LUND UNIVERSITY

Vegetation Observation in the Big Data Era

Sentinel-2 data for mapping the seasonality of land vegetation

Cai, Zhanzhang

2019

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Cai, Z. (2019). *Vegetation Observation in the Big Data Era: Sentinel-2 data for mapping the seasonality of land vegetation.* [Doctoral Thesis (compilation), Dept of Physical Geography and Ecosystem Science]. Lund University, Faculty of Science, Department of Physical Geography and Ecosystem Science.

Total number of authors:
1

Vegetation Observation in the Big Data Era

Sentinel-2 data for mapping the seasonality of land vegetation

ZHANZHANG CAI
DEPARTMENT OF PHYSICAL GEOGRAPHY AND ECOSYSTEM SCIENCE | LUND UNIVERSITY
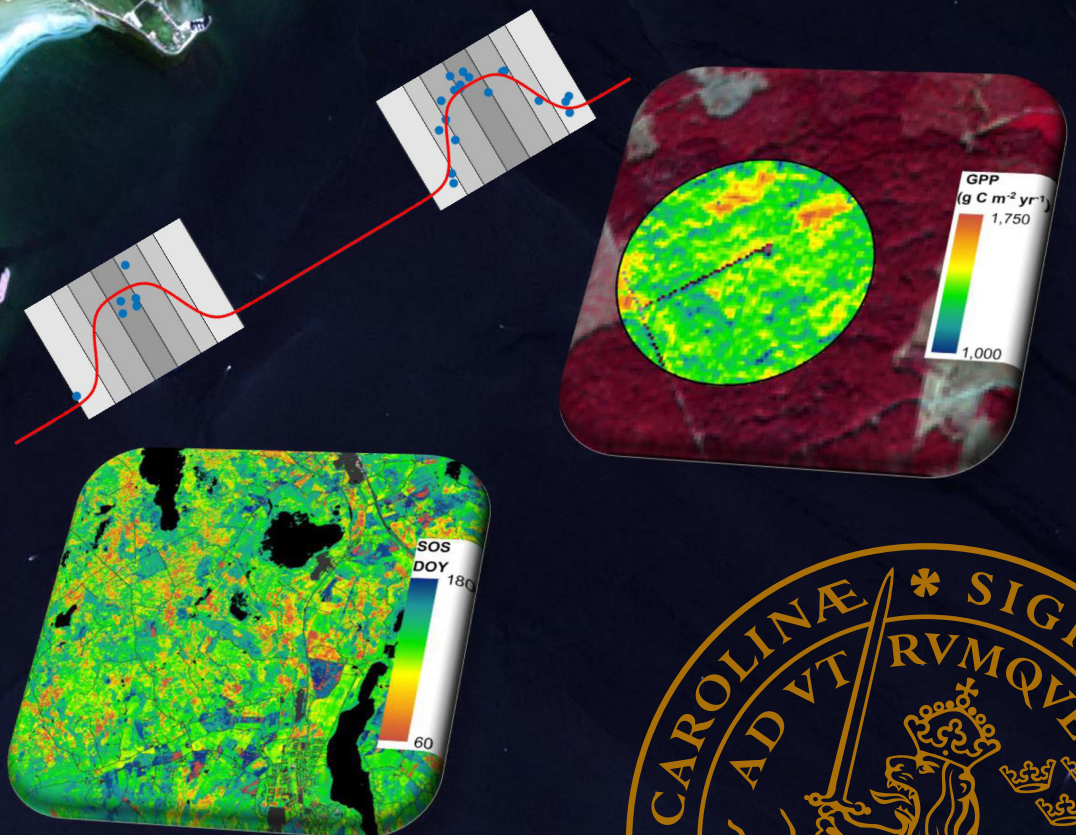
Vegetation Observation in the Big Data Era

Vegetation Observation in the Big Data Era

Sentinel-2 data for mapping the seasonality of land vegetation

Zhanzhang Cai

LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Science, Lund University, Sweden.
To be defended at Pangea auditorium, Geocentrum II, Sölvegatan 12, Lund.
Friday, February 22, 2019, at 10:00 am.

Faculty opponent
Professor Kirsten de Beurs
Geography and Environmental Sustainability
The University of Oklahoma, Norman, USA.

Organization	Document name
LUND UNIVERSITY	DOCTORAL DISSERTATION
Department of Physical Geography and Ecosystem Science	Date of issue
	2019-01-18
Author: Zhanzhang Cai	Sponsoring organization: Swedish National Space Agency

Title and subtitle
Vegetation Observation in the Big Data Era: Sentinel-2 data for mapping the seasonalilty of land vegetation

Abstract

Using satellite remote sensing data for observing vegetation seasonality is an important approach to estimate phenology and carbon uptake of land vegetation. The successful launch of Sentinel-2B in 2017 initiated full operation of the Sentinel-2 twin satellites, and they now provide 10 - 60 m spatial resolution satellite data at 5 days temporal resolution worldwide, releasing approximately 3.2 TB of image data per day. With Sentinel-2's huge amount of high spatial resolution and high temporal resolution data, Earth observation is facing new opportunities and challenges.

To adapt to the characteristics of Sentinel-2 MSI data, the existing time-series analysis methods used for vegetation seasonality studies with regular time step data (e.g., from the MODIS sensor) require modification and improvements. In this thesis, a new time-series analysis method, based on the currently available methods, was developed for estimating vegetation seasonality from high spatial resolution Sentinel-2 data. The new method is applied to Sentinel-2 data to estimate vegetation phenology and photosynthetic carbon uptake, and the outputs are evaluated based on ground reference data and compared to MODIS products.

By comparing with ground reference data (in-situ NDVI time-series, flux tower GPP time-series, and elevation), function fitting methods (e.g., double logistic function fitting) provide the most robust description of the seasonal dynamics for MODIS NDVI time-series among five tested smoothing methods. Based on this finding, we developed box constrained separable least squares fits to double logistic functions with seasonal shape priors, and tested the robustness of the method on six years of simulated Sentinel-2 data by use of MODIS data. The results show that the new method is flexible enough to simulate interannual variations and robust enough when data are sparse.

The box constrained function fitting method applied to Sentinel-2 MSI 2-band Enhanced Vegetation Index (EVI2) data was further used to estimate vegetation phenology and gross primary productivity (GPP) across diverse Nordic vegetation types. The results indicate that daily EVI2 time-series derived from Sentinel-2 is more accurate than from MODIS, with an RMSE of 0.08 for Sentinel-2 and 0.13 for MODIS versus the ground spectral data. With reference to the dates of greenness rising estimated from digital cameras, the dates estimated from Sentinel-2 (RMSE: 8.1 days) are closer than those from MODIS (RMSE: 14.4 days). Sentinel-2 data also generate more phenological details along elevation gradients and land cover variations than MODIS. However, Sentinel-2 does not show any advantage in estimating GPP, when comparing with data from flux towers. The average error between the modelled GPP from Sentinel-2 EVI2 and the GPP derived from flux tower data was similar to that from MODIS. This result partly reflects inabilities in the flux tower data to resolve variation at the same high resolution as Sentinel-2, and further studies will be required to fully evaluate the capability of the sensor in this respect.

In conclusion, the new method, box constrained separable least squares fits to double logistic functions with seasonal shape priors, is useful and computationally efficient for robustly reconstructing daily vegetation index time-series and estimating vegetation phenology from Sentinel-2 data. In addition, by applying the new method to Sentinel-2 data is useful for describing the spatial variation of GPP in the footprint area, although Sentinel-2 did not show improvements in estimating GPP compared with MODIS data. The developed time-series methods will be implemented in a subsequent version of the TIMESAT software package for processing of irregular time step data.

Key words: Sentinel-2, Time-series analysis, Phenology, GPP.

Classification system and/or index terms (if any)

Supplementary bibliographical information	Language: English

ISSN and key title	ISBN (print): 978-91-85793-99-0
	ISBN (PDF): 978-91-985016-0-5

| Recipient's notes	Number of pages 170	Price
	Security classification	

Signature _Zhanzhang Cai_ Date 2019-01-15

Vegetation Observation in the Big Data Era

Sentinel-2 data for mapping the seasonality of land vegetation

Zhanzhang Cai

LUND

UNIVERSITY

"或生而知之，或学而知之，或困而知之，及其知之，一也。
或安而行之，或利而行之，或勉强而行之，及其成功，一也。"

《中庸》

"Some are born with the knowledge; some know them by study; and some acquire the knowledge after a painful feeling of their ignorance. But the knowledge being possessed, it comes to the same thing.

Some do it with a natural ease; some from a desire for their advantages; and some by strenuous effort. But the achievement being made, it comes to the same thing."

-Doctrine of the Mean

List of papers

I. Cai, Z., Jönsson, P., Jin, H. & Eklundh, L. (2017). Performance of Smoothing Methods for Reconstructing NDVI Time-Series and Estimating Vegetation Phenology from MODIS Data. *Remote Sensing,* 9**,** 1271.

II. Jönsson, P., Cai, Z., Melaas, E., Friedl, M. & Eklundh, L. (2018). A Method for Robust Estimation of Vegetation Seasonality from Landsat and Sentinel-2 Time Series Data. *Remote Sensing,* 10**,** 635.

III. Cai, Z., Jin, H., Jönsson, P., Ardö, J. & Eklundh, L. (2018). Estimating vegetation phenology from Sentinel-2 MSI data across diverse Nordic vegetation types. Submitted to *Remote Sensing of Environment*.

IV. Cai, Z., Junttila, S., Holst, J., Karamihalaki, M., Ardö, J., Jin, H., Jönsson, P. & Eklundh, L. (2018). Modelling Daily GPP with Sentinel-2 Data in the Nordic Region – Comparison with Data from MODIS. *Manuscript.*

Contributions

Paper I: Z.C., conceived and designed the experiments, performed the experiments, analysed the data and led the writing.

Paper II: Z.C., contributed to the study design, developed the software, implemented the method and performed all data analysis, participated in the discussion and editing of the manuscript.

Paper III: Z.C., conceived and designed the experiments, performed the experiments, analysed the data and led the writing.

Paper IV: Z.C., conceived and designed the experiments, performed the experiments, analysed the data and led the writing.

Abstract

Using satellite remote sensing data for observing vegetation seasonality is an important approach to estimate phenology and carbon uptake of land vegetation. The successful launch of Sentinel-2B in 2017 initiated full operation of the Sentinel-2 twin satellites, and they now provide 10 - 60 m spatial resolution satellite data at 5 days temporal resolution worldwide, releasing approximately 3.2 TB of image data per day. With Sentinel-2's huge amount of high spatial resolution and high temporal resolution data, Earth observation is facing new opportunities and challenges.

To adapt to the characteristics of Sentinel-2 MSI data, the existing time-series analysis methods used for vegetation seasonality studies with regular time step data (e.g., from the MODIS sensor) require modification and improvements. In this thesis, a new time-series analysis method, based on the currently available methods, was developed for estimating vegetation seasonality from high spatial resolution Sentinel-2 data. The new method is applied to Sentinel-2 data to estimate vegetation phenology and photosynthetic carbon uptake, and the outputs are evaluated based on ground reference data and compared to MODIS products.

By comparing with ground reference data (in-situ NDVI time-series, flux tower GPP time-series, and elevation), function fitting methods (e.g., double logistic function fitting) provide the most robust description of the seasonal dynamics for MODIS NDVI time-series among five tested smoothing methods. Based on this finding, we developed box constrained separable least squares fits to double logistic functions with seasonal shape priors, and tested the robustness of the method on six years of simulated Sentinel-2 data by use of MODIS data. The results show that the new method is flexible enough to simulate interannual variations and robust enough when data are sparse.

