

Monitoring crop phenology using a smartphone based near-surface remote sensing approach

Koen Hufkens^{a,b,*}, Eli K. Melaas^c, Michael L. Mann^d, Timothy Foster^e, Francisco Ceballos^f, Miguel Robles^f, Berber Kramer^f

^a Department of Applied Ecology and Environmental Biology, Ghent University, Belgium

^b INRA, UMR ISPA, Villenave d'Ornon, France

^c Department of Earth & Environment, Boston University, Boston, MA, USA

^d Department of Geography, The George Washington University, Washington DC, USA

^e School of Mechanical, Aerospace and Civil Engineering, University of Manchester, Manchester, UK

^f Markets, Trade, and Institutions Division, International Food Policy Research Institute, Washington DC, USA

ARTICLE INFO

Keywords:

Crowdsourcing
Remote sensing
Winter wheat
Insurance
India

ABSTRACT

Smallholder farmers play a critical role in supporting food security in developing countries. Monitoring crop phenology and disturbances to crop growth is critical in strengthening farmers' ability to manage production risks. This study assesses the feasibility of using crowdsourced near-surface remote sensing imagery to monitor winter wheat phenology and identify damage events in northwest India. In particular, we demonstrate how streams of pictures of individual smallholder fields, taken using inexpensive smartphones, can be used to quantify important phenological stages in agricultural crops, specifically the wheat heading phase and how it can be used to detect lodging events, a major cause of crop damage globally. Near-surface remote sensing offers granular visual field data, providing detailed information on the timing of key developmental phases of winter wheat and crop growth disturbances that are not registered by common satellite remote sensing vegetation indices or national crop cut surveys. This illustrates the potential of near-surface remote sensing as a scalable platform for collecting high-resolution plot-specific data that can be used in supporting crop modeling, extension and insurance schemes to increase resilience to production risk and enhance food security in smallholder agricultural systems.

1. Introduction

Smallholder farmers are estimated to produce 30–50% of global food supply (Ricciardi et al., 2018), and account for over 80% of farms worldwide (Lowder et al., 2016). The success of smallholder farmers therefore is critical to global and national food security, in particular developing countries across SSA and Asia (IFAD and UNEP, 2013). Yet, many smallholder farmers are highly vulnerable to production risks posed by unpredictable weather events (e.g. droughts, floods, heat waves). Extreme weather shocks can force smallholder farmers to adopt costly coping mechanisms such as liquidating productive assets or reducing food consumption, limiting growth potential. Furthermore, the threat of negative weather events may also reduce the overall willingness to invest in climate-smart adaptive practices or other profitable technologies to improve long-term agricultural productivity and livelihoods (Barnett et al., 2008). Critically,

extreme weather events are likely to be exacerbated in the future with climate change, both in terms of their frequency and magnitude, in both tropical and sub-tropical regions where most smallholder farmers are concentrated (Auffhammer et al., 2012; Harvey et al., 2014; Morton, 2007). This highlights the urgent need for solutions to improve small producers' resilience to weather-related production risks throughout the growing season.

Extreme weather events do not affect all farmers equally. Differences in individual adaptive capacity (e.g. input usage, preventive and coping mechanisms) and the highly localized nature of some extreme weather events (e.g. heavy rainfall, hailstorms) can result in highly heterogeneous crop losses even within small geographic areas (Below et al., 2012; Fishman, 2016; IFAD and UNEP, 2013; Jain et al., 2015). Of particular interest to this study are variations in the timing of crop development or phenology among farmers in the same area, which may explain heterogeneity in crop damage. For example, heat stress has

* Corresponding author.

