PFG (2018) 86:15–27
https://doi.org/10.1007/s41064-018-0043-6

ORIGINAL ARTICLE

Detection of Phenology-Defined Data Acquisition Time Frames For Crop Type Mapping

Henning Gerstmann[1] · **Cornelia Gläßer**[1] · **Detlef Thürkow**[1] · **Markus Möller**[2]

Received: 11 October 2017 / Accepted: 22 February 2018 / Published online: 25 April 2018

Abstract Agricultural monitoring and assessment based on satellite data increasingly gains importance due to the growing number of available satellite sensors with high geometric and temporal resolution. Such tasks often require multiple images acquired on specific dates that among others account for inter-annual phenological variations to provide accurate results. This contribution presents an approach that links peaks of spectral separability profiles to crop phenological phases. The phases are spatially interpolated using a phenological model and ground observations. The profiles show the respective temporal development of the F-measure which is used as indicator for class-wise separability. It originates from binary classifications of vegetation indices computed for each set of a satellite data archive covering multiple years. Acquisition dates, which repeatedly show a separability maximum define phenological indicator phases. Potential alternative phases can be also defined. Experiments based on multi-temporal RapidEye satellite imagery were performed for three crops at two German test sites under different environmental conditions. The results showed that the phases yellow ripeness, heading and flowering can function as indicator phases for high spectral separability of winter barley, winter wheat and winter rapeseed. We could identify at least two identical, stable indicator phases per crop type for both test sites, which suggests the transferability and robustness of the presented approach.

✉ Henning Gerstmann
henning.gerstmann@geo.uni-halle.de

[1] Institute for Geosciences and Geography, Martin Luther University Halle-Wittenberg, Von-Seckendorff-Platz 4, 06120 Halle (Saale), Germany

[2] Institute for Strategies and Technology Assessment, Julius Kühn-Institute, Federal Research Centre for Cultivated Plants, Stahnsdorfer Damm 81, 14532 Kleinmachnow, Germany

Keywords Spectral separability · Phenological phases · Multi-temporal · RapidEye · Germany

Zusammenfassung *Detektion von phänologisch definierten Datenaufnahmezeiträumen für die Klassifikation von Feldfrüchten*
Auf Satellitendaten basierendes landwirtschaftliches Monitoring gewinnt durch die wachsende Anzahl verfügbarer Sensoren mit hoher zeitlicher und geometrischer Auflösung zunehmend an Bedeutung. Für solche Anwendungen werden oftmals Satellitendaten von verschiedenen Aufnahmetagen benötigt, deren Auswahl inter-annuelle phänologische Variationen berücksichtigen muss, um exakte Ergebnisse zu liefern. Dieser Beitrag präsentiert einen Ansatz, um Maxima von spektralen Trennbarkeitsprofilen mit phänologischen Phasen von Feldfrüchten zu verbinden. Diese Phasen werden unter Nutzung eines phänologischen Modelles und Beobachtungsdaten räumlich interpoliert. Die Trennbarkeitsprofile zeigen den zeitlichen Verlauf des F-Maß, das als Indikator für klassenspezifische Trennbarkeit genutzt wird. Dieses stammt von binär klassifizierten Vegetationsindizes, die für jeden Datensatz einer mehrjährigen, multi-temporalen Zeitserie von Satellitenbilddatensätzen berechnet wurden. Zeitpunkte, während denen wiederholt das Trennbarkeitsmaximum beobachtet werden konnte, weisen die Indikatorphasen aus. Potentielle Alternativphasen können ebenso bestimmt werden. Die Untersuchungen wurden für drei Fruchtarten in zwei Untersuchungsgebieten in Deutschland unter verschiedenen Umweltbedingungen auf Basis von RapidEye-Satellitendaten durchgeführt. Die Ergebnisse zeigen, dass die Phasen Gelbreife, Ährenschieben und Blüte als Indikatoren für hohe spektrale Trennbarkeit von Wintergerste, Winterweizen und Winterraps dienen können. Für jede untersuchte Fruchtart konnten wenigstens zwei, für beide Untersuchungsgebiete übereinstimmende, stabile Indi-

katorphasen ausgewiesen werden, was die Übertragbarkeit und Robustheit des gezeigten Verfahrens belegt.

Schlagwörter Spektrale Trennbarkeit · Phänologische Phasen · Multitemporal · RapidEye · Deutschland

1 Introduction

The accuracy and efficiency of remote sensing applications in agriculture based on optical satellite imagery is strongly affected by the variability of crop type phenology. The phenology of crops, describing the timing and the sequence of developmental stages, varies significantly between crop types as well as under different agronomic practices and climates. This has to be considered when remote sensing methods are applied for crop and land use mapping (e.g., Schmidt et al. 2014), in-season yield estimations for food security assessment (e.g., Meroni et al. 2014) or efficient image compositing (Frantz et al. 2017). All these applications require an exactly timed data set selection to provide the most accurate results.

Crop type mapping is one key application in agricultural remote sensing. To derive accurate land use maps that are essential for many purposes, crop types can be efficiently distinguished during specific phenological phases. These phenological windows require dynamic adjustments for annual variability in crop development, which has been continuously studied during the last few decades (Murakami et al. 2001; van Niel and McVicar 2004; Förster et al. 2012; Conrad et al. 2014; Schmidt et al. 2014; Azar et al. 2016; Möller et al. 2017).

The phenology-aided selection of the time steps based on expert knowledge is widely studied (Guerschman et al. 2003; Peña-Barragàn et al. 2011; Conrad et al. 2014; Schmidt et al. 2014). However, these studies are often based on static crop calendars that do not reflect weather-induced annual shifts in phenology (Meroni et al. 2014). Crop-specific responses to such variations account for inter-annual differences between the times of highest separability among crop types (Nitze et al. 2015), which points at the necessity for a dynamic derivation of spatial phenological information. For this purpose, the analysis of temporal profiles that consist of satellite-based vegetation indices (e.g., Frantz et al. 2017) proved to be a successful strategy. For instance, phenological metrics like maximum and minimum vitality derived from such temporal profiles enabled the direct extraction of a small number of key growth stages such as green up, heading and senescence (Xu et al. 2017). Phenological in situ observations, which usually provide a much higher number of individual stages, can be alternatively used to interpret variations of separability profiles of crops (Förster et al. 2012). Furthermore, they can be used to validate phenological stages derived from vegetation index profiles (Xu et al. 2017) which are often highly affected by noise. The spatial modelling of point observations of the crops' growth status (Gerstmann et al. 2016a) combines the advantages of spatially explicit and in situ phenological data sets and can thus improve the knowledge about critical days for crop type classification.

Currently, the majority of modelling and classification approaches aggregates crop types of similar spectral and phenological behaviour to wider classes such as winter cereals or root crops. While these aggregations are sufficient for most applications, information on crop species level is essential for detailed yield predictions (Nitze et al. 2015), water management (Conrad et al. 2013), or subsidy control. However, this class aggregation is often unavoidable, since data gaps caused by cloud coverage or sensor-specific characteristics limit the capability to separate spectrally similar species.

These data gaps are globally an issue for almost all applications based on optical remote sensing imagery. Hence, the consideration of broader time frames increases the chance of having usable data available during high separability periods. Narrow alternative time frames can also be found by inter-annual analyses of species-specific phenological behaviour, but these time frames have only limited relevance due to the mentioned high probability of data gaps.

The main objective of this study is to identify phenological phases that suit for optimal class separation in crop mapping. Therefore, a framework was developed that systematically combines interpolated phenological ground observations with satellite image acquisition dates and spectral separability patterns. In doing so, phenological phases are analysed in terms of their suitability as indicators for optimal acquisition time frames.

The approach is applied to multi-temporal RapidEye data acquired for an agriculturally used test site in Central Germany. At this site, indicator phases are defined based on separability profile and phenological phases. The transferability of these indicator phases is evaluated on a validation site characterised by different growing conditions.

