Variability of wheat phenology from Sentinel-1 and -2 time series: a case study for Brandenburg, Germany

ATIKUL HOQUE
June 2022

SUPERVISORS:
Dr. M. Schlund
Dr. A. Vrieling

Variability of wheat phenology from Sentinel-1 and -2 time series: a case study for Brandenburg, Germany

ATIKUL HOQUE
Enschede, The Netherlands, June 2022

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.

Specialization: Natural Resource Management

SUPERVISORS:
Dr. M. Schlund
Dr. A. Vrieling

THESIS ASSESSMENT BOARD:
Dr. R. Darvishzadeh Varchehi (Chair)
Dr. S. Erasmi (External Examiner, Thünen Institute of Farm Economics, Germany)

ABSTRACT

Crop phenology plays a vital role in regulating agricultural practices and managing natural resources. Accurate estimation of different phenological stages of crops can help in timely irrigation, yield prediction and fertilization. Remote sensing has proven to be a viable technique to monitor phenology over traditional field measurements. High spatial and temporal resolution of two European satellites, Sentinel-1 and -2 provide a unique opportunity to monitor phenology of crops. However, accurate estimation of crop-specific phenology is lacking at a large scale. This study estimated wheat phenology from Sentinel-1 and -2 time series for Brandenburg state of Germany for 2017 to 2021 using parcel information from Land Parcel Identification System (LPIS) data and ground truth information from the German Meteorological Station (DWD). The spatial and temporal variability of wheat phenology were explained with temperature and precipitation data. Time series of three SAR parameters from Sentienl-1 (VV, VH and VH/VV (CR)) and Enhanced Vegetation Index (EVI) from Sentinel-2 were generated for wheat fields in Google Earth Engine. LOESS and Double Logistics (DL) methods were tested to fit the data to reduce noise. As LOESS performed better in terms of representation of the temporal trajectory of wheat, LOESS was retained and used for subsequent analysis. Both CR and EVI were proven to be sensitive to wheat growing stages. Breakpoints and threshold methods were tested on LOESS fitted CR and EVI to match the DWD reported in-situ phenology phases. Third breakpoint from CR had median values of 0 to 3 days (different depending on the year) with milk ripeness phase, while the fourth breakpoint had median values of 1 to 5 days with the harvest phase. Peak of season (POP) estimated from EVI was correlated with the ears emerge phase with a median difference of 0 to 10 days. In general, the estimated dates were earlier than the field dates. For most cases, median values were less than the revisit period of both the sensors signifying the agreement between the in-situ and remote sensing estimates. However, due to outlier profiles, the RMSE values were between eight to twelve days. Coefficient of determination (R^2) between mean estimated phenology and DWD reported phenology were 0.24 for ears emerge, 0.18 for milk ripeness, and 0.24 for harvest. It is noted that the low R^2 values do not coincide with the achieved RMSE and median values. For most cases, a south-north spatial pattern was noticed where the south had earlier retrievals of phases than the north. A similar gradient was noticed in temperature, where the southern part experienced higher temperature than the northern counterpart. Interannual variability was evident from the findings where warmer temperatures led to earlier retrievals (R^2 0.24 for ears emerge, 0.02 for milk ripeness and 0.25 for harvest). Weaker positive correlations were found between precipitation and estimated stages of crops. This study demonstrates that Sentinel-1 and -2 have a high potential to monitor crop-specific phenology of wheat for a large area which can be used in crop productivity monitoring. The information can help in risk damage assessment in agricultural systems.

Keywords: crop phenology, winter wheat, Sentinel-1, Sentinel-2, time series, LOESS, breakpoint, threshold

ACKNOWLEDGEMENTS

Foremost, I would like express my sincere gratitude and thanks to both my supervisor Dr. Michael Schlund and Dr. Anton Vrieling for your continuous guidance and support throughout the MSc research period. It has been an incredible journey for me from the beginning of writing the research proposal to the final submission of the thesis, from which I learned a lot. I sincerely appreciate the assistance and advises I had from both of you during every discussions we had. Your insightful feedback and constructive criticism pushed me to coherently articulate my ideas and present them in scientific manner. I really enjoyed both of your supervision.

I would also like to acknowledge Dr. Roshanak Darvishzadeh, my research chair, for helping me with valuable suggestions and providing feedback during proposal defense and mid-term phase. I took several classes with her throughout my MSc curriculum and I really enjoyed her teaching.

My appreciation to Martina Wenzl, my internship supervisor at DLR, with whom I had lively discussions about several aspect of my research. I gained in-depth knowledge and ideas about the study area and in-situ data from her. Furthermore, it was an amazing experience working with her during the whole internship period which was fruitful for articulating my MSc research.

I would also like to thank Xinyan Fan, who helped me in several technical part of the work. She got me through difficulties in working with Google Earth Engine. I am grateful to her.

I owe humble gratitude to ITC for providing me the opportunity to pursue this MSc. I would like to extend my gratitude towards Orange Knowledge Programme (OKP) for funding my studies. Without the fund, I would not be able to pursue my MSc in the Netherlands.

I want to thank all of my friends and colleagues at ITC for their company during every class, project, discussions, fieldworks, and hangouts. I truly learned a lot and grew as a person with your company.

Finally, my sincerest thanks to my parents for their unconditional love and support throughout my whole MSc life away from home.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

API	Application Programming Interface
AVHRR	Advanced Very High Resolution Radiometer
BBCH	Biologische Bundesanstalt, Bundessortenamt und CHemische Industrie
CR	(Cross Ratio, VH/VV)
CSV	Comma-Sperated Values
DL	Double Logistics
DOY	Day of Year
DWD	Deutscher Wetterdienst
EOS	End of Season
POP	Peak of Season
EVI	Enhanced Vegetation Index
EVI2	Enhanced Vegetation Index 2
EW	Extra-Wide swath
FAO	Food and Agriculture
GEE	Google Earth Engine
GRD	Ground Range Detected
IACS	Integrated Agriculture and Control System
IDW	Inverse Distance Weighting
InSAR	Interferometric Synthetic Aperture Radar
IW	Interferometric Wide swath
LAI	Leaf Area Index
LOESS	locally estimated scatterplot smoothing
LOS	Length of Season
LPIS	Land Parcel Identification System
LSP	Land Surface Phenology
MODIS	Moderate Resolution Imaging Spectroradiometer
NDSI	Normalized Difference Snow Index
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NIR	Near-Infrared
SWIR	Short Wavelength Infrared
PPI	Plant Phenology Index
PROBA-V	PROBA Vegetation
QN	QUALITAETS_NIVEAU
RMSE	Root Mean Square Error
SAR	Synthetic Aperture Radar
S1	Sentinel-1
S2	Sentinel-2
SM	StripMap
SOS	Start of Season
SRTM	Shuttle Radar Topography Mission
VH	Vertical Send, Horizontal Receive
VI	Vegetation Indices
VIIRS	Visible Infrared Imaging Radiometer Suite
VV	Vertical Send, Vertical Receive
WV	Wave

1. INTRODUCTION

1.1. Background

Global food demand is increasing with the unprecedented growth of the world population (FAO, 2007). Agricultural production is becoming increasingly intensive to meet the demand of food. Increased food production harms the environment through emission and jeopardizes the ecosystem services (Foley et al., 2005). Hence, it is becoming increasingly important to consider the environmental impact while promoting the advances in agricultural production (Calicioglu et al., 2019). Accurate spatial and temporal information regarding crop growth and understanding their patterns and processes are of high demand for the decision-makers to regulate the agricultural practices and manage the natural resources simultaneously (Bargiel, 2017). Knowing phenological information of crops can be of great importance in this regard.

Phenology is the periodical recurring of biological events throughout the year (Lieth, 1974). Crop phenology refers to the seasonal growth stages (or phenophases), such as sowing, emergence, tillering, flowering, ripening, and harvest. Crop planting dates may change from year to year depending on the weather conditions and farming practices. Moreover, location, climate change, local weather conditions, soil quality, and management activities influence crop phenology (Gao and Zhang, 2021). The importance of crop phenology is manifold in agricultural monitoring. For example, identifying if and when emergence occurs can be a relevant first indicator of success for crop establishment and development (Sakamoto et al., 2013). Spatial information about the emergence and harvest dates can facilitate early crop mapping and crop monitoring throughout the season (Gao and Zhang, 2021). In addition, precise information about different crop phenophases is crucial for yield estimation and crop growth modelling (Yang et al., 2018). For example, the effect of water stress on crop yield depends on the phenopase during which this stress occurs (Anderson et al., 2016). For instance, extreme heat during the flowering phase results in a significant reduction in yields of maize and wheat (Lobell et al., 2011). Due to heavy rainfall and wind occurring in the later stage of wheat development, lodging can occur (Berry et al., 2003).

Precise temporal information can help timely irrigation. Furthermore, crop phenology can provide helpful information about fertilization, time of harvest, and pest management of the crops (Bouchet et al., 2016; Gao et al., 2017). Zheng et al. (2016) noted that applying fertilizers at suitable stages stimulates the leaves' nitrogen and chlorophyll content, leading to increased yield for rice. Pest infestation and diseases of plants are also essential considerations, and the growth phases of the crops largely determine the magnitude of their impacts. For instance, wheat midges attack the plants only in its flowering phase, which can significantly reduce the final production of wheat. In contrast, the yellow dwarf virus can attack barley between the booting and emergence stage (Sivertsen et al., 1999). Therefore, spatial and temporal information on the timing of crop phenophases is vital for monitoring crops.

Phenophases of plants like emergence and senescence are intimately related to the climate, as their variability is strongly related to environmental conditions (Reed et al., 1994). The frequency and intensity of extreme weather events (heatwaves, droughts, floods) resulting from climate change have increased recently in Europe (Weilnhammer et al., 2021). In 2018 for example, most parts of Europe faced extreme temperature and dry conditions while the southern part received excess rainfall (Beillouin et al., 2020; Buras et al., 2020). These instances and their growing frequency pose a massive toll on agriculture since the climate is the biggest driver of variability in agricultural production (Ray et al., 2015). The overall increased temperature and changes in precipitation have caused significant changes in the phenological cycles of crops around the world. Mo et al. (2016) noted that an increased temperature led to the advancement of heading and maturity dates, thus shortening the duration of the vegetative and the entire growing season. Rezaei et al. (2018) also found that the sowing date of winter wheat has advanced in western Germany because of the changes in mean temperature. The shifting of phenology impacts the ecosystem functions through energy balance, hydrological cycles, and carbon cycles (Richardson et al., 2010). Hence, accurate representation of crop specific phenology is needed over large geographic region which we lack currently.

Plant phenology can be captured through ground-based monitoring like frequent field visits by trained personnel or volunteers (Schwartz et al., 2012). These conventional field measurements can provide accurate information about crop phenophases. However, they are highly time-consuming, suffer from subjectivity, and might not be feasible to routinely implement, mainly to cover wider geographical regions (Richardson et al., 2018). In this case, remote sensing-based estimation of phenology is proven to be a viable alternative. Previous studies utilized multi-temporal satellite imagery to extract land surface phenology (LSP) at the landscape level (Beurs and Henebry, 2005; Delbart et al., 2008; Zhang et al., 2017). Coarse to moderate resolution optical sensors ranging from AVHRR, MODIS, and SPOT-VEGETATION have been used to estimate LSP because of their frequent temporal observations, long term records, and ready to use Vegetation Indices (VI) products (Zhang et al., 2003; Zhang and Zhang, 2016). Other coarse-resolution sensors, i.e. PROBA-V, and VIIRS were used to map phenology over a large scale and greatly enhance the understanding of vegetation phenology and how it responds to global change (Guzmán Q. et al., 2019; Zhang et al., 2018a, 2018b). However, based on their coarse spatial resolutions (from 250 m to 1 km), the sensors provide mixed signals because mostly different plant species occur together in one pixel, especially in heterogeneous landscapes. Given that agricultural fields are often smaller in size compared to the spatial resolution, LSP derived from these sources cannot accurately represent crop phenology. This results also in the fact that phenology estimates from these sensors are difficult to compare with ground-based measurements (Nietupski et al., 2021; Nijland et al., 2016). Hence, satellite data of both high spatial and temporal records are needed.

With the advent of satellite imagery of finer spatial and temporal resolution such as Sentinel-2, it has become possible to extract different phenological phases (or phenometrics) of vegetation at finer scale (Fan et al., 2020; Vrieling et al., 2018). As such, more precise information can be obtained for heterogenous landscapes where the phenology varies substantially due to (small) fields with different crops, trees or roads alongside fields, management practices, crop growing times, cultivars, and other land covers. Sentinel-2 (S2) satellites provide multispectral imagery with a spatial resolution of 10 m in the visible and near-infrared bands. The two satellites (S2A and S2B) have a revisit period of five days at the equator and shorter near the poles, where the overlap between the swaths of different orbits increase. These imagery have been extensively used for studying vegetation phenology at a finer scale (Bolton et al., 2020; Cheng et al., 2020; Jönsson et al., 2018; Moon et al., 2021; Tian et al., 2021; Vrieling et al., 2018). Bolton et al. (2020) retrieved vegetation phenology at continental scale with using two-band vegetation indices (EVI2) (Jiang et al., 2008) from harmonized S2 and Landsat product at 30 m resolution. They have achieved accuracies of 41 and 49 days between estimated and observed LSP from camera at agricultural sites. Their results were largely impacted by outliers. Tian et al. (2021) used S2 data to extract vegetation phenology across Europe using Normalized Difference Vegetation Index (NDVI) (Tucker, 1979), EVI2, and Plant Phenology Index (PPI) (Jin and Eklundh, 2014). However, these studies estimated general vegetation phenology without looking at specific crops.

Several studies attempted to estimate crop-specific phenology using fine-resolution imagery (Misra et al., 2020). For example, Gao et al. (2020a) noted that S2 imagery could be used to effectively retrieve start of season (SOS) of maize and soybean within two weeks of crop emergence. The end of season (EOS) of cover crops can be well determined through S2-derived NDVI when frequent observations without cloud cover are available (Gao et al., 2020b). Pan et al. (2021) accurately classified the planting area of wheat using a phenology-based algorithm. However, a spatial and temporal representation of crop-specific phenology on a larger scale and considering multiple years is still lacking. Furthermore, cloud cover remains a significant drawback for estimating phenology with optical satellite data alone. It may result in no or few observations during critical moments of the crop cycle.

Synthetic aperture radar (SAR) is less affected by weather conditions and is known for its all-day capabilities. It is sensitive to structural and dielectric properties of vegetation and soil, including biomass, water content, etc., which varies concerning crop types, growth stages, and conditions of crops (Inoue et al., 2014). Therefore, it can be a reliable data source for monitoring the phenological stages of crops. Sentinel-1 (S1) C-band sensors (both Sentinel-1A and Sentinel-1B) collect imagery with a temporal resolution of six days at the equator with both the sensors. The temporal coverage increases closer to the poles, where the orbital overlap increases. It further increases by combining both the ascending and descending orbits. This is a considerable advantage for estimating phenometrics because of its frequent and reliable observations, given that cloud cover does not prevent the imaging of the land surface (Meroni et al., 2021; Stendardi et al., 2019; Veloso et al., 2017). However, as of December 2021, Sentiel-1B stopped working due to an anomaly. Although European Space Agency (ESA) is trying to fix the anomaly, long-term data unavailability (several months) can be assumed.

Nasrallah et al. (2019) used S1 (VV, VH, and VV/VH) to monitor winter wheat phenology and detect germination, heading, soft dough, and harvesting. Khabbazan et al. (2019) noted changes in S1 backscatter in different phenophases as plant structure and water content change. Several authors use the backscatter cross-polarization ratio VH/VV (CR) as CR reduces the effect of soil moisture and the interaction of soil and vegetation (Vreugdenhil et al., 2018). Schlund and Erasmi (2020) indicated a high potential of the CR of S1 to detect winter wheat phenology. A few studies used SAR S1 and optical S2 combined to monitor crop phenology (d'Andrimont et al., 2020; Mercier et al., 2020; Meroni et al., 2021). Veloso et al. (2017) evaluated the temporal behaviour of S1 backscatter for various summer and winter crops in Southwest France. They noted that CR has a high potential for monitoring crops similar to S2 derived NDVI. D'Andrimont et al. (2020) utilized the strength and consistency of S1 and S2 time series for detecting flower phenology of oilseed rape over Germany. They concluded that both sensors could accurately capture the flowering dates with a five-day delay between the sensors and the reference. Meroni et al. (2021) explained the potential of using S1 backscatter for deriving crop phenology and how it showed good agreement with NDVI for winter crops. These studies confirmed that backscatter and VIs could vary temporally depending on the soil properties and crop structure, which needs further exploration. Mercier et al. (2020) suggested using both S1 and S2 data for retrieving phenology because of higher accuracy than using each sensor alone. Although these studies focused specifically on the field level, they either considered small areas or only a limited number of fields in a large extent. The current study attempts to fill this knowledge gap by studying all the fields of a specific crop in a larger region.

This study aimed to accurately derive crop-specific phenology of winter wheat using S1 and S2 time series data and compare the results with in-situ field data. Winter wheat is considered in this study as it is one of the most important cereal crops in Europe and worldwide (Trnka et al., 2014). The Brandenburg state of Germany is selected as the study area because Germany has a dense network of in-situ data for phenology reported by the German Meteorological Service (Deutscher Wetterdienst - DWD). Furthermore, Brandenburg experienced different climatic conditions with extreme weather events during the past few years. For example, part of 2018 was extremely dry, whereas 2017 was much wetter than usual. This study explored how these different weather conditions affect the phenology of the crops. VV, VH and CR were obtained from S1, whereas the Enhanced Vegetation Index (EVI) was retrieved from S2. EVI is an improved version of the widely used NDVI which reportedly starts saturating in dense vegetation canopies (Huete et al., 1999, 1997). Incorporating both the soil and atmosphere adjustments, EVI performs better in dense vegetation conditions and has been used widely in past phenological studies (Zeng et al., 2020). The spatial and temporal variability of retrieved phenometrics were further investigated over the study area for a period of five years (2017 to 2021). Furthermore, the relationship between the retrieved phenology with gridded weather data from local climate stations was investigated. The spatial and temporal variations of the estimated phenology were explained with temperature and precipitation data.

1.2. Research objectives and research questions

a) **To explore the temporal profile of S1 backscatter and S2 EVI for wheat and assess their correspondence with the expected phenophases from fields across Brandenburg.**

Q1) Do field-specific temporal patterns of Sentinel-1 derived VV, VH, and VH/VV backscatter match the expected phenology of the crop?

Q2) Do field-specific temporal patterns of Sentinel-2 derived EVI provide a clear temporal signal that matches the expected phenology of the crop?

Q3) What are the observed differences between S1 and S2 derived temporal patterns?

b) **To retrieve phenometrics for wheat fields from S1 and S2 temporal profiles and calibrate these against reported DWD phenophases.**

Q4) Which in-situ observed phenophases correspond with phenometrics derived from S1 and S2 temporal profile?

Q5) How well do the retrieved phenometrics correspond to DWD phenophases?

c) **To validate and map crop specific-phenometrics using the performed calibration and explain their spatio-temporal variability with gridded weather data**

Q6) How do the phenometrics vary spatially and temporally?

Q7) Can the retrieved spatial and temporal variability of the phenometrics be explained by the weather data?

2. STUDY AREA

The area of interest for this study is the Brandenburg State of Germany. It is located in the eastern part of the country with a total area of 29,654 km². Agriculture is the dominant land use covering approximately 45% of the whole land surface. 77% of the agricultural land is arable, and 23% is permanent grassland (Ihinegbu and Ogunwumi, 2021). Winter wheat is one of the most cultivated crops in Germany, accounting for 46% of the total cereal area in 2020 (DESTATIS- 2020).

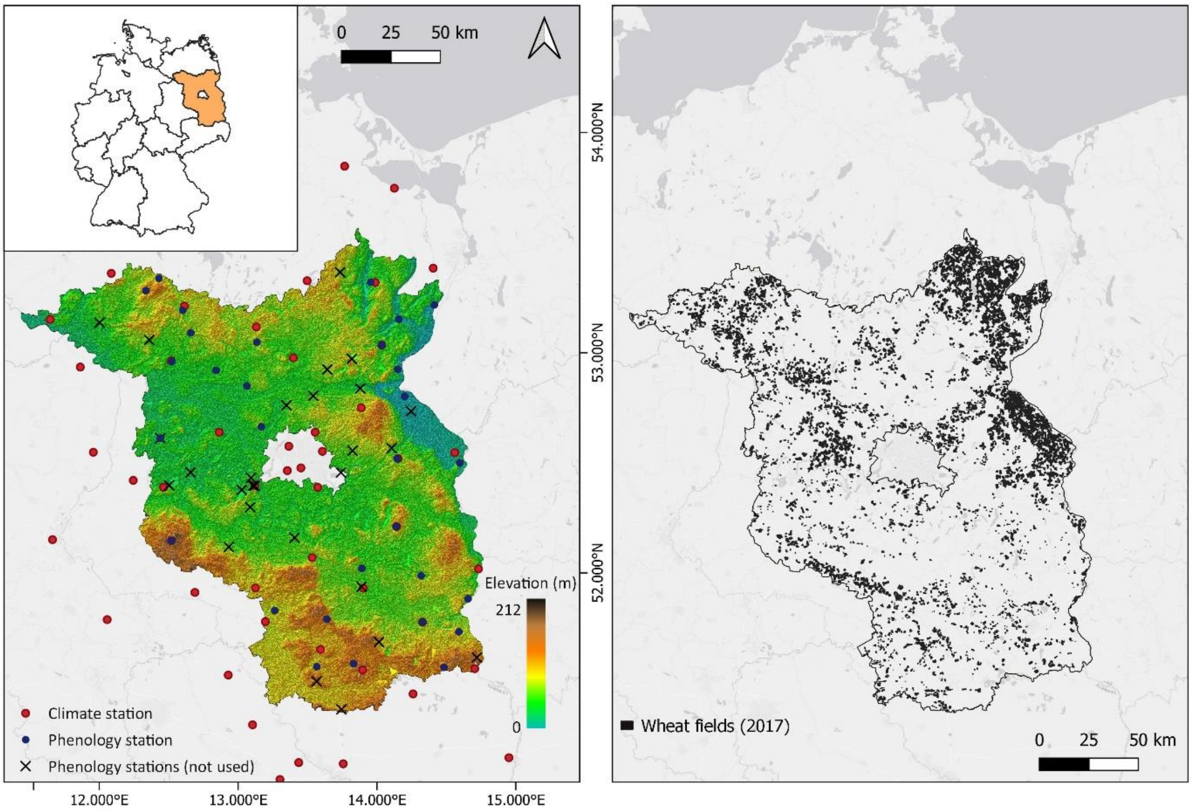


Figure 1 Overview of the study area showing DWD weather stations in and around the state, phenology stations from DWD, elevation in meters from SRTM digital elevation model, and wheat fields from 2017. Phenology stations which are not used for this study are shown with cross marks (see Section 3.2). 2017 is selected randomly to visualize the locations of wheat fields (n = 9806).