E-mail address: koen.hufkens@gmail.com (K. Hufkens).

damaging effects on yields, especially during crop anthesis (flowering). Elevated temperatures during this growth stage have been linked to reductions in maize and wheat yields of 3.8 and 5.5% respectively (Lobell et al., 2014), where temperatures above 30 °C during wheat anthesis cause complete sterility (Farooq et al., 2011; Saini and Aspinall, 1982). Similarly, lodging of wheat through the toppling of stems, which is a major cause of wheat damage globally, is known to be heavily influenced by occurrence of heavy rain and/or high wind only in later stages of wheat development (Berry et al., 2003; Gent and Kiyomoto, 1998; Vera et al., 2012). Consequently, widespread assessment of crop phenology between individual fields can provide critical inputs to support monitoring of agricultural productivity in smallholder systems and, in doing so, provide the necessary data to support targeted interventions to improve farmers' climate resilience and livelihoods (Auffhammer et al., 2012; Carletto et al., 2015; Harvey et al., 2014; Morton, 2007).

To date, efforts to monitor and map weather impacts on agricultural production have been limited by a lack of systematic field-level crop (yield) data collection and reporting (e.g. via farmer surveys) in most developing countries. Similarly, while advances have been made in the use of remote sensing imagery to map crop yields at field-scales (Azzari et al., 2017; Burke and Lobell, 2017; Jain et al., 2016), there remain significant challenges to the application of satellite-based methods for reliable crop yield assessment in smallholder systems due to the mismatch between the spatial resolution of openly-accessible satellite imagery, small plot sizes, highly heterogeneous cropping patterns, and high levels of cloud cover during crop growth seasons (Duncan et al., 2015; Jain et al., 2017; Mann and Warner, 2017).

Motivated by these challenges, there is potential to link satellite remote sensing with near-surface imagery (e.g. from unmanned aerial vehicles, smartphones, or fixed cameras) to support phenological monitoring and rapid assessments of field-level impacts of extreme weather events on smallholder agricultural productivity. Near-surface remote sensing techniques and, in particular, fixed cameras or "phenocams", have shown promise as a cost-effective way to bridge the gap between landscape-scale satellite remote sensing and a stand level monitoring of vegetation dynamics (Hufkens et al., 2012; Keenan et al., 2014; Klosterman et al., 2014). Phenocam-derived data products have served as calibration and validation data for large scale land-surface models of phenology for forest (Chen et al., 2016; Melaas et al., 2016) and grasslands (Hufkens et al., 2016). Given the strong dependence of crop yields on intra-seasonal weather variability and management practices, such data products could potentially enable significant improvements in field-level crop growth status and yield loss estimation in agricultural systems. Yet, despite this, there has so far been little research assessing the ability of phenocam-style in-situ data collection to support phenological monitoring in agricultural regions of developing countries.

In this study we address this important knowledge gap by quantifying the capacity of near-surface remote sensing imagery to accurately monitor crop phenology and physical disturbances to crop growth at field scales. Our analysis focuses on a pilot case study of winter wheat production in the states of Punjab and Haryana in northwest India, one of the world's most important smallholder agricultural systems in terms of wheat production. Near-surface imagery was collected for 508 individual one-acre fields using inexpensive smartphone cameras operated throughout the growing season by farmers.

2. Methods

2.1. Smallholder farmer site selection

The study focused on 50 villages in selected districts in India's Punjab (Fatehgarh, Ludhiana, Patiala) and Haryana (Fatehabad, Sirsa, Yamunanagar) states. These states are part of the Indo-Gangetic Plains, a zone of importance for crop production, that accounts for around 30%

of India's total wheat production. The climate across both states is mostly dominated by a hot arid steppe climate, according to the Koeppen-Geigen scale (Appendix Table 1). Regions in southern parts of both provinces are characterized by warm temperate conditions with dry winters and hot summers, while northern regions show arid hot desert climates. Most selected villages were located in a hot arid steppe climate.

Arable land in the region is extensively cropped and largely mechanized. Wheat crops grown during the (dry) Rabi season are irrigated fully, predominantly using groundwater from the underlying Indo-Gangetic Plain aquifer. While groundwater levels in the area have fallen significantly over recent decades due to over-abstraction for irrigation (Rodell et al., 2009), farmers retain widespread access to groundwater resources due to the subsidization of electricity for pumping by the Indian government (Shah et al., 2006).