Finally, a web-based tool is presented that provides almost real-time phenological raster data covering Germany to apply the framework in practise.

2 Study Sites and Data

2.1 Site Descriptions

The two study sites have been intensively investigated by the Terrestrial Environmental Observatories (TERENO) research network that focuses on observations of long-term climate change impacts on regional scales (Bogena 2016). They were used for method development and testing, respectively.

The training site is located in Central Germany (Fig. 1), approximately 30 km north of the city of Halle (Saale), with the town Hettstedt being its centre (Fig. 2, left). The site consists of two sub-sites: a western sub-site covering the eastern parts of the Harz mountain range (HM) and an eastern sub-site representing the adjacent lowlands (HL). Both sub-sites cover an area of 25 km × 25 km each. Annual mean temperatures and precipitation are 8.5 °C and 630 mm for HM and 9.2 °C and 540 mm for HL, respectively. The area is dominated by agricultural land use on fertile chernozemic soils that allow the cultivation of demanding species. The dominating crop types are winter wheat, cropped on ≈ 45% of the agricultural area, winter rapeseed (≈ 20%), perennial grassland (≈ 10%) and winter barley (≈ 10%).

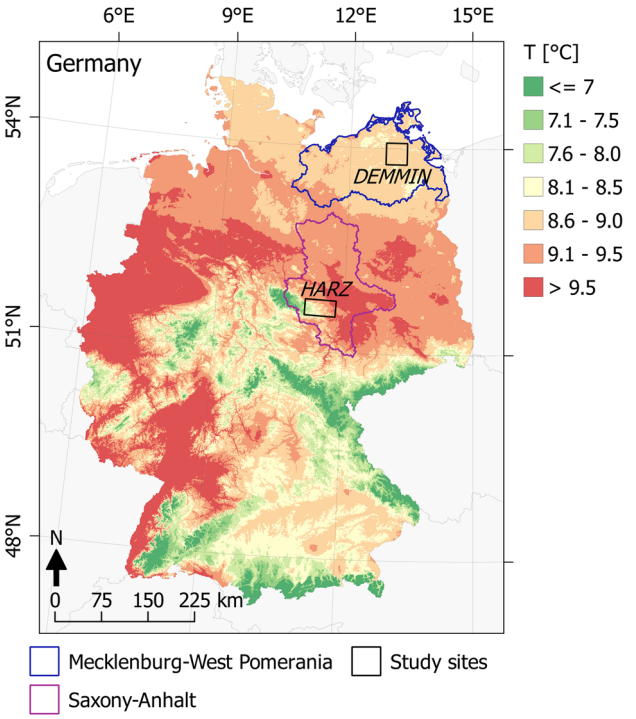

Fig. 1 Locations of the study sites in Germany and annual mean temperature (DWD Climate Data Center 2017b). Projection: WGS84/UTM Zone 33N

The second site used for the transferability study is located in the federal state of Mecklenburg-West Pomerania, in the surrounding of the city Demmin (DM, Fig. 2, right). The site covers an area of 35 km × 35 km. Annual mean temperature and precipitation in DM are 8.7 °C and 590 mm, respectively. Here, sandy soils of reduced fertility require more frequent cultivation of less demanding crops, which is represented by the lower percentage of winter wheat (≈ 30%) and higher percentages of the area covered by perennial grassland (≈ 15%) and by less-demanding winter rye (≈ 5%). Gradients between comparatively warm temperatures in the HL and relatively cool temperatures in HM are apparent in Fig. 1. Contrary to this strong gradient, the DM site is characterized by a homogeneous annual mean temperature which ranges between the two extremes of the Harz sub-sites.

Two aspects determined the selection of the study sites. First, natural conditions, e.g., soil fertility, should differ among the regions to assess the robustness of the presented framework. For instance, precipitation sums also varied between the sites, especially in the strong easterly gradient of decreasing precipitation between the Harz sub-sites (> 700 mm in the west and < 500 mm in the east). Second, crop types of the training and validation sites should show

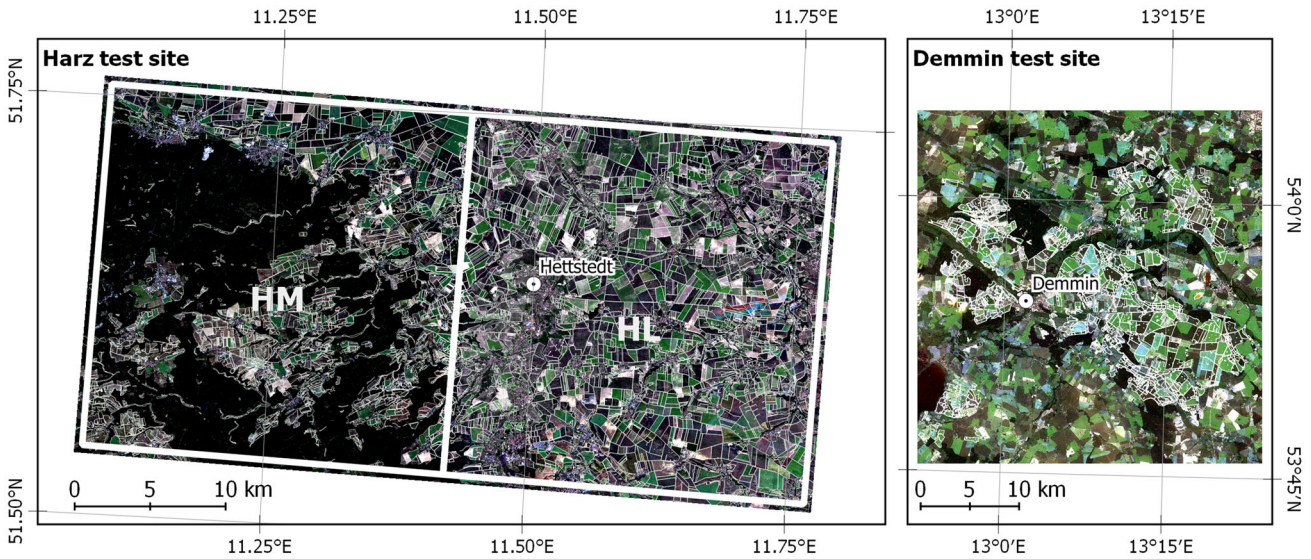

Fig. 2 RapidEye images of the study sites Harz (left) and Demmin (right) acquired on June 6, 2011 (HL), June 2, 2011 (HM) and July 3, 2010 (DM). Band combination RGB 3-2-1 (true colour). Fields with crop cultivation data are accentuated. Projection: WGS84/UTM Zone 33N

some similarities to sustain the comparability of the results. So, the field sizes should ensure that a sufficient number of pure (not mixed) pixels is available for the calculation of the reflectance means per field. Both study sites fulfilled these requirements.

2.2 Satellite Data

Multi-temporal RapidEye data sets covering the HM and HL sites between 2010 and 2015 were made available through the RapidEye Science Archive (RESA; Borg et al. 2013, grant no: 653). The RapidEye satellite constellation collects imagery of the Earth's surface in five spectral bands (blue: 440–510 nm, green: 520–590 nm, red: 530–685 nm, red edge: 690–730 nm, near infra-red/NIR: 760–850 nm, Tyc et al. 2005). The data were obtained at processing level 3A with a spatial resolution of 5 m × 5 m. Atmospheric correction and cloud masking were performed using the software ATCOR 2© (Richter and Schläpfer 2015). For the DM site, RapidEye data sets for the years 2010–2012 were available at processing level 1B with 6.5 m × 6.5 m resolution. Preprocessing of these data sets was performed using the software CATENA© (Krauß et al. 2013) that also includes ATCOR 2 ©. The effects of differing sensor viewing angles (Nagol et al. 2015) on vegetation reflectance are eliminated by application of the ATCOR processing chain. The temporal distribution of the data sets is displayed in Fig. 3 and highlights inter-annual differences in temporal coverage and density, which are typical for optical imagery of high temporal and geometric resolution (Whitcraft et al. 2015).

