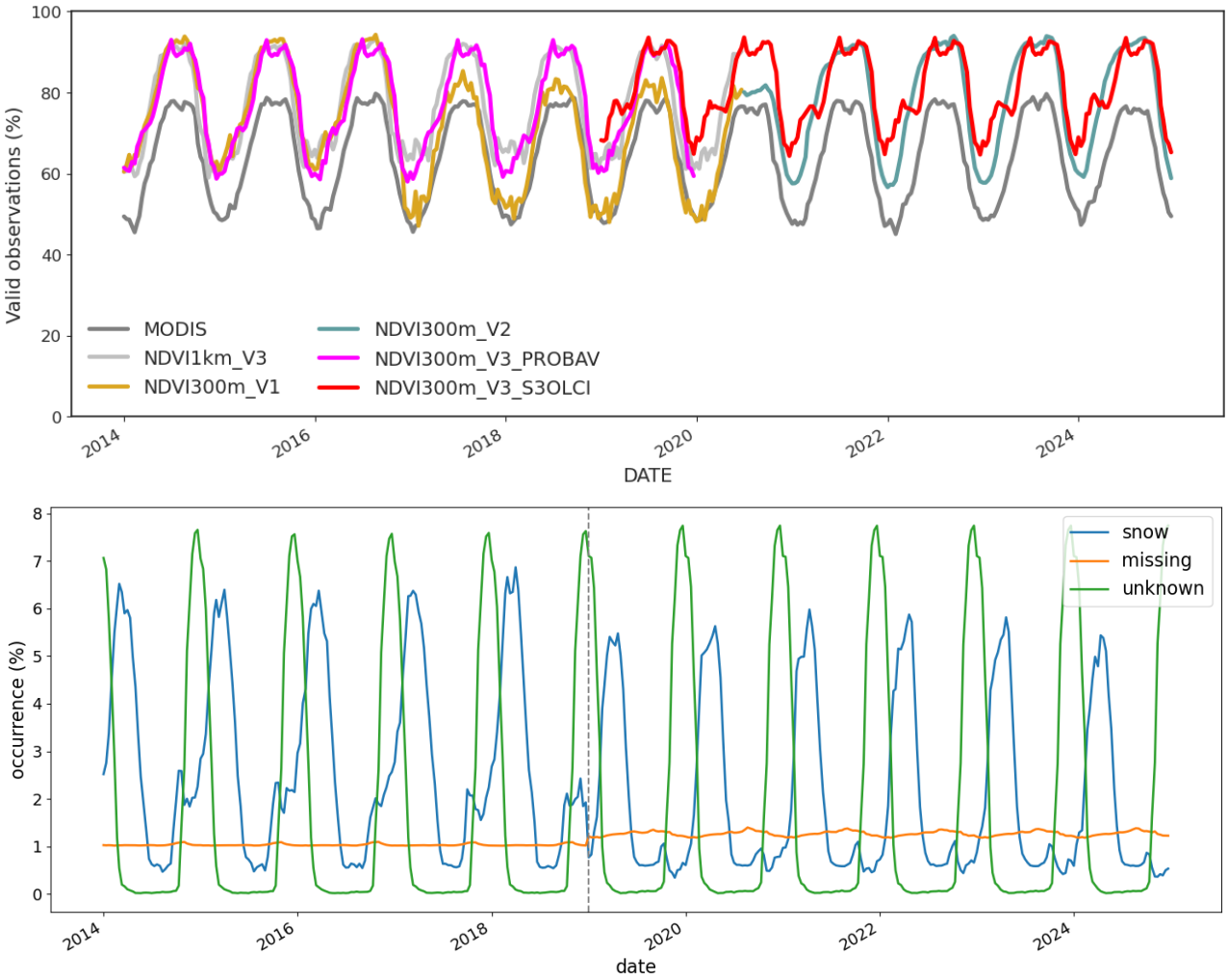


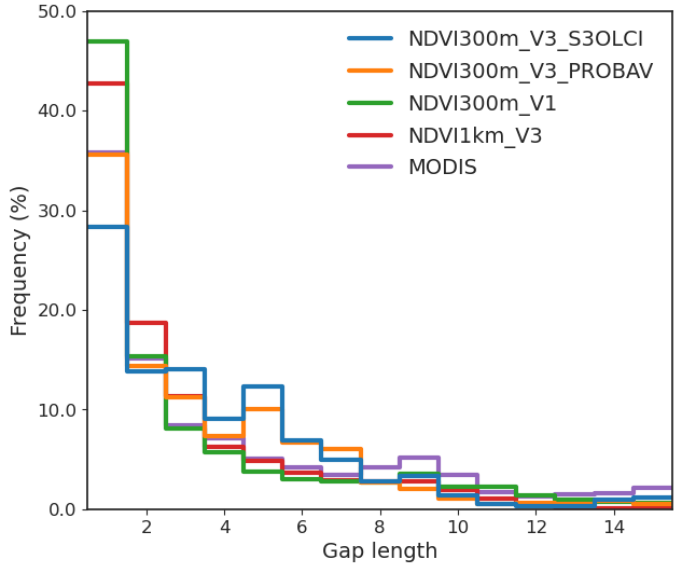
Figure 15: *Top*: temporal evolution of the number of valid observations over land (%) of NDVI 300m V3 Sentinel-3/OLCI (red), NDVI 300m V3 PROBA-V (pink), NDVI 300m V1 (yellow), NDVI 300m V2 (blue), NDVI 1km V3 (light grey), and MODIS MCD43A4 V6.1 (grey). *Bottom*: temporal evolution of occurrences of snow (blue), missing (orange), and unknown (green) flag values present in the NDVI layer of NDVI 300m V3. The vertical grey dashed line indicates the switch between PROBA-V and Sentinel-3/OLCI.

5.2.4 Gap length frequency

A data gap is defined as a single missing data point or consecutive dates with missing data in the extractions over the LANDVAL V2 sites of the NDVI 300m V3 dataset and reference datasets. The frequency of data gap occurrences per length of the data gap is plotted in Figure 16 for the year 2019 and in Figure 17 for the year 2023.

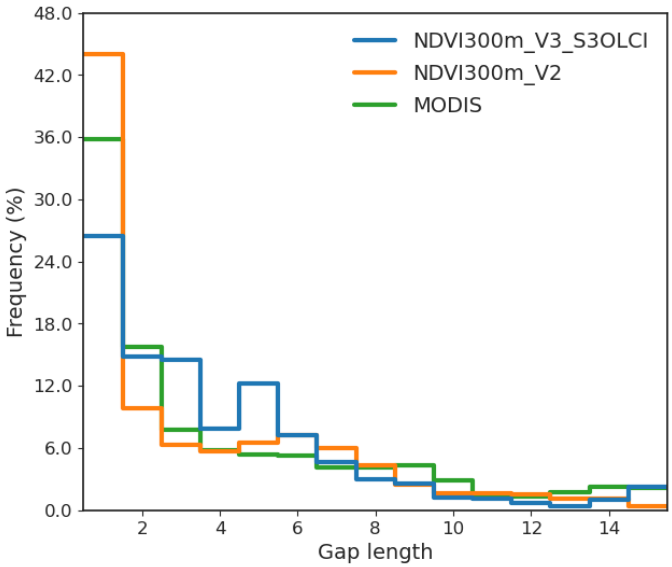
These figures show that gaps of one dekad are most common for all products in the comparison and that its frequency is lowest for the NDVI 300m V3 dataset. Larger data gaps are less frequent. There are two peaks, one around three and one around five dekads for the NDVI 300m V3 dataset. These peaks are related to high latitude observations (latitude > 55°) for which NOBS is equal to zero (bit 0 is set) and for which the MODIS BRDF prior is interpolated (bit 7 is set). For these observations a gap is expected during the Northern winter period, the exact length depends on the location of each site and the amount of dekads for which the MODIS BRDF interpolated prior flag is applied at that location.

The mean gap length is lowest for the NDVI 300m V3 PROBA-V and Sentinel-3/OLCI datasets, related to the longer accumulation period. The gap length frequency of NDVI 300m V3 is more comparable to the other datasets when the NRT processing mode is used (see Annex 2: Back-processing vs NRT processing).



Dataset	Mean gap length (dekads)
NDVI 300m V3 Sentinel-3/OLCI	1.8
NDVI 300m V3 PROBA-V	2.0
NDVI 300m V1	3.5
NDVI 1km V3	2.1
MODIS	3.4

Figure 16: Histogram of gap length (left) and mean gap length (right) over all 2000 LANDVAL V2 sites for NDVI 300m V3 Sentinel-3/OLCI (blue), NDVI 300m V3 PROBA-V (orange), NDVI 300m V1 (green), NDVI 1km V3 (red), and MODIS MCD43A4 V6.1 (purple) for the year 2019.



Dataset	Mean gap length (dekads)
NDVI 300m V3 Sentinel-3/OLCI	3.3
NDVI 300m V2	3.3
MODIS	5.9

Figure 17: Histogram of gap length (left) and mean gap length (right) over all 2000 LANDVAL V2 sites for NDVI 300m V3 Sentinel-3/OLCI (blue), NDVI 300m V2 (orange), and MODIS MCD43A4 V6.1 (green) for the year 2023.

5.2.5 Conclusion

What are the temporal variation and spatial distribution of the product completeness of NDVI 300m V3 in comparison to MCD43A4? How do the results based on NDVI 300m V3 compare to those of NDVI 300m V1, NDVI 300m V2, and NDVI 1km V3?

The product completeness of NDVI 300m V3 is remarkably high, compared to MCD43A4 V6.1 and the other reference datasets. This is related to the long accumulation period that is used for NDVI 300m V3. The completeness in NRT processing is lower and more comparable to the completeness of the reference datasets, as shown in [Annex 2: Back-processing vs NRT processing]. For all datasets, the product completeness is highest across desert sites and lower at higher latitudes during the Northern hemisphere winter period and in areas with persistent cloud cover.

5.3 Q3 VISUAL COMPARISON BETWEEN NDVI 300M V3 AND REFERENCE DATASETS?

5.3.1 Introduction

The scientific question to be answered is the following: how do the NDVI 300m V3 Sentinel-3/OLCI and PROBA-V datasets compare visually to the NDVI 300m V1, NDVI 300m V2, NDVI 1km V3, and MODIS MCD43A4 V6.1 datasets? The analyses are performed over two one-year periods (2019 and 2023) and are based on a selection of 1° x 1° and 10° x 10° full resolution tiles (see §4.3.3). The results are shown for a few dekads, quicklooks for the whole period can be found in [Annex 1: Digital Annex].

5.3.2 Visual comparison at 1° x 1°

To compare the sharpness of NDVI 300m V3 with previous versions V1 and V2, quicklooks of two 1° x 1° mini tiles are visually inspected, considering only values according to the flag use recommendations in §5.1. Figure 18 and Figure 20 compare V3 with V1 and Sentinel-3/OLCI V3 with PROBA-V V3 for a region in Belgium and in South Sudan respectively. Figure 19 and Figure 21 inspect the same regions but for comparison between V3 and V2. For some dekads, quicklooks of the QFLAG layer are also provided, in which the various QFLAG bits are overplotted on top of each other starting with the one at the bottom of the colour table (for example NOBS=0 is plotted first, then INTERP. PRIOR, and so on).

V3 shows reduced sharpness, but higher completeness compared to V1, especially for the tile in Belgium. BRDF normalisation – included in V3 but not in V1 – accumulates TOC reflectances over a certain period and makes use of a climatology prior in case no (good quality) observations are available. While this reduces gaps, it also lowers the sharpness. The striped gaps in V1 are also gone in V3, thanks to the use of an updated collection of PROBA-V input data. The quicklooks of V2 and V3 are very similar with respect to sharpness, but there are more gaps in V2. These gaps are pixels that are flagged with an extreme warning or because of TOC-r is out of range. The differences between V2 and V3 are hence related to differences in NDVI uncertainties, with higher uncertainties in V3 leading to less extreme warnings (see §5.1), and to the difference in accumulation period and ReBeLS version, with better outlier detection in V3 and a longer period resulting in less out of range TOC-r values.

The difference between Sentinel-3/OLCI V3 and PROBA-V V3 is small. Any difference in sharpness is likely related to differences in ground sampling distance between PROBA-V (< 100m at nadir) and Sentinel-3/OLCI (300m at nadir) and the resampling steps in the processing chain (see §4.2.5.4). There are more gaps in the PROBA-V product, because more of the data is flagged as snow, warning,

or extreme warning than for Sentinel-3/OLCI (see §5.1). This is related to different cloud/snow classification algorithms and higher uncertainties in Sentinel-3/OLCI.

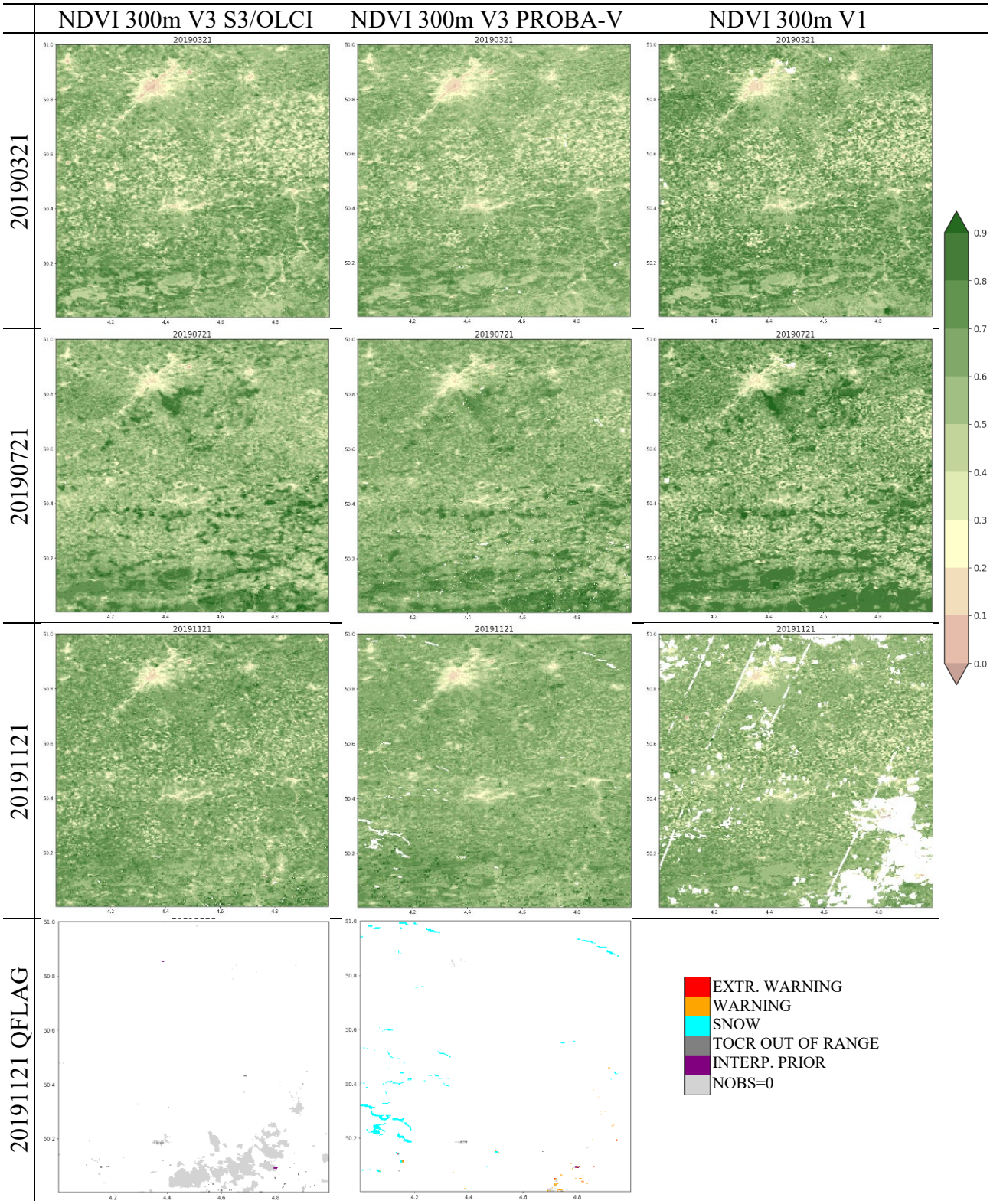


Figure 18: Visual comparison of NDVI for the Western Europe 1°x1° mini tile for the year 2019.

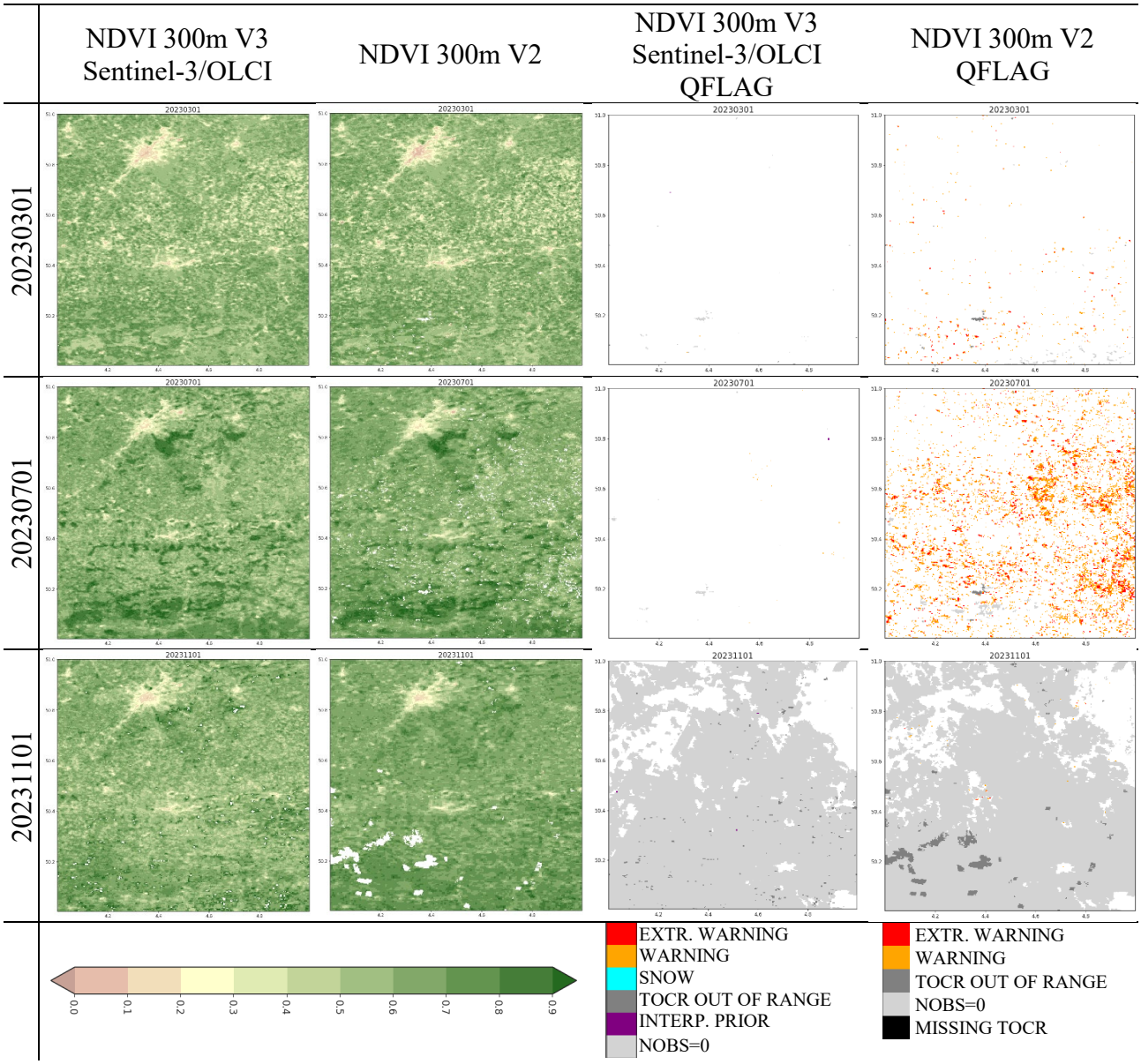


Figure 19: Visual comparison of NDVI for the Western Europe 1°x1° mini tile for the year 2023.

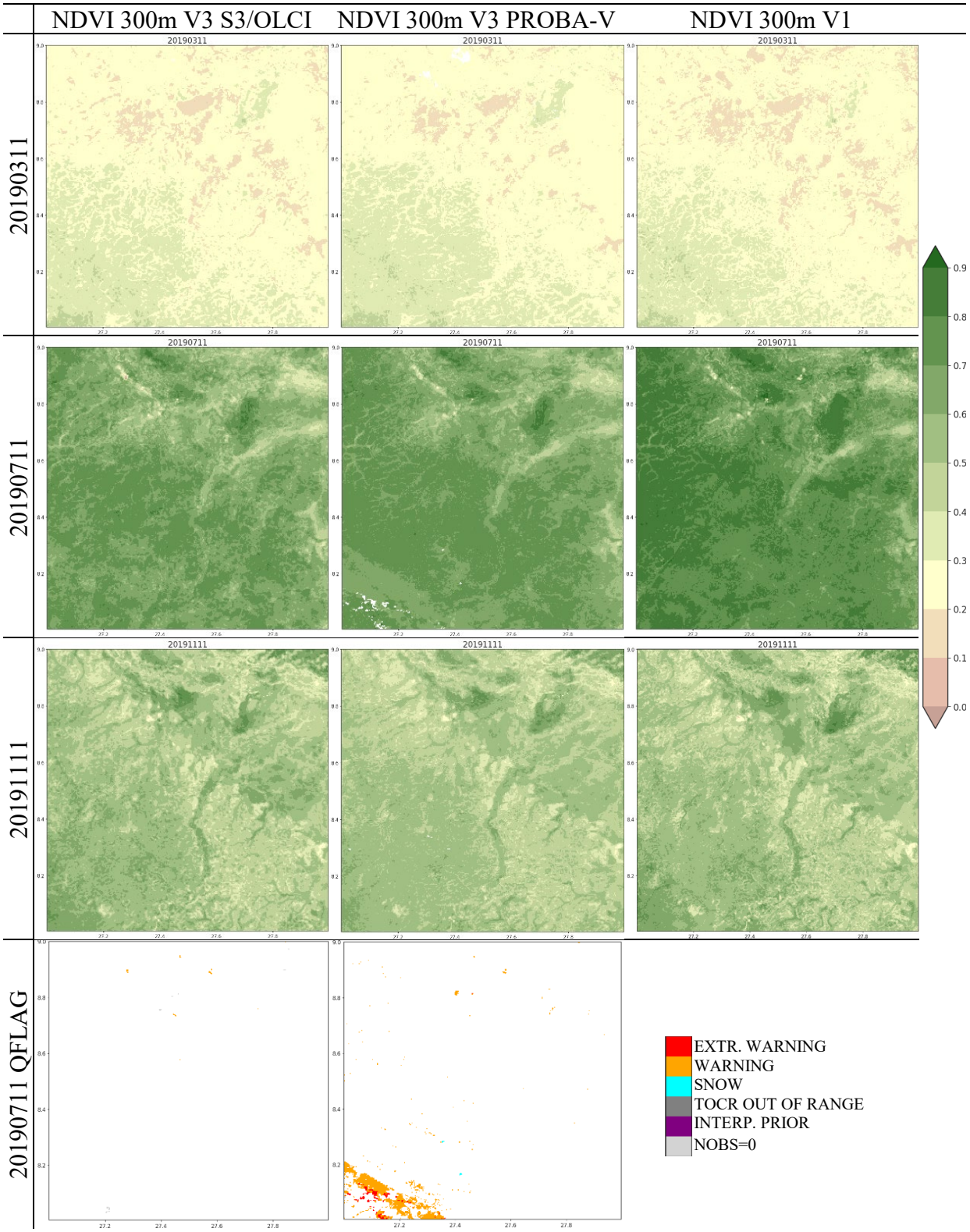


Figure 20: Visual comparison of NDVI for the Central Africa 1°x1° mini tile for the year 2019.

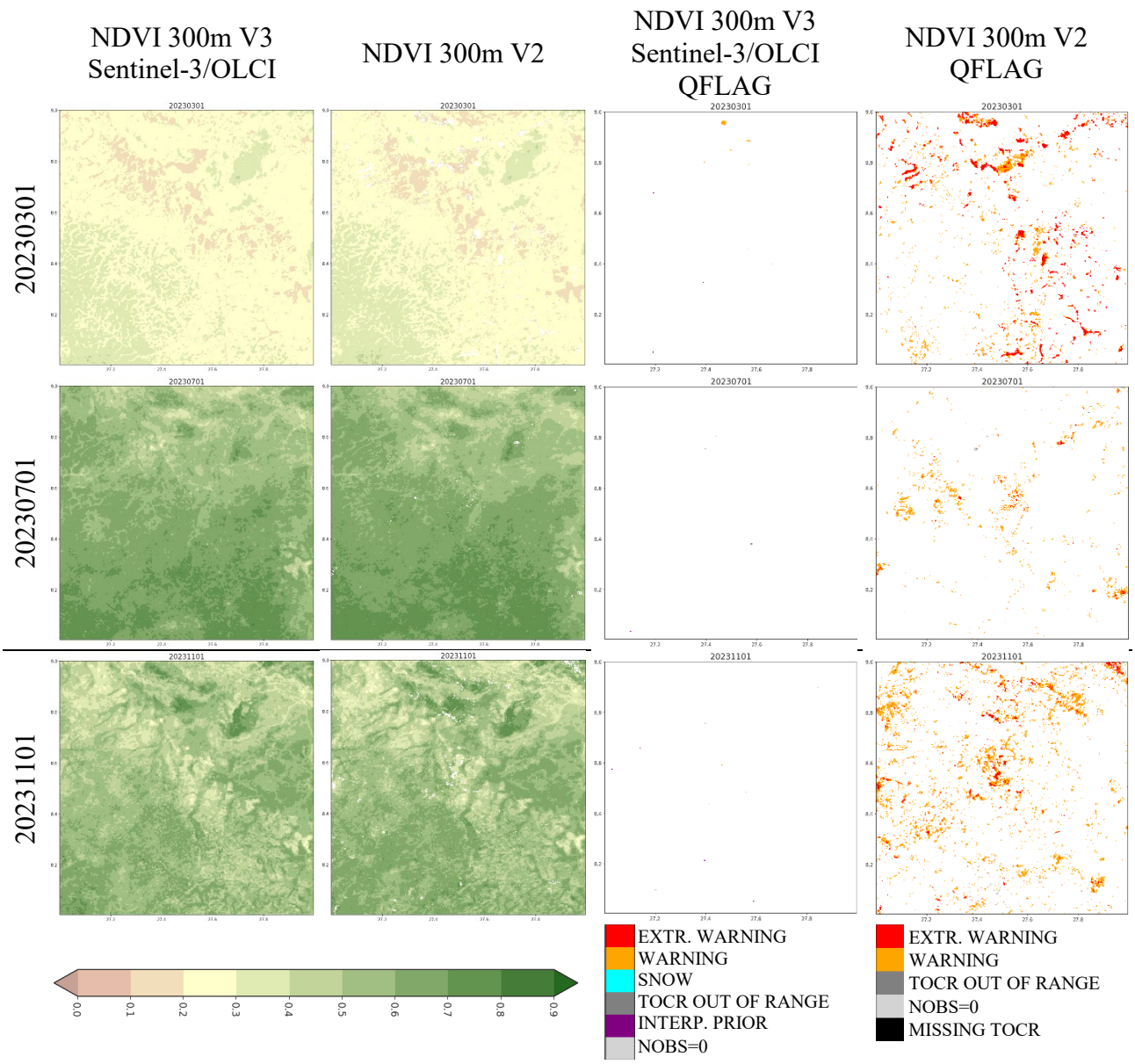


Figure 21: Visual comparison of NDVI for the Central Africa 1°x1° mini tile for the year 2023.

5.3.3 Visual comparison at 10° x 10°

Full resolution quicklooks of 10° x 10° are visually inspected to determine possible (completeness) differences between the NDVI 300m V3 product and external reference MODIS MCD43A4 V6.1. NDVI 300m V2 quicklooks are shown as well to compare with the previous version of this product. Figure 22, Figure 24, Figure 26, and Figure 28 contain the visual comparison of valid NDVI values, flagged according to the recommendations in §5.1, for a tile in Northern Europe, Western Europe, South America, and Central Africa, respectively. Figure 23, Figure 25, Figure 27, and Figure 29 show the QFLAG layer for the same tiles in the same way as in §5.3.2.

There are no artefacts or major irregularities visible in NDVI 300m V3. The NDVI patterns are similar for all three datasets.

Compared to MCD43A4 V6.1, the NDVI 300m V3 dataset has a much higher completeness. Apart from differences in snow classification, due to different algorithms being used, it is not immediately clear why the MCD43A4 V6.1 dataset has so much lower completeness.

NDVI 300m V3 has a much lower completeness than NDVI 300m V2, especially for the tile in Northern Europe and in South America. The gaps in NDVI 300m V3 are related to pixels flagged as snow (QFLAG bit 1) and as interpolated prior (QFLAG bit 7). V2 does not have a snow flag in the quality layer, some snow pixels are masked because the TOC-r is out of range, but the number is much lower than the snow pixels masked in V3. There is also no interpolated prior flag in V2. The low completeness of V3 in the South America tile is caused by this flag and is due to better outlier detection due to a different ReBeLS version and because a longer accumulation window is used. Although a snow flag is available in V3, the snow classification is not perfect, as illustrated by the snow in the Alpes which is not correctly identified (see Figure 24). Although not flagged as such, snow-influenced NDVI values over the Alpes are discarded in V2, where those pixels are removed by the ‘TOC-r out of range’ quality flag. It is expected that this effect will also be present in NRT processing of V3 where a smaller accumulation window causes more TOC-r out of range values (see Annex 2: Back-processing vs NRT processing).

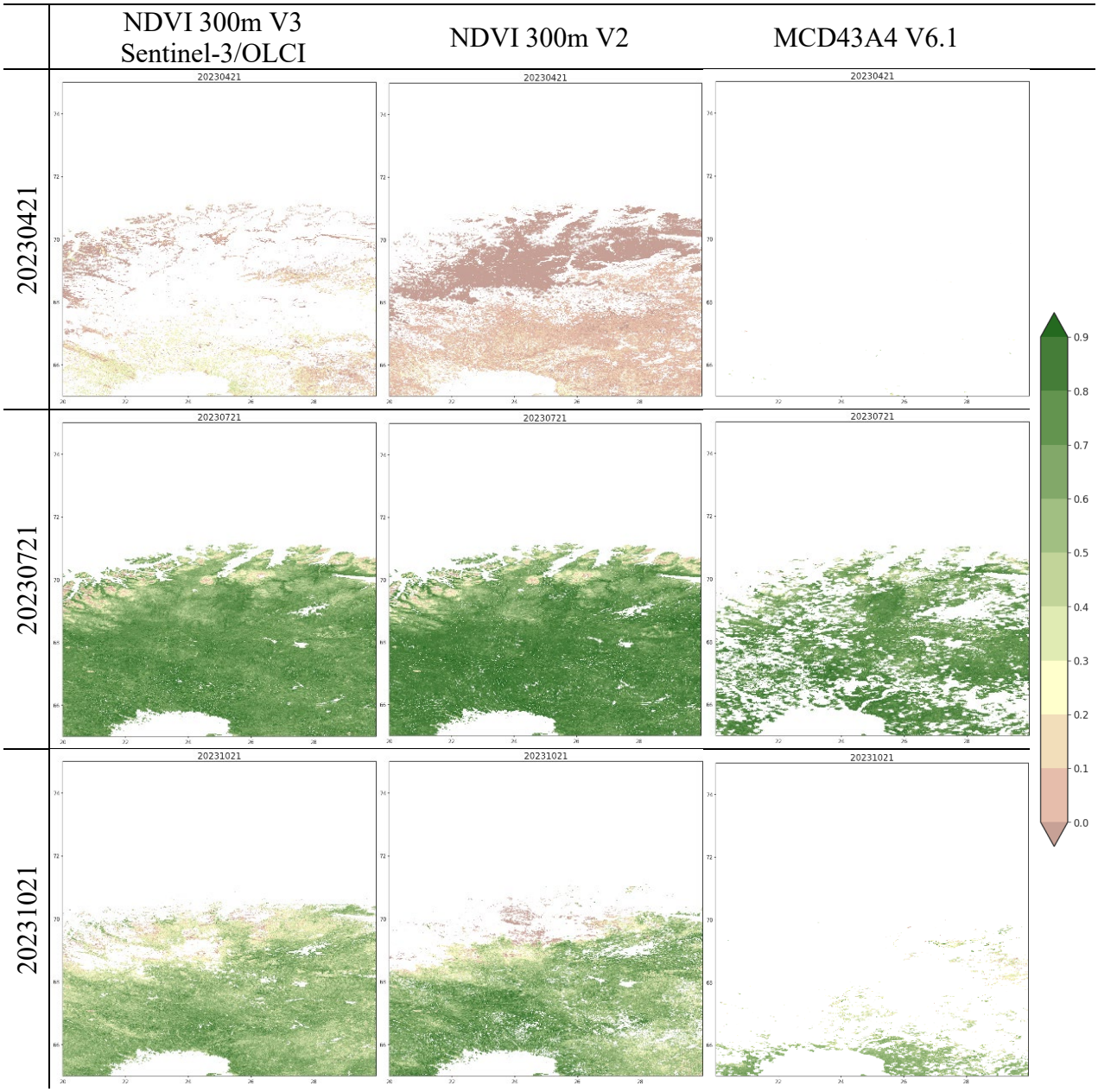


Figure 22: Visual comparison of NDVI for the Northern Europe 10° x 10° tile for the year 2023.

