

ORIGINAL ARTICLE

Special Section: Affordable Phenomics

Comparison of PhenoCams and drones for lean phenotyping of phenology and senescence of wheat genotypes in variety testing

Simon Treier^{1,2}  | Nicolas Vuille-dit-Bille¹  | Margot Visse-Mansiaux¹  |
 Frank Liebisch³  | Helge Aasen⁴  | Lukas Roth²  | Achim Walter²  |
 Juan M. Herrera¹ 

¹Cultivation Techniques and Varieties in Arable Farming Group, Plant-Production Systems Division, Agroscope, Nyon, Switzerland

²ETH Zürich, Institute of Agricultural Sciences, Universitätstrasse 2, Zurich, Switzerland

³Water Protection and Substance Flows Group, Agroecology and Environment Division, Agroscope, Zürich, Switzerland

⁴Earth Observation of Agroecosystems Team, Agroecology and Environment Division, Agroscope, Zürich, Switzerland

Correspondence

Simon Treier, Cultivation Techniques and Varieties in Arable Farming group, Plant-Production Systems Division, Agroscope, Route de Duillier 60, Nyon, 1260, Switzerland.
 Email: simon.treier@agroscope.admin.ch

Assigned to Associate Editor Jennifer A Lachowiec.

Funding information

Agroscope; European Union's Horizon 2020 research and innovation program projects; InnoVar — Next generation variety testing for improved cropping on European farmland, Grant/Award Number: 818144; Invite - INnovations in plant VarIety Testing in Europe to foster the introduction of new

Abstract

In variety testing and breeding of wheat (*Triticum aestivum* L.), it is crucial to know the timing of phenological stages and the senescence behavior of genotypes to select for locally adapted varieties. Sound knowledge of the timing of phenological stages also allows for a more meaningful interpretation of measurements such as yield, quality, or disease ratings. In the presence of stresses, only a combined characterization of phenology and environmental conditions can allow for insights into unraveling stress resistance and stress avoidance. Capturing these traits visually in the field is very time-consuming. Here, a semimobile PhenoCam setup was used to track phenology and senescence from ear emergence to full maturity. PhenoCams mounted on field masts took images of wheat plot trials on a daily basis. In a partial least squares regression analysis, the temporal features of multiple vegetation indices were combined in one model to track phenology and senescence. The method was compared with visual reference methods and repeated drone flights with a multispectral camera. The Pearson's correlation between visual reference methods and PhenoCam predictions was stronger than 0.8, often above 0.9, for most stages. An economic analysis showed that PhenoCams are economically interesting, especially for observing remote experimental sites. Thus, PhenoCams offer a cost-effective replacement for visual ratings of phenology and senescence, particularly in the context of multienvironment trials.

Plain Language Summary

When testing and breeding wheat varieties, it is crucial to know when important stages of plant development occur. In different environments, early, mid or late varieties may be better adapted to local conditions. Visual assessment of plant development in the field is very time consuming, requiring multiple field visits during the

Abbreviations: DAS, days after sowing; DN, digital number; loess, locally estimated scatter plot smoothing; RGB, red-green-blue; RGBVI, red-green-blue vegetation index; RMSE, root mean square error; PLSR, Partial least squares regression; SIFT, scale-invariant feature transformation; VI, vegetation index.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *The Plant Phenome Journal* published by Wiley Periodicals LLC on behalf of American Society of Agronomy and Crop Science Society of America.

varieties better adapted to varying biotic and abiotic conditions and to more sustainable crop management practices, Grant/Award Number: 817970

growing season. Here, a PhenoCam setup was used to track plant development from the time when wheat ears emerge to full maturity. Solar-powered cameras mounted on field poles captured daily images of wheat plot trials. The dynamic color changes of the different varieties on the images were used to track the development of the plants. We could show that PhenoCams are suitable for tracking plant development. They offer a cost-effective substitute to visual reference methods, especially for distant trial fields, as it is not necessary to visit the fields several times during the growing season when using PhenoCams.

1 | INTRODUCTION

In variety testing, breeding, and research of bread wheat genotypes (*Triticum aestivum* L.), it is crucial to know the timing of phenological stages and the senescence behavior of the individual genotypes. The temporal characterization of plant development allows the selection of genotypes better adapted to local climates and soils. For example, early flowering behavior allows plants to escape early summer drought and heat stress during the sensitive stage around meiosis (e.g., Rogger et al., 2021), whereas later flowering behavior allows plants to escape late frosts around meiosis (Langer et al., 2014). As the climatic conditions in Central Europe are changing, the use of adapted wheat genotypes can be a strategy to mitigate adverse effects on yield and reduce production risks (Holzkämper et al., 2015; Rogger et al., 2021). According to Asseng et al. (2013), wheat yield is more prone to uncertainty with increasing levels of CO₂ and temperature. For every 1°C increase in temperature above the temperature optimum, there is an estimated decrease of 6% in wheat yield, and yield becomes more variable in space and time (Asseng et al., 2015). Consequently, a diversity of wheat germplasms must be maintained and developed to provide a diverse set of adaption strategies to grow wheat in future climate conditions (Kahiluoto et al., 2019).

In field experiments, genotypes with different phenological development may not be exposed to the same stresses in the same year. Accurate knowledge of the timing of the phenological stages thus allows for a more meaningful interpretation of other measurements, such as yield. For example, low radiation at jointing, booting (Jia et al., 2021), young microspore stage (10–12 days before heading; Yang et al., 2020), ear emergence (Welbank et al., 1968), anthesis (Ford & Thorne, 1975), and throughout grain filling (Jia et al., 2021) can significantly reduce yield, mainly due to a lower number of grains per ear and, therefore, a reduced sink size for carbon accumulation. The young microspore stage is generally sensitive to stresses (Dong et al., 2017; Yang et al., 2020), and heat or drought during anthesis also adversely affect yield (Farooq et al., 2014; Mahrookashani et al., 2017). Thus, in years with low radiation, heat, or drought conditions, some

genotypes could avoid adverse conditions by an earlier or later phenological development.

Regarding the characterization of disease resistance of different genotypes, specific weather conditions are conducive to various wheat diseases during specific phenological stages (e.g., Ferrigo et al., 2016). Only a combined characterization of phenology and environmental conditions, that is, a thorough envirotyping, will permit a disentanglement of disease resistance from disease avoidance due to different phenological development. As a complement to standardized agronomic measurements, such as yield, baking quality, overwintering, plant height, 1000-kernel weight, or disease ratings (WBF, 2021), vegetation indices (VI), obtained from spectral measurements, are increasingly being used to estimate crop productivity. VIs have been shown to be best correlated with yield at specific phenological stages, usually shortly after flowering in the case of wheat. In this context, knowing the phenology is also critical for comparing VIs (e.g., Longchamps & Philpot, 2023; Naito et al., 2017; D. Wang et al., 2022). Such comparisons allow for a temporal normalization of spectral measurements, as the spectral signatures depend not only on genotypes but also on the phenological stages.

Similar to phenology, the senescence behavior of wheat has been shown to be a selection criterion for higher yielding genotypes (Hund et al., 2019). Genotypes showing a late onset of senescence followed by a rapid progression of senescence produce higher grain yields under water-limited conditions. The so-called stay-green behavior combines a prolonged photosynthetic activity with a rapid and efficient translocation of nutrients and sugars from other plant organs to the grain (Anderegg et al., 2020; Cao et al., 2021; Christopher et al., 2014, 2016). By contrast, maintaining a green canopy late into the growing season, but without a rapid and efficient translocation of sugars and nutrients, can be associated with a lower yield in the absence of water-limited conditions (Anderegg et al., 2020; Kipp et al., 2014). Knowing the end of senescence is important, as the varieties in breeding and variety testing trials do not mature or reach senescence at the same time but are typically all harvested on the same date. A genotype that has been senescent for, for example, 10 days before

harvest but remains in the field in humid conditions could yield grains of low quality and have higher loads of mycotoxins due to black head molds (Hershman, 2011; Lorenz, 1986; Poursafar et al., 2016). Furthermore, particularly when humid conditions occur in combination with cooler temperatures, the breaking of seed dormancy could lead to preharvest sprouting (Gao et al., 2013; Zhou et al., 2017), and consequently to the degradation of starch, lipids, and proteins in grains (Yan et al., 2023). Genotypes senescent for a longer period before harvest can also be prone to grain shedding (Aasen et al., 2020).

Finally, it is also important to know the maturity behavior of a genotype to plan optimal crop rotations (Montazeaud et al., 2016). For example, a wheat genotype with early maturity may allow for subsequent legume cover crop to develop more biomass (Blackshaw et al., 2010). In double-crop systems with wheat and soybean, as is widely used in the United States, a wheat harvested earlier can be followed by an earlier soybean sowing, increasing both the growth and yield of the latter (Parvej et al., 2020). For all these reasons, it is crucial not only to perform an adequate envirotyping but also to associate information on environmental conditions with the phenological characterization of genotypes (Costa-Neto et al., 2023; Elmerich et al., 2023) for a comprehensive view of differences in yield and quality.

Estimating the timing of the phenological and senescence stages visually in the field requires frequent field visits of experts during the period when these stages usually occur. This is very time-consuming and therefore expensive (Montazeaud et al., 2016; Velumani et al., 2020), especially as breeding and variety testing trials are usually conducted in several locations to account for genotype by environment ($G \times E$) interactions. These visual assessments also suffer from observation bias in case the assessments are done by different experts. To overcome the drawbacks of visual field ratings, methods are being developed to screen the progression of plant development in a more automated and objective manner. Adamsen et al. (1999) used a digital camera to describe wheat senescence 20 years ago. Sadeghi-Tehran et al. (2017) used digital images generated with a field scanner to detect wheat heading and flowering. Christopher et al. (2014, 2016) and Montazeaud et al. (2016) used a hand-held Greenseeker to measure the normalized difference vegetation index (NDVI) and applied dynamic models to describe the stay-green properties of wheat, such as delayed onset of senescence and an accelerated senescence rate. In the study by Anderegg et al. (2020), the senescence dynamics of more than 300 winter wheat varieties were tracked with a radio spectrometer. The authors stated that compared with spectral tracking, visual assessment remains the gold standard method, as it showed a closer correlation with yield than the VIs derived with the spectral methods, but the latter offer the potential for upscaling to very large breeding trials, where visual ratings are no longer feasible.

Core Ideas

- PhenoCams tracked phenology and senescence in a wheat variety testing trial for three consecutive seasons.
- Phenology from heading onward and plant/flag leaf senescence was predicted with high accuracy ($r > 0.8$).
- The PLSR prediction was based on multiple temporal features of vegetation indices (VI) instead of VI values.
- Drone based approaches are slightly more accurate, but PhenoCams are cheaper, as less field visits are needed.
- PhenoCams are cost-effective lean-phenotyping method for phenology and senescence in multi-environment trials.

All the methods mentioned so far still require measuring each plot individually and on several dates. To measure multiple plots at once, Burkart et al. (2018) extracted simple VI dynamics from single images taken with a drone 100m above a barley field throughout the growing season and then compared these dynamics with the timing of the phenological stages. Cao et al. (2021) compared the ability of more expensive drone-based multispectral cameras with cheaper drone-based red-green-blue (RGB) cameras to track senescence and stay-green. They concluded that however multispectral sensors allow for a more accurate characterization of senescence parameters (e.g., early, middle, and late senescence and senescence rate), cheaper RGB sensors also allow for tracking senescence behavior. Although these approaches showed promising results, they all still need frequent field visits. In addition, radio spectrometers and multispectral drone-based sensors are expensive, and for drone-based approaches, the images have to be processed with specific photogrammetric software.

The need for frequent visits to the study site can be overcome by monitoring plants with fixed-position cameras that take images at a high frequency (typically several times a day). Such fixed-position systems were applied to derive information on dynamic traits of plants (e.g., timing of phenological stages); yet, most of these PhenoCam studies focused on forests, ecology, or ecophysiology of larger systems. Typically, these studies are based on camera platforms installed above tree canopies (e.g., Ahrends et al., 2009; Keenan et al., 2014; Richardson et al., 2007, 2009), inside canopies (e.g., Kurc & Benton, 2010), or opportunistically profit from webcams pointing at relevant vegetation (e.g., Graham et al., 2010; Ide and Oguma, 2010). Richardson et al.

(2013, 2018) based their work on the PhenoCam Network (<https://phenocam.nau.edu/>).

In agriculture, there is vast amount of research describing protocols for obtaining information on crop state (e.g., Adamsen et al., 1999; Hunt et al., 2013), morphology (e.g., Hasan et al., 2019), and performance (e.g., Jensen et al., 2007; Gracia-Romero et al., 2017; Yue et al., 2019; H. Wang et al., 2020) based on digital images derived from different sources. The use of fixed-position digital repeat PhenoCams is also gaining interest in agriculture. Naito et al. (2017) used PhenoCams mounted on masts at 8m above rice fields, which combined RGB and NDVI images and took images daily from the late vegetative stage to the dough stage. They estimated traits related to rice yield, such as shoot biomass and grain weight under different nitrogen treatments. Bhatti et al. (2024) installed an NDVI PhenoCam at 6m above ground for gap filling of satellite-based NDVI time series. Thereby, they improved the classification of crops on satellite maps.

A limited number of studies have applied PhenoCams to track phenology in agricultural experiments. Most of these phenology studies featured one genotype of one species per image, and the literature on the application of digital repeat photography in the context of variety characterization is sparse (Aasen et al., 2020). Taylor and Browning (2021) used opportunistic images of the PhenoCam network to estimate different phenological stages of corn, wheat/barley, soybean, and alfalfa. Guo et al. (2022) tracked maize phenology in RGB images from masts of different heights. Liu et al. (2022) installed RGB timelapse cameras on sticks 1.5 m above the canopy to estimate the effects of cropping systems on crop phenology. On wheat, Zhu et al. (2016) established a fixed-position digital repeat imaging workflow on three varieties sown in three environments. In their approach, cameras were installed 5 m above the ground, and high-resolution images were analyzed with computer vision algorithms to detect ears upon emergence. Velumani et al. (2020) installed 47 fixed-position cameras in four different environments. With each camera covering only a relatively small area of one single variety, the authors generated high-resolution images that were analyzed with deep learning algorithms to detect heading and flowering. Brocks et al. (2016) and Brocks and Bareth (2018) mounted a pair of two RGB PhenoCams on a platform at 10 m above the ground and applied stereo vision to create three-dimensional surface models to estimate above-ground biomass on nine barley cultivars but not to track plant development.

The platforms mentioned so far were permanently installed or limited in height to several meters above ground (~ 6 m) when a more mobile setup was used. Field trials, especially in the context of multilocation trials, are usually conducted in different fields in subsequent seasons to allow for adequate crop rotation. Therefore, a PhenoCam should be ideally mounted on masts that allow for an easy set up and dismount-

ing. The masts should also be relatively high to cover a larger area with multiple genotypes at once. To our knowledge, only Aasen et al. (2020) used PhenoCams to describe the timing of the phenological stages of multiple genotypes of a crop that all appeared on the same image instead of individual images per genotype. However, they used cable-connected cameras on permanent masts. Thus, PhenoCams have been applied to track crop status, phenology, and senescence but not in setups with day-to-day applicability for variety testing. There are promising drone-based approaches, but they still need field visits for every measurement, which is time-consuming and can complicate seasonal measurements logistics.

Scenarios with more distance between the camera and the plot call for cameras with better spatial resolution. In recent years, there has been significant improvement in digital imaging technology, with increases in the signal-to-noise ratio and spatial resolution of cameras. Moreover, solar-powered cameras are now available, and the data storage capacity has increased. Thus, there are commercially available compact and fully autonomous time-lapse cameras that can work for weeks up to months without any intervention. Such cameras can be used for digital repeat imagery, mounted on easily dismountable, yet relatively high masts, and combined with the application of bespoke image analysis protocols, thus allowing for deriving information on plants with a high temporal resolution at relatively low hardware costs and without the need for frequent field visits. Such a PhenoCam setup can therefore offer breeders and examination offices opportunities for continuous information on the dynamics of crop growth and senescence with high temporal resolution and precision (Aasen et al., 2020).

Previous studies have often focused on one VI at a time to track plant development and senescence (e.g., Anderegg et al., 2020; Cao et al., 2021). For PhenoCam examinations, images were typically taken throughout the growing season at high temporal resolution, that is, daily or several times per day (Aasen et al., 2020). The greenness, or in general, the dynamic of the color changes of the plant canopy, is then tracked with the help of VIs (Hufkens et al., 2016; Richardson et al., 2009), such as the green chromatic coordinate (GCC), and the VI dynamics are analyzed (e.g., Ide and Oguma, 2010; Migliavacca et al., 2011; Browning et al., 2017). The progression of phenology and senescence usually leads to different changes in plants at different stages. For example, chlorophyll breakdown may be difficult to visually detect at the beginning of senescence, where 50% of leaf chlorophyll can be lost before visual yellowing and chlorosis (Chapman et al., 2021). At the same time, chlorophyll breakdown is a very dominant visual feature during later stages of senescence. It may be well tracked with visible band VIs, such as GCC, as well as with multispectral VIs, such as NDVI, which are both related to chlorophyll absorbance. Anthocyanins, carotenoids, and sometimes colorless chlorophyll breakdown products (Fischer

& Feller, 1994; Hörtensteiner, 2006) or changes in water content can lead to a change of the spectral signature beyond changes in greenness. Therefore, VIs that focus not only on greenness and chlorophyll, but also on changes in other spectral bands in the visible and non-visible spectrum, could confer complementary information on the development of the plant (Anderegg et al., 2020; Cao et al., 2021). As an example of a VI that reflects the dynamics of the breakdown of two pigment types, the plant senescence reflectance index (PSRI) uses chlorophyll/carotenoid ratio (Anderegg et al., 2020; Merzlyak et al., 1999).

VIs can also be combined. For example, Anderegg et al. (2020) used a random forest regression based on multiple VI-derived senescence dynamics parameters to predict yield and grain protein content. Guo et al. (2022) integrated textural and spectral dynamics of RGB images into a single analysis to track the phenology of maize. Longchamps and Philpot (2023) applied pairs of normalized difference metrics of two VIs, one primarily related to chlorophyll concentration and one primarily related to water content, to track the phenology of corn and soybean. VIs are often highly correlated, and random forests are prone to overfitting in such cases (e.g., Gregorutti et al., 2016). Therefore, as an alternative approach to combine multiple correlated VI dynamics at once, partial least squares regression (PLSR) was used in this study instead of random forest regression.

The aim of this study is to establish and evaluate a PhenoCam-based lean phenotyping workflow to monitor wheat phenology and senescence. Overall, the hypothesis tested is that such a workflow can offer superior performance compared to conventional approaches in variety testing. Specifically, (I) we introduce a non-stationary mast-based PhenoCam setup that is suitable for high-throughput field phenotyping of phenology and senescence of entire wheat variety testing trials. (II) We suggest a method for preprocessing and analyzing multiple VI dynamics at once, thus supporting the prediction of the timing of different stages of plant development. (III) We compare the PhenoCam method with different types of visual reference field ratings and VIs based on drone images (RGB and multispectral). (IV) The cost effectiveness of the different methods is analyzed using a simple economic calculation example.

2 | MATERIALS AND METHODS

This study was carried out on a wheat variety testing trial over three consecutive seasons (2020–2021, 2021–2022, 2022–2023). During this period, the fields were observed with PhenoCams, and visual field ratings were collected as reference measurements. As technical benchmark methods, additional drone flights were conducted. The methods were

compared to each other in terms of performance and cost. A conceptual overview of the study is provided in Figure 1.

2.1 | Field experiments

The winter wheat variety testing experiment (Figure 2a) was sown at Agroscope's agricultural research station, Changins, Switzerland (46°23'55.4" N, 6° 14' 20.4" E, 425 m.a.s.l., the World Geodetic System [WGS] 84). The soil of the experimental site is a shallow Calcaric Cambisol (Baxter, 2007; de Cárcer et al., 2019). The trial consisted of 30 modern registered European winter wheat varieties and is further referred to as the EuVar trial. The same varieties were sown in three different treatment regimes for the three seasons. In the "maximal" regimen, one growth regulator and one fungicide treatment were applied. In the "medium" regimen, there was only the growth regulator application and not the fungicide application. In the "minimal" regimen, neither a growth regulator nor a fungicide was applied. Tables S1 and S2 give a detailed overview of the different treatments. Fertilizers and herbicides were applied in three splits and at equal rates to all treatments according to the Proof of Ecological Performance certification guidelines (Swiss Federal Council, 2013), which represent a minimal standard of good practice for agriculture in Switzerland. Each variety-treatment combination was repeated on three plots. Within single plots, a wheat genotype was sown in eight rows, with a spacing of 15cm between them, resulting in an observable canopy of about 1.25 m × 6.7 m each. Within blocks of 3 by 10 plots, the genotypes were randomly distributed, and these blocks were randomly nested within three treatment replicates. Each replicate contained three blocks, and each block was treated with one of the three treatments. The 270 plots of the experiment spanned 27 rows (which followed the tractor track direction) and 10 columns (Figure S1). In total, the experiments were 79 m long (in tractor track direction) and about 55 m wide. The first two seasons of this experiment were first described by Treier et al. (2024), but the main characteristics are also described here for clarity.

2.2 | PhenoCam setup

Tikee PRO 2 / 2+ (Enlaps SAS) solar powered autonomous time lapse camera systems (Figure 2b) were installed in the field on TekMast VMS-21-M mobile field masts (Teksam Company NV) at 12 m aboveground (Figure 2a). The masts were stabilized with ropes from three sides and the anchor pins were reinforced with ground screws (Figure S2). Each camera system carried two cameras with CMOS RGB sensors (4608 × 3456 pixels), which had a fixed horizontal angle of 90° between them. According to full width at half maximum

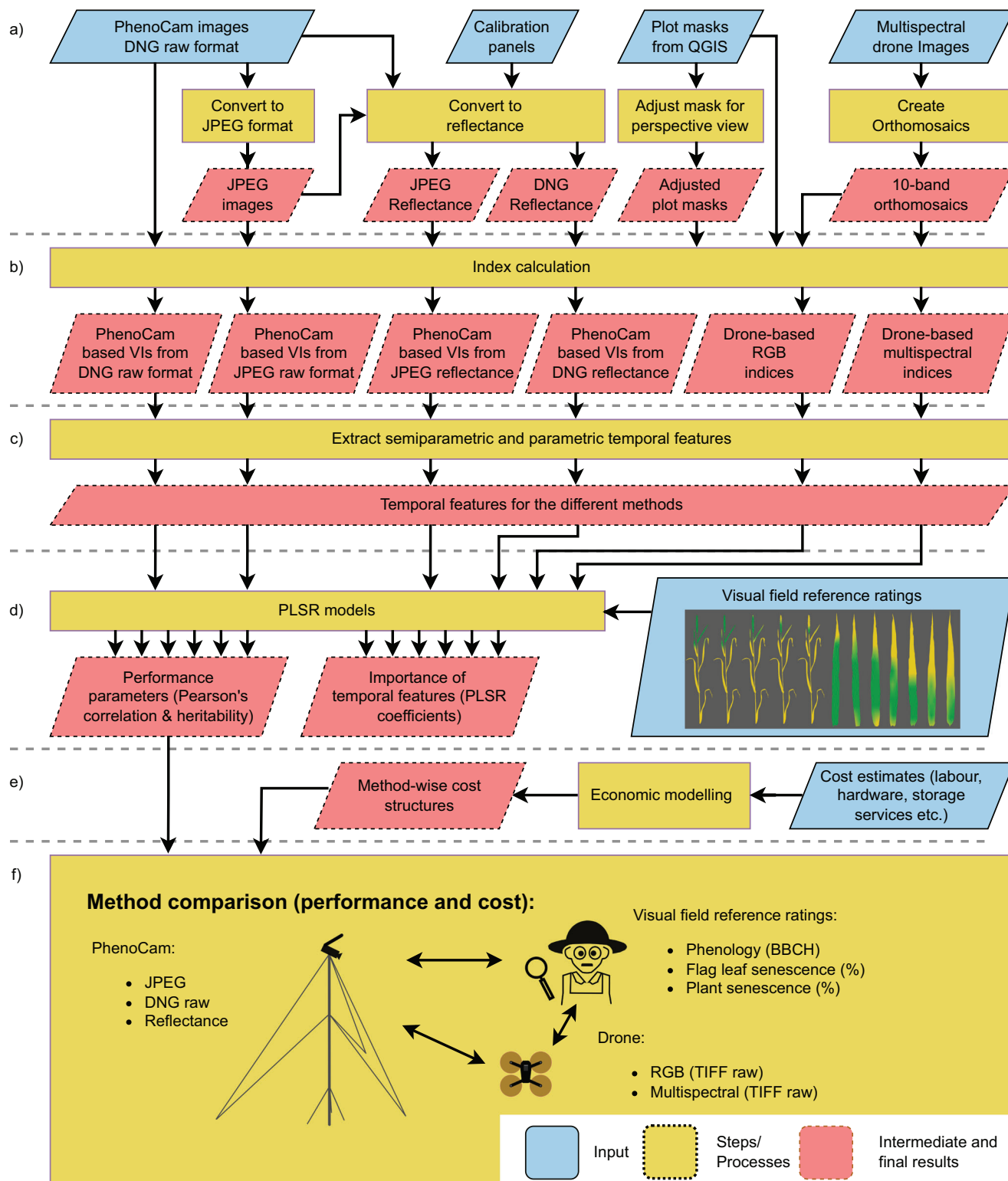


FIGURE 1 Overview of the workflow of the study. (a) Image data were acquired with PhenoCams and converted to different image and data types. Drone images were aligned into orthomosaics. For PhenoCams and drone data, plot masks were created. (b) From all plots on all images, various vegetation indices (VIs) were calculated. (c) From VI values, semiparametric and parametric temporal features were derived. (d) Visual field reference ratings were carried out for three types of ratings (phenology, flag leaf senescence, and plant senescence). VI based temporal features were then used to predict the timing of visual ratings in partial least squares regression (PLSR) models. These models also allowed to determine the most relevant temporal features for PLSR prediction. (e) To also compare the cost structure of the different methods, conceptual economic modeling was carried out. (f) The methods were compared to each other in terms of performance and cost.