FIELDS in Punjab and Haryana are typically double-cropped with rice or (to a smaller extent) cotton planted during the Kharif monsoon (June - October), and wheat planted in the Rabi season (October - March). Across sites, mean annual temperatures and precipitation are $24.45 \pm 0.2^\circ\text{C}$ and $712 \pm 187\text{ mm}$, respectively (Hijmans et al., 2005). Differences in the mean minimum temperature for the coldest or the mean maximum temperature for the warmest month are limited, with temperature differences among sites of 2.1°C and 1.9°C , respectively. Yet, precipitation between sites varies from 333 to 1219 mm.

Data were collected through the Picture-Based Insurance (PBI) project, a research initiative aiming to develop high-quality crop insurance products using a stream of smartphone pictures of insured fields for claim verification (Ceballos et al., 2018). All PBI protocols and tools, including those used in this paper, were developed by the authors to advance this objective. In our study we selected study sites following a clustered sampling approach. First, villages were randomly chosen within a radius of 5 km from each of 25 third-party weather stations in the area. Second, within each village, 15 farmers satisfying a number of criteria were randomly selected to be invited for study participation. The criteria required that the smallholder farmer (1) had less than 15 acres of operational land; (2) was in possession of an Android smartphone; and (3) was planning to grow wheat during the upcoming growing season (running from November - April). In total 508 farmers (or approximately 10 farmers per village) agreed to participate in the PBI study. For these farmers, the study team listed all plots on which the farmer was planning to grow wheat, and randomly selected one field for each farmer to be included in the study.

2.2. Near-surface remote sensing smartphone application and image acquisition protocol

To participate in the study, farmers were asked to take three repeat pictures per week of the insured site throughout the entire growing season, between 10:00 h and 14:00 h from approximately the same location as an initial northward oriented picture, and with approximately the same view angle every time. The outlined protocol adheres as closely as possible to the phenocam guidelines established by Sonnentag et al. (2012), describing the preferred camera settings and the orientation during image capture, while taking into consideration the constraints imposed by the use of a smartphone and the farmer's time availability.

To facilitate image acquisition, a custom Android smartphone application was developed. The application allowed the farmer to set up an observation site by taking an initial geo-referenced image of a field. Subsequent images were referenced relative to the initial image and location. The application included an interactive map to check if the farmer's position in the field corresponded to the location of the initial picture, and a visual aid in the form of a "ghost" image (a mildly transparent image of the initial picture) that allowed the farmer to align static features in the landscape (such as distant trees or structures) and one or two installed reference poles with the initial picture (Fig. 1a).

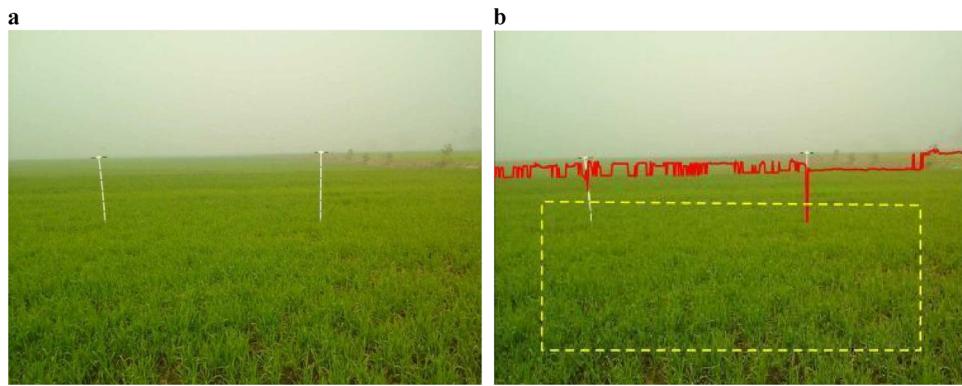


Fig. 1. a. Image as acquired by a smallholder farmer of a wheat field during the early growing season. The location of image acquisition location was constrained by the reference pole(s) in the image and the GPS location of the farmer's smartphone. b. The same image as processed to automatically delineate a region of interest (dashed yellow polygon), relative to the horizon as detected by a change point algorithm (red full line) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

The smartphone application helped ensure an almost identical view frame throughout the season. A fixed white balance between images was used to obtain comparable RGB ratios at a given site throughout the season. All pictures were uploaded to a server for further processing.