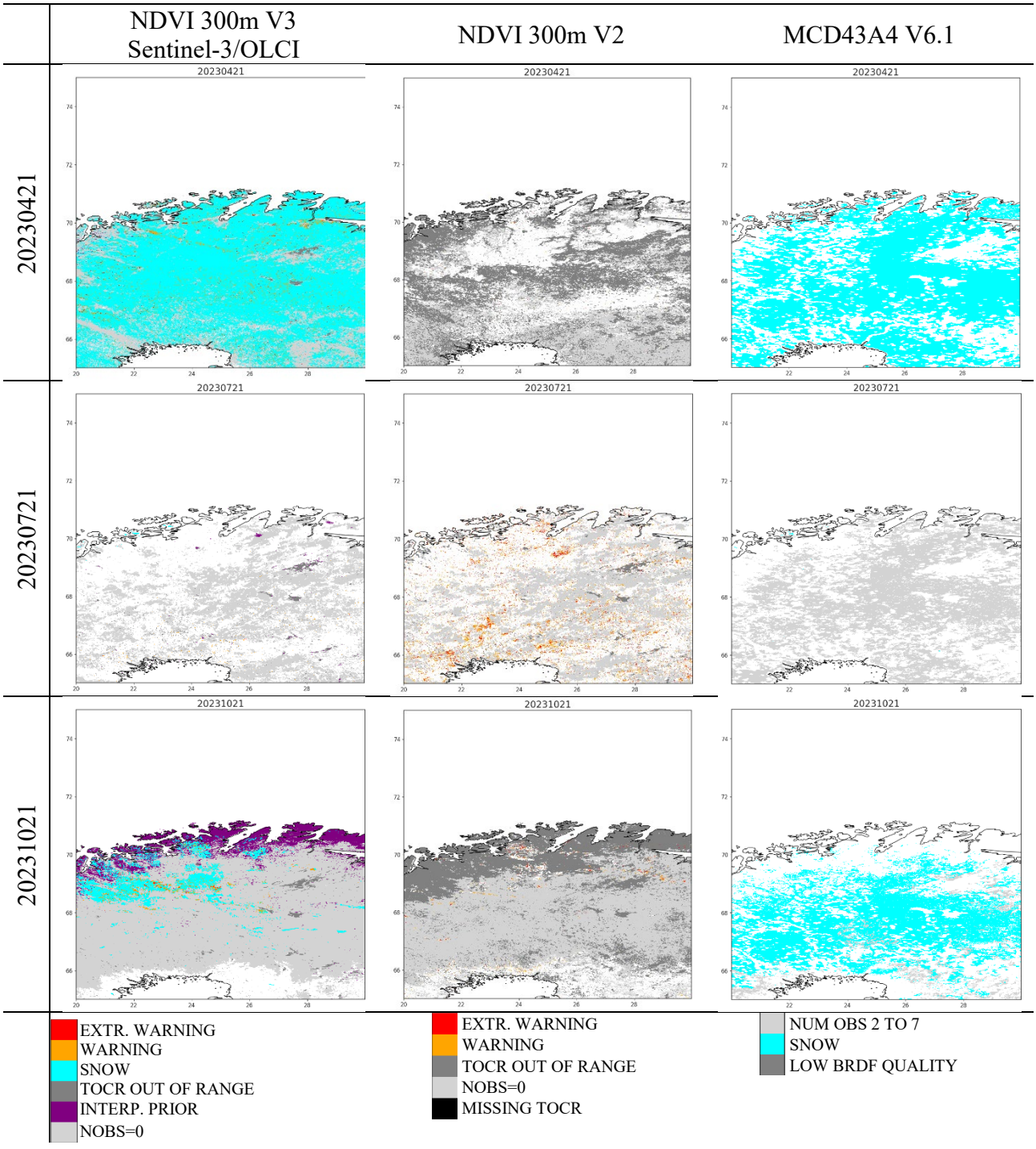


Figure 23: Visual comparison of QFLAG for the Northern Europe 10° x 10° tile for the year 2023.

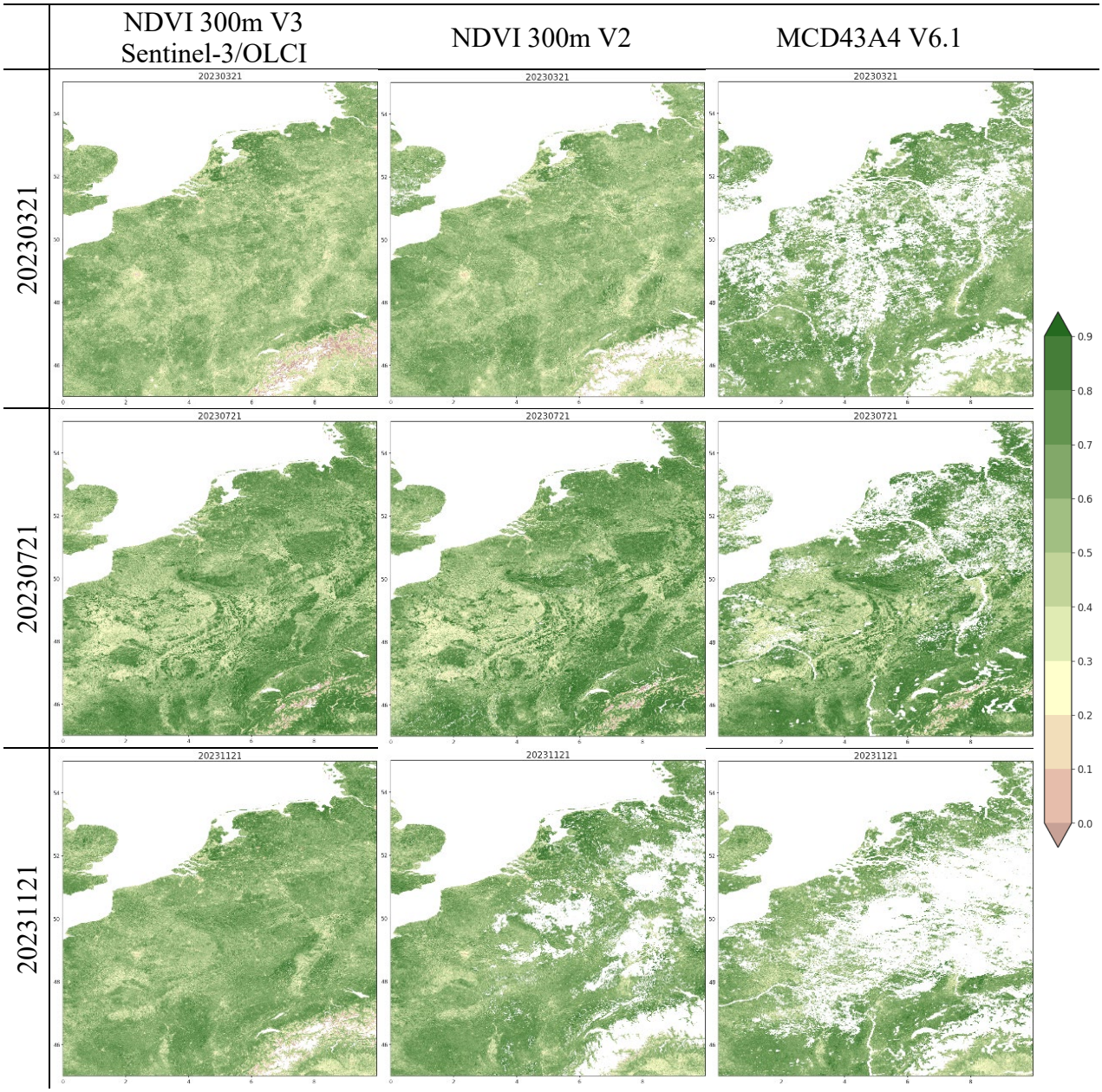


Figure 24: Visual comparison of NDVI for the Western Europe 10° x 10° tile for the year 2023.

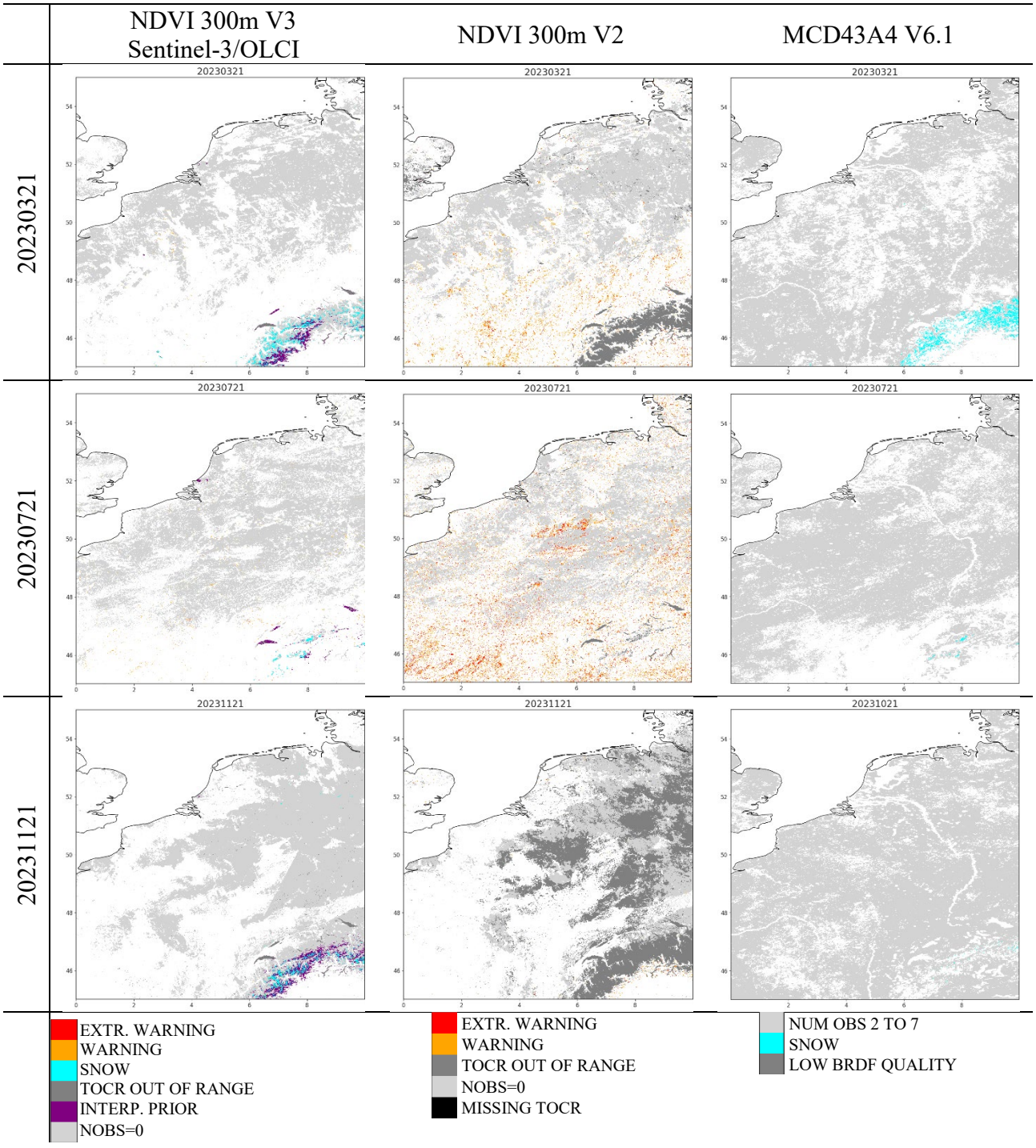


Figure 25: Visual comparison of QFLAG for the Western Europe 10° x 10° tile for the year 2023.

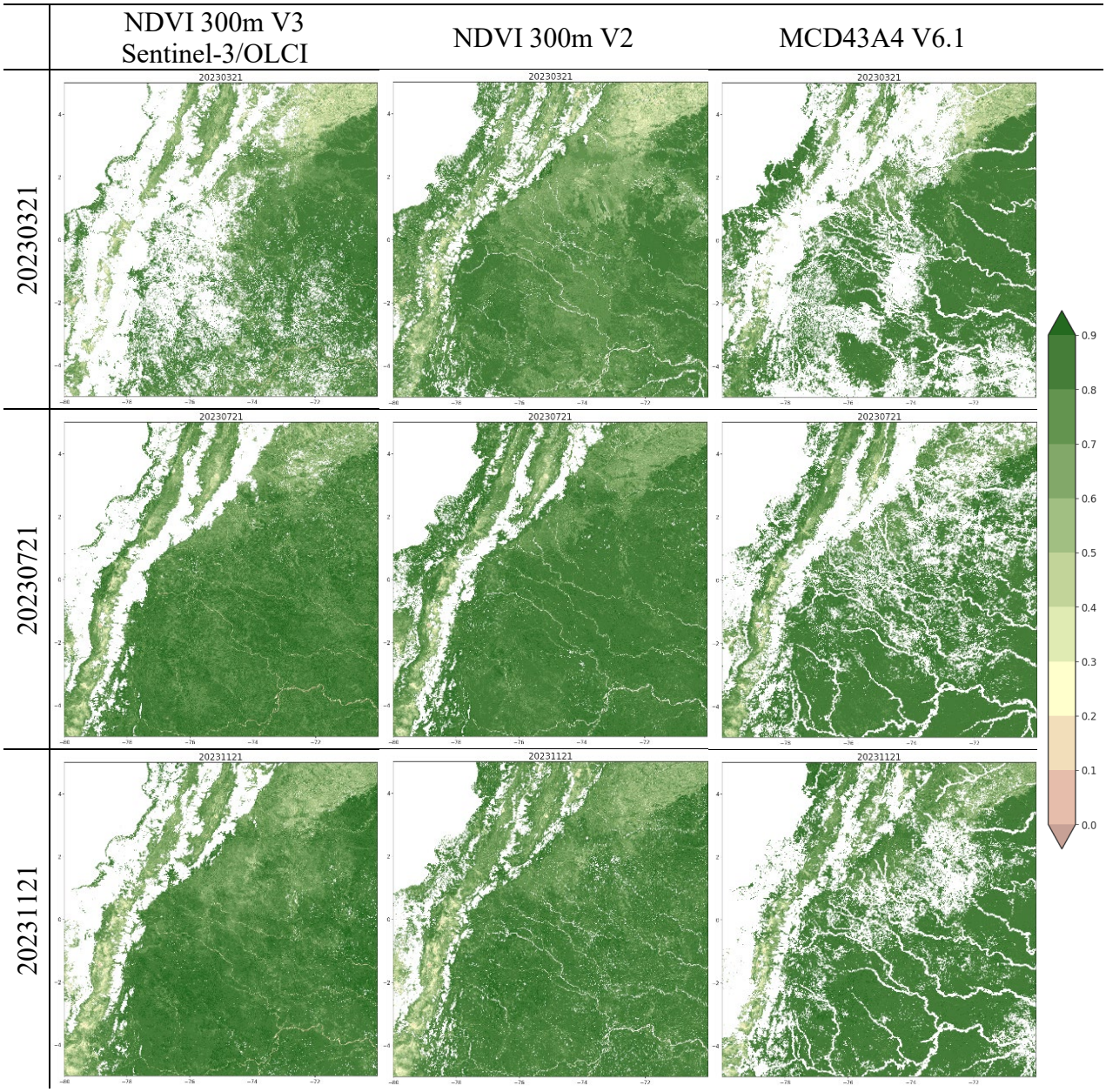


Figure 26: Visual comparison of NDVI for the South America 10° x 10° tile for 2023.

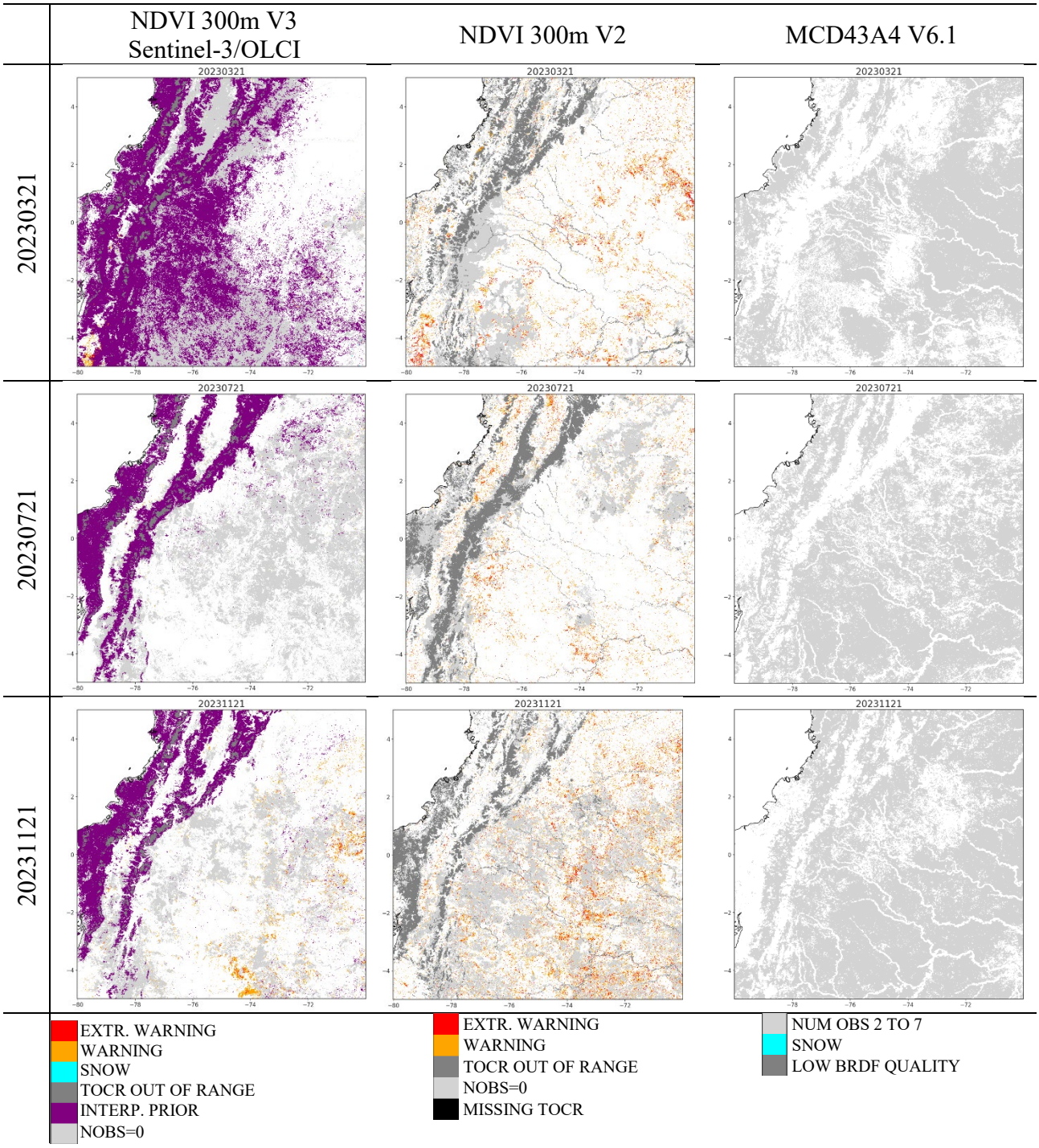


Figure 27: Visual comparison of QFLAG for the South America 10° x 10° tile for 2023.

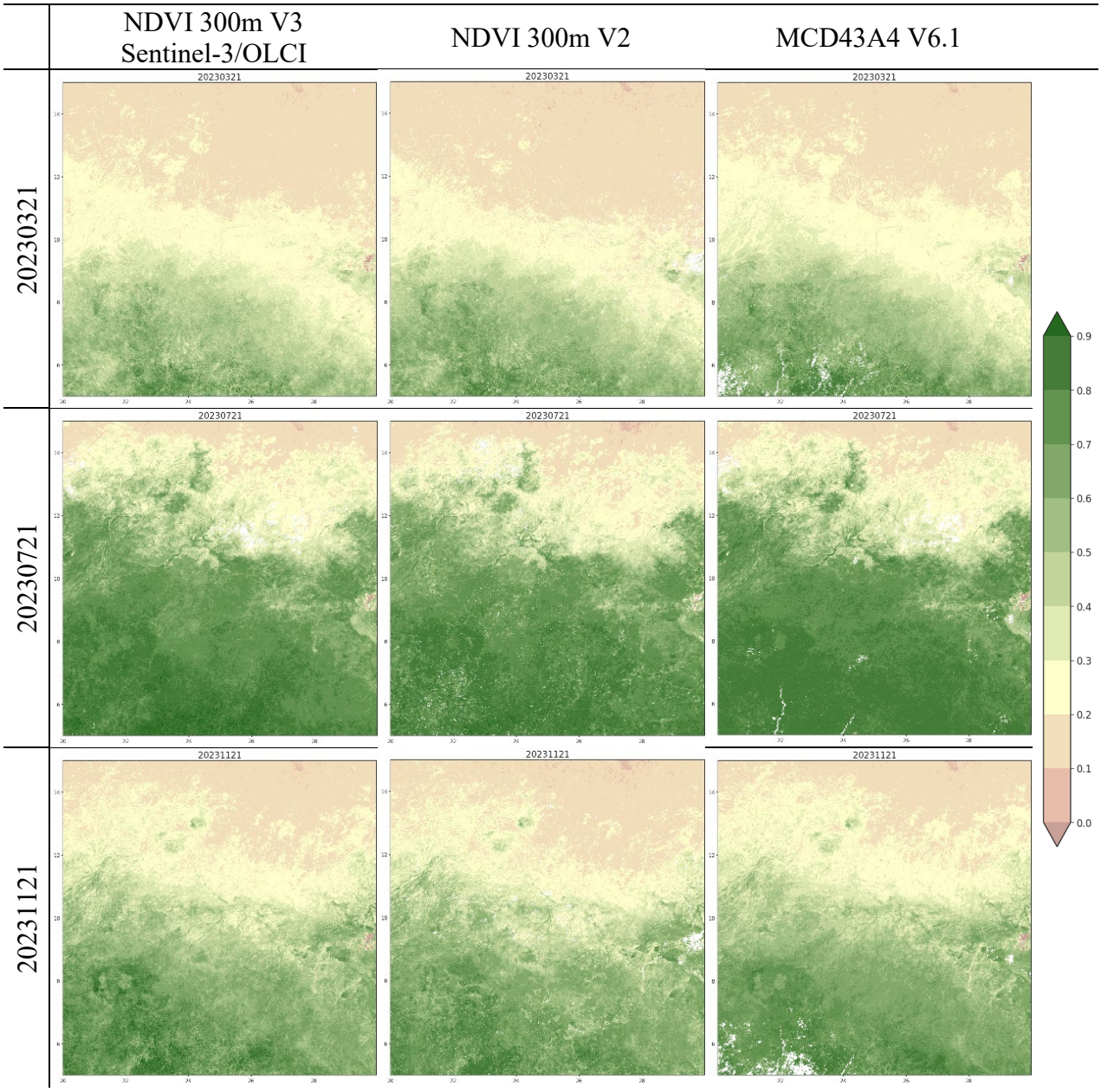


Figure 28: Visual comparison of NDVI for the Central Africa 10° x 10° tile for 2023.

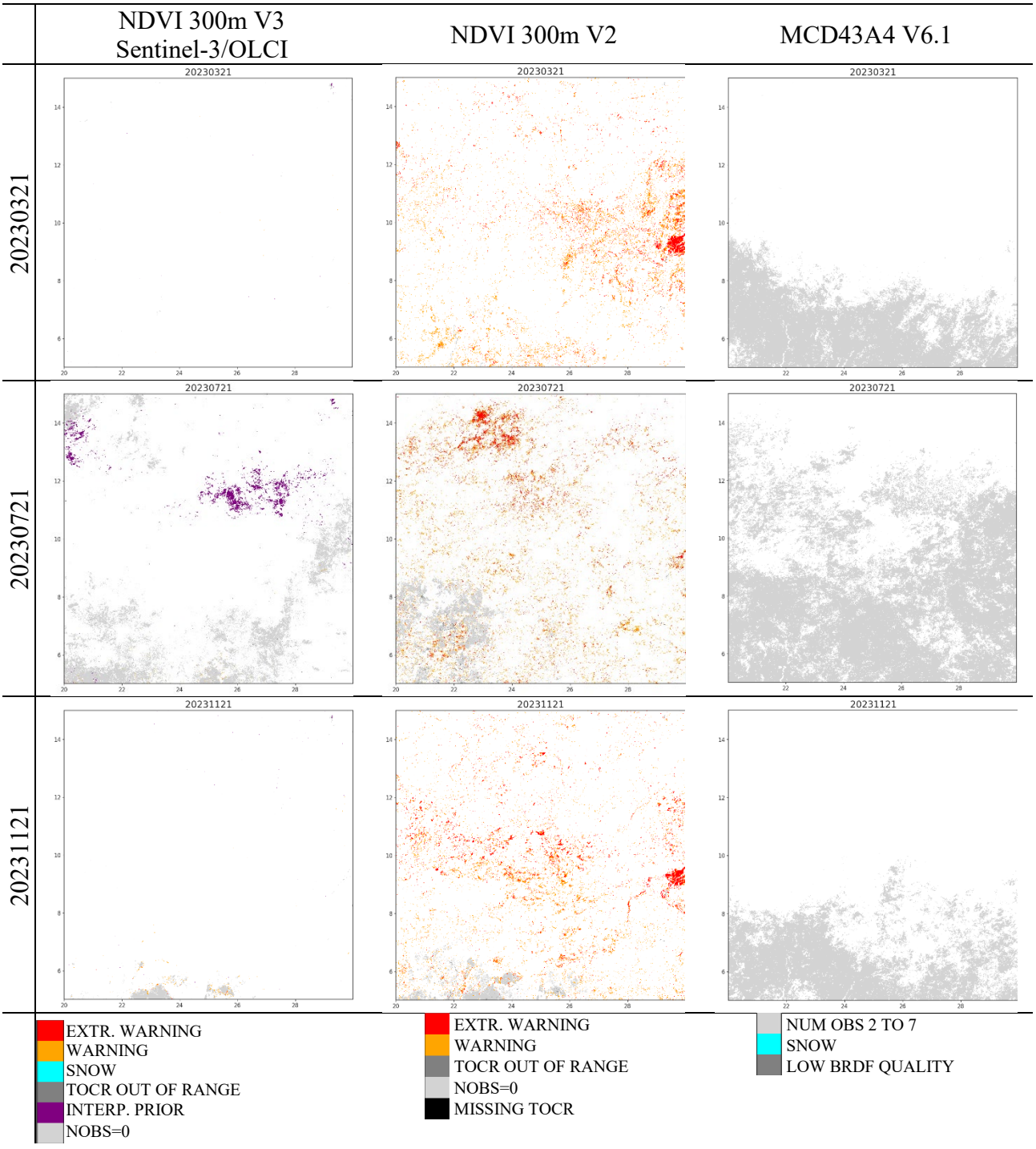


Figure 29: Visual comparison of QFLAG for the Central Africa 10° x 10° tile for 2023.

5.3.4 Conclusion

How do the NDVI 300m V3 Sentinel-3/OLCI and PROBA-V datasets compare visually to the NDVI 300m V1, NDVI 300m V2, NDVI 1km V3, and MODIS MCD43A4 V6.1 datasets?

The NDVI 300m V3 quicklooks are similar in sharpness and in NDVI values and patterns to the reference datasets. The completeness of NDVI 300m V3 is higher than that of the external reference dataset MCD43A4 V6.1 and NDVI 300m V1. Differences in completeness with NDVI 300m V2 are caused by the inclusion of snow and interpolated prior quality flags in V3, which filter out some of the data. To have a more complete product, users might opt to include pixels classified with the interpolated prior flag, but this comes at the cost of having more uncertain results. In NRT processing, flagging is slightly different, and more gaps are to be expected (see Annex 2: Back-processing vs NRT processing).

5.4 Q4: Statistical and spatial consistency between PROBA-V and Sentinel-3/OLCI for NDVI 300m V3?

5.4.1 Introduction

The scientific question to be answered is the following: what is the statistical and spatial consistency between PROBA-V and Sentinel-3/OLCI based NDVI 300m V3?

This is an important question to answer, because it will demonstrate the consistency of the NDVI 300m V3 product over the whole time series, including the period 2014 – 2018 based on PROBA-V data and the period 2019 – 2024 (and future NRT observations) based on Sentinel-3/OLCI.

The analyses are performed over a one-year period, i.e. 2019 the year for which PROBA-V and Sentinel-3/OLCI observations overlap. For PROBA-V, a test dataset over the year 2019 is only available for the quality assessment but it is not included in the public product. In the disseminated product the switch between PROBA-V and Sentinel-3 is made at the start of 2019. The datasets used in these analyses are based on a LANDVAL V2 extractions for the statistical consistency analysis (see §4.3.4) and on global systematic subsample created with 300m resolution pixels for the spatial consistency analysis (see §4.3.2).

The analyses include:

- Statistical consistency analysis:
  - Histograms of overall bias
  - GM regression analysis, overall and per biome
- Spatial consistency analysis:
  - Histograms per biome
  - Spatial distribution of validation metrics
- Comparison of uncertainty values

5.4.2 Statistical consistency

5.4.2.1 Overall bias

Figure 30 shows the frequency histogram of the overall bias between NDVI 300m V3 Sentinel-3/OLCI and NDVI 300m V3 PROBA-V. The bias histogram is symmetrical with a peak at 0.0, indicating that in general, Sentinel-3/OLCI values and PROBA-V values are close.

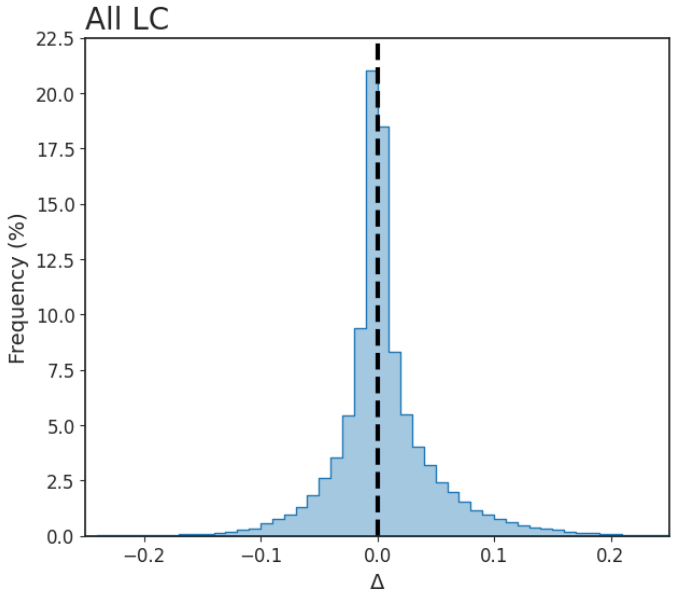


Figure 30: Frequency histogram of the overall bias between NDVI 300m V3 Sentinel-3/OLCI and NDVI 300m V3 PROBA-V for all LANDVAL V2 sites.

5.4.2.2 GM regression

The statistical consistency between NDVI 300m V3 Sentinel-3/OLCI and NDVI 300m V3 PROBA-V is investigated based on a GM regression analysis. Figure 31 shows scatter density plots per biome and, in the lower right corner, for all land cover types together. Statistics related to the GM regression over all land cover types are given in Table 8. These figures show a linear relationship between both datasets with regression line close to the 1:1 line. The slope for all land covers is slightly lower than one (slope=0.978), indicating that NDVI 300m V3 Sentinel-3/OLCI values are slightly higher than NDVI 300m V3 PROBA-V. The scatter plots per biome also show nearly one-on-one relationships between the two datasets. The spread is largest for the EBF, DBF, NLF, and MXF biome types, probably caused by differences is cloud, cloud shadow, and snow detection, and the influence of the MODIS prior when there are no valid TOC input reflectances available (NOBS=0).

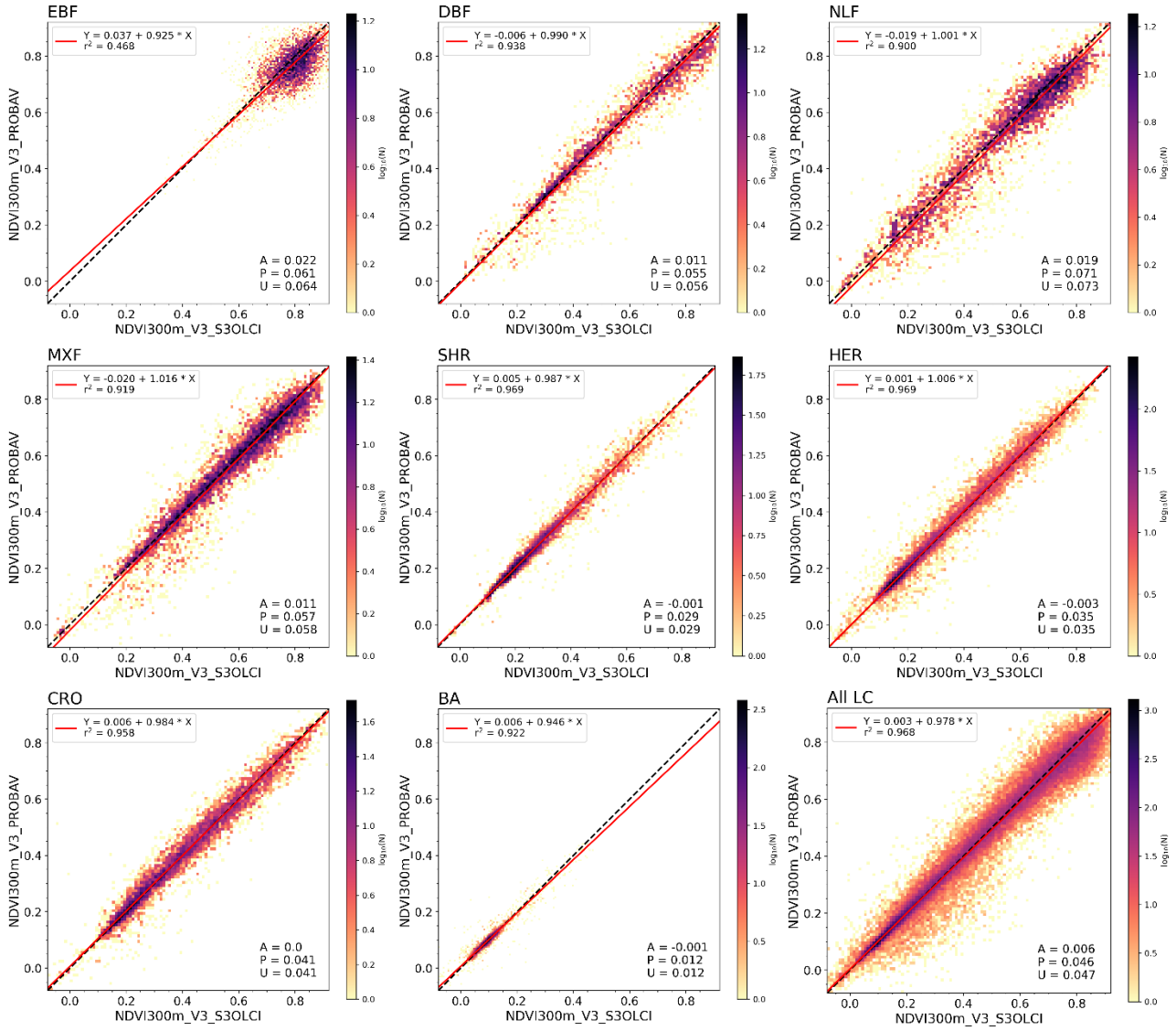


Figure 31: Scatter density plots between NDVI 300m V3 Sentinel-3/OLCI and NDVI 300m V3 PROBA-V over all LANDVAL V2 sites for the year 2019, per biome and for all land cover types.

Table 8: Statistics of comparison between NDVI 300m V3 Sentinel-3/OLCI and NDVI 300m V3 PROBA-V over all LANDVAL V2 sites for the year 2019.

N	R	Bias (%)	MD (%)	STD (%)	MAD (%)	RMSD (%)	offset / slope
56393	0.98	0.006 (1.4%)	0.0004 (0.11%)	0.05 (12%)	0.015 (4%)	0.05 (12%)	0.003 / 0.98

Figure 32 compares the APU statistics for all land cover types and per biome type, for all LANDVAL V2 sites. All three statistics are largest for the forest biomes (NLF, EBF, MXF, and DBF) and smallest for BA. The mean bias is very small for all land cover types combined indicating large agreement between the Sentinel-3/OLCI and PROBA-V datasets.