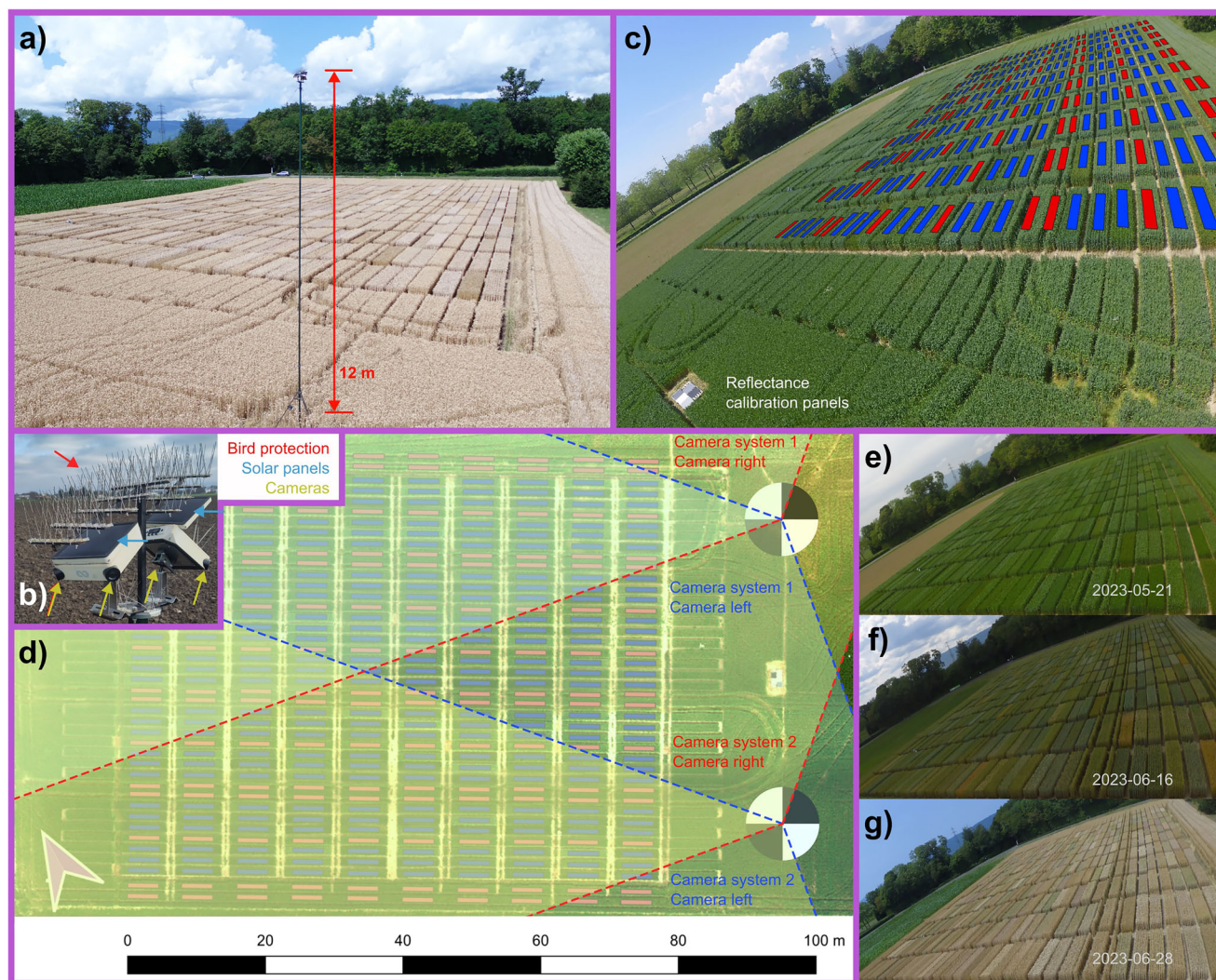


FIGURE 2 Overview of the PhenoCam setup and data. Masts (a) were installed in the field carrying camera systems (b) to create at least five images a day (c and e–g). (d) Two masts with two camera systems pointing toward the experiment of interest were installed at the narrow side of the experiment, partially covering the same plots from different angles as indicated by the opaque areas. In the back of (d), an orthomosaic is shown as created for every drone flight. Plot masks were created and adjusted for perspective view for PhenoCam images (c) or adjusted to field plots for drone-based orthomosaics (d). The colors of the masks in (c) and (d) indicate whether a plot was part of the experiment (blue) or a border plot to separate different treatments (red).

specifications, the spectral sensitivity of the sensors was highest from 430 nm to 500 nm for the blue spectral band (B), from 475 nm to 600 nm for the green band (G) and from 580 nm to 660 nm for the red band (R). The two cameras of one system together covered an angle of 220° horizontally and overlapped for central parts of these 220° regions (Figure 2d). We had two masts with two camera systems on each mast (eight cameras in total), which were set up on the narrow side of the experiment at a distance of 30 m (in 2022) to 45 m (in 2021 and 2023) from each other, but only four cameras covered different parts of the EuVar experiments. These four cameras were oriented from north-east to south-west. With a vertical opening angle of 90°, the cameras were installed obliquely pointing toward the ground, covering at least the region from the base

of the masts to the edge of the fields in the direction of the horizon.

The masts were installed in the field shortly after sowing and uninstalled when the wheat was mature, 1 day before harvest or after harvest. The cameras were programmed to take images every 2 h in the period from 7:00 a.m. to 5:00 p.m. every day to cover the period of daylight from spring onward. Images were saved in DNG (Digital Negative) raw format on SD (Secure Digital) cards plugged into the cameras. As the Tikee PRO 2 camera model allowed for a maximal memory size of 128 GB, the cards had to be replaced once during the duration of the experiment, for which the masts had to be lowered. This was done about 2 weeks before the expected heading date.

Images were saved in 16-bit DNG raw format. This format is data-heavy (34 MB/image), and to test whether the lighter 8-bit JPEG (Joint Photographic Experts Group) format (15.9 MB/image) also allowed for similar quality, DNG images were transformed to JPEG format in Python.

2.3 | Multispectral measurements

In parallel to mast recordings, the trials were also monitored with an airborne MicaSense RedEdge-MX Dual multispectral camera (MicaSense Inc.). The camera was carried by a DJI Inspire 2 drone (SZ DJI Technology Co. Ltd.). The flight height was 60 m in 2021 and 40 m in 2022 and 2023, resulting in a ground sampling distance of 3.98 and 2.71 cm, respectively. The side overlap was set to 80%, the flight speed was limited to 5 m s⁻¹ and an image was taken at an interval of 2 s in 2021 and 1 s in 2022 and 2023, resulting in a front overlap of approximately 80% for the two flight configurations. Images of a calibrated MicaSense reflectance panel were taken at the beginning and the end of each flight. Flights were conducted throughout the growing seasons (Figure 3). From shortly before heading (BBCH 59; Lancashire et al., 1991) to the end of senescence, the flights were flown at higher temporal intervals of weekly to several times a week. The images were saved in a raw TIFF image format.

Agisoft Metashape Professional software (Agisoft LLC) was used to align images to generate 10 band orthomosaics (Figure 2d) that covered the whole experiment. Details on the spectral properties of the 10 bands of the sensor are described in Table S3. The reflectance panel used for calibration featured a QR code, and Agisoft provided the functionality to detect this code and conduct a calibration of the targets autonomously.

2.4 | Mask creation

To define regions of interest on images, plot masks (Figure 2c) were created for each plot appearing on each camera in each year. First, orthogonal masks were created based on a CSV file, specifying row and column position of the plot and corresponding meta information (e.g., genotypes, treatments, etc.) of the plots. This was achieved with a Python 3.8 script (van Rossum & Drake, 2009), using the “ofgr” module of the “GDAL” library (GDAL/OGR Contributors, 2024) and defining approximate plot dimensions in the image coordinate system (that is, pixel coordinates) directly in the script. Then, a homography transform was applied to the shape coordinates to achieve a perspective view. The homography matrix was estimated based on four corresponding points between orthog-

onal masks and perspective images, using a Python script provided by Socretquiliqaa Lee (<https://gist.github.com/Socret360/bcefb0f95cfc20800ea3409f40b8bb58>). The transformed coordinates were calculated as the dot product of the orthogonal shape coordinates with the homography matrix. The masks were then manually adjusted in QGIS (QGIS Development Team, 2022) to match a base image. To account for border effects in the field and for inaccuracies of referencing and superimposition of different images, buffers were applied to masks.

As a consequence of perspective, the masks had very different sizes, and shape buffers between masks were adjusted based on individual visual judgment, resulting in shapes corresponding to approximately 50% of the surface of the plots. The masts shook slightly in the wind, resulting in differences in the position and orientation of the cameras. Thus, the masks needed adjustment over time. To that end, well-illuminated reference images were selected throughout the growing season, which were taken between 10:30 a.m. and 12:30 p.m. The masks were then manually adjusted for these reference images in QGIS based on the masks from the base image, saved to GeoJSON format again, and used as reference masks for the respective reference image and all subsequent images until a new reference image was available. Georeferenced masks for drone data analysis (Figure 2d) were created similarly as masks for PhenoCams, but in the Swiss CH1903+/LV95—EPSG:2056 coordinate reference system and without homography transform. Border buffers of 25 cm and up to 1 m were left on plot width and length, respectively.

2.5 | Index calculation and index value extraction

A large number of color VIs were proposed for different applications in agronomy and were compiled and compared in various publications (e.g., Anderegg et al., 2020, 2023; Cao et al., 2021; Hasan et al., 2019; Hunt et al., 2013; Li et al., 2023; D. Wang et al., 2022). In this study, a multitude of described VIs were calculated from color bands (Figure 1b) in the visual RGB color space (Table 1) and for multispectral drone images also from the infrared range (Table 2). Each pixel within each plot featured a value for each VI. To aggregate these values to single values per plot and VI, zonal statistics were applied by calculating the mean, the 50th percentile (or median), the 90th, the 98th, and the 99th percentiles of the pixel values within the individual masks. Higher percentiles were included, as for most VIs, higher values are associated with plants, and choosing a high percentile can help avoid exposed soil within plots that affects VI values (Deery et al., 2016). The 98th and the 99th percentiles are close to maximal values (i.e., the 100th percentile) but not as sensitive to artifacts and disturbances.

TABLE 1 Red-green-blue (RGB) vegetation indices (VI). For some VIs, the literature provides multiple, sometimes significantly different, formulas. For excess red index (ExR), excess green minus excess red (ExGR), and triangular greenness index (TGI), this is indicated as subscript at the end of the index names. The numbers behind the band names in the formulas indicate the wavelengths of the spectral bands used.

Index	Full name	Formula	Reference
ExB	Excess blue index	$\frac{(1.4 \cdot \text{Blue}_{444}) - \text{Green}_{531}}{\text{Red}_{650} + \text{Green}_{531} + \text{Blue}_{444}}$	Lu et al. (2019); Mao et al. (2003); Xu et al. (2022)
ExG	Excess green index	$(2 \cdot \text{Green}_{531}) - \text{Red}_{650} - \text{Blue}_{444}$	Woebbecke et al. (1995)
ExR _b	Excess red index	$(1.4 \cdot \text{Red}_{650}) - \text{Blue}_{444}$	Meyer et al. (1998)
ExR _g	Excess red index	$(1.4 \cdot \text{Red}_{650}) - \text{Green}_{531}$	Cao et al. (2021); Lu et al. (2019); Meyer et al. (1998); Xu et al. (2022); Zhang et al. (2021)
ExR _{g-norm}	Normalized excess red index	$\frac{(1.4 \cdot \text{Red}_{650}) - \text{Green}_{531}}{\text{Red}_{650} + \text{Green}_{531} + \text{Blue}_{444}}$	Lu et al. (2019)
ExGR _{Meyer}	Excess green minus excess red	$\text{ExG} - \text{ExR}_b = (2 \cdot \text{Green}_{531}) - (2.4 \cdot \text{Red}_{650})$	Meyer and Neto (2008)
ExGR _{Zhang}	Excess green minus excess red	$\text{ExG} - \text{ExR}_g = (3 \cdot \text{Green}_{531}) - (2.4 \cdot \text{Red}_{650}) - \text{Blue}_{444}$	Zhang et al. (2021)
ExGR _{Lu}	Excess green minus excess red	$\text{ExG} - \text{ExR}_{g-norm} = (2 \cdot \text{Green}_{531}) - \text{Red}_{650} - \text{Blue}_{444} - \frac{(1.4 \cdot \text{Blue}_{444}) - \text{Green}_{531}}{\text{Red}_{650} + \text{Green}_{531} + \text{Blue}_{444}}$	Lu et al. (2019)
G _B Ratio	Green-blue ratio index	$\frac{\text{Green}_{531}}{\text{Blue}_{444}}$	Sellaro et al. (2010)
GLI	Green leaf ratio	$\frac{(2 \cdot \text{Green}_{531}) - \text{Red}_{650} - \text{Blue}_{444}}{(2 \cdot \text{Green}_{531}) + \text{Red}_{650} + \text{Blue}_{444}}$	Louhaichi et al. (2001)
G _R Ratio	Green-red ratio index	$\frac{\text{Green}_{531}}{\text{Red}_{650}}$	Tucker (1979)
IKAW	Kawashima index	$\frac{\text{Red}_{650} - \text{Blue}_{444}}{\text{Red}_{650} + \text{Blue}_{444}}$	Kawashima (1998)
MGRVI	Modified green-red vegetation index	$\frac{(\text{Green}_{531})^2 - (\text{Red}_{650})^2}{(\text{Green}_{531})^2 + (\text{Red}_{650})^2}$	Bendig et al. (2015)
MNVI	Meyer–Neto vegetation index	$(2 \cdot \text{Green}_{531}) - (2 \cdot \text{Blue}_{444}) - (2.4 \cdot \text{Red}_{650})$	Jin et al. (2017); Meyer and Neto (2008)
NGBDI	Normalized green-blue difference index	$\frac{\text{Green}_{531} - \text{Blue}_{444}}{\text{Green}_{531} + \text{Blue}_{444}}$	Meyer and Neto (2008); Xu et al. (2022)
NGRDI	Normalized green-red difference index	$\frac{\text{Green}_{531} - \text{Red}_{650}}{\text{Green}_{531} + \text{Red}_{650}}$	Meyer and Neto (2008)
RBRI	Red-blue ratio index	$\frac{\text{Red}_{650}}{\text{Blue}_{444}}$	Hasan et al. (2019); Segal (1982); Sellaro et al. (2010)
RGBVI	Red-green-blue vegetation index	$\frac{(\text{Green}_{531})^2 - (\text{Blue}_{444} \cdot \text{Red}_{650})}{(\text{Green}_{531})^2 + (\text{Blue}_{444} \cdot \text{Red}_{650})}$	Bendig et al. (2015)
TGI _{fixed}	Triangular greenness index (simplified)	$\text{Green}_{531} - (0.39 \cdot \text{Red}_{650}) - (0.61 \cdot \text{Blue}_{444})$	Kavaliauskas et al. (2023); Hunt et al. (2011) Segarra et al. (2023)
TGI	Triangular greenness index	$-0.5 \cdot [(\lambda_{650} - \lambda_{444})(\text{Red}_{650} - \text{Green}_{531}) - (\lambda_{650} - \lambda_{531})(\text{Red}_{650} - \text{Blue}_{444})]$	Hunt et al. (2011)
VARI	Visible atmospherically resistant index	$\frac{\text{Green}_{531} - \text{Red}_{650}}{\text{Green}_{531} + \text{Red}_{650} - \text{Blue}_{444}}$	Gitelson et al. (2002)
RCC	Red chromatic coordinate	$\frac{\text{Red}_{650}}{\text{Red}_{650} + \text{Green}_{531} + \text{Blue}_{444}}$	Gillespie et al. (1987)
GCC	Green chromatic coordinate	$\frac{\text{Green}_{531}}{\text{Red}_{650} + \text{Green}_{531} + \text{Blue}_{444}}$	Gillespie et al. (1987)
BCC	{\mathrm{Blue}} chromatic coordinate	$\frac{\text{Blue}_{444}}{\text{Red}_{650} + \text{Green}_{531} + \text{Blue}_{444}}$	Gillespie et al. (1987)
R	Red	Red_{650}	–
G	Green	Green_{531}	–
B	Blue	Blue_{444}	–

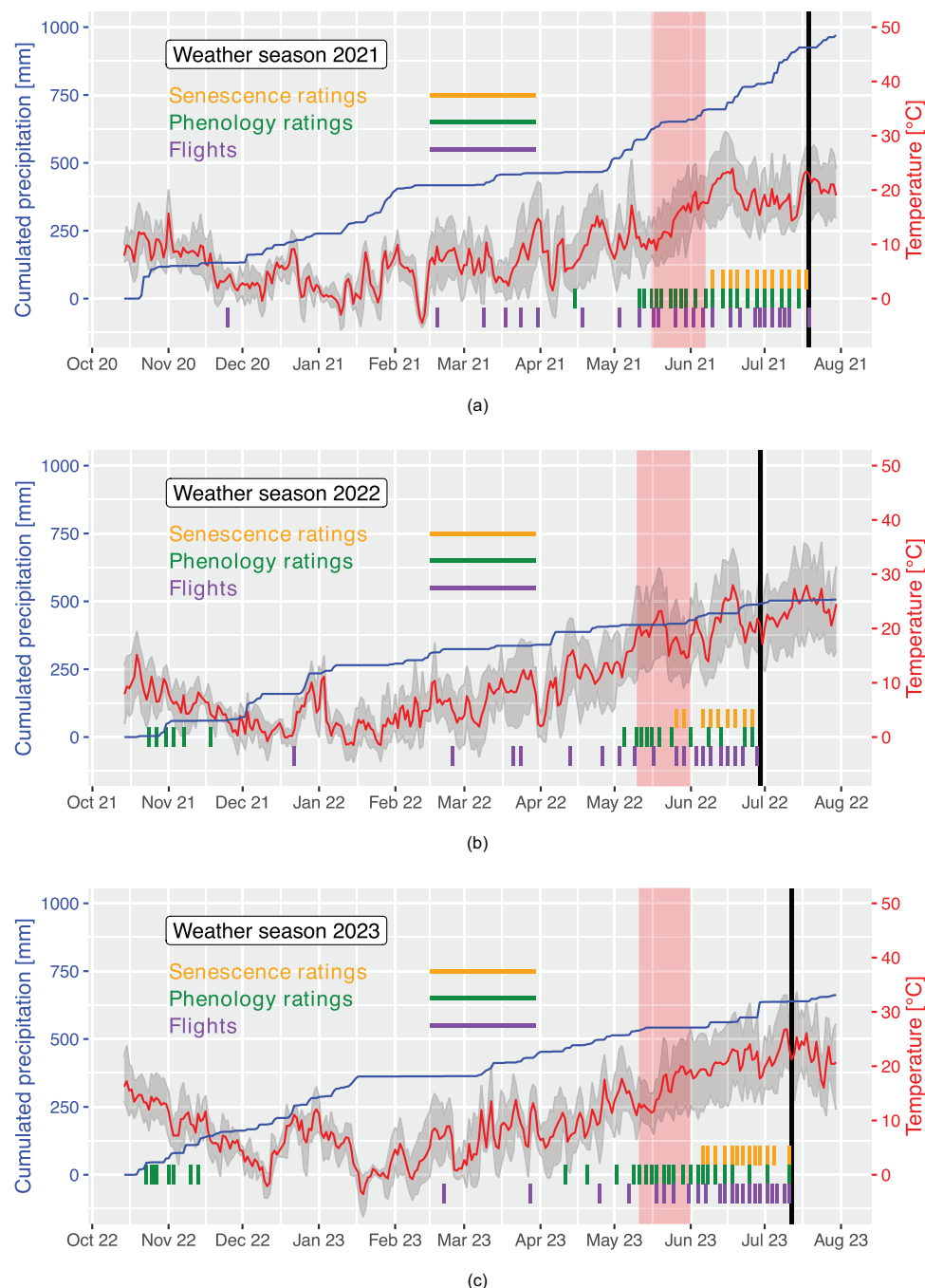


FIGURE 3 Overview of the measurements and meteorological conditions in the three experimental seasons. (a–c) General weather conditions during the growing seasons 2021–2023 from sowing until after harvest. Red shows the mean air temperature, and the shades indicate daily temperature minima and maxima. The vertical purple lines indicate the dates of flights, the green and orange lines the dates of phenology and senescence ratings, respectively. Cumulative precipitation is shown as a rising blue line. During the period shaded in red, heading was observed in the field. Harvest dates are marked by black lines.

2.6 | Extraction of temporal features

The values of these VIs often follow characteristic dynamics throughout the growing season (e.g., Figure 2e–g). For very early stages such as emergence (BBCH 09; Lancashire et al., 1991) or the three-leaf stage (BBCH 13), no corresponding temporal behavior of VIs, for example, a sudden

and pronounced increase in VI values, was observed based on visual inspection. Therefore, the analysis conducted in this study focused on the stages from heading (BBCH 59) to senescence. Heading occurs typically after mid-May or around 200 days after sowing (DAS) and to reduce the amount of data to be handled, as well as to simplify the automatic feature extraction, the data were limited to the relevant period.

TABLE 2 Multispectral vegetation indices. The numbers behind the band names in the formulas indicate the wavelengths of the spectral bands used.

