



# globdiversity

rs-enabled ebvs for  
terrestrial ecosystems

## Algorithm Theoretical Basis Document

### Land Surface Phenology

Document Ref:	ATBD-LSP
Version:	3.4
Contract No:	4000120011/17/I-NB
Creation Date:	2018-05-03
Last Modified:	2020-09-30

Submitted by



University of  
Zurich<sup>UZH</sup>



UNIVERSITY OF TARTU



WAGENINGEN  
UNIVERSITY & RESEARCH



This page intentionally left blank

### Distribution List

This GlobDiversity document is currently stored at the Dropbox are under the folder  
Dropbox\03\_Tasks\03\_Task3\01\_ATBD

This document is accessible to all participants of the GlobDiversity project.

### Summary

This document refers to the activities of task 2 and 3, including sub packages. This document is a draft and the version 4.0 of the Algorithm Theoretical Baseline Document (ATBD).

### Document Release Sheet

Authors:	Claudia Rösli (UZH) Rogier de Jong (UZH) Vladimir Wingate (UZH) David Marshall (DLR) Uta Heiden (DLR) Rubens van de Kerchove (VITO)	
Approval (internal)	Claudia Rösli (UZH) Vladimir Wingate (UZH)	
Approval (ESA)	Marc Paganini (ESA)	
Distribution:		

## Change Record

#	Date	Page(s)	Change record	Release
0	2018-05-04		First document draft	0.1
1	2018-09-10		V1	1.0
2	2018-10-12		Added detection n.o. growing seasons and approach for uncertainty estimation after MTR	1.1
3	2019-01-11		Added comments/changes to timing	2.0
4	2020-01-30		Edited Chapters 4-7 to reflect the prototype design in Python3	3.0
5	2020-03-17		Edited and updated all sections to reflect latest prototype. Added validation sections to chapter 6	3.1
6	2020-03-29		Formatting	3.2
7	2020-06-17		Complete Review and update with Results and summary	3.3
8	2020-09-30		Integration of the Upscaling results – FINAL version	3.4

## Terms, Definitions and Abbreviations

<b>ATBD</b>	Algorithm Theoretical Baseline Document
<b>BoA</b>	Bottom of Atmosphere
<b>CBD</b>	Convention on Biodiversity
<b>DOY</b>	Day of year
<b>ESA</b>	European Space Agency
<b>GCC</b>	Green Chromatic Coordinates
<b>GEO BON</b>	Group on Earth Observations Biodiversity Observation Network
<b>IPBES</b>	Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services
<b>LSP</b>	Land Surface Phenology
<b>NDVI</b>	Normalized Difference Vegetation Index
<b>NIR</b>	Near-infrared
<b>RS-enabled EBV</b>	Remotely Sensed Essential Biodiversity Variables
<b>ToA</b>	Top of Atmosphere
<b>VI</b>	Vegetation Index
<b>ITC</b>	Faculty of Geo-Information Science and Earth Observation (University of Twente)
<b>WEnR</b>	Wageningen Environmental Research, Wageningen University & Research
<b>UZH</b>	University of Zurich
<b>UTM</b>	Universal Transverse Mercator (coordinate system)
<b>DLR</b>	German Aerospace Center
<b>EVI</b>	Enhanced Vegetation Index
<b>fAPAR</b>	fraction of Absorbed Photosynthetically Active Radiation
<b>MODIS</b>	Moderate Resolution Imaging Spectroradiometer
<b>Mp</b>	Midpoint

## Table of Contents

<b>1.</b>	<b>Introduction</b> .....	<b>1</b>
<b>2.</b>	<b>Scientific background</b> .....	<b>3</b>
<b>3.</b>	<b>Input Data</b> .....	<b>4</b>
3.1.	Sentinel-2 input data set.....	5
3.2.	Landsat 8 input data set .....	6
3.3.	Land cover map .....	6
3.4.	Data fusion of different data sets – Landsat 8 and Sentinel-2 satellite images .....	6
<b>4.</b>	<b>Algorithm description Land Surface Phenology</b> .....	<b>8</b>
4.1.	Overview .....	8
4.2.	LSP core algorithm.....	9
4.2.1.	Extraction of LSP metrics .....	9
4.2.2.	Estimation of uncertainty .....	11
4.2.3.	Detection number of growing seasons.....	12
4.2.4.	Flagging .....	13
4.2.5.	Output .....	14
<b>5.</b>	<b>Product</b> .....	<b>16</b>
5.1.	Product description .....	16
5.2.	Accuracy analysis .....	17
5.3.	Validation Method .....	18
5.4.	Validation results and discussion .....	19
<b>6.</b>	<b>Remarks on prototyped algorithm development and processing</b> .....	<b>24</b>
6.1.	Fitting procedure .....	24
6.2.	Cloud and invalid pixel detection algorithm and high latitude areas of interest .....	24
6.3.	Memory requirements .....	25
6.4.	Constraints from neighboring pixels .....	25
6.5.	Availability of L2-images.....	25
6.6.	Number of available observations .....	25
6.7.	Time stamp .....	25
6.8.	Reduction to area of interest, naturally vegetated areas.....	26
<b>7.</b>	<b>Upscaling results</b> .....	<b>27</b>

7.1. Introduction .....	27
7.2. Results.....	28
<b>8. References.....</b>	<b>33</b>
<b>9. Appendix.....</b>	<b>35</b>

## List of Figures

Figure 1: Flow chart pre-processing and merging Sentinel-2 and Landsat 8 into NDVI time series as input for the LSP processor. Red boxes indicate data products and blue boxes indicate processing steps. The processing steps were completed using the gdal tools from the Open Source GEOSpatial Foundation. ....	5
Figure 2: Double-logistic curve fitted onto NDVI-pixel values. Explanation of parameters as defined in Equation ( 3 ).....	8
Figure 3: Flow-chart of main processing procedure from pixel based time series until the extraction of phenological parameters and drawing of maps.....	9
Figure 4: Definitions of different parameter used for phenology description.....	11
Figure 5: Double-logistic curve fitted onto NDVI-pixel values. Explanation for measure of quality. The RMSE during green-up and senescence state is based on the residuals in days of year between observation and modelled curve. RMSE for dormancy and peak state is based on residuals in NDVI between model and observations. ....	12
Figure 6: Schematic detection of double growing season based on the midpoint value. If several consecutive points lye lower than the midpoint value with afterwards consecutive points lying above, a multi-growing season time-series is detected. Picture from Garonna (2016).....	13
Figure 7: The print screen shows the statistics for the output file with 23 layers (or bands). As here, different layers that can be illustrated in a histogram. The histogram for Band 9, the growing season length, is displayed with a mean of 121 days and a range from 3 to 184. The band list is the same as listed in Table 4. ....	16
Figure 8: The overview image shows the results of SOS, EOS, Amplitude and RMSE for a part of the Kytalyk NP. The colorscale is adjusted depending on the histogram and existing values. In the image the structures of topography and vegetation are well visible. White pixels correspond to invalid pixels, mostly related to water. ....	17
Figure 9: LSP metrics for the four sites (Laegern, Kytalyk NP, Toolik Lake and Bavaria NP), computed using the phenocam GCC (column 1), satellite GCC (column 2) and satellite NDVI time series (column 3).....	20
Figure 10: Bar plots of LSP metrics SOS, EOS, GSL, computed using satellite NDVI, Satellite GCC, and phenocam GCC time series, for each of the four sites (Laegern, Kytalyk, Toolik Lake and Bavaria NP), for the years 2017 (all sites) and 2018 (Toolik Lake and Bavaria NP).....	22
Figure 11: 2018 Corine land cover product for Finland at 20 m resolution available from <a href="https://www.syke.fi/en-US/Open_information/Spatial_datasets/Downloadable_spatial_dataset#C">https://www.syke.fi/en-US/Open_information/Spatial_datasets/Downloadable_spatial_dataset#C</a> .....	27
Figure 12: CCI 2016 20 m land cover prototype over Senegal available from <a href="http://2016africallandcover20m.esrin.esa.int/">http://2016africallandcover20m.esrin.esa.int/</a> .....	28
Figure 13: Per pixel LSP metrics over Finland: Start of Season (SOS, DoY), End of Season (EOS, DoY) dates, Growing Season Length (GSL, DoY) and Amplitude (AMP). The visible diagonal is due to the overlap in neighboring UTM zones. ....	30
Figure 14: Per pixel LSP metrics over Senegal: Start of Season (SOS, DoY), End of Season (EOS, DoY) dates, Growing Season Length (GSL, DoY) and Amplitude (AMP).....	31

Figure 15: Per pixel LSP metrics over a region (Karkkila, 60°31'57.9"N 24°12'42.9"E) in Finland (see Figure 13 for the legend): Start of Season (SOS), End of Season (EOS) dates, Growing Season Length (GSL) and Amplitude (AMP). Additionally also the Corine Land Cover (see Figure 11 for the legend) and a Bing maps image is given as reference.....32

### List of Tables

Table 1: Number keys for SCL-layer of Sentinel-2 image files classified with sen2cor .....	5
Table 2: Excluded flags for processing for the Landsat 8 images, derived from pixel-qa layer .....	6
Table 3: Binary number corresponding to the flagging system.....	14
Table 4: Layers stored in the output GeoTiff file and their definitions. ....	14
Table 5 Values for each metric (SOS, EOS, and GSL), computed with each time series (phenocam GCC, satellite GCC and satellite NDVI), and the difference between phenocam GCC and satellite GCC, as well as phenocam GCC and satellite NDVI, , for sites in 2017 and 2018.....	20
Table 6: summary of the CCC upscaling processing results .....	28

## 1. Introduction

This Algorithm Theoretical Baseline Document (ATBD) describes all technical issues from the prototyping of the Land Surface Phenology (LSP) in the context of the Remotely Sensed Essential Biodiversity Variables (RS-enabled EBVs) product of the ESA funded GlobDiversity Project. This document shall specify the process flow of the prototyped algorithm and the associated program in more detail.

GlobDiversity is the first large-scale project explicitly designed to develop and engineer RS-enabled EBVs. This project initiated and funded by the European Space Agency (ESA) supports the efforts of the Convention on Biological Diversity (CBD) and Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), among others, and is adopted under the umbrella of the Group on Earth Observations Biodiversity Observation Network (GEO BON). The GlobDiversity project shall support the initiative to build a global knowledge of biological diversity of terrestrial ecosystems (= on land) and of relevance for society.

There are three RS-enabled EBVs designed as part of the GlobDiversity project with each algorithm documented by such an ATBD:

- Fragmentation (Wageningen Environmental Research WEnR, Wageningen University & Research)
- Canopy chlorophyll concentration (Faculty of Geo-Information Science and Earth Observation ITC, University of Twente)
- Land surface phenology (Dept. of Geography, University of Zurich (UZH), the hereby documented algorithm)

Within the project, these three variables were investigated in detail to contribute to an observation system to assess the variable in an efficient and effective way, covering extensive areas at a fine spatial and temporal resolution. The definition and selection, name and definition of the three RS-enabled EBVs was based on the expertises existing within the project consortium and independent from any efforts of defining and prioritising possible candidate EBVs and RS-enabled EBVs that might have existed at the time of the project's start in 2018.

In the following, the algorithm of the processing chain to derive Land Surface Phenology (LSP) is described in detail. The algorithm was chosen and developed by the University of Zurich and then transmitted to the German Aerospace Center (DLR) to be translated into a code suitable for cloud computing of larger areas of interest. The algorithm has been chosen and developed with the goal of a potential future global application based on high-resolution satellite data (10-30m) and with a computational efficient implementation. The ATBD includes a description of the necessary pre-processing steps and the processing step of the core algorithm. In addition, results from the project performed on few test sites globally distributed are presented with a chosen validation approach. In addition, the last chapter presents restrictions of the current implementation and modifications that might be necessary for a potential global processing.

The organization of this document is structured in 9 chapters as shown in the table below.

	<b>Content</b>
Chapter 1	Introduction
Chapter 2	Describes the scientific background, and addresses the current standard processing schemes
Chapter 3	Provides information about the input data and pre-processing needed
Chapter 4	Describes the implementation of the core algorithm with the processing chain
Chapter 5	Provides information and a summary about the data product and the validation
Chapter 6	Discusses practical considerations for implementation
Chapter 7	Upscaling results for Senegal and Finland
Chapter 8	Includes the references
Chapter 9	Includes an appendix

## 2. Scientific background

Land Surface Phenology (LSP) was first defined by Friedl et al. (2006) as “the study of seasonal patterns of vegetated land surfaces as observed from remote sensing”. The core term, “Phenology”, refers to the timing of recurring events in the life cycle of in particular plants as influenced by the environment (Lieth 1974). Typical studies on phenology focused on the timing of observable phases of plant life cycles – such as flowering, fruiting or leaf development, coloration and fall (Menzel and Sparks 2006). Single plant detection to estimate plant phenology requires very high spatial resolution data (<2m) that is not yet available free and open from satellite remote sensing. Even with such data, it is unclear if the phenological processes commonly understood by ecologist and plant-physiologists would be possible to be observed. Therefore, LSP focuses on the timing of seasonal activity of plants at ecosystem scale (Schwartz 2013) as observed from satellites. This measure is not identical to plant phenology but is strongly influenced linked, and its observational feasibility and maturity render LSP a valuable remotely sensed enabled Essential Biodiversity Variable (RS-enabled EBV).