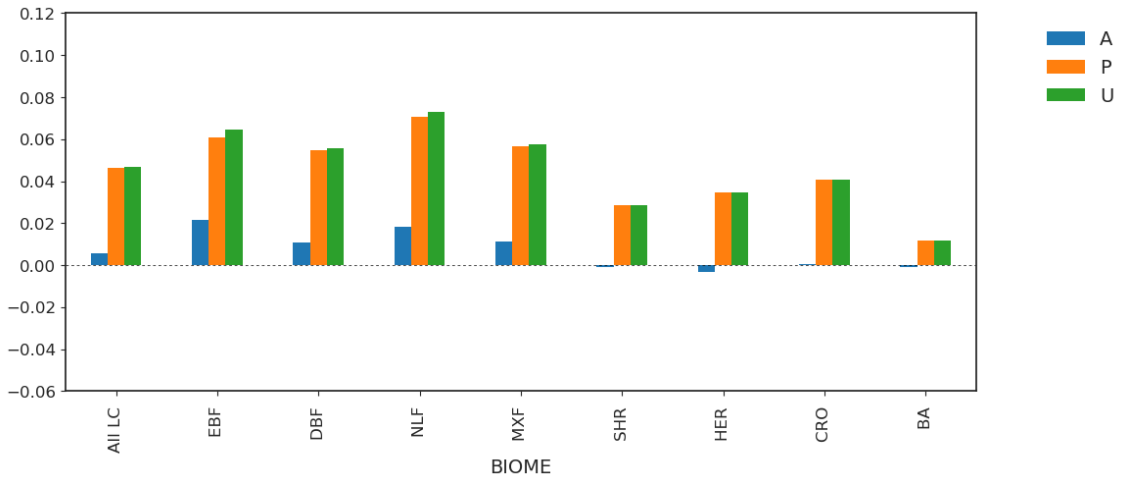


Figure 32: APU between NDVI 300m V3 Sentinel-3/OLCI and NDVI 300m V3 PROBA-V over all land cover types and for the individual biome types, computed over all LANDVAL V2 sites for the year 2019.

5.4.3 Spatial consistency

5.4.3.1 Histograms per biome

In Figure 33, the NDVI frequency distributions of NDVI 300m V3 Sentinel-3/OLCI and NDVI 300m V3 PROBA-V are compared over different biomes. For BA, HER, and SHR the differences are negligible, while the forest biomes (EBF, DBF, NLF, and MXF) and to a lesser extent the CRO biome show a small shift of the Sentinel-3/OLCI histograms towards higher NDVI values compared to PROBA-V. The shift is thus largest for the biomes with more dense vegetation. There is also more saturation at the maximum NDVI for these forest biomes in NDVI 300m V3 Sentinel-3/OLCI. The difference between PROBA-V and Sentinel-3/OLCI NDVI values is due to the MODIS prior, which has a larger influence on the Sentinel-3/OLCI data since, for this sensor, less valid input TOC reflectances are available for BRDF inversion (selection based on input TOC reflectance uncertainties is more stringent and NOBS=0 occurs more often, see §5.1).

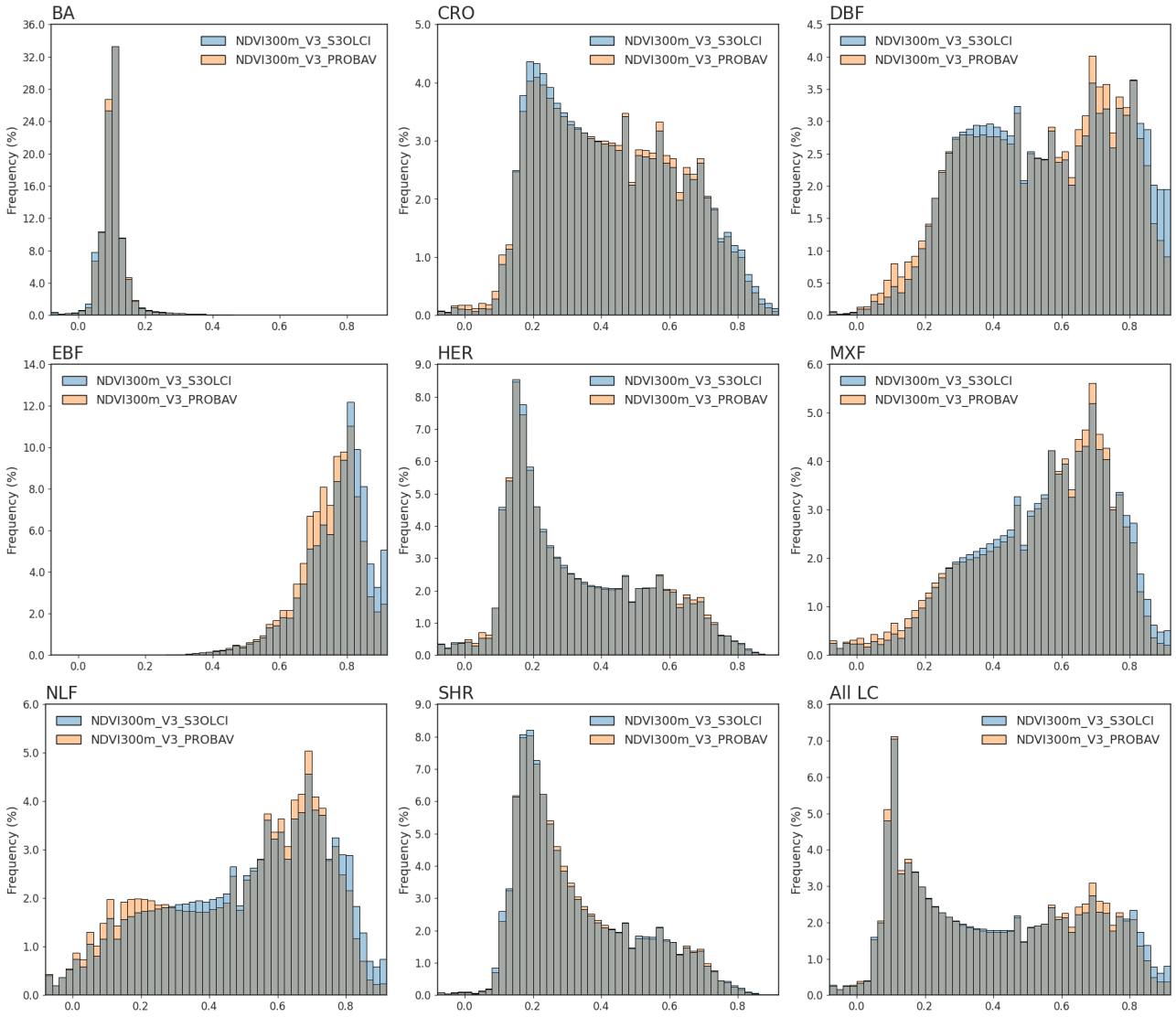


Figure 33: Frequency histograms per biome and for all land cover types comparing NDVI 300m V3 Sentinel-3/OLCI (blue) and NDVI 300m V3 PROBA-V (orange). Grey colour indicates where both histograms overlap.

5.4.3.2 Spatial distribution of validation metrics

The spatial distributions of accuracy, precision, and uncertainty between NDVI 300m V3 Sentinel-3/OLCI and NDVI 300m V3 PROBA-V are shown in Figure 34. The spatial patterns are strongly linked to vegetation densities, with higher bias and uncertainties in areas with dense vegetation (e.g. Amazon, central Africa, South-East Asia, etc.), which could already be concluded from the regression analysis in §5.4.2. The difference between both datasets is very small over deserts and other bare areas, e.g. the Sahara, the Arabian Peninsula, and Australia. The larger difference between the

Sentinel-3/OLCI and PROBA-V data in regions with higher vegetation density is probably related to the fact that a limited number of valid observations is available over e.g. the wet tropics, differences in spatial resolution and thus image sharpness, and to the influence of the MODIS prior.

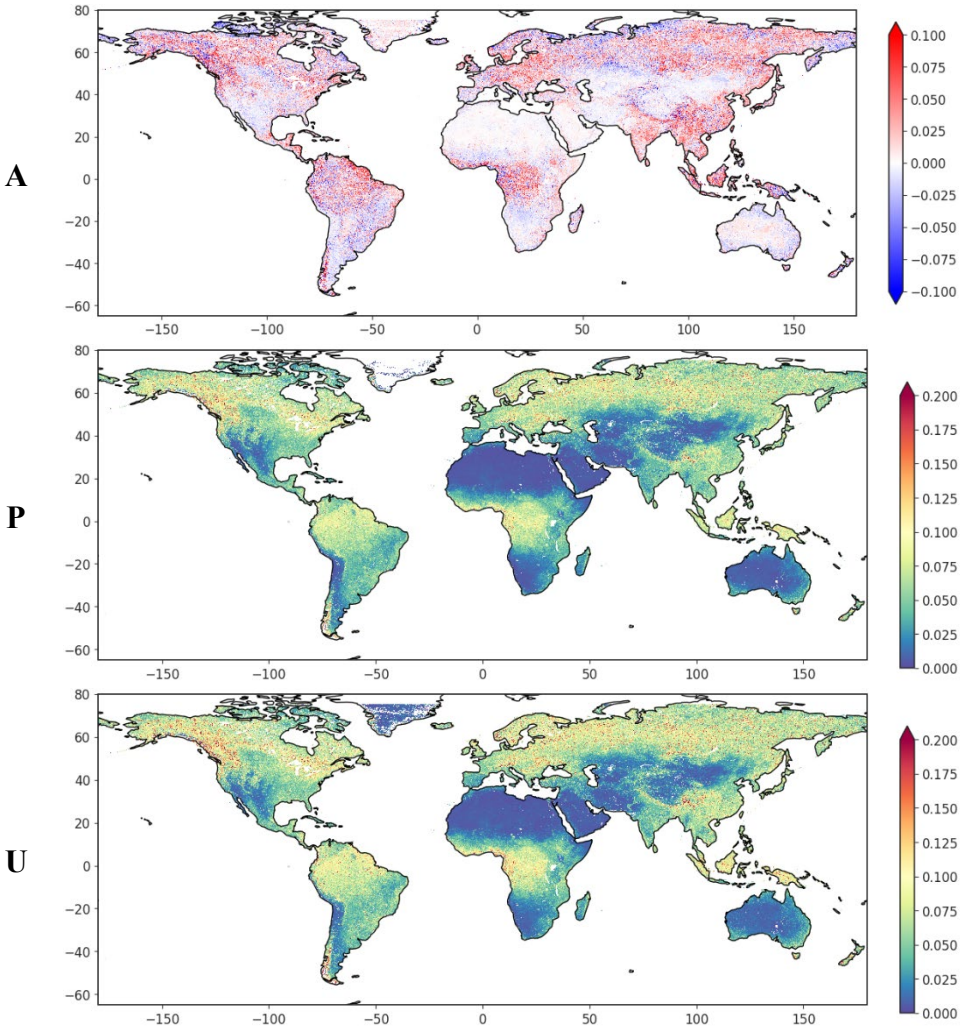


Figure 34: Spatial patterns of pixel-wise APU between NDVI 300m V3 Sentinel-3/OLCI and NDVI 300m V3 PROBA-V for all pairwise valid NDVI values for the year 2019.

5.4.4 Uncertainties

Figure 35 compares the frequency histograms of the NDVI uncertainty of PROBA-V and Sentinel-3/OLCI, while the spatio-temporal patterns of the mean NDVI uncertainty of the two sensors is shown in Figure 36. The Sentinel-3/OLCI uncertainties, with a mean value of 0.06, are generally larger than those of PROBA-V, with mean value of 0.03. This is related to proper error propagation in the

Sentinel-3/OLCI dataset, which includes uncertainties linked to atmospheric correction, while for PROBA-V default TOC reflectance uncertainties are used (see §4.2.5.7 and CGLOPS1_ATBD_BRDFCorrection300m-V1).

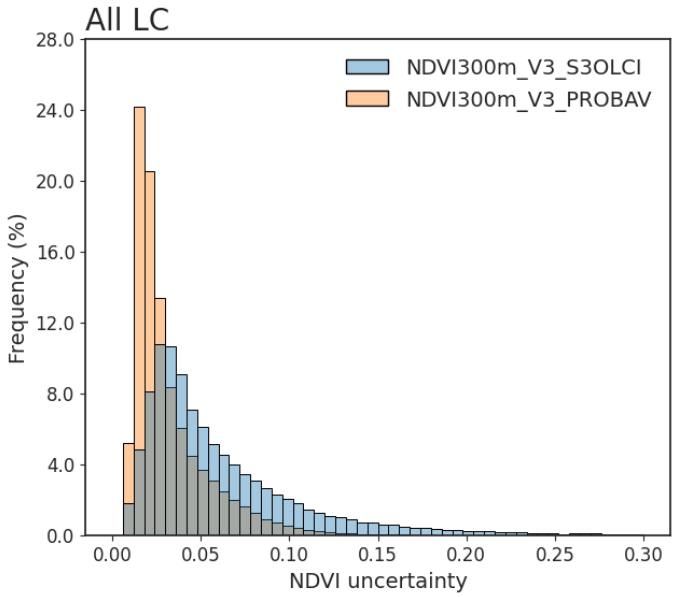


Figure 35: Frequency histograms of NDVI uncertainty for NDVI 300m V3 Sentinel-3/OLCI (blue) and NDVI 300m V3 PROBA-V (orange).

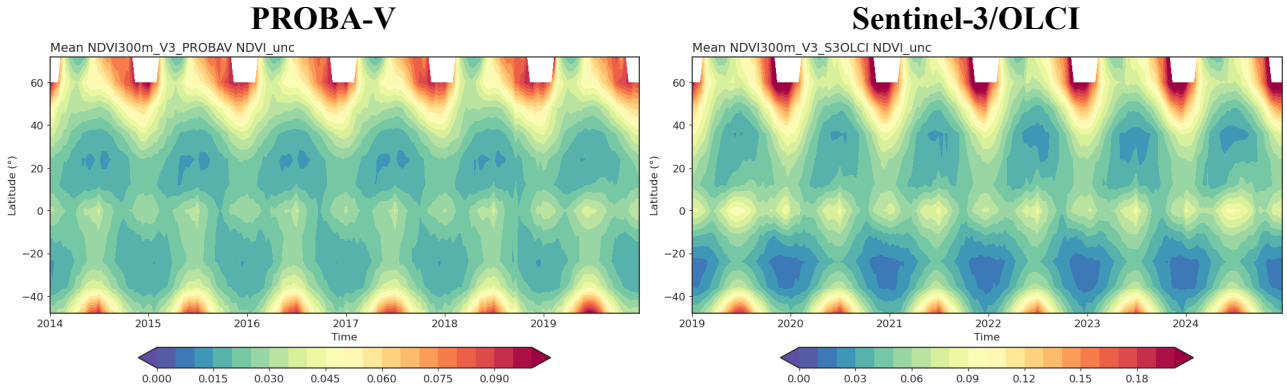


Figure 36: Hovmöller plots of the mean uncertainty of NDVI 300m V3 PROBA-V (left) and NDVI 300m V3 Sentinel-3/OLCI (right). Note that the range of the colour bar is twice as large for Sentinel-3/OLCI.

PROBA-V and Sentinel-3/OLCI mean NDVI uncertainty have similar spatio-temporal patterns as can be seen in Figure 36. In this figure, the range of the colour bar for Sentinel-3/OLCI is twice as large as for PROBA-V for easier comparison of the spatial patterns. In general, uncertainties are twice as large for Sentinel-3/OLCI. In the Southern hemisphere, the uncertainties for both sensors are more

comparable, while in the Northern hemisphere, and especially at high latitudes, the difference in uncertainties is larger.

5.4.5 Conclusion

What is the statistical and spatial consistency between PROBA-V and Sentinel-3/OLCI for NDVI 300m V3?

There is a very small positive bias (mean bias = 0.005, median bias = 0.000) between Sentinel-3/OLCI and PROBA-V for NDVI 300m V3, with Sentinel-3/OLCI NDVI values being slightly higher than PROBA-V values. The differences are strongly linked to biome types and are more pronounced (but small, up to 0.02 for EBF) for forest biomes, in the wet tropics, and at high latitudes. This is related to cloud/snow detection algorithms and the impact of the MODIS priors on the NDVI data.

The uncertainties for Sentinel-3/OLCI are on average a factor two larger than those of PROBA-V. This is because full error propagation is applied for the Sentinel-3/OLCI dataset, while, for PROBA-V S1 TOC inputs, default TOC reflectance uncertainties are applied [CGLOPS1_ATBD_BRDFCorrection300m-V1].

Overall, there is good consistency between the two datasets, considering that they are derived from different sensors with different spectral characteristics and different ground resolution. This indicates that the NDVI 300m V3 product is consistent over the whole period 2014 – NRT. This is investigated further in Section 5.8.

5.5 Q5: Statistical and spatial consistency between NDVI 300m V3, NDVI 300m V2, and NDVI 300m V1?

5.5.1 Introduction

The scientific question to be answered is the following: what is the difference between NDVI 300m V3, and NDVI 300m V1 and NDVI 300m V2 in terms of statistical and spatial consistency?

By comparing these three datasets, it is possible to quantify the differences between the different algorithms used. This is important to determine the effect of different input TOC data and different ReBeLS versions. To clarify the differences between product versions, the most important aspects that influence the NDVI product are summarized in Table 9.

Table 9: Summary of differences between CGLOPS NDVI 300m product versions for several aspects.

	NDVI 300m V3 (Sentinel-3 period)	NDVI 300m V3 (PROBA-V period)	NDVI 300m V2 (Sentinel-3 period)	NDVI 300m V1 (PROBA-V period)
Input data	Sentinel-3/OLCI TOC reflectance V2.3	PROBA-V Collection 2 daily TOC reflectance composites (S1_TOC)	Sentinel-3/OLCI TOC V1 (up to February/2025)	PROBA-V Collection 1 10-daily TOC reflectance composites (S10) (since December 2016)
	Per-pixel uncertainties propagated from actual TOA uncertainties	No per-pixel uncertainties	Per-pixel uncertainties propagated from default (2%) TOA uncertainty	No per-pixel uncertainties
Atmospheric correction	SMAC with uncertainty propagation, incl. aerosol model uncertainty	SMAC	SMAC with uncertainty propagation (no aerosol model uncertainty included)	SMAC
	AOT from CAMS	AOT from MERRA-2	AOT from CAMS	Image-retrieved AOT

BRDF correction	ReBeLS v1.6 with improved outlier detection and filtering of extreme geometries	ReBeLS v1.6 with improved outlier detection and filtering of extreme geometries	ReBeLS v1.3 30-days composite window	N/A
NDVI algorithm	NDVI based on normalized TOC reflectance V2.3 from OLCI Oa7 and Oa8 (red) and Oa16 and Oa18 (NIR)	NDVI based on normalized TOC reflectance from red and NIR bands	NDVI based on normalized TOC reflectance V1 from OLCI Oa7, Oa8, Oa9, and Oa10 (red) and Oa16, Oa17, and Oa18 (NIR)	Maximum NDVI composite based on surface directional reflectances

The analyses are performed over two one-year periods (2019 for the comparison with NDVI 300m V1, and 2023 for the comparison with NDVI 300m V2) and are based on LANDVAL V2 extractions for the statistical consistency (see §4.3.4) and on pairwise comparison of valid NDVI values for a global systematic subsample created with 300m resolution pixels (see §4.3.2) for the spatial consistency analysis.

The analyses include:

- Statistical consistency analysis:
  - Histograms of overall bias
  - GM regression analysis, overall and per biome
- Spatial consistency analysis:
  - Histograms per biome
  - Spatial distribution of validation metrics
- Comparison of uncertainty values between V3 and V2

5.5.2 Comparison of NDVI 300m V3 with NDVI 300m V1

5.5.2.1 Statistical consistency

5.5.2.1.1 Overall bias

The frequency histogram of the overall bias between NDVI 300m V3 PROBA-V and NDVI 300m V1 is shown in Figure 37. The bias histogram is slightly skewed towards negative values with a peak at -0.008. This indicates that the NDVI values in V3 are a bit lower than those in V1. The two versions differ in the applied atmospheric correction and in handling of BRDF effects. This explains the differences in NDVI values. Indeed, the atmospheric correction of TOC reflectances in V1 is based

on image-retrieved AOT and this has been shown to lead to higher NDVI values (see §4.2.5.7). The inclusion of BRDF correction in V3 also influences the comparison with V1. BRDF effects are influenced by vegetation density and SZA (see §0).

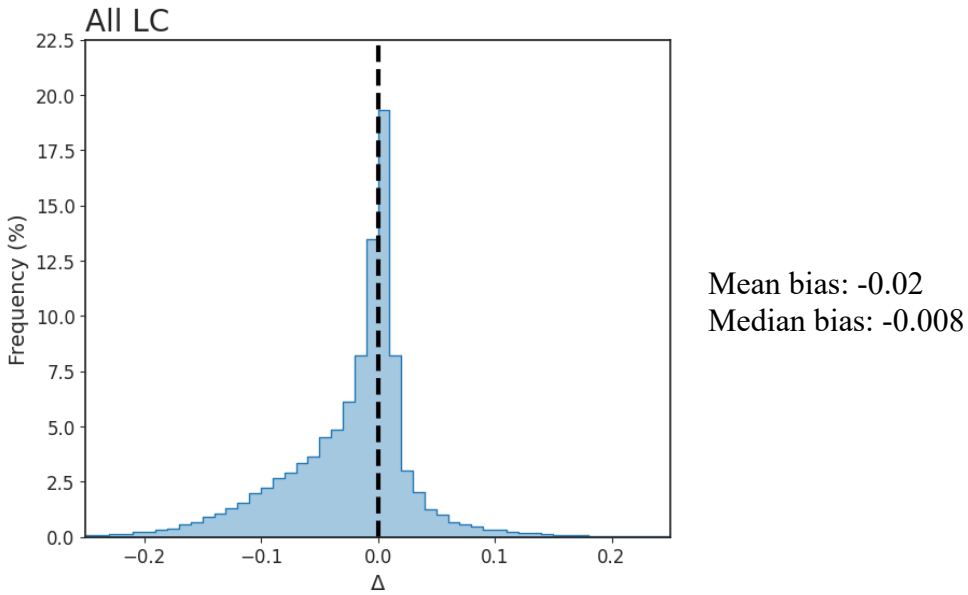


Figure 37: Frequency histogram of the overall bias between NDVI 300m V3 PROBA-V and NDVI 300m V1 for the year 2019 for all LANDVAL V2 sites.

5.5.2.1.2 GM regression

Scatter density plots demonstrating the statistical consistency between NDVI 300m V3 PROBA-V and NDVI 300m V1 per biome type and for all land cover types are shown in Figure 38 and Figure 39. For all biome types, the GM regression line deviates quite a bit from the 1:1 line. The slope of the GM regressions is positive, again indicating that the NDVI values in V1 are higher than those in V3 and the differences are larger for higher NDVI values. The scatter is largest for forest biome types, and especially for EBF. This is an effect of anisotropy coming from different illumination and viewing angles (see §0), which clearly shows the difference between V3 which is based on BRDF corrected reflectances while V1 is not. There is a small accumulation of negative NDVI values in the HER scatter plot. These are likely undetected snow pixels.

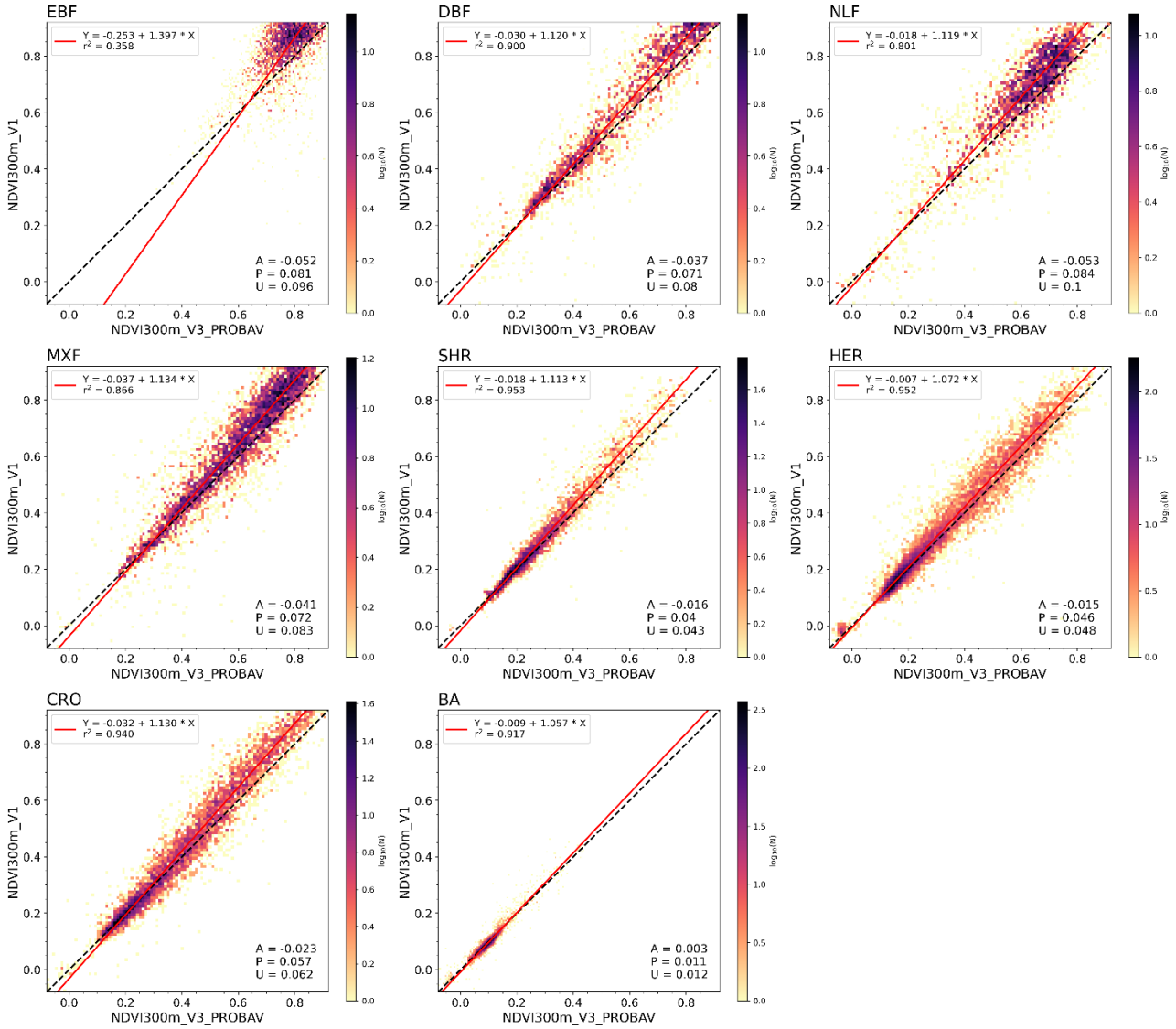


Figure 38: Scatter density plots between NDVI 300m V3 PROBA-V and NDVI 300m V1 per biome type over all LANDVAL V2 sites for the year 2019.

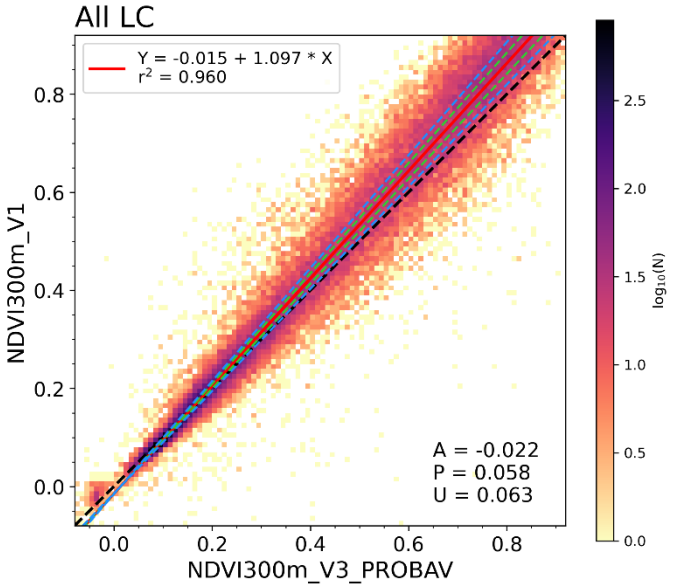


Figure 39: Scatter density plot between NDVI 300m V3 PROBA-V and NDVI 300m V1 over all LANDVAL V2 sites for the year 2019. Green and blue dashed lines indicate the goal and threshold uncertainty requirement.

Table 10 contains statistics related to the GM regression over all land cover types. The last column in the table are the percentages of LANDVAL V2 sites for which the uncertainty of NDVI V3 compared to NDVI V1 is within the goal and threshold limits. These are the LANDVAL V2 sites for which the NDVI lies within the green and blue dashed lines of Figure 39, respectively.

Table 10: Statistics of comparison between NDVI 300m V3 PROBA-V and NDVI 300m V1 over all LANDVAL V2 sites for the year 2019. The last column contains the percentage of pixels with uncertainty within the goal (2.5%) and threshold (5%) range.

N	R	Bias (%)	MD (%)	STD (%)	MAD (%)	RMSD (%)	Offset / slope	% Goal / Threshold
44995	0.98	-0.02 (-6%)	-0.008 (-2%)	0.06 (14%)	0.02 (5%)	0.06 (15%)	-0.015 / 1.1	23 / 42

Figure 40 compares the metrics of accuracy, precision, and uncertainty between NDVI 300m V3 PROBA-V and NDVI 300m V1 for all LANDVAL V2 sites and per biome type. The statistics are highest for forest biomes and smallest for BA. The mean bias (accuracy) is overall negative, indicating again that NDVI values in V3 are lower than in V1.

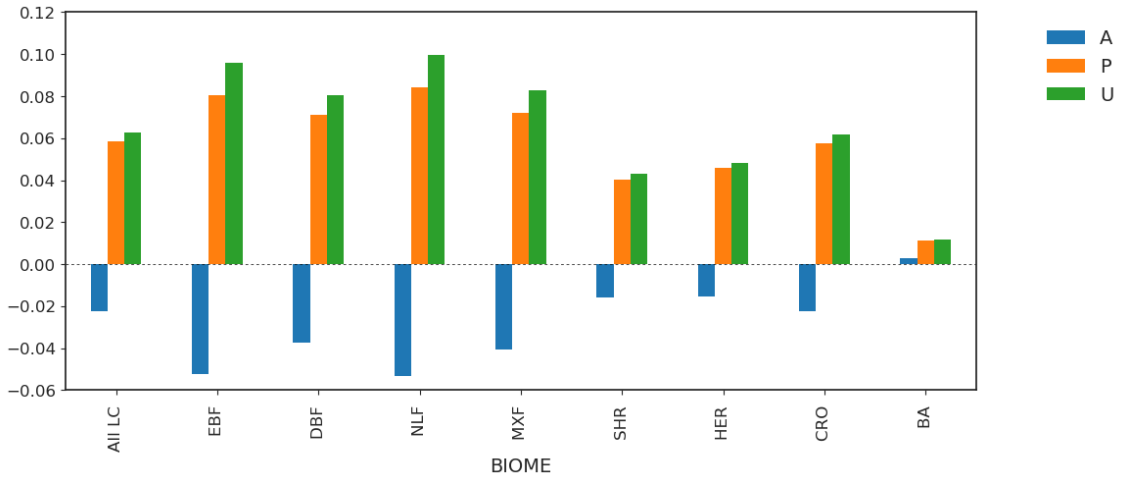


Figure 40: APU statistics between NDVI 300m V3 PROBA-V and NDVI 300m V1 for the year 2019.

5.5.2.2 Spatial consistency

5.5.2.2.1 Histograms per biome

The NDVI frequency distributions of NDVI 300m V3 PROBA-V and NDVI 300m V1 are compared per biome type in Figure 41. The differences are minimal for BA, while the forest biomes show large NDVI 300m V3 shifts towards lower values. The histograms of CRO, HER, and SHR are also shifted to lower values, but to a lesser extent. Where NDVI 300m V1 tends to saturate at 0.92, this is no longer the case for NDVI 300m V3 PROBA-V.

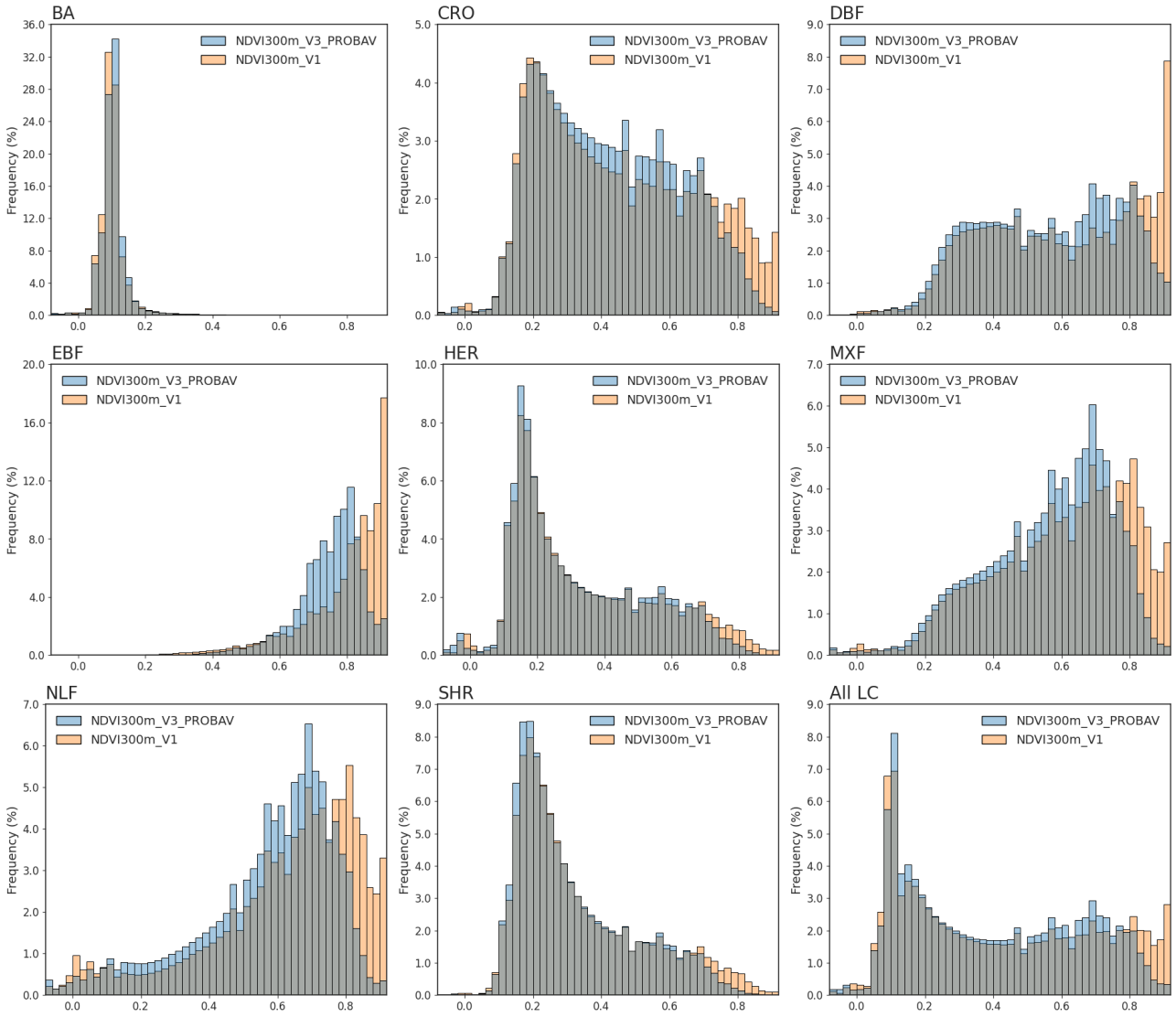


Figure 41: Frequency histograms per biome and for all land cover types comparing NDVI 300m V3 PROBA-V (blue) and NDVI 300m V1 (orange) for the year 2019. Grey colour indicates where both histograms overlap.