The box constrained function fitting method applied to Sentinel-2 MSI 2-band Enhanced Vegetation Index (EVI2) data was further used to estimate vegetation phenology and gross primary productivity (GPP) across diverse Nordic vegetation types. The results indicate that daily EVI2 time-series derived from Sentinel-2 is more accurate than from MODIS, with an RMSE of 0.08 for Sentinel-2 and 0.13 for MODIS versus the ground spectral data. With reference to the dates of greenness rising estimated from digital cameras, the dates estimated from Sentinel-2 (RMSE: 8.1 days) are closer than those from MODIS (RMSE: 14.4 days). Sentinel-2 data

also generate more phenological details along elevation gradients and land cover variations than MODIS. However, Sentinel-2 does not show any advantage in estimating GPP, when comparing with data from flux towers. The average error between the modelled GPP from Sentinel-2 EVI2 and the GPP derived from flux tower data was similar to that from MODIS. This result partly reflects inabilities in the flux tower data to resolve variation at the same high resolution as Sentinel-2, and further studies will be required to fully evaluate the capability of the sensor in this respect.

In conclusion, the new method, box constrained separable least squares fits to double logistic functions with seasonal shape priors, is useful and computationally efficient for robustly reconstructing daily vegetation index time-series and estimating vegetation phenology from Sentinel-2 data. In addition, by applying the new method to Sentinel-2 data is useful for describing the spatial variation of GPP in the footprint area, although Sentinel-2 did not show improvements in estimating GPP compared with MODIS data. The developed time-series methods will be implemented in a subsequent version of the TIMESAT software package for processing of irregular time step data.

Table of Contents

1 Introduction

1.1 Motivation

Land vegetation plays a crucial role in the Earth's ecosystems due to its major influences on the cycles of carbon, energy, and hydrology between atmosphere and ecosystems (IPCC 2014). Satellite remote sensing technology provides an efficient and economical platform for global land vegetation observation.

The first Landsat satellite was launched in 1972, and since then, humans have begun to observe global vegetation from space data. This also led to a revolution in vegetation observation: the observations were no longer random spatial samples but covered the entire land surface. In a broad sense, vegetation observations have entered the era of big data since then. Although Landsat data can cover the whole earth, its low revisit frequency (about 15 days) limits its performance in estimating vegetation seasonality (e.g. Fisher et al. 2006, Melaas et al. 2013). Since the Advanced Very High-Resolution Radiometer (AVHRR) and the Moderate Resolution Imaging Spectroradiometer (MODIS) data were later available, seasonal observations of vegetation on a global scale began. AVHRR and MODIS provide daily images covering all the land surface in the temporal domain.

Scientists have established a variety of methods to analyse the growth curve of vegetation based on these daily data (Menenti et al. 1993, Olsson and Eklundh 1994, Jönsson and Eklundh 2002, Chen et al. 2004, Beck et al. 2006), and the products from these methods have been used for studying the vegetation phenology and carbon cycle of vegetation (Turner et al. 2006, Zhang et al. 2006). However, the low spatial resolution of AVHRR and MODIS with over hundreds of meters limits the observation of vegetation, especially the fact that each pixel contains a mixture of signals and noise, such as different vegetation types, different land types, clouds and cloud shadows, etc., affecting the comparison with ground reference data.

Sentinel-2 is an Earth Observation mission in the European Union's Copernicus programme. The European Space Agency (ESA) launched the Sentinel-2A and Sentinel-2B satellites in 2015 and 2017 respectively, equipped with the Multispectral Instrument (MSI) on-board (Drusch et al. 2012), providing global coverage of $10 – 60$ m spatial resolution, 5 days temporal resolution, and 13 spectral bands of terrestrial observations, and they currently distribute an average of 3.2 TB of satellite image data per day. This combination allows for the development of a range of new applications and products which are invaluable for vegetation observation. Differently from the AVHRR and MODIS data, the Sentinel-2 MSI data are not regular time-step data and in a lower frequency. Therefore, how to

efficiently and accurately extract seasonal information of vegetation from Sentinel-2 data will be the key to effective use of these data in the future.

In this context, this thesis, funded by the Swedish National Space Agency (SNSA) and Lund University, plans to develop and test a methodology for estimating seasonality of vegetation using Sentinel-2 high spatial resolution data and to evaluate the scientific value of the product regarding vegetation phenology and photosynthetic carbon uptake.

1.2 Vegetation seasonality from satellite data

1.2.1 Radiative properties of vegetation

Central to the development of remote sensing for monitoring vegetation is an understanding of the interaction of radiation with vegetation. The radiative properties of vegetation are mainly described from the canopy as a whole and its other components, including leaves, stems, soil, or water (Jones and Vaughan 2010). Among these components, the leaves, essentially the chlorophyll contained therein, are considered as the most crucial components in phenology and carbon studies, since they indicate vegetation growing season and vegetation carbon uptake ability. While radiation from the sun hits a leaf, the magnitudes of spectral reflectance, spectral absorptance and spectral transmittance depend not only on the wavelength but also on a range of chemical and structural characteristics, such as chemical composition, leaf water content, and leaf structure (Curran 1989, Slaton et al. 2001).

The sensors on-board satellites, e.g., AVHRR and MODIS, are not yet able to observe reflection of radiation from leaves of the vegetation only, but always record combined spectral reflection from canopies in a pixel which consist of a number of components, such as green leaves, stems, background soils, background water, etc. Therefore, we need quantitative indices to extract information about the green leaves from the mixed components.

Due to the influence of the atmosphere, what we obtain from the remote sensing satellite sensor is the top-of-atmosphere reflectance. Atmospheric correction is required to obtain surface reflection that carries vegetation information. The correction of the impacts of oxygen and ozone is relatively easy to perform since oxygen and ozone are stable on both spatial and temporal scale. The remaining components, aerosol and water vapour, are more varying and thus the main challenge of atmospheric correction (Liang 2005).

1.2.2 Vegetation indices

Vegetation indices are commonly used in vegetation observation, especially in assessing biophysical and biochemical variables, such as canopy chlorophyll content (Gitelson et al. 2005), leaf area index (LAI) (Tillack et al. 2014), fraction of photosynthetically active radiation absorbed by the vegetation (fAPAR) (Gitelson et al. 2014), and gross primary productivity (GPP) (Xiao et al. 2004). The establishment of most vegetation indices is based on two facts: (1) the chlorophyll accounts for almost all the absorption in the red band and much of that in the blue band, and (2) there is high reflectance in near-infrared band due to leaf scattering.

The most widely used vegetation index is the Normalized Difference Vegetation Index (NDVI, Tucker, 1979):

$$NDVI = (NIR - R)/(NIR + R), \tag{1}$$

where NIR and R are reflectance in the near infrared and red wavelength bands respectively. NDVI can enhance the difference of reflectance information between vegetation and other background information, and it is used as a variable to estimate fAPAR in the light use efficiency (LUE) model (Monteith 1972, 1977). However, in the past thirty years of use, some problems have been observed in using NDVI, such as the saturation in dense vegetation (Jackson et al. 2004) and the sensitivity to background reflectance (Huete 1988, van Leeuwen and Huete 1996, Rocha et al. 2008).

The Enhanced Vegetation Index (EVI, Huete et al. 2002):

$$EVI = 2.5 \times (NIR - R)/(NIR + 6 \times R - 7.5 \times B + 1), \tag{2}$$

where B is reflectance in the blue wavelength band. EVI avoids some of the problems of NDVI, for example, it does not become saturated as quickly as NDVI.

The 2-band Enhance Vegetation Index (EVI2, Jiang et al. 2008):

$$EVI2 = 2.5 \times (NIR - R)/(NIR + 2.4 \times R + 1), \tag{3}$$

was specifically designed for MODIS, does not require the blue band, and it is close to EVI.

A novel development, the physically-based plant phenology index (PPI, Jin and Eklundh 2015), showed a linear relationship with canopy LAI and strong correlation with GPP. Snow influence on PPI is much smaller than on NDVI and EVI.

1.2.3 Time-series smoothing methods

Vegetation seasonality can be observed by establishing a time-series of vegetation indices. However, the accuracy and continuity of the time-series of vegetation indices are affected by noise in the signal received by the satellite sensor due to geometric misregistration, anisotropic reflectance effects, electronic errors, artefacts due to data resampling, clouds, cloud shadows, snow, and aerosols (Goward et al. 1991).

In order to reduce the impacts of noise and to reconstruct continuously seasonal curves, a multitude of time-series processing methods have been used, e.g. Fourier transforms (Menenti et al. 1993, Olsson and Eklundh 1994), wavelet smoothing (Galford et al. 2008), statistical filters (Reed et al. 1994), Savitzky-Golay filtering (Chen et al. 2004), least-squares fits to asymmetric Gaussian functions (Jönsson and Eklundh 2002) and double logistic functions (Zhang et al. 2003, Beck et al. 2006, Fisher et al. 2006), and variations of spline smoothing (Bradley et al. 2007, Hermance et al. 2007, Atzberger and Eilers 2011). However, there is no final conclusion about which method is always the best (Hird and McDermid 2009, White et al. 2009, Atkinson et al. 2012, Geng et al. 2014).

1.2.4 TIMESAT

TIMESAT is a software program developed by P. Jönsson and L. Eklundh (2002, 2004) for time-series analyses of satellite data. The latest TIMESAT version 3.3 includes trend analysis, time-series smoothing and extraction of seasonality parameters. These analyses can only be used for regular time step data, e.g., products from AVHRR and MODIS. The method used for trend analysis in TIMESAT is the seasonal-trend decomposition procedure based on Loess (STL) (Cleveland et al. 1990), which decomposes trends, seasons and noise in the time series. TIMESAT includes Savitzky-Golay filtering (SG), asymmetric Gaussian functions (AG), and double logistic functions (DL). To improve the accuracy of data fitting, a weighting system and upper envelope adaptation are available. The weighting system allows assigning a weight to each data point so that the smoothing curve prefers to follow the high-quality data values and decreases the influence from low-quality data values. The upper envelope adaptation adds a positive bias to the fits and has been used with many smoothing algorithms to minimise the effects of cloud contamination that generally decreases the estimations of vegetation indices such as NDVI (Chen et al. 2004, Jönsson and Eklundh 2004). The last step of the TIMESAT processing is extracting key time points or key values from smoothed time-series, i.e. seasonality parameters (Figure 1). Start-of-season (SOS) and end-of-season (EOS) dates are generally specified as the point in time when a defined fraction (e.g. 20%) of the seasonal amplitude has been reached.
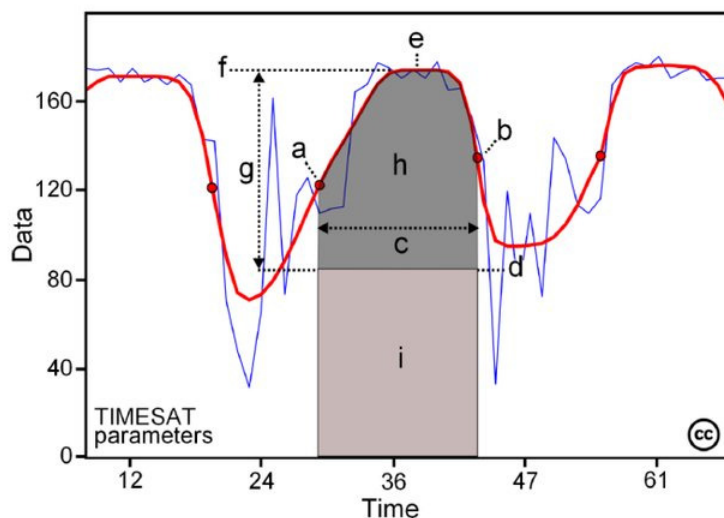
Figure 1. Some of the seasonality parameters generated by TIMESAT: (a) beginning of season, (b) end of season, (c) length of season, (d) base value, (e) time of middle of season, (f) maximum value, (g) amplitude, (h) small integrated value, (h+i) large integrated value. The red and blue lines represent the filtered and the original data, respectively (Eklundh and Jönsson 2017).