2.3 Auxiliary Data

Parcel-based cropping information for 2010–2015 for the Harz sites were provided by the Ministry of Agriculture and Environment of Saxony-Anhalt and for DM via the TERENO long-term research programme. As part of the European Commission's Land Parcel Identification System (LPIS), the parcels are determined by ownership structure and homogeneity regarding the cultivated crop type (Inan et al. 2010). Germany-wide volunteer-based phenological point observations and meteorological measurements were obtained from the German Weather Service (DWD Climate Data Center 2017a, c) via FTP.[1] The names of the available phenological phases for winter wheat, winter rapeseed and winter barley are listed in Table 1 along with their numeric code according to the DWD observation programme (Kaspar et al. 2014).

Furthermore, a digital elevation model of 1 km × 1 km resolution was generated from the Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM, Rabus et al. 2003).

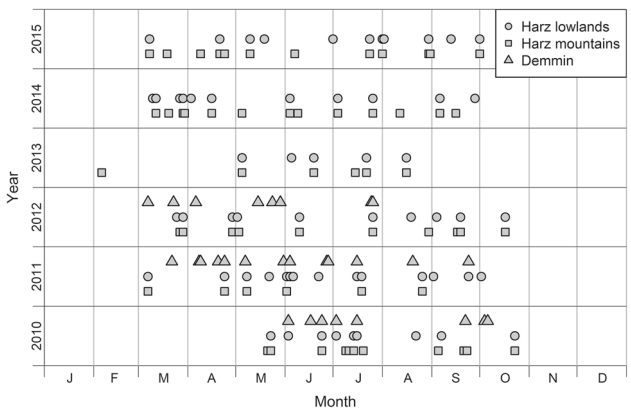

Fig. 3 Temporal distribution of the RapidEye data sets

Table 1 Observed phenological phases for winter wheat (WW), winter barley (WB) and winter rapeseed (WR) and corresponding phase IDs

Name	ID	Crop
Sowing	10	WW, WB, WR
Emergence	12	WW, WB, WR
Fourth leaf unfolded	14	WR
Shooting	15	WW, WB
Heading	18	WW, WB
Stem elongation	67	WR
Bud formation	17	WR
Beginning of flowering	5	WR
End of flowering	7	WR
Milk ripening	19	WW, WB
Yellow ripening	21	WR
Full ripening	22	WR
Harvest	24	WW, WB, WR

3 Methods

The approach is based on crop-specific spectral separability profiles for six subsequent years. These profiles were computed from the RapidEye data sets by applying separability analysis on spectral features (Gerstmann et al. 2016b) calculated from parcel averages of reflectances. The profiles were, analogous to Möller et al. (2017), coupled with modelled phenological phases. The separability maxima over time exhibit the indicator phases, i.e., phenological phases that optimally suit for crop separation. However, before finally assigning the indicator phases, a reliability check was included, which among others targets at the question, if the maxima of separability occur at the same position over several years (here:

[1] ftp://ftp-cdc.dwd.de/pub/CDC/observations_germany.

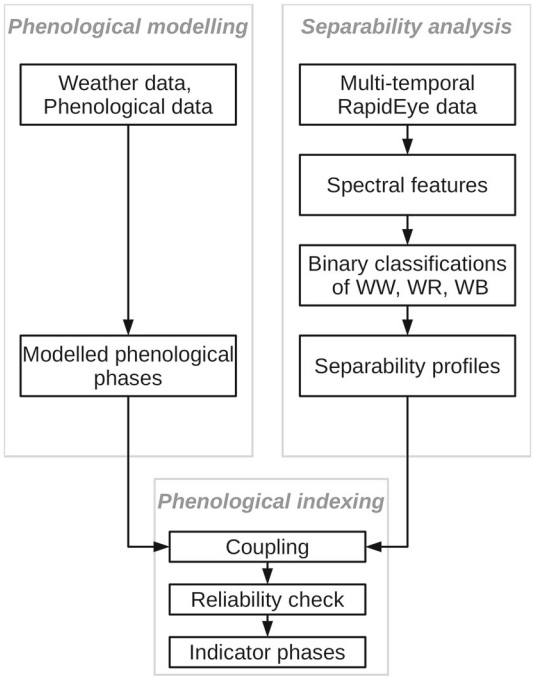

Fig. 4 Workflow for the detection of indicator phases for the optimal separability of the crop types winter wheat (WW), winter barley (WB) and winter rapeseed (WR)

2010–2015). The workflow is summarized in Fig. 4. The components of the framework are described in detail in the following sections.

3.1 Modelling of Phenological Phases

Phenological modelling was performed using the PHASE model according to Gerstmann et al. (2016a). PHASE is a statistical model based on the growing-degree-days approach that relates phenological events to accumulated temperatures from a defined starting point. The model sums up accumulated daily mean temperatures between the day of sowing and the observation of the phenological phase. The temperatures have been interpolated from meteorological measurements on the 1 km × 1 km DEM, for the period between sowing of the crop and the observed date of a phenological event on a phenological station. The temperatures are adjusted for photoperiod on the specific location, which is dependent on latitude and day of year. A specific temperature sum is defined as the 45% quantile of the distribution of the temperature sums for all phenological stations in Germany that reported the target phase. Finally, the day on which this temperature sum is exceeded at each station is determined and interpolated on the Germany-wide grid.

The PHASE model as implemented in the *PHASE Analyser*[2] web application was utilized to extract all DWD-observed phenological phases of winter wheat, winter barley and winter rapeseed (see Table 1). This service is a *GeoServer*[3]-based open source map server application that provides daily updated phenological raster data. These data sets were created using the PHASE model that was implemented as a daily modelling routine. The results are distributed under strict application of *Open Geospatial Consortium* (OGC) standards via a *Drupal*[4]-based web interface. In Fig. 5, the subsequent processing steps are displayed, beginning with the automatic modelling, followed by the creation of the web services and visualisation. The underlying infrastructure utilises *php*-based parser applications to integrate the required input data automatically into a *PostgreSQL/PostGIS* database. These back-end procedures are required to call the *PHASE* model at a daily temporal interval ("Cronjob").

3.2 Spectral Separability Assessment

The separability analysis was exemplarily implemented for the three crop types winter wheat, winter barley and winter rapeseed, which dominate the Harz study site. These species are most relevant for regional modelling approaches or yield estimations, as they cover around 60% of the total agricultural areas in the study sites. The high number of available classes (> 60) would lead to inaccurate results, because they include numerous crop types that are only cultivated on a small number of fields. Thus, the analyses were performed only on fields that were covered by the target crops and other crop types cultivated at least on 25 fields on average over the study period, which includes summer-cropped wheat, durum wheat, barley and oats, as well as winter rye, maize, sugar beet, potatoes and perennial grassland. Consequently, the used fields covered more than 75% of all fields in the study area, but the number of classes was reduced by more than 80%.

Vegetation index profiles show specific (inter-) annual patterns that can be traced back to plant phenological phases (Förster et al. 2012) and are, along with single band reflectance values, powerful features for crop classification (Löw et al. 2013). Thus, a setting similar to the configuration presented by Löw et al. (2013) was implemented. It combined single-band reflectances and four well-established vegetation indices as spectral features (see Table 2). The Normalized Difference Vegetation Index (NDVI, Rouse et al. 1974), which is the most frequently applied vegetation index for agricultural remote sensing applications, shows a variety of issues that can be solved by including of other spectral

[2] http://phase.geo.uni-halle.de/phase-wms-dienste.

[3] http://www.geoserver.org.

[4] http://www.drupal.org.