2.3. Near-surface remote sensing image processing

2.3.1. Site selection

Among all 508 farmers, participation was high, with producers providing approximately one image a week during the peak of the growing season (0.7 ± 0.3 images per week across the whole season, Appendix Fig. A1). To test our proposed methodology for monitoring crop phenology, we restrict our analysis to sites with picture coverage from the start of the growing season until harvest, and a minimum of 30 images roughly evenly spaced throughout the season. Time series not adhering to these standards were removed from the final dataset, so as to ensure our methodology could be tested against the best available data to minimize uncertainties introduced by lower levels of participation from some producers. The final dataset, used for all subsequent analyses, retains 52 sites (Appendix, Table 1), and includes 2,527 images. Henceforth, all references to “sites” indicate this selected subset. The retained sites recorded up to 2.9 images a week during the peak of the growing season and averaged 1.9 ± 0.8 images per week across the wheat growing season (Rabi). A map of the location of the selected smallholder farmer field sites is provided in Fig. 2.

2.3.2. Region of interest delineation and vegetation index calculation

A 90th percentile Green Chromatic Coordinates (Gcc) was calculated for a fixed region-of-interest (ROI) for every image to increase the stability of the greenness signal over time (Sonnenstag et al., 2012). The Gcc is defined as the ratio of the green Digital Number (DN) to the sum of all digital numbers (or image brightness), or:

$$\text{Gcc} = \text{Green DN} / (\text{Red DN} + \text{Green DN} + \text{Blue DN}) \quad (1)$$

where DN is defined as the constituent values of a given color in red-green-blue (RGB) color space.

The absence of a fixed vantage point made delineating a fixed ROI to calculate a color vegetation index (Eq. (1)) impossible. To automatically delineate a ROI on an image-by-image basis, a horizon detection algorithm was implemented. The horizon detection algorithm finds change points in the blue channel along the vertical axis of the images using the Pruned Exact Linear Time (PELT) method (Killick et al., 2011). These change points approximate the location of the horizon. Top corner points of a final trapezoid ROI were then defined by the median horizon locations for the left and right half of the image, padded by 10% of the image width and height. Similarly, the two bottom corner points were defined by padding the bottom and sides of the image by 10% of the image width and height. For all images the ROI was outlined. An example of an automatically delineated ROI is provided in Fig. 1b.