The areas cultivated for winter wheat vary from state to state. A brief overview of the crops in Brandenburg with the largest cultivation areas is presented in **Figure 2**. In addition to winter wheat, maize (silage and biogas) and rye were two most cultivated crops in the state for the studied years. Winter wheat in Germany is usually sown around late September to mid-October, and the crop emerges around late October, followed by a winter dormancy period till March. After the dormancy, newly formed leaves unfold, and crops grow in size and volume. It peaks around May and is harvested between late July and early August. Brandenburg surrounds Berlin, Germany's capital city (**Figure 1**), where high urbanization significantly impacts land use. Due to the increasing demand for food products in Berlin, crop production increased substantially in Brandenburg (Gutzler et al., 2015). Agriculture in Brandenburg is characterized by low soil fertility dominated by sandy and sandy-loamy soils, large farm enterprises with high technology levels, and yield limitations due to lack of water (Gutzler et al., 2015; Wolff et al., 2021). Low soil fertility and low precipitation rate (600 mm/ year on average) make it challenging for agricultural production. Hence, farmers rely on intensive usage of fertilizer and pesticides along with heavy machinery employing advanced technology (Wolff et al., 2021).

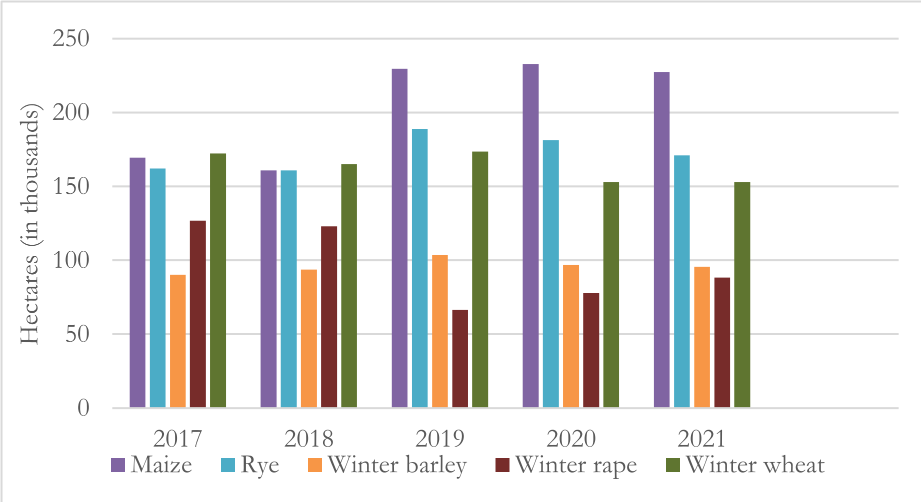


Figure 2 Temporal graph showing top five growing crops in terms of harvested areas for Brandenburg for the studied period. Data is retrieved from the Land Parcel Identification System (web: https://geobasis-bb.de/lgb/de/)

With an average daily maximum temperature of 14°C, Brandenburg is one of Germany's warmest areas (DWD, 2021a). Brandenburg has a temperate climate with warm summers and relatively warm winters. The mean temperature is around 19°C in summer, and 4.7°C in winter, which are above the national average (DWD, 2021b). A brief overview of the spatial variability of spring temperature and precipitation of Brandenburg is shown in **Figure 3**.

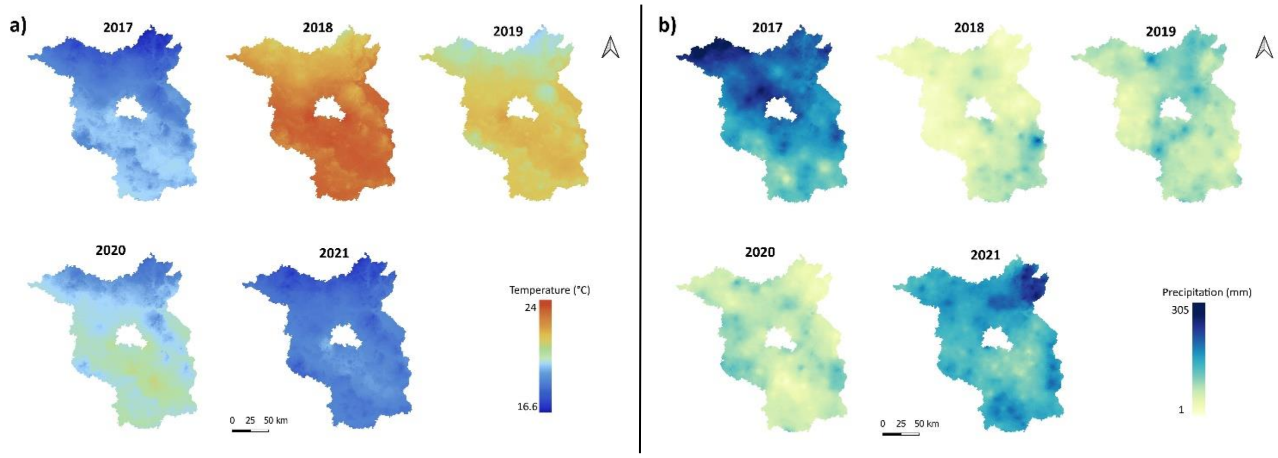


Figure 3 weather condition of the study area for the growing season (April, May, and June) of the studied years **a)** average of daily maximum temperature) **b)** total precipitation

In terms of precipitation, on average, Brandenburg receives less than 600 mm per year (1991 – 2020), making it one of Germany's driest locations. Furthermore, precipitation highly varies spatially and over time. For instance, the state received 721 mm of annual rainfall in 2017, about 20% more than the average of 30 years. In the preceding year, it only received 393 mm, 30% lower than the average, making it the driest year in Brandenburg to date (Reinermann et al., 2019).

3. DATA AND PRE-PROCESSING

For the study, remote sensing data from two satellite constellations are utilized, i.e. S1 and S2. In addition, ground phenology observation data and gridded weather data (rainfall and temperature) from DWD are used. To focus the analysis on wheat fields, annual crop type maps for the study area are obtained for the corresponding years from the government of Brandenburg (https://geobasis-bb.de/lgb/de/).

3.1. Remote Sensing data

The Sentinel-1 mission provides SAR data at C-band (5.405 GHz/ 5.54 cm wavelength). It includes two satellite constellations: Sentinel-1A, launched in April 2014, followed by the launch of Sentinel-1B in April 2016. It provides data at four polarization combinations: VV, VH, VV+VH, and HH+HV, depending on the mode in which the satellite is operating. Datasets are acquired in four modes: Interferometric Wide swath (IW), StripMap (SM), ExtraWide swath (EW), and Wave (WV). This study uses data from both the ascending and descending orbits at IW mode. The Ground Range Detected (GRD) product level of the IW mode provides dual polarization data (both VV and VH) at a pixel spacing of 10 m with a swath of 250 km (Torres et al., 2012). In this study, the dataset is extracted from the Google Earth Engine (GEE) cloud computing platform. It offers an application programming interface (API) containing an extensive archive of freely available satellite remote sensing datasets, including S1 and S2. This study used S1 GRD scenes from the collection "COPERNICUS/S1_GRD_FLOAT" of GEE, which is in linear scale and gets updated faster than the collection that is logarithmically scaled ("COPERNICUS/S1_GRD"). The datasets are processed using the Sentinel-1 toolbox to remove the thermal noise and generate a calibrated and ortho-corrected backscatter coefficient σ°. To reduce the effect of terrain on the radiometry, it is converted to γ° (Small, 2011) using the following equation:

$$\gamma^{\circ} = \sigma^{\circ} / \cos\theta \qquad \textbf{(Eq. 1)}$$

where θ represents the local incidence angle of the sensor. Because in the analysis, all pixels within an individual field were aggregated through averaging, further speckle filtering was not performed. Based on the γ° values, CR was calculated per pixel as the ratio of VH and VV backscatter and then aggregated at field level.

$$\gamma^{\circ}_{CR} = \gamma^{\circ}_{VH} / \gamma^{\circ}_{VV} \qquad \textbf{(Eq. 2)}$$

Sentinel-2 (S2) is a multispectral imaging mission with a wide orbital swath of 290 km. It consists of two satellites; Sentinel-2A was launched on 23 June 2015, followed by Sentinel-2B on 7 March 2017. Together, they provide imagery with a revisit time of five days at the equator and shorter (2 – 3 days) near the poles where the swaths of different orbits overlap. S2 offers 13 spectral bands: four bands at 10 m in the visible and NIR region, four and two in red edge and Short Wave Infrared (SWIR) region respectively at 20 m, and the other three bands at 60 m which are primarily used for atmospheric corrections (aerosols, water vapor, and cloud mask). Level-1C imagery, top of atmosphere (TOA) was used in GEE from the collection "COPERNICUS/S2". TOA was chosen instead of the bottom of atmosphere reflectance (Level-2A) because Level-2A products are only available from April 2017 onwards in the GEE platform, thus not fully covering the period of interest. Additionally, VIs like the EVI can reduce the adverse impact of atmospheric influences (Matsushita et al., 2007).

The S2 dataset was screened for clouds, cloud shadows, and snow following the cloud and snow score method used by Chamnan (2021) and Meroni et al. (2021), adapted from Housman et al. (2018).

During the process, brightness levels in blue, cirrus and the three visible bands were used to assign a cloud score. As clouds are reasonably brighter in these bands, a higher score is given to the brighter pixels. Additionally, the Normalized Difference Water Index (NDWI) (Gao, 1996) was also considered to generate the cloud scoring. As clouds are moist, higher cloud scores are provided for pixels with higher NDWI values. Combining both the indicators of cloudiness, a threshold is used to classify the pixels as cloud. Cloud shadows were identified using cloud ground projection from the solar geometry, cloud height, and dark pixels. Finally, snow scores were assigned using the Normalized Difference Snow Index (NDSI) (Hall et al., 2002). Finally, a specific threshold (0.2 for cloud and snow, 0.35 for cloud shadows) is applied to each factor to define cloud shadow and snow, whereas pixels above the threshold were masked out from the images. Removing the contaminated observations, the remaining pixels were taken for the calculation of VI. EVI was calculated using the following formula:

$$EVI = G(\frac{NIR-RED}{NIR+(C1*RED-C2*BLUE)+L}) \qquad \textbf{(Eq. 3)}$$

where NIR, RED, and BLUE are reflectance in Near Infrared (832 nm, band 8), Red (664 nm, band 4), and Blue (492 nm, band 2) regions. G is determined by the C value which is calculated by linear fitting (RED = c * NIR). C1 and C2 are coefficients to correct aerosol scattering, and L is a soil adjustment factor. The following values were used for these parameters: G = 2.5, C1 = 6.0, and L = 1 which were commonly used in previous literature (Huete et al., 1997).

3.2. In-situ phenology observation data

Datasets for ground phenological observation are collected from DWD. Data can be freely obtained from the DWD Climate Data Centre (CDC, ftp://ftp-cdc.dwd.de). Phenological observations of a wide variety of crops are collected for an extended period for the whole of Germany. Volunteer surveyors gather these observations at designated observation points throughout the agricultural season. The observations are made two or three times each week, over the same fields within 5 km of the station's nominal geolocation (DWD, 2015). The datasets are then quality controlled by the agency (Kaspar et al., 2015). The quality control is done in three steps: provided with codes, i.e. QN (QUALITAETS_NIVEAU) = 1 means only formal control is done; QN = 2 refers to second control before correction QN = 10 means that quality control and routine corrections are finished. When duplicate records are found with different QN values for a specific observation, the highest one is recommended. No duplicate values are found for the studied years; hence all the observations were used, regardless of the QN values.

Table 1 reported phenophases of winter wheat from DWD with the description of each phases, corresponding DWD code and more widely used BBCH code of each phases

Name of the stages	Description	DWD code	BBCH code
Beginning of emergence	the cotyledons, which are still rolled up, have pierced the surface of the earth and reached a height of approx. 1 cm	12	10
Start of shooting	about 50% of the plants clearly grow in length and the first node of the culm can be felt above the ground	15	31
Ears emerge	On about half of the stalks the tips of the ears push out of the leaf sheaths	18	51

Milk ripeness	the first grains in several ears have reached their final size, but are still soft and green	19	75
Yellow ripeness	the first grains in about half of the ears have finished their discoloration from green to yellow and can be easily detached from the panicle	21	87
Harvest	the observed field is harvested	24	99

The reported phenophases are labelled with a DWD code and a description. A lookup table is also provided, which compares the DWD scale with the more frequently used Biologische Bundesanstalt, Bundessortenamt und CHemische Industrie (BBCH) scale. The BBCH scale is a numeric classification system developed to code different phenophases of crops (Meier, 2001). It contains two-digit codes; the first is the principal growth stage (e.g. 6 = flowering) while the second represents the secondary growth stage. The BBCH scale offers the possibility of defining the stages of the crop precisely. The reported phenological phases are shown in **Table 1** with their corresponding DWD and BBCH code, along with their description provided by DWD.

There are 18 to 25 phenology observation stations (varies depending on the year) that recorded the phases for the study period. The number of phenology stations which reported the studied phases along with the total parcels within their buffers are shown in **Annex 6**. Stations that did not record the studied phases were not considered for calibration and validation processes. Half the number of stations from 2019 and 2021 were taken for the calibration process. The two years were chosen because of their contrasting meteorological conditions (**Figure 3**). The other half of these two years as well as all the stations of the left out years (2017, 2018, 2020) were taken for validation. The location of the retained phenology stations are shown in **Figure 1.**

3.3. Crop type data

Agricultural crop information was retrieved freely from the government website of Brandenburg (https://geobasis-bb.de/lgb/de/) which is derived from the Land Parcel Identification System (LPIS). LPIS is part of the Integrated Administration and Control System (IACS) which manages the payment of agricultural subsidies to the farmers. It provides parcel-level spatial information about agricultural land use annually. In this study, wheat fields *(Winterweizen)* are taken from the dataset for 2017-2021 and used as a mask layer to retrieve temporal profiles of wheat fields. To avoid that field edges and associated mixed pixels influence the signal, an internal 10 m buffer was taken, thus reducing the field size for which the profile was extracted. This was done to keep the signals purely coming from the specified fields while avoiding trees, shadows, ditches, and nearby roads from the surroundings or neighbouring fields. To ascertain that each retained field is sufficiently large and contains a decent number of pixels for averaging, only fields bigger than 1 ha after taking the internal buffer were considered in this study. **Table 2** provides insights about the number of wheat fields, and their areas before and after the pre-processing.

Table 2 Statistical summary of wheat fields for the studied years before and after the pre-processing steps

Year	Total parcels	Max area (ha)	Parcels bigger than 1 ha after taking internal buffer	Mean area of retained parcels (ha)
2017	9806	242	8485	18
2018	9240	205	8024	18.2
2019	9890	225	8550	18

2020	8610	195	7460	18.2
2021	8939	222	7781	17.4

Finally, the polygons of respective years were used as the area of interest for extracting time series from S1 and S2 data.

3.4. Weather data

Gridded weather (temperature and rainfall) data was collected from the DWD portal (https://opendata.dwd.de/climate_environment/CDC/grids_germany) for the respective years. Both dataset have a spatial resolution of 1 km, are in ESRI-ascii-grid-format and are freely downloadable from the website. Observations are collected at two meters above the ground using Laser Precipitation Sensors and Ulatrasonic Anemometers. The gridded datasets are interpolated from the DWD station data considering DEM and using height regression and Inverse Distance Weight (IDW) methods (Kaspar et al., 2013). Point locations of the DWD climate stations of Brandenburg are visualized in **Figure 1**.

4. METHODS

Figure 4 shows the methodology flowchart of the study. First of all, temporal profiles from the three SAR parameters and EVI were explored and compared against DWD in-situ phenology data. Secondly, phenometrics were retrieved from the profiles and after calibration and validation using DWD data, they were mapped to assess their spatial and temporal variations. Finally, retrieved phenometrics were aggregated at a similar grid as the weather data to see if the variations can be explained with weather parameters.

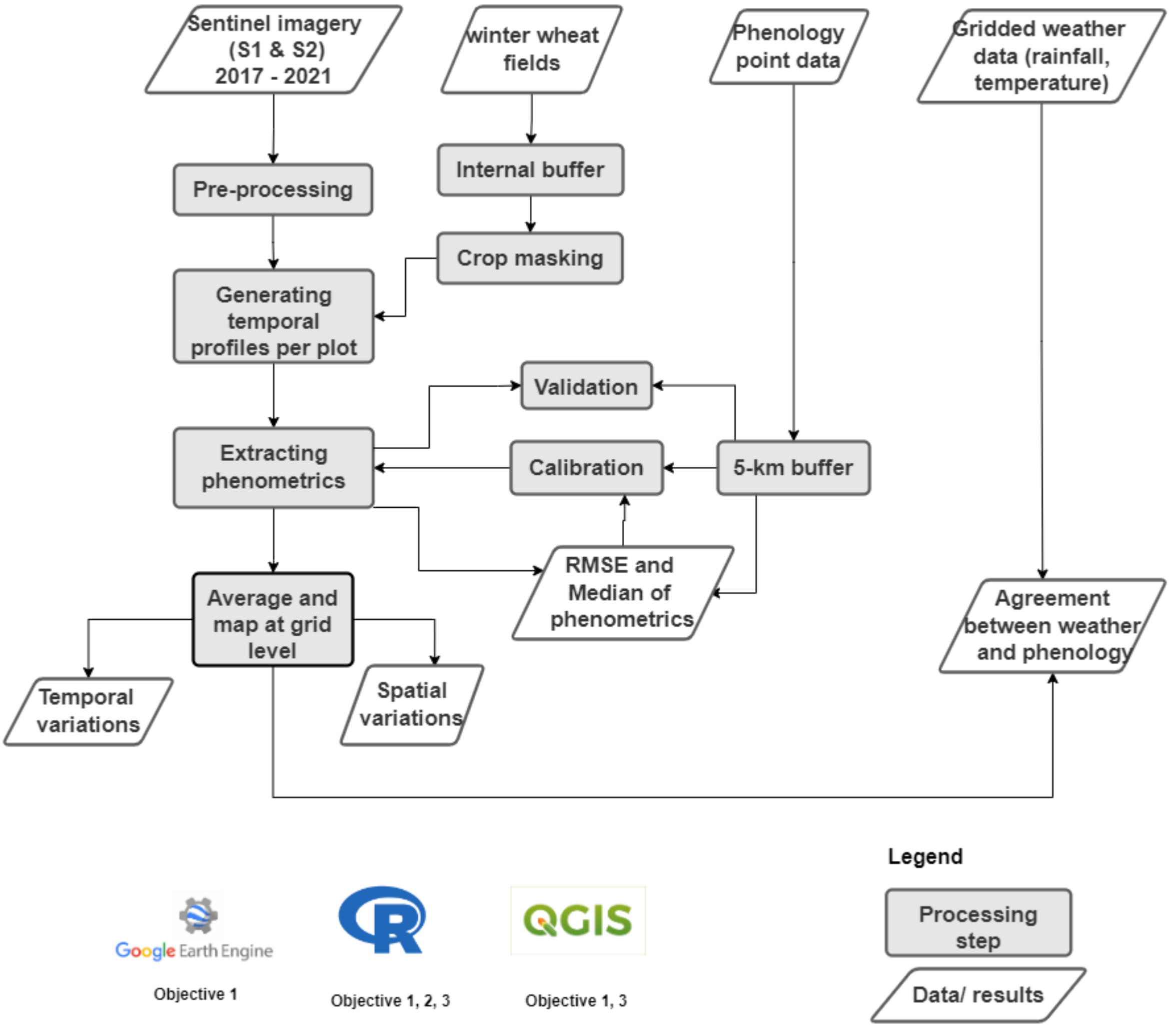


Figure 4 Overview of the methods

4.1. Exploration of temporal profiles

Time series data for the respective years are retrieved from the GEE platform as mean EVI and SAR parameters (CR, VH, and VV) per parcel from both pre-processed S2 and S1 imagery respectively. Each parcel is given a unique id to retrieve the time series and to later link outcomes of the analysis back to the parcel. Before retrieval of the time series, an empty list of no data values (-9999) for a daily time step ranging from 1 September (from previous year) to 31 October was created inside GEE. Then the empty list is filled with valid observations from the sensors. Due to the multi-day revisit periods of both sensors and the cloud masking of S2, no data values remain in each time series. Since S2 data is distributed as granules or tiles, there are overlaps between the neighbouring tiles. Since each tile is processed differently, the values can be slightly different for a parcel that shares different tiles. Hence, for the time series to be consistent, the maximum values among the tiles is chosen (e.g., in case of a single observation on a specific day that is "artificially" duplicated in two tiles, the observation with the highest EVI is retained). This was done at parcel level. Finally, the time series were exported as CSV files for each SAR parameter and EVI per year.

Temporal profiles were generated by smoothing the raw values from EVI and SAR parameters using locally weighted scatterplot smoothing (LOESS), a local polynomial fitting (Cleveland, 1979). Local polynomial regression fitting like LOESS is commonly used in remote sensing time series analysis and has proven to perform quite well for smoothing by providing sufficient generalization of the time series while retaining the important temporal characteristics (Cai et al., 2017). LOESS was executed in the R environment (R Core Team, 2021). It fits a polynomial regression for each value and its neighbourhood, whereby the neighbourhood size defines the smoothness of the fitting specified by span values. Higher span values result in smoother fittings. Hence, an optimal span value is required to provide a fair trade-off between the original and smoothed time series. Generalized cross validation, visual inspection as well as a good correspondence with in-situ phenology data (which span value retrieves accurate phenometrics during the calibration period) were used to find the optimal span value. After careful inspection, a span value of 0.40 was selected for the SAR parameters and EVI. In addition to LOESS, Double logistic (DL) function fitting of the non-smoothed data series (Elmore et al., 2012) was also tested for fitting. It was implemented using the 'greenbrown' package in the R environment (Forkel et al., 2015) and can be written as Fan et al. (2020):

$$f(t) = m_1 + (m_2 - m_7 t)\left(\frac{1}{1+e^{(m3-m4t)}} - \frac{1}{1+e^{(m5-m6t)}}\right) \qquad \textbf{(Eq. 4)}$$

where t denotes time in days. Meaning of the model parameters: $m_1$is the base minimum value of EVI or the SAR parameters; m_2 denotes the amplitude from base to the maximum value; m_3 $[m_5]$ are the inflection point for onset [offset] of greenness; m_4 $[m_6]$ are the slope of local time series and m_7 is an additional linear slope which can consider factors like decreased wheat before harvest. Both the LOESS and DL were tested to assess which one fits the data better and yields more accurate phenometrics in the later part of the study.

