Charles University

Faculty of Science

Department of Applied Geoinformatics and Cartography
Study programme: Geography

Branch of study: Cartography and geoinformatics

Tatsiana Danilchyk

ASSESSMENT OF VEGETATION PHENOLOGY USING SENTINEL-2 TIME SERIES DATA

HODNOCENÍ FENOLOGIE VEGETACE POMOCÍ ČASOVÝCH ŘAD DAT SENTINEL-2

Diploma thesis

Supervisor: doc. RNDr. Přemysl Štych, Ph.D.

Prague 2022

University: Charles University **Faculty:** Faculty of Science

Department: Applied Geoinformatics and Cartography **Academic year:** 2021/2022

Diploma thesis assignment

for Tatsiana Danilchyk

study programme: Cartography and Geoinformatics

Thesis title: Assessment of vegetation phenology using Sentinel-2 time series data

Principles for the accomplishment

The thesis will deal with the problematics of using vegetation indices to detect temporal changes in vegetation phenology using Sentinel-2 satellite data in period of 2018-2020. There will be studied 6 species of selected perennial vegetation species in different locations: spruce, pine, oak, hornbeam, beech and ash. The aim of this study is to investigate the potential of time series of selected vegetation indices of Sentinel-2 satellite data for detecting different phenological phases of perennial vegetation. An algorithm will be developed in the open access cloud platform Google Earth Engine to generate relevant indices for determining phenological changes. Another objective will be to develop an algorithm in the R environment for the time series analysis of vegetation indices using selected functions to detect individual phenological phases.

From a geoinformation point of view, the work addresses the following tasks:

- selection of appropriate data and methods to find a suitable way to assess phenology using Sentinel-2 time series, including selection of appropriate vegetation indices;
- to investigate methods and create a workflow of the preprocessing steps of Sentinel-2 in cloud-based platform and evaluate an accuracy (e.g. cloud masking)
- to investigate and test appropriately packages for the phenological phases monitoring and create an analytical workflow using programming language R in the RStudio software
- to evaluate the phenology of the selected monitored vegetation species based on phenometric methods and time series of vegetation indices to detect relevant phenological phases (SOS - start of season, EOS - end of season)
- to evaluate a relevancy of selected vegetation indices for detection of relevant phenological phases

- to create a method/algorithm enables to generate the time series of the selected indices and analyze them and to detect SOS and EOS based on time series analysis and phenological tools (packages)
- to validate the acquired results based on Sentinel-2 data by its correlation with in-situ measurements from the CHMI, to assess a relevancy remote sensed data with in-situ data.

Scope of graphic materials: as necessary

Scope of text report: approx. 60–100 pages

References:

BOORI, M. S., CHOUDHARY, K., PARINGER, R., SHARMA, A. K., KUPRIYANOV, A., CORGNE, S. (2019): Monitoring crop phenology using NDVI time series from sentinel 2 satellite data. In: 2019 5th International Conference on Frontiers of Signal Processing, ICFSP 2019.

DE BEURS, K. M., HENEBRY, G. M. (2004): Land surface phenology, climatic variation, and institutional change: Analyzing agricultural land cover change in Kazakhstan. Remote Sensing of Environment, 4, 89.

DE BEURS, K. M., HENEBRY, G. M. (2010): Spatio-temporal statistical methods for modelling land surface phenology. In: Phenological Research: Methods for Environmental and Climate Change Analysis. Springer Netherlands, 177–208.

HAMUNYELA, E., VERBESSELT, J., ROERINK, G., HEROLD, M. (2013): Trends in spring phenology of western European deciduous forests. Remote Sensing, 12, 5.

JENSEN, J. R. (2007): Remote sensing of the environment: An earth resource perspective. Pearson Prentice Hall, Upper Saddle River, NJ.

JÖNSSON, P., EKLUNDH, L. (2004): TIMESAT - A program for analyzing time-series of satellite sensor data. Computers and Geosciences, 8, 30.

VRIELING, A., MERONI, M., DARVISHZADEH, R., SKIDMORE, A. K., WANG, T., ZURITA-MILLA, R., OOSTERBEEK, K., O'CONNOR, B., PAGANINI, M. (2018): Vegetation phenology from Sentinel-2 and field cameras for a Dutch barrier island. Remote Sensing of Environment, 215.

Diploma thesis supervisor: doc. RNDr. Přemysl Štych, Ph.D.

Diploma thesis assignment date: 8. 1. 2020

Diploma thesis submission date: 9. 12. 2021

…… .....……

Diploma thesis supervisor Guarantor of the study programme

In Prague

Prohlášení

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze dne ……

Tatsiana Danilchyk

Acknowledgement

I would like to express my sincere gratitude to my supervisor doc. RNDr. Přemysl Štych, Ph.D., to Josef Laštovička, Ph.D., to Mgr. Jan Svoboda and to Mgr. Daniel Paluba for their support of my diploma thesis, for their valuable advice, comments, and time devoted to the elaboration of this diploma thesis. Also, I would like to thank the specialists from the Global Change Research Centre Academy of Sciences of the Czech Republic and the Czech Hydrometeorological Institute of the Czech Republic, in particular Ing. Lenka Hájková, Ph.D., Ing. Lenka Bartošová, Ph.D., prof. Mgr. Ing. Miroslav Trnka, Ph.D. and Ing. Petra Dížková, for the provided materials, overall recommendations and discussions on the topic. Special deepest thanks go to my parents, for their patience, moral support and encouragement during the preparation of this work and my studies.

Abstract

This work aims to evaluate the detection of phenological phases of vegetation based on phenometric parameters according to archival Sentinel-2 data in the selected areas over the period 2018-2020. The first part of the work describes literature review of the relevant publications, which is followed by the description of the suggested methodology. Then, there are the results with the graphic material and description for each monitored site. In the final part of the work, advantages and disadvantages of the developed algorithm are discussed followed up by suggestions for future research and improvement. The developed algorithm consists of two parts. Masking out cloudy and cloud shadow pixels and generation on the vegetation indices time series is done in the GEE platform. The time series analysis and detection of SOS and EOS as well as statistical analysis are done in the R environment. The study areas of size 20 x 20 m represent different species of perennial vegetation across the Czech Republic. For the assessment of the phenophases detection are selected NDVI, RENDVI, NDRE, NDMI and MCARI. The Asymmetric Gaussian function and Double Logistic function are fitted to the time series of each vegetation season in each tested site, the phenology metrics are derived based on threshold or derivatives methods. The results are validated by the in-situ data provided by the CHMI. NDMI showed the highest accuracy in SOS detection when Asymmetric Gaussian function is applied (5-7 days mean absolute standard deviation). NDVI, RENDVI and NDRE detected SOS accurately with fitted Double Logistic function depending on the monitored species (RENDVI for ash – 1 day, NDVI for beech – 5 days, RENDVI for hornbeam – 8 days). For EOS RENDVI showed the highest reliability for all tested sites, whereas NDMI and MCARI failed to detect senescence.

Key words: time series, phenology, Google Earth Engine, Sentinel-2, phenophase

Abstrakt

Cílem této práce je vyhodnotit detekci fenologických fází vegetace na základě fenometrických parametrů podle archivních dat Sentinel-2 ve vybraných oblastech v období 2018-2020. V první části práce je uveden literární přehled relevantních publikací, na který navazuje popis navržené metodiky. Poté jsou uvedeny výsledky s grafickými materiály a popisem pro jednotlivé sledované lokality. V závěrečné části práce jsou diskutovány výhody a nevýhody vytvořeného algoritmu, na které navazují návrhy na budoucí výzkum a zlepšení. Vyvinutý algoritmus se skládá ze 2 částí. Odmaskování oblačných pixelů a generování na časové řadě vegetačních indexů se provádí v prostředí GEE. Analýza časových řad a detekce SOS a EOS a statistická analýza se provádí v prostředí R. Studované plochy 20 x 20 m reprezentují různé druhy trvalé vegetace na celém území České republiky. Pro hodnocení detekce fenofází jsou zvoleny hodnoty NDVI, RENDVI, NDRE, NDMI a MCARI. Asymetrická Gaussova funkce a Dvojitá logistická funkce jsou aplikovany na časové řady jednotlivých vegetačních období v každé testované lokalitě, fenologické parametry jsou odvozeny na základě prahových hodnot nebo derivací. Výsledky jsou ověřeny na základě in-situ dat poskytnutých ČHMÚ. NDMI vykázal nejvyšší přesnost při detekci SOS při použití Asymetrické Gaussovy funkce (průměrná absolutní směrodatná odchylka 5-7 dní). NDVI, RENDVI a NDRE detekovaly SOS přesně u aplikované Double Logistic funkcí v závislosti na sledovaném druhu (RENDVI pro jasan - 1 den, NDVI pro buk - 5 dní, RENDVI pro habr - 8 dní). Pro EOS vykazovalo RENDVI nejvyšší spolehlivost pro všechny testované lokality, zatímco NDMI a MCARI senescenci nedetekovaly.

Klíčová slova: časové řady, fenologie, Google Earth Engine, Sentinel-2, fenofáze

Table of contents

List of figures and tables

Figures

Tables

List of abbreviations used

AC	Atmospheric correction
AG	Asymmetric Gaussian function
API	Application programming interface
AVHRR	Advanced Very High Resolution Radiometer
BFAST	Breaks For Additive Seasonal and Trend
CIre	Chlorophyll Index Red-Edge
CHMI	Czech Hydrometeorological Institute
CUZK	Český Úřad Zeměměřický a katastrální
DEM	Digital elevation model
DL	Double Logistic function
DOY	Day of year
EOS	End of season
EVI	Enhanced Vegetation Index
GEE	Google Earth Engine
GP	Ground phenology
LAI	Leaf Area Index
LSP	Land Surface Phenology
MCARI	Modified Chlorophyll Absorption in Reflectance Index
MSI	Multispectral Instrument
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	Normalized Difference Vegetation Index
NDMI	Normalized Difference Moisture Index
NDRE	Normalized Difference Red-Edge Index
NDWI	Normalized Difference Water Index
NIR	Near infrared spectral band
RE1	Red-Edge spectral band (704 nm)
RE2	Red-Edge spectral band (740 nm)
RENDVI	Red-Edge Normalized Difference Vegetation Index
S2	Sentinel-2
SCL	Scene Classification
SNR	Signal to Noise Ratio
SOS	Start of season
SWIR	Short-wave Infrared spectral band
ÚHÚL	Ústav Hospodářské Úpravy Lesa
VI	Vegetation index

1 Introduction

The phenology monitoring plays an important role in the ecosystem's dynamics, which are strongly linked to climate changes on local and global scales. Remote sensing data has been a valuable tool for the assessment of land surface changes. Copernicus Earth observation programme provides information on different thematic services, one of which is land monitoring. On the pan-European level in the coordination of the European Environmental Agency there is systematically performed monitoring of biophysical parameters and CORINE Land Cover datasets are provided. The Copernicus Land Monitoring service produces yearly metrics on vegetation phenology and productivity parameters continuously at European extent (Copernicus 2021a).

Land surface phenology is the longstanding study of seasonal patterns in plant phenophases based on the vegetation indices time series analysis derived from remote sensing data (de Beurs, Henebry 2004). Sentinel-2 multispectral instruments provide a better temporal, spectral and spatial resolution compared to the satellites used in the majority of studies on phenology monitoring, phenological phases specifically. One of the main advantages of Sentinel-2 are the additional red-edge and short-wave infrared (SWIR) spectral bands, which can be used to a better potential phenophase detection of plants. Nevertheless, there still has been not much research carried out to state if they are more capable of phenology monitoring than traditional vegetation indices (VI), like NDVI.

This work aims to evaluate the detection of phenological phases of vegetation based on phenometric parameters according to archival Sentinel-2 data in the selected areas. Thanks to the collaboration with the project "Fenologické Fáze", which is curated by the Czech Hydrometeorological Institute and Global Change Research Institute of the Czech Academy of Sciences (Fenologické fáze 2021), the studied methodology was applied on the vegetation species that are monitored at the phenological stations on the long-term basis.

The main objectives to achieve this goal are the following:

1. To create an algorithm in the open access cloud-based platform Google Earth Engine (GEE) to generate relevant indicators for phenology changes detection (e.g., vegetation indices);
2. To preprocess and use open access Sentinel-2 satellite data (missions Sentinel-2A and Sentinel-2B L2A atmospherically corrected data) in the observed period of time from 2018;

3. To investigate methods and create a workflow of the preprocessing steps in cloud-based platform GEE and to evaluate an accuracy of preprocessing (e.g. cloud masking);
4. To investigate and test appropriately packages for the phenological phases monitoring and to create an analytical workflow using programming language R in the RStudio software;
5. To evaluate the phenology of the selected monitored vegetation species based on phenometric methods and time series of vegetation indices to detect relevant phenological phases (SOS - start of season, EOS - end of season);
6. To evaluate a relevancy of selected vegetation indices for detection of relevant phenological phases;
7. To contribute to the creation of the method/algorithm which analyzes Sentinel-2 time series for the purposes of phenology assessment (in the GEE platform and RStudio software) with these functionalities: 1) algorithm computes vegetation indices for any selected point or polygon in a chosen period with thoroughly masked out cloudy and shadow pixels; 2) algorithm enables to generate the time series of the selected indices and analyze them and 3) to detect SOS and EOS based on time series analysis and phenological tools (packages);
8. To validate the acquired results based on Sentinel-2 data by its correlation with in-situ measurements from the CHMI, to assess a relevancy remote sensed data with in-situ data.

This work brings a newly developed open access algorithm that enables one to generate and evaluate vegetation time series curves using Sentinel-2 satellite data for the purpose of detection of individual phenological phases of the selected vegetation species.

2 Literature Research and Problem Statement

2.1 Phenology detection

According to one of the first definitions, plant phenology is a study that deals with the detection of the significant stages in the development of vegetation during the year (Novák 1922). Flowering, leaf unfolding, canopy growth and leaf fall are all examples of the life-cycle biological events of vegetation. In the last few decades, phenology has been recognized as an environmental science with the potential for addressing a variety of ecosystem problems (Schwartz 2003). Prediction of plant life cycle phases has been established by a network of phenological observations since the beginning of systematic phenology monitoring and has been widely used in agricultural and forestry practice (Bergmeier 1998).

The definition of phenology has been modified and expanded over time and on the basis of other studies. According to the definition of Páník (2008), phenology is the study of the time dynamics of periodically recurring signs of vitality, i.e., evaluative stages of plants and animals, and the study of the connections of these so-called phenological phases to climatic and soil conditions during individual years. The publications of the researchers at that time have proven the potential phenology analysis can provide for better understanding of climate change's impacts on ecosystems (Schwartz 2013). The essence of phenology is the monitoring of life phenomena in nature, which occur annually, but in unequal terms, or with different intensity, because they reflect time-varying environmental conditions. Based on the observations, the observer (phenologist) determines the beginning and end of the phenological phases. The phenological phase (phenophase) is a specific biological event that occurs during the annual life cycle of plants, such as germination, sprouting, leaf budburst, flowering, (e.g., spring phenophases), changes in coloring or leaf fall (e.g., autumn phenophases).

Plant phenology is the study of the timing (and quantity) of annually recurring plant growth and reproductive phenomena, as well as the drivers of these events associated with endogenous and exogenous forces. Within a year, the timing and quantity of plant phenological phenomena show distinct seasonality. Regarding plant phenology of ecosystem processes, studies deal with individual plants, populations, and communities. From the perspectives of natural landscape dynamics, the field of study also includes different spatial scales, namely, local, regional, hemispherical, and global scales. Therefore, plant phenology provides a unified measure of the timing and quantity of plant growth and reproductive phenomena at different vegetation units and spatial scales. A central purpose of plant phenological studies is to discover temporal

rhythmicity and spatial patterns within the timing and quantity of plant phenological phenomena, as well as their physical, chemical and biological mechanisms. Basic tasks of plant phenological studies include the following aspects: (1) observing and recording the timing and quantity of plant phenological phenomena visually, as well as with a digital camera, and satellite remote sensing, etc.; (2) examining and correcting visual phenological records, as well as phenological data from photography and satellite remote sensing; (3) revealing spatiotemporal characteristics of the timing and quantity of plant phenological phenomena and their endogenous and exogenous causes; and (4) simulating and predicting spatiotemporal patterns of the timing and quantity of plant phenological phenomena (White et al. 2009).

Already at the beginning of the phenology study, phenological observations were compared and correlated with long-term climate time-series and the interrelations between the parameters were studied as well. The observed phenological data are an expression of the climate of the area. The results of phenological observations of plants and trees provide relevant information about the course and length of their growing season. The developmental phases of all biota occur regularly every year, but always at different dates and with different intensities. This is caused by time-varying environmental conditions, especially the course of weather conditions in the area in individual years. In addition to genetic factors, meteorological factors also contribute to the duration of phenological phases and the variability of their onset. The onset of phenological phases, such as leaf budburst, leaf unfolding or flowering, is usually only possible when the air and soil temperature exceeds a certain critical point characteristic of each phase of the plant life cycle (Larcher 1988).

Monitoring of phenological phases and their evaluation also functions as a bio indicator of climate change. The temperature requirements of plant species for the onset of individual phenological phases are most accurately expressed by the cumulative sum of effective temperatures (Havlíček 1986). With the trend of continuous global warming, the interest in phenological research and applications has been growing (de Beurs, Henebry 2010). Depending on the changes in the weather in the given climatic conditions, the tendencies of climate change can also be assessed according to the changes in the phenological phases. Expected climate change and related negative factors may affect the course of basic plant life cycle events. Due to global warming, there may also be changes in the development of woody plants and herbs. In recent years, phenological observations have been used to determine the interrelations and responses of vegetation to ongoing climate change and its response to significant weather fluctuations in different years (Vitasse et al. 2011).

The onset or duration of phenological phases (they divide the entire vegetation period of plants into several parts) can be used to characterize the weather conditions of individual years. On the contrary, a long series of phenological observations can be used to study climate fluctuations, because the trends of the onset of certain phenological phases must naturally correlate with the trends of air temperature (Vávra 2014).

Many European countries have a long tradition in systematic observations and collection of phenological data. These records are a suitable and significant source for investigating phenology changes at the country scale or for providing reliable validation of satellite phenology data with in-situ measurements.

2.2 Phenology Detection in Remote Sensing

In recent years, satellite observations have been recognized as a powerful tool in determining plant phenology. Satellites equipped with multispectral sensors run on the orbits around the Earth providing data with high spatial and temporal resolution. Whereas land surveys and in-situ measurements would be limited to representing only local conditions, remote sensing allows to detect changes in vegetation properties at different ranges - from local to continental and global levels (Zhang et al. 2012).

Vegetation contains chlorophyll, the presence of which is one of the fundamental conditions for plants to perform photosynthesis. A healthy green leaf intercepts the incident radiant flux directly from the Sun or from scattered light. This incident electromagnetic energy interacts with the pigments, water, and intercellular air spaces within the plant leaf. The total interaction of spectral radiant flux on and within the leaf can be described as the summary of radiant flux reflected from the leaf, the amount of radiant flux absorbed by the leaf and the amount of radiant flux transmitted through the leaf (Jensen 2007).

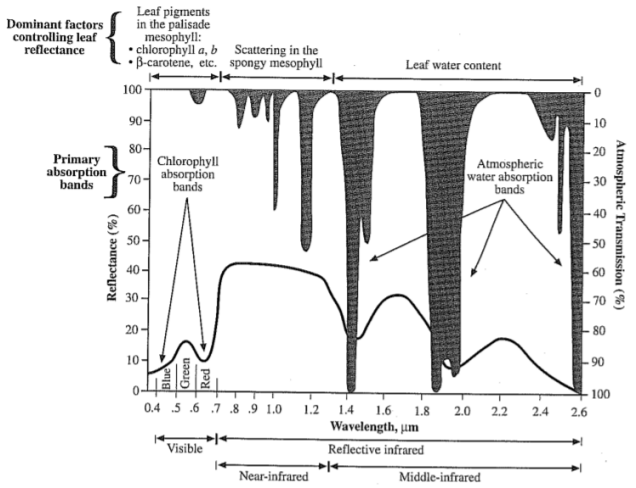


Fig. 1: Spectral reflectance characteristics of healthy green vegetation for the wavelength interval 0.4-2.6 μm
Source: Jensen (2007), p.358

In the near-infrared region, healthy green vegetation is generally characterized by high reflectance (40-60%), high transmittance (40-60%) through the leaf onto the leaves lying below it, and relatively low absorptance (5-10%) (fig. 1) (Jensen 2007). It leads to the process called leaf additive reflectance. It means that a leaf reflects 50% of the incident near-infrared energy back to the atmosphere, while the remaining 50% is transmitted through the leaf onto the leaf below it. Changes in the near-infrared spectral properties of healthy green vegetation may provide information about plant senescence and/or stress. A healthy green leaf exhibited strong chlorophyll absorption in the blue and red wavelengths, a reasonable increase in green reflectance, and approximately 76% reflectance in the near-infrared part of the spectrum. After a certain point, near-infrared reflectance decreased as the leaves senesced. It has been known that there is a direct relationship between response in the near-infrared region and biomass parameters, whereas an inverse relationship has been recorded between the red region and biomass parameters. The distribution of vegetation pixels during growing season (fig. 2) in a scene in the red and near-infrared parts of the wavelength reports low red and low near-infrared reflectance at the beginning of the growing season, but when the complete canopy closure is reached, the reflected near-infrared radiant flux would be high and the red reflectance would be low. This relationship between two parts of the spectrum has prompted to the development of various remote sensing vegetation indices, which can be more highly correlated with biomass than these parameters alone (Jensen 2007).

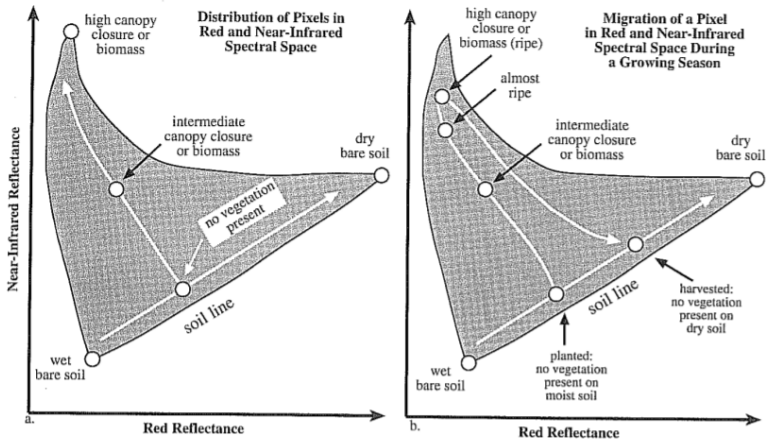


Fig. 2: The distribution of pixels in Red and Near Infrared spectral regions
Source: Jensen (2007), p.364

Chlorophyll absorbs 70 to 90% of incident radiation in the blue and red parts of the spectrum; as a result, the vegetation has low values of reflectance in these parts of the spectrum. There is a local maximum reflectance in the green part of the visible spectrum, where the value reaches up to 10%, it is the reason why vegetation appears green. The spectral performance of vegetation is characterized mainly by a significant increase of reflectance values in the near-infrared part of the spectrum - at a wavelength of 0.7 μm to about 50% for coniferous vegetation and about 55% for deciduous vegetation, and therefore it is possible to distinguish these two types of vegetation in that band (Červená 2016). The reflectance also changes with the seasonal change in the leaf canopy. Phenological data are the parameters obtained by observing the pace of seasonal development and the change of selected plants. The development speed of a plant is determined not only by the inherited characteristics of the species but also by variations of environment conditions. Practically speaking, the speed of plant development changes not only within a vegetation season, when changes in the environment are mainly affected by weather changes, but also over long-term periods depending on fluctuations in climatic conditions. The rate of plant development is usually assessed based on the knowledge of the start time of phenological phases.

The methodology process of the thesis aims to identify important phenological stages, which can be accurately identified using remote sensing data for each selected plant species. The methodology of CHMI defines the following phenophases: leaf budburst, leaf unfolding 10%, full leaf unfolding, start of flowering (10%), peak flowering, leaf coloring (senescence) – 10% and 100%, leaf fall 10% and 100%.

Plant phenology obtained by remote sensing is defined as land surface phenology (LSP) (de Beurs, Henebry 2004). It is defined as the study of vegetation phenology from satellite-

derived VI (or satellite-derived biophysical variables), which could be used as proxy for vegetative vigour and amount. LSP uses satellite observations to consistently estimate phenological phases. The research purposes of LSP studies can be classified into three main categories: studies that focus on spatial or spatiotemporal description of LSP (37%); studies that analyze vegetation dynamics and their and how they correspond to climate factors (52%); studies, which use phenometrics for other applications, such as land cover classification and land cover change detection (Caparros-Santiago et al. 2021).

One of the main phenology metrics defined in LSP are start of the growing season (SOS), end of the growing season (EOS), length of the growing season (LOS), peak of the growing season (POS). SOS is defined as the point in time extracted from the ascending phase (pre-maximum phase) of the curve of the vegetation index (or biophysical variable), which could be associated with the timing of spring phenophases. Also, it is conceptually defined as a rapid sustained increase in remotely sensed greenness after the longest annual period of photosynthetic senescence (for SOS definition, the term 'greenup' can also be used). Critically, in addition to radiation absorption by vegetation canopies, LSP includes the aggregate, confounding influence of atmospheric contamination, cloud cover, snow cover, soil wetness, and bidirectional viewing effects, Therefore, LSP must be related, but not identical to plant phenology (White et al. 2009). EOS is defined as the point in time extracted from the descending phase (post-maximum phase) of the curve of the vegetation index (or biophysical variable), which could be related to the timing of autumn phenophases. In addition, it can be referred to in the literature as end of senescence, end of greenness, dormancy or autumn phenology. LOS is the difference between SOS and EOS, which could be associated with the complete temporal extension of the period of vegetative development of plants (Caparros-Santiago et al. 2021). Since GP and LSP have different definitions, an absolute match in terms of specific day of year is unlikely to occur. However, the general behaviour of trends in the SOS from ground and satellite observations can be assumed to be fairly related, since both observe various starting points in the vegetation growth cycle (Misra et al. 2016).

Vegetation indices describe the reflectance of a certain land cover type depending on the wavelength of the incident radiation. Remote sensing of vegetation is mainly performed by obtaining the electromagnetic wave reflectance information from canopies using spectral sensors. It is well known that the reflectance of light spectra from plants changes with plant type, water content within tissues, and other intrinsic factors (Xue, Su 2017). The reflectance characteristics of vegetation to the electromagnetic spectrum are determined by the chemical and morphological characteristics of the surface of leaves. One of the most used and

implemented indices calculated from multispectral information as a normalized ratio between the red and near infrared bands is the Normalized Difference Vegetation Index (NDVI). The more advanced EVI considers the parameter to simultaneously correct soil and atmospheric effects, which includes the values of NIR, red and , which are corrected by the atmosphere, and the soil adjustment parameter. To minimize atmospheric effects is to use the differences in the blue and red reflectance bands as an estimate of the level of atmospheric influence. When the aerosol concentration is higher, the difference in the two bands also increases (Batten 1998).

Modern remote sensing techniques provide a promising alternative and new opportunities for phenological studies (Badeck et al. 2004), a move forward from the traditional ground-based observations of phenology. In comparison to ground phenology (GP), remote sensing techniques provide a global coverage of data at various temporal and spatial scales. Nevertheless, there have been only a few successful attempts to temporarily match satellite observations with ground-based phenological observations. It can be explained by temporal and spatial resolution issues. One of the main problems is the discrepancy between spatially continuous pixel-based phenological information derived from satellite imagery and spatially discontinuous point-based species-specific GP. Another problem scientists are facing is a mismatch between the definitions of LSP and GP itself. GP is the visual interpretation of species-specific phenological phases such as budburst, leafing, flowering, etc., whereas LSP is defined in terms of area averaged intensity of dominant vegetation or canopy greenness and cover, including the ground cover beneath it such as soil, grass (Reed et al. 1994; Fisher et al. 2007). Furthermore, it is important to mention that LSP estimates are influenced by the methods used for corrections, smoothing, phenology detection, and the accuracy or homogeneity of the area of interest. Despite the mentioned limitations, satellite data nevertheless may provide valuable and spatially continuous information about the LSP (Tucker et al. 2001; Han 2013).

Misra et al. (2020) performed a comparative analysis of the existing approaches to detect phenological parameters. He emphasizes how significantly Sentinel-2 missions have contributed to knowledge on vegetation over the last three years. However, despite the additional red-edge and SWIR bands available on the Sentinel-2 multispectral instruments, with improved vegetation species detection capabilities, there has been very little research on their efficiency to track vegetation cover and its phenology.

Most of the existing LSP studies have relied on medium or coarse spatial resolution satellite sensors, such as MODIS, Landsat and AVHRR, or fusion of MODIS and Landsat (Tian et al. 2013; Walker et al. 2014). Newer satellite missions, such as Sentinel-2, RapidEye, and

Planet, provide data at much finer spatial and higher temporal resolutions, although currently over a shorter time series than Landsat and MODIS. With a high geometric and spectral consistency with Landsat, and additional information in the SWIR and red-edge wavelengths, Sentinel-2 provides opportunities to monitor and study plant phenology in much greater detail (Misra et al. 2020; Astola et al. 2019). In fig. 3 one can see the contribution of satellites, VIs and methods to phenology monitoring in the relevant studies.

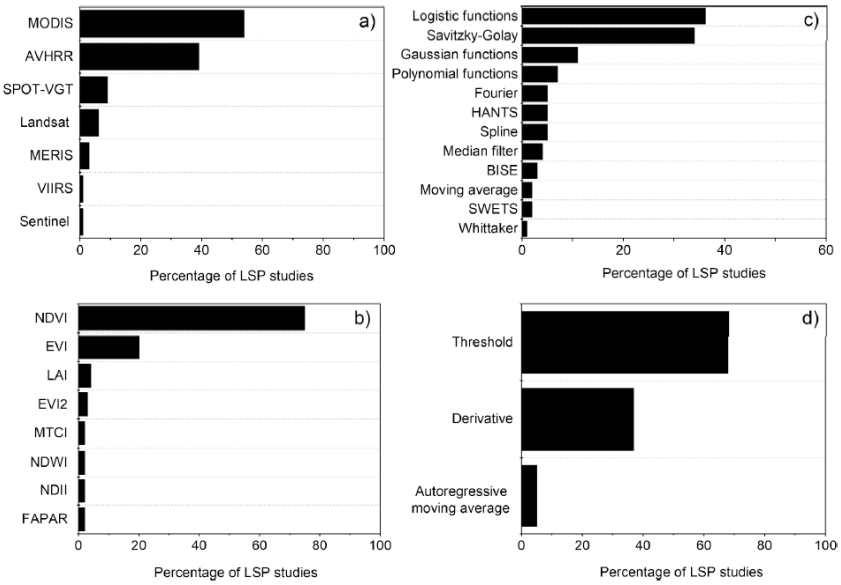


Fig. 3: Percentage of LSP studies using: a) sensors; b) VI or biophysical variables; c) smoothing techniques; d) phenometric extraction methods.
Source: Caparros-Santiago et al. (2021)

2.2.1 Vegetation Indices

To extract the phenological parameters as precisely as possible information from more than one spectral band is often combined to enhance the vegetation properties within a pixel. Vegetation indices are dimensionless, radiometric measures extracted and modeled using remotely sensed data. They indicate the relative abundance and activity of green vegetation, including leaf area index (LAI), percentage green cover, chlorophyll content, green biomass, absorbed photosynthetically active radiation (Huete et al. 2002).

VIs are one of the most extensive and effective variables for vegetation parameters estimation because they are dimensionless indicators that can be used to depict the density, growth or health status of vegetation by using the contrast of spectral reflectance in different wavelengths, especially in the red and near-infrared (NIR) ranges. Among these indices, the NDVI is regarded as the most popular vegetation index for estimating vegetation biophysical characteristics (Sun et al. 2020).

There are different variations of vegetation indices. Apart from the conventional red and NIR bands, the potential of using the red-edge spectral information for retrieval of vegetation parameters has also been proved (Zhu et al. 2017; Xie et al. 2018). The red-edge of a vegetated spectrum is defined as the sharp change in the reflectance curve between 680 and 750 nm. As an important component of the Copernicus program, Sentinel-2 satellites from the ESA provide imagery with three red-edge bands (Sun et al. 2020).

In this work were studied and tested conventional VI using Green, Red, NIR and SWIR bands: Normalized Difference Vegetation Index (NDVI) and Normalized Difference Moisture Index (NDMI). Apart from that, to assess the potential of Sentinel 2 red-edge bands the following indices were tested: Red-Edge Normalized Difference Vegetation Index (RENDVI), Normalized Difference Red-Edge (NDRE) and Modified Chlorophyll Absorption Ratio Index (MCARI). Also, as a part of the research, there were tested Leaf Chlorophyll Index, Chlorophyll Index Green, Chlorophyll Index Red-Edge and Tasseled Cap Wetness, but after practical consideration it was decided to focus on the first five mentioned above indices. The chlorophyll absorption is very high in the visible spectrum, where the reflectance is the highest for green wavelengths. In shifting from the range of visible wavelengths to invisible towards the NIR, the reflectance starts to increase for red edge wavelengths as the chlorophyll absorption decreases significantly (Otsu et al. 2019).

