TERRASCOPE SENTINEL-2

QUALITY ASSESSMENT REPORT (QAR)

S2 – BIOPAR and NDVI inter-comparison – V200-V102

Reference: ***Terrascope Sentinel-2 Quality Assessment Report S2 – BIOPAR and NDVI inter-comparison – V200-V102***

Author(s): Else Swinnen, Liesbeth De Keukelaere

Version: 1.0
Date: 5/10/2020

DOCUMENT CONTROL

Signatures

Authors	Else Swinnen, Liesbeth De Keukelaere
Reviewers	Liesbeth De Keukelaere
Approvers	Dennis Clarijs, Jurgen Everaerts
Issuing authority	VITO

Change record

Release	Date	Updates	Approved by
1.0	05/10/2020	Initial external version	Dennis Clarijs, Jurgen Everaerts, Erwin Wolters

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ACRONYM	EXPLANATION
ATBD	Algorithm Theoretical Basis Document
BIOPAR	Biophysical Parameter
CCC	Canopy Chlorophyll Content
CGS	Collaborative Ground Segment
CWC	Canopy Water Content
ESA	European Space Agency
FAPAR	Fraction of absorbed photosynthetic active radiation
FCOVER	Fraction of vegetation cover
GM	Geometric mean
iCOR	Image Correction for atmospheric effects
LAI	Leaf Area Index
MAE	Mean Absolute Error
MBE	Mean Bias Error
MEP	Mission Exploitation Platform
MSI	Multi-Spectral Instrument
NDVI	Normalized Difference Vegetation Index
OGC	Open Geospatial Consortium
PDF	Product Distribution Portal
PROBA-V	Project for On-Board Autonomy – Vegetation
PUM	Product User Manual
QAR	Quality Assessment Report
RAA	Relative Azimuth Angle
RD	Related Document
RMPDs	Root Mean Product Difference systematic: systematic differences
RMPDu	Root Mean Product Difference unsystematic: random differences
RMSE	Root Mean Squared Error
S2	Sentinel-2
Sen2Cor	Sentinel-2 Correction
SZA	Solar Zenith Angle
TOC	Top-Of-Canopy
VITO	Vlaams Instituut voor Technologisch Onderzoek
VZA	View Zenith Angle
WVP	Water vapour

1 Introduction

1.1 TERRASCOPE explained

Terrascope is the Belgian platform for Copernicus, PROBA-V, and SPOT-VEGETATION satellite data, products, and services. It provides easy, full, free and open access to all users without restrictions. This allows non-specialist users to explore the wealth of remote sensing information and build value-added products and services.

The following data are included:

- The SPOT-VEGETATION archive
- The PROBA-V archive
- Sentinel-1 SAR data over Belgium and its surroundings
- Sentinel-2 optical data over Europe and soon to be expanded to Africa
- Sentinel-3 optical and thermal Synergy (SYN) – Vegetation (VGT) data

Normalized Difference Vegetation Index (NDVI) and biophysical indicators (BIOPARS) derived from Sentinel-2 data are offered alongside the Top-Of-Canopy (TOC) reflectance products. The biophysical parameters are: fraction of absorbed photosynthetically active radiation (FAPAR), leaf area index (LAI), fraction of vegetation cover (FCOVER), canopy chlorophyll content (CCC), and canopy water content (CWC). The latter two products are not visible in the Terrascope Viewer, but the data can be downloaded and/or further explored using the Terrascope Virtual Machine environment (see below).

Users have the possibility to build derived information products to their own specification, using the Terrascope processing cluster through provided Virtual Machines or Notebooks. This eliminates the need for data download (and consequential storage costs), because the cluster holds all of the data and it is directly accessible. Integration of data or products in your own application is facilitated through Open Geospatial Consortium (OGC) web services.
Terrascope is user-centered, so any suggestions for new or enhanced functionality are welcome. Feel free to contact us: info@terrascope.be.

1.2 Scope of the document

This quality assessment report (QAR) documents the results of the inter-comparison of the most recent version (V200) of the TERRASCOPE BIOPAR and NDVI products against the former version (V102).
The document is organized as following:

- Section 2 describes the method of inter-comparison, the metrics and sample strategy used, and the ancillary data.
- The next 4 sections discuss the results of the inter-comparison:
    - Section 3: FAPAR
    - Section 4: FCOVER
    - Section 5: LAI
    - Section 6: NDVI
- Section 7 summarizes the results for all BIOPARS and NDVI.

The sections describing the results are written as stand-alone sections for readers interested in only one variable.

1.3 Description

The TERRASCOPE S2 BIOPAR and NDVI product are generated based on the S2 Level-2A (L2) Top-Of-Canopy (TOC) products that are downloaded from the COPERNICUS Science Hub, or, for historic data, generated using the Sen2Cor atmospheric correction tool (Louis et al., 2017).

From the TOC reflectance products, the NDVI and BIOPARs are derived using the methodology developed by INRA-EMMAH. It mainly consists in simulating a comprehensive data base of canopy (TOC) reflectances based on vegetation characteristics and observation and illumination geometry. Neural networks are then trained to estimate a number of these canopy characteristics (BIOPARs) from the simulated TOC reflectances along with a set of corresponding angles defining the observational configuration. For Sentinel-2, two neural networks (NN) were developed by INRA-EMMAH and implemented in Terrascope:

- one based on 10 m input bands only (R3-NN)
- one based on a combination of 10 m and 20 m input bands (R8-NN). This NN is also implemented in the Sentinel-2 Toolbox.

The output BIOPARs are:

- Leaf Area Index (LAI),
- Fraction of Absorbed Photosynthetically Active Radiation (FAPAR),
- Fraction of Vegetation Cover (FCOVER),
- Canopy Chlorophyll Content (CCC)
- Canopy Water Content (CWC).

This document compares the FAPAR, LAI, FCOVER and NDVI data of the V200 with V102 for one tile (31UFS).

1.3.1 Changes between the processing of V200 and V102

All changes between V200 and V102 are described in Table 1. The changes that are expected to have an impact on the BIOPAR and NDVI data values are indicated in blue.
For the scene classification, an updated version of Sen2Cor is used (Mueller-Wilm et al., 2019). This can lead to different detections of the classes of the scene classification layer.
Sen2Cor (Louis et al., 2017) is used in V200 for the atmospheric correction of the reflectances instead of iCOR in V102. This has several impacts: (1) the atmospheric correction is different, (2) the aerosol optical thickness (AOT) estimation is done differently resulting in different AOT values used in the correction, and (3) Sen2Cor performs standard a topographic correction which was not done in V102. This latter change will have an influence on areas with more relief. If the AOT retrieval fails in Sen2Cor, an arbitrary value of AOT=0.2 is used for the atmospheric correction. In iCOR (V102), a fallback AOT climatology dataset (from ECMWF) was used.
In addition, in V102 the geolocation of each image was checked and where necessary a geometric shift was applied. Because this also introduced errors in the dataset, it was decided not to do this anymore for V200. As a consequence, when comparing images of V200 with V102, some are affected by a geolocation shift.
All angles are now generated per pixel and used as input for the neural networks to retrieve the BIOPARS (V200).

More information on the processing of the V102 and V200 data sets can be found in [RD-1], [RD-2], [RD-4] and [RD-5].

Table 1: List of changes between TERRACOPE Sentinel-2 v200 and v102

Adaptations between v200 and V102	Clarification
Scene selection	
Tiles covered with more than 95% clouds are not processed.	To optimise data storage, scenes almost fully covered with clouds will no longer be processed.
Scene classification	
Update of Sen2Cor version (v2.5.5 à v2.8)	On 20.02.2019 a new version of Sen2Cor was released (v2.8).
Atmospheric correction	
Download of Sentinel-2 L2A Near-Real-Time products	Since March 2018, the Payload Data Ground Segment (PDGS) has processed L2A products over Europe, and production was extended to global in December 2018. [RD1]
Atmospheric correction with Sen2Cor instead of iCOR.	To be in line with the ESA processing, we have switched from iCOR to SEN2COR processing. Both are valid processors and have been included in various inter-comparison exercises (e.g. ACIX-I (Doxani et al., 2018) and ACIX-II (publication in preparation)).
Biopar processing	
Sun and sensor angle for each pixel in calculated and included in the BIOPAR processing	On 06/11/2018 ESA deployed a new Production Baseline (02.07) which includes an accurate detector footprint. With this information it is possible to link one pixel to one detector.
Output products	
Additional layers at 60 m: - WVP - Angles: RAA, VZA and SZA	Users can use these layers to derive other products or for quality control.
Removal of the: - Cloudmask layers (10m, 20m and 60m) - Shadowmask layers (10m, 20m, 60m)	This information can also be found in the SCENECLASSIFICATION layer.
Update of the INSPIRE Metadata	The metadata xml file has been updated to be compliant with ISO-19115-2 standards.
Pixels identified in the SCENECLASSIFICATION layer as cloud, cloud shadow, snow, cirrus or saturated pixels are masked in the NDVI and BIOPAR products.	Pixels belonging to one of these groups will retrieve inaccurate NDVI and BIOPAR values. To avoid wrong interpretations, these pixels are masked in the end products.

Adaptations between v200 and V102	Clarification
Data Archive	
Switch from Product Distribution Portal (PDF) to the Terrascope Catalogue with accompanying GeoJSON metadata information.	The new catalogue is easier to maintain.
Change in folder structure at TERRASCOPE platform, i.e. on the Virtual Machines (VMs) and Notebooks.	The folder structure has been revised, made more concise and user-friendly.

1.3.2 Research questions

The research questions driving this analysis are the following:

- What is the magnitude of the agreement and the difference between the V200 and V102 products?
- Is there a bias between V200 and V102 products and what is its magnitude?
- How do the differences behave over time?
- What are the main drivers for the differences, which changes have the most impact on the analysis.

1.4 Related documents

Table 2 lists the related documents that are used as input to this document.

Table 2: List of related documents.

	Document and link
RD-1	TERRASCOPE Sentinel-2 ATBD S2 – TOC – V200
RD-2	TERRASCOPE Sentinel-2 ATBD S2 – TOC – V102
RD-3	TERRASCOPE Sentinel-2 QAR S2 - TOC inter-comparison V200-V102
RD-4	TERRASCOPE Sentinel-2 ATBD S2 – NDVI & BIOPAR – V200
RD-5	TERRASCOPE Sentinel-2 ATBD S2 – NDVI & BIOPAR – V102

2 Methods

2.1 Overall procedure

The time series of 20150706 until 20200328 for tile 31UFS (Figure 1) is used in the inter-comparison analysis. A sample of paired observations from V200 and V102 are statistically compared. The general sampling method, as well as the specific sampling methods per analysis are described in detail in section 2.2.

In order to answer the research questions, the following criteria are assessed on the sample of paired observations: (1) product completeness, (2) statistical consistency, (3) temporal consistency, and (4) spatial consistency. These are further explained in section 2.3.

A range of inter-comparison metrics and plots are used to compared the V200 against V102 data. These are detailed in section 2.4.

The analysis is done for FAPAR, FCOVER, LAI, and NDVI. For the former 3, both 10 m and 20 m resolution data are available and these are discussed together.

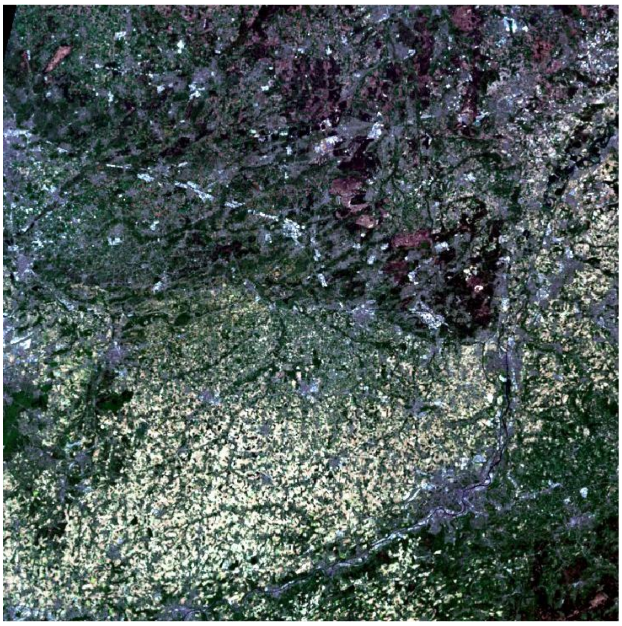


Figure 1: True color image of tile 31UFS for 22/04/2020.

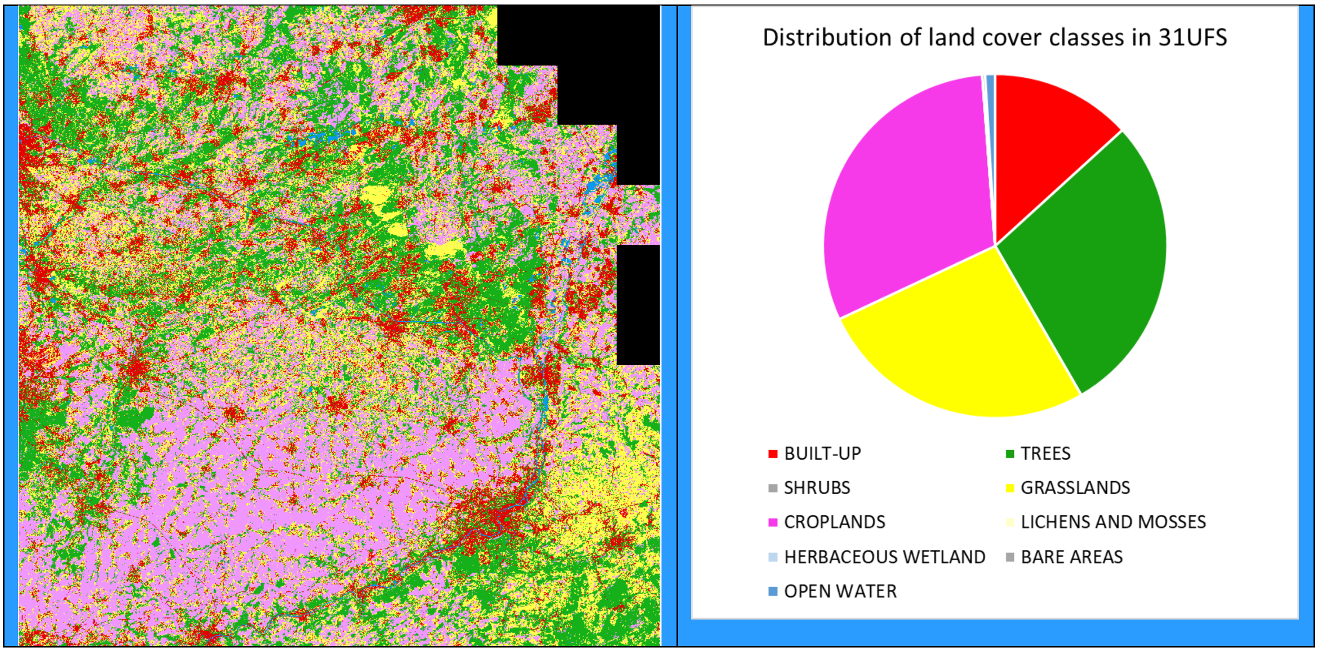
2.2 Sampling strategies

The inter-comparison is always performed on the entire time series, but not on all pixels. The image pairs per date (V200 and V102) are systematically subsampled. Here the middle pixel in a grid is selected. Simultaneously, the same is done for the associated SCENECLASSIFICATION layer. Only pixels belonging to the following classes are retained for the analysis: dark area pixels (DN=2), vegetation (DN=4), bare soil (DN=5) and water (DN=6). Only when both V200 and V102 SCENECLASSIFICATIONs label the pixel as one of these classes, the corresponding pixel values from V200 and V102 of the NDVI or BIOPAR is retained in the sample.

The prototype 10 m land cover map generated in the framework of the [ESA WorldCover project](#) is used to differentiate the analyses per land cover class (Figure 2). The 4 classes 'built-up', 'croplands', 'grasslands' and 'trees' are the dominant land cover types for tile 31UFS. Therefore, only these classes are analysed separately.

While conducting the analysis, it was observed that for a number of dates, the V200 and V102 images have a geolocation shift. In V102, the images were assessed for geolocation accuracy, and if necessary a shift was applied. The V200 processing does not include this assessment and shift anymore, because it was observed that it also introduced errors. When comparing the images for these dates, the differences can be large, because we are not comparing the same pixels. Therefore, these dates are excluded in the overall and spatial analysis. In the temporal analysis, these dates are kept in order to show the effect on the results.

Figure 2: Prototype 10 m land cover map for tile 31UFS (left) and distribution of the classes (right), generated in the framework of the ESA WorldCover project.



2.2.1 Overall analysis

The systematic subsample is defined as the middle pixel in an arbitrary grid of 21 × 21 pixels for the 20 m parameters, and the 21st pixel in a grid of 42 × 42 pixels for the 10 m products. The total sample is the combination of all subsampled image pairs. All images from the entire time series are subsampled, except those with a geometric shift in V102.
The statistics are calculated on the entire sample. The outputs are statistics and graphs for the entire dataset. This is calculated for the entire sample or per land cover class.

2.2.2 Temporal analysis

The systematic subsample is defined the same as for the overall analysis, i.e. the middle pixel in a grid of 21 × 21 pixels for the 20 m parameters, and the 21st pixel in a grid of 42 × 42 pixels for the 10 m products. There is one sample per image pair, i.e., per date. Dates with insufficient data were not considered in the analysis (< 250 good pixels).

The statistics are calculated on the sample per date. The output are temporal graphs of the statistics, such that their evolution over time can be evaluated (e.g. seasonal trends). The analysis are performed for all land cover types, but also per land cover type.

2.2.3 Spatial analysis

The sampling of the spatial analysis is done differently. Here, the statistics are calculated over the time series of selected pixels. Again, a systematic subsample is used, but with a finer grid size: for the 20 m data the middle pixel of a 11 × 11 pixel grid is taken to allow more spatial detail, and for the 10M data the 11th pixel within a 22 × 22 pixels grid. The dates for which the V102 images have a geometric shift are excluded from the analysis.
The output are images, one for each statistic. No distinction in land cover type is made.

2.3 Criteria

Table 3 lists the criteria that are assessed on the data sets.

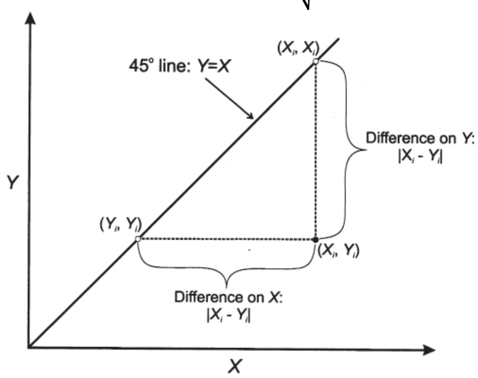
Table 3: Description of the criteria under evaluation

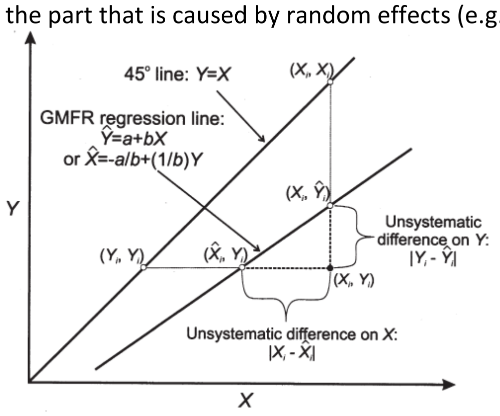
Criterium	Description
Statistical consistency	The evaluation of statistical consistency is based on (i) histograms of overall bias between datasets, and (ii) the geometric mean regression between datasets. This is done for all land cover and per land cover.
Temporal consistency	Temporal consistency is analysed through evaluation of the temporal variations of validation metrics (i.e. metrics calculated per time step)
Spatial consistency	Spatial consistency refers to the realism and repeatability of the spatial distribution of retrievals, including the absence of artefacts (e.g., missing data, stripes, unrealistic values, etc.), based on expert knowledge. The analysis is based on spatial distribution of validation metrics at tile level .

2.4 Inter-comparison metrics

The validation metrics that are used for the assessment of statistical, temporal and spatial consistency are listed in Table 4. A selection of these validation metrics can be calculated per pixel over time in order to evaluate spatial consistency, per time step over grouped pixels in order to evaluate temporal consistency or per pixel over time to evaluate the spatial consistency. In the equations below, X is V200 and Y is V102.

Table 4: Validation metrics used for the inter-comparison of the Sentinel-2 V102 and V200 TOC reflectance and BIOPARs.

Metric	Description
R^2	The coefficient of determination (R^2) indicates agreement or covariation between two data sets with respect to a linear regression model. It summarizes 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 data sets. $$R^2 = \left(\frac{\sigma(X,Y)}{\sigma(X)\cdot\sigma(Y)}\right)^2$$ with $\sigma(X)$ and $\sigma(Y)$ the standard deviation of X and Y and $\sigma(X,Y)$ 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 orthogonal geometric mean (GM) regression model is used to identify the relationship between two data sets of remote sensing measurements, with both data sets subject to noise. The GM regression model minimizes the sum of the products of the vertical and horizontal distances (errors on Y and X): $$Y = a + b\cdot X$$ with $b = \frac{\lambda_1 - \sigma_X{}^2}{\sigma_{XY}}$ (GMR slope) $a = \bar{Y} - b\cdot\bar{X}$ (GMR intercept) σ_X and σ_{XY} : the variance of X and the covariance of XY λ_1: from Duveiller et al. (2016) $\bar{X}$: the mean value of X $\bar{Y}$: the mean value of Y
RMSE	The Root Mean Squared Difference (RMSE) measures how far the difference between the two data sets deviates from 0 (Figure 3). $$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2}$$  *Figure 3: 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)*

Metric	Description
Random or Unsystematic difference (RMPDu)	The Root of the Mean Random or Unsystematic Product Difference $$RMPD_u = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\lvert X_i - \hat{X}_i\rvert\right)\left(\lvert Y_i - \hat{Y}_i\rvert\right)}$$ with $\hat{X}_i$ and $\hat{Y}_i$ estimated using the GM regression line and n the number of samples. It expresses the part of the difference that is not explained by the systematic difference (Figure 4) which is the geometric mean regression line, but the part that is caused by random effects (e.g. viewing angle dependencies).  *Figure 4: The unsystematic difference is measured by the distance between the point ($\boldsymbol{X_i}$, $\boldsymbol{Y_i}$) and the GM regression line on X and Y axis (Ji and Gallo, 2006)*
Systematic difference (RMPDs)	The Root of the Mean Systematic Product Difference $$RMPD_s = \sqrt{MSD - MPD_u}$$ with $$MSD = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2$$ It expresses the part of the difference that is explained by the systematic difference (Figure 4), this is the difference of the geometric mean regression line and the 1-1 line.
MBE	The Mean Bias Error (MBE) is defined as the actual difference between two data sets and positive and negative differences between observations: $$MBE = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i) = \bar{X} - \bar{Y}$$
MAE	The Mean Absolute Error (MAE) is defined as the absolute deviation between two data sets: $$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert X_i - Y_i\rvert$$

3 FAPAR 10M and 20M: RESULTS

This chapter discusses the results of the inter-comparison between V200 and V102 for the FAPAR products at 10 m and 20 m resolution.

3.1 Overall statistical analysis

The Mean Bias Difference (MBE), Mean Absolute Difference (MAE), Root Mean Squared Difference (RMSE), which is further split into random and systematic differences, were calculated on the overall sample (see section 2.2.1) for all land cover classes and per land cover class. The results are summarized in Table 5.

Comparing the statistics between the two resolutions, it is observed that the overall MBE for 20M is negative (-0.014) and almost 0 for the 10 m resolution FAPAR. This suggests that the difference is the largest for the FAPAR_20M, but the opposite is the case. The MAE and RMSE are both higher for FAPAR_10M than for FAPAR_20M. The reason why the MBE is so small is that positive and negative values are almost equally present. This is further investigated in the next sections. For FAPAR_10M, the random differences are always larger than the systematic ones. This is not the case for FAPAR_20M, where for built-up and cropland classes, the random difference is of the same magnitude as the systematic differences.

All difference statistics are largest for ‘built-up’ areas. The MBE is larger negative for ‘built-up’, shifts to smaller negative or even positive values when vegetation content increases.

Table 5: Mean Bias Error (MBE), Mean Absolute Error (MAE), Root Mean Squared Difference (RMSD), Random difference and Systematic difference between V200 and V102 of FAPAR_20M and FAPAR_10M for all land cover classes (All LC) or per biome.

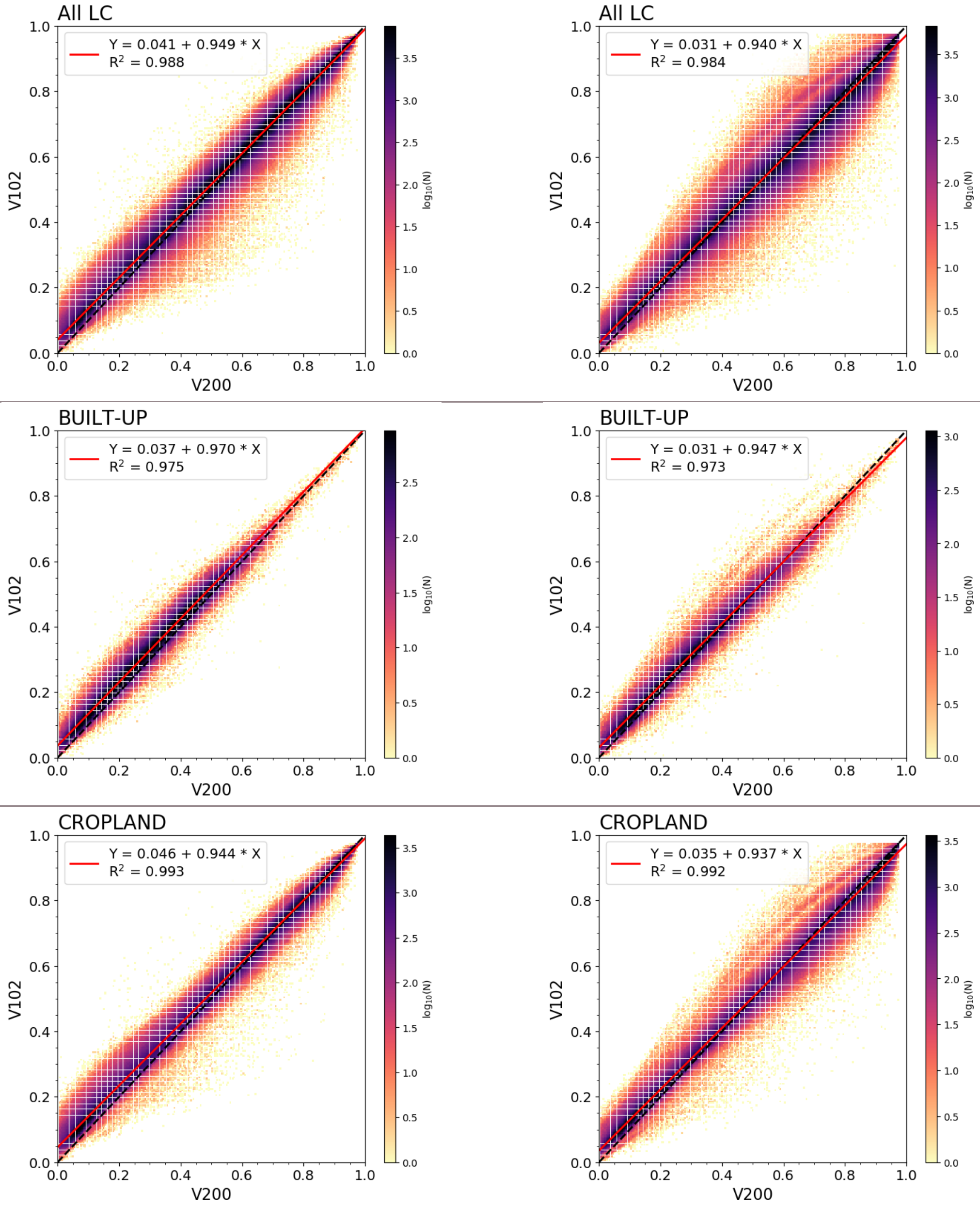
	FAPAR_20M					FAPAR_10M				
	MBE	**MAE**	**RMSE**	**Rand. diff**	**Sys. diff.**	**MBE**	**MAE**	**RMSE**	**Rand. diff**	**Sys. diff.**
All LC	-0.014	0.023	0.031	0.025	0.018	0.001	0.026	0.035	0.032	0.015
BUILT-UP	-0.028	0.031	0.037	0.023	0.028	-0.017	0.024	0.031	0.025	0.019
CROPLAND	-0.018	0.025	0.033	0.023	0.024	-0.004	0.025	0.033	0.027	0.019
GRASSLAND	-0.011	0.022	0.029	0.024	0.016	0.005	0.025	0.034	0.032	0.013
TREES	-0.008	0.021	0.029	0.027	0.009	0.009	0.028	0.039	0.038	0.012

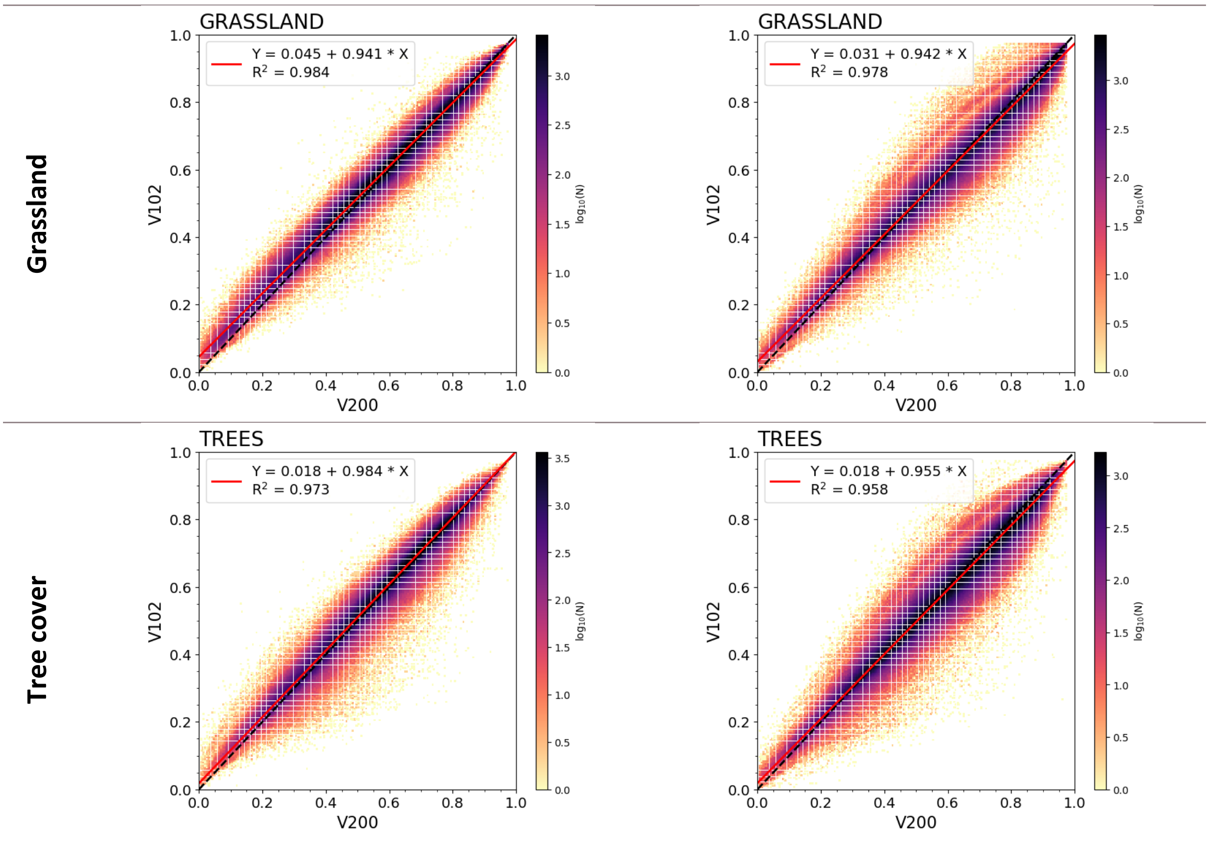
Figure 5 shows the corresponding scatterplots with the orthogonal geometric mean regression line and the coefficient of determination. The V200 and V102 are overall linearly related, with FAPAR_20M having less scatter. For all land covers and both resolutions, the V200 FAPAR has more smaller values as is shown by the positive offset of the regression line. For FAPAR_20M, the regression line does not cross the 1:1 line, but for FAPAR_10M it does.