It should be noted that, time series which had less than ten observations during the growing period (March till August) were not considered while fitting the data. This was done to have confidence about the profiles as fewer observations might affect the fitting. The number of retained parcels (temporal profiles) after the screening are shown in **Annex 6**. Unexpected high values (> 0.15 for EVI and >-11 for CR) at the beginning of the season are removed. The reason behind removing those high values is that those could potentially be coming from a previous crop as very high cross ratio and EVI values are not expected around the cultivation period. Finally, the smoothed temporal profiles are plotted together with the reported phenophases from DWD. Ten representative fields are chosen from a 5 km buffer of Neuküstrinchen phenology station (station id: 12173) to explore if the temporal profile matches the expected phenology of the crop.

4.2. Retrieval of phenometrics from temporal profiles

For the retrieval of phenometrics from the time series, two different standard methods (breakpoints and threshold) were tested and used. For each method, calibration with the field data is performed to tune the parameters. Through the comparison, phenometrics, which had the closest correspondence with the DWD phenophases were identified. Hence, those identified phenometrics were retrieved and validated in the later part. The retrieval methods are discussed in the following sections:

Breakpoints

Breakpoints were calculated using the 'strucchange' package in R environment using the smoothed datasets.

The algorithm of the package is based on Bai and Perron (2003) and Zeileis et al. (2003). These breakpoints split the time series iteratively into different portions. If two successive segments have different regression coefficients, a breakpoint is assumed in this approach and the best position of a breakpoint is determined by minimizing an information criterion (Verbesselt et al., 2012). In this way, different phenometrics can be detected from the time series where abrupt changes are visible. This method has been used widely in several time-series studies to identify forest trend shifts (Forkel et al., 2013) and to detect phenology of crops (Harfenmeister et al., 2021; Löw et al., 2021; Schlund and Erasmi, 2020). Only the growing season (March – August) is considered for calculating the breakpoints.

Threshold
The threshold method is often used in satellite remote sensing because of its proven robustness in retrieving phenometrics (Huang et al., 2019; You et al., 2013). The method can extract four different metrics in a time series, such as: start of season (SOS), end of season (EOS), Length of season (LOS) and point of peak (POP). Except for the POP, the timing of the other metrics are dependent of the threshold values. For instance, 25% threshold can extract SOS earlier and EOS later than 50% threshold. On the other hand, POP is determined at the peak of the time series. For this study, different thresholds are tested to extract the metrics (20% to 50% at an interval of 5%) and to retain those that best correspond to specific phenophases. The threshold method was tested and implemented on both EVI and CR using the '*greenbrown*' package of the R environment (Forkel et al., 2015).

4.3. Calibration and validation of phenometrics against DWD phenophases

The identified breakpoints and the threshold-derived metrics were compared against the field observations to assess which metric could best explain a phenological stage. Temporal differences (Δ_t) between the DWD observed and remote sensing estimated phenological stages were calculated as (**Eq. 5**). Root Mean Square Error (RMSE) and median values were calculated from the difference for each field inside the station buffers for all years. RMSE and median values were used as measures of accuracy of estimated phenometrics.

$$\Delta_t = \Delta_{obs} - \Delta_{est} \qquad (\textbf{Eq. 6})$$

DWD does not provide the actual field location but rather a nominal location within 5 km from the observed field. This makes it impossible to directly link the in-situ observed phenophases and phenometrics of a specific field. One solution could be interpolation of the field observed data then relating them with the field-specific phenometrics. However, Meroni et al. (2021) found that different phenophases observed by DWD have a large spatial variability within their small scale and do not display a clear spatial pattern. Therefore, interpolation was assumed to be not appropriate. Hence, they compared the average and spread of DWD phenophases in a region and compared this with the LSP metrics found for sample fields within the same region. Lacking precise location of ground data, Tian et al. (2021) used a buffer around the ground data points and took average of the plots to match with their estimated S2 phenometrics.

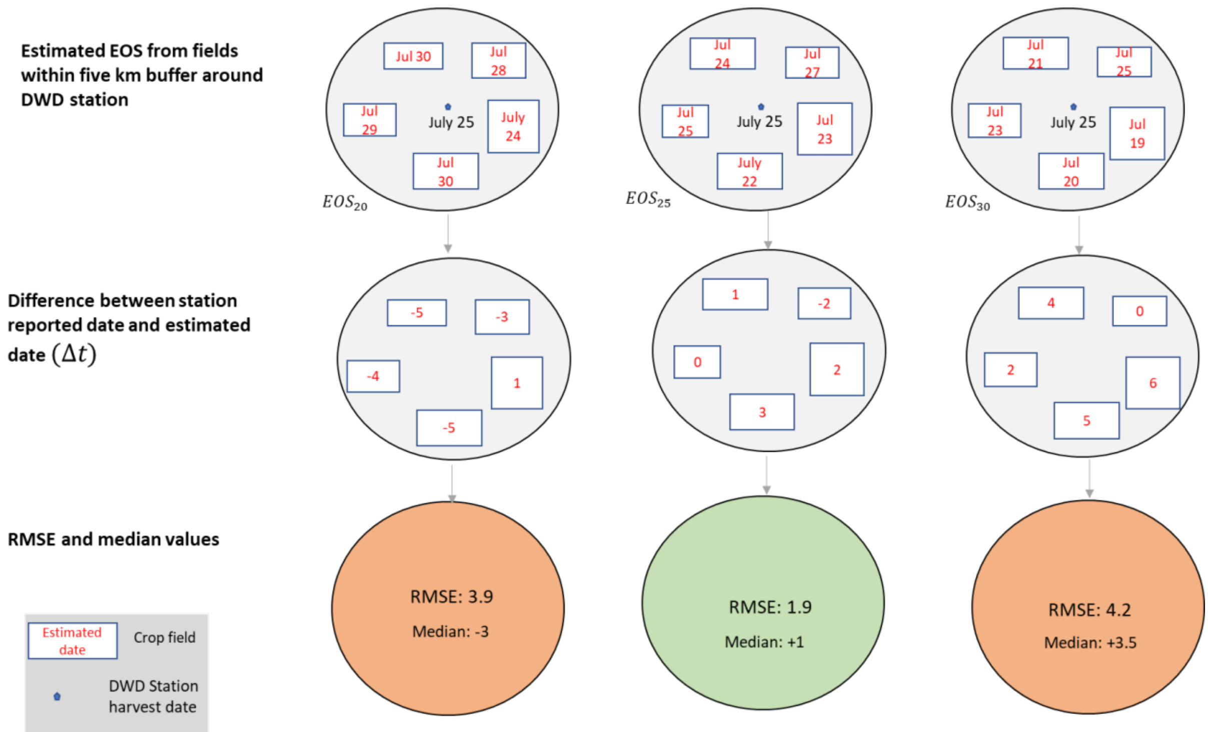


Figure 5 A graphical representation of the calculation of RMSE and median values from estimated phenometrics and reported phenophase data

This study generates buffers of five kilometres around the DWD phenology stations (as suggested in **Figure 5**). Within those buffers, temporal differences (Δt), RMSE, and median values between observed and estimated phenological phases are calculated (**Figure 5**). As indicated earlier, half of the stations of 2021 and 2019 were used for calibration. The left-out fields from these two years as well as the other three years (2017, 2018, and 2020) were used for validation. To gain more confidence about the results, the coefficient of determination (R^2) is calculated between the observed phenophases and mean phenometrics of all the fields per stations.

4.4. Relationship with gridded weather data

The retained metrics chosen from calibration process are calculated for all the wheat fields for all five years and visualized in maps to show their spatial and temporal variations. Linear regression was performed between the weather variables (temperature and precipitation) and phenometrics. For each phenometric, a regression was performed with monthly weather data for the month(s) in the growing season, potentially corresponding to that phase. For example, as ears emerge occurs around mid-May, weather data of March, April, and May are tested individually and collectively with different combinations (March-April, April-May, March-April-May) to link with this phase. This analysis aimed to find relationships with a high R2 to understand which weather parameters drive variability in phenometrics. A similar grid representing the phenology is needed for relating the phenometrics with gridded weather data. A 3 × 3 km grid is taken by resampling the original 1 km grid from weather data. This resampling is performed because a lot of overlap between the field boundaries occur in the 1 km grid. Therefore, grid-wise relationships were built between the two variables for each year and combining all the years. Scatterplots with a linear trendline were drawn showing while showing the R^2 values and the significance of the relationship (p-value) between the estimated phases and weather variables.

5. RESULTS

5.1. Assessment of temporal profiles of EVI and SAR parameters of wheat

The assessment of temporal profile is divided into three subcategories: first, a visual comparison between the two tested fitting methods for a few exemplary wheat fields were done. Secondly, temporal profiles of wheat from three different years (2017, 2018, and 2020) were analysed with their temporal variations. These three years were selected because of their nature in different meteorological conditions. Finally, ten different temporal profiles of wheat are studied with their correspondence against DWD reported phenophases.

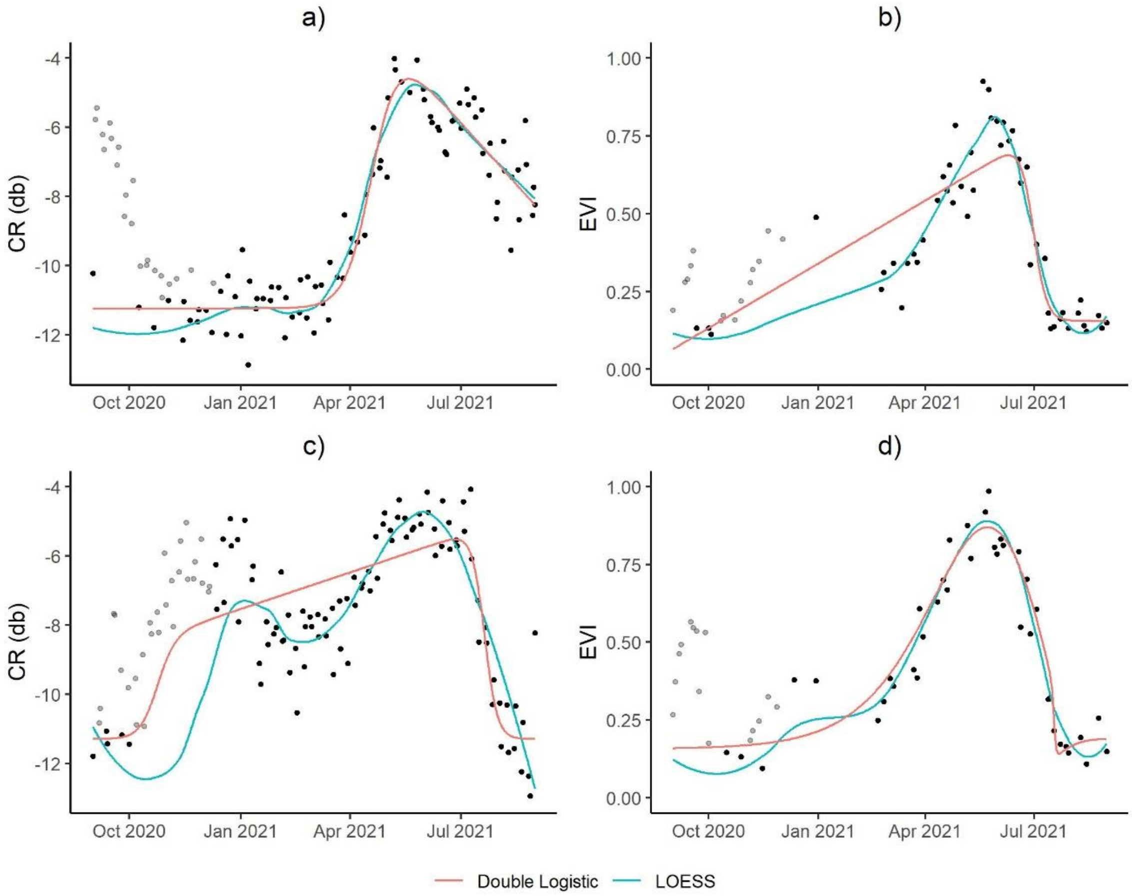


Figure 6 CR and EVI observations of 4 different wheat fields with LOESS and double logistic smoothing. Extreme high values until mid-December are shown in grey dots that are not considered while fitting. Valid observations are shown in black dots.

As expected, there are more observations in S1 CR than in S2 EVI. In general, very few observations can be noticed between mid-December to mid-February in S2. In terms of fitting, DL fits the data well when there is less noise in the data (**Figure 6a** and **d**). However, where the observations show variability (**Figure 6b** and **c**), it underfits the data. Moreover, winter green-up could not be modelled with DL. On the other hand, LOESS performs well in better representing of the data's temporal trajectories for both the sensors. It seemed to deal better with noise than DL, and the initial winter green-up was evident. Hence for further analysis, only LOESS smoothing is used.

Figure 7 shows that temporal profiles of both VV and VH follow a similar pattern during the winter dormancy period. From late March till mid-June, VV values continues to drop and then increase again. However, VH behavior were less pronounced and showed more variability over the three years than VV.

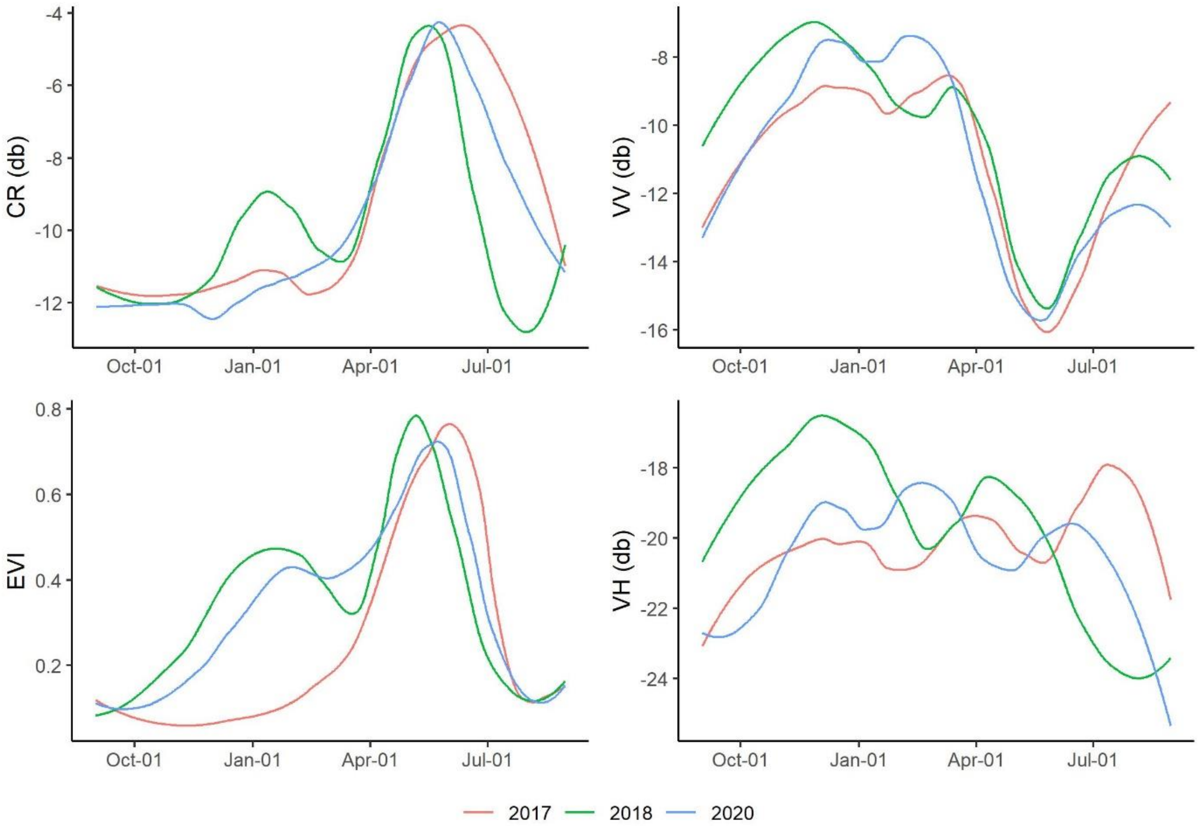


Figure 7 LOESS fitted temporal profiles of wheat for 2017, 2018 and 2020 of EVI and the three SAR parameters. Representative fields are chosen within the buffer of 'Lunow' phenology station (station id: 15469; Lat: 52.922, Long: 14.122).

CR and EVI follow a similar pattern where the values continue to increase during the first half of the growing season and then start to decree as the crops reach their maturity. Generally, CR reached their peaks and dropped later than EVI for all three years. Exploring the temporal profile of wheat for three years, it can be concluded that the season for 2018 was shorter than the other two years. On the other hand, 2017 experienced the most extended season among the three while also reaching the peak latest. In general, CR and EVI depict the wheat season better than VV or VH alone and remain stable in space and time. Hence, to estimate phenometrics, CR and EVI were considered out of the four parameters.

As **Figure 8** suggests, CR and EVI values increased after March and reached their peaks around mid-May to early June when the ears emerge phase (BBCH 51) was reported. After the ears emerge stage, the values of EVI and CR started to decrease. EVI drops abruptly during the senescence period till harvest, while for CR, a gradual decrease is noticed (**Figure 8**).

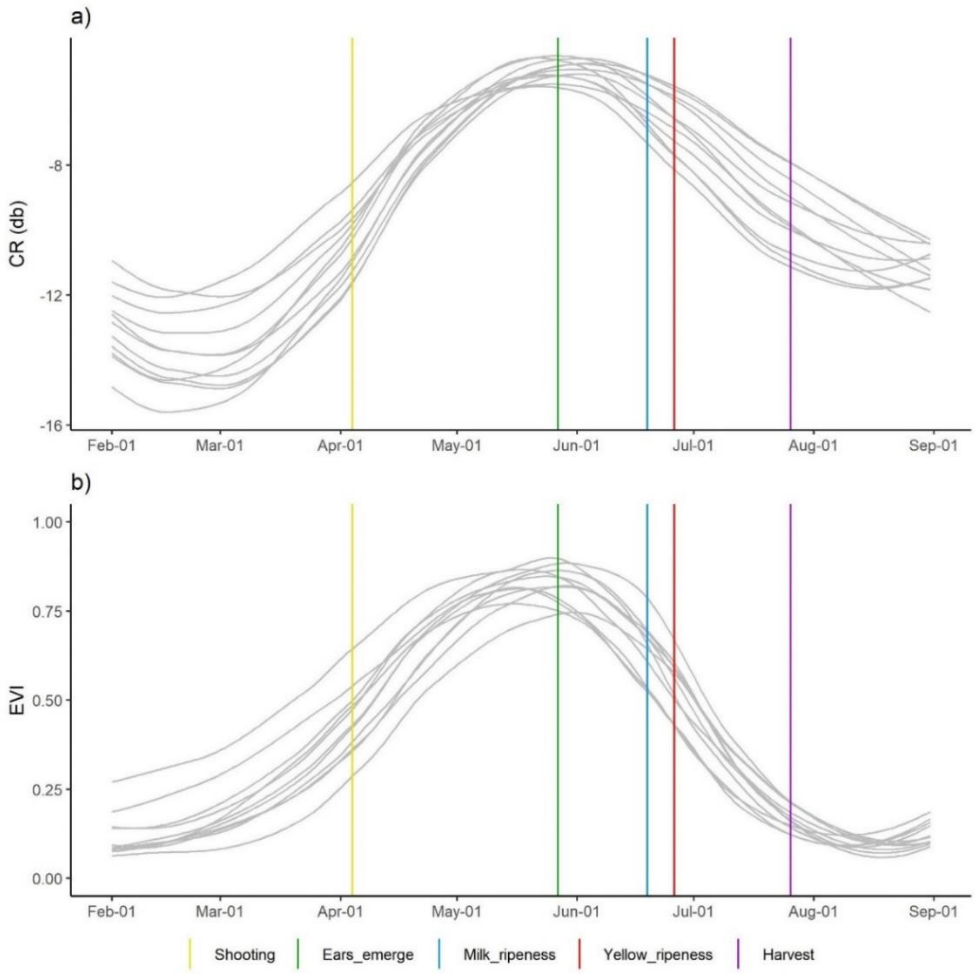


Figure 8 Temporal profiles from CR and EVI of a few wheat fields showing the correspondence with DWD reported phenophases. The profiles are shown in grey lines and the reported phases are illustrated with coloured vertical lines. The fields are chosen within the phenology station Neuküstrinchen (station id: 12173; Lat: 52.799, Long: 14.166)

5.2. Calibration and validation of phenometrics

5.2.1. Calibration

Both breakpoints and threshold methods were applied to the selected EVI and CR profiles (see Section 4.3) to calibrate and find the phenophases that best match the estimated metrics. Different parameters for breakpoint method were tested as well (DL does not allow to change any model parameter). Generally, four breakpoints were detected between March and the end of August from both sensors. The third breakpoint from CR showed RMSE and median values of 11.2 and 2 days respectively with milk ripeness. The final breakpoint had closest match with harvest with RMSE of 9.2 and median values of 5 days. On the other hand, POP from EVI corresponds well with ears emerge with RMSE of 11 days and median of 6 days. The other two reported phases from DWD (shooting and yellow ripeness) were not considered in this study as they could not be accurately retrieved using the two methods (**Table 3**). The accuracy was defined in terms of RMSE and median values (**Table 3**) where RMSE less than 12 days was considered accurate. The last two reported stages (yellow ripeness and harvest) showed RMSE of 14 days with EVI EOS50 and RMSE 12 days with EVI EOS25 respectively. SOS50 from EVI had RMSE 19 days with shooting dates, however, in most cases, retrieved dates are earlier (median 5.5 days) than the field observed dates. On the other hand, the second breakpoint from CR revealed RMSE 16 days with shooting, but the estimated dates are usually later than the actual dates (median -8 days). Threshold method could not detect SOS and EOS from CR.

Based on RMSE and median values, milk ripeness and harvest from CR breakpoint and ears emerge from EVI threshold were retained and used for validation and final retrievals for the whole state.

Table 3 Result of the calibration between the remote sensing metrics from both the methods and sensors and their links with DWD phenophases. Approximately 50 % of 2019 and 2021 data are used during the process. The retained metrics which are chosen from calibration and later used for validation and subsequent analysis are shown in bold. Dash (-) signifies the estimation was not possible. RMSE and median values are in days.