Normalized Difference Vegetation Index

NDVI was initially proposed by Rouse et al. (1974). The NDVI is a dimensionless index that describes the difference between the visible and near-infrared reflectance of vegetation cover and can be used to estimate the density of greenness on an area of land. The NDVI relies on the fact that healthy vegetation reflects a lot of light in the NIR spectrum, in distinctive contrast with most non-plant objects. On the contrary, the vegetation that is dehydrated and unhealthy reflects less light in the NIR, but the same amount in the visible range. Therefore, NDVI can help to highlight vegetation from other land features, and even help differentiate healthy vegetation from unhealthy vegetation (Evangelides et al. 2020). It is calculated as follows (1):

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

NDVI is applied in the assessment of biomass, canopy structure, vegetation stress, calculating LAI, and other plant leaves characteristics (Gitelson 2004; Xue, Su 2017). However, studies based on NDVI revealed that the existing methods for retrieval of LSP might involve

uncertainties. NDVI happened to be less sensitive to the dynamics of photosynthetic activity in densely vegetated areas (Huete et al. 2002). Since the index is calculated through a normalization procedure, the range of NDVI values is [-1, 1], having a sensitive response to green vegetation even for low vegetation covered areas. It is assumed that only positive values correspond to vegetated zones; the higher the index, the greater the chlorophyll content of the plant canopy of interest. However, NDVI is sensitive to the effects of soil brightness, soil color (the NDVI values are particularly high with darker backgrounds), atmosphere, clouds (Huete et al. 2002).

Normalized Difference Moisture Index

The NDMI is a normalized difference moisture index, which is derived from NIR (B8A) and SWIR (B11) bands to display moisture. Although most of the VI used to extract phenological characteristics are based primarily on the Red and NIR regions of the electromagnetic spectrum, which are linked to plant chlorophyll content and leaf structure, some other indices utilize spectral data from the SWIR, which is associated with water content. The SWIR band reflects changes in both the vegetation water content and the spongy mesophyll structure in vegetation canopies, while the NIR reflectance is affected by leaf cell structure and leaf dry matter content but not by water content (Sentinel Hub 2021). The combination of NIR with SWIR (2) eliminates variations caused by leaf internal structure and leaf dry matter content, improving the accuracy in retrieving the vegetation water content. The amount of water available in the leaf structure largely restricts the spectral reflectance in the SWIR wavelengths (Jin et al. 2005). Due to these properties, the NDMI is widely used to detect and track any changes in plant health conditions and its water stress level.

$$NDMI = \frac{NIR - SWIR}{NIR + SWIR} \quad (2)$$

Wilson and Sader (2002) compared NDMI and NDVI to detect forest changes at different Landsat acquisition intervals in Maine, USA. In all classification trials, the NDMI change maps had a higher overall accuracy than the NDVI change maps. They concluded that it was due to an increased ability to detect lighter disturbances including fragmentary deforestation. Laštovička et al. (2020) found that NDMI is useful for the detection of the disturbed forest and forest recovery after bark beetle outbreaks and provides relevant information regarding the health status of the forest. Based on the NDMI characteristics, it can be anticipated that the index might be able to detect changes in leaves properties in the beginning of the growing season, when the water and moisture content is the highest, although the end of the season might

be detected earlier than the actual leaf coloring phase due to the decreasing water content in vegetation in summer months.

Red-Edge Normalized Difference Vegetation Index

The Red Edge Normalized Difference Vegetation Index (RENDVI or NDVI705) is a slight alteration to the traditional NDVI and is adopted for use with high spectral resolution reflectance data which includes Red Edge bands of the electromagnetic spectrum (such as data from Sentinel-2). RENDVI is more affected by chlorophyll content when compared to the NDVI and common applications include precision agriculture, forest monitoring, forest fires and vegetation stress detection (Evangelides et al. 2020). RENDVI using Sentinel-2 imagery can be calculated by using bands RE1 (B5) and RE2 (B6) and it ranges between -1 to +1, where RENDVI decreases when the vegetation stress level increases. RENDVI is calculated from the following formula (3):

$$RENDVI = \frac{RE2 - RE1}{RE2 + RE1} \quad (3)$$

Evangelides and Nobajas (2020) tested RENDVI in assessment of post-fire regeneration. Despite the fact the study was focused on a short-term monitoring period (14 months), the authors claim the results to be precise and successful in detection of revegetation dynamics (L3 Harris Geospatial documentation center 2021).

Normalized Difference Red-Edge Index

The NDRE is sensitive to chlorophyll content in leaves, variability in leaf area, and soil background effects. High values of NDRE represent high levels of leaf chlorophyll content. Soil normally has the lowest values, unhealthy plants have intermediate values, and healthy plants have the highest values (MicaSense 2020). Leaf chlorophyll content is known to accurately indicate photosynthesis activity, mutations, stress and nutritional state of vegetation. The index showed high sensitivity to a wide range of chlorophyll content and is most suitable for the remote estimation of chlorophyll content in higher plants. The reflectance near the RE1 (B5) band is generally more sensitive to chlorophyll content than the reflectance near the green band, where it is affected by other pigments. Wavelengths beyond 900 nm are also not suitable reference bands due to the presence of minor water absorption bands near 970 nm and 1200nm.

NDRE is considered a better marker of plant conditions than NDVI for middle and late season crops that have already accumulated a large amount of chlorophyll. The reason is that red-edge part of the spectrum can pass through the leaves far deeper than red. For this reason,

the index can be considered more relevant than NDVI for intensive use during the entire vegetation season, as NDVI often becomes inaccurate after plants accumulate a maximum amount of chlorophyll content (Earth Observation System 2021). NDRE formula is very similar to NDVI, but instead of Red, the red-edge band is used in the ratio (4):

$$NDRE = \frac{NIR - RE1}{NIR + RE1} \quad (4)$$

The red-edge wavelength provides a sensitive indicator of chlorophyll content in leaves, density of leaves, and soil background effects. Chlorophyll has maximum absorption in the red band wave, and so red light cannot penetrate beyond a few layers of leaf cells. Leaves are more translucent to red-edge light than red light: the red-edge wavelength penetrates a leaf much more deeply than red or blue wavebands. Therefore, NDRE is more suitable for middle and late growth stages, when crops have accumulated high concentrations of chlorophyll in the leaves and red light will penetrate poorly (Boiarskii 2019).

Modified Chlorophyll Absorption in Reflectance Index

MCARI is very sensitive to variations in chlorophyll concentrations and it measures the depth of chlorophyll absorption. MCARI values are not affected by illumination conditions, the background reflectance from soil and other non-photosynthetic materials. Kim et al. (1994) developed the chlorophyll absorption ratio index (CARI) which measures the depth of chlorophyll absorption at 670 nm relative to the green reflectance peak at 550 nm and the reflectance at 700 nm. CARI was designed to reduce the variability of the photosynthetically active radiation due to the presence of different types of non-photosynthetic materials. It uses bands corresponding to the minimum absorption of the photosynthetic pigments, centered at 550 nm and 700 nm, together with the maximum chlorophyll absorption band (670 nm). The modified chlorophyll absorption ratio index (MCARI) which was defined by Daughtry et al. (2000) as following (5):

$$MCARI = (RE1 - RED) - 0.2 * (RE1 - GREEN) * \frac{RE1}{RED} \quad (5)$$

2.2.2 Smoothing and Breakpoint Detection Algorithms

This part of the work focuses on a summary of the available relevant studies and a proposal of the methodology. The approach described in this work is relatively new for the Sentinel-2 satellite data; the existing methodology of extraction of the phenometric parameters and smoothing of the vegetation time series was applied mostly to Landsat and MODIS data.

Zheng et al. (2016) proposed an approach to detect crop phenology using high spatiotemporal resolution data obtained using SPOT5 and MODIS. This approach consists of combining the data of both satellites to compute NDVI with the spatial resolution of SPOT5 and the temporal resolution of MODIS. The algorithm of phenological data extraction is based on the analysis of the NDVI time series. In this work, the authors tried to answer two questions: if the proposed algorithm is suitable for the time series calculation using multiple combinations of SPOT5 and MODIS data; is the generated data suitable for detecting phenological changes of highly fragmented landscapes? The authors state that to generate high spatial and temporal resolution data, various fusion algorithms have been proposed and proven practicable. Gao et al. (2006) developed a Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) algorithm for blending Landsat ETM+ and MODIS data to generate daily surface reflectance data at ETM+ spatial resolutions firstly. The STARFM algorithm was used directly to combine the vegetation indices obtained from SPOT5 and MODIS data, as it has already been proved that the STARFM method performs better when directly fusing VIs rather than when the reflectance is fused and then the VIs are calculated (Tian et al. 2013). The NDVI was chosen because it was validated as an effective indicator for phenological data extraction by many studies (Jönsson, Eklundh 2002). Due to the fact that undesirable phenomena (noise) can affect the accuracy of the results of the extraction of phenological data, it is therefore important to remove these noises. Recently, various algorithms have been designed and implemented to smooth the time series. However, there is no single optimal algorithm so far. In this research, three different filtering methods, asymmetric Gaussian function, double logistic function and Savitzky-Golay filtering in the TIMESAT software, were tested to smooth the time-series NDVI. As a result, it turned out that Savitzky-Golay had better performances for smoothing the NDVI time series. Furthermore, to determine the beginning and the end of phenological phases, a threshold method was used, which assumes that a certain phenological phase begins when the NDVI value exceeds a predefined threshold. In general, threshold settings are usually based on the characteristics of the NDVI curve; however, these properties vary depending on changes in crop type, and different crops have different growing seasons during their vegetation periods (Pan et al. 2015). The authors adopted the method developed by Pan et al. based on the ratio of NDVImin and NDVImax (NDVIratio) over a specific period for crops to determine the thresholds of the SOS and EOS. Here, specific periods were defined as from the beginning to the peak and from the peak to the end of the season to identify the NDVImin for the definitions of the SOS and EOS. Since the NDVImin may be different at the start and end of season, the threshold values are also varied for SOS and EOS. The results show that the STARFM

algorithm is able to combine SPOT5 and MODIS data and the fused datasets have the potential to detect crop phenology at a high spatial resolution. The accuracy of phenology estimation that involved blended data was better than that using MODIS data, indicating the strong feasibility and reliability of the proposed framework for phenology monitoring. This research demonstrates the potential of using high spatial and temporal resolution data by combining the STARFM algorithm to detect phenological phases. In the work, SPOT5 and MODIS data were tested, but the STARFM method can be implemented on other satellite data for accurate detection of phenology with higher spatial resolution.

De Beurs and Henebry (2010) described 12 spatiotemporal statistical methods to determine the beginning and the end of phenological phases using satellite imagery. The extraction of phenometric parameters can be performed using several quantitative methods. These methods can be divided into four groups: thresholds, derivatives, smoothing functions, and fitted models. There are described advantages, potential limitations, and case studies of the use of the methods for each category. The main disadvantage of observing changes in phenological phases using satellite imagery is the complexity of validation of the results by ground observations, which usually measure something quite different (Schwartz 2002; Fisher et al. 2007). As a result, it is often not clear what exactly the LSP values detect. For example, at higher latitudes, the largest increase in NDVI values which are used to identify the beginning and the end of phenological phases, is often due to snow melt (Reed et al. 1994; Delbart et al. 2005). On the other hand, the end of the phenological phase can also be confused when low NDVI values are shown due to a longer cloud cover period. Moreover, due to the ambiguity and diversity of relationships between satellite LSP measurements and actual phenological events, new methods of statistical analysis for determining phenological phases are still being developed.

The simplest and most commonly used method to determine SOS and EOS is to use thresholds. The threshold values are set randomly at a certain level (e.g., a range of values from 0.1 to 0.35). The beginning of the phase is then determined as the day of the year (DOY) when the NDVI crosses the threshold upwards; similarly, EOS is designated as a DOY such that the NDVI crosses the same threshold in downward direction. To determine when the DOY threshold is reached, time series are filtered and clouded pixels are removed. Karlsen et al. (2007) proposed a variation on the extraction of NDVI values. It consists in setting threshold values on the average NDVI curve over a period of several years.

The first category of VI thresholds is based on long-term mean VI. Karlsen et al. (2007) presented the method, where they calculated a 21-year mean value for each pixel only considering pixels with positive values of NDVI. SOS for each year was then considered to be the date when the NDVI value passed the long-term mean value. This threshold was chosen because it showed the highest correlation with the onset of leafing in birch as observed at ground level. EOS was determined by the date when NDVI passed below 70% of the 21-year mean. Peak timing was determined as the date with maximum NDVI.

The second approach determines the SOS and EOS by comparing years among each other (Shabanov et al. 2002). It takes a baseline year (the median year in the time series) and sets the NDVI values for DOY 120 and DOY 270 as the SOS and EOS thresholds for this year, respectively. The DOYs at which these predefined NDVI thresholds were reached in each other year determine the SOS and EOS for that year.

The third threshold-based approach is based on the NDVI value ratio, so-called normalization (White et al. 1997). The SOS is determined as the day when the NDVI ratio exceeds 50% in upward direction, thus EOS is defined as exceeding 50% of the value in downward direction. The advantage of long-term average minimum and maximum NDVI values is that they are usually not strongly influenced by outliers. The disadvantage is that the minimum and maximum NDVI may not be stable over time and could change significantly, for example, due to disturbance processes or other changes in the landscape.

The last threshold-based approach is defined on NDWI assessment (de Beurs, Henebry 2010). Delbart et al. (2005) argued that NDVI is not the optimal index for measuring SOS and EOS in snow-covered areas because the onset of NDVI growth corresponds to the onset of snow melting. Trends in SOS may not be due to the actual beginning of vegetation, but rather to a reduction in the extent of snow cover. The Normalized Difference Water Index (NDWI), which is based on reflections in the NIR and SWIR bands, may be more effective in estimating the SOS for areas where heavy snow cover can be expected.

The next category is the derivatives. Tateishi and Ebata (2004) defined SOS as the time of the largest increase in NDVI. They hypothesized, as White et al. (1997) found that SOS is characterized by the fastest increase in greenery during the growing season. The maximum positive value of the derived function determines the rapid increase of NDVI and is therefore referred to as SOS. Similarly, the minimum negative value is considered EOS. The method is very simple and easily determines the fastest increase and decrease in NDVI. Balzter et al. (2007) introduced the Camelback Phenology algorithm, which uses a moving window

consisting of five time series composites for each pixel. In each window, the slope of NDVI regression against time is calculated. In the second step, a second-order derivative is calculated with the same moving window. SOS is determined as the time point when the second derivative of the regression reaches a local maximum and the slope is positive. EOS is determined when the second order derivative reaches a local maximum and the slope is negative. Unlike the threshold method, the derived function methods do not have statistical errors. Thus, it is very difficult to understand whether the observed changes fall within the natural variability of the data or are the result of a significant change. These methods have problems determining SOS and EOS when the NDVI signal cannot follow a sharp and rapid increase or decrease.

Another category is based on smoothing functions and fitted models. Reed et al. (1994) determined SOS and EOS as the dates that a smoothed time series crosses a curve set from moving average models. Based on 14-day composites, they found that a moving average curve based on nine composite periods produces the best results. The disadvantage of this method is that it does not deal with disturbances or other changes on the land surface. Another method in this category is the Fourier analysis. It approximates complicated curves with a sum of sinusoidal waves at multiple frequencies (de Beurs, Henebry 2010).

The last category for the extraction of phenological characteristics described by de Beurs and Henebry (2010) based on satellite imagery is model fitting. Tucker et al. (2001) updated the method to parameterize NDVI temporal profiles, which are divided at the peak NDVI into two time series and with 4 parameters, which need to be estimated for each part of the curve. Zhang et al. (2004) fitted a logistic model of vegetation growth to the Enhanced Vegetation Index (EVI). Like Tucker et al. (2001), the authors divided the annual EVI curve in two parts, vegetation growth and senescence, and fit the models separately.

Jönsson and Eklundh (2002) developed an approach based on Gaussian local functions, which fits a number of local model functions that are merged in a global function. This merging of multiple local functions increases the flexibility of the fittings and allows the fitted function to follow the complex behavior of the time series that is not possible with a simple Gaussian model.

Geng et al. (2019) aimed to investigate the spatiotemporal characteristics of abrupt vegetation changes and to reveal their potential stimulating factors using MODIS NDVI data with a resolution of 1 km for a period from 2000 to 2017. The Breaks for Additive Season and Trend (BFAST) algorithm was adopted to detect the phenological phase change dynamics using remote sensing data.

Methods for detecting abrupt changes include Breaks for Additive Season and Trend (BFAST) (Verbesselt et al. 2010). BFAST is an effective algorithm to integrate the iterative decomposition of time series into seasonal, trend and remainder components, and using this method we can detect changes within a long time series. BFAST iteratively estimates the time and number of abrupt changes in time series and characterizes the change by magnitude and direction. It can be used to analyze different types of time series (e.g., Landsat, MODIS) and can be applied to other disciplines, such as hydrology or climatology. Detection of the phenological changes is affected by the noise in the time series. The amplitude of changes in time series varies between different types of land cover. The method has shown that it is able to detect the timing of phenological changes within the time series even if sudden disturbances and noises occur. BFAST identifies phenological changes independently of phenometric parameters by using the entire time series. The method analyzes each pixel separately without setting thresholds for detecting changes in the time series. Long-term phenological changes can be detected within NDVI time series of a large range of land surface types (e.g., meadows, forests and deciduous forests) that have a seasonal amplitude greater than the noise level. The results of the study by Geng et al. (2019) indicated that approximately 80% of the vegetation areas experienced at least one abrupt change during the study period. The main factors of these changes were anthropogenic activity and environmental impact.

Hamunyela et al. (2013) compared the phenological data of the Pan-European Phenological Database with SOS estimates derived from MODIS NDVI to assess for trends in the timing of deciduous forest leafing in Western Europe. In addition, he performed a spearman's correlation analysis of the NDVI values with air temperature during each season. The authors calculated the NDVI, removed the outliers by applying quality flags or replacing them with the average value of its two immediate temporal neighbours Then, to smoothen the NDVI time series of each pixel, they applied to applied Savitzky-Golay filter. The smoothed NDVI time series (with 16-days interval) was then interpolated to daily values using the cubic spline method. To extract SOS values the threshold method of NDVI ratio exceeding 50% (White et al. 1997). The time changes of the onset of the phenological phases were evaluated using the trend linear regression analysis. Phenological changes are largely influenced by global warming. In that study, a significant shift of up to 3 weeks to the early start of the growing season was found in ground-based observations and satellite SOS estimates. It also appears that the magnitude and trajectories of shifts in satellite SOS estimates are well comparable to the in-situ observations, what stresses the importance of satellite imagery in monitoring phenology in constantly changing climate conditions.

Vrieling et al. (2018) compared the phenological characteristics obtained from the NDVI Sentinel-2 time series with data obtained from a GCC camera. Phenological parameters from the Sentinel-2 time series can be obtained with a spatial resolution of 10 m, which provides new opportunities for research in the field of landscape phenology. For areas with high fragmentation and diversity of landscape components (e.g., the Schiermonnikoog island is studied in this article), Sentinel-2 allows to analyze vegetation phenology with higher accuracy compared to commonly used medium resolution sensors such as MODIS. A double hyperbolic tangent function, which corresponds to a double logistic function, was used to correct the NDVI time series curve and determine the phenological phases (Beck et al. 2006). Statistical evaluation of the results was performed using the standard deviation, mean square error and correlation coefficient with the camera results. The authors confirm the reliability of Sentinel-2 high-resolution data for the purpose of determining phenological characteristics and state that future research in other vegetation types monitored by phenological cameras will allow an advanced evaluation of Sentinel-2 data as a key tool for studying vegetation phenology.

Jönsson et al. (2018) introduced a new method for detecting changes in vegetation. The method is based on a box-constrained separable least squares fitting to logistic model or other functions (de Beurs, Henebry 2004), as none of the existing methods has been proved to show higher total efficiency than the others, different methods have been found to be optimal under different conditions). First, the authors define the minimum value of NDVI outside the vegetation period - a so-called base level, the common amplitude, the common rise time, the common fall time, and the shifted inflexion points - a shape prior. Then, the model is applied on the NDVI time series data. The authors state that if there is a significant amount of the satellite observations available in a certain period of time, the corresponding local parameters determining the shape of the model function in that period can be independently changed. If there are no observations in the study period, the corresponding local parameters are strictly dependent on the shape prior. The method is flexible enough for a year-to-year analysis, but robust enough when the number of available observations is low. It has been tested with Landsat, Sentinel-2 and MODIS data in a forested area in Sweden, demonstrating the feasibility and potential of the method for efficient modeling of vegetation period curves.

Boori et al. (2019) described the dynamics of changes in maximum, minimum and average NDVI values over a one-year period and define the dependences on weather, temperature, precipitation and snow conditions. The authors conclude that the existing correlation between these conditions and remote sensing phenology mapping can be an efficient, reliable and traceable tool in detecting the dates of tillering, jointing and maturity of crops.

Motohka et al. (2010) suggested a method to detect phenology using the Green-Red Vegetation Index (GRVI). GRVI was evaluated as a phenological indicator based on a time series of observations of spectral reflectance and phenology at several representative ecosystems (deciduous forest, coniferous forest, meadows and rice fields). The results showed a correlation between GRVI values and the seasonal changes of vegetation and ground surface with high temporal resolution. It has been found that GRVI has the following advantages as a phenological indicator: "GRVI = 0" can be an independent threshold for detecting the early greenup phase and the autumn phase of color change, and GRVI can clearly respond to vegetation disturbances and differentiate ecosystem types. Unlike NDVI, where it is difficult to track autumn color change using the NDVI time series. One simple and possible reason for this problem is that NDVI is not sensitive enough to change the color of the leaves from green to yellow or red, because green reflectance is not used in the calculation of NDVI. In addition, due to the balance between changes in the reflectivity of the green and red bands in response to greenup and senescence, the threshold value "GRVI = 0" should also be effective for phenology detection. Using this feature, a simple algorithm for determining the timing of leaf coloration and autumn coloration may be possible.

Zheng et al. (2016) reported another possible approach for the detection of phenological changes of crops. All data were acquired using terrestrial spectrometers. The authors investigated the ability of the Chlorophyll Index Red-Edge (CIre) to detect vegetation. CIre and NDVI were used to represent the dynamics of vegetation changes during the year. A double logistic function was applied to smooth the vegetation curve. The SOS and EOS dates were extracted using the points of maximum increase and minimum decrease of the vegetation curve. The authors claim that CIre detects the initial spring phenological phases with higher accuracy (leaf unfolding, flowering) than NDVI.

Misra et al. (2016) discussed the limitations and mismatches in comparing LSP and GP. In this study, MODIS NDVI time series data were smoothed using two filtering techniques for comparison. First, a weighted Gaussian filter was applied to the raw time series to remove obvious outliers due to high differences to the precedent and subsequent values. Then, the filtered NDVI time series was smoothed again using Gaussian or Double Log function to remove undetected outliers. As it was already mentioned, different LSP phenological phases detection methods have been defined and tested, but still there was no single method determined to be the best to describe the phenology from satellite NDVI data. In this study, the methods used to derive phenological information can be classified into three broad categories, namely thresholds of amplitude, delayed moving average (DMA) and rate of change (derivatives). The

thresholds of amplitude (20%, 50%, 60% and 75%) determine the specific day of the year on which the smoothed NDVI time series crosses the corresponding value of the NDVI amplitude of a given year. The delayed moving average method determines SOS. The derivatives (namely, 1st, 2nd and 3rd derivatives) determine SOS as the date of the maximum increase in the respective NDVI derivative curve. The local maxima of 1st derivative correspond to the maximum rate of increase of the greenup, whereas, the local maxima of 2nd and 3rd derivative corresponds to the beginning of greenup. In particular SOS from the 2nd derivative indicates the timing when the majority of a pixel is turning green, and the 3rd derivative indicates where the change of green up rate is greatest. The Gaussian filter was able to follow the winter troughs better than the Double Log function. It should be mentioned that the different LSP SOS methods showed more variability and differentiation when applied to the Gaussian smoothed series, whereas the Double Log did not differentiate well among the derivatives. The different LSP SOS estimates when compared with species-rich GP dataset revealed that different LSP SOS extraction methods agree better with specific phases of GP, and the choice of data processing or smoothing strongly affects the LSP SOS extracted.

According to the literature research, no final most efficient approach for detecting multi temporal dynamics of phenological changes has not yet been determined. Relevant articles offer diverse approaches to statistical analysis of time series of vegetation indices and determination of breakpoint dates in phenological phases using several quantitative methods. However, most studies focus on MODIS data, Landsat, while the Sentinel data application is yet relatively sparse. In this way, it can be assumed that the use of Sentinel-2 data will allow to explore a new approach for detecting the multi-temporal dynamics of phenological changes and to achieve better results.

2.3 Phenological packages in R

The R programming environment provides a collection of packages explicitly dedicated to spatial and temporal analysis of phenology and detection of phenological phases in time series. Every package has a set of functions, which perform certain computations based on scientific approaches in order to extract phenological metrics.

2.3.1 Package “phenex”

The phenex package provides a collection of functions for spatial and temporal analysis of phenological datasets and satellite observations of vegetation (Lange, Doktor 2017). To process and analyze VI time series the first step is to create an object of NDVI-class by applying

functions *new* or *modelNDVI* to the vector containing raw VI observations with values in the interval [-1, 1]. Not only NDVI, but any vegetation index can be processed by this package as long as raw input values satisfy the condition of the [-1, 1] interval. The calculations are performed for each year separately. The next step determines the correction method for reducing noise in VI time series. The phenex package provides 2 methods: *bise* and *runningAvg*.

bise - Best index slope extraction after Viovy et al. (1992). The algorithm requires a set of parameters. The *slidingperiod* (the default value is 40) is the value, according to which the algorithm is searching forward within daily VI observations over 1 year, decreases are only accepted if no higher value is found within the period. optimized for the area of Central Europe. The parameter has to be modified according to the study area and its climate conditions. The next parameter is *growthFactorThreshold,* which defines a maximum increase threshold with a default to 0.1 (10% increase per day allowed). And the last one is *cycleValues* which determines whether the end of the VI time series is combined with its beginning or not (the default value is true).

runningAvg - a method to reduce noise in time series through running averaging. As input parameters, it takes an object of NDVI-class and the window size of the running averaging algorithm (the default value is 7).

Then, the *modelValues* function is applied to determine which model will be fitted to the corrected VI time series. The available methods in the package are linear interpolation, spline interpolation, double sigmoidal function, Double Logistic function, symmetric or asymmetric Gaussian function, fitted growth model, Fast Fourier transfusion, Savitzky-Golay filter.

The last step is the extraction of phenological metrics from the modelled VI values. The function allows to determine which phase will be extracted: *max* (day of the year with the highest VI value), *maxval* (the highest modelled VI value and its standard deviation), *min* (day of the year (before day with maximum VI value) with the lowest VI value), *minval* (the lowest modelled VI value and its standard deviation), *greenup* (day of the year at which greenup occurs and its standard deviation), *senescence* (day of the year at which senescence occurs and its standard deviation). Also, this function requires to set whether a local (VI values first reach the value of the *threshold* between the lowest and the highest VI value of time series) or a global (VI values are first equal or higher as the value of *threshold*) threshold will be set to extract the certain phase. The number of normal distributed values around the threshold for estimation of standard deviation is also a required parameter.

2.3.2 Package “greenbrown”

The package “greenbrown: Land Surface Phenology and Trend Analysis” has been developed at the Max Planck Institute for Biogeochemistry, Jena, Germany (Forkel, Wutzler 2015). It presents a collection of functions to analyze trends, trend changes and phenology events in gridded time series like from satellite observations or climate model simulations. The package provides access to different methods for 1) trend and breakpoint analysis, 2) time series smoothing and interpolation, and 3) analysis of land surface phenology. Phenological metrics are often derived from VI time series to map the spatial variability of vegetation phenology or to analyze temporal changes in vegetation.

The package provides a function *Phenology* that allows to calculate all phenometric characteristics in one step setting the default parameters in the function. The package performs temporal smoothing and gap filling using different methods: linear interpolation, splines, double logistic functions, season-trend model and fitting methods as provided in the *greenProcess* function of the phenopix package.

2.3.3 Package “phenopix”

The package provides functions to process digital images, depict greenness index trajectories and extract relevant phenological stages (Filippa et al. 2020). It processes the input VI time series without previous image preprocessing. Most of the relevant functions are considered rather internal functions of the function *greenProcess* but might be used separately as well. The extraction of phenometric characteristics in this package is based on applying fitted models. For example, SOS and EOS are defined as the dates that a smoothed time series crosses a curve set from a certain model. The phenopix package contains built-in functions for determining spline and double logistic functions for the time series.

BeckFit - the function estimates parameters of the double logistic equation from Beck et al. (2006) and provides an uncertainty estimation. Parameters are estimated by a call to the function FitDoubleLogBeck from the greenbrown package. Uncertainty is computed by adding noise to the raw data and by estimating again the parameters. Noise is added according to the standard deviation of the residuals (fitted - observed). The procedure is repeated nrep times.

ElmoreFit - this function fits a double logistic curve to observed values using the function according to the approach described in (Elmore et al. 2012). Other specifications of the function are similar to the BeckFit.

PhenoExtract - This function takes fitted data as an input and extracts phenological metrics according to different methods. The methods include 'trs', 'derivatives', 'klosterman', 'gu'. Depending on the chosen method, all thresholds and parameters in the functions are defined accordingly.

PhenoBP - the function extracts breakpoints of the time series. The difference of this approach from other thresholding approaches, like PhenoExtract, is that PhenoBP takes raw data as an input, but fitted data (e.g., after performing the BeckFit function) can also be used.

3 Methodology

A general workflow of this study is presented in fig. 4. The described workflow has been focused on the creation of an application tool to generate time series for arbitrarily selected polygons/points. The selected image collection "Sentinel-2 MSI: Multispectral Instrument, Level-2A" was linked to the collection "S2 - Cloud Probability". The cloud and shadow pixels were thoroughly masked out.

The selected vegetation indices were computed over every image in the corresponding image collection and added to the final image dataset. Then, after the time series is generated, the VI values are adjusted for the export into a CSV table, which is an input for further time series processing and analysis in RStudio. The input data in RStudio is presented as a “data frame” data type, which is used for storing data tables. A data frame is a list of vectors of equal length. The data frame was adjusted to the correct formatting and divided into a few data frames according to the year of the observation.

R offers elaborate sets of functions - packages, for the purpose of any statistical analysis. For phenological data processing, there are certain packages available that allow VI time series analysis (noise reduction, breakpoint detection) based on specific functions. Based on Ulsig et al. (2017), Berra et al. (2021), Sun et al. (2021), Garcia et al. (2019) and the own research on the functionality of phenology packages, the following were selected: phenex, greenbrown and phenopix (apart from that, the phenofit and phenology R packages were tested). Each package has built-in functions to perform noise reduction, curve smoothing (e.g., Double Logistic Function, Gaussian asymmetric function, Savitzky-Golay filter - these functions are claimed to show the best performance in processing VI time series based on the literature research) and the breakpoint detection (SOS and EOS). After that, the results obtained from the satellite images processing were compared and validated by the in-situ phenological phases measurements provided by the CHMI.

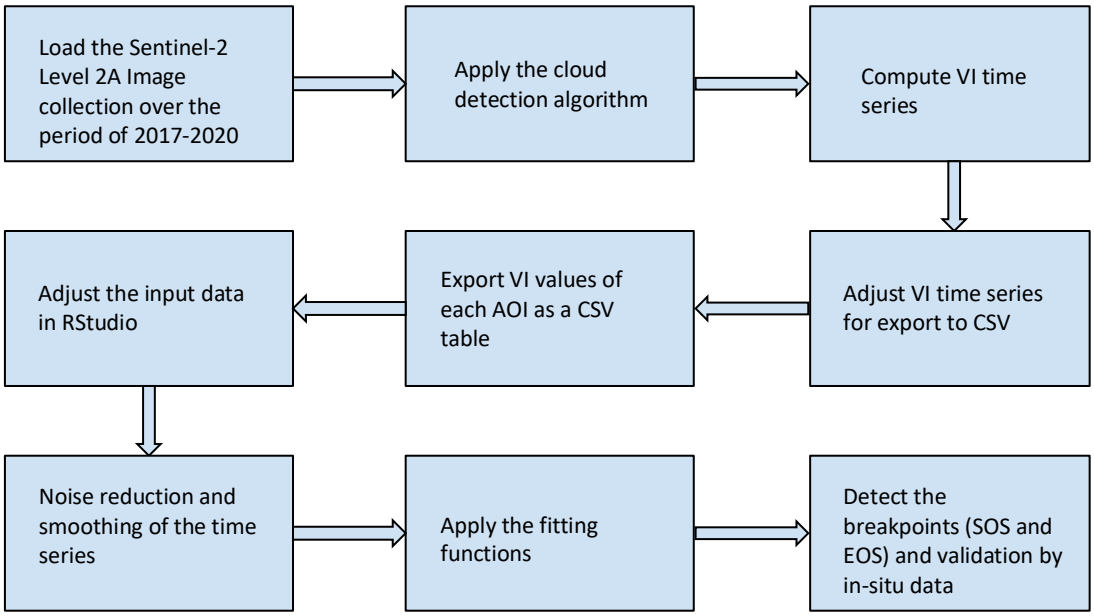


Fig. 4: The methodology workflow
Source: author's work

3.1 Data

3.1.2 Satellite data

In this work, Sentinel-2 Level-2A images were used. The Copernicus Sentinel-2 mission comprises a constellation of two polar-orbiting satellites placed in the same sun-synchronous orbit, phased at 180° to each other with a wide swath of 290 km. Due to that the mission gives a higher revisit time: both satellites temporally complement each other: 10 days at the equator with one satellite and with two satellites all regular areas are revisited every five days under the same viewing conditions (in the case of scene overlap, the temporal coverage for the Czech Republic is less than 3 days, according to Laštovička et al. 2020).. Sentinel constellation consists of two twin satellites – Sentinel-2A launched on 23 June 2015 and Sentinel-2B launched on 7 March 2017.