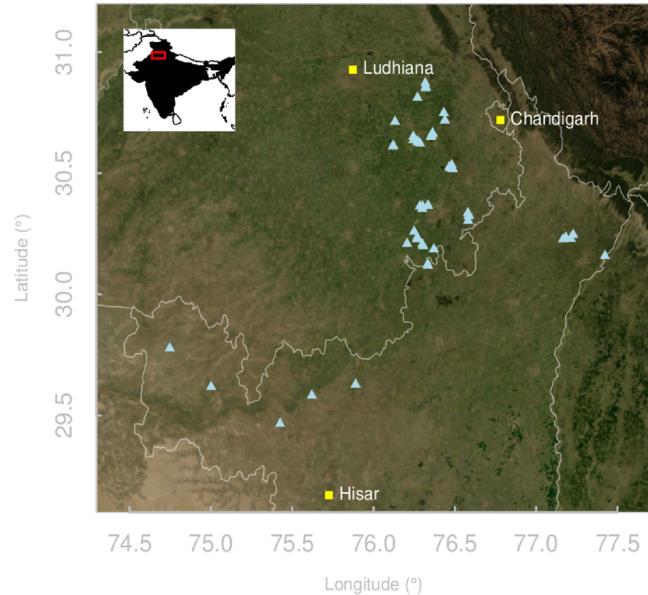


Fig. 2. Overview of the 52 selected sites (blue triangles) locations in the Punjab and Haryana provinces (marked with full grey lines), Northern India. Cities with more than 500 000 inhabitants are marked with yellow squares (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Individual Gcc time series were smoothed and normalized between 0 and 1 using a fitted loess model with a fixed span of 0.4. Normalization between 0 and 1 accounts for the fact that cameras are not radiometrically calibrated, therefore absolute Gcc values would be affected by differences in image rendering between devices, making the absolute values incomparable across sites. Normalized values allow us to create data that is meaningfully comparable. In addition, the smoothing routine removes occasional outliers but retains the overall trajectory of the Gcc time series. Due to uneven time steps in image acquisition we use daily interpolated smoothed data in all subsequent analyses.

2.4. Defining Gcc thresholds for phenology and lodging

Most remote sensing derived phenological indicators are defined as ad-hoc thresholds, such as a fixed amplitude (e.g. 50%) of a smoothed time series, or based on inflection points on fitted curves (MODIS and Landsat products) (Jönsson and Eklundh, 2004; Zhang et al., 2003). These thresholds often have limited relation to particular physiological or developmental stages. However, in wheat and other crops, the timing of growth stages has a profound influence on the impacts of extreme weather on crop yields (Gent and Kiyomoto, 1998; Saini and Aspinall,