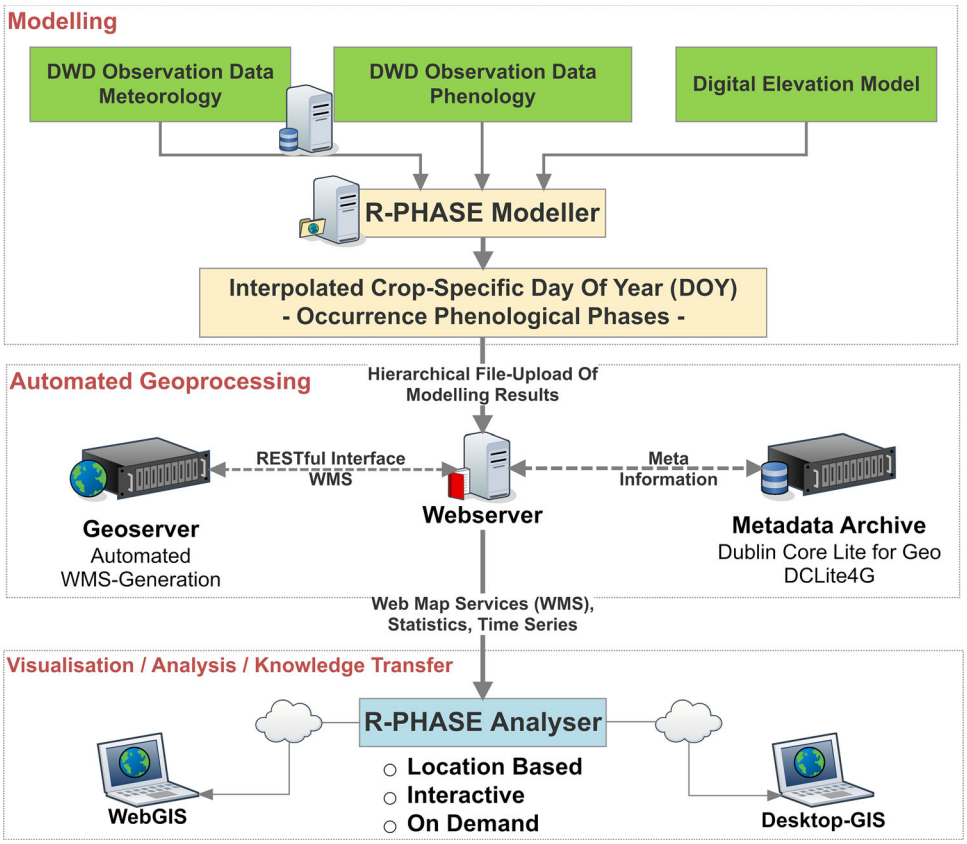

Fig. 5 Workflow of the automatic generation of modelled phenological phases

Table 2 Vegetation indices and other spectral features used for the calculation of separability profiles

Index	Abbreviation	Formula	References
Normalized Difference Vegetation Index	NDVI	$\frac{\rho_{NIR}-\rho red}{\rho_{NIR}+\rho red}$	Rouse et al. (1974)
Green NDVI	GNDVI	$\frac{\rho_{NIR}-\rho green}{\rho_{NIR}+\rho green}$	Buschmann and Nagel (1993)
Red edge NDVI	RE_NDVI	$\frac{\rho_{NIR}-\rho red\ edge}{\rho_{NIR}+\rho red\ edge}$	Gitelson and Merzlyak (1994)
Wide-Dynamic Range Vegetation Index	WDRVI	$\frac{0.15\times\rho_{NIR}-\rho red}{0.15\times\rho_{NIR}+\rho red}$	Gitelson (2004)
Blue band	B1	ρ_{blue}	Tyc et al. (2005)
Green band	B2	ρ_{green}	
Red band	B3	ρ_{red}	
Red edge band	B4	$\rho_{red\ edge}$	
Near infrared band	B5	ρ_{NIR}	

ρ_{region} reflectance in the specified spectral region

bands or mathematical modifications. Three of these modifications (Table 2) were incorporated into the classification scheme. Each of the modifications is more sensitive to a specific shortcoming of the NDVI, specifically the Red Edge NDVI (RE_NDVI) for the red edge region as a vitality indicator, the Green NDVI (GNDVI) for green chlorophyll content and the Wide-Dynamic Range Vegetation Index (WDRVI) for saturation of the band reflectances.

Classifications were applied at the field level, because object-based classifications are reported to be superior to pixel-based approaches when high resolution imagery is used (De Wit and Clevers 2004; Blaschke 2010). A single-date classification approach was chosen to reflect the high probability that in some parts of the world, e.g., cloud cover limits the availability of multiple data sets useful for classification (Nitze et al. 2015; Frantz et al. 2017). For each acquisition date and parcel, mean values of the spectral bands and indices extracted from the RapidEye data were calculated.

Next, a classification scheme was set up that performs the widely used supervised random forest classification (Breiman 2001) implemented by Liaw and Wiener (2002) in the statistical computing environment *R* (R Core Team 2016).

The critical parameters node size and number of trees was set to 2 and 1000, respectively. This ensures stable results and reduces the chance of building over-fitted trees (Kuhn and Johnson 2013).

Three feature evaluations were performed per acquisition date, to find the best-suited feature for the separation of each target crop. Here, each spectral feature was classified sequentially. Evaluations were performed as binary classifications, i.e., features of fields covered by one out of the three target crops, winter wheat, winter barley and winter rapeseed, were classified against those from all other fields (see right box in Fig. 4). Finally, the $F1$- or F-measure, introduced by van Rijsbergen (1979), was calculated as a metric for the assessment of the class-wise classification accuracy. The F-measure is defined as the harmonic mean of the positive predictive value (precision) and the true positive rate (recall) of the classification confusion matrix. This statistic is especially meaningful for non-parametric classifiers and proved to be a good accuracy measure for random forest-based crop classifications (Löw et al. 2015). Precision, recall and F-measure were calculated based on the contingency table produced by the random forest implementation.

The specific optimal index per crop and acquisition date was then defined as the spectral feature with the highest F-measure. Based on the F-measure of all acquisition dates, a temporal profile of separability was computed (see Sect. 3.3).

3.3 Phenological Indexing of Inter-Annual Separability Patterns

All available modelled phases (see Table 1) for the three target crops for the years 2010–2015 were processed. Following Möller et al. (2017), a test site-specific phenological phase was considered as the period between the crop-specific medians of two subsequently beginning phases. The period between the first DOY and the first observed phase of a year was named after the last observed phase of the previous year.

The phenological indexing of separability patterns was started with the coupling of the separability profiles and the corresponding phenological phases. The duration of the last observed phase before harvest was prolonged by 1 week to address the stronger small-scale variations in harvesting due to the sequence of operations within farming cooperatives. The shift of the modelled harvesting ensured that the majority of all fields were actually harvested on the acquisition date. Furthermore, since the phases were linked to spectral response of the crops, the prolongation also addressed the fact that freshly harvested fields are often hard to distinguish from fields with fully ripened cereals using the spectral regions recorded by RapidEye. This similarity is due to the fact that crop residues remaining on the fields spectrally resemble ripened crops before harvest.

A score value R^P was calculated to assess the reliability of a phase to be optimal for separation of the corresponding crop type according to the following equation.

$$R^P = N^P_{\text{obs}} \times (F^P_{\text{max}} + F^P_{\text{mean}}) \quad (1)$$

N^P_{obs} describes how often a phase has been identified as optimal for a year and study site. For these observations, F^P_{mean} and F^P_{max} represent the maximum and mean of the F-measures. This score value was required to account for differences of the temporal distribution of the satellite images. Such temporal gaps might possibly lead to undetected periods of high separability (see Fig. 3 and Sects. 1 and 2.2).

4 Results

4.1 Phenological Development

The phenological differences for all phases of the target crops between 2010 and 2015 vary among the sites, which is exemplarily shown for winter wheat in Fig. 6. For winter barley and winter rapeseed, these patterns are similar. In general, it could be observed that the growing season in DM starts with a delay of 1–4 days compared to HM and 5–12 days compared to HL, with the exception of 2015 when shooting started earliest in DM. The end of the growing season showed an opposite tendency, because the model predicted the start of the harvest period in DM 3–12 days earlier than in HL.