Sentinel-2 is equipped with an optical Multispectral Instrument (MSI) payload that consists of 13 spectral bands with a range of 0.443-2.19 μm in the electromagnetic spectrum: four bands at 10 m, six bands at 20 m and three bands at 60 m spatial resolution (tables 1-3). The Sentinel-2 satellite provides data from the same area every five days and at a resolution of 10, 20 and 60 m for each band. Original Sentinel-2 products are provided by ESA and are stored at the Copernicus Open Access Hub (Copernicus 2021b). Also, Sentinel products are distributed by the USGS, Planet Explorer, Remote Pixel platform and some other platforms. The images are stored and repackaged into tile-based bundles there. Copernicus Open Access Hub stores both level 1C and level 2A (after atmospheric correction from 2017) products.

Atmospherically corrected products are computed by using the Sen2Cor processor (Laštovička et al. 2020).

Table 1: 10 m Spatial Resolution Bands and associated SNR

Band number	S2A		S2B		Lref (reference radiance) (W m−2 sr−1 µm−1)	SNR
	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)		
2	492.4	66	492.1	66	128	154
3	559.8	36	559.0	36	128	168
4	664.6	31	664.9	31	108	142
8	832.8	106	832.9	106	103	174

Source: ESA 2021

Table 2: 20 m Spatial Resolution Bands and associated SNR

Band number	S2A		S2B		Lref (reference radiance) (W m−2 sr−1 µm−1)	SNR
	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)		
5	704.1	15	703.8	16	74.5	117
6	740.5	15	739.1	15	68	89
7	782.8	20	779.7	20	67	105
8a	864.7	21	864.0	22	52.5	72
11	1613.7	91	1610.4	94	4	100
12	2202.4	175	2185.7	185	1.5	100

Source: ESA 2021

Table 3: 60 m Spatial Resolution Bands and associated SNR

Band number	S2A		S2B		*Lref* (reference radiance) (W m−2 sr−1 µm−1)	SNR @ *Lref*
	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)		
1	442.7	21	442.2	21	129	129
9	945.1	20	943.2	21	9	114
10	1373.5	31	1376.9	30	6	50

Source: ESA 2021

The atmospheric correction processor Sen2Cor was developed by Telespazio VEGA Deutschland GmbH on behalf of ESA. Sen2Cor is a Level-2A processor which main purpose is to correct single-date Sentinel-2 Level-1C Top-Of-Atmosphere products from the effects of the atmosphere in order to obtain a Level-2A Bottom-Of-Atmosphere reflectance product (Main-Knorn et al., 2017). Additionally, Aerosol Optical Thickness, Water Vapour and Scene Classification (SCL) maps with Quality Indicators for cloud and snow probabilities can be generated.

The basic workflow of the Sen2Cor processor (fig. 5) consists of five modules of reading and processing data, providing the configuration parameters, creating an internal temporary database, and converting the products into the final format. Two main modules are designed to process the S-2 Level-1C input data, the SCL module and the AC module. The full description of the Sen2Cor retrieval methods is published in the paper by Main-Knorn et al. (2017).

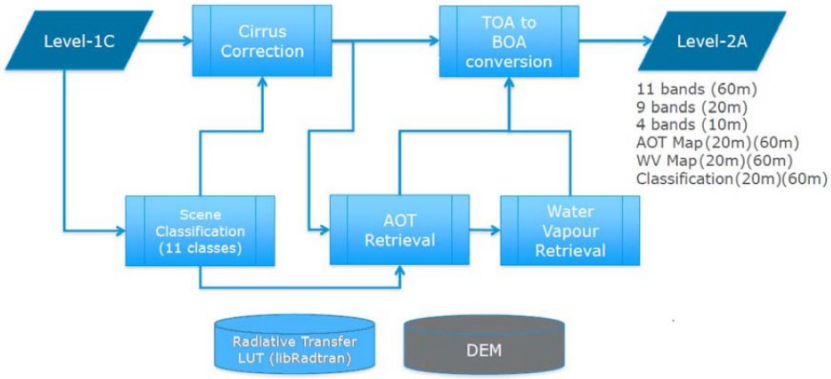


Fig. 5: Level-2A processing schema with Sen2Cor.
Source: Main-Knorn et al. (2017)

The Level-2A product provides Bottom-of-Atmosphere reflectance images with scene classification derived from the associated Level-1C products. Each Level-2A product is composed of 100 x 100 km tiles in cartographic geometry (UTM/WGS84 projection) (ESA 2021).

Cloud platform Google Earth Engine provides a "Sentinel-2 MSI: Multispectral Instrument, Level-2A" dataset, available from March 28th, 2017. The assets contain 12 UINT16 spectral bands representing atmospherically corrected Bottom-of-Atmosphere reflectance scaled by 10000, scene classification table (11 classes) and some more layers containing information about clouds. The scenes can be filtered by cloud cover metadata, e.g., setting the maximum at 20%, but this does not help much if the filter is set per scene of 100 x 100 km and the 20% just might cover the whole area of interest.

3.1.2 Auxiliary data from the CHMI and in-situ measurements

To verify the results obtained by processing satellite images, in this project there were used in-situ measurements. The in-situ measurements are observed and indicated by specialists from the CHMI, university professors and students and amateurs; what helps to get a high number of annual observation measurements. The methodology of ground phenology observations developed by the CHMI includes the correct choice of the monitored location and the tree individuals. The monitored tree has to be healthy (it is important to exclude plants irreversibly damaged or dying due to diseases, pests or severe mechanical damage (windfalls, thunderstorms etc.), but it does not apply to the seasonal occurrence of these conditions). It was necessary for the given plant species to be represented by five or, at least, by three individuals of a given species, which were in the age group between 20 and 50 years. Furthermore, when choosing the location of the monitored species one has to avoid very urbanized areas, in case it was not possible, to choose the areas of interest only on the border of the town; avoid places clearly disturbed by urbanization, industrial areas, constructions, mining or agricultural activities (surroundings of cement plants or other significant sources of dust, active quarries and pits, soil excavations, storage areas, landfills, areas affected by seepage of silage juices and storage of agrochemical materials or fuels, surroundings of granaries, barns, threshers and stacks); prioritize biomes richer in species (e.g. prefer mixed forests over cultivated spruce or forest edges over their denser and with less species diversity centers); give preference to the locations with open landscapes, when locations in deep basins, valleys and ravines were less suitable (Český hydrometeorologický ústav 2019).

Phenology information was obtained by observing the rate of seasonal development of selected plants suitable for this purpose, which specifically represent the monitored species. The rate of the plant development is determined not only by the seasonal changes, but it is also affected by fluctuations in climatic conditions over several years. The rate of the plant development is usually assessed based on knowledge of the time of onset of phenophases. According to the CHMI (Český hydrometeorologický ústav 2019), the description of the phenophase is a list of features whose current occurrence characterizes the phenophase and allows it to be recognized as a certain period in the process of plant development. The onset of the phenophase is a time step determining when the development has just reached the level given by the description of the phenophase. It is usually expressed in a calendar date determined according to the rules of phenological observations. In general, there are 1 to 3 levels of phenophase onset (10, 50, 100%). There is a set of rules, which have to be followed during phenology observations:

- to monitor selected plant individuals only,
- the onset of the phenophase is always determined by the description of the phenophase,
- the onset date is considered to occur only when at least half of the monitored individuals achieved the current phenophase.

The CHMI phenological stations provide information on the dates of all declared and monitored phenophases, but for the purposes of this work the following in-situ phenological phases monitored dates (depending on a certain species) were taken leaf budburst, leaf unfolding 10%, leaf unfolding 100%, leaf coloring 10% and leaf coloring 100% (Fenologické fáze 2021).

The phenophases leaf budburst and leaf unfolding 100% are observed for all selected tree species. All leaves or needles are considered when observing. The leaf blade is already deployed; depending on the phenophase, leaves start unfolding (10%) or have already unfolded completely (100%). The leaf has a characteristic mature shape and size. Only the 100% level is recorded, when most of leaves correspond to the description of the phenophase.

The phenophases leaf coloring 10% and leaf coloring 100% are observed for the majority of selected tree species (except pine and spruce). It must not be confused with pathological yellowing, which is a signal of the disease occurrence, pests or other adverse environmental factors. Similar to the greenup phenophase, there are 2 levels of the phenophase onset:

- 10% of the leaves on the tree have turned yellow (browning or otherwise discolored);

- 100% of the leaves on the tree have turned yellow (browning or otherwise discolored); when estimating the 100% level, it is also necessary to include leaves that have already fallen.

The in-situ data from the CHMI were provided in a format of an excel spreadsheet, which contained the information about the name of the phenological station, its geographical location and altitude above sea level, the monitored species and the dates of the onset of the first 3 phenological phases (leaf budburst, leaf unfolding 10% and leaf unfolding 100%) and 2 phenophases of EOS (leaf coloring 10% and leaf coloring 100%) for deciduous forests (fig. 6). The dates are presented as a serial day of the year, a spreadsheet transforming these serial numbers into the "dd-mm" format is attached.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Rok/ fáze	**RA**	**PL10**	**LX**	**ZL10**	**ZL100**		**Stanice: Chřibská (384 m n. m.)**					
2	**2016**	125	128	151	283	293		**Geografické souřadnice lokality sledované dubu letního:**					
3	**2017**	122	130	148	273	293		50°51'58'					
4	**2018**	111	116	132	282	289		14°28'54''					
5	**2019**	120	122	153	283	303							
6	**2020**	116	118	150	293	303							
7													
8	**RA= rašení**												
9	**PL10=první listy 10%**												
10	**LX= plné olistění**												
11	**ZL10= žloutnutí 10%**												
12	**ZL100= žloutnutí 100%**												
13	**Údaje jsou uvedeny v pořadovém dni roku.**												
14													

Fig. 6: The in-situ data provided by the CHMI for each monitored station
Source: The CHMI

The altitude of each area of interest was verified whether by the WMS layer "ZABAGED - výškopis - 3D vrstevnice" of ČÚZK geoportal or by GPS field surveys. The altimetry contour line intervals of the WMS layer differ depending on the character of the relief (5, 2 or 1 m).

3.2 Study Areas

The study areas were selected according to the cooperation of the Faculty of Science of the Charles University and the phenological department of the CHMI, which provided detailed information and measurements of certain phenophases of selected tree species over the monitored period. The selected areas differ not only in vegetation cover type, deciduous or coniferous, but also in geographical and climate conditions: lowlands, flatlands, mountainous regions, in the north, northwest, southeast and southwest of Czechia. Therefore, 6 study areas with homogeneous vegetation cover were chosen based on the location of the monitored phenological stations. The bands used for the time series analysis of vegetation have 10 m respectively 20 m spatial resolution. For this reason, the MMU was set as the area of 400 m^2. It

ensured the homogenous representation of the studied vegetation type. The accuracy of the results obtained by remote sensing observations was verified and correlated with the in-situ measurements of the same location.

To identify the pixels of interest GPS measurements were conducted with the Trimble GeoXH device. The list of coordinates of areas was provided by the CHMI for the following case study areas: Běleč nad Orlicí, Modrava, Chřibská, Frýdlant, Měděnec and Vranovice. The provided coordinates had to be verified to find a homogeneous location of the evaluated species. First, there were used the maps/spatial information provided by the Phenological Department of the CHMI as well as the tree species composition maps (from 2013 to 2019) in the geoportal of the Institute for Forest Management (ÚHÚL), which maintains a central database with information on forests land cover of the Czech Republic, forestry and hunting (ÚHÚL 2021). The species were located either in heterogeneous vegetation areas or on the edge of a forest, or they were presented by single trees. These areas required precision measurements by the GPS. For purpose of location within a grid of Sentinel-2 pixels, a vector grid of Sentinel-2 image of 10 and 20 m spatial resolution were created and imported into the QGIS project (CRS is WGS 84 / UTM Zone 33N). For in-situ measurements was this project exported to the mobile application QField. The rasters were transformed into vector grids using the QGIS tool *Raster > Conversion > Polygonize (Raster to Vector).* To locate the pixels of interest as precisely as possible and retrieve additional relevant information from WMS layers provided by ČÚZK, the following WMS layers were added to the project: Hill Shaded Digital Terrain Model of ČR (DMR 5G), Orthophoto, CIR Orthophoto.

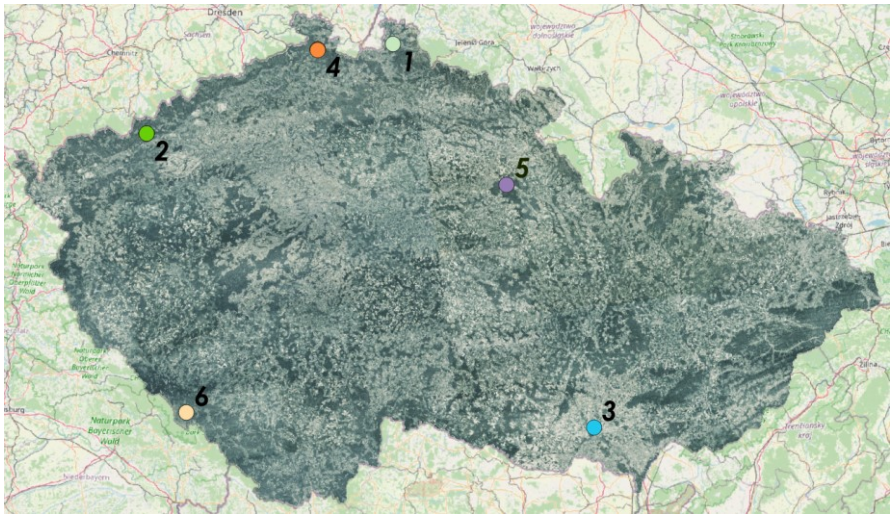


Fig. 7: The location of the case study areas (1 – Frýdlant, 2 – Měděnec, 3 – Vranovice, 4 – Chřibská, 5 – Běleč nad Orlicí, 6 – Modrava)
Source: author's work in QGIS

3.2.1 Frýdlant

The monitored area is located on a plain land with the soil stratum of sands, 423 m above sea level, with an inclination angle of 5° orientated to the southwest. According to the data provided by the CHMI, the area of interest represents a 90x30 m site of hornbeams (Carpinus betulus) surrounded by arable land (fig. 8).

Fig. 8: The location of the pixels of interest in the Sentinel-2 10x10m grid (Frýdlant)
Source: author's work in QGIS

The figure 9 demonstrates a positive trend in average temperature values over the last 3 years compared to the 30-year average. Unlike the other monitored locations, the long-term climate observations in Frýdlant reveal lower mean precipitation in the last 3 years compared to the 30-year average (580.8 mm in 2018, 649.2 mm in 2019, 781.5 mm in 2020, when the overall average is 845.2 mm).

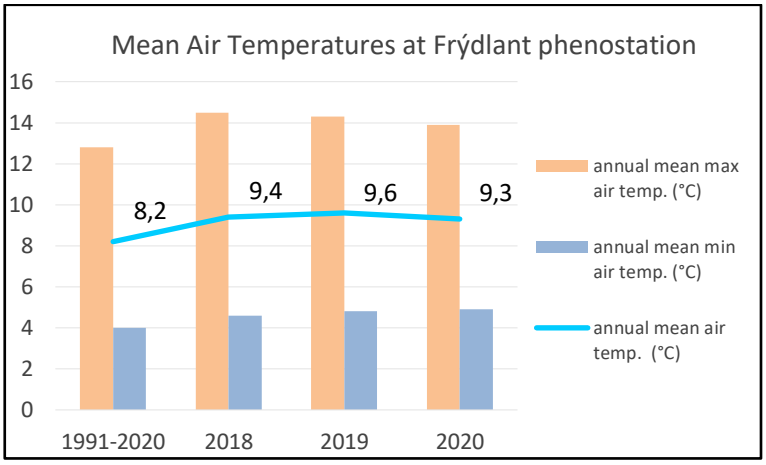


Fig. 9: Temperature dynamics over monitored years in Frýdlant
Source: The CHMI

3.2.2 Měděnec

The study area is located on a highland (823 m above sea level) with an inclination angle of 5°, orientated to the northwest. The area has undergone a significant change due to systematic deforestation in recent years. Strands of spruce, beech, maple, and cherry tree have been cut

down since 2017. A patch of a few beech trees (Fagus sylvatica) is monitored at this site (fig. 10, 11).

Fig. 10: The location of the pixels of interest in the Sentinel-2 10x10m grid (Měděnec)
Source: author's work in QGIS

Fig. 11: The monitored beech trees in the area of interest in Měděnec
Source: author's photos

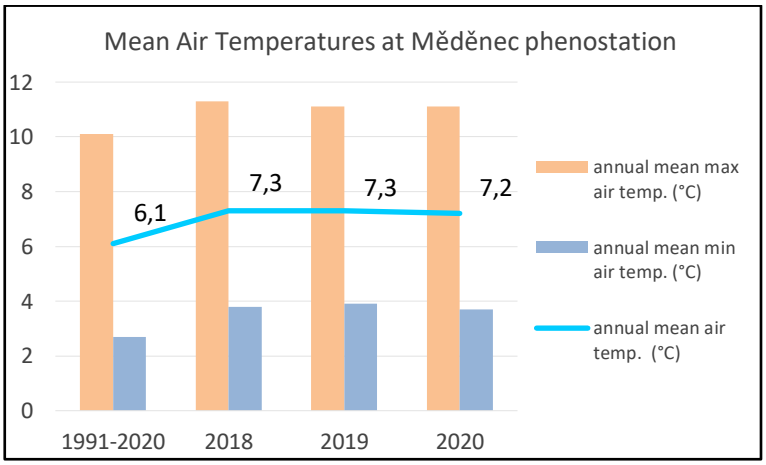


Fig. 12: Temperature dynamics over monitored years in Měděnec
Source: The CHMI

From the fig. 12, it is evident that in recent years the average year temperature has increased by 1°C compared to the entire monitoring period. On the contrary, the average

precipitation level has decreased in recent years (the average over the last 3 years is 661.9 mm and 805.5 mm in 1991-2020).

3.2.3 Vranovice

The monitored area is located in the lowlands (173 m above sea level). It is represented by a homogenous ash tree forest with an underbrush vegetation level of a size of 0.036 km^2 (fig. 13, 14).

Fig. 13: The location of the pixels of interest in the Sentinel-2 10x10m grid (Vranovice)
Source: author's work in QGIS

Fig. 14: The monitored ash trees in the area of interest in Vranovice
Source: the CHMI

The area of Vranovice is characterized by higher annual mean air temperatures compared to the other study areas. Annual mean maximum air temperature is 17°C for the monitored period of 2018-2020 (fig. 15).

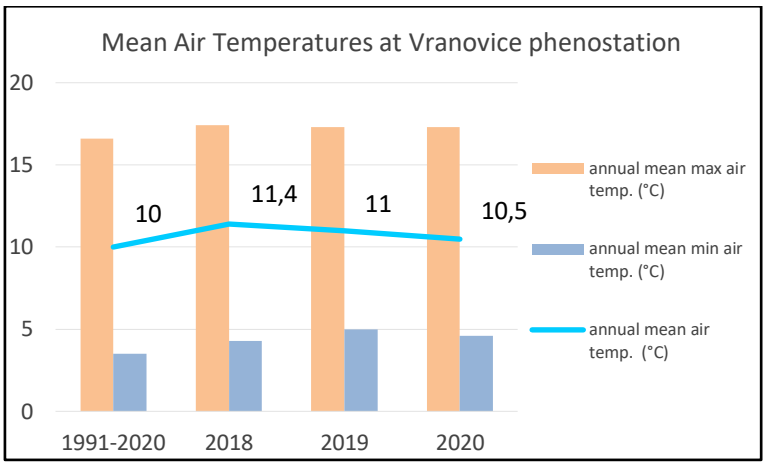


Fig. 15: Temperature dynamics over monitored years in Vranovice
Source: The CHMI

3.2.4 Chřibská

In the area of Chřibská, the oak tree (Lucanus cervus) is monitored (fig. 17). Compared to the previous stations, in Chřibská only 2 individual oak trees come down to the marked out area of 20 x 20 m^2. Due to the dense and wide tree canopy, the prevailing area of the pixel is covered by the chosen species. The monitored area is located on a moderate slope with an inclination angle of 10°, orientated to the west, 370 m above sea level (fig. 16).

Fig. 16: The location of the pixels of interest in the Sentinel-2 10x10m grid (Chřibská)
Source: author's work in QGIS

Fig. 17: The monitored oak tree in the area of interest in Chřibská
Source: author's photos

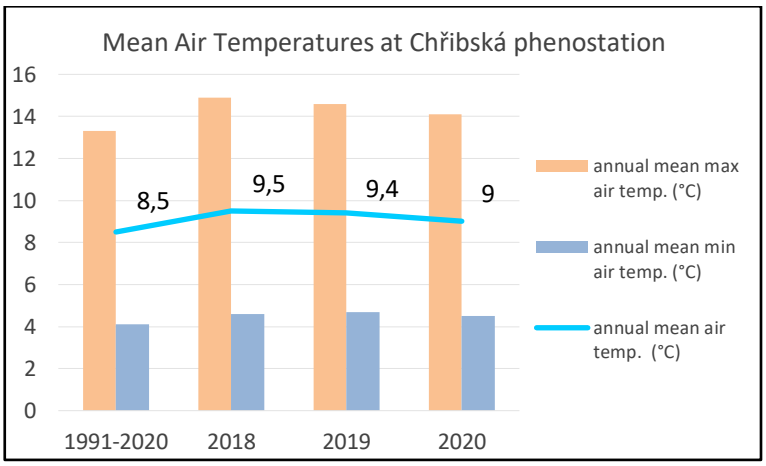


Fig. 18: Temperature dynamics over monitored years in Chřibská
Source: The CHMI

The average temperatures in Chřibská have been over 9°C in the last three years compared to the 30-year average of 8.5°C (fig. 18). Therefore, there is positive dynamics in precipitation level: from 488.1 mm in 2018 to 907 mm in 2020, while the 30-year average is 824.1 mm.

3.2.5 Běleč nad Orlicí

At the Běleč nad Orlicí phenological station pine trees were monitored (fig. 19). The station is located on a plain land with the soil stratum of sands, 241 m above sea level, with an inclination angle of 5°. The study area is represented by light pine (Pinus sylvestris) forest, no mixed species affect the monitoring.

Fig. 19: The location of the pixels of interest in the Sentinel-2 10x10m grid (Běleč nad Orlicí)
Source: author's work in QGIS

According to the data available at ÚHÚL (2021) and ČÚZK (2021) geoportals, the vegetation structure has been stable in this area (the data is available from 2013). The average minimum and maximum year temperatures over the monitored years are presented in fig. 20. The precipitation level shows a high amplitude over these years: 394.3 mm in 2018, 556.4 mm in 2019, 710.2 mm in 2020 compared to 587.4 mm over the period of last 30 years.

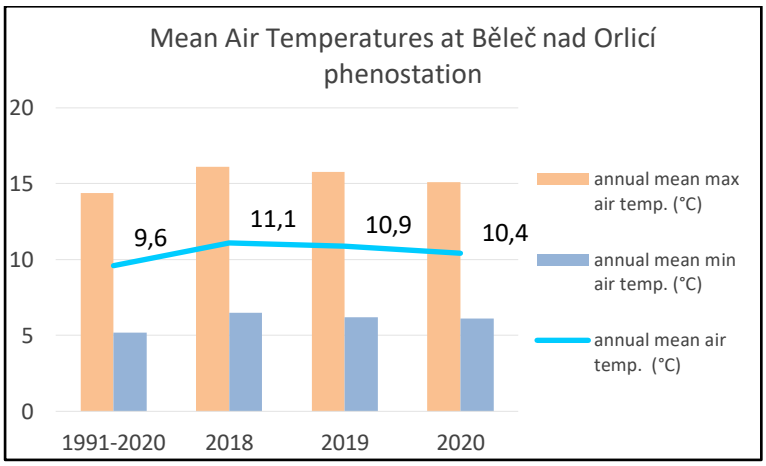


Fig. 20: Temperature dynamics over monitored years in Běleč nad Orlicí
Source: The CHMI

3.2.6 Modrava, Filipova Huť

The phenological station is located on a plain land 1002 m above sea level with an inclination angle of 5° orientated to the south (fig. 21). The area is represented by pure spruce (Picea abies) forest.

Fig. 21: The location of the pixels of interest in the Sentinel-2 10x10m grid (Modrava)
Source: author's work in QGIS

The average minimum and maximum year temperatures in Modrava are significantly lower than the other locations (fig. 22). The precipitation level does not show any significant dynamics in the last 3 years: 1053.1 mm in 2018, 932.4 mm in 2019, and 1070.5 mm in 2020 with 1092.4 mm average over the period of last 30 years.

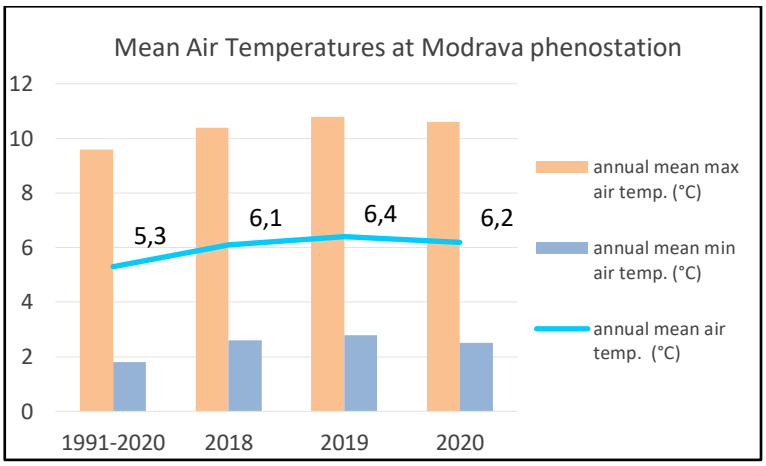


Fig. 22: Temperature dynamics over monitored years in Modrava
Source: The CHMI

3.3 Google Earth Engine data processing

According to the OGC statistics, the total volume of generated remote sensing data is increasing significantly every month. The current trend is the usage of free accessible open data satellite imagery sources (Google Earth Engine, Copernicus Sci-Hub, USGS, Sentinel Hub, Amazon Web Services). Large amount of diverse freely available information brings new possibilities for scientific research, complex analysis, and accurate modeling of different natural phenomena.

Google Earth Engine is an open access cloud-based platform for scientific analysis and visualization of geospatial datasets. It has a searchable data catalog, including the Sentinel, Landsat or MODIS datasets, precipitation data, sea surface temperature data, climate data, and elevation data. Earth Engine hosts satellite imagery and stores it in a public data archive. Besides the library of satellite imagery, the platform contains an API for accessing and processing remote sensing data from the archive (supports JavaScript, Python) and an infrastructure for parallel calculation of large volumes of data. Parallel computing in a cloud environment provides a significant increase in data processing speed compared to a single computer (Paluba et al. 2021). The second significant advantage is quick and easy access to the platform using a web browser with the ability to work with the data catalog and the use of analysis algorithms. The Google Earth Engine is a platform for analyzing geospatial information, where forest and water coverage, land use change and other applications can be analyzed.

The developed algorithm enables the user to generate and export the time series of VI values of a specified polygon or a point over a certain period of time. Primarily, two datasets

provided by the GEE cloud platform were required for the further data processing: “Sentinel-2 MSI: Multispectral Instrument, Level-2A” and “Sentinel-2: Cloud Probability”. Sentinel-2 data in GEE has been stored since March 28, 2017. Firstly, both Sentinel-2 L2A and Sentinel-2 Cloud Probability image collections were loaded into the code editor and filtered by the date (from March 28th, 2017 to December 1th, 2020) and by the area of interest (Copernicus 2021b; GEE 2021a). The GEE allows to draw points or polygons directly in the map window of the cloud platform or to import shapefiles.

If the areas/pixels of interest were located, the shapefiles were imported to the GEE *Assets > New > Table Upload (Shape files).* These cloud assets allow storing data in the form of GeoTIFF images, Shape files or CSV tables. After filtering both images collections, they were merged into one in order to join the *probability* band from the Sentinel-2 Cloud Probability collection to the Sentinel-2 L2A collection. To perform this, the GEE functions *ee.Join.saveFirst* and *apply()* were used. As a result, the *probability* band was added to the Sentinel-2 L2A collection as an image property, based on the condition that the property *‘system: index’*, which is unique for every image, pairs an image in the second collection.

The Cloud Probability dataset is created by applying a machine learning-based cloud detection algorithm s2cloudless. This algorithm reportedly significantly outperforms other methods, such as Sen2Cor, SenL1C, and FMask (López-Puigdollers et al. 2021). The approach itself was based on taking cloud masks coming from the multitemporal MAJA processor and training a simple gradient boosted decision trees algorithm in a way that it works on pixel level and that one can run it even at different resolutions. All bands are resampled using bilinear interpolation to 10m resolution before the gradient boost base algorithm is applied. The resulting 0..1 floating point probability is scaled to 0..100 and stored as UINT8. Areas missing any or all of the bands are masked out. Higher values are more likely to be clouds or highly reflective surfaces (e.g., roof tops or snow) (GEE 2021b).

To mask out cloudy and shadowed pixels, the following cloud masking algorithm using the Sentinel-2 Cloud Probability dataset was applied. Clouds were identified from the Sentinel-2 Cloud Probability dataset (s2cloudless) and shadows were defined by cloud projection intersection with low-reflectance near infrared (NIR) pixels (Spatial Thoughts 2020). Firstly, the user has to define a set of the required parameters:

cloud_filter - maximum allowed image cloud cover percentage;

cld_prb_thresh - cloud probability (extracted from the probability band; values greater than the threshold are considered clouds);

nir_drk_thresh - reflectance in the NIR band (values less than are considered potential cloud shadows);

cld_prj_dist - maximum allowed distance (km) to search for cloud shadows from cloud edges;

buffer - distance (m) to dilate the edge of the objects identified as clouds.

In this work, after practical consideration and visual comparison, the above-mentioned parameters were defined as following: *cloud_filter* = 60%, *cld_prb_thresh* = 40%, *nir_drk_thresh* = 0.15, *cld_prj_dist* = 1 km and *buffer* = 45 m. To define the buffer parameter is an important step, which minimizes cloud and cloudy shadow pixels uncertainty. The objective of this specific cloud detection is to remove messy pixels from the follow-up processing. After that, there were defined the functions, first, to get the probability band from the final image collection and apply the cloud probability threshold; second, to identify and exclude water pixels using the *Scene Classification Map* band from Sentinel-2 L2A dataset (the value is 6), to identify dark pixels based on the NIR band and the *nir_drk_thresh parameter*, to project shadows based on the *'EAN_SOLAR_AZIMUTH_ANGLE'* image property and the *cld_prj_dist* parameter; third, to combine cloud and cloud shadow components to produce the final mask (fig. 23).

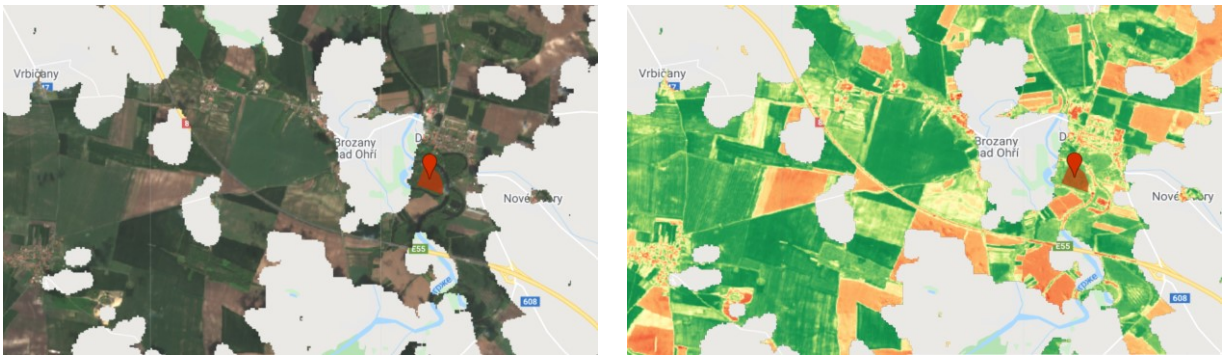


Fig. 23: Masking out cloudy and shadow pixels using the s2cloudless algorithm
Source: author's work in GEE

The next step after masking out cloudy and shadowed pixels was to compute VI values for each area of interest. There were defined functions to generate NDVI, RENDVI, NDRE, NDMI and MCARI according to the corresponding determinate formulas. They were generated using either the *image.normalizedDifference()* or *image.expression()* functions. Both functions require the band names as input parameters, in addition to that, the formula itself has to be

defined in the second case. Then, to add the new bands to the image collection and generate the VI time series, the created functions were mapped over each image of the collection by applying the function *ee.ImageCollection.map()*. As a result, every image in the final image collection included 27 bands and 83 properties. To generate time series plots the *ui.Chart.image.seriesByRegion()* was applied.

Then, the data had to be adjusted for export as a CSV table. First, the VI band was selected by the function *image.select()*, to which the *reduceRegions()* function was applied to reduce the area of extracted pixels to the defined 20 x 20 m polygon, and filter the values which do not equal to zero, as in many images relevant pixels could be masked out by the cloud detection algorithm. Another issue, which has to be taken into account when exporting Sentinel-2 time series as a CSV, is a possible overlap of the neighboring tiles (fig. 24). It means that the same raw pixel can be present in up to 4 tiles. In addition, since each tile is processed independently, the output pixel values can be slightly different. As a solution to that problem, these overlapping pixels need to be harmonized by taking all VI values for the same day (generated from the same raw pixels) and assign the maximum value as an export value. This results in a clean output with one NDVI value per observation.