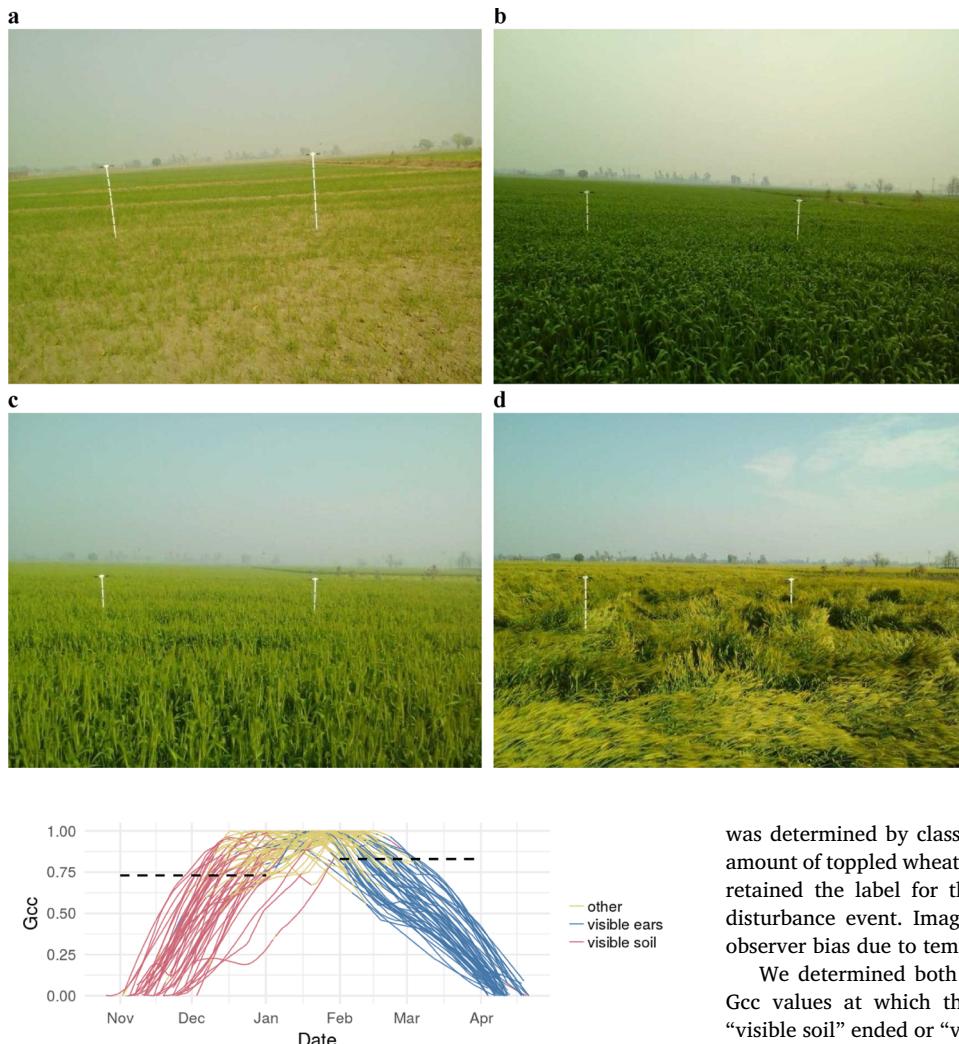


Fig. 3. Developmental stages of wheat as documented using the smartphone near-surface remote sensing PBI field trial. **a.** Initial growth after sowing with soil showing through the seedlings, generally marked as “visible soil” in our visible assessment. **b.** a closed canopy with no visible soil. **c.** wheat showing “visible (wheat) ears” in the image, our second classification label **d.** an image showing lodging, with the wheat plant toppled by wind and rain.

1982). We analyzed to what extent the average amplitude value at which crops visibly change growth stages (a ‘common threshold’) is more meaningful than these arbitrary thresholds, focusing on the onset of two critical wheat growth stages, namely: (1) canopy closure, or the absence of visible bare soil, indicative of the end of the leaf and tiller development stage, and (2) visible wheat ears (i.e. flowering bodies of wheat) representing onset of the heading growth stage.

To link Gcc values with the various growth phases of wheat and to derive the two common thresholds, we manually classified all images for the presence of “visible soil” (Fig. 3a) or “visible (wheat) ears” (Fig. 3c). The images not corresponding to either of these two classifications represent the growth phase between the end of tillering and the development of flowering bodies, i.e. the stem extension developmental phase of wheat (Fig. 3b). In addition, all images were also assessed for the presence of lodging damage closer to harvest (e.g. Fig. 3d). Lodging

was determined by classifying all images as damaged where a visible amount of toppled wheat or with stems deviating from vertical. We only retained the label for the first occurrence of lodging, i.e. the main disturbance event. Images were classified in random order to avoid observer bias due to temporal autocorrelation in the presented images.

We determined both the dates and the smoothed and normalized Gcc values at which the longest continuous series of data labeled “visible soil” ended or “visible ears” started for each site. Here, the end of “visible soil” labels and the start of “visible wheat ears” labels correspond to the wheat phenological phases of wheat tillering and wheat flowering or heading, respectively. As such, the variability in dates at which these events occur determine the variability in crop development among sites, while the Gcc values indicate the value at which these changes took place. Using the longest continuous series of labeled data means that images labeled outside the main phenological phase, such as visible soil not associated with tillering, were discarded from the analysis to ensure a conservative estimate of crop phenology and Gcc values. This conservative estimate only takes into account images that do not show ambiguity by imposing a discrete label on a continuous process. Extracted Gcc values were used to determine common (mean \pm standard deviation) Gcc amplitude thresholds (%) for both phenological phases across all sites (Fig. 4) and validated using a 1000-fold leave-p-out cross validation (omission of $p = 10$ sites, randomly drawn without replacement). We use these common thresholds for the end of tillering and start of heading to estimate the phenological dates from the smoothed and normalized Gcc time series across all sites. Estimated dates, using a common Gcc threshold, constitute an alternative set of information to the dates stemming from the manual classification above (Fig. 5).

2.5. Satellite remote sensing data pre-processing

We compare the dates at which individual, smartphone-based ground observations identify the transition between growth stages to those identified from satellite remote sensing products using common