The scatterplot of FAPAR_10M seemingly shows a separate population above the regression line. This is visible in the plot for all land cover, but more pronounced for ‘built-up’, ‘cropland’, and ‘grassland’. This was further investigated by looking at the scatterplots of single dates. For most of the dates, the scatterplots are narrow, linear and close to the 1:1 line. But when the Aerosol Optical Thickness (AOT) used to correct the data for atmospheric scattering differs largely between V200 and V102, the relationship drifts from the 1:1 line and becomes non-linear or shows more scatter.

Figure 5: Scatterplots between V200 (X) and V102 (Y) FAPAR_20M (left) and FAPAR_10M (right) products for the period 01/07/2015 – 26/03/2020. Red line is the orthogonal geometric mean regression line between the two versions.

	FAPAR_20M	**FAPAR_10M**
All land cover		
Built-up		
Cropland		





Examples are shown in Figure 6 for 3 specific cases: (1) AOT between V200 and V102 is similar, (2) AOT is much lower for V200 and V102, and in both versions the AOT range in the image is small, and (3) the AOT is very different between both versions, has a large range in V102, but a small range in V200. In the first case, the FAPAR is very similar between V200 and V102. Their difference is almost the same for 10 m and 20 m resolution FAPAR. A large difference in AOT as in case (2) results in a non-linear relationship with not much scatter. The impact is much larger for FAPAR_10M than for FAPAR_20M. The third case results in large scatter between the 2 versions, which is most pronounced for FAPAR_10M. These results suggest that AOT differences between the two versions have a much larger impact on the FAPAR_10M. This is probably caused by the use of only 3 bands in the estimation, of which 2 bands have shorter wavelengths and are thus more affected by AOT. The FAPAR_20M is based on 8 bands that include also two SWIR bands that are less perturbed by AOT (see also [RD-3] for more details). Hence, the FAPAR_10M is more sensitive to accurate AOT estimation.

It should be noted that the inter-comparison merely demonstrates differences and analyses their causes, but it does not indicate which data set is more accurate.

Figure 6: Comparison of scatterplots for 3 dates with variable difference in AOT estimated in V200 and V102 (columns). Top row: FAPAR_10M, middle row: FAPAR_20M, bottom row: AOT.

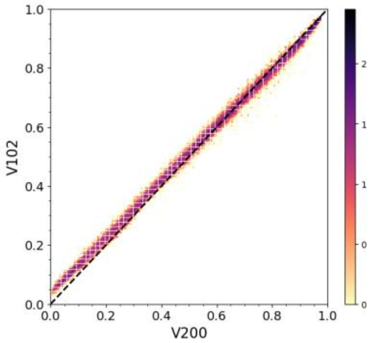
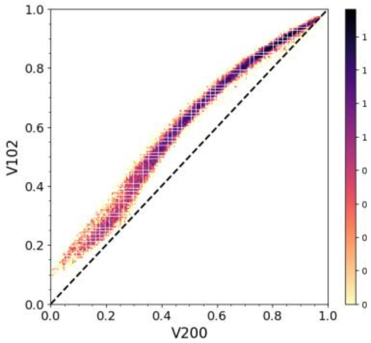
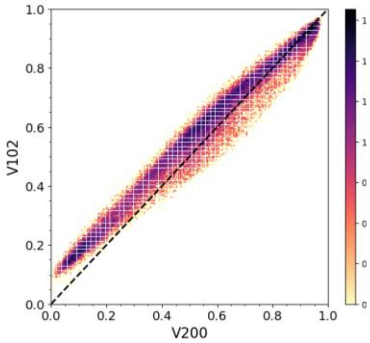
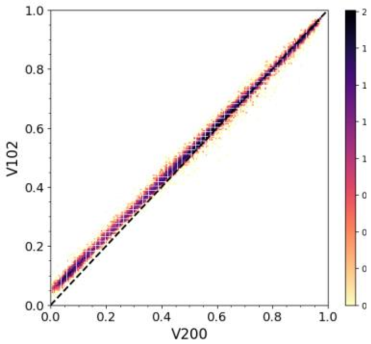
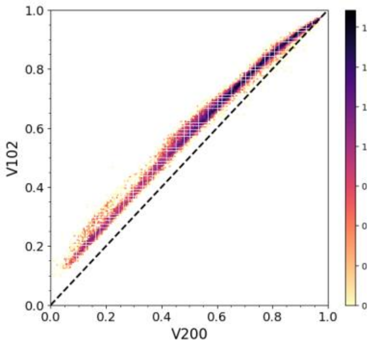
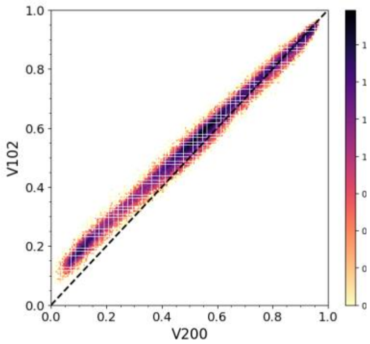
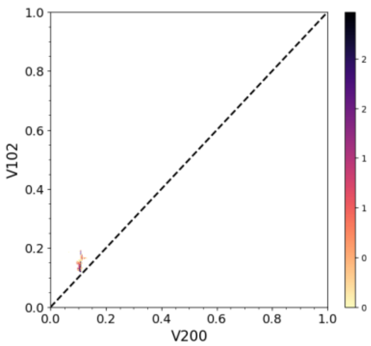
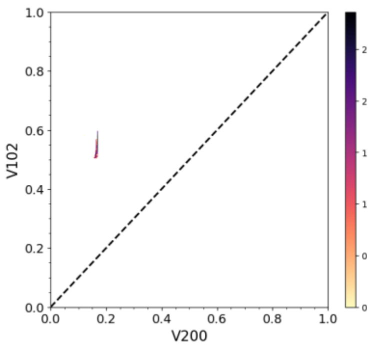
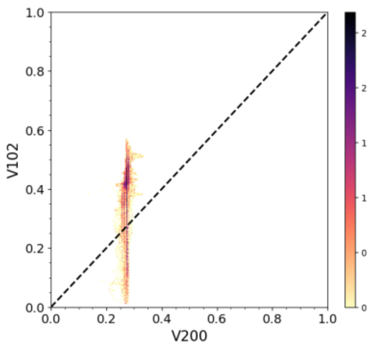
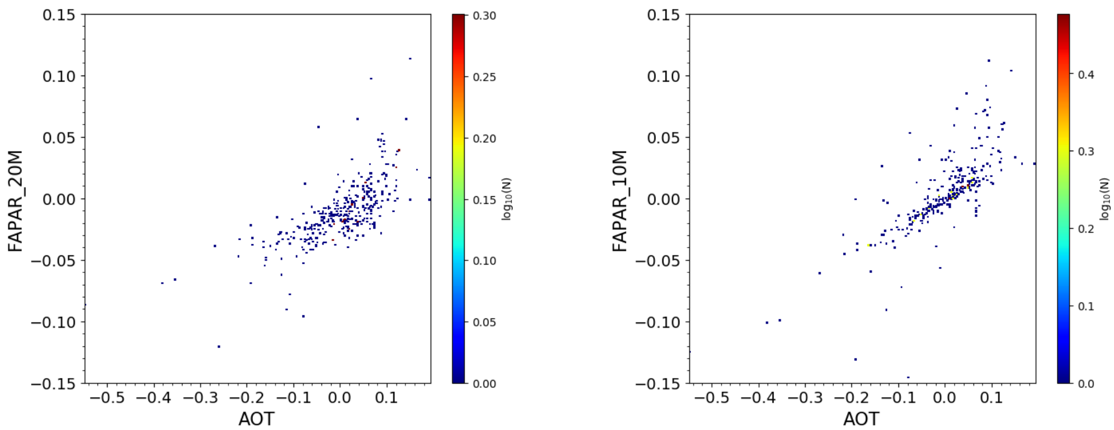
	Small difference in AOT (23/04/2018)	**Large difference in AOT (28/05/2016)**	**Large range of difference in AOT (21/04/2018)**
Scatterplot FAPAR_10M			
Scatterplot FAPAR_20M			
Scatterplot AOT			

Figure 7 shows the scatterplot between the MBE calculated between V200 and V102 (MBE=V200-V102) of AOT and FAPAR for both resolutions. It demonstrates that if the AOT of V200 is lower than V102 (MBE AOT < 0), then the MBE of FAPAR is also lower in V200 (MBE FAPAR < 0). This relationship is stronger for FAPAR_10M.

Figure 7: Scatterplot between Mean Bias Error between V200 and V102 of AOT and FAPAR_20M (left) and FAPAR_10M (right). MBE = V200-V102.

FAPAR_20M	FAPAR_10M



3.2 Histograms and bias histograms

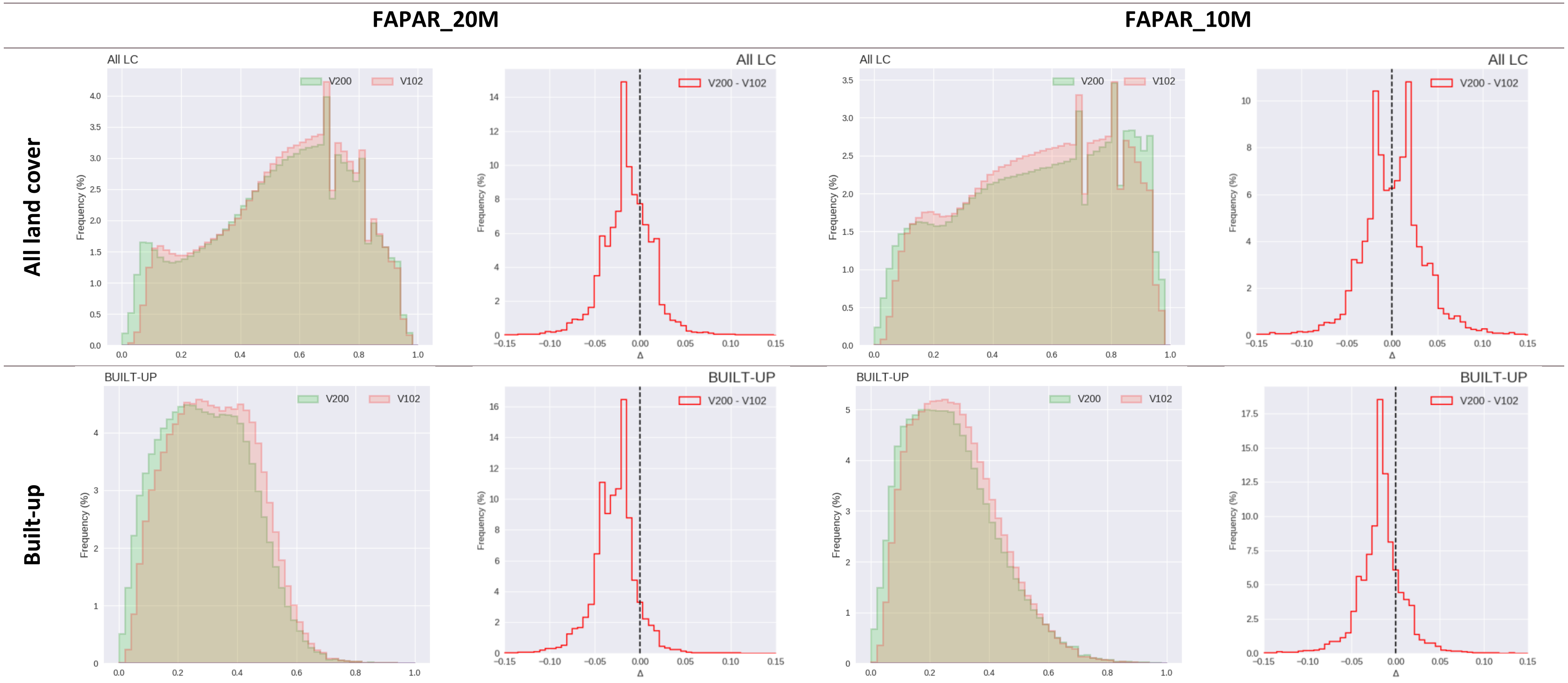
Histograms and bias histograms for V200 and V102 FAPAR_20M and FAPAR_10M are shown in Figure 8 for all land cover in 31UFS and for the 4 major land cover classes. The bias is V200 minus V102, thus if the peak of the bias is negative, then V102 has overall higher values compared to V200.

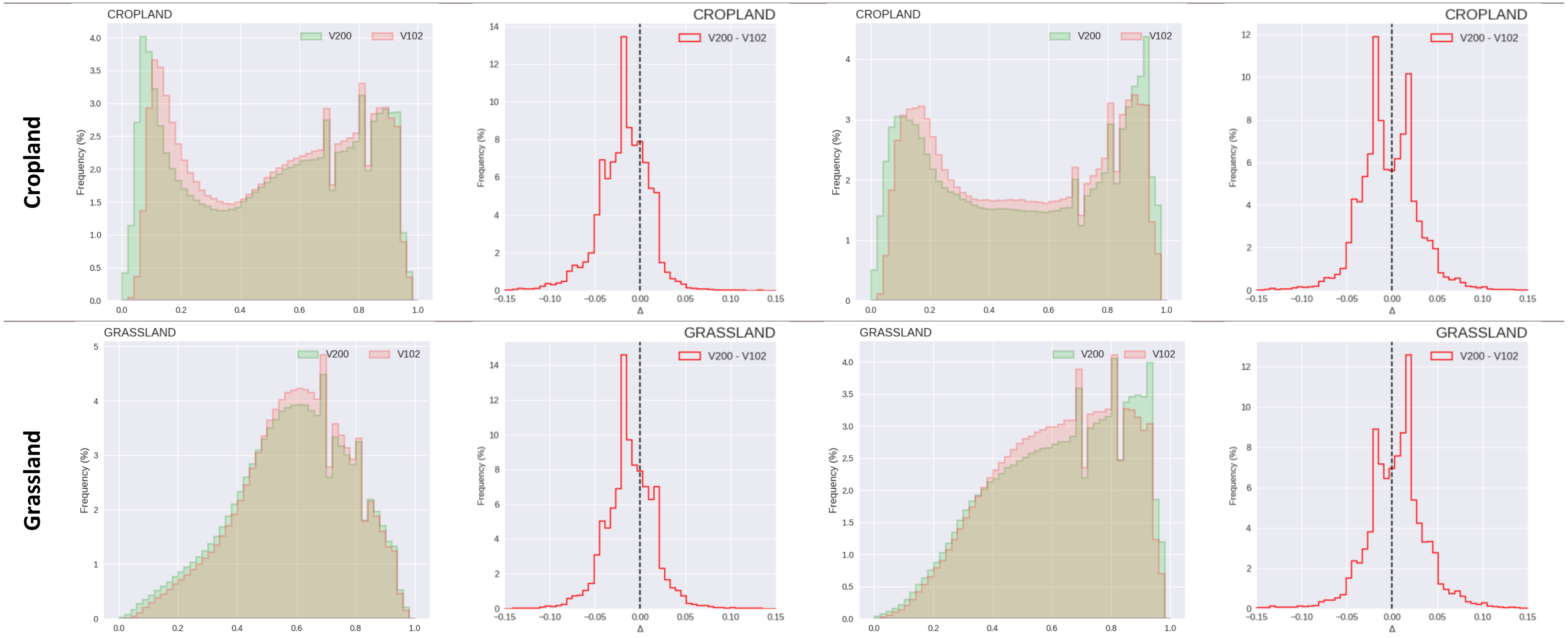
Looking at all land cover, the range of the FAPAR V200 values is higher compared to V102. For the 20M data, this is mainly because the range is extended to lower values, whereas for 10M FAPAR, also more higher values are present. In the bias histogram of FAPAR_20M, the peak is around -0.018, with 81% of the pixels between [-0.051, 0.015] , and 95% between [-0.075, 0.039]. For FAPAR_10M, there are two peaks of same magnitude in the bias histograms, one around -0.018 and one around 0.018, with 79% of the pixels between [-0.039, 0.039] and 94.5% between [-0.069, 0.069].

When looking at the results per land cover class, the following is observed:

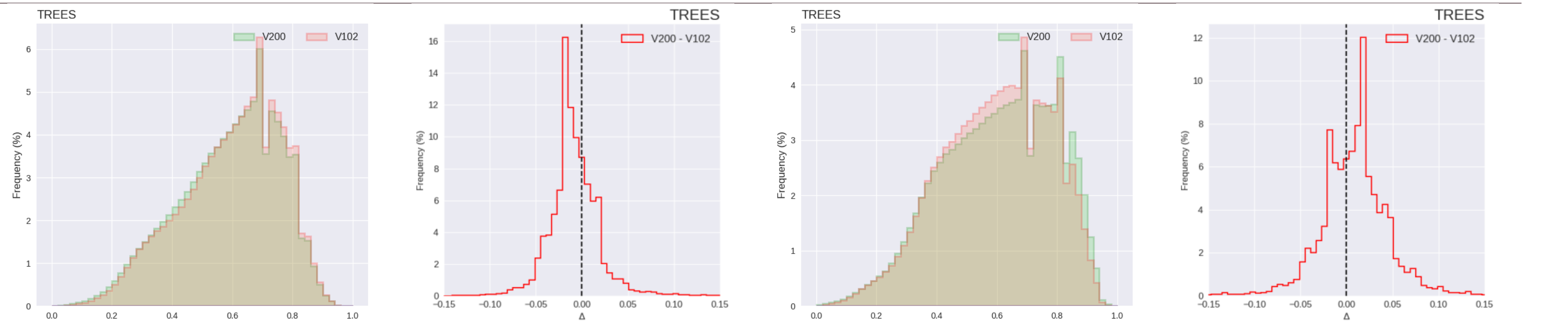
- 'Built-up': there is a clear shift to lower FAPAR values for both resolutions. The shift is most pronounced for FAPAR_20M. For FAPAR_10M, the shift is only towards lower values (only 1 peak in bias histogram).
- 'Cropland': the results of the bias histogram are very similar to the 'all land cover' results, i.e. mainly a shift to lower values for FAPAR_20M with a negative peak around -0.018 in the bias histogram, and two peaks around -0.018 and 0.018 for FAPAR_10M, with the latter peak having a lower magnitude.
- 'Grassland': the peak of the bias histogram for FAPAR_20M is also negative and around -0.018, but the distribution is different than for 'all land cover', with slightly more pixels having a positive bias. For FAPAR_10M, the two peaks are again located around -0.018 and 0.018, but the latter has a higher magnitude. For grasslands, the FAPAR_10M V200 values range is thus predominantly extended to the higher values, which is opposite to the FAPAR_20M.
- 'Tree cover': the results look very similar as those of 'grassland' with even a larger disagreement of the FAPAR_20M and FAPAR_10M in their difference between V200 and V102. FAPAR_20M shows an overall negative bias, and FAPAR_10M an predominant positive bias (V102 smaller than V200).

Figure 8: Histograms and bias histograms for V200 and V102 FAPAR_20M and FAPAR_10M products over the period 01/07/2015 – 26/03/2020. Bias is calculated as V200-V102.




Cropland
CROPLAND
V200
V102
Frequency (%)
CROPLAND
V200 - V102
Frequency (%)
Δ
CROPLAND
V200
V102
Frequency (%)
CROPLAND
V200 - V102
Frequency (%)
Δ
Grassland
GRASSLAND
V200
V102
Frequency (%)
GRASSLAND
V200 - V102
Frequency (%)
Δ
GRASSLAND
V200
V102
Frequency (%)
GRASSLAND
V200 - V102
Frequency (%)
Δ

Tree cover


TREES
V200
V102
Frequency (%)
TREES
V200 - V102
Frequency (%)
Δ
TREES
V200
V102
Frequency (%)
TREES
V200 - V102
Frequency (%)
Δ

The above results indicate that FAPAR_20M shows approximately the same behaviour between V200 and V102 irrespective of the land cover class. This is not the case for FAPAR_10M, where areas with less vegetation cover exhibit mainly a negative bias (V200 is lower than V102), and opposite for areas with high vegetation cover (V200 higher than V102).

3.3 Temporal analysis of agreement and differences

This section describes the result of the temporal analysis. Per date, the FAPAR image of V200 is compared to that of V102 according to the sampling strategy described in section 2.2.2. For each date, a number of statistics is calculated and these are presented as temporal plots. This is done for all land cover types and per biome and presented for MBE in Figure 9 and for RMSE in Figure 10.

For the majority of the dates, the MBE of FAPAR 20M and 10M varies between -0.05 and 0.05. There are more dates outside this range for FAPAR_10M. The MBE range and its magnitude is similar among land cover types, except for 'built-up' the MBE is lower and ranges between [-0.08, 0.02].

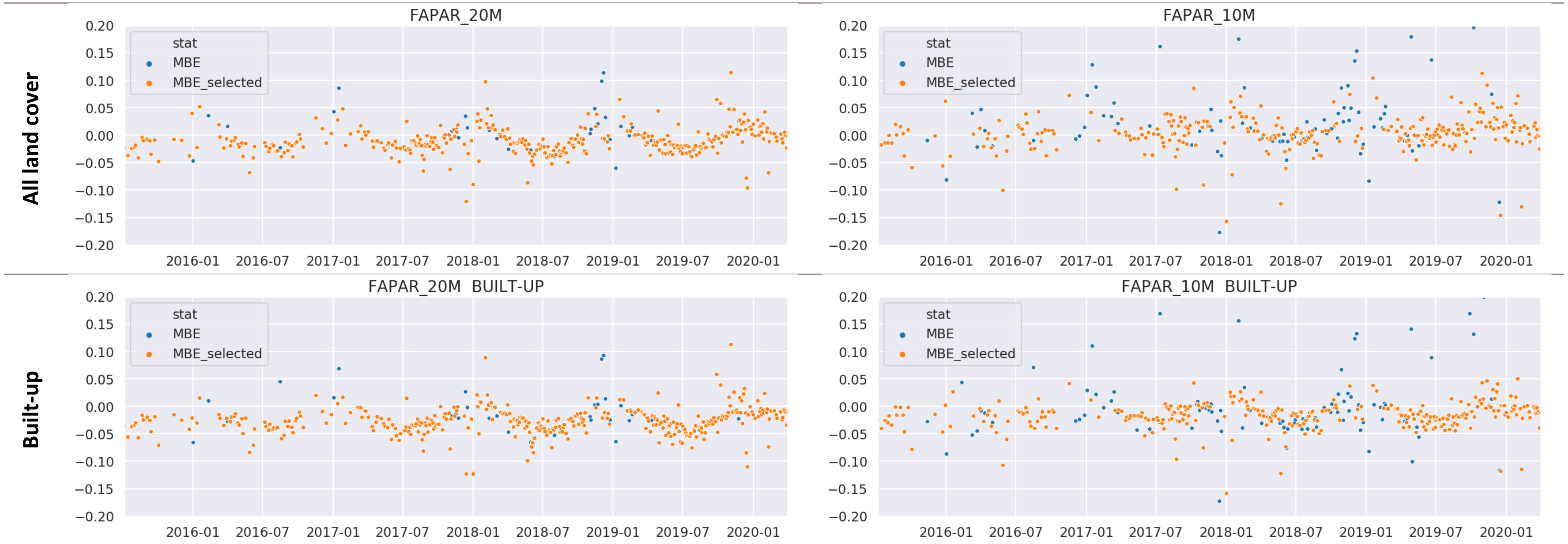
The MBE shows a seasonal pattern, with positive MBE in winter and negative MBE in summer. This is the case for all land cover types. When looking at the temporal evolution of the MBE between the AOT V200 and V102, the same seasonality is observed. The majority of the AOT MBE ranges between [-0.15, 0.15], with positive values in winter and negative values in summer (see [RD-3]). This is consistent with the analysis shown in Figure 7.

The RMSE between V200 and V102 of FAPAR 20M and 10M varies roughly between [0.02, 0.07]. For FAPAR_20M, the magnitude of the RMSE is similar among land cover types, whereas for FAPAR_10M the RMSE range increases with more vegetation cover. There is a substantial amount of dates with RMSE above 0.07. For FAPAR_20M, there are dates for which the RMSE is up to 0.15 and for FAPAR_10M even up to 0.25. Most of these dates are cloudy (small sample size) or the V102 data have been manually corrected for a geolocation shift.

A weak seasonal pattern is also visible in the RMSE temporal plots. This is the most expressed for 'built-up' in FAPAR_20M.

Figure 9: Temporal plots of Mean Bias Error (MBE) between V200 and V102 of FAPAR_20M (left) and FAPAR_10M (right) for all land cover types (top row) and per biome (next rows). MBE are all dates and MBE_selected are all dates with sufficient sample size and where a geometric shift exists between V200 and V102.