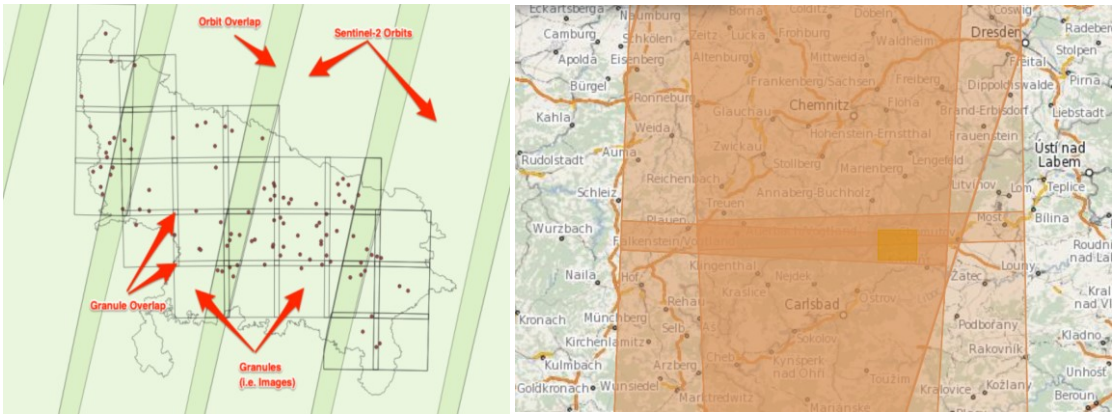


Fig. 24: The overlap of the adjacent Sentinel-2 tiles
Source: Spatial Thoughts (2020); Copernicus (2021b)

All these steps have to be repeated not only in one image, but also over the whole image collection, for this purpose the *ee.ImageCollection.map()* function was applied again.

Then the time series results were exported to a csv table. The obtained VI time series were corrected and adjusted by applying fitted model functions in the R environment (RStudio). Based on the literature review (Misra et al. 2016), it was proved to achieve better results using the asymmetric Gaussian function and the double logistic function before other alternative filters. Both techniques showed an approximately balanced noise reduction ability while

maintaining the proper integrity of the VI value. After applying fitted model functions certain phenology information such as SOS, LOS and EOS of the growing season were retrieved.

3.4 RStudio data processing

This chapter of the thesis is focused on the methods, which were used to filter and smooth VI time series and to extract phenology information from raw and corrected time series values. The general literature research of existing algorithms is provided in chapter 2 of this work.

For the purpose of the VI analysis, programming language R was chosen. R is a programming language and free software environment for statistical computing and graphics and data analysis. One of the most popular user interfaces is RStudio, an open access integrated development environment for desktop. The advantage of the R language is the fact that any user can create a set of functions, which allow specialized statistical techniques, graphical devices, import/export capabilities, reporting tools etc. These sets are called R packages, extensions to the R statistical programming language. R packages contain code, data, and documentation in a standardized collection format that can be installed by users of R, typically via a centralized software repository such as CRAN (the Comprehensive R Archive Network) or R-Forge. In this work, there were used 3 packages, which provide certain functionality for LSP analysis, specifically for phenology metrics extraction.

The methodology of phenology metrics extraction is based on model fitting. To allow for the variable nature of the VI curves, several models have been developed that can easily be adjusted to the available satellite data. The fitted model can be applied on raw or corrected data. In this work, two fitting model methods were tested - Asymmetric Gaussian function and Double Logistic function (Beck et al. 2006). These two methods were chosen based on the literature research, described in detail in chapter 2.

The package phenex provides a collection of functions for spatial and temporal analysis of phenological datasets and satellite observations of vegetation. For the noise reduction of the time series, it uses the Moving Average method. In statistics, a moving average (rolling average or running average) is a calculation to analyze data points by creating a series of averages of different subsets of the full data set (Wikipedia 2021). The mean is normally taken from an equal number of data on either side of a central value. This ensures that variations in the mean are aligned with the variations in the data rather than being shifted in time. An example of a simple equally weighted running mean (*SMAk*) is the mean over the last k entries of a dataset containing n entries (6).

$$SMA_k = \frac{p_{n-k+1} + p_{n-k+2} \cdots + p_n}{k} = \frac{1}{k} \sum_{i=n-k+1}^{n} p_i \quad (6)$$

Then the algorithm applies the Asymmetric Gaussian Function on the corrected by Moving Average method (Lange, Doktor 2017).

Jönsson, Eklundh (2002) developed a model fit existing of a number of local model functions that are merged in a global function. The authors proved the increasing flexibility to the complex behavior of the time series compared to a simple Gaussian model or lower order Fourier estimates. The Gaussian-type local functions are as follows (7):

$$\text{NDVI} = c_1 + c_2 \begin{cases} \exp\left[-\left(\frac{t - a_1}{a_2}\right)^{a_3}\right], & \text{if } t > a_1 \\ \exp\left[-\left(\frac{a_1 - t}{a_4}\right)^{a_5}\right], & \text{if } t < a_1 \end{cases} \quad (7)$$

The base parameters c_1 and c_2 determine the intercept and the amplitude of the curves, respectively. The parameter a1 determines the timing of the maximum (measured in time units). The upper part of the equation is fitted to the right half of the time series (time is after the peak a_1 is reached). The lower part of the equation fits to the left half of the time series. The parameters a_2 and a_4 determine the width of the curves on the right and left side, respectively. The parameters a_3 and a_5 determine the flatness of the curves on the right and left side respectively. The fit of this local function alone already requires the estimation of seven parameters. The authors provided the option of fitting a centerpiece to the model, in case of a flat plateau for a peak, introducing even more parameters. After this rather complicated function has been fitted to the VI time series, SOS is determined from the global model as the point in time for which the value has increased 10% above the base level. The authors use a threshold method to determine SOS (de Beurs, Henebry 2010).

Fig. 25 shows how the local functions change based on the varying parameters. In (a) parameter a_2, which determines the width of the right function half, has been decreased (solid line) and increased (dashed line) compared to the value of the left half. In (b) parameter a_3, which determines flatness of the right function half, has been decreased (solid line) and increased (dashed line) compared to the value of the left half (Jönsson, Eklundh 2004).

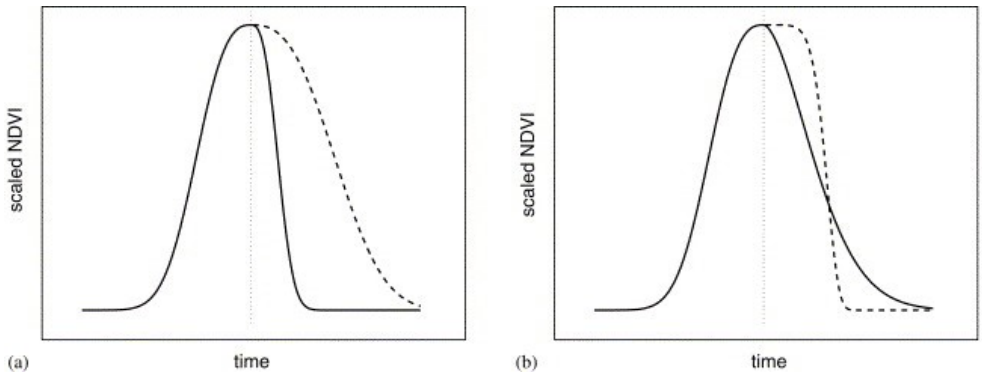


Fig. 25: The effects of varying the parameters in AG function fitting
Source: Jönsson, Eklundh (2004)

Another model fitting method tested for phenology information extraction in this work is Double Logistic function (Beck et al. 2006). The authors propose a Double Logistic function to describe NDVI time series spanning one year. The Double Logistic function models the NDVI as a function of time (t) using six parameters (fig. 26): the winter NDVI (wNDVI); the maximum NDVI (mNDVI) during the growing year; two inflection points, one as the curve rises (S) and one as it drops (A); and the rate of increase or decrease (mS and mA) of the curve at the inflection points (8).

$$\mathrm{NDVI}(t) = \mathrm{wNDVI} + (\mathrm{mNDVI} - \mathrm{wNDVI}) \times \left(\frac{1}{1+\exp(-mS \times (t-S))} + \frac{1}{1+\exp(mA \times (t-A))} - 1 \right) \quad (8)$$

For each pixel and year, the authors used the earlier derived winter NDVI and estimated the remaining five parameters using an iterative non-linear least squares procedure. This resulted in a double logistic curve fitted to the NDVI time series. Based on the residuals produced by this fit, a weighting scheme for the observations is set up and in a second iteration reestimated all parameters except the winter NDVI. The weighting scheme considers observed NDVI values that are initially overestimated by the fitted curve and assigns lower weights to them for the second iteration. This type of weighting is commonly applied in NDVI modelling algorithms and rests on the assumption that anomalies in the NDVI dataset are negatively biased (Beck et al. 2006). The example in figure 26 is defined by six parameters: the winter and maximum NDVI (wNDVI and mNDVI), the spring and autumn inflection points (S and A), and the rate of increase or decrease in NDVI at S and A, respectively (mS and mA).

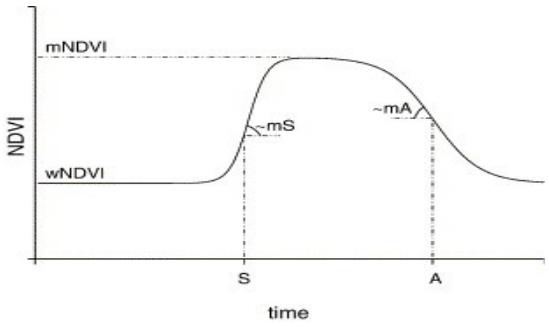


Fig. 26: Example of the Double Logistic function used to model the yearly NDVI time series. Source: Beck et al. (2006)

The R packages for phenology analysis, which provide the above-mentioned functionality, were used. These are Phenex, Phenopix and Greenbrown packages.

In the **phenex package** (Lange, Doktor 2017), the *Asymmetric Gaussian function* was applied to the one-year VI time series of available satellite observations. The fitted model method was applied to both raw time series data and corrected time series values after smoothing it by the *Running Average* filter. The first approach consists of the sequential application of the following package functions: *new(), modelValues(), phenoPhase()*.

new() – generates a newly allocated object from the class identified by the first argument. Corresponding to the specified class, named arguments have to be passed. In the case of the Phenex package, the arguments are *values*, the one-year VI time series, and *year*, corresponding year to the values.

modelValues() – modeled VI values by fitting Asymmetric Gaussian function to the time series. By applying *phenoPhase()* phenology metrics were extracted based on modelled VI time series values. The required arguments were: *x* – an object of class "NDVI" containing modelled VI values; *phase* – determines which phase will be extracted: greenup and senescence were chosen in this work; *method* – was set to *local* with the threshold defined as the first VI value between lowest and highest VI value of time series. One of the main tasks was to correctly define the input parameters for the *phenoPhase()* function. For the purposes of this work, a method based on local threshold was chosen (Walker et al. 2014). The threshold was defined individually for every VI calculation based on the max/min VI values difference. As the functions requires the time series values to be in the [0, 1] interval, the MCARI time series values had to be normalized. The number of normal distributed values around the threshold for the standard deviation estimation was set to 1000. The *Running Average* correction method was chosen to smooth the VI time series values. The code for the phenophase detection in the phenex package can be found in the appendix 9. As an output the function produces a list containing

the index corresponding to the day of the year at which the phenological phase occurs as a list entry *mean* and its standard deviation as a list entry *sd*.

In the case of the **phenopix and greenbrown packages** (Filippa et al. 2020; Forkel, Wutzler 2015) the extraction of phenology metrics was computed by the application of the Double Logistic function. Both sequentially fit a Double Logistic curve to observed values and extract phenological metrics according to the specified method. As an input, an object of the class 'time-series' is required. For this purpose, a *ts()* function from the system package *stats* has to be called. As input parameters, only the vector of observed time series values was declared, the parameters like *start*, *end* and *frequency* were set automatically to the first index of the data frame, to the last index of the data frame and the frequency was set to 1, respectively. The indices are aligned with the date of the observation and the obtained value of the VI time series.

The next step after transforming the vector of the time series values is to fit a double logistic curve to the values using the function as described in (Beck et al. 2006) (7), which estimates parameters of the double logistic equation and provides an uncertainty estimation. Uncertainty might be computed by adding noise to the raw data and by estimating again the parameters. Noise is added according to the standard deviation of the residuals. The required argument for this function is an input of an object of the class 'time-series'. The argument 'uncert' defines whether uncertainty should be estimated or not (default is set to FALSE), others specify the uncertainty parameters in case the uncertainty is calculated. For the purpose of efficiency and due to computation time, the uncertainty was not computed in this project. After performing the *BeckFit()* function, a list with fitted values, parameters, fitting formula and a list containing uncertainty predicted values if the uncertainty was computed. The developed code is in appendix 9.

After the Double Logistic curve was fitted to the observed time series values, phenology metrics were extracted using the *PhenoExtract()* function. As inputs, the function required the following arguments:

data – an output from the fitting *BeckFit()* function;

method – implements the derivative method for phenology (White et al., 1997);

uncert – defines whether uncertainty should be estimated or not;

envelope – "quantiles" or "min-max" might be chosen. For "quantiles", the uncertainty envelope will be computed as quantiles, for "min-max" as min, max and mean.

quantiles - specifies quantiles in the quantile function;

plot - defines whether a plot should be returned with annotated thresholds;

As an output, there is a list of the calculated phenology metrics (table 4) and a graph (they can be found in the appendix 8 for each VI).

Table 4: Extracted phenology metrics in the phenopix package.

sos	*start of the season*
eos	*end of the season*
los	*length of the season*
pop	*position of peak value (maximum)*
mgs	*mean growing season value*
rsp	*rate of spring greenup*
rau	*rate of autumn senescence rates*
peak	*peak value (maximum)*
msp	*mean spring value*
mau	*mean autumn value*

Source: Filippa et al. (2020)

In contrast to the corrected time series values algorithm, to extract phenology information from a raw data time series the *PhenoBP()* function was used. The input parameters are similar to the previous functions: an object of the class "time-series", a maximum number of breaks to be detected in the time series (was set to 2, because the goal was to extract SOS and EOS), confidence level for the uncertainty computation (was set to 0.95, same as in uncertainty computations in previous methods) and whether the plot should be computed or not. Unlike the other threshold approaches, *PhenoBP()* is developed to work with raw data. It therefore does not require fitting an equation. However, the function can also be used with fitted data.

Phenology information extraction is performed according to the breakpoint analysis. The function calls *breakpoints* from the package *strucchange*, testing, monitoring and dating structural changes in classical linear regression models. Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered an explanatory variable, and the other is considered a dependent variable (Yale University Courses 2021). A linear regression model has an equation of the form (9):

$$Y = a + bX, \qquad (9)$$

where X is the explanatory variable and Y is the dependent variable. The slope of the line is b, and a is the intercept (the value of y when x = 0).

To compare with the efficiency of the phenopix package, the Double Logistic function was fitted to the time series observations in the greenbrown package. It refers to the same

equation described in Beck et al. (2006) but requires different input arguments. Besides an object of the "time series" class, the function *FitDoubleLogBeck()* declares the time steps of the input and output time series, as it was described in the phenopix package, where the frequency was set to 1 and every time step referred to a certain obtained observation. Apart from that, the weighting scheme can be applied to the observed values as described in (Beck et al. 2006) (fig. 26). This is useful for NDVI observations because higher values will get a higher weight in the estimation of the double logistic function than lower values. Apart from that, the following parameters can be specified: whether to compute standard errors of parameters based on the Hessian matrix or to plot the Double Logistic curve and the number of initial parameter sets from which to start optimization.

The next step is to implement a derivative method to extract phenology metrics using the function *PhenoDeriv()*. This function was applied to the raw data. As an output a list of phenology metrics was obtained (table 5), the graphs for each VI are in the appendix 8.

Table 5: Extracted phenology metrics in the greenbrown package.

sos	*start of the season*
eos	*end of the season*
los	*length of the season*
peak	*peak value (maximum)*
pop	*position of peak value (maximum)*
mgs	*mean growing season value*
rsp	*rate of spring greenup*
rau	*rate of autumn senescence rates*
msp	*mean spring value*
mau	*mean autumn value*
trough	*trough value (minimum)*
pot	*position of trough value (minimum)*

Source: Forkel, Wutzler (2015)

4 Results

In this chapter there were evaluated the results of SOS and EOS detection of certain vegetation species based on the methodology proposed in the previous chapter. The short-term VI time series were calculated from the VI values (a mean value taken from four 10 x 10 m pixels or from one 20 x 20 m pixel depending on the Sentinel-2 band resolution) over the period of 3 years (2018-2020). As vegetation indices there were used NDVI, RENDVI, NDRE, NDMI, and MCARI. For the selected phenological stations, the phenological variables (start and end of the growing season) of each dataset were also compared with ground phenological data. One of the main objectives of the work was to evaluate the data sources of vegetation indices in terms of extracting the start and end of the growing season based on ground measurements.

The following parts of this chapter describe the results obtained from each monitored area: Běleč nad Orlicí (pine tree), Modrava (spruce), Chřibská (oak tree), Frýdlant (hornbeam), Měděnec (beech tree), and Vranovice (ash tree). To detect SOS and EOS, the phenology analysis algorithms were applied to the VI time series in RStudio. The phenophases were calculated for the corrected VI time series values (the results presented in a form of graphs can be found in the appendix 8). There were used three phenology detection packages in R: Phenex, Phenopix and Greenbrown. The Asymmetric Gaussian function and Double Logistic function (according to Beck et al.) were fitted to annual VI time series. To determine the breakpoints on the curve, the derivatives method was applied (de Beurs, Henebry 2010).

The chapter is divided into 7 sections of the corresponding monitored areas and the last one presenting the overall statistical analysis of the VIs performance. The first part of each section describes available Sentinel-2 images and their distribution over monitored years. The second part visualizes the VI time series generated in GEE along with linear regression trend line over the period. The third part of the section provides detailed information on the results obtained after processing VI time series using phenology packages in R, along with the set of graphs representing the DOY correspondence. Apart from that, standard deviations for each VI and its reference value, and the absolute mean standard deviations were calculated.

As it was described in previous chapters, there were generated time series of a selection of vegetation indices in the GEE: NDVI, NDRE, RENDVI, NDMI and MCARI. Then, the VI time series were evaluated and compared to define which index is the most relevant to use in detection of phenophases depending on the vegetation type, geographical location and climate conditions. The linear regression trend line is generated for every index and it shows whether

mean VI values increase or decrease over the monitored period. All VI time series generated in GEE for the study areas are present in the appendices 1-6, in the appendix 7 the maximum VI values for each year are calculated. The map depicting this trend can be generated in the GEE code for each area of interest as a part of the algorithm.

The format of the input data in the R algorithms is a dataframe. Each dataframe consists of three columns: *index*, *date* and *value*. After running the algorithm, as an output the user receives an index number of the corresponding observation of SOS or EOS, respectively. It is of importance to note that originally the index does not refer to a DOY because the available observations within each month are distributed unevenly, due to filtering most of cloudy and shadowed pixels out. In this chapter, both results obtained from the ground and remote sensing observations were transformed to the DOY format.

4.1 Frýdlant

The Sentinel-2 image collection after filtering by the area extent, time period and cloud coverage consisted of 365 images. Then, to avoid null values in the area of interest for further time series generating, the image collection was reduced to the cloudless pixels with VI values not equal to zero, overlapping pixels were eliminated. The final collection contained 135 images in total. The distribution of the available observations per month over the period of interest is shown in table 6. The total number of the acquired images in 2018 is 47, in 2019 and 2020 there were 38 and 36 images, respectively. The distribution of the observations in 2019 is similar to 2018, except the fact that there are no available observations in August 2019. In 2020, the maximum number of the available observations was registered in August (8 images).

Table 6: The number of the obtained observations per each month in Frýdlant.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
2018	1	5	2	6	6	2	4	7	4	6	3	1	47
2019	1	6	3	6	2	6	4	-	5	2	3	-	38
2020	1	2	4	6	2	2	4	8	4	2	1	-	36

Source: author's work

The outputs generated in GEE represent the VI time series and the linear trendline with the coefficient of determination (R-squared) for the corresponding vegetation index (fig. 27, 28). All VI time series graphs generated in GEE can be found in Appendix 1. NDVI showed the best accuracy in SOS detection in the phenex package (4 day mean standard deviation for

all years in total), whereas NDRE detected EOS with the same accuracy in the phenopix and greenbrown packages. The more detailed description of the results can be found below.

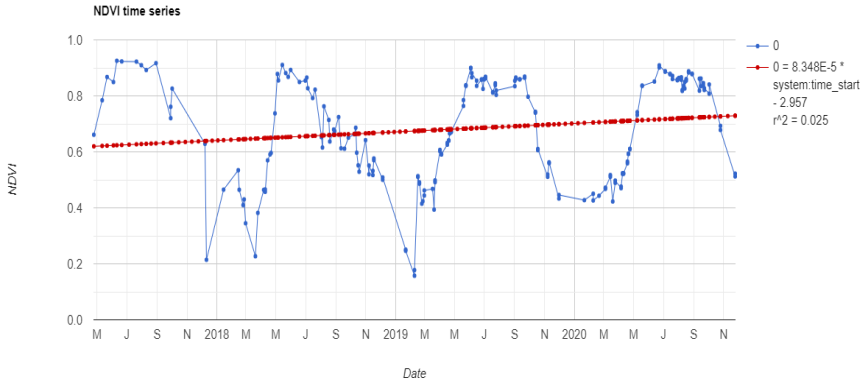


Fig. 27: The NDVI time series of 20x20m area (Frýdlant) in 2017-2020
Source: author's work in GEE

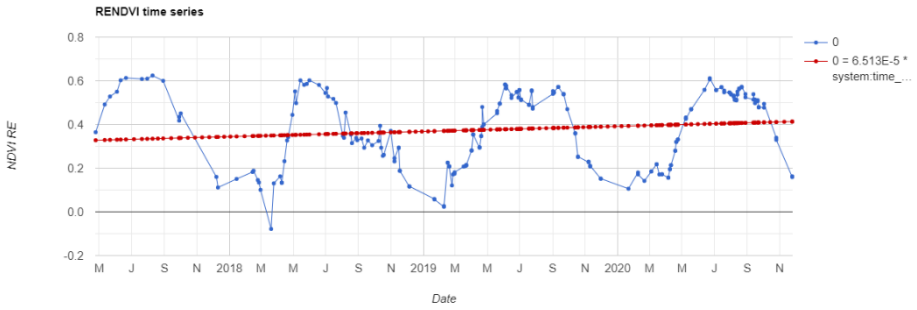


Fig. 28: The RENDVI time series of 20x20m area (Frýdlant) in 2017-2020
Source: author's work in GEE

According to the data provided by the CHMI, the leaf budburst of the hornbeam started on 12.4. for all the years monitored in this work. The leaf unfolding 10% was defined as 18.4. in both 2018 and 2019, and 2 days later, 20.4. in 2020. The leaf coloring 10% started on 5.10. in 2018, 12.10. in 2019 and 18.10. in 2020.

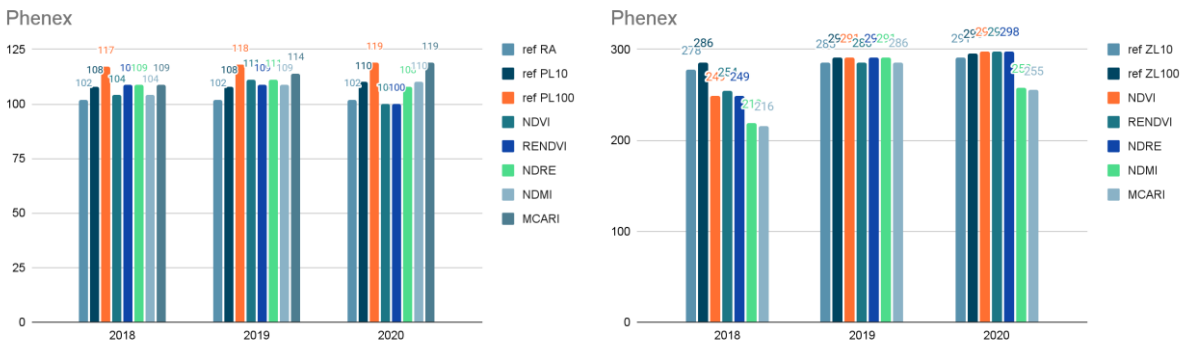


Fig. 29: The SOS and EOS of the hornbeam in the phenex package
(Y axis – DOY, ref – ground phenological data, RA – leaf budburst, PL10 – leaf unfolding 10%, PL100 – leaf unfolding 100%, ZL10 – leaf coloring 10%, ZL100 – leaf coloring 100%)
Source: CHMI, author's work

After the application of the Asymmetric Gaussian function in the phenex package, the SOS was detected the most accurately for NDVI and NDMI (14.4.), 19.4. for NDRE and MCARI, and 21.4. for RENDVI (table 7, fig. 29). The start of the leaf coloring was registered by the ground observations on 5.10. 2018. The results obtained by the remote sensing observations detected EOS a few weeks earlier than the actual start of the corresponding phenophase: NDVI was 6.9., RENDVI was 11.9., NDRE was 9.9., and NDMI and MCARI, similar to the other monitored vegetation, failed to detect EOS (EOS breakpoints for these indices were 7.8. and 4.8.). In 2019, the NDVI, RENDVI, NDRE and NDMI SOS was detected 7-9 days later than from ground measurements, for MCARI the start of the season was on 24.4. EOS for all VI was registered either on 13.10. or on 18.10., whereas the actual start of leaf coloring was on 12.10. In 2020 the SOS detected from remote sensing time series was 2 days earlier (NDVI, RENDVI), 6-8 days later (NDRE, NDMI) or 17 days later (MCARI) than the actual leaf budburst. The NDVI, RENDVI and NDRE time series registered EOS on the same date, 25.10., which was 7 days later than the start of leaf coloring monitored by the ground observations.

Therefore in the phenex package, SOS in 2018 was detected most accurately by NDVI and NDMI (2 days later than the reference date), in 2019 it RENDVI and NDMI (7 days later), and in 2020 it was NDVI and RENDVI, which detected SOS 2 days earlier. In 2018 EOS was not detected precisely with the best result of 24 days earlier by RENDVI. For 2019 and 2020, RENDVI also showed the best accuracy, detecting EOS 1 day later in 2019 and 7 days later in 2020.

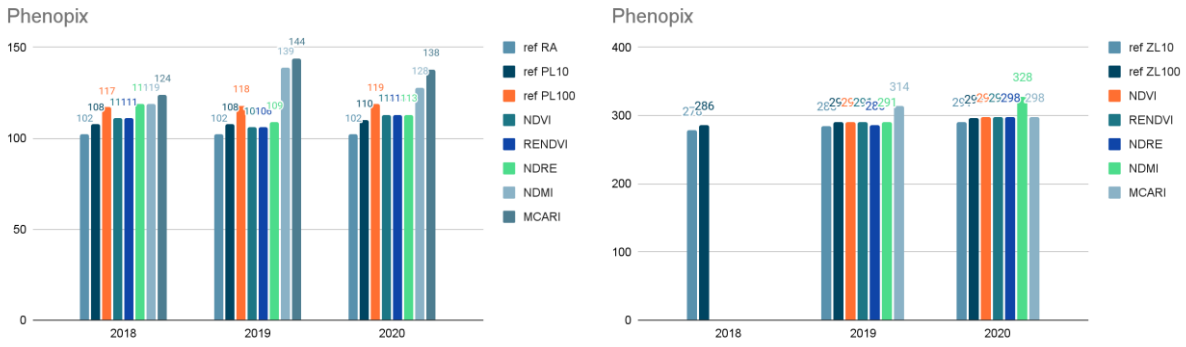


Fig. 30: The SOS and EOS of the hornbeam in the phenopix package
(Y axis – DOY, ref – ground phenological data, RA – leaf budburst, PL10 – leaf unfolding 10%, PL100 – leaf unfolding 100%, ZL10 – leaf coloring 10%, ZL100 – leaf coloring 100%)
Source: CHMI, author's work

In the phenopix package in 2018, after applying the Double Logistic function, NDVI and RENDVI time series detected SOS on 21.4., NDRE and NDMI on 29.4., MCARI on 4.5. (fig. 30). For EOS in that case, the algorithm failed to detect the date accurately: the VI curve breakpoints were in June or July. In 2019 the NDVI and RENDVI SOS was 4 days after the

monitored leaf budburst, for the NDRE 7 days later, and for NDMI and MCARI it was on 19.5. and 24.5., respectively. EOS for all vegetation indices, except MCARI, was between 13.10. and 18.10., whereas for MCARI the actual date was 7.11. In October 2019 there were only 2 observations obtained from the Sentinel-2 image collection, while there were 3 in November, where the first one was 7.11., which explains this gap between the EOS dates. In 2020, NDVI, RENDVI and NDRE defined SOS as 23.4., NDMI and MCARI - 8.5. and 18.5., respectively. EOS in 2020 for all VIs, except NDMI, was detected on 25.10., whereas for NDMI it was a month later, 24.11., which was the only available observation for this month in 2020.

To sum it up, in the phenopix package, SOS for all monitored years was detected most accurately by NDVI and RENDVI (9 days later in 2018, 4 days later in 2019 and 11 days later in 2020). But important to note, that the results obtained from the phenopix package for hornbeam corresponded better with leaf unfolding 10% (-2/+3 days from the reference date). For EOS, NDRE showed the highest accuracy detecting it 1 day later in 2019 and 7 days later in 2020.

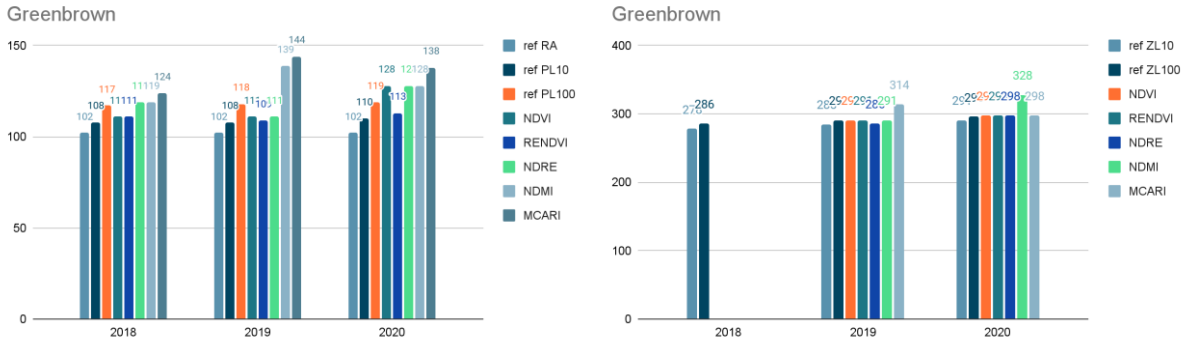


Fig. 31: The SOS and EOS of the hornbeam in the greenbrown package
(Y axis – DOY, ref – ground phenological data, RA – leaf budburst, PL10 – leaf unfolding 10%, PL100 – leaf unfolding 100%, ZL10 – leaf coloring 10%, ZL100 – leaf coloring 100%)
Source: CHMI, author's work

The functions applied in the greenbrown package showed almost identical results of SOS and EOS to the ones from the phenopix package except for SOS detection in 2020 (fig. 31). In the phenopix package, 23.4. was SOS for NDVI, RENDVI and NDRE, whereas for NDVI and NDRE in the greenbrown package it was 8.5. Similarly, the reason for these fluctuations is the gap in the availability of the observations in the most relevant month for phenophase detection – 23.4. and 8.5.

Table 7: The standard deviation of the SOS of the hornbeam in reference to the leaf budburst.

	The SOS standard deviation (leaf budburst)											
	Phenex				Phenopix				Greenbrown			
	2018	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*
NDVI	-2	-9	2	4	-9	-4	-11	8	-9	-9	-26	15
RENDVI	-7	-7	2	5	-9	-4	-11	8	-9	-7	-11	9
NDRE	-7	-9	-6	7	-17	-7	-11	12	-17	-9	-26	17
NDMI	-2	-7	-8	6	-17	-37	-26	27	-17	-37	-26	27
MCARI	-7	-12	-17	12	-22	-42	-36	33	-22	-42	-36	33

Source: author's work

Table 8: The standard deviation of the SOS of the hornbeam in reference to the leaf unfolding 10%.

	The SOS standard deviation (leaf unfolding 10%)											
	Phenex				Phenopix				Greenbrown			
	2018	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*
NDVI	4	-3	10	6	-3	2	-3	3	-3	-3	-18	8
RENDVI	-1	-1	10	4	-3	2	-3	3	-3	-1	-3	2
NDRE	-1	-3	2	2	-11	-1	-3	5	-11	-3	-18	11
NDMI	4	-1	0	2	-11	-31	-18	20	-11	-31	-18	20
MCARI	-1	-6	-9	5	-16	-36	-28	27	-16	-36	-28	27

Source: author's work

Table 9: The standard deviation of the EOS of the hornbeam in reference to the leaf coloring 10%.