5.5.2.2.2 Spatial distribution of validation metrics

The spatial distribution of differences between NDVI 300m V3 PROBA-V and NDVI 300m V1, shown in Figure 42, is strongly linked with vegetation densities. The APU statistics are much higher in areas with dense vegetation such as the Amazon, central Africa, and South-East Asia, while they are close to zero over the Sahara, the Arabian Peninsula, and Australia. The mean bias with V1 is overall quite high and mostly negative.

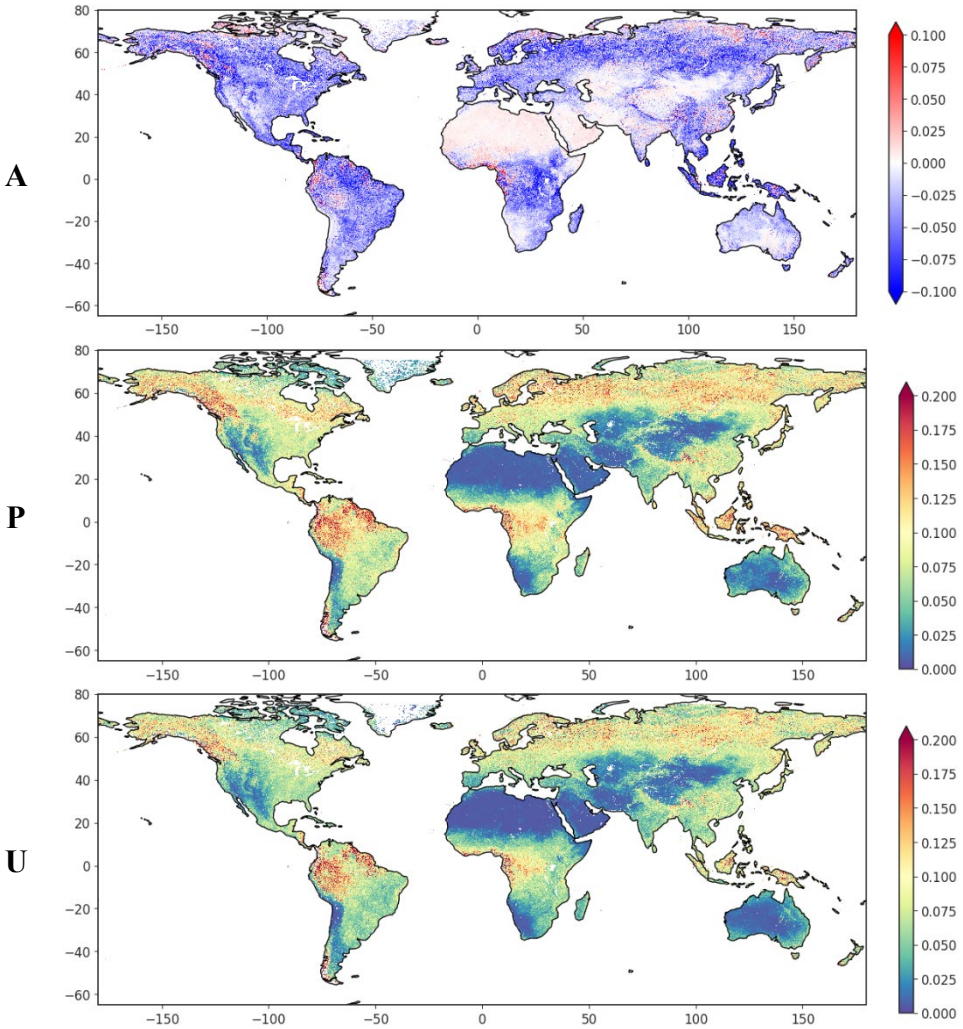


Figure 42: Spatial pattern of pixel wise APU between NDVI 300m V3 PROBA-V and NDVI 300m V1 for all pairwise valid NDVI values for the year 2019.

5.5.3 Comparison of NDVI 300m V3 with NDVI 300m V2

5.5.3.1 Statistical consistency

5.5.3.1.1 Overall bias

The bias histogram for comparison between NDVI 300m V3 Sentinel-3/OLCI and NDVI 300m V2 is shown in Figure 43. It peaks around 0.003, a bit higher than zero, and has a somewhat larger tail towards negative values. Therefore, NDVI values in V3 are slightly lower than those in V2. Differences between the two versions are the spectral band selection, differences in the input per-pixel uncertainties, difference in the BRDF accumulation window, and some changes in the BRDF

correction algorithm related to outlier detection. While the difference in the BRDF accumulation window (see e.g. Annex 2: Back-processing vs NRT processing) and the band selection [CGLOPS1_ATBD_NDVI300m-V3] should not lead to systematic bias, the differences in input uncertainties and ReBeLS algorithm have an influence on outlier detection, which might explain the skewedness in the bias histogram.

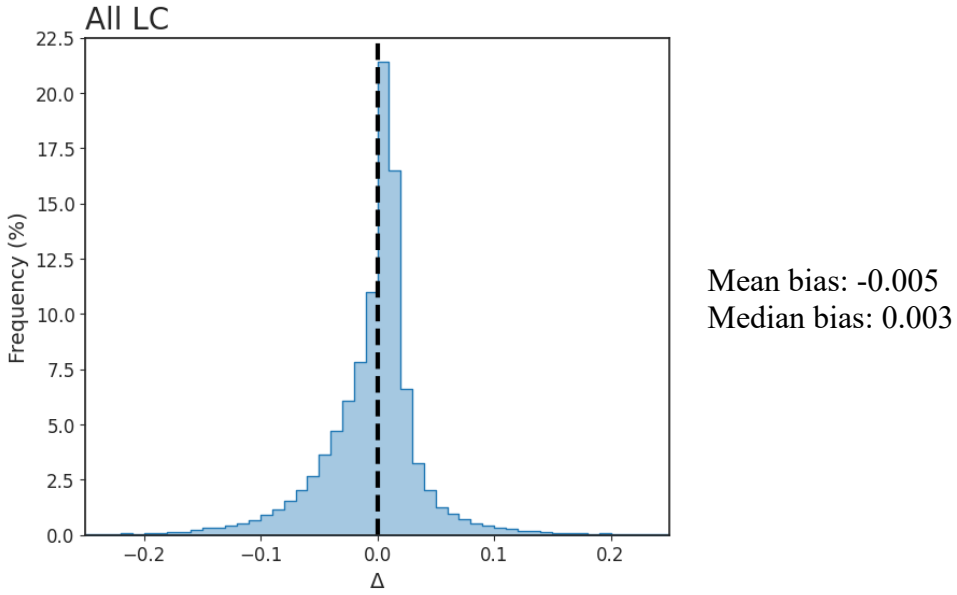


Figure 43: Frequency histogram of the overall bias between NDVI 300m V3 Sentinel-3/OLCI and NDVI 300m V2 for the year 2023 for all LANDVAL V2 sites.

5.5.3.1.2 GM regression

Scatter plots with GM regressions for statistical comparison between NDVI 300m V3 Sentinel-3/OLCI and NDVI 300m V2 are shown in Figure 44. The correlation between these two datasets is very good, with regression lines close to the 1:1 line. Even for forest biome types, the scatter is minimal because both datasets have been corrected for BRDF effects. Differences that remain are caused by different TOC input data and small algorithm differences.

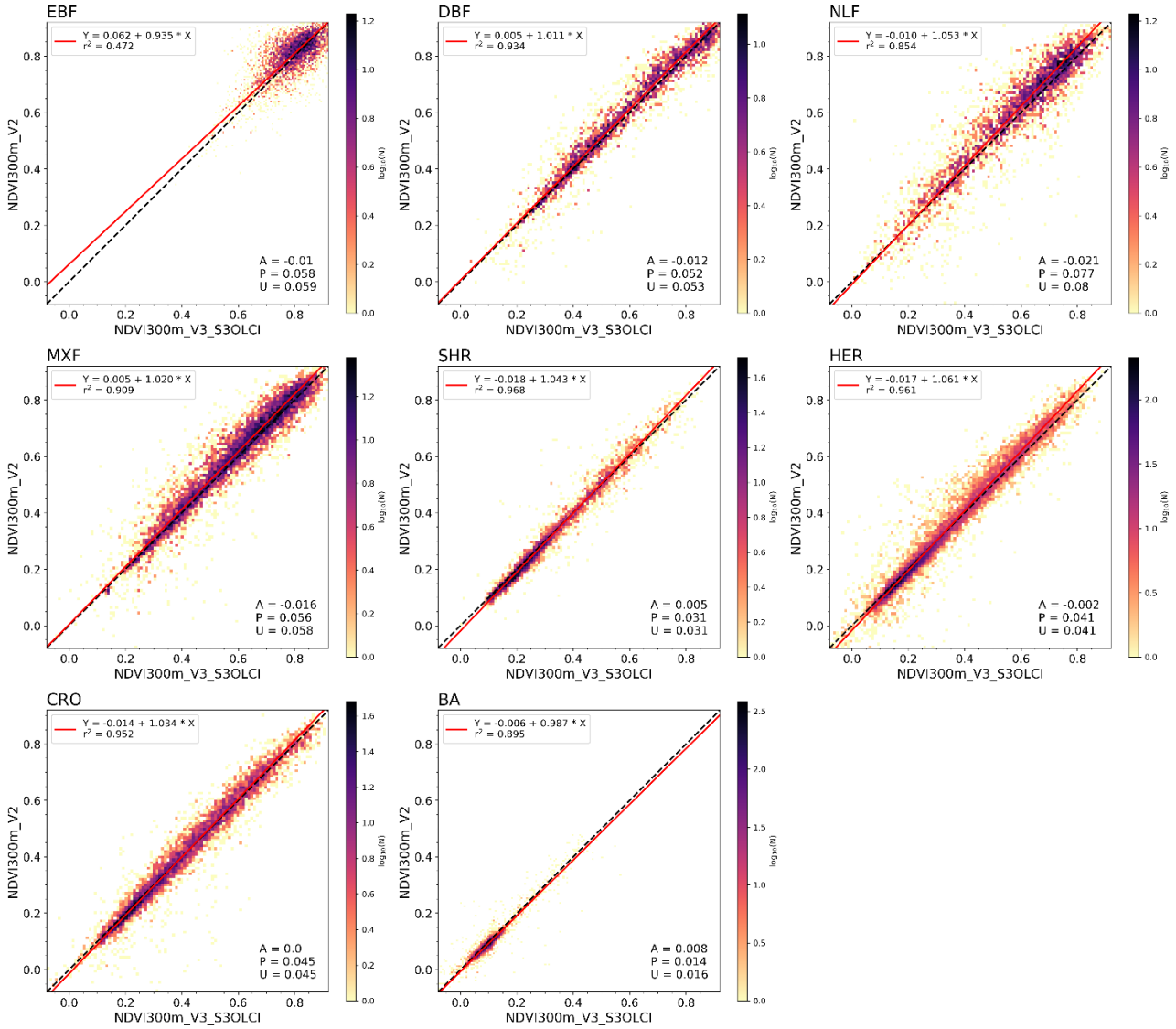


Figure 44: Scatter density plots between NDVI 300m V3 Sentinel-3/OLCI and NDVI 300m V2 per biome type over all LANDVAL V2 sites for the year 2023.

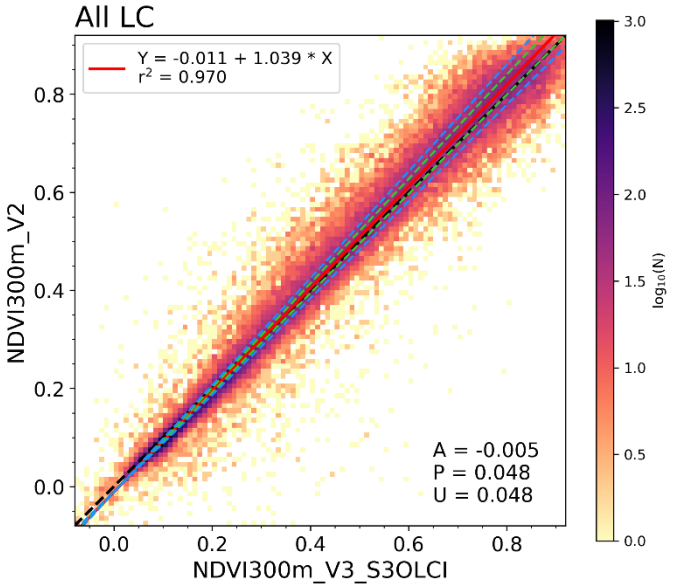


Figure 45: Scatter density plot between NDVI 300m V3 Sentinel-3/OLCI and NDVI 300m V2 over all LANDVAL V2 sites for the year 2023. Green and blue dashed lines indicate the goal and threshold uncertainty requirement.

Table 11 contains statistics related to the GM regression over all land cover types. The last column in the table are the percentages of LANDVAL V2 sites for which the uncertainty of NDVI V3 compared to NDVI V2 is within the goal and threshold limits. These are the LANDVAL V2 sites for which the NDVI lies within the green and blue dashed lines of Figure 45, respectively.

Table 11: Statistics of comparison between NDVI 300m V3 Sentinel-3/OLCI and NDVI 300m V2 over all LANDVAL V2 sites for the year 2023. The last column contains the percentage of pixels with uncertainty within the goal (2.5%) and threshold (5%) range.

N	R	Bias (%)	MD (%)	STD (%)	MAD (%)	RMSD (%)	Offset / slope	% Goal / Threshold
53929	0.98	-0.005 (-1.1%)	0.003 (0.8%)	0.05 (11%)	0.016 (4%)	0.05 (12%)	-0.011 / 1.04	28 / 51

Figure 46 shows the APU statistics between NDVI 300m V3 Sentinel-3/OLCI and NDVI 300m V2 for the year 2023 for all LANDVAL V2 sites and per biome. As expected, the mean bias, the precision metric, and the uncertainty metric are lower than for the comparison between V3 and V1.

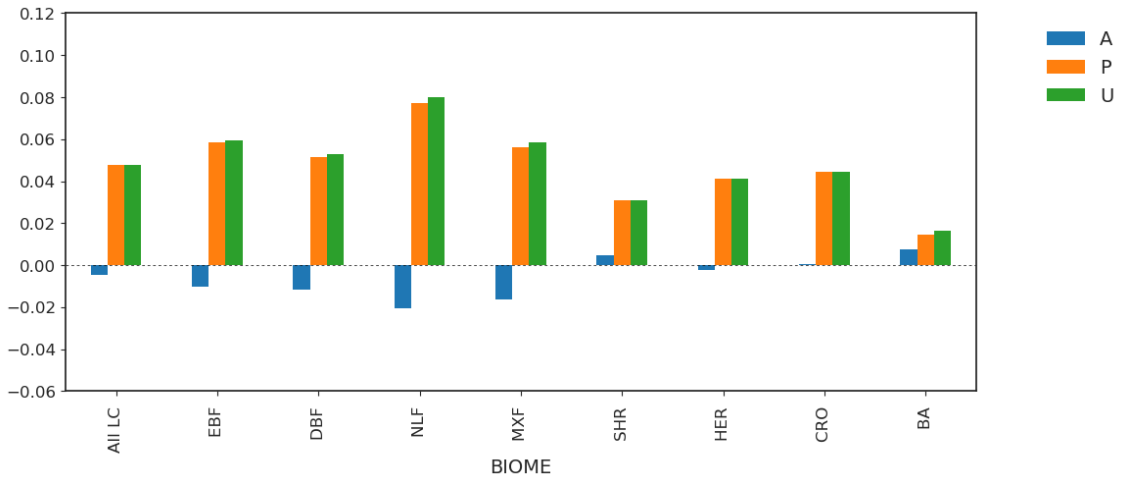


Figure 46: APU statistics between NDVI 300m V3 Sentinel-3/OLCI and NDVI 300m V2 for the year 2023.

5.5.3.2 Spatial consistency

5.5.3.2.1 Histograms per biome

The comparison of NDVI frequency distributions between NDVI 300m V3 Sentinel-3/OLCI and NDVI 300m V2 is shown in Figure 47. The histograms mostly overlap but some small differences are visible although the histograms are not shifted in a particular direction. The NDVI 300m V3 dataset contains fewer extreme values and therefore the histograms are narrower.

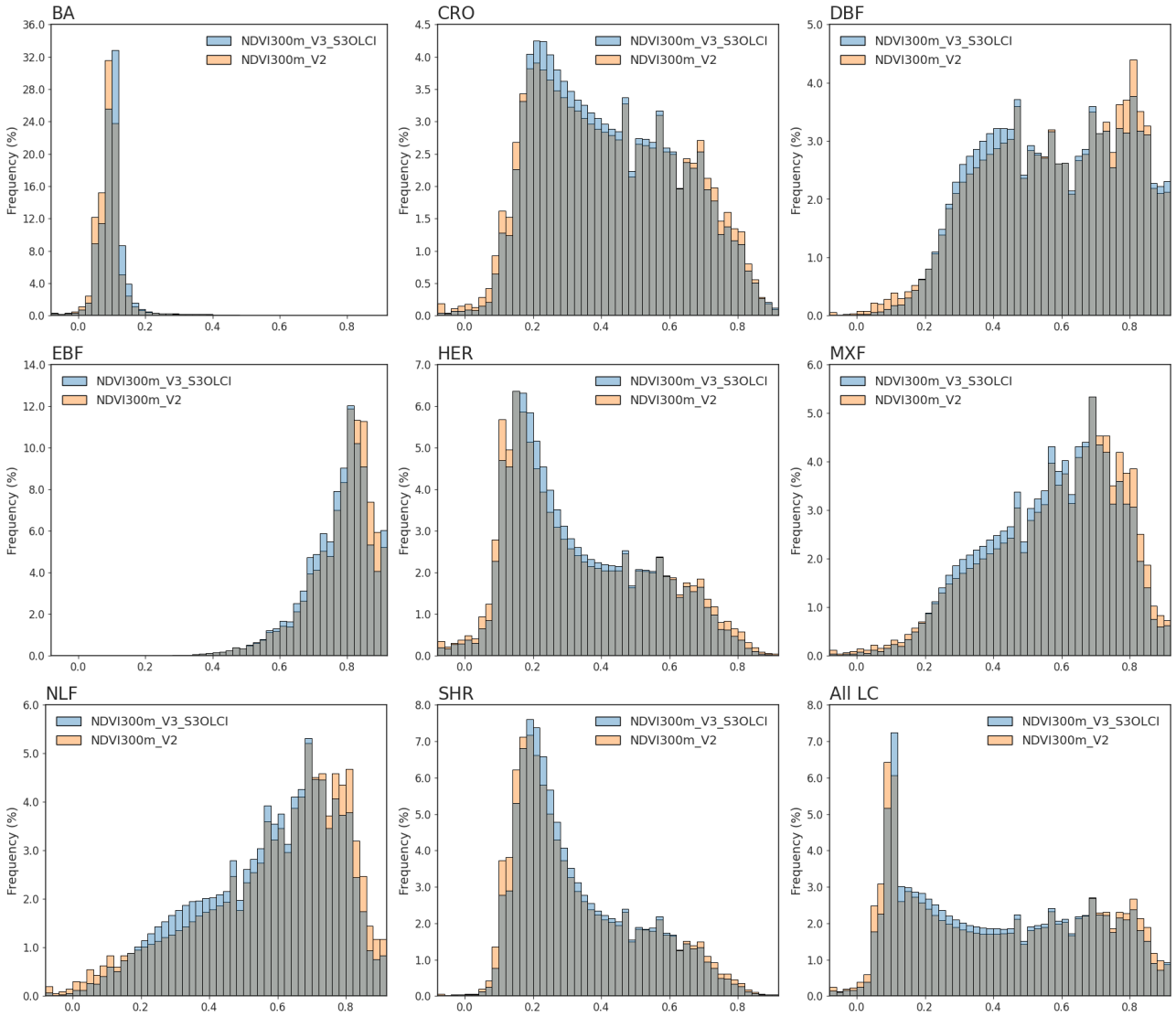


Figure 47: Frequency histograms per biome and for all land cover types comparing NDVI 300m V3 Sentinel-3/OLCI (blue) and NDVI 300m V2 (orange) for the year 2023. Grey colour indicates where both histograms overlap.

5.5.3.2.2 Spatial distribution of validation metrics

The spatial distributions of the APU statistics between NDVI 300m V3 Sentinel-3/OLCI and NDVI 300m V2 are shown in Figure 48. The difference between both datasets is highest at high latitudes and in tropical regions. These regions are mostly affected by clouds and high SZA and thus differences in outlier rejection methodology are most visible here. The difference between back

processing and NRT processing also influences these patterns (see Annex 2: Back-processing vs NRT processing (Scientific Quality Monitoring 2025)).

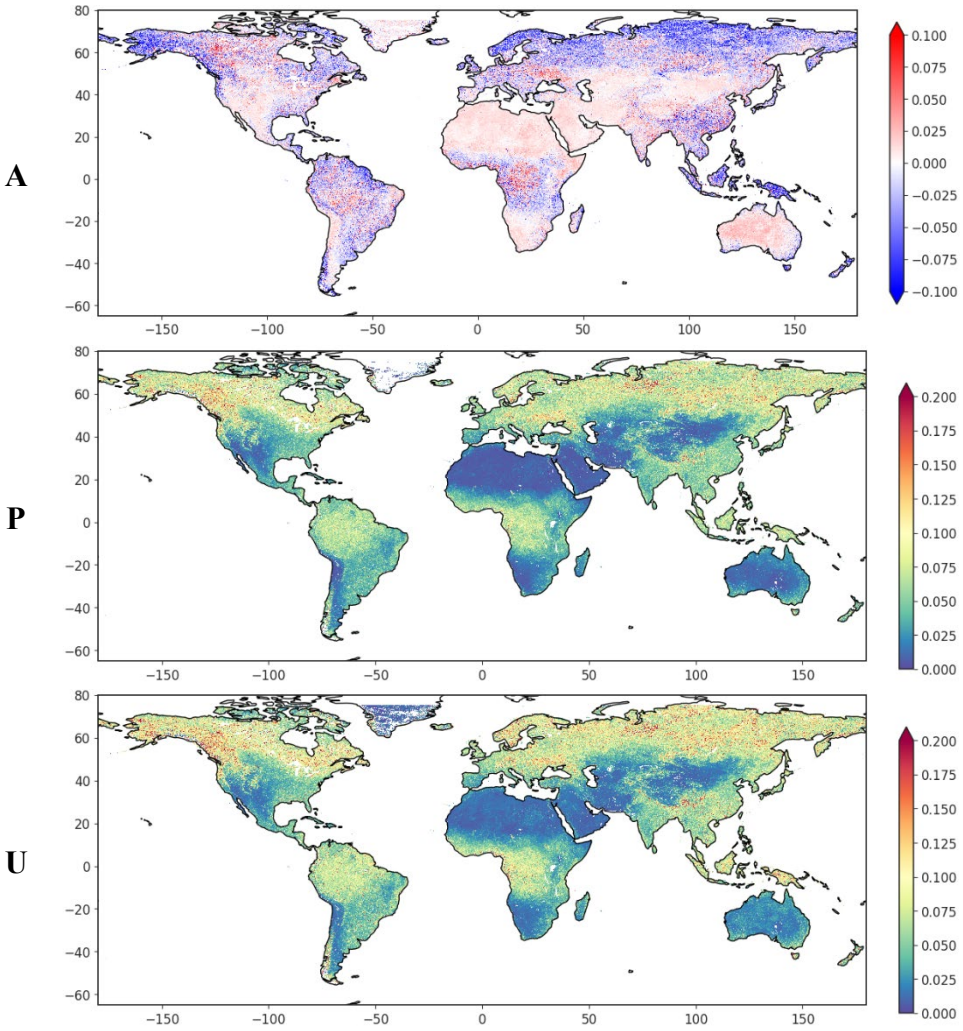


Figure 48: Spatial pattern of pixel wise APU between NDVI 300m V3 Sentinel-3/OLCI and NDVI 300m V2 for all pairwise valid NDVI values for the year 2023.

5.5.3.3 Uncertainties

Figure 49 compares the histograms of uncertainty associated to the NDVI 300m V3 Sentinel-3/OLCI and NDVI 300m V2, while spatio-temporal plots are shown in Figure 50. The patterns in

time and space are similar for the two datasets, but the V3 uncertainties are higher due to higher input TOC reflectance uncertainties.

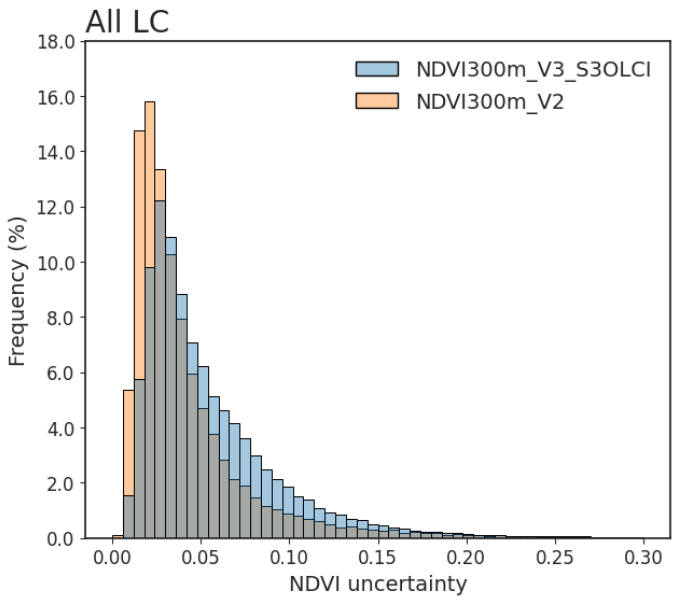


Figure 49: Frequency histograms of NDVI uncertainty for NDVI 300m V3 Sentinel-3/OLCI (blue) and NDVI 300m V2 (orange).

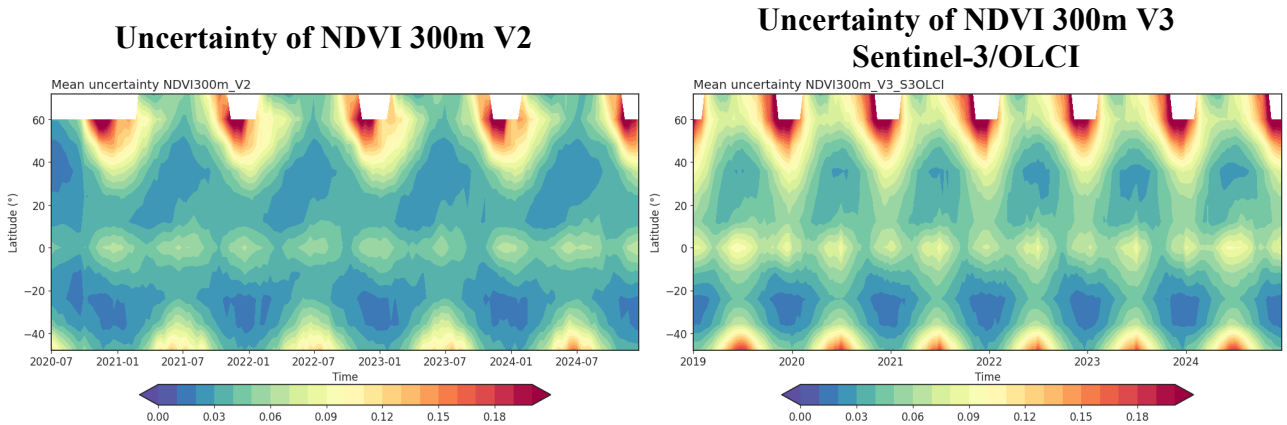


Figure 50: Hovmöller plots of the mean uncertainty of NDVI 300m V2 (left) and NDVI 300m V3 Sentinel-3/OLCI (right).

5.5.4 Conclusion

What is the difference between NDVI 300m V3 and NDVI 300m V1 in terms of statistical and spatial consistency?

There is a small negative bias between NDVI 300m V3 and NDVI 300m V1 (mean bias = -0.02, median bias = -0.008). The bias histogram is largely skewed towards negative values indicating that V3 NDVI values are smaller than V1 values. The differences are larger for higher NDVI values, and the spatial distribution of the differences is strongly linked to vegetation densities.

Differences between both datasets are expected because V3 is based on BRDF normalised reflectances, while anisotropy effects are still present in V1, and the atmospheric correction is performed based on different AOT inputs.

What is the difference between NDVI 300m V3 and NDVI 300m V2 in terms of statistical and spatial consistency?

A small bias exists between NDVI 300m V3 and NDVI 300m V2 (mean bias = -0.005, median bias = 0.003) but the datasets are much more consistent than V1 versus V3. The bias between V2 and V3 is small for all biome types, and the spatial distribution is more random.

Small differences are expected due to differences in the ReBeLS BRDF correction algorithm version and difference in the input data (mainly regarding TOC reflectance uncertainties).

5.6 Q6: STATISTICAL AND SPATIAL CONSISTENCY BETWEEN NDVI 300M V3 AND NDVI 1KM V3?

5.6.1 Introduction

The scientific question to be answered is the following: what is the statistical and spatial consistency with NDVI 1km V3?

This is a relevant question for users that require a longer time series, i.e. including the time series of SPOT/VGT1, SPOT/VGT2, and PROBA-V at 1km resolution, for example for generation of vegetation anomaly indices where actual observations are compared to long term statistics.

The analyses are performed over a one-year period (2019) and are based on LANDVAL V2 extractions for the statistical consistency analysis (see §4.3.4) and on a global systematic subsample created with 1km resolution pixels (see §4.3.2) for the spatial consistency analysis.

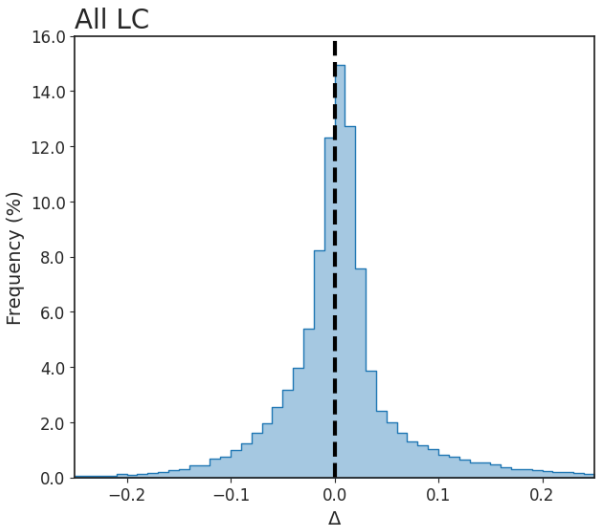
The analyses include:

- Statistical consistency analysis:
  - Histograms of overall bias
  - GM regression analysis
- Spatial consistency analysis:
  - Histograms per biome
  - Spatial distribution of validation metrics

5.6.2 Statistical consistency

5.6.2.1 Overall bias

Figure 51 shows the frequency histogram of the overall bias between NDVI 300m V3 Sentinel-3/OLCI and NDVI 1km V3 for all pairwise valid NDVI values. The histogram is symmetrical with a peak at 0.003 and almost no skewedness. The 300m NDVI data is therefore slightly higher than the 1km NDVI values.



Mean bias: 0.007
Median bias: 0.003

Figure 51: Frequency histogram of the overall bias between NDVI 300m V3 Sentinel-3/OLCI and NDVI 1km V3 for all LANDVAL V2 sites for the year 2019.

5.6.2.2 GM regression

A scatterplot between NDVI 300m V3 and NDVI 1km V3 for all LANDVAL V2 sites is shown in Figure 52 for all pairwise valid NDVI values. The regression line is very close to the 1:1 line, but there is quite some scatter present, especially for high NDVI values. The statistics related to this regression are given in Table 12. The last column in the table are the percentages of LANDVAL V2 sites for which the uncertainty of NDVI 300m V3 compared to NDVI 1km V3 is within the goal and threshold limits. These are the LANDVAL V2 sites for which the NDVI lies within the green and blue dashed lines of Figure 52, respectively.

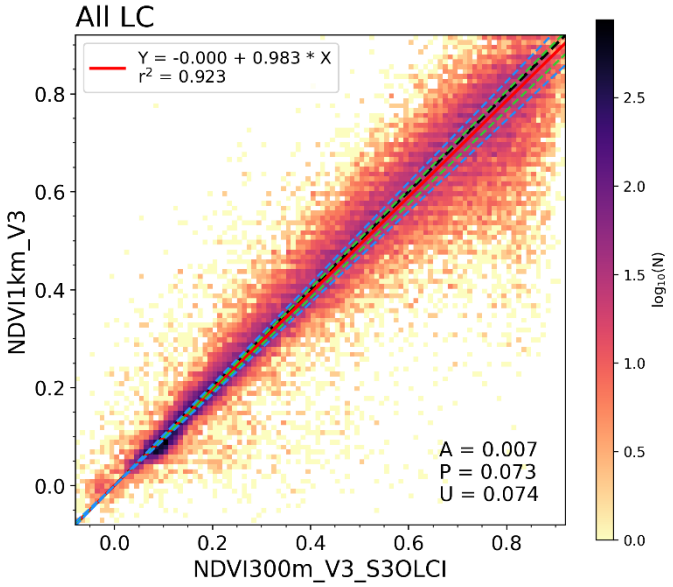


Figure 52: Scatter density plot between NDVI 300m V3 Sentinel-3/OLCI and NDVI 1km V3 over all LANDVAL V2 sites for the year 2019. Green and blue dashed lines indicate the goal and threshold uncertainty requirement.

Table 12: Statistics of comparison between NDVI 300m V3 Sentinel-3/OLCI and NDVI 1km V3 over all LANDVAL V2 sites for the year 2019. The last column contains the percentage of pixels with uncertainty within the goal (2.5%) and threshold (5%) range.

N	R	Bias (%)	MD (%)	STD (%)	MAD (%)	RMSD (%)	Offset / slope	% Goal / Threshold
52408	0.96	0.007 (1.9%)	0.004 (0.9%)	0.07 (19%)	0.02 (5%)	0.07 (19%)	-0.0004 / 0.98	18 / 35

5.6.3 Spatial consistency

5.6.3.1 Histograms per biome

The histograms per biome for NDVI 300m V3 and NDVI 1km V3 (Figure 53) confirm the good consistency between both datasets. There is some deviation for forest biome types, and especially for EBF for which the NDVI 300m V3 dataset has a much narrower distribution than NDVI 1km V3. A possible reason is the absence of a prior and the shorter accumulation window in the BRDF correction of NDVI 1km V3 data. This is especially important in regions with EBF biome type that have lots of temporal gaps due to persistent cloud cover since BRDF inversion yields more stable results when the accumulation period is longer.