Index	Full name	Formula	Reference
ARI1	Anthocyanin reflectance index	$\frac{1}{\text{Green}_{560}} - \frac{1}{\text{RedEdge}_{705}}$	Gitelson et al. (2001)
ARI2	Anthocyanin reflectance index	$\text{NIR}_{842} \cdot \left(\frac{1}{\text{Green}_{560}} - \frac{1}{\text{RedEdge}_{705}} \right)$	Gitelson et al. (2001)
DVI	Difference vegetation index	$\text{NIR}_{842} - \text{Red}_{668}$	Tucker (1979)
EVI	Enhanced vegetation index	$2.5 \cdot \frac{\text{NIR}_{842} - \text{Red}_{650}}{\text{NIR}_{842} + 6 \cdot \text{Red}_{650} - 7.5 \cdot \text{Blue}_{444} + 1}$	Huete et al. (2002)
NDRE	Normalized difference red edge index	$\frac{\text{NIR}_{842} - \text{RedEdge}_{717}}{\text{NIR}_{842} + \text{RedEdge}_{717}}$	Gitelson and Merzlyak (1994); Barnes et al. (2000); Tang et al. (2022)
NDVI	Normalized difference vegetation index	$\frac{\text{NIR}_{842} - \text{Red}_{668}}{\text{NIR}_{842} + \text{Red}_{668}}$	Rouse et al. (1974)
NDVI ₇₁₇	Normalized difference vegetation index	$\frac{\text{RedEdge}_{717} - \text{Red}_{668}}{\text{RedEdge}_{717} + \text{Red}_{668}}$	Rouse et al. (1974)
PSRI ₇₀₅	Plant senescence reflectance index	$\frac{\text{Red}_{650} - \text{Green}_{531}}{\text{RedEdge}_{705}}$	Merzlyak et al. (1999)
PSRI ₇₁₇	Plant senescence reflectance index	$\frac{\text{Red}_{668} - \text{Green}_{560}}{\text{RedEdge}_{717}}$	Merzlyak et al. (1999)
PSRI ₇₄₀	Plant senescence reflectance index	$\frac{\text{Red}_{650} - \text{Green}_{531}}{\text{RedEdge}_{740}}$	Merzlyak et al. (1999)
SAVI	Soil adjusted vegetation index	$1.5 \cdot \frac{\text{NIR}_{842} - \text{Red}_{650}}{\text{NIR}_{842} + \text{Red}_{650} + 0.5}$	Huete (1988)
SR	Simple ratio	$\frac{\text{NIR}_{842}}{\text{Red}_{668}}$	Birth and McVey (1968); Jordan (1969)

Thus, to extract semiparametric temporal features from VIs (Figure 1c), only VI values from 20 days before expected heading, that is, 180 DAS, up to 1 or 2 days before harvest were considered. To derive smooth VI dynamics, either the rolling mean, a Savitzky–Golay filter, spline smoothing, and locally estimated scatter plot smoothing (loess) were applied to the data (e.g., Bhatti et al., 2024; Guo et al., 2022; Hufkens et al., 2019; Klosterman et al., 2014). The maximal and minimal values of the smoothed dynamics from the four different smoothing types were defined as 100% and 0%, respectively, and two temporal features were extracted as the time in DAS when the value reached 10% and 2%, respectively, similar to Christopher et al. (2014), where 10% was defined as conclusion of senescence. An overview of the temporal feature types is presented in Table 3. For VIs with increasing values toward maturity and senescence, the VI values were reflected over the DAS axis (i.e., the x-axis) so that the maximum always appeared earlier than the minimum.

These semiparametric approaches allow capturing very dynamic seasons but do not imply growth dynamics and are more prone to overfitting if the measurement noise is very high (Roth et al., 2021). Thus, we also derived parametric temporal features from Gompertz models (Figure 1c; e.g., Anderegg et al., 2020; Chapman et al., 2021). First, for each measurement, the accumulated thermal time from sowing was calculated in growing degree days (GDD)

as:

$$T_{\text{therm},h} = \sum_{h=1}^n \begin{cases} \frac{T_{\text{max},h} + T_{\text{min},h}}{2 \cdot 24} - \frac{T_{\text{base}}}{24}, & \text{if } T_{\text{min},h} > T_{\text{base}}, \\ 0, & \text{if } T_{\text{min},h} \leq T_{\text{base}}, \end{cases} \quad (1)$$

where $T_{\text{max},h}$ and $T_{\text{min},h}$ are the maximum and minimum temperatures of the n th hour h after sowing. T_{base} was assumed to be 0°C (McMaster, 1997). Twenty-four hourly means sum up to the GDD of 1 day.

We then selected data for the relevant growth period by just considering data from 180 DAS to harvest. VI values were reflected over the DAS axis again where necessary, but this time, we ensured that the minimum appeared earlier than the maximum. The data were smoothened with a loess function to extract the minimum (LoessMin) and maximum (LoessMax) of the smoothened VI data. The original VI values that appeared before LoessMin were set to the LoessMin value, and the values appearing after LoessMax were set to the LoessMax value. These restricted VI values should ensure that the model captures the main slope and is not dampened by high VI values before LoessMin or a possible decrease in VI after LoessMax. In addition, the data were shifted in such a way that LoessMin was 0 and the restricted VI data started to increase around 0 GDD. This translation of the restricted VI data ensured that the value range was suitable for fitting a Gompertz model,

TABLE 3 Overview and description of temporal features. The temporal features were calculated for all vegetation indices (VIs) and data aggregation methods (i.e., mean and different percentiles).

Temporal feature	Description	Feature class
D1 _{LocMax_1}	first derivative of Gompertz, first local maximum	Gompertz derivative
D2 _{LocMax_1}	second derivative of Gompertz, first local maximum	
D2 _{LocMin_1}	second derivative of Gompertz, first local minimum	
D3 _{LocMax_1}	third derivative of Gompertz, first local maximum	
D3 _{LocMax_2}	third derivative of Gompertz, second local maximum	
D3 _{LocMin_1}	third derivative of Gompertz, first local minimum	
D4 _{LocMax_1}	fourth derivative of Gompertz, first local maximum	
D4 _{LocMin_1}	fourth derivative of Gompertz, first local minimum	
D4 _{LocMax_2}	fourth derivative of Gompertz, second local maximum	
D4 _{LocMin_2}	fourth derivative of Gompertz, second local minimum	
Loess _{Max}	Loess smoothed curve at maximum	Loess smoothing with threshold
Loess _{Min}	Loess smoothed curve at minimum	
Loess _{0.02}	Loess smoothed curve at 2% of max-min range	
Loess _{0.1}	Loess smoothed curve at 10% of max-min range	
Rolling _{0.02}	Rolling mean at 2% of max-min range	Rolling mean smoothing with threshold
Rolling _{0.1}	Rolling mean at 10% of max-min range	
Sav.Gol _{0.02}	Savitzky–Golay smoothed curve at 2% of max-min range	Savitzky–Golay smoothing with threshold
Sav.Gol _{0.1}	Savitzky–Golay smoothed curve at 10% of max-min range	
Spline _{0.02}	Smoothing spline at 2% of max-min range	Spline smoothing with threshold

(Continues)

TABLE 3 (Continued)

Temporal feature	Description	Feature class
Spline _{0.1}	Smoothing spline at 10% of max-min range	
Spline _{Max}	Smoothing spline at maximum	
Spline _{Min}	Smoothing spline at minimum	

Abbreviation: Loess, locally estimated scatter plot smoothing.

$$I = ae^{-be^{-ct}}, \quad (2)$$

with the package “nls.multstart” (Padfield & Matheso, 2020) in R (R Development Core Team, 2022). In the model, I represents the VI value at time t . a is the asymptote and was restricted to values from 0.9 to $1.1 \cdot \text{LoessMax}$. b is a location parameter that mainly affects the starting point of the curve. Parameter c has an impact on the slope and the starting point. The Gompertz parametrization allowed for a monotonously increasing dynamic, from which temporal features were derived by calculating the first four derivatives of the fitted Gompertz model. Local minima and maxima of the derivatives were determined and the timing of the local minima and the maxima as well as of LoessMin and LoessMax were extracted. The timing was then transformed from GDD back to DAS to use the same temporal unit as for the semiparametric method. These procedures were performed individually for the mean and the different percentiles used for aggregation of the VI values for each VI in each plot.

This procedure was also adopted for drone-based VI. VI values were smoothed with a penalized smoothing spline in the R package “pspline” (Ripley & Ramsey, 2024), with degrees of freedom set to two-thirds of the number of measurements and interpolated for single days. From daily VI values, semiparametric and parametric temporal features were then extracted as for PhenoCam data but without the different smoothing approaches.

2.7 | In-field calibration panels

Color VIs can be based on raw digital numbers (DN) of images, or reflectance values can be derived from DNs with calibration panels of a known reflectance. To this end, five calibration panels were installed in the field (Figure 2c) throughout the growing season (but only from May 26, 2021 for the first year). The calibration panels were detected on the individual images by a semi-autonomous pipeline. When all five panels were detected on an image, the DN values were transformed into reflectance values using the empirical line

method. Please refer to the Supporting Information Section S4 for more details on the installation of the calibration panels and their semi-autonomous detection.

In total, 14,051 images were taken, and for 3,929 images, calibration data were directly available. When calibration data were not directly available, they were obtained from other images based on the following criteria. For the 3,929 images of the second camera that had calibration data available for the first camera of the same camera system, the calibration equation was taken from the first camera. Images of the two cameras were taken with the same ISO setting and exposure time. For 2,592 images with no calibration available from the same camera system, the calibration equation of an image of another camera was used if taken within the same period (± 15 min), and with the same ISO setting and exposure time. For 3,601 images, no suitable calibration information was found, and no reflectance was calculated for those cases, reducing the temporal resolution of reflectance-based VI values to 81% compared to DN-based VI values.

2.8 | Visual field reference ratings

Three types of in-field reference ratings were conducted in parallel to PhenoCam observations and flights (Figure 1d) and recorded using the Field Book app (Rife & Poland, 2014): Phenology (“BBCH”), flag leaf senescence (“Sen-Leaf”), and plant senescence (“SenPlant”). Phenology was rated according to the BBCH scale (Lancashire et al., 1991). The rating interval was 2–4 days during the heading period and decreased to approximately weekly toward complete maturity. The senescence rating of the flag leaf was carried out according to the scale reported by Pask et al. (2012), with 0% corresponding to a fully green leaf and 100% to a fully senescent leaf (Figure 4a).

The plant senescence rating (Figure 4b) was inspired by the plot senescence of Anderegg et al. (2020) and the peduncle and ear senescence rating provided by Chapman et al. (2021). The plot rows were opened manually, and the senescence of the whole plant was rated from 0% (fully green plant) to 100% (completely senescent plant). Field ratings were performed mostly in 5% steps except for very late ratings from 95% onward. The last 5% were rated in smaller steps and mainly related to changes around the ear bases and peduncles. Rating values of 100% were only achieved when the ears base as well as the peduncles were completely senescent.

The single rating events for phenology and both types of senescence are visualized in Figure 3. All visual ratings were in DAS. To allow for daily resolution, the values were smoothed with a penalized smoothing spline in the “pspline” package (Ripley & Ramsey, 2024), with degrees of freedom set to two-thirds of the number of measurements, and interpolated for single days. The rating of a specific reference level

was given the DAS value of the day on which the interpolated value exceeded a specific reference level for the first time.

2.9 | PLSR models to predict phenology and senescence from temporal features

As the temporal features used as predictors were expected to be highly correlated, PLSR was preferred over other approaches such as random forest regression. Random forests are prone to overfitting when using highly correlated data (e.g., Gregorutti et al., 2016), and PLSR was shown to produce more generalizable predictions than random forests (e.g., Lee et al., 2017). Recursive feature elimination (RFE) can be applied to increase generalizability of random forests and decrease the risk of overfitting (Darst et al., 2018; Gregorutti et al., 2016). However, computational capacity requirements for PLSR are moderate, but repeated RFE can be very computationally intensive.

2.9.1 | Feature selection

The temporal features extracted previously for each plot were used as input data in a PLSR analysis (Figure 1d) with the R package “PLS” (Mevik & Wehrens, 2007). Given that many VIs were tested based on multiple aggregation percentiles and the mean aggregation, the number of features as predictor variables in PLSR initially exceeded the number of observations. Although PLSR analysis is suitable for this situation (Carrascal et al., 2009), the number of temporal features was reduced based on the magnitudes of the relative PLSR coefficients $\beta_{\text{rel},i}$, which were calculated for each temporal feature type as:

$$\beta_{\text{rel},i} = \frac{|\beta_i|}{\sum_{i=1}^n |\beta_i|}, \quad (3)$$

where β_i denotes the PLSR coefficient of the i th of n temporal features. PLSR started with the full set of temporal features available for all plots. The features with the lowest $\beta_{\text{rel},i}$ were skipped in a backward feature elimination until the most predictive features were left in the model (similar to methods summarized in Mehmood et al., 2012). At the beginning, 200 temporal features were dismissed at each iteration. With the decrease in features, the number of dismissed features continuously decreased. When for example 345 features were left, five temporal features were dismissed at each iteration, only two features were dismissed when 280 features were left, and only one feature was eliminated for the last iterations, from 100 features down to 15 features. The PLSR used 10 components and 10 segments for cross-validation. With each model, the different levels of the reference rating types

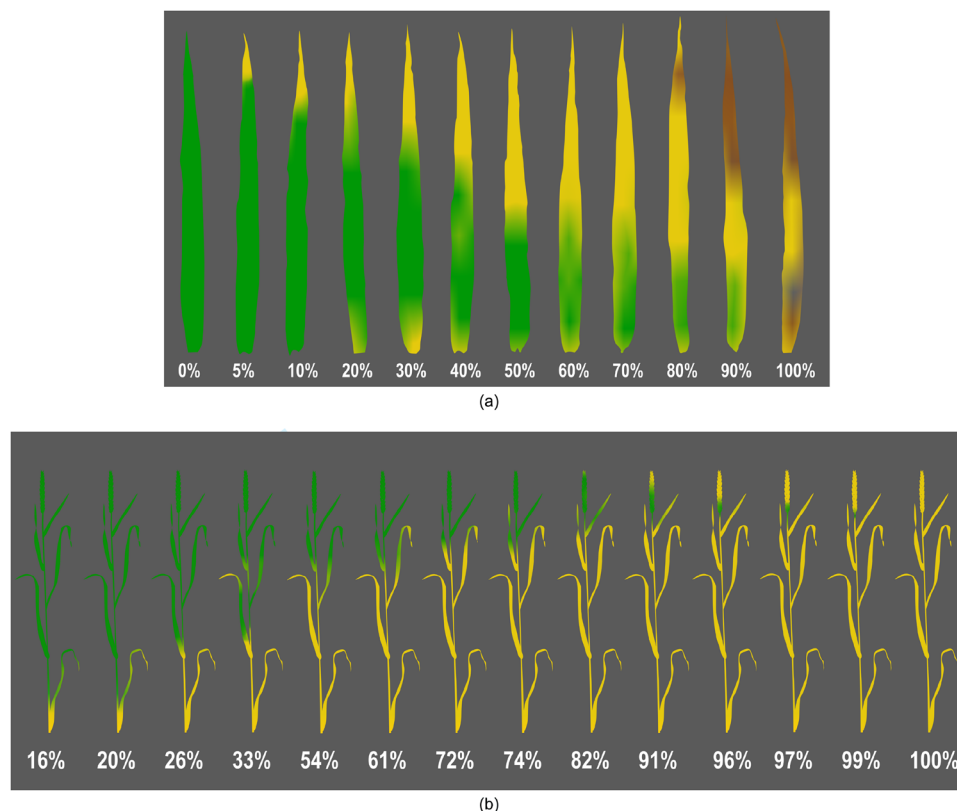


FIGURE 4 Senescence rating scales. (a) Flag leaf senescence according to Joshi et al. (2007) and Pask et al. (2012). (b) Plant senescence scale. The percentage depicts senescent proportion of all pixels. Figure in (a) was inspired by an image of the John Innes Centre and the University of Nottingham.

were predicted using the full set of observations and were correlated with the reference rating. With this procedure, it was determined that the correlations were relatively stable above 200 temporal features but started to weaken below, and the number of temporal features was set to 200 for the next step (Figure S3).

2.9.2 | Hundred times repeated cross-validation

For each level of each type of reference measurement (BBCH, SenLeaf, SenPlant), in the previous step, a set of 200 temporal features was selected. This reduction in features allowed for a computationally efficient cross-validation. For each reference level, the data were now randomly split by two split approaches. For the first approach, the PLSR models were trained with 75% of randomly chosen observations and tested with the remaining 25% of the observations. For the second approach, PLSR models were trained with the observations of 19 randomly chosen genotypes and tested on the observations of the remaining 11 genotypes. The maximum number of components of the PLSR model was set at 15, and the optimal number of components in the range of 1–15 was selected for each model individually with the “selectNcom” function. As

model accuracy metric, Pearson’s correlation between predictions and reference measurements was used. This procedure was repeated 100 times for each reference level, with the random split into training and validation data repeated each time. To characterize the quality of the PLSR models, the mean correlation coefficients and standard deviation of the coefficients within 100 repetitions were used as well as genotype-wise differences from visual reference values.

2.9.3 | Importance estimation of VIs and feature types

Temporal features were extracted on the basis of multiple semiparametric and parametric methods to capture dynamics, further named feature types. In addition, temporal features were based on various zonal statistics (mean and multiple percentiles) for pixel-value aggregation, and on various VIs. To estimate the importance of the different feature types, aggregation methods and VIs, the sums of magnitudes of the relative PLSR coefficients were calculated within three reference classes. The classes were generically defined for the three reference types as early, when reference levels, that is, values that the different reference types could assume,

ranged from 10 to 25, intermediate for values 40 to 70, and late for values 85 to 100. To achieve robust importance estimates, coefficients were summed up for the different reference values of the reference classes for all iterations of cross-validation and the different groups of comparison. Comparison groups were either feature types, aggregation methods, or VIs. $\beta_{\text{rel,sum,ref.class,group}}$ was the sum of the 100 relative PLSR coefficients $\beta_{\text{rel},i}$ of the 100 repetitions k of cross-validation (only 75%/25% train/test data split) for all reference levels j within one reference class, and one comparison group,

$$\beta_{\text{rel,sum,ref.class,group}} = \sum_{j=1}^n \sum_{k=1}^{100} \sum_{i=1}^{200} \beta_{\text{rel},i,j;k}, \quad (4)$$

where $\beta_{\text{rel},i,j;k}$ was the relative PLSR coefficient of the i th out of 200 temporal features, of the k th out of 100 repetitions of the j th reference level within a reference class. Finally, coefficient sums were normalized to the range from 0 to 1.

2.10 | Heritability and genotypic ranking consistency of predicted values

In addition to Pearson's correlation, heritability of PLSR-predicted values was calculated as quality criterion, using the R package "SpATS" (Rodríguez-Álvarez et al., 2018). This package allows for providing information on the location of measurements within the experiments as row and column coordinates to a mixed model, which are then used for a spatial correction. While VI-based predictions from PLSR models were trained with random train/test data splits, the SpATS predictions were applied to the complete dataset to have balanced and complete spatial data. After spatial correction, the generalized heritability according to Oakey et al. (2006) could be calculated.

In generalized heritability, the effective dimensions ED_g are divided by the difference between the number of genotypes m_g and the number of zero eigenvalues ζ_g :

$$H_{\text{genal}}^2 = \frac{ED_g}{(m_g - \zeta_g)}, \quad (5)$$

with

$$ED_g = (m_g - 1) \frac{\sigma_g^2}{(\sigma_g^2 + \frac{\sigma_e^2}{r})}. \quad (6)$$

Heritability was calculated for every 10th iteration in the previous cross-validation for each year individually.

For every iteration where heritability was estimated, also best linear unbiased estimators (BLUEs) of genotypes within individual treatments were calculated with SpATS. Genotype rankings were then compared with visual field ratings by means of Spearman rank correlation. To examine the consistency of the ranking for individual genotypes, the ranks of the BLUEs were plotted for visual examination.

2.11 | Method cost comparison

For a schematic comparison of the economic cost of the different methods (Figure 1e), different cost components were estimated based on personal experience. The cost components were, for example, material costs, staff cost, operation/processing cost, transportation cost, as well as continuous data storage costs (e.g., Huang et al., 2024; Marinello, 2023). A detailed listing of the components is shown in Table 4. The values of the components are dependent on the number of measurements $n_{\text{measurement}}$ and the costs were estimated for the four methods "Visual Rating," "Drone RGB," "Drone Multispectral," and "PhenoCam". The total costs for one method correspond to the sum of the different components applicable for the different methods, as indicated by \times in the "Methods" columns in Table 4. Note that 15 measurements were assumed to correspond to one season. Some costs were associated with field visits, which were necessary for every measurement of "Visual Rating," "Drone RGB," "Drone Multispectral," but just twice for the "PhenoCam" method for setup and dismounting. Note that 35 PhenoCam images were taken each week. Once a week was about the average measurement frequency of visual ratings and drone flights, although this frequency can vary from three times a week to biweekly, depending on the phenological stage. Thus, 35 PhenoCam images were assumed to correspond to one measurement of the other methods. For each measurement or field visit, it was assumed that two times 0.4 h would be needed to load and unload the equipment to/from the car for each round trip. Otherwise, a travel speed of 80 km h⁻¹ was assumed, which was relevant for the calculation of the travel costs, depending on the distance covered. Calculations were conducted for two scenarios with 700 plots and 1,400 plots and three different distances of the experiments from the research station, which changed the equipment needed and the time necessary for traveling and rating or flying. The terms used and assumed values for the two plot number scenarios are shown in Table 5. Calculations were performed for 1–90 measurements or zero to six seasons. The storage cost values were estimated to be 0.023 CHF GB/month based on the official price listing of standard Google cloud storage on Swiss-based servers (Google, 2024), assuming a storage duration of 10 years.

TABLE 4 Equations used for cost estimations. The cost of the methods was calculated by summing the equations in this table as indicated with \times in the “Method” columns. The terms used in the equations are described in Table 5.