A number of approaches have been used to derive metrics from time series of Vegetation Indices (VIs) to provide information about vegetation activity over large areas, in a systematic and continuous way. VIs are combinations of multiple spectral values used to infer the amount of vegetation within a pixel (Campbell and Wynne. 2011). Since the 1980s, many VIs have been put forward including: the Normalized Difference Vegetation Index (NDVI); the Enhanced Vegetation Index (EVI); and vegetation cover percentage or fraction of Absorbed Photosynthetically Active Radiation (fAPAR). Since then, many LSP studies have used VI time series to derive information about the annual timing of vegetation growth, senescence and dormancy at resolutions ranging from moderate (>250m) to coarse (>25km) (Friedl et al. 2006; White et al. 2009). Many studies (e.g., Myneni et al. 1997; Zeng, Jia, and Epstein 2011) have shown the asset of using the moderate-resolution imaging spectroradiometer (MODIS) for global phenology monitoring, however, recent studies show a better agreement between satellite-based phenology in comparison with in-situ data when using Sentinel-2 data in comparison with MODIS time series (e.g., Lange et al. 2017). Higher spatial-resolution satellite observations thus allow for a reduction of the gap between LSP and plant phenology.

For the current project, we use combined high-resolution optical information from Sentinel-2 satellites and Landsat 8. We base the computations on the widely used NDVI as it is independent of the sensor (unlike EVI) and simple to calculate (in comparison to fAPAR). In order to extract phenological timing, we will use a model to fit the yearly profile. We use the mathematical model of the double logistic function that is also widely used in phenology studies. The model is stable against noisy observations and can reproduce steep green-up and senescence phases. Because of these advantages, the model is suitable for global use as even Arctic tundra can be represented. Note, we developed the LSP algorithm for natural vegetation only and the results are not tested or validated for non-natural vegetation areas such as agricultural land. The output of the algorithm is a pixel based analysis of the phenology with a measure for accuracy. Results are tested against several criteria to ensure reliability of the result. The results are represented in a multi-layer, georeferenced map (GeoTiff) for the region of interest. The results were validation with ground-based observation from so-called phenocam observations, a time series of canopy images, using the same algorithm as for the satellite-based LSP products.

### 3. Input Data

LSP product is based on a pixel-wise analysis of the optical images of high-resolution satellites of Landsat 8 (30m) and Sentinel-2A&B (10m). The vegetation index (VI) of NDVI is used for the extraction of the phenological profile of a year and is calculated from the RED and the near-infrared (NIR) channel (see Equation ( 1 ) ). The BLUE and GREEN channel are additionally needed for the proposed validation approach that uses the Green Chromatic Coordinates (GCC, see Equation ( 2 ) ).

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

$$GCC = \frac{GREEN}{GREEN + BLUE + RED} \quad (2)$$

In order to reduce atmospheric influences on NDVI, bottom of atmosphere (BoA) reflectance is required for processing. Using BoA is crucial, in particular for the interpretation of the amplitude of vegetation activity and for high latitudes. When correcting L1C to L2A, influences of atmosphere are reduced.

A flow chart for the current implementation of the pre-processing is shown in Figure 1. The input data for the LSP core algorithm is the NDVI timeseries or more specifically a data cube with the NDVI time series together with the corresponding time vector. Therefore, the flagging of invalid pixels (see Chapter 3.1 to 3.3) and the merging of different data sets (see Chapter 3.4)

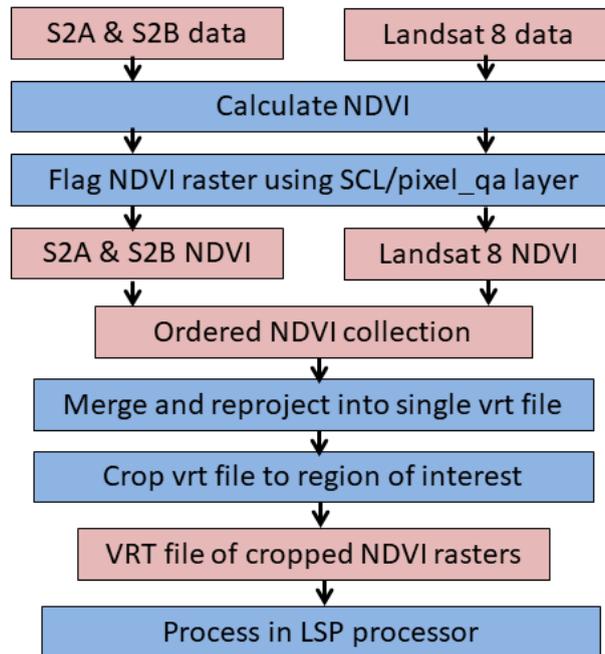


Figure 1: Flow chart pre-processing and merging Sentinel-2 and Landsat 8 into NDVI time series as input for the LSP processor. Red boxes indicate data products and blue boxes indicate processing steps. The processing steps were completed using the gdal tools from the Open Source GEOspatial Foundation.

### 3.1. Sentinel-2 input data set

For Sentinel-2 channels B02 (blue), B03 (green), B04 (red), and NIR (B08) with 10m spatial resolution are used. L2 processing is done with sen2cor as ESA does not yet provide L2A data sets before end 2018 globally<sup>1</sup> and the current project is focused on the years 2017 and 2018. In addition, flagging of sen2cor is used. The quality band SCL (to be found under /S2x\_MSIL2A\_\*.SAFE/GRANULE/L2A\_\* /IMG\_DATA/R20m/L2A\*\_SCL\_20m.jp2) is used for flagging and excluding invalid data. Pixels with the flags of 0, 1, 2, 3, 6, 8, 9, 11 are not used in the processing (Table 1). Note, the SCL is available for 20m resolution only, whereas the NDVI is calculated for 10m resolution. One 20m pixel, however, corresponds exactly to four 10m-pixels as the layers have aligned coordinate grids.

Sen2cor has been chosen due to the fact that ESA provides the L2A images on an operational basis starting from December 2018 and will be therefore more and more used. In addition, it is expected that the (cloud) flagging algorithm will be further developed and improved as more and more will be using this data set.

Table 1: Number keys for SCL-layer of Sentinel-2 image files classified with sen2cor

<sup>1</sup> <https://earth.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-2a-processing>

0	No_Data	6	water
1	Saturated_defective	7	unclassified
2	Dark_area	8	cloud_med_prob
3	cloud_shadow	9	cloud_high_prob
4	vegetation	10	thin_cirrus
5	bare_soil	11	snow/ice

### 3.2. Landsat 8 input data set

For Landsat 8, B2-B5 are in use (blue, green, red, NIR) as well as «pixel\_qa» for the cloud flagging. L2A data from the LaSCR processor are provided by NASA and can be downloaded for instance from Google Earth Engine. In Table 2, the flags that are excluded from the LSP processing are given:

*Table 2: Excluded flags for processing for the Landsat 8 images, derived from pixel-qa layer*

Excluded Flags					Reason
480	992				#high confid cloud
928	416				#mid confid cloud
904	392	840			#cloud shadow
324	388	836	900	1348	#water

### 3.3. Land cover map

For additional exclusion and reduction of processed data, land cover data (building and streets) could be used to flag the data (not implemented in the prototyped pre-processing steps). Ideally, a land cover map could distinguish between non-vegetated areas (e.g., buildings and streets), crop areas (not target area of this project) and water areas. Nevertheless, the land cover map should be in a similar resolution as the original data set of 10m. A high quality global data set does not exist yet but might exist in future (e.g., land cover project within ESA's Climate Change Initiative CCI). Streets and building can be extracted for instance from the open street map data base, and rasterized with the raster of the sentinel-scenes to flag all pixels that touch a vector object. This approach has been chosen from UZH to do the preprocessing, however, is not part of the core LSP algorithm presented here.

### 3.4. Data fusion of different data sets – Landsat 8 and Sentinel-2 satellite images

Within the here presented work, Landsat 8 and Sentinel-2A and -2B were used together to have the maximum possible number of observations. Landsat 8 is available with 30m resolution and Sentinel-2 with 10m for the red, green, blue (RGB) and NIR bands. Landsat 8 images are available with a 16-day temporal resolution in comparison with 5-6 days for the tandem of Sentinel-2A&B. We therefore decided to be based on the higher resolution of the Sentinel-2 images and down-sampled the Landsat 8 pixels with a nearest-neighbor approach. The

downsampling and reprojection of the Landsat 8 images to that of the Sentinel-2 images is done during the creation of the vrt file. This step, performed using the gdal command *gdalbuildvrt*, allows the user to set the desired resolution (in this case the Sentinel-2 resolution) and to stack the rasters in a datacube. Each row of the datacube then contains an NDVI image and the columns act as the required time series for the phenological calculations. The result showed that the effect of mixed pixels of the few available Landsat 8 observations can be neglected due to the dominating effect of the atmosphere in the time series. In addition, it is expected that misregistration between Landsat 8 and Sentinel-2 will be corrected and same image registration will be used (Storey et al., 2016). The harmonization, however, of the two data set also through project, such as proposed by Claverie et al. (2018) were not ready yet and in particular the 10m resolution of the Sentinel-2 is higher weighted than the difference in NDVI value caused by the misalignment. Nevertheless, a future user can choose the preprocessing steps as the core algorithm expects an NDVI time series or any time series (i.e., also fAPAR or EVI) per pixel for the processing.

## 4. Algorithm description Land Surface Phenology

### 4.1. Overview

Land surface phenology (LSP) describes the yearly activity profile of vegetation and the characteristics of this curve can be extracted and visualized in a map. A dense time series is required in order to reliably fit the vegetation activity. Within this project, the trade-off analysis showed that a double-logistic model is best suited to represent the yearly LSP (from yearly data sets). Beforehand, double vegetation seasons need to be detected, and the time of interest for one season (one double-logistic curve) will be used. Consecutive years are treated and processed separately.

So far, high-resolution ( $\leq 30\text{m}$ ) LSP assessment were rarely performed as with Landsat-only the time series is too sparse (16 days temporal resolution on average) to give stable and reliable results. Only in combination with Sentinel-2A and 2B is a yearly phenology curve dense enough to give reliable fitting results for most regions of the Earth. This also means that a high-resolution yearly data set is required for LSP monitoring. The combination of Sentinel-2 and Landsat 8 observation is ideal for these purposes.

We decided to use the double logistic representation with 6 unknown variables as in Equation ( 3 ) and depicted in Figure 2. Within the double logistic function represents  $v_1$  the vertical offset of the curve,  $v_2$  represents the amplitude of the function (high activity – dormancy), and  $v_3$  &  $v_5$  represent the slope of the curve at the turning points with the x co-ordinates  $v_4$  and  $v_6$  (the times of each turning point).

$$f(x) = v_1 + \frac{v_2}{1+e^{-v_3*(t-v_4)}} - \frac{v_2}{1+e^{-v_5*(t-v_6)}}, \quad (3)$$

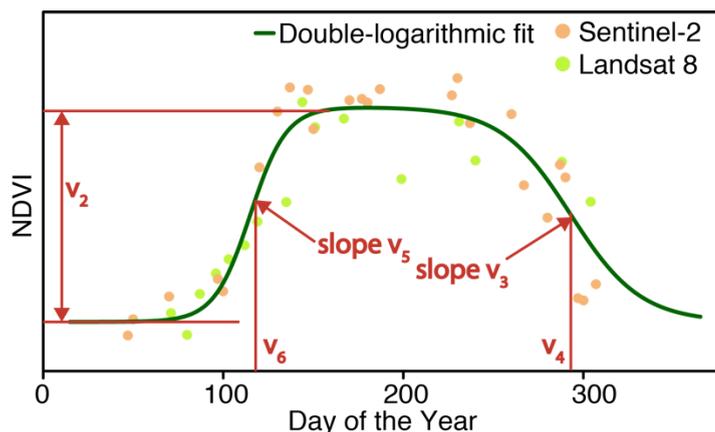


Figure 2: Double-logistic curve fitted onto NDVI-pixel values. Explanation of parameters as defined in Equation ( 3 ).

This same representation of the double-logistic curve has been used in several studies (e.g., Beck et al. 2006; Meroni et al. 2014). In comparison, there are also studies using 7 unknowns in order

to account for a summer decrease in canopy greenness (Elmore et al. 2012; Wu et al. 2017), where the double-logistic function is combined with a linear representation of the slope between the two branches of the logarithmic functions. Another possibility would be to use a seventh parameter to account for different baselines in each dormancy phase, as used by Fischer 1994. Nevertheless, we assume that the dormancy level (or winter NDVI) is constant between two consecutive years. This should be validated in future studies when a reasonable time series is available for Sentinel-2 observations. This assumption also precludes using “winter NDVI” values as used for instance in Beck et al. 2006 for MODIS time series.

## 4.2. LSP core algorithm

The NDVI time series per pixel (pre-processing see Chapter 3) will be processed individually according to the flow diagram in Figure 3. The processing includes the iterative double-logistic fit (Chapter 4.2.1), detection of double-growing season (Chapter 4.2.3), extraction of measures of uncertainty (Chapter 4.2.2), flagging procedure (Chapter 4.2.4) and extraction of the final product (Chapter 4.2.5).