	FAPAR_20M	**FAPAR_10M**



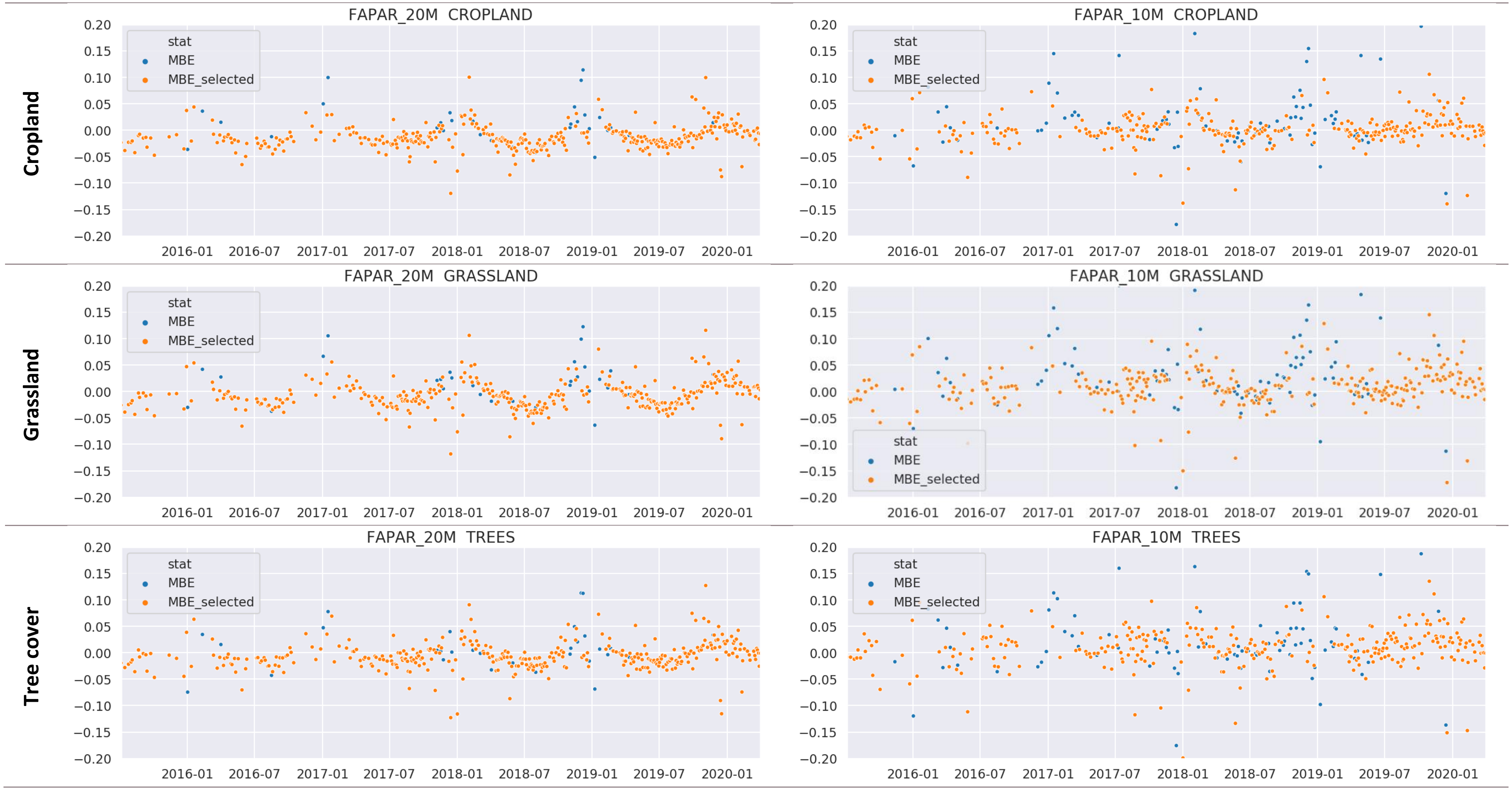
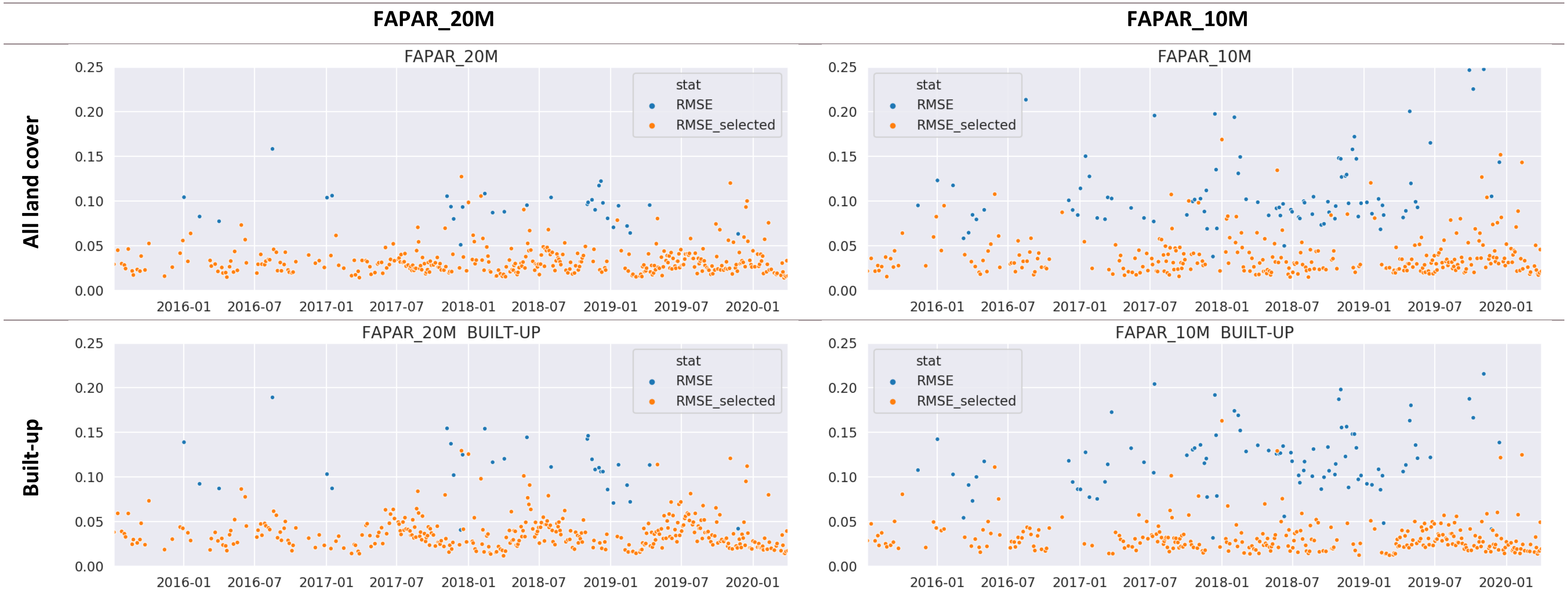
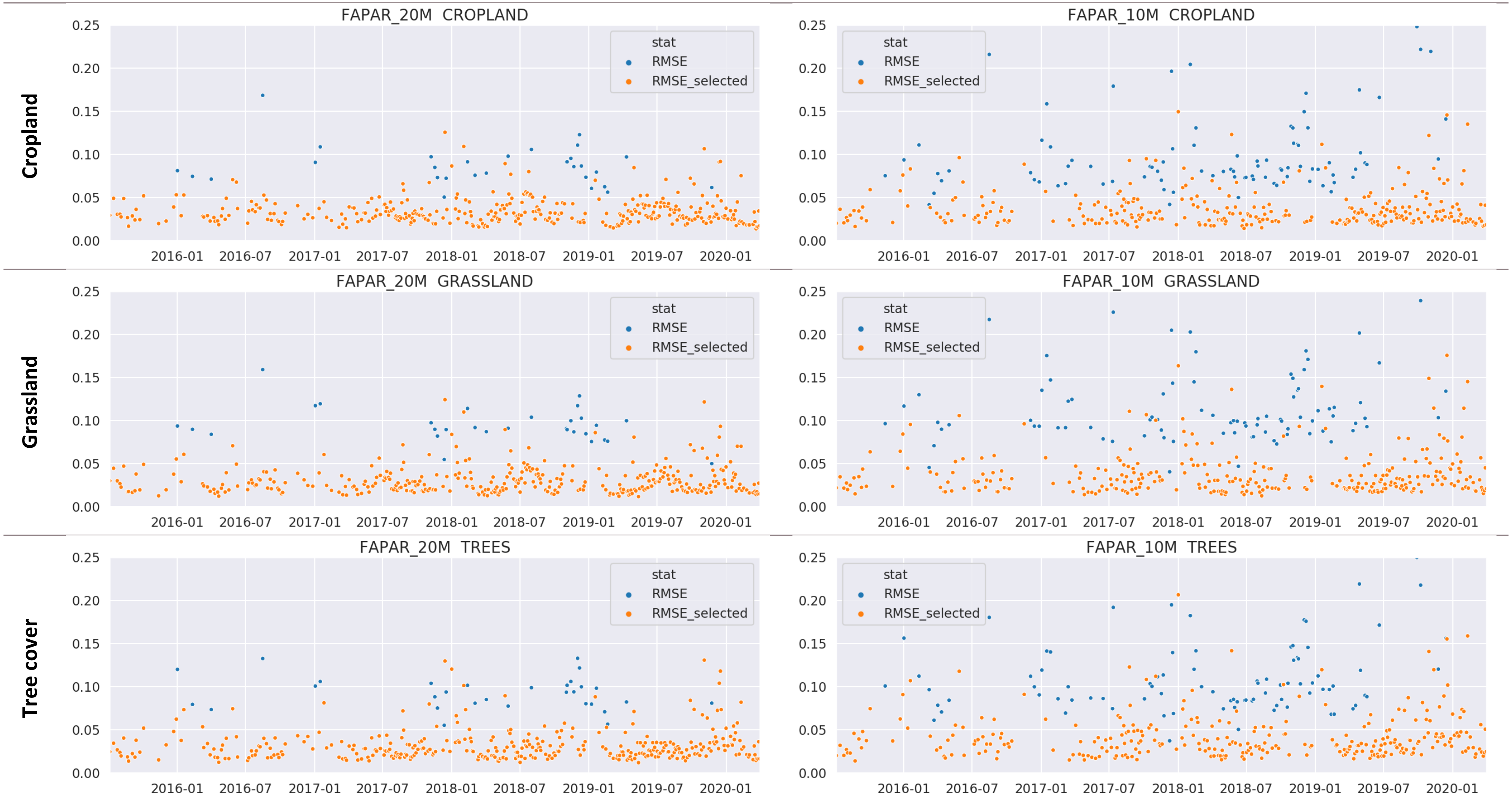

FAPAR_20M CROPLAND
FAPAR_10M CROPLAND
FAPAR_20M GRASSLAND
FAPAR_10M GRASSLAND
FAPAR_20M TREES
FAPAR_10M TREES
Cropland
Grassland
Tree cover
stat
MBE
MBE_selected
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.20
2016-01
2016-07
2017-01
2017-07
2018-01
2018-07
2019-01
2019-07
2020-01

Figure 10: Temporal plots of Root Mean Squared Error (RMSE) between V200 and V102 of FAPAR_20M (left) and FAPAR_10M (right) for all land cover types (top row) and per biome (next rows). RMSE are all dates and RMSE_selected are all dates with sufficient sample size and where a geometric shift exists between V200 and V102.




Cropland
FAPAR_20M CROPLAND
FAPAR_10M CROPLAND
Grassland
FAPAR_20M GRASSLAND
FAPAR_10M GRASSLAND
Tree cover
FAPAR_20M TREES
FAPAR_10M TREES
stat
RMSE
RMSE_selected
0.00
0.05
0.10
0.15
0.20
0.25
2016-01
2016-07
2017-01
2017-07
2018-01
2018-07
2019-01
2019-07
2020-01

3.4 Spatial patterns of differences

The spatial patterns of the per-pixel MBE, RMSE, and systematic and random differences between V200 and V102 FAPAR_20M and 10M are shown in Figure 11.

A number of patterns can be discerned in the spatial plots of all the metrics for both the 20 m and 10 m FAPAR. The differences are largest in the southeast corner, which is an area that is characterized by a more pronounced relief. Similarly, the river valleys also have larger differences between V200 and V102. This is probably because in V200 Sen2Cor applies a topographic correction on the images which results in brighter or less bright slopes. Such a correction was not done in V102. Although differences can be locally large, the majority shows moderate differences.

The MBE is negative for most pixels in 31UFS for FAPAR_20M, whereas for FAPAR_10M it is a mixture of negative and positive values. This is in agreement with the bias histograms (see Figure 8).

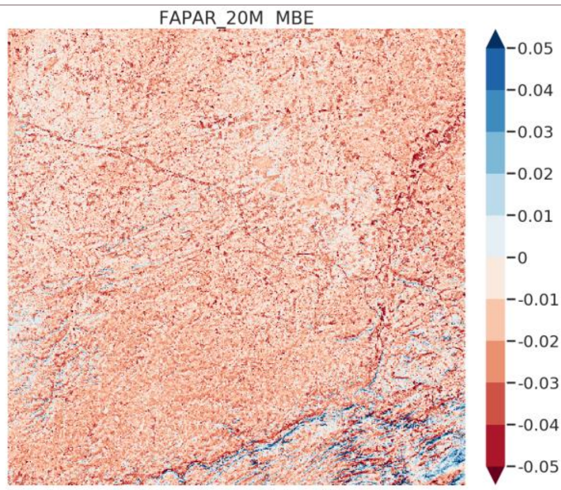
The RMSE is higher for FAPAR_10M than for FAPAR_20M. The spatial patterns are also slightly different between the resolutions. In the FAPAR_10M a number of patches with very low RMSE can be found. Comparing these with the land cover map (Figure 2) these are large grassland areas and built-up areas. The river valleys are also les discernable in the RMSE between the 10M FAPAR versions.

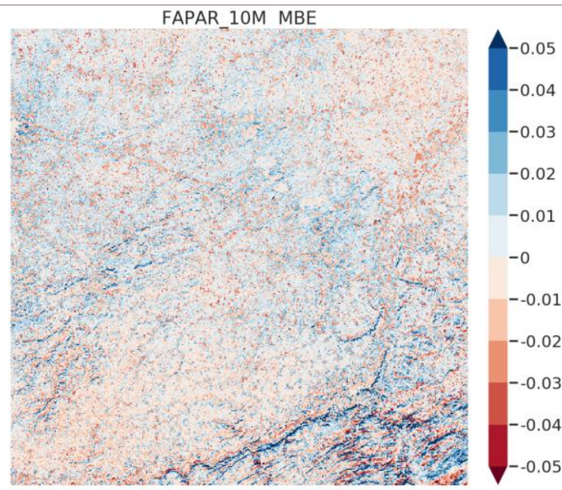
When looking at the split between systematic and random differences, it is observed that the two FAPAR_10M versions mainly differ from random differences rather than systematic ones. For FAPAR_20M this is almost equally distributed for some regions that belong to the 'cropland' or 'built-up' (see Figure 2). This is in line with the results of the overall statistics in Table 5.

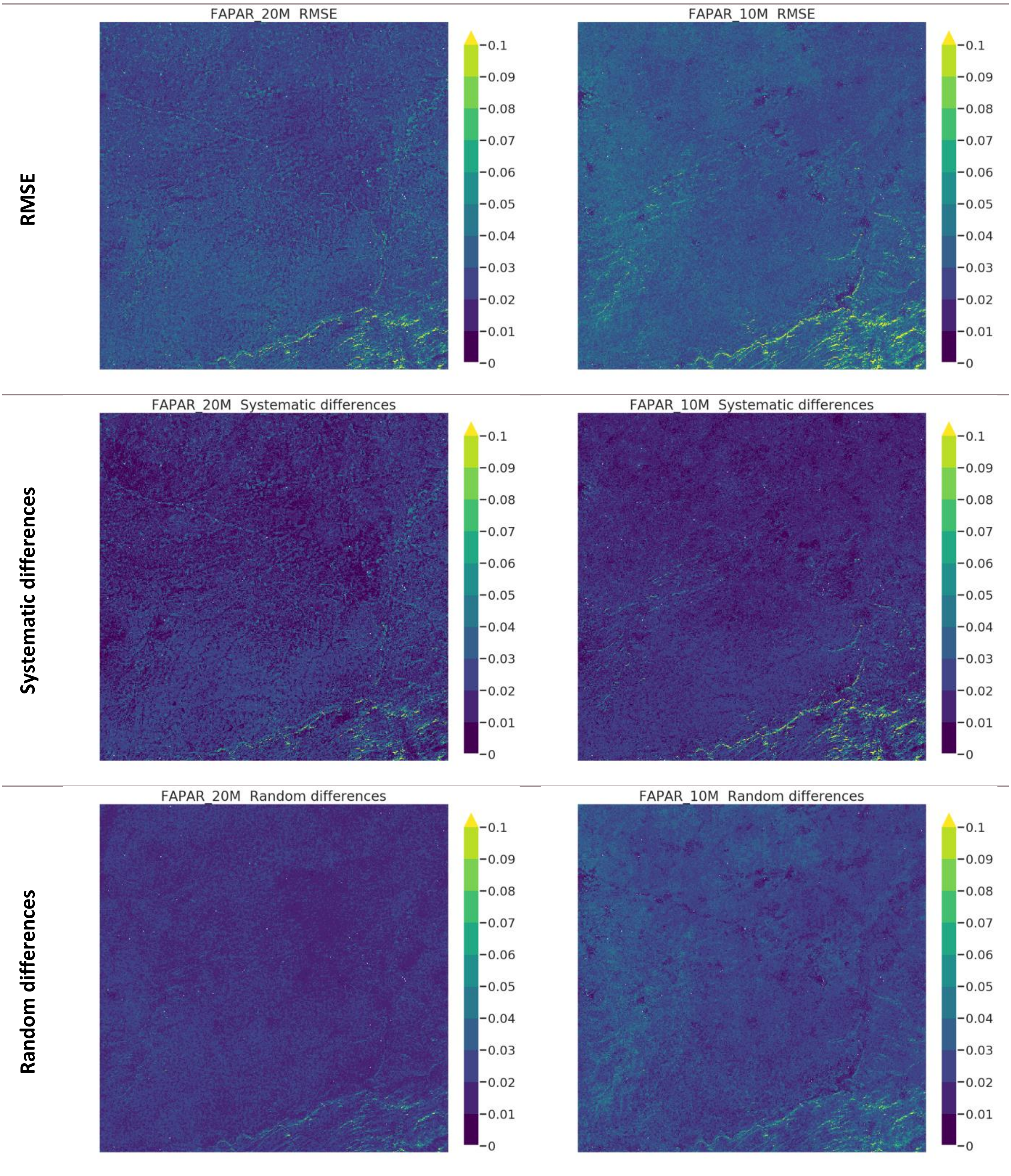
Figure 11: Spatial pattern of the per-pixel Mean Bias Error (MBE), Root Mean Squared Error (RMSE), Systematic difference and Random difference between V200 and V201 of FAPAR_20M (left) and FAPAR_10M (right).

FAPAR_20M | **FAPAR_10M**

MBE






RMSE
Systematic differences
Random differences
FAPAR_20M RMSE
FAPAR_10M RMSE
FAPAR_20M Systematic differences
FAPAR_10M Systematic differences
FAPAR_20M Random differences
FAPAR_10M Random differences
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0

3.5 Conclusion for FAPAR_20M and FAPAR_10M

The full time series until 31/03/2020 of tile 31UFS of the TERRASCOPE FAPAR_20M and FAPAR_10m V200 was compared to the previous version V102 in order to demonstrate the agreement and the differences.
Overall, FAPAR V200 and V102 agree linearly, and with a reasonable amount of scattering. The agreement is better for the FAPAR_20M, which has a small negative bias meaning that V200 FAPAR values are lower than V102. The range of FAPAR_20M values is larger towards lower values.
The FAPAR_10M also exhibits a linear relationship between V200 and V102, with a bias depending on land cover type: negative for 'built-up' and 'cropland', positive for 'grassland' and 'tree cover'. There is more scatter compared to FAPAR_20M in the difference.

A large influence of the difference in AOT was observed. Large differences in AOT used in V200 and V102 are responsible for:

- Non-linear relationship between the two versions
- More scatter in the relationship
- Seasonal pattern in difference

The spatial analysis demonstrated that the topographic correction used in the atmospheric correction of V200 has an impact on the data: larger differences for areas with more topography. As the analysis focused on the differences on the two versions, it is not possible to judge which version is more accurate. We can assume that the topographic correction should be an improvement in the processing.

The use of per-pixel Sun angles instead of mean Sun angles per image in the retrieval of FAPAR is not discernable from the results.

It should be mentioned that the impact of geometric errors was not assessed. This is usually expressed in random differences rather than systematic ones. Large differences between V200 and V102 were observed for those dates for which a geolocation shift was applied in V102 and not in V200. The difference metrics were significantly larger than the ones observed for the images that have the same geolocation in both versions.

It is advised not to mix the two version of the FAPAR in the same analysis, because this could lead to erroneous conclusions.

4 FCOVER 10M and 20M: RESULTS

This chapter discusses the results of the inter-comparison between V200 and V102 for the FCOVER products at 10 m and 20 m resolution.

4.1 Overall statistical analysis

The Mean Bias Difference (MBE), Mean Absolute Difference (MAE), Root Mean Squared Difference (RMSE), which is further split into random and systematic differences, were calculated on the overall sample (see section 2.2.1) for all land cover classes and per land cover class. The results are summarized in Table 6.

Comparing the statistics between the two resolutions, it is observed that the overall MBE of FCOVER for 20 m is negative (-0.010) and small positive (0.004) for the 10M resolution FCOVER. The MBE is negative for all land cover classes for FCOVER_20M, but only for 'built-up' for FCOVER_10M. All other land cover classes have positive MBE for FCOVER_10M. The MAE and RMSE are both higher for FCOVER_10M than for FCOVER_20M, except for 'built-up'. For FCOVER_20M, the RMSE and MAE are very similar among the land cover classes, opposed to FCOVER_10M that shows more variation in difference among the land cover classes.

For FCOVER_10M, the random differences are always larger than the systematic ones. This is not the case for FCOVER_20M, where for built-up the random difference is smaller than the systematic differences. For both resolutions for 'cropland' and 'grassland', the systematic difference is much lower than the random.

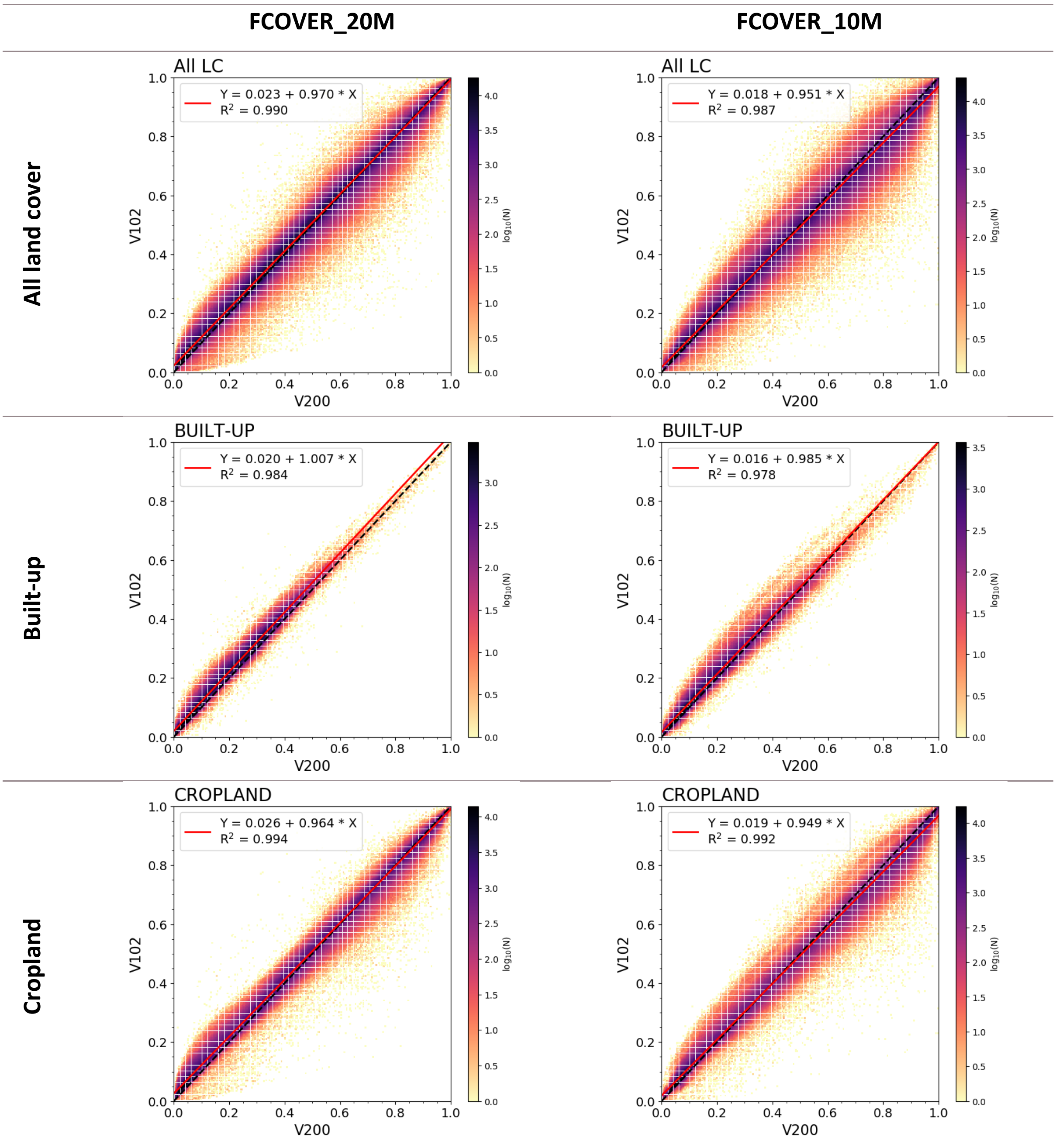
Table 6: Mean Bias Error (MBE), Mean Absolute Error (MAE), Root Mean Squared Difference (RMSD), Random difference and Systematic difference between V200 and V102 of FCOVER_20M and FCOVER_10M for all land cover classes (All LC) or per biome.

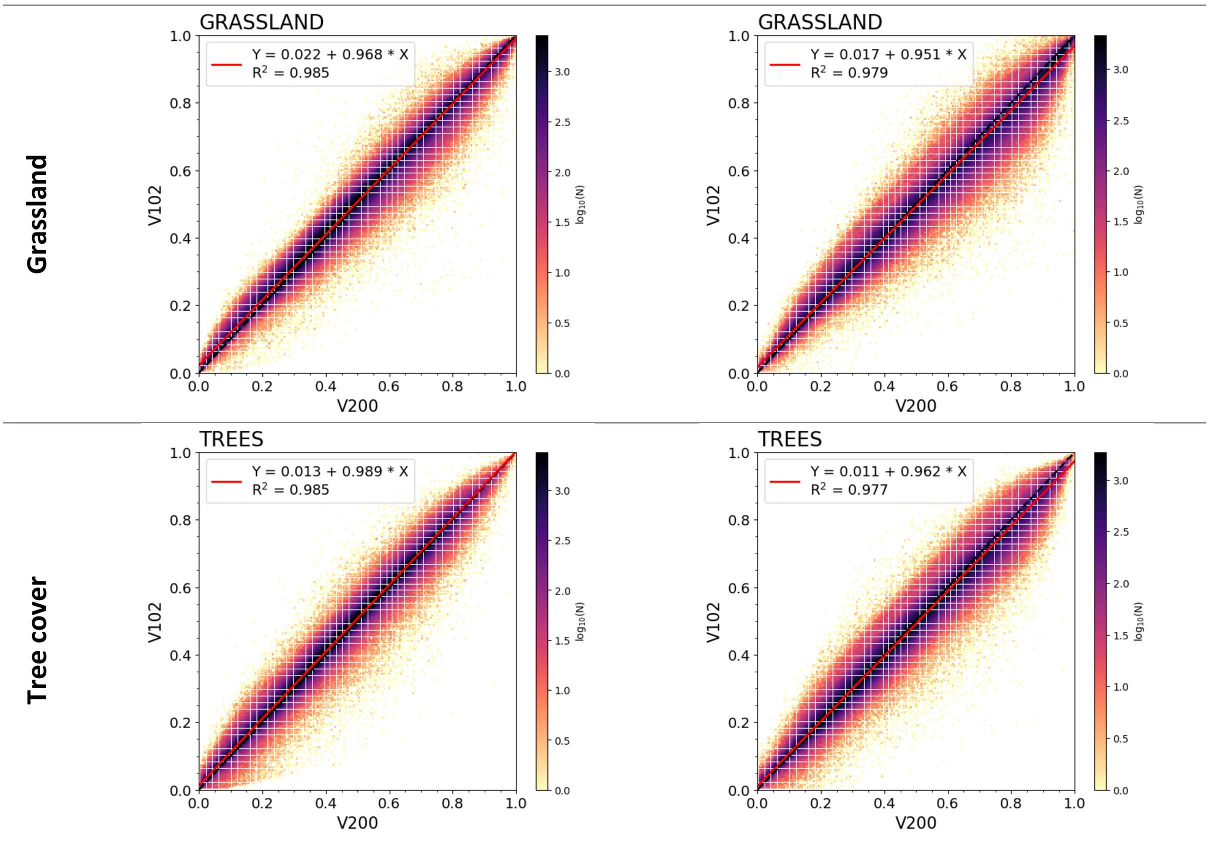
	FCOVER_20M					FCOVER_10M				
	MBE	MAE	RMSE	Rand. diff	Sys. diff.	MBE	MAE	RMSE	Rand. diff	Sys. diff.
All LC	-0.010	0.020	0.027	0.024	0.012	0.004	0.024	0.035	0.032	0.014
BUILT-UP	-0.021	0.022	0.027	0.017	0.021	-0.014	0.018	0.024	0.020	0.014
CROPLAND	-0.008	0.019	0.027	0.025	0.009	0.006	0.024	0.035	0.033	0.011
GRASSLAND	-0.007	0.021	0.028	0.027	0.010	0.009	0.027	0.039	0.036	0.015
TREES	-0.011	0.021	0.027	0.023	0.015	0.004	0.023	0.033	0.028	0.017

Figure 12 shows the corresponding scatterplots with the orthogonal geometric mean regression line and the coefficient of determination. The V200 and V102 are overall linearly related, with FCOVER_20M having less scatter. For all land covers and both resolutions, the V200 FCOVER has more smaller values as is shown by the positive offset of the regression line. For FCOVER_20M, the regression line crosses the 1:1 line at a higher FCOVER value than for FCOVER_10M.