4.2 Spectral Separability Profiles

For each test site and each of the three investigated crop types, spectral separability profiles were computed from the

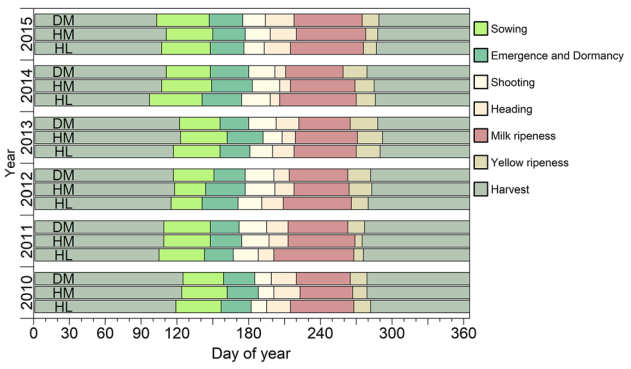

Fig. 6 Modelled phenological phases of winter wheat in the three study sites for the years 2010 to 2015. The beginning of a phase is defined by the modelled average day of year of phenological event in the study site

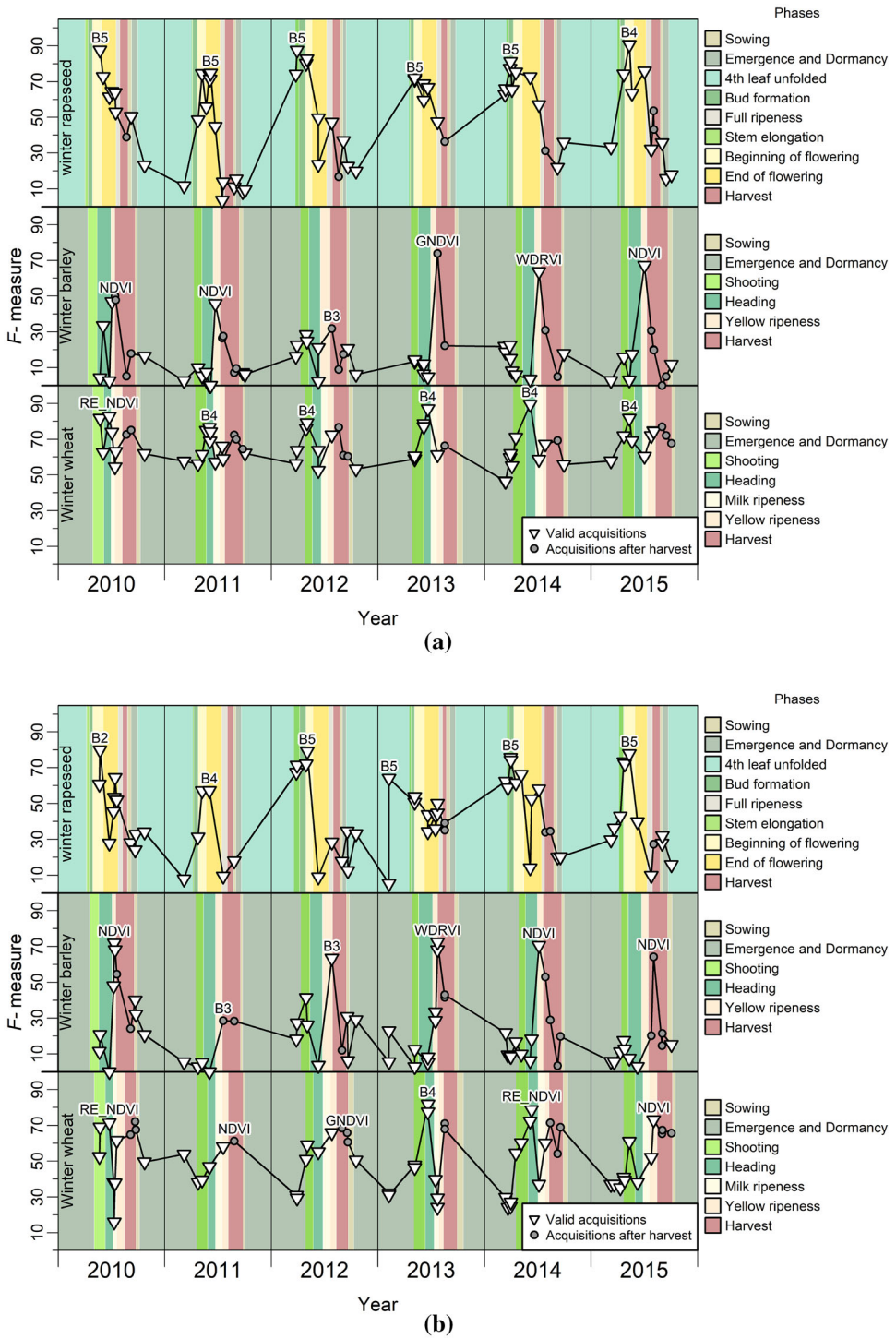

Fig. 7 Spectral separability profiles based on RapidEye data and modelled phenological phases for winter wheat, winter barley and winter rapeseed for the HL (**a**) and the HM (**b**) sites between 2010 and 2015. The spectral feature with the highest F-measure per season is noted above the respective peak in the profiles. Valid acquisitions are timed between sowing and harvest. *NDVI* Normalised Difference Vegetation Index, *WDRVI* Wide-dynamic Range Vegetation Index, *RE_NDVI* Red Edge NDVI, *GNDVI* Green NDVI

highest F-measure of all classified spectral features per random forest classification. The profiles are displayed as black lines in Fig. 7 separately for the two Harz sub-sites. Intra-annual patterns are visible that are characterized by peaks and depressions of the calculated F-measure profiles. For instance, winter wheat and winter rapeseed showed higher F-measures at their annual maximum and less intra-annual variations than winter barley.

4.3 Phenological Indexing of High Separability Time Frames

The RapidEye imagery of the HL and HM test sites, for which the annual F-measure was maximal, were coupled to the corresponding phenological phases (Fig. 7). The phase with the highest reliability value R^P (Eq. 1) value is considered as the optimal phenological indicator phase. All other phases,

Table 3 Indicator phases

Crop	Phase	N^P_{obs}	DOY_{min}	DOY_{max}	F^P_{max}	F^P_{mean}	R^P	Spectral feature
Winter wheat	Heading	7	153	175	0.89	0.81	11.94	Band 4
	Shooting	4	122	130	0.82	0.80	6.48	Band 4
	Yellow ripening	3	200	213	0.73	0.66	4.16	GNDVI
Winter barley	Yellow ripening	8	173	207	0.73	0.63	10.80	WDRVI
	Shooting	3	111	125	0.28	0.20	1.44	NDVI
Winter rapeseed	Beginning of flowering	6	123	143	0.91	0.81	10.31	Band 4
	Stem elongation	2	88	88	0.88	0.88	3.50	Band 5
	Bud formation	2	88	88	0.81	0.78	3.19	Band 5
	Fourth leaf unfolded	3	274	37	0.64	0.33	2.90	Band 5
	End of flowering	2	153	157	0.75	0.66	2.81	Band 5

Number of observations (N^P_{obs}), earliest (DOY_{min}) and latest (DOY_{max}) observation, maximum (F^P_{max}) and average F-measure (F^P_{mean}), reliability (R^P) and spectral variable with highest F-measure (phases: see Table 1). The phases are ordered decreasingly according to their R^P value

during which the annual separability reaches its maximum at least twice, are referred to as alternative phenological indicator phases. Table 3 summarizes the resulting indicator phases and corresponding separability metrics.