	The EOS standard deviation (leaf coloring 10%)											
	Phenex				Phenopix				Greenbrown			
	2018	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*
NDVI	29	-6	-7	14		-6	-7	7		-6	-7	7
RENDVI	24	-1	-7	11		-6	-7	7		-6	-7	7
NDRE	29	-6	-7	14		-1	-7	4		-1	-7	4
NDMI	59	-6	33	33		-6	-37	22		-6	-37	22
MCARI	62	-1	36	33		-29	-7	18		-29	-7	18

Source: author's work

The SOS detected by the algorithm in the phenex package showed lower standard deviations in regards to both monitored GP phenophases. NDVI defined SOS the earliest, which was the closest to the actual leaf budburst, NDRE and NDMI detected the date the closest to the leaf unfolding phenophase. Nevertheless, all indices, except MCARI, had the standard deviation less than 7 days to both phenophases of the beginning of the season (tables 7, 8). On the contrary, for EOS it cannot be defined which package would be the most appropriate to use based on the absolute mean standard deviation, because the phenopix and greenbrown failed to detect EOS from the available observations in 2018. One can suppose that the functions used

in these packages are sensitive to the number of observations and its distribution throughout the months. Nevertheless, RENDVI, NDRE along with NDVI showed lower values of the standard deviation in all packages (table 9).

Summing up the above-mentioned description of the results, one can state that in the phenex package SOS was detected most accurately from NDVI, RENDVI and NDMI time series, EOS firmly from RENDVI. For the phenopix package, the results obtained from the packages as well as the following calculation of the absolute mean standard deviation show that RENDVI and NDVI were the most accurate in SOS detection, whereas in the greenbrown it was decisively RENDVI (tables 7-9). In regards to EOS, for both packages NDRE had the highest precision (4 days mean standard deviation).

4.2 Měděnec

The Sentinel-2 image collection over the period 2018-2020, reduced to the extent of the monitored beech area, included 492 images. After the elimination of duplicated and zero pixel values, the collection counted 115 available images in total. The distribution of the available observations per month over the period of interest is shown in table 10.

Table 10: The number of the obtained observations per each month in Měděnec.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
2018	-	6	1	6	5	1	4	4	4	3	2	1	37
2019	1	5	3	7	2	6	4	3	3	4	-	-	38
2020	2	-	4	8	3	3	1	5	6	-	3	-	35

Source: author's work

The acquired images are evenly distributed for every year, but the number of observations per month varies significantly. The minimum number of observations (1-2 images) were registered in March and June 2018, in May 2019, in July 2020, and missing images in October 2020. The maximum number of observations was recorded in April in each year. The winter months are not relevant for the phenophase detection and analysis because vegetation greenness is at the minimum level or not present at all.

Figures 32, 33 represent generated in GEE VI time series for the period of 2017-2020 with the calculated linear regression trend of the corresponding index (the graphs representing the other vegetation indices can be found in Appendix 2).

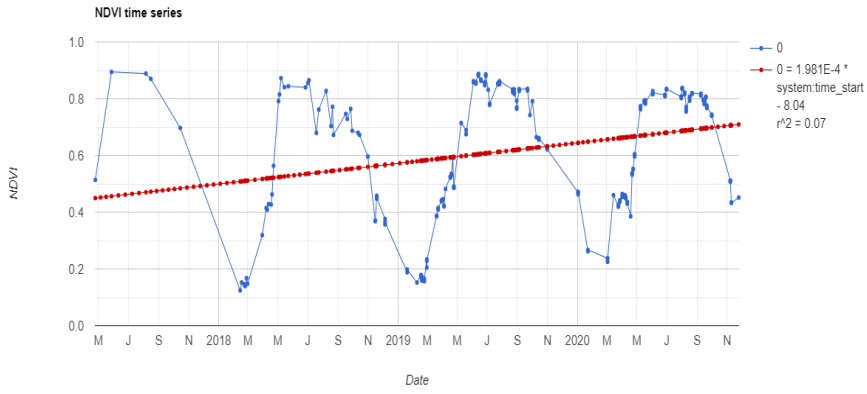


Fig. 32: The NDVI time series of 20x20m area (Měděnec) in 2017-2020
Source: author's work in GEE

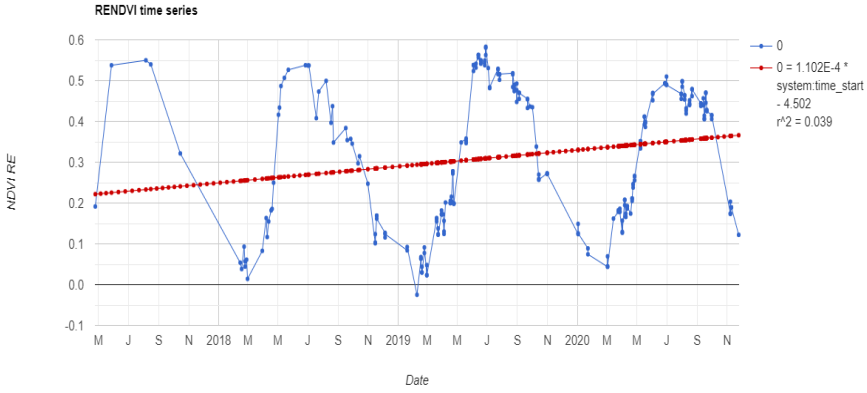


Fig. 33: The RENDVI time series of 20x20m area (Měděnec) in 2017-2020
Source: author's work in GEE

The bar charts (fig. 34-36) represent the distribution of the DOY for the GP and LSP observations. In 2018 and 2020, the start of the beech leaf budburst monitored by ground observations was on 22.4. In 2019, the reference leaf budburst started on 28.4. The leaf coloring 10% according to the data provided by the CHMI started 20.9. in 2018, 8.10. in 2019 and 28.9. in 2020.

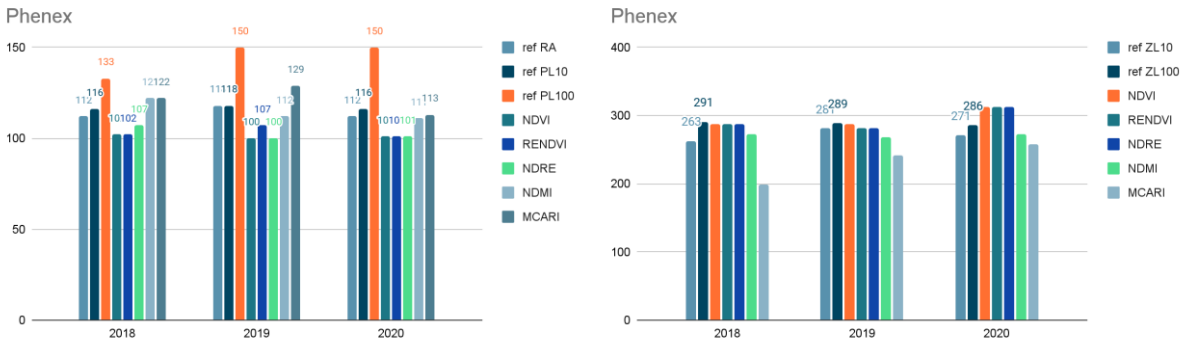


Fig. 34: The SOS and EOS of the beech in the phenex package
(Y axis – DOY, ref – ground phenological data, RA – leaf budburst, PL10 – leaf unfolding 10%, PL100 – leaf unfolding 100%, ZL10 – leaf coloring 10%, ZL100 – leaf coloring 100%)
Source: CHMI, author's work

In the phenex package (fig. 34), for NDVI and RENDVI SOS was 10 days earlier, on 12.4.; for NDRE it was on 17.4.; for both NDMI and MCARI SOS occurred on 2.5. For NDVI, RENDVI and NDMI EOS was detected on 14.10., whereas for NDMI it was 29.9. and MCARI failed at the detection of EOS with the final date 18.7. In 2019 NDVI and NDRE detected SOS as 9.4., RENDVI as 17.4.; for NDMI it was 22.4. and MCARI detected it 11 days later than the reference date. NDVI, RENDVI and NDRE detected EOS 1-6 days later than the actual start of leaf coloring 10% (8.10.). For the NDMI EOS was 26.9. and MCARI defined the actual EOS too early, on 30.8. In 2020, the NDVI, RENDVI and NDRE SOS was detected on 11.4., whereas for NDMI and MCARI it happened 1-2 days later than the actual leaf budburst. Due to the fact that in this year there were no available observations in October, the results cannot be reliable. Nevertheless, NDVI, RENDVI and NDRE detected EOS on 4.11.; for NDMI it was 30.9., and MCARI - 2 weeks earlier, 15.9.

Therefore, in the phenex package SOS for beech in 2018 was detected most accurately by NDRE (5 days earlier), in 2019 and 2020 by NDMI (6 and 1 day earlier, respectively). In regards to EOS in 2018 and 2020 NDMI results were 9 and 2 days later than start of leaf coloring, respectively, whereas in 2019 RENDVI and NDRE detected it with just 1 day difference from the reference.

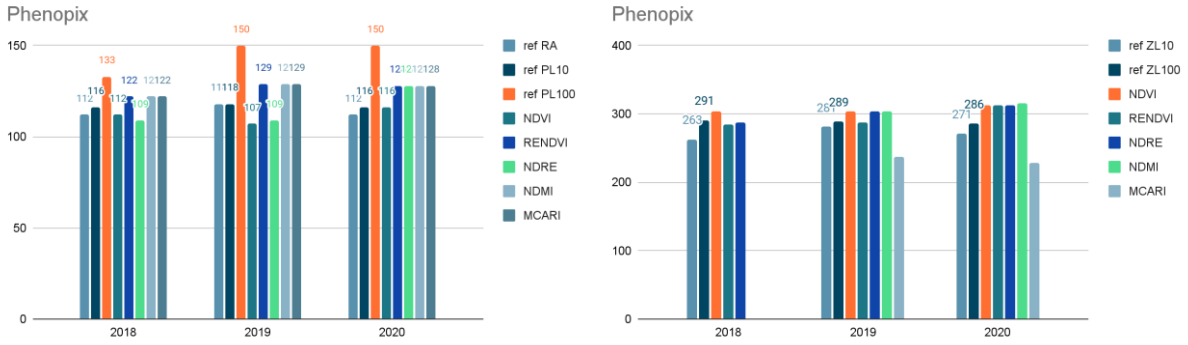


Fig. 35: The SOS and EOS of the beech in the phenopix package
(Y axis – DOY, ref – ground phenological data, RA – leaf budburst, PL10 – leaf unfolding 10%, PL100 – leaf unfolding 100%, ZL10 – leaf coloring 10%, ZL100 – leaf coloring 100%)
Source: CHMI, author's work

In the phenopix package (fig. 35), the results for 2018 are the following: the NDRE detected SOS 3 days earlier than the actual date, RENDVI, NDMI and MCARI 10 days later, whereas NDVI defined SOS precisely as the reference date. NDMI and MCARI failed to detect EOS; for NDVI it was on 31.10., for RENDVI and NDRE on 11.10. and 14.10., respectively. In 2019, NDVI and NDRE detected SOS 9-11 days earlier than the referred leaf budburst; the other vegetation indices, on the contrary, defined 11 days later, 9.5. EOS in 2019 was detected on 31.10. by NDVI, NDRE and NDMI; RENDVI defined it as 14.10., whereas MCARI was

25.8. In 2020, SOS detected by NDVI was 4 days later than the leaf budburst, 26.4. All other monitored indices defined it as 8.5. For EOS, again due to the lack of observations in October, the obtained results could not be accurate. For all indices, except MCARI, EOS was on 7.11. or 9.11., whereas for MCARI it was detected in the middle of August.

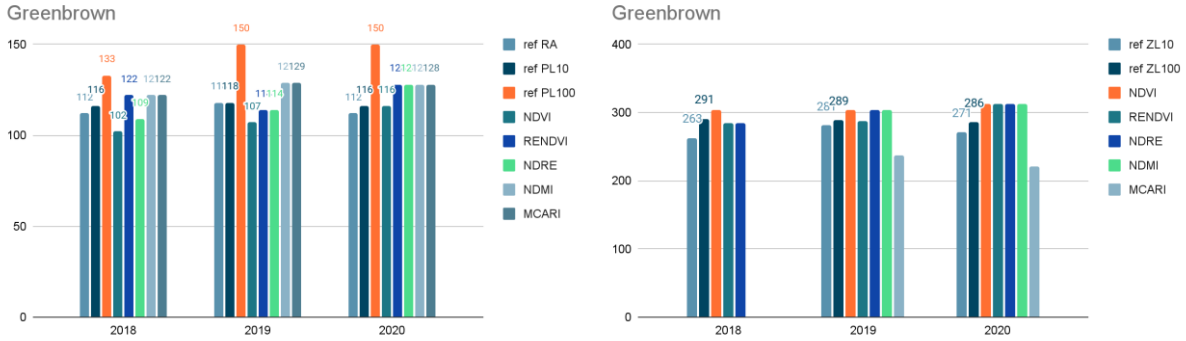


Fig. 36: The SOS and EOS of the beech in the greenbrown package
(Y axis – DOY, ref – ground phenological data, RA – leaf budburst, PL10 – leaf unfolding 10%, PL100 – leaf unfolding 100%, ZL10 – leaf coloring 10%, ZL100 – leaf coloring 100%)
Source: CHMI, author's work

The results obtained from the greenbrown package (fig. 36) were very similar to the phenopix package. In 2018, SOS was detected as follows: for NDVI on 12.4., for NDRE on 19.4., and for the other indices it was on 2.5. At the detection of EOS NDMI and MCARI failed as in the previous cases, whereas for the NDVI it was 31.10., and for the RENDVI and NDRE on 11.10. In 2019, the NDVI detected the start of the season on 17.4., RENDVI and NDRE on 24.4., NDMI and MCARI on 9.5. EOS was registered between 14.10. and 31.10. for all indices, except MCARI, which was in the middle of August. In 2020, the detection of SOS and EOS was identical to the results obtained from the phenopix package.

In the phenopix, in 2018 the most precise results were achieved by NDVI (same as the reference date), NDRE in 2019 (9 days earlier) and NDVI again in 2020 (4 days later). In the greenbrown, in its turn, NDRE showed the best accuracy (3-4 days earlier in 2018 and 2019), whereas RENDVI detected EOS relatively better than the other vegetation indices. NDRE and REDNVI showed very similar results in detecting EOS in both phenopix and greenbrown packages (table 13).

Table 11: The standard deviation of the SOS of the beech in reference to the leaf budburst.

	The SOS standard deviation (leaf budburst)											
	Phenex				Phenopix				Greenbrown			
	2018	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*
NDVI	10	18	11	13	0	11	-4	5	10	11	-4	8
RENDVI	10	11	11	11	-10	-11	-16	12	-10	4	-16	10
NDRE	5	18	11	11	3	9	-16	9	3	4	-16	8
NDMI	-10	6	1	6	-10	-11	-16	12	-10	-11	-16	12
MCARI	-10	-11	-1	7	-10	-11	-16	12	-10	-11	-16	12

Source: author's work

Table 12: The standard deviation of the SOS of the beech in reference to the leaf unfolding 10%.

	The SOS standard deviation (leaf unfolding 10%)											
	Phenex				Phenopix				Greenbrown			
	2018	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*
NDVI	14	18	15	16	4	11	0	5	14	11	0	8
RENDVI	14	11	15	13	-6	-11	-12	10	-6	4	-12	7
NDRE	9	18	15	14	7	9	-12	9	7	4	-12	8
NDMI	-6	6	5	6	-6	-11	-12	10	-6	-11	-12	10
MCARI	-6	-11	3	7	-6	-11	-12	10	-6	-11	-12	10

Source: author's work

In 2018 and 2020 the difference between the dates, when the leaf budburst and leaf unfolding 10% were registered, is 4 days, 22.4. and 26.4., respectively, whereas in 2020 both phenophases occurred on the same day (28.4.) according to the ground observations. NDVI and NDRE showed the most accurate results with the absolute mean standard deviation less than 10 days in the phenopix and greenbrown packages, whereas in the phenex package NDMI and MCARI had higher accuracy (tables 11, 12).

Table 13: The standard deviation of the EOS of the beech in reference to the leaf coloring 10%.

	The EOS standard deviation (leaf coloring 10%)											
	Phenex				Phenopix				Greenbrown			
	2018	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*
NDVI	-24	-6	-42	24	-41	-23	-42	35	-41	-23	-42	35
RENDVI	-24	-1	-42	22	-21	-6	-42	23	-21	-6	-42	23
NDRE	-24	-1	-42	22	-24	-23	-42	30	-21	-23	-42	29
NDMI	-9	12	-2	8		-23	-44	34		-23	-42	33
MCARI	64	39	13	39		44	43	44		44	50	47

Source: author's work

The leaf coloring phenophase starting dates at the Měděnec monitored area, registered by the ground observations, varied significantly from year to year: in 2018 it was 20.9., in 2019 -

8.10., in 2020 - 28.9. The most consistent and accurate results in all packages were calculated from the RENDVI time series, with the absolute mean standard deviation of 22 days (table 13).

4.3 Vranovice

The original Sentinel-2 image collection for the ash in Vranovice, filtered by the defined parameters, included 200 images. Then, the zero pixel values and duplicates were excluded resulting in 137 images available. The distribution of the available observations per month over the period of interest is shown in table 14. The total number of available observations in 2020 is 44 images, when in 2018 and 2019 it is 39 and 38 images accordingly. In 2018, the observations are evenly distributed throughout every month (it varies from 3 to 6 observations from April to October). Despite the largest total number of observations in 2020, they significantly differ from month to month, which may affect the phenophase detection. There were 10 observations in April 2020, following only 1 observation in June and October, and other months varied from 4 to 7 observations.

Table 14: The number of the obtained observations per each month in Vranovice.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
2018	-	2	1	3	6	3	5	6	6	6	1	-	39
2019	1	3	2	5	2	4	2	7	6	4	-	1	38
2020	1	1	4	10	5	1	7	5	6	1	3	-	44

Source: author's work

Figures 37, 38 represent the VI time series with linear regression trend for the monitored area of the ash tree in Vranovice over the period 2017-2020 (the graphs representing the other vegetation indices can be found in Appendix 3).

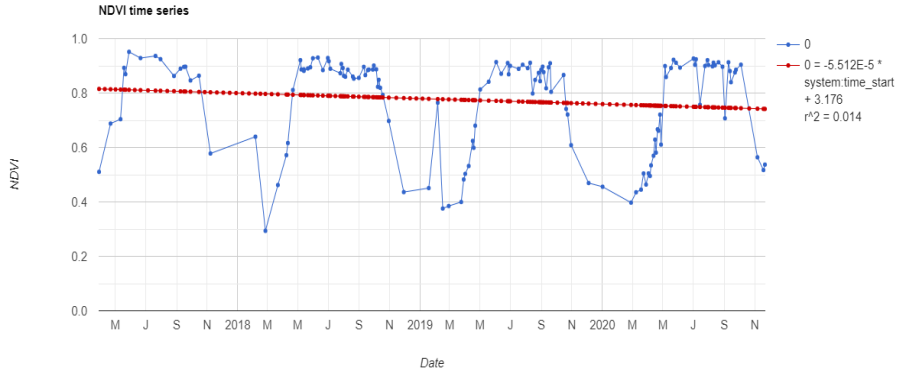


Fig. 37: The NDVI time series of 20x20m area (Vranovice) in 2017-2020
Source: author's work in GEE

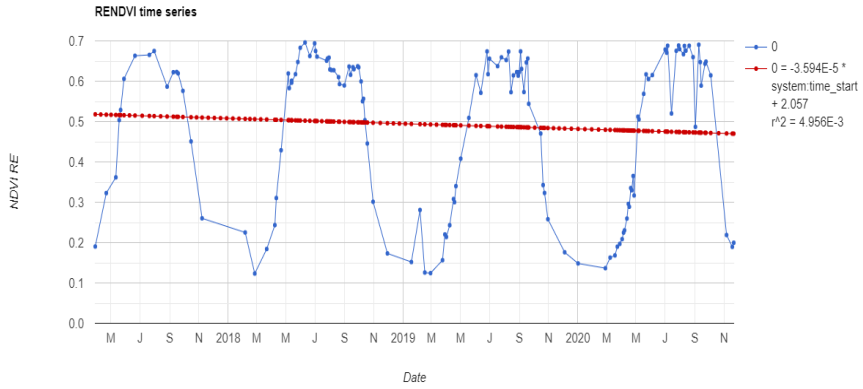


Fig. 38: The RENDVI time series of 20x20m area (Vranovice) in 2017-2020
Source: author's work in GEE

The monitored area is represented not only by the ash, but also has a thick underbrush cover, which affects the chlorophyll content in the pixel, and, as a result, the detection of SOS. The reference ground observations provided by the CHMI do not include any information on leaf coloring, because these phenophases are not monitored by phenologists on a long-term basis. The bar charts (fig. 39-41) in the Vranovice study area do not provide any observations of either the leaf unfolding 10% or the leaf coloring phenophases. The leaf budburst of the ash tree in Vranovice started on 20.4. in 2018, 1.5. in 2019, and 24.4. in 2020.

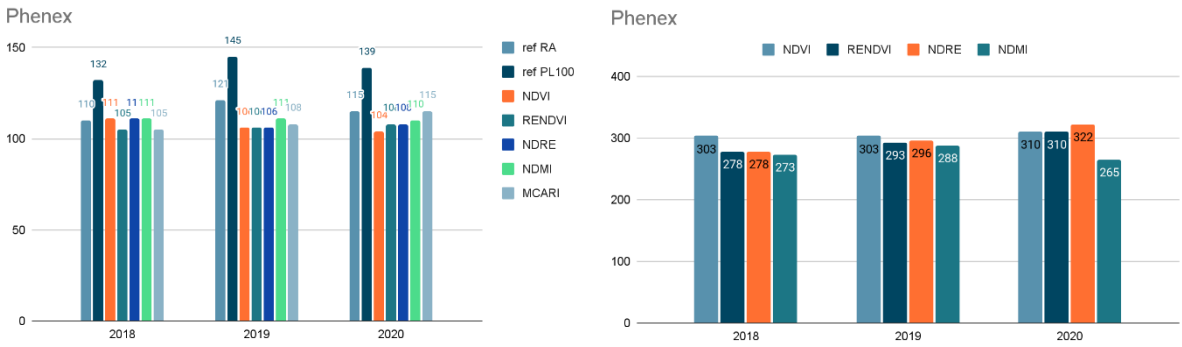


Fig. 39: The SOS and EOS of the ash in the phenex package
(Y axis – DOY, ref – ground phenological data, RA – leaf budburst, PL100 – leaf unfolding 100%)
Source: CHMI, author's work

In the phenex package in 2018 NDVI, NDRE and NDMI (fig. 39) detected SOS on 21.4., whereas RENDVI and MCARI registered 9 days earlier than the GP observation, on 11.4. In 2019 for all of the monitored VIs SOS date was registered up to 2 weeks earlier. The NDVI, RENDVI and NDRE detected it on 16.4., MCARI on 18.4. and NDMI on 21.4. In 2020 NDVI and NDRE defined it as 15.4., RENDVI 2 days later, NDMI as 20.4. and for MCARI it was 1 day later than the reference date. For the start of the leaf coloring, there is no reference information available from the ground observations. Nevertheless, the remote sensing observations reveal the following results: in 2018 the NDVI EOS was detected on 30.10.,

NDRE and RENDVI on 5.10., for the NDMI it was 22.9. The developed in R algorithm failed to detect any relevant breakpoint from the MCARI time series.

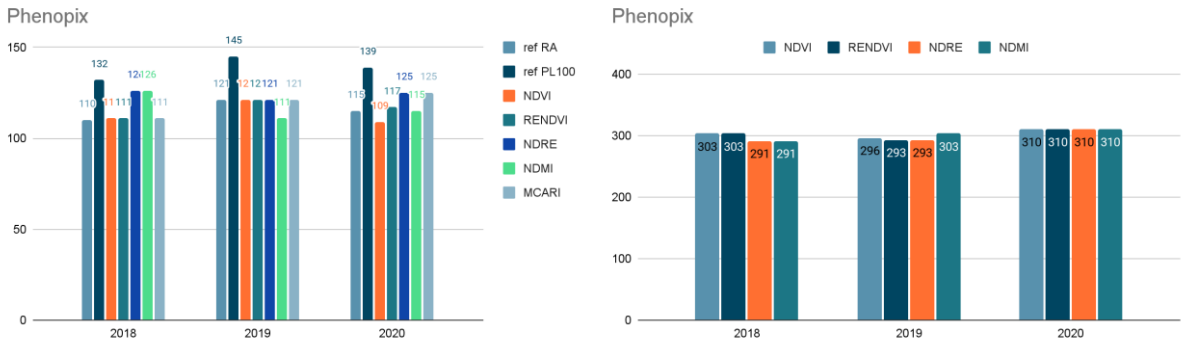


Fig. 40: The SOS and EOS of the ash in the phenopix package
(Y axis – DOY, ref – ground phenological data, RA – leaf budburst, PL100 – leaf unfolding 100%)
Source: CHMI, author's work

In the phenopix package (fig. 40) for 2018 NDVI, RENDVI and MCARI SOS started on 21.4., whereas for the NDRE and NDMI on 5.5. This gap can be explained by the fact that there are no available observations between these two dates. In 2019, the NDVI, RENDVI, NDRE and MCARI SOS coincide with the start of the leaf coloring, whereas for the NDMI it was registered 10 days earlier, on 21.4. In 2020, the NDVI, RENDVI and NDMI detected SOS with a 2-3 day difference from the GP observations. For the NDRE and MCARI SOS was registered on 5.5. As for EOS, similarly to the phenex package, MCARI failed to detect the date correctly. In 2018, the VIs registered EOS between 18.10. and 30.10., in 2019 between 20.10. and 30.10., whereas in 2020, due to the fact that there was only one available observation in this month, it was defined as 6.11. for all monitored indices.

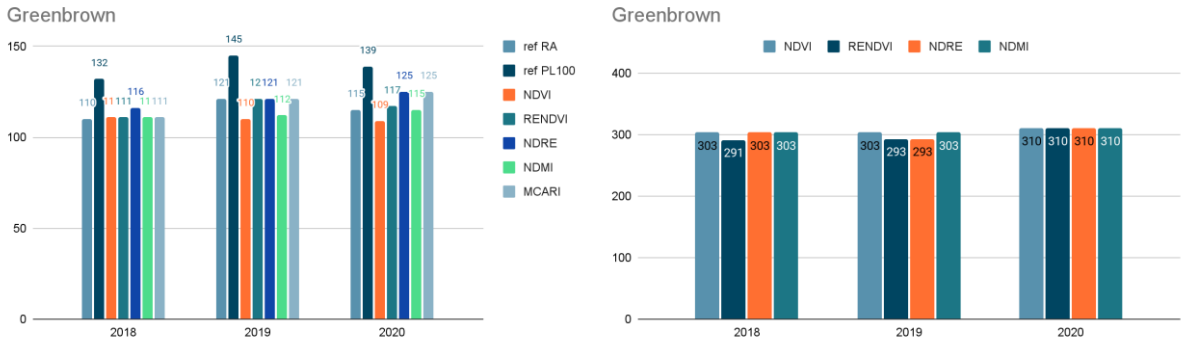


Fig. 41: The SOS and EOS of the ash in the greenbrown package
(Y axis – DOY, ref – ground phenological data, RA – leaf budburst, PL100 – leaf unfolding 100%)
Source: CHMI, author's work

The vegetation season detected by the remote sensing measurements started on 21.4. for all monitored indices, except NDRE, which defined it 6 days later (fig. 41). In 2019, the RENDVI, NDRE and MCARI SOS was equal to the reference date. The NDVI and NDMI

detected it 9 and 11 days earlier, respectively. In 2020, the results from the greenbrown package were equal to the results from the phenopix package. The EOS calculated from the greenbrown package coincide with the results from the phenopix package.

Table 15: The standard deviation of the SOS of the ash in reference to the leaf budburst.

	The SOS standard deviation (leaf budburst)											
	Phenex				Phenopix				Greenbrown			
	2018	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*
NDVI	-1	15	11	9	-1	0	6	2	-1	11	6	6
RENDVI	5	15	7	9	-1	0	-2	1	-1	0	-2	1
NDRE	-1	15	7	8	-16	0	-10	9	-6	0	-10	5
NDMI	-1	10	5	5	-16	10	0	9	-1	9	0	3
MCARI	5	13	0	6	-1	0	-10	4	-1	0	-10	4

Source: author's work

Table 16: The standard deviation of the SOS of the ash in reference to the leaf unfolding 100%.

	The SOS standard deviation (leaf unfolding 100%)											
	Phenex				Phenopix				Greenbrown			
	2018	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*
NDVI	21	39	35	32	21	24	30	25	21	35	30	29
RENDVI	27	39	31	32	21	24	22	22	21	24	22	22
NDRE	21	39	31	30	6	24	14	15	16	24	14	18
NDMI	21	34	29	28	6	34	24	21	21	33	24	26
MCARI	27	37	24	29	21	24	14	20	21	24	14	20

Source: author's work

The fact that the senescence phenophases have not been registered by the phenologists in the area of South Moravia, where the study area Vranovice is located, makes it impractical to calculate standard deviations of the acquired results with the missing reference date. Nevertheless, the results of the detected SOS calculated from the VI time series indicate high correspondence with the leaf budburst phenophase. The calculations in the phenex package in 2018 did not exceed the SOS date over 5 day from the reference date for all vegetation indices. In 2019 NDMI performed comparatively better than the other calculations (detected 10 days before leaf budburst), whereas in 2020 NDMI detected SOS 5 days earlier and MCARI defined it on the same day as the reference. In the phenopix and greenbrown packages despite quite different results for all indices, the most accurate and stable one throughout the whole monitored period was RENDVI, which detected SOS 0-2 days later than the start of leaf budburst. The lowest absolute mean standard deviation was registered for the NDVI and RENDVI results (1-2 days) in the phenopix package, for the RENDVI in the greenbrown

package (1 day), whereas in the phenex package it was for NDMI and equaled 5 days (tables 15, 16).

4.4 Chřibská

After running the GEE algorithm, there was an image collection of 184 Sentinel-2 images for the monitored area of Chřibská. Then, the image collection was reduced and the duplicates and zero values were eliminated, the final collection counted 120 images. The distribution of the available observations per month over the period of interest is shown in table 17. The total number of the obtained observations per year gradually decreased over the monitored period, from 42 in 2018 to 31 in 2020. There are no available observations in March 2018. In April, there are 5-6 observations each year, 2-4 in May, 3-4 in September and from 3 to 5 in July. In June, August and October the values differ from 2 to 7 observations per month.

Table 17: The number of the obtained observations per each month in Chřibská.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
2018	2	3	-	5	4	2	5	7	4	6	3	1	42
2019	2	6	1	5	2	6	4	2	3	5	1	-	37
2020	-	-	2	6	3	2	3	7	4	2	2	-	31

Source: author's work

The outputs acquired after applying the developed algorithm represent a VI time series with the value of corresponding vegetation index and linear trendline (fig. 42, 43) for oak tree monitoring. The graphs representing the other vegetation indices can be found in Appendix 4.

Fig. 42: The NDVI time series of 20x20m area (Chřibská) in 2017-2020
Source: author's work in GEE

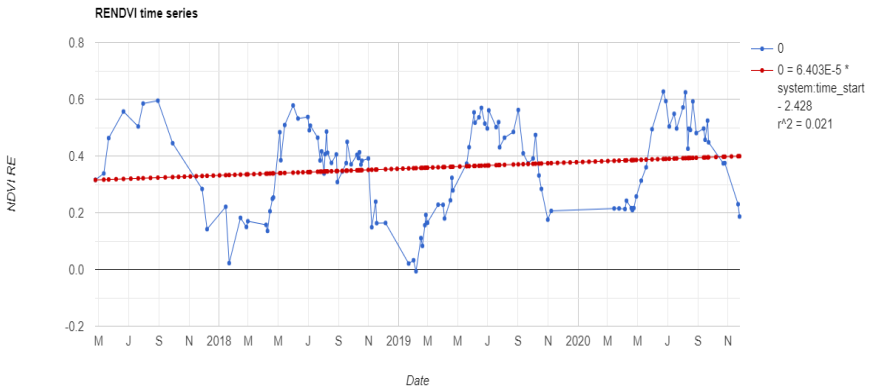


Fig. 43: The RENDVI time series of 20x20m area (Chřibská) in 2017-2020
Source: author's work in GEE

The bar charts (fig. 44-46) represent the distribution of the DOY for the GP and LSP observations. According to the GP observations provided by the CHMI, leaf budburst started 21.4. in 2018, 30.4. in 2019 and 26.4. in 2020, following leaf unfolding 10% was 2-5 days later. In regards to leaf coloring 10%, it was registered 9.10. in 2018, 10.10. in 2019, and 20.10. in 2020.