The scatterplot of FCOVER_10M shows seemingly a separate population above the regression line. This is visible in the plot for all land cover, but more pronounced for 'built-up', 'cropland' and 'grassland'. This was further investigated by looking at the scatterplots of single dates. For most of the dates, the scatterplots are narrow, linear and close to the 1:1 line. But if the Aerosol Optical Thickness (AOT) used to correct the data for atmospheric scattering differs largely between V200 and V102, the relationship drifts from the 1:1 line and becomes non-linear or shows more scatter.

Figure 12: Scatterplots between V200 (X) and V102 (Y) FCOVER_20M (left) and FCOVER_10M (right) products for the period 01/07/2015 – 26/03/2020. Red line is the orthogonal geometric mean regression line between the two versions.





Examples of this are shown in Figure 13 for 3 specific cases: (1) AOT between V200 and V102 is similar, (2) AOT is much lower for V200 and V102, in both versions the AOT range in the image is small, and (3) the AOT is very different between both version and has a large range in V102 but a small range in V200. In the first case, the FCOVER is very similar between V200 and V102, except that the 10 m FCOVER is less linearly related between the two versions. Their difference is almost the same for 10 m and 20 m resolution FCOVER. A large difference in AOT as in case (2) results in a clear non-linear relationship with not much scatter. The impact is much larger for FCOVER_10M than for FCOVER_20M. The third case results in large scatter between the 2 versions, which is most pronounced for FCOVER_10M. These results suggest that AOT differences between the two versions have a much larger impact on the FCOVER_10M. This is probably caused by the use of only 3 bands in the retrieval of FCOVER, of which 2 bands have shorter wavelengths and are thus more affected by the difference in AOT. The FCOVER_20M is based on 8 bands that include also two SWIR bands that are less perturbed by AOT differences (see also [RD-3] for more details). Hence, the FCOVER_10M is more sensitive to accurate AOT estimation.
It should be noted that the inter-comparison merely demonstrates differences and analyses their causes, but it does not indicate which data set is more accurate.

Figure 13: Comparison of scatterplots for 3 dates with variable difference in AOT estimated in V200 and V102 (columns). Top row: FCOVER_10M, middle row: FCOVER_20M, bottom row: AOT.

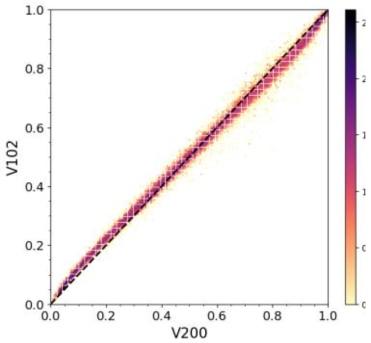
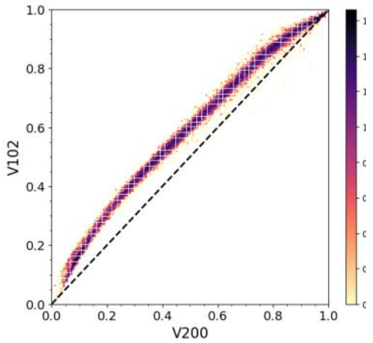
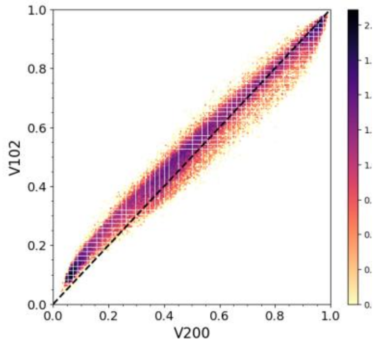
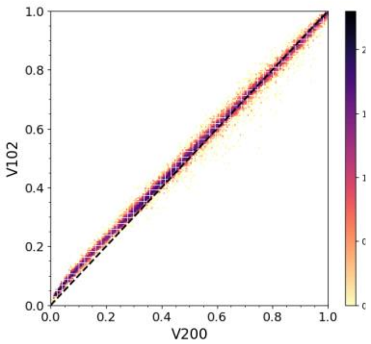
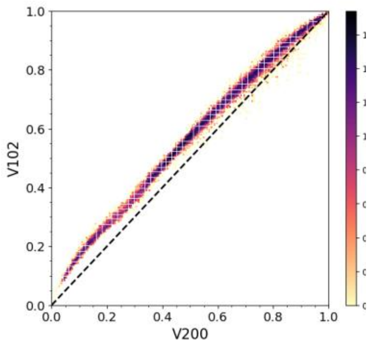
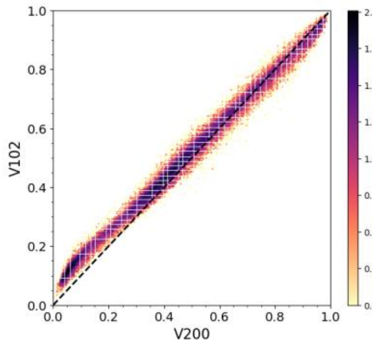
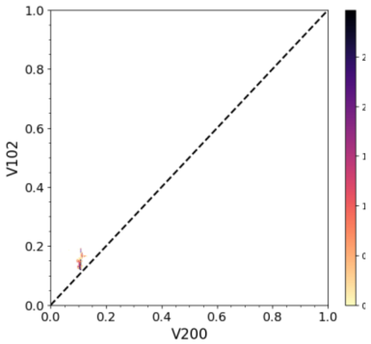
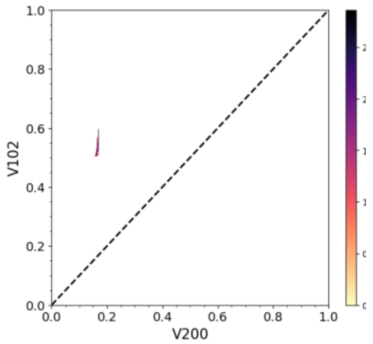
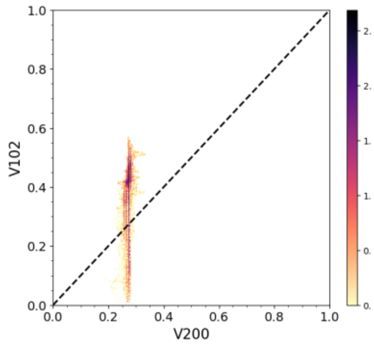
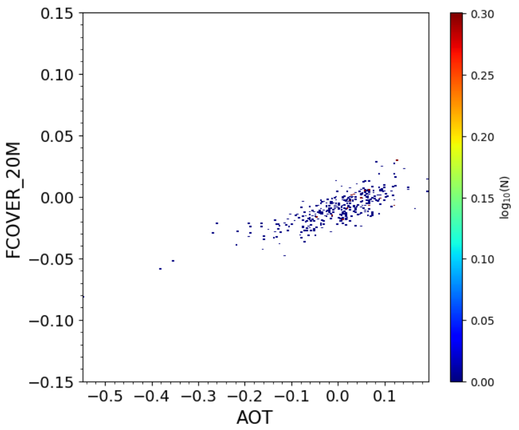
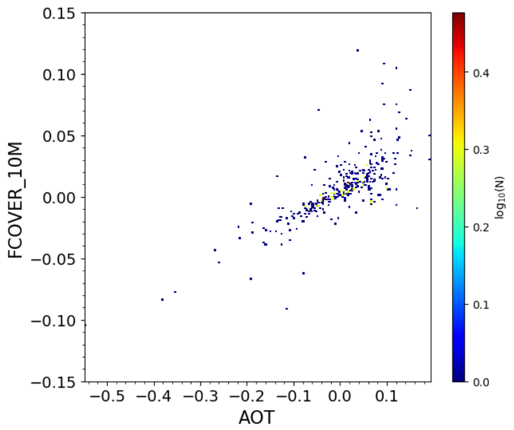
	Small difference in AOT (23/04/2018)	**Large difference in AOT (28/05/2016)**	**Large range of difference in AOT (21/04/2018)**
Scatterplot FCOVER_10M			
Scatterplot FCOVER_20M			
Scatterplot of AOT			

Figure 14 shows the scatterplot between the MBE calculated between V200 and V102 (MBE=V200-V102) of AOT and FCOVER for both resolutions. It demonstrates that if the AOT of V200 is lower than V102 (MBE AOT < 0), then the MBE of FCOVER is also lower in V200 (MBE FCOVER < 0). This relationship is stronger for FCOVER_10M.

Figure 14: Scatterplot between Mean Bias Error between V200 and V102 of AOT and FCOVER_20M (left) and FCOVER_10M (right). MBE = V200-V102.

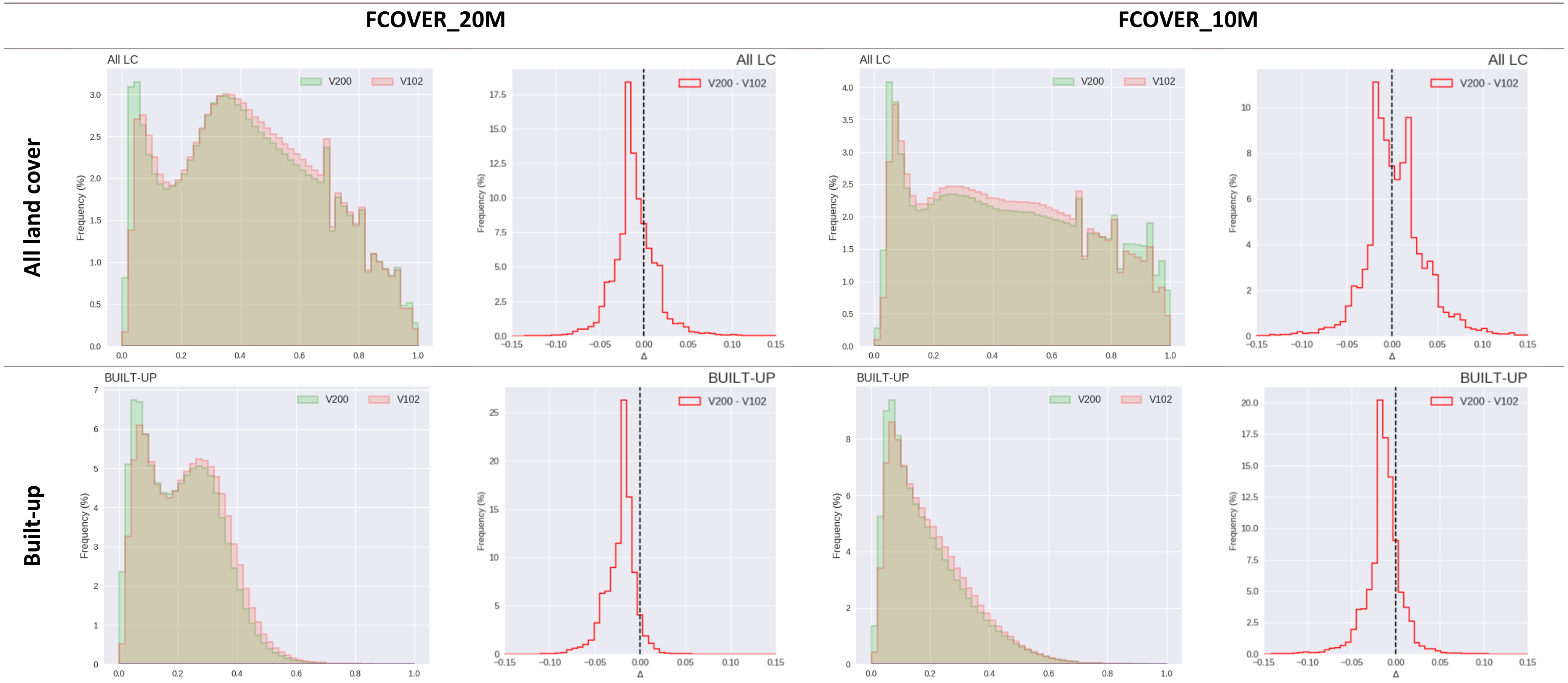
FCOVER_20M	FCOVER_10M
	

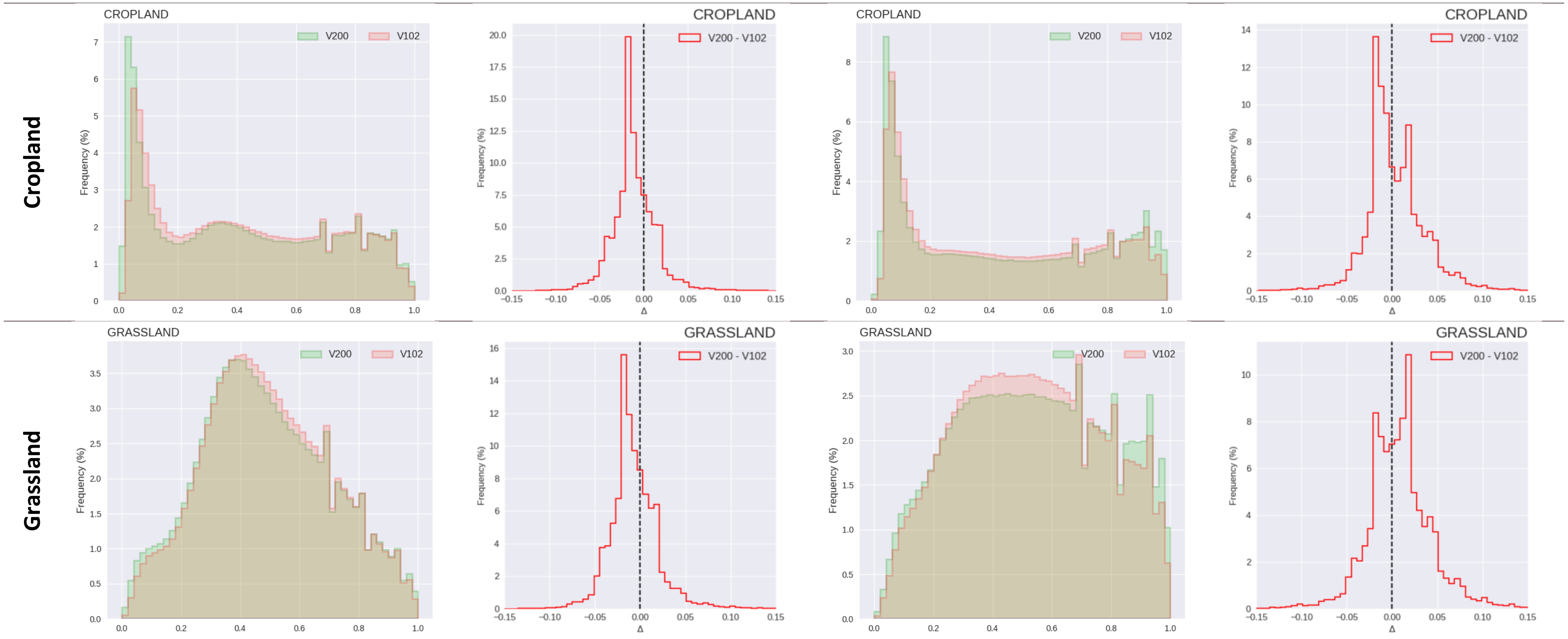
4.2 Histograms and bias histograms

Histograms and bias histograms for V200 and V102 FCOVER_20M and FCOVER_10M are shown in Figure 15 for all land cover in 31UFS and for the 4 major land cover classes. The bias is V200 minus V102, thus if the peak of the bias is negative, then V102 has overall higher values compared to V200. Looking at all land cover, the range of the FCOVER V200 values is similar compared to V102. For the 20M data, more low values are observed, whereas for 10M FCOVER, also more high values are present. In the bias histogram of FCOVER_20M, the peak is around -0.018, with 84% of the pixels between [-0.051, 0.015] , and 94.5% between [-0.069, 0.033]. For FCOVER_10M, there are two peaks of same magnitude in the bias histograms, one around -0.018 and one around 0.018, with 80% of the pixels between [-0.039, 0.039] and 94.5% between [-0.069, 0.069].
When looking at the results per land cover class, the following is observed:

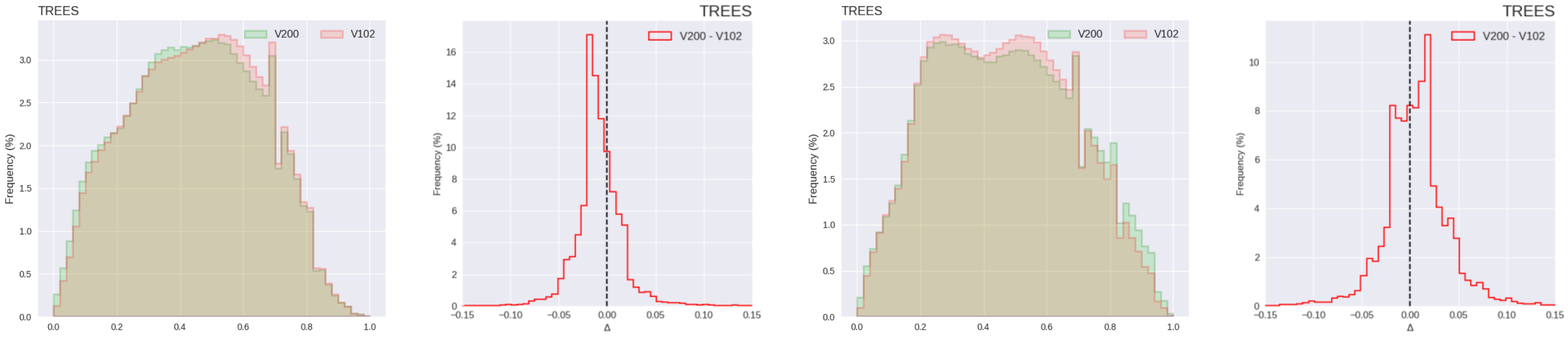
- 'Built-up': there is a clear shift to lower FCOVER values for both resolutions. The shift is slightly more pronounced for FCOVER_20M. For FCOVER_10M, the shift is only towards lower values (only 1 peak in bias histogram).
- 'Cropland': the results of the bias histogram are very similar to the 'all land cover' results, i.e. mainly a shift to lower values for FCOVER_20M with a negative peak around -0.018 in the bias histogram, and two peaks around -0.018 and 0.018 for FCOVER_10M, with the latter peak having a smaller magnitude.
- 'Grassland': the peak of the bias histogram for FCOVER_20M is also negative and around -0.018. For FCOVER_10M, the two peaks are again located around -0.018 and 0.018, but the latter has a higher magnitude. For grasslands, the FCOVER_10M V200 values range is thus in the majority of the pixels extended to the higher values, which is opposite to the FCOVER_20M.

Figure 15: Histograms and bias histograms for V200 and V102 FCOVER_20M and FCOVER_10M products over the period 01/07/2015 – 26/03/2020. Bias is calculated as V200-V102.




Cropland
CROPLAND
V200
V102
Frequency (%)
CROPLAND
V200 - V102
Frequency (%)
Δ
CROPLAND
V200
V102
Frequency (%)
CROPLAND
V200 - V102
Frequency (%)
Δ
Grassland
GRASSLAND
V200
V102
Frequency (%)
GRASSLAND
V200 - V102
Frequency (%)
Δ
GRASSLAND
V200
V102
Frequency (%)
GRASSLAND
V200 - V102
Frequency (%)
Δ

Tree cover


TREES
V200
V102
Frequency (%)
TREES
V200 - V102
Frequency (%)
Δ
TREES
V200
V102
Frequency (%)
TREES
V200 - V102
Frequency (%)
Δ

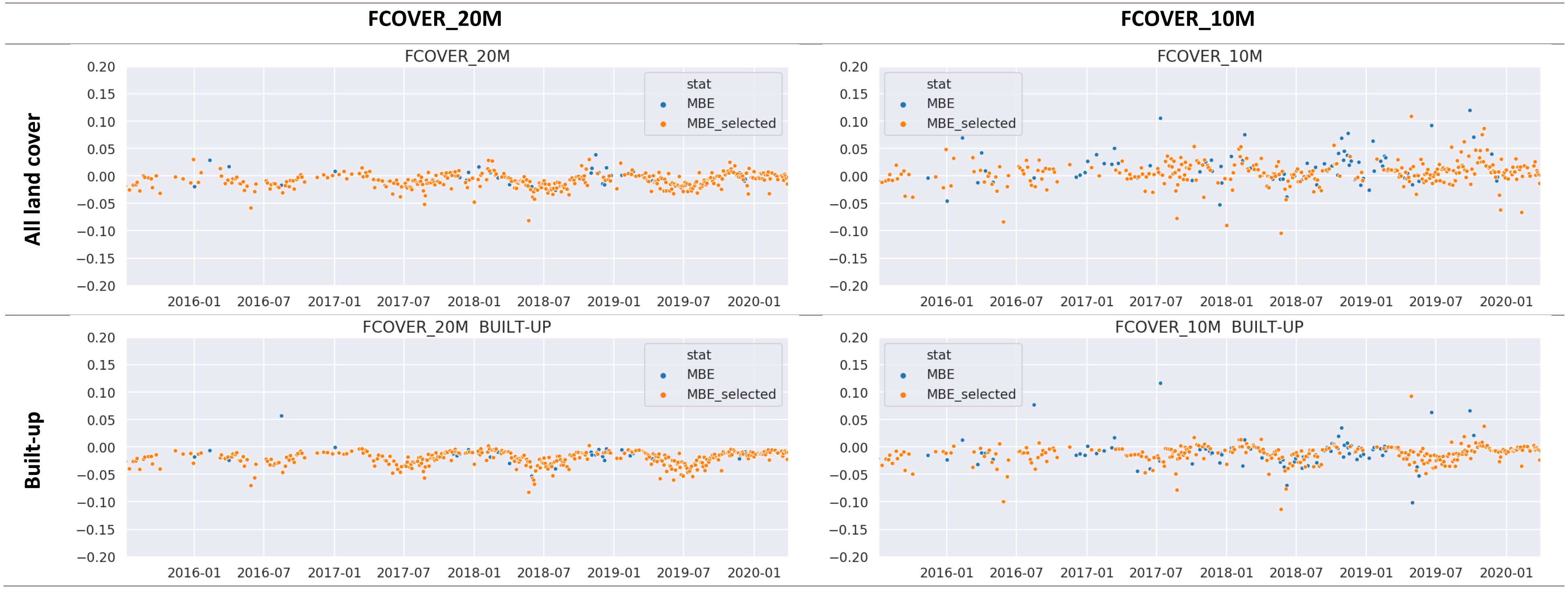
- 'Tree cover': the results look very similar as those of 'grassland' with even a larger disagreement of the FCOVER_20M and FCOVER_10M in their difference between V200 and V102. FCOVER_20M shows an overall negative bias, and FCOVER_10M a predominant positive bias (V102 smaller than V200).

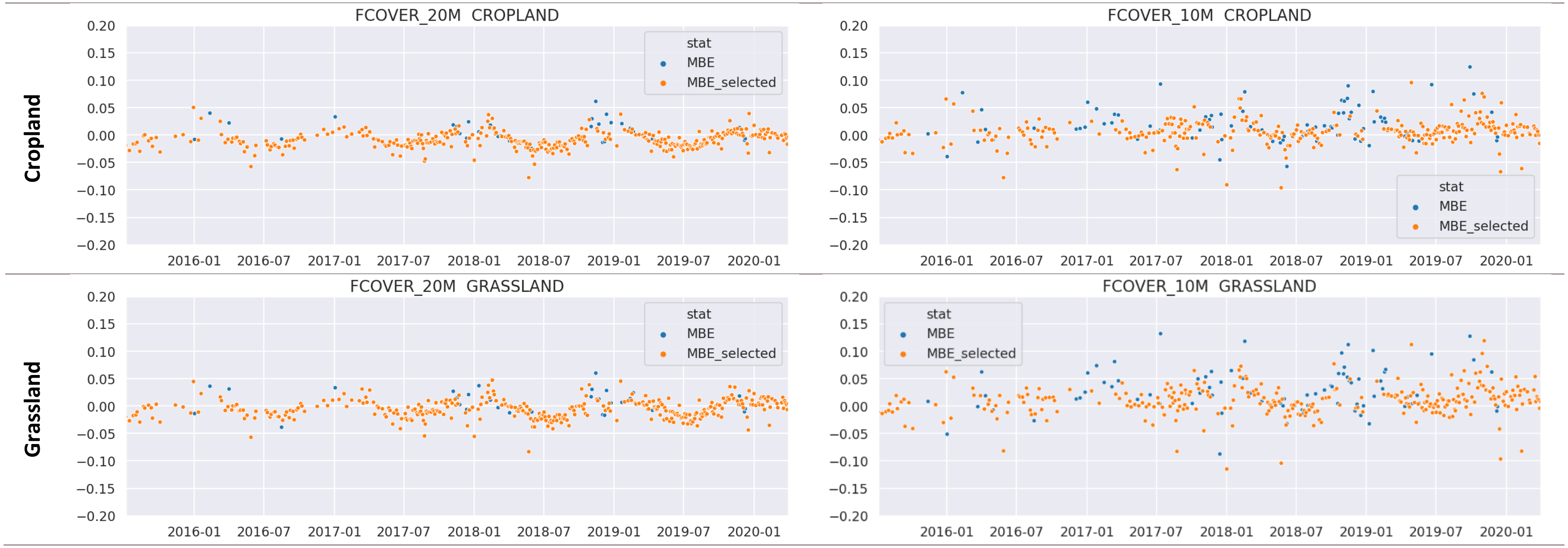
The above results indicate that FCOVER_20M shows approximately the same behaviour between V200 and V102 irrespective of the land cover class. This is not the case for FCOVER_10M, where areas with less vegetation cover exhibit mainly a negative bias (V200 is lower than V102), and opposite for areas with high vegetation cover (V200 higher than V102).

4.3 Temporal analysis of agreement and differences

This section describes the result of the temporal analysis. Per date, the FCOVER image of V200 is compared to that of V102 according to the sampling strategy described in section 2.2.2. For each date, a number of statistics is calculated and these are presented as temporal plots. This is done for all land cover types and per biome and presented for MBE in Figure 16 and for RMSE in Figure 17.
For the majority of the dates, the MBE of FCOVER 20M and 10M varies between -0.05 and 0.05. There are more dates outside this range for FCOVER_10M. The MBE range and its magnitude is similar among land cover types, except for 'built-up' the MBE is lower and ranges between [-0.08, 0.01].
The MBE shows a seasonal pattern, with positive MBE in winter and negative MBE in summer. This is the case for all land cover types, except for 'tree cover'. When looking at the temporal evolution of the MBE between the AOT V200 and V102, the same seasonality is observed (see [RD-3]). The majority of the AOT MBE ranges between [-0.15, 0.15], with positive values in winter and negative values in summer. This is consistent with the analysis shown in Figure 16.
The RMSE between V200 and V102 of FCOVER 20M and 10M varies roughly between [0.01, 0.07]. For FCOVER_20M, the magnitude of the RMSE is similar among land cover types, whereas for FCOVER_10M the RMSE range increases with more vegetation cover. There is a substantial amount of dates with RMSE above 0.07. For FCOVER_20M, there are dates for which the RMSE is up to 0.20 and for FCOVER_10M even up to 0.25. Most of these dates are cloudy (small sample size) or the V102 data have been manually corrected for a geolocation shift.
A weak seasonal pattern is also visible in the RMSE temporal plots. This is the most expressed for 'built-up' at both resolutions.