For winter barley, only the phases "yellow ripening" and "shooting" were detected to be optimal at least twice. "Yellow ripening" showed the highest R^P value, since it was considered as optimal a total of eight times. In addition, both F^P_{max} and F^P_{mean} are significantly higher than for "shooting", which was selected three times. The WDRVI reached the highest F-measure during "yellow ripening", while during "shooting", the NDVI outperformed all other spectral features.

Three indicator phases could be found for winter wheat. "Heading" ($N^P_{obs} = 7$) outperformed "yellow ripening" ($N^P_{obs} = 3$) and "shooting" ($N^P_{obs} = 4$). F^P_{max} and F^P_{mean} are comparable for "shooting" and "heading", while for "yellow ripening", the F-measure metrics were significantly lower. The red edge band was the best-performing spectral feature both for "heading" and "shooting".

The phases usable as indicators for rapeseed classification are, as expected, dominated by "beginning of flowering", with a R^P value of 10.30. During flowering, the NIR reflectance of rapeseed starts to decrease while the reflectance measured by the green spectral band, that is nearest to the yellow wavelength region, increases as a consequence of the intense yellow colour. These changes are visible in the reflectance spectrum and unique during that temporal period compared to the other crops that are still highly vital. Furthermore, four time frames are potential alternatives, namely the phases "stem elongation", "bud formation", "end of flowering" and "fourth leaf unfolded", which are characterized by very low reliabilities. The F^P_{max} of 0.88 and 0.81 for "stem elongation" and "bud formation" almost reach the values for "beginning of flowering" (0.91). It is also noteworthy that "stem elongation" showed a higher F^P_{mean} value than "beginning of flowering", but is observed as optimal much rarer.

4.4 Regional Transferability and Validation of Indicator Phases

The F-measure separability profiles of the DM study site, for which both environmental conditions and temporal distribution of the data sets are different compared to the two Harz sites (see Sect. 2), were also calculated and are visualised in Fig. 8.

The profiles clearly show a dependency on data availability, since not all indicator phases are represented at least by one data set in every year. Accordingly, the separability maximum for winter wheat was observed during "heading" in 2010 and 2011, which was found to be the optimal indicator phase for winter wheat separation based on the Harz sites analyses. Also, the annual maxima of the F-measure separability profiles for winter barley in 2010 and for winter rapeseed in 2012 exactly correspond to the detected optimal phases.

In 2012, no data set was available that represented "heading" for winter wheat. There, the separability maximum was observed during the alternative phase "shooting". However, this maximum is found six days before the optimal phase "heading" starts, which corresponds to the uncertainty caused by modelling, observation and small-scale differences between parcels (Gerstmann et al. 2016a).

The maximum separability for winter barley in 2011 was found during the alternative phase "shooting". In 2012, no data set was available during "yellow ripening" or "shooting". In general, the F-measures for winter barley are low (< 0.5) due to relatively small sample sizes (< 20) and thus the findings are only of limited reliability.

No data set was available during the optimal phase of winter rapeseed ("beginning of flowering") for 2010. The detected separability maximum occurred during the alternative phase "end of flowering". "End of flowering" was also detected as optimal for 2011, although a data set rep-

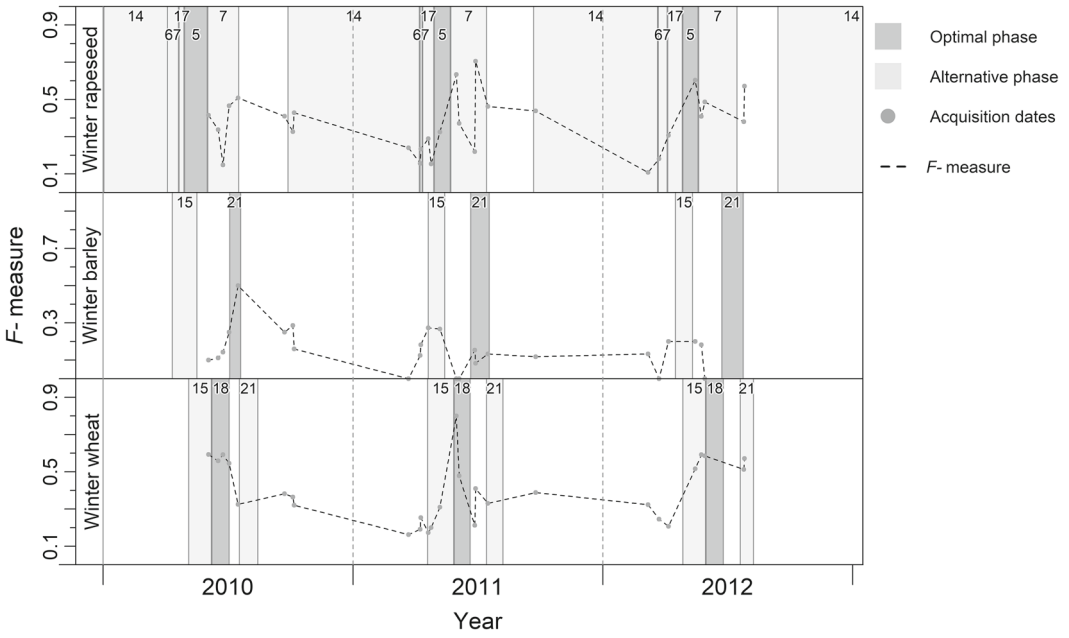

Fig. 8 *F*-measure separability profiles for winter wheat, winter barley and winter rapeseed for the DM site and the phenological indicator phases detected through the analysis of the HL and HM test sites (see Fig. 7; Table 3). The numbers represent the phenological phases (see Table 1)

resenting the indicator phase "beginning of flowering" was available.

In summary, the separability maxima coincide with the defined phenological indicator phases (optimal or alternative) in eight out of nine crop-year-combinations. For three crop-year-combinations, no data set was available representing the optimal phase, which underlines the necessity for definition of alternative phases.

5 Discussion

5.1 Phenological Uncertainty

As the results showed, phenological phases can be used as indicators for acquisition time frames suitable for spectral discrimination of crops. However, as described in Sect. 4.4, sources of potential inaccuracy have to be considered when interpreting the results. In this study, this especially concerns the phenological interpolation results of the PHASE model, which also provides the kriging standard deviation as spatial inaccuracy metric (Gerstmann et al. 2016a). This metric can be used for additional interpretation of the separability profiles. Accordingly, the model uncertainty is expected to be between 3 and 8 days (Möller et al. 2017), with the majority of phases less than 5 days. Based on data of 2011 Gerstmann et al. (2016a) reported accuracies for winter wheat between < 4 and 11 days, for winter barley between 3.8 and 5 days and for winter rapeseed between 5 and 7 days. Modelling earlier phases was associated with higher uncertainty than that of the later phases. Due to small-scale variations in plant phenology, the period of phase transition must be understood as a gradient rather than a sharp turnover. Phenological variations in turn are caused by management or cultivar-specific differences that cannot be observed by the monitoring network. Additionally, the reliability of volunteer-based observations is complicated to assess due to the subjectiveness of the observation design (Flanagin and Metzger 2008; Mehdipoor et al. 2015).

5.2 Sensor-borne Uncertainty

The defined phenological patterns also depend on the temporal composition of the satellite data sets. Time frames of high separability may remain undetected in situations of reduced image availability. Since the image acquisitions for the Harz data set are almost synchronous for both sub-sites, this factor is assumed to be negligible for this site. However, it becomes more important, e.g., for winter rapeseed at the Demmin site, for which the first acquisition in 2010 was available in the "end of flowering" phase only. In other words, this data set was acquired after the detected optimal phase "beginning of flowering", which underlines the necessity of having alternative indicator phases (see Sect. 5.4).

5.3 Separability Profiles

The separability profiles are based on a small set of spectral features which are well-established for a large variety of crop types. Other sensors such as Sentinel-2 provide additional

bands especially in the red edge and short-wave infra-red wavelength ranges useful for crop mapping (e.g., Immitzer et al. 2016). As the separability profiles are partly characterized by strong variations, optimized spectral indices, as proposed by Rivera et al. (2014) or Gerstmann et al. (2016b), could lead to more stable profiles.