		Remote sensing metrics	DWD phenophase	RMSE	Median
Breakpoints	CR	Second breakpoint	Shooting	16	-8
		Third breakpoint	**Milk ripeness**	**11.2**	**2**
		Fourth breakpoint	**Harvest**	**9.2**	**5**
	EVI	Second breakpoint	Shooting	18	-8
		Third breakpoint	Milk ripeness	14	-4
		Fourth breakpoint	Harvest	12	5
Threshold	CR	SOS25	Shooting	-	-
		SOS50	Shooting	-	-
		POP	Ears emerge	13	6
		EOS25	Harvest	-	-
		EOS50	Yellow ripeness	-	-
	EVI	SOS25	Shooting	42	22
		SOS50	Shooting	19	5.5
		POP	**Ears emerge**	**11**	**6**
		EOS25	Harvest	12	7.5
		EOS50	Yellow ripeness	14	-2

5.2.2. Validation

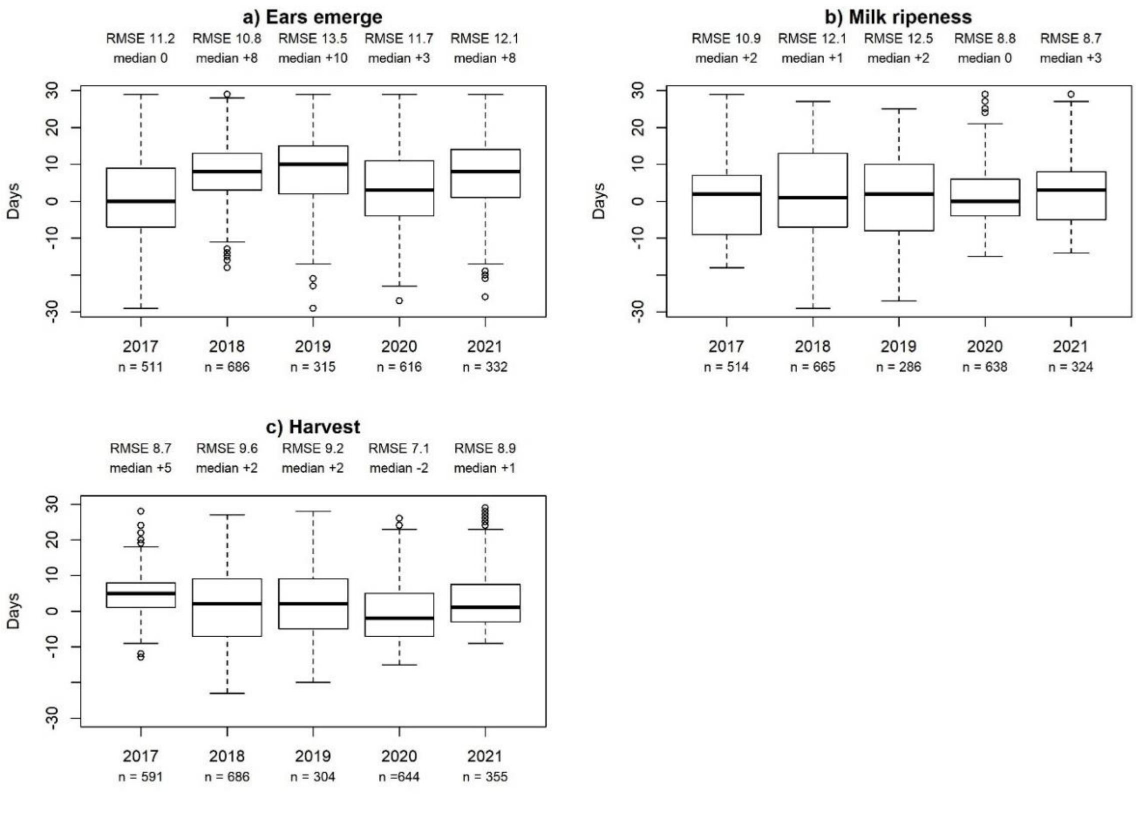


Figure 9 Boxplot showing the temporal difference between remote sensing metrics and DWD reported phases for the validation data. RMSE and median values are in days. n denotes to the number of plots used for the validation

All retained phases from the three metrics (except for ears emerge 2019 and 2021) showed median values of less than five days. The outlier profiles (shown in **Annex 2**) in both CR and EVI influence the RMSE values (8 to 12 days). It can be noted that no general pattern was found among the outliers from both CR and EVI. Hence it was not possible to remove those before calculating the metrics.

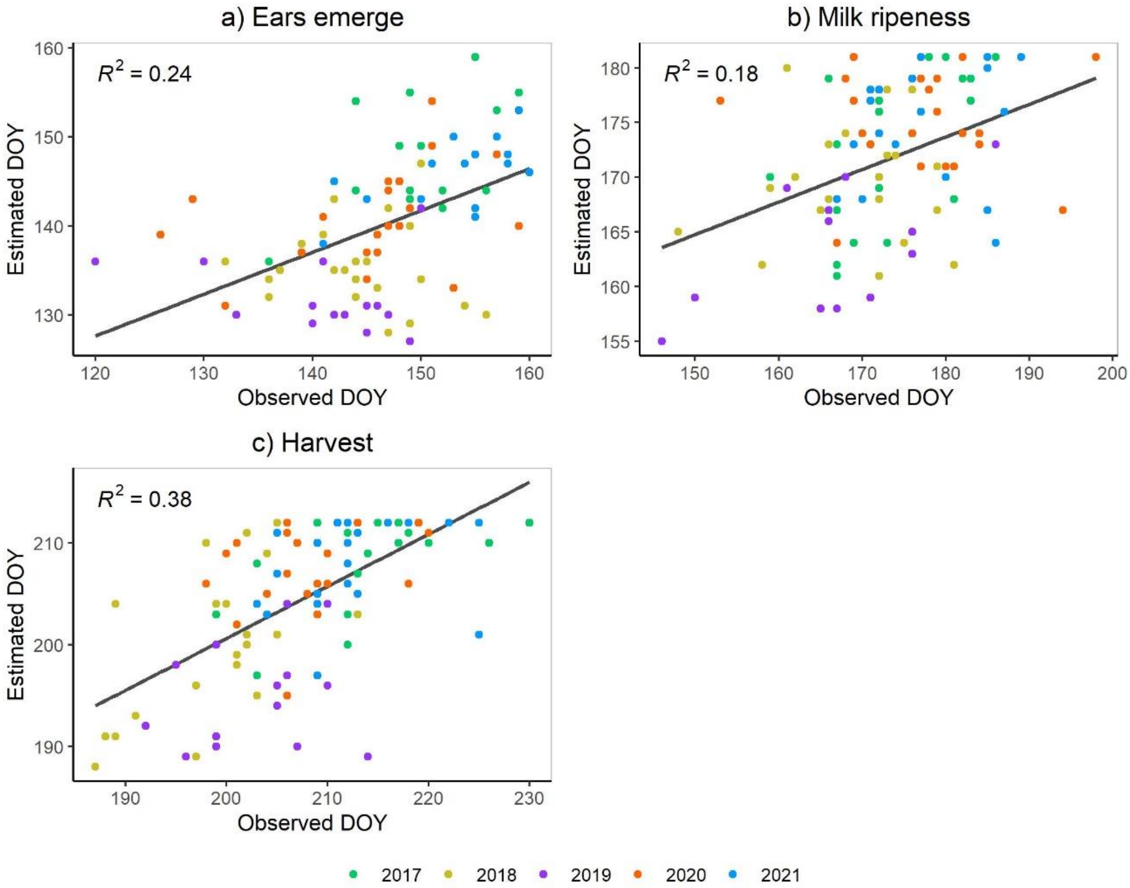


Figure 10 Scatterplot between the mean phenometrics and DWD phenophases of the phenology stations. Ears emerge is estimated with EVI POP. Milk ripeness and harvest were estimated with third and fourth breakpoints respectively

Among the three estimated phenometrics, harvest showed the lowest median and RMSE across all years (**Figure 9**). On the other hand, ears emerge showed highest RMSE values among the three phases (between 10.8 to 13.5 days), whereas for milk ripeness, RMSE values were between 8.7 and 12.5 days. Median values from both the milk ripeness and harvest phases are under five days showing the consistency of the calculated metrics from CR breakpoints. In the case of ears emerge, the median dates are higher than the other phases (between eight and ten days for 2018, 2020, and 2021). With only one exception, the median values from all the phases are positive, indicating that the calculated dates are generally earlier than the actual dates. The R^2 value was highest (0.38) for harvest detection (**Figure 10**) than the other two metrics (0.24 for milk ripeness and 0.18 for ears emerge). Exemplary temporal profiles together with estimated metrics and reported phases for 2017, 2018, and 2020 are shown in **Figure 11** and 2019 and 2021 are shown in **Annex 1**.

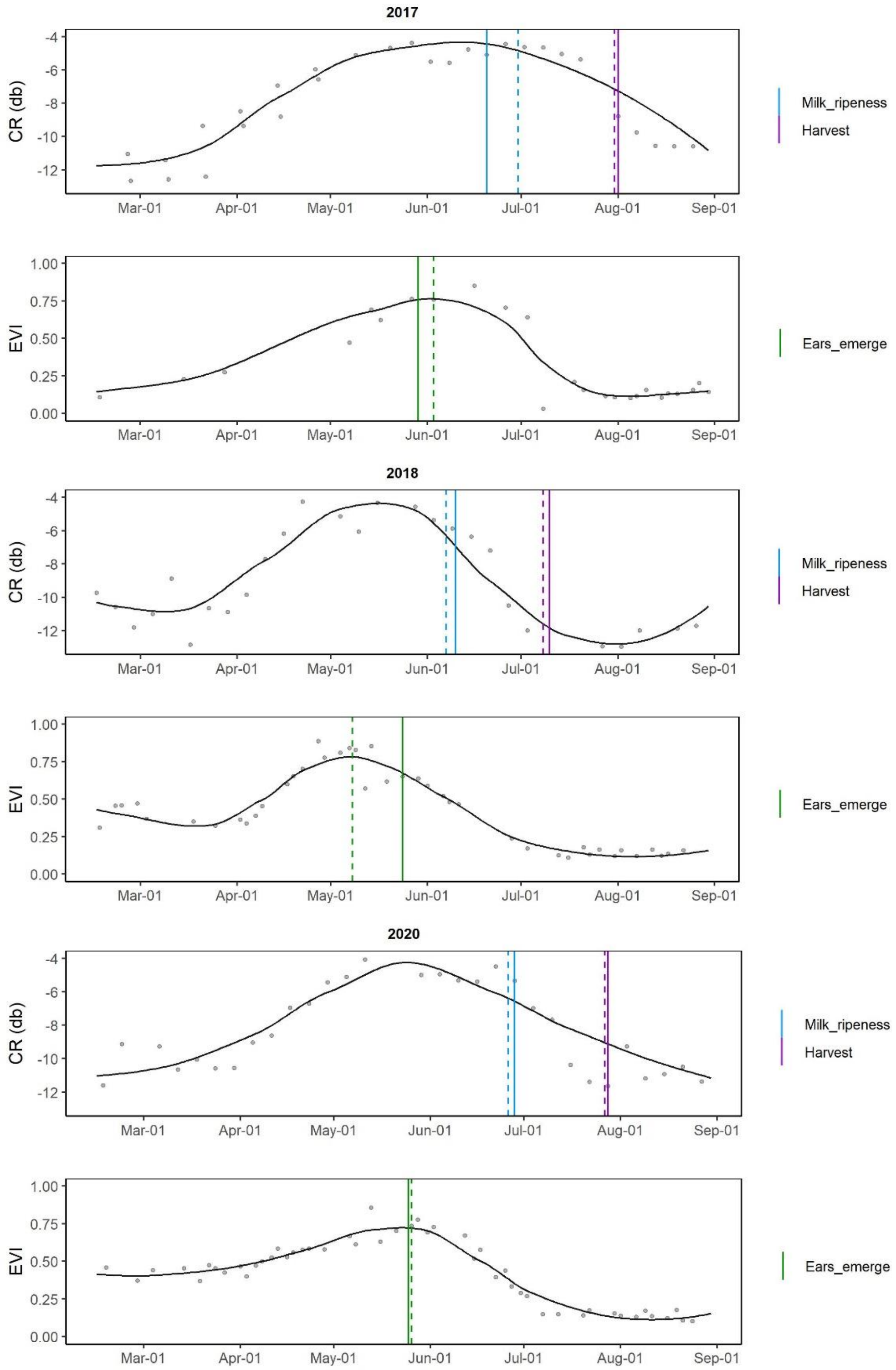


Figure 11 Exemplary temporal profiles with calculated metrics for 2017, 2018, and 2020. Remote sensing metrics are shown with dashed line and reported phases are represented by solid line.

5.3. Retrieval and mapping of the derived phenometrics

The retained phenometrics from the calibration and validation process are retrieved for all the wheat fields for the five studied years. Results in this section are presented for each of the three phenological phases individually. The phases are mapped showing all five years side by side to show the spatial and temporal variability of each phase.

5.3.1. Ears Emerge – BBCH 51

Figure 12 shows the variability of ears emerge in terms of day of year (DOY) across the five years. In general, ears emerge for wheat occurs from beginning of May till mid-June. As **Figure 12a** suggests, ears emerge detected by POP (EVI) have high spatial and temporal variability. The extreme high (higher than DOY 170/June 19) or low value (less than DOY 120/April 30) in the range might be the outliers that were discussed in the previous section (5.2) and could be flagged as wrong. For most of the fields of 2017 and 2021, the phase occurs later than the other three years (**Figure 12b**). On the other hand, 2018 and 2019 experienced this phase earlier than the other three years for almost all the fields.

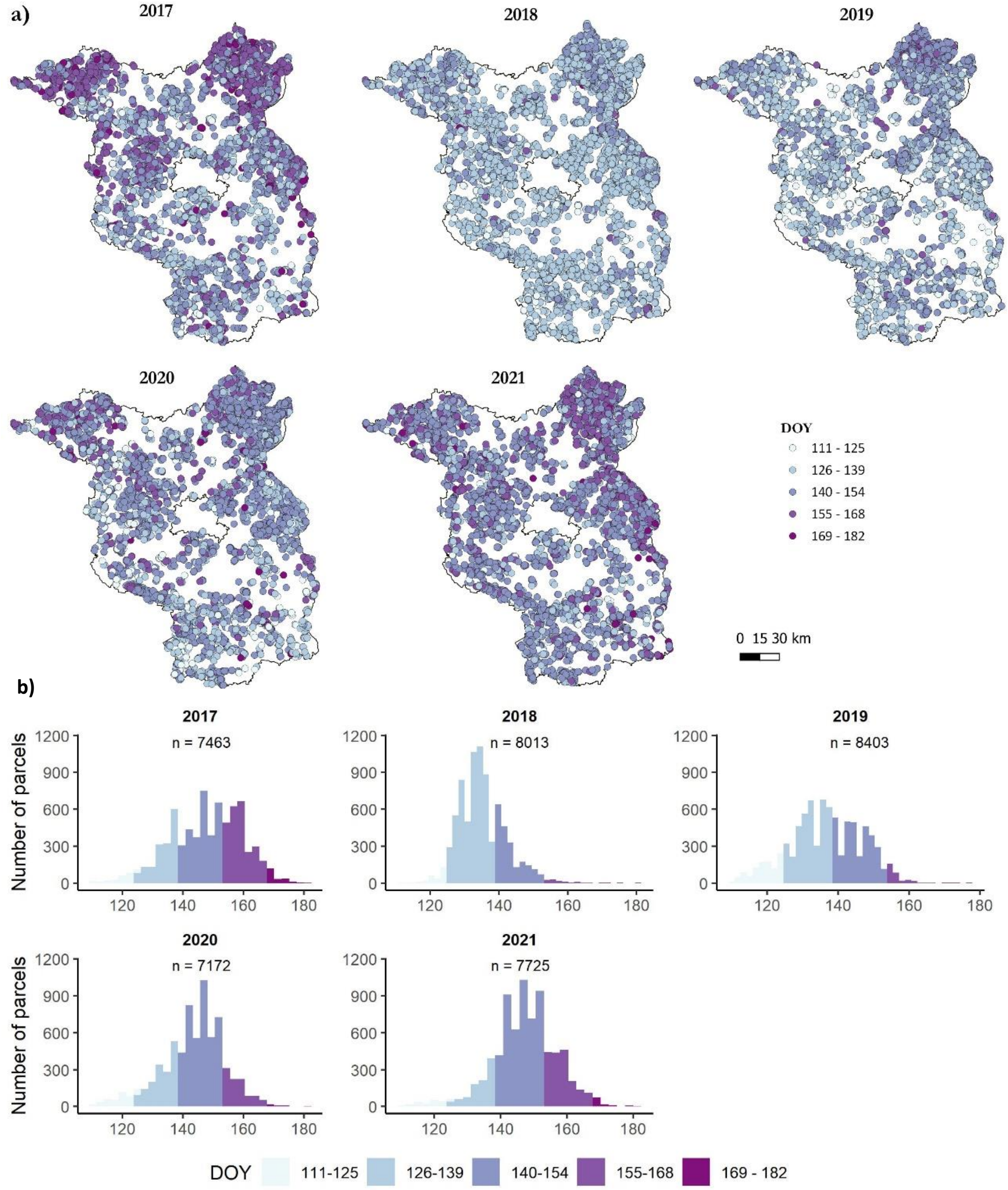


Figure 12 Ears emerge (Day of Year; DOY) as obtained from POP EVI **a)** shows DOY of all the classified parcels (shown with points) **b)** histogram showing the number of parcels belong to each class. n denotes the number of parcels

A subtle south to north gradient is present in 2017, 2020, and 2021 where more darker points appeared in the north than in the south. No clear pattern is visible for the other two years.

5.3.2. Milk ripeness – BBCH 75

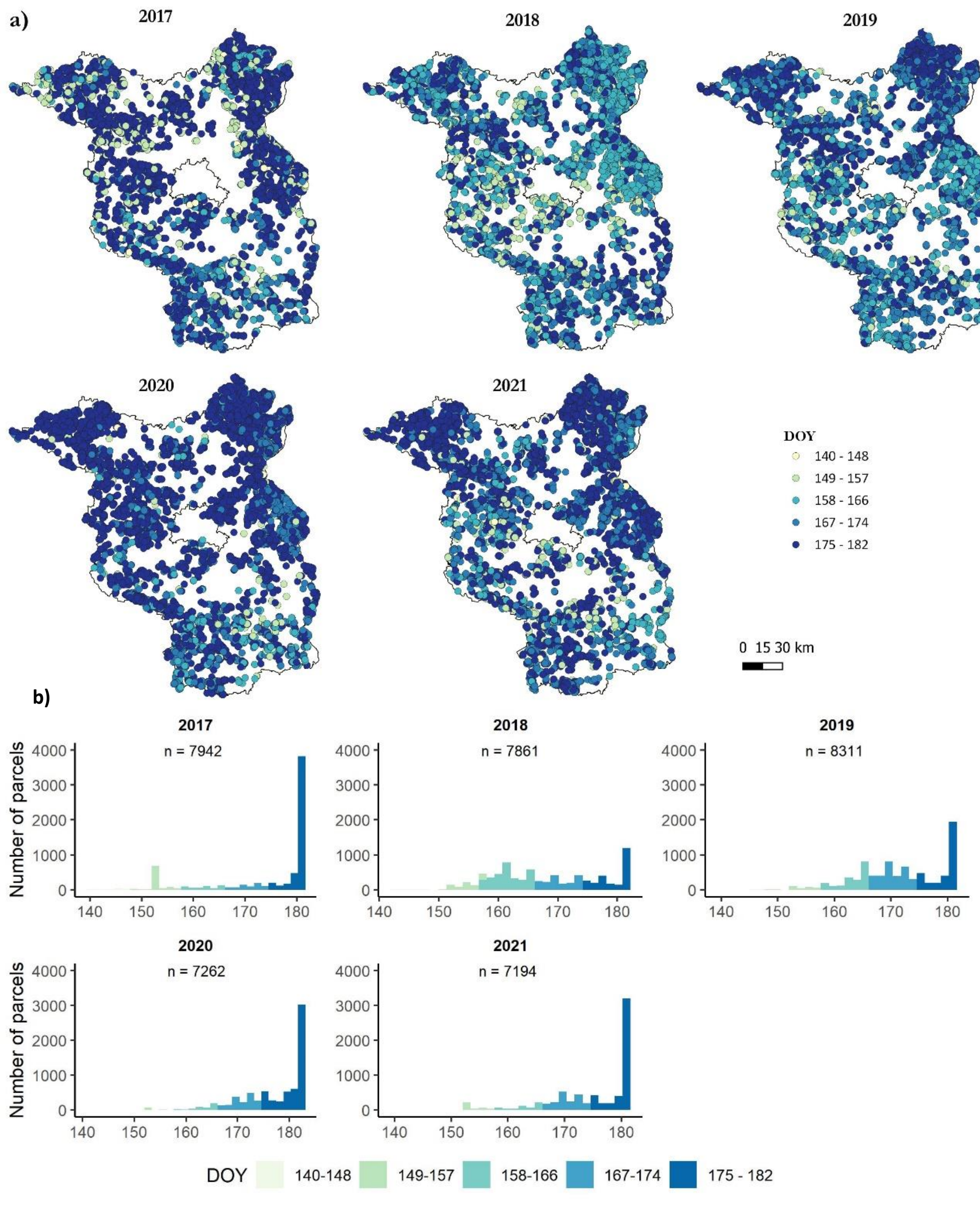


Figure 13 Milk ripeness (DOY) as obtained from third breakpoint of CR **a)** shows DOY of all the classified parcels (shown with points) **b)** histogram showing the number of parcels belong to each class

The third breakpoint generally detected this phase during the first week of June (DOY 152) until the end of the month. Over half of the plots across 2017, 2020, and 2021 estimated breakpoint at DOY 181/June 30 (**Figure 13b**). For 2018 and 2019, more retrievals were recorded during mid-June than in the other three years. No clear south-north gradient was visible for 2017 and 2018. The other three years showed a similar south-north gradient as the ears emerge phase, with the south having earlier retrievals than the north (**Figure 13a**).

5.3.3. Harvest – BBCH 99

The majority of the fields of 2017 are estimated to be harvested at the end of July (DOY 212/ July 31). Similar to the milk ripeness phase, 2018 and 2019 showed more spread than the other three years with more fields with earlier harvest retrievals (beginning and mid-July). The years 2020 and 2021 showed a similar pattern, with more than half of the fields estimated to be harvested at the end of July (**Figure 14**).