Figure 16: Temporal plots of Mean Bias Error (MBE) between V200 and V102 of FCOVER_20M (left) and FCOVER_10M (right) for all land cover types (top row) and per biome (next rows). MBE are all dates and MBE_selected are all dates with sufficient sample size and where a geometric shift exists between V200 and V102.




Cropland
FCOVER_20M CROPLAND
FCOVER_10M CROPLAND
Grassland
FCOVER_20M GRASSLAND
FCOVER_10M GRASSLAND
stat
MBE
MBE_selected
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.20
2016-01
2016-07
2017-01
2017-07
2018-01
2018-07
2019-01
2019-07
2020-01

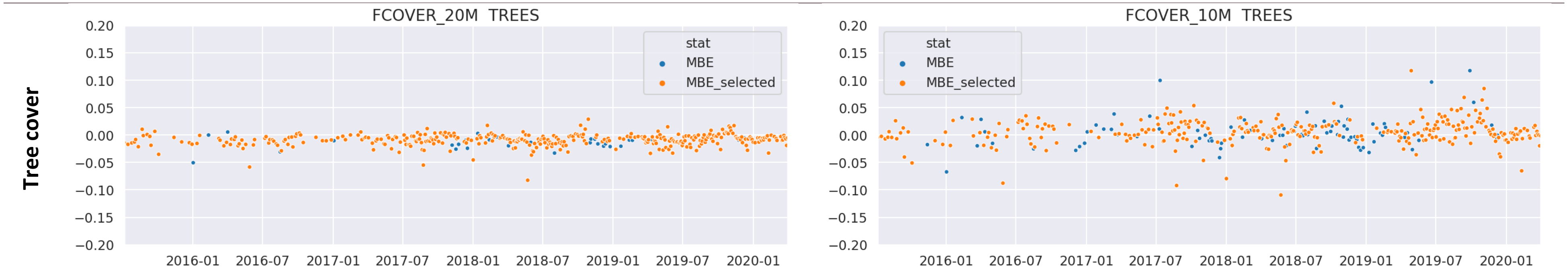
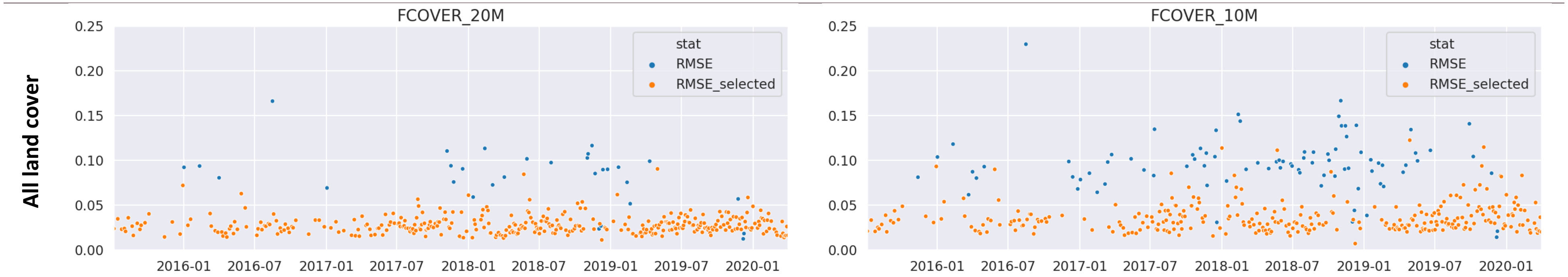
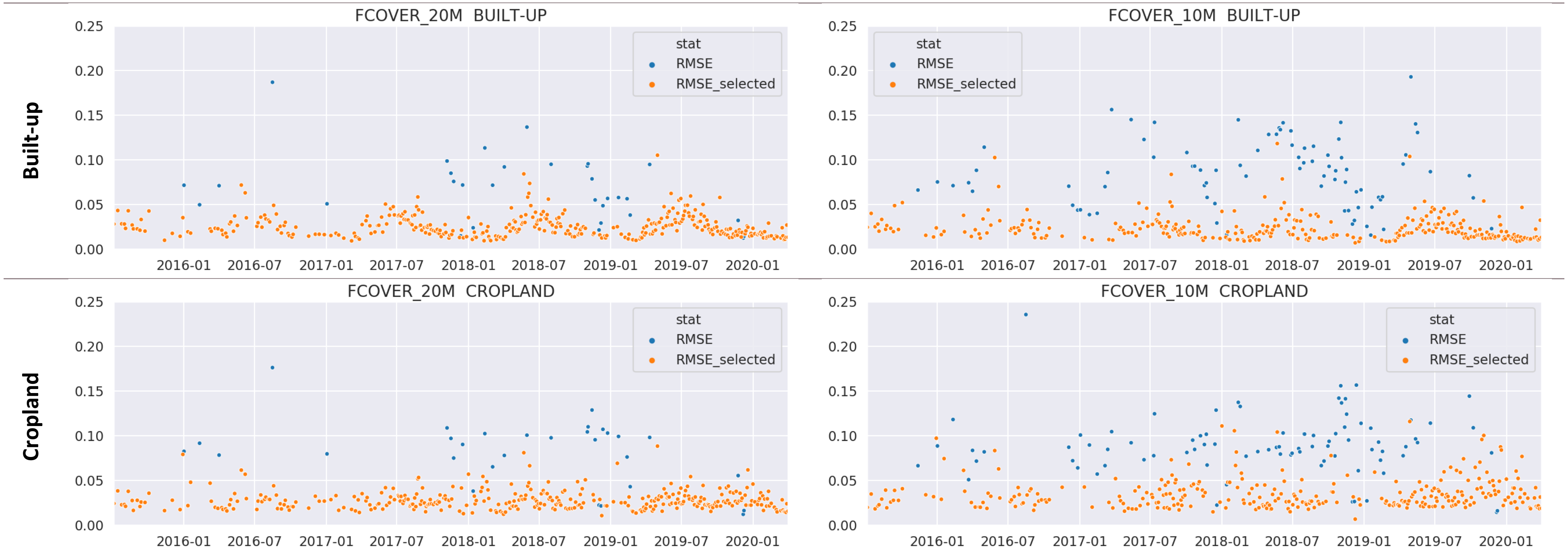
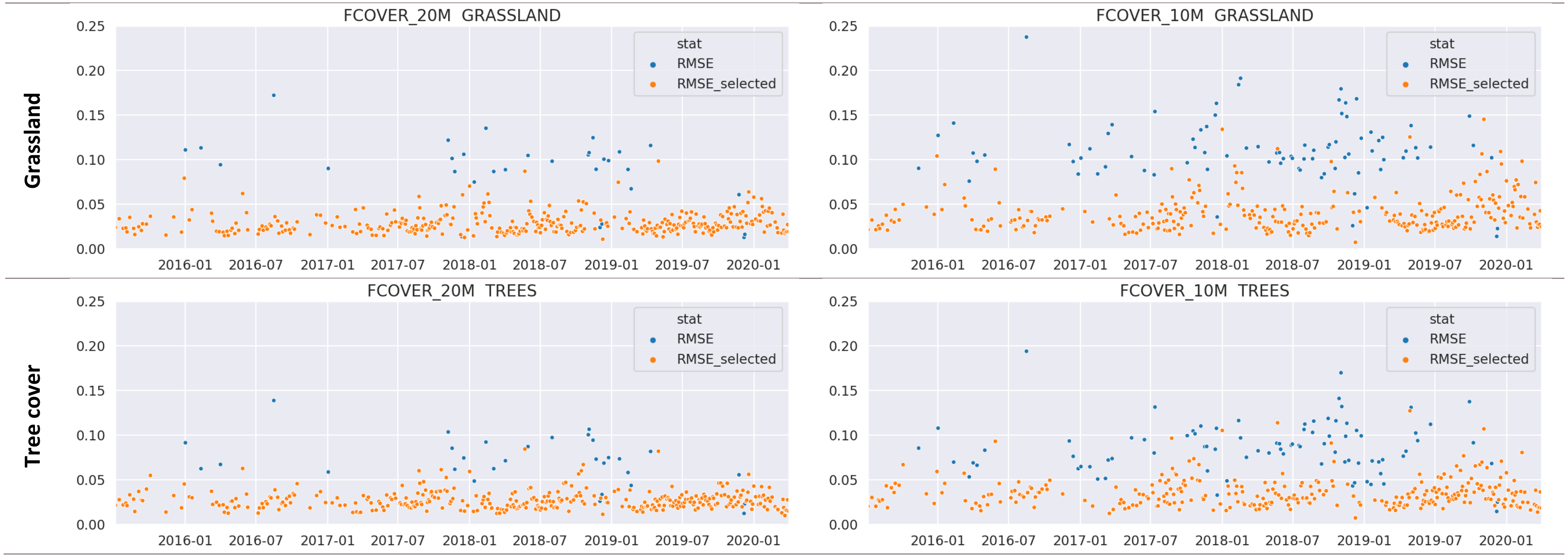


Figure 17: Temporal plots of Root Mean Squared Error (RMSE) between V200 and V102 of FCOVER_20M (left) and FCOVER_10M (right) for all land cover types (top row) and per biome (next rows). MBE are all dates and MBE_selected are all dates with sufficient sample size and where a geometric shift exists between V200 and V102.

FCOVER_20M	**FCOVER_10M**



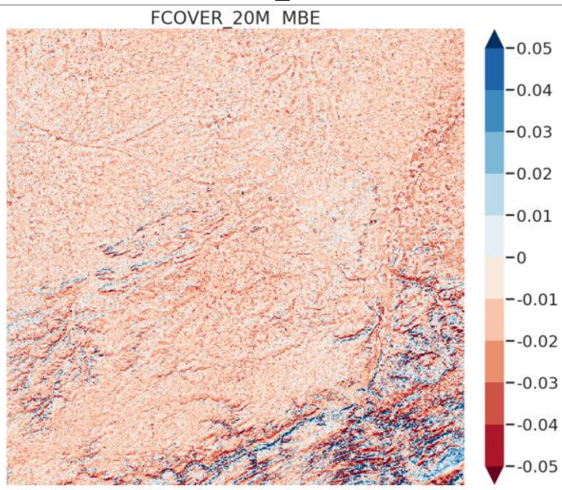
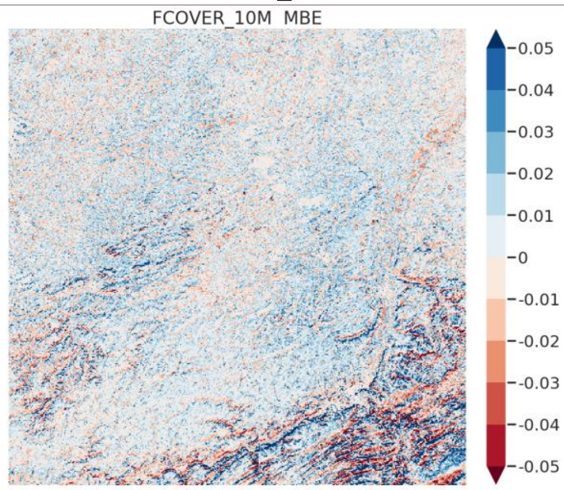

FCOVER_20M BUILT-UP
FCOVER_10M BUILT-UP
FCOVER_20M CROPLAND
FCOVER_10M CROPLAND
Built-up
Cropland
stat
RMSE
RMSE_selected
0.25
0.20
0.15
0.10
0.05
0.00
2016-01
2016-07
2017-01
2017-07
2018-01
2018-07
2019-01
2019-07
2020-01

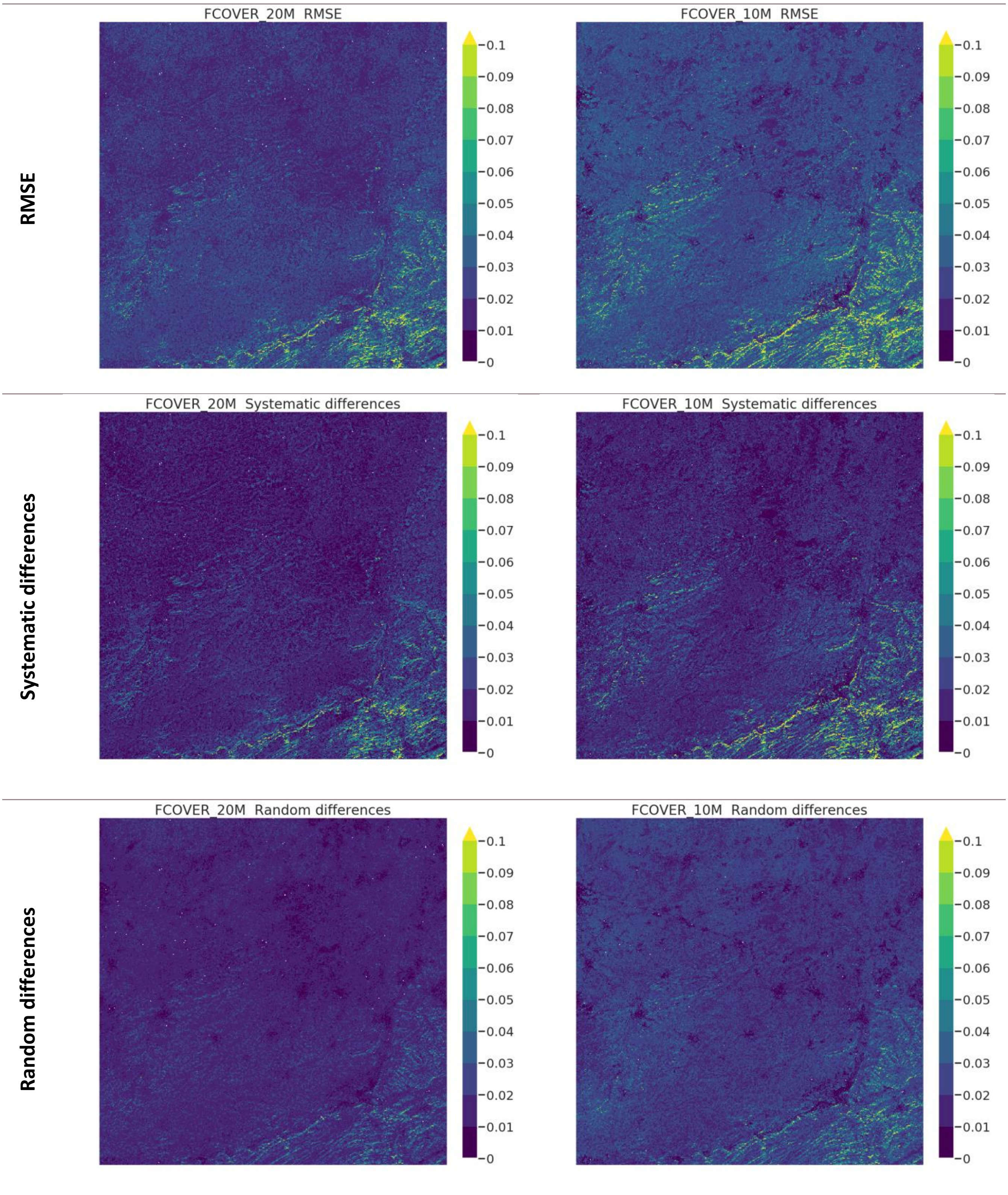

FCOVER_20M GRASSLAND
FCOVER_10M GRASSLAND
FCOVER_20M TREES
FCOVER_10M TREES
Grassland
Tree cover
stat
RMSE
RMSE_selected
0.25
0.20
0.15
0.10
0.05
0.00
2016-01
2016-07
2017-01
2017-07
2018-01
2018-07
2019-01
2019-07
2020-01

4.4 Spatial patterns of differences

The spatial patterns of the per-pixel MBE, RMSE, systematic and random differences between V200 and V102 FCOVER_20M and 10M are shown in Figure 18.
A number of spatial patterns can be discerned in the spatial plots of all the metrics for both the 20M and 10M FCOVER. The differences are the largest in the southeast corner, which is an area that is characterized by a more pronounced relief. Similarly, the river valleys also have larger differences between V200 and V102. This is probably because in V200, Sen2Cor applies a topographic correction on the images which results in brighter or less bright slopes. Such a correction was not done in V102. Although differences can be locally large, the majority shows moderate differences.
The MBE is negative for most pixels in 31UFS for FCOVER_20M, whereas for FCOVER_10M it is a mixture of negative and positive values. This is in agreement with the bias histograms (see Figure 15). The RMSE is higher for FCOVER_10M than for FCOVER_20M. The spatial patterns are also slightly different between the resolutions. In the FCOVER_10M a number of patches with very low RMSE can be found. Comparing these with the land cover map (Figure 2) these are large grassland areas and built-up areas.
When looking at the split between systematic and random differences, it is observed that the two FCOVER_10M versions mainly differ in random differences rather than systematic ones for areas with no pronounced topography. For FCOVER_20M the random differences are smallest.

Figure 18: Spatial pattern of the per-pixel Mean Bias Error (MBE), Root Mean Squared Error (RMSE), Systematic difference and Random difference between V200 and V201 of FCOVER_20M (left) and FCOVER_10M (right).

	FCOVER_20M	**FCOVER_10M**
MBE		


RMSE
FCOVER_20M RMSE
FCOVER_10M RMSE
Systematic differences
FCOVER_20M Systematic differences
FCOVER_10M Systematic differences
Random differences
FCOVER_20M Random differences
FCOVER_10M Random differences
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0

4.5 Conclusions for FCOVER_20M and FCOVER_10M

The full time series until 31/03/2020 of tile 31UFS of the TERRASCOPE FCOVER_20M and FCOVER_10m V200 was compared to the previous version V102 in order to demonstrate the agreement and the differences.
Overall, FCOVER V200 and V102 agree linearly, and with a reasonable amount of scattering. The agreement is better for the FCOVER_20M, which has a small negative bias irrespective of land cover type. A negative bias means that V200 FCOVER values are lower than those of V102. The range of FCOVER_20M values is similar between versions, but V200 has more lower values.
The FAPAR_10M also exhibits a linear relationship between V200 and V102, with a bias depending on land cover type: mainly negative for 'built-up' and 'cropland', and mainly positive for 'grassland' and 'tree cover'. There is more scatter compared to FCOVER_20M in the difference. The range of the FCOVER values is similar among versions, but FCOVER_10M V200 has more low and more high values.

A large influence of the difference in AOT was observed. Large differences in AOT used in V200 and V102 are responsible for:

- Non-linear relationship between the two versions
- More scatter in the relationship
- Seasonal pattern in difference

The spatial analysis demonstrated that the topographic correction used in the atmospheric correction of V200 TOC reflectances has an impact on the data: larger differences for areas with more topography. As the analysis focused on the differences on the two versions, it is not possible to judge which version is more accurate. We can assume that the topographic correction should be an improvement in the processing.

The use of per-pixel Sun angles instead of mean Sun angles per image in the retrieval of FCOVER is not discernable from the results.

It should be mentioned that the impact of geometric errors was not assessed. This is usually expressed in random differences rather than systematic ones. Large differences between V200 and V102 FCOVER were observed for those dates for which a geolocation shift was applied in V102 and not in V200. The difference metrics were significantly larger than the ones observed for the images that have the same geolocation in both versions.

It is advised not to mix the two FCOVER versions in the same analysis, because this could lead to erroneous conclusions.

5 LAI 10M AND 20M: RESULTS

This chapter discusses the results of the inter-comparison between V200 and V102 for the LAI products at 10 m and 20 m resolution.

5.1 Overall statistical analysis

The Mean Bias Difference (MBE), Mean Absolute Difference (MAE), and Root Mean Squared Difference (RMSE), which is further split into random and systematic differences, were calculated on the overall sample (see section 2.2.1) for all land cover classes and per land cover class. The results for LAI are summarized in Table 7.
Comparing the statistics between the two resolutions, it is observed that the overall MBE of LAI for 20M is negative (-0.043) and positive (0.073) for the 10M resolution LAI. The MBE is negative for all land cover classes for LAI_20M, but only for 'built-up' for LAI_10M. All other land cover classes have positive MBE for LAI_10M. The MAE and RMSE are both higher for LAI_10M than for LAI_20M, except for 'built-up' and both metrics vary among the land cover classes.
For LAI_10M, the random differences are always larger than the systematic ones, but not in the same magnitude for all land covers. This is not the case for LAI_20M, where for built-up the random difference is smaller than the systematic differences.

Table 7: Mean Bias Error (MBE), Mean Absolute Error (MAE), Root Mean Squared Difference (RMSD), Random difference and Systematic difference between V200 and V102 of LAI_20M and LAI_10M for all land cover classes (All LC) or per biome.

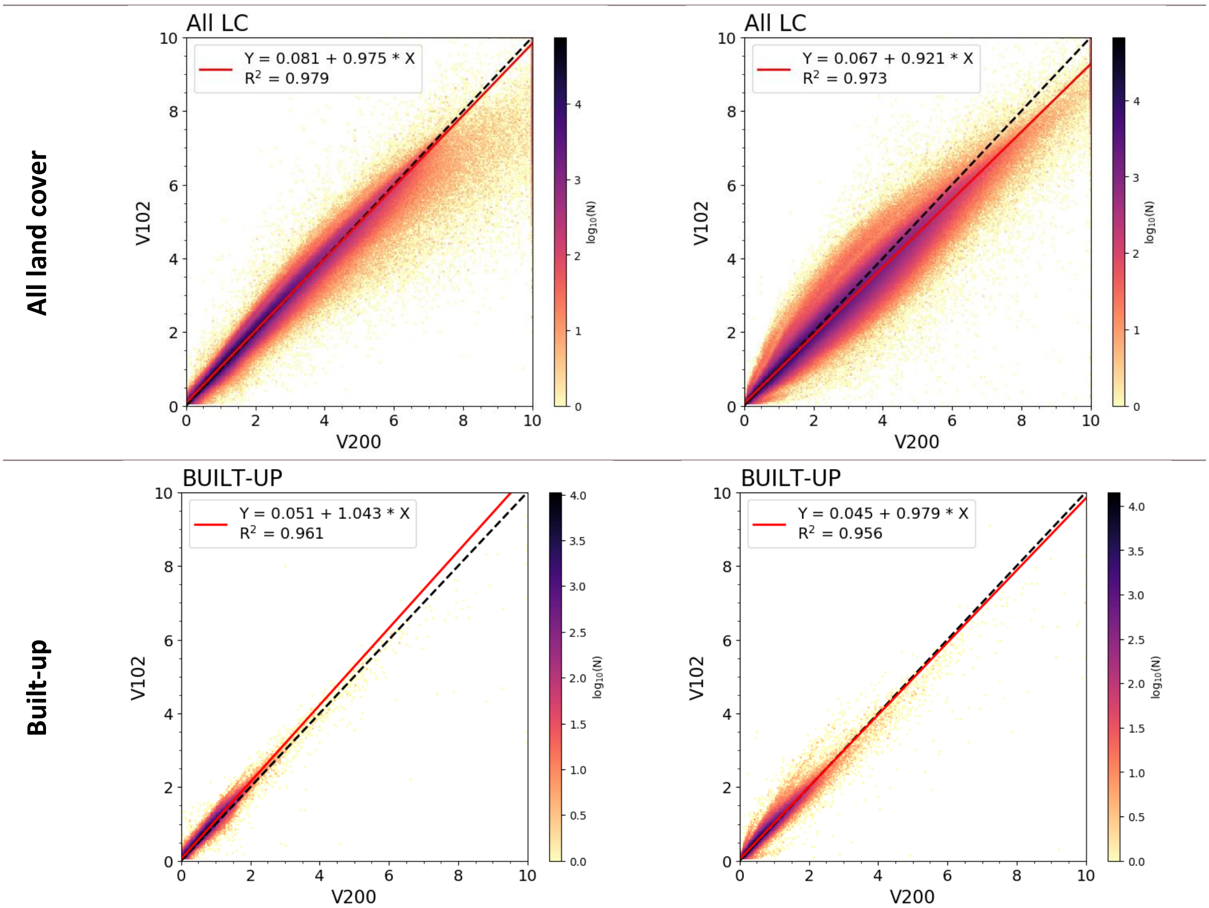
	LAI_20M					LAI_10M				
	MBE	MAE	RMSE	Rand. diff	Sys. diff.	MBE	MAE	RMSE	Rand. diff	Sys. diff.
All LC	-0.043	0.107	0.179	0.171	0.052	0.073	0.165	0.271	0.234	0.137
BUILT-UP	-0.078	0.085	0.109	0.075	0.079	-0.034	0.063	0.099	0.092	0.035
CROPLAND	-0.059	0.097	0.137	0.121	0.065	0.071	0.167	0.251	0.233	0.093
GRASSLAND	-0.043	0.112	0.176	0.167	0.055	0.087	0.189	0.298	0.263	0.141
TREES	-0.020	0.116	0.226	0.216	0.064	0.092	0.172	0.295	0.230	0.184

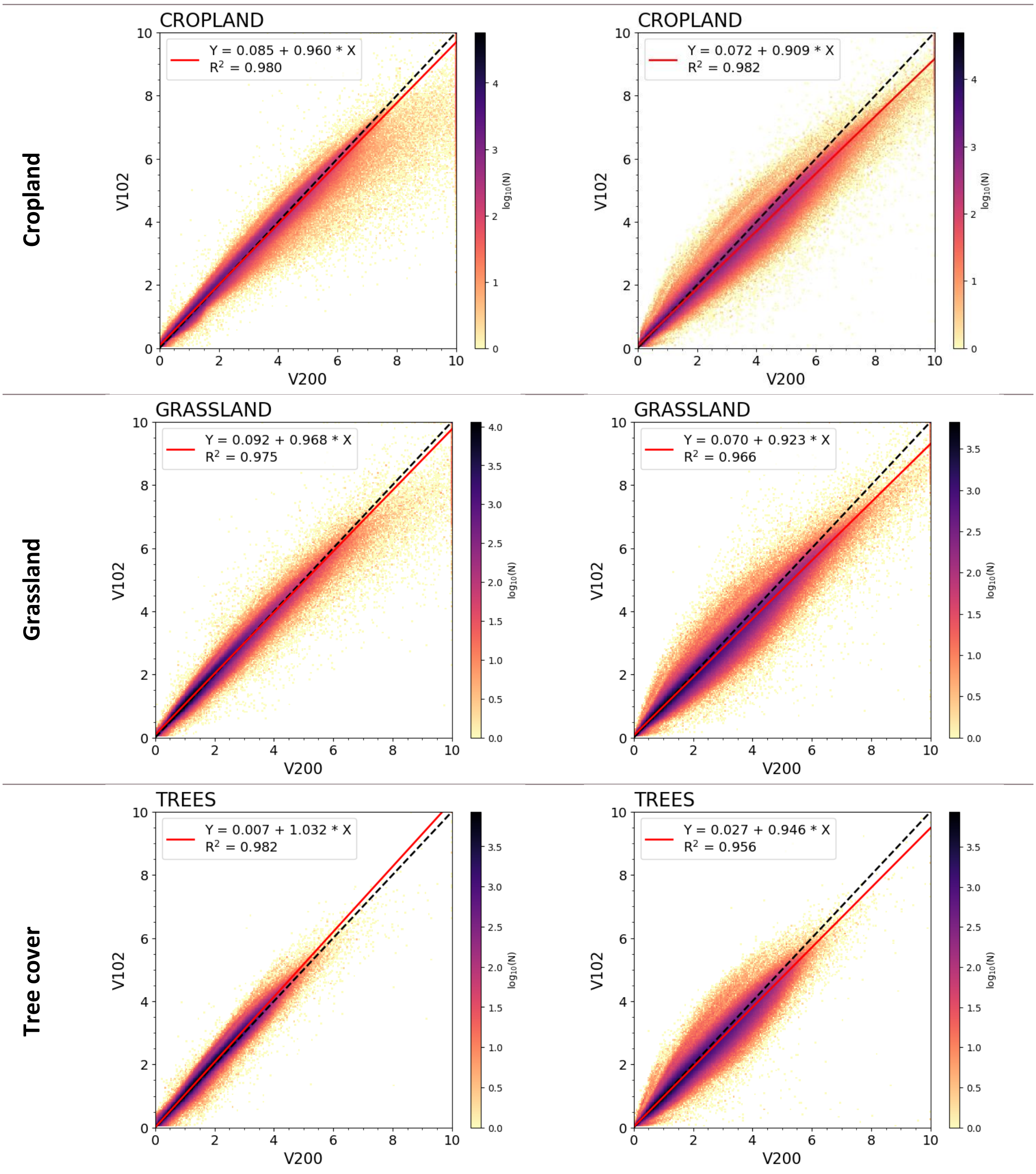
Figure 19 shows the corresponding scatterplots with the orthogonal geometric mean regression line and the coefficient of determination. The V200 and V102 are overall linearly related up to LAI=5 and the difference between the 20M LAI versions shows less scattering. For higher LAI the relationship changes: the V102 LAI values in the range [5,8] are stretched over the range [5,10] in V200 which is more pronounced for LAI_20M. In this range, the 20M LAI difference shows the most scatter than LAI_10M between the 2 versions.
For all land covers and both resolutions, the V200 LAI has more smaller values as is shown by the positive offset of the regression line. The slope of the regression line is < 1 for all land cover of LAI_10M. For the 20M LAI, this depends on the land cover type.
The scatterplot of LAI_10M shows seemingly a separate population above the regression line. This is visible in the plot for all land cover, but more pronounced for 'built-up', 'cropland' and 'grassland'.