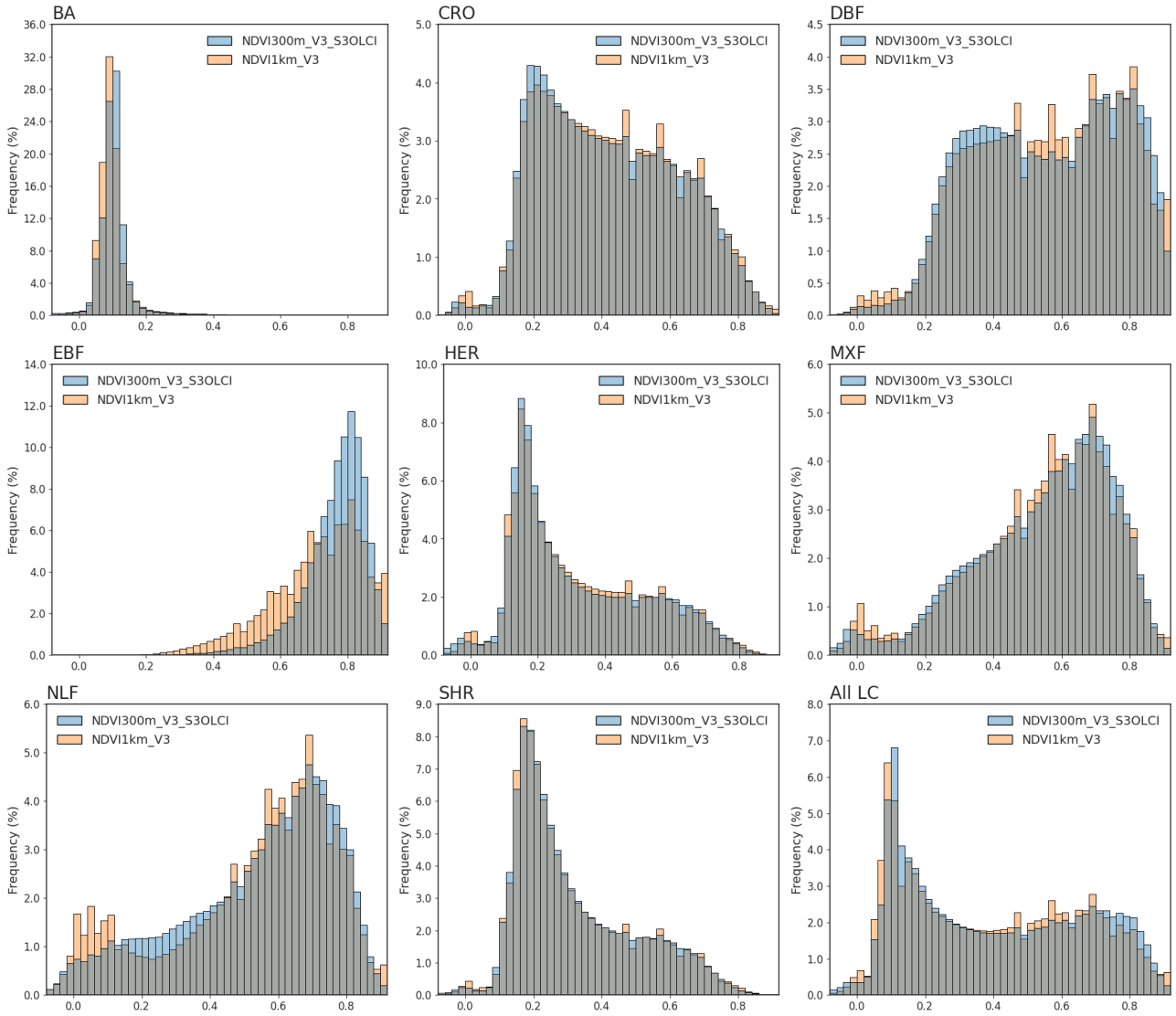


Figure 53: Frequency histograms per biome and for all land cover types comparing NDVI 300m V3 Sentinel-3/OLCI (blue) and NDVI 1km V3 (orange) for the year 2019. Grey colour indicates where both histograms overlap.

5.6.3.2 Spatial distribution of validation metrics

The spatial patterns of APU statistics between NDVI 300m V3 and NDVI 1km V3 are shown in Figure 54. Here again, a large difference is visible over tropical areas with permanent cloud cover.

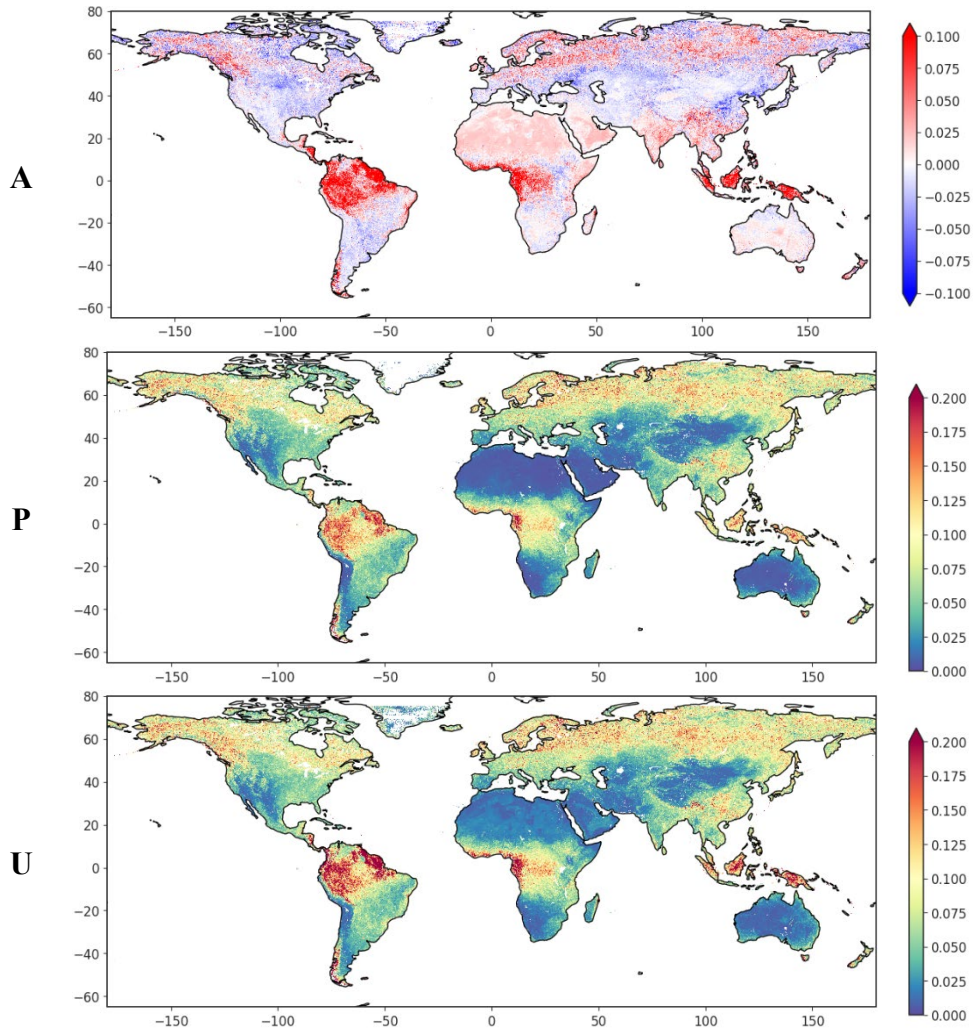


Figure 54: Spatial pattern of pixel APU between NDVI 300m V3 Sentinel-3/OLCI and NDVI 1km V3 for all pairwise valid NDVI for the year 2019.

5.6.4 Conclusion

What is the statistical and spatial consistency with NDVI 1km V3?

NDVI 300m V3 shows overall large statistical and spatial consistency with NDVI 1km V3. The differences are mostly unsystematic, except for the EBF biome and to lesser extent the other forest biomes, probably related to the higher temporal smoothness in NDVI 300m V3 thanks to the longer accumulation period in the BRDF inversion (see §5.8.3).

5.7 Q7: STATISTICAL AND SPATIAL CONSISTENCY BETWEEN NDVI 300M V3 AND MCD43A4?

5.7.1 Introduction

The scientific question to be answered is the following: what is the spatial and statistical consistency of NDVI 300m V3 in comparison to MCD43A4 V6.1?

The analyses are performed over a one-year period (2019) and are based on LANDVAL V2 extractions for the statistical consistency analysis (see §4.3.4) and on a global systematic subsample created with 1km resolution pixels (see §4.3.2) for the spatial consistency analysis.

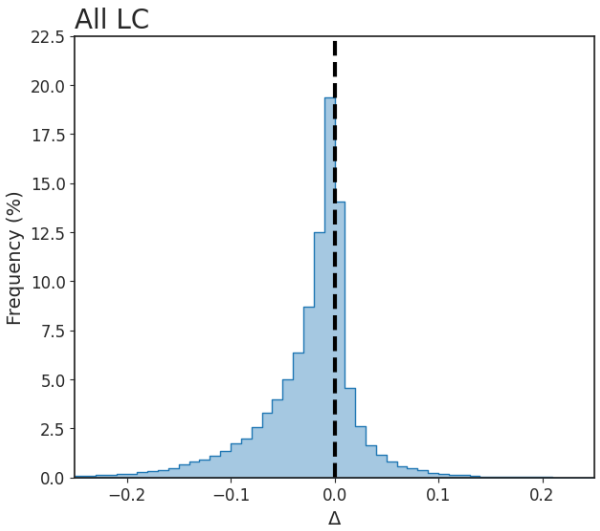
The analyses include:

- Statistical consistency analysis:
  - Histograms of overall bias
  - GM regression analysis
- Spatial consistency analysis:
  - Histograms and bias histograms per biome
  - Spatial distribution of validation metrics

5.7.2 Statistical consistency

5.7.2.1 Overall bias

The bias frequency histogram in Figure 55 indicates that, in general, NDVI 300m V3 is slightly lower than MCD43A4 V6.1. The bias histogram peaks around -0.012 and is skewed towards negative values. Differences in NDVI are expected based on differences between satellites, sensors, and processing (see §4.2.5), and although the bias histogram is slightly skewed, it is quite narrow.



Mean bias: -0.02
Median bias: -0.012

Figure 55: Frequency histogram of the overall bias between NDVI 300m V3 Sentinel-3/OLCI and MCD43A4 V6.1 for all LANDVAL V2 sites for the year 2019.

5.7.2.2 GM regression

Figure 56 shows the scatterplot between NDVI 300m V3 and MCD43A4 V6.1. The regression line deviates from the 1:1 line, especially for higher NDVI values, but the intercept is close to zero. For higher NDVI values, MODIS MCD43A4 V6.1 values are therefore higher than NDVI 300m V3 NDVI values.

The statistics related to the GM regression are given in Table 13. The last column in the table are the percentages of LANDVAL V2 sites for which the uncertainty of NDVI 300m V3 compared to MCD43A4 V6.1 is within the goal and threshold limits. These are the LANDVAL V2 sites for which the NDVI lies within the green and blue dashed lines of Figure 56, respectively.

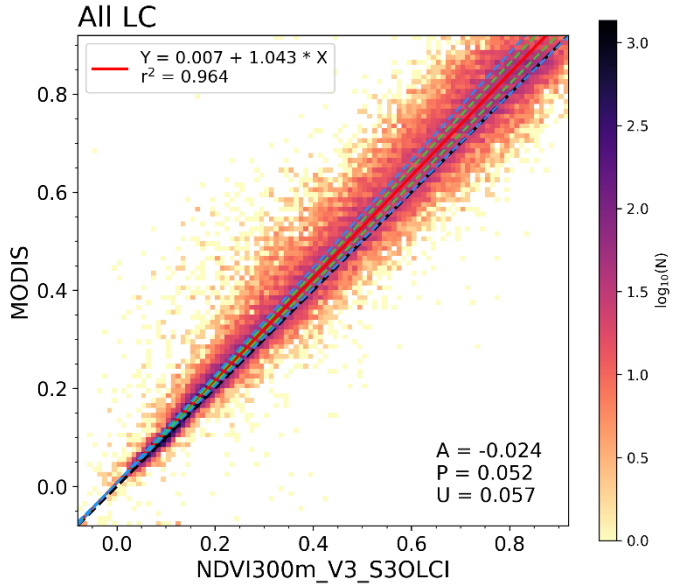


Figure 56: Scatter density plot between NDVI 300m V3 Sentinel-3/OLCI and MCD43A4 V6.1 for all LANDVAL V2 sites for the year 2019. Green and blue dashed lines indicate the goal and threshold uncertainty requirement.

Table 13: Statistics of comparison between NDVI 300m V3 Sentinel-3/OLCI and MODIS MCD43A4 V6.1 over all LANDVAL V2 sites for the year 2019. The last column contains the percentage of pixels with uncertainty within the goal (2.5%) and threshold (5%) range.

N	R	Bias (%)	MD (%)	STD (%)	MAD (%)	RMSD (%)	Offset / slope	% Goal / Threshold
47619	0.98	-0.02 (-6%)	-0.012 (-3%)	0.05 (12%)	0.02 (5%)	0.06 (14%)	0.007 / 1.04	21 / 41

5.7.3 Spatial consistency

5.7.3.1 Histograms per biome

Histograms per biome (Figure 57) show relatively good consistency between NDVI 300m V3 and MCD43A4 V6.1. Especially for BA, HER, and SHR the histograms largely overlap. For the forest biome types, the difference is larger, due to spectral differences between the sensors. Especially the spectral coverage of the NIR bands is very different between Sentinel-3/OLCI and MODIS (see §4.2.5.5) leading to large differences in NDVI.

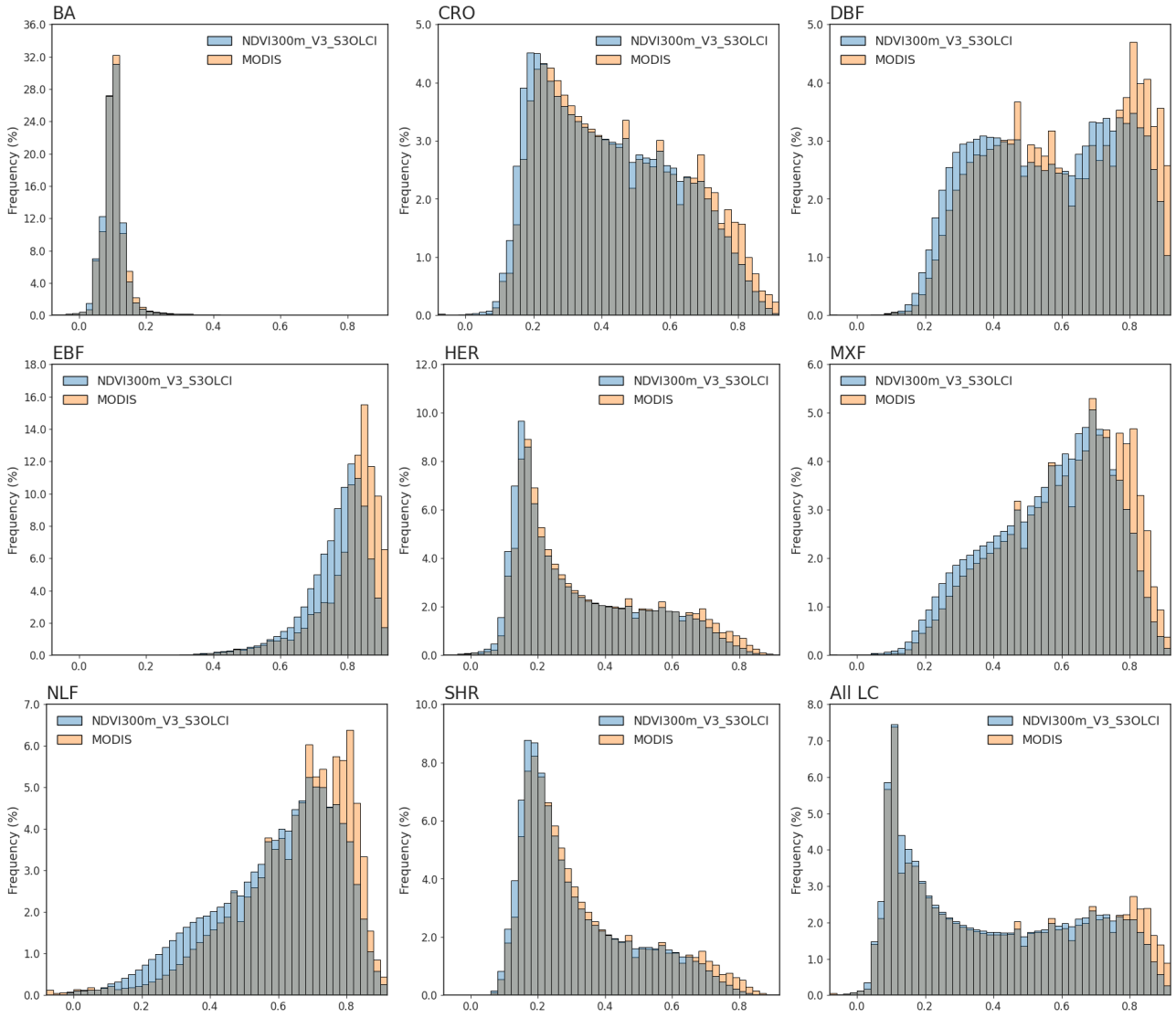


Figure 57: Frequency histograms per biome and for all land cover types comparing NDVI 300m V3 Sentinel-3/OLCI (blue) and MCD43A4 V6.1 (orange) for the year 2019. Grey colour indicates where both histograms overlap.

5.7.3.2 Spatial distribution of validation metrics

Spatial patterns of accuracy, precision, and uncertainty between NDVI 300m V3 and MCD43A4 V6.1 (see Figure 58) show that differences are distributed globally. The bias is mostly negative, indicating that MODIS values are higher than NDVI 300m V3. In desert regions, with very low vegetation densities, the bias is close to zero.

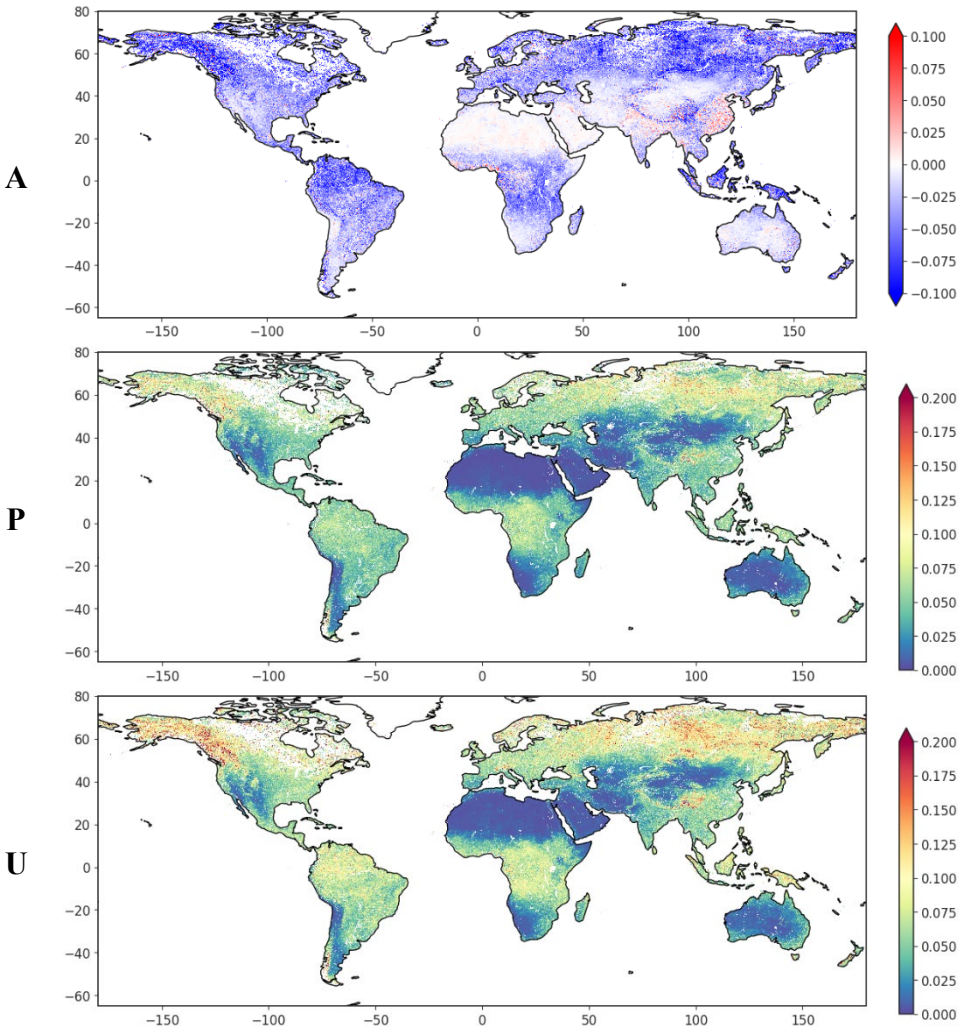


Figure 58: Spatial pattern of pixel wise APU between NDVI 300m V3 Sentinel-3/OLCI and MCD43A4 V6.1 for all pairwise valid NDVI values for the year 2019.

5.7.4 Conclusion

What is the spatial and statistical consistency of NDVI 300m V3 in comparison to MCD43A4?

There is a small negative bias between NDVI 300m V3 and MCD43A4 V6.1 (mean bias = -0.02, median bias = -0.012). Since both datasets are BRDF corrected, the differences are likely also related to differences in satellite, sensor, spectral response function, and processing techniques.

5.8 Q8: TEMPORAL CONSISTENCY OF NDVI 300M V3?

5.8.1 Introduction

The scientific questions to be answered are the following: What is the temporal consistency of NDVI 300m V3 in comparison to MCD43A4? How do the results based on NDVI 300m V3 compare to those of NDVI 300m V1, NDVI 300m V2, and NDVI 1km V3? What is the stability of the NDVI 300m V3 product?

The analysis of the temporal smoothness is performed over two one-year periods (2019 and 2023), while the other analyses use the full time series (January 2014 – December 2024). The temporal pixel profiles, temporal smoothness, and stability analyses based on inter-annual precision are performed on extractions of valid NDVI values for the LANDVAL V2 sites (see §4.3.4). The spatio-temporal evolution of the mean NDVI and of validation metrics of the comparison with an external reference is performed on a global systematic subsample (see §4.3.2) created with 1km resolution extractions.

The analyses include:

- Temporal pixel profiles over LANDVAL V2 sites
- Temporal smoothness
- Stability analysis
- Spatio-temporal evolution of validation metrics (Hovmöller plots) of the comparison with an external reference

5.8.2 Temporal pixel profiles

Figure 59 shows temporal pixel profiles for NDVI 300m V3 and all reference datasets for a selection of sites in different biomes. Temporal profiles for all LANDVAL V2 sites can be found in [Annex 1: Digital Annex]. The NDVI 300m V3 dataset shows similar temporal behaviour as the reference datasets. It is more stable than NDVI 300m V1, with less fluctuations in the patterns (see for example the EBF, DBF, and MXF profiles in Figure 59), and in some cases also more stable than NDVI 300m V2.

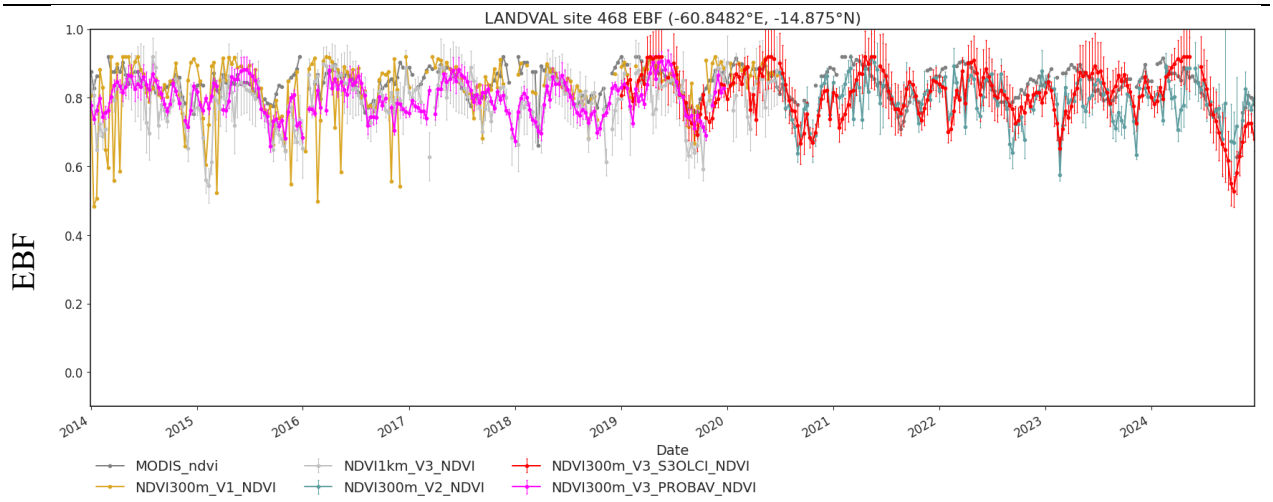
In some cases (e.g. site 376 and site 1356), the temporal completeness of NDVI 300m V3 is higher than that of the reference datasets. This is mostly the case when, for a certain site and dekad, there is no valid input TOC reflectance data, but NDVI is still produced due to the use of the MODIS prior climatology in the BRDF normalisation. These values should be treated with caution and can be filtered based on information contained in the QFLAG layer. In NDVI 300m V2, these data points are flagged with 'TOC-r out of range' because of the shorter accumulation window. The NDVI 300m

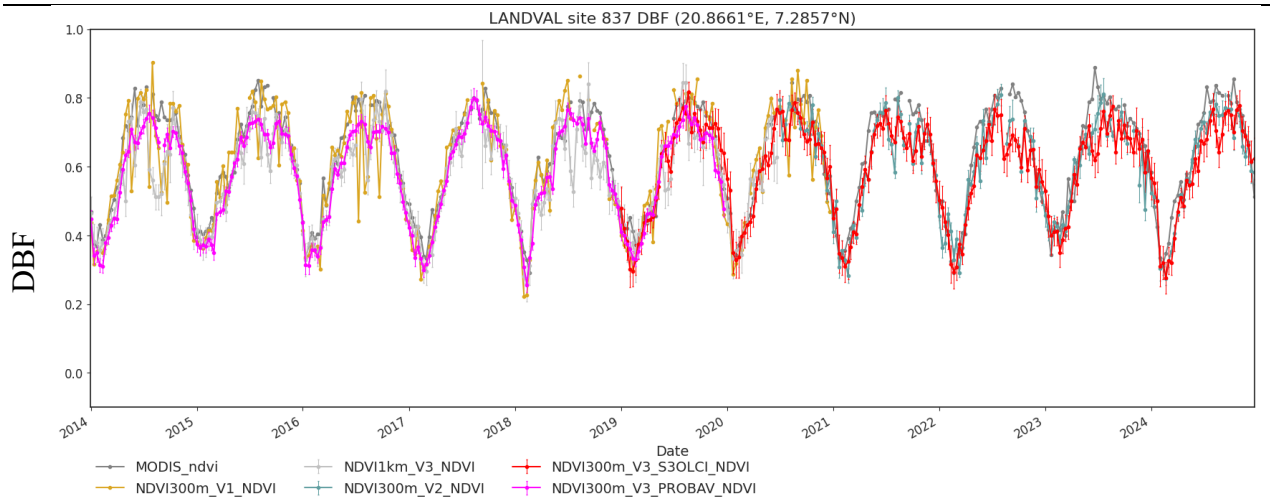
V3 completeness for these LANDVAL sites decreases in NRT processing (see Annex 2: Back-processing vs NRT processing).

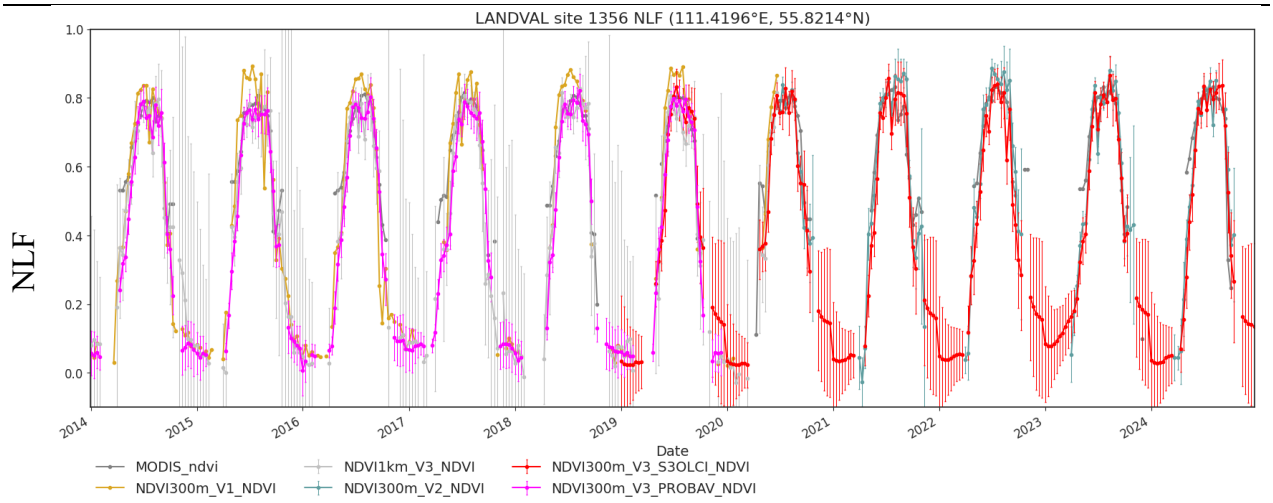
NDVI 300m V3 sometimes also shows lower completeness than the other datasets, and this also differs between PROBA-V and Sentinel-3/OLCI (see also §5.2). For missing periods in Northern hemisphere wintertime (e.g. site 211 and site 514), this is caused by flagging of snow pixels, which happens differently for the two sensors. Where Sentinel-3/OLCI often shows long gaps in winter, there are sometimes still data points present in PROBA-V. These points have been flagged as snow in Sentinel-3/OLCI, but are identified as cloud (over snow) in PROBA-V. Therefore, very low NDVI values are still outputted for the PROBA-V dataset. Sometimes gaps also occur during the summer period (e.g. site 1163 and site 1223). These gaps can be traced back to pixels for which QFLAG bit 7 is flagged, thus for which input data is missing and values are based on interpolated MODIS priors. In general, as already discussed above, the NDVI uncertainties are slightly larger for Sentinel-3/OLCI NDVI300m V3 compared to the PROBA-V NDVI300 V3.

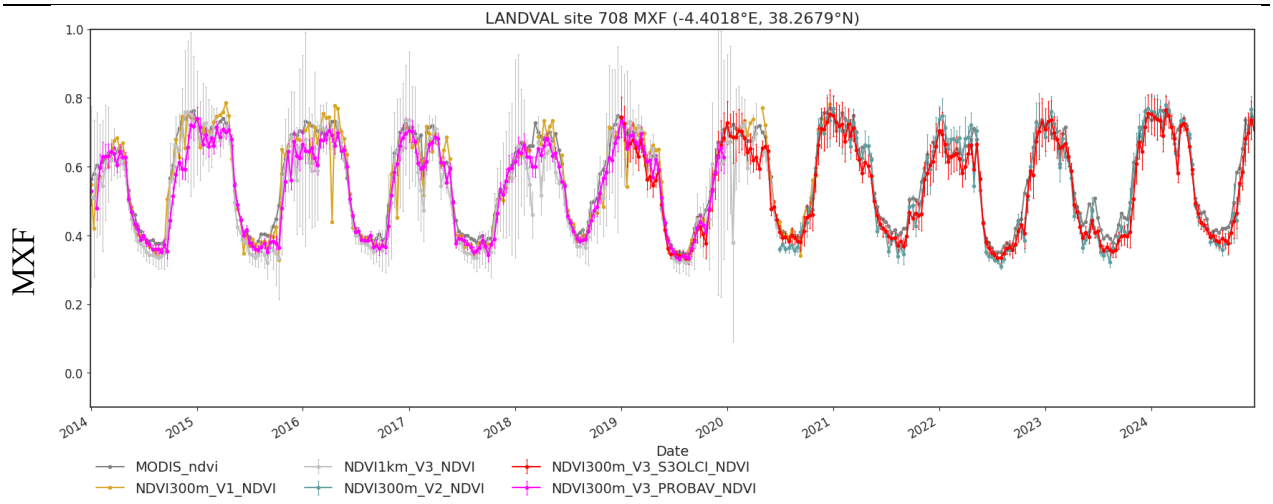
The use of the MODIS prior, when NOBS=0, in some cases also leads to an amplitude difference between Sentinel-3/OLCI and PROBA-V (e.g. site 837 and site 1276). Especially, when NOBS=0 for one of the sensors only.

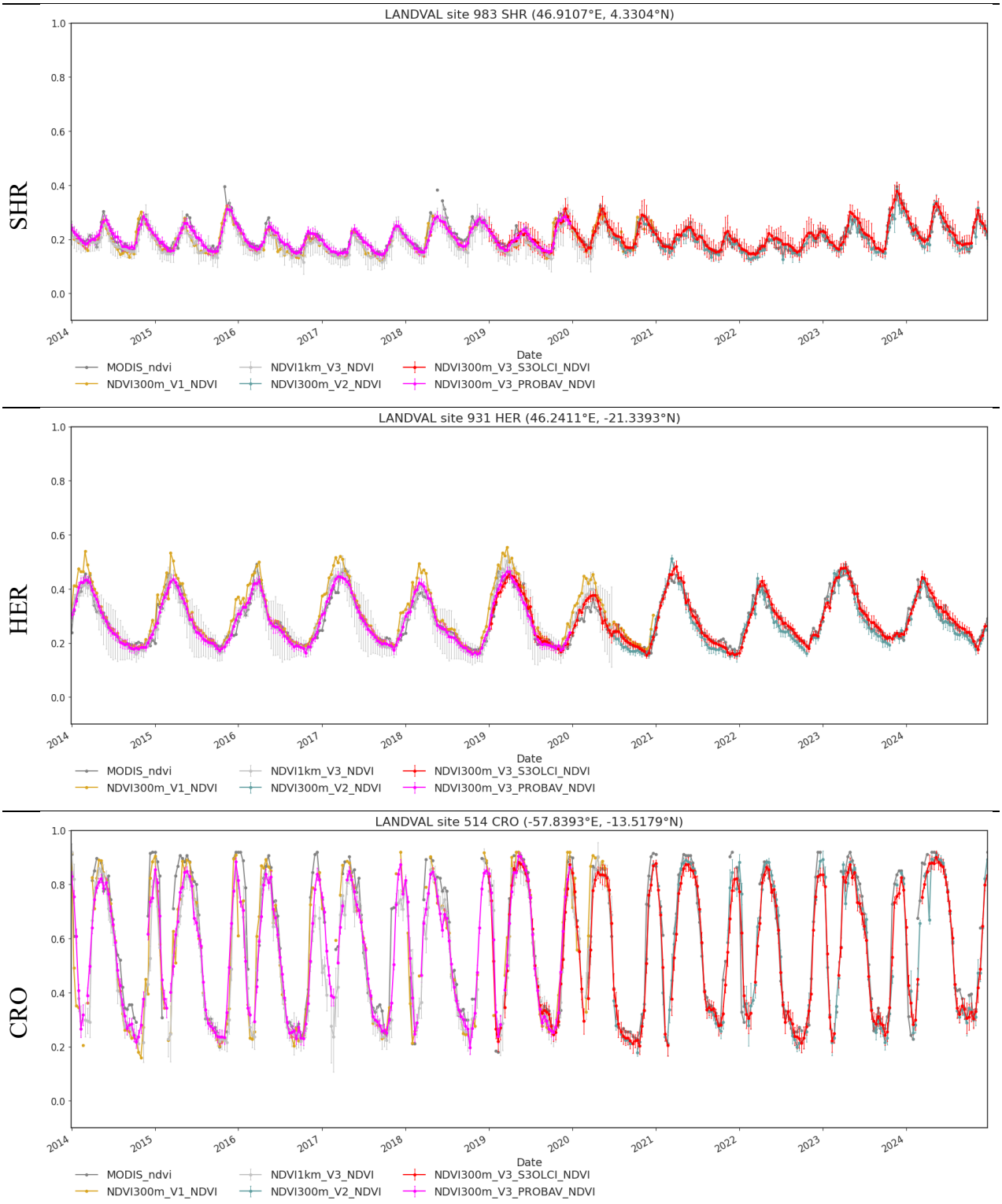
Overall, the temporal plots show good agreement between all sensors, especially when taking the uncertainties into account.