The accuracies achieved by application of machine learning algorithms such as random forest generally tend to increase with the number of included predictor variables. Hence, this would be the preferred strategy instead of classifying each feature separately if the approach is applied in practise. The performance analysis of the spectral features could then be assessed by their variable importance, a parameter that ranks the predictor variables according to their contribution to the prediction accuracy. In doing so, both classification results and time frame reliability should increase. As a side-effect, the probability to produce an overfitted model would be reduced, although Breiman (2001) reported that random forest classifiers are relatively robust against overfitting to the training data compared to other classification approaches. However, in this study, we refrained from this option and applied a single-variable random forest approach instead to reduce the complexity for the demonstration of the approach and the results. Nevertheless, overfitting might be an issue when only one predictor variable is used as it is the case in this study, but several variations of the parametrisation (tree size, number of trees, number of terminal nodes) of the random forest classifier only slightly affected the results.

5.4 Alternative Indicator Phases

Alternative phases, that have been detected as optimal at least twice, can be valuable indicators for data set selection when no satellite data are available during the optimal phenological time frame. However, alternative phases can produce additional inaccuracies. For instance, the alternative phases "bud formation" and "stem elongation" for winter rapeseed showed comparably low R^P values, since they represent short phases only. Thereby, exact start and end dates are complicated to define due to the modelling uncertainty of the beginning phenological phases (see Sects. 4.1 and 5.1).

From the botanical perspective, "bud formation" and "stem elongation" are actually more reliable than indicated by R^P. The selection of these early spring phases is due to shorter winter dormancy of rapeseed compared to the dormancy of cereal crops. This difference is measurable by vegetation indices as an earlier increase of photosynthetic activity. The alternative indicator phase "fourth leaf unfolded" occurs before winter dormancy and is characterized by significantly higher green vegetation coverage compared to winter cereals, which is also measurable by vegetation indices.

5.5 Validation and Regional Transferability

The regional transferability was proven for most of the investigated crops and years. However, as mentioned earlier in Sects. 4.4 and 5.2, no data set was available during the optimal phase for some crop-year combinations, which means that the expected separability maximum could not be found.

The bad performance ($F < 0.5$) for winter barley could be due to the fact that barley was cultivated on between 20 and 40 fields only. This number is remarkably lower than for winter wheat and winter rapeseed. Consequently, data gaps due to clouds or other factors had a stronger influence on statistical analysis than for wheat and barley, since sample sizes can easily fall below the lower threshold of approximately 30 samples for reliable statistical analyses.

5.6 Application Strategies

Both mono- and multi-temporal approaches can benefit from the presented methodology, once time frames have been defined also for other relevant crop types. For mono-temporal classification, a single data set should be selected for a point in time when the growth status of all target crops is within an optimal or alternative phenological indicator phase.

Multi-temporal classification should include at least one data set acquired during the optimal phase for all relevant crops at the study site. Thus, if the modelled indicator phases are available in the *PHASE Analyser* distribution system (see Sect. 3.1), they can be obtained and used for the user-specific application without the necessity of having a continuous temporal coverage of the satellite data over the complete growing cycle.

For multi-temporal classification, the approach would most likely lead to some congruent acquisition dates that are optimal for most of the crop types. For instance, the alternative phase "shooting" of winter wheat and winter barley is congruently timed with the optimal phase "beginning of flowering" for winter rapeseed. This supports the findings by, for instance, Schmidt et al. (2014) and Murakami et al. (2001). In their studies, a total of three to four images was found to be necessary to achieve a sufficient classification accuracy. However, the proof if the identified indicator phases are also valid in multi-temporal classifications is still an outstanding task.

The presented approach combined with the described web interface can assist data-intensive classification approaches based on satellite data of high spatial, spectral and temporal resolution like Landsat 8, Sentinel-2 or HJ-1, since it represents an effective strategy to reduce the calculation effort. This reduction is due to the focus on the most significant spectral features instead of all spectral bands provided by these sensors and on the reduction of acquisition dates to

be included in the classification during the most distinctive temporal periods.

Other applications which might benefit from phenological indicator phases and pre-assessment are hierarchical classification frameworks (e.g., Förster et al. 2012; Forkuor et al. 2015) and phenology-adaptive algorithms for compositing (Frantz et al. 2017). In the latter approach, the sub-images that are merged into phenology-adaptive, pixel-based image composites have to be selected to preferably represent an identical growth status of the land use classes. Especially in regions of persistent cloud coverage like Zambia, which Frantz et al. (2017) investigated in their study, the approach presented in this study can improve the timing of such target dates. Lastly, crop yield estimations for food security especially in regions of large inter-annual phenological variations (Meroni et al. 2014) might also be improved by the application of phenological indicator phases.

All these applications focus on the retrospective selection of already acquired remote sensing data. Beyond this, the spatially explicit knowledge of the phenological indicator phases can be applied to accurately schedule airborne flight campaigns or tasked satellite image acquisitions.

6 Conclusions and Outlook

This study presented a methodology that combines multi-annual crop separability profiles and automatically modelled phenological phases to derive indicator phases. These phases represent periods of time within a growing season during which spectral separability of a crop type is maximal, at least for the selected classification approach.

We tested the methodology for three frequently cultivated crops in Germany at two study sites under different environmental conditions and different data characteristics in terms of pre-processing and satellite image acquisition dates. The results showed that for each of the investigated crop types a minimum of two stable indicator phases exist.

The approach is currently spatially limited due to Germany's unique phenological observation network. However, the promising results suggest to put further research on its transferability to other regions of similar natural conditions.

Apart from crop mapping, the general framework of linking systematically collected phenological ground truth observations to vegetation reflectance patterns is ready to be tested for a large variety of other possible applications within an agricultural context. Thereby, the web interface providing phenological raster data for entire Germany enables various user groups to apply this approach of using phenological indicator phases to their own specific study regions.

Acknowledgements This study was supported by the German Ministry of Economics and Energy (BMWi) and the German Aerospace Center (DLR) under grant 50EE1263. The authors want to thank Dr. Daniel Doktor and Xingmei Xu (Helmholtz Centre for Environmental Research—UFZ) for the atmospheric correction of the RapidEye data sets for the Harz study site and Dr. Patrick Knöfel (Julius Maximilians University Würzburg) for organizing the preprocessing of the RapidEye data sets of Demmin. The authors want also express their gratitude to Dr. Erik Borg (German Aerospace Center—DLR) for the preparation and provision of the land use information of the Demmin site.

References

Azar R, Villa P, Stroppiana D, Crema A, Boschetti M, Brivio PA (2016) Assessing in-season crop classification performance using satellite data: a test case in Northern Italy. Eur J Remote Sens 49(1):361–380

Blaschke T (2010) Object based image analysis for remote sensing. ISPRS J Photogramm Remote Sens 65(1):2–16

Bogena HR (2016) Tereno: German network of terrestrial environmental observatories. J Large-scale Res Facil JLSRF 2:52

Borg E, Daedelow H, Apel M, Missling KD (2013) Rapideye science archive: Remote sensing data for the German scientific community. In: Borg E, Daedelow H, Johnson R (eds) RESA, GITO mbH Verlag, Berlin, RESA Workshop, Neustrelitz, vol 3, pp 5–20. http://elib.dlr.de/81718/

Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324

Buschmann C, Nagel E (1993) In vivo spectroscopy and internal optics of leaves as basis for remote sensing of vegetation. Int J Remote Sens 14(4):711–722. https://doi.org/10.1080/01431169308904370

Conrad C, Rahmann M, Machwitz M, Stulina G, Paeth H, Dech S (2013) Satellite based calculation of spatially distributed crop water requirements for cotton and wheat cultivation in fergana valley, uzbekistan. Glob Planet Change 110:88–98