1.2.5 Applications of vegetation seasonality

1.2.5.1 Mapping phenology

'*Phenology is the study of timing of recurring biological events, the causes of their timing with regard to biotic and abiotic forces, and the interrelation among phases of the same or different species.*' (Lieth 1974)

Seasonality parameters extracted by time series of vegetation index is an important and widely used approach of large-scale phenological mapping (Schwartz 2003, Delbart et al. 2006, Fisher et al. 2007), and it has also been used to produce a global phenology product, e.g., MODIS MCD12Q2 product (https://lpdaac.usgs.gov) (Zhang et al. 2003). Phenology estimated from satellite remote sensing is generally called land surface phenology (de Beurs and Henebry 2004). The concept of land surface phenology is widely used to study global climate, large-scale ecosystem, and forest management (Heumann et al. 2007, Lioubimtseva and Henebry 2009, de Beurs and Henebry 2010). Due to seasonality parameters are normally estimated from the coarse resolution satellite remote sensing data, one pixel usually describes the aggregate temporal behaviours of multiple vegetation species and land cover types.

Gross primary productivity (GPP) stands for the overall rates of photosynthetic uptake of carbon by leaves (Bonan 2015). GPP is an important variable for studying the global carbon cycle (Prince 1991, Running et al. 2004). Vegetation index time-series from satellite data is a crucial variable for modelling regional and global GPP because the content of chlorophyll is directly related to GPP and the vegetation index responds to radiation absorption of chlorophyll. The accuracy of GPP modelling is related to the spatial and temporal resolution of satellite data. There are two main types of top-down models commonly used to estimate GPP in remote sensing: simple statistic models (e.g., Schubert et al. 2010, Schubert et al. 2012) and light use efficiency (LUE) models (Monteith 1972, Monteith John et al. 1977). Both statistic and LUE models have been used in many studies to estimate GPP (e.g., Prince 1991, Ruimy et al. 1994, Running et al. 2004, Xiao et al. 2004, Olofsson et al. 2008, Wu et al. 2010, Sjöström et al. 2011, McCallum et al. 2013). The vegetation index is the key driving variable in both models. Statistic and LUE GPP models are all empirical models, acting as useful complementary to the process-based GPP models, like modelling GPP bottom-up in DGVM (Smith et al. 2001) using Farquhar photosynthesis model (Farquhar et al. 1980).

1.3 Ground reference data

Ground-based measurements using multispectral sensors provide an efficient approach for more precise observation of vegetation reflectance (Fensholt et al. 2004, Gamon et al. 2006, Eklundh et al. 2011, Lange et al. 2017). After regular sensor calibration following Jin and Eklundh (2015), the vegetation reflectance of each spectral band can be accurately measured. Lange et al. (2017) used ground-based multispectral sensor data to validate Sentinel-2 data at a temperate deciduous forest site.

Phenological cameras (PhenoCams) have been widely used for the observation of vegetation phenology, and consequently to evaluate the accuracy of satellite estimated vegetation phenology (Richardson et al. 2007, Sonnentag et al. 2012, Melaas et al. 2016, Baumann et al. 2017, Richardson et al. 2018b, Vrieling et al. 2018). Red-Green-Blue (RGB) images from PhenoCams provide detailed information of overstorey and understorey seasonal development, species distribution, and weather conditions, e.g., sunny, cloudy, snowy. To compare with the satellite NDVI, Petach et al. (2014) used an infrared-enabled security camera to capture ground-based NDVI time-series. However, the difference in viewing angles between satellites and PhenoCams may cause non-negligible differences in phenological parameters estimation (Vrieling et al. 2018).

1.4 Sentinel-2

ESA's Sentinel-2 MSI data have been used in monitoring vegetation phenology (Vrieling et al. 2018) and vegetation species classification (Immitzer et al. 2016, Persson et al. 2018). Sentinel-2 MSI data are significantly different from AVHRR and MODIS data regarding space, time and spectral resolution. First, Sentinel-2's spatial resolution of 10 – 60 meters enables clearly observing the contours of land cover, clouds and cloud shadow, allows more accurate matching of ground observation footprint areas, and enables more precise classification of land cover. Second, the lower return interval (5 days at the equator) means that reliable cloud free maximum value composites of 8-10 days, similar to AVHRR and MODIS products, cannot be generated. Using all available cloud-free observations of Sentinel-2 generates irregular and sometimes wide time-steps, which means that to obtain a smooth vegetation index time series it is necessary to improve the previous smoothing methods or explore new methods. Third, Sentinel-2 MSI provide observations in red edge bands that adds new opportunities for capturing the spectral characteristics of the vegetation. Due to the huge volume of data and the spatial/temporal characteristics of Sentinel-2, the use of Sentinel-2 data to observe vegetation at global or regional scale requires consideration of data processing, data management and data application methods (Figure 2).

1.5 Objectives

The overall aim of the thesis is to explore the utilisation of Sentinel-2 data in science as well as in applied fields. More specifically, the project aims at developing and testing methodology for estimating terrestrial vegetation seasonality at high spatial resolution based on Sentinel-2 data and assessing the scientific value of these high-resolution products within the fields of phenology and photosynthetic carbon uptake. The dynamic capabilities of Sentinel-2 data will be assessed and compared with similar products based on coarse-resolution data. Specific aims are:

- To evaluate the performance of smoothing algorithms in representing seasonal vegetation growth with regular time-step satellite vegetation index time-series data by employing a variety of reference data sets.

- To present a data processing method feasible for generating seasonal data from sparse and irregular time-step data.

- To investigate if vegetation phenology estimated from Sentinel-2 data by the new method improves the agreement with in-situ observed vegetation phenology in comparison to MODIS data for Nordic vegetation types.

- To investigate if the spatial and temporal resolution of Sentinel-2 data are sufficient for estimating GPP and if they can improve the estimation of GPP in comparison to MODIS data.

Figure 2. Mind map of vegetation observation in the Big Data Era: Sentinel-2 data for mapping the seasonality of land vegetation. Created with iMindMap (www.iMindMap.com).

2 Material and Methods

2.1 Satellite data

2.1.1 Sentinel-2 MSI data

The Sentinel-2A and 2B MSI level-1C top of atmosphere reflectance data (Drusch et al. 2012) were used in this thesis. The study sites in the thesis were covered by eight 100km×100km tiles (Paper II, Paper III, and Paper IV). A total of 1,489 available Sentinel-2A and -2B MSI images (978 GB) from 2016 to 2017 were downloaded from the ESA Copernicus Sentinels Scientific Data Hub. The data volume has been doubled since June 16, 2017, when the Sentinel-2B became fully operational. These images were atmospherically corrected using the Sen2Cor processor (2.4.0) (Louis et al. 2016) to obtain level-2A land surface reflectance. We finally generated 1.1 TB of level-2A data since Sen2Cor processor can only process entire tiles.

For observing vegetation greenness, we applied Normalized Difference Vegetation Index (NDVI, Tucker 1979) in Paper II and 2-band Enhanced Vegetation Index (EVI2, Jiang et al. 2008) in Paper III and Paper IV. In paper II, we developed a robust vegetation time-series fitting method based on NDVI. We used EVI2 in Paper III to evaluate whether the developed fitting method for processing Sentinel-2 MSI data improved the phenology retrieval. In addition to EVI2 being superior to NDVI in less sensitive to background reflectance and presenting more details over dense vegetation (Rocha and Shaver 2009), another reason for using EVI2 in Paper III was that EVI2 had been widely used in phenological studies (Zhang et al. 2018). In Paper IV, we continued to use EVI2 to estimate GPP, since Enhanced Vegetation Index (EVI), which is similar to EVI2, has been proven to be superior to NDVI in estimating GPP (Xiao et al. 2005).

We calculated each pixel's vegetation index value in Paper II and Paper IV and vegetation index values of ground observations footprints in Paper III. Since each footprint of ground observations in Paper III covered more than one 10 m resolution Sentinel-2 MSI pixel, we calculated the vegetation index value of the footprint area by averaging the red reflectance and the NIR reflectance at the footprint pixels respectively.

In addition to generating land surface reflectance through the processing of Sen2Cor, this process generates 20 m resolution scene classification information. The processed pixels were divided into 12 categories: no data, saturated or defective,

dark area pixels, cloud shadow, vegetation, bare soil/desert, water, cloud (low probability), cloud (medium probability), cloud (high probability), thin cirrus, and snow or ice (Louis et al. 2016). In this thesis, the pixels from the scene classification were resampled down to 10 m resolution to match the spatial resolution of red and NIR bands. Only pixels classified as vegetation and bare soil/desert were considered as high-/good-quality pixels. The proportion of high-quality pixels of the total footprint's pixels were used to represent the quality of the footprint vegetation index in Paper III.

2.1.2 Landsat data

Images from Landsat satellites were used in the development and testing of the new fitting method in Paper II, since its long historical time-series records. The main reason for using these images was that they had long historical records and irregularly temporal samples similar to Sentinel-2. A side-lap area of a Landsat tile in the Norunda region was used to build up a relatively dense observation data set. We used Landsat-5 Thematic Mapper (TM) and Landsat-7 Enhanced Thematic Mapper Plus (ETM+) surface reflectance data from 2000 to 2014 in the development, and used Landsat-8 Operational Land Imager (OLI) NBAR data from the Harmonized Landsat Sentinel-2 (HLS) surface reflectance product (Claverie et al. 2018) during 2013 to 2017 in the testing. These data were atmospherically corrected by using the LEDAPS algorithm (Masek et al. 2006) and the LaSRC algorithm (Vermote et al. 2016) respectively. Quality assessment (QA) information was generated by FMASK (function of mask), which marked per-pixel land, cloud, cloud shadow, snow, and water (Zhu and Woodcock 2014).

NDVI was calculated for each Landsat 30 m pixel in keeping with the data processing of Sentinel-2. The quality label of each pixel was created from the outputs from FMASK. To exclude some unrealistically high NDVI values in winter, the Landsat 8 OLI data recorded at solar zenith angle larger > 75° was marked as low-quality data.

2.1.3 MODIS data

We used MODIS data in this thesis for the following three purposes: first, they were used to evaluate the performance of the fitting methods in the past studies on the smoothing vegetation index time-series in Paper I; second, they were used for testing the performance of the newly developed fitting method in Paper II; and third, they were used for comparing to Sentinel-2 MSI data in Paper III and Paper IV.

In Paper I, the MOD09Q1 V005 MODIS/Terra 8-day interval 250 m land surface reflectance dataset (Vermote 2015b) was used to generate NDVI in order to evaluate the performance of most commonly used fitting methods. It is the highest spatial resolution MODIS product, which can maximise correspondence with ground observations. The 8-day interval product has less noise than daily MODIS products, and it expresses more seasonal detail than the 16-day interval and monthly products. The choice also aimed to minimise the impact of the quality and quantity of satellite data on the results when comparing different smoothing methods. The quality weights for each observation were generated from binary MODIS quality flags.

In Paper II, the MOD09GA V006 MODIS/Terra daily 500 m land surface reflectance product (Vermote and Wolfe 2015) from 2011 to 2016 was used to evaluate the performance of the newly developed fitting method. Since ground observations were not used in this paper, fine spatial resolution was not required. The key task in the evaluation procedure was to evaluate the robustness of the method in fitting irregular time step data. MOD09GA was used as a reference dataset, and it was also used for generating a simulated Sentinel-2 dataset from 2011-2016. The simulated dataset only included the MODIS observations that corresponded exactly in date with the observations from Sentinel-2 observed in 2016. The quality weights were created from binary MODIS quality flags for excluding low-quality observations.

Since ground observation data were used to evaluate and compare satellite data of different spatial resolution in Paper III and Paper IV, we used two different spatial resolution MODIS datasets: the MOD09Q1 V006 MODIS/Terra 8-day interval 250m land surface reflectance dataset (Vermote 2015b) and the MOD09A1 MODIS/Terra 8-day interval 500m land surface reflectance dataset (Vermote 2015a). We applied the same pre-processing method as in Paper I for these two datasets.