This was further investigated by looking at the scatterplots of single dates. For most of the dates, the scatterplots are narrow, linear and close to the 1-1 line. But if the Aerosol Optical Thickness (AOT) used to correct the data for atmospheric scattering differs largely between V200 and V102, the relationship drifts from the 1-1 line and becomes non-linear or shows more scatter.

Figure 19: Scatterplots between V200 (X) and V102 (Y) LAI_20M (left) and LAI_10M (right) products for the period 01/07/2015 – 26/03/2020. Red line is the orthogonal geometric mean regression line between the two versions.

LAI_20M **LAI_10M**

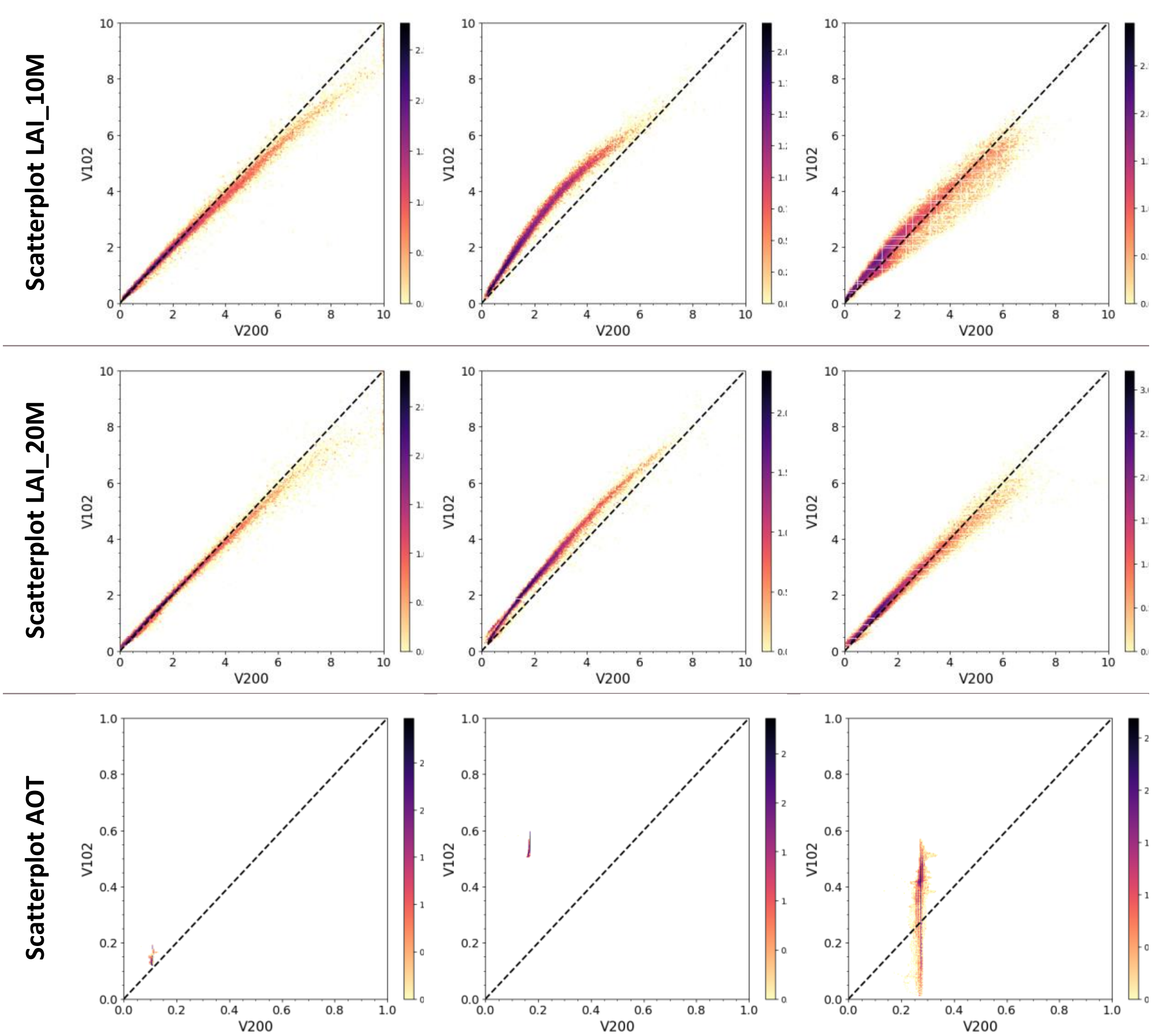




Examples of this are shown in Figure 20 for 3 specific cases: (1) AOT between V200 and V102 is similar, (2) AOT is much lower for V200 and V102, and in both versions the AOT range in the image is small, and (3) the AOT is very different between both version and has a large range in V102 but a small range in V200. In the first case, the LAI is very similar between V200 and V102 up to LAI=5, and above skewed to higher values in V200. Their difference is almost the same for 10 m and 20 m resolution LAI. A large difference in AOT as in case (2) results in a clear non-linear relationship with not much scatter. The impact is much larger for LAI_10M than for LAI_20M. The third case results in large scatter between the 2 versions which is most pronounced for LAI_10M. These results suggest that AOT differences between the two versions have a much larger impact on the LAI_10M. This is probably caused by the use of only 3 bands in the retrieval of LAI, of which 2 bands have shorter wavelengths and are thus more affected by the difference in AOT. The LAI_20M is based on 8 bands that include also two SWIR bands that are less perturbed by AOT difference (see also [RD-3] for more details). Hence, the LAI_10M is more sensitive to accurate AOT estimation.

Figure 20: Comparison of scatterplots for 3 dates with variable difference in AOT estimated in V200 and V102 (columns). Top row: LAI_10M, middle row: LAI_20M, bottom row: AOT.

	Small difference in AOT (23/04/2018)	**Large difference in AOT (28/05/2016)**	**Large range of difference in AOT (21/04/2018)**

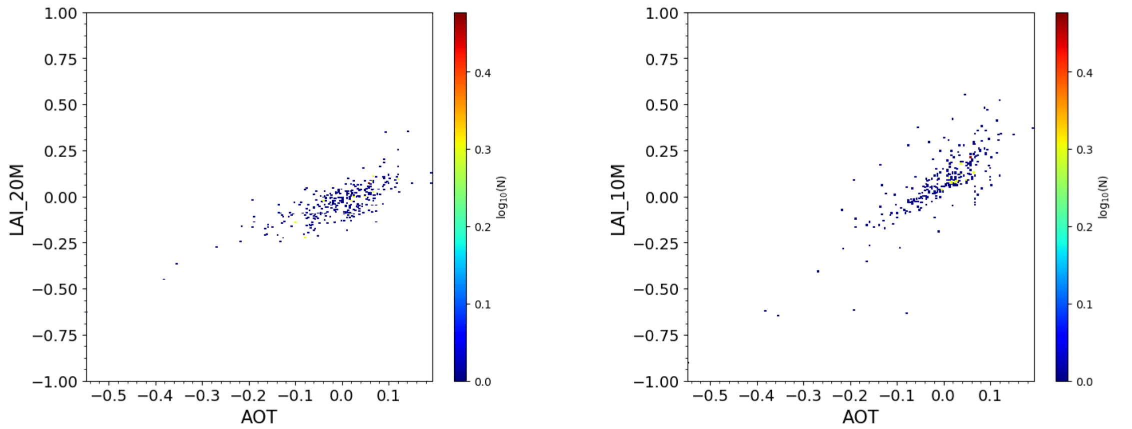


It should be noted that the inter-comparison merely demonstrates differences and analyses their causes, but it does not indicate which data set is more accurate.

Figure 21 shows the scatterplot between the MBE calculated between V200 and V102 (MBE=V200-V102) of AOT and LAI for both resolutions. It demonstrates that if the AOT of V200 is lower than V102 (MBE AOT < 0), then the MBE of LAI is also lower in V200 (MBE LAI < 0). This relationship is stronger for LAI_10M.

Figure 21: Scatterplot between Mean Bias Error between V200 and V102 of AOT and LAI_20M (left) and LAI_10M (right). MBE = V200-V102.

LAI_20M	**LAI_10M**

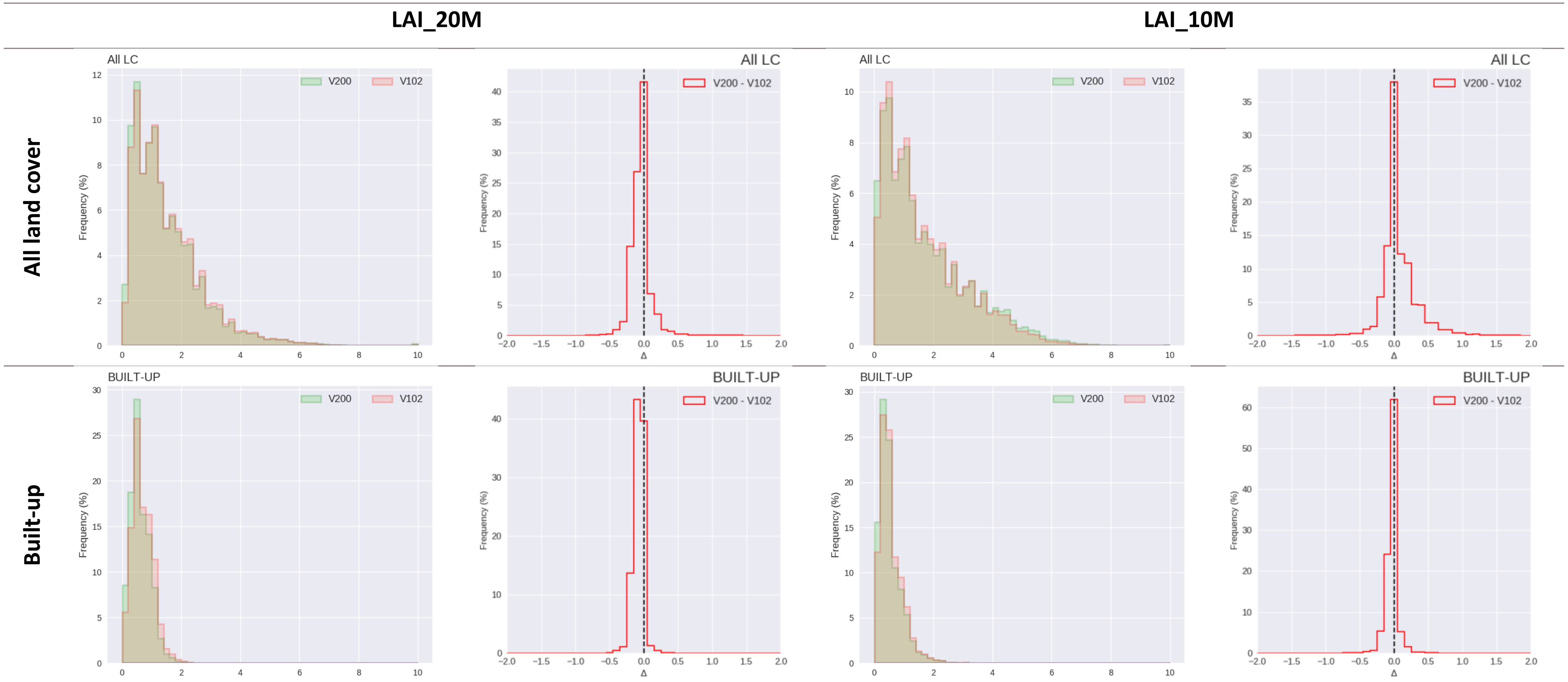


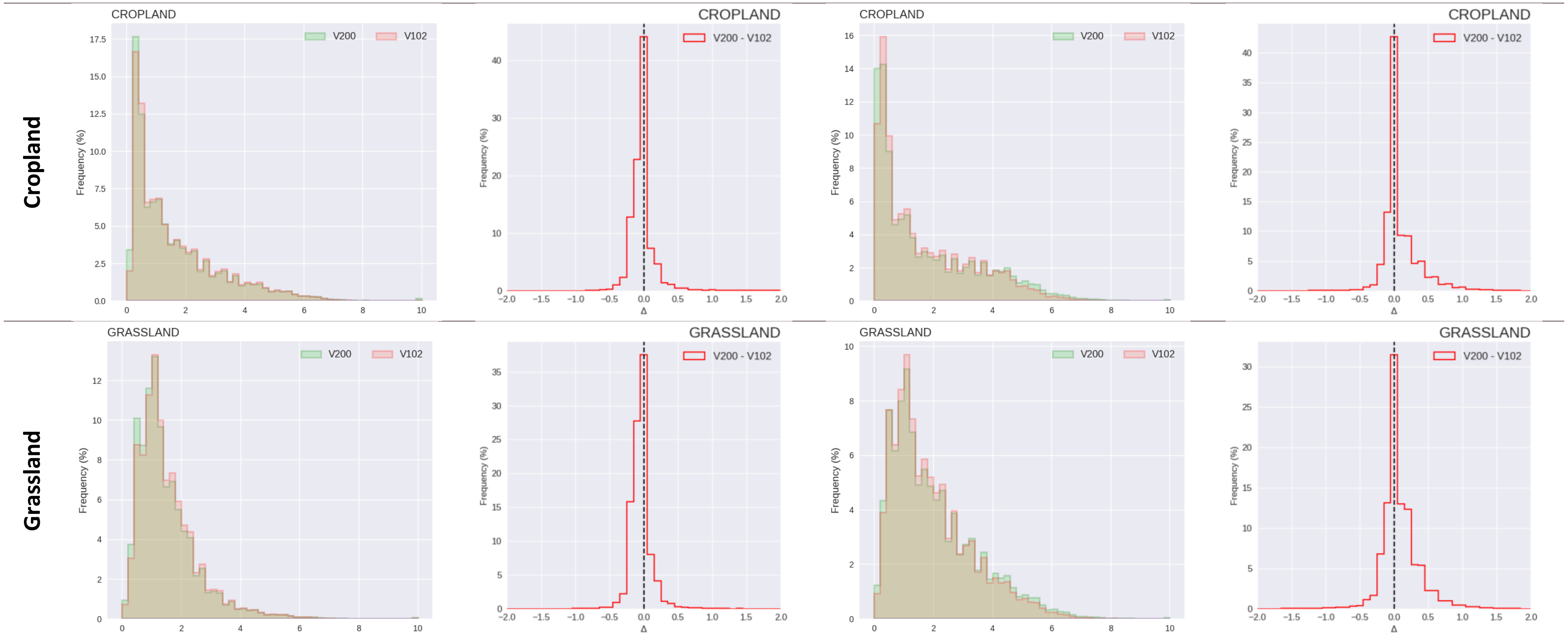
5.2 Histograms and bias histograms

Histograms and bias histograms for V200 and V102 LAI_20M and LAI_10M are shown in Figure 22 for all land cover in 31UFS and for the 4 major land cover classes. The bias is V200 minus V102, thus if the peak of the bias is negative, then V102 has overall higher values compared to V200.
Looking at all land cover, the range of the LAI V200 values is similar compared to V102. For the 20M data, slightly more low values are observed, whereas for 10M LAI, also slightly more high values are present. In the bias histogram of LAI_20M, the peak is around 0, with 75% of the pixels between [-0.15, 0.15] , and 93.5% between [-0.25, 0.25]. For LAI_10M, there is also one peak 0, with 80% of the pixels between [-0.25, 0.25] and 92% between [-0.45, 0.45].
When looking at the results per land cover class, the following is observed:

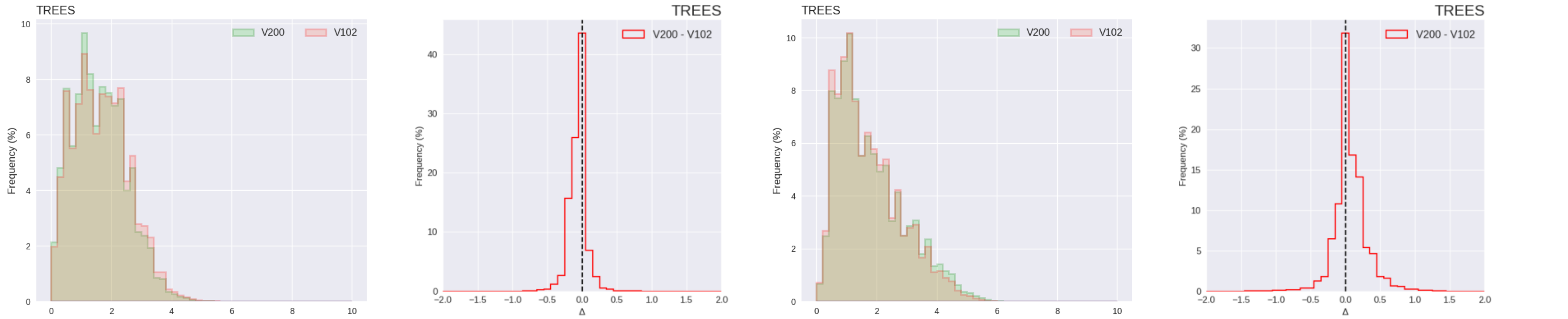
- 'Built-up': there is a shift to lower LAI values for both resolutions. The shift is slightly more pronounced for LAI_20M. For LAI_10M, the peak in the bias histogram is still around 0, whereas for the 20M LAI it is around -0.1.
- 'Cropland': the results of the bias histogram are very similar to the 'all land cover' results, i.e. a peak around 0 and a skewed histogram toward negative values for LAI_20M. For the 10M LAI, the range of positive bias values is larger than the negative ones, and the peak is clearly around 0.

Figure 22: Histograms and bias histograms for V200 and V102 LAI_20M and LAI_10M products over the period 01/07/2015 – 26/03/2020. Bias is calculated as V200-V102.




Cropland
Grassland
CROPLAND
GRASSLAND
V200
V102
V200 - V102
Frequency (%)
Δ

Tree cover


TREES
V200
V102
Frequency (%)
TREES
V200 - V102
Frequency (%)
Δ
TREES
V200
V102
Frequency (%)
TREES
V200 - V102
Frequency (%)
Δ

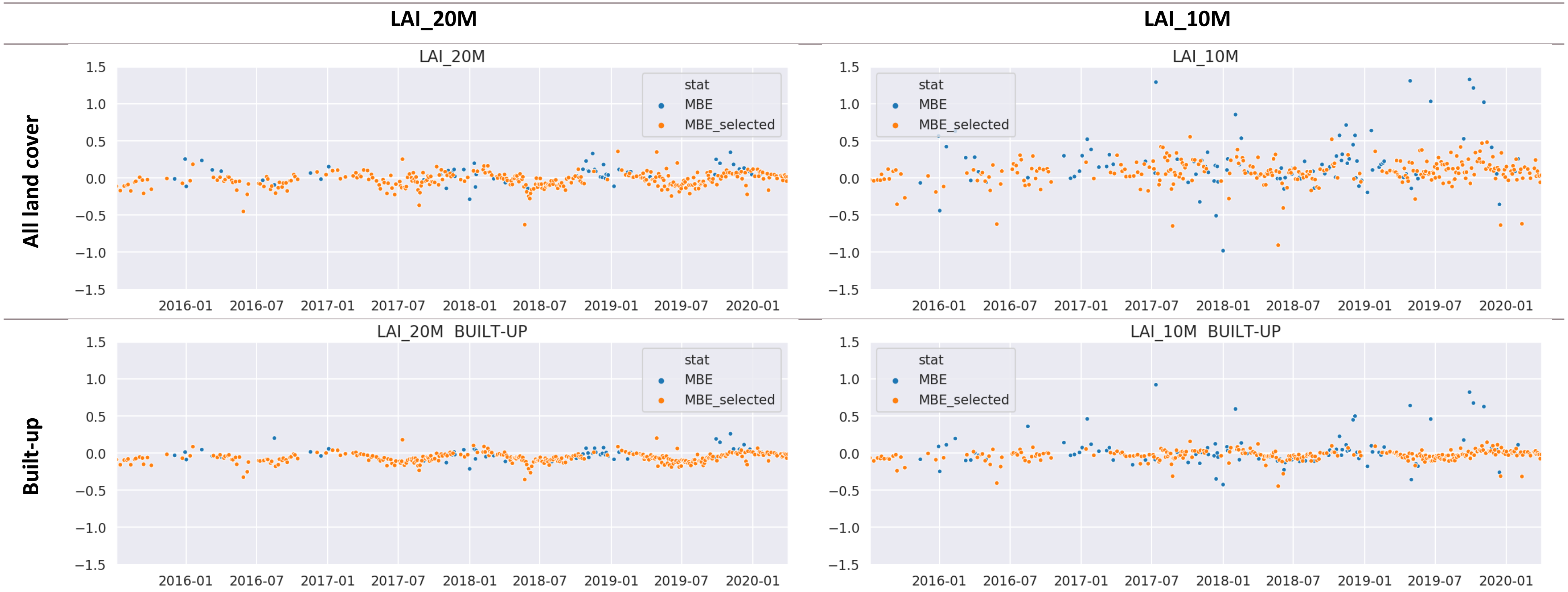
- 'Grassland': the peak of the bias histogram for both LAI_20M and LAI_10M is also around 0. Again, the bias histogram is skewed to negative values for the 20M LAI and to positive values for 10M.
- 'Tree cover': the results look very similar as those of 'grassland' with even a larger disagreement of the LAI_20M and LAI_10M in their difference between V200 and V102. LAI_20M is skewed to a negative bias, and LAI_10M to a predominant positive bias (V102 smaller than V200). Both have a clear peak around 0.

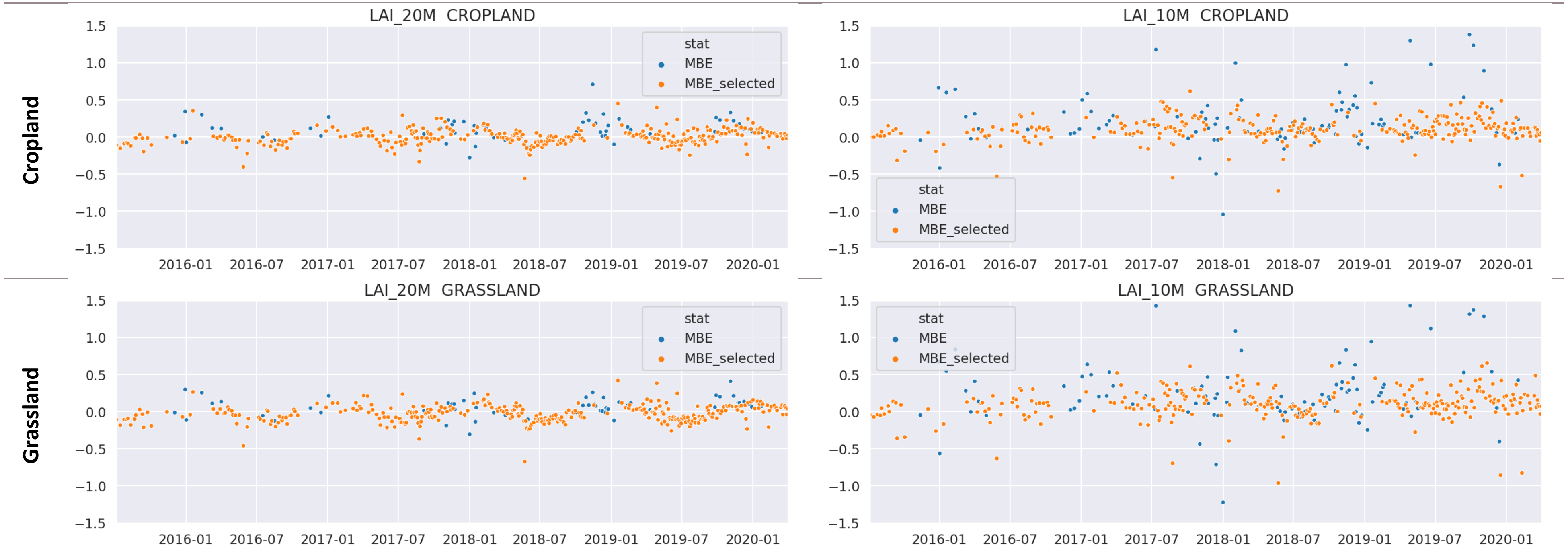
The above results indicate that LAI_20M shows approximately the same behaviour between V200 and V102, irrespective of the land cover class. This is not the case for LAI_10M, where the bias histogram becomes more skewed to positive values with higher vegetation cover.

5.3 Temporal analysis of agreement and differences

This section describes the result of the temporal analysis. Per date, the LAI image of V200 is compared to that of V102 according to the sampling strategy described in section 2.2.2. For each date, a number of statistics is calculated and these are presented as temporal plots. This is done for all land cover types and per biome and presented for MBE in Figure 23 and for RMSE in Figure 24Figure 17.
For the majority of the dates, the MBE of LAI 20M and 10M varies between -0.5 and 0.5. There are more dates outside this range for LAI_10M. The MBE range and its magnitude is similar among land cover types, except for 'built-up' the MBE is lower and ranges between [-0.2, 0.2].
The MBE shows a seasonal pattern, with positive MBE in winter and negative MBE in summer. This is the case for all land cover types. When looking at the temporal evolution of the MBE between the AOT V200 and V102, the same seasonality is observed (see [RD-3]). The majority of the AOT MBE ranges between [-0.15, 0.15], with positive values in winter and negative values in summer. This is consistent with the analysis shown in Figure 23.
The RMSE between V200 and V102 of LAI 20M and 10M varies roughly between [0.0, 0.5]. The RMSE range is largest for 'grassland' and 'cropland', and smallest for 'built-up' for the 20M LAI. The RMSE range of 'tree cover' is equally large as for 'cropland' and 'grassland' for the 10M LAI.
There is a substantial amount of dates with RMSE above 0.5. For LAI_20M, there are dates for which the RMSE is up to 0.8 and for LAI_10M even up to 1.6. Most of these dates are cloudy (small sample size) or the V102 data have been manually corrected for a geolocation shift.
A weak seasonal pattern is also visible in the RMSE temporal plots. This is the most expressed for 'built-up' at both resolutions.

Figure 23: Temporal plots of Mean Bias Error (MBE) between V200 and V102 of LAI_20M (left) and LAI_10M (right) for all land cover types (top row) and per biome (next rows). MBE are all dates and MBE_selected are all dates with sufficient sample size and where a geometric shift exists between V200 and V102.