Description	Equation	Method			
		Visual rating	Drone RGB	Drone multi-spectral	Pheno-Cam
Staff cost round trips	$n_{\text{measurement}} \cdot \text{Cost}_{\text{rating,staff}} \cdot (2 \times t_{\text{drive}}) \cdot n_{\text{days,rating}}$	\times	—	—	—
Vehicle cost round trips	$n_{\text{measurement}} \cdot \text{Cost}^{-\text{dist}} \cdot \text{dist}^{-t} \cdot (2 \times (t_{\text{drive}} - 0.4)) \cdot n_{\text{days,rating}}$	\times	—	—	—
Staff cost rating	$n_{\text{measurement}} \cdot \text{Cost}_{\text{rating,staff}} \cdot t_{\text{rating}}$	\times	—	—	—
Cost drone and sensor	$\text{Cost}_{\text{drone,sensor}}$	—	\times	—	—
Initial processing cost	$n_{\text{seasons}} \cdot (\text{Cost}_{\text{tech,staff}}^{-t} + \text{Cost}_{\text{comput.}}^{-t}) \cdot t_{\text{proc.init.}}$	—	\times	\times	\times
Staff cost round trips	$n_{\text{measurement}} \cdot \text{Cost}_{\text{tech,staff}}^{-t} \cdot (2 \times t_{\text{drive}})$	—	\times	\times	—
Vehicle cost round trips	$n_{\text{measurement}} \cdot \text{Cost}^{-\text{dist}} \cdot (2 \times (t_{\text{drive}} - 0.4)) \cdot \text{dist}^{-t}$	—	\times	\times	—
Staff cost drone piloting	$n_{\text{measurement}} \cdot \text{Cost}_{\text{tech,staff}}^{-t} \cdot t_{\text{flight}}$	—	\times	\times	—
Storage cost images	$n_{\text{measurement}} \cdot \text{Cost}_{\text{storage}}^{-\text{GB}} \cdot \text{Size}_{\text{data,images}}$	—	\times	\times	—
Storage cost photogrammetry	$n_{\text{measurement}} \cdot \text{Cost}_{\text{storage}}^{-\text{GB}} \cdot \text{Size}_{\text{data,photogrammetry}}$	—	\times	\times	—
Cost image handling	$n_{\text{measurement}} \cdot (\text{Cost}_{\text{tech,staff}}^{-t} + \text{Cost}_{\text{comput.}}^{-t}) \cdot t_{\text{proc.}}$	—	\times	\times	—
Computation cost processing	$n_{\text{measurement}} \cdot \text{Cost}_{\text{comput.}}^{-t} \cdot t_{\text{comput.}}$	—	\times	\times	—
Cost drone	$\text{Cost}_{\text{drone}}$	—	—	\times	—
Cost drone sensor	$\text{Cost}_{\text{sensor}}$	—	—	\times	—
Cost field masts	$\text{Cost}_{\text{masts}}$	—	—	—	\times
Cost time lapse cameras	$\text{Cost}_{\text{cameras}}$	—	—	—	\times
Initial processing cost	$n_{\text{seasons}} \cdot (\text{Cost}_{\text{tech,staff}}^{-t} + \text{Cost}_{\text{comput.}}^{-t}) \cdot t_{\text{proc.Init}}$	—	—	—	\times
Staff cost round trips	$n_{\text{seasons}} \cdot n_{\text{visits,PhenoCam}} \cdot n_{\text{persons}} \cdot \text{Cost}_{\text{tech,staff}}^{-t} \cdot (2 \times t_{\text{drive}})$	—	—	—	\times
Staff cost vehicle	$n_{\text{seasons}} \cdot n_{\text{visits,PhenoCam}} \cdot \text{Cost}^{-\text{dist}} \cdot (2 \times (t_{\text{drive}} - 0.4)) \cdot \text{dist}^{-t}$	—	—	—	\times
Staff cost setup & dismounting	$n_{\text{seasons}} \cdot n_{\text{visits,PhenoCam}} \cdot \text{Cost}_{\text{tech,staff}}^{-t} \cdot t_{\text{setup,dismounting}} \cdot n_{\text{person}}$	—	—	—	\times
Storage cost images	$n_{\text{measurement}} \cdot n_{\text{cameras}} \cdot 35_{\text{images}} \cdot \text{Cost}_{\text{storage}}^{-\text{GB}} \cdot \text{Size}_{\text{data}}$	—	—	—	\times
Cost image handling	$n_{\text{measurement}} \cdot n_{\text{cameras}} \cdot 35_{\text{images}} \cdot (\text{Cost}_{\text{tech,staff}}^{-t} + \text{Cost}_{\text{comput.}}^{-t}) \cdot t_{\text{proc,image}}$	—	—	—	\times
Computation cost processing	$n_{\text{measurement}} \cdot n_{\text{cameras}} \cdot 35_{\text{images}} \cdot \text{Cost}_{\text{comput.}}^{-t} \cdot t_{\text{comput.}}$	—	—	—	\times

2.12 | Weather data recording

The air temperature and daily precipitation were obtained from a Meteoswiss (Federal Office of Meteorology and Climatology, <https://www.meteoswiss.admin.ch>) weather station, which was located about 800 m from the experimental site at Changins (46°24′03.7″ N, 6° 13′ 39.6″ E, 458 m.a.s.l., WGS 84).

3 | RESULTS

3.1 | Index dynamics from different sensors and image formats

We calculated 27 RGB VIs from PhenoCam and drone-based data. For drone-based data, 12 additional multispectral VIs were calculated. Sample data of two RGB VIs, visible atmospherically resistant index (VARI) and excess green minus

excess red (ExGR_{Zhang}), are shown in Figure 5, based on JPEG DNs and JPEG-based reflectance. In general, the temporal dynamics appeared very similar between JPEG DN and JPEG reflectance data. The same VIs based on DNG raw DNs instead of JPEG DNs could look different with more variability and less pronounced temporal dynamics (Figure S4). VIs based on drone data appeared to be smoother than the ones derived from JPEG PhenoCam data, but have similar temporal dynamics (e.g., Figure 5c,d). However, previously described patterns were not found for all VIs. For example, Kawashima index (IKAW) VI dynamics were more similar between PhenoCam DNG raw data and drone data than between PhenoCam JPEG and drone data. In addition, IKAW showed less pronounced temporal dynamics, especially for JPEG-based VIs (Figure S5). The maintenance of PhenoCams resulted in small changes in PhenoCam orientation. Consequently, the VI time series of plots at image borders could be interrupted at maintenance (e.g., Plot_102, Figure 5).

TABLE 5 Explanation of terms in cost estimation equations of Table 4. The terms are grouped by methods. Terms in method “Universal” are used in two or more methods. Where applicable, values used in the cost estimation are provided for two scenarios (700 and 1,400 plots) and different methods. Values are in Swiss francs (CHF). In December 2024, 1 CHF corresponded to 1.06 € and 1.11 \$ (www.xe.com).

Method	Term	Description	Value (if fixed)	
			700 plots	1400 plots
Universal	$n_{\text{measurements}}$	Number of measurements.	–	–
	n_{seasons}	Number of seasons, corresponds to $n_{\text{measurements}}$ divided by 15 and rounded down.	–	–
	dist^{-t}	Distance covered within 1 h of drive.	80 km h ⁻¹	80 km h ⁻¹
	t_{drive}	Time to get to the experimental site (one-way). For the time of round trips, this time is multiplied by 2.	–	–
	$(t_{\text{drive}} - 0.4)$	The negative offset of 0.4 h penalizes the distance covered during the first hour of driving for loading and unloading the equipment.	–	–
	$\text{Cost}^{-\text{dist}}$	Cost per kilometer driven.	0.6 CHF km ⁻¹	0.6 CHF km ⁻¹
	$\text{Cost}_{\text{tech,staff}}^{-t}$	Cost of one technical staff for 1 h.	78 CHF h ⁻¹	78 CHF h ⁻¹
	$\text{Cost}_{\text{comput.}}^{-t}$	Cost of 1 h computing.	3 CHF h ⁻¹	3 CHF h ⁻¹
	$\text{Cost}_{\text{storage}}^{-\text{GB}}$	Cost to store 1 GB for 10 years.	2.76 CHF GB ⁻¹	2.76 CHF GB ⁻¹
	$\text{Size}_{\text{data,images}}$	Size of image data per measurement generated with drones or of single images for the PhenoCams.	5 GB ^a /17 GB ^b / 0.016 GB ^c	10 GB ^a /34 GB ^b / 0.016 GB ^c
	$\text{Size}_{\text{data,photogrammetry}}$	Only applies to drone flights: Size of photogrammetric projects of the drone measurements per measurement.	11 GB ^a /18 GB ^b	22 GB ^a /36 GB ^b
	$t_{\text{proc.init.}}$	Time for initial processing, for example, creating georeferenced image masks, setting up the processing pipeline, etc.	7 h ^a /8 h ^b /16 h ^c	14 h ^a /16 h ^b /32 h ^c
	$t_{\text{proc.}}$	Processing time after initial processing.	1 h ^a /3 h ^b /1 s ^c	2 h ^a /6 h ^b /1 s ^c
	$t_{\text{comput.}}$	Computation time of data per measurement (drones) or per image (PhenoCam).	2 h ^a /5 h ^b /1 min ^c	4 h ^a /10 h ^b /1 min ^c
	t_{flight}	Time needed to cover 700 and 1,400 plots, respectively, with droneflights.	2 h	4 h
$\text{Cost}_{\text{rating}}$	$n_{\text{days,rating}}$	Number of days to complete the rating for one measurement, but not necessarily full days. Determines the number of round trips per measurement.	1	2
	$\text{Cost}_{\text{rating,staff}}$	Cost of one rating staff for 1 h.	61 CHF h ⁻¹	61 CHF h ⁻¹
	t_{rating}	Time needed for one measurement of all plots.	5 h	10 h
$\text{Cost}_{\text{RGB,drone}}$	$\text{Cost}_{\text{drone,sensor}}$	Cost of drone with integrated camera system, for example, DJI Mavic 3 Pro (SZ DJI Technology Co. Ltd.).	1700 CHF	1,700 CHF
$\text{Cost}_{\text{Multispec.,drone}}$	$\text{Cost}_{\text{drone}}$	Cost of drone that can carry a Micasense RedEdge-MX DUAL sensor, for example, DJI Matrice 350 RTK (SZ DJI Technology Co. Ltd.).	9300 CHF	9300 CHF
	$\text{Cost}_{\text{sensor}}$	Cost of Micasense RedEdge-MX DUAL sensor (MicaSense Inc.).	9500 CHF	9500 CHF
$\text{Cost}_{\text{PhenoCam}}$	$n_{\text{days,setup,dismounting}}$	Number of days to either set up or dismounting the PhenoCams.	1	2
	$n_{\text{visits,PhenoCam}}$	Number of times to visit the experimental site for the setup and for the dismounting of the PhenoCams.	2	2

(Continues)

TABLE 5 (Continued)

Method	Term	Description	Value (if fixed)	
			700 plots	1400 plots
	n_{cameras}	Number of cameras (two cameras per system, four per mast in our setup).	8	16
	n_{persons}	Two people are needed for the setup and dismantling of the PhenoCams.	2	2
	$\text{Cost}_{\text{masts}}$	Cost of two or four Teksam field masts.	24,000 CHF	48,000 CHF
	$\text{Cost}_{\text{cameras}}$	Cost of four or eight autonomous time lapse camera systems (eight and 16 cameras in total), for example, Enlaps Tikee 3 Pro+ (Enlaps SAS).	6800 CHF	13,600 CHF
	35_{images}	35 images of the PhenoCam correspond to one measurement of the other methods.	—	—
	$t_{\text{proc.,image}}$	Time needed to handle and process one image.	1 s	1 s
	$t_{\text{setup,dismounting}}$	Time needed to setup or dismount the cameras.	6 h	12 h

^aRGB drone, ^bMultipsectral drone, ^cPhenoCam.

TABLE 6 Number of temporal features and observations for different sensors, image formats, and data treatments.

Method	No. of temporal features	No. of observations
JPEG DN	1,005	2,101
JPEG reflectance	1,226	2,101
DNG DN	758	2,092
DNG reflectance	837	2,092
Drone RGB	1,452	810
Drone multispectral	2,217	810

3.2 | Temporal feature count overview

We derived 758 to 2,217 different temporal features from the VI data, depending on the different sensors, image formats, and data treatments (Table 6). Most temporal features were found in the “Drone Multispectral” data, followed by “Drone RGB,” which represents a subset of the “Drone Multispectral” set. This was followed by the PhenoCam-based JPEG and finally the DNG methods. For the latter two, more temporal features were found for the reflectance option than for the DN option. Only features that were available for a large proportion of observations were included in PLSR. Automated feature extraction was more effective on JPEG data than on DNG data, which led to a decrease in features from JPEG to DNG.

For drone-based methods, 810 observations were available for PLSR modeling, as 270 plots were observed in each year. By contrast, plots appeared on multiple PhenoCams in the same season, and, on average, each plot was recorded by 2.6 cameras, although plots on the edges of the field were just recorded from one camera. Camera-plot combinations with 1500 or more missing temporal features were excluded. 2,101

observations were available for PhenoCam JPEG data and 2,092 for DNG.

Using PLSR in a first feature selection round, we selected 200 temporal features for each field reference type and reference level. When we used less than 200 features, correlations between PLSR prediction and field reference levels started to decrease (Figure S3).

3.3 | Comparison of PLSR prediction performance of different methods

Eight methods to predict plant development, which depended on the different sensors, image formats, and data treatments, were compared individually by Pearson’s correlation for the different reference rating types. The mean correlations for all cross-validation data with reference rating values across 100 iterations for each reference level are summarized in boxplots (Figure 6). Six of the methods were based on PhenoCam data, three each for JPEG and DNG data formats. For both formats, a method was based on calibrated reflectance data, one on DN and one on both data types. The remaining two methods were based on calibrated drone data in the multispectral and RGB color space.

Models based on multispectral drone data were best correlated with reference levels for all reference rating types (Figure 6a), followed by RGB drone data. PhenoCam methods showed slightly weaker correlations compared to drone-based VIs, when “All” data types (DN and Reflect) were used, but almost no difference was observed between JPEG and DNG. The JPEG DN method was slightly inferior to using both data types but superior to the remaining three methods “Reflect_JPEG,” “DNG_DN,” and “Reflect_DNG.” The standard deviation of these correlations σ_r (Figure 6b) revealed that higher correlations were also more consistent

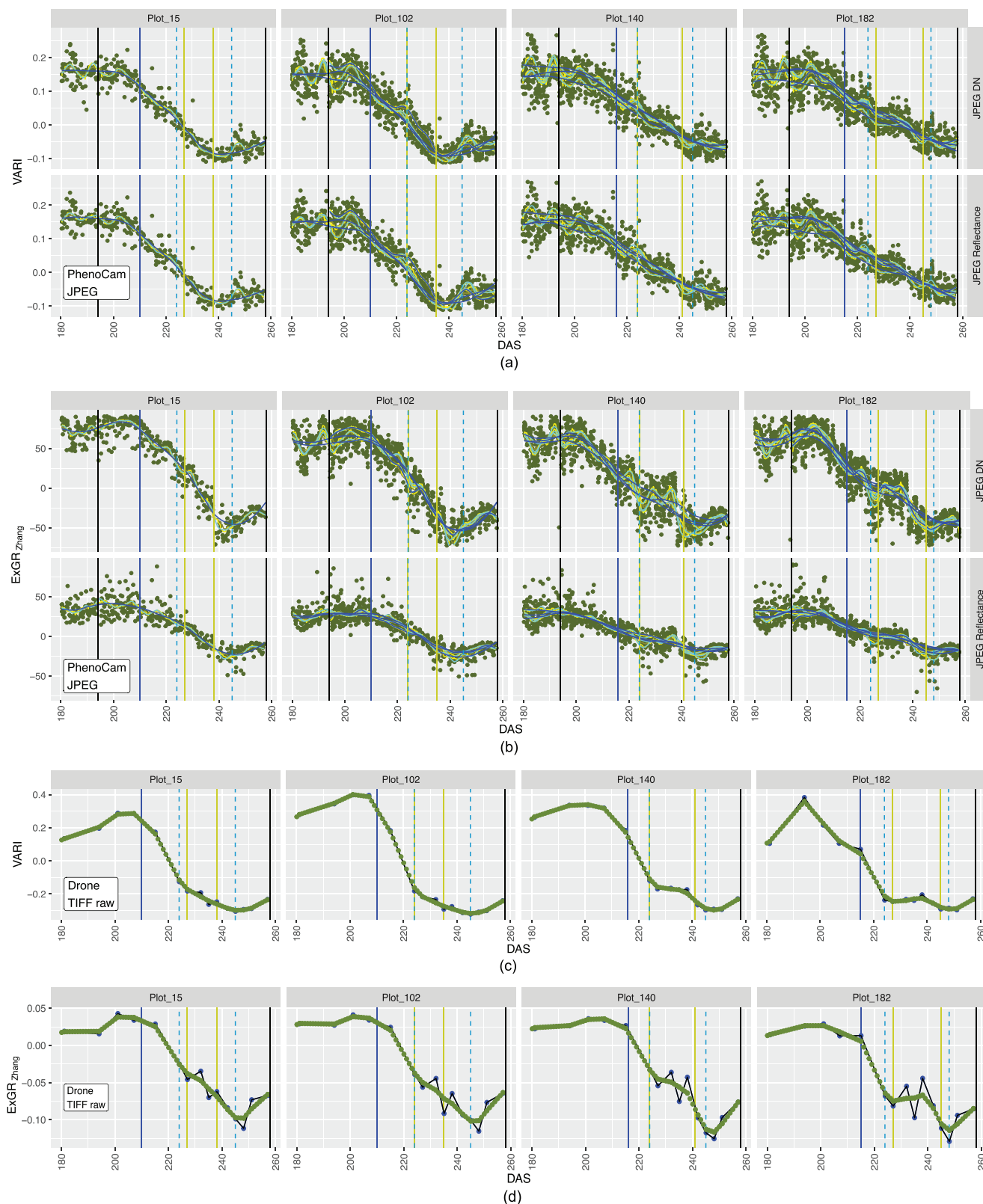


FIGURE 5 Sample vegetation index (VI) data derived from PhenoCams (a and b) and a drone-based camera (c and d) for two VIs (VARI and $ExGR_{Zhang}$) and four randomly chosen plots during the seasons 2022. The temporal axis is in days after sowing (DAS). For the PhenoCam data, green points in the PhenoCam image are initial VI values and lines represent smoothed data of different smoothing methods (dark yellow: rolling mean; dark blue: loess smoothing; yellow: Savitzky–Golay; light blue: spline smoothing). In plots with multiple lines of the same color, multiple cameras observed the same plot. Data are shown for unprocessed data (“DN”) and for calculated reflectance values. For the drone data, the blue dots

(Continues)

FIGURE 5 (Continued)

are the initial VI values. Green lines represent a smoothed spline interpolated for a daily temporal resolution. The colored vertical lines indicate specific levels of visual field reference ratings as observed at the respective plots: Solid blue line indicates the heading date (BBCH 59), the dashed blue lines indicate plant senescence levels of 10% and 90% respectively. The yellow lines correspond to flag leaf senescence at 10% and 90%. The black lines toward the end mark the harvest date. The first vertical black line for the PhenoCam data shows the date of PhenoCam maintenance.

as σ_r was lower for higher correlations, except for the correlation between SenLeaf reference type and drone-based RGB methods.

3.4 | Detailed comparison of selected methods

As the JPEG based PhenoCam methods seemed to perform slightly better than the DNG based models, they were compared to the drone-based methods in more detail. The correlations of the PLSR predictions for all reference levels of the different reference types (Figure 7a) were very strong ($r > 0.8$) and even stronger than 0.9 for the later levels of the BBCH scale and the intermediate levels of SenLeaf and SenPlant in 100 times repeated cross-validation. Early BBCH stages and early as well as late senescence stages showed weaker correlations in general. The correlations were consistently higher for drone-based methods, except for later stages for SenLeaf, in which the drone-based predictions in the RGB color space showed high variability. Applying PLSR models trained on the training data to predict all observations resulted in stronger and more consistent correlations than solely predicting and correlating the test data set in cross-validation. When using the 75%/25% train/test data split in cross-validation, correlations were slightly higher (0.02 on average) compared to 19 genotypes/11 genotypes train/test cross-validation with standard deviation increasing by only 0.01 between the two. Therefore, most of the remaining analysis was conducted on only cross-validation data based on a 75%/25% data split.

Separate correlations for the 3 years were weaker in general (Figure 7b), but the trends remained similar. The correlations were weak to very weak for early SenPlant and weak to strong for late stages of all reference types. For SenLeaf and early stages of SenPlant, the correlations were weaker in 2023 than in the other 2 years. SenPlant showed weaker overall correlations in 2022. The correlation of BBCH did not show a distinct year-wise pattern except for weak correlations for the latest BBCH levels in 2022. The root mean square error (RMSE) was similarly low for both senescence rating types in 2021 and 2023, with slightly higher RMSE for earlier stages in 2022 (Figure S6). As for these correlations, no distinct year-wise pattern was found for RMSE of BBCH.

To better understand the reason behind varying correlations in dependence of the different years, we examined the temporal density of the reference measurements (Figure 8). Later

stages of BBCH occurred in a short period in 2022 compared to the other 2 years. Stages of senescence and especially SenLeaf occurred in a shorter interval for most reference levels in 2023.

3.5 | Predictive performance with relation to genotype

Genotype-wise mean differences from visual reference values were calculated for PhenoCam data based on DN values of JPEG images and drone data based on all multispectral VIs. The predictions were based on PLSR models trained with 75% of the data, but the prediction was applied to the complete dataset. The differences ranged from - 2.96 days to 3.15 days, with a negative difference indicating that the predicted date was too late and a positive difference indicating that the prediction was too early.

In general, the absolute differences were smaller for drone-based predictions compared to PhenoCam predictions, but the differences generally pointed in the same direction for the two imaging platforms.

A group-wise summary of absolute differences was calculated for a general overview (Table 7). The data were grouped by reference type, reference classes, and image acquisition platforms. The largest deviation occurred for the early stages of SenLeaf, with a mean difference of 0.91 days for PhenoCam predictions and 0.49 days for predictions based on multispectral drone data. The best predictions with the PhenoCam were achieved for the intermediate stages of SenPlant (0.36 days) and for late stages of SenPlant with the multispectral drone (0.23 days). The differences were consistently lower for the drone-based estimates.

To assess temporal bias, the differences were compared with the visual reference ratings (Figure S7). In all reference classes, a temporal bias was observed for both methods, but the bias was consistently more pronounced for PhenoCam predictions compared to drone predictions, for which the bias was minimal.

3.6 | Genotype ranking consistency

When comparing the BLUEs based on VI-based predictions with BLUEs based on visual reference values by means of Spearman's rank correlations (Figure 9), general trends were

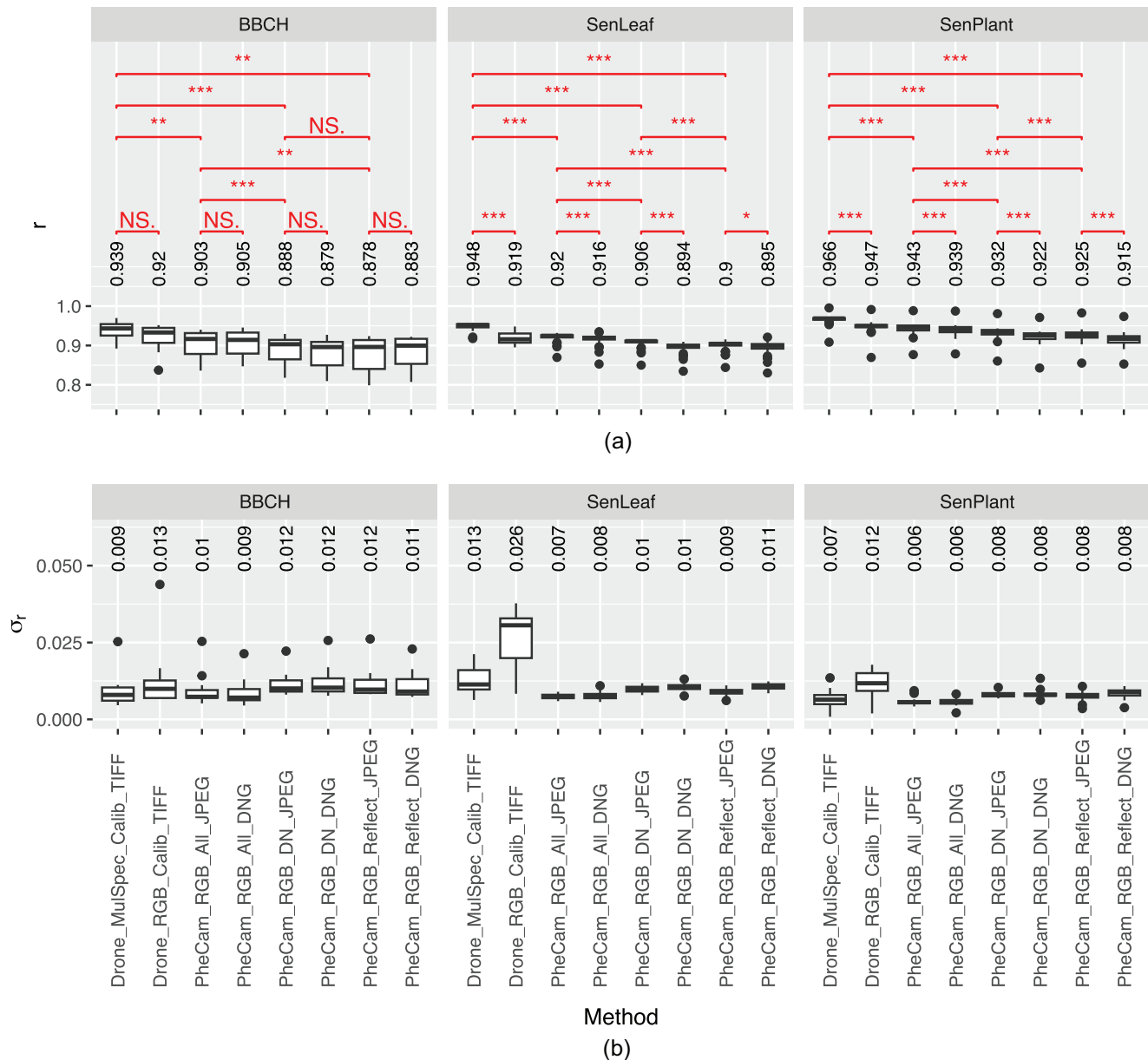


FIGURE 6 Overview of the Pearson's correlation of PhenoCam- and drone-based predictions of timing of phenology (BBCH), flag leaf senescence (SenLeaf), and plant senescence (SenPlant) with field reference measurements. Values above boxplots indicate mean values. (a) Mean correlations r for all cross-validation data (only 75%/25% train/test data split in cross-validation) with reference rating values across 100 iterations for each reference level. (b) The mean standard deviations of these correlations across 100 iterations for each reference level σ_r . The methods compared are indicated by composite names in the x -axis labels: "PheCam" and "Drone" indicate the platform of image acquisition, and "RGB" and "MulSpec" indicate the color space of the features. "Reflect" indicates that only calibrated reflectance data (reference panels) were used as opposed to "DN" for the use of digital numbers. "All" means that "Reflect" and "DN" data were used. "JPEG" and "TIFF" indicate the image data format used. Pairwise t -tests were applied to examine whether the different methods produced significantly different results. Pairing was by reference levels of the three reference types. Significance levels: NS, not significant. $p > 0.05$, $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$.