2.2 Analysing the seasonality

2.2.1   Smoothing methods on regular time step data

Among the many time-series smoothing methods applied on analysing vegetation seasonality from regular time step data, we selected five common methods in Paper I: adaptive Savitzky-Golay filtering (SG) (Savitzky and Golay 1964, Chen et al. 2004), adaptive LOESS filtering (LO) (Cleveland and Devlin 1988), spline smoothing (SP) (Craven and Wahba 1979, Woltring 1986), least-squares fits to asymmetric Gaussian functions (AG) (Jönsson and Eklundh 2002) and double logistic functions (DL) (Zhang et al. 2003, Beck et al. 2006). These five methods

can be generally characterised as belonging to two broad categories: local methods (SG and LO) and global methods (AG and DL). SP, according to different settings, can be characterised as local or global. Each method has different combinations of parameter setting. We tested all these settings and used ground observation data to evaluate the methods (see 2.3). The result of these experiments in Paper I guided the direction of the method development in the following papers.

2.2.2   Method development

Based on the results of Paper I and the characteristics of irregular time step satellite data, we chose double logistic functions (Fischer 1994) as the basis in Paper II to develop a new smoothing method, box constrained separable least squares fitting to double logistic model functions with shape priors, for reconstructing continuous time-series. A shape prior is a general average shape over all growing seasons, and it was computed per pixel. The method was implemented in the TIMESAT platform (Jönsson and Eklundh 2004, Eklundh and Jönsson 2017), and applied to process Sentinel-2 MSI, MODIS, and ground data in Papers III and IV.

In this method the output time-series y_m is a sum of n basis functions, one for each season

$$y_m = c_0 + \sum_{i=1}^{n} c_i b\left(\boldsymbol{x}^i, t\right), \tag{4}$$

where c_0 is the base level, c_i is the amplitude factor for season i, and basis functions are taken as double logistic functions (Fischer 1994)

$$b\left(\boldsymbol{x}^i, t\right) = \frac{1}{1+\exp\left(\frac{x_1^i - t}{x_2^i}\right)} - \frac{1}{1+\exp\left(\frac{x_3^i - t}{x_4^i}\right)} , \tag{5}$$

where x_1^i, x_2^i, x_3^i and x_4^i are the four parameters for determining double logistic functions' left inflexion points, the time period of increase, the right inflexion points and the time period of falling for season i respectively.

The detailed process for estimating these parameters in TIMESAT is shown in Figure 3:

1. extract single time-series from the image stack, including vegetation index values, weights, and dates;
2. keep only good quality data for the following processes; end if the number of remaining data lower than a certain value, e.g., at least 5 data points are required to fit a double logistic function (Equation 5);
3. determine the base level c_0, using a low percentile of the clear observation histogram for the fall time period, e.g., 5%;

24

4. apply sinusoidal functions to detect the seasons' rising and falling positions, as the initial values of x_1^i and x_3^i;
5. determine a shape prior as a common shape for all seasons, with initial values for x_1^i and x_3^i;
6. divide the shape prior for each season into seven regions and then detect if there are sufficient points in the regions to determine function parameters in the season (Paper II Table 2);
   a. if no, lock the parameter to those from the shape prior;
   b. if yes, set upper and lower boundaries for the parameters, defining a 'box' (Coleman and Li 1996);
7. determine double logistic functions for each season based on the constrained boundaries in step 6, and reconstruct daily vegetation index.
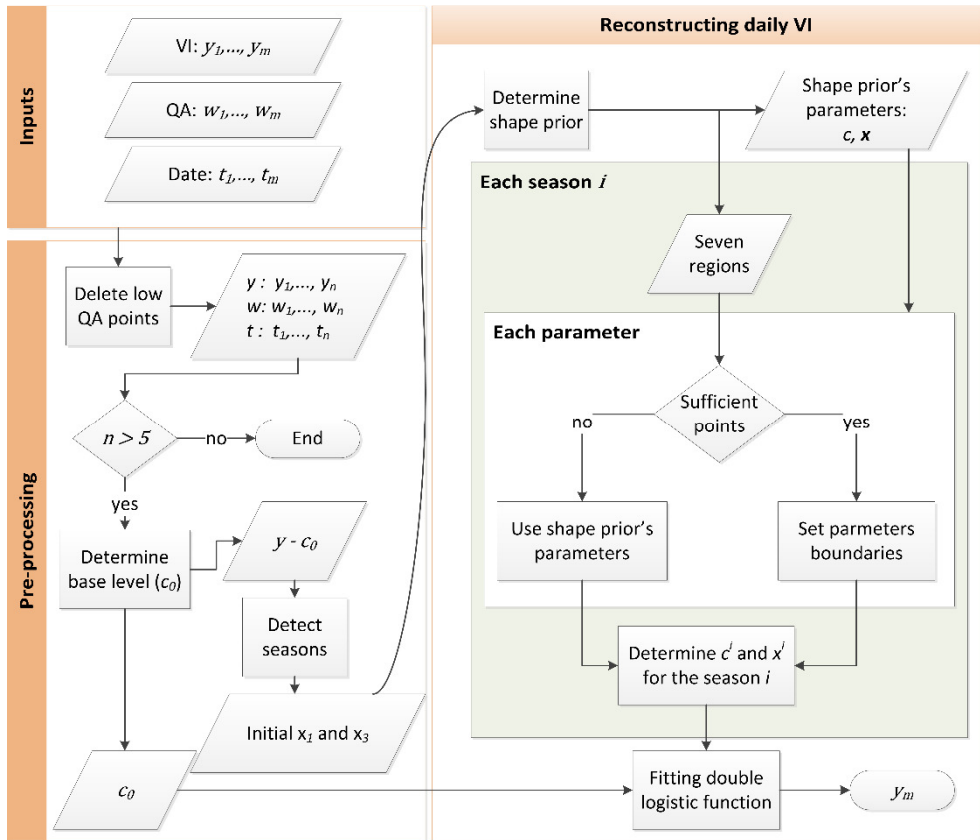

Figure 3. Flowchart of the shape prior and box constrained least squares fit in TIMESAT.

Using the new method to process two years of full size (10980×10980 pixels) Sentinel-2 images at one tile (186 images on average) requires an average of 10 core hours.

2.3 Ground reference data

Use of reliable ground data is critical for the validation and evaluation of satellite data and products. We used various ground data in this thesis to evaluate satellite driven vegetation index time-series, phenology, and GPP.

2.3.1 Ground spectral data

The continuously measured multispectral data at ten sites were used in Paper I and III, in comparison with vegetation index time-series reconstructed from MODIS and Sentinel-2 data. The sensors used to measure these multispectral data were calibrated following the method by Jin and Eklundh (2015), and their footprint areas were estimated based on the sensor height above the canopy, view azimuth angle, field of view (FOV), and off-nadir angle (Eklundh et al. 2011). These multispectral data were converted to red and NIR reflectance, and then NDVI and EVI2 were calculated for comparing to the satellite data. Since the comparison between ground spectral data and satellite data was direct, we did not apply any smoothing method to the ground spectral data. The comparison of ground spectral data and satellite data was limited to the growing seasons for reducing background noise effects, particularly snow, on the results.

2.3.2 PhenoCam

The PhenoCam data in 2016 and 2017 from five Swedish ecosystem stations belonging to ICOS Sweden (Integrated Carbon Observation System) (ICOS Sweden 2018) were used in Paper III as reference data to evaluate greenness rising dates and greenness falling dates. We extracted Green Chromatic Coordinates (GCC) time-series from the PhenoCam images, a similar approach used by Sonnentag et al. (2012) and Richardson et al. (2018a) to process the PhenoCam images. The GCC time-series were further processed using the shape prior and box constrained fitting method for extracting the greenness rising dates and greenness falling dates.

2.3.3　In-situ environment data and flux-tower derived GPP

In Paper IV, we used daily EVI2 estimated from Sentinel-2 MSI, MODIS and in-situ environmental variables (photosynthetic photon flux density or air temperature) to drive empirical linear regression GPP models at eight flux tower sites located in Nordic countries and belonging to the ICOS infrastructure (https://www.icos-ri.eu/). The outcomes were compared to flux data derived daily GPP, which was estimated from partitioning net ecosystem exchange (NEE) with in-situ measured environmental variables, e.g., global radiation, air and soil temperature, and vapour pressure deficit by using the REddyProc tool (Reichstein et al. 2005, Wutzler et al. 2018). The footprint area of flux data was defined as a static circle from the centre of the flux tower with a radius of 10 times the altitude of the measurement, which approximately reflected flux footprint under strong convection (Weil and Horst 1992).

2.3.4　Elevation

Elevation data, at 50 m spatial resolution in Paper I and at 2 m spatial resolution in Paper III, were used to evaluate phenology parameters derived from smoothed vegetation index data. The elevation data were obtained from the Swedish mapping, cadastral and land registration authority, Lantmäteriet. In the analysis, these elevation data were resampled to match MODIS 250 m spatial resolution and Sentinel-2 10 m spatial resolution respectively.

3 Results: Research paper summaries

3.1 Summary of paper I

Title: Performance of smoothing methods for reconstructing NDVI time-series and estimating vegetation phenology from MODIS data

Introduction: In this study, we investigated the performance of five commonly used smoothing methods, Savitzky-Golay filtering (SG), locally weighted regression scatterplot smoothing (LO), spline smoothing (SP), least-squares fitting to asymmetric Gaussian functions (AG), and least-squares fitting to double logistic functions (DL), with all 1092 possible parameter settings (simulations) in smoothing MODIS derived NDVI. We used ground spectral tower measured NDVI at 10 sites and carbon flux tower estimated GPP at 4 sites to evaluate the smoothed satellite-derived NDVI time-series, and the elevation data over the mountainous Ammar area was used to evaluate phenology parameters estimated from smoothed NDVI.

Research highlights:

- The smoothing methods reduced the error between MODIS NDVI and ground-measured NDVI in 89% of the simulations, with the average root mean square error (RMSE) decreasing from 0.14 to 0.08.

- All the smoothing methods increased the average Spearman's rank correlation coefficient (ρ) between GPP and NDVI from 0.34 to 0.51 and up to 0.64 with optimal parameters.

- Generally, differences between methods were small and no single method always performed better than the others.

- Cross-validation was useful for selecting parameters for SG, LO, and SP. It improved the fits and gave fairly good results; however, in some cases the method failed.

- The function fitting methods (AG and DL) derived phenological parameters that always showed the strongest and most robust relationships with elevation across a topographical gradient.

- The function fitting methods were found to generally reduce the risk of achieving very poor results, making them safer than the other methods to be used when it is not possible to carry out any calibration against ground measurements.

3.2 Summary of paper II

Title: A method for robust estimation of vegetation seasonality from Landsat and Sentinel-2 time series data

Introduction: The developed method was based on the finding in paper I that double logistic fitting function was more robust than other methods. We presented a data processing method based on double logistic functions, shape prior, and box constrained separable least squares fits to logistic model functions. The design and initial testing of this method was done on 15 years of Landsat TM/ETM+ NDVI time-series data. The method aimed to fit continuous seasonal functions to time-series of irregular satellite remote sensing data, e.g. Landsat and Sentinel-2, and to be robust in handling data gaps. For a detailed description of the method, see 2.2.2. Once the method was developed, we tested it for extracting phenological parameters from Landsat OLI and Sentinel-2 MSI data. In addition, we tested the robustness of the method by using simulated Sentinel-2 data from MODIS data for the period 2011-2016 (data description in 2.1.3). We generated two sets of data: one fitted from simulated data in 2016 with shape priors, and the other fitted from simulated data in 2016 without shape priors. These two output data sets were compared to daily MODIS data as a reference.

Research highlights:

- We developed a flexible and robust method, the shape prior and box constrained separable least squares fitting to logistic model functions, for modelling the phenology of growing seasons with data from optical satellites like Landsat and Sentinel-2 at irregular time step.

- Using the shape prior can add robustness to the function fitting. With shape prior, the RMSE between simulated start of season (SOS) and reference SOS was reduced from 24.5 days to 8.5 days, and the RMSE between simulated end of season (EOS) and reference EOS was reduced from 18.4 days to 13.2 days.

- The method relies on accurate labelling of pixel quality and the availability of data from long time series in order to obtain stable parameters.

- For Sentinel-2, the proposed method allows extending the time series backwards using the Landsat records for the first years of operation.