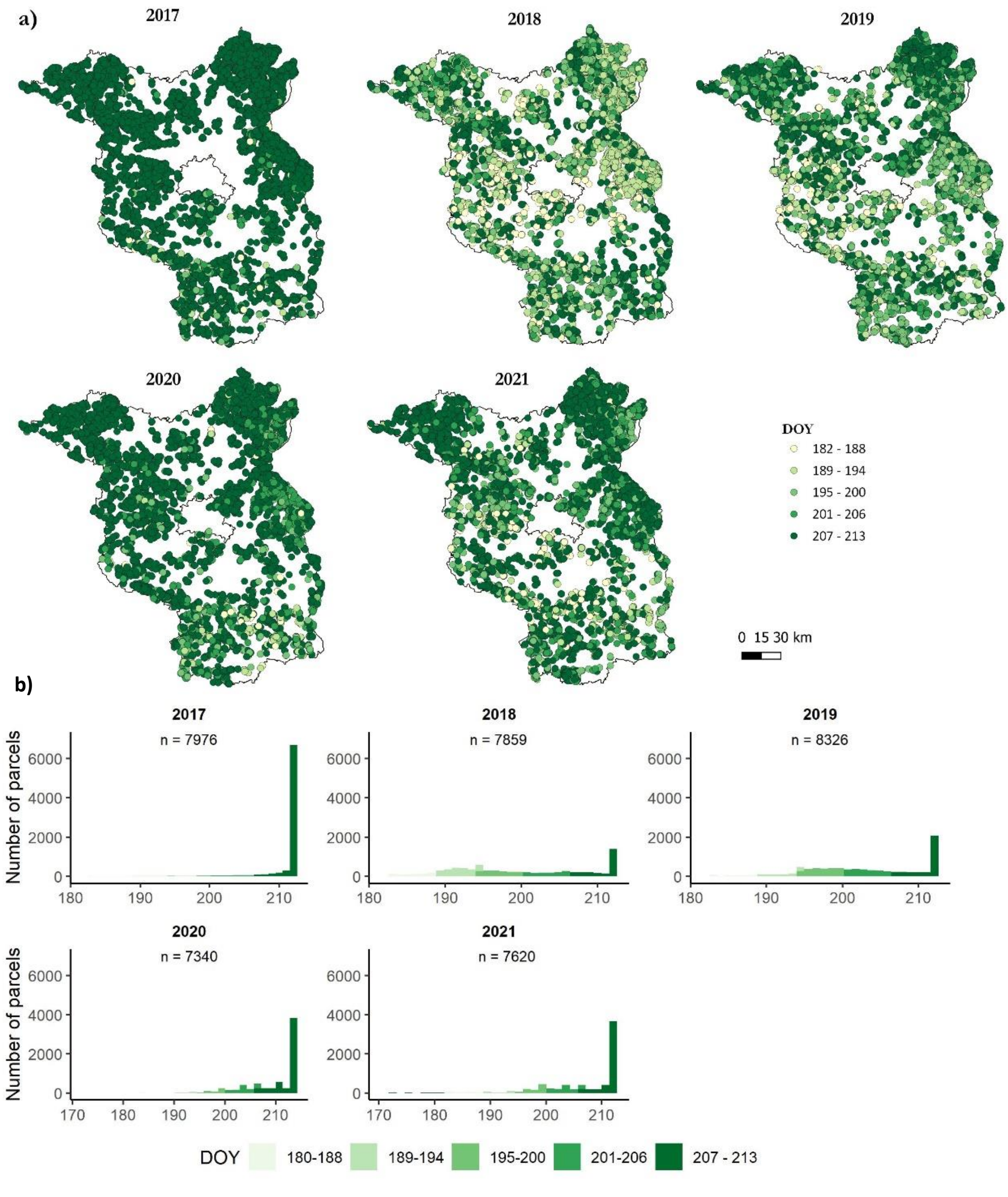


Figure 14 Harvest (DOY) as obtained from fourth breakpoint of CR **a)** shows DOY of all the classified parcels with points **b)** histogram showing the number of parcels belong to each class

5.4. Spatial and temporal variations at station level

To understand the spatial variability of wheat within the buffer of DWD stations, a station is selected randomly which reported the studied phases for the five years, and the temporal differences are mapped. Maps of 2017, 2018, and 2020 are shown in **Figure 15**, and the other two years (2019 and 2021) are shown

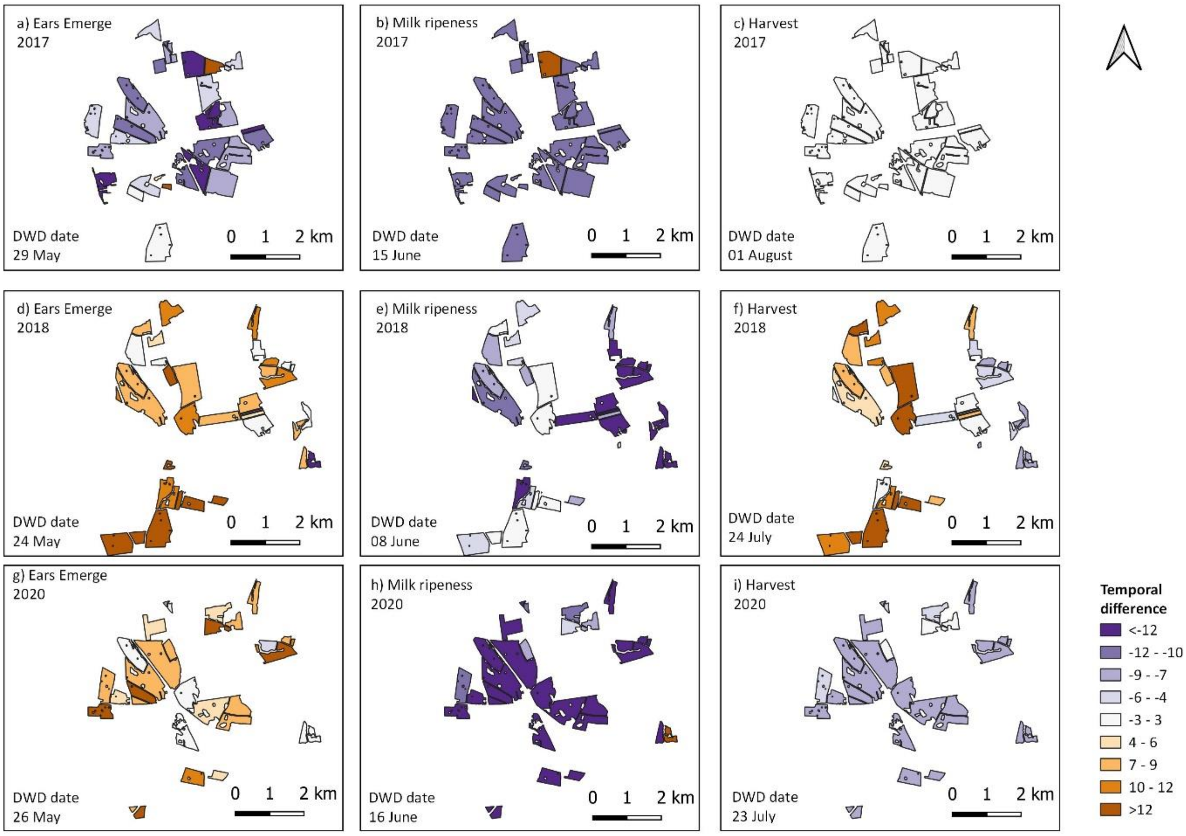


Figure 15 Temporal differences in days between the remote sensing estimated phenology dates and station reported days. Fields are selected around 5-km buffer of 'Lunow' phenology station (station id: 15469) which is selected randomly. 2017, 2018 and 2020 are shown because of their distinct nature in meteorological conditions. DWD date refers to the station reported date of the corresponding phase

in **Annex 7**. Except for milk ripeness and harvest phase of 2017, the other stages show spatial variability (**Figure 15**). The spatial patterns vary from station to station without following any clear pattern. As the fields change more often due to crop rotation every year, temporal variations were not clearly seen at this scale.

5.5. Relationship between the three retreived phases and weather conditions

A summary table of the results of regression analysis between different combination of months of weather and phenometrics are shown in **Annex 5**. As it suggests, temperature and precipitation data of April had the highest R^2 values with ears emerge; hence April was taken to model the relationship between ears emerge and weather. For milk ripeness, aggregation of April, May, and June were chosen, and June-July were taken for harvest. Overall, a negative correlation ($p < 0.001$) is seen between the phases and temperature (R^2 0.24, 0.02, and 0.25 for ears emerge, milk ripeness and harvest, respectively, as shown in **Table 4**). For example, the majority of the grids of 2017 and 2021 reported milk ripeness later than 2018 and 2019. 2018 and 2019 experienced higher April temperatures than 2017 and 2021 (**Figure 17**). A similar temporal variability among the phenometrics and temperature is evident for milk ripeness and harvest phase (**Annex 3** and **Annex 4**). An opposite but weaker relationship ($p<0.001$) is noticed for the year-to-year variability among the phases and precipitation where high precipitation leads to delay in stages (R^2 0.02, 0.03, and 0.13 for ears emerge, milk ripeness, and harvest, respectively).

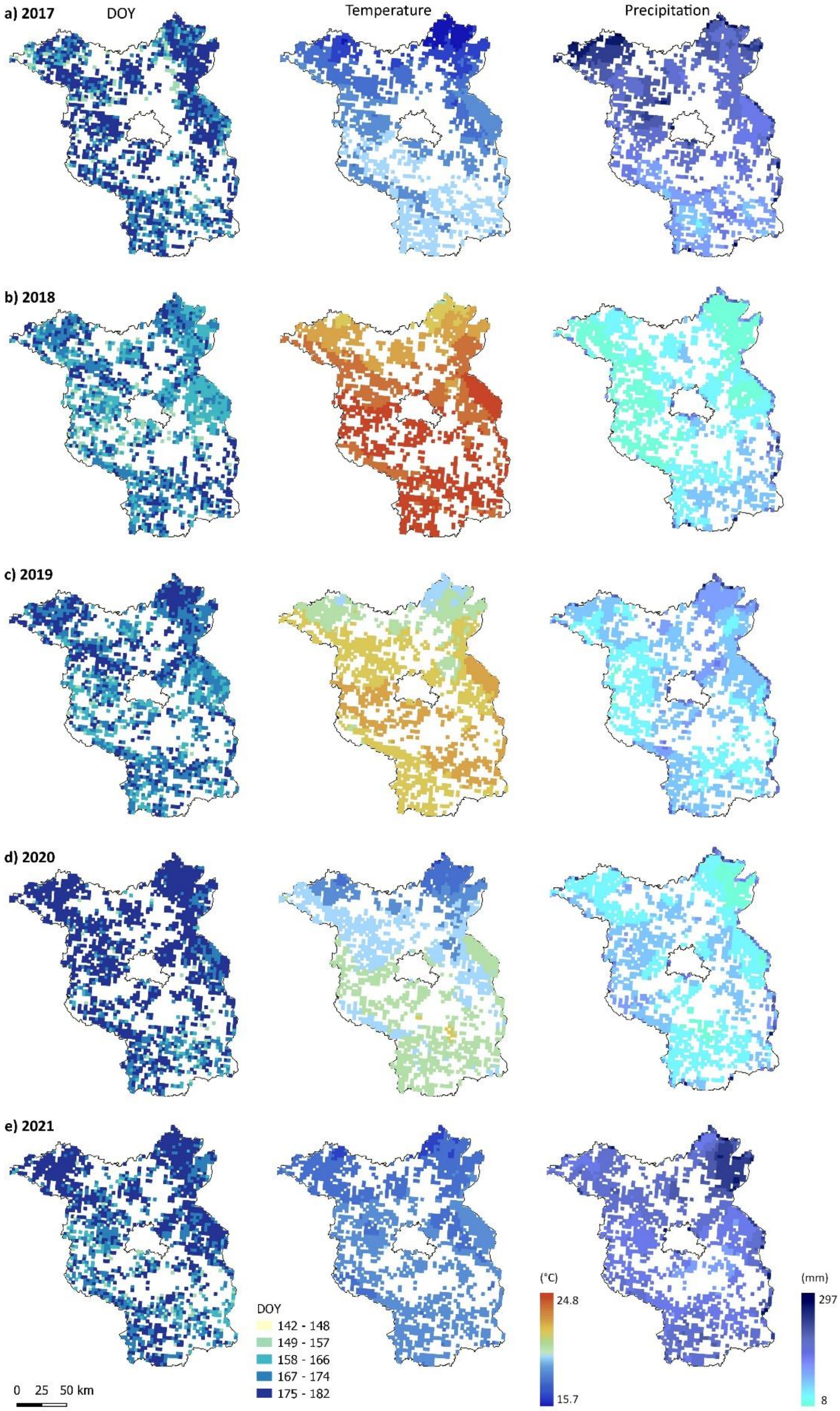


Figure 16 Spatial and temporal variability of milk ripeness (BBCH 75) with corresponding temperature and precipitation. Temperature is shown as the monthly average of daily maximum for April, May, and June. And precipitation is shown as sum of the three months.

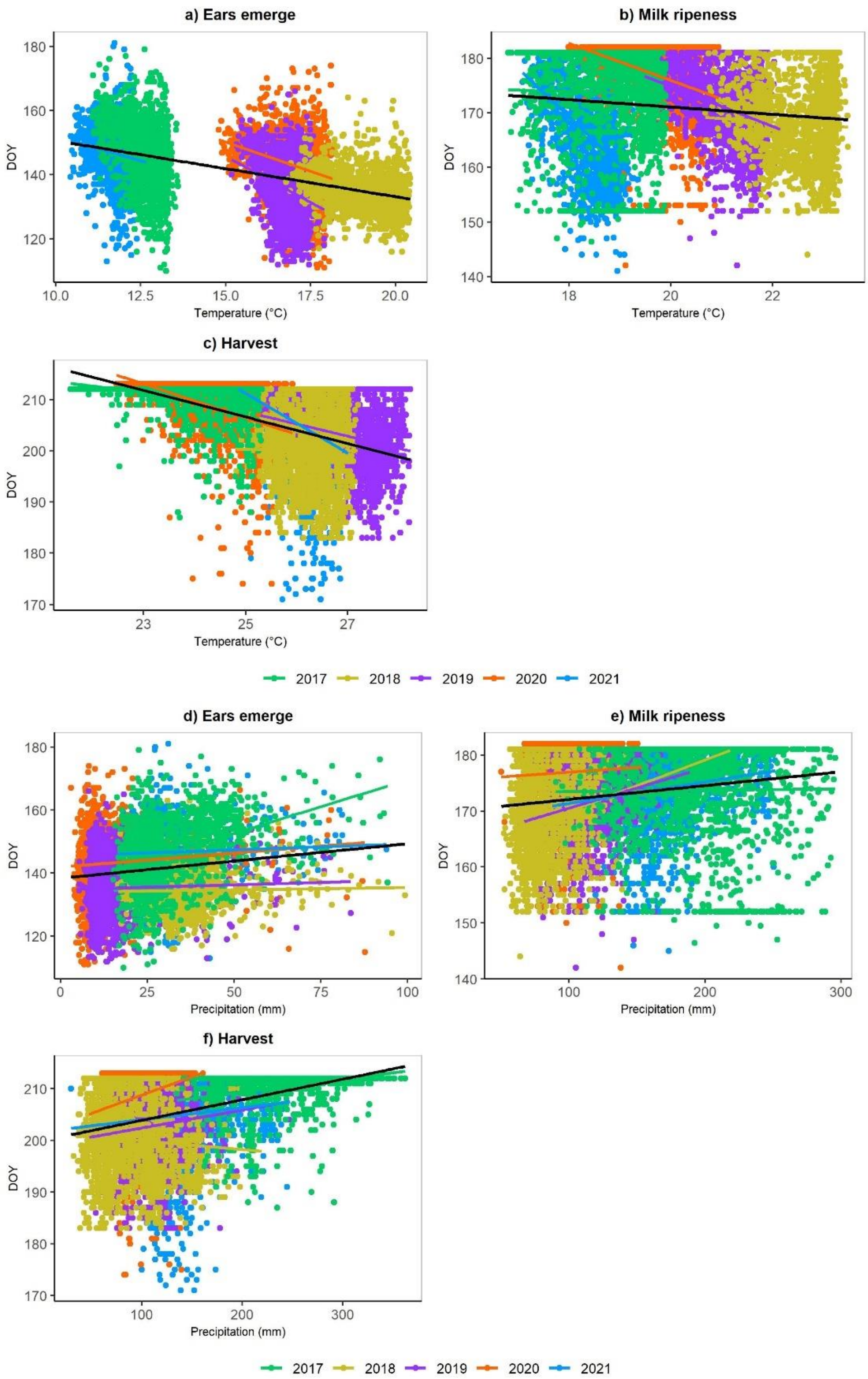


Figure 17 Scatterplot showing the relationship between estimated phases and Temperature (**a, b, c**) and precipitation (**d, e, f**). An overall (shown in black line) and individual regression lines are fitted for each years (shown with the corresponding color of the year).

For ears emerge and harvest, the overall relationship between the two variables is stronger than induvial years (**Table 4**). This shows that within-year variability was modelled poorly with temperature and precipitation data. Although looking at the spatial variability of the phenometrics and temperature, a similar south-north gradient is evident for ears emerge 2017, 2018, 2019, and 2020 (**Annex 3**), where the north experienced cooler temperatures than the south leading to more earlier retrievals in the south. A similar pattern emerged for harvest (2018, 2019, 2020, and 2021 shown in **Annex 4**) and milk ripeness (2019, 2020 and 2021, shown in **Figure 16**).

Table 4 Summary statistics of the relationship between estimated phenometrics and the weather parameters (temperature and precipitation). R^2 is the coefficient of determination, p is significance level of the linear regression between the dependent (phenometrics) and independent (weather parameters) variables. Non-significant regressions ($p > 0.01$) are shown in bold

		Temperature		Precipitation	
	Years	R^2	p	R^2	p
Ears emerge	Overall	0.24	<0.001	0.02	<0.001
	2017	0.15	<0.001	0.14	<0.001
	2018	0.09	<0.001	**<0.01**	**0.329**
	2019	0.08	<0.001	**<0.01**	**0.215**
	2020	0.08	<0.001	<0.01	<0.001
	2021	0.02	<0.001	**<0.01**	**0.095**
Milk ripeness	Overall	0.02	<0.001	0.03	<0.001
	2017	**< 0.01**	**0.426**	**<0.01**	**0.492**
	2018	<0.01	<0.001	0.03	<0.001
	2019	0.05	<0.001	0.03	<0.001
	2020	0.11	<0.001	**<0.01**	**0.521**
	2021	0.11	<0.001	0.02	<0.001
Harvest	Overall	0.25	<0.001	0.13	<0.001
	2017	0.11	<0.001	0.09	<0.001
	2018	**<0.01**	**0.748**	<0.01	<0.001
	2019	0.06	<0.001	0.01	<0.001
	2020	0.16	<0.001	0.04	<0.001
	2021	0.10	<0.001	<0.01	<0.001

6. DISCUSSION

S1 and S2 dataset has gained popularity in phenology studies because of their sensitivity to vegetation growth and their high spatial and temporal resolutions. Previous studies used these dataset to retrieve phenology in big scale without looking at specific crops. Various studies retrieved crop specific phenology either at a small scale or at big scale with only taking handful of parcels. This study aimed to filling this knowledge gap by considering all the crop parcels of Brandenburg state of Germany for five years. It demonstrates that S1 and S2 timeseries can be used to accurately derive wheat phenology for a large area. This section discusses the findings of the research in light of previous literature followed by the limitations of this study and future recommendations.

6.1. Fitting of the data

Visual observations from the exemplary fields show that LOESS fits the data better than DL. As the data are fitted locally in LOESS, more detailed changes were noticed than fitting a global method like DL. For example, early green-up around late December and early January were modelled properly with LOESS (**Figure 6d**) where DL did not incorporate that. LOESS is better suited to identify small changes in a timeseries and can be tuned with ground data during calibration as it offers flexibility to set fitting parameters (Cai et al., 2017). Curve details were not pronounced with DL when there was more noise, and it tended to underfit the data, which is in line with the findings of Cai et al. (2017). Cai et al. (2017) argued that DL is designed to provide better results than local fitting approaches when there are noise and frequent data gaps. However, underfitting were observed with DL in this study and the peaks were not pronounced well when there were large variabilities within the observations. However, results from Cai et al. (2017) are based on MODIS 8-day interval NDVI data (250 m spatial resolution) which might not be transferrable for more frequent S1 and S2 observations, and the noise present in them could be the reason behind the mismatch of the findings. It is worth noting that DL fitting with upper envelope is proven to provide better results making the observations less biased towards (potentially cloud affected for the optical sensor) lower values (Chen et al., 2004). However, the upper envelope function was not used because of its complexity while fitting DL in this study. To conclude, LOESS was observed to capture fine details from time series to detect crop phenology and take advantage of the flexibility to tune the parameters during the calibration process. Hence, LOESS was used to fit the data and detect phenometrics.

6.2. Assessment of the temporal profiles of wheat

Both VV and VH backscatter behaved relatively similarly from the beginning of October until the end of March (before the shooting phase). Wheat plants remain short during this winter dormancy period, and both the SAR backscatter are dominated by soil backscatter (mainly soil moisture and texture) and surface roughness (Veloso et al., 2017). Due to the increase in biomass after the dormancy period, VV and VH backscatter drop between the end of March and mid-May. The decrease in VH was less pronounced than VV, similar to the findings of Mercier et al. (2020). The direct ground and canopy contributions dominate VV backscatter, and the increase in attenuation of the crop's vertical structure is linked to the decrease in VV during this growing stage (Larranaga et al., 2013; Mattia et al., 2003). On the other hand, VH is influenced by volume scattering and the double bounce effect between wheat stems and the ground (Brown et al., 2003; Picard et al., 2003). VH backscatter during the growing period is noisier than VV and substantially different under the three different meteorological conditions (**Figure 7**). This could be explained by the dominance of volume scattering over the double bounce due to increase in fresh biomass (Khabbazan et al., 2019). Nonetheless, the CR constantly increased in this phase, confirming the suitability of using CR as a proxy of fresh biomass. VV backscatter started to increase during the latter half of the season (from the end of May to the end of July) with the decrease of water content of wheat (Veloso et al., 2017; Vreugdenhil et al., 2018). Differences between the years could be explained by their meteorological conditions, as VH backscatter increases more under drier conditions (Veloso et al., 2017). The sudden drop of VV and VH backscatter from late July could be attributed to the harvest of the crop, post-harvest management at the field, and, again, the increased influence of soil as the biomass decreases significantly.