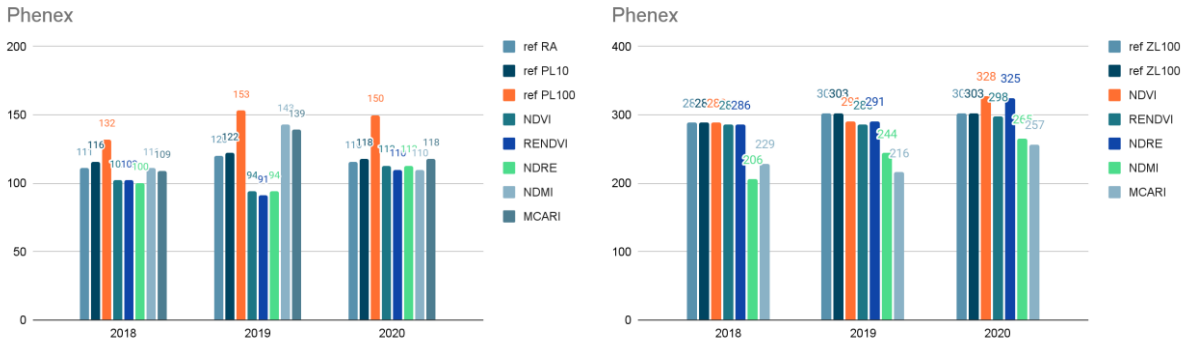


Fig. 44: The SOS and EOS of the oak detected in the phenex package
(Y axis – DOY, ref – ground phenological data, RA – leaf budburst, PL10 – leaf unfolding 10%, PL100 – leaf unfolding 100%, ZL10 – leaf coloring 10%, ZL100 – leaf coloring 100%)
Source: CHMI, author's work

The algorithm developed in the phenex package detected SOS in 2018 as follows: NDVI and RENDVI detected it 9 days earlier than the reference date (12.4.), NDRE 11 days earlier, whereas NDMI and MCARI were the most precise – 0-2 days earlier than leaf budburst. For EOS RENDVI and NDRE performed better than the other VIs (it was detected 4 days later than the actual leaf coloring 10%). In 2019 detection of SOS was significantly less accurate: NDMI and MCARI SOS were 23 and 19 days earlier, respectively (table 18). EOS in 2019 was defined by RENDVI 3 days earlier than the leaf coloring 10% date, which was the best result (fig. 44), as well as in 2020, which was 5 days earlier than the reference date. In regards to SOS in 2020, NDMI and MCARI were the most accurate ones, 6 days earlier and 2 days later than the reference date, respectively. NDMI and MCARI failed to detect EOS for all monitored years.

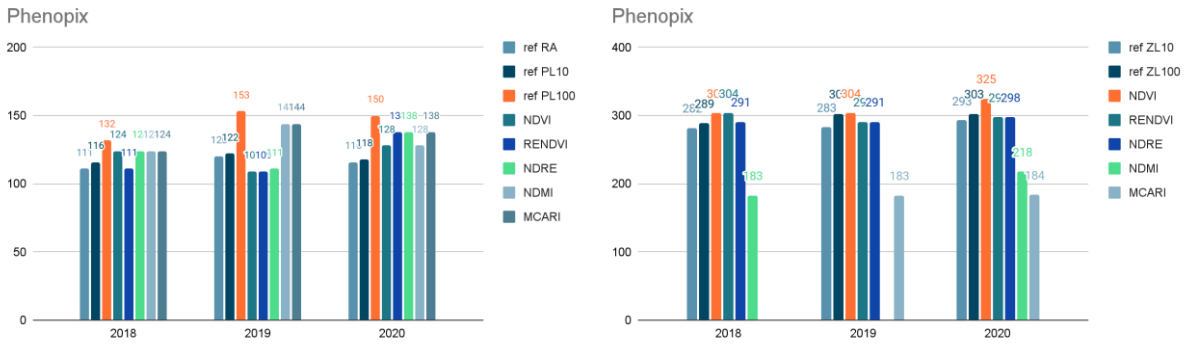


Fig. 45: The SOS and EOS of the oak detected in the phenopix package
(Y axis – DOY, ref – ground phenological data, RA – leaf budburst, PL10 – leaf unfolding 10%, PL100 – leaf unfolding 100%, ZL10 – leaf coloring 10%, ZL100 – leaf coloring 100%)
Source: CHMI, author's work

For SOS both the phenopix and greenbrown packages showed almost identical results in reference to the start of leaf budburst (fig. 45, 46). In 2018 RENDVI detected SOS on the exact date as the reference date in both packages, whereas the other indices detected it 13 days later. As for EOS, NDRE performed better in the phenopix, detecting it 9 days later than the actual leaf coloring 10% started, whereas in the greenbrown NDVI, RENDVI and NDRE performed equally inaccurate –EOS was defined 29 days later. NDMI and MCARI failed to detect EOS for all indices for all monitored years. In 2019 NDVI, RENDVI and NDRE detected SOS 9-11 days earlier than the reference date, whereas EOS was the most accurate in RENDVI and NDRE time series – 8 days later than the start of leaf coloring 10%. In 2020 the performance of NDVI and NDMI in SOS detection was the most accurate compared to the other vegetation indices – 8 days earlier than the registered start date. For EOS detection, the best results were obtained by RENDVI and NDRE – 5 days later than the start of leaf coloring 10%.

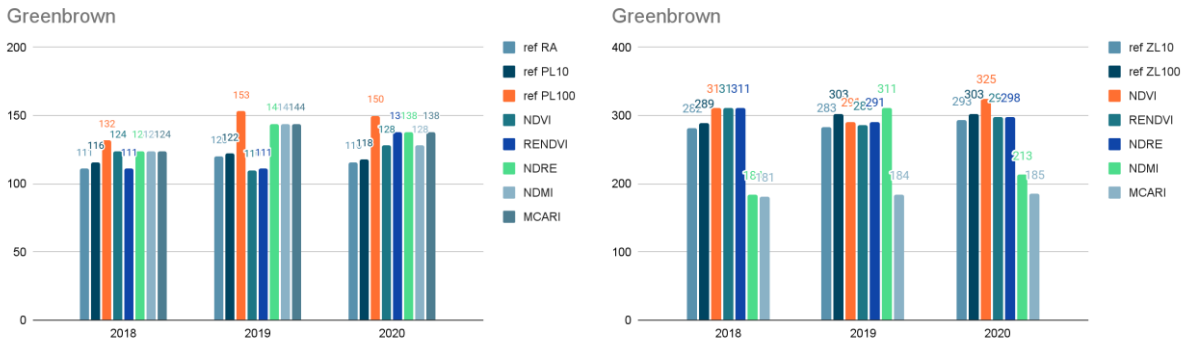


Fig. 46: The SOS and EOS of the oak detected in the greenbrown package
(Y axis – DOY, ref – ground phenological data, RA – leaf budburst, PL10 – leaf unfolding 10%, PL100 – leaf unfolding 100%, ZL10 – leaf coloring 10%, ZL100 – leaf coloring 100%)
Source: CHMI, author's work

In 2018, similar to the results obtained in the phenopix package, RENDVI in the greenbrown (fig. 46) detected SOS on the exact same date as the registered leaf budburst (table

18). In 2019 RENDVI again performed better in SOS detection compared to the other indices – 9 days earlier. As well as for EOS detection which was defined only 3 days later by the RENDVI time series. In 2020 the results were identical to the ones in the phenopix package with the best performance of NDVI and NDMI for SOS (12 days later) and RENDVI and NDRE for EOS season (5 days later).

Table 18: The standard deviation of the SOS of the oak in reference to the leaf budburst.

	The SOS standard deviation (leaf budburst)											
	Phenex				Phenopix				Greenbrown			
	2018	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*
NDVI	9	26	3	13	-13	11	-12	12	-13	10	-12	12
RENDVI	9	29	6	15	0	11	-22	11	0	9	-22	10
NDRE	11	26	3	13	-13	9	-22	15	-13	-24	-22	20
NDMI	0	-23	6	10	-13	-24	-12	16	-13	-24	-12	16
MCARI	2	-19	-2	8	-13	-24	-22	20	-13	-24	-22	20

Source: author's work

Table 19: The standard deviation of the SOS of the oak in reference to the leaf unfolding 10%.

	The SOS standard deviation (leaf unfolding 10%)											
	Phenex				Phenopix				Greenbrown			
	2018	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*Mean*
NDVI	14	28	5	16	-8	13	-10	10	-8	12	-10	10
RENDVI	14	31	8	18	5	13	-20	13	5	11	-20	12
NDRE	16	28	5	16	-8	11	-20	13	-8	-22	-20	17
NDMI	5	-21	8	11	-8	-22	-10	13	-8	-22	-10	13
MCARI	7	-17	0	8	-8	-22	-20	17	-8	-22	-20	17

Source: author's work

Table 20: The standard deviation of the EOS of the oak in reference to the leaf coloring 10%.

	The EOS standard deviation (leaf coloring 10%)											
	Phenex				Phenopix				Greenbrown			
	2018	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*Mean*
NDVI	-7	-8	-35	17	-22	-21	-32	25	-29	-8	-32	23
RENDVI	-4	-3	-5	4	-22	-8	-5	12	-29	-3	-5	12
NDRE	-4	-8	-32	15	-9	-8	-5	7	-29	-8	-5	14
NDMI	76	39	28	48	99		75	87	98	-28	80	69
MCARI	53	67	36	52		100	109	155	101	99	108	103

Source: author's work

Tables 18 and 19 represent the standard deviations of SOS of all analyzed VI time series in reference to the leaf budburst and leaf unfolding 10%. In the phenex package the best results show the NDMI and MCARI time series (absolute mean standard deviation is 8-10 days),

whereas in the phenopix and greenbrown packages this is RENDVI (absolute mean standard deviation is 10-11 days). There is not much difference observed in the standard deviations regarding the leaf budburst and leaf unfolding because the actual difference between these two phenophases is 4 days in 2018 (21.4. and 26.4.), and 2 days in 2019 (30.4. and 2.5.) and in 2020 (26.4. and 28.4.). The NDMI and MCARI time series failed to detect EOS using both fitting functions, whereas RENDVI and NDRE showed the least absolute mean standard deviation compared to the other indices (table 20).

4.5 Běleč nad Orlicí

The Sentinel-2 image collection was filtered by the extent of the monitored area and 74 images were available. In the end, to eliminate the overlapping pixels, the image collection was harmonized to have an output with one VI value per observation. The final collection size was reduced to 73 images in total. The distribution of the available observations per month over the period of interest is shown in table 21. The Běleč nad Orlicí monitored area contains 27 observations in total for 2018, 23 observations for 2019 and 16 observations for 2020. The average number of the obtained observations for each month is 2-3 images.

Table 21: The number of the obtained observations per each month in Běleč nad Orlicí.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
2018	-	2	-	4	3	3	5	3	2	2	3	-	*27*
2019	2	1	-	4	2	4	1	4	1	3	1	-	23
2020	2	1	-	3	-	1	3	3	3	-	-	-	16

Source: author's work

The outputs acquired after applying the developed algorithm in GEE represent a VI time series (fig. 47, 48). The graphs representing the other vegetation indices are in Appendix 5.

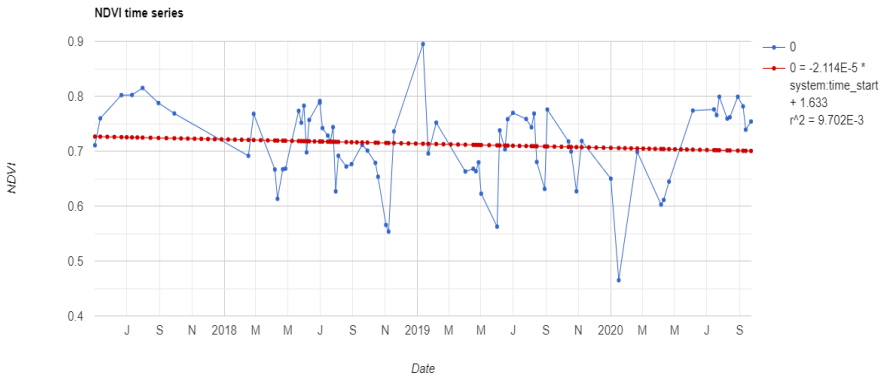


Fig. 47: The NDVI time series of 20x20m area (Běleč nad Orlicí) in 2017-2020
Source: author's work in GEE

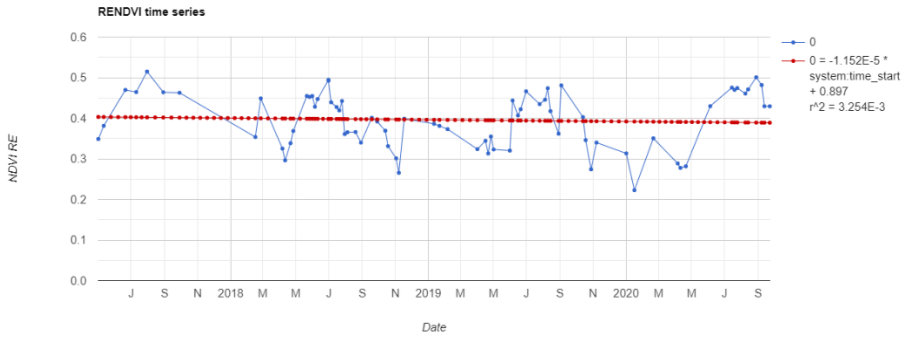


Fig. 48: The RENDVI time series of 20x20m area (Běleč nad Orlicí) in 2017-2020
Source: author's work in GEE

According to the in-situ measurements provided by the CHMI, leaf budburst started 21.4. in 2018, 28.4. in 2019 and 2020, full leaf unfolding started 25.5. in 2018, 10.6. in 2019 and 9.6. in 2020. In the studied packages in most cases, the acquired results were closer to the full leaf unfolding phenophase than to the leaf budburst. Therefore, in 2018 in the phenex package the best performance in SOS detection showed NDVI and RENDVI (6 days later).

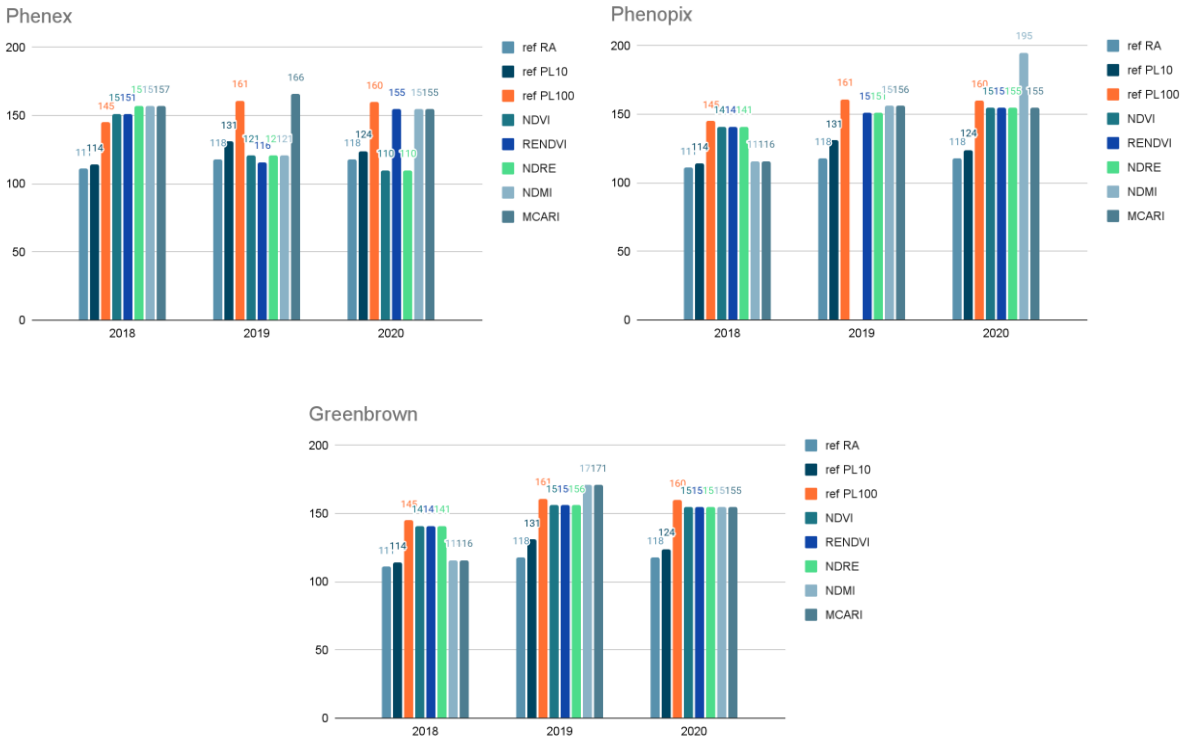


Fig. 49: The SOS of the pine detected in all packages
(Y axis – DOY, ref – ground phenological data, RA – leaf budburst, PL10 – leaf unfolding 10%, PL100 – leaf unfolding 100%)
Source: CHMI, author's work

Whereas the obtained results were closer to the leaf budburst phenophase: NDVI and NDRE detected it 3 days after the reference date and RENDVI 2 days before. In 2020, RENDVI, NDMI and MCARI detected SOS 5 days before full leaf unfolding (tables 22, 23). Similar to the phenex, the results from phenopix and greenbrown packages in 2018 were quite

different from other monitored years (fig. 49). NDMI and MCARI detected SOS 5 days after the start of leaf budburst, whereas NDVI, RENDVI and NDRE 4 days before leaf unfolding 100%. In other monitored years all of the acquired results were much closer to full leaf unfolding, among which NDVI, RENDVI and NDRE showed the best performance (absolute mean standard deviation was 5-6 days).

Table 22: The standard deviation of the SOS of the pine in reference to the leaf budburst.

	The SOS standard deviation (leaf budburst)											
	Phenex				Phenopix				Greenbrown			
	2018	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*
NDVI	-40	-3	8	17	-30	-	-37	34	-30	-38	-37	35
RENDVI	-40	2	-37	26	-30	-33	-37	33	-30	-38	-37	35
NDRE	-46	-3	8	19	-30	-33	-37	33	-30	-38	-37	35
NDMI	-46	-3	-37	29	-5	-38	-77	40	-5	-53	-37	32
MCARI	-46	-48	-37	44	-5	-38	-37	27	-5	-53	-37	32

Source: author's work

Table 23: The standard deviation of the SOS of the pine in reference to the full leaf unfolding.

	The SOS standard deviation (full leaf unfolding)											
	Phenex				Phenopix				Greenbrown			
	2018	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*
NDVI	-6	40	50	32	4	-	5	5	4	5	5	5
RENDVI	-6	45	5	19	4	10	5	6	4	5	5	5
NDRE	-12	40	50	34	4	10	5	6	4	5	5	5
NDMI	-12	40	5	19	29	5	-35	23	29	-10	5	15
MCARI	-12	-5	5	7	29	5	5	13	29	-10	5	15

Source: author's work

The SOS obtained by the remote sensing observations show higher correspondence with the full leaf unfolding phenophase than with the leaf budburst, when the chlorophyll content is lower. The fitting of the Double Logistic function to the NDVI, RENDVI and NDRE time series detected SOS with a 5-day mean standard deviation, the Asymmetric Gaussian function fitted to the MCARI time series had a 7-day mean standard deviation, whereas the NDMI in both cases showed lower accuracy.

4.6 Modrava

After filtering the Sentinel-2 image collection according to the defined parameters, 58 images were obtained. The distribution of the available observations per month over the period of interest is shown in table 24. The area of Modrava is characterized by a small total number

of acquired observations per year. The average number of acquired observations is 1-2 images per month.

Table 24: The number of the obtained observations per each month in Modrava.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
2018	-	-	3	1	1	-	2	4	2	3	-	-	16
2019	1	4	2	2	-	1	3	1	2	3	-	-	19
2020	-	1	1	2	2	1	1	5	2	1	1	-	17

Source: author's work

The outputs in figures 50, 51 represent filtered VI time series with the corresponding trend lines of the VI across the monitored years (the graphs representing the other vegetation indices can be found in Appendix 6).

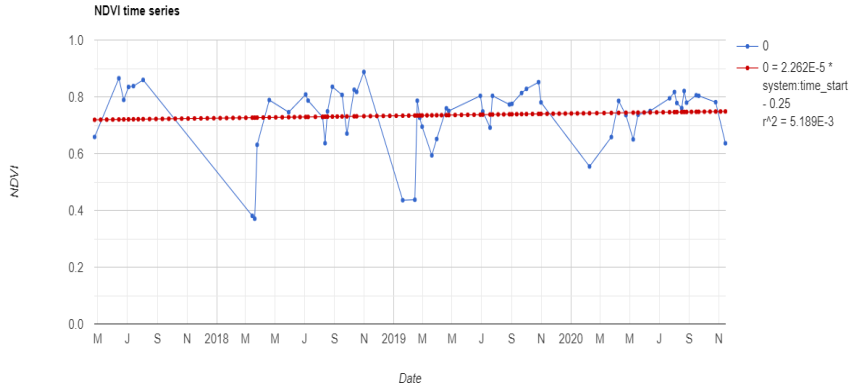


Fig. 50: The NDVI time series of 20x20m area (Modrava) in 2017-2020
Source: author's work in GEE

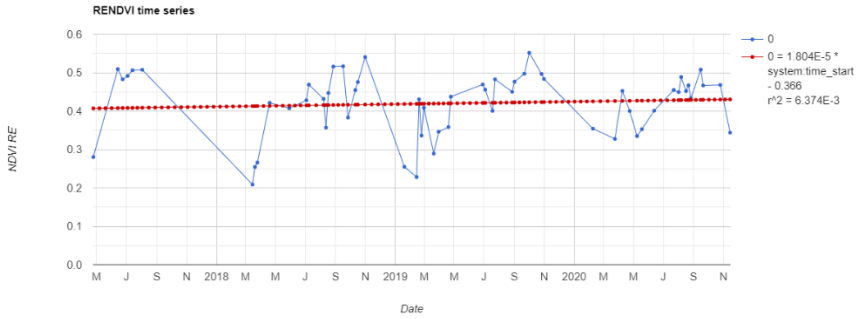


Fig. 51: The RENDVI time series of 20x20m area (Modrava) in 2017-2020
Source: author's work in GEE

At the Modrava phenological station in-situ measurements registered leaf budburst 14.5. in 2018, 2.6. in 2019 and 30.5. in 2020, leaf unfolding 100% was on 24.5. in 2018, 14.6. in 2019 and 1.7 in 2020 (fig. 52). In 2018 there were only 1-2 images per month (19.4., 29.5.), which caused a significant gap in the time series values. A similar gap was observed in 2019: 24.4. was followed only by 28.6.

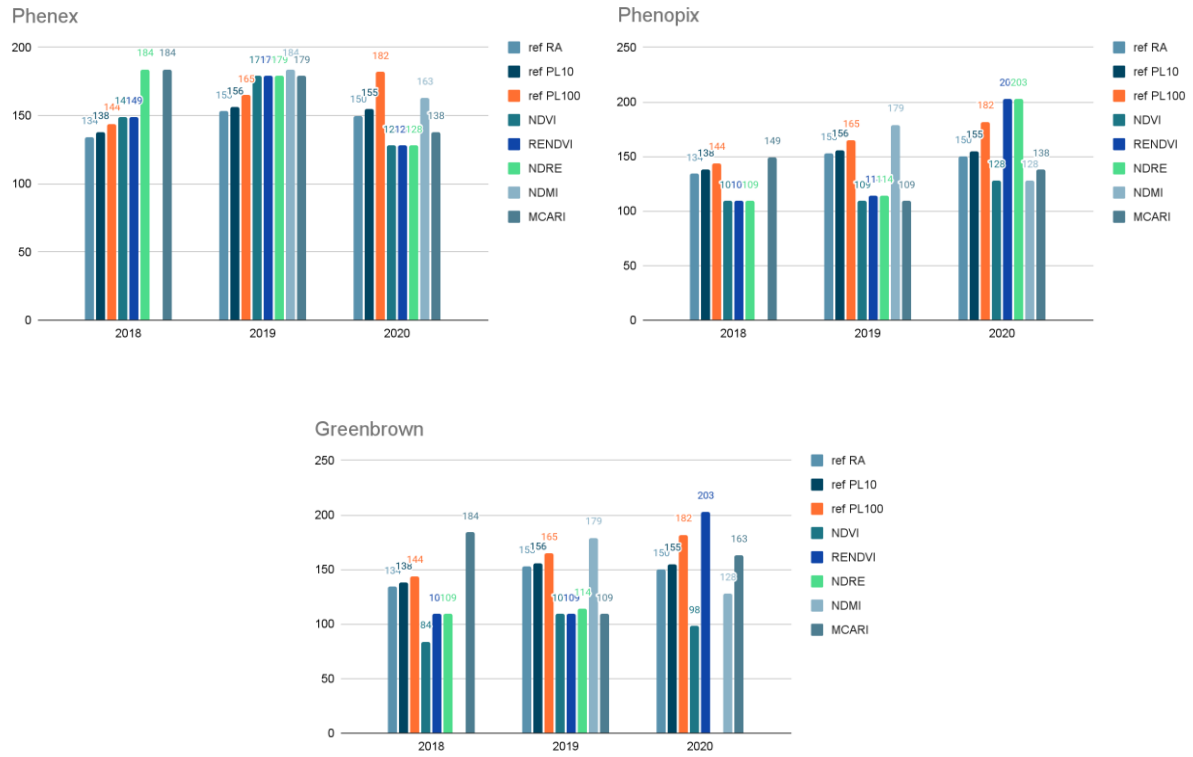


Fig. 52: The SOS of the spruce detected in all packages
(Y axis – DOY, ref – ground phenological data, RA – leaf budburst, PL10 – leaf unfolding 10%, PL100 – leaf unfolding 100%)
Source: CHMI, author's work

The most accurate results obtained after fitting the AG function to the selected VI time series in 2018 defined SOS 5 days after of the full leaf unfolding (NDVI and RENDVI). In 2019 all VIs performed quite similar detecting SOS 14-19 days later full leaf unfolding, whereas in 2020 the results were closer to leaf budburst phenophase with the best performance of NDMI 13 days after the reference GP observation (tables 25, 26). In the phenopix package SOS in 2018 was registered 25 days before leaf budburst for NDVI, RENDVI and NDRE, whereas MCARI was the closest to the full leaf unfolding (5 days after the phenophase). NDMI performed better compared to the other indices, defining SOS in 2019 14 days after the leaf unfolding 100%. In 2020 MCARI detected SOS -12/+13 days in regards to leaf budburst.

Table 25: The standard deviation of the SOS of the spruce in reference to the leaf budburst.

	The SOS standard deviation (leaf budburst)											
	Phenex				Phenopix				Greenbrown			
	2018	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*
NDVI	-15	-26	22	21	25	44	22	30	50	44	52	49
RENDVI	-15	-26	22	21	25	39	-53	39	25	44	-53	41
NDRE	-50	-26	22	33	25	39	-53	39	25	39		32
NDMI		-31	-13	22		-26	22	24		-26	22	24
MCARI	-50	-26	12	29	-15	44	12	24	-50	44	-13	36

Source: author's work

Table 26: The standard deviation of the SOS of the spruce in reference to the full leaf unfolding.

	The SOS standard deviation (full leaf unfolding)											
	Phenex				Phenopix				Greenbrown			
	2018	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*	*2018*	*2019*	*2020*	*mean*
NDVI	-5	-14	54	24	35	56	54	48	60	56	84	67
RENDVI	-5	-14	54	24	35	51	-21	36	35	56	-21	37
NDRE	-40	-14	54	36	35	51	-21	36	35	51		43
NDMI		-19	19	19		-14	54	34		-14	54	34
MCARI	-40	-14	44	33	-5	56	44	35	-40	56	19	38

Source: author's work

In the phenex package the minimum absolute mean standard deviation had NDMI with 19 days in regards to full leaf unfolding phenophase. In the phenopix, NDMI and MCARI performed identically, the minimum absolute mean standard deviation was 24 days from leaf budburst and 34-35 days from full leaf unfolding, respectively, whereas in the greenbrown only NDMI showed the same results.

4.7 Statistical assessment of VIs

To statistically assess the overall performance of the selected VIs, a paired Student's t-test was applied to the obtained data. A paired t-test is a statistical test that compares the means of two related groups to determine if there is a significant difference between the two groups, if the mean difference between pairs of measurements is 0. There are two samples: observations in one sample can be paired with observations in the other sample. To determine whether this difference is statistically significant, the p-value is compared to the significance level (Spector 2021). A significance level of 0.05 is accepted as the common value, but it can differ (like 0.01 or 0.1) depending on the data. A significance level of 0.05 indicates a 5% risk of concluding that a difference exists when there is no actual difference. In this work, the paired t-test was performed on three datasets to estimate how statistically significant the difference between the groups is. The test can be computed if the dataset has normal distribution. The first dataset (12 samples) compiled SOS dates detected from each VI time series from each station (Běleč nad Orlicí and Modrava were not included due to the lack of available observations and unreliability of the obtained results) together for each package. The table 27 represents the p-values computed by the test from this dataset in reference to the ground measurements of the leaf budburst. In the phenex package the least statistically significant difference was between NDMI and the reference value, whereas NDVI and RENDVI showed statistically significant difference from the reference. This corresponds with the acquired results for the selected study areas,

where NDMI showed the highest accuracy of SOS detection in the phenex package. This means that LSP results were very similar to the in-situ measurements of leaf budburst. On the contrary, in the phenopix and greenbrown packages, NDMI has a p-value less than 0.05, which means that the mean difference between the paired observations is significantly different from 0.

Table 27: P-values of Student's t-Test per each VI for the leaf budburst phenophase

	Phenex	*Phenopix*	*Greenbrown*
NDVI	0.01792	**0.3888**	**0.5451**
RENDVI	0.01922	0.0304	0.05502
NDRE	0.07595	0.0290	0.00275
NDMI	**0.47850**	0.00168	0.00370
MCARI	0.09811	0.00065	0.00065

Source: author's work

The next dataset (9 samples) included EOS detected by VI time series and the reference date as leaf coloring 10% phenophase from in-situ data. It contained the results only from Chřibská, Frýdlant and Měděnec phenostations because in Vranovice there are no available in-situ data for EOS. In the phenex, RENDVI samples showed the least statistically significant difference, in the phenopix and greenbrown it was NDVI. In all three packages, NDVI, RENDVI and NDRE did not show any statistically significant difference between the samples, which means that the mean difference between the paired observation was close to 0 (table 28).

Table 28: P-values of Student's t-Test per each VI for the leaf coloring 10% phenophase

	Phenex	*Phenopix*	*Greenbrown*
NDVI	0.1245	**0.7898**	**0.775**
RENDVI	**0.2756**	0.6	0.6053
NDRE	0.1613	0.605	0.6461
NDMI	0.03238	0.06843	0.1647
MCARI	0.00077	0.01927	0.0225

Source: author's work

The final dataset (21 sample), which the paired t-test was applied to, contained both the SOS and EOS dates from three locations (Chřibská, Frýdlant and Měděnec) in reference to the leaf budburst and leaf coloring 10% phenophases in order to evaluate the means of the groups for the growing season in total. NDVI, RENDVI and NDRE in three packages overall confirmed that the selected vegetation indices do not have statistically significant difference from the reference in-situ measurement for all monitored years in total.

Table 29: P-values of Student's t-Test per each VI for the growing season

	Phenex	*Phenopix*	*Greenbrown*
NDVI	**0.8754**	**0.8471**	**0.8323**
RENDVI	**0.6938**	**0.7713**	**0.7509**
NDRE	**0.7856**	**0.8191**	**0.9841**
NDMI	0.07795	0.1567	0.3153
MCARI	0.02761	0.0842	0.1086

Source: author's work

4 Discussion

The main objective of the work was to create a method/ algorithm, which generates VI time series for an arbitrary point or polygon over a certain period of time, and analyzes the time series in order to retrieve phenometric parameters, i.e. phenological phases. For these purposes, open access cloud-based platform Google Earth Engine and RStudio environment were chosen. GEE is legitimately considered one of the best tools for remote sensing and geoinformation solutions. This cloud-based platform stores vast amounts of satellite image collections, various vector datasets, what is significant advantage of the GEE compared to desktop environments. Another very relevant advantage of the GEE is that any calculations are computed very fast in contrast to the majority of desktop software. Apart from that, the platform requires the JavaScript programming language skills, or it can be set as an API Python package. These facts make the work process very flexible and convenient, when the user can create a code suitable to his own purposes or adjust the parameters. Codes as well as all datasets and image collections are stored on the server and can be reached from any computer and shared with other users. For processing phenological data, RStudio was used. The RStudio is a desktop software environment developed for the R programming language. The R is considered one of the best tools for statistical data analysis; it is often used in natural science to handle datasets of different sizes (correlation, standard deviation, regressions, means etc.). R offers elaborate sets of functions – packages, for the purposes of various types of statistical analysis and certain packages are available that allow time series analysis (noise reduction, breakpoint detection) based on certain sets of functions.

The Sentinel-2 2A and 2B satellites with their high spatial and spectral resolution and revisit time of 2-3 days in middle latitudes have been a significant source for vegetation monitoring and LSP analysis over the last four years. One of the most important advantages is the additional red-edge and SWIR bands available on the Sentinel-2 multispectral instruments, with improved phenology detection capabilities. The majority of previous studies on LSP is based on satellites with medium or coarse spatial resolution, such as MODIS, Landsat and AVHRR, or fusion of MODIS and Landsat. Other satellites, compared to Sentinel-2 missions, have several shortcomings in terms of spatial and temporal resolution: longer revisit time (8 days for Landsat), coarse spatial resolution (250 m for MODIS). Vrieling et al. (2018) pointed out that, even with the higher spatial resolution and frequent revisit time, the data availability from Landsat-8 is limited due to its single orbit and is not fine enough to capture fast-occurring changes such as cropland green-up. These facts make the vegetation monitoring very limited

over heterogeneous landscapes, not suffering from mixed or cloudy pixels. Hamunyela, Verbesselt, et al. (2013) compared the SOS estimates of European deciduous forests derived from MODIS NDVI with in-situ measurements. They found the significant correlation in the dynamics in the shifts of SOS estimates, but highlighted the importance of using sensors of high spatial and temporal resolution, especially at highly fragmented homogenous ecosystems, where significant shift might be concealed. The data Sentinel-2 might help to overcome some of these limitations.