Cropland
Grassland
LAI_20M CROPLAND
LAI_10M CROPLAND
LAI_20M GRASSLAND
LAI_10M GRASSLAND
stat
MBE
MBE_selected
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
2016-01
2016-07
2017-01
2017-07
2018-01
2018-07
2019-01
2019-07
2020-01

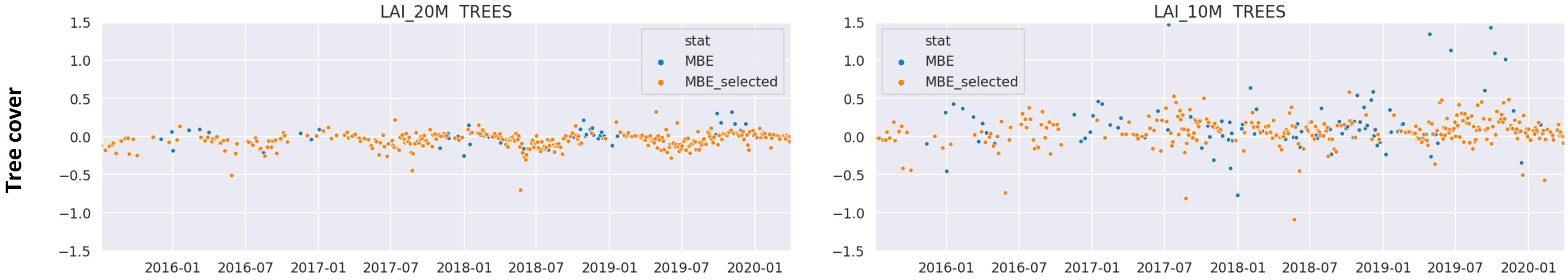
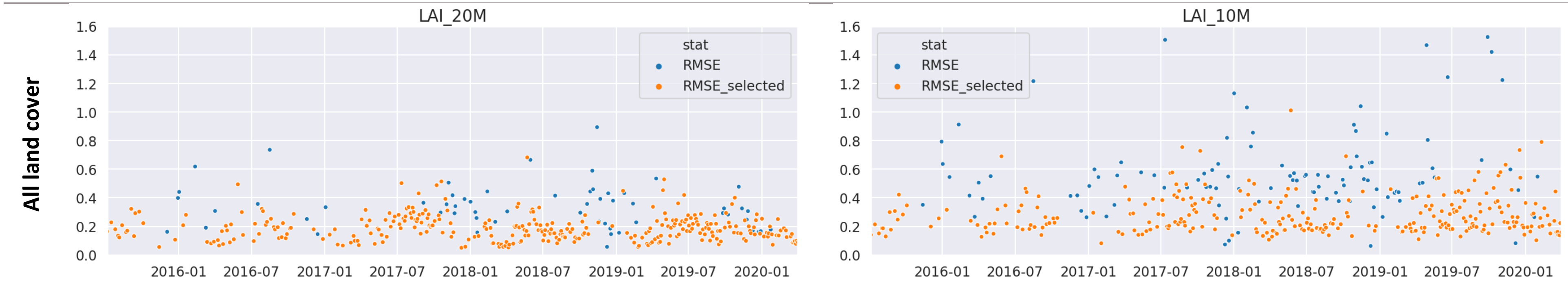
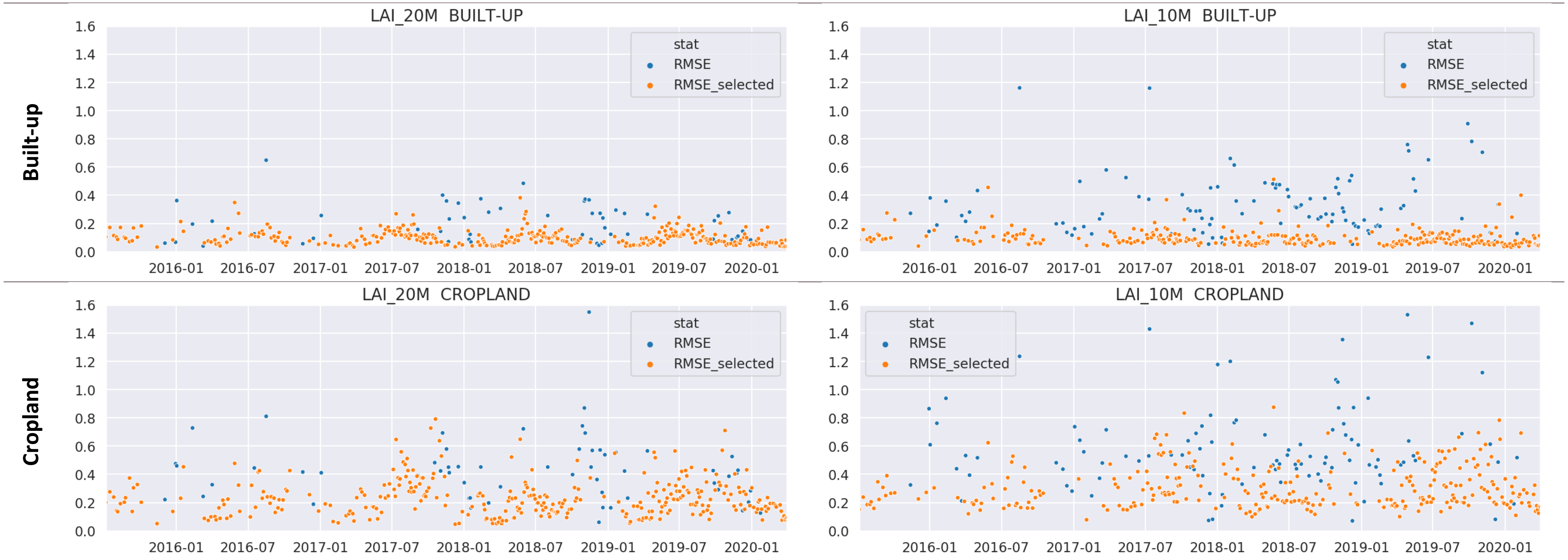
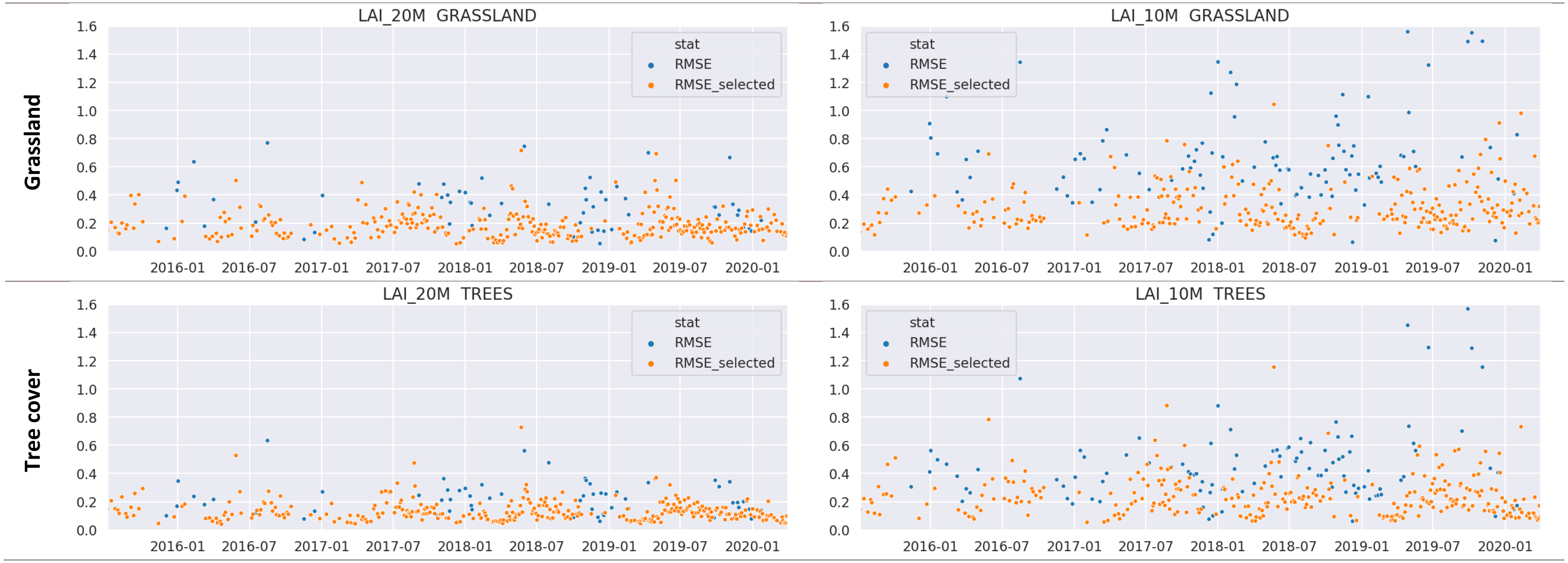


Figure 24: Temporal plots of Root Mean Squared Error (RMSE) between V200 and V102 of LAI_20M (left) and LAI_10M (right) for all land cover types (top row) and per biome (next rows). MBE are all dates and MBE_selected are all dates with sufficient sample size and where a geometric shift exists between V200 and V102.

	LAI_20M	**LAI_10M**



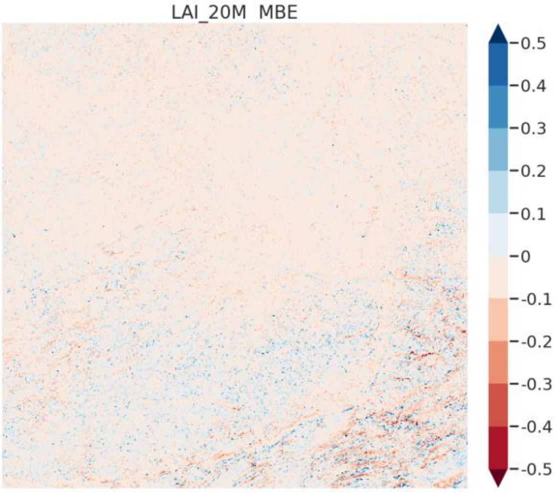
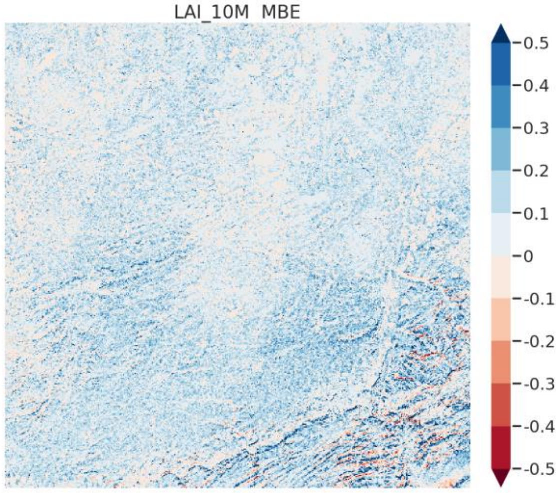
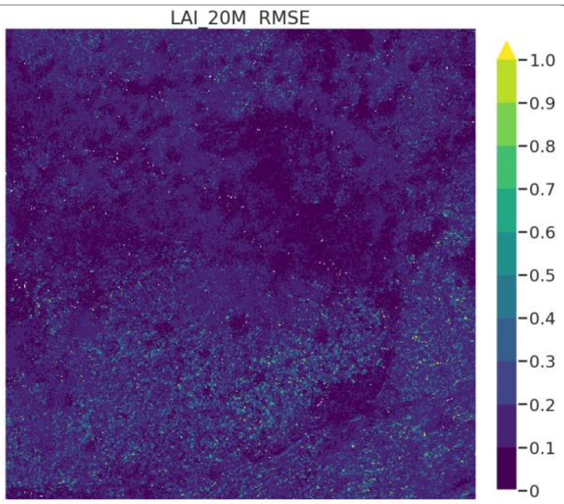
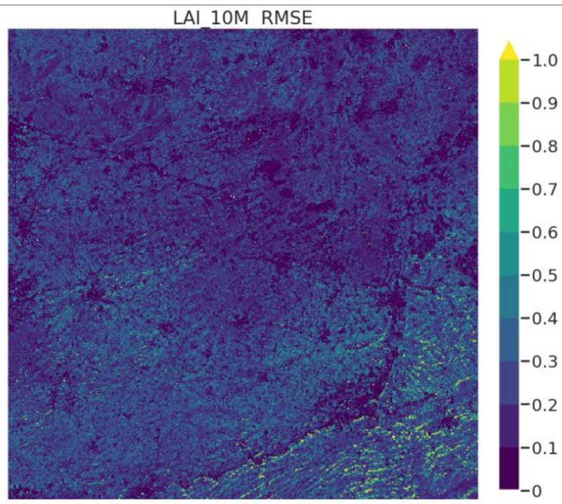

Built-up
Cropland
LAI_20M BUILT-UP
LAI_10M BUILT-UP
LAI_20M CROPLAND
LAI_10M CROPLAND
stat
RMSE
RMSE_selected
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
2016-01
2016-07
2017-01
2017-07
2018-01
2018-07
2019-01
2019-07
2020-01

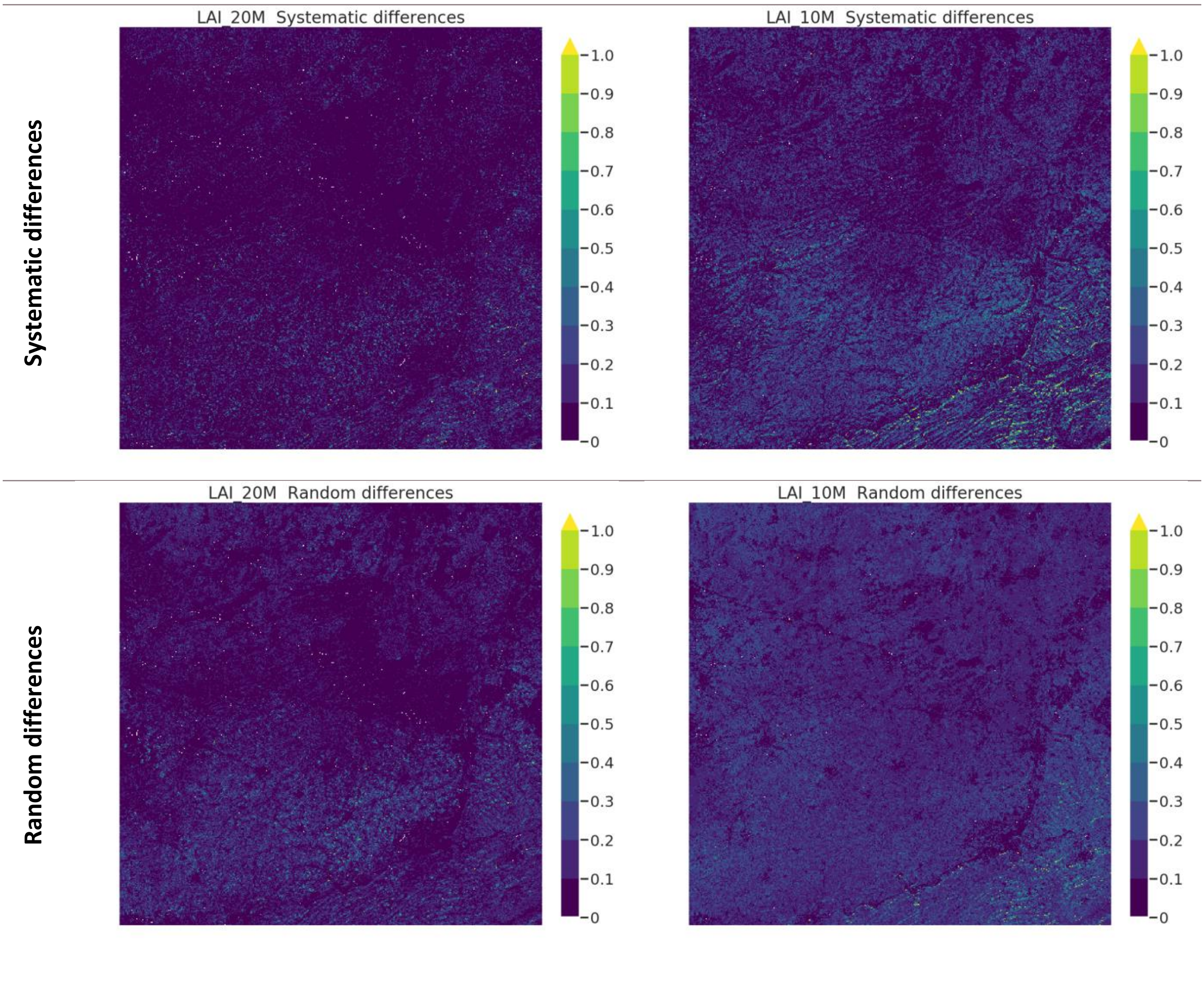

Grassland
Tree cover
LAI_20M GRASSLAND
LAI_10M GRASSLAND
LAI_20M TREES
LAI_10M TREES
stat
RMSE
RMSE_selected
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
2016-01
2016-07
2017-01
2017-07
2018-01
2018-07
2019-01
2019-07
2020-01

5.4 Spatial patterns of differences

The spatial patterns of the per-pixel MBE, RMSE, systematic and random differences between V200 and V102 LAI_20M and 10M are shown in Figure 25.
A number of spatial patterns can be discerned in the spatial plots of all the metrics for both the 20 m and 10 m LAI. The differences are largest in the southeast corner, which is an area that is characterized by a more pronounced topography. This is probably because in V200, Sen2Cor applies a topographic correction on the images which results in brighter or less bright slopes. Such a correction was not done in V102. Although differences can be locally large, the majority of the image shows moderate differences.
The MBE is negative for most pixels in 31UFS for LAI_20M, whereas the opposite is true for LAI_10M. This is largely in agreement with the bias histograms (see Figure 15).
The RMSE is clearly higher for LAI_10M than for LAI_20M. The spatial patterns are also slightly different between the resolutions. In the LAI_10M a number of patches with very low RMSE can be found. Comparing these with the land cover map (Figure 2) these are mainly built-up areas. The impact of topography is less discernable in the RMSE of LAI_20M compared to that of LAI_10M.
The differences associated with the topographic correction are expressed mainly in the systematic difference, which is as expected.
When looking at the split between systematic and random differences outside the areas with pronounced topography, it is observed that the random differences are slightly higher. For LAI_20M the random differences are higher for areas with more relief.

Figure 25: Spatial pattern of the per-pixel Mean Bias Error (MBE), Root Mean Squared Error (RMSE), Systematic difference, and Random difference between V200 and V201 of LAI_20M (left) and LAI_10M (right).

	LAI_20M	**LAI_10M**
MBE		
RMSE		



5.5 Conclusions for LAI_20M and LAI_10M

The full time series until 31/03/2020 of tile 31UFS of the TERRASCOPE LAI_20M and LAI_10m V200 was compared to the previous version V102 in order to demonstrate the agreement and the differences.

Overall, LAI V200 and V102 agree linearly and with a reasonable amount of scattering up to LAI=5. For larger LAI values, there is shift towards larger LAI values in V200. The agreement is better for the LAI_20M, for which the peak of the bias histogram is around 0, but skewed to negative values. This is irrespective of land cover type. A negative bias means that V200 LAI values are lower than those of V102. The range of LAI_20M values is similar between versions, but V200 has slightly more lower values.

The LAI_10M also exhibits a linear relationship between V200 and V102 up to LAI=5, and a shift towards higher values for V200 for LAI>5 but to a lesser extent as for LAI_20M. The observed bias peaks around 0, but the skewedness of the distribution depends weakly on land cover type: mainly to negative for 'built-up' and 'cropland', and mainly to positive for 'grassland' and 'tree cover'. There is more scatter compared to LAI_20M in the difference. The range of the LAI values is similar among versions, but LAI_10M V200 has more low and more high values.

A large influence of the difference in AOT was observed. Large differences in AOT used in V200 and V102 are responsible for:

- Non-linear relationship between the two LAI versions
- More scatter in the relationship between the two LAI versions
- Seasonal pattern in difference the two LAI versions

The spatial analysis demonstrated that the topographic correction used in the atmospheric correction of V200 TOC reflectances has an impact on the data: larger differences for areas with more topography. As the analysis focused on the differences on the two versions, it was not possible to judge which version is more accurate. We can assume that the topographic correction should be an improvement in the processing.

The use of per-pixel Sun angles instead of mean Sun angles per image in the LAI retrieval is not discernable from the results.

It should be mentioned that the impact of geometric errors was not assessed. This is usually expressed in random differences rather than systematic ones. Large differences between V200 and V102 LAI were observed for those dates for which a geolocation shift was applied in V102 and not in V200. The difference metrics were significantly larger than the ones observed for the images that have the same geolocation in both versions.

It is advised not to mix the two version of the LAI in the same analysis, because this could lead to erroneous conclusions.

6 NDVI 10M: RESULTS

This chapter discusses the results of the inter-comparison between V200 and V102 for the NDVI products at 10 m resolution.

6.1 Overall statistical analysis

The Mean Bias Difference (MBE), Mean Absolute Difference (MAE), Root Mean Squared Difference (RMSE), which is further split into random and systematic differences were calculated on the overall sample (see section 2.2.1) for all land cover classes and per land cover class. The results for NDVI are summarized in Table 7.

It is observed that the overall MBE of NDVI is slightly negative (-0.003). The MBE is negative for 'built-up' and 'tree cover', and positive for 'cropland', and 0 for 'grassland'. The overall MAE and RMSE are0.023 and 0.033 resp. and are both highest for 'built-up' and lowest for 'grassland'.

The random differences are always larger than the systematic ones, but not in the same magnitude for all land covers.

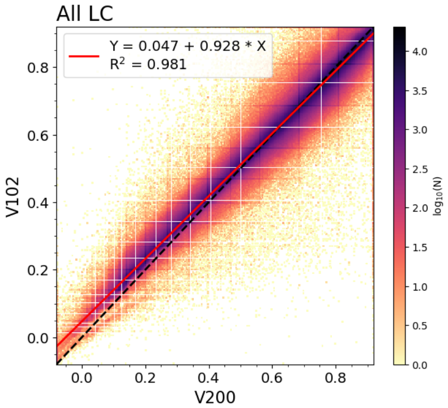
Table 8: Mean Bias Error (MBE), Mean Absolute Error (MAE), Root Mean Squared Difference (RMSD), Random difference and Systematic difference between V200 and V102 of NDVI_10M for all land cover classes (All LC) or per biome.

	NDVI_10M				
	MBE	**MAE**	**RMSE**	**Rand. diff**	**Sys. diff.**
All LC	-0.003	0.023	0.033	0.029	0.016
BUILT-UP	-0.022	0.029	0.040	0.031	0.026
CROPLAND	0.008	0.024	0.036	0.035	0.010
GRASSLAND	0.000	0.019	0.028	0.026	0.011
TREES	-0.009	0.024	0.032	0.025	0.021

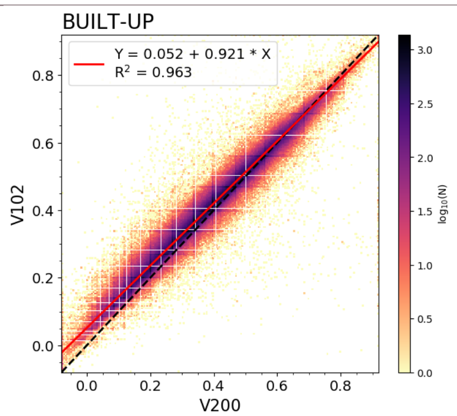
Figure 26 shows the corresponding scatterplots with the orthogonal geometric mean regression line and the coefficient of determination. The V200 and V102 are overall linearly related with a relatively high amount of scattering. For all land covers, the V200 NDVI has more smaller values as is shown by the positive offset of the regression line. The slope of the regression line is < 1 for all land cover types. The scatterplot of NDVI_10M does not show a separate population above the regression line as is the case for the BIOPARs in previous sections. But the amount of scatter is much higher for NDVI than for the BIOPARs. The NDVI was further investigated similar as for the BIOPARs by looking at the scatterplots of single dates. For most of the dates, the scatterplots are narrow, linear and close to the 1-1 line. But if the Aerosol Optical Thickness (AOT) used to correct the data for atmospheric scattering differs largely between V200 and V102, the relationship drifts from the 1-1 line and becomes slightly non-linear or shows more scatter.

Figure 26: Scatterplots between V200 (X) and V102 (Y) NDVI_10M products for the period 01/07/2015 – 26/03/2020. Red line is the orthogonal geometric mean regression line between the two versions.

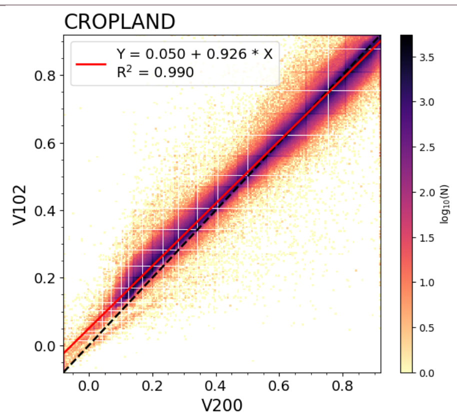
All land cover



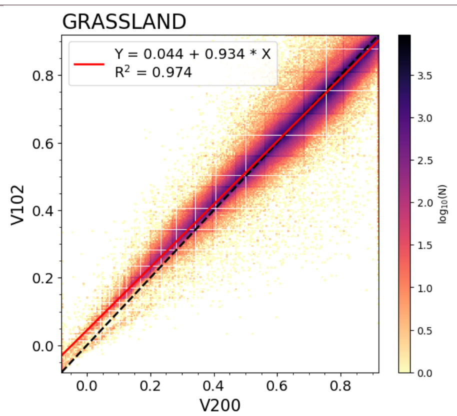
Built-up



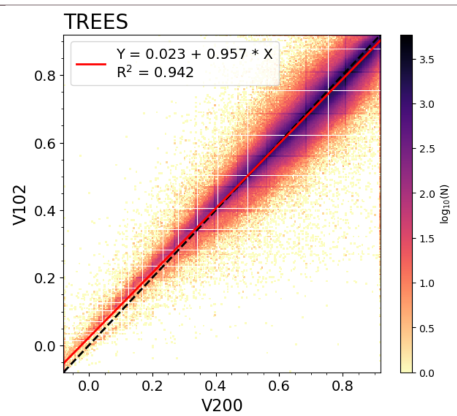
Cropland



Grassland



Tree cover



Examples of this are shown in Figure 27 for 3 specific cases: (1) AOT between V200 and V102 is similar, (2) AOT is much lower for V200 and V102, and in both versions the AOT range in the image is small,

and (3) the AOT is very different between both version and has a large range in V102 but a small range in V200. In the first case, the NDVI is very similar between V200 and V102 and follows the general pattern as discussed above. A large difference in AOT as in case (2) results in a constant offset of the linear relationship and in more scatter. The third case results in larger scatter between the 2 versions, but does not alter the general relationship. These results suggest that AOT differences between the two versions have an impact on NDVI_10M, which results in more scatter and a constant offset between the NDVI versions. It demonstrates that NDVI is sensitive to accurate AOT estimation.
It should be noted that the inter-comparison merely demonstrates differences and analyses their causes, but it does not indicate which data set is more accurate.

Figure 27: Comparison of scatterplots for 3 dates with variable difference in AOT estimated in V200 and V102 (columns). Top row: NDVI_10M, bottom row: AOT.

	Small difference in AOT (23/04/2018)	**Large difference in AOT (28/05/2016)**	**Large range of difference in AOT (21/04/2018)**

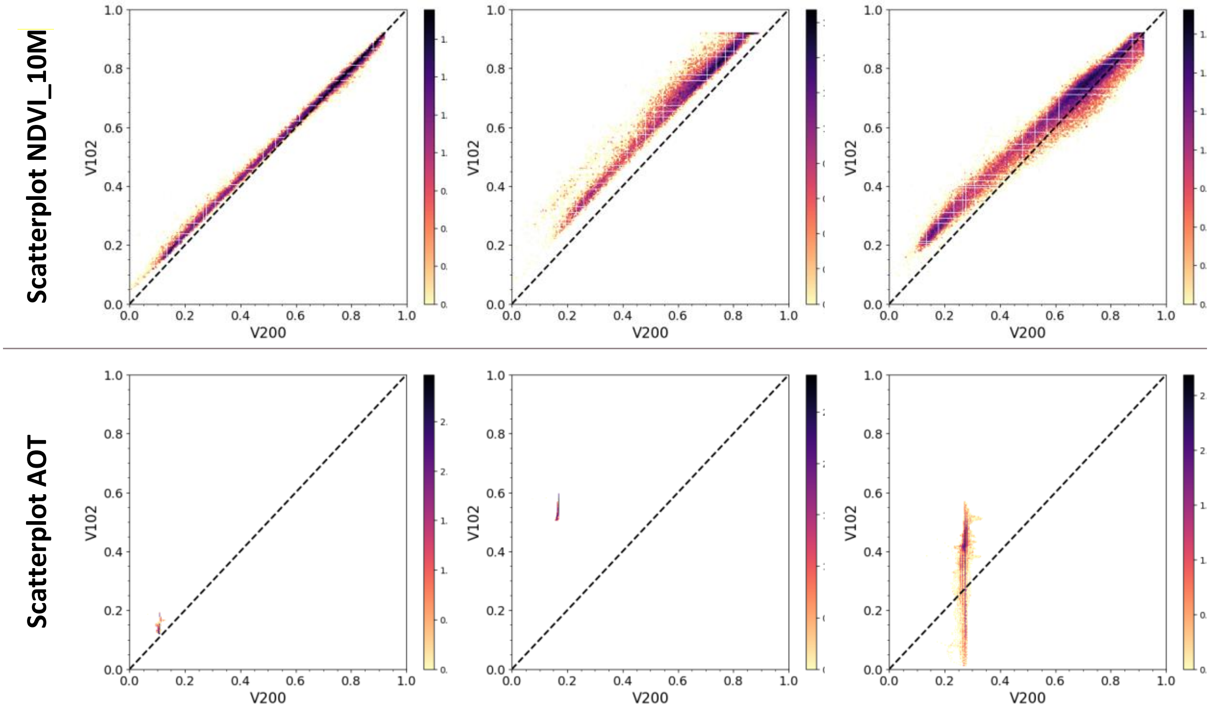
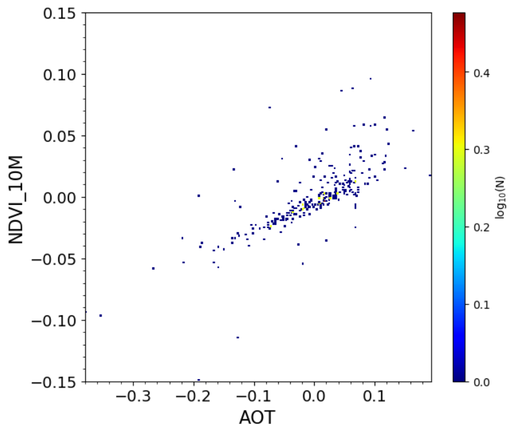


Figure 28 shows the scatterplot between the MBE calculated between V200 and V102 (MBE=V200-V102) of AOT and NDVI_10M. It demonstrates that if the AOT of V200 is lower than V102 (MBE AOT < 0), then the MBE of NDVI is also lower in V200 (MBE NDVI < 0).