- This method requires testing in different biomes to better understand how to choose parameters for base level determination and parameter constraints.

- The proposed method is implemented in the TIMESAT software package and available for parallel processing.

3.3 Summary of paper III

Title: Estimating vegetation phenology from Sentinel-2 MSI data across diverse Nordic vegetation types

Introduction: Based on the proposed method in the previous study (paper II), we aimed to further investigate if vegetation phenology estimated from Sentinel-2 data by this method improves the agreement with ground observed vegetation phenology in comparison to MODIS data. We compared the reconstructed daily time-series of EVI2 and phenology estimations from Sentinel-2 MSI and MODIS datasets to ground measured EVI2 at five sites and phenology estimations from PhenoCam GCC at six sites. Elevation and land cover map data were used to demonstrate the ability of Sentinel-2 data to represent the spatial details of phenology. At the same time, the above experiments also tested the ability of the proposed method in precisely extracting vegetation phenological information from satellite data.

Research highlights:

- The method produced satisfactory results across all the vegetation types, and due to the higher spatial resolution, Sentinel-2 generated data that more accurately matched ground measurements of EVI2 than what was achieved with MODIS data, with an RMSE of 0.08 for Sentinel-2 and 0.13 for MODIS versus the ground spectral data.

- With PhenoCam GCC estimations as the reference, Sentinel-2 generated smaller RMSEs for greenness rising (8.1 days) than for greenness falling (17.3 days). Sentinel-2 greenness rising had smaller RMSE than MODIS greenness rising, but the result of greenness falling did not show that Sentinel-2 was better than MODIS.

- This study could not verify if PhenoCam GCC data as reference data was accurate enough for estimating phenological dates.

- The 10 m resolution of Sentinel-2 could effectively present phenological variations along an elevation gradient. The rates of greenness rising and falling changes along rising elevation for deciduous forest were 0.22 day m^{-1} and -0.11 day m^{-1} ($p < 0.00$), and for heath were 0.29 day m^{-1} and -0.29 day m^{-1} ($p < 0.00$).

- Sentinel-2 generated clear phenological details in land cover variations. Each vegetation type showed different characteristics of greenness rising dates.

- Processing of Sentinel-2 data with the box constrained data smoothing method for producing 10 m vegetation phenology maps and other dynamic vegetation products was successful in the different Nordic ecosystems.

3.4 Summary of paper IV

Title: Modelling daily GPP with Sentinel-2 data in the Nordic region – comparison with data from MODIS

Introduction: In this study, we evaluated the performance of the proposed box-constrained function fitting method and Sentinel-2 MSI data for modelling GPP. Empirical linear regression GPP models driven by daily EVI2 and environmental variables (air temperature/photosynthetic photon flux density) were created at eight Nordic ecosystem stations for simulating daily GPP. We used flux tower estimated GPP as ground reference data. As a continuation of the previous studies, we compared Sentinel-2 to MODIS and investigated if Sentinel-2 MSI data can improve the accuracy of GPP estimation.

Research highlights:

- The errors between the satellites estimated GPP and the flux towers estimated GPP varied among sites (RMSE: 0.63 - 2.69 g C m^{-2} d^{-1}).

- In comparison to flux towers GPP, there were small differences between Sentinel-2 MSI GPP (RMSE: 1.60 g C m^{-2} d^{-1}) and MODIS GPP (RMSE: 1.61 g C m^{-2} d^{-1}).

- The usage of static footprint area significantly limited the accuracy of GPP estimation from satellite data.

- The quantitative differences of GPP estimation due to different spatial resolution EVI2 inputs were smaller than due to the different GPP model formulations used.

- Sentinel-2 MSI 10 m data can reveal strong spatial differences within the flux footprint area.

- A combination of improved processing methodology and input data preparation is required to improve the accuracy and precision of GPP estimations.

4 Discussion

By summarizing results from the four papers of this thesis, the advantages of the high spatial resolution exhibited by the Sentinel-2 MSI are obvious for extracting seasonality of vegetation compared to MODIS. This advantage allows Sentinel-2 data and its products to show more surface detail and more accurately match the footprint area of ground reference data. We developed the 'box constrained separable least squares fit to double logic function with shape priors' to effectively reconstruct the continuous vegetation index time-series from Sentinel-2 irregular time step observations. In this section, the results of the four papers will be discussed in the context of vegetation observations in the big data era (Figure 2), which will focus on the contribution and limitation of the new fitting method and the approach of satellite-ground evaluation.

4.1 Data processing

Data processing is the core of the field of vegetation observation in the big data era. It is a key step in the use of Sentinel-2 satellite data to provide products to downstream studies. The Sentinel-2 data processing flow used in papers II, III, and IV approximately follows the processing flow of AVHRR and MODIS data: atmospheric correction, extracting pixel-wise vegetation index, and time-series analysis. However, the differences of Sentinel-2 MSI data from MODIS and AVHRR data in temporal and spatial resolution lead differences in each processing step.

4.1.1 Processing of spectral data

Atmospheric correction and pixel quality labelling are important pre-processing steps before time-series analysis of vegetation index can be performed. For Sentinel-2 data, Sen2Cor is used for correcting atmospheric effects and labelling pixel qualities (Louis et al. 2016). The main difference of pixel quality labelling between fine spatial resolution and coarse spatial resolution is that more details, e.g., cloud shadow, can be labelled by in fine spatial resolution pixel qualities (Zhu et al. 2015). Although cloud shadows are much clearer seen in Sentinel-2 and Landsat images than in MODIS images, automatically identifying cloud shadows is still difficult (Zhu and Woodcock 2012). The cloud fundamentally changes the spectral properties of the pixel, since it is essentially composed of aerosols and water droplets. The cloud shadow weakens the light intensity. When estimating the

reflectance, all the bands are underestimated, but the relationships between the bands do not change much, so the spectral characteristics of the pixel are largely retained. For this reason, we can still distinguish the land surface features under the cloud shadows, resulting in some difficulty to identify the shadow. This difficulty causes some error in the outputs. As the example in Figure 4 shows, the classification scene generated from Sen2Cor did not accurately identify the cloud shadow. The pixels that should have been marked as cloud shadows are labelled as vegetation, which leads to a reduction in the accuracy of the results. Cloud shadows add noise to the time-series analysis and may affect the detection of phenological parameters.
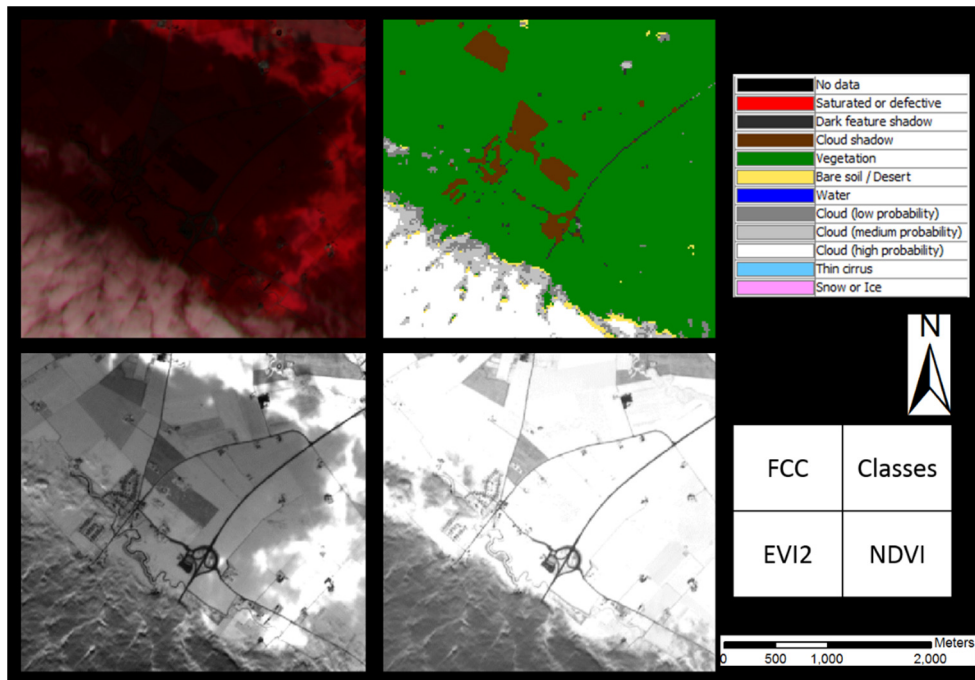
Figure 4. An example of vegetation indices and data quality from 33VUE area, taken at 10:30 am on June 29, 2017. The top left image shows Sentinel-2 false-colour composite (FCC), where the white part is the top of the cloud and the black part is the shadow of the cloud. The top right image is the preliminary classification result generated by the Sen2Cor atmospheric calibration program, which represents the quality of the data. The lower left and lower right maps represent the EVI2 and NDVI vegetation indices (EVI2 and NDVI); high values are bright shades, while the lower values are dark shades.

In addition, we can see that the impacts of cloud shadows on EVI2 and NDVI are different. EVI2 has a significantly lower value under the shadow, while NDVI shows little effect that is hard to see from the images (Figure 4).

34

4.1.2   Time-series processing

One of the difficulties in smoothing satellite remote sensing time-series data is that the noise in the vegetation index time-series does not conform to any distribution, but it is normally a negative bias. From the example in 4.1.1. (Figure 4), the cloud layer causes an underestimation of the vegetation index, but the cloud shadow makes the NDVI overestimated. The original intention of the upper envelope function was to reduce the effects of atmospheric interference (Chen et al. 2004), and the upper envelope function needs to be cautious on Sentinel-2 or Landsat data due to possible cloud shadow effects. Due to the low spatial resolution of MODIS and AVHRR, cloud shadows are difficult to distinguish; therefore errors caused by cloud shadows are difficult to quantify. Using 10 m resolution Sentinel-2 satellite data, cloud shadows can be identified, although the quality labels are not completely correct but still can help identifying noise to a certain extent.

The development of the new fitting method takes into account the uncertainty of the data quality and the inherent error of the vegetation index time series. Paper I showed that smoothing the vegetation index by the function fitting methods (e.g., double logistic function fitting) was more reliable than other smoothing methods. This conclusion is supported by the findings of Hird and McDermid (2009) and Atkinson et al. (2012). The function in the function fitting can be seen as a shape prior. The expression of the fitted function is designed in advance, and each parameter in the expression is then estimated by the least squares method. This is the reason for the function fitting method being more robust than other smoothing methods. When the data quality is high and the data time interval is very small, a local fitting method (such as SG and LO in paper I) can be used to achieve good results. However, at an evergreen forest site, e.g., Norunda site, the ground vegetation index time-series shows relatively small amplitude and large noise, and the noise fluctuations even exceed the amplitude variation (e.g., PhenoCam GCC in Figure 4 in paper III). In this case, the data smoothing needs to be constrained in a way that determines the shape of the function in advance. This is one of the reasons that we chose function fitting method as the basis of our new method (paper II).

The development and initial testing of the new fitting method are based on data from Norunda, which is located at high latitude with large solar zenith angles and snow cover from November to January and, in consequence, most of or all observations are lost during this time. The loss of observations affects the estimation of the baseline of the vegetation index time-series. A 180-day data gap between two seasons is difficult for interpolation. Therefore, the new method determines the base value in a statistical way (paper II). In the Norunda region between 2016 and 2017,

low-quality data such as clouds, cloud shadows, and snow observed by Sentinel-2 accounted for 84.4% of the overall observations (Table 3 in paper III). There are only 28 high-quality observations available for smoothing in these two years, and the time distribution of these data is uneven. Therefore a priori shape and box constraints are used to further limit the impact of noise or gaps to the final result. In addition, this method is useful for estimating the parameters of common seasonal shapes (e.g., Olsson et al. 2016, Baumann et al. 2017).