CR appeared to be less sensitive to soil and atmospheric effect than VV or VH backscatter confirming the similar findings of previous literature (Khabbazan et al., 2019; Schlund and Erasmi, 2020; Veloso et al., 2017; Vreugdenhil et al., 2018). It can be assumed that soil and atmospheric effects are similar in both polarizations because of their similar acquisition time. Therefore, the ratio (CR) cancels out the impact of soil or atmosphere. Different temperature and precipitation conditions between the three years were pronounced in their CR and EVI temporal profiles. For instance, the earlier decrease in the 2018 profile after reaching the peak could be attributed to the warmer spring and summer temperature of 2018 than in the other two years. High temperature during the growing season can lead to earlier green-up and senescence of plants leading to a shorter season than usual (Chmielewski et al., 2004; Estrella et al., 2007; Siebert and Ewert, 2012). On the other hand, 2017 reached its peak at the end of June, later than the other two years, which could be attributed to the wet conditions of 2017. EVI followed a similar pattern as CR during the spring season till the ears emerge period. However, EVI started to decrease earlier than CR (**Figure 8**). Being sensitive to chlorophyll content and the greenness of plants, EVI tend to decrease with the drying up of crops while CR remains sensitive with the standing crop biomass confirming the findings of Veloso et al. (2017) and Meroni et al. (2021). Nonetheless, these differing sensitivities of both CR and EVI confirm their suitability for identifying different phenological phases.

6.3. Retrieval of the phenometrics and their links with ground measurement

The breakpoint method showed better results with CR in terms of finding breakpoints closer to the field dates than EVI (**Table 3**). This is consistent with the findings of Harfenmeister et al. (2021), where they had found better accuracies using breakpoints with CR than NDVI. Due to high RMSE and median values, detecting the shooting phase (BBCH 31) was not considered for retrievals in this study (**Table 3**). However Schlund and Erasmi (2020) accurately derived the shooting period with CR by taking a single phenology station. It can be argued that the accuracy decreases on a large scale with more station data. It should be noted that observing the actual start of shooting at field can be difficult, causing subjectivity and error in the data (Siebert and Ewert, 2012). Moreover, higher variability was observed among station data reporting shooting phase (from end of March till the beginning of May), which did not always coincide with the temporal profiles and the estimated metrics.

Previous literature could not detect the ripening phase with CR breakpoint with high consistency (Harfenmeister et al., 2021; Schlund and Erasmi, 2020). Löw et al. (2021) detected later stage of the ripening phase with InSAR coherence and Alpha from S1, which were not used in this study. Harfenmeister et al. (2021) noted that the water content of the plants start to decrease at this phase with very little structural change in the temporal profiles. Hence, the breakpoint from CR was detected later when there were explicit temporal profiles change, resulting in larger deviations between breakpoint and ripening phase (Harfenmeister et al., 2021). Schlund and Erasmi (2020) also concluded that the ripening stage could not be detected with high certainty. However, this study could relate the third breakpoint with the milk ripeness phase. The breakpoint was detected when there was abundant structural change in the temporal profile during the senescence period, when the phase is reported (**Figure 11**). For all five years, less than three days in median values were observed between the third breakpoint and milk ripeness phase (**Figure 9**). However, the R^2 value was only 0.18, which does not coincide with the achieved RMSE or median values for this phase. It is worth mentioning that R^2 is calculated between mean phenometrics from fields within station buffers and DWD phenophases. The poor correlation of determination could be explained by the outlier profiles (**Annex 2**), which makes the mean phenometrics higher resulting in larger deviations.

The most drastic change of the appearance of the fields occur in the harvest date as all the plants in a field are harvested in a day. Generally, the harvest date can vary from field to field depending on the weather conditions, availability of machinery and labour. During this period, backscatter from soil dominates the VV polarization, while very low volume scattering contributes to VH resulting in a breakpoint (Khabbazan et al., 2019). The potential of the final breakpoint from CR to detect the harvest period was well documented by Schlund and Erasmi (2020) with similar degree of accuracies acquired in this study. The final breakpoint could accurately detect the harvest period from CR with the lowest RMSE values among the three phases (8.7, 9.6, 9.2, 7.1, and 8.9 days for 2017 to 2021, respectively). It is worth noting that breakpoint detected both harvest and milk ripeness phase around a single date (June 30 for milk ripeness and July 31 for harvest) for the vast majority of fields (**Figure 13b** and **Figure 14b**). This makes the breakpoint method stable in space and time, which is also consistent with the findings of Harfenmeister et al. (2021) and Schlund and Erasmi (2020). However, the consistency of the phases to be retrieved around a single date is dependent on the LOESS span value. The lower span value of LOESS (i.e. 0.3 or lower) makes the retrieved dates from breakpoints more normally distributed. As the low span value introduces more errors (high RMSE and median values), 0.4 was chosen. However, it can be argued that phenological phases to be reported in a single date for such a high number of fields are not realistic as more variations are noticed due to management decisions and other factors.

Threshold method could not detect SOS and EOS from CR. The reason could be LOESS fitting retaining more curve details and a more complex CR profile than EVI. A higher LOESS span value could solve the problem which was not tested in this study and can be a potential area of future research. However, the EOS metrics (EOS25 and EOS50) and POP with threshold method from EVI had a good match with harvest, yellow ripeness and ears emerge, respectively (**Table 3**). Only the best metric (POP) in terms of RMSE and median values is retained in this study to avoid estimating too many metrics. It is noted from the calibration process that EOS50 is correlated with yellow ripeness while EOS25 corresponds with harvest. Meroni et al. (2021) reported that EOS50 from S2 NDVI had good agreement with ripening stage (BBCH 83), while EOS50 from S1 CR is well correlated with harvest. In this study EOS from CR could not be detected; however, it can be assumed that CR EOS50 could be attributed to a later stage than ripeness (i.e. harvest). This is because while VIs from optical sensors are sensitive to vegetation greenness and chlorophyll content, CR is sensitive to standing biomass (Veloso et al., 2017). Similarly, EVI started to drop early in values while CR remained stable and decreased later than EVI, making the EOS50 more suitable for detecting the harvest period. However, as EOS could not be retrieved from CR, it could not be proved.

During the ears emerge phase (BBCH 51) in the field, a big change can be noticed in wheat plants with emerging heads. Plant height with their biomass and Leaf Area Index (LAI) reach their maximum limit in this period making the peak of the temporal profile a potential metric to detect this phase. Harfenmeister et al. (2021) have found global maxima from both CR and NDVI to be well correlated with this stage (approximately four days for NDVI and 7-9 days for CR). This finding is consistent with the outcome of peak (POP) from EVI of this study. However, RMSE is approximately 12 days for all five years, which could be attributed to the outlier profiles and variabilities of reported dates between the phenology stations. It should be noted that POP from CR and EVI reports similar median values (6 days) and very close RMSE (13 and 11, respectively) while detecting this phase. This signifies the potential of both sensors to detect this phase. The positive median values from 2018, 2020 and 2021 (+8, +3, and +8) suggest that POP detection generally occurs earlier than actual dates. From a similar standpoint, Harfenmeister et al. (2021) explained the suitability of their identified maxima to detect ‘development of flag leave’ (BBCH 47), an earlier stage than BBCH 51. Hence, retrieving this flag leave phase accurately, the timing of ears emerge could be predicted well in advance, which could help farmers in planning ahead to apply fertilizers.

6.4. Linking estimated phenology with weather observations

Spatial and temporal pattern of wheat phenology can vary depending on temperature, precipitation, photoperiod, soil type, and field management practices. In this study, monthly gridded data of precipitation and temperature were used to explain the variations of wheat in space and time. In general, both the weather parameters could explain the inter-annual variations of the estimated phenology. For example, the warmer conditions of 2018 and 2019 can be attributed to the earlier detected dates for all three phases. On the other hand, the wet conditions of 2017 and 2021 can cause the delay in reaching the phenology phases. Schlund and Erasmi (2020) found similar temporal variability when comparing wheat phenology between 2017, 2018, and 2019. Pasqui et al. (2022) studied maize phenology in the Netherlands and concluded that 2018's season was shortened by around 30 days than the preceding year due to the drought of 2018. D'Andrimont et al. (2020) concluded that flowering stage of winter wheat of Germany occurred earlier in 2018 than usual, confirming the findings of this study. However, in this study, spatial variability of estimated phenology could not be well explained by the two weather parameters. Although in some cases, a subtle south-north gradient was noticed among the temperature and phenology maps. For example, a south-north gradient can be seen for all three phase of several years, where more earlier retrievals are present in the south than in the north. The south-north gradient of phenology could be attributed to the similar gradient of temperature where the southern part experienced higher temperature than the northern counterpart. No clear pattern is visible in precipitation maps. Despite the visual gradient, the R^2 values between the estimated phenometrics and temperature for individual years are very low (**Table 4**).

Coarse temporal resolution of the weather data can be the reason behind the low correlation in space (**Figure 17**). Daily data on both precipitation and temperature could be used in future studies. In addition, correlating phenology with growing degree days (GDD) could explain the variability more prominently. GDD is calculated yearly by accumulating daily total temperature which occurs above the base temperature required for the plant growth. It accounts for both temporal and spatial variation of phenology and constraints the thermal variation within which crop development is feasible (Wang et al., 2022). Hence, understanding the spatial variability of estimated phenometrics with GDD could be a possible direction for future research. Other than climatic parameters, soil type, hydrography, topography, and geology could drive crop phenological variability, which were not considered in this study. For example, clayey loam and loamy soils are considered the optimum soil types for wheat growing because their good maintenance between moisture release and retention (Wong and Asseng, 2006). On the other hand, sandy soils have poor water retention rates, which can lead to lower plant productivity (Wong and Asseng, 2006). These factors might impact plant phenology, which can be a potential area for future research. Hence, with these factors being involved, it is difficult to classify that climate is solely responsible for the phenology variations.

6.5. Limitations and recommendations

The study identified three phenological phases with a median value of less than the revisit period of both S1 and S2 sensors (five days) for most of the cases. However, the higher RMSE values could be attributed to several factors which are discussed in the following sections.

First, the outlier profiles (**Annex 2**) significantly influenced the RMSE values. As these outliers do not follow any specific pattern, it was impossible to detect them and remove them prior to the phenology estimation. Although a cloud, cloud shadow, and snow masking algorithm was applied to S2 before retrieving the profiles, there might still be some contaminated pixels which could alter the profiles (Chamnan, 2021). Bolton et al. (2020) described a secondary cloud screening techniques to remove affected observations for optical sensors which could potentially decrease the noise from S2 data. However, to the best of my knowledge, no specific method exists to detect abnormal CR profiles from SAR data for crop. Hence, more research is needed to identify noisy temporal profiles of specific crops. Furthermore, taking data from both the viewing geometries could potentially add more noise to the S1 CR data. The descending mode passes over Europe in the early morning and ascending in the evening. Dew in the morning can affect the backscatter values (Khabbazan et al., 2019) of the descending mode. However, Meroni et al. (2021) argued that no significant differences were observed while using single viewing geometries than using both. The effect of the combined use of both the viewing geometries on phenology of crops should be studied further.

In addition to the outlier temporal profiles, biased or wrong reported information from DWD could contribute to the higher RMSE. Siebert and Ewert (2012) found that 21% of the reported observations were questionable for the whole of Germany between 1959 and 2009. According to the instructions from DWD, the observations should preferably be made from fields within one to two kilometres while not exceeding five kilometres. Although the number of fields to be observed is not explicitly mentioned, Harfenmeister et al. (2021) argued that the observations are based on one to four fields in most cases. Hence, the transferability of a few observations to a 5-km buffer area can be questionable due to different management decisions and soil types per fields. Furthermore, **Figure 15** confirms that high spatial variability exists within the 5-km buffer of a station. Hence this mismatch could contribute to increasing the RMSE values. Taking a smaller buffer (one to two km) around the station could reduce the uncertainty while significantly decreasing the number of calibration and validation fields. Another possible workaround would be to use remote sensing metrics within buffer to further make a quality check of the DWD observations. Stations showing high variability from the mean estimated metrics could be flagged as uncertain. This could be taken into consideration while using DWD observations for future research.

Another source of errors might be coming from the crop type map from LPIS data. False claims from farmers and digitization errors of the data are not uncommon in this dataset which might lead to a wrong crop type map (Blickensdörfer et al., 2022). Hence, some outlier profiles might be attributed to a different crop than winter wheat. Furthermore, LPIS covers lands only when the landowner or the farmer applies for subsidies. As not all the farmers or landowners might do that, LPIS does not cover all the fields within an area. Future research could look into a possible cross-quality check of crop parcel by comparing it from a different sourced dataset.

Another possible source of error might be the varying number of estimated breakpoints per field. In general, four breakpoints were retrieved during the growing season (from March to August). The breakpoints were numbered in ascending order (from one to four), while the third and fourth breakpoints were assigned to relate them to the in-situ phases. However, in some cases, an additional or only three breakpoints were retrieved due to outlier profiles. These anomalies confused the numbering scheme and added errors to the retrievals because a wrong breakpoint was classified as a specific phase. However, it should be noted that only three cases like this were identified during the calibration process when a wrong breakpoint was assigned to a stage. Nonetheless, future work should programmatically set the numbering based on the known period where a breakpoint is expected to occur.

In this study, wheat phenology is estimated using S1 and S2 sensors separately. Future studies could combine both sensors using data fusion, enabling a single data source to seamlessly retrieve crop phenology that remains relatively underexplored to date. Furthermore, S2 offers three bands (band 5, 6, and 7) in the red edge region (i.e. the sharp slope between low reflectance in the red and high reflectance in NIR region) in the spectral reflectance curve at 20 m spatial resolution. The red edge bands have proven to be significantly correlated with LAI and the growth of vegetation while being less sensitive to atmosphere and soil effects

(Baret et al., 1992; Darvishzadeh et al., 2009). Hence, future studies could focus on the rich spectral information from the red edge bands of S2 to detect the phenology of crop parcels.

7. CONCLUSION

This study retrieved crop-specific phenology of winter wheat from S1 and S2 time series for Brandenburg, Germany, from 2017 to 2021. First of all, temporal profiles of three SAR parameters from S1 (VH, VV, and CR) and EVI from S2 were explored, and their physical characteristics were explained against the DWD in-situ phenology data. CR and EVI matched the expected phenology of the crops, and their profiles behaved more consistently than VH or VV alone. Secondly, breakpoint and threshold method were used with LOESS fitted data to estimate phenometrics. Calibration and validation of the phenometrics were done prior to the retrievals using parcels from a 5-km buffer of DWD phenology stations. The retrieved phenometrics were mapped to show their spatial and temporal variations. Finally, phenometrics and weather data (temperature and precipitation) were aggregated at a 3 × 3 km grid to relate to each other. Linear regression was performed between the two variables to set the relations. The conclusion of the research based on the research questions are presented below:

Q1) Do field-specific temporal patterns of Sentinel-1 derived VV, VH, and VH/VV (CR) backscatter match the expected phenology of the crop?

Three of the SAR parameters proved to be sensitive to wheat growth stages. CR showed less sensitivity to soil and atmospheric effect than VH or VV, making it suitable for tracking vegetation development. Different phases of wheat, i.e. development, maturity, and senescence could be identified from the CR temporal profile.

Q2) Do field-specific temporal patterns of Sentinel-2 derived EVI provide a clear temporal signal that matches the expected phenology of the crop?

EVI shows a clear pattern, increasing with the crops chlorophyll content and reaching its peak around the ears emerge phase. Then it starts to decline till the crop is harvested.

Q3) What are the observed differences between S1 and S2 derived temporal patterns?

EVI followed similar patterns as CR, increasing during the first half of the growing season until the crop reaches the ears emerge phase. However, after reaching the peak, EVI starts to drop earlier than CR as EVI is sensitive to the chlorophyll content of the crop while CR is sensitive to standing biomass. Moreover, EVI drops faster and abruptly, while for CR, the decline is slower and more subtle.

Q4) Which in-situ observed phenophases correspond with phenometrics derived from S1 and S2 temporal profile?

From the calibration, it was found that ears emerge correspond with S2 EVI (RMSE 11, median 6). Milk ripeness and harvest correspond with S1 CR (RMSE 11.2, median 2 for milk ripeness; RMSE 9.2, median 5 for harvest).

Q5) How well do the retrieved phenometrics correspond to DWD phenophases?

The retrieved phenometrics had median values less than five for most of the phases. However, due to the presence of outlier temporal profiles RMSE values were between eight and twelve. The R^2 values between the mean phenometrics and DWD phenophases were 0.24 for ears emerge, 0.18 for milk ripeness, and 0.38 for harvest.

Q6) How do the phenometrics vary spatially and temporally?

In most of the cases, a south-north gradient was evident in the phenometrics maps (ears emerge 2017, 2018, 2020, 2021; milk ripeness 2019, 2020, 2021; harvest 2018, 2019, 2020, 2021). In these cases, south had more earlier retrievals than the north. Temporal variations were very clear for all three phases. For 2018 and 2019, earlier retrievals were observed than usual, where for 2017 and 2021, late retrievals were noticed.

Q7) Can the retrieved spatial and temporal variability of the phenometrics be explained by the weather data?

A overall negative correlation was evident between the retrieved phenometrics and temperature (R^2 0.24 for ears emerge, 0.02 for milk ripeness, and 0.25 for harvest). Although the R^2 for the spatial variability was very low, a similar south-north gradient was visible for the temperature and phenometrics maps. The interannual variability between the phenometrics and precipitation was a weak positive correlation (R^2 0.02 for ears emerge, 0.03 for milk ripeness and 0.13 for harvest).

In conclusion, S1 CR and S2 EVI can be used to accurately retrieve winter wheat phenology over a large geographical region. While the outlier temporal profiles significantly altering the accuracies in terms of RMSE values, median values remain under five days for most of the cases. The extracted results can be used as inputs for crop growth monitoring, yield estimation which are considered very crucial for food security. The extracted results can improve our understanding of variability of crop phenology and their drivers under the climate change scenario.

LIST OF REFERENCES

Anderson, M.C., Zolin, C.A., Sentelhas, P.C., Hain, C.R., Semmens, K., Tugrul Yilmaz, M., Gao, F., Otkin, J.A., Tetrault, R., 2016. The Evaporative Stress Index as an indicator of agricultural drought in Brazil: An assessment based on crop yield impacts. *Remote Sensing of Environment 174*, 82–99. https://doi.org/10.1016/J.RSE.2015.11.034

Bai, J., Perron, P., 2003. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics 18*, 1–22. https://doi.org/10.1002/JAE.659

Baret, F., Jacquemoud, S., Guyot, G., Leprieur, C., 1992. Modeled analysis of the biophysical nature of spectral shifts and comparison with information content of broad bands. *Remote Sensing of Environment 41*, 133–142. https://doi.org/10.1016/0034-4257(92)90073-S

Bargiel, D., 2017. A new method for crop classification combining time series of radar images and crop phenology information. *Remote Sensing of Environment 198*, 369–383. https://doi.org/10.1016/J.RSE.2017.06.022

Beillouin, D., Schauberger, B., Bastos, A., Ciais, P., Makowski, D., 2020. Impact of extreme weather conditions on European crop production in 2018. *Philosophical Transactions of the Royal Society B 375*. https://doi.org/10.1098/RSTB.2019.0510

Berry, P.M., Sterling, M., Baker, C.J., Spink, J., Sparkes, D.L., 2003. A calibrated model of wheat lodging compared with field measurements. *Agricultural and Forest Meteorology 119*, 167–180. https://doi.org/10.1016/S0168-1923(03)00139-4

Beurs, K.M. De, Henebry, G.M., 2005. Land surface phenology and temperature variation in the International Geosphere–Biosphere Program high-latitude transects. *Global Change Biology 11*, 779–790. https://doi.org/10.1111/J.1365-2486.2005.00949.X

Blickensdörfer, L., Schwieder, M., Pflugmacher, D., Nendel, C., Erasmi, S., Hostert, P., 2022. Mapping of crop types and crop sequences with combined time series of Sentinel-1, Sentinel-2 and Landsat 8 data for Germany. *Remote Sensing of Environment 269*, 112831. https://doi.org/10.1016/J.RSE.2021.112831

Bolton, D.K., Gray, J.M., Melaas, E.K., Moon, M., Eklundh, L., Friedl, M.A., 2020. Continental-scale land surface phenology from harmonized Landsat 8 and Sentinel-2 imagery. *Remote Sensing of Environment 240*, 111685. https://doi.org/10.1016/J.RSE.2020.111685

Bouchet, A.-S., Laperche, A., Bissuel-Belaygue, C., Snowdon, R., Nesi, N., Stahl, A., 2016. Nitrogen use efficiency in rapeseed. A review. *Agronomy for Sustainable Development 36*, 1–20. https://doi.org/10.1007/S13593-016-0371-0

Brown, S.C.M., Quegan, S., Morrison, K., Bennett, J.C., Cookmartin, G., 2003. High-resolution measurements of scattering in wheat canopies - Implications for crop parameter retrieval. *IEEE Transactions on Geoscience and Remote Sensing 41*, 1602–1610. https://doi.org/10.1109/TGRS.2003.814132

Buras, A., Rammig, A., S. Zang, C., 2020. Quantifying impacts of the 2018 drought on European ecosystems in comparison to 2003. *Biogeosciences 17*, 1655–1672. https://doi.org/10.5194/BG-17-1655-2020

Cai, Z., Jönsson, P., Jin, H., Eklundh, L., 2017. Performance of smoothing methods for reconstructing NDVI time-series and estimating vegetation phenology from MODIS data. *Remote Sensing 9*(12), 1271. https://doi.org/10.3390/RS9121271

Calicioglu, O., Flammini, A., Bracco, S., Bellù, L., Sims, R., 2019. The future challenges of food and agriculture: an integrated analysis of trends and solutions. *Sustainability 11*(1), 222. https://doi.org/10.3390/SU11010222

Chamnan, K., 2021. Yield monitoring with Sentinel-2: A first assessment for The Netherlands. In *University of Twente Faculty of Geo-Information and Earth Observation (ITC).*

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., Eklundh, L., 2004. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote Sensing of Environment 91*, 332–344. https://doi.org/10.1016/J.RSE.2004.03.014

Cheng, Y., Vrieling, A., Fava, F., Meroni, M., Marshall, M., Gachoki, S., 2020. Phenology of short vegetation cycles in a Kenyan rangeland from PlanetScope and Sentinel-2. *Remote Sensing of Environment 248*, 112004. https://doi.org/10.1016/J.RSE.2020.112004

Chmielewski, F.M., Müller, A., Bruns, E., 2004. Climate changes and trends in phenology of fruit trees and field crops in Germany, 1961–2000. *Agricultural and Forest Meteorology 121*, 69–78. https://doi.org/10.1016/S0168-1923(03)00161-8

Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, *74*(368), 829–836. https://doi.org/10.1080/01621459.1979.10481038

d'Andrimont, R., Taymans, M., Lemoine, G., Ceglar, A., Yordanov, M., van der Velde, M., 2020. Detecting flowering phenology in oil seed rape parcels with Sentinel-1 and -2 time series. *Remote Sensing of Environment 239*, 111660. https://doi.org/10.1016/J.RSE.2020.111660

Darvishzadeh, R., Atzberger, C., Skidmore, A.K., Abkar, A.A., 2009. Leaf Area Index derivation from hyperspectral vegetation indicesand the red edge position. *International Journal of Remote Sensing 30*, 6199–6218. https://doi.org/10.1080/01431160902842342

Delbart, N., Picard, G., le Toan, T., Kergoat, L., Quegan, S., Woodward, I., Dye, D., Fedotova, V., 2008. Spring phenology in boreal Eurasia over a nearly century time scale. *Global Change Biology 14*, 603–614. https://doi.org/10.1111/J.1365-2486.2007.01505.X

DESTATIS-Statistisches Bundesamt. Anbaufläche von Winterweizen im Jahr 2020 um 7% Gesunken, 2020. URL https://www.destatis.de/DE/Presse/Pressemitteilungen/2020/05/PD20_168_412.html (accessed 5.28.22).