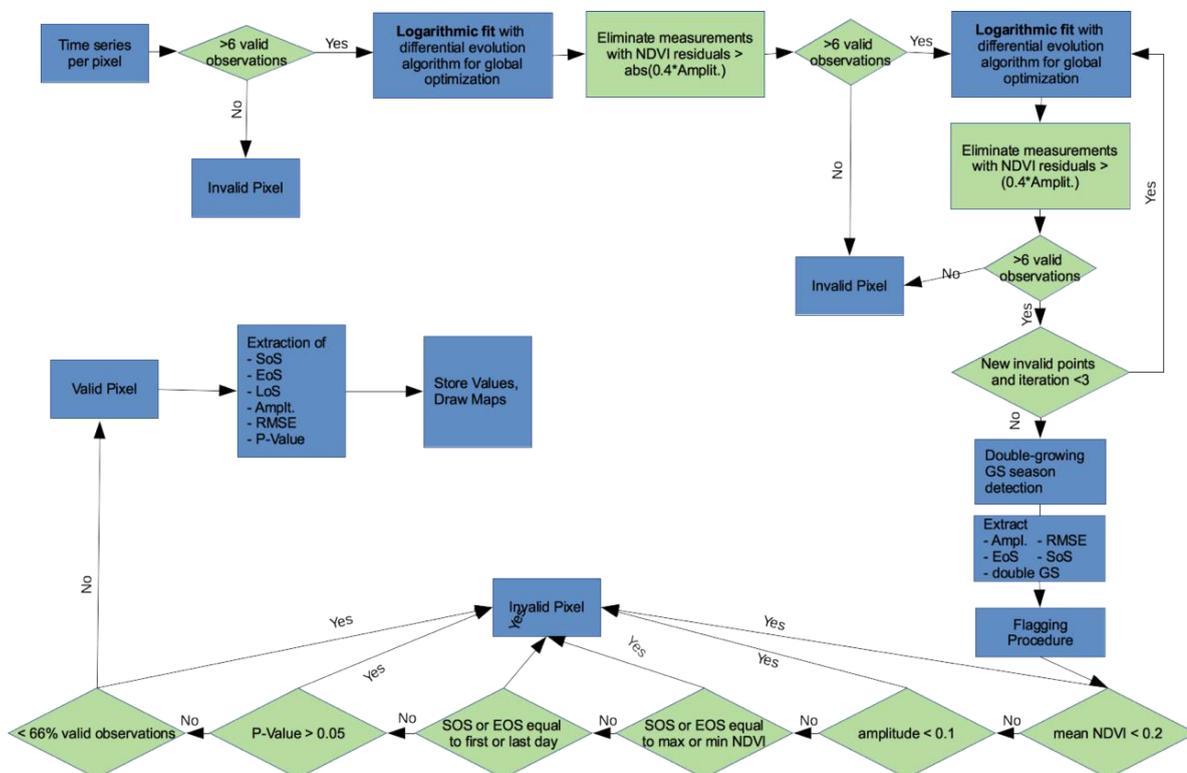


Figure 3: Flow-chart of main processing procedure from pixel based time series until the extraction of phenological parameters and drawing of maps.

### 4.2.1. Extraction of LSP metrics

The core of the algorithm is the double-logistic fit. As the classification algorithm from Sentinel-2 and Landsat 8 L2-products still has some undetected outliers, e.g., from undetected clouds, an iterative process is needed. After each individual processing iteration, the algorithm tests whether there are more than 6 valid observations available, as this is the number of unknowns in the double-logistic model. If this criterion is not met, the pixel is directly flagged as invalid in the final product and will not be processed further.

The fit procedure uses the python package *scipy optimize curve fit* with the *Trust Region Reflective algorithm* for the optimization. The values for the six variables are given with bounding regions to constrain the fit. If convergence is not successful, a nearest neighbor constraint can be implemented, whereby the fitting values for a successful neighboring pixel are used as a starting point for the curve fit. This can help to overcome some local minima in the parameter space but it can also lead to artefacts in the calculations with results smearing out across the calculation space. This feature is therefore optional. Nevertheless, due to the 6 unknowns of the double-logistic function and the high noise in the satellite observations, results of the fit can vary between different runs and convergence is sometimes slow or not stable. For more a more detailed discussion and recommendation regarding the used fitting procedure, see the outlook in Chapter 6.1.

The fitting is done in four iterations. The first iteration includes a fit of all valid observations after the L2-quality flagging. For observation with resulting residuals 40% larger than the absolute magnitude of the amplitude of the fitted curve ( $v_2$ ) are then flagged as invalid observations. In a next step, the fitting is performed up to three more times as long as there are still outliers, but only residuals that are below the phenological curve are eliminated (positive residual values). We only expect a few outliers with a higher NDVI value than the fitted curve because most effects lower the NDVI value (e.g., clouds, snow, cirrus clouds, shadow etc.). The remaining observations are the final observations used in the fitting procedure. When the last iteration is performed, the start of season (SOS) and end of season (EOS) are extracted from the fitted curve with the midpoint procedure. The midpoint ( $Mp$ ) is defined as the average of the maximum and minimum NDVI in the phenology curve so that

$$Mp = 0.5 * (NDVI_{max} - NDVI_{min}) + NDVI_{min} \quad (4)$$

The amplitude is defined as  $NDVI_{max} - NDVI_{min}$  and the growing season length (GSL) as  $EoS - SoS$  in days. The code uses the longest group of consecutive days above the midpoint to calculate the dedicated days of start and end of season. In addition, the RMSE of the residuals is extracted as

$$\sqrt{\frac{\sum_{i=1}^n Residuals^2}{n - 1}} \quad (5)$$

with  $n$  the number of valid observations.

The different types of phenological description are visualized in Figure 4.

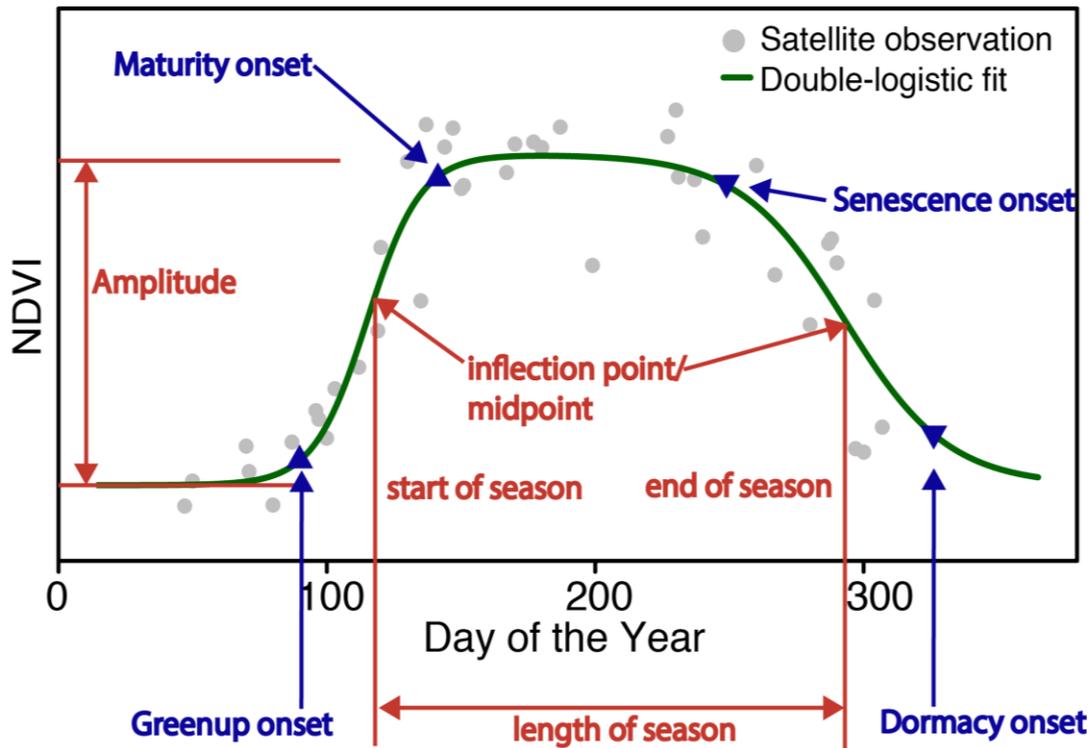


Figure 4: Definitions of different parameter used for phenology description.

#### 4.2.2. Estimation of uncertainty

The double-logistic curve fitting algorithm has 6 unknowns and the measurement noise of the L2-data is high and therefore the curve is under-determined. Tests have shown that the fitted parameters are highly correlated and a traditional uncertainty analysis based for instance on a covariance matrix is not useful. Therefore, RMSE per vegetation state (Dormancy, Green-up, Peak and Senescence) is chosen as the measure of quality. The dormancy states before spring and after autumn are taken as one state based on our assumption that the dormancy NDVI does not change between two consecutive seasons.

For the dormancy and peak states, the RMSE is based on residual NDVI, whereas for the green-up and senescence phases, the RMSE is based on days. The phase transitions are estimated with the turning points (third derivative equals zero) as shown in Figure 5. Every RMSE per vegetation state comes together with the available number of observations for this time period. For the Green-up and Senescence phases, it is possible that some points fall in ill-defined regions where the RMSE cannot be calculated. For example, an observed NDVI during the Greenup phase may be higher than the maximum NDVI of the fitted curve and therefore the RMSE cannot be found for the observation as it lies beyond the limits of the curve in the vertical direction (y-axis). In this case, the NDVI value is neglected when calculating the RMSE. As a result, RMSE values in days may be underestimated.

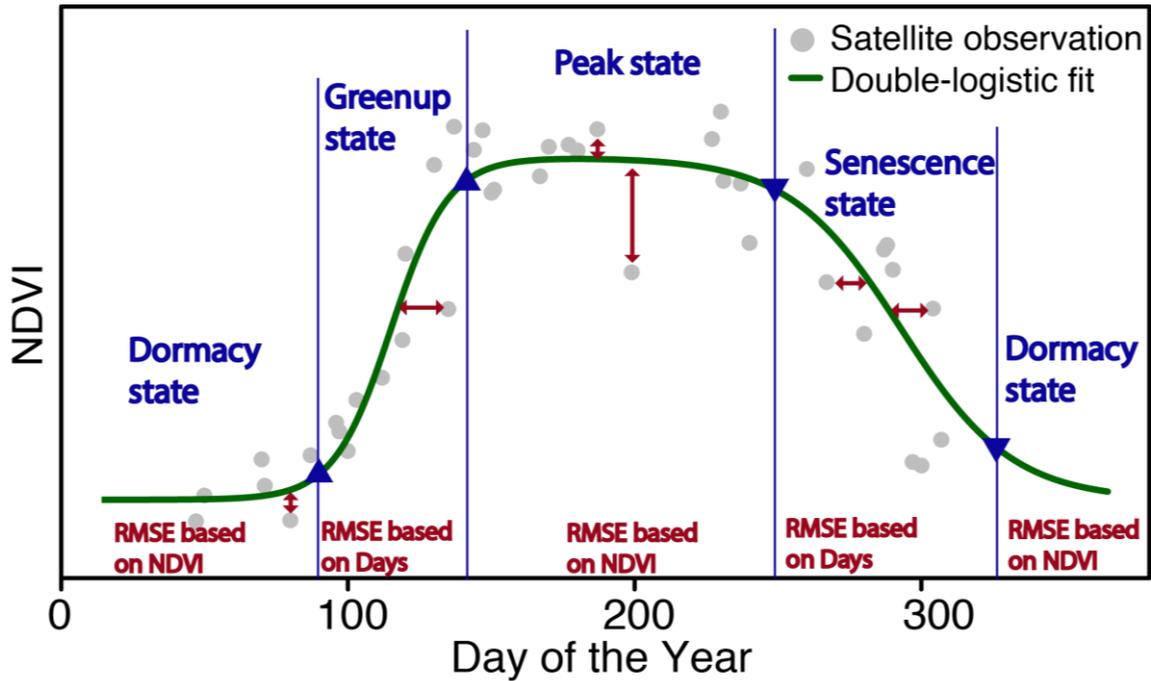


Figure 5: Double-logistic curve fitted onto NDVI-pixel values. Explanation for measure of quality. The RMSE during green-up and senescence state is based on the residuals in days of year between observation and modelled curve. RMSE for dormancy and peak state is based on residuals in NDVI between model and observations.

#### 4.2.3. Detection number of growing seasons

The detection of the number of growing seasons requires a flexible smoothing approach without a-priori constraints on growing-season number or shape. Harmonics have been found to suit this purpose well. The double-logistic function fit does not qualify because it requires a-priori knowledge on the number of growing seasons. Splines or local smoothers tend to be more sensitive than harmonics to data gaps which are expected due to sustained winter cloudiness.

The here used approach to derive the phenological metrics assumes a single growing season. The algorithm can a) fit the more dominant single growing season, b) interpolate over multiple growing seasons or c) flag the time series as invalid. The multi-season instances thus will not necessarily be flagged out and a detection of the number of growing seasons thus has been implemented as an uncertainty indicator for the phenological metrics. The users can therefore decide to include or exclude instances flagged with more than one growing season detected. Note that the algorithm is not suited for agricultural areas where commonly more than two vegetation seasons are expected.

The number of growing seasons is detected as follows:

- 1) Fitting of harmonic function based on the mean NDVI ( $NDVI_{mean}$ ) in addition with three components representing 1, 2 and 3 vegetation cycles, respectively, per year or frequency:

$$NDVI_{ds} = NDVI_{mean} + \sum_{i=FC_1}^{FC_{max}} A_i \cdot \cos(i \cdot x + \Phi_i) \quad (6)$$

In Equation ( 6 ),  $FC$  represents the Fourier component where  $FC_{max} = 3$ ,  $NDVI_{mean}$  is the mean NDVI,  $A_i$  is the amplitude at time  $i$ ,  $\Phi_i$  is the phase shift and  $x$  is the day number represented in radians.