4.1.3   Spatial processing

Spatial aggregation was applied to Sentinel-2 EVI2 pixels in paper III and IV from smaller to larger spatial scales to capture the footprint of ground observations. In the case of spatially aggregated satellite remote sensing data, the uncertainties from non-linearity and heterogeneity are easily overlooked (Jones and Vaughan 2010). There are three different scenarios of dealing with land surface reflectance to regional average daily vegetation index: (1) averaging the reflectance for the region – calculating EVI2 – smoothing; (2) calculating each pixel's EVI2 – averaging EVI2 – smoothing; and (3) calculating each pixel's EVI2 – smoothing each pixel's EVI2 time-series – averaging. Since both EVI2 and double logistic are nonlinear functions, the results from these three scenarios are different from each other unless the region is completely homogeneous. The choice of scenario depends on the purpose of the processing.

We used two different approaches in these two papers respectively. Scenario (1) was used in paper III to compare satellite-derived vegetation seasonality and ground multispectral derived vegetation seasonality because the outputs of ground-based multispectral sensor observation are the overall reflectance of the R band and the overall reflectance of the NIR band in the footprint area. In addition, the pixels from PhenoCam images are difficult to match with the corresponding Sentinel-2 pixels, so using the average reflectance for the region of interest is more reasonable for comparing PhenoCam and satellite data. Scenario (3) was used in paper IV for developing empirical linear GPP models because the purpose of the processing was to keep the original spatial resolution so that each pixel's GPP could be estimated.

4.1.4   Ground validation

Ground reference data were used in papers I, III, and IV. The site observation data included ground-based multispectral sensor data, PhenoCam data, and GPP derived from flux towers. These data were used to evaluate the time-series and phenological parameters from the smoothing methods. Ground-based multispectral sensors provide the same or similar wavelength to the bands of satellite sensors (Eklundh et

al. 2011, Jin and Eklundh 2015), so the same vegetation index could be applied in the comparisons. We emphasize in papers I and III that ground-based multispectral sensor data are the most direct reference data in comparison with ground data from PhenoCams and flux towers. PhenoCam is a low-cost and convenient equipment compared to the other techniques. However, the phenological camera usually estimates the plant phenology, while the satellite data estimates the land surface phenology. In paper III, the end of season values estimated by PhenoCams are considerably different from those estimated by the satellite (RMSE: 16.7 – 28.4 days, Table 6 in paper III). We give five explanations that may cause this difference: 1) different observation angles; 2) difficulty in capturing the region of interest of PhenoCam data; 3) different vegetation indices lead to different time-series shape; 4) the double logic function does not well fit the GCC time series; and 5) the amplitude of GCC data is too small in some vegetation types. The flux data derived GPP is a variable that directly expresses the rate of photosynthesis. However, it is sometimes difficult to use for evaluating satellite data because the variations of GPP are affected by not only the leaf chlorophyll contents but also by many environment variables, i.e. temperature, light, and humidity (Monteith 1972). Furthermore, rapid variation in the flux footprint area adds uncertainty to the comparisons (Gelybó 2013).

4.2 Data management

The size of a Sentinel-2 MSI image is large, approximate 600 Mb for 1 tile L1C image. The L2A surface reflectance image is even larger, over 900 Mb for 1 tile. To store all the daily EVI2 time-series outputs from the processing requires extremely large space. To facilitate data storage and data transfer, we have implemented a method in the new TIMESAT software whereby only five function parameters, i.e., the base level and the parameters of the double logistic functions, are stored, instead of 365 layers for one year's data. This improvement saves a lot of resource in data storage and data transfer. The process of reconstructing the time-series from these parameters is also efficient, and the user can choose to output values corresponding to any date or at any given time step.

4.3 Data application

The application of data in Figure 2 is divided into two parts: scientific applications and societal applications. These two part are closely related to each other since many of these technologies and data are common. While the thesis focuses only phenology and carbon cycling, the importance and usefulness of the methodology in generating data for a large number of societal and other science applications is recognized.

4.3.1 Monitoring of vegetation phenology

The concept of vegetation phenology was used in paper I, II, and III. In fact, vegetation phenology here all belongs to the concept of land surface phenology (LSP, de Beurs and Henebry 2004). Helman (2018) raised three questions about monitoring land surface phenology by time-series smoothing method using satellite remote sensing data: (1) What do we really detect? (2) Could methods impact the results? (3) Are observed phenology changes caused by climate or the used method? These questions can help us to thoroughly evaluate whether the new method is competent for vegetation phenology estimates.

The difference between LSP and plant phenology (PP) is that PP describes the phenology of plants, while LSP describes the overall phenology of satellite pixels. The satellite is located above the vegetation, so also non-vegetation, e.g., soil or water bodies, is inevitably observed. Furthermore, the seasonal profile of different plant communities become mixed when coarse-resolution data are used. The ground multispectral observations in this thesis focus particular vegetation types, thus they show vegetation phenology. The viewing angles of PhenoCams are usually close to parallel to the ground. For example in a forest, the ROI is usually defined to include mainly the canopy, and this prevents much of the ground from being viewed. In this way, PhenoCam and satellite phenology do not represent the same target areas.

The selection of a specific smoothing method indeed impacts the results (paper I). However, the final result depends on not only the smoothing method but also the selection of atmospheric correction and the chosen vegetation index. In addition to the time-series smoothing method, there are other aspects of the estimation of the phenological parameters, e.g. using thresholds (Piao et al. 2006, Philippon et al. 2007) or derivatives (Tateishi and Ebata 2004) of the seasonal curve. de Beurs and Henebry (2010) mentioned the number of data points and the temporal resolutions as general limitations to the parameter estimations. The shortcomings mentioned above can be compensated to some extent by shape prior and box constrained method. Helman (2018) suggested to use several methods in the analysis and then present a range in phenological changes. However, this neglects the fact that certain methods may be less suitable for certain types of data or in specific geographical areas.

Another important factor is the ability of the vegetation index to separate between canopy reflection and reflection from the ground. Use of efficient vegetation indices may radically improve the ability to model phenology. For example, the recently developed plant phenology index (PPI, Jin and Eklundh 2014) was developed from radiative transfer theory to be linear with green leaf area index and to reduce the

influence of background colour variations. This index has shown to be efficient for regional phenology mapping based on MODIS NBAR data (Jin et al. 2017, Karkauskaite et al. 2017). However, the degree to which this index is sensitive to viewing geometry variations in Sentinel-2 data has not yet been evaluated, which precluded its use in this thesis.

The results in this thesis support the observation that different fitting methodology may generate different results, but that there is strong sensitivity to variations in the data. With sparse data from Sentinel-2 and Landsat, it is important that the chosen method is robust enough to handle data gaps and irregularity in the data.

4.3.2 GPP estimation

The empirical linear regression GPP model used in paper IV is similar to that used in the study by Schubert et al. (2012), and the research areas are all located in northern Europe. The obtained results are also similar between our study (RMSE: $0.63 – 2.69$ g C m^{-2} d^{-1}) and the study by Schubert et al. (2012) (RMSE: $0.63 – 2.38$ g C m^{-2} d^{-1}). The similarity of the results indicates that the choice of satellite data may have limited impact on the accuracy of GPP estimates. Robinson et al. (2018) showed that by using a LUE model Landsat 30 m improved the accuracy of GPP estimation in comparison with MODIS 250 m, however, MODIS 250 m data performed better at deciduous forest, shrubland, and grassland.

Results from paper IV suggest that the improvement in using high-resolution satellite data from Sentinel-2 may be marginal for GPP estimation. However, this conclusion is strongly dependent on the choice of flux tower GPP as the modelling target. These data aggregate fluxes over a rather large footprint area and are not able to resolve the variation observed in the high-resolution Sentinel-2 data. Footprint models (Kljun et al. 2015) may enable separation of the flux components for better matching with the satellite data (Gelybó et al. 2013). Thus, it cannot be precluded that the high resolution of Sentinel-2 will give us a better understanding of the spatial variations of CO_2 fluxes.

40

5 Conclusions

This thesis presents four papers that demonstrate two major findings: (1) the progress of developing a new time-series analysis method, box constrained separable least squares fitting to logistic model functions with shape prior for reconstructing continuous Sentinel-2 vegetation index time-series; (2) and the abilities of Sentinel-2 with the new method to estimate vegetation phenology and GPP. The thesis contributes to Sentinel-2 data processing, data management, and data application of vegetation seasonality in the context of vegetation observation in the big data era.

Before developing the new method, the performance of existing time-series analysis methods in reconstructing NDVI time-series and extracting land surface phenology from MODIS data was investigated. Comparing with the ground observation data, it was found that these methods performed very well after precisely adjusting the parameters, and the function fitting methods (e.g., fits to double logistic function) were safer and more robust than other methods when ground calibration was not available. Based on this finding, we used double logistic function as the base model for the new method, and added box constraints and shape prior operations to improve the robustness of the new method in processing irregular time step data, e.g., Landsat and Sentinel-2. This new method improves the processing of Sentinel-2 from level-2 surface reflectance to level-3 vegetation seasonality products.

Ground reference data are necessary for evaluating the performance of Sentinel-2 and the new method in estimating vegetation phenology and GPP. The EVI2 time-series generated by the new method satisfactorily matched the ground measured EVI2. Sentinel-2 with its higher spatial resolution is, therefore, able to better capture footprint areas of in-situ measurement and better present spatial variations in comparison to MODIS. PhenoCams are more economical than multispectral sensors but are difficult to match to satellite observations. Both Sentinel-2 EVI2 and MODIS EVI2 processed by the new method can fairly accurately simulate daily GPP. While aggregated results do not differ much, Sentinel-2 data enables the spatial variation of GPP in the footprint area to be presented.

The processing scheme developed for Sentinel-2 time-series data in this thesis is general in nature and can thus be extended to other biomes or climate regions.

6 Outlook

In the past year, 2018, the Sentinel-2 satellites provided approximately 1.25 PB of global terrestrial imagery data (Soille et al. 2018). Despite the advances in automatic and efficient extraction of the seasonality of land vegetation from Sentinel-2 MSI data, the new fitting method in this thesis still needs further development. Important room for improvement is to enable processing of complex seasonal changes, e.g., shifting cultivation; disturbance, e.g., insect defoliation (Olsson et al. 2012); and vegetation trends (Jamali et al. 2014). DBEST by Jamali et al. (2015) and Kalman filter with seasonal prior by Olsson et al. (2016) raise the possibility to improve the method developed in this thesis or integrate these methods together for analysis of trends and disturbances. In addition, there is no mature technology to robustly detect strongly varying annual seasonality from Sentinel-2 data. I am particularly interested in developing a season scanner, which is able to automatically detect the positions of irregularly placed season, such as in rotation cropland areas, from Sentinel-2 or Landsat time series data.

Having the red edge bands is a main difference between Sentinel-2 MSI and other satellite remote sensing sensors. The red edge bands have been shown to be valuable for assessment of vegetation chlorophyll status and early stress detection (Horler et al. 2007). This is a valuable research direction, especially for the study of vegetation changes under extreme drought.

Ground reference data are important for assessing the capability of satellite data and fitting methods. I plan to test the new method in other environments, such as temperate region and tropic; therefore ground reference data will be indispensable.

Acknowledgement

Reference

Atkinson, P. M., C. Jeganathan, J. Dash & C. Atzberger (2012) Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *Remote Sensing of Environment,* 123**,** 400-417.

Atzberger, C. & P. H. Eilers (2011) A time series for monitoring vegetation activity and phenology at 10-daily time steps covering large parts of South America. *International Journal of Digital Earth,* 4**,** 365-386.

Baumann, M., M. Ozdogan, A. D. Richardson & V. C. Radeloff (2017) Phenology from Landsat when data is scarce: Using MODIS and Dynamic Time-Warping to combine multi-year Landsat imagery to derive annual phenology curves. *International Journal of Applied Earth Observation and Geoinformation,* 54**,** 72-83.

Beck, P. S., C. Atzberger, K. A. Høgda, B. Johansen & A. K. Skidmore (2006) Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. *Remote sensing of Environment,* 100**,** 321-334.

Bonan G (2015) Ecological Climatology: Concepts and Applications. 3 edn. Cambridge University Press, Cambridge. doi:10.1017/cbo9781107339200

Bradley, B. A., R. W. Jacob, J. F. Hermance & J. F. Mustard (2007) A curve fitting procedure to derive inter-annual phenologies from time series of noisy satellite NDVI data. *Remote Sensing of Environment,* 106**,** 137-145.