similar as for the plot-wise Pearson's correlations (Figure 7) but correlations were weaker. Overall, the mean BLUE rank correlation for the 11 genotypes used in cross-validation (random 19 genotypes/11 genotypes train/test data split at each iteration) was 0.87 compared to 0.92 when correlating the BLUEs of all genotypes, and the standard deviation of correlations was larger for the cross-validation (0.16 compared to

0.12 when including all genotypes). Mean correlations were very strong ($r > 0.8$) for most reference levels of the different reference types but they were lower for early and late reference levels. Especially for the early levels of plant senescence there were only very weak rank correlations for the PhenoCam methods but better rank correlations for the drone methods. The mean rank correlation was thus weaker for the

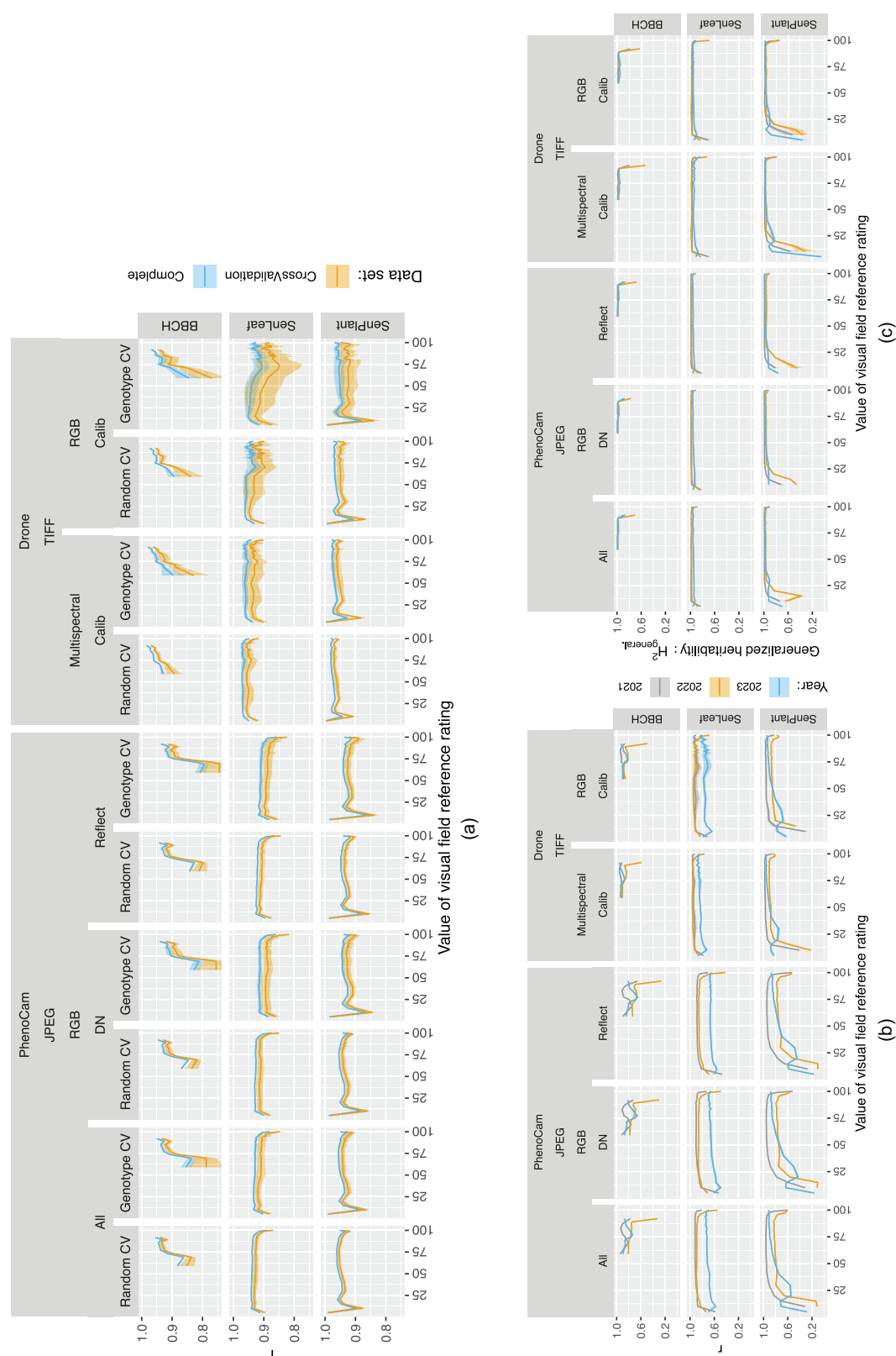


FIGURE 7 Quality metrics of partial least squares regression (PLSR) models. (a) The Pearson's correlation r between PLSR-based predictions and field reference ratings. Data of the different sensors, processing methods and cross-validation (CV) types are arranged in columns. Random CV refers to a 75%/25% train/test data split. Genotype CV refers to a 19/11 genotype train/test data split. The rows represent the three types of field reference measurements: Phenology (BBCH), flag leaf senescence (SenLeaf), and plant senescence (SenPlant). "PhenoCam" and "Drone" indicate the platform for image acquisition, "RGB" and "Multispectral" indicate the color space of the features. "Reflect" indicates that only calibrated reflectance data (reference panels) were used as opposed to "DN" for the use of digital numbers. "All" means that "Reflect" and "DN" data were used. "Calib" for drone data indicates the use of Micasense calibration panel. "JPEG" and "TIFF" indicate the image data format used. Yellow indicates the correlations only for test data from the test-training split, and blue indicates the correlation for all data (train + test). The shaded areas indicated the mean \pm standard deviation in 100 repetitions of cross-validation, while the lines are the means. (b) Correlations are based on all data but analyzed for individual years. (c) Heritability of PLSR-based predictions for 10 repetitions per reference level based on PLSR predictions for all data as input in the SpATS models. Data in (b) and (c) are only shown for 75%/25% train/test data split in cross-validation, and the colors indicate the different years.

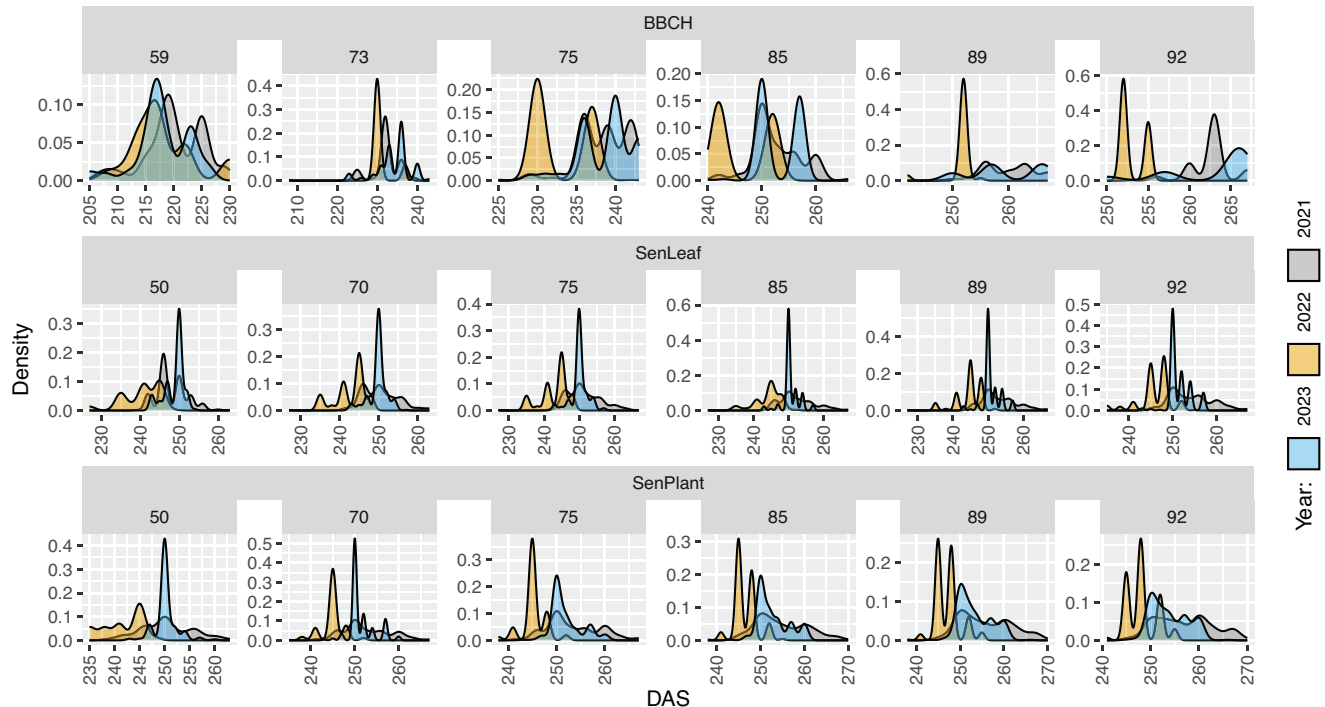


FIGURE 8 Temporal density of interpolated reference observations of all three reference rating types by year for generically selected reference levels. The rows represent the three types of field reference measurements: Phenology (BBCH), flag leaf senescence (SenLeaf), and plant senescence (SenPlant). The temporal axis is in days after sowing (DAS).

TABLE 7 Mean absolute differences between visual reference ratings and vegetation index (VI)-based predictions in days summarized across all genotypes. Mean absolute differences and standard deviations (SD) are shown, grouped by reference types (phenology: BBCH; flag leaf senescence: SenLeaf; plant senescence: SenPlant), reference classes (early, intermediate, late), and image acquisition platforms (PhenoCam and drone). For the drone-based predictions, multispectral VIs are included. Predictions are based on PLSR models trained with 75% of the data, but the prediction is applied to the complete dataset.

Mean absolute differences of prediction from references (days)							
Reference type	Platform	Reference classes					
		Early		Intermediate		Late	
		mean	SD	mean	SD	mean	SD
BBCH	PhenoCam	–	–	0.76	0.56	0.51	0.41
	Drone	–	–	0.41	0.37	0.26	0.20
SenLeaf	PhenoCam	0.91	0.68	0.52	0.42	0.43	0.35
	Drone	0.49	0.38	0.31	0.23	0.24	0.19
SenPlant	PhenoCam	0.54	0.49	0.36	0.21	0.41	0.31
	Drone	0.35	0.27	0.26	0.21	0.23	0.17

PhenoCams compared to drones (0.88 and 0.91) with a larger standard deviation (0.16 and 0.13 respectively).

The ranking of individual genotypic BLUEs was consistent for the genotypes, that is, early genotypes were predicted to be early, late genotypes were predicted to be late for most reference classes and reference types (Figure 10). Consistent with the plot-wise (Figure 7) and BLUE rank correlations (Figure 9), ranking of plant senescence was less consistent with more variability of ranking for individual genotypes.

3.7 | VI and feature type importance

We analyzed normalized relative PLSR coefficient sums $norm.\beta_{rel,sum,ref.class,group}$ to determine the importance, that is, predictiveness of VIs, and feature types. For PhenoCam “JPEG DN” format (Figure 11a), the VIs of the ExGR type and especially ExGR_{Zhang} were the dominant features. For BBCH and early SenLeaf, VARI was also important. For BBCH, GCC also played a crucial role, and the G_R_Ratio

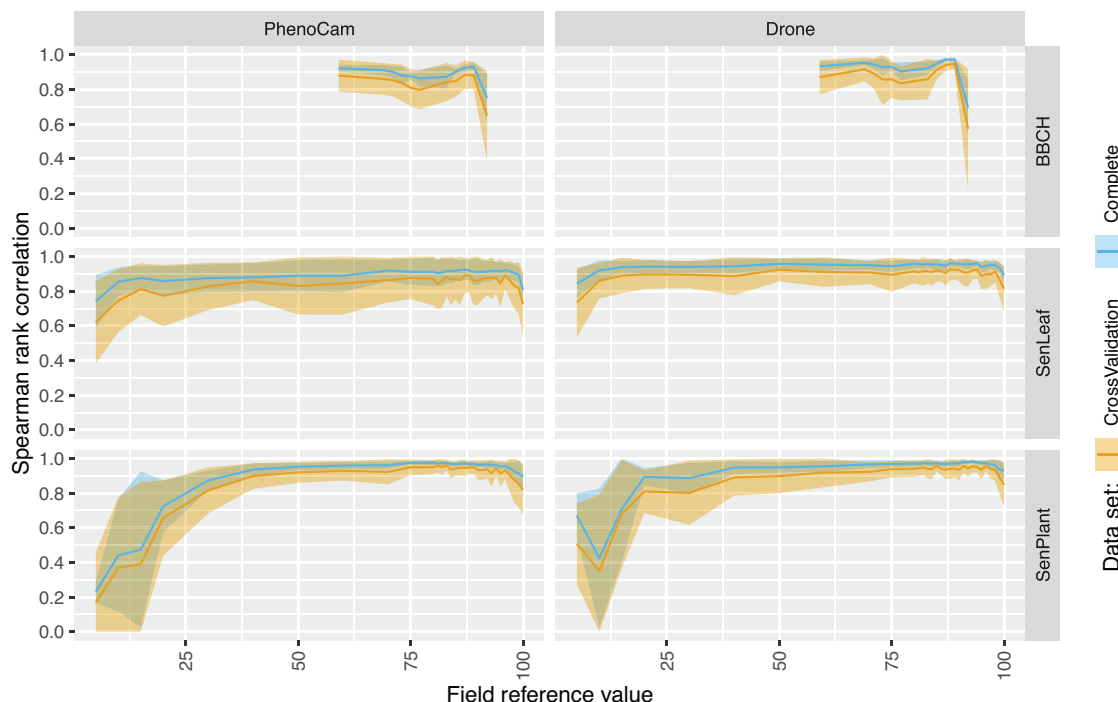


FIGURE 9 Spearman's rank correlation of best linear unbiased estimators (BLUE) of genotypes within treatments between visual reference ratings and vegetation index (VI)-based predictions. The BLUEs were calculated with SpATS for visual reference ratings and VI-based predictions. The data are arranged by reference types (phenology: BBCH; flag leaf senescence: SenLeaf; plant senescence: SenPlant). "PhenoCam" and "Drone" represent the image acquisition platform. For PhenoCams, data are based on DN values of JPEG image format. For drone-based predictions, multispectral VIs are included. VI-based predictions from PLSR models were trained with 19 randomly chosen genotypes in cross-validation of the PLSR models, but the SpATS prediction was applied to the complete dataset to have balanced and complete spatial data. The Spearman's correlations are displayed in blue for the complete dataset (30 genotypes within each treatment), while yellow shows the correlations only for the 11 genotypes from the test-training split that were not included in training. The shaded areas indicate the mean \pm standard deviation in 10 repetitions of cross-validation for which a model was fitted in SpATS, while the lines are the means.

was prominent for SenLeaf. In SenPlant, green leaf ratio (GLI) was an important feature. By contrast, for the PhenoCam "DNG raw DN" format (Figure 11c), $\text{ExGR}_{\text{Zhang}}$, although still important, was the dominant feature only for late BBCH and intermediate to late SenPlant. Furthermore, GCC and red-green-blue vegetation index (RGBVI) were important for BBCH. Excess red index (ExR_g), modified green-red vegetation index (MGRVI), Meyer-Neto vegetation index (MNVI), and RGBVI were predictive for SenLeaf and ExR_g , and MNVI was also important for SenPlant. The predictiveness of VIs varied from early to late reference classes.

The local maximum of the first derivative of the VIs $D1_{\text{LocMax}_1}$ was an important feature type in the prediction of all reference types and values for "JPEG DN" and "DNG raw DN" data (Figure 11b,d). By contrast, both the first local minimum of the second derivative $D2_{\text{LocMin}_1}$ and the first local maximum of the third derivative $D3_{\text{LocMax}_1}$ were increasingly important from early to late reference stages. For other feature types, a similar but less pronounced trend from early to late could be observed. For the different data aggregation methods, no clear trend was found, but the 50th percentile

and/or the mean were generally among the most important aggregation methods (Figure S8).

For drone-based data from the RGB colorspace (Figure S9a), blue chromatic coordinate (BCC), GCC, IKAW, MGRVI, MNVI, red chromatic coordinate (RCC), and RGBVI were especially predictive and the same feature types ($D1_{\text{LocMax}_1}$, $D2_{\text{LocMin}_1}$, $D3_{\text{LocMax}_1}$) but also $D4_{\text{LocMax}_1}$ and the semiparametric temporal features $\text{Spline}_{0.02}$ and $\text{Spline}_{0.1}$ were important (Figure S9b).

In the multispectral color space (Figure S10a), in addition to BCC and IKAW, the most dominant VIs were multispectral VIs, such as difference vegetation index (DVI), normalized difference red edge index (NDRE), NDVI, PSRI_{717} , PSRI_{740} , and soil adjusted vegetation index (SAVI). Most predictive feature types were similar to drone-based RGB data (Figure S10b).

3.8 | Method cost comparison

The cost estimates for the different methods were highly dependent on the number of plots observed, the number of

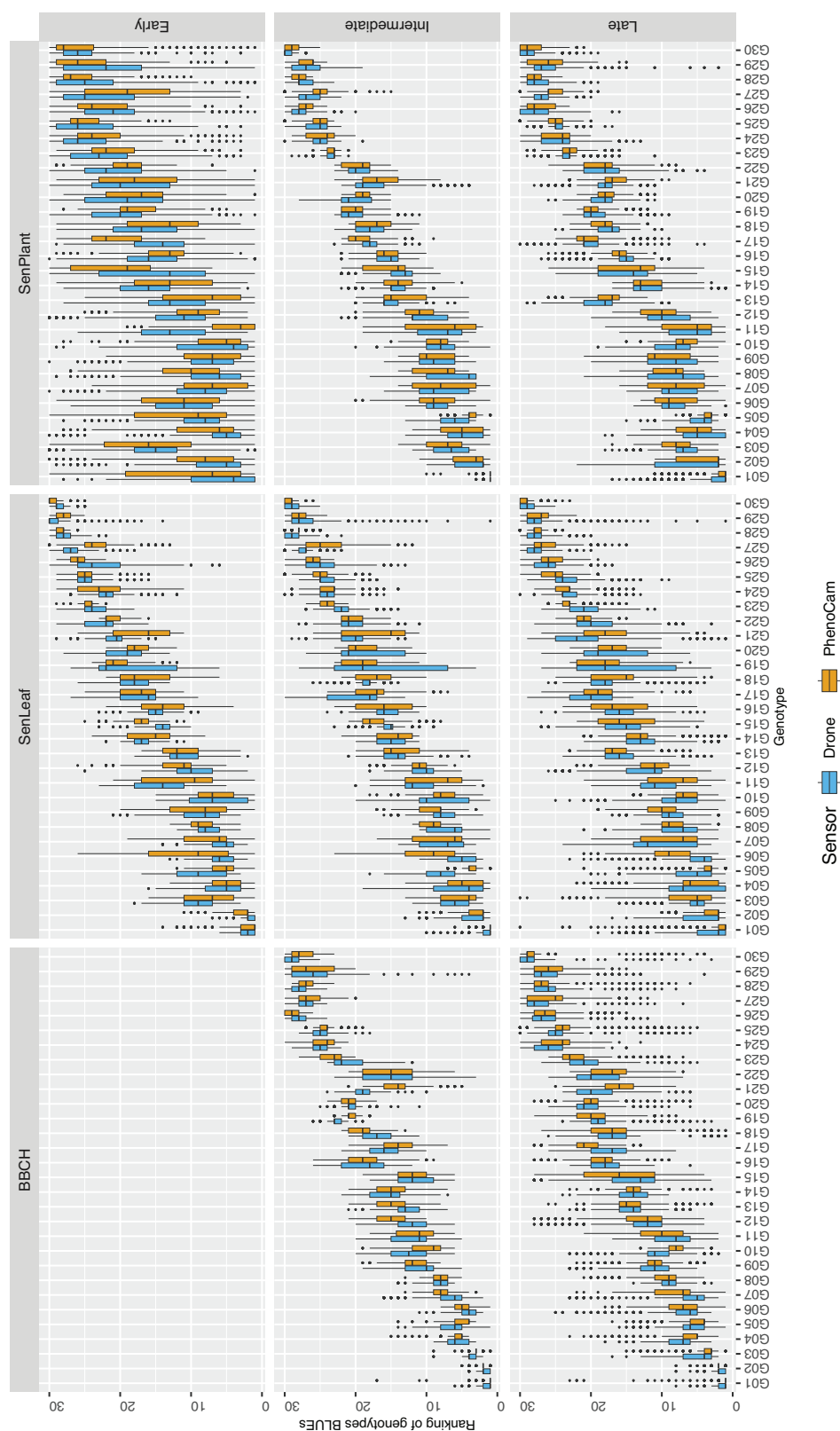


FIGURE 10 Rankings of genotype best linear unbiased estimators (BLUEs). The data are arranged by reference types (Phenology: BBCH; Flag leaf senescence: SenLeaf; Plant senescence: SenPlant), reference classes (early, intermediate, late) and genotypes (G01, ..., G30). “PhenoCam” and “Drone” represent the image acquisition platform and are indicated by colors. For drone-based predictions, multispectral VIs are included. Predictions are based on PLSR models trained with 75% of the data, but the prediction is applied to the complete dataset. Each boxplot contains 240–2,880 data points, depending on how many reference values were summarized within one reference class.



FIGURE 11 Importance of vegetation indices (VIs) and temporal feature types in partial least squares regression (PLSR) models based on PhenoCam-derived JPEG (a and b) and DNG raw (c and d) images. The data are arranged by reference types (phenology: BBCH; flag leaf senescence: SenLeaf; plant senescence: SenPlant), and reference classes (early, intermediate, late). (a and c) Importance of the different VIs as the normalized sum of the relative PLSR coefficients $norm.\beta_{rel,sum}$ over 100 repetitions of cross-validation. (b and d) Importance of the different feature types as $norm.\beta_{rel,sum}$ over the 100 repetitions of cross-validation. D1–D4 indicate the first four derivatives of the Gompertz function. LocMax and LocMin refer to local maxima and minima, and the number after LocMax and LocMin represents the order along increasing days after sowing (DAS) when there were multiple local maxima/minima of the same type. The remaining features correspond to semiparametric temporal features of the smoothing types loess, rolling mean, Savitzky–Golay, and spline.