DWD, 2021a. Wetter und Klima - Deutscher Wetterdienst - Berlin und Brandenburg. URL https://www.dwd.de/EN/weather/weather_climate_local/berlin-brandenburg/bbb_node.html (accessed 10.20.21).

DWD, 2021b. Wetter und Klima - Deutscher Wetterdienst - Leistungen - Klimakarten des Dürreindex. URL https://www.dwd.de/DE/leistungen/rcccm/int/rcccm_int_spi.html (accessed 10.20.21).

DWD, 2015. Wetter und Klima - Deutscher Wetterdienst - Daten Deutschland DWD - Anleitung zur phänologischen Beobachtung. URL https://www.dwd.de/DE/klimaumwelt/klimaueberwachung/phaenologie/daten_deutschland/beobachtersuche/beobachteranleitung_download.pdf (accessed 10.20.21).

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 18*, 656–674. https://doi.org/10.1111/J.1365-2486.2011.02521.X

Estrella, N., Sparks, T.H., Menzel, A., 2007. Trends and temperature response in the phenology of crops in Germany. *Global Change Biology 13*, 1737–1747. https://doi.org/10.1111/J.1365-2486.2007.01374.X

Fan, X., Vrieling, A., Muller, B., Nelson, A., 2020. Winter cover crops in Dutch maize fields: Variability in quality and its drivers assessed from multi-temporal Sentinel-2 imagery. *International Journal of Applied Earth Observation and Geoinformation 91*, 102139. https://doi.org/10.1016/j.jag.2020.102139

FAO, 2017. The future of food and agriculture–Trends and challenges. *Annual Report*, *296*, 1-180.

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N., Snyder, P.K., 2005. *Global consequences of land use. Science (1979) 309*, 570–574. https://doi.org/10.1126/SCIENCE.1111772/SUPPL_FILE/FOLEY_SOM.PDF

Forkel, M., Carvalhais, N., Verbesselt, J., Mahecha, M.D., Neigh, C.S.R., Reichstein, M., 2013. Trend change detection in NDVI time series: effects of inter-annual variability and methodology. *Remote Sensing 5*, 2113-2144. https://doi.org/10.3390/RS5052113

Forkel, M., Migliavacca, M., Thonicke, K., Reichstein, M., Schaphoff, S., Weber, U., Carvalhais, N., 2015. Codominant water control on global interannual variability and trends in land surface phenology and greenness. *Global Change Biology 21*, 3414–3435. https://doi.org/10.1111/GCB.12950

Gao, B.C., 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment 58*, 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3

Gao, F., Anderson, M., Daughtry, C., Karnieli, A., Hively, D., Kustas, W., 2020a. A within-season approach for detecting early growth stages in corn and soybean using high temporal and spatial resolution imagery. *Remote Sensing of Environment 242*, 111752. https://doi.org/10.1016/J.RSE.2020.111752

Gao, F., Anderson, M.C., Hively, W.D., 2020b. Detecting Cover Crop End-Of-Season Using VENμS and Sentinel-2 Satellite Imagery. *Remote Sensing 12*, 3524. https://doi.org/10.3390/RS12213524

Gao, F., Anderson, M.C., Zhang, X., Yang, Z., Alfieri, J.G., Kustas, W.P., Mueller, R., Johnson, D.M., Prueger, J.H., 2017. Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. *Remote Sensing of Environment 188*, 9–25. https://doi.org/10.1016/J.RSE.2016.11.004

Gao, F., Zhang, X., 2021. Mapping crop phenology in near real-time using satellite remote sensing: challenges and opportunities. *Journal of Remote Sensing 2021*, 1–14. https://doi.org/10.34133/2021/8379391

Gutzler, C., Helming, K., Balla, D., Dannowski, R., Deumlich, D., Glemnitz, M., Knierim, A., Mirschel, W., Nendel, C., Paul, C., Sieber, S., Stachow, U., Starick, A., Wieland, R., Wurbs, A., Zander, P., 2015. Agricultural land use changes – a scenario-based sustainability impact assessment for Brandenburg, Germany. *Ecological Indicators 48*, 505–517. https://doi.org/10.1016/J.ECOLIND.2014.09.004

Guzmán Q., J.A., Sanchez-Azofeifa, G.A., Espírito-Santo, M.M., 2019. MODIS and PROBA-V NDVI products differ when compared with observations from phenological towers at four tropical dry forests in the Americas. *Remote Sensing 11*, 2316. https://doi.org/10.3390/RS11192316

Hall, D.K., Riggs, G.A., Salomonson, V. v., DiGirolamo, N.E., Bayr, K.J., 2002. MODIS snow-cover products. *Remote Sensing of Environment 83*, 181–194. https://doi.org/10.1016/S0034-4257(02)00095-0

Harfenmeister, K., Itzerott, S., Weltzien, C., Spengler, D., 2021. Detecting phenological development of winter wheat and winter barley using time series of Sentinel-1 and Sentinel-2. *Remote Sensing 13*, 5036. https://doi.org/10.3390/RS13245036

Housman, I.W., Chastain, R.A., Finco, M. v., 2018. An evaluation of forest health insect and disease survey data and satellite-based remote sensing forest change detection methods: case studies in the United States. *Remote Sensing 10*, 1184. https://doi.org/10.3390/RS10081184

Huang, X., Liu, J., Zhu, W., Atzberger, C., Liu, Q., 2019. The optimal threshold and vegetation index time series for retrieving crop phenology based on a modified dynamic threshold method. *Remote Sensing 11*, 2725. https://doi.org/10.3390/RS11232725

Huete, A.R., Didan, K, Huete, A., Didan, Kamel, Leeuwen, W. van, Jacobson, A., Solanos, R., Laing, T., 1999. MODIS vegetation index (MOD 13) algorithm theoretical basis document principal investigators development team MODIS product ID: MOD13.

Huete, A.R., Liu, H.Q., Batchily, K., van Leeuwen, W., 1997. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sensing of Environment 59*, 440–451. https://doi.org/10.1016/S0034-4257(96)00112-5

Ihinegbu, C., Ogunwumi, T., 2021. Multi-criteria modelling of drought: a study of Brandenburg Federal State, Germany. *Modeling Earth Systems and Environment 1*, 1–15. https://doi.org/10.1007/S40808-021-01197-2

Inoue, Y., Sakaiya, E., Wang, C., 2014. Capability of C-band backscattering coefficients from high-resolution satellite SAR sensors to assess biophysical variables in paddy rice. *Remote Sensing of Environment 140*, 257–266. https://doi.org/10.1016/j.rse.2013.09.001

Jiang, Z., Huete, A.R., Didan, K., Miura, T., 2008. Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment 112*, 3833–3845. https://doi.org/10.1016/J.RSE.2008.06.006

Jin, H., Eklundh, L., 2014. A physically based vegetation index for improved monitoring of plant phenology. *Remote Sensing of Environment 152*, 512–525. https://doi.org/10.1016/J.RSE.2014.07.010

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 10*, 635. https://doi.org/10.3390/RS10040635

Kaspar, F., Müller-Westermeier, G., Penda, E., Mächel, H., Zimmermann, K., Kaiser-Weiss, A., Deutschländer, T., 2013. Monitoring of climate change in Germany – data, products and services of Germany's National Climate Data Centre. *Advances in Science and Research 10*, 99–106. https://doi.org/10.5194/ASR-10-99-2013

Kaspar, F., Zimmermann, K., Polte-Rudolf, C., 2015. An overview of the phenological observation network and the phenological database of Germany's national meteorological service (Deutscher Wetterdienst). *Advances in Science and Research 11*, 93–99. https://doi.org/10.5194/ASR-11-93-2014

Khabbazan, S., Vermunt, P., Steele-Dunne, S., Arntz, L.R., Marinetti, C., van der Valk, D., Iannini, L., Molijn, R., Westerdijk, K., van der Sande, C., 2019. Crop monitoring using Sentinel-1 data: A case study from The Netherlands. *Remote Sensing 11*, 1887. https://doi.org/10.3390/rs11161887

Larranaga, A., Alvarez-Mozos, J., Albizua, L., Peters, J., 2013. Backscattering behavior of rain-fed crops along the growing season. *IEEE Geoscience and Remote Sensing Letters 10*, 386–390. https://doi.org/10.1109/LGRS.2012.2205660

Lieth, H., 1974. Purposes of a Phenology Book. *Springer*, Berlin, Heidelberg, pp. 3–19. https://doi.org/10.1007/978-3-642-51863-8_1

Lobell, D.B., Schlenker, W., Costa-Roberts, J., 2011. Climate Trends and Global Crop Production Since 1980. *Science (1979)* 333, 616–620. https://doi.org/10.1126/SCIENCE.1204531

Löw, J., Ullmann, T., Conrad, C., 2021. The impact of phenological developments on interferometric and polarimetric crop signatures derived from Sentinel-1: examples from the DEMMIN study site (Germany). *Remote Sensing 13*, 2951. https://doi.org/10.3390/RS13152951.

Matsushita, B., Yang, W., Chen, J., Onda, Y., Qiu, G., 2007. Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to topographic effects: a case study in high-density cypress forest. *Sensors 7*, 2636–2651. https://doi.org/10.3390/S7112636.

Mattia, F., le Toan, T., Picard, G., Posa, F.I., D'Alessio, A., Notarnicola, C., Gatti, A.M., Rinaldi, M., Satalino, G., Pasquariello, G., 2003. Multitemporal C-band radar measurements on wheat fields.

IEEE Transactions on Geoscience and Remote Sensing 41, 1551–1560. https://doi.org/10.1109/TGRS.2003.813531.

Meier, U., 2001. Growth stages of mono- and dicotyledonous plants. URL https://agris.fao.org/agris-search/search.do?recordID=US201300311612 (accessed 10.20.21).

Mercier, A., Betbeder, J., Baudry, J., Le Roux, V., Spicher, F., Lacoux, J., Roger, D., Hubert-Moy, L., 2020a. Evaluation of Sentinel-1 & 2 time series for predicting wheat and rapeseed phenological stages. *ISPRS Journal of Photogrammetry and Remote Sensing 163*, 231–256. https://doi.org/10.1016/J.ISPRSJPRS.2020.03.009

Mercier, A., Betbeder, J., Baudry, J., le Roux, V., Spicher, F., Lacoux, J., Roger, D., Hubert-Moy, L., 2020b. Evaluation of Sentinel-1 & 2 time series for predicting wheat and rapeseed phenological stages. *ISPRS Journal of Photogrammetry and Remote Sensing 163*, 231–256. https://doi.org/10.1016/J.ISPRSJPRS.2020.03.009

Meroni, M., d'Andrimont, R., Vrieling, A., Fasbender, D., Lemoine, G., Rembold, F., Seguini, L., Verhegghen, A., 2021. Comparing land surface phenology of major European crops as derived from SAR and multispectral data of Sentinel-1 and -2. *Remote Sensing of Environment 253*, 112232. https://doi.org/10.1016/j.rse.2020.112232

Misra, G., Cawkwell, F., Wingler, A., 2020. Status of phenological research using Sentinel-2 data: a review. *Remote Sensing 12*, 2760. https://doi.org/10.3390/RS12172760

Mo, F., Sun, M., Liu, X.Y., Wang, J.Y., Zhang, X.C., Ma, B.L., Xiong, Y.C., 2016. Phenological responses of spring wheat and maize to changes in crop management and rising temperatures from 1992 to 2013 across the Loess Plateau. *Field Crops Research 196*, 337–347. https://doi.org/10.1016/j.fcr.2016.06.024

Moon, M., Richardson, A.D., Friedl, M.A., 2021. Multiscale assessment of land surface phenology from harmonized Landsat 8 and Sentinel-2, PlanetScope, and PhenoCam imagery. *Remote Sensing of Environment 266*, 112716. https://doi.org/10.1016/J.RSE.2021.112716

Nasrallah, A., Baghdadi, N., Hajj, M. El, Darwish, T., Belhouchette, H., Faour, G., Darwich, S., Mhawej, M., 2019. Sentinel-1 data for winter wheat phenology monitoring and mapping. *Remote Sensing 11*, 2228. https://doi.org/10.3390/RS11192228

Nietupski, T.C., Kennedy, R.E., Temesgen, H., Kerns, B.K., 2021. Spatiotemporal image fusion in Google Earth Engine for annual estimates of land surface phenology in a heterogenous landscape. *International Journal of Applied Earth Observation and Geoinformation 99*, 102323. https://doi.org/10.1016/J.JAG.2021.102323

Nijland, W., Bolton, D.K., Coops, N.C., Stenhouse, G., 2016. Imaging phenology; scaling from camera plots to landscapes. *Remote Sensing of Environment 177*, 13–20. https://doi.org/10.1016/J.RSE.2016.02.018

Pan, L., Xia, H., Zhao, X., Guo, Y., Qin, Y., 2021. Mapping winter crops using a phenology algorithm, time-series Sentinel-2 and Landsat-7/8 images, and Google Earth Engine. *Remote Sensing 13*, 2510. https://doi.org/10.3390/RS13132510

Pasqui, M., Magno, R., Shorachi, M., Kumar, V., Steele-Dunne, S.C., 2022. Sentinel-1 SAR backscatter response to agricultural drought in The Netherlands. *Remote Sensing 14*, 2435. https://doi.org/10.3390/RS14102435

Picard, G., le Toan, T., Mattia, F., 2003. Understanding C-band radar backscatter from wheat canopy using a multiple-scattering coherent model. *IEEE Transactions on Geoscience and Remote Sensing 41*, 1583–1591. https://doi.org/10.1109/TGRS.2003.813353

R Core Team, 2021. R: A language and environment for statistical computing. R *Foundation for Statistical Computing*, Vienna, Austria.

Ray, D.K., Gerber, J.S., Macdonald, G.K., West, P.C., 2015. Climate variation explains a third of global crop yield variability. *Nature Communications 6*, 1–9. https://doi.org/10.1038/ncomms6989

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*, 703–714. https://doi.org/10.2307/3235884

Reinermann, S., Gessner, U., Asam, S., Kuenzer, C., Dech, S., 2019. The effect of droughts on vegetation condition in Germany: an analysis based on two decades of satellite earth observation time series and crop yield statistics. *Remote Sensing 11*, 1783. https://doi.org/10.3390/RS11151783

Rezaei, E.E., Siebert, S., Hüging, H., Ewert, F., 2018. Climate change effect on wheat phenology depends on cultivar change. *Scientific Reports 8*, 1–10. https://doi.org/10.1038/s41598-018-23101-2

Richardson, A.D., Black, T.A., Ciais, P., Delbart, N., Friedl, M.A., Gobron, N., Hollinger, D.Y., Kutsch, W.L., Longdoz, B., Luyssaert, S., Migliavacca, M., Montagnani, L., Munger, J.W., Moors, E., Piao, S., Rebmann, C., Reichstein, M., Saigusa, N., Tomelleri, E., Vargas, R., Varlagin, A., 2010. Influence of spring and autumn phenological transitions on forest ecosystem productivity. *Philosophical Transactions of the Royal Society B: Biological Sciences 365*, 3227–3246. https://doi.org/10.1098/rstb.2010.0102

Richardson, A.D., Hufkens, K., Milliman, T., Frolking, S., 2018. Intercomparison of phenological transition dates derived from the PhenoCam Dataset V1.0 and MODIS satellite remote sensing. *Scientific Reports 8*, 1–12. https://doi.org/10.1038/s41598-018-23804-6

Sakamoto, T., Gitelson, A.A., Arkebauer, T.J., 2013. MODIS-based corn grain yield estimation model incorporating crop phenology information. *Remote Sensing of Environment 131*, 215–231. https://doi.org/10.1016/j.rse.2012.12.017

Schlund, M., Erasmi, S., 2020. Sentinel-1 time series data for monitoring the phenology of winter wheat. *Remote Sensing of Environment 246*, 111814. https://doi.org/10.1016/j.rse.2020.111814

Schwartz, M.D., Betancourt, J.L., Weltzin, J.F., 2012. From Caprio's lilacs to the USA National Phenology Network. *Frontiers in Ecology and the Environment 10*, 324–327. https://doi.org/10.1890/110281

Siebert, S., Ewert, F., 2012. Spatio-temporal patterns of phenological development in Germany in relation to temperature and day length. *Agricultural and Forest Meteorology 152*, 44–57. https://doi.org/10.1016/J.AGRFORMET.2011.08.007

Sivertsen, T.H., Nejedlik, P., Oger, R., Sigvald, R., 1999. The phenology of crops and the development of pests and diseases Literature, research, models and future operational integration.

Small, D., 2011. Flattening gamma: Radiometric terrain correction for SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing 49*, 3081–3093. https://doi.org/10.1109/TGRS.2011.2120616

Stendardi, L., Karlsen, S.R., Niedrist, G., Gerdol, R., Zebisch, M., Rossi, M., Notarnicola, C., 2019. Exploiting time series of Sentinel-1 and Sentinel-2 imagery to detect meadow phenology in mountain regions. *Remote Sensing 11,* 542. https://doi.org/10.3390/RS11050542

Tian, F., Cai, Z., Jin, H., Hufkens, K., Scheifinger, H., Tagesson, T., Smets, B., van Hoolst, R., Bonte, K., Ivits, E., Tong, X., Ardö, J., Eklundh, L., 2021. Calibrating vegetation phenology from Sentinel-2 using eddy covariance, PhenoCam, and PEP725 networks across Europe. *Remote Sensing of Environment 260*, 112456. https://doi.org/10.1016/J.RSE.2021.112456

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B.Ö., Floury, N., Brown, M., Traver, I.N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, M., Croci, R., Pietropaolo, A., Huchler, M., Rostan, F., 2012. GMES Sentinel-1 mission. *Remote Sensing of Environment 120*, 9–24. https://doi.org/10.1016/J.RSE.2011.05.028

Trnka, M., Rötter, R.P., Ruiz-Ramos, M., Kersebaum, K.C., Olesen, J.E., Žalud, Z., Semenov, M.A., 2014. Adverse weather conditions for European wheat production will become more frequent with climate change. *Nature Climate Change 4*, 637–643. https://doi.org/10.1038/nclimate2242

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment 8*, 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

Veloso, A., Mermoz, S., Bouvet, A., le Toan, T., Planells, M., Dejoux, J.F., Ceschia, E., 2017. Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. *Remote Sensing of Environment 199*, 415–426. https://doi.org/10.1016/j.rse.2017.07.015

Verbesselt, J., Zeileis, A., Herold, M., 2012. Near real-time disturbance detection using satellite image time series. *Remote Sensing of Environment 123*, 98–108. https://doi.org/10.1016/J.RSE.2012.02.022

Vreugdenhil, M., Wagner, W., Bauer-Marschallinger, B., Pfeil, I., Teubner, I., Rüdiger, C., Strauss, P., 2018. Sensitivity of Sentinel-1 backscatter to vegetation dynamics: An Austrian case study. *Remote Sensing 10*, 1396. https://doi.org/10.3390/rs10091396

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*, 517–529. https://doi.org/10.1016/j.rse.2018.03.014

Wang, X., Liu, Y., Li, X., He, S., Zhong, M., Shang, F., 2022. Spatiotemporal variation of osmanthus fragrans phenology in China in response to climate change from 1973 to 1996. *Frontiers in Plant Science 12*, 716071. https://doi.org/10.3389/FPLS.2021.716071/FULL

Weilnhammer, V., Schmid, J., Mittermeier, I., Schreiber, F., Jiang, L., Pastuhovic, V., Herr, C., Heinze, S., 2021. Extreme weather events in Europe and their health consequences – A systematic review. *International Journal of Hygiene and Environmental Health 233*, 113688. https://doi.org/10.1016/J.IJHEH.2021.113688

Wolff, S., Hüttel, S., Nendel, C., Lakes, T., 2021. Agricultural landscapes in Brandenburg, Germany: an analysis of characteristics and spatial patterns. *International Journal of Environmental Research 15*, 487–507. https://doi.org/10.1007/S41742-021-00328-Y/TABLES/4

Wong, M.T.F., Asseng, S., 2006. Determining the causes of spatial and temporal variability of wheat yields at sub-field scale using a new method of upscaling a crop model. *Plant and Soil 283*, 203–215. https://doi.org/10.1007/S11104-006-0012-5

Yang, Yang, Anderson, M.C., Gao, F., Wardlow, B., Hain, C.R., Otkin, J.A., Alfieri, J., Yang, Yun, Sun, L., Dulaney, W., 2018. Field-scale mapping of evaporative stress indicators of crop yield: An application over Mead, NE, USA. *Remote Sensing of Environment 210*, 387–402. https://doi.org/10.1016/J.RSE.2018.02.020

You, X., Meng, J., Zhang, M., Dong, T., 2013. Remote sensing based detection of crop phenology for agricultural zones in China using a new threshold method. *Remote Sensing 5*, 3190–3211. https://doi.org/10.3390/RS5073190

Zeileis, A., Kleiber, C., Walter, K., Hornik, K., 2003. Testing and dating of structural changes in practice. *Computational Statistics & Data Analysis 44*, 109–123. https://doi.org/10.1016/S0167-9473(03)00030-6

Zeng, L., Wardlow, B.D., Xiang, D., Hu, S., Li, D., 2020. A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. *Remote Sensing of Environment 237*, 111511. https://doi.org/10.1016/J.RSE.2019.111511

Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C.F., Gao, F., Reed, B.C., Huete, A., 2003. Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment 84*, 471–475. https://doi.org/10.1016/S0034-4257(02)00135-9

Zhang, X., Jayavelu, S., Liu, L., Friedl, M.A., Henebry, G.M., Liu, Y., Schaaf, C.B., Richardson, A.D., Gray, J., 2018a. Evaluation of land surface phenology from VIIRS data using time series of