Chen, J., P. Jönsson, M. Tamura, Z. Gu, B. Matsushita & L. Eklundh (2004) A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote sensing of Environment,* 91**,** 332-344.

Claverie, M., J. Ju, J. G. Masek, J. L. Dungan, E. F. Vermote, J.-C. Roger, S. V. Skakun & C. Justice (2018) The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sensing of Environment,* 219**,** 145-161.

Cleveland, R. B., W. S. Cleveland, J. E. McRae & I. Terpenning (1990) STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics,* 6**,** 3-73.

Cleveland, W. S. & S. J. Devlin (1988) Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association,* 83**,** 596-610.

Coleman, T. F. & Y. Li (1996) An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on optimization,* 6**,** 418-445.

Craven, P. & G. Wahba (1979) Smoothing noisy data with spline functions. *Numerische Mathematik,* 31**,** 377-403.

Curran, P. J. (1989) Remote sensing of foliar chemistry. *Remote Sensing of Environment,* 30**,** 271-278.

de Beurs, K. M. & G. M. Henebry (2004) Land surface phenology, climatic variation, and institutional change: Analyzing agricultural land cover change in Kazakhstan. *Remote Sensing of Environment,* 89**,** 497-509.

de Beurs, K. M. & G. M. Henebry (2005) Land surface phenology and temperature variation in the International Geosphere–Biosphere Program high-latitude transects. *Global Change Biology,* 11**,** 779-790.

de Beurs, K. M. & G. M. Henebry (2010) Spatio-Temporal Statistical Methods for Modelling Land Surface Phenology. In *Phenological Research: Methods for Environmental and Climate Change Analysis,* eds. I. L. Hudson & M. R. Keatley, 177-208. Dordrecht: Springer Netherlands.

Delbart, N., T. Le Toan, L. Kergoat & V. Fedotova (2006) Remote sensing of spring phenology in boreal regions: A free of snow-effect method using NOAA-AVHRR and SPOT-VGT data (1982–2004). *Remote Sensing of Environment,* 101**,** 52-62.

Drusch, M., U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese & P. Bargellini (2012) Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment,* 120**,** 25-36.

Eklundh, L., H. Jin, P. Schubert, R. Guzinski & M. Heliasz (2011) An optical sensor network for vegetation phenology monitoring and satellite data calibration. *Sensors,* 11**,** 7678-7709.

Eklundh, L. & P. Jönsson (2017) Timesat 3.3 with seasonal trend decomposition and parallel processing - software manual. 88. Lund University, Lund, Sweden.

Farquhar, G. v., S. v. von Caemmerer & J. Berry (1980) A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species. *Planta,* 149**,** 78-90.

Fensholt, R., I. Sandholt & M. S. Rasmussen (2004) Evaluation of MODIS LAI, fAPAR and the relation between fAPAR and NDVI in a semi-arid environment using in situ measurements. *Remote Sensing of Environment,* 91**,** 490-507.

Fischer, A. (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**,** 220-230.

Fisher, J., J. Mustard & M. Vadeboncoeur (2006) Green leaf phenology at Landsat resolution: Scaling from the field to the satellite. *Remote Sensing of Environment,* 100**,** 265-279.

Fisher, J. I., A. D. Richardson & J. F. Mustard (2007) Phenology model from surface meteorology does not capture satellite‐based greenup estimations. *Global Change Biology,* 13**,** 707-721.

Galford, G. L., J. F. Mustard, J. Melillo, A. Gendrin, C. C. Cerri & C. E. Cerri (2008) Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil. *Remote Sensing of Environment,* 112**,** 576-587.

Gelybó, G., Z. Barcza, A. Kern & N. Kljun (2013) Effect of spatial heterogeneity on the validation of remote sensing based GPP estimations. *Agricultural and forest meteorology,* 174**,** 43-53.

Gamon, J. A., A. F. Rahman, J. L. Dungan, M. Schildhauer & K. F. Huemmrich (2006) Spectral Network (SpecNet)—What is it and why do we need it? *Remote Sensing of Environment,* 103**,** 227-235.

Geng, L., M. Ma, X. Wang, W. Yu, S. Jia & H. Wang (2014) Comparison of eight techniques for reconstructing multi-satellite sensor time-series NDVI data sets in the Heihe river basin, China. *Remote Sensing,* 6**,** 2024-2049.

Gitelson, A. A., Y. Peng & K. F. Huemmrich (2014) Relationship between fraction of radiation absorbed by photosynthesizing maize and soybean canopies and NDVI from remotely sensed data taken at close range and from MODIS 250m resolution data. *Remote Sensing of Environment,* 147**,** 108-120.

Gitelson, A. A., A. Viña, V. Ciganda, D. C. Rundquist & T. J. Arkebauer (2005) Remote estimation of canopy chlorophyll content in crops. *Geophysical Research Letters,* 32.

Goward, S. N., B. Markham, D. G. Dye, W. Dulaney & J. Yang (1991) Normalized difference vegetation index measurements from the Advanced Very High Resolution Radiometer. *Remote sensing of environment,* 35**,** 257-277.

Helman, D. (2018) Land surface phenology: What do we really 'see' from space? *Science of The Total Environment,* 618**,** 665-673.

Hermance, J. F., R. W. Jacob, B. A. Bradley & J. F. Mustard (2007) Extracting Phenological Signals From Multiyear AVHRR NDVI Time Series: Framework for Applying High-Order Annual Splines With Roughness Damping. *IEEE Transactions on Geoscience and Remote Sensing,* 45**,** 3264-3276.

Heumann, B. W., J. W. Seaquist, L. Eklundh & P. Jönsson (2007) AVHRR derived phenological change in the Sahel and Soudan, Africa, 1982–2005. *Remote Sensing of Environment,* 108**,** 385-392.

Hird, J. N. & G. J. McDermid (2009) Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sensing of Environment,* 113**,** 248-258.

Horler, D. N. H., M. Dockray & J. Barber (1983) The red edge of plant leaf reflectance. *International Journal of Remote Sensing,* 4**,** 273-288.

Huete, A., K. Didan, T. Miura, E. P. Rodriguez, X. Gao & L. G. Ferreira (2002) Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote sensing of environment,* 83**,** 195-213.

Huete, A. R. (1988) A soil-adjusted vegetation index (SAVI). *Remote sensing of environment,* 25**,** 295-309.

ICOS Sweden. 2018. http://www.icos-sweden.se/.

Immitzer, M., F. Vuolo & C. Atzberger (2016) First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sensing,* 8**,** 166.

IPCC. 2014. Climatee Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Asepects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, United Kingdom and New York, NY, USA.

Jackson, T. J., D. Chen, M. Cosh, F. Li, M. Anderson, C. Walthall, P. Doriaswamy & E. R. Hunt (2004) Vegetation water content mapping using Landsat data derived normalized difference water index for corn and soybeans. *Remote Sensing of Environment,* 92**,** 475-482.

Jamali, S., J. Seaquist, L. Eklundh & J. Ardö (2014) Automated mapping of vegetation trends with polynomials using NDVI imagery over the Sahel. *Remote Sensing of Environment,* 141**,** 79-89.

Jamali, S., P. Jönsson, L. Eklundh, J. Ardö & J. Seaquist (2015) Detecting changes in vegetation trends using time series segmentation. *Remote Sensing of Environment,* 156**,** 182-195.

Jiang, Z., A. R. Huete, K. Didan & T. Miura (2008) Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment,* 112**,** 3833-3845.

Jin, H. & L. Eklundh (2014) A physically based vegetation index for improved monitoring of plant phenology. *Remote Sensing of Environment,* 152**,** 512-525.

Jin, H. & L. Eklundh (2015) In Situ Calibration of Light Sensors for Long-Term Monitoring of Vegetation. *IEEE Transactions on Geoscience and Remote Sensing,* 53**,** 3405-3416.

Jones, H. G. & R. A. Vaughan (2010) *Remote Sensing of Vegetation: Principles, Techniques, and Applications*. Oxford university press.

Jönsson, P. & L. Eklundh (2002) Seasonality extraction by function fitting to time-series of satellite sensor data. *Geoscience and Remote Sensing, IEEE Transactions on,* 40**,** 1824-1832.

Jönsson, P. & L. Eklundh (2004) TIMESAT—a program for analyzing time-series of satellite sensor data. *Computers & Geosciences,* 30**,** 833-845.

Karkauskaite, P., Tagesson, T., & Fensholt, R. (2017). Evaluation of the Plant Phenology Index (PPI), NDVI and EVI for start-of-season trend analysis of the Northern Hemisphere boreal zone. *Remote Sensing*, *9*(5), 485.

Kljun, N., P. Calanca, M. Rotach & H. P. Schmid (2015) A simple two-dimensional parameterisation for Flux Footprint Prediction (FFP). *Geoscientific Model Development,* 8**,** 3695-3713.

Lange, M., B. Dechant, C. Rebmann, M. Vohland, M. Cuntz & D. Doktor (2017) Validating MODIS and Sentinel-2 NDVI Products at a Temperate Deciduous Forest Site Using Two Independent Ground-Based Sensors. *Sensors,* 17**,** 1855.

Liang, S. (2005) *Quantitative remote sensing of land surfaces*. John Wiley & Sons.

Lieth, H. (1974) *Phenology and seasonality modeling*. Springer Science & Business Media.

Lioubimtseva, E. & G. M. Henebry (2009) Climate and environmental change in arid Central Asia: Impacts, vulnerability, and adaptations. *Journal of Arid Environments,* 73**,** 963-977.

Louis, J., V. Debaecker, B. Pflug, M. Main-Korn, J. Bieniarz, U. Mueller-Wilm, E. Cadau & F. Gascon (2016) Sentinel-2 Sen2Cor: L2A Processor for Users. In *Living Planet Symposium*, 91.

Masek, J. G., E. F. Vermote, N. E. Saleous, R. Wolfe, F. G. Hall, K. F. Huemmrich, G. Feng, J. Kutler & L. Teng-Kui (2006) A Landsat surface reflectance dataset for North America, 1990-2000. *IEEE Geoscience and Remote Sensing Letters,* 3**,** 68-72.

McCallum, I., O. Franklin, E. Moltchanova, L. Merbold, C. Schmullius, A. Shvidenko, D. Schepaschenko & S. Fritz (2013) Improved light and temperature responses for light-use-efficiency-based GPP models. *Biogeosciences,* 10**,** 6577-6590.

Melaas, E. K., M. A. Friedl & Z. Zhu (2013) Detecting interannual variation in deciduous broadleaf forest phenology using Landsat TM/ETM+ data. *Remote Sensing of Environment,* 132**,** 176-185.

Melaas, E. K., D. Sulla-Menashe, J. M. Gray, T. A. Black, T. H. Morin, A. D. Richardson & M. A. Friedl (2016) Multisite analysis of land surface phenology in North American temperate and boreal deciduous forests from Landsat. *Remote Sensing of Environment,* 186**,** 452-464.

Menenti, M., S. Azzali, W. Verhoef & R. Van Swol (1993) Mapping agroecological zones and time lag in vegetation growth by means of Fourier analysis of time series of NDVI images. *Advances in Space Research,* 13**,** 233-237.

Monteith, J. L. (1972) Solar Radiation and Productivity in Tropical Ecosystems. *Journal of Applied Ecology,* 9**,** 747-766.

Monteith John, L., C. J. Moss, W. Cooke George, W. Pirie Norman & H. Bell George Douglas (1977) Climate and the efficiency of crop production in Britain. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences,* 281**,** 277-294.

Olofsson, P.; Lagergren, F.; Lindroth, A.; Lindström, J.; Klemedtsson, L.; Kutsch, W.; Eklundh, L. (2008) Towards operational remote sensing of forest carbon balance across northern europe. *Biogeosciences*, 5, 817-832.