- 2) Use the midpoint (NDVI) value detected before (see Chapter 4.2.1, equation (4)) to determine the threshold between dormancy and growing season.
- 3) Calculate number of clusters of observations above the midpoint level using run length encoding as visualized in Figure 6. The number of clusters represents the number of growing seasons with an expected maximum of 3.

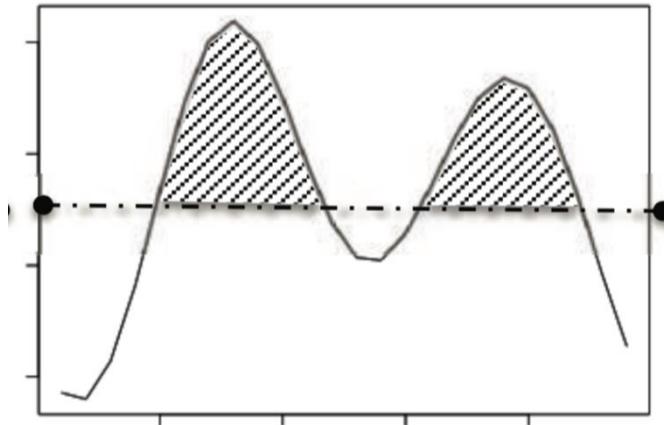


Figure 6: Schematic detection of double growing season based on the midpoint value. If several consecutive points lie lower than the midpoint value with afterwards consecutive points lying above, a multi-growing season time-series is detected. Picture from Garonna (2016)

#### 4.2.4. Flagging

After the calculation of the LSP metrics, several criteria need to be met in order to exclude unreasonable and wrong results also excluding fitting results without a strong enough vegetation cycle.

Several criteria define if a result is invalid:

1.  $< 7$  valid observations (see Chapter 4.2.1)
2. Mean of all observations is  $< 0.2 \rightarrow$  NDVI is too low for vegetation signal
3. Amplitude is  $< 0.1 \rightarrow$  amplitude is too low for a phenological signal
4. SOS equals maximum or EOS equals minimum NDVI  $\rightarrow$  no phenology
5. SOS or EOS equals to first or last day of observation  $\rightarrow$  no convergence
6. More than 34% of the points were flagged in the iterative fitting procedure  $\rightarrow$  no phenology
7. P-Value of F-test against the assumption “the fit is better than using the mean(NDVI)” is  $> 0.05 \rightarrow$  fit not better than using mean value

In the appendix, a list of the used flags in the binary system is shown. The binary system corresponds to the decimal number as listed in Table 3.

Table 3: Binary number corresponding to the flagging system.

Flag type	Corresponding Binary Number
1) < 7obs	1
2) Mean NDVI <0.2	2, 6, 10, 14, 26, 30, 34, 38, 42, 46, 58, 62, 66, 70, 74, 78, 90, 94, 98, 102, 106, 110, 122, 126
3) Ampl <0.1	4, 6, 12, 14, 28, 30, 36, 38, 44, 46, 60, 62, 68, 70, 76, 78, 92, 94, 100, 102, 108, 110, 124, 12
4) No phenological curve	8, 10, 12, 14, 24, 26, 28, 30, 40, 42, 44, 46, 56, 58, 60, 62, 72, 74, 76, 78, 88, 90, 92, 94, 104, 106, 108, 110, 120, 122, 124, 126
5) No dormancy	24, 26, 28, 30, 56, 58, 60, 62, 88, 90, 92, 94, 120, 122, 124, 126
6) <66% valid observations after fit	32, 34, 36, 38, 40, 42, 44, 46, 56, 58, 60, 62, 96, 98, 100, 102, 104, 106, 108, 110, 120, 122, 124, 126
7) F-test >0.05	64, 66, 68, 70, 72, 74, 76, 78, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 120, 122, 124, 126

#### 4.2.5. Output

After the flagging procedure, the map with all valid pixels can be drawn. The output is stored as GeoTiff with the output variables as individual layers (SOS, EOS, GSL, amplitude, and quality measures of RMSE (one per state and one general), nValid observations, P-Value, number of growing seasons). SOS and EOS are presented in day of year (DOY) and GSL in days. All other variables are unitless. In total 23 layers are available and listed in Table 4. The Tiff files are georeferenced in the Universal Transverse Mercator (UTM) coordinate system, with the x/y coordinates starting from the upper left hand corner.

Table 4: Layers stored in the output GeoTiff file and their definitions.

Layer #	Output layer	Abbreviation	Reference Chapter
1	x-coordinate in UTM (East)	x	-
2	y-coordinate in UTM (North)	y	-
3	Pixelindex	Ind	-
4	Number of total observations (available images)	nobs	3.4
5	Number of valid observations (after L2-flagging)	nobsvalid	3.1 - 3.3
6	Number of final observations used in curve fit	nobsfinal	4.2.1

7	Start of Season (DOY)	SOS	4.2.1
8	End of Season (DOY)	EOS	4.2.1
9	Growing Season Length (days)	GSL	4.2.1
10	P-value	P-Value	4.2.4
11	Phenology quality flag	phenoflag	4.2.4
12	Curve fit RMSE	dlogrmse	4.2.1
13	Number of iterations	niter	4.2.1
14	Curve amplitude (dimensionless NDVI)	dlogampl	4.2.2
15	Growing season count	gscount	4.2.2
16	Dormancy phase RMSE	DormRMSE	4.2.2
17	Number of dormancy phase observations	DormNobs	4.2.2
18	Peak phase RMSE	PeakRMSE	4.2.2
19	Number of peak phase observations	PeakNobs	4.2.2
20	Green-up phase RMSE	GreenuRMSE	4.2.2
21	Number of green-up phase observations	GreenuNobs	4.2.2
22	Senescence phase RMSE	ScenRMSE	4.2.2
23	Number of senescence phase observations	ScenNobs	4.2.2

## 5. Product

### 5.1. Product description

The output product is a GeoTiff file with 23 layers corresponding to the layers described in Table 4 in Chapter 4.2.5, including the coordinates of the pixel, SOS, EOS, GSL, amplitude, RMSE quality measures, number of final observations used in the fitting, and P-Value. At each layer, a map can be generated. The individual pixel values can at a later stage be used for comparison with observation of different years. Figure 7 and Figure 8 show an example output from the algorithm.

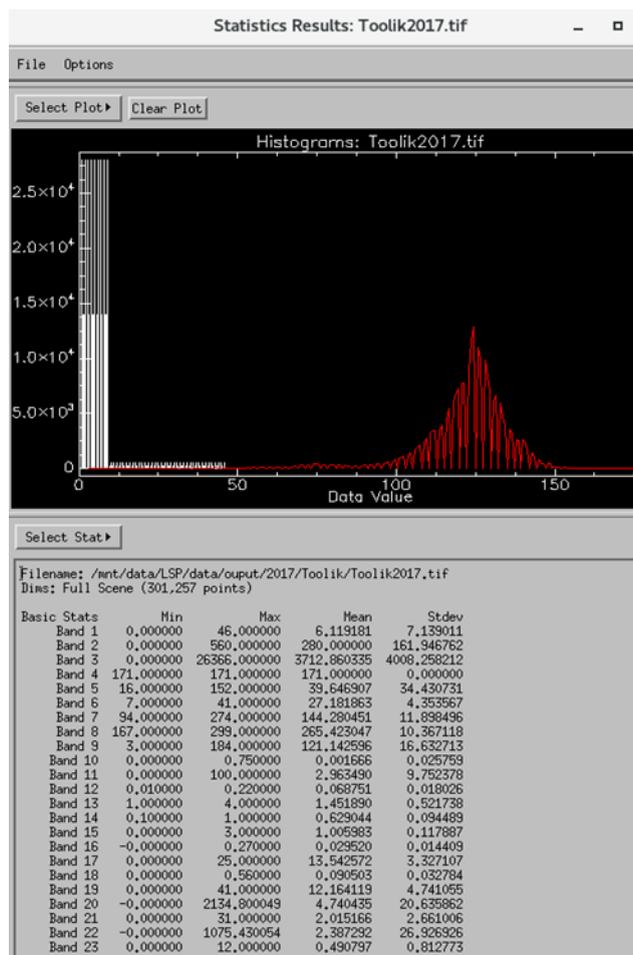


Figure 7: The print screen shows the statistics for the output file with 23 layers (or bands). As here, different layers that can be illustrated in a histogram. The histogram for Band 9, the growing season length, is displayed with a mean of 121 days and a range from 3 to 184. The band list is the same as listed in Table 4.

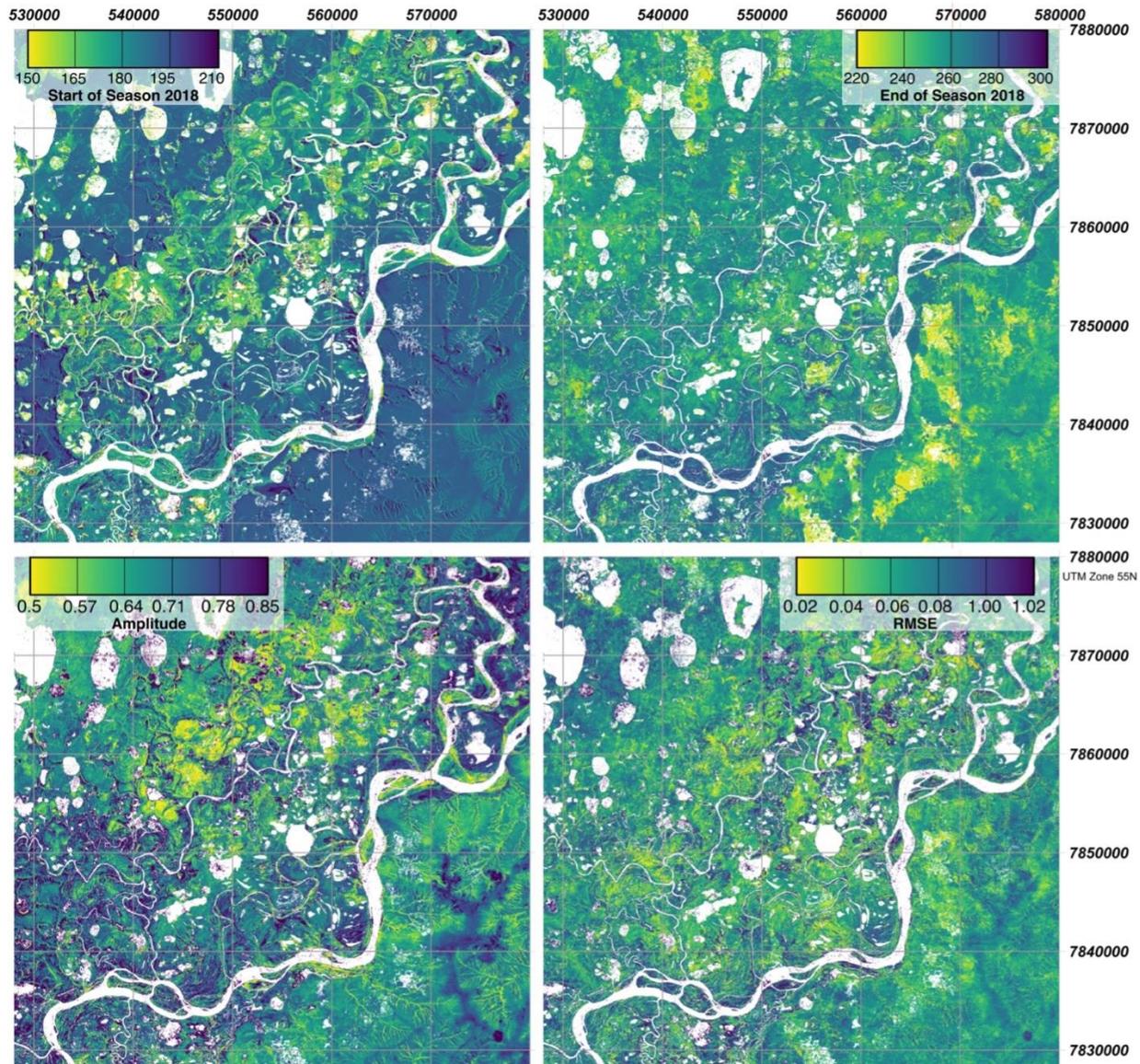


Figure 8: The overview image shows the results of SOS, EOS, Amplitude and RMSE for a part of the Kytalyk NP. The colorscale is adjusted depending on the histogram and existing values. In the image the structures of topography and vegetation are well visible. White pixels correspond to invalid pixels, mostly related to water.

## 5.2. Accuracy analysis

We provide the RMSE of the full fitted curve as measure for the quality of the fit. In addition, the P-Value also gives an additional measure for the reliability of the result. For the Green-up, Peak, Senescence and Dormancy phases, we also provide an RMSE value (in NDVI for Peak and Dormancy phases, and in Days for Green-up and Senescence) and the number of observations in each phase.