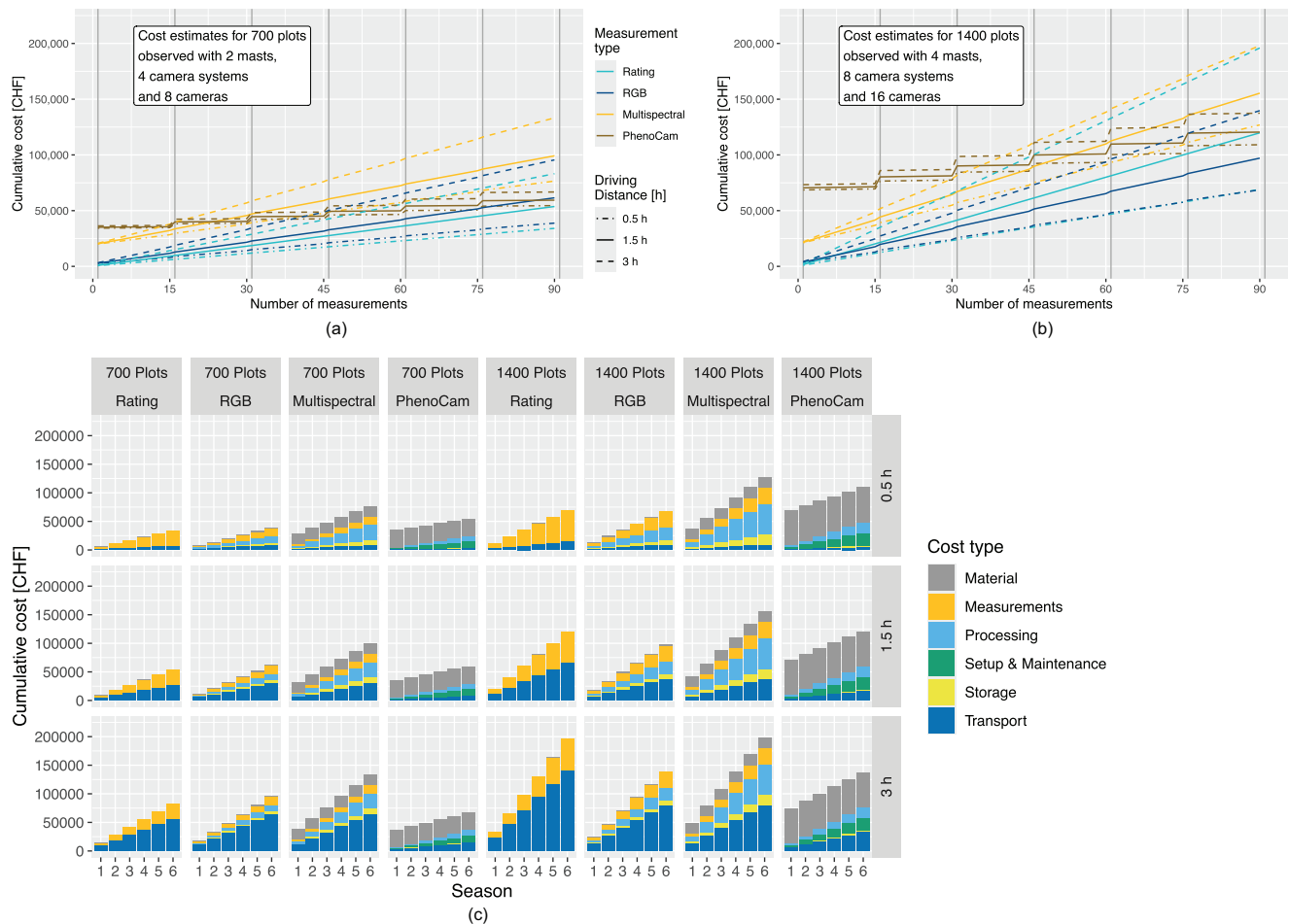


FIGURE 12 Overview of cost development and cost types in dependence of number of measurements and seasons. (a and b) Estimated total costs with respect to number of measurements. Note that 15 measurements were assumed to correspond to 1 year, and years are marked with vertical gray lines. Line types indicate different distances of experimental site from research station in hours (one-way). Line colors indicate different methods that were compared to each other in this study. Measurement types were as follows: Rating = visual field ratings, RGB = RGB drone, Multispectral = multispectral drone and PhenoCams. (c) Cost types for one to six seasons of 15 measurements each are summarized.

measurements carried out, the number of seasons considered, and the distance of the experimental site from the research station. The visual field reference rating, closely followed by the RGB drone method, had the lowest initial costs (Figure 12a,b). The costs then increased almost linearly with the number of measurements. For the 700-plot scenario, visual rating was cheaper than the RGB drone for all driving distances, and this difference increased with the number of measurements. In the scenario including 1,400 plots, for 0.5 h of driving, the visual rating was slightly cheaper than the RGB drone, but for 1.5–3 h of driving time, the RGB drone method was cheaper, and this difference increased with the number of measurements. The method with the next-highest initial cost was multispectral drones, for which the costs increased almost linearly with the number of measurements. The marginal costs for additional measurements were higher for multispectral measurement compared to the RGB drone method and

the visual rating (Table 8). The PhenoCam method had the highest initial cost. By contrast, additional measurements had only a slight effect on costs. New costs arose, particularly for the setup and dismantling of the camera systems. However, these seasonal costs were higher than with the other methods. For non-PhenoCam methods, the costs were relatively low at the beginning of a new season. Only new flight plans had to be created for new seasons and new plot masks had to be drawn for the analysis of the images of drone-based methods. The different travel times had a relatively little effect on the PhenoCam method, while they led to significantly different costs for drone flights and visual ratings. In the 700-plot scenario, at around 30 measurements or three seasons, the PhenoCams became cheaper compared to multispectral drones, and around 60 flights or four seasons, they became cheaper compared to RGB drones except for the 0.5 h driving distance. These general patterns were

TABLE 8 Marginal cost of different methods after 90 measurements for the 700-plot and the 1400-plot scenarios.

Method	Driving distance (h/one-way)	Marginal costs (CHF)	
		700 plots scenario	1,400 plots scenario
Rating	0.5	376	764
	1.5	594	1,328
	3	921	2,174
PhenoCam	0.5	227	456
	1.5	273	562
	3	342	721
Drone—RGB	0.5	407	732
	1.5	659	1,048
	3	1,037	1,522
Drone—Multispectral	0.5	635	1,188
	1.5	887	1,504
	3	1265	1978

similar for the 1,400-plot scenario, but more measurements were needed before the PhenoCam method became cheaper than drone-based approaches.

Considering cost structures, transport was an important cost factor for a driving time of 1.5 h or more and the most significant cost driver for visual field reference ratings, especially in the 1,400 plot scenario, as the field needed to be visited twice to complete the rating of all plots. Another scenario would be to stay overnight in a hotel, which would also lead to higher costs, but this scenario was not included here. The sum of measurements cost (i.e., drone piloting) and the image processing cost for the RGB drone was similar compared to the measurement cost of visual ratings in the 700-plot scenario. For 1,400 plots, the RGB drone was cheaper for 1.5 h or more driving time, as it could cover more plots in a shorter period without the need of an overnight stay or a double visit to complete the measurements. The multispectral drone method came with higher processing costs, and as large data volumes were produced in multispectral imaging, the storage of the images became an important cost driver, in addition to higher initial material costs. It was the most expensive method after six seasons in all scenarios.

PhenoCams had by far the highest initial material costs. By contrast, the transportation costs were low, as only two field visits were necessary for set up and dismounting, assuming no technical incidents occurred. Although each PhenoCam could capture many images per day, the total amount of data was very manageable compared to drone measurements, especially multispectral. Even though the initial processing costs were relatively high, as plot masks had to be corrected for perspective and adapted to shaking cameras and growing vegetation, the overall processing costs remained relatively low.

Thus, once the material was acquired, the cost for additional measurements was relatively cheap, making this the most economical method for scenarios with 3 h driving time after six seasons. In the scenario with 1.5 h driving time, PhenoCam costs were comparable to those of RGB drones and to visual field ratings after six seasons. Only for the 1,400-plot scenario did PhenoCams remain more expensive compared to RGB drones. Yet, due to the low marginal costs of additional measurements (Table 8), PhenoCams would become the cheapest method with additional seasons, which would eventually also be the case for scenarios with 1 h driving time and a high number of seasons to measure.

4 | DISCUSSION

4.1 | Ability of PhenoCams to track phenology and senescence in field conditions

The PhenoCams allowed for the tracking of phenology and senescence over three seasons with high reliability and genotype specificity under field conditions in a real variety testing experiment. Field phenotyping is known to be one of the most challenging phenotyping settings due to confounding effects, such as spatial and temporal variability of traits due to, for example, heterogeneous field conditions, changing weather throughout a measurement campaign, or disruptive weather events, which can make data acquisition and interpretation very challenging (e.g., Araus et al., 2018; Aasen et al., 2020; Reynolds et al., 2020). For example, in this study, shaking cameras, dust and dirt, changing illumination, noncontinuous VI dynamics due to drought/rain interplay, memory restrictions of autonomous cameras, trade-offs in the experimental setup due to the needs of field operations, and animal interference (e.g., Figure S11) made it complicated to operate the cameras and analyze the images. Nevertheless, strong correlations with visual field reference ratings and high heritabilities for PhenoCam-based traits were attained in field conditions. The quality of the predictions derived from PhenoCams was slightly inferior to drone-based predictions, and especially to multispectral predictions. Nevertheless, PhenoCams are a promising tool for the field phenotyping of dynamic traits. They increase the temporal resolution of image acquisition considerably, even at remote experimental sites.

4.2 | Quality of predictions from PhenoCam and drone setups

A comparison of PhenoCam data with drone-based multispectral data was conducted to estimate the benefit of bands from near-infrared and red-edge regions to track plant

development. Comparison with drone-based RGB data allowed for the estimation of effects related to viewing geometries. The better performance of the drone-based VIs compared to the PhenoCams may be largely due to the near-optimal conditions under which the drone measurements took place. Plots were observed at a close to nadir view, with homogeneous viewing geometries, while in PhenoCams, plots were captured from extremely different viewing geometries and distances. In plots close to the PhenoCam, single plants, even plant organs, were distinguishable in a nadir-oriented view. For the most distant plots, a single pixel corresponded to several plants, and only plot-wise mean color properties could be tracked from a very lateral view on the upper part of the canopy. This added variability to the optical signal beyond phenotypic differences of different plots, which makes the analysis more challenging. For flying with the multispectral camera, the timing was, whenever possible, close to noon, allowing for relatively homogeneous illumination of images between flights. This increased the signal-to-noise ratio compared to PhenoCams, which captured images at a much higher frequency but with a large variability of illumination and viewing geometries.

PhenoCams covered many plots with two or even three cameras. Although the same plot was observed, the viewing geometry and distance from the camera were often very different, especially with regard to the path of light from the sun via the plot to the camera. The plots were subjected to bidirectional reflectance changes (Nicodemus, 1977; Schaepman-Strub et al., 2006) for the different cameras. Nonetheless, based on a visual comparison (Figure 5; Figure S5) and high heritability (Figure 7c), the different cameras tracked similar VI dynamics for the same plot, as VIs normalize and reduce the effects of bidirectional reflectance changes (Aasen et al., 2020; Sonnentag et al., 2012).

Genotype predictions were often within a range of ± 1 day from the visual reference rating, which is fairly sufficient for differentiating early, intermediate, and late genotypes in variety testing. This was confirmed by Spearman's rank correlations of genotype BLUEs and when plotting the ranking of BLUEs of individual genotypes across different reference classes and reference types. Nevertheless, the PhenoCam predictions showed a temporal bias toward the mean dates across all measurements, highlighting a limitation of the PhenoCam approach. By contrast, the temporal bias for drone-based methods was minimal.

4.3 | Quality of predictions in different years

Although the prediction correlations were lower for SenLeaf in 2023 than for the other years (Figure 7b), the RMSE of the predictions was lower or similar than in the other years (Figure S6). At the same time, correlations for SenLeaf in

2022 were strong, despite a relatively high RMSE. This could be explained by the extended duration during which specific SenLeaf levels occurred in 2022 and the short duration in 2023, as shown by the temporal density of selected rating levels (Figure 8). When levels occur in a relatively short period, strong correlations are more difficult to attain and weaker for the same RMSE compared to situations with a more temporally dispersed occurrence of the same rating level.

This highlights that the quality of the method is also affected by $G \times E$ interactions, as meteorological conditions that promote rapid plant development and senescence lead to weaker correlations. This is also the rationale for why data from the 3 years were used to train the PLSR models. The meteorological conditions and, therefore, the development of the plants contrasted strongly over the 3 years. A low predictive performance would be expected when predicting a wet year from dry years, and the two dryer seasons also had very different phenological development.

In addition, VI dynamics are directly affected by meteorological conditions. For example, 2022 was a hot and dry year that caused the flag leaves to roll. After rain events, the leaves were able to recover slightly, which could lead to a short-term flattening of the temporal dynamics toward maturity. This may be a valid explanation of noncontinuous trends, for example, for the VIs VARI, ExGR_{Zhang} and NDVI in 2022, in which data in June showed the slope of declining VIs flattened out after significant rains or even increased again (Figure S12).

4.4 | Quality of predictions in different stages

The prediction accuracy of PLSR models for the early or late phenology and senescence stages in this study was often low. This may be related to the small phenotypic changes with which these early and late stages are associated, which may be too small to be detected from cameras at a distance. In addition, in this study, multiple raters conducted the ratings over years but also within years, inevitably leading to rater bias. Later stages of phenology are tedious to track, as they require the manual inspection of some grains in each plot and are limited in precision to track small changes between rating events (Anderegg et al., 2020). This is also true especially in the early and late stages of senescence. Visual scoring methods are subjective and can be affected by foliar diseases, abiotic foliar damage, and other confounding influences, leading to phenotypic heterogeneity within plots (e.g., Chapman et al., 2021; Christopher et al., 2016; Kipp et al., 2014). Thus, the precision limitation of visual scoring itself is likely to limit the precision of the approach (Anderegg et al., 2020).

Later stages of phenology do not address the external phenotypes of the plants but the state of the grains, which cannot

be visually seen without opening the husk. The high predictiveness of the RGB, as well as the multispectral VIs, is thus surprising and most likely the result of a relationship between external visual features and internal processes in the grain.

4.5 | Comparing VIs from different image formats and data treatments

Even within data from the same sensor, VIs can show very distinct patterns depending on the image format and data treatment (e.g., Figure S5). The JPEG DN method was shown to be superior to the JPEG reflectance and both DNG raw methods in this study, while it was inferior to the combined use of DN and reflectance in the same PLSR analysis, although the differences were rather minor (e.g., Figure 6).

JPEG format images are derived from raw images after Bayer matrix decomposition by multiple transformations, such as white balance application, gamma correction, and dynamic range compression. These transformations increase the contrast in the images and lead to a visually enhanced nonlinear representation of light intensities. The nonlinear nature of these transformations also leads to changes in VI dynamics. With gamma correction exponents <1 , which make images appear brighter, changes for high values in the linear format represent smaller changes for nonlinear JPEG. For lower values, changes in the linear DNG raw format lead to larger changes in nonlinear JPEG. With ratios including high/low values and especially both, the ratios can be very different for linear and nonlinear formats. If VI formulas include sums/differences, the VIs can even change from positive to negative or vice versa. Indices such as VARI and ExGR did not seem to be affected significantly, while other indices, such as IKAW, were affected. These transformations can thus amplify VI dynamics, which may explain why automated feature extraction was more effective on JPG data than on DNG raw data in this study.

4.6 | Comparing color spaces and RGB sensors

Cao et al. (2021) showed the superiority of multispectral VIs over RGB-based VIs and that they could reveal more detailed phenotype changes but were also more sensitive to rainfall than RGB VIs, which also seemed to be the case in our study (e.g., Figure S12). In the study at hand, the drone-based PLSR prediction from RGB VIs was often almost as strongly correlated with visual reference ratings as those from multispectral measurements, which is in line with Cao et al. (2021), who showed that multispectral and RGB VIs together only slightly outperformed the pure RGB VIs. However, the robustness of

the prediction (σ_r), seemed more affected, especially for later stages of SenLeaf.

This study only used RGB bands from a narrowband multispectral camera for RGB-based VIs, but Cao et al. (2021) compared a low-cost integrated RGB camera of a DJI Phantom 4 drone with more expensive MicaSense multispectral narrowband sensors for their ability to track stay-green phenotypes in wheat. RGB VIs based on the cheaper sensor better classified senescence types than RGB VIs from the more expensive narrowband sensor in their study. Thus, it is highly likely that the method presented herein would lead to results of similar or better quality when applied to drone-based low-cost RGB camera data.

4.7 | Combining multiple temporal features in a PLSR analysis

Differences in absolute values of spectral bands or VIs depend not only on the phenology of a plant, but also on morphology and canopy structure (Anderegg et al., 2020), leaf pigments, and epicuticular waxes (Tafesse et al., 2022), and viewing geometry (Aasen et al., 2020). However, relative changes over time, that is, the dynamics of the VIs, are stable and suitable for the extraction of temporal features (Aasen et al., 2020; Anderegg et al., 2020).

Pigments such as chlorophyll, anthocyanins, and carotenoids are formed and degraded at specific times during plant growth (Fischer & Feller, 1994; Hörtensteiner, 2006) and these changes are temporally correlated with dynamic changes in VIs. At different developmental stages, the different VI-based temporal features of different VIs are best correlated with the physiological processes of the plant. Thus, it is reasonable to base the analysis not on absolute VI values but on their dynamics and not to use a single VI for all stages of phenology and senescence but to combine multiple VIs in an analysis. With this rationale, using temporal features in PLSR analysis is a promising new approach. PLSR can be used to analyze a dataset with a higher number of predictor variables than the number of observations and in which the predictor variables are strongly correlated (Carrascal et al., 2009), which can be expected for the different temporal features used as predictor variables in this study.

Thus, selecting the 200 most predictive features was not meant to avoid overfitting but to reduce computational effort in the 100 times repeated cross-validation. In PLSR analysis, overfitting can be avoided by choosing a number of PLSR components that is significantly smaller than the number of predictor variables. With a maximum number of 15 components in our PLSR models, the ratio predictors/components was ≥ 139 for PhenoCam data (2092 or 2101 observations and 15 or fewer components), and ≥ 54 for drone data

(810 observations and 15 or fewer components). Thus, the number of observations was much larger than the number of components, and no overfitting was expected.

Although the 75%/25% train/test data split in cross-validation led to slightly better prediction accuracy, 19 genotypes/11 genotypes train/test cross-validation allowed for high correlations between predictions and visual reference measurements. The development of 11 randomly chosen genotypes, and thus 99 plots, unseen in training, was accurately predicted by PLSR models in 100 repetitions for each reference level, which demonstrates the generalizability of the method and argues against overfitting. An increased set of genotypes in training could further increase the generalizability of the PLSR predictions reported in this study.

4.8 | Cost and measurement frequency of different methods

In this study, PhenoCams were the cheapest method for tracking phenology and senescence after five seasons in the 700-plot scenario when considering an experimental site with a driving distance of 3 h (one-way). Although the initial costs of PhenoCams for hardware and the efforts for setup were fairly high, they allowed for an almost unlimited increase in the temporal resolution of image acquisition to hourly or even beyond without significantly increasing data acquisition costs. However, for visual field ratings and drone-base approaches, the round trips necessary for each measurement became the main driver of cost, and every additional measurement came at considerable marginal costs.

By contrast, PhenoCams need—assuming no technical incidents occur—just two field visits for setup and dismounting and if images could be transmitted via mobile networks, SD cards would not need to be changed when full. The DNG raw format used in this study was too data-heavy to be transmitted to a server via a mobile network. However, we demonstrated that the JPEG format-based VIs allowed the tracking of senescence and phenology to be even slightly better than those based on the DNG raw format. The 8-bit JPEG format is lighter and can be transmitted to servers; thus, no SD card change would be necessary in such a setup, and no storage limitation would hinder frequent image acquisition. An image in JPEG could be transmitted every 10 min, maximizing the probability of capturing good quality images on many days. Such a JPEG-based setup would also make it possible to follow the seamless operation of the cameras almost in real time without the need to visit the PhenoCams in person. Using PhenoCams with the JPEG format therefore offers many benefits without a major loss in quality.

Tschurr et al. (2024) argued that at higher temporal resolution, RGB-based VIs can compensate for lower spectral

resolution. Such comparisons are difficult for multiple reasons. For example, Tschurr et al. (2024) did not include the DVI, PSRI, and SAVI multispectral VIs in their study, which all showed high predictiveness in this study, and PSRI was approximating visual senescence ratings best also in the study by Anderegg et al. (2020). In addition, our study confirmed the number of field visits as an important cost driver in the context of remote field experiments (Barreto et al., 2024; Montazeaud et al., 2016; Velumani et al., 2020). In the 700-plot scenario, the marginal costs for the RBG drone at 0.5, 1.5, and 3 h driving distance were 407, 659, and 1,037 CHF, respectively. For the 1,400-plot scenario, the cost was, again, significantly higher (Table 8). Thus, additional measurements came at a price, and if multispectral sensors need to be flown less often, this could lead to multispectral VIs being the economically more favorable option, despite higher marginal costs, depending on the difference in the number of flights required compared to RGB-based VIs. Multispectral imaging becomes particularly interesting if the sensor has already been procured to measure other plant traits.

Furthermore, drone flights must be organized along with other activities, and the logistics of a field season can be very demanding, as many tasks can be completed only in good weather conditions. Due to time constraints, it is often not possible to fly in optimal conditions, and flying, especially in rainy periods and for distant experiments, may be impossible. Multispectral drones, therefore, also have an advantage in these aspects, due to the lower number of flights required in dense field seasons, thereby allowing for more flexibility in planning and a lower workload. PhenoCams, on the contrary, can capture an image during that rare 20 min of a day under suitable conditions without the need for intervention. Further improvements in the PhenoCam setup and pipeline could lead to an additional shift in balance in favor of PhenoCams. However, following similar reasoning, the availability of autonomous drone systems may shift the balance in favor of drones again. Although these systems were strongly restricted by regulation some years ago in many countries (Aasen et al., 2020), the legislation changed in some countries, and such systems can be operated, drastically facilitating the logistics of distant experiments and increasing the probability of a high frequency of flights in fair meteorological conditions.

The cost comparison did not include the additional benefits of the methods. Especially multispectral VIs provide additional information on the plant state, such as general health and nitrogen content. It is challenging to put a price tag on this type of information, but this aspect should not be neglected in such considerations. Finally, the cost considerations presented herein are meant to serve as a conceptual framework that allows us to approximate real costs and to reason about the most relevant cost structures. They are not meant to be precise representations of true costs, which are even more complex.

4.9 | VI and feature type importance in PLSR modeling without ex ante knowledge on phenology

Normalized relative PLSR coefficient sums $norm.\beta_{rel,sum,ref.class,group}$ were used to describe the importance of different VIs and feature types in respective PLSR models. They are an integrative measure, as they are also impacted by how often a VI or feature type is included in the input data of the PLSR models. This depended a lot on how well the VIs could be smoothed with the different smoothing functions or fit with a Gompertz function. Many studies usually normalize dynamics with ex ante knowledge about phenology, for example, by calculating the days after anthesis (e.g., Anderegg et al., 2020; Christopher et al., 2014, 2016; Cao et al., 2021). This requires the preceding tracking of phenology. The methods developed and examined in this study are meant to work without the need for supplement ratings. The methods rely on clear VI dynamics from an early minimum to a late maximum or vice versa. VIs that did not follow such a clear dynamic may have failed in the automated analysis procedure, may thus be underrepresented in the PLSR input data, and may never reach high $norm.\beta_{rel,sum,ref.class,group}$ values. Yet, $norm.\beta_{rel,sum,ref.class,group}$ remains a valid metric to describe the overall usefulness of a VI to the process presented.

As previously described, different sensors, image formats, and data treatments can lead to different VI dynamics. Hence, different VIs and feature types were most predictive among the various methods. Furthermore, the input data in this study were highly correlated, and with correlated data, small changes in the data could lead to the preference of one feature over the other. Within one data type or image format, the temporal features of the input data in the PLSR models were the same for the PLSR models of the different reference rating types (BBCH, SenLeaf, and SenPlant) and classes (early, intermediate, and late). Changes in $norm.\beta_{rel,sum}$ for VIs and feature types between reference rating types and classes showed that for different phenotypic processes and stages, different temporal features were most useful, indicating that the tracking of such processes should not rely on one but on multiple VIs.