LANDVAL site 837 DBF (20.8661°E, 7.2857°N)
DBF
Date
MODIS_ndvi
NDVI300m_V1_NDVI
NDVI1km_V3_NDVI
NDVI300m_V2_NDVI
NDVI300m_V3_S3OLCI_NDVI
NDVI300m_V3_PROBAV_NDVI


LANDVAL site 1356 NLF (111.4196°E, 55.8214°N)
NLF
Date
MODIS_ndvi
NDVI300m_V1_NDVI
NDVI1km_V3_NDVI
NDVI300m_V2_NDVI
NDVI300m_V3_S3OLCI_NDVI
NDVI300m_V3_PROBAV_NDVI


LANDVAL site 708 MXF (-4.4018°E, 38.2679°N)
MXF
Date
MODIS_ndvi
NDVI300m_V1_NDVI
NDVI1km_V3_NDVI
NDVI300m_V2_NDVI
NDVI300m_V3_S3OLCI_NDVI
NDVI300m_V3_PROBAV_NDVI


LANDVAL site 983 SHR (46.9107°E, 4.3304°N)
SHR
LANDVAL site 931 HER (46.2411°E, -21.3393°N)
HER
LANDVAL site 514 CRO (-57.8393°E, -13.5179°N)
CRO
Date
MODIS_ndvi
NDVI300m_V1_NDVI
NDVI1km_V3_NDVI
NDVI300m_V2_NDVI
NDVI300m_V3_S3OLCI_NDVI
NDVI300m_V3_PROBAV_NDVI

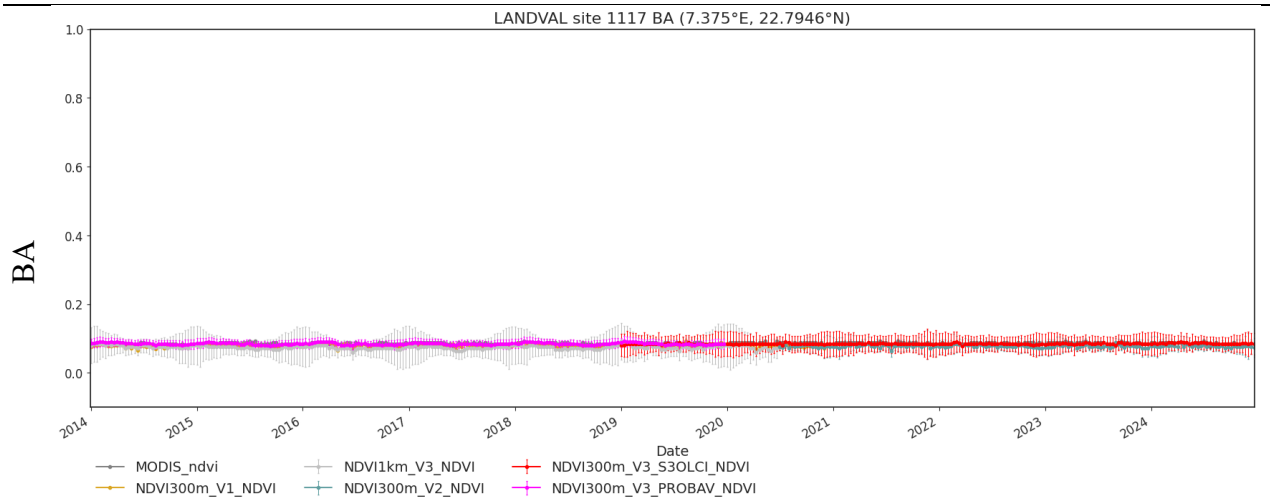
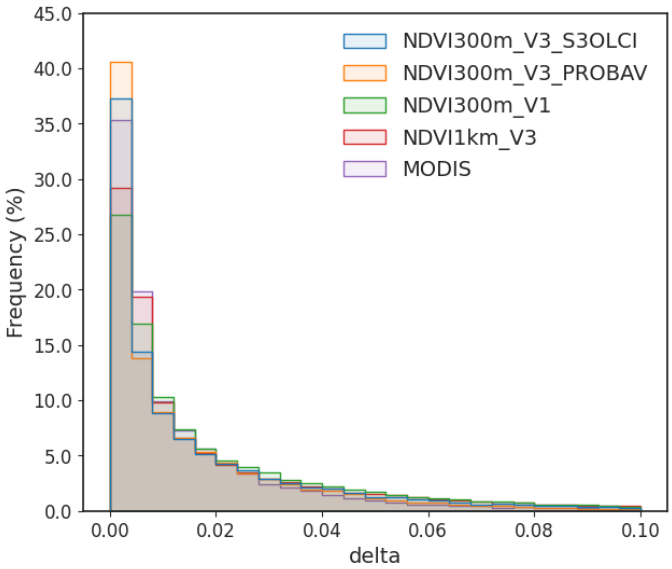


Figure 59: Temporal profiles over a selection of LANDVAL sites of NDVI 300m V3 Sentinel-3/OLCI (red), NDVI 300m V3 PROBA-V (pink), NDVI 300m V1 (yellow), NDVI 300m V2 (blue), NDVI 1km V3 (light grey), MODIS MCD43A4 V6.1 (grey). Temporal profiles for all LANDVAL V2 sites can be found in [Annex 1: Digital Annex].

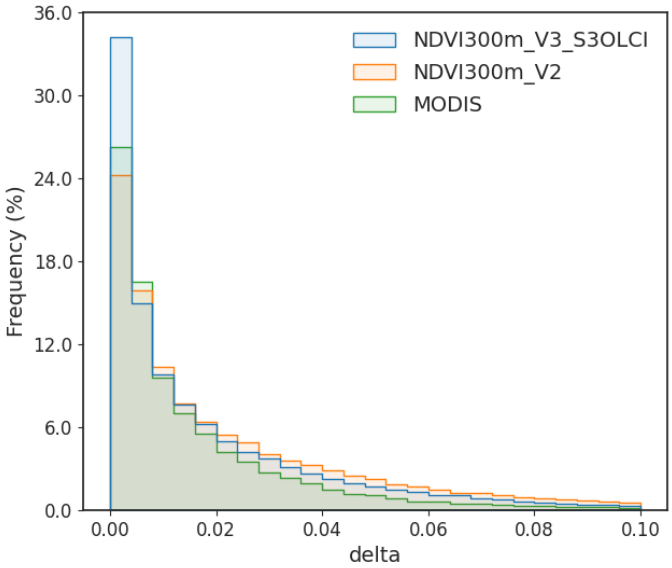
5.8.3 Temporal smoothness

The frequency histograms of the temporal smoothness (also called the intra-annual precision, or δ) of the various NDVI datasets are shown in Figure 60 for the year 2019 and in Figure 61 for the year 2023. Median values of δ and the time series relative noise are provided as an indication of the overall intra-annual precision of the products. The NDVI 300m V3 dataset is smoother than previous versions NDVI 300m V1 and NDVI 300m V2, i.e. the median smoothness and relative noise are lower for V3. The smoothness of MODIS MCD43A4 V6.1 is slightly higher that of NDVI 300m V3. This is due to the higher aggregation time (MODIS data are aggregated over 16 days, while the NDVI 300m V3 input data are 10-daily aggregations), and the fact that MODIS has two overpasses per day from the Terra and Aqua satellites.



	Median δ	Relative noise
NDVI 300m V3 Sentinel-3/OLCI	0.008	8.8
NDVI 300m V3 PROBA-V	0.007	7.4
NDVI 300m V1	0.01	10.9
NDVI 1km V3	0.008	11.1
MODIS	0.006	7.5

Figure 60: Frequency histograms of temporal smoothness (left), and median and time series relative noise per dataset (right) for the year 2019, for NDVI 300m V3 Sentinel-3/OLCI (blue), NDVI 300m V3 PROBA-V (orange), NDVI 300m V1 (green), NDVI 1km V3 (red), and MCD43A4 V6.1 (purple).



	Median δ	Relative noise
NDVI 300m V3 Sentinel-3/OLCI	0.010	8.4
NDVI 300m V2	0.014	10.9
MODIS	0.008	7.1

Figure 61: Frequency histograms of temporal smoothness (left), and median and time series relative noise per dataset (right) for the year 2023, for NDVI 300m V3 Sentinel-3/OLCI (blue), NDVI 300m V2 (orange), and MCD43A4 V6.1 (green).

5.8.4 Stability analysis

To evaluate the general stability of the NDVI 300m V3 product, a Hovmöller plot depicting the spatio-temporal evolution of the mean NDVI is shown in Figure 62. The figure shows the normal seasonal patterns, as expected, and – more importantly – the transition from PROBA-V to Sentinel-3/OLCI inputs (January 2019) is not discernible. This supports previous analyses that showed very good consistency between PROBA-V and Sentinel-3/OLCI based NDVI 300m V3 (see §5.4) and supports the general conclusion of the good temporal consistency of NDVI 300m V3.

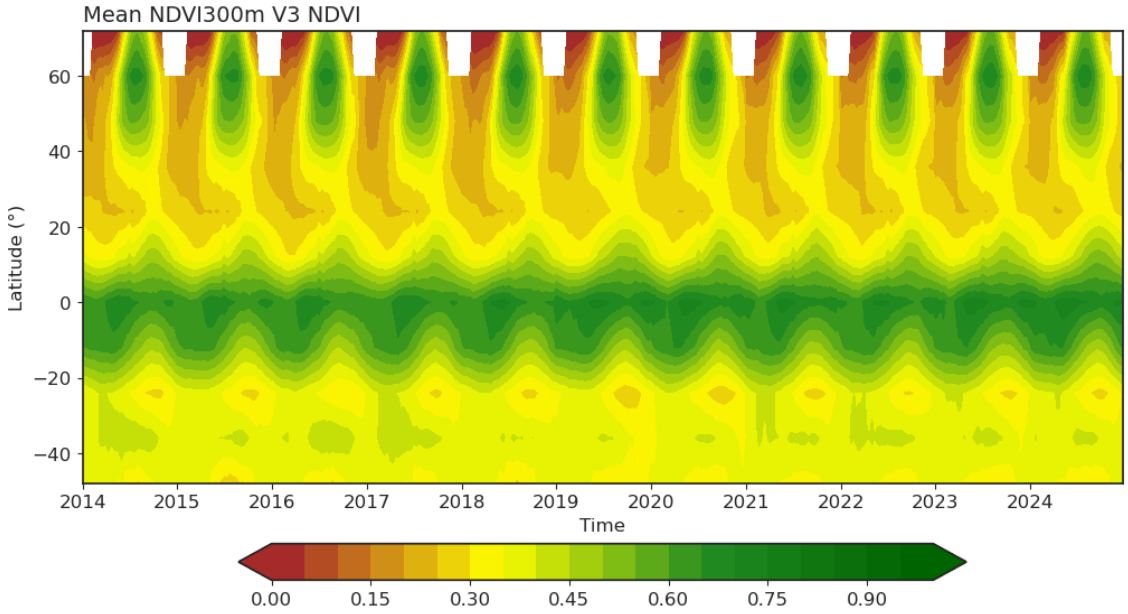


Figure 62: Spatio-temporal evolution of mean NDVI of the NDVI 300m V3 dataset for the period 2014 – 2024.

The inter-annual precision (i.e., dispersion of NDVI 300m V3 values from year to year) is assessed over all non-EBF and non-CRO LANDVAL V2 sites with completeness above 90%. CRO and EBF biome sites are excluded because these contain non-natural variability due to agricultural practices and are affected by cloud coverage, respectively. Sites with completeness below 90% are excluded because gaps also cause changes in the stability statistic, while this is not related to changes in the NDVI values themselves. For every quarter, the year up to the end of that quarter is compared to the year before that, using anomalies of the 5th and 95th percentile. Thus, the first precision value that can be computed for the NDVI 300m V3 dataset is the last quarter of 2015, using the year 2015 as current dataset and the year 2014 as reference dataset.

The stability of the NDVI 300m V3 product is determined as the slope of the linear regression of all quarterly precision values over time. It is expressed as percentage change per decade (10 years). This is illustrated in Figure 63. Overall, the stability of the product is very good, with a value of 0.003 (0.8%), well within the goal requirement of 1.5% (see §2.1).

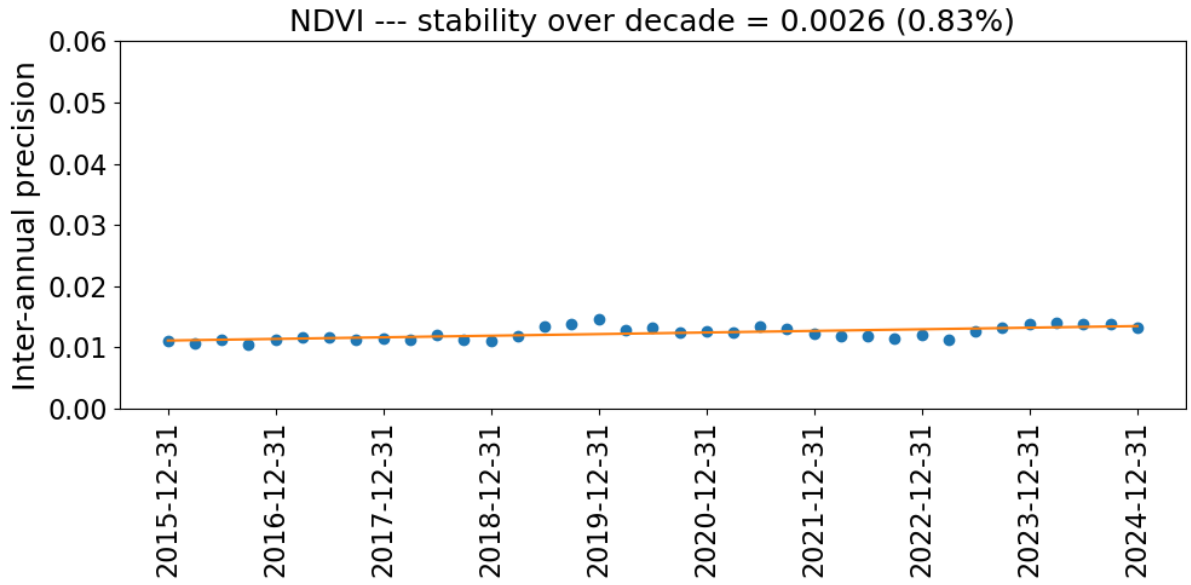


Figure 63: Temporal evolution of inter-annual precision of the NDVI 300m V3 product.

5.8.5 Spatio-temporal agreement with MCD43A4

The spatio-temporal variation of the Accuracy metric (MBE) and the Uncertainty metric (RMSD) is illustrated in the Hovmöller plots of Figure 64. This figure shows the comparison of resp. NDVI 300m V3, a combination of NDVI 300m V1 and NDVI 300m V2, and a combination of NDVI 1km V3 and NDVI 300m V3 with MCD43A4 V6.1 over the full time series. Both the MBE and RMSD show yearly patterns, related to seasonal variations in vegetation growth. In general, differences are largest around the equator (i.e. wet tropics, regions with persistent cloud cover) and at high latitudes (i.e. regions with bad illumination conditions and a limited number of good quality observations in winter).

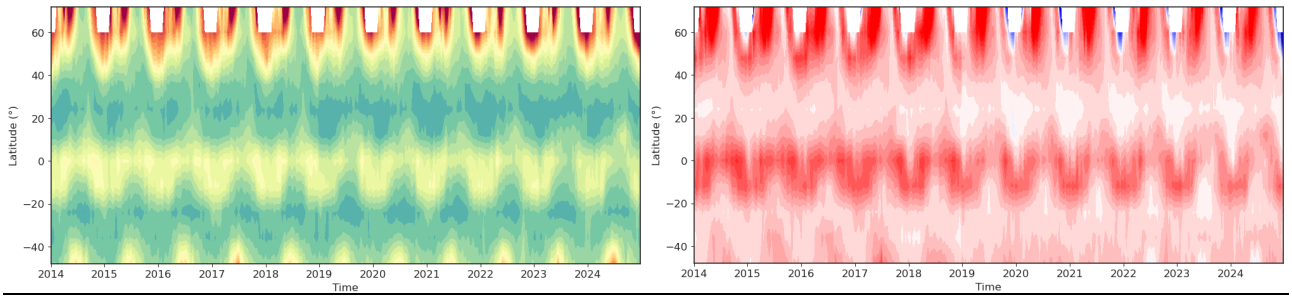
The switch between PROBA-V and Sentinel-3/OLCI for NDVI 300m V3 at the start of 2019 is visible in the upper Hovmöller plots, but the transition is much smoother than the switch between NDVI 300m V1 and NDVI 300m V2 (see Figure 40 of CGLOPS1_QAR_NDVI300m-V2 and middle plots of Figure 64). In NDVI 300m V1, another discontinuity is visible around December 2016, which is related to the switch from PROBA-V Collection 0 to Collection 1. Overall, the consistency over time

is much improved in NDVI 300m V3. The temporal consistency of the new dataset can also be determined from Figure 62, which shows the mean NDVI per latitude band over time. There is no mayor discontinuity visible and thus agreement between PROBA-V and Sentinel-3/OLCI is very high.

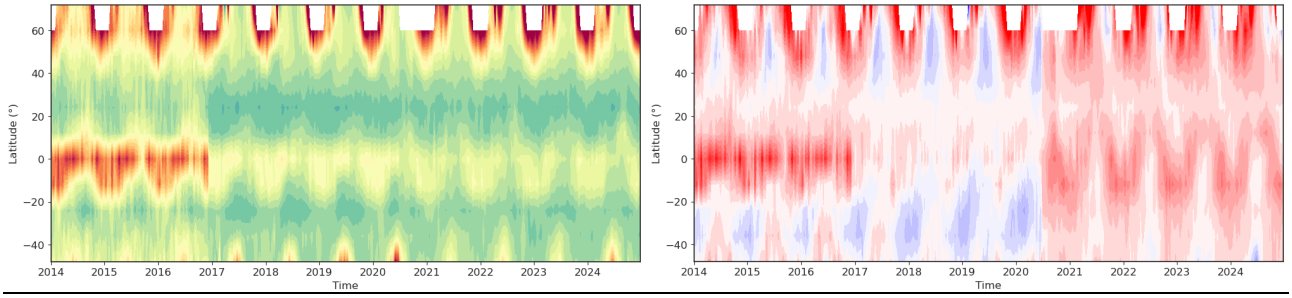
Sometimes a time series over a long period is needed and therefore the 1km (available since 1999) and 300m NDVI data can be combined. To investigate the temporal consistency over this combined time series, Hovmöller plots were created for the combination of NDVI 1km V3 and NDVI 300m V3 in comparison with MCD43A4 V6.1 over the same period (bottom plots in Figure 64). The switch between NDVI 1km V3 and NDVI 300m V3 is clearly visible, but most pronounced around the equator. This is consistent with the results analysed in Section 5.6.3. Despite the switch being visible, the consistency between the NDVI 1km V3 and NDVI 300m V3 datasets is very good (see §5.6). Therefore, the NDVI 1km V3 dataset can be used to extend the NDVI 300m V3 to the past, which is useful for certain applications that require a long time series.

RMSD	MBE

NDVI 300m V3 PROBA-V (2014 – 2018) and Sentinel-3/OLCI (2019 – 2024) vs MCD43A4 V6.1



NDVI 300m V1 (2014 – June 2020) and NDVI 300m V2 (July 2020 – 2024) vs MCD43A4 V6.1



NDVI 1km V3 (2014 – 2019) + NDVI 300m V3 (2019 – 2024) vs MCD43A4 V6.1

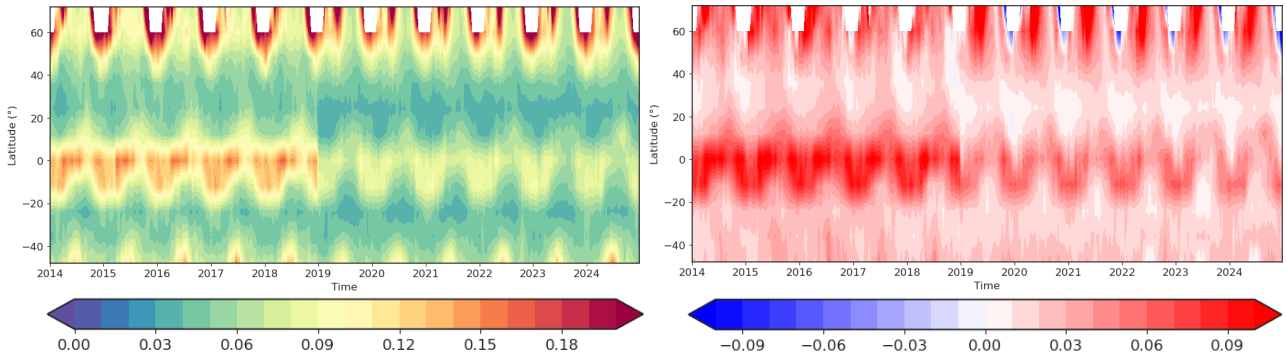


Figure 64: Hovmöller plots of the RMSD (left) and MBE (right) between NDVI 300m V3 and MCD43A4 V6.1 for the period 2014 – 2024 with switch between PROBA-V and Sentinel-3/OLCI at the start of 2019 (top), between NDVI 300m V1 (2014 – June 2020) + NDVI 300m V2 (July 2020 – 2024) and MCD43A4 V6.1 (middle), and between NDVI 1km V3 (2014 – 2019) + NDVI 300m V3 (2019 – 2024) and MCD43A4 V6.1 (bottom).

5.8.6 Conclusion

What is the temporal consistency of NDVI 300m V3 in comparison to MCD43A4?

The temporal profiles of NDVI 300m V3 over all LANDVAL V2 sites are in good agreement with the temporal profiles of the MCD43A4 V6.1 dataset. The smoothness of both datasets is comparable,

even though the smoothness of MODIS is slightly better. The transition between PROBA-V and Sentinel-3/OLCI for NDVI 300m V3 does not lead to discontinuities in the time series.

How do the results based on NDVI 300m V3 compare to those of NDVI 300m V1, NDVI 300m V2, and NDVI 1km V3?

The temporal behaviour of NDVI 300m V3 is in general similar to that of the other CGLOPS NDVI datasets. However, NDVI 300m V3 is smoother than the NDVI 300m V1 dataset, especially for EBF sites. For some sites, a higher completeness is achieved during Northern hemisphere wintertime, than for NDVI 300m V2. This is related to the longer accumulation window in the BRDF normalisation. In some cases, the completeness of the NDVI 300m V3 dataset is lower than for the previous versions. This is caused by flagging of snow pixels and pixels where no recent observations are available and the prior information is interpolated.

The Hovmöller plots show that the NDVI 300m V3 dataset has overall better temporal consistency than the combination of NDVI 300m V1 and NDVI 300m V2. They also indicate that the NDVI 1km V3 dataset is suitable to extend the NDVI 300m V3 dataset to the past.

What is the stability of the NDVI 300m V3 product?

The analyses show that in general the stability of the NDVI 300m V3 product is high, also during the switch from PROBA-V to Sentinel-3/OLCI. The product is very stable, with the stability criterion computed over the 2014 – 2024 period equal to 0.003 (0.8%), well below the goal value of 1.5%.

6 CONCLUSIONS, PERSPECTIVES, AND RECOMMENDATIONS

This quality assessment report for the NDVI 300m V3 dataset focussed on the questions listed in §3.2 to evaluate the quality of the product. Here we give a short summary for each question of the results that were obtained.

1. *What is the occurrence of quality flags (and NDVI uncertainties) in the NDVI 300m V3 product in terms of temporal and spatial distribution? Which flags are recommended for users and are applied in the subsequent analyses?*

   The NOBS=0 quality flag and high NDVI uncertainties occur mainly during the Northern hemisphere wintertime, at high latitudes, and in areas with persistent cloud cover (e.g. the wet tropics). There is no clear spatial or temporal pattern for the 'Warning' and 'Extreme warning' occurrences. There are more NOBS=0 and less 'Warning', 'Extreme warning', and snow flag occurrences for Sentinel-3/OLCI compared to PROBA-V, related to higher NDVI uncertainties in Sentinel-3/OLCI and differences in cloud/snow masking algorithm. There is no relation between NDVI and NDVI uncertainty, but there is a clear exponentially decreasing relation between NDVI uncertainty and NOBS.

   It is recommended to exclude NDVI values with 'Extreme warning' BRDF inversion quality flag, NDVI values for which MCD43P BRDF priors are based on linearly interpolated values, and NDVI values that are classified as snow.

2. *What are the temporal variation and spatial distribution of the product completeness of NDVI 300m V3 in comparison to MCD43A4 V6.1? How do the results based on NDVI 300m V3 compare to those of NDVI 300m V1, NDVI 300m V2, and NDVI 1km V3?*

   The product completeness of the NDVI 300m V3 back-processed dataset is very high, related to the long accumulation period and the use of a climatology in the BRDF inversion. In NRT processing, the spatial and temporal completeness are more comparable to that of the reference products (see Annex 2: Back-processing vs NRT processing).

3. *How do the NDVI 300m V3 Sentinel-3/OLCI and PROBA-V datasets compare visually to the NDVI 300m V1, NDVI 300m V2, NDVI 1km V3, and MODIS MCD43A4 V6.1 datasets?*

   NDVI 300m V3 visual quicklooks are similar to the reference products with respect to sharpness and spatial patterns. Differences in completeness between NDVI 300m V3 and

NDVI 300m V2 can be explained by the inclusion of a snow and interpolated prior quality flag in V3, and the difference in accumulation window.

4. *What is the statistical and spatial consistency between PROBA-V and Sentinel-3/OLCI for NDVI 300m V3?*

   There is good statistical and spatial consistency between PROBA-V and Sentinel-3/OLCI for NDVI 300m V3. There are some differences present which are largest for forest biome types, related to differences in cloud/snow detection and the impact of the MODIS prior when no valid input TOC reflectances are available. The NDVI uncertainties of Sentinel-3/OLCI are on average a factor two larger than those of PROBA-V.

5. *What is the difference between NDVI 300m V3, and NDVI 300m V1 and NDVI 300m V2 in terms of statistical and spatial consistency?*

   Overall, there is good consistency between NDVI 300m V3 and the previous versions. There is a small negative bias with NDVI 300m V1, strongly linked to vegetation densities and caused by anisotropy effects and differences in atmospheric correction. The consistency with NDVI 300m V2 is much better, and small differences are explained by differences in the ReBeLS BRDF algorithm.

6. *What is the statistical and spatial consistency with NDVI 1km V3?*

   Large statistical and spatial consistency is found between NDVI 300m V3 and NDVI 1km V3. Differences are mostly unsystematic because both versions are based on BRDF-normalised reflectances.

7. *What is the spatial and statistical consistency of NDVI 300m V3 in comparison to MCD43A4 V6.1?*

   There is a small negative bias between NDVI 300m V3 and MCD43A4 V6.1, most likely related to differences in sensor and processing techniques. The effects are largest for forest biomes and at high latitudes.

8. *What is the temporal consistency of NDVI 300m V3 in comparison to MCD43A4? How do the results based on NDVI 300m V3 compare to those of NDVI 300m V1, NDVI 300m V2, and NDVI 1km V3? What is the stability of the NDVI 300m V3 product?*

The temporal profiles of NDVI 300m V3 over LANDVAL V2 sites indicate similar temporal behaviour compared to the reference datasets. The temporal profiles are more stable than NDVI 300m V1 and show higher completeness in winter due to the larger accumulation window in the BRDF normalisation. The temporal smoothness is large. The Hovmöller plots show that the temporal consistency between PROBA-V and Sentinel-3/OLCI is improved in NDVI 300m V3 compared to the combination of NDVI 300m V1 and NDVI 300m V2. They also indicate that the NDVI 1km V3 dataset is suitable to extend the NDVI 300m V3 dataset to the past in order to have a longer time series. The stability of the NDVI 300m V3 product is within requirements.

It can be concluded that the NDVI 300m V3 product shows high product completeness, good stability, and large statistical, spatial, and temporal agreement with the reference datasets. Some differences are expected because of differences in sensor, processing approaches (especially BRDF normalisation), and flagging of bad data. Most discrepancies are minimal and unsystematic. Since the PROBA-V data from 2014 – 2018 and the Sentinel-3/OLCI data from 2019 – onwards are now both corrected for BRDF effects, NDVI differences between the two sensors are minimal. A homogeneous NDVI time series at 300m resolution from 2014 to present is now available.

7 REFERENCES

Ackerman, S., Frey, R., Strabala, K., Liu, Y., Gumley, L., Baum, B., Menzel, P., 2010. Discriminating clear-sky from cloud with MODIS - Algorithm Theoretical Basis Document (MOD35).

Breunig, F.M., Galvão, L.S., Formaggio, A.R., Epiphanio, J.C.N., 2011. Directional effects on NDVI and LAI retrievals from MODIS: A case study in Brazil with soybean. International Journal of Applied Earth Observation and Geoinformation 13, 34–42. https://doi.org/10.1016/j.jag.2010.06.004

Buchhorn, M., Smets, B., Bertels, L., Lesiv, M., Tsendbazar, N.-E., Herold, M., Fritz, S., 2019. Copernicus Global Land Service: Land Cover 100m: epoch 2015: Globe. https://doi.org/10.5281/zenodo.3939038

Cihlar, J., Chen, J., Li, Z., Huang, F., Latifovic, R., Dixon, R., 1998. Can interannual land surface signal be discerned in composite AVHRR data? J Geophys Res 103, 23163.

Cihlar, J., Latifovic, R., Chen, J., Trishchenko, A., Du, Y., Fedosejevs, G., Guindon, B., 2004. Systematic corrections of AVHRR image composites for temporal studies. Remote Sens Environ 89, 217–233. https://doi.org/10.1016/j.rse.2002.06.007

Claverie, M., Vermote, E.F., Weiss, M., Baret, F., Hagolle, O., Demarez, V., 2013. Validation of coarse spatial resolution LAI and FAPAR time series over cropland in southwest France. Remote Sens Environ 139, 216–230. https://doi.org/10.1016/j.rse.2013.07.027

Dawson, T.P., Curran, P.J., 1998. Technical note A new technique for interpolating the reflectance red edge position. Int J Remote Sens 19, 2133–2139. https://doi.org/10.1080/014311698214910

Doelling, D.R., Wu, A., Xiong, X., Scarino, B.R., Bhatt, R., Haney, C.O., Morstad, D., Gopalan, A., 2015. The radiometric stability and scaling of collection 6 terra- and aqua-MODIS VIS, NIR, and SWIR spectral bands. IEEE Transactions on Geoscience and Remote Sensing 53, 4520–4535. https://doi.org/10.1109/TGRS.2015.2400928

Duveiller, G., Fasbender, D., Meroni, M., 2016. Revisiting the concept of a symmetric index of agreement for continuous datasets. Sci Rep 6, 1–14. https://doi.org/10.1038/srep19401

Fougnie, B., Henry, P., Cabot, F., Meygret, A., Laubies, M.-C., 2000. VEGETATION : In-flight multi-angular calibration, in: Proceedings of SPIE. pp. 331–338.

Francois, M., Santandrea, S., Mellab, K., Vrancken, D., Versluys, J., 2014. The PROBA-V mission: the space segment. Int J Remote Sens 35, 2548–2564. https://doi.org/10.1080/01431161.2014.883098

Galvao, L.S., Ponzoni, F.J., Epiphanio, J.C.N., Rudorff, B.F.T., Formaggio, A.R., 2004. Sun and view angle effects on NDVI determination of land cover types in the Brazilian Amazon region with hyperspectral data. Int J Remote Sens 25, 1861–1879. https://doi.org/10.1080/01431160310001598908

Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A.M., Gu, W., Kim, G.K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J.E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S.D., Sienkiewicz, M., Zhao, B., 2017. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). J Clim 30, 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Gómez-Chova, L., Mateo-García, G., Muñoz-Marí, J., Camps-Valls, G., 2017. Cloud detection machine learning algorithms for PROBA-V, in: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, pp. 2251–2254. https://doi.org/10.1109/IGARSS.2017.8127437

Gonsamo, A., Chen, J.M., 2013. Spectral Response Function Comparability Among 21 Satellite Sensors for Vegetation Monitoring. IEEE Transactions on Geoscience and Remote Sensing 51, 1319–1335. https://doi.org/10.1109/TGRS.2012.2198828

Govaerts, Y., Sterckx, S., Adriaensen, S., 2013. Use of simulated reflectances over bright desert target as an absolute calibration reference. Remote Sensing Letters 4, 523–531. https://doi.org/10.1080/2150704X.2013.764026

Gutman, G., Csiszar, I., Romanov, P., 1999. Using NOAA / AVHRR Products to Monitor El Niño Impacts : Focus on Indonesia in 1997 – 98 1189–1205.

Henry, P., Meygret, A., 2000. Calibration of VEGETATION cameras on board SPOT 4. Proceedings of the VEGETATION 2000 conference Belgirate, Italy, 3–6 April 2000 1–9.

Ji, L., Gallo, K., 2006. An agreement coefficient for image comparison. Photogramm Eng Remote Sensing 72, 823–833. https://doi.org/10.14358/PERS.72.7.823

Lachérade, S., Fougnie, B., Henry, P., Gamet, P., Lacherade, S., Fougnie, B., Henry, P., Gamet, P., 2013. Cross calibration over desert sites: Description, methodology, and operational implementation. IEEE Transactions on Geoscience and Remote Sensing 51, 1098–1113. https://doi.org/10.1109/TGRS.2012.2227061

Martínez-Sánchez, E., Sánchez-Zapero, J., Camacho, F., 2024. LAND VALidation (LANDVAL) V2: Representative global sampling for satellite product intercomparison and calibration. Dataset.

Meyer, D., Verstraete, M., Pinty, B., 1995. The effect of surface anisotropy and viewing geometry on the estimation of NDVI from AVHRR. Remote Sensing Reviews 12, 3–27.