### 5.3. Validation Method

For validating the satellite-derived LSP results, we adopt and adapt the methodology proposed by Zhang et al. (2018) and authors therein using phenocam time series observations. Hereby, the time series of canopy images is used to calculate the greenness over the season. To do so, we selected an area of interest defined by the operator, as in the methodology suggested by (Filippa et al. 2016), and calculated using RGB images for a given time of the day, using the GCC computed from the red, green and blue channels of the image (see Equation ( 2 ). The area of interest is chosen to be representative of the main vegetation or land cover type; disturbing signals, including non-vegetation areas, such as bare ground and sky and saturation effects of the camera (often seen in large distances) were be avoided to ensure a homogenous time series for the computation of LSP metrics. Often, phenocam cameras take up to one picture per hour; therefore, in order to reduce the data volume and to be consistent with the overflight time of the satellite, one image per day was select, in this case, we have selected images taken at 11 am local time from the phenocam time series, to be consistent with the acquisition of Sentinel-2 and Landsat 8 scenes.

LSP metrics are computed for the spatially averaged phenocam time series using the same algorithm as that used to process the satellite GCC and NDVI image time series. Here, the RGB phenocam images were converted to the GCC time series that then was averaged over a polygon or region of interest, representative of a homogenous vegetation or land cover type, defined by the operator, as in the methodology suggested by Filippa et al. (2016). Subsequently, LSP metrics were computed for this averaged time series and area of interest. Polygons or regions of interest were selected in such a way as to reduce the influence of non-vegetation areas, such as bare ground and sky, minimize a possible saturation of the sensor for objects at large distances, and select on the main vegetation or land cover type and hence ensure a homogenous time series for the computation of LSP metrics.

Subsequently, the satellite NDVI and GCC time-series, were clipped to the field of view (FOV) of the phenocams, the resulting time-series were averaged to a single value and LSP metrics computed. To address the effect of outliers affecting the averaging process, a threshold of a minimum number of valid observations for a pixel before averaging can be used. Many phenocams follow a unified phenocam installation protocol such as that used by the EuroPhen<sup>2</sup> or the North America's PhenoCam<sup>3</sup> network. They often use a StarDot<sup>4</sup> NetCam camera pointing North with a 40° opening angle. We additionally assume a maximum viewing distance of 300m corresponding to the field of view. The maximum distance, however, may change depending on the observation site.

Note that quality control measures, such as minimal amplitude or P-Value, need adaption for the GCC time series as the absolute GCC values differ from NDVI values. Some phenocams also offer a Near Infra-Red (NIR) channel for calculating the NDVI, nevertheless, we intentionally exclude this channel as not all cameras have it available, in addition, few have calibrated the NIR with the red channel and the images are often not captured simultaneously.

---

<sup>2</sup> Europhen.org

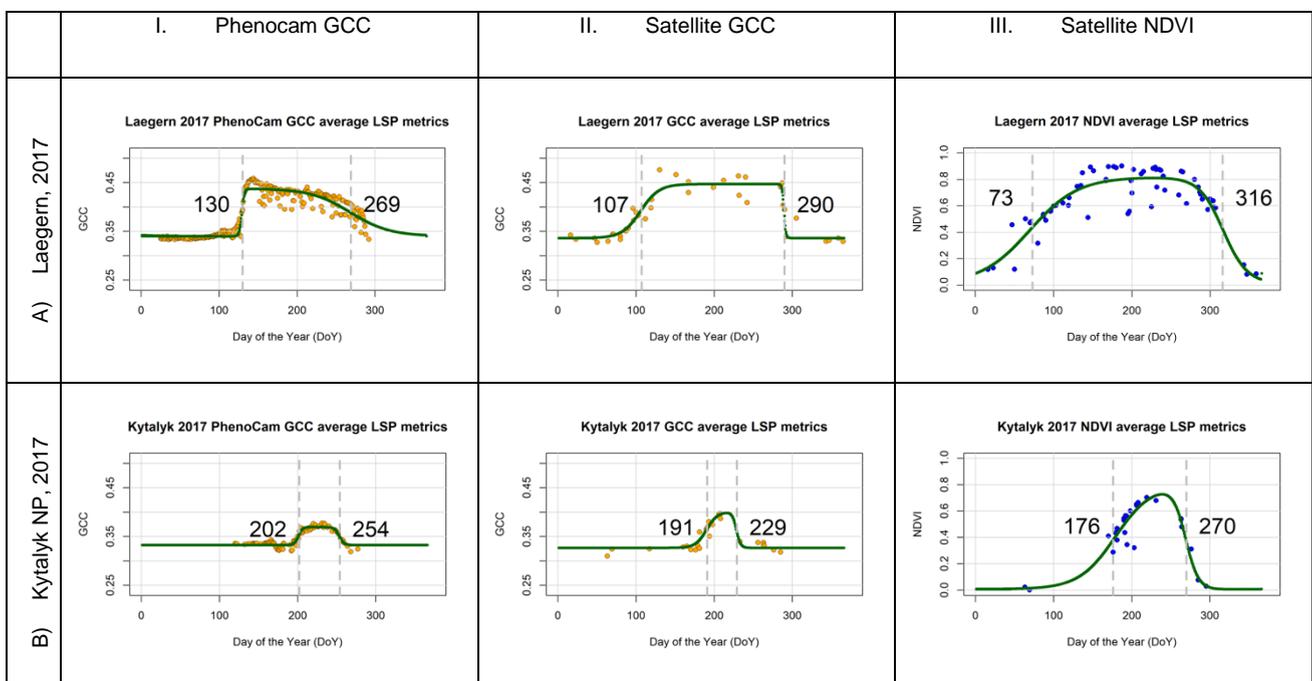
<sup>3</sup> <https://phenocam.sr.unh.edu/webcam/>

<sup>4</sup> [www.stardot-tech.com](http://www.stardot-tech.com)

In general, GCC is expected to saturate earlier than the NDVI values, resulting in an earlier SOS and later EOS, respectively (e.g., Keenan et al. 2014; Browning et al. 2017; Filippa et al. 2018). This occurs since GCC is a measure of the “greenness” – the green light – and the NDVI is connected directly to the photosynthetic activity via the NIR channel. Photosynthetic activity can change when the vegetation stays “green”. In general, it has been shown that LSP metrics derived from GCC tends to have later SOS and earlier EOS because photosynthesis is assumed to start earlier and ends later than the change of greenness indicates (Keenan et al. 2014; Walther et al. 2016) resulting in general in a longer GSL. We therefore compare for the validation the GCC of the two independent data sets (satellite images and ground-based phenocam) with each other and in addition with the NDVI derived phenology from satellite images.

#### 5.4. Validation results and discussion

We present here the results of the validation for four different phenocams available within the GlobDiversity project. The phenocams are located in two different biomes, the Arctic tundra (Toolik Lake (Alaska, USA) and Kytalyk National Park (NP) (Siberia, RUS)) and temperate forest (Laegern forest (CH), Bavarian National Park (GER)). To validate the NDVI-based LSP metrics, we compare the phenocam GCC time series and its derived LSP metrics, phenocam with the satellite GCC and NDVI time series and derived LSP metrics. These time series are plotted in Figure 9 together with two derived metrics (SOS and EOS). Shown are the results for Laegern and Kytalyk NP for 2017, and the results for Toolik Lake and Bavaria NP for 2018. In general, it is well visible that the phenocam time series (column 1, Figure 9) contains a dense time series, while in particular the satellite-derived time series for 2017 (columns 2&3, Figure 9) are based on fewer observation as Sentinel-2B was launched end of 2017 only. In general for satellite images, green-up phases and senescence phases are often dominated by unstable and cloudy weather conditions resulting in a poorly defined green-up and senescence phases in the NDVI and GCC time series.



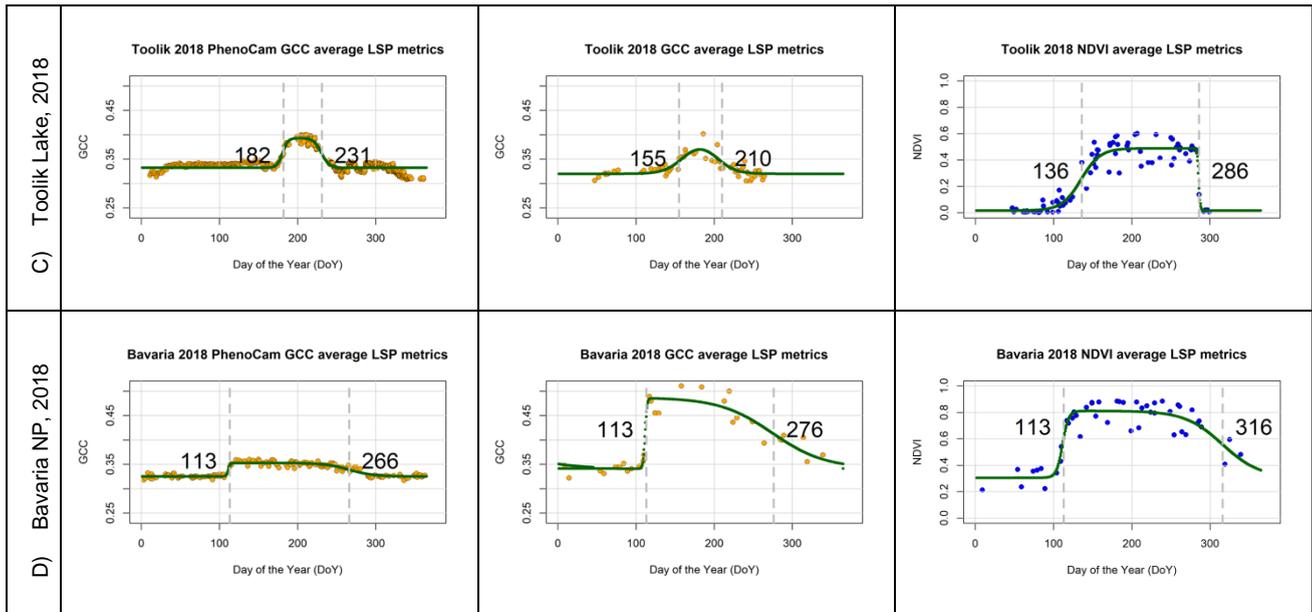


Figure 9: LSP metrics for the four sites (Laegern, Kytalyk NP, Toolik Lake and Bavaria NP), computed using the phenocam GCC (column 1), satellite GCC (column 2) and satellite NDVI time series (column 3).

The values for each metric (SOS, EOS, and GSL), computed with each time series (phenocam GCC, satellite GCC and satellite NDVI), and the difference between phenocam GCC and satellite GCC metrics as well as phenocam GCC metrics with satellite NDVI metrics for sites in 2017 and 2018 are include in Table 5. We find that the differences between phenocam GCC and satellite GCC metrics are smaller than between phenocam GCC and satellite NDVI metrics, thereby confirming that satellite GCC metrics are closer to those derived from phenocam GCC. Nevertheless, the effect is more visible for higher latitudes (Toolik Lake and Kytalyk NP, column B&C, Figure 9) than for temperate forest (Bavarian NP and Laegern, column A&D, Figure 9) as also described by Walther et al. (2016). Nevertheless, the phenological curve from the Bavarian NP forest camera (column D, Figure 9) stands out with a very low amplitude in comparison with the phenocam time series of other pilot sites. This can be explained with the fact that this camera is a Bushnell animal camera trap, which is different than those used for the three other sites (Netcam SC), and therefore the sensors' light sensitivity is different and results in a low amplitude of the GCC time series phenological curve.

In addition, we find that for the temperate forest sites of Laegern and Bavaria, the results for GSL derived from the 2017 phenocam GCC, satellite GCC and NDVI time series, are within the expected range cited in literature (Figure 4). For example, Schieber et al., (2013) find that, on average, the GSL of beech species ranges from 128 to 181 days along a 200-1400 m a.s.l altitudinal gradient.

Table 5 Values for each metric (SOS, EOS, and GSL), computed with each time series (phenocam GCC, satellite GCC and satellite NDVI), and the difference between phenocam GCC and satellite GCC, as well as phenocam GCC and satellite NDVI, , for sites in 2017 and 2018.

LSP	phenocam	Satellite	phenocam GCC –	phenocam GCC –
-----	----------	-----------	----------------	----------------

		Metric	GCC	GCC	NDVI	satellite GCC	satellite NDVI
2017	Laegern	SOS	130	107	73	23	57
		EOS	269	290	316	-21	-47
		GSL	139	183	243	-44	-104
2017	Kytalyk NP	SOS	202	191	176	11	26
		EOS	254	229	270	25	-16
		GSL	52	38	94	14	-42
2017	Toolik Lake	SOS	140	123	125	17	15
		EOS	217	199	262	18	-45
		GSL	77	76	137	1	-60
2017	Bavaria NP	SOS	136	136	131	0	5
		EOS	264	262	313	2	-49
		GSL	128	126	182	2	-54
2018	Toolik Lake	SOS	182	155	136	27	46
		EOS	231	210	286	21	-55
		GSL	49	55	150	-6	-101
2018	Bavaria NP	SOS	113	113	113	0	0
		EOS	266	276	316	-10	-50
		GSL	153	163	203	-10	-50

In Figure 10 the LSP metric (SOS, EOS, and GSL) values are plotted for each site and each year evaluated. The bar graphs serve to highlight the differences and similarities between metrics. For example, for Bavaria 2017 and 2018, SOS is remarkably consistent between the different time series, supporting the precision of the product (i.e. the closeness of the measurements to each other). In contrast, EOS and GSL for Bavaria 2017 and 2018 are more variable; this is to be expected, as senescence in temperate forests tends to occur disparately (Keenan et al. 2014). Furthermore, GCC and NDVI tend to measure different characteristics leaf reflectance. In effect, our results further agree with published literature in that the response of NDVI-derived metrics was distinct from GCC-derived metrics, with the former (NDVI), consistently resulting longer growing seasons (GSL), as can be clearly distinguished in Figure 10.