4.10 | Practical limitations of the PhenoCam approach

Although PhenoCams reduce the need for frequent field visits, they have some practical limitations. Importantly, the footprints of the mast ropes occupy a significant area, with a diameter of 20 m, which poses an obstacle to field management. Then, the masts should be high to cover a large area with top-down viewing geometry. However, there are certain

risks involved in leaving tall masts unattended in fields for long periods of time, as the anchors could loosen in the wind, and the masts could fall. Tall masts also lead to slightly moving cameras, which means that the image sections between images can change. Producing the adjusted masks is laborious and time-consuming, unless reliable digital image stabilization is used, which would simplify processing. More detailed practical aspects of applying PhenoCams are described in the Supporting Information Section S5. Although PhenoCam was tested on real variety testing trials, only one location was tested each year. To test the applicability of the method in multi-environment trials, the next step would be to carry out trials at several locations in one year.

4.11 | Phenotypic interpretation of most relevant VIs

To link the most relevant VIs of PLSR prediction to specific physiological metrics and environmental conditions, what the color bands represent must be understood. Green light is generally strongly reflected by chlorophyll, while large proportions of red and blue light are absorbed (Bendig et al., 2015; Tucker, 1979). Other pigments also affect the color of plants. Carotenoids are generally red, orange, or yellow, whereas anthocyanins contribute to a pink, scarlet, purple, or blue appearance (Młodzieńska, 2009; Sudhakar et al., 2016). Although these pigments are less visible than chlorophyll, they still contribute to the spectral signatures of wheat canopies and their changes over time. The differences in reflectance for healthy vegetation are used in RGB-based VIs. G_R_Ratio and MGRVI, which contain only the red and green bands, are mostly driven by chlorophyll content, with MGRVI using squaring to amplify reflectance differences between bands. RCC, GCC, BCC, GLI, VARI, MNVI, RGBVI, and different ExGR type VIs also contain the blue band and are more complex to interpret. Higher blue values can be related to non-leaf objects, such as woody parts (Moore et al., 2016).

For VARI, the idea of the green-red ratio used in normalized green-red difference index (NGRDI) was complemented by the introduction of the blue band for atmospheric auto-correction, and VARI has also been shown to be more sensitive to fractional canopy cover than NGRDI (Gitelson et al., 2002). RGBVI captures the reflectance differences due to chlorophyll α -absorption and chlorophyll β -absorption (Bendig et al., 2015). IKAW does not even use the green band to estimate the chlorophyll content of the leaves. Although the inclusion of the green band allows for a better correlation with chlorophyll under specific meteorological conditions, IKAW was shown to be more robust in more diverse weather in a previous study by Kawashima (1998). Multispectral CIs bands from near-infrared and red-edge regions generally show a high reflectance from both stressed and healthy

vegetation. They are therefore typically combined with the red or green band to build VIs that amplify the difference between near-infrared or red-edge and red or green. NDVI and NDRE are very similar, but the latter uses a red-edge band in contrast to the near-infrared band of NDVI. Compared to NDRE, NDVI tends to saturate more easily with a high chlorophyll content (Morlin Carneiro et al., 2020), but both are driven mainly by chlorophyll content, as is the simpler DVI (Tucker, 1979). With SAVI, a soil influence correction is applied, but otherwise it is very similar to NDVI and driven by chlorophyll content (Huete, 1988). Finally, PSRI uses the chlorophyll/carotenoid ratio, which appears to be an important indicator of the concurrent pigment breakdown processes during senescence (Anderegg et al., 2020; Merzlyak et al., 1999).

In summary, the ensemble of the most useful VIs in this study appeared to be driven mainly by chlorophyll, but the concurrent breakdown processes of anthocyanins and carotenoids also seemed important. Most RGB-based VIs included the blue band, which supposedly helped to auto-correct for the influence of soil background and changing weather conditions. The multispectral VIs were most likely more sensitive to chlorophyll content than RGB-based VIs.

5 | CONCLUSION

This study proposed PhenoCams as an alternative to field ratings to track phenology from heading onward and senescence at plant and flag leaf levels in wheat variety testing. A mobile PhenoCam setup was demonstrated, and a PLSR approach was applied to analyze multiple temporal features of different VIs. PhenoCam-based predictions were only slightly inferior to a technical benchmark based on multispectral drone images. For phenology, the mean genotype-wise absolute deviation from visual reference ratings was 0.62 days for PhenoCams and 0.32 days for multispectral drones. Flag leaf senescence deviated by 0.51 and 0.29 days, and plant senescence deviated by 0.42 and 0.25 days for PhenoCams and drones, respectively. Uncalibrated JPEG images were sufficient to track plant development, and images could be directly transmitted via mobile networks from remote experimental sites in future setups. PhenoCams needed a higher initial investment but enabled an increased temporal resolution at lower marginal costs, whereas transfers to experimental sites were shown to be important cost drivers for ratings and drone flights. Thus, these findings present PhenoCams as a promising option when a high temporal resolution is needed. They could be a cost-effective lean phenotyping method to replace visual ratings of phenology and senescence in distant experiments, for example, multi-environment trials.

AUTHOR CONTRIBUTIONS

Simon Treier: Conceptualization; methodology; software; formal analysis; visualization; writing—original draft. **Juan M. Herrera:** Project administration; funding acquisition; conceptualization; supervision; methodology; acquisition; writing—review and editing. **Lukas Roth:** Writing—review and editing. **Nicolas Vuille-dit-Bille:** Support in data acquisition and index curation; writing—review and editing. **Margot Visse-Mansiaux:** Support in data acquisition; review and editing. **Helge Aasen:** Practical advice; review and editing. **Frank Liebisch:** Data acquisition; review and editing. **Achim Walter:** Review and editing.

ACKNOWLEDGMENTS

We thank Johanna Antretter, Fernanda Arelmann Steinbrecher, Ulysse Schaller, Matthias Schmid, and Julien Vaudroz for rating the phenology, and Nicolas Widmer and his team, as well as Yann Imhoff, for field management.

Open access publishing facilitated by Agroscope, as part of the Wiley - Agroscope agreement via the Consortium Of Swiss Academic Libraries.









CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

A detailed description of the mask creation and adjustment procedure suggested in this publication, including source code and example data, is provided on GitHub (<https://github.com/TreAgron/ShapeFromCSVHomographyTransform/tree/main>).

ORCID

Simon Treier  <https://orcid.org/0009-0002-1333-2197>
Nicolas Vuille-dit-Bille  <https://orcid.org/0009-0005-9515-3062>
Margot Visse-Mansiaux  <https://orcid.org/0000-0002-7844-3413>
Frank Liebisch  <https://orcid.org/0000-0003-0000-7491>
Helge Aasen  <https://orcid.org/0000-0003-4343-0476>
Lukas Roth  <https://orcid.org/0000-0003-1435-9535>
Achim Walter  <https://orcid.org/0000-0001-7753-9643>
Juan M. Herrera  <https://orcid.org/0000-0002-9398-7224>

REFERENCES

- Aasen, H., Kirchgessner, N., Walter, A., & Liebisch, F. (2020). PhenoCams for field phenotyping: Using very high temporal resolution digital repeated photography to investigate interactions of growth, phenology, and harvest traits. *Frontiers in Plant Science*, 11, 1–16. <https://doi.org/10.3389/fpls.2020.00593>
- Adamsen, F. J., Pinter, P. J., Barnes, E. M., LaMorte, R. L., Wall, G. W., Leavitt, S. W., & Kimball, B. A. (1999). Measuring wheat senescence

- with a digital camera. *Crop Science*, 39, 719–724. <https://doi.org/10.2135/cropsci1999.0011183X003900030019x>
- Ahrends, H. E., Etzold, S., Kutsch, W. L., Stöckli, R., Brügger, R., Jeanneret, F., Wanner, H., Buchmann, N., & Eugster, W. (2009). Tree phenology and carbon dioxide fluxes: Use of digital photography for process-based interpretation at the ecosystem scale. *Climate Research*, 39, 261–274.
- Anderegg, J., Tschurr, F., Kirchgessner, N., Treier, S., Schmucki, M., Streit, B., & Walter, A. (2023). On-farm evaluation of UAV-based aerial imagery for season-long weed monitoring under contrasting management and pedoclimatic conditions in wheat. *Computers and Electronics in Agriculture*, 204, 107558. <https://doi.org/10.1016/j.compag.2022.107558>
- Anderegg, J., Yu, K., Aasen, H., Walter, A., Liebisch, F., & Hund, A. (2020). Spectral vegetation indices to track senescence dynamics in diverse wheat germplasm. *Frontiers in Plant Science*, 10, 1–20. <https://doi.org/10.3389/fpls.2019.01749>
- Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., & Cairns, J. E. (2018). Translating high-throughput phenotyping into genetic gain. *Trends in Plant Science*, 23, 451–466. <https://doi.org/10.1016/j.tplants.2018.02.001>
- Asseng, S., Ewert, F., Martre, P., Rötter, R. P., Lobell, D. B., Cammarano, D., Kimball, B. A., Ottman, M. J., Wall, G. W., White, J. W., Reynolds, M. P., Alderman, P. D., Prasad, P. V. V., Aggarwal, P. K., Anothai, J., Basso, B., Biernath, C., Challinor, A. J., De Sanctis, G., & Zhu, Y. (2015). Rising temperatures reduce global wheat production. *Nature Climate Change*, 5, 143–147. <https://doi.org/10.1038/nclimate2470>
- Asseng, S., Ewert, F., Rosenzweig, C., Jones, J. W., Hatfield, J. L., Ruane, A. C., Boote, K. J., Thorburn, P. J., Rötter, R. P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal, P. K., Angulo, C., Bertuzzi, P., Biernath, C., Challinor, A. J., Doltra, J., ... Wolf, J. (2013). Uncertainty in simulating wheat yields under climate change. *Nature Climate Change*, 3, 827–832. <https://doi.org/10.1038/nclimate1916>
- Barnes, E. M., Clarke, T. R., Richards, S. E., Colaizzi, P. D., Haberland, J., Kostrzewski, M., Waller, P., Choi, C., Riley, E., & Thompson, T. (2000). Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. In *Proceedings of the Fifth International Conference on Precision Agriculture* (pp. 1356). ASA-CSSA-SSSA.
- Barreto, C. A. V., Das Graças Dias, K. O., De Sousa, I. C., Azevedo, C. F., Nascimento, A. C. C., Guimarães, L. J. M., Guimarães, C. T., Pastina, M. M., & Nascimento, M. (2024). Genomic prediction in multi-environment trials in maize using statistical and machine learning methods. *Scientific Reports*, 14, 1062. <https://doi.org/10.1038/s41598-024-51792-3>
- Baxter, S. (2007). *World reference base for soil resources* (World Soil Resources Report 103). Food and Agriculture Organization of the United Nations(2006) <https://doi.org/10.1017/S0014479706394902>
- Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., Gnyp, M. L., & Bareth, G. (2015). Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *International Journal of Applied Earth Observation and Geoinformation*, 39, 79–87. <https://doi.org/10.1016/j.jag.2015.02.012>
- Bhatti, M. T., Gilani, H., Ashraf, M., Iqbal, M. S., & Munir, S. (2024). Field validation of NDVI to identify crop phenological signatures. *Precision Agriculture*, 25, 2245–2270. <https://doi.org/10.1007/s11119-024-10165-6>
- Birth, G. S., & McVey, G. R. (1968). Measuring the color of growing turf with a reflectance spectrophotometer. *Agronomy Journal*, 60, 640–643. <https://doi.org/10.2134/agronj1968.00021962006000060016x>
- Blackshaw, R. E., Molnar, L. J., & Moyer, J. R. (2010). Suitability of legume cover crop-winter wheat intercrops on the semi-arid Canadian prairies. *Canadian Journal of Plant Science*, 90, 479–488. <https://doi.org/10.4141/CJPS10006>
- Bradski, G., & Kaehler, A. (2000). OpenCV. *Dr. Dobb's Journal of Software Tools*, 120, 122–125.
- Brocks, S., & Bareth, G. (2018). Estimating barley biomass with crop surface models from oblique RGB imagery. *Remote Sensing*, 10, 268. <https://doi.org/10.3390/rs10020268>
- Brocks, S., Bendig, J., & Bareth, G. (2016). Toward an automated low-cost three-dimensional crop surface monitoring system using oblique stereo imagery from consumer-grade smart cameras. *Journal of Applied Remote Sensing*, 10, 046021. <https://doi.org/10.1117/1.JRS.10.046021>
- Browning, D. M., Karl, J. W., Morin, D., Richardson, A. D., & Tweedie, C. E. (2017). Phenocams bridge the gap between field and satellite observations in an arid grassland ecosystem. *Remote Sensing*, 9, 1071. <https://doi.org/10.3390/rs9101071>
- Burkart, A., Hecht, V. L., Kraska, T., & Rascher, U. (2018). Phenological analysis of unmanned aerial vehicle based time series of barley imagery with high temporal resolution. *Precision Agriculture*, 19, 134–146. <https://doi.org/10.1007/s11119-017-9504-y>
- Cao, X., Liu, Y., Yu, R., Han, D., & Su, B. (2021). A comparison of UAV RGB and multispectral imaging in phenotyping for stay green of wheat population. *Remote Sensing*, 13, 5173. <https://doi.org/10.3390/rs13245173>
- Carrascal, L. M., Galván, I., & Gordo, O. (2009). Partial least squares regression as an alternative to current regression methods used in ecology. *Oikos*, 118, 681–690. <https://doi.org/10.1111/j.1600-0706.2008.16881.x>
- Chapman, E. A., Orford, S., Lage, J., & Griffiths, S. (2021). Capturing and selecting senescence variation in wheat. *Frontiers in Plant Science*, 12, 638738. <https://doi.org/10.3389/fpls.2021.638738>
- Christopher, J. T., Christopher, M. J., Borrell, A. K., Fletcher, S., & Chenu, K. (2016). Stay-green traits to improve wheat adaptation in well-watered and water-limited environments. *Journal of Experimental Botany*, 67, 5159–5172. <https://doi.org/10.1093/jxb/erw276>
- Christopher, J. T., Veyradier, M., Borrell, A. K., Harvey, G., Fletcher, S., & Chenu, K. (2014). Phenotyping novel stay-green traits to capture genetic variation in senescence dynamics. *Functional Plant Biology*, 41, 1035–1048. <https://doi.org/10.1071/FP14052>
- Costa-Neto, G., Crespo-Herrera, L., Fradgley, N., Gardner, K., Bentley, A. R., Dreisigacker, S., Fritsche-Neto, R., Montesinos-López, O. A., & Crossa, J. (2023). Envirome-wide associations enhance multi-year genome-based prediction of historical wheat breeding data. *G3 Genes Genomes Genetics*, 13, jkac313. <https://doi.org/10.1093/g3journal/jkac313>
- de Cárcer, P. S., Sinaj, S., Santonja, M., Fossati, D., & Jeangros, B. (2019). Long-term effects of crop succession, soil tillage and climate on wheat yield and soil properties. *Soil and Tillage Research*, 190, 209–219. <https://doi.org/10.1016/j.still.2019.01.012>
- Darst, B. F., Malecki, K. C., & Engelman, C. D. (2018). Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC Genetics*, 19, 65. <https://doi.org/10.1186/s12863-018-0633-8>

- Deery, D. M., Rebetzke, G. J., Jimenez-Berni, J. A., James, R. A., Condon, A. G., Bovill, W. D., Hutchinson, P., Scarrow, J., Davy, R., & Furbank, R. T. (2016). Methodology for high-throughput field phenotyping of canopy temperature using airborne thermography. *Frontiers in Plant Science*, 7, 1808. <https://doi.org/10.3389/fpls.2016.01808>
- Dong, B., Zheng, X., Liu, H., Able, J. A., Yang, H., Zhao, H., Zhang, M., Qiao, Y., Wang, Y., & Liu, M. (2017). Effects of drought stress on pollen sterility, grain yield, abscisic acid and protective enzymes in two winter wheat cultivars. *Frontiers in Plant Science*, 8, 1008. <https://doi.org/10.3389/fpls.2017.01008>
- Elmerich, C., Faucon, M. P., Garcia, M., Jeanson, P., Boulch, G., & Lange, B. (2023). Envirotyping to control genotype environment interactions for efficient soybean breeding. *Field Crops Research*, 303, 109113. <https://doi.org/10.1016/j.fcr.2023.109113>
- Farooq, M., Hussain, M., & Siddique, K. H. M. (2014). Drought stress in wheat during flowering and grain-filling periods. *Critical Reviews in Plant Sciences*, 33, 331–349. <https://doi.org/10.1080/07352689.2014.875291>
- Ferrigo, D., Raiola, A., & Causin, R. (2016). Fusarium toxins in cereals: Occurrence, legislation, factors promoting the appearance and their management. *Molecules*, 21, 627. <https://doi.org/10.3390/molecules21050627>
- Fischer, A., & Feller, U. (1994). Senescence and protein degradation in leaf segments of young winter wheat: Influence of leaf age. *Journal of Experimental Botany*, 45, 103–109. <https://doi.org/10.1093/jxb/45.1.103>
- Ford, M. A., & Thorne, G. N. (1975). Effects of variation in temperature and light intensity at different times on growth and yield of spring wheat. *Annals of Applied Biology*, 80, 283–299. <https://doi.org/10.1111/j.1744-7348.1975.tb01634.x>
- Gao, X., Hu, C. H., Li, H. Z., Yao, Y. J., Meng, M., Dong, J., Zhao, W. C., Chen, Q. J., & Li, X. Y. (2013). Factors affecting pre-harvest sprouting resistance in wheat (*Triticum aestivum* L.): A review. *Journal of Animal & Plant Science*, 23, 556–565.
- GDAL/OGR Contributors. (2024). *Geospatial data abstraction software library*. Open Source Geospatial Foundation.
- Gillespie, A. R., Kahle, A. B., & Walker, R. E. (1987). Color enhancement of highly correlated images. II. Channel ratio and “chromaticity” transformation techniques. *Remote Sensing of Environment*, 22, 343–365. [https://doi.org/10.1016/0034-4257\(87\)90088-5](https://doi.org/10.1016/0034-4257(87)90088-5)
- Gitelson, A. A., Kaufman, Y. J., Stark, R., & Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment*, 80, 76–87. [https://doi.org/10.1016/S0034-4257\(01\)00289-9](https://doi.org/10.1016/S0034-4257(01)00289-9)
- Gitelson, A. A., & Merzlyak, M. N. (1994). Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. *Journal of Photochemistry and Photobiology B: Biology*, 22, 247–252.
- Gitelson, A. A., Merzlyak, M. N., & Chivkunova, O. B. (2001). Optical properties and nondestructive estimation of anthocyanin content in plant leaves. *Photochemistry and Photobiology*, 74, 38–45.
- Google. (2024)(2024). *Cloud storage—Prices*. <https://cloud.google.com/storage/pricing?hl=de>
- Gracia-Romero, A., Kefauver, S. C., Vergara-Díaz, O., Zaman-Allah, M. A., Prasanna, B. M., Cairns, J. E., & Araus, J. L. (2017). Comparative performance of ground vs. aerially assessed RGB and multispectral indices for early-growth evaluation of maize performance under phosphorus fertilization. *Frontiers in Plant Science*, 8, 1–13. <https://doi.org/10.3389/fpls.2017.02004>
- Graham, E. A., Riordan, E. C., Yuen, E. M., Estrin, D., & Rundel, P. W. (2010). Public internet-connected cameras used as a cross-continental ground-based plant phenology monitoring system. *Global Change Biology*, 16, 3014–3023. <https://doi.org/10.1111/j.1365-2486.2010.02164.x>
- Gregorutti, B., Michel, B., & Saint-Pierre, P. (2016). Correlation and variable importance in random forests. *Statistics and Computing*, 27, 659–678. <https://doi.org/10.1007/s11222-016-9646-1>
- Guo, Y., Chen, S., Fu, Y. H., Xiao, Y., Wu, W., Wang, H., & Beurs, K. (2022). Comparison of multi-methods for identifying maize phenology using PhenoCams. *Remote Sensing*, 14, 244. <https://doi.org/10.3390/rs14020244>
- Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., & Smith, N. J. (2020). Array programming with NumPy. *Nature*, 585, 357–362.
- Hasan, U., Sawut, M., & Chen, S. (2019). Estimating the leaf area index of winter wheat based on unmanned aerial vehicle RGB-image parameters. *Sustainability*, 11, 1–11. <https://doi.org/10.3390/su11236829>
- Hershman, D. E. (2011). *Black “sooty” head mold of wheat* [Fact Sheet PPFS-AG-SG-07]. <https://plantpathology.ca.uky.edu/files/ppfs-ag-sg-07.pdf>
- Holzschläger, A., Fossati, D., Hiltbrunner, J., & Fuhrer, J. (2015). Spatial and temporal trends in agro-climatic limitations to production potentials for grain maize and winter wheat in Switzerland. *Regional Environmental Change*, 15, 109–122. <https://doi.org/10.1007/s10113-014-0627-7>
- Huang, Q., Wang, X., Gao, Q., Carraro, A., Sozzi, M., & Marinello, F. (2024). Indicators to digitization footprint and how to get digitization footprint (Part 2). *Computers and Electronics in Agriculture*, 224, 109206. <https://doi.org/10.1016/j.compag.2024.109206>
- Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25, 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)
- Huete, A. R., Didan, K., Miura, T., Rodriguez, E., Gao, X., & Ferreira, L. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83, 195–213. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2)
- Hufkens, K., Keenan, T. F., Flanagan, L. B., Scott, R. L., Bernacchi, C. J., Joo, E., Brunsell, N. A., Verfaillie, J., & Richardson, A. D. (2016). Productivity of North American grasslands is increased under future climate scenarios despite rising aridity. *Nature Climate Change*, 6, 710–714.
- Hufkens, K., Melaas, E. K., Mann, M. L., Foster, T., Ceballos, F., Robles, M., & Kramer, B. (2019). Monitoring crop phenology using a smartphone based near-surface remote sensing approach. *Agricultural and Forest Meteorology*, 265, 327–337. <https://doi.org/10.1016/j.agrformet.2018.11.002>
- Hund, A., Kronenberg, L., Anderegg, J., Yu, K., & Walter, A. (2019). Non-invasive field phenotyping of cereal development. In F. Ordon (Ed.), *Burleigh dodds series in agricultural science* (pp. 249–292). Burleigh Dodds Science Publishing.
- Hunt, E. R., Daughtry, C. S. T., Eitel, J. U. H., & Long, D. S. (2011). Remote sensing leaf chlorophyll content using a visible band index. *Agronomy Journal*, 103, 1090–1099. <https://doi.org/10.2134/agronj2010.0395>
- Hunt, E. R., Doraiswamy, P. C., McMurtrey, J. E., Daughtry, C. S. T., Perry, E. M., & Akhmedov, B. (2013). A visible band index for remote sensing leaf chlorophyll content at the canopy scale.