Olsson, L. & L. Eklundh (1994) Fourier series for analysis of temporal sequences of satellite sensor imagery. *International Journal of Remote Sensing,* 15**,** 3735-3741.

Olsson, P. O., A. M. Jonsson & L. Eklundh (2012) A new invasive insect in Sweden - Physokermes inopinatus: Tracing forest damage with satellite based remote sensing. *Forest Ecology and Management,* 285**,** 29-37.

Olsson, P.-O., J. Lindström & L. Eklundh (2016) Near real-time monitoring of insect induced defoliation in subalpine birch forests with MODIS derived NDVI. *Remote Sensing of Environment,* 181**,** 42-53.

Persson, M., E. Lindberg & H. Reese (2018) Tree Species Classification with Multi-Temporal Sentinel-2 Data. *Remote Sensing,* 10**,** 1794.

Petach, A. R., M. Toomey, D. M. Aubrecht & A. D. Richardson (2014) Monitoring vegetation phenology using an infrared-enabled security camera. *Agricultural and Forest Meteorology,* 195-196**,** 143-151.

Philippon, N., L. Jarlan, N. Martiny, P. Camberlin & E. Mougin (2007) Characterization of the Interannual and Intraseasonal Variability of West African Vegetation between 1982 and 2002 by Means of NOAA AVHRR NDVI Data. *Journal of Climate,* 20**,** 1202-1218.

Piao, S., J. Fang, L. Zhou, P. Ciais & B. Zhu (2006) Variations in satellite-derived phenology in China's temperate vegetation. *Global Change Biology,* 12**,** 672-685.

Prince, S. (1991) A model of regional primary production for use with coarse resolution satellite data. *International Journal of Remote Sensing,* 12**,** 1313-1330.

Reed, B. C., J. F. Brown, D. VanderZee, T. R. Loveland, J. W. Merchant & D. O. Ohlen (1994) Measuring phenological variability from satellite imagery. *Journal of vegetation science,* 5**,** 703-714.

Reichstein, M., E. Falge, D. Baldocchi, D. Papale, M. Aubinet, P. Berbigier, C. Bernhofer, N. Buchmann, T. Gilmanov, A. Granier, T. Grünwald, K. Havránková, H. Ilvesniemi, D. Janous, A. Knohl, T. Laurila, A. Lohila, D. Loustau, G. Matteucci, T. Meyers, F. Miglietta, J.-M. Ourcival, J. Pumpanen, S. Rambal, E. Rotenberg, M. Sanz, J. Tenhunen, G. Seufert, F. Vaccari, T. Vesala, D. Yakir & R. Valentini (2005) On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biology,* 11**,** 1424-1439.

Richardson, A. D., K. Hufkens, T. Milliman, D. M. Aubrecht, M. Chen, J. M. Gray, M. R. Johnston, T. F. Keenan, S. T. Klosterman & M. Kosmala (2018a) Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery. *Scientific data,* 5**,** 180028.

Richardson, A. D., K. Hufkens, T. Milliman & S. Frolking (2018b) Intercomparison of phenological transition dates derived from the PhenoCam Dataset V1. 0 and MODIS satellite remote sensing. *Scientific reports,* 8**,** 5679.

Richardson, A. D., J. P. Jenkins, B. H. Braswell, D. Y. Hollinger, S. V. Ollinger & M.-L. Smith (2007) Use of digital webcam images to track spring green-up in a deciduous broadleaf forest. *Oecologia,* 152**,** 323-334.

Rocha, A. V., D. L. Potts & M. L. Goulden (2008) Standing litter as a driver of interannual $CO2$ exchange variability in a freshwater marsh. *Journal of Geophysical Research: Biogeosciences,* 113.

Rocha, A. V. & G. R. Shaver (2009) Advantages of a two band EVI calculated from solar and photosynthetically active radiation fluxes. *Agricultural and Forest Meteorology,* 149**,** 1560-1563.

Ruimy, A., B. Saugier & G. Dedieu (1994) Methodology for the estimation of terrestrial net primary production from remotely sensed data. *Journal of Geophysical Research: Atmospheres,* 99**,** 5263-5283.

Running, S. W., R. R. Nemani, F. A. Heinsch, M. Zhao, M. Reeves & H. Hashimoto (2004) A continuous satellite-derived measure of global terrestrial primary production. *Bioscience,* 54**,** 547-560.

Savitzky, A. & M. J. Golay (1964) Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry,* 36**,** 1627-1639.

Schubert, P., L. Eklundh, M. Lund & M. Nilsson (2010) Estimating northern peatland CO_2 exchange from MODIS time series data. *Remote Sensing of Environment,* 114**,** 1178-1189.

Schubert, P., F. Lagergren, M. Aurela, T. Christensen, A. Grelle, M. Heliasz, L. Klemedtsson, A. Lindroth, K. Pilegaard, T. Vesala & L. Eklundh (2012) Modeling GPP in the Nordic forest landscape with MODIS time series data—Comparison with the MODIS GPP product. *Remote Sensing of Environment,* 126**,** 136-147.

Schwartz, M. D. (2003) *Phenology: an integrative environmental science*. Springer.

Sjöström, M., J. Ardö, A. Arneth, N. Boulain, B. Cappelaere, L. Eklundh, A. De Grandcourt, W. Kutsch, L. Merbold & Y. Nouvellon (2011) Exploring the potential

of MODIS EVI for modeling gross primary production across African ecosystems. *Remote Sensing of Environment,* 115**,** 1081-1089.

Slaton, M. R., E. Raymond Hunt Jr. & W. K. Smith (2001) Estimating near-infrared leaf reflectance from leaf structural characteristics. *American Journal of Botany,* 88**,** 278-284.

Smith, B., I. C. Prentice & M. T. Sykes (2001) Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space. *Global Ecology and Biogeography,* 10**,** 621-637.

Soille, P., A. Burger, D. De Marchi, P. Kempeneers, D. Rodriguez, V. Syrris & V. Vasilev (2018) A versatile data-intensive computing platform for information retrieval from big geospatial data. *Future Generation Computer Systems,* 81**,** 30-40.

Sonnentag, O., K. Hufkens, C. Teshera-Sterne, A. M. Young, M. Friedl, B. H. Braswell, T. Milliman, J. O'Keefe & A. D. Richardson (2012) Digital repeat photography for phenological research in forest ecosystems. *Agricultural and Forest Meteorology,* 152**,** 159-177.

Tateishi, R. & M. Ebata (2004) Analysis of phenological change patterns using 1982–2000 Advanced Very High Resolution Radiometer (AVHRR) data. *International Journal of Remote Sensing,* 25**,** 2287-2300.

Tillack, A., A. Clasen, B. Kleinschmit & M. Förster (2014) Estimation of the seasonal leaf area index in an alluvial forest using high-resolution satellite-based vegetation indices. *Remote Sensing of Environment,* 141**,** 52-63.

Tucker, C. J. (1979) Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment,* 8**,** 127-150.

Turner, D. P., W. D. Ritts, W. B. Cohen, S. T. Gower, S. W. Running, M. Zhao, M. H. Costa, A. A. Kirschbaum, J. M. Ham, S. R. Saleska & D. E. Ahl (2006) Evaluation of MODIS NPP and GPP products across multiple biomes. *Remote Sensing of Environment,* 102**,** 282-292.

van Leeuwen, W. J. D. & A. R. Huete (1996) Effects of standing litter on the biophysical interpretation of plant canopies with spectral indices. *Remote Sensing of Environment,* 55**,** 123-138.

Vermote, E. (2015a) MOD09A1 MODIS/Terra Surface Reflectance 8-Day L3 Global 500m SIN Grid V006 [Data set]. NASA EOSDIS LP DAAC. doi: 10.5067/MODIS/MOD09A1.006. https://doi.org/10.5067/modis/mod09a1.006.

Vermote, E. (2015b) MOD09Q1 MODIS/Terra Surface Reflectance 8-Day L3 Global 250m SIN Grid V006 [Data set]. NASA EOSDIS LP DAAC. doi: 10.5067/MODIS/MOD09Q1.006. https://doi.org/10.5067/modis/mod09q1.006.

Vermote, E., C. Justice, M. Claverie & B. Franch (2016) Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment,* 185**,** 46-56.

Vermote, E. & R. Wolfe (2015) MOD09GA MODIS/Terra Surface Reflectance Daily L2G Global 1kmand 500m SIN Grid V006 [Data set]. NASA EOSDIS LP DAAC. doi: 10.5067/MODIS/MOD09GA.006. https://doi.org/10.5067/modis/mod09ga.006.

Vrieling, A., M. Meroni, R. Darvishzadeh, A. K. Skidmore, T. Wang, R. Zurita-Milla, K. Oosterbeek, B. O'Connor & M. Paganini (2018) Vegetation phenology from

Sentinel-2 and field cameras for a Dutch barrier island. *Remote sensing of environment*.

Vrieling, A., A. K. Skidmore, T. Wang, M. Meroni, B. J. Ens, K. Oosterbeek, B. O'Connor, R. Darvishzadeh, M. Heurich & A. Shepherd (2017) Spatially detailed retrievals of spring phenology from single-season high-resolution image time series. *International journal of applied earth observation and geoinformation,* 59**,** 19-30.

Weil, J. & T. Horst (1992) Footprint estimates for atmospheric flux measurements in the convective boundary layer. *Precipitation Scavenging and Atmosphere-Surface Exchange,* 2**,** 717-728.

White, M. A., D. Beurs, M. Kirsten, K. Didan, D. W. Inouye, A. D. Richardson, O. P. Jensen, J. O'keefe, G. Zhang & R. R. Nemani (2009) Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982–2006. *Global Change Biology,* 15**,** 2335-2359.

Woltring, H. J. (1986) A Fortran package for generalized, cross-validatory spline smoothing and differentiation. *Advances in Engineering Software (1978),* 8**,** 104-113.

Wu, C., J. W. Munger, Z. Niu & D. Kuang (2010) Comparison of multiple models for estimating gross primary production using MODIS and eddy covariance data in Harvard Forest. *Remote Sensing of Environment,* 114**,** 2925-2939.

Wutzler, T., A. Lucas-Moffat, M. Migliavacca, J. Knauer, K. Sickel, L. Šigut, O. Menzer & M. Reichstein (2018) Basic and extensible post-processing of eddy covariance flux data with REddyProc. *Biogeosciences,* 15**,** 5015-5030.

Xiao, X., Q. Zhang, B. Braswell, S. Urbanski, S. Boles, S. Wofsy, B. Moore & D. Ojima (2004) Modeling gross primary production of temperate deciduous broadleaf forest using satellite images and climate data. *Remote Sensing of Environment,* 91**,** 256-270.

Xiao, X., Q. Zhang, D. Hollinger, J. Aber & B. Moore III (2005) Modeling gross primary production of an evergreen needleleaf forest using modis and climate data. *Ecological Applications,* 15**,** 954-969.

Zhang, X., M. A. Friedl & C. B. Schaaf (2006) Global vegetation phenology from Moderate Resolution Imaging Spectroradiometer (MODIS): Evaluation of global patterns and comparison with in situ measurements. *Journal of Geophysical Research: Biogeosciences,* 111.

Zhang, X., M. A. Friedl, C. B. Schaaf, A. H. Strahler, J. C. Hodges, F. Gao, B. C. Reed & A. Huete (2003) Monitoring vegetation phenology using MODIS. *Remote sensing of environment,* 84**,** 471-475.

Zhang, X., S. Jayavelu, L. Liu, M. A. Friedl, G. M. Henebry, Y. Liu, C. B. Schaaf, A. D. Richardson & J. Gray (2018) Evaluation of land surface phenology from VIIRS data using time series of PhenoCam imagery. *Agricultural and Forest Meteorology,* 256-257**,** 137-149.

Zhu, Z., S. Wang & C. E. Woodcock (2015) Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sensing of Environment,* 159**,** 269-277.

Zhu, Z. & C. E. Woodcock (2012) Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment,* 118**,** 83-94.

Zhu, Z. & C. E. Woodcock (2014) Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change. *Remote Sensing of Environment,* 152**,** 217-234.

LUND
UNIVERSITY