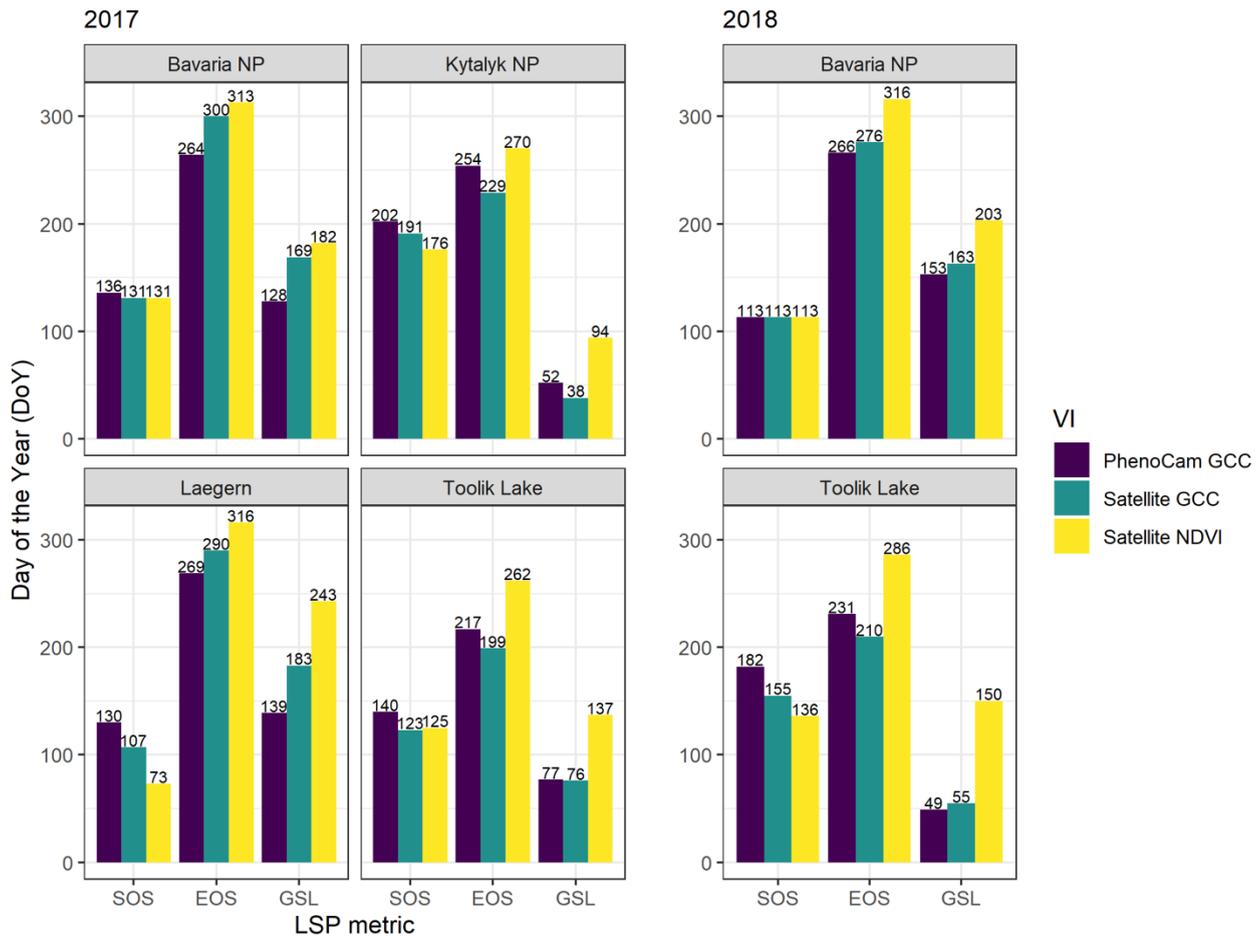


Figure 10: Bar plots of LSP metrics SOS, EOS, GSL, computed using satellite NDVI, Satellite GCC, and phenocam GCC time series, for each of the four sites (Laegern, Kytalyk, Toolik Lake and Bavaria NP), for the years 2017 (all sites) and 2018 (Toolik Lake and Bavaria NP).

In conclusion, we are able to validate our results with the ground-based phenocam observations, nevertheless, the comparison of the time series highlight the restrictions imposed by a limited number of observation available when using satellite GCC and NDVI time series, which are often due to cloud cover. The results also highlight that the GCC time series retrieved from satellite and phenocam are comparable and can be used for validation; however, satellite-based NDVI results diverge when compared to GCC time series. Several other factors are in addition responsible for the differences observed in phenocam and satellite LSP metrics; specifically, phenocam GCC is derived from an RGB time series captured at an oblique angle across a vegetation canopy, and is further extracted and averaged across a region of interest, defined by the operator; in contrast, the satellite NDVI and GCC time series are based on Landsat 8 and Sentinel-2 surface reflectance measurements (in the green, red, blue and NIR bands), acquired from near-nadir view angles. Thus, it is expected that the LSP metrics derived from these distinct time series show some variation. Furthermore, and as identified by Keenan et al. (2014), changes in surface spectral properties, especially in spring, are primarily controlled by changes in leaf area; thus, the LSP metric SOS derived from phenocam GCC, satellite GCC and satellite NDVI time series, often reveal a strong agreement, especially for sites with a homogenous land cover or vegetation type and temperate deciduous character. This pattern was clearly identified for the Laegern and

Bavaria NP temperate forest sites, where SOS values were remarkably consistent (Figure 10). In contrast, autumn shifts in surface reflectance properties are mainly caused by decreases in leaf chlorophyll and the subsequent appearance of pigments, followed by leaf senescence. This process tends to be gradual and fluctuating across a given landscape, and as such it is reflected in the more variable EOS values we observe; furthermore, it is consistent the current literature (Bolton et al. 2020). Similarly, as a result of the oblique viewing angle of the phenocam, these sensors often capture the growth of sparse understory vegetation, such as herb and grass species emerging before forest canopy leaves during spring green-up; this process is especially prevalent in temperate forests. In contrast, satellites have near-nadir viewing angles which capture mainly canopy-level leaf emergence (understory vegetation emergence is mostly masked by the tree canopy), and therefore contribute towards the different dates of LSP metrics.

Finally, our validation method does not consider a possible inhomogeneity in the vegetation types within the phenocam, FOV and corresponding area covered by the satellite. Estimation of the camera's FOV projected on the satellite image is also an approximation only and strongly depends on the observation site topography and vegetation. The former effect is minimized by the operator by selecting carefully the area of interest in the phenocam image, while the latter is address by carefully selecting the corresponding satellite FOV.

In light of these issues, the agreement between both the phenocam and satellite derived time series is acceptable and demonstrates the overall quality of the different time series used and the final derived LSP product.

## 6. Remarks on prototyped algorithm development and processing

This chapter shall highlight some experiences made when prototyping the algorithm and running the algorithm on the different test sites. These points might need to be addressed when using the algorithm and when further developing the procedures.

### 6.1. Fitting procedure

The prototype contains two functions for the fitting procedure implemented in the code, however, significantly different in their performance. Originally, the development team had chosen the differential evolution based DEoptim fitting function in R that had a significantly more stable performance in the fitting procedure, in particular with the remaining cloudy observations, than any other used procedure and in addition does not need a starting value. In general, the time series is ill-posed and therefore a stable algorithm is crucial.

The implementation in the prototype in python, however, showed significant draw-backs in runtime. Next to DEoptim, the 'scipy optimize curve fit' function is also implemented in python. The scipy curve fit function is much faster than the DEoptim function and was therefore chosen for the processing of the pilot sites. The scipy curve fit function uses a Trust Region Reflective algorithm for the optimization. Bounds are given for each variable in the double logistic curve fit. If the fit is unsuccessful, the neighboring pixel implementation (described above) can be used, where the fitted values from successful fits are used as a starting point (see Chapter 6.4). In most cases, the fitting converges on an answer but in some cases, the fitting is unsuccessful and gives a Runtime Error.

The main drawback of the second fitting function is the need of a good starting value to get the absolute maximum. This fact, however, prohibits an automated implementation for larger areas or even a global implementation. It is rather recommended to optimize the DEoptim implementation for future projects.

### 6.2. Cloud and invalid pixel detection algorithm and high latitude areas of interest

As mentioned above, the standard pixel classification from ESA and NASA was used, respectively. Nevertheless, the algorithm is not yet always and everywhere reliable. Therefore, one needs to expect remaining outliers and that is why the LSP algorithm uses an iterative procedure. After some testing, Landsat 8 is used in a more conservative way and more observations are kept (e.g. connected to indicated probability of cloud cover). This might lead to a bias in the time series, however, as many more observations from Sentinel-2 are available, the effect might be negligible.

Regarding Sentinel-2, a high number of "unclassified" pixels have been observed in particular in the Kytalyk area. These pixels are generally kept within the procedure, however, are often connected to cloud or snow cover and resulted to more outliers in the curve. It is hoped that this influence will be reduced with further improvement of the Sen2cor algorithm leading to a more reliable phenological curve. It is also worth mentioning the handling of the validity of snow-cover pixels: in areas such as Laegern forest, snowy pixels certainly contaminate the time series and should be excluded. However, in high-latitude areas, snow is the dormancy state and green up follows directly the first snow melt. Therefore in these areas, the snowy pixels should be kept in

the time series. The algorithm presented here excludes the as snow classified pixels. Nevertheless, the influence is marginal due to the above described effect of “unclassified” pixels in high latitudes that are often snow-covered pixels.

If a systematic effect is detected in the resulting LSP products one could check individual time series and fitted curves. For instance, a few not-flagged observations can bias a large part of an image due to too few observations being available, in particular during green-up and senescence phase. This could need manual intervention to flag these wrong observations. However, such effects can be reduced by using all available observations (Sentinel-2A, -2B and Landsat 8) as well as a high-quality cloud detection algorithm.

### **6.3. Memory requirements**

Run time and memory are the biggest obstacles for the full phenology processing. For small areas of interest though, neither aspect is an insurmountable problem, and the pixel-wise calculations can be run in a few hours or less. However, in the current implementation large areas need considerably more time to run and more memory to load the NDVI time series datacube. The first problem, run time, can be addressed by splitting the datacube into smaller subsets to be run in parallel. Even then, large amounts of memory may still be required to load the datacube. A tradeoff between run time and memory is a key consideration for optimizing the algorithm.

### **6.4. Constraints from neighboring pixels**

The time series for each pixel is handled independently. It is possible that neighboring pixels have similar fit values and the output of one pixel could be used for a faster convergence of the fitting process. The algorithm can implement a nearest neighbor fitting procedure where the results for neighboring pixels (one above and one before in the datacube) can be used as a starting point for the curve fitting of the current pixel. Nevertheless, when using the nearest neighbor fitting, a systematic “striping” for certain areas, in particular with few available data points, was detected and therefore this option was not used. This option again requires storing data and memory allocation could constrain performance.

### **6.5. Availability of L2-images**

All Sentinel-2 and Landsat 8 observations need to be available in Level 2 (bottom of atmosphere). Note, that both, Nasa and ESA, give a limitation of the validity of their L2-algorithm to a certain latitude. Kytalyk certainly lies above this limit, however, due to the very fast and steep green-up, no limitation due to atmospheric effects were detected.

### **6.6. Number of available observations**

Time series can contain more than 200 observations per pixel, however, due to cloud cover, this data can be reduced by around 50% on average (e.g., for temperate Europe). Observation areas in high latitudes (e.g. Arctic regions), however, can easily have double the observation due to more overflights but also have higher cloud cover.

### **6.7. Time stamp**

The time stamp for each observation is converted to a float value of days since 0AD, or 0CE. After the fitting procedure, the final values for Start and End of Season have a magnitude of 736587

and 736495 in 2017, for example. The fitting is not performed by day of year but by days since OAD. After the fitting, the Start and End of Season are converted to DOY values. An advantage of processing in days since OAD is that phenological seasons can span December to January without causing any problems to the fitting since the time domain has no discontinuities. In addition, values for the estimated Start and End of Season can be given as initial parameters to help with the convergence. Nevertheless, some fitting procedures can have issues in convergence when dealing with large numbers in the strongly underestimated double-logistic curve. A temporarily reduction of the numbers (i.e., all time-stamps minus time-stamp of first observation) might be necessary for a faster and more stable convergence.

Time stamp needs to include the time of overflight (not only DOY) in order to allow more than one observation per day from different satellites. It can be assumed that two different satellite do not have the identical overflight time (in seconds).

### **6.8. Reduction to area of interest, naturally vegetated areas**

The calculation region can be reduced so that only pixels within a region of interest are processed. For instance, crop fields or covered areas can be excluded, if reliable land cover data exists. Open street map could be used to exclude streets and houses (= urban areas), or any other land cover map. Excluding non-vegetated areas reduces the number of to be processed pixels.

## 7. Upscaling results

### 7.1. Introduction

LSP metrics were retrieved for two pre-defined areas of interests, i.e. Finland and Senegal, for the year 2019 based on NDVI time series.

Only vegetated areas were considered in this analysis. For Finland, the 2018 20 m Corine Land Cover product was used to distinguish vegetated from non-vegetated areas (Figure 11).