Figure 28: Scatterplot between Mean Bias Error between V200 and V102 of AOT and NDVI_10M. MBE = V200-V102.

NDVI_10M



6.2 Histograms and bias histograms

Histograms and bias histograms for V200 and V102 NDVI_10M are shown in Figure 29 for all land cover in 31UFS and for the 4 major land cover classes. The bias is V200 minus V102, thus if the peak of the bias is negative, then V102 has overall higher values compared to V200.
Looking at all land cover, the range of the NDVI V200 values is similar compared to V102, but more low and high values are observed. In the bias histogram of NDVI_10M, the peak is around 0, with 83% of the pixels between [-0.035, 0.035] , and 95.6% between [-0.055, 0.055].
When looking at the results per land cover class, the following is observed:

- 'Built-up': there is a shift to lower NDVI values. The peak of the bias histogram is negative and around -0.02.
- 'Cropland': the bias histogram shows a peak around 0 and one around -0.02. The majority of the pixels have a negative bias (V200 NDVI < V102 NDVI).
- 'Grassland': the peak of the bias histogram is also around 0, with a rather symmetrical distribution at both sides.
- 'Tree cover': the results look very similar, but opposite as those of 'cropland' with a peak around 0 and 0.02.

The above results indicate that difference between the V200 and V102 of NDVI_10M shows different biases for the various land cover types.

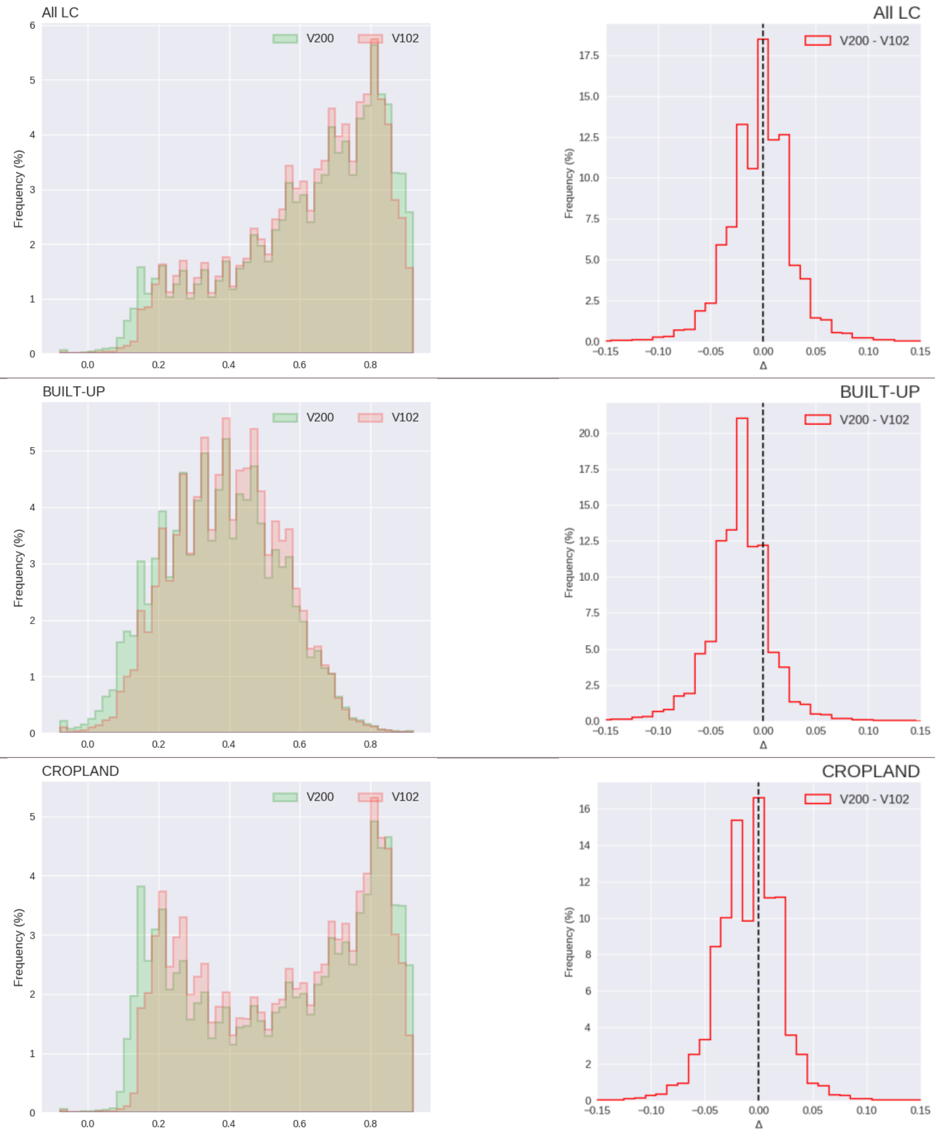
Figure 29: Histograms and bias histograms for V200 and V102 NDVI_10M products over the period 01/07/2015 – 26/03/2020. Bias is calculated as V200-V102.

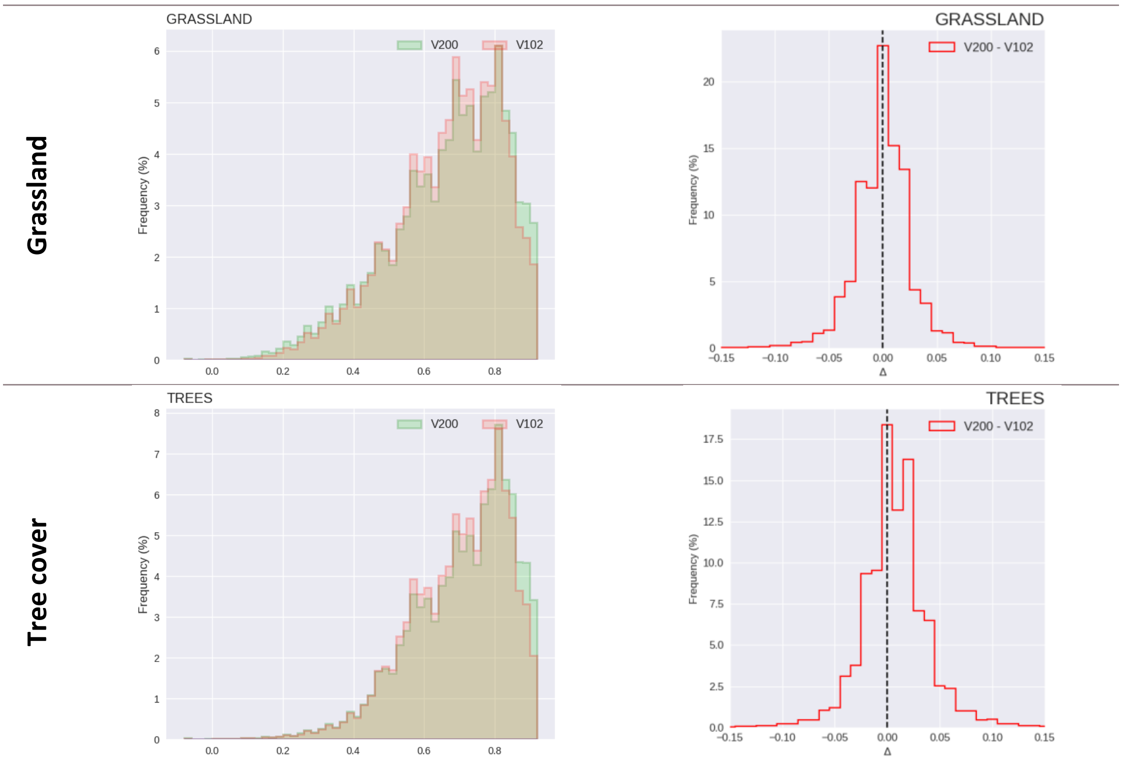
NDVI_10M

All land cover

Built-up

Cropland





6.3 Temporal analysis of agreement and differences

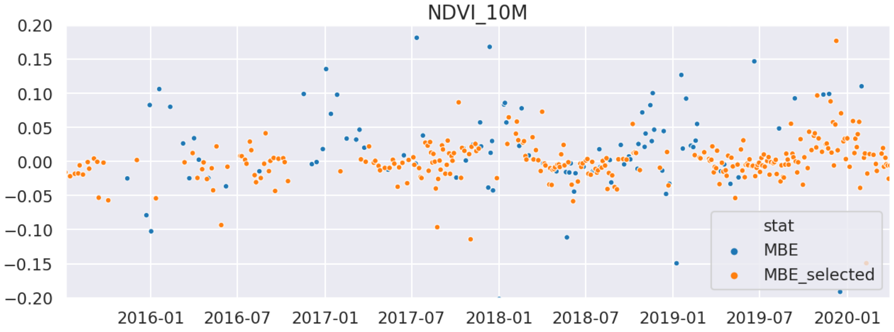
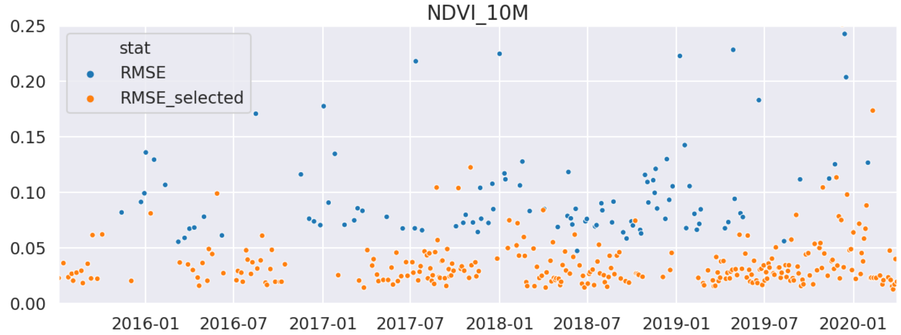
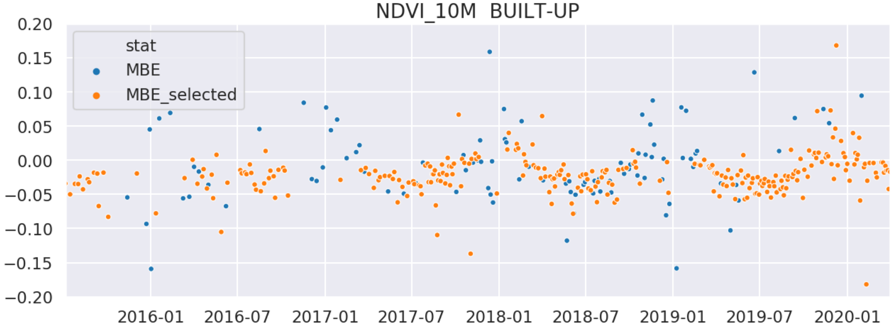
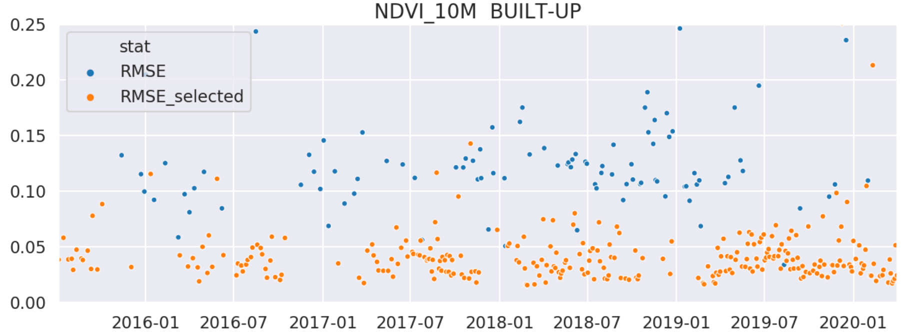
This section describes the result of the temporal analysis. Per date, the NDVI image of V200 is compared to that of V102 according to the sampling strategy described in section 2.2.2. For each date, a number of statistics is calculated and these are presented as temporal plots. This is done for all land cover types and per biome and presented for MBE and for RMSE in Figure 30Figure 17.

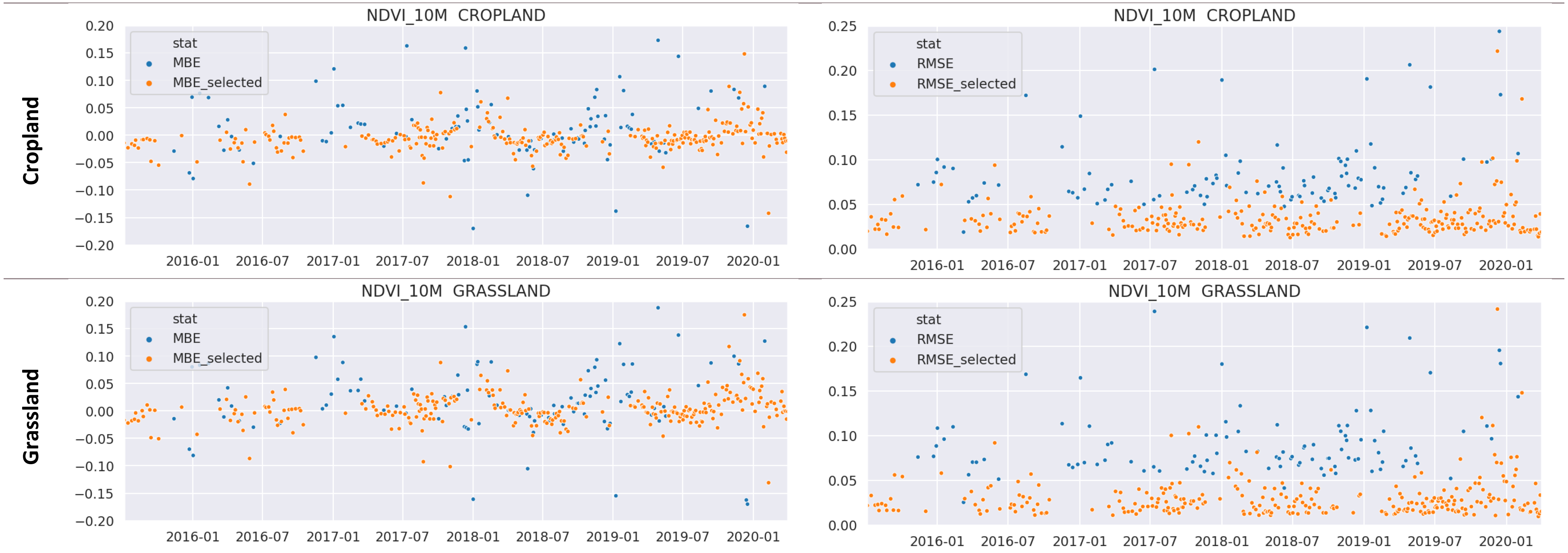
For the majority of the dates, the MBE of NDVI_10M varies between -0.05 and 0.05. There are however a significant amount of dates for which the MBE is beyond that range. The predominant MBE range is smallest for 'built-up' and increases with more vegetation cover.

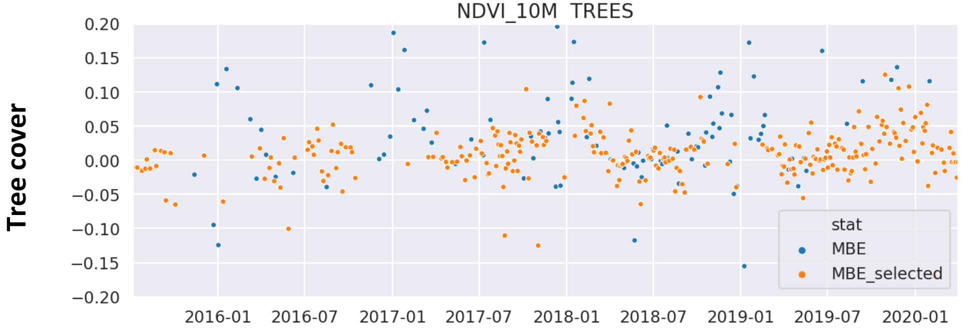
The MBE shows a seasonal pattern, with positive MBE in winter and negative MBE in summer. This is the case for all land cover types. When looking at the temporal evolution of the MBE between the AOT V200 and V102, the same seasonality is observed (see [RD-3]). The majority of the AOT MBE ranges between [-0.15, 0.15], with positive values in winter and negative values in summer. This is consistent with the analysis shown in Figure 30.

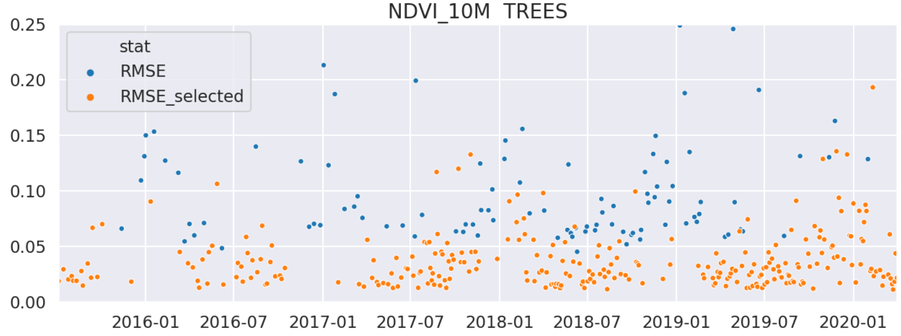
The RMSE between V200 and V102 of NDVI_10M varies roughly between [0.0, 0.08]. The RMSE range is largest for 'grassland' and 'tree cover, and smallest for 'cropland'. There is a substantial amount of dates with RMSE above 0.08, even up to 0.25. Most of these dates are cloudy (small sample size) or the V102 data have been manually corrected for a geolocation shift. A weak seasonal pattern is also visible in the RMSE temporal plots.

Figure 30: Temporal plots of Mean Bias Error (MBE) (left) and Root Mean Squared Error (RMSE) (right) between V200 and V102 NDVI_10M (right) for all land cover types (top row) and per biome (next rows). MBE are all dates and MBE_selected are all dates with sufficient sample size and where a geometric shift exists between V200 and V102.

	MBE NDVI_10M	**RMSE NDVI_10M**
All land cover		
Built-up		


Cropland
NDVI_10M CROPLAND
stat
MBE
MBE_selected
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.20
2016-01
2016-07
2017-01
2017-07
2018-01
2018-07
2019-01
2019-07
2020-01
NDVI_10M CROPLAND
stat
RMSE
RMSE_selected
0.25
0.20
0.15
0.10
0.05
0.00
Grassland
NDVI_10M GRASSLAND
stat
MBE
MBE_selected
NDVI_10M GRASSLAND
stat
RMSE
RMSE_selected

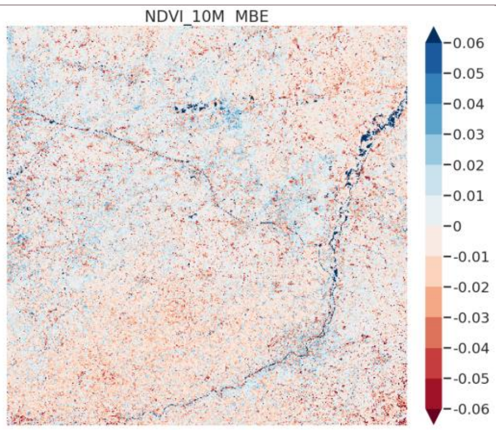
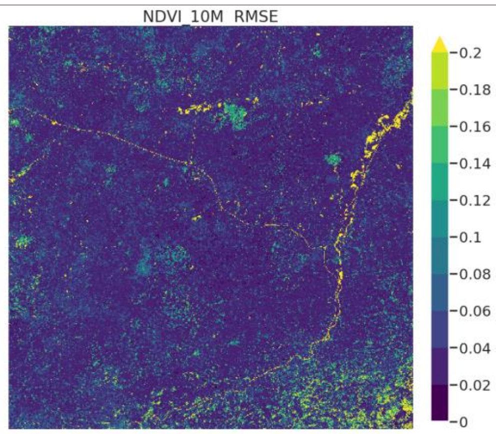
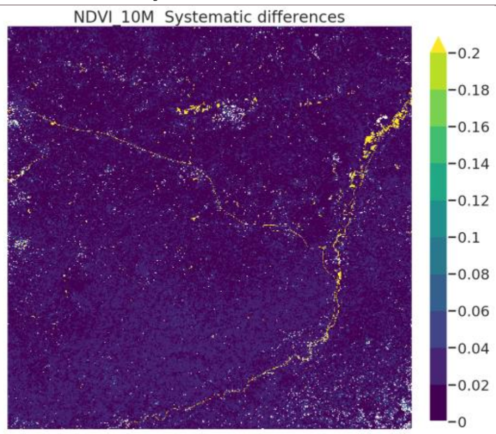
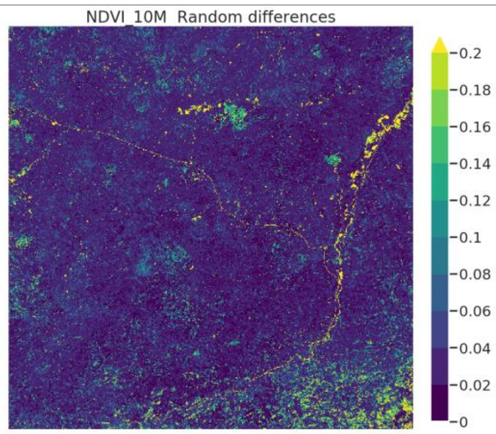

NDVI_10M TREES
Tree cover
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.20
2016-01
2016-07
2017-01
2017-07
2018-01
2018-07
2019-01
2019-07
2020-01
stat
MBE
MBE_selected


NDVI_10M TREES
0.25
0.20
0.15
0.10
0.05
0.00
2016-01
2016-07
2017-01
2017-07
2018-01
2018-07
2019-01
2019-07
2020-01
stat
RMSE
RMSE_selected

6.4 Spatial patterns of differences

The spatial patterns of the per-pixel MBE, RMSE, systematic and random differences between V200 and V102 LAI_20M and 10M are shown in Figure 31.
A number of spatial patterns can be discerned in the spatial plots of all the metrics for the 10M NDVI. The majority of the pixels show moderate differences. The MBE is negative for areas that are dominated by ‘cropland’ and ‘built-up’ and positive for ‘tree cover’. The class ‘grassland’ shows both positive and negative values. The largest differences (MBE and RMSE) are associated with rivers and river valleys (positive MBE) and in the southeast corner of the image (negative bias).
When looking at the split between systematic and random differences, it is observed that the random differences prevail over the systematic ones, except for the water bodies (rivers) where both differences are high.

Figure 31: Spatial pattern of the per-pixel Mean Bias Error (MBE), Root Mean Squared Error (RMSE), Systematic difference and Random difference between V200 and V201 of NDVI_10M.

MBE	**RMSE**
	
Systematic diff.	**Random diff.**
	

6.5 Conclusion for NDVI_10M

The full time series until 31/03/2020 of tile 31UFS of the TERRASCOPE NDVI_10m V200 was compared to the previous version V102 in order to demonstrate the agreement and the differences. Overall, NDVI V200 and V102 agree linearly and with a reasonable amount of scatter. The peak of the bias histogram is around 0 when considering all pixels. This however depends largely on land cover type: negative for 'built-up' and then moving towards positive bias when vegetation amount increases. The range of the NDVI values is similar among versions, but NDVI_10M V200 has more low and more high values.

A large influence of the difference in AOT was observed. Large differences in AOT used in V200 and V102 are responsible for:

- A systematic bias between the two NDVI versions
- More scatter in the relationship between the two NDVI versions
- Seasonal pattern in difference the two NDVI versions

The spatial analysis demonstrated that the topographic correction used in the atmospheric correction of V200 TOC reflectances has an impact on the data: larger differences for areas with more topography. As the analysis focused on the differences between the two versions, it was not possible to judge which version is more accurate. We can assume that the topographic correction should be an improvement in the processing.

It should be mentioned that the impact of geometric errors was not assessed. This is usually expressed in random differences rather than systematic ones. Large differences between V200 and V102 NDVI were observed for those dates for which a geolocation shift was applied in V102 and not in V200. The difference metrics were significantly larger than the ones observed for the images that have the same geolocation in both versions.

It is advised not to mix the two versions of the NDVI in the same analysis, because this could lead to erroneous conclusions.

7 General conclusion

The full time series until 31/03/2020 of tile 31UFS of the TERRASCOPE BIOPARs and NDVI V200 were compared to their previous version V102 in order to demonstrate the agreement and the differences.

Overall, all BIOPARs and NDVI show a good linear agreement between V200 and V102. For LAI, this was the case for LAI values up to 5. For higher LAI values, V200 was significant larger than V102. The amount of scatter or noise in the comparison is the highest for NDVI. The 10 m BIOPARs consistently had larger scatter compared to the 20M data.
The bias between the two versions was assessed. For the 20 m BIOPARs, the peak was slightly negative and the bias did not change among different land cover type. The opposite was observed for the 10 m data sets: the bias shifted from negative values to positive ones with increasing vegetation cover.

A large influence of the difference in AOT was observed. Large differences in AOT used in V200 and V102 are responsible for:

- Non-linear relationship between the two versions of the BIOPARs
- More scatter in the relationship between the two versions of BIOPARs and NDVI
- Seasonal pattern in the difference between the two BIOPARs and NDVI versions
- A systematic bias between V200 and V102 for NDVI

The spatial analysis demonstrated that the topographic correction used in the atmospheric correction of V200 TOC reflectances has an impact on the BIOPAR data: larger differences for areas with more topography. For NDVI, this was not clearly observed. As the analysis focused on the differences between the two versions, it is not possible to judge which version is more accurate. We can assume that the topographic correction should be an improvement in the processing.

The use of per-pixel Sun angles instead of mean Sun angles per image in the BIOPARs retrieval is not discernable from the results.

It should be mentioned that the impact of geometric errors was not assessed. This is usually expressed in random differences rather than systematic ones. Large differences between V200 and V102 data were observed for those dates for which a geolocation shift was applied in V102 and not in V200. The difference metrics were significantly larger than the ones observed for the images that have the same geolocation in both versions.

It is advised not to mix the two version of the BIOPARs or NDVI in the same analysis, because this could lead to erroneous conclusions.

8 References

De Keukelaere, L., Sterckx, S., Adriaensen, S., Knaeps, E., Reusen, I., Giardino, C., Bresciani, M., Hunter,P., Neil, C., Van der Zande, D. & Vaiciute, D. (2018). Atmospheric correction of Landsat-8/OLI and Sentinel-2/MSI data using iCOR algorithm: validation for coastal and inland waters. European Journal of Remote Sensing., Vol 51, 525-542. https://doi.org/10.1080/22797254.2018.1457937

Doxani, G.; Vermote, E.; Roger, J.-C.; Gascon, F.; Adriaensen, S.; Frantz, D.; Hagolle, O.; Hollstein, A.; Kirches, G.; Li, F.; Louis, J.; Mangin, A.; Pahlevan, N.; Pflug, B.; Vanhellemont, Q. Atmospheric Correction Inter-Comparison Exercise. *Remote Sens.* **2018**, *10*, 352. https://www.mdpi.com/2072-4292/10/2/352

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

Louis, J., Devignot, O., Pessiot, L. (2017) S2 MPC Level 2A Product Format Specification. Ref. S2-PDGS-MPC-L2A-PFS-V14.2. https://sentinel.esa.int/documents/247904/685211/Sentinel-2-MSI-L2A-ProductFormat-Specifications.pdf

Mueller-Wilm, U., Devignot, O., Pessiot, L. (2018) S2 MPC Sen2Cor Configuration and User Manual. Ref. S2-PDGS-MPC-L2A-SUM-V2.5.5. http://step.esa.int/thirdparties/sen2cor/2.5.5/docs/S2-PDGS-MPC-L2A-SUMV2.5.5_V2.pdf

Mueller-Wilm, U., Devignot, O., Pessiot, L. (2019) S2 MPC Sen2Cor Software Release Note. Ref. S2-PDGS-MPC-L2A-SRN-V2.8.0. http://step.esa.int/thirdparties/sen2cor/2.8.0/docs/S2-PDGS-MPC-L2A-SRNV2.8.pdf