A principal issue, the author faced in the study, was the data preprocessing. The Sentinel-2 Level 2A datasets stored in the GEE are already atmospherically corrected by running the Sen2Cor processor and are available since March 2017. Eliminating cloudy and shadow pixels is a fundamental task to retrieve clean, without noise vegetation index time series. Pixels containing clouds and cloud shadows were identified from the combination of the S2 cloud probability dataset (s2cloudless) presented in the GEE from Sentinel Hub and shadows defined by cloud projection intersection with low-reflectance NIR pixels. The code available on the GEE platform requires the set of input parameters, which values may vary depending on the geographical location and climate conditions of the study area. The most critical parameters, changes of which affected the final number of images, were the cloud probability threshold (set to 40%) and the buffer of cloud-identified objects (was set to 45m). The selections of these values were based on the visual analysis. According to the obtained images after application of the algorithm, it can be claimed that the cloud and cloudy shadow masking was successful. However, it should be noted that the algorithm does not work 100%, in some cases it is not possible to capture the edges of the shadows, which are a few meters wide, and occasionally single cirrus clouds. The influence of these factors can be eliminated by increasing the buffer zone parameter, but on the other hand, it can capture even more clean pixels, which will also impede the time series analysis.

An essential task was to choose the study areas and monitored vegetation species. The reference in-situ data were provided by the CHMI for a various number of the phenological stations monitored on a long-term basis. In order to make the validation of the acquired results in this work feasible, the areas of interest were chosen close to the monitored phenological stations. In this work, were studied 6 test sites with different vegetation species (two of which represented coniferous forests and the others were deciduous species), geographical location and climate conditions. At each phenological station there was monitored one vegetation species in 2018 – 2020. The goal was to select as homogenous pixel as possible to retrieve the information. To achieve it, the Orthophoto and CIR Orthophoto, the forest monitoring portal

(ÚHÚL), the GPS surveys executed by the author, the information from the CHMI specialists were used. The pixels of interest for the pine, spruce and ash tree monitoring were easier to define as the areas were presented by homogenous forests of the corresponding vegetation species. Whereas for the oak, hornbeam and beech the precise identification of the pixel of interest was crucial, because they were presented by a group of several single trees. For the GPS surveys the Sentinel-2 10 x 10 m grids were uploaded to the QField mobile application, the GPS coordinates of the corresponding vegetation species were measured by the Trimble GeoXH device borrowed from the department. Another important aspect, which had to be taken into consideration for the area analysis, was the presence of the underbrush or thicket vegetation, which made the pixel heterogeneous. Because of it, the VI time series values might contain the signal reflected from the vegetation species under the monitored trees. In the pine and spruce forests, no underbrush layer was registered. At the oak site, the area of interest (20 x 20 m) contained two single trees and non-photosynthetic material presented by an unpaved road, which affects the VI values when the leaf canopy is not yet dense.

The location factors affected the total number of obtained observations (images) as well. For example, in the spruce site in Modrava there was an evident lack of available images (which might be also explained by its location in mountainous region (1002 m), high average precipitation level of 1092.4 mm and therefore cloud cover), what resulted in impracticability of the algorithm application. The number of cloudless observations for the area of interest was 16-19 images per year with 1-2 per month or missing observation in April, May. These months are most important for the detection of growing season as well as September and October, when the senescence occurs. On the other hand, uneven distribution of images throughout the year also had a negative influence on the phenophase detection.

In this study, there was developed the algorithm for the detection of start and end of growing season tested on certain vegetation of tree species. One of the crucial decisions was to select an appropriate method for phenology assessment by the means of the vegetation index time series analysis and adjustment. In contrast to GP observations, the LSP phenophases represents the timings of the changes observed in the VI times series from the vegetated surface aggregated at the pixel scale. The phenological information has been modelled using fitting functions and extracted (thresholds, derivatives and moving averages methods) from time series based on satellite data. The literature research (e.g. de Beurs, Henebry 2010) pointed that the most efficient methods for the retrieval of phenological information are Asymmetric Gaussian functions, Savitzky-Golay filter, and Double Logistic functions among a wide range of other possible methods. Caparros-Santiago, J. A. et al. (2021) stated that Double Logistic function

and Asymmetric Gaussian function handled the phenophase detection better for vegetation with a marked seasonal behaviour (e.g. deciduous broadleaf forests). However, for evergreen needle leaf forests, where seasonality is less marked, Savitzky-Golay algorithm reportedly showed better results than other methods (Caparros-Santiago, J. A. et al. 2021). Another important point is whether to derive phenology metrics from raw or smoothed time series. There has been carried out many studies on phenology estimation for both smoothed and non-smoothed VI time series. The application of a smoothing algorithm might exclude important data and therefore alter the phenological results. Therefore, the problem whether to use noise removal smoothing methods and which or to work with not corrected time series values could be tested in the future research.

The vegetation indices tested in this work can be divided into 2 categories: conventional VI using green, red, NIR and SWIR bands (NDVI, NDMI) and the indices chosen to assess the potential of Sentinel-2 red-edge bands (RENDVI, NDRE, MCARI). Apart from that, at the research stage, there were tested Leaf Chlorophyll Index, Chlorophyll Index Green, Chlorophyll Index Red-Edge, Tasseled Cap Wetness. These indices are represented by combinations of spectral bands green, red, NIR and RE1; and they are used to assess the total chlorophyll content in the leaves. Nevertheless, Gitelson (2004) found that the red reflectance saturates significantly at moderate-to-high vegetation density, which therefore affects VI (e.g., NDVI) values. Using a wavelength that absorbs less, e.g. the green or the red-edge bands (as RENDVI, NDRE), could be a solution to this problem. The efficacy of red-edge bands was proved by the results of RENDVI, which represents the combination of 2 red-edge Sentinel-2 spectral bands. In majority of the study areas, RENDVI outperformed the other vegetation indices in both SOS and EOS with less than 10-day difference from the reference date. In Vranovice for ash tree it detected SOS 0-2 days from the actual leaf budburst for all monitored years. In Měděnec for beech the mean absolute standard deviation of RENDVI was 7 days. In Frýdlant, where hornbeam was monitored, the remote sensing based results were closer to leaf unfolding 10% phenophase with the best performance of RENDVI (3 days from the reference date). The vegetation indices, which consist of combination of red-edge and red spectral bands, significantly outperformed NDMI and MCARI, when fitting Double Logistic function. In regards to EOS, both RENDVI and NDRE performed better than the other indices: for oak and hornbeam mean absolute standard deviation varied 4-7 days, for beech it was 22-23 days, which was nevertheless the best achieved result.

The LSP SOS detected by NDMI and MCARI time series showed significant correspondence with the onset of the leaf budburst phenophase, when Asymmetric Gaussian

function was fitted: the absolute mean standard deviation of the VI results was 6-7 days shift for beech, 5-6 days for ash, 8-10 days for oak. Although hornbeam MCARI results were worse than the other indices (12-day shift for MCARI against 4 days in NDVI). For all deciduous species test sites, the results obtained from the phenex package imply that NDMI and MCARI are more efficient in detection of leaf budburst, whereas VI, which use red-edge and NIR spectral bands, show significantly better results in the phenopix and greenbrown packages. Overall, the ash trees test site (Vranovice) can be denoted as the most successful in the SOS detection with the least mean difference from the in-situ measurements of 1 day for RENDVI (phenopix and greenbrown). Important to note, that Vranovice has the highest mean annual air temperature among the other monitored stations, whereas Modrava, which in a mountainous region and has the lowest mean annual air temperature, showed the lowest accuracy in SOS detection.

In regards to the study areas, where coniferous species were monitored (Běleč nad Orlicí, Modrava), the remote sensing observations showed higher correspondence with full leaf unfolding phenophase than with leaf budburst, when the chlorophyll content is lower. Pines and spruces are coniferous trees, which stay green throughout the whole year. The needles can stay on trees through several winters; they fall off only due to age and new needles quickly replenish them. In evergreen vegetation systems VI is often insensitive to changes in leaf growth and canopy development displaying only major variations in the VI. This fact makes the accurate detection of phenological phases using sparse remote sensing observations quite difficult to achieve, whereas temporal resolution of Sentinel-2 has this potential. SOS detection for pine in Běleč nad Orlicí has similar tendency as SOS detection for deciduous sites: when Asymmetric Gaussian function is applied, MCARI (7 days standard deviation) performed better, whereas in Double Logistic function fit NDVI, RENDVI and NDRE performed with 5 days standard deviation from full leaf unfolding.

For remote sensing based detection of SOS for coniferous species, it is crucial to have a sufficient number of observations, especially in the most important months for vegetation growth (April, May) in order to monitor changes in chlorophyll content. This can be observed in Běleč nad Orlicí, where the number of relevant observation is 3-4 per month, and in Modrava, where there are on average one observation per month available. The best-achieved result for spruce (Modrava) was in the phenex package for NDMI with 19 days of mean standard deviation from full leaf unfolding, whereas SOS for pine was detected with 5-day standard deviation. The lack of the observations makes the detection and analysis of the individual phenological phases less accurate and reliable.

The statistical t-Test showed that NDMI had the least significant difference between the results and the input in the phenex package, whereas NDVI in the phenopix and greenbrown. The statistical test showed that NDVI, RENDVI and NDRE had significant correspondence with the onset of the leaf coloring 10% phenophase, which aligns with the actual results, where RENDVI outperformed other indices. MCARI reportedly reduces the effects from the background, however, it was found to be the most sensitive index (e.g., showing the highest difference between the dry and the wet samples). It was expected for this VI to outperform other indices in Chřibská, where the role of the background, country road, is significant. Nonetheless, the VI showed average results – 10-day mean standard deviation. Gitelson et al. (2006) proposed Chlorophyll Index Red-Edge that significantly improved the accuracy of chlorophyll content prediction and is less sensitive. For further research this index can be tested for its efficiency in phenophase detection. NDMI was unable to detect the EOS correctly because the index indicates the moisture content in leaves, i.e. water stress. As Laštovička et al. (2020) proved in their study, NDMI provides relevant information regarding the health of the forest and is useful for the detection of the forest disturbances and recovery.

The AG function fitted to the NDMI and MCARI time series showed the best results in SOS detection in regards to leaf budburst phenophase, whereas the DL function in both packages was considerably more successful when applied to the NDVI, RENDVI and NDRE time series. The red-edge spectral bands are very sensitive to medium to high levels of chlorophyll content. They are capable of penetrating the leaf better than the red band that has saturation issue in the top of canopy. Unlike NDVI, RENDVI takes into account a narrower waveband at the edge of the chlorophyll absorption. RENDVI, similarly to NDRE, showed fewer fluctuations for EOS retrieval for all monitored locations. For oak the best-achieved result by RENDVI for EOS was 4-day absolute mean standard deviation (AG function), for hornbeam it was 7-day standard deviation (DL function; NDRE had 4-day standard deviation).

An important point, which should be taken into account, is detailed information what the studied pixel contains. MCARI is known to minimize the combined effects of soil and non-photosynthetic surfaces. This was proved by the results as well, where the MCARI time series analysis showed the best result in detecting SOS (phenex package) with 8-day shift regarding the in-situ measurements. At the Měděnec and Vranovice sites, the pixel value might be mixed with the signal reflected from the underbrush layer, where its productivity exerts a strong influence on vegetation indices (Loranty et al. 2018).

For the purposes of the future research, the author of this work would suggest testing the methods in other case studies and to focus on a detailed phenology analysis of the other monitored species. Moreover, it would be useful to be compared and in the detail evaluated the results from smoothed and non-smoothed time series. For comprehensive analysis of the vegetation dynamics, the author would also suggest (if the data are available) to implement the comparison between LSP and climate factors, such as temperature, precipitation. In many studies, these factors have shown strong correlation with phenological cycle dynamics (Caparros-Santiago et al., 2021). Apart from that, to a form a clear and detailed report on correlated factors with phenology dynamics, it is essential to include other non-climatic drivers, such as soil conditions and dynamics, geographical location and others. Another potential suggestion could be to try to fill temporal gaps by applying linear interpolation (Spatial Thoughts 2021). This could be useful in cases, when the majority pixels are masked out due to the cloud cover. This method would be appropriate to test, but the researcher has to be cautious when applying it, as the interpolated new results might negatively affect the phenophase detection.

One of the main disadvantages of the proposed algorithm is that it is divided into two parts, which are developed in two different environments: GEE and R. In the following studies it would be reasonable to compile these workflows into a single one. As it was already mentioned, R provides an extensive repository of constantly updated packages with different functionality. A new package “rgee”, first public release of which was in January 2021. It is an R binding package for calling GEE API from within R (Aybar 2020). Thank to this developer’s contribution, one can integrate the data collections and functionality of GEE in R environment. However, this tool is still in a very new developing stage, and several bugs and limitations have yet been reported. Another potential approach to create a unified user-friendly algorithm is to design the AG and DL fitting functions and subsequent phenology information retrieval methods in the GEE. Currently, this approach also has its limitations. Nevertheless, one of the main objectives for future improvement of the algorithm is to visualize the obtained results over large area extents comparing several vegetation species.

5 Conclusion

This work aimed to evaluate the detection of phenological phases of vegetation based on the vegetation indices generated from the Sentinel-2 archival data in the period of 2018-2020. The main objective was to create a method/ algorithm, which generates VI time series with eliminated cloudy pixel values for any point, or polygon over certain period of time, the second part of which is computed in the R environment to retrieve relevant phenological metrics, such as SOS and EOS. The results were verified by in-situ measurements provided by the CHMI.

In this work, the author studied 6 test sites, presented by different homogeneous forest vegetation species. The VI time series analysis was performed on conventional vegetation indices, such as NDVI and NDMI, as well as the indices, which use Sentinel-2 red-edge bands (RENDVI, NDRE and MCARI). The developed algorithm consisted of two parts: VI time series generation in GEE and VI time series in R. In GEE the image collection over each area of interest in 2018-2020 was generated, the clouds and cloud shadow pixels were masked out. Then, the selected VI time series were generated and were exported in the CSV format for the following processing in R. The input data in R contained the date of the acquired observation and corresponding VI value. In the R environment, the Asymmetric Gaussian and Double Logistic functions were fitted to smoothed or non-smoothed VI time series, and 50% vegetation index ratio threshold or derivatives method were used to retrieve the LSP breakpoints, such as SOS and EOS.

The results refer that coniferous sites require a sufficient number of observation especially, as chlorophyll content dynamics in these vegetation species has less amplitude throughout the year. On the other hand, the deciduous forest test sites, SOS showed the highest correspondence with leaf budburst phenophase. The best results were obtained at Vranovice test site, where ash trees were monitored, in the phenopix package the standard deviation was 1-3 days for NDVI, RENDVI and NDRE, in the greenbrown RENDVI had 1-day shift from the mean values, NDMI referred 5-day mean absolute standard deviation. In regards to EOS, at all tested locations (oak, beech and hornbeam) RENDVI and NDRE were the most accurate vegetation indices. For oak it was 4-day mean absolute standard deviation in the phenex package for RENDVI, for hornbeam it was 4 days for NDRE and 7 days for RENDVI in the greenbrown and packages. In both these packages, Double Logistic function failed to retrieve EOS from the NDMI and MCARI time series. The main conclusion of this work supports previous findings of efficiency of the Asymmetric Gaussian function Double Logistic function for satellite-based phenology metrics extraction. Another hypothesis that vegetation indices using Sentinel-2 red-edge

spectral bands might perform better and be more reliable in the LSP phenophase detection was also proved. Therefore, the method/ algorithm proposed in this work with the consideration of the appropriate vegetation index depending on the type of vegetation species and its geographical location can be useful for phenologists in detection of phenophases in any selected area in the Czech Republic.

For the future research and improvement of the developed algorithm, the author would recommend visualizing the obtained SOS and EOS values for each pixel in regards to other adjacent homogenous vegetation cover to study the trends of the phenophase onsets and spatial relations of different species in the area. Another potential improvement future research could be focused on is more detailed time series analysis of the Sentinel-2 red-edge spectral bands. For this reason, future research could be focused on analyzing RENDVI, CIre and other vegetation indices to optimize the parameters and methods used to retrieve precise phenology metrics on homogenous vegetation cover. Moreover, it would be a good approach to evaluate the phenology of coniferous vegetation using another algorithm, e.g. Savitzky-Golay filter, which is reportedly better for coniferous vegetation time series analysis. Apart from that, the author would suggest making the developed algorithm more agile and user-friendly.

7 References

ASTOLA, H., HÄME, T., SIRRO, L., MOLINIER, M., KILPI, J. (2019): Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region. Remote Sensing of Environment, 223.

AYBAR, C. (2020): rgee: Google Earth Engine for R, https://r-spatial.github.io/rgee/index.html (10. 11. 2021).

BADECK, F. W., BONDEAU, A., BÖTTCHER, K., DOKTOR, D., LUCHT, W., SCHABER, J., SITCH, S. (2004): Responses of spring phenology to climate change. 162.

BALZTER, H., GERARD, F., GEORGE, C., WEEDON, G., GREY, W., COMBAL, B., BARTHOLOME, E., BARTALEV, S., LOS, S. (2007): Coupling of vegetation growing season anomalies and fire activity with hemispheric and regional-scale climate patterns in central and east Siberia. Journal of Climate, 15, 20, 3713-3729.

BATTEN, G. D. (1998): Plant analysis using near infrared reflectance spectroscopy: The potential and the limitations. Australian Journal of Experimental Agriculture, 7, 38.

BECK, P. S. A., ATZBERGER, C., HØGDA, K. A., JOHANSEN, B., SKIDMORE, A. K. (2006): Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. Remote Sensing of Environment, 3, 100.

BERGMEIER, E. (1998): The phenological approach in Mediterranean landscape analysis. Phytocoenologia, 2, 28.

BERRA, E. F., GAULTON, R. (2021): Remote sensing of temperate and boreal forest phenology: A review of progress, challenges and opportunities in the intercomparison of in-situ and satellite phenological metrics. Forest Ecology and Management, 480, 118663.

BOIARSKII, B. (2019): Comparison of NDVI and NDRE Indices to Detect Differences in Vegetation and Chlorophyll Content. JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES, 4, spl1.

BOORI, M. S., CHOUDHARY, K., PARINGER, R., SHARMA, A. K., KUPRIYANOV, A., CORGNE, S. (2019): Monitoring crop phenology using NDVI time series from sentinel 2 satellite data. In: 2019 5th International Conference on Frontiers of Signal Processing, ICFSP 2019.

CAPARROS-SANTIAGO, J. A., RODRIGUEZ-GALIANO, V., DASH, J. (2021): Land surface phenology as indicator of global terrestrial ecosystem dynamics: A systematic review. ISPRS Journal of Photogrammetry and Remote Sensing, 171, 330–347.

ČERVENÁ, L. (2016): Řízená klasifikace. Výukový list, https://www.natur.cuni.cz/geografie/geoinformatika-kartografie/ke-stazeni/projekty/moderni-geoinformacni-metody-ve-vyuce-gis-kartografie-a-dpz/rizena-klasifikace/view (2.12. 2020).

ČESKÝ HYDROMETEOROLOGICKÝ ÚSTAV (2019): Návod pro činnost fenologických pozorovatelů pro projekt www.fenofaze.cz. Lesní dřeviny. Praha.

COPERNICUS (2021a): Copernicus Land Monitoring, https://land.copernicus.eu/ (27. 11. 2021).

COPERNICUS (2021b): Copernicus Open Access Hub, https://scihub.copernicus.eu/ (20. 7. 2021).

ČÚZK (2021): Geoprohlížeč ČÚZK, https://ags.cuzk.cz/geoprohlizec/?k=12232 (12. 9. 2021).

DAUGHTRY, C. S. T., WALTHALL, C. L., KIM, M. S., DE COLSTOUN, E. B., MCMURTNEY, J. E. (2000): Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. Remote Sensing of Environment, 2, 74.

DE BEURS, K. M., HENEBRY, G. M. (2004): Land surface phenology, climatic variation, and institutional change: Analyzing agricultural land cover change in Kazakhstan. Remote Sensing of Environment, 4, 89.

DE BEURS, K. M., HENEBRY, G. M. (2010): Spatiotemporal statistical methods for modelling land surface phenology. In: Phenological Research: Methods for Environmental and Climate Change Analysis. Springer Netherlands, 177–208.

DELBART, N., KERGOAT, L., LE TOAN, T., LHERMITTE, J., PICARD, G. (2005): Determination of phenological dates in boreal regions using normalized difference water index. Remote Sensing of Environment, 1, 97.

EARTH OBSERVATION SYSTEM (2021): Normalized Difference Red-Edge, https://eos.com/industries/agriculture/ndre/ (02.12.2020).

ELMORE, A. J., GUINN, S. M., MINSLEY, B. J., RICHARDSON, A. D. (2012): Landscape controls on the timing of spring, autumn, and growing season length in mid-Atlantic forests. Global Change Biology, 2, 18.

ESA (2021): Sentinel-2 MSI Technical Guide, https://sentinels.copernicus.eu/web/sentinel/technicalguides/sentinel-2-msi (22.11.2021).

EVANGELIDES, C., NOBAJAS, A. (2020): Red-Edge Normalized Difference Vegetation Index (NDVI705) from Sentinel-2 imagery to assess post-fire regeneration. Remote Sensing Applications: Society and Environment, 17.

FENOLOGICKÉ FÁZE (2021): Fenologické fáze: Sledované druhy, https://www.fenofaze.cz/cz/sledovane-druhy (28.11.2021).

FILIPPA, G., CREMONESE, E., MIGLIAVACCA, M., GALVAGNO, M. (2020): phenopix: Process Digital Images of a Vegetation Cover, https://cran.r-project.org/web/packages/phenopix/index.html (29. 8. 2021).

FISHER, J. I., MUSTARD, J. F. (2007): Cross-scalar satellite phenology from ground, Landsat, and MODIS data. Remote Sensing of Environment, 3, 109.

FORKEL, M., WUTZLER, T. (2015): greenbrown - land surface phenology and trend analysis. A package for the R software. Version 2.5, https://rdrr.io/rforge/greenbrown/ (29. 8. 2021).

GAO, F., MASEK, J., SCHWALLER, M., HALL, F. (2006): On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance. IEEE Transactions on Geoscience and Remote Sensing, 8, 2207-2218.

GARCIA, M., MOUTAHIR, H., CASADY, G., BAUTISTA, S., RODRIGUEZ, F. (2019): Using hidden Markov models for land surface phenology: An evaluation across a range of land cover types in southeast Spain. Remote Sensing (Basel, Switzerland), 5, 11, 507.

GEE (2021a): Earth Engine Data Catalog: Sentinel-2, https://developers.google.com/earthengine/datasets/catalog/sentinel-2 (12. 9. 2021).

GEE (2021b): Sentinel-2 Cloud Masking with s2cloudless, https://developers.google.com/earth-engine/tutorials/community/sentinel-2-s2cloudless (20. 11. 2021).

GENG, L., CHE, T., WANG, X., WANG, H. (2019): Detecting spatiotemporal changes in vegetation with the BFAST model in the Qilian Mountain region during 2000-2017. Remote Sensing, 2, 11.

GITELSON, A. A. (2004): Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation. Journal of Plant Physiology, 2, 161.

GITELSON, A. A., KEYDAN, G. P., MERZLYAK, M. N. (2006): Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. Geophysical Research Letters, 11, 33.

HAMUNYELA, E., VERBESSELT, J., ROERINK, G., HEROLD, M. (2013): Trends in spring phenology of western European deciduous forests. Remote Sensing, 12, 5.

HAN, Q. (2013): Remote sensing-based quantification of spatial variation in canopy phenology of four dominant tree species in Europe. Journal of Applied Remote Sensing, 1, 7.

HAVLÍČEK, V. (1986): Agrometeorologie. Praha: SZN Praha.

HUETE, A., DIDAN, K., MIURA, T., RODRIGUEZ, E. P., GAO, X., FERREIRA, L. G. (2002): Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment, 1–2, 83.

JENSEN, J. R. (2007): Remote sensing of the environment: An earth resource perspective. Pearson Prentice Hall, Upper Saddle River, NJ.

JIN, S., SADER, S. A. (2005): Comparison of time series tasseled cap wetness and the normalized difference moisture index in detecting forest disturbances. Remote Sensing of Environment, 3, 94.

JÖNSSON, P., EKLUNDH, L. (2002): Seasonality extraction by function fitting to time-series of satellite sensor data. IEEE Transactions on Geoscience and Remote Sensing, 8, 40.

JÖNSSON, P., EKLUNDH, L. (2004): TIMESAT – A program for analyzing time-series of satellite sensor data. Computers and Geosciences, 8, 30.

JÖNSSON, P., CAI, Z., MELAAS, E., FRIEDL, M. A., EKLUNDH, L. (2018): A method for robust estimation of vegetation seasonality from Landsat and Sentinel-2 time series data. Remote Sensing, 4, 10.

KARLSEN, S. R., SOLHEIM, I., BECK, P. S. A., HØGDA, K. A., WIELGOLASKI, F. E., TØMMERVIK, H. (2007): Variability of the start of the growing season in Fennoscandia, 1982-2002. International Journal of Biometeorology, 6, 51.

KIM, M. S., DAUGHTRY, C. S. T., CHAPPELLE, E. W., MCMURTNEY, J. E., WALTHALL, C. L. (1994): The use of high spectral resolution bands for estimating absorbed photosynthetically active radiation (Apar). Proceedings of the 6th Symposium on Physical Measurements and Signatures in Remote Sensing.

L3 HARRIS GEOSPATIAL DOCUMENTATION CENTER. (2021): Narrowband Greenness. https://www.l3harrisgeospatial.com/docs/NarrowbandGreenness.html#red_edge_NDVI (28.11.2021).

LANGE, M., DOKTOR, D. (2017): phenex: Auxiliary Functions for Phenological Data Analysis, https://cran.r-project.org/web/packages/phenex/index.html (29. 8. 2021).

LARCHER, W. (1988): Fyziologická ekologie rostlin. Praha: Academia.

LAŠTOVIČKA, J., ŠVEC, P., PALUBA, D., KOBLIUK, N., SVOBODA, J., HLADKÝ, R., ŠTYCH, P. (2020): Sentinel-2 data in an evaluation of the impact of the disturbances on forest vegetation. Remote Sensing, 12, 12.

LÓPEZ-PUIGDOLLERS, D., MATEO-GARCÍA, G., GÓMEZ-CHOVA, L. (2021): Benchmarking deep learning models for cloud detection in landsat-8 and sentinel-2 images. Remote Sensing, 5, 13.

LORANTY, M. M., DAVYDOV, S. P., KROPP, H., ALEXANDER, H. D., MACK, M. C., NATALI, S. M., ZIMOV, N. S. (2018): Vegetation indices do not capture forest cover variation in Upland Siberian larch forests. Remote Sensing, 11, 10.

MAIN-KNORN, M., PFLUG, B., LOUIS, J., DEBAECKER, V., MÜLLER-WILM, U., GASCON, F. (2017): Sen2Cor for Sentinel-2.

MICASENSE. (2020): What is NDRE?, https://micasense.com/what-is-ndre/ (28.11.2021).

MISRA, G., BURAS, A., MENZEL, A. (2016): Effects of different methods on the comparison between Land Surface and Ground Phenology - A methodological case study from South-Western Germany. Remote Sensing, 9, 8.

MISRA, G., CAWKWELL, F., WINGLER, A. (2020): Status of phenological research using sentinel-2 data: A review. Remote Sensing, 17, 12.

MOTOHKA, T., NASAHARA, K. N., OGUMA, H., TSUCHIDA, S. (2010): Applicability of Green-Red Vegetation Index for remote sensing of vegetation phenology. Remote Sensing, 10, 2.

NOVÁK, V. (1922): Phenologická pozorování: Jejich význam a organizace. Moravský zemský výzkum. ústav zemědělský, Brno.

OTSU, K., PLA, M., DUANE, A., CARDIL, A., BROTONS, L. (2019): Estimating the threshold of detection on tree crown defoliation using vegetation indices from UAS multispectral imagery. Drones, 4, 3, 1–23.

PALUBA, D., LAŠTOVIČKA, J., MOURATIDIS, A., ŠTYCH, P. (2021): Land cover-specific local incidence angle correction: A method for time-series analysis of forest ecosystems. Remote Sensing (Basel, Switzerland), 9, 13, 1743.

PAN, Z., HUANG, J., ZHOU, Q., WANG, L., CHENG, Y., ZHANG, H., BLACKBURN, G. A., YAN, J., LIU, J. (2015): Mapping crop phenology using NDVI time-series derived from HJ-1 A/B data. International Journal of Applied Earth Observation and Geoinformation, 1, 34.

PÁNÍK, M., TRNKA. M., MOŽNÝ, M., ČERNÁ., H. (2008): Methods of compiling long-term phenological series in southern Moravia. MendelNet ’08 Agro, 153–162.

REED, B. C., BROWN, J. F., VANDERZEE, D., LOVELAND, T. R., MERCHANT, J. W., OHLEN, D. O. (1994): Measuring phenological variability from satellite imagery. Journal of Vegetation Science, 5, 5.

ROUSE, W., HAAS, R. H., DEERING, D. W. (1974): Monitoring vegetation systems in the Great Plains with ERTS, NASA SP-351. Third ERTS-1 Symposium, Vol. 1.

SCHWARTZ, M. D., REED, B. C., WHITE, M. A. (2002): Assessing satellite-derived start-of-season measures in the conterminous USA. International journal of climatology, 14, 22, 1793–1805.

SCHWARTZ, M. D., AULT, T. R., BETANCOURT, J. L. (2013): Spring onset variations and trends in the continental United States: Past and regional assessment using temperature-based indices. International Journal of Climatology, 13, 33.

SENTINEL HUB (2021): Normalized Difference Moisture Index (NDMI). Custom Scripts. A repository of custom scripts that can be used with Sentinel-Hub services. https://custom-scripts.sentinel-hub.com/sentinel-2/ndmi/ (28.11.2021).

SHABANOV, N. v., ZHOU, L., KNYAZIKHIN, Y., MYNENI, R. B., TUCKER, C. J. (2002): Analysis of interannual changes in northern vegetation activity observed in AVHRR data from 1981 to 1994. IEEE Transactions on Geoscience and Remote Sensing, 1, 40.

SPATIAL THOUGHTS (2020): Extracting Time Series using Google Earth Engine, https://spatial thoughts.com/2020/04/13/extracting-time-series-ee/ (14. 11. 2021).

SPATIAL THOUGHTS (2021): Temporal Gap-Filling with Linear Interpolation in GEE, https://spatial thoughts.com/2021/11/08/temporal-interpolation-gee/#more-6094 (27. 11. 2021).

SPECTOR, P. (2021): Berkeley Statistics: Using t-tests in R, https://statistics.berkeley.edu/computing/rt-tests (22. 10. 2021).

SUN, Y., QIN, Q., REN, H., ZHANG, T., CHEN, S. (2020): Red-Edge Band Vegetation Indices for Leaf Area Index Estimation from Sentinel-2/MSI Imagery. IEEE Transactions on Geoscience and Remote Sensing, 2, 58, 826–840.

TATEISHI, R., EBATA, M. (2004): Analysis of phenological change patterns using 1982-2000 Advanced Very High Resolution Radiometer (AVHRR) data. International Journal of Remote Sensing, 12, 25.

TIAN, F., WANG, Y., FENSHOLT, R., WANG, K., ZHANG, L., HUANG, Y. (2013): Mapping and evaluation of NDVI trends from synthetic time series obtained by blending landsat and MODIS data around a coalfield on the loess plateau. Remote Sensing, 9, 5.

TUCKER, C. J., SLAYBACK, D. A., PINZON, J. E., LOS, S. O., MYNENI, R. B., TAYLOR, M. G. (2001): Higher northern latitude normalized difference vegetation index and growing season trends from 1982 to 1999. International Journal of Biometeorology, 4, 45.

ULSIG, L., NICHOL, C., HUEMMRICH, K., LANDIS, D., MIDDLETON, E., LYAPUSTIN, A., MAMMARELLA, I., LEVULA, J., PORCAR-CASTELL, A. (2017): Detecting inter-annual variations in the phenology of evergreen conifers using long-term MODIS vegetation index time series. Remote Sensing (Basel, Switzerland), 1, 9, 49.

ÚHÚL (2021): ÚHÚL: Informace o lesním hospodářství, http://geoportal.uhul.cz/mapy/mapylhpovyst.html (27. 4. 2021).

VÁVRA, A. (2014): Fenologické mapování krajiny s podporou GIT. Olomouc. Doktorská disertační práce. Univerzita Palackého v Olomouci, Přírodovědecká fakulta, Katedra geoinformatiky.

VERBESSELT, J., HYNDMAN, R., NEWNHAM, G., CULVENOR, D. (2010): Detecting trend and seasonal changes in satellite image time series. Remote Sensing of Environment, 1, 114.

VIOVY, N., ARINO, O., BELWARD, A. S. (1992): The best index slope extraction (Bise): A method for reducing noise in NDVI time-series. International Journal of Remote Sensing, 8, 13.

VITASSE, Y., FRANÇOIS, C., DELPIERRE, N., DUFRÊNE, E., KREMER, A., CHUINE, I., DELZON, S. (2011): Assessing the effects of climate change on the phenology of European temperate trees. Agricultural and Forest Meteorology, 7, 151.

VRIELING, A., MERONI, M., DARVISHZADEH, R., SKIDMORE, A. K., WANG, T., ZURITA-MILLA, R., OOSTERBEEK, K., O'CONNOR, B., PAGANINI, M. (2018): Vegetation phenology from Sentinel-2 and field cameras for a Dutch barrier island. Remote Sensing of Environment, 215.

WALKER, J. J., DE BEURS, K. M., WYNNE, R. H. (2014): Dryland vegetation phenology across an elevation gradient in Arizona, USA, investigated with fused MODIS and landsat data. Remote Sensing of Environment, 144.

WHITE, M. A., DE BEURS, K. M., DIDAN, K., INOUYE, D. W., RICHARDSON, A. D., JENSEN, O. P., O'KEEFE, J., ZHANG, G., NEMANI, R. R., VAN LEEUWEN, W. J. D., BROWN, J. F., DE WIT, A., SCHAEPMAN, M., LIN, X., DETTINGER, M., BAILEY, A. S., KIMBALL, J., SCHWARTZ, M. D., BALDOCCHI, D. D., LEE, J. T., LAUENROTH, W. K. (2009): Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982-2006. Global Change Biology, 10, 15, 2335–2359.