Conrad C, Dech S, Dubovyk O, Fritsch S, Klein D, Löw F, Schorcht G, Zeidler J (2014) Derivation of temporal windows for accurate crop discrimination in heterogeneous croplands of Uzbekistan using multitemporal RapidEye images. Comput Electron Agric 103:63–74

De Wit A, Clevers J (2004) Efficiency and accuracy of per-field classification for operational crop mapping. Int J Remote Sens 25(20):4091–4112

DWD Climate Data Center (2017a) Historical daily station observations (temperature, pressure, precipitation, sunshine duration, etc.) for Germany, version v005

DWD Climate Data Center (2017b) Multi-annual means of grids of air temperature (2m) over Germany, 1981-2010, version v1.0

DWD Climate Data Center (2017c) Phenological observations of crops from sowing to harvest (annual reporters, historical), Version v003

Flanagin AJ, Metzger MJ (2008) The credibility of volunteered geographic information. GeoJournal 72(3):137–148. https://doi.org/10.1007/s10708-008-9188-y

Forkuor G, Conrad C, Thiel M, Landmann T, Barry B (2015) Evaluating the sequential masking classification approach for improving crop discrimination in the Sudanian Savanna of West Africa. Comput Electron Agric 118(Suppl C):380–389. https://doi.org/10.1016/j.compag.2015.09.020

Förster S, Kaden K, Förster M, Itzerott S (2012) Crop type mapping using spectral-temporal profiles and phenological information. Comput Electron Agric 89:30–40

Frantz D, Röder A, Stellmes M, Hill J (2017) Phenology-adaptive pixel-based compositing using optical earth observation imagery. Remote Sens Environ 190:331–347

Gerstmann H, Doktor D, Gläßer C, Möller M (2016a) PHASE: a geostatistical model for the kriging-based spatial prediction of crop phenology using public phenological and climatological observations. Comput Electron Agric 127:726–738

Gerstmann H, Möller M, Gläßer C (2016b) Optimization of spectral indices and long-term separability analysis for classification of cereal crops using multi-spectral RapidEye imagery. Int J Appl Earth Obs Geoinf 52:115–125

Gitelson AA (2004) Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. J Plant Physiol 161(2):165–173. https://doi.org/10.1078/0176-1617-01176

Gitelson AA, Merzlyak MN (1994) Quantitative estimation of chlorophyll-*a* using reflectance spectra: experiments with autumn chestnut and maple leaves. J Photochem Photobiol B 22(3):247–252. https://doi.org/10.1016/1011-1344(93)06963-4

Guerschman J, Paruelo J, Bella CD, Giallorenzi M, Pacin F (2003) Land cover classification in the argentine pampas using multi-temporal landsat tm data. Int J Remote Sens 24(17):3381–3402

Immitzer M, Vuolo F, Atzberger C (2016) First experience with sentinel-2 data for crop and tree species classifications in central europe. Remote Sens 8(3):166

Inan H, Sagris V, Devos W, Milenov P, van Oosterom P, Zevenbergen J (2010) Data model for the collaboration between land administration systems and agricultural land parcel identification systems. J Environ Manag 91(12):2440–2454

Kaspar F, Zimmermann K, Polte-Rudolf C (2014) An overview of the phenological observation network and the phenological database of Germany's national meteorological service (Deutscher Wetterdienst). Adv Sci Res 11:93–99

Krauß T, d'Angelo P, Schneider M, Gstaiger V (2013) The fully automatic optical processing system CATENA at DLR. ISPRS Hannover Workshop 1:177–181

Kuhn M, Johnson K (2013) Applied predictive modeling. Springer, New York

Liaw A, Wiener M (2002) Classification and regression by randomforest. R News 2(3):18–22. http://CRAN.R-project.org/doc/Rnews/

Löw F, Michel U, Dech S, Conrad C (2013) Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using support vector machines. ISPRS J Photogramm Remote Sens 85:102–119. https://doi.org/10.1016/j.isprsjprs.2013.08.007

Löw F, Knöfel P, Conrad C (2015) Analysis of uncertainty in multi-temporal object-based classification. ISPRS J Photogramm Remote Sens 105:91–106. https://doi.org/10.1016/j.isprsjprs.2015.03.004

Mehdipoor H, Zurita-Milla R, Rosemartin A, Gerst K, Weltzin J (2015) Developing a workflow to identify inconsistencies in volunteered geographic information: a phenological case study. PLoS One 10(10):1–14

Meroni M, Rembold F, Verstraete MM, Gommes R, Schucknecht A, Beye G (2014) Investigating the relationship between the inter-annual variability of satellite-derived vegetation phenology and a proxy of biomass production in the Sahel. Remote Sens 6(6):5868–5884. https://doi.org/10.3390/rs6065868, http://www.mdpi.com/2072-4292/6/6/5868

Möller M, Gerstmann H, Feng G, Dahms T, Förster M (2017) Coupling of phenological information and $NDVI$ time series: limitations and potentials for the assessment and monitoring of soil erosion risk. CATENA 150:192–205

Murakami T, Ogawa S, Ishitsuka N, Kumagai K, Saito G (2001) Crop discrimination with multitemporal SPOT/HRV data in the Saga Plains, Japan. Int J Remote Sens 22(7):1335–1348. https://doi.org/10.1080/01431160151144378

Nagol JR, Sexton JO, Kim DH, Anand A, Morton D, Vermote E, Townshend JR (2015) Bidirectional effects in landsat reflectance estimates: is there a problem to solve? ISPRS J Photogramm Remote Sens 103:129–135

Nitze I, Barrett B, Cawkwell F (2015) Temporal optimisation of image acquisition for land cover classification with random forest and modis time-series. Int J Appl Earth Obs Geoinf 34:136–146

Peña-Barragàn JM, Ngugi MK, Plant RE, Six J (2011) Object-based crop identification using multiple vegetation indices, textural features and crop phenology. Remote Sens Environ 115(6):1301–1316. https://doi.org/10.1016/j.rse.2011.01.009

R Core Team (2016) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/

Rabus B, Eineder M, Roth A, Bamler R (2003) The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar. ISPRS J Photogramm Remote Sens 57(4):241–262

Richter R, Schläpfer D (2015) Atmospheric/topographic correction for satellite imagery (ATCOR-2/3 User Guide, Version 9.0.0, June 2015). DLR report DLR-IB, pp 565–01

Rivera J, Verrelst J, Delegido J, Veroustraete F, Moreno J (2014) On the semi-automatic retrieval of biophysical parameters based on spectral index optimization. Remote Sens 6(6):4927–4951. https://doi.org/10.3390/rs6064927

Rouse J, Jr RH, Schell JA, Deering D (1974) Monitoring vegetation systems in the Great Plains with ERTS, NASA SP-351. In: Third ERTS-1 symposium, vol 1. NASA, Washington, DC, pp 309–317

Schmidt T, Schuster C, Kleinschmit B, Förster M (2014) Evaluating an intra-annual time series for grassland classification—how many acquisitions and what seasonal origin are optimal? IEEE J Sel Top Appl Earth Obs Remote Sens 7(8):3428–3439

Tyc G, Tulip J, Schulten D, Krischke M, Oxfort M (2005) The RapidEye mission design. Acta Astronaut 56(1–2):213–219

van Niel TG, McVicar TR (2004) Determining temporal windows for crop discrimination with remote sensing: a case study in south-eastern Australia. Comput Electron Agric 45(1–3):91–108. https://doi.org/10.1016/j.compag.2004.06.003

van Rijsbergen C (1979) Information retrival, 2nd edn. Springer, Berlin

Whitcraft A, Vermote E, Becker-Reshef I, Justice C (2015) Cloud cover throughout the agricultural growing season: impacts on passive optical earth observations. Remote Sens Environ 156:438–447

Xu X, Conrad C, Doktor D (2017) Optimising phenological metrics extraction for different crop types in Germany using the moderate resolution imaging spectrometer (MODIS). Remote Sens 9(3):254. https://doi.org/10.3390/rs9030254