Meyer, D.J., 1996. Estimating the effective spatial resolution of an AVHRR time series. Int J Remote Sens 17, 2971–2980. https://doi.org/10.1080/01431169608949122

Mica, S., Galli, L., Duhoux, G., Livens, S., Jovanovic, V., Giustiniani, A., Dries, J.C., Zender, J., Santandrea, S., 2012. PROBA-V geometric calibration. International Geoscience and Remote Sensing Symposium (IGARSS) 1034–1037. https://doi.org/10.1109/IGARSS.2012.6351222

Rahman, H., Dedieu, G., 1994. SMAC: a simplified method for the atmospheric correction of satellite measurements in the solar spectrum. Int J Remote Sens 15, 123–143. https://doi.org/10.1080/01431169408954055

Schaaf, C.B., Gao, F., Strahler, A.H., Lucht, W., Li, X., Tsang, T., Strugnell, N.C., Zhang, X., Jin, Y., Muller, J.P., Lewis, P., Barnsley, M., Hobson, P., Disney, M., Roberts, G., Dunderdale, M., Doll, C., D'Entremont, R.P., Hu, B., Liang, S., Privette, J.L., Roy, D., 2002. First operational BRDF, albedo nadir reflectance products from MODIS. Remote Sens Environ 83, 135–148. https://doi.org/10.1016/S0034-4257(02)00091-3

Schaaf, C.B., Liu, J., Gao, F., Strahler, A.H., 2011. Aqua and terra MODIS albedo and reflectance anisotropy products. Remote Sensing and Digital Image Processing 11, 549–561. https://doi.org/10.1007/978-1-4419-6749-7_24

Sellers, P.J.J., 1985. Canopy reflectance, photosynthesis and transpiration. Int J Remote Sens 6, 1335–1372. https://doi.org/10.1080/01431168508948283

Sterckx, S., Adriaensen, S., 2019. PROBA-V radiometric calibration, PROBA-V QWG #10, 23-24 October 2019. https://earth.esa.int/web/sppa/activities/qwgs/proba-v-qwg-meetings/proba-v-qwg-meeting-10.

Sterckx, S., Adriaensen, S., Dierckx, W., Bouvet, M., 2016. In-Orbit Radiometric Calibration and Stability Monitoring of the PROBA-V Instrument. Remote Sens (Basel) 8, 546. https://doi.org/10.3390/rs8070546

Sterckx, S., Livens, S., Adriaensen, S., 2013. Rayleigh, deep convective clouds, and cross-sensor desert vicarious calibration validation for the PROBA-V mission. IEEE Transactions on Geoscience and Remote Sensing 51, 1437–1452. https://doi.org/10.1109/TGRS.2012.2236682

Swinnen, E., Dierckx, W., 2015. Validation Report - Normalized difference Vegetation Index - Products: LSA-410 and LSA-453 (ENDVI10). Issue 2.1.

Sylvander, S., Albert-Grousset, I., Henry, P., 2004. Geometrical performance of the VEGETATION products 00, 573–575. https://doi.org/10.1109/igarss.2003.1293846

Toté, C., Swinnen, E., Sterckx, S., Adriaensen, S., Benhadj, I., Iordache, M.-D., Bertels, L., Kirches, G., Stelzer, K., Dierckx, W., Van den Heuvel, L., Clarijs, D., Niro, F., 2018. Evaluation of PROBA-V Collection 1: Refined Radiometry, Geometry, and Cloud Screening. Remote Sens (Basel) 10, 1375. https://doi.org/10.3390/rs10091375

Toté, C., Swinnen, E., Sterckx, S., Benhadj, I., Dierckx, W., Gomez-Chova, L., Ramon, D., Stelzer, K., Van den Heuvel, L., Clarijs, D., Niro, F., 2021. The Reprocessed Proba-V Collection 2: Product Validation, in: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. IEEE, pp. 8084–8086. https://doi.org/10.1109/IGARSS47720.2021.9553376

Toté, C., Swinnen, E., Sterckx, S., Clarijs, D., Quang, C., Maes, R., 2017. Evaluation of the SPOT/VEGETATION Collection 3 reprocessed dataset: Surface reflectances and NDVI. Remote Sens Environ 201, 219–233. https://doi.org/10.1016/j.rse.2017.09.010

Toté, C., Swinnen, E., Van Den Heuvel, L., Clarijs, D., 2023. Evaluation of PROBA-V C2 products Final Report.

Trishchenko, A.P., 2009. Effects of spectral response function on surface reflectance and NDVI measured with moderate resolution satellite sensors: Extension to AVHRR NOAA-17, 18 and METOP-A. Remote Sens Environ 113, 335–341. https://doi.org/10.1016/j.rse.2008.10.002

Trishchenko, A.P., Cihlar, J., Li, Z., 2002. Effects of spectral response function on surface reflectance and NDVI measured with moderate resolution satellite sensors. Remote Sens Environ 81, 1–18. https://doi.org/10.1016/S0034-4257(01)00328-5

Tucker, C., Pinzon, J., Brown, M., Slayback, D., Pak, E., Mahoney, R., Vermote, E., El Saleous, N., 2005. An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. Int J Remote Sens 26, 4485–4498. https://doi.org/10.1080/01431160500168686

van Leeuwen, W.J.D., Orr, B.J., Marsh, S.E., Herrmann, S.M., 2006. Multi-sensor NDVI data continuity: Uncertainties and implications for vegetation monitoring applications. Remote Sens Environ 100, 67–81. https://doi.org/10.1016/j.rse.2005.10.002

Verhoef, W., 1984. Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. Remote Sens Environ 16, 125–141. https://doi.org/10.1016/0034-4257(84)90057-9

Vermote, E., Justice, C.O., Bréon, F.M., 2009. Towards a generalized approach for correction of the BRDF effect in MODIS directional reflectances. IEEE Transactions on Geoscience and Remote Sensing 47, 898–908. https://doi.org/10.1109/TGRS.2008.2005977

Vermote, E.F., Kotchenova, S., 2008. Atmospheric correction for the monitoring of land surfaces. Journal of Geophysical Research Atmospheres 113. https://doi.org/10.1029/2007JD009662

Vermote, E.F., Saleous, E., Z, N., Justice, C.O., 2002. Atmospheric correction of MODIS data in the visible to middle infrared: first results. Remote Sensing of. Environment 83, 97–111.

Weiss, M., Baret, F., Garrigues, S., Lacaze, R., 2007. LAI and fAPAR CYCLOPES global products derived from VEGETATION. Part 2: validation and comparison with MODIS collection 4 products. Remote Sens Environ 110, 317–331. https://doi.org/10.1016/j.rse.2007.03.001

Wevers, J., Müller, D., Danne, O., Brockmann, C., Paperin, M., Stelzer, K., Preusker, R., 2019. Sentinel-3 OLCI Level 2 cloud flagging status and potential improvements, in: 5th S3 Validation Team Meeting. https://nikal.eventsair.com/QuickEventWebsitePortal/sentinel-3-validation-team-s3vt/esa/ExtraContent/ContentSubPage?page=10&subPage=1.

Wolfe, R.E., Nishihama, M., Fleig, A.J., Kuyper, J.A., Roy, D.P., Storey, J.C., Patt, F.S., 2002. Achieving sub-pixel geolocation accuracy in support of MODIS land science. Remote Sens Environ 83, 31–49. https://doi.org/10.1016/S0034-4257(02)00085-8

Wolters, E., Dierckx, W., Iordache, M.-D., Swinnen, E., 2018. PROBA-V Products User Manual v3.0. VITO.

Wolters, E., Swinnen, E., Toté, C., Sterckx, S., 2016. SPOT-VGT Collection 3 Products User Manual. VITO.

Xiong, X., Barnes, W., 2006. An overview of MODIS radiometric calibration and characterization. Adv Atmos Sci 23, 69–79. https://doi.org/10.1007/s00376-006-0008-3

ANNEX 1: DIGITAL ANNEX

The Digital Annex can be downloaded from the CLMS website: https://land.copernicus.eu/en/technical-library

It contains the following folders:

- QUICKLOOKS:
  Maps over Northern Europe (NEUR), Western Europe (WEUR), South America (SAMI), and Central Africa (CAFR) 10° x 10° tiles, and maps over Brussels (BRU) and Sudan (SUD) 1° x 1° tiles.
  Subfolders per tile (as specified in §4.3.3).
  Filenames for NDVI: "<DATASET>_<YYYYMMDD>_<TILE>_NDVI.png"
  Filenames for quality flags: "<DATASET>_<YYYYMMDD>_<TILE>_QFLAG.png"

- TEMPORAL_PROFILES:
  Temporal profiles for the LANDVAL sites.
  Subfolders per biome (as specified in §4.3.4).
  Filenames: "site<SITE_ID>.png"

ANNEX 2: BACK-PROCESSING VS NRT PROCESSING (SCIENTIFIC QUALITY MONITORING 2025)

OBJECTIVE

Starting from 2025, the NDVI 300m V3 products are generated in a different processing configuration compared to the dataset under evaluation in the main document. The NDVI300m V3 products from 2014 to 2024 are produced in back-processing mode, with an accumulation period of 14 months for the BRDF correction, whereas the data from 2025 onwards are produced in near-real time (NRT) mode with an accumulation window of only 30 days. The accumulation window defines the period of TOC reflectances and uncertainties that are used as input to the BRDF correction algorithm (see CGLOPS1_ATBD_BRDFCorrection300m-V1). Because of this change in accumulation window, a difference is expected between back-processed and NRT processed NDVI data, with the back-processed data yielding more precise results.

This annex summarises the investigation of the differences between both processing modes. It focusses on the differences between back-processed data and NRT data for the year 2025. It also serves as the scientific quality monitoring of the year 2025. The analyses are done for the period of January – December 2025 with a short-term statistic (STS) as reference dataset. The STS is created for all 36 dekads in a year and is computed as the median per-pixel value of the NDVI 300m V3 product over the years 2020 to 2024. The median is chosen to minimise the influence of extreme outliers on the STS. The period 2020 – 2024 is chosen to have a reference dataset of the most recent five years, representing the average state of NDVI at present, that has been fully validated. We did not include all 11 years (2014 - 2024) that were validated in the main QAR, because NDVI is expected to change over time due to changes in climate and land cover/land use.

The goal of the scientific quality monitoring is to verify the quality of the product and its stability over time. This is achieved by repeating some of the analyses in the main document. The results are compared with those of the overall quality assessment in the main document.

METHODOLOGY

To define the differences between back-processing and NRT processing mode, various procedures applied in the main QAR are repeated for the NDVI 300m V3 products covering the period January − December 2025. The same sampling strategy is used as in the main QAR (global subsamples and LANDVAL V2 extractions), but instead of comparing to other datasets, the products of 2025 are compared to the STS based on NDVI 300m V3 products covering the period 2020 – 2024. The January − December 2025 period is also called the NRT period or the ‘ACT’ period.

RESULTS

Quality flags

The temporal distribution of quality flag occurrences in NDVI 300m V3 during January – December 2025 is shown in Figure 65. The five plotted quality flags (NOBS=0, warning, extreme warning, snow observations, and interpolated prior) show the same temporal behaviour as in the right panel of Figure 7 in the main QAR (for Sentinel-3/OLCI based NDVI data), but the number of occurrences is lower for 2025. This is due to the difference in accumulation window between back-processing in the main QAR and NRT processing for the 2025 data. For NOBS=0, the difference is largest during the Northern hemisphere winter period. NRT processing in the winter suffers more from missing observations due to bad illumination conditions, cloud cover, and snow. Since the accumulation window in the BRDF inversion is smaller than for back-processing, there is a higher chance of missing TOC-r reflectances and thus NDVI values are more likely to be flagged as missing. Since Figure 65 only shows flag occurrences for valid NDVI pixels, the number of NOBS=0 at high latitudes is lower in NRT mode. The lower number of occurrences of warning and extreme warning pixels is related to the outlier screening. The outlier screening is less reliable in NRT mode because the accumulation window is smaller, and the resulting uncertainties are larger, leading to a smaller number of pixels classified with a warning or extreme warning. The temporal evolution of the occurrences of interpolated prior quality flag shown in Figure 65 is slightly different than in the right panel of Figure 7. This is also because only occurrences over valid NDVI pixels are considered and these valid pixels differ between the two datasets.

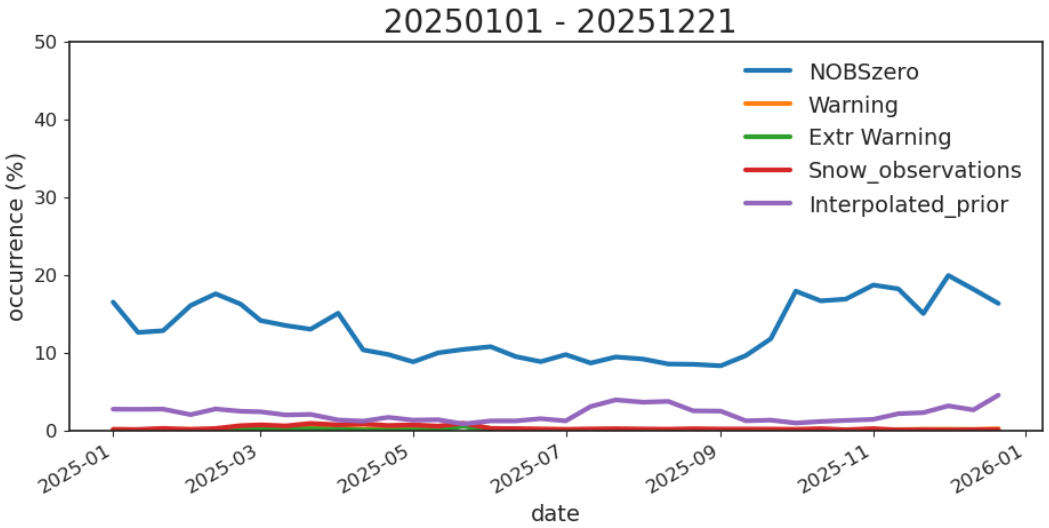


Figure 65: Temporal distribution of flag occurrences over land pixels with valid NDVI. NOBS=0 (blue), warning (orange), extreme warning (green), snow observations (red), and interpolated prior (purple) are shown for the period January – December 2025.

The spatial distributions of the five quality flags of NRT 2025 products are shown in Figure 66. The spatial patterns are very similar to those for the back-processing method in the main QAR (see Figure 8) but Figure 66 also confirms the lower number of pixels with NOBS=0, warning, or extreme warning quality flag in NRT. There are less datapoints with NOBS=0 at high latitudes, while the number is similar in tropical areas with permanent cloud cover. This can also be attributed to invalid NDVI pixels which are not included in the quality flag analysis.

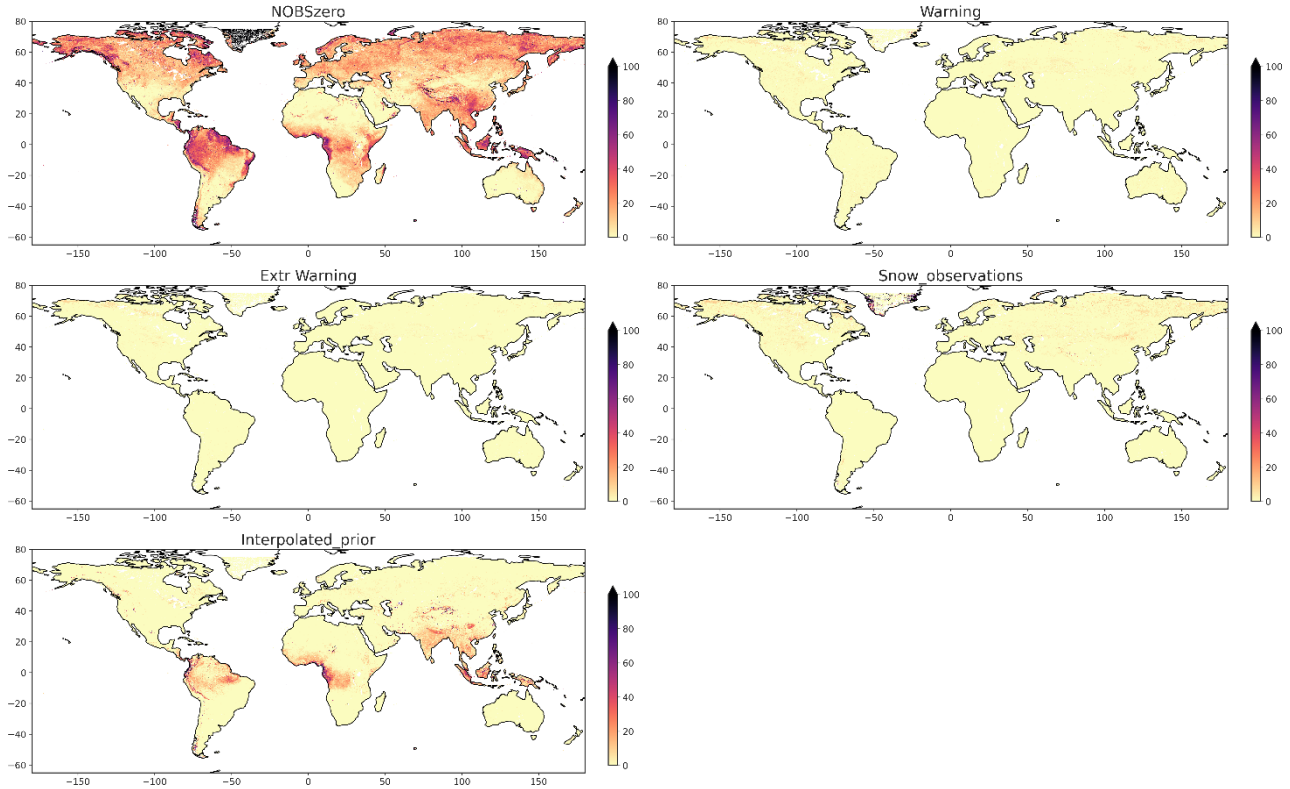


Figure 66: Spatial distribution of flag occurrences over land pixels with valid NDVI (January – December 2025). From upper left to bottom right NOBS=0, warning, extreme warning, snow observations, and interpolated prior are shown.

Uncertainties

The spatial pattern of mean NDVI 300m V3 and its associated uncertainty for the period January – December 2025 (Figure 67) are similar to the spatial patterns in the main QAR (Figure 11). The main difference is that the magnitude of the uncertainties in NRT processing is around twice as large (notice that the range of the colour bar is twice as large as in Figure 11), since the accumulation window is smaller and therefore less input observations are used in the BRDF inversion**.** The uncertainties are highest at high latitudes and in the Tropics, due to persistent cloud cover and/or poor illumination

conditions in winter. Lowering the accumulation period from 14 months to 30 days can result in four times higher NDVI uncertainties.

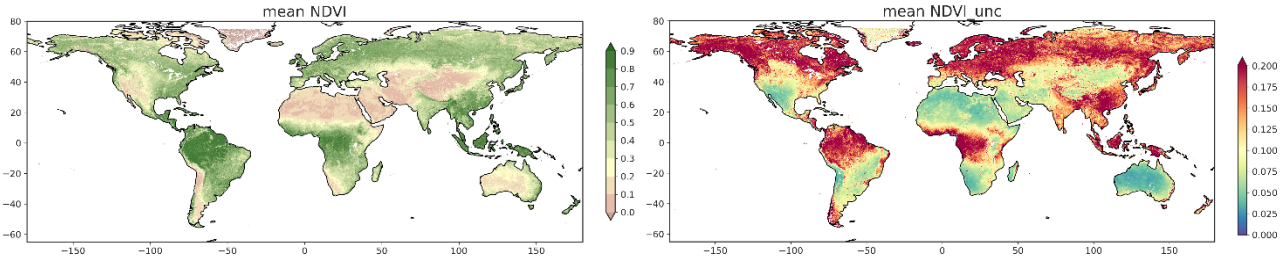


Figure 67: Spatial pattern of mean NDVI 300m V3 (left) and mean NDVI uncertainty (right) for the period January – December 2025.

From Figure 68 it is also clear that uncertainties are a factor ~2 higher than in the back-processing mode described in the main QAR (Figure 12, top-right). There is no relation between NDVI and NDVI uncertainty visible.

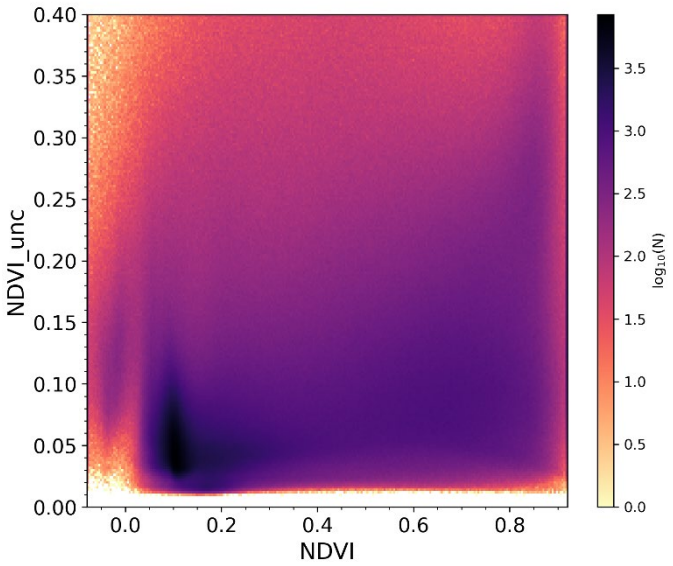


Figure 68: Scatter density plot showing the relation between NDVI and NDVI uncertainty for NDVI 300m V3 January – December 2025.

Product completeness

Figure 69 shows the spatial distribution of invalid observations during January – December 2025 (left) and the temporal evolution of the percentage of valid observations over land during this period and for the STS (right). The percentage of valid observations, and thus product completeness, is lower for the current period than for the STS, especially for the Northern hemisphere winter period. This is

related to the shorter accumulation period in NRT processing in the BRDF correction. A shorter accumulation period means less valid pixels, especially at high latitudes in winter, and thus to lower completeness. The spatial distribution of completeness is similar to that of the reference datasets in the main QAR (see Figure 13 and Figure 14).

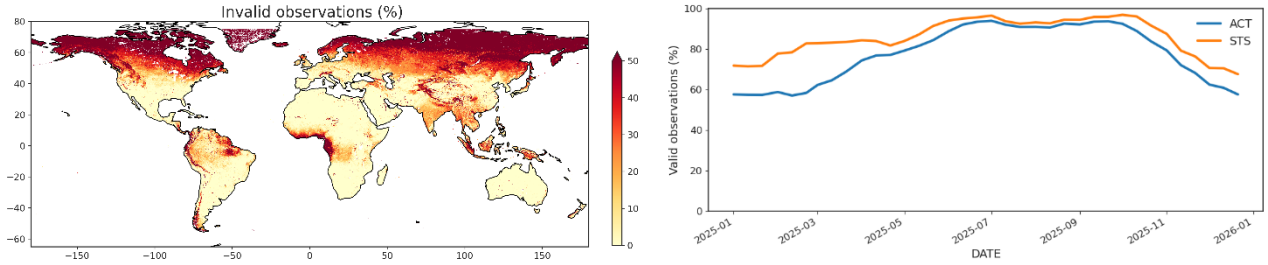
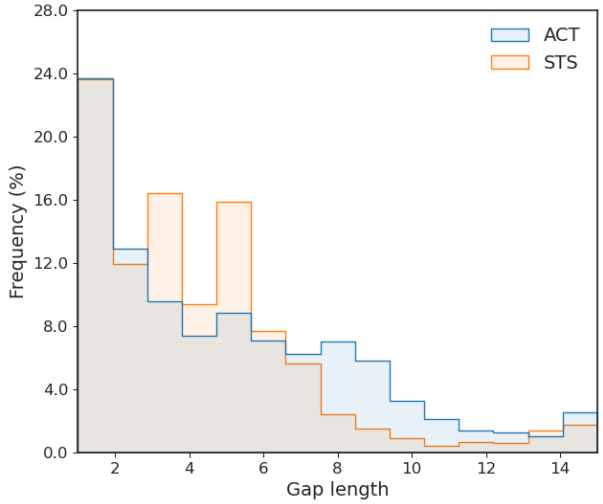


Figure 69: Spatial distribution of invalid observations (left) and temporal evolution of the percentage of valid observations over land (right) of NDVI 300m V3 for the period January – December 2025 (ACT, blue) and the STS (orange).

A similar conclusion can be drawn from Figure 70, where the frequency distribution of gap lengths is shown (i.e. the number of consecutive dekads with missing data points). The January – December 2025 dataset contains a higher frequency of long gaps compared to the STS and to the back-processed dataset (see Figure 16 and Figure 17). The STS has two clear peaks at a gap length of three and of five dekads. As explained in §5.2.4, these gaps are caused by high latitude observations for which there are no observations (NOBS=0) and for which the MODIS BRDF priors are interpolated.



Dataset	Mean gap length (dekads)
2025 (ACT)	5.7
STS	3.6

Figure 70: Histograms of gap length over all 2000 LANDVAL V2 sites and mean gap length for the NDVI 300m V3 January – December 2025 dataset ('ACT', blue) and the STS (orange).

Statistical consistency

The frequency histogram of the overall bias and the GM regression over all LANDVAL V2 sites between the NDVI 300m V3 January – December 2025 dataset and the STS are shown in Figure 71. The bias histogram is centred close to zero with mean bias = 0.003 and median bias = 0.003. It is a little bit skewed towards positive values, indicating that the NDVI data for 2025 is in general slightly higher than the STS. The same conclusion can be drawn from the scatterplot: although the regression line is very close to the 1:1 line and R^2 is large, the slope is slightly below one (slope = 0.987) which indicates that the 'ACT' values for 2025 are a bit higher than the STS values. These figures show very good agreement between the two datasets; some differences are however expected since different years are compared and natural variation in vegetation growth is expected. Overall, the differences between NDVI 300m V3 January – December 2025 and the STS are much lower than the differences between the NDVI 300m V3 dataset and the reference datasets in the main QAR (see §5.5, §5.6, and §5.7).

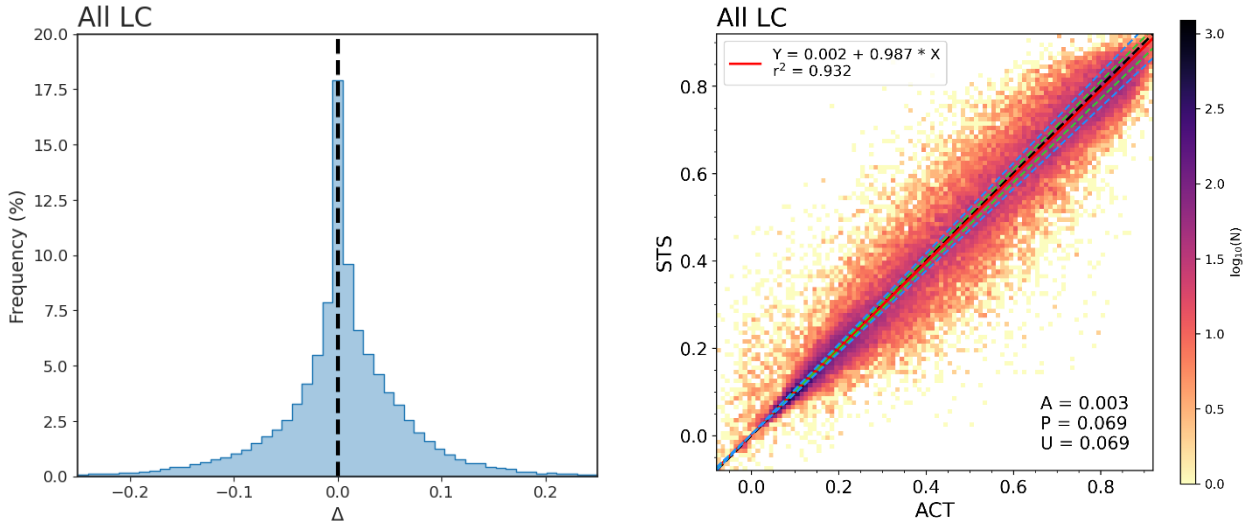


Figure 71: Frequency histogram of the overall bias (left) and scatter density plot (right) between the NDVI 300m V3 January – December 2025 and the STS for all pairwise samples.

Table 14 contains the statistics related to the GM regression over all land cover types. Values for all the statistics are close to zero. The last column are the percentages of LANDVAL V2 sites with uncertainty within the goal and threshold requirements. The percentages are comparable to those in the main QAR (see Table 10 and Table 11).

Table 14: Statistics of comparison between NDVI 300m V3 January – December 2025 and the STS over all LANDVAL V2 sites. The last column contains the percentage of pixels with uncertainty within the goal (2.5%) and threshold (5%) range.

N	R	Bias (%)	MD (%)	STD (%)	MAD (%)	RMSD (%)	Offset / slope	% Goal / Threshold
55496	0.97	0.003 (0.8%)	0.003 (0.6%)	0.07 (17%)	0.03 (7%)	0.07 (17%)	0.002 / 0.99	20 / 38

Figure 72 presents the Accuracy (A), Precision (P), and Uncertainty (U) metrics for all LANDVAL V2 sites combined and per biome. The mean bias (or accuracy statistic) is very low for all biome types, indicating good agreement between the 2025 dataset and the STS. The uncertainty is largest for the CRO, NLF, and MXF biomes, probably linked to crop rotation and variations in vegetation growth between years. The best agreement is found for the BA biome. The mean bias is slightly lower, but the precision and uncertainty metric are up to twice as large compared to the APU statistics for PROBA-V and Sentinel-3/OLCI in Section §5.4..

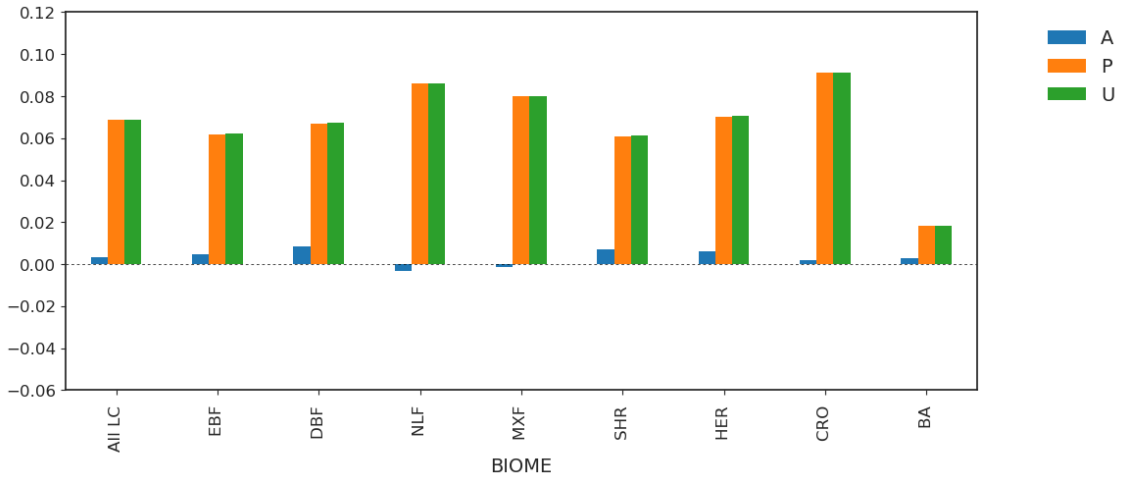


Figure 72: APU validation metrics between NDVI 300m V3 January- December 2025 and the STS for all samples (left) and per biome type (see Table 7).

Spatial consistency

The spatial consistency between NDVI 300m V3 January – December 2025 and the STS is analysed with histograms over different biomes (Figure 73). The histograms show good agreement between the two datasets. The differences are lowest for Bare Areas (BA), and are largest at higher latitudes and for forest biomes with dense vegetation (EBF, DBF, NLF, and MXF) for which the 2025 dataset has higher values than the STS.

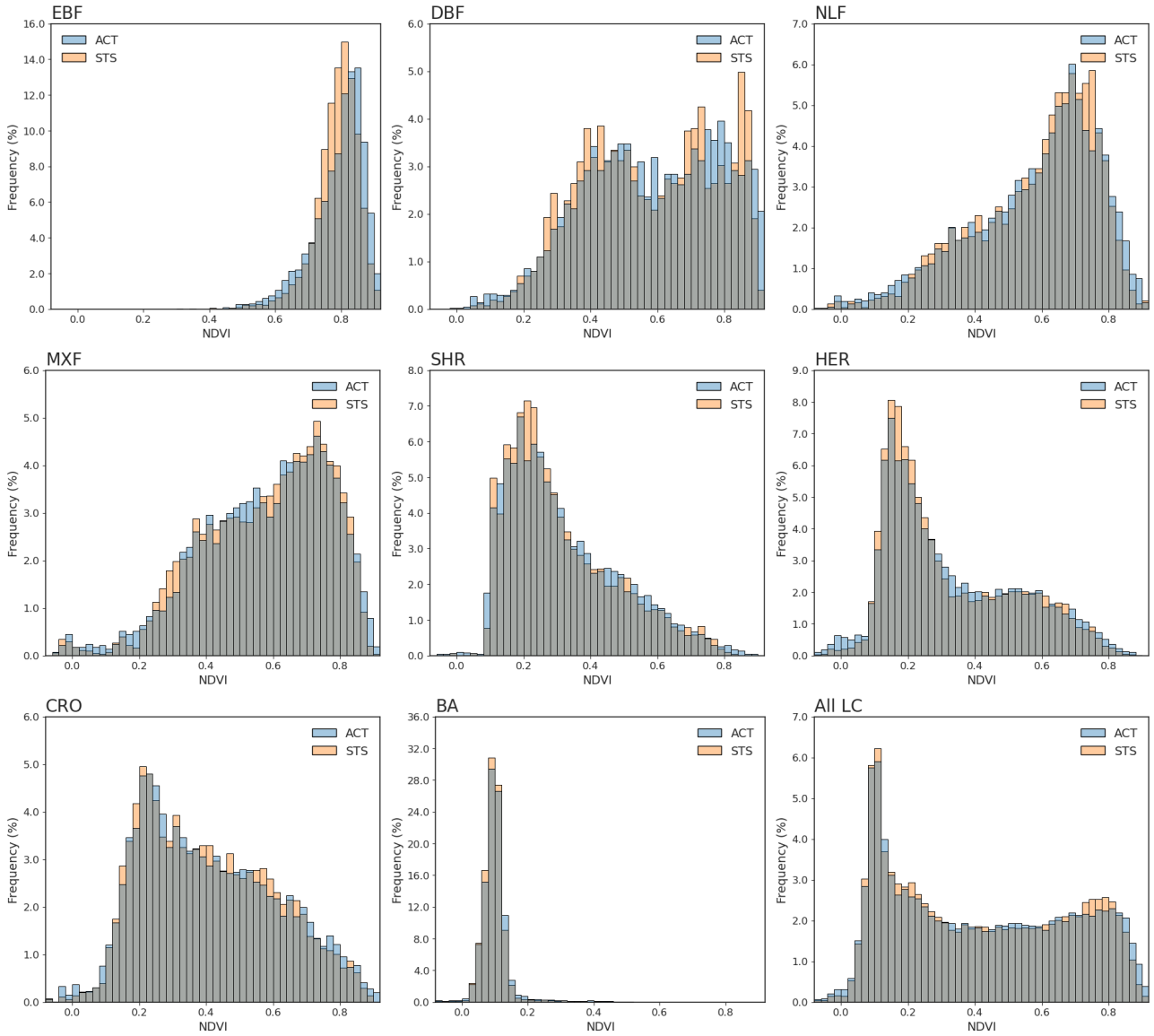


Figure 73: Frequency histograms per biome and for all land cover types comparing the January – December 2025 period ('ACT', blue) with the STS (orange). Grey colour indicates where both histograms overlap.