WHITE, M. A., THORNTON, P. E., RUNNING, S. W. (1997): A continental phenology model for monitoring vegetation responses to interannual climatic variability. Global Biogeochemical Cycles, 2, 11.

WIKIPEDIA (2021): Moving Average, https://en.wikipedia.org/wiki/Moving_average#Cumulative_moving_average (7. 9. 2021).

WILSON, E. H., SADER, S. A. (2002): Detection of forest harvest type using multiple dates of Landsat TM imagery. Remote Sensing of Environment, 3, 80.

XIE, Q., DASH, J., HUANG, W., PENG, D., QIN, Q., MORTIMER, H., CASA, R., PIGNATTI, S., LANEVE, G., PASCUCCI, S., DONG, Y., YE, H. (2018): Vegetation Indices Combining the Red and Red-Edge Spectral Information for Leaf Area Index Retrieval. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5, 11.

XUE, J., SU, B. (2017): Significant remote sensing vegetation indices: A review of developments and applications, 1, 17.

YALE UNIVERSITY COURSES (2021): Linear Regression, http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm (29.11.2021).

ZHANG, X., GOLDERBERG, M. D., YU, Y. (2012): Prototype for monitoring and forecasting fall foliage coloration in real time from satellite data. Agricultural and Forest Meteorology, 158-159, 21-29.

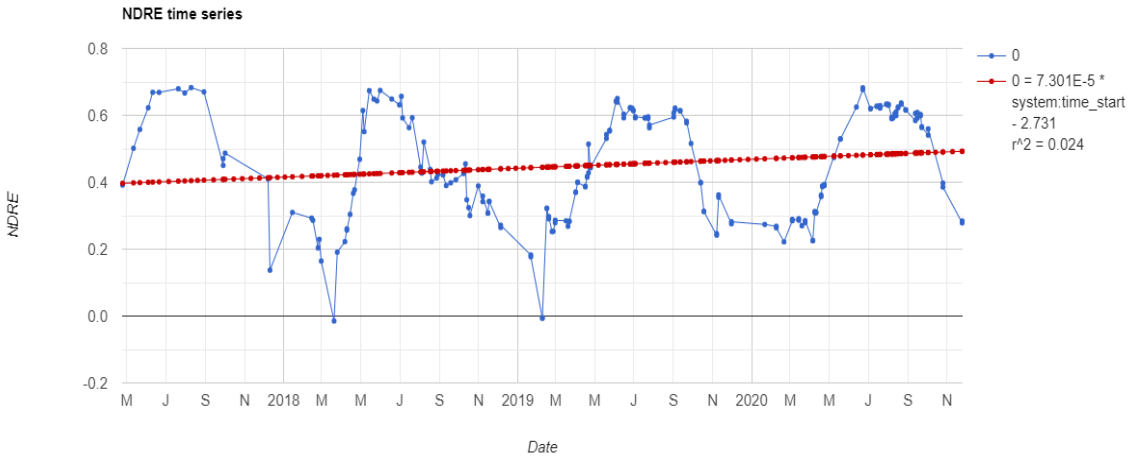
ZHANG, X., FRIEDL, M. A., SCHAAF, C. B., STRAHLER, A. H. (2004): Climate controls on vegetation phenological patterns in northern mid- and high latitudes inferred from MODIS data. Global Change Biology, 7, 10.

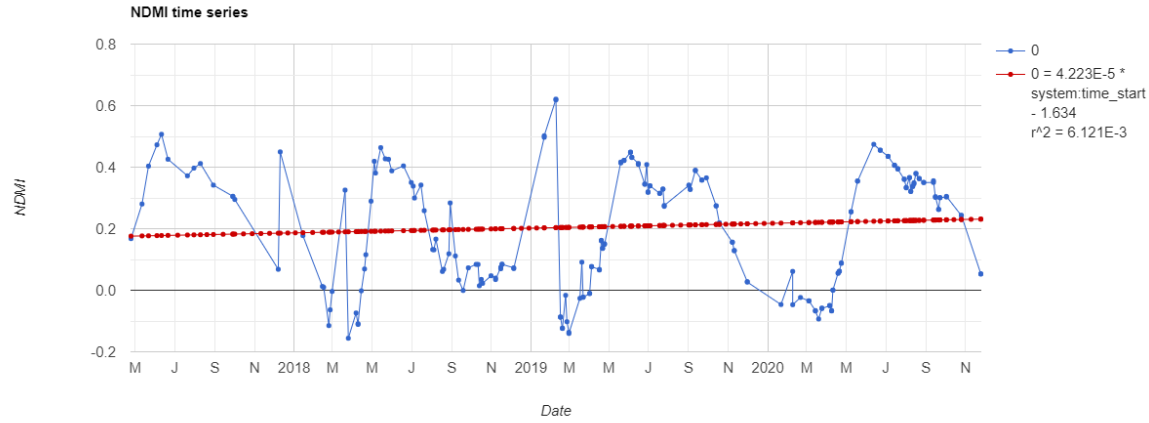
ZHENG, Y., WU, B., ZHANG, M., ZENG, H. (2016): Crop phenology detection using high spatiotemporal resolution data fused from SPOT-5 and MODIS products. Sensors (Switzerland), 12, 16, 1–21.

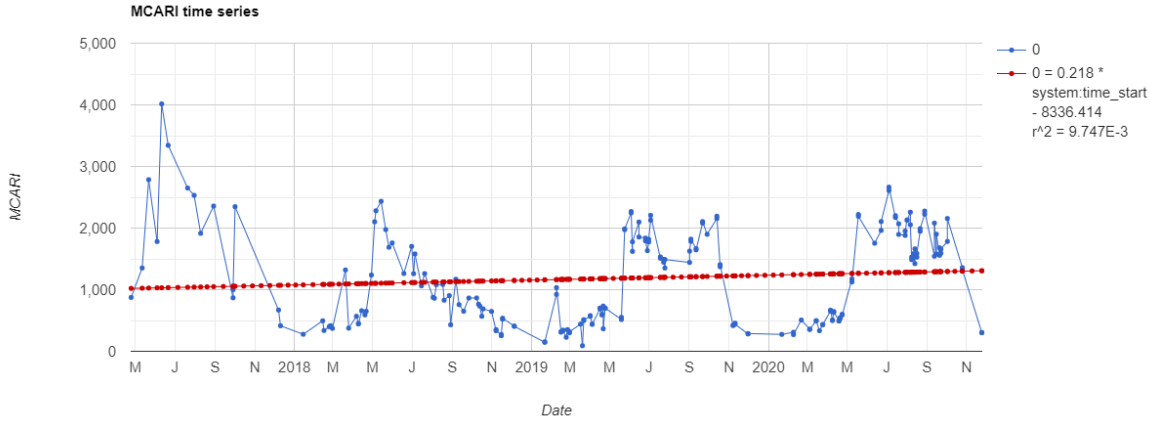
ZHU, Y., LIU, K., LIU, L., MYINT, S. W., WANG, S., LIU, H., HE, Z. (2017): Exploring the potential of world view-2 red-edge band-based vegetation indices for estimation of mangrove leaf area index with machine learning algorithms. Remote Sensing, 10, 9.

Appendices

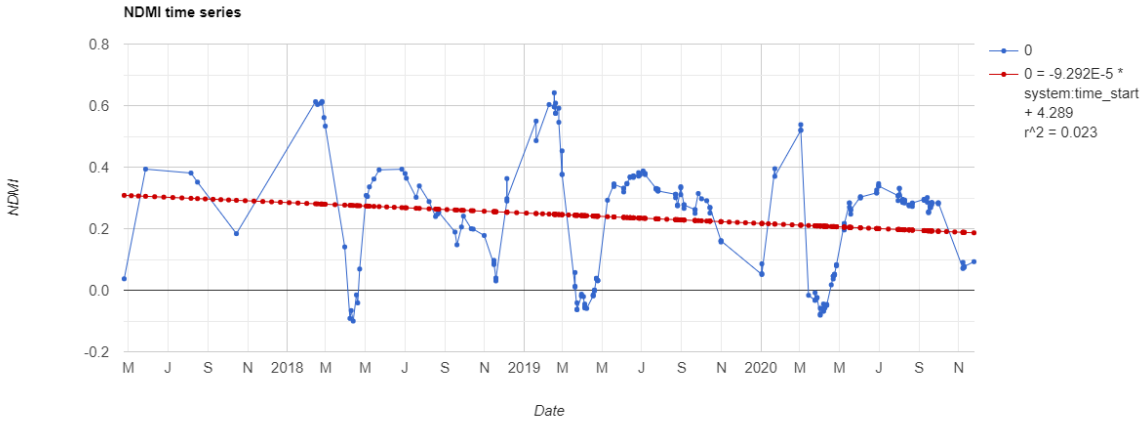
Appendix 1 VI time series for 2017-2020 in Frýdlant

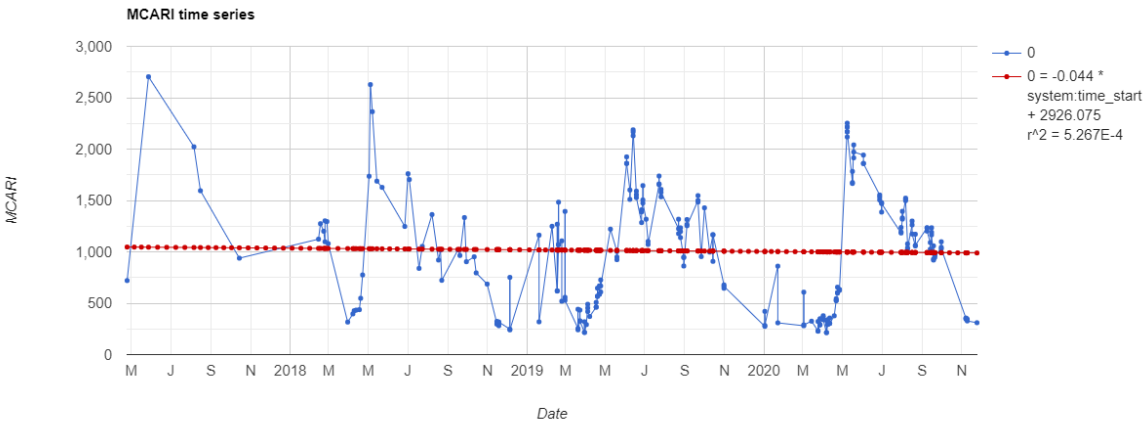




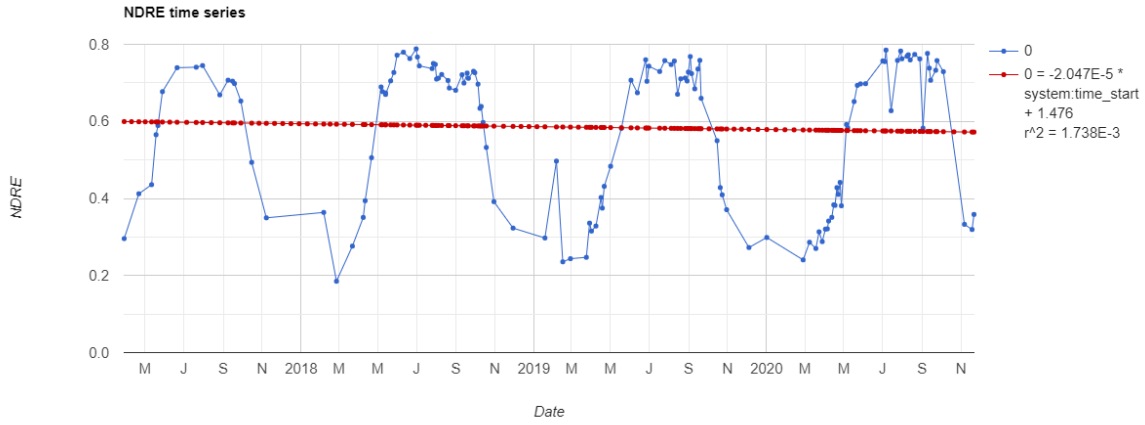


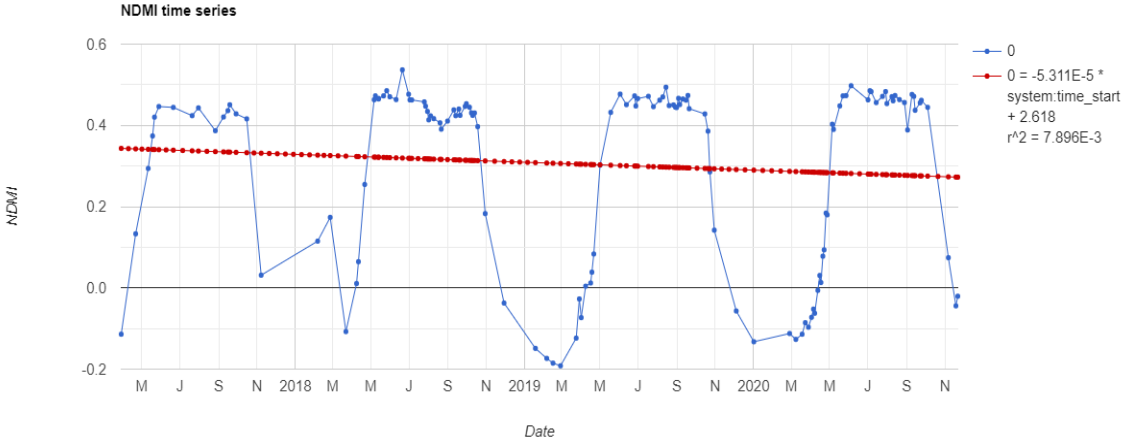
Appendix 2 VI time series for 2017-2020 in Měděnec

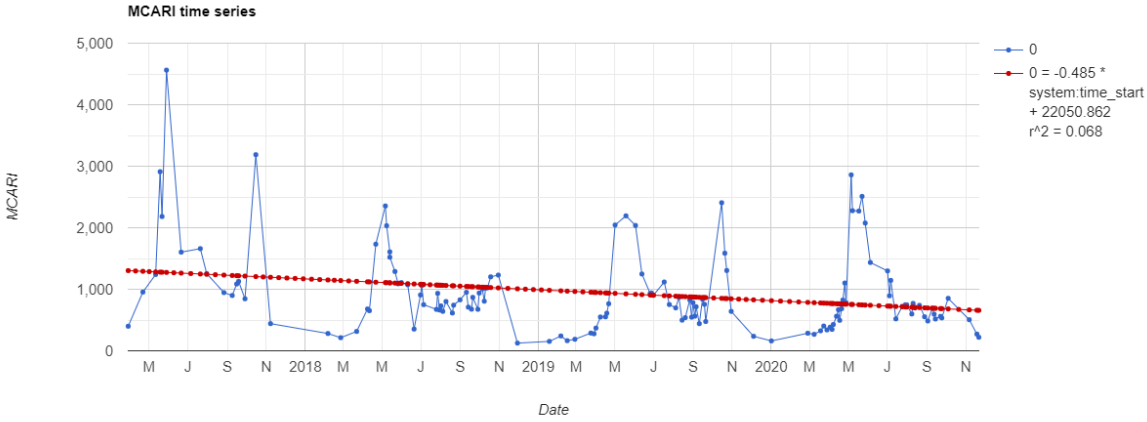




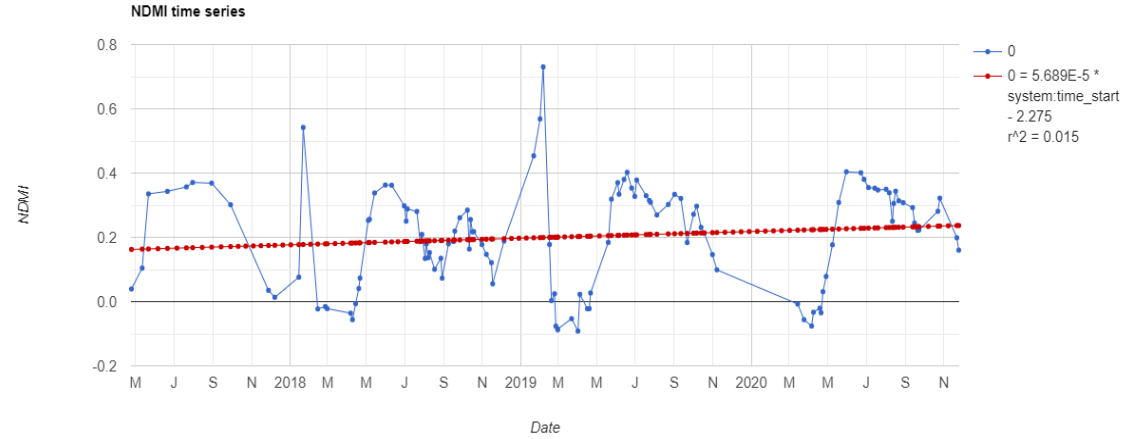
Appendix 3 VI time series for 2017-2020 in Vranovice

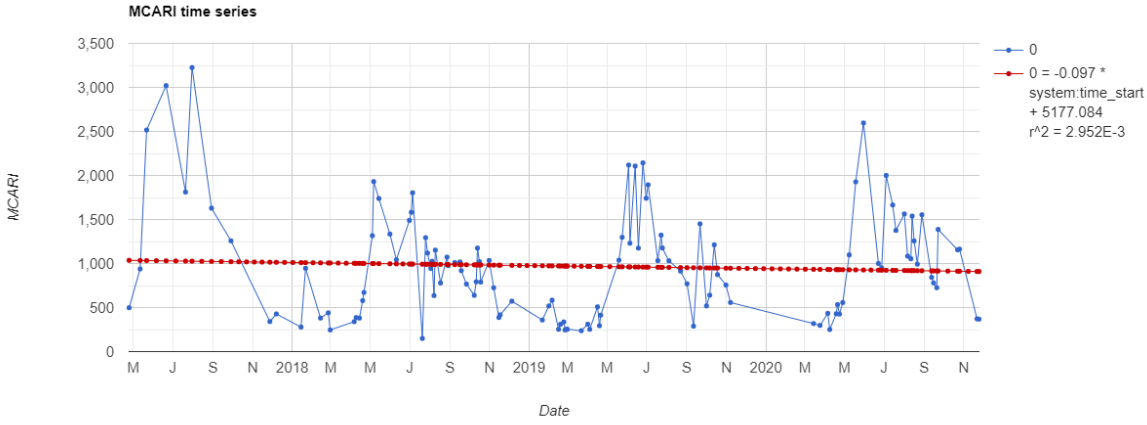




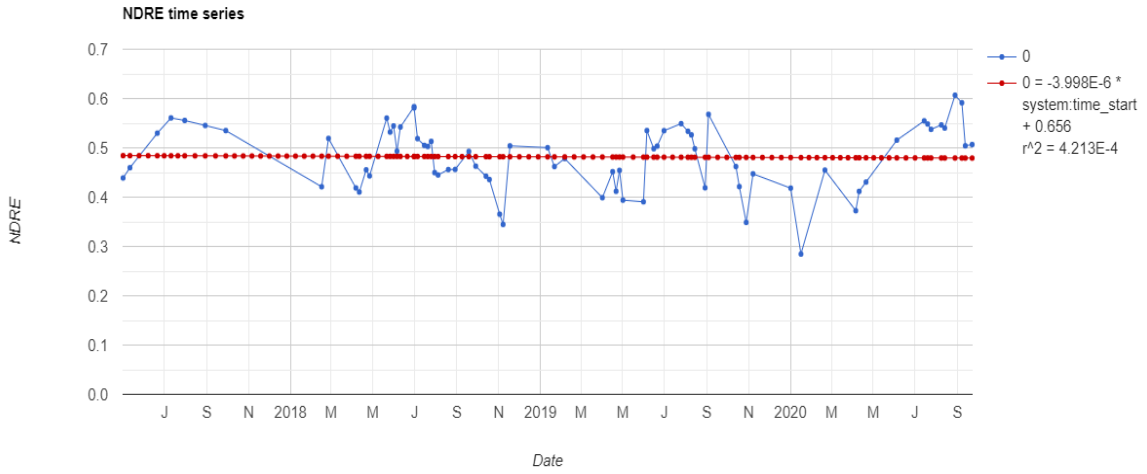


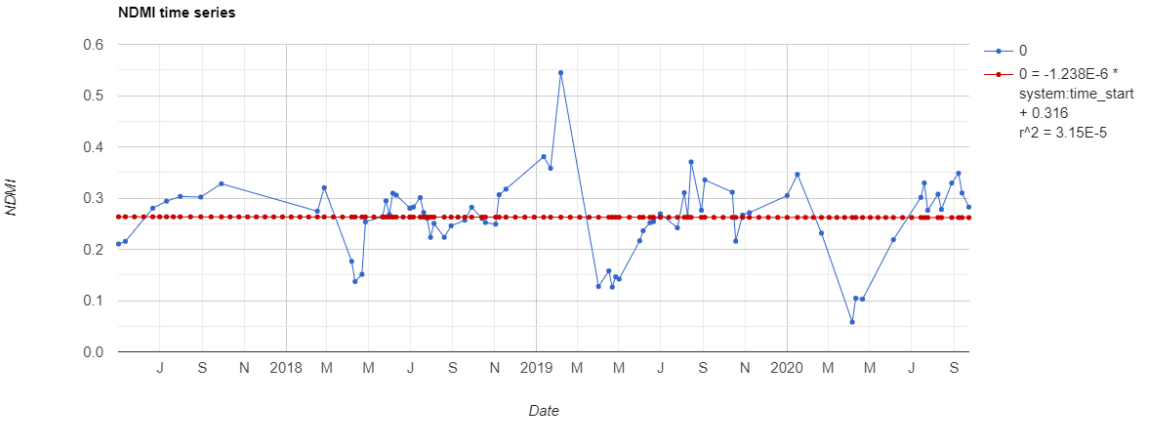
Appendix 4 VI time series for 2017-2020 in Chřibská

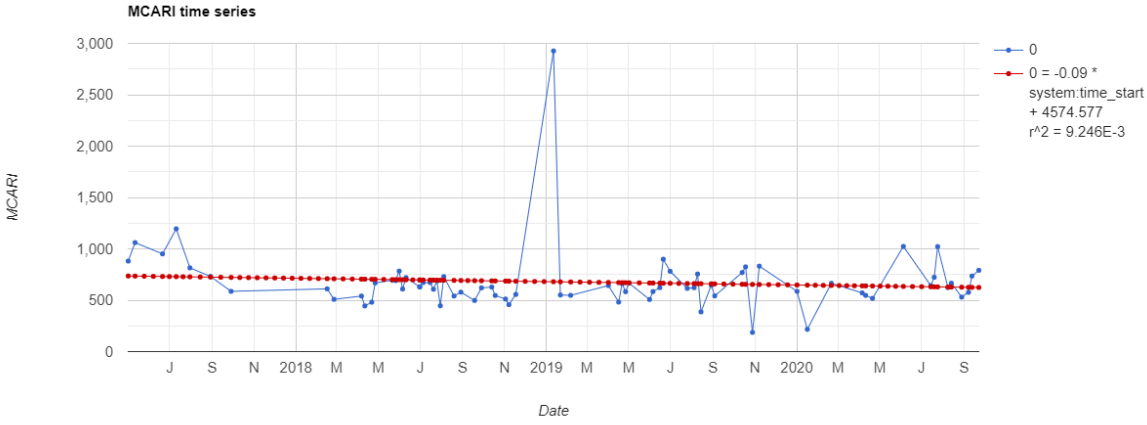




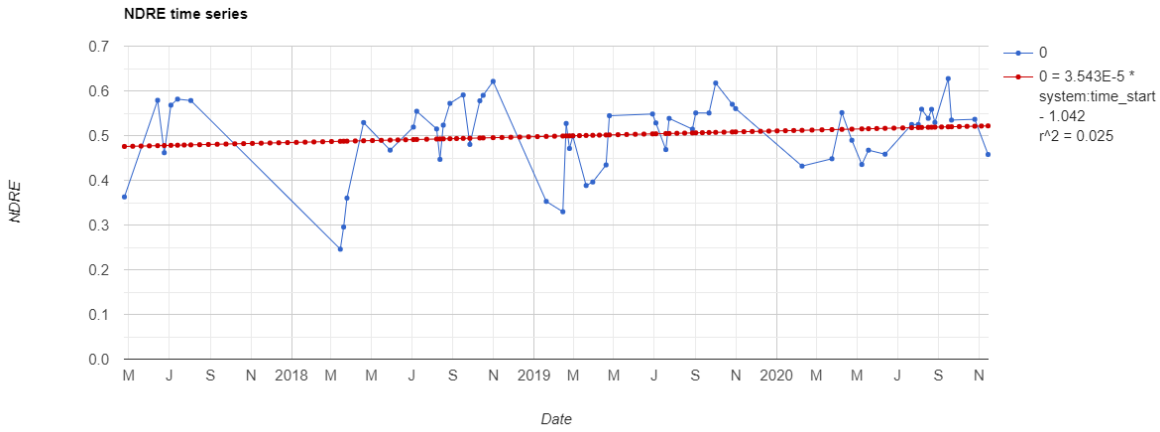
Appendix 5 VI time series for 2017-2020 in Běleč nad Orlicí

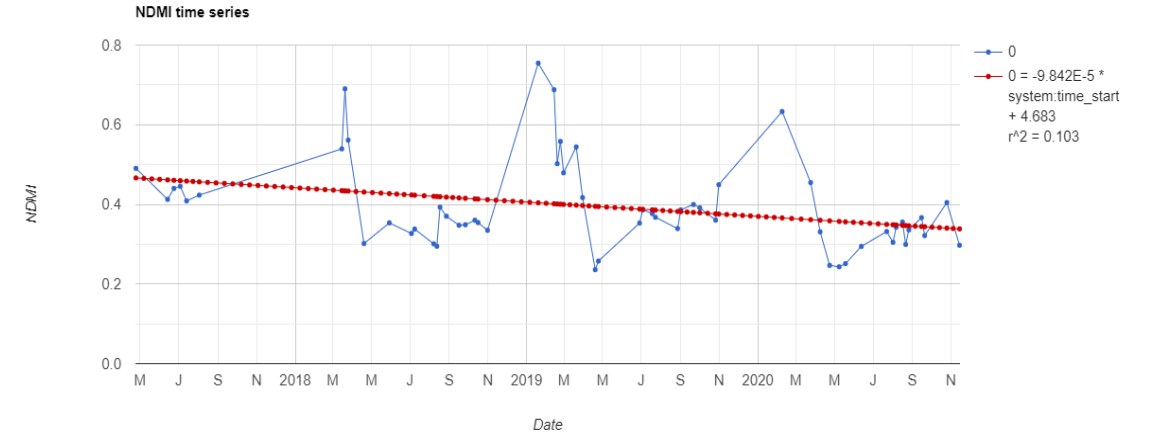


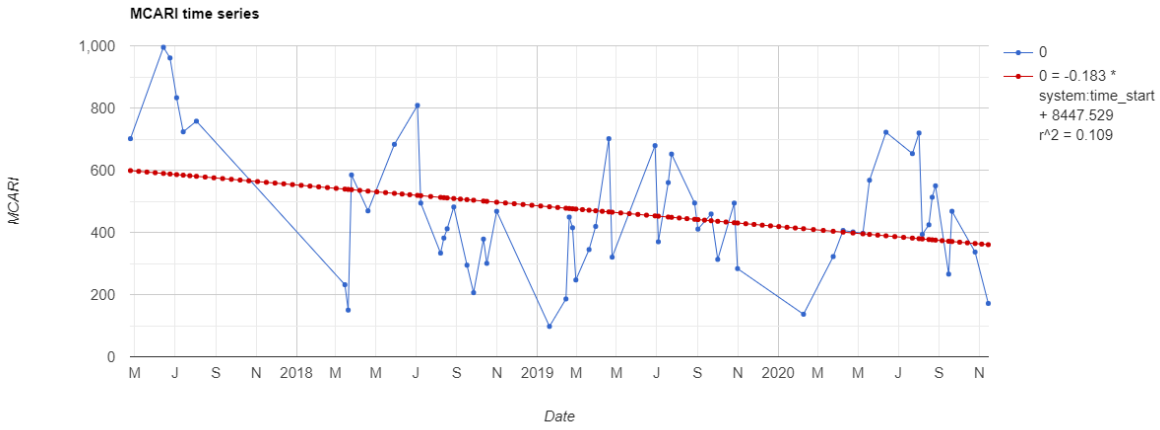




Appendix 6 VI time series for 2017-2020 in Modrava







Appendix 7 The maximum VI time series values for each year

Frýdlant					
	NDVI	**RENDVI**	**NDRE**	**NDMI**	**MCARI**
2018	2018-05-14 *0.9194*	2018-05-14 *0.6078*	2018-05-14 *0.6793*	2018-05-14 *0.4648*	2018-05-14 *2681.386*
2019	2019-06-03 *0.9001*	2019-06-03 *0.5832*	2019-06-05 *0.6504*	2019-02-08 *0.6210*	2019-06-03 *2269.153*
2020	2020-06-22 *0.9089*	2020-06-22 *0.6118*	2020-06-22 *0.6823*	2020-06-12 *0.4742*	2020-07-04 *2664.256*
Měděnec					
	NDVI	**RENDVI**	**NDRE**	**NDMI**	**MCARI**
2018	2018-05-07 *0.8767*	2018-07-03 *0.5448*	2018-07-03 *0.6338*	2018-02-23 *0.6133*	2018-05-04 *2728.171*
2019	2019-06-13 *0.8870*	2019-06-28 *0.5829*	2019-06-28 *0.6674*	2019-02-16 *0.6416*	2019-06-13 *2187.430*
2020	2020-08-01 *0.8376*	2020-06-30 *0.5098*	2020-06-30 *0.5916*	2020-03-02 *0.5380*	2020-05-08 *2252.222*
Vranovice					
	NDVI	**RENDVI**	**NDRE**	**NDMI**	**MCARI**
2018	2018-06-10 *0.9304*	2018-06-10 *0.6960*	2018-06-30 *0.7882*	2018-06-20 *0.5367*	2018-05-06 *2354.096*
2019	2019-06-02 *0.9138*	2019-09-03 *0.6739*	2019-09-03 *0.7682*	2019-08-14 *0.4939*	2019-10-15 *2406.106*
2020	2020-07-01 *0.9268*	2020-09-09 *0.6904*	2020-07-06 *0.7850*	2020-06-04 *0.4976*	2020-05-05 *2859.608*
Chřibská					
	NDVI	**RENDVI**	**NDRE**	**NDMI**	**MCARI**
2018	2018-05-31 *0.8727*	2018-05-31 *0.5781*	2018-05-31 *0.6608*	2018-01-21 *0.5421*	2018-05-06 *1930.417*
2019	2019-06-18 *0.9003*	2019-06-18 *0.5695*	2019-06-18 *0.6498*	2019-02-05 *0.7304*	2019-06-25 *2143.035*
2020	2020-08-01 *0.8884*	2020-06-22 *0.6269*	2020-08-06 *0.6756*	2020-05-30 *0.4041*	2020-05-30 *2595.817*
Běleč nad Orlicí					
	NDVI	**RENDVI**	**NDRE**	**NDMI**	**MCARI**
2018	2018-06-30 *0.7912*	2018-06-30 *0.4945*	2018-06-30 *0.5835*	2018-02-25 *0.3202*	2018-05-31 *783.430*
2019	2019-01-11 *0.8949*	2019-09-03 *0.4809*	2019-09-03 *0.5679*	2019-02-05 *0.5446*	2019-02-05 *2926.014*
2020	2020-07-24 *0.7990*	2020-08-28 *0.5007*	2020-08-28 *0.6067*	2020-09-07 *0.3483*	2020-06-04 *1023.384*
Modrava					
	NDVI	**RENDVI**	**NDRE**	**NDMI**	**MCARI**
2018	2018-10-31 *0.8876*	2018-10-31 *0.5407*	2018-10-31 *0.6212*	2018-03-20 *0.6902*	2018-07-03 *808.273*
2019	2019-10-26 *0.8516*	2019-10-01 0.5519	2019-10-01 *0.6174*	2019-01-19 *0.7546*	2019-04-19 *701.268*
2020	2020-08-21 *0.8206*	2020-09-15 *0.5080*	2020-09-15 *0.6274*	2020-02-08 *0.6331*	2020-06-12 *721.946*

Source: author's work in GEE

Appendix 8 The graphs of SOS and EOS detection in the packages

The graphs are uploaded on the CD and shared in the Google Drive here:

https://drive.google.com/file/

Graph 1: Frýdlant, phenex package
Graph 2: Frýdlant, phenopix package
Graph 3: Frýdlant, greenbrown package
Graph 4: Měděnec, phenex package
Graph 5: Měděnec, phenopix package
Graph 6: Měděnec, greenbrown package
Graph 7: Vranovice, phenex package
Graph 8: Vranovice, phenopix package
Graph 9: Vranovice, greenbrown package
Graph 10: Chřibská, phenex package
Graph 11: Chřibská, phenopix package
Graph 12: Chřibská, greenbrown package
Graph 13: Běleč nad Orlicí, phenex package
Graph 14: Běleč nad Orlicí, phenopix package
Graph 15: Běleč nad Orlicí, greenbrown package
Graph 16: Modrava, phenex package
Graph 17: Modrava, phenopix package
Graph 18: Modrava, greenbrown package

Appendix 9 The codes in GEE and R

The links to the developed codes:

Github: https://github.com/tanjased/DP_phenophase_detection.git

GEE: https://code.earthengine.google.com/