PhenoCam imagery. *Agricultural and Forest Meteorology 256–257*, 137–149. https://doi.org/10.1016/J.AGRFORMET.2018.03.003

Zhang, X., Liu, L., Liu, Y., Jayavelu, S., Wang, J., Moon, M., Henebry, G.M., Friedl, M.A., Schaaf, C.B., 2018b. Generation and evaluation of the VIIRS land surface phenology product. *Remote Sensing of Environment 216*, 212–229. https://doi.org/10.1016/J.RSE.2018.06.047

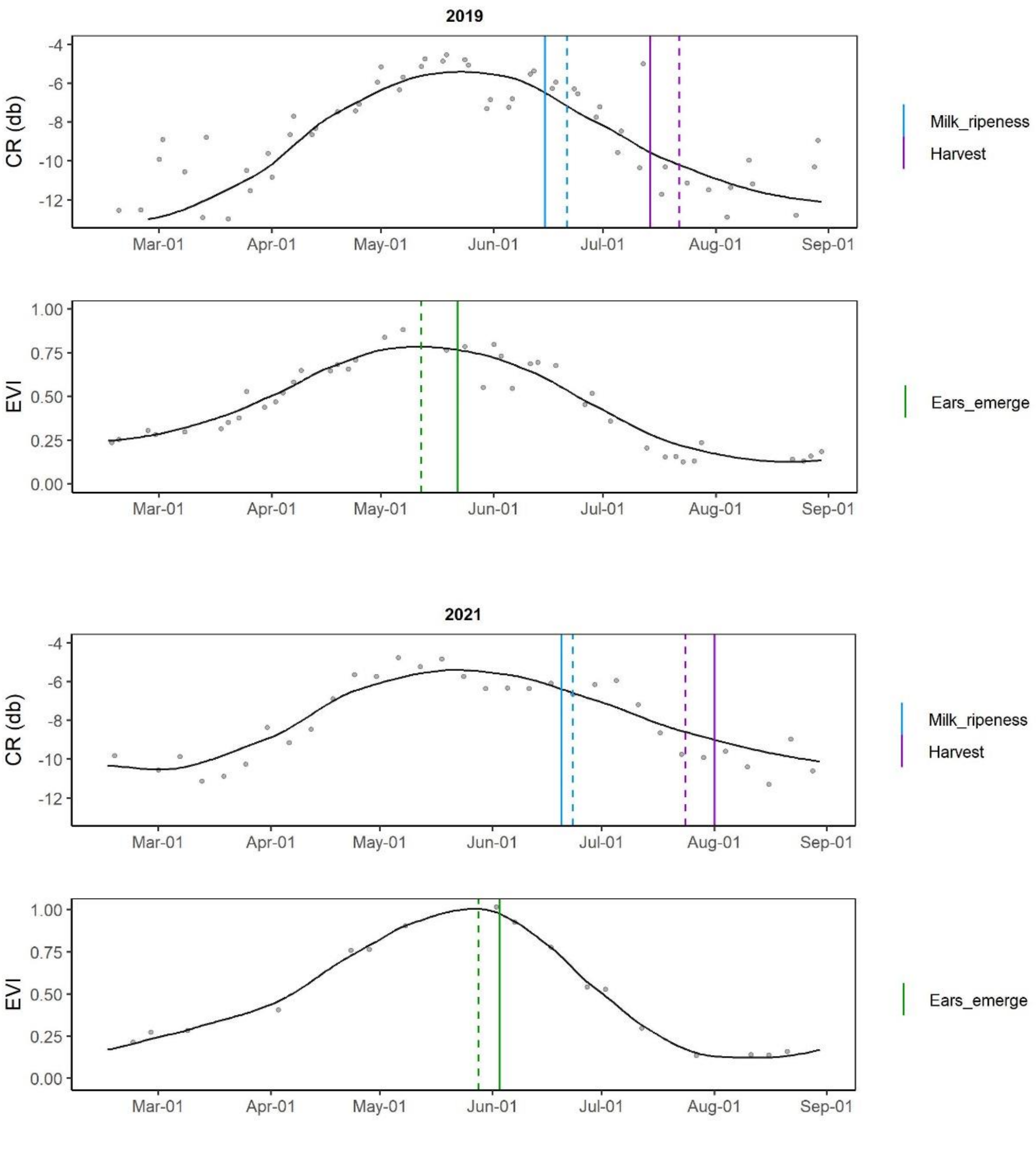
Zhang, X., Wang, J., Gao, F., Liu, Y., Schaaf, C., Friedl, M., Yu, Y., Jayavelu, S., Gray, J., Liu, L., Yan, D., Henebry, G.M., 2017. Exploration of scaling effects on coarse resolution land surface phenology. *Remote Sensing of Environment 190*, 318–330. https://doi.org/10.1016/J.RSE.2017.01.001

Zhang, X., Zhang, Q., 2016. Monitoring interannual variation in global crop yield using long-term AVHRR and MODIS observations. *ISPRS Journal of Photogrammetry and Remote Sensing 114*, 191–205. https://doi.org/10.1016/j.isprsjprs.2016.02.010

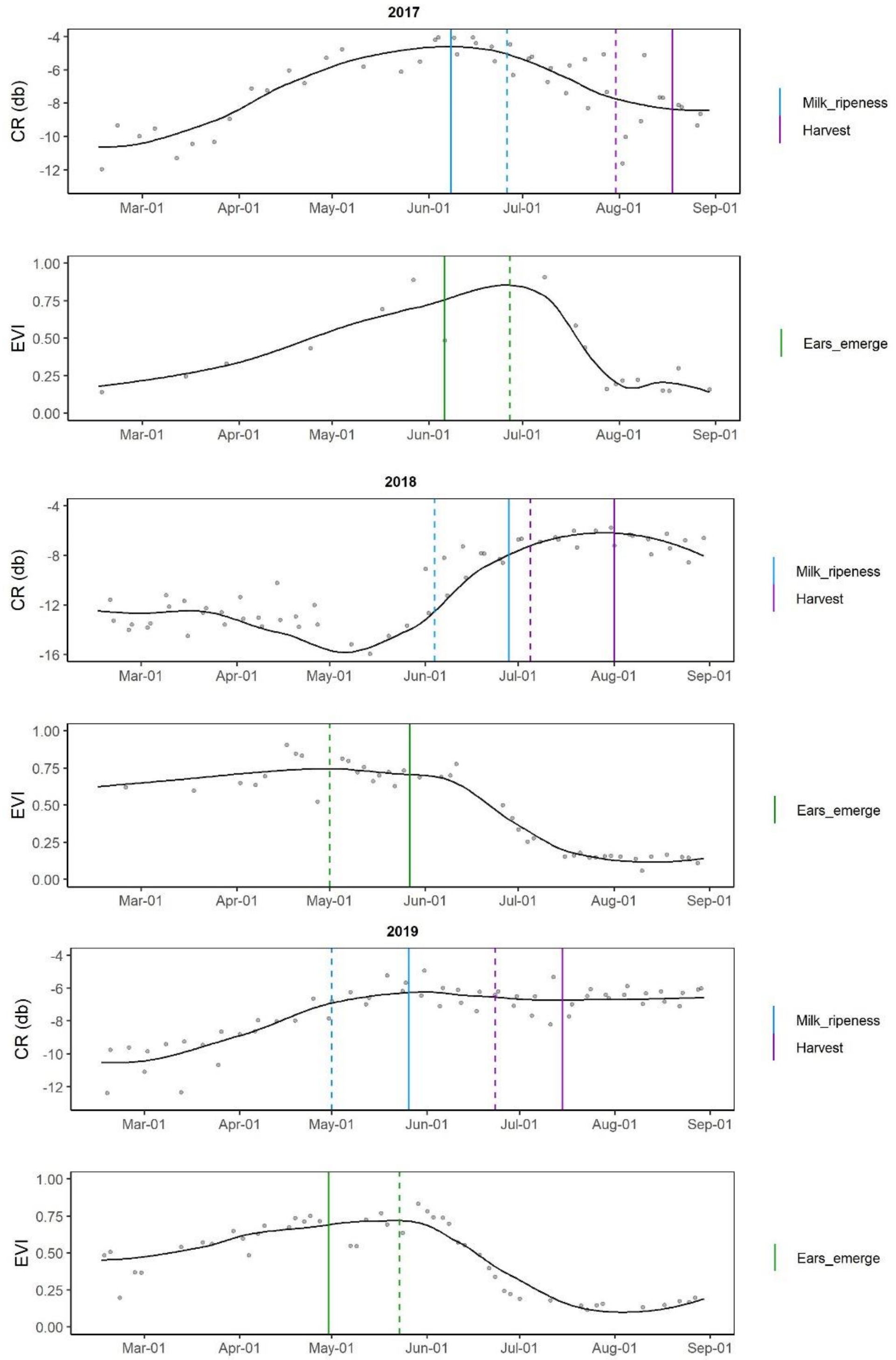
Zheng, H., Cheng, T., Yao, X., Deng, X., Tian, Y., Cao, W., Zhu, Y., 2016. Detection of rice phenology through time series analysis of ground-based spectral index data. *Field Crops Research 198*, 131–139. https://doi.org/10.1016/j.fcr.2016.08.027

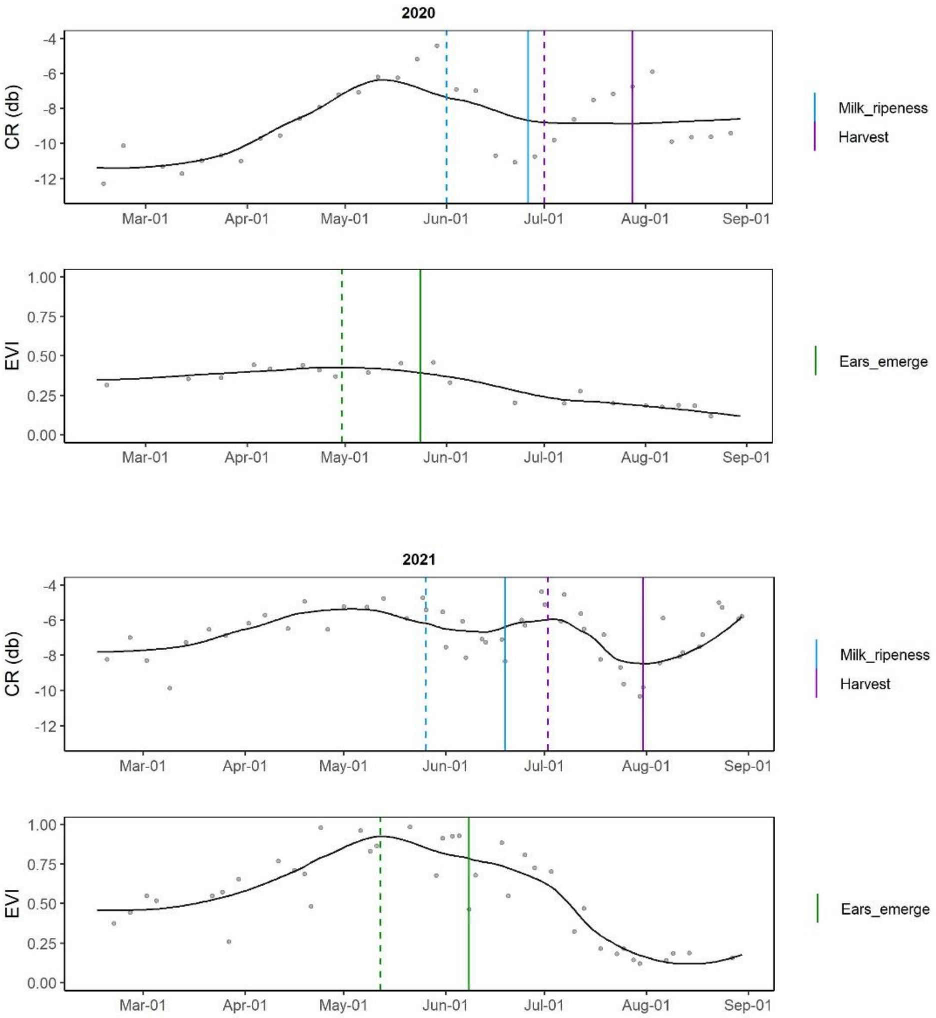
8. APPENDIX

Annex 1: Exemplary temporal profiles from 2019 and 2021 of CR and EVI showing the remote sensing estimated metrics with their corresponding phenophases. Estimated metrics are shown in dashes line and DWD phenophases are shown in solid line.

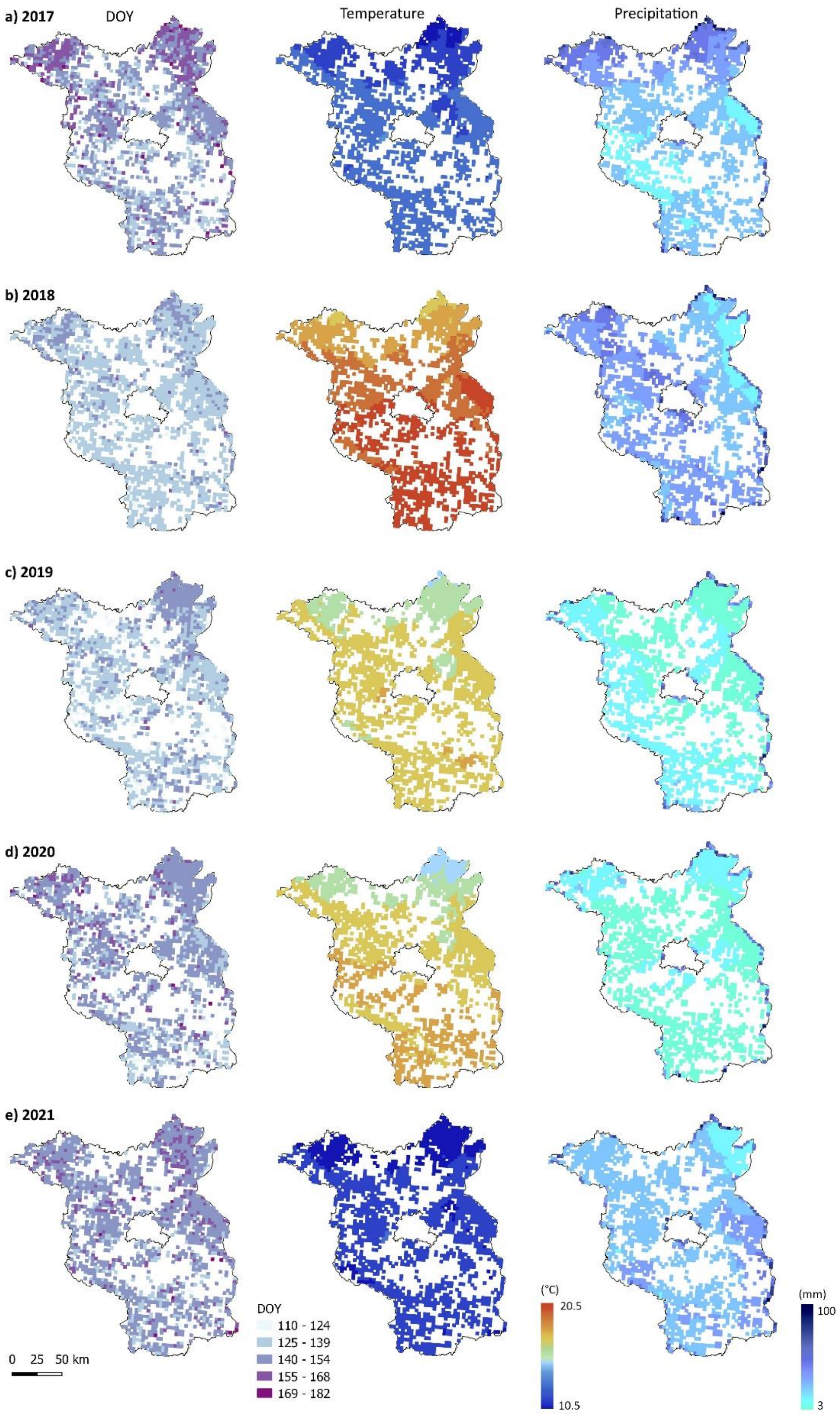


Annex 2: Outlier temporal profiles which have large deviations with the DWD reported phenophases. One profile is chosen per year and per sensor. For description of figure elements, refer to **Annex 1**.

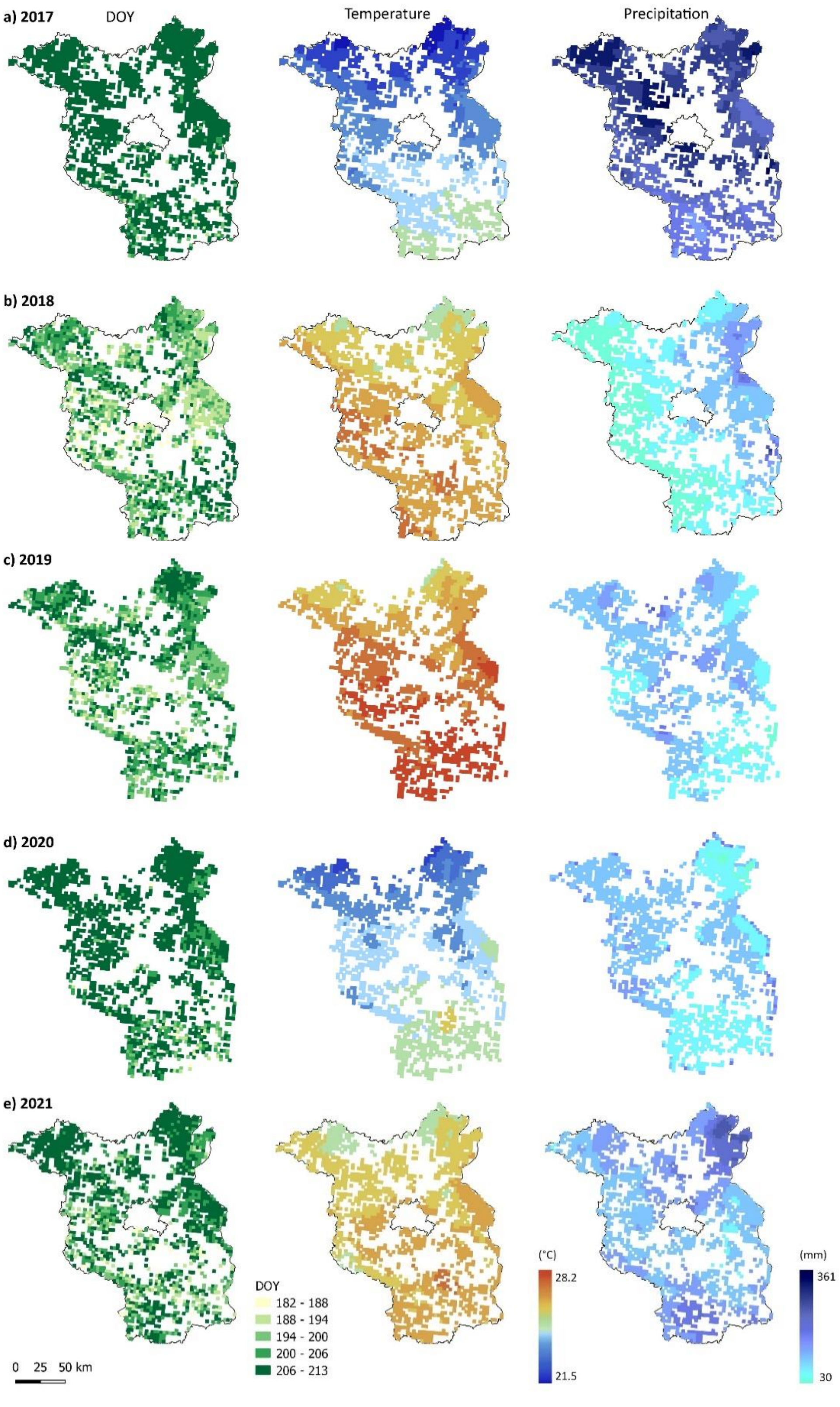



2020
CR (db)
-4
-6
-8
-10
-12
Mar-01
Apr-01
May-01
Jun-01
Jul-01
Aug-01
Sep-01
Milk_ripeness
Harvest
EVI
1.00
0.75
0.50
0.25
0.00
Ears_emerge
2021

Annex 3: Spatial and inter-annual variability of ears emerge (BBCH 51) with corresponding temperature and precipitation (April).



Annex 4: Spatial and inter-annual variability of harvest (BBCH 99) with corresponding temperature and precipitation (June-July).



Annex 5: Regression results of tested weather parameters against the estimated metrics. The retained metrics are shown in bold.

		Temperature		Precipitation	
	Years	R^2	p	R^2	p
Ears emerge	March	0.12	<0.001	0.02	<0.001
	April	**0.24**	**<0.001**	**0.02**	**<0.001**
	May	0.09	<0.001	<0.01	0.329
	March-April	0.08	<0.001	<0.01	0.215
	April-May	0.08	<0.001	<0.01	<0.001
	March- April-May	0.02	<0.001	<0.01	0.095
Milk ripeness	April	0.02	<0.001	0.01	<0.001
	May	0.01	<0.001	0.01	<0.001
	June	<0.01	0.043	<0.001	<0.001
	April-May	<0.01	<0.001	<0.001	<0.001
	May-June	0.06	<0.001	<0.01	<0.001
	April-May-June	**0.02**	**<0.001**	**0.03**	**<0.001**
Harvest	May	<0.01	<0.001	0.03	<0.001
	June	0.21	<0.001	0.11	<0.001
	July	0.18	<0.001	0.13	<0.001
	May-June	0.06	<0.001	0.06	<0.001
	April-May-June	0.12	<0.001	<0.01	<0.001
	June-July	**0.25**	**<0.001**	**0.13**	**<0.001**
	May-June-July	0.08	<0.001	0.07	<0.001

Annex 6: Table showing total number of retained parcels used for phenometrics estimatation (per phase per year), number of phenology stations reported per phase per year, number of total retained parcels belongs to the buffer of station (per phase per year)

	Years	Number of total retained parcels used for phenometrics estimation	Number of DWD phenology stations	Number of retained parcels within 5-km buffer of stations
Ears emerge	2017	7463	24	511
	2018	8013	24	686
	2019	8403	20	622
	2020	7172	22	616
	2021	7725	25	682
Milk ripeness	2017	7942	22	514
	2018	7861	20	665
	2019	8311	18	593
	2020	7262	22	638
	2021	7194	23	669
Harvest	2017	7976	25	591
	2018	7859	24	686
	2019	8326	24	611
	2020	7340	24	644
	2021	7620	26	675

Annex 7: Table showing total number of retained parcels used for phenometrics estimation (per phase per year), number of phenology stations reported per phase per year, number of total retained parcels belongs to the buffer of station (per phase per year)