Figure 74 shows the spatial patterns of pixel wise APU validation metrics between NDVI 300m V3 January – December 2025 and the STS. The validation metrics are closest to zero for deserts and other bare areas (e.g. the Sahara and the Arabian Peninsula) and higher further from the equator. The regions popping out in red in the bias histogram (i.e. South America, South Africa, and Central Asia) are known to have experienced severe droughts during 2025.

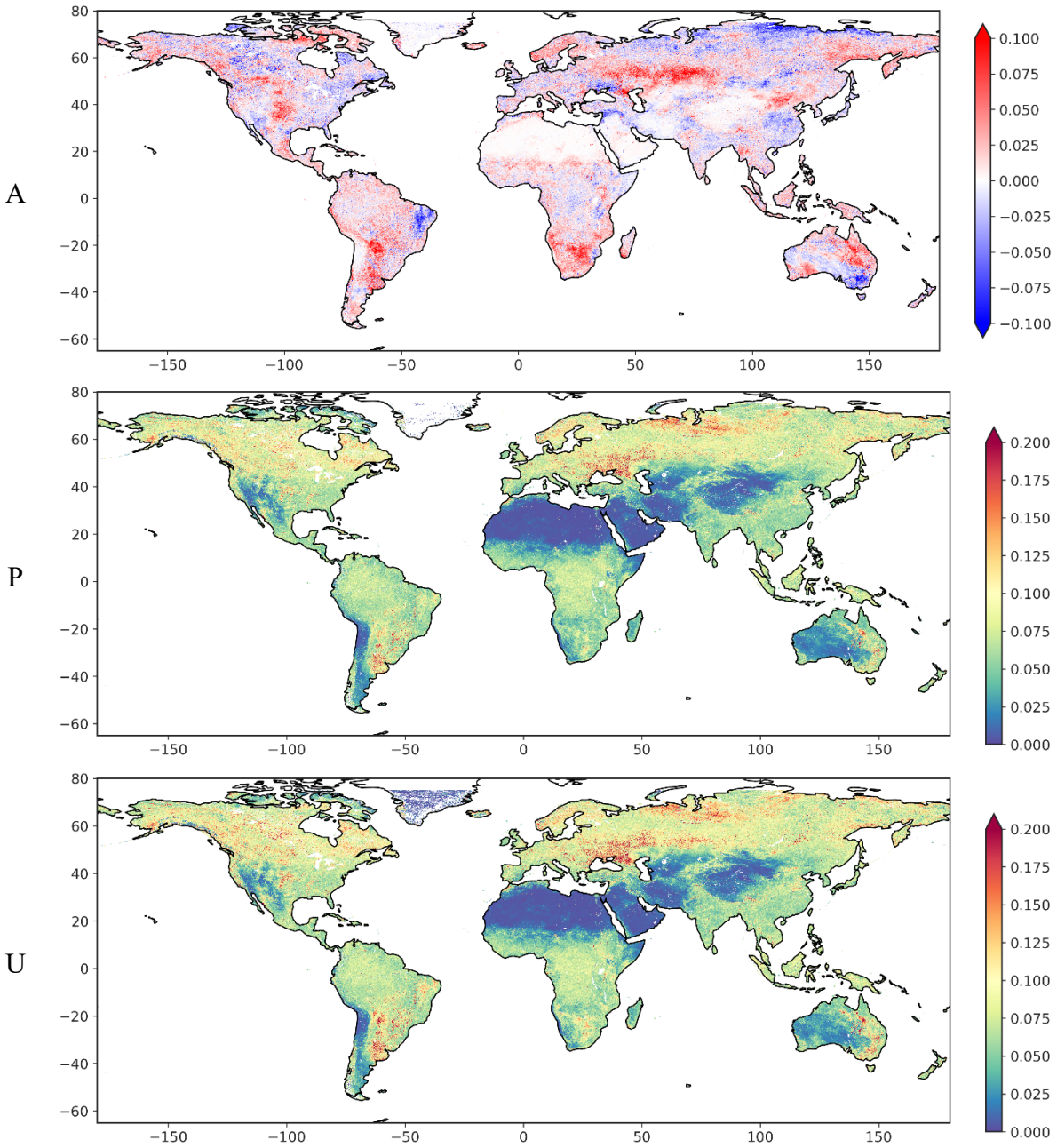


Figure 74: Spatial patterns of pixel-wise APU between the 2025 NRT period and the STS for all pairwise valid NDVI values.

Temporal consistency

Figure 75 and Figure 76 show the temporal evolution of percentiles (P5, P25, P50, P75, and P95) for the NDVI 300m V3 January – December 2025 dataset and the STS over all LANDVAL V2 sites for

landcover types and per biome, respectively. Pairwise masking is applied, in such a way that the percentiles for each dekad are based on the same set of pixel extractions in both datasets. The temporal evolution of 2025 is in good agreement with that of the STS.

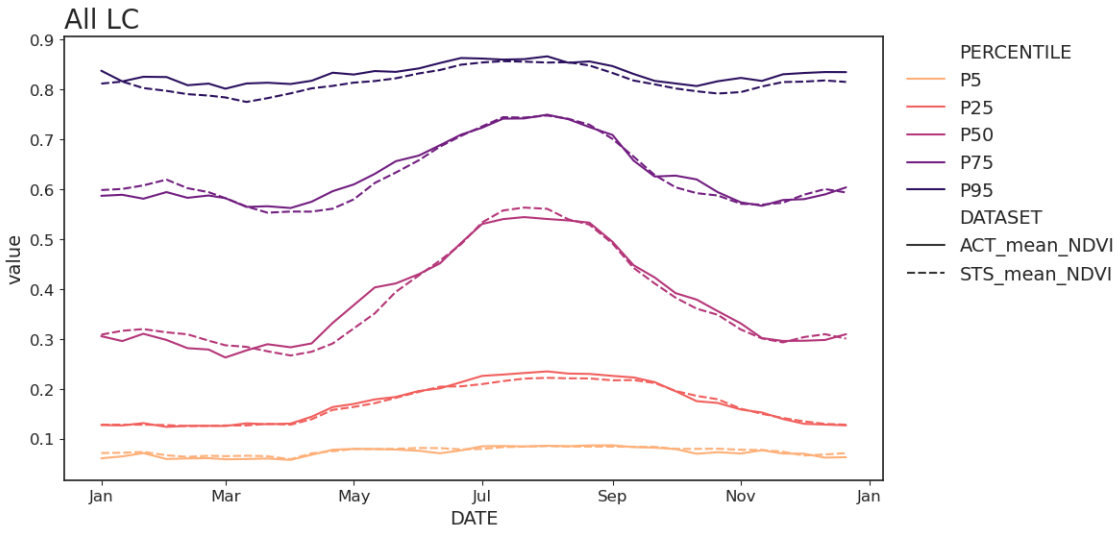


Figure 75: Temporal evolution of percentiles (P5, P25, P50, P75, and P95) for NDVI 300m V3 January – December 2025 (ACT, solid lines) and the STS (dashed lines) over all biome types. Pairwise masking between ACT and STS is applied per dekad.

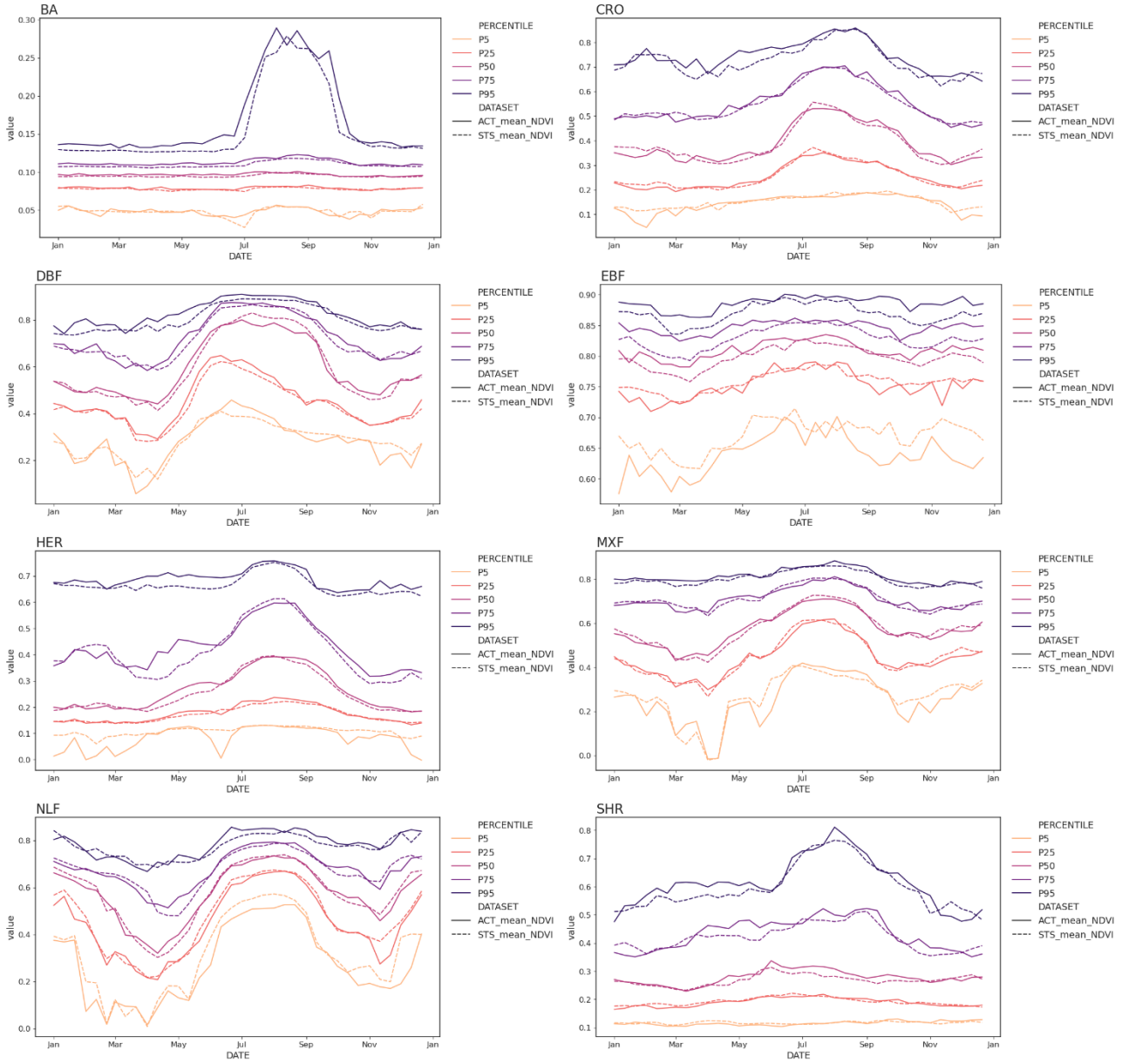
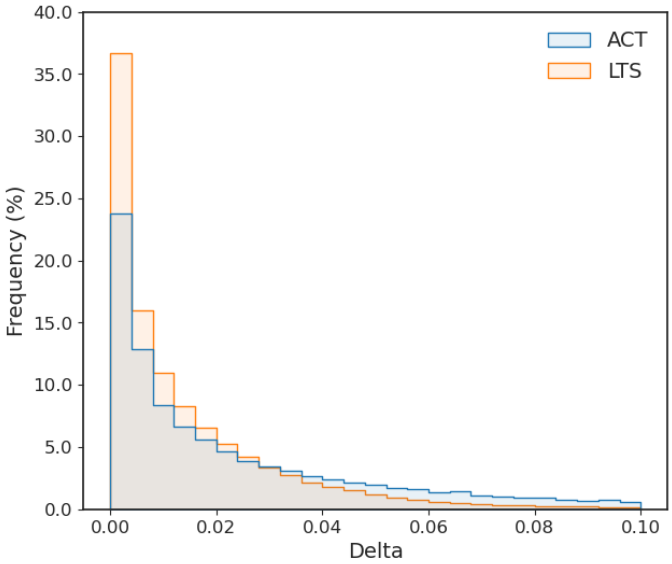


Figure 76: Temporal evolution of percentiles (P5, P25, P50, P75, and P95) for NDVI 300m V3 January – December 2025 (ACT, solid lines) and the STS (dashed lines) per biome type.

The frequency histograms of the temporal smoothness (or delta, δ) for the NDVI 300m V3 January – December 2025 dataset and the STS dataset are presented in Figure 77. The STS dataset has a higher smoothness because observations are averaged over five years and a longer accumulation window has been used in the BRDF processing. The values of the median temporal smoothness and the relative noise are comparable to those of the reference datasets in §5.8.3.

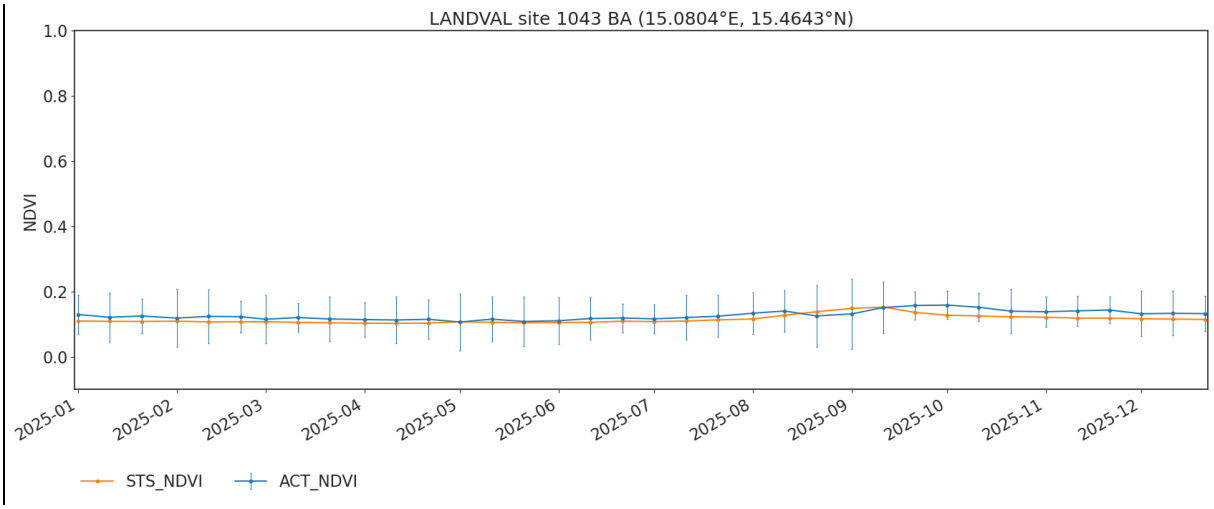


	Median δ	Relative noise
2025 (ACT)	0.015	11
STS	0.008	7

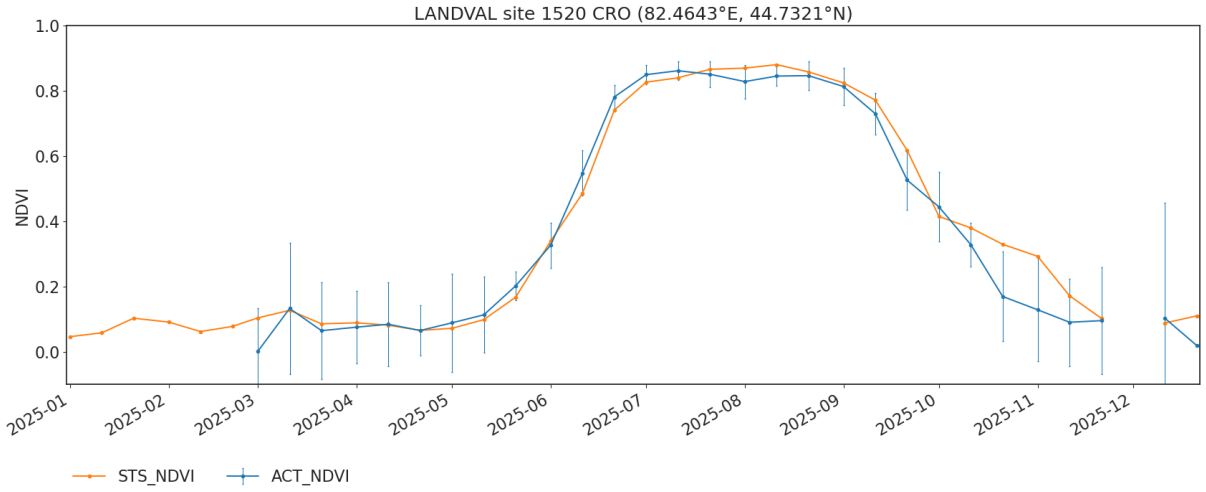
Figure 77: Frequency histograms of temporal smoothness (δ) for NDVI 300m V3 January – December 2025 ('ACT', blue) and the STS (orange) on the left, and median temporal smoothness and relative noise on the right.

For each biome type, a temporal profile of a selected LANDVAL V2 site is shown in Figure 78. The NDVI 300m V3 dataset for the year 2025 is shown in blue and the STS is shown in orange. Overall, the profiles show similar behaviour. Some differences are expected since two different periods are compared with each other, thus discrepancies from natural vegetation differences are expected. The back-processed dataset (STS) tends to be smoother and has larger completeness, especially during winter months. The error bars of the NRT dataset are larger than for the back-processed datasets in the main QAR (see Figure 59) since the NDVI uncertainties are larger

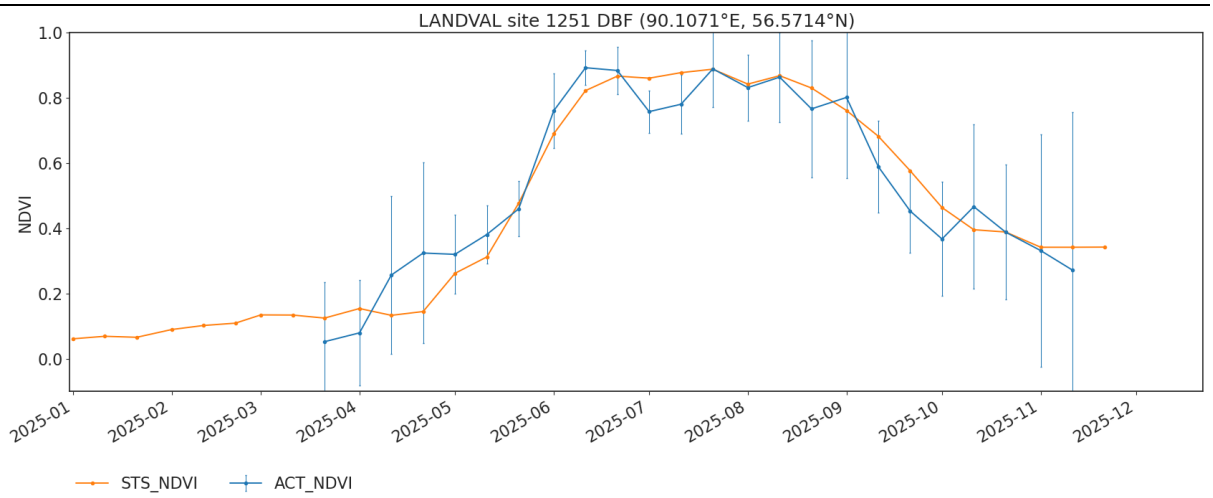
BA



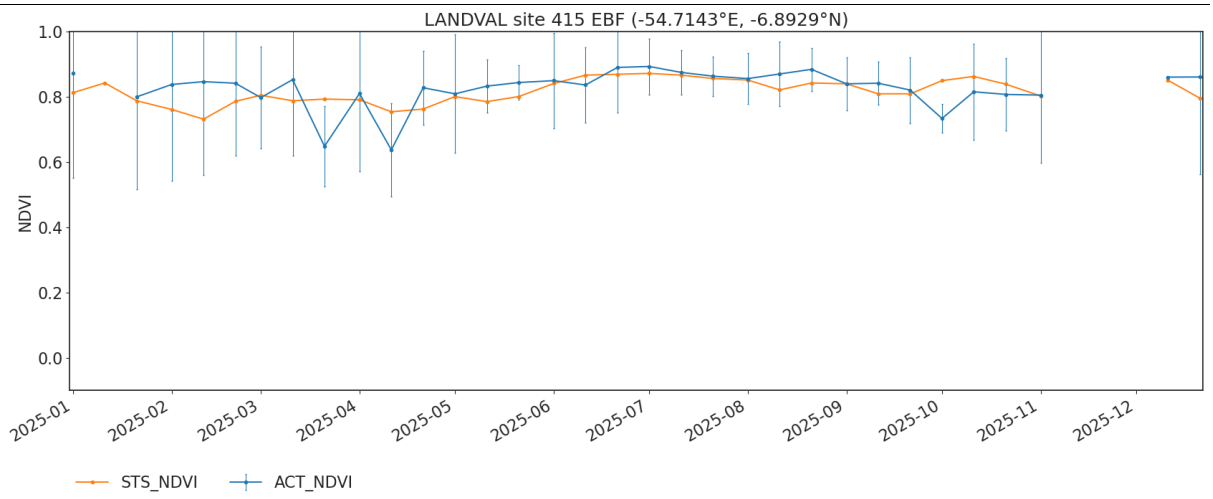
CRO


LANDVAL site 1520 CRO (82.4643°E, 44.7321°N)
NDVI
1.0
0.8
0.6
0.4
0.2
0.0
2025-01
2025-02
2025-03
2025-04
2025-05
2025-06
2025-07
2025-08
2025-09
2025-10
2025-11
2025-12
STS_NDVI
ACT_NDVI

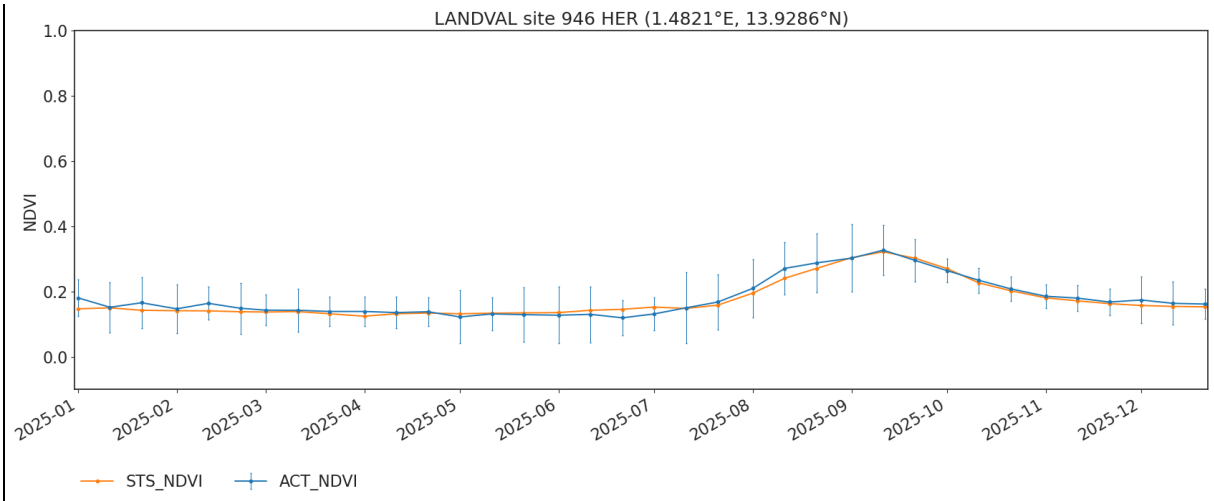
DBF


LANDVAL site 1251 DBF (90.1071°E, 56.5714°N)
NDVI
1.0
0.8
0.6
0.4
0.2
0.0
2025-01
2025-02
2025-03
2025-04
2025-05
2025-06
2025-07
2025-08
2025-09
2025-10
2025-11
2025-12
STS_NDVI
ACT_NDVI

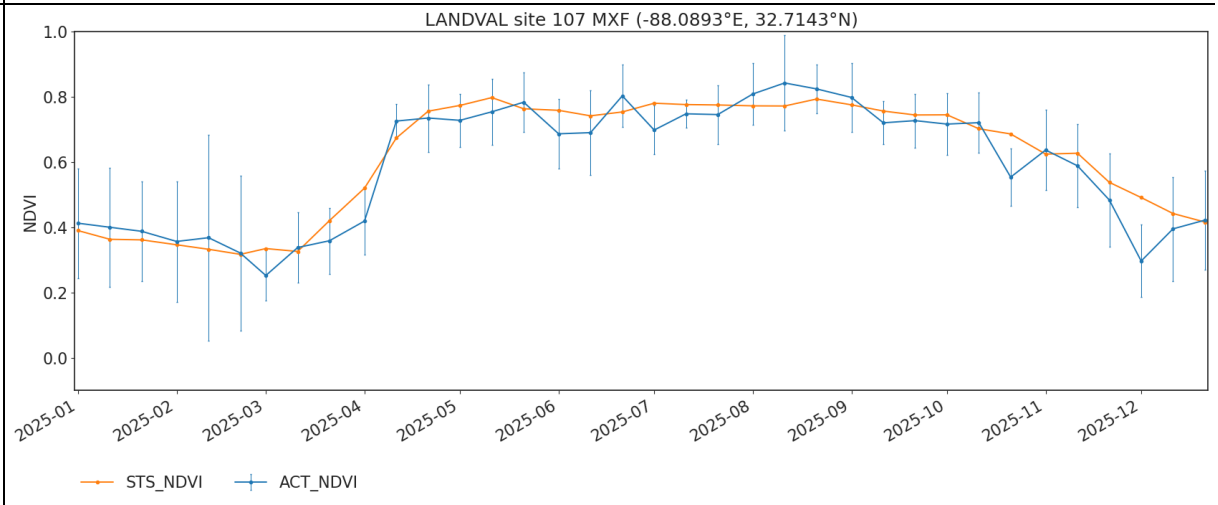
EBF


LANDVAL site 415 EBF (-54.7143°E, -6.8929°N)
NDVI
1.0
0.8
0.6
0.4
0.2
0.0
2025-01
2025-02
2025-03
2025-04
2025-05
2025-06
2025-07
2025-08
2025-09
2025-10
2025-11
2025-12
STS_NDVI
ACT_NDVI

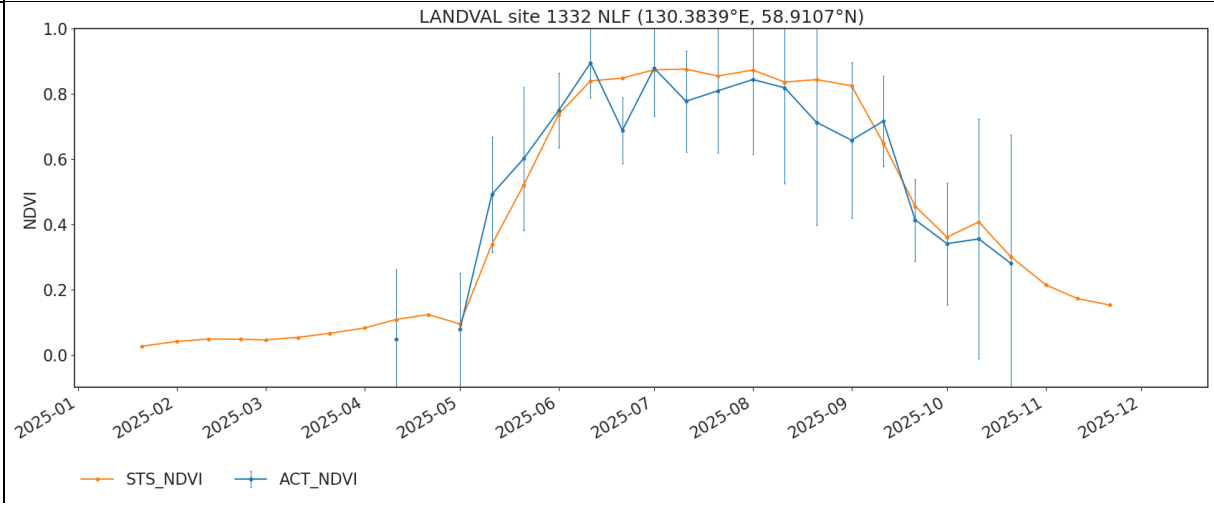
HER


LANDVAL site 946 HER (1.4821°E, 13.9286°N)
NDVI
1.0
0.8
0.6
0.4
0.2
0.0
2025-01
2025-02
2025-03
2025-04
2025-05
2025-06
2025-07
2025-08
2025-09
2025-10
2025-11
2025-12
STS_NDVI
ACT_NDVI

MXF


LANDVAL site 107 MXF (-88.0893°E, 32.7143°N)
NDVI
1.0
0.8
0.6
0.4
0.2
0.0
2025-01
2025-02
2025-03
2025-04
2025-05
2025-06
2025-07
2025-08
2025-09
2025-10
2025-11
2025-12
STS_NDVI
ACT_NDVI

NLF


LANDVAL site 1332 NLF (130.3839°E, 58.9107°N)
NDVI
1.0
0.8
0.6
0.4
0.2
0.0
2025-01
2025-02
2025-03
2025-04
2025-05
2025-06
2025-07
2025-08
2025-09
2025-10
2025-11
2025-12
STS_NDVI
ACT_NDVI

SHR

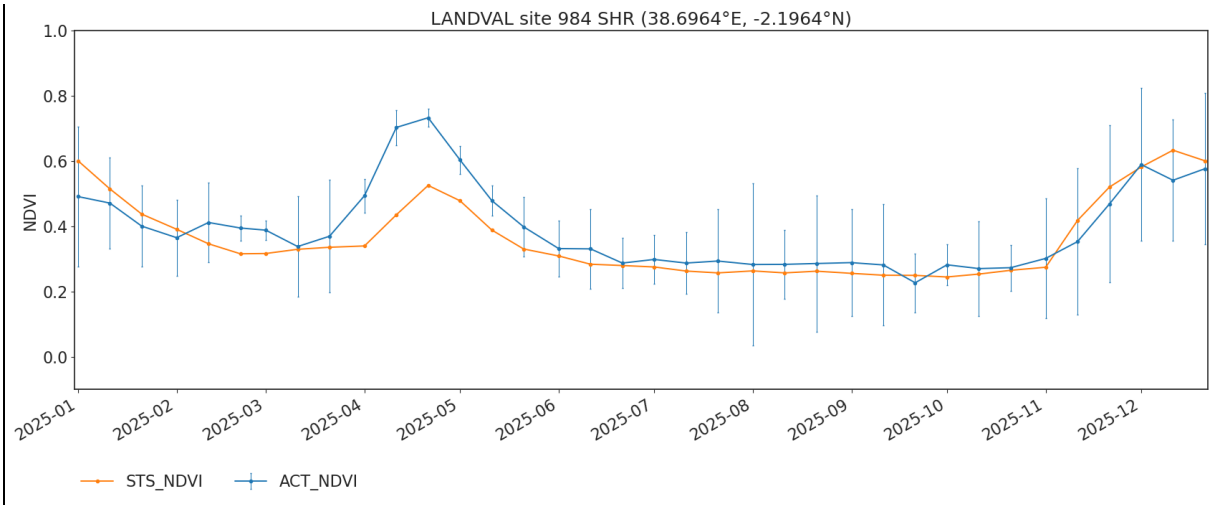


Figure 78: Temporal profiles over a selection of LANDVAL V2 sites for NRT data from 2025 (blue) and the STS (orange).

The stability analysis from §5.8.4 is extended to include the NRT period from January – December 2025. The temporal evolution of inter-annual precision is shown in Figure 79. The stability over the entire time frame is equal to 0.003 (0.8%), well within the goal requirement of 1.5% (see §2.1).

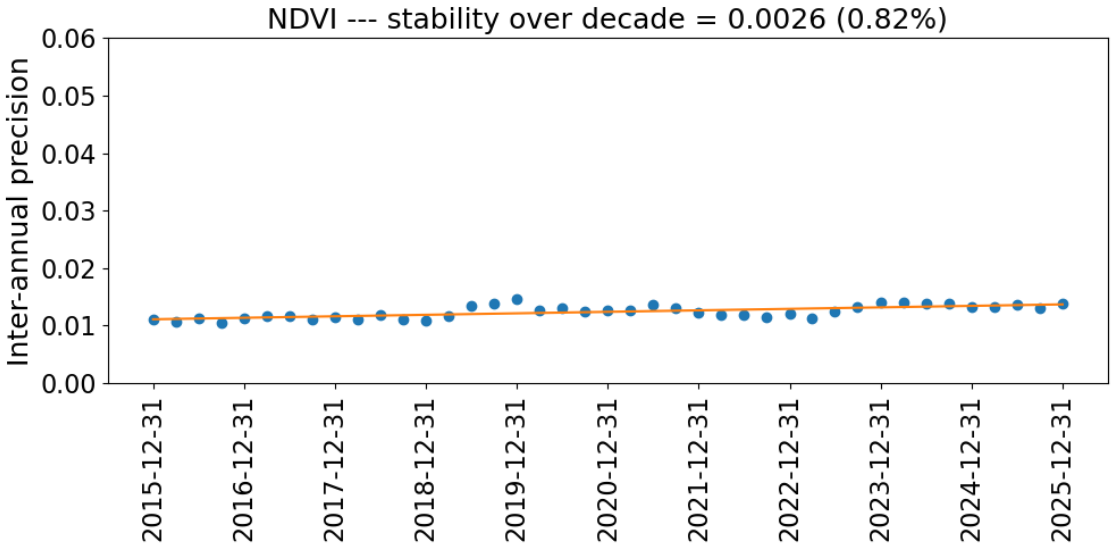


Figure 79: Temporal evolution of the inter-annual precision of the NDVI 300m V3 product from 2014 to 2025

CONCLUSION

The comparison analysis between back-processing and NRT processing of the NDVI 300m V3 product showed overall very good agreement, both statistically, spatially, and temporally. However,

some discrepancies are found in product completeness, the occurrences of quality flags, and in the magnitude of uncertainties, as expected. The NRT product has lower completeness (especially at high latitudes), less occurrences of the NOBS=0, warning, and extreme warning quality flags, and overall higher uncertainties. This can all be traced back to the difference in BRDF accumulation window, which is equal to 14 months in back-processing, but only spans 30 days in NRT processing. A smaller accumulation window is more affected by missing observations due to snow, cloud/cloud shadow cover, or bad illumination conditions. Therefore, it is more likely that there are missing TOC-r reflectances during the 30-day accumulation period and thus more NDVI values will be flagged. This results in lower product completeness, more and longer gaps in the product, and, somewhat counterintuitively, less pixels that are flagged with NOBS=0. A shorter accumulation window also results in higher NDVI uncertainties, and related to that, less pixels are flagged with a warning or extreme warning.

The analyses in this annex also serve as SQM of the January – December 2025 NDVI 300m V3 product. There is overall good consistency between the 2025 period and the STS. There are small differences between both datasets, but they are acceptable since different periods are compared for which natural variation in vegetation of the surface is expected, and because the accumulation period for the BRDF correction is different in back-processing and NRT processing. The product also has high stability with stability metric equal to 0.8%, well within the goal requirement of 1.5%.