- International Journal of Applied Earth Observation and Geoinformation*, 21, 103–112. <https://doi.org/10.1016/j.jag.2012.07.020>
- Hörtensteiner, S. (2006). Chlorophyll degradation during senescence. *Annual Review of Plant Biology*, 57, 55–77. <https://doi.org/10.1146/annurev.arplant.57.032905.105212>
- Ide, R., & Oguma, H. (2010). Use of digital cameras for phenological observations. *Ecological Informatics*, 5, 339–347. <https://doi.org/10.1016/j.ecoinf.2010.07.002>
- Jensen, T., Apan, A., Young, F., & Zeller, L. (2007). Detecting the attributes of a wheat crop using digital imagery acquired from a low-altitude platform. *Computers and Electronics in Agriculture*, 59, 66–77. <https://doi.org/10.1016/j.compag.2007.05.004>
- Jia, J., Xu, M., Bei, S., Zhang, H., Xiao, L., Gao, Y., Zhang, Y., Sai, L., Xue, L., Lei, J., & Qiao, X. (2021). Impact of reduced light intensity on wheat yield and quality: Implications for agroforestry systems. *Agroforestry Systems*, 95, 1689–1701. <https://doi.org/10.1007/s10457-021-00668-w>
- Jin, X., Liu, S., Baret, F., Hemerlé, M., & Comar, A. (2017). Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. *Remote Sensing of Environment*, 198, 105–114. <https://doi.org/10.1016/j.rse.2017.06.007>
- Jordan, C. F. (1969). Derivation of leaf-area index from quality of light on the forest floor. *Ecology*, 50, 663–666. <https://doi.org/10.2307/1936256>
- Joshi, A. K., Kumari, M., Singh, V. P., Reddy, C. M., Kumar, S., Rane, J., & Chand, R. (2007). Stay green trait: Variation, inheritance and its association with spot blotch resistance in spring wheat (*Triticum aestivum* L.). *Euphytica*, 153, 59–71. <https://doi.org/10.1007/s10681-006-9235-z>
- Kahiluoto, H., Kaseva, J., Balek, J., Olesen, J. E., Ruiz-Ramos, M., Gobin, A., Kersebaum, K. C., Takáč, J., Ruget, F., Ferrise, R., Bezak, P., Capellades, G., Dibari, C., Mäkinen, H., Nendel, C., Ventrella, D., Rodríguez, A., Bindí, M., & Trnka, M. (2019). Decline in climate resilience of European wheat. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 123–128. <https://doi.org/10.1073/pnas.1804387115>
- Kavaliauskas, A., Žydelis, R., Castaldi, F., Auškalnienė, O., & Povilaitis, V. (2023). Predicting maize theoretical methane yield in combination with ground and UAV remote data using machine learning. *Plants*, 12, 1823. <https://doi.org/10.3390/plants12091823>
- Kawashima, S. (1998). An algorithm for estimating chlorophyll content in leaves using a video camera. *Annals of Botany*, 81, 49–54. <https://doi.org/10.1006/anbo.1997.0544>
- Keenan, T. F., Darby, B., Felts, E., Sonnentag, O., Friedl, M. A., Hufkens, K., O'Keefe, J., Klosterman, S., Munger, J. W., Toomey, M., & Richardson, A. D. (2014). Tracking forest phenology and seasonal physiology using digital repeat photography: A critical assessment. *Ecological Applications*, 24, 1478–1489. <https://doi.org/10.1890/13-0652.1>
- Kipp, S., Mistele, B., & Schmidhalter, U. (2014). Identification of stay-green and early senescence phenotypes in high-yielding winter wheat, and their relationship to grain yield and grain protein concentration using high-throughput phenotyping techniques. *Functional Plant Biology*, 41, 227. <https://doi.org/10.1071/FP13221>
- Klosterman, S. T., Hufkens, K., Gray, J. M., Melaas, E., Sonnentag, O., Lavine, I., Mitchell, L., Norman, R., Friedl, M. A., & Richardson, A. D. (2014). Evaluating remote sensing of deciduous forest phenology at multiple spatial scales using PhenoCam imagery. *Biogeosciences*, 11, 4305–4320. <https://doi.org/10.5194/bg-11-4305-2014>
- Kurc, S., & Benton, L. (2010). Digital image-derived greenness links deep soil moisture to carbon uptake in a creosotebush-dominated shrubland. *Journal of Arid Environments*, 74, 585–594.
- Lancashire, P. D., Bleiholder, H., Boom, T. V. D., Langelüddeke, P., Stauss, R., Weber, E., & Witzengerber, A. (1991). A uniform decimal code for growth stages of crops and weeds. *Annals of Applied Biology*, 119, 561–601. <https://doi.org/10.1111/j.1744-7348.1991.tb04895.x>
- Langer, S. M., Longin, C. F. H., & Würschum, T. (2014). Flowering time control in European winter wheat. *Frontiers in Plant Science*, 5, 1–11. <https://doi.org/10.3389/fpls.2014.00537>
- Lee, S. Y., Mediani, A., Maulidiani, M., Khatib, A., Ismail, I. S., Zawawi, N., & Abas, F. (2017). Comparison of partial least squares and random forests for evaluating relationship between phenolics and bioactivities of *Neptunia oleracea*. *Journal of the Science of Food and Agriculture*, 98, 240–252. <https://doi.org/10.1002/jsfa.8462>
- Li, W., Li, D., Liu, S., Baret, F., Ma, Z., He, C., Warner, T. A., Guo, C., Cheng, T., Zhu, Y., Cao, W., & Yao, X. (2023). RSARE: A physically-based vegetation index for estimating wheat green LAI to mitigate the impact of leaf chlorophyll content and residue-soil background. *ISPRS Journal of Photogrammetry and Remote Sensing*, 200, 138–152. <https://doi.org/10.1016/j.isprsjprs.2023.05.012>
- Liu, Y., Bachofen, C., Wittwer, R., Duarte, G. S., Sun, Q., Klaus, V. H., & Buchmann, N. (2022). Using PhenoCams to track crop phenology and explain the effects of different cropping systems on yield. *Agricultural Systems*, 195, 103306. <https://doi.org/10.1016/j.agsy.2021.103306>
- Longchamps, L., & Philpot, W. (2023). Full-season crop phenology monitoring using two-dimensional normalized difference pairs. *Remote Sensing*, 15, 5565. <https://doi.org/10.3390/rs15235565>
- Lorenz, K. (1986). Effects of blackpoint on grain composition and baking quality of New Zealand wheat. *New Zealand Journal of Agricultural Research*, 29, 711–718. <https://doi.org/10.1080/00288233.1986.10430468>
- Louhaichi, M., Borman, M. M., & Johnson, D. E. (2001). Spatially located platform and aerial photography for documentation of grazing impacts on wheat. *Geocarto International*, 16, 65–70. <https://doi.org/10.1080/10106040108542184>
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60, 91–110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
- Lu, N., Zhou, J., Han, Z., Li, D., Cao, Q., Yao, X., Tian, Y., Zhu, Y., Cao, W., & Cheng, T. (2019). Improved estimation of aboveground biomass in wheat from RGB imagery and point cloud data acquired with a low-cost unmanned aerial vehicle system. *Plant Methods*, 15, 17. <https://doi.org/10.1186/s13007-019-0402-3>
- Mahrookashani, A., Siebert, S., Hüging, H., & Ewert, F. (2017). Independent and combined effects of high temperature and drought stress around anthesis on wheat. *Journal of Agronomy and Crop Science*, 203, 453–463. <https://doi.org/10.1111/jac.12218>
- Mao, W., Wang, Y., & Wang, Y. (2003). Real-time detection of between-row weeds using machine vision. In *Proceedings of 2003 ASAE annual meeting*. American Society of Agricultural and Biological Engineers.
- Marinello, F. (2023). Digitization footprint. In Q. Zhang (Ed.), *Encyclopedia of digital agricultural technologies* (pp. 356–363). Springer International Publishing.
- McMaster, G. (1997). Growing degree-days: One equation, two interpretations. *Agricultural and Forest Meteorology*, 87, 291–300. [https://doi.org/10.1016/S0168-1923\(97\)00027-0](https://doi.org/10.1016/S0168-1923(97)00027-0)

- Mehmood, T., Liland, K. H., Snipen, L., & Sæbø, S. (2012). A review of variable selection methods in partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 118, 62–69. <https://doi.org/10.1016/j.chemolab.2012.07.010>
- Merzlyak, M. N., Gitelson, A. A., Chivkunova, O. B., & Rakitin, V. Y. (1999). Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiologia Plantarum*, 106, 135–141. <https://doi.org/10.1034/j.1399-3054.1999.106119.x>
- Mevik, B. H., & Wehrens, R. (2007). The pls package: Principal component and partial least squares regression in R. *Journal of Statistical Software*, 18, 1–23. <https://doi.org/10.18637/jss.v018.i02>
- Meyer, G. E., Mehta, T., Kocher, M. F., Mortensen, D. A., & Samal, A. (1998). Textural imaging and discriminant analysis for distinguishing weeds for spot spraying. *Transactions of the ASAE*, 41, 1189–1197.
- Meyer, G. E., & Neto, J. C. (2008). Verification of color vegetation indices for automated crop imaging applications. *Computers and Electronics in Agriculture*, 63, 282–293. <https://doi.org/10.1016/j.compag.2008.03.009>
- Migliavacca, M., Galvagno, M., Cremonese, E., Rossini, M., Meroni, M., Sonnentag, O., Cogliati, S., Manca, G., Diotri, F., Busetto, L., Cescatti, A., Colombo, R., Fava, F., Morra di Cella, U., Pari, E., Siniscalco, C., & Richardson, A. D. (2011). Using digital repeat photography and eddy covariance data to model grassland phenology and photosynthetic CO₂ uptake. *Agricultural and Forest Meteorology*, 151, 1325–1337. <https://doi.org/10.1016/j.agrformet.2011.05.012>
- Montazeaud, G., Karatoğma, H., Öztürk, I., Roumet, P., Ecartot, M., Crossa, J., Özer, E., Özdemir, F., & Lopes, M. S. (2016). Predicting wheat maturity and stay-green parameters by modeling spectral reflectance measurements and their contribution to grain yield under rainfed conditions. *Field Crops Research*, 196, 191–198. <https://doi.org/10.1016/j.fcr.2016.06.021>
- Moore, C. E., Brown, T., Keenan, T. F., Duursma, R. A., Van Dijk, A. I. J. M., Beringer, J., Culvenor, D., Evans, B., Huete, A., Hutley, L. B., Maier, S., Restrepo-Coupe, N., Sonnentag, O., Specht, A., Taylor, J. R., Van Gorsel, E., & Liddell, M. J. (2016). Reviews and syntheses: Australian vegetation phenology: new insights from satellite remote sensing and digital repeat photography. *Biogeosciences*, 13, 5085–5102. <https://doi.org/10.5194/bg-13-5085-2016>
- Morlin Carneiro, F., Angeli Furlani, C. E., Zerbato, C., Candida De Menezes, P., Da Silva Gírio, L. A., & Freire De Oliveira, M. (2020). Comparison between vegetation indices for detecting spatial and temporal variabilities in soybean crop using canopy sensors. *Precision Agriculture*, 21, 979–1007. <https://doi.org/10.1007/s11119-019-09704-3>
- Młodzińska, E. (2009). Survey of plant pigments: Molecular and environmental determinants of plant colors. *Acta Biologica Cracoviensia Series Botanica*, 51, 7–16.
- Naito, H., Ogawa, S., Valencia, M. O., Mohri, H., Urano, Y., Hosoi, F., Shimizu, Y., Chavez, A. L., Ishitani, M., Selvaraj, M. G., & Omasa, K. (2017). Estimating rice yield related traits and quantitative trait loci analysis under different nitrogen treatments using a simple tower-based field phenotyping system with modified single-lens reflex cameras. *ISPRS Journal of Photogrammetry and Remote Sensing*, 125, 50–62. <https://doi.org/10.1016/j.isprsjprs.2017.01.010>
- Nicodemus, F. E. (1977). *Geometrical considerations and nomenclature for reflectance* (NBS Monograph 160). US Department of Commerce, National Bureau of Standards. <https://doi.org/10.6028/NBS.MONO.160>
- Oakey, H., Verbyla, A., Pitchford, W., Cullis, B., & Kuchel, H. (2006). Joint modeling of additive and non-additive genetic line effects in single field trials. *Theoretical and Applied Genetics*, 113, 809–819. <https://doi.org/10.1007/s00122-006-0333-z>
- Padfield, D., & Matheso, G. (2020). Package nls. multstart. CRAN.
- Parvej, M. R., Holshouser, D. L., Kratochvil, R. J., Whaley, C. M., Dunphy, E. J., Roth, G. W., & Faé, G. S. (2020). Early high-moisture wheat harvest improves double-crop system: I. Wheat yield and quality. *Crop Science*, 60, 2633–2649. <https://doi.org/10.1002/csc2.20172>
- Pask, A. J. D., Pietragalla, J., Mullan, D. M., & Reynolds, M. P. (2012). *Physiological breeding II: A field guide to wheat phenotyping*. CIMMYT.
- Poursafar, A., Ghosta, Y., & Javan-Nikkhah, M. (2016). A taxonomic study on *Stemphylium* species associated with black (sooty) head mold of wheat and barley in Iran. *Mycologia Iranica*, 3, 99–109. <https://doi.org/10.22043/mi.2017.26183>
- QGIS Development Team. (2022). QGIS geographic information system. <https://qgis.org>
- R Development Core Team. (2022). *R: A language and environment for statistical computing*. CRAN. <http://www.r-project.org>
- Reynolds, M. P., Chapman, S. C., Crespo-Herrera, L., Molero, G., Mondal, S., Pequeno, D. N., Pinto, F., Pinera-Chavez, F. J., Poland, J., Rivera-Amado, C., Saint Pierre, C., & Sukumaran, S. (2020). Breeder friendly phenotyping. *Plant Science*, 295, 110396. <https://doi.org/10.1016/j.plantsci.2019.110396>
- Richardson, A. D., Braswell, B. H., Hollinger, D. Y., Jenkins, J. P., & Ollinger, S. V. (2009). Near-surface remote sensing of spatial and temporal variation in canopy phenology. *Ecological Applications*, 19, 1417–1428.
- Richardson, A. D., Hufkens, K., Milliman, T., Aubrecht, D. M., Chen, M., Gray, J. M., Johnston, M. R., Keenan, T. F., Klosterman, S. T., Kosmala, M., Melaas, E. K., Friedl, M. A., & Froking, S. (2018). Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery. *Scientific Data*, 5, 180028. <https://doi.org/10.1038/sdata.2018.28>
- Richardson, A. D., Jenkins, J. P., Braswell, B. H., Hollinger, D. Y., Ollinger, S. V., & Smith, M. L. (2007). Use of digital webcam images to track spring green-up in a deciduous broadleaf forest. *Oecologia*, 152, 323–334. <https://doi.org/10.1007/s00442-006-0657-z>
- Richardson, A. D., Keenan, T. F., Migliavacca, M., Ryu, Y., Sonnentag, O., & Toomey, M. (2013). Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agricultural and Forest Meteorology*, 169, 156–173.
- Rife, T. W., & Poland, J. A. (2014). Field book: An open-source application for field data collection on android. *Crop Science*, 54, 1624–1627.
- Ripley, B., & Ramsey, J. (2024). Package ‘pspline’. CRAN. <https://cran.r-project.org/web/packages/pspline/>
- Rodríguez-Álvarez, M. X., Boer, M. P., van Eeuwijk, F., & Eilers, P. H. C. (2018). Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spatial Statistics*, 23, 52–71. <https://doi.org/10.1016/j.spasta.2017.10.003>
- Rogger, J., Hund, A., Fossati, D., & Holzkämper, A. (2021). Can Swiss wheat varieties escape future heat stress?. *European Journal of Agronomy*, 131, 126394. <https://doi.org/10.1016/j.eja.2021.126394>
- Roth, L., Rodríguez-Álvarez, M. X., van Eeuwijk, F., Piepho, H. P., & Hund, A. (2021). Phenomics data processing: A plot-level model for repeated measurements to extract the timing of key stages and quantities at defined time points. *Field Crops Research*, 274, 108314. <https://doi.org/10.1016/j.fcr.2021.108314>

- Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *NASA Special Publication*, 351, 309.
- Sadeghi-Tehran, P., Sabermanesh, K., Virlet, N., & Hawkesford, M. J. (2017). Automated method to determine two critical growth stages of wheat: Heading and flowering. *Frontiers in Plant Science*, 8, 1–14. <https://doi.org/10.3389/fpls.2017.00252>
- Schaepman-Strub, G., Schaepman, M., Painter, T., Dangel, S., & Martonchik, J. (2006). Reflectance quantities in optical remote sensing—definitions and case studies. *Remote Sensing of Environment*, 103, 27–42. <https://doi.org/10.1016/j.rse.2006.03.002>
- Segal, D. (1982). Theoretical basis for differentiation of ferric-iron bearing minerals, using Landsat MSS data. In *Proceedings of Symposium for Remote Sensing of environment, 2nd thematic conference on remote sensing for exploratory geology* (p. 951). Office of Scientific and Technical Information, U.S. Department of Energy.
- Segarra, J., Rezzouk, F. Z., Aparicio, N., González-Torralba, J., Aranjuelo, I., Gracia-Romero, A., Araus, J. L., & Kefauver, S. C. (2023). Multiscale assessment of ground, aerial and satellite spectral data for monitoring wheat grain nitrogen content. *Information Processing in Agriculture*, 10, 504–522. <https://doi.org/10.1016/j.inpa.2022.05.004>
- Sellaro, R., Crepy, M., Trupkin, S., Karayekov, E., Buchovsky, A. S., Rossi, C., & Casal, J. J. (2010). Cryptochrome as a sensor of the blue/green ratio of natural radiation in Arabidopsis. *Plant Physiology*, 154, 401–409. <https://doi.org/10.1104/pp.110.160820>
- Sonnentag, O., Hufkens, K., Teshera-Sterne, C., Young, A. M., Friedl, M., Braswell, B. H., Milliman, T., O’Keefe, J., & Richardson, A. D. (2012). Digital repeat photography for phenological research in forest ecosystems. *Agricultural and Forest Meteorology*, 152, 159–177. <https://doi.org/10.1016/j.agrformet.2011.09.009>
- Sudhakar, P., Latha, P., & Reddy, P. (2016). Plant pigments. In P. Sudhakar, P. Latha, & P. Reddy (Eds.), *Phenotyping crop plants for physiological and biochemical traits* (pp. 121–127). Academic Press. <https://doi.org/10.1016/B978-0-12-804073-7.00015-6>
- Swiss Federal Council. (2013). Verordnung über die Direktzahlungen an die Landwirtschaft (Direktzahlungsverordnung, dzv).
- Tafesse, E. G., Warkentin, T. D., Shirliff, S., Noble, S., & Bueckert, R. (2022). Leaf pigments, surface wax and spectral vegetation indices for heat stress resistance in pea. *Agronomy*, 12, 739. <https://doi.org/10.3390/agronomy12030739>
- Tang, Z., Jin, Y., Alsina, M. M., McElrone, A. J., Bambach, N., & Kustas, W. P. (2022). Vine water status mapping with multispectral UAV imagery and machine learning. *Irrigation Science*, 40, 715–730. <https://doi.org/10.1007/s00271-022-00788-w>
- Taylor, S. D., & Browning, D. M. (2021). Classification of daily crop phenology in PhenoCams using deep learning and hidden Markov models. *Remote Sensing*, 14, 286.
- Treier, S., Herrera, J. M., Hund, A., Kirchgessner, N., Aasen, H., Walter, A., & Roth, L. (2024). Improving drone-based uncalibrated estimates of wheat canopy temperature in plot experiments by accounting for confounding factors in a multi-view analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 218, 721–741. <https://doi.org/10.1016/j.isprsjprs.2024.09.015>
- Tschurr, F., Roth, L., Storni, N., Zumsteg, O., Walter, A., & Anderegg, J. (2024). Temporal resolution trumps spectral resolution in UAV-based monitoring of cereal senescence dynamics. *Plant Methods*, 20, 188. <https://doi.org/10.1186/s13007-024-01308-x>
- 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](https://doi.org/10.1016/0034-4257(79)90013-0)
- van Rossum, G., & Drake, F. L. (2009). Python 3 reference manual. CreateSpace.
- Velumani, K., Madec, S., De Solan, B., Lopez-Lozano, R., Gillet, J., Labrosse, J., Jezequel, S., Comar, A., & Baret, F. (2020). An automatic method based on daily in situ images and deep learning to date wheat heading stage. *Field Crops Research*, 252, 107793. <https://doi.org/10.1016/j.fcr.2020.107793>
- Wang, D., Li, R., Zhu, B., Liu, T., Sun, C., & Guo, W. (2022). Estimation of wheat plant height and biomass by combining UAV imagery and elevation data. *Agriculture*, 13, 9. <https://doi.org/10.3390/agriculture13010009>
- Wang, H., Jia, G., Epstein, H. E., Zhao, H., & Zhang, A. (2020). Integrating a PhenoCam-derived vegetation index into a light use efficiency model to estimate daily gross primary production in a semi-arid grassland. *Agricultural and Forest Meteorology*, 288–289, 107983. <https://doi.org/10.1016/j.agrformet.2020.107983>
- WBF. (2021). Verordnung des WBF über Vermehrungsmaterial von Ackerpflanzen-, Futterpflanzen- und Gemüsearten (Technical Report). Das Eidgenössische Departement für Wirtschaft, Bildung und Forschung (WBF).
- Welbank, P. J., Witts, K. J., & Thorne, G. N. (1968). Effect of radiation and temperature on efficiency of cereal leaves during grain growth. *Annals of Botany*, 32, 79–95. <https://doi.org/10.1093/oxfordjournals.aob.a084201>
- Woebbecke, D. M., Meyer, G. E., Barga, K. V., & Mortensen, D. A. (1995). Color indices for weed identification under various soil, residue, and lighting conditions. *Transactions of the ASAE*, 38, 259–269. <https://doi.org/10.13031/2013.27838>
- Xu, X., Liu, L., Han, P., Gong, X., & Zhang, Q. (2022). Accuracy of vegetation indices in assessing different grades of grassland desertification from UAV. *International Journal of Environmental Research and Public Health*, 19, 16793. <https://doi.org/10.3390/ijerph192416793>
- Yan, S., -N., Yu, Z.-Y., Gao, W., Wang, X.-Y., Cao, J.-J., Lu, J., Ma, C. X., Chang, C., & Zhang, H. P. (2023). Dissecting the genetic basis of grain color and pre-harvest sprouting resistance in common wheat by association analysis. *Journal of Integrative Agriculture*, 22, 2617–2631. <https://doi.org/10.1016/j.jia.2023.04.017>
- Yang, H., Dong, B., Wang, Y., Qiao, Y., Shi, C., Jin, L., & Liu, M. (2020). Photosynthetic base of reduced grain yield by shading stress during the early reproductive stage of two wheat cultivars. *Scientific Reports*, 10, 14353. <https://doi.org/10.1038/s41598-020-71268-4>
- Yue, J., Yang, G., Tian, Q., Feng, H., Xu, K., & Zhou, C. (2019). Estimate of winter-wheat above-ground biomass based on UAV ultrahigh-ground-resolution image textures and vegetation indices. *ISPRS Journal of Photogrammetry and Remote Sensing*, 50, 226–244. <https://doi.org/10.1016/j.isprsjprs.2019.02.022>
- Zhang, J., Qiu, X., Wu, Y., Zhu, Y., Cao, Q., Liu, X., & Cao, W. (2021). Combining texture, color, and vegetation indices from fixed-wing UAS imagery to estimate wheat growth parameters using multivariate regression methods. *Computers and Electronics in Agriculture*, 185, 106138. <https://doi.org/10.1016/j.compag.2021.106138>
- Zhou, Y., Tang, H., Cheng, M. P., Dankwa, K. O., Chen, Z. X., Li, Z. Y., Gao, S., Liu, Y. X., Jiang, Q. T., Lan, X. J., Pu, Z. E., Wei, Y. M., Zheng, Y. L., Hickey, L. T., & Wang, J. R. (2017). Genome-wide association study for pre-harvest sprouting resistance in a large germplasm

collection of Chinese wheat landraces. *Frontiers in Plant Science*, 8, 401. <https://doi.org/10.3389/fpls.2017.00401>

Zhu, Y., Cao, Z., Lu, H., Li, Y., & Xiao, Y. (2016). In-field automatic observation of wheat heading stage using computer vision. *Biosystems Engineering*, 143, 28–41. <https://doi.org/10.1016/j.biosystemseng.2015.12.015>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Treier, S., Vuille-dit-Bille, N., Visse-Mansiaux, M., Liebisch, F., Aasen, H., Roth, L., Walter, A., & Herrera, J. M. (2025). Comparison of PhenoCams and drones for lean phenotyping of phenology and senescence of wheat genotypes in variety testing. *The Plant Phenome Journal*, 8, e70039. <https://doi.org/10.1002/ppj2.70039>