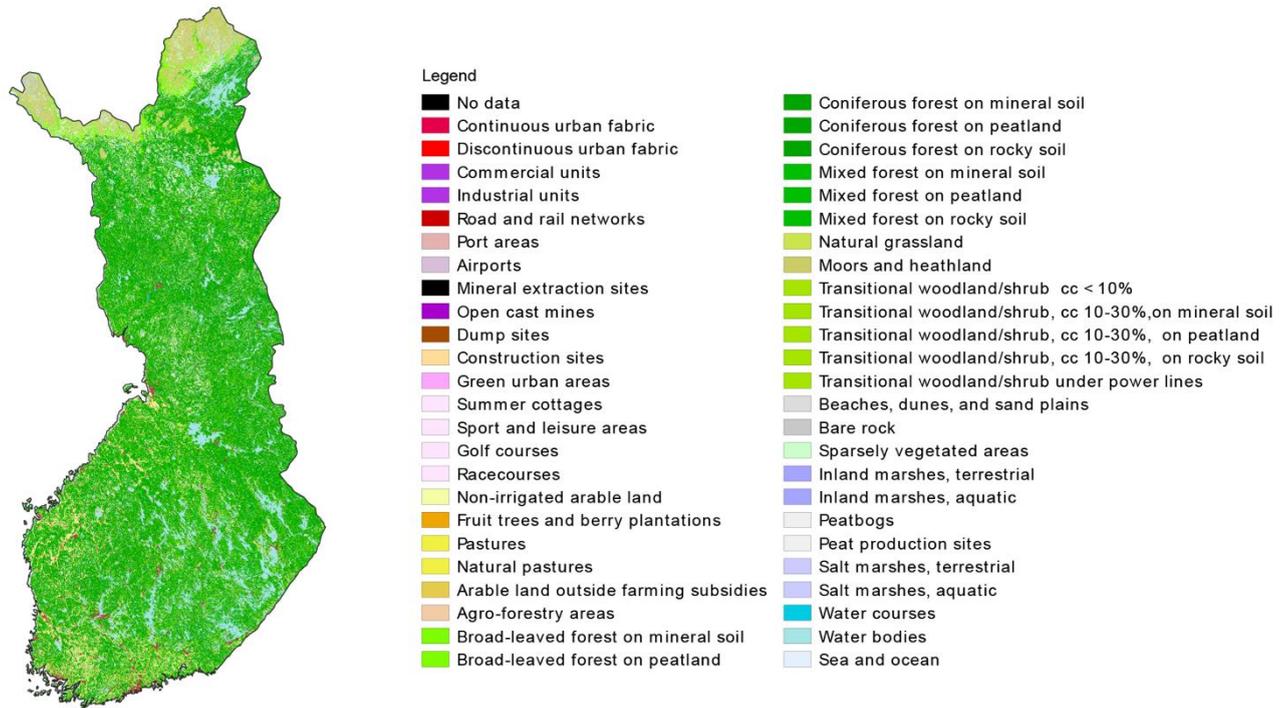


Figure 11: 2018 Corine land cover product for Finland at 20 m resolution available from [https://www.syke.fi/en-US/Open information/Spatial\\_datasets/Downloadable\\_spatial\\_dataset#C](https://www.syke.fi/en-US/Open_information/Spatial_datasets/Downloadable_spatial_dataset#C)

For Senegal, the 2016 20 m ESA CCI land cover product for Africa was used (see Figure 12).

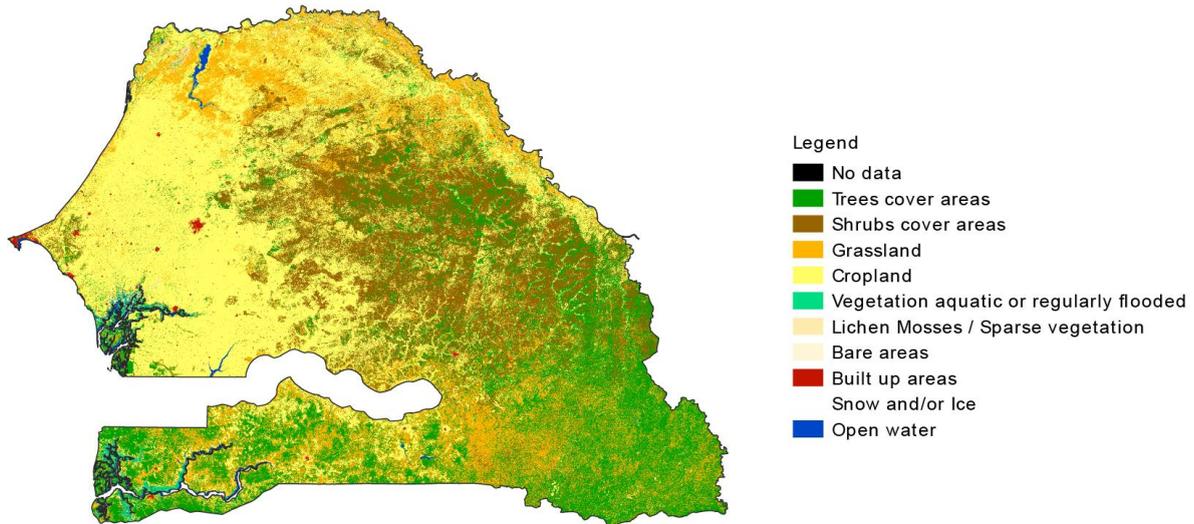


Figure 12: CCI 2016 20 m land cover prototype over Senegal available from <http://2016africalandcover20m.esrin.esa.int/>

As a preprocessing step, both land cover products were first resampled to the Sentinel-2 UTM tiling grid at 10 m resolution using nearest neighbor resampling and reclassified into vegetated and non-vegetated areas.

Next, NDVI time series were calculated at 10 m resolution for all S2 products available over the 2 countries. Herein non-vegetated areas detected by the SCL layer were set to No Data (see section 3.1)

Subsequently based on those NDVI time series, LSP metrics were calculated using the prototype LSP algorithm. In order to run the LSP algorithm at country scale however, some changes to the implementation of the prototype algorithm were needed in order to prevent out of memory errors. In contrast to the pilot sites which consist of a single S2 tile (or only a couple of tiles) it was no longer possible to use a single vrt containing all the S2 input products. Instead, LSP was calculated by S2 tile, which were further split into small blocks of 256 x 256 pixels. In this effort, only blocks with vegetated pixels were processed. The overlap in the S2 grid between neighboring S2 tiles and UTM zones results in overlapping LSP metrics. Post-processing of the results into a common projection system (e.g. latlon or webmercator) is therefore advised but was not performed in this experiment as this depends on the user requirements.

## 7.2. Results

Table 6, summarizes the total number of LSP products and processing parameters, derived from single Sentinel-2 images for both Senegal and Finland.

Table 6: Summary of the LSP upscaling processing results

Parameter	Senegal	Finland
Number of processed S2 tiles	40	65
Number of processed NDVI images	4394	15546

Number of processed LSP blocks	37404	79772
--------------------------------	-------	-------

Based on the NDVI time series LSP metrics were calculated and visualized showing the spatio-temporal variation in LSP. Figure 13 and Figure 14 demonstrate the Start of Season (SOS), End of Season (EOS) dates, Growing Season Length (GSL) and Amplitude (AMP) for whole Finland and Senegal respectively. Figure 15 finally, illustrates these layers zoomed in on a particular region in Finland (village of Karkkila).

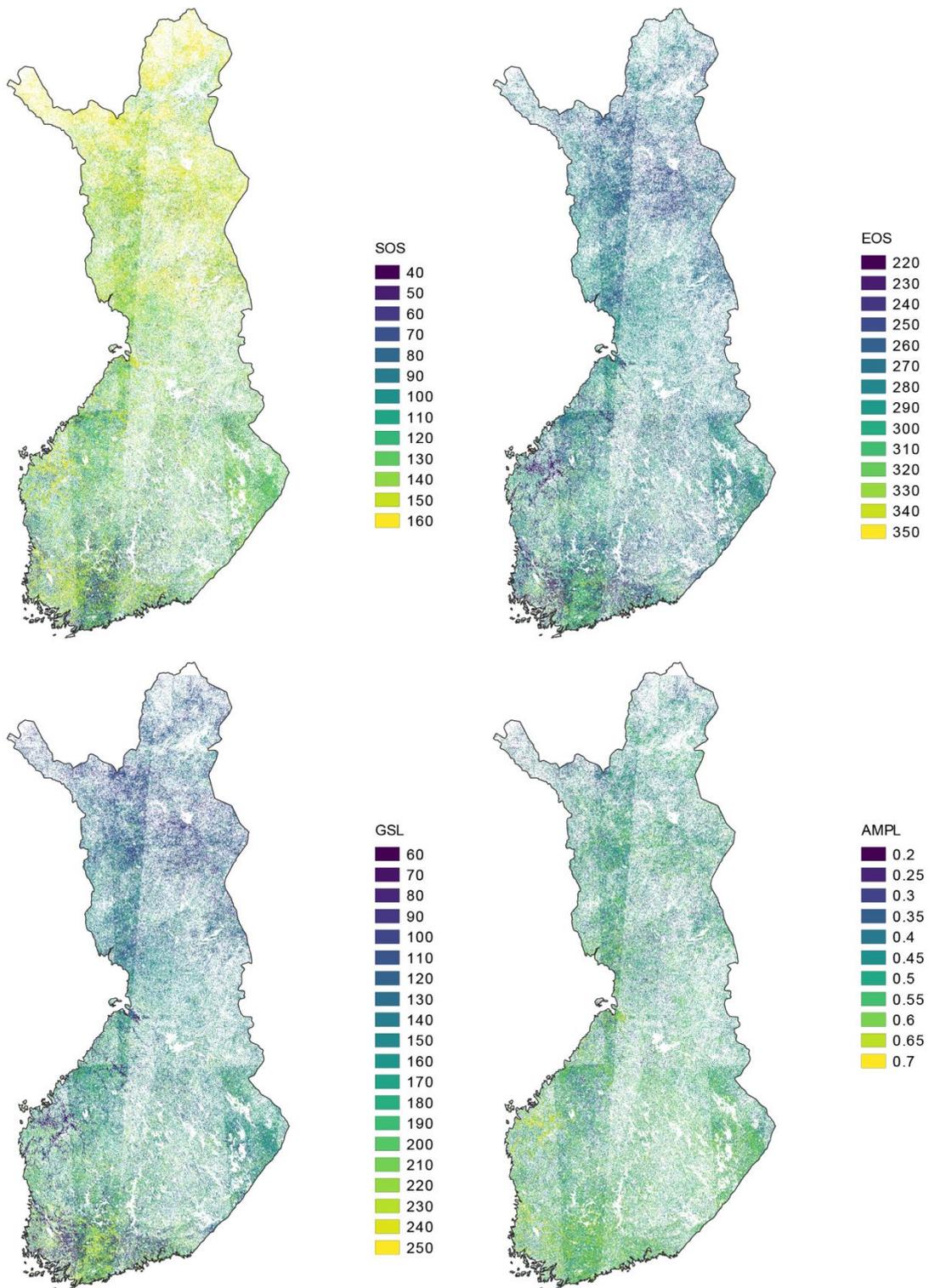


Figure 13: Per pixel LSP metrics over Finland: Start of Season (SOS, DoY), End of Season (EOS, DoY) dates, Growing Season Length (GSL, DoY) and Amplitude (AMP). The visible diagonal is due to the overlap in neighboring UTM zones.

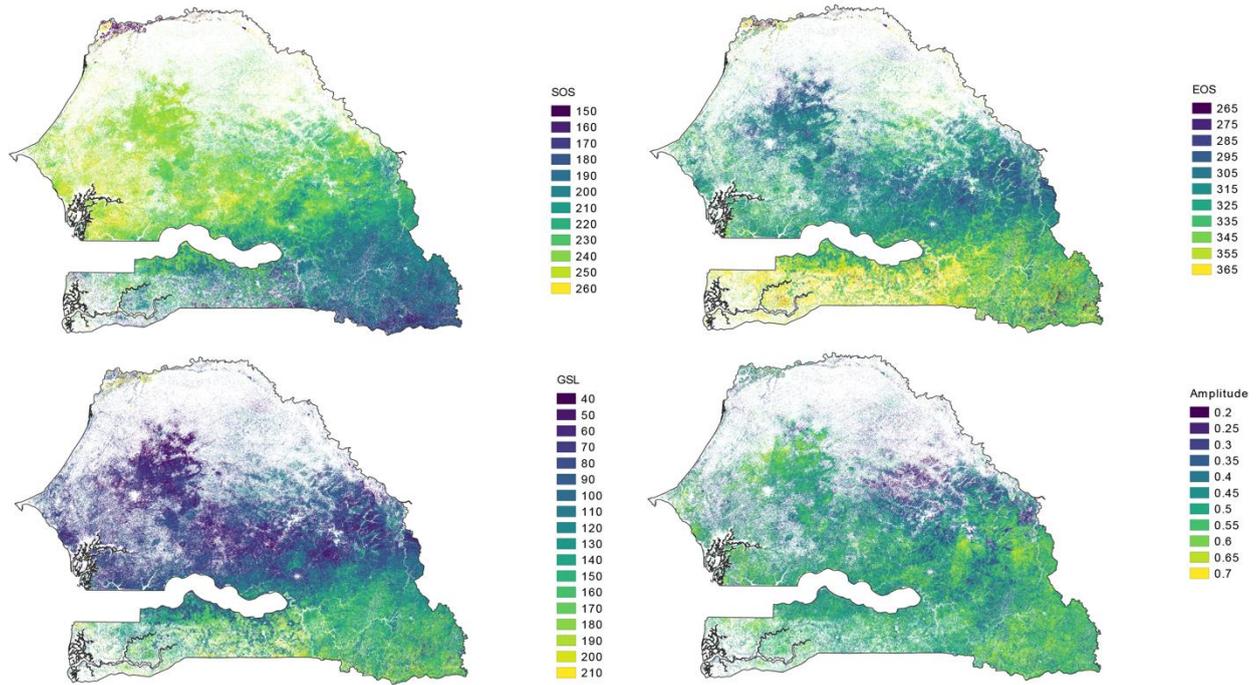


Figure 14: Per pixel LSP metrics over Senegal: Start of Season (SOS, DoY), End of Season (EOS, DoY) dates, Growing Season Length (GSL, DoY) and Amplitude (AMP)

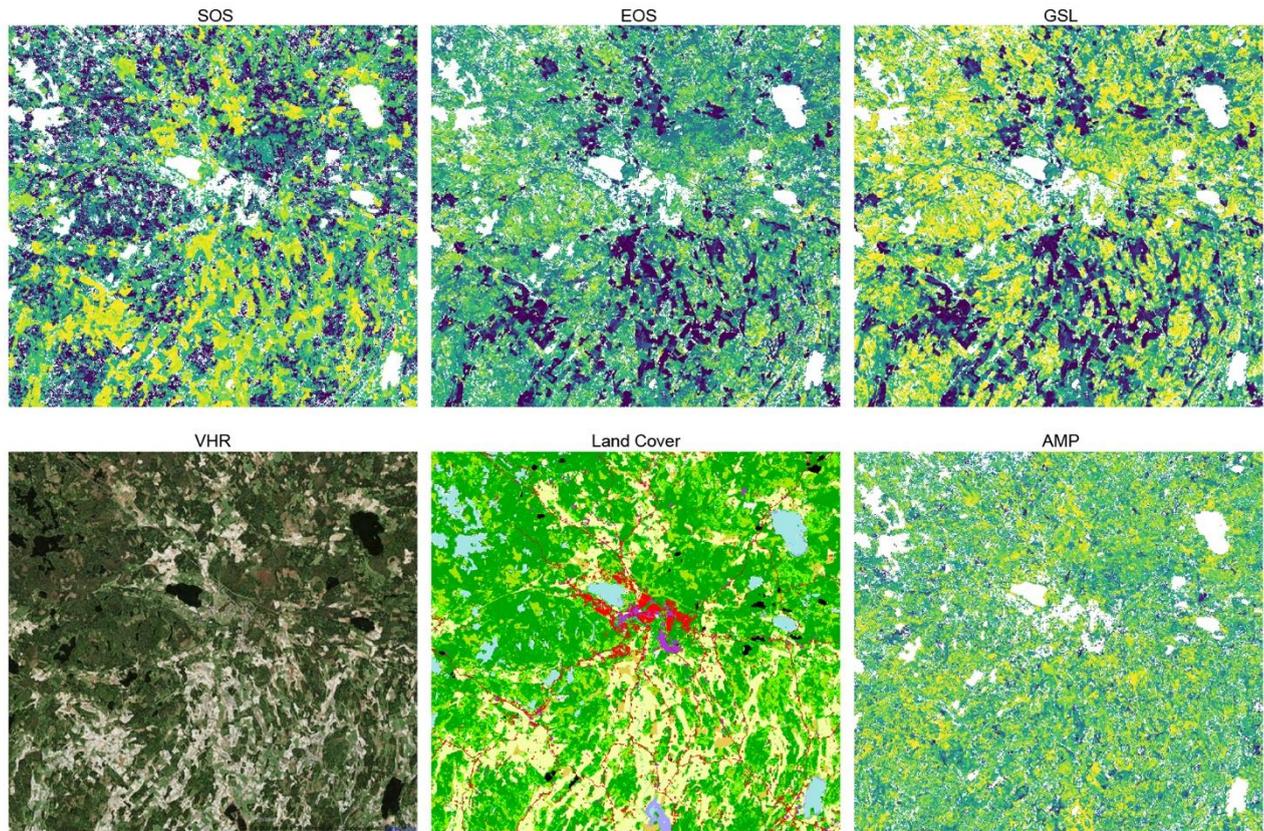


Figure 15: Per pixel LSP metrics over a region (Karkkila, 60°31'57.9"N 24°12'42.9"E) in Finland (see Figure 13 for the legend): Start of Season (SOS), End of Season (EOS) dates, Growing Season Length (GSL) and Amplitude (AMP). Additionally, also the Corine Land Cover (see Figure 11 for the legend) and a Bing maps image is given as reference

## 8. References

- Beck, P. S. A., C. Atzberger, K. A. H. Ilgda, B. Johansen, and A. K. Skidmore. 2006. "Improved Monitoring of Vegetation Dynamics at Very High Latitudes: A New Method Using MODIS NDVI." *Remote Sensing of Environment* 100(3):321-34. Retrieved (<http://www.scopus.com/inward/record.url?eid=2-s2.0-31344451662&partnerID=40&md5=056dff10d968cc51e07f63e8f9c1c049>).
- Browning, Dawn M., Jason W. Karl, David Morin, Andrew D. Richardson, and Craig E. Tweedie. 2017. "PhenoCams Bridge the Gap between Field and Satellite Observations in an Arid Grassland Ecosystem." *Remote Sensing* 9 (10). <https://doi.org/10.3390/rs9101071>.
- Campbell, James B. and Randolph H. Wynne. 2011. *Introduction to Remote Sensing FIFTH EDITION*. New York, London,: Guilford Press.
- Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J.-C., Skakun, S. V., & Justice, C. (2018). The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sensing of Environment*, 219, 145-161.
- Elmore, A. J., S. M. Guinn, B. J. Minsley, and A. D. Richardson. 2012. "Landscape Controls on the Timing of Spring, Autumn, and Growing Season Length in Mid-Atlantic Forests." *Global Change Biology* 18(2):656–74. Retrieved (<http://www.scopus.com/inward/record.url?eid=2-s2.0-84855856944&partnerID=40&md5=5691a2f207e30d544716388848103c10>).
- Filippa, Gianluca, Edoardo Cremonese, Mirco Migliavacca, Marta Galvagno, Matthias Forkel, Lisa Wingate, Enrico Tomelleri, Umberto Morra Di Cella, and Andrew D Richardson. 2016. "Phenopix: AR Package for Image-Based Vegetation Phenology." *Agricultural and Forest Meteorology* 220: 141–50.
- Filippa, Gianluca, Edoardo Cremonese, Mirco Migliavacca, Marta Galvagno, Oliver Sonnentag, Elyn Humphreys, Koen Hufkens, et al. 2018. "NDVI Derived from Near-Infrared-Enabled Digital Cameras: Applicability across Different Plant Functional Types." *Agricultural and Forest Meteorology* 249 (November): 275–85. <https://doi.org/10.1016/j.agrformet.2017.11.003>.
- Fischer, Alberte. 1994. "A Model for the Seasonal Variations of Vegetation Indices in Coarse Resolution Data and Its Inversion to Extract Crop Parameters." *Remote Sensing of Environment* 48(2):220–30.
- Friedl, M. H. et al. 2006. "Land Surface Phenology." A Community White Paper Requested by NASA April 10. Retrieved ([ftp://ftp.iluci.org/Land\\_ESDR/Phenology\\_Friedl\\_whitepaper.pdf](ftp://ftp.iluci.org/Land_ESDR/Phenology_Friedl_whitepaper.pdf)).
- Garonna I (2016) Large-scale Dynamics of Land Surface Phenology and its Climatic Constraints. Phd thesis University of Zurich, Switzerland.
- Keenan, T. F., B. Darby, E. Felts, O. Sonnentag, M. A. Friedl, K. Hufkens, J. O'Keefe, et al. 2014. "Tracking Forest Phenology and Seasonal Physiology Using Digital Repeat Photography: A Critical Assessment." *Ecological Applications* 24 (6): 1478–89. <https://doi.org/10.1890/13-0652.1>.
- Lange, Maximilian et al. 2017. "Validating MODIS and Sentinel-2 NDVI Products at a Temperate Deciduous Forest Site Using Two Independent Ground-Based Sensors." *Sensors (Switzerland)* 17(8):1–18.
- Lieth, H. 1974. "Purposes of a Phenology Book." Pp. 3–19 in *Phenology and Seasonality*

Modeling, edited by J. IL Morison and M. D. Morecroft.

- Menzel, Annette and Tim Sparks. 2006. "Temperature and Plant Development: Phenology and Seasonality." Pp. 70–95 in *Plant Growth and Climate Change*. Oxford: Blackwell Publishing.
- Meroni, Michele, Michel M. Verstraete, Felix Rembold, Ferdinando Urbano, and François Kayitakire. 2014. "A Phenology-Based Method to Derive Biomass Production Anomalies for Food Security Monitoring in the Horn of Africa." *International Journal of Remote Sensing*.
- Myneni, R. B., C. D. Keeling, C. J. Tucker, G. Asrar, and R. R. Nemani. 1997. "Increased Plant Growth in the Northern High Latitudes from 1981 to 1991." *Nature* 386(6626):698–702. Retrieved (<http://dx.doi.org/10.1038/386698a0>).
- Schieber, B, R Janík, and Z Snopková. 2013. "Phenology of Common Beech (*Fagus Sylvatica* L.) along the Altitudinal Gradient in Slovakia (Inner Western Carpathians)." *Journal of Forest Science* 59 (4): 176–84.
- Schwartz, Mark D. 2013. *Phenology: An Integrative Environmental Science*.
- Storn, Rainer and Kenneth Price. 1997. "Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces." *Journal of Global Optimization* 11(4):341–59.
- Storey, James, David P. Roy, Jeffrey Masek, Ferran Gascon, John Dwyer, and Michael Choate. 2016. "A Note on the Temporary Misregistration of Landsat-8 Operational Land Imager (OLI) and Sentinel-2 Multi Spectral Instrument (MSI) Imagery." *Remote Sensing of Environment* 186: 121–22. <https://doi.org/10.1016/j.rse.2016.08.025>.
- Walther, S., Voigt, M., Thum, T., Gonsamo, A., Zhang, Y., Köhler, P., ... Guanter, L., 2016. Satellite chlorophyll fluorescence measurements reveal large-scale decoupling of photosynthesis and greenness dynamics in boreal evergreen forests. *Global Change Biology*, 22(9), 2979–2996. <https://doi.org/10.1111/gcb.13200>
- White, Michael A. et al. 2009. "Intercomparison, Interpretation, and Assessment of Spring Phenology in North America Estimated from Remote Sensing for 1982-2006." *Global Change Biology* 15(10):2335–59.
- Wu, C. et al. 2017. "Land Surface Phenology Derived from Normalized Difference Vegetation Index (NDVI) at Global FLUXNET Sites." *Agricultural and Forest Meteorology* 233:171–82. Retrieved (<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84999106832&doi=10.1016%2Fj.agrformet.2016.11.193&partnerID=40&md5=9185a580aff8e4c7b666dee1f8184ed6>).
- Zeng, H., G. Jia, and H. Epstein. 2011. "Recent Changes in Phenology over the Northern High Latitudes Detected from Multi-Satellite Data." *Environmental Research Letters* 6(4):45508. Retrieved (<http://www.scopus.com/inward/record.url?eid=2-s2.0-84555190803&partnerID=40&md5=d5d5fe61dc321832c893c2f665197caa>).

## 9. Appendix

Decimal Flag	#initial obs <7	mean NDVI < 0.2	Amp. < 0.1	non-phenologica l curve	no dormancy	< 66% valid obs after fit	F-test >0.05
0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0
4	0	0	1	0	0	0	0
6	0	1	1	0	0	0	0
8	0	0	0	1	0	0	0
10	0	1	0	1	0	0	0
12	0	0	1	1	0	0	0
14	0	1	1	1	0	0	0
24	0	0	0	1	1	0	0
26	0	1	0	1	1	0	0
28	0	0	1	1	1	0	0
30	0	1	1	1	1	0	0
32	0	0	0	0	0	1	0
34	0	1	0	0	0	1	0
36	0	0	1	0	0	1	0
38	0	1	1	0	0	1	0
40	0	0	0	1	0	1	0
42	0	1	0	1	0	1	0
44	0	0	1	1	0	1	0
46	0	1	1	1	0	1	0
56	0	0	0	1	1	1	0
58	0	1	0	1	1	1	0
60	0	0	1	1	1	1	0
62	0	1	1	1	1	1	0
64	0	0	0	0	0	0	1
66	0	1	0	0	0	0	1

68	0	0	1	0	0	0	1
70	0	1	1	0	0	0	1
72	0	0	0	1	0	0	1
74	0	1	0	1	0	0	1
76	0	0	1	1	0	0	1
78	0	1	1	1	0	0	1
88	0	0	0	1	1	0	1
90	0	1	0	1	1	0	1
92	0	0	1	1	1	0	1
94	0	1	1	1	1	0	1
96	0	0	0	0	0	1	1
98	0	1	0	0	0	1	1
100	0	0	1	0	0	1	1
102	0	1	1	0	0	1	1
104	0	0	0	1	0	1	1
106	0	1	0	1	0	1	1
108	0	0	1	1	0	1	1
110	0	1	1	1	0	1	1
120	0	0	0	1	1	1	1
122	0	1	0	1	1	1	1
124	0	0	1	1	1	1	1
126	0	1	1	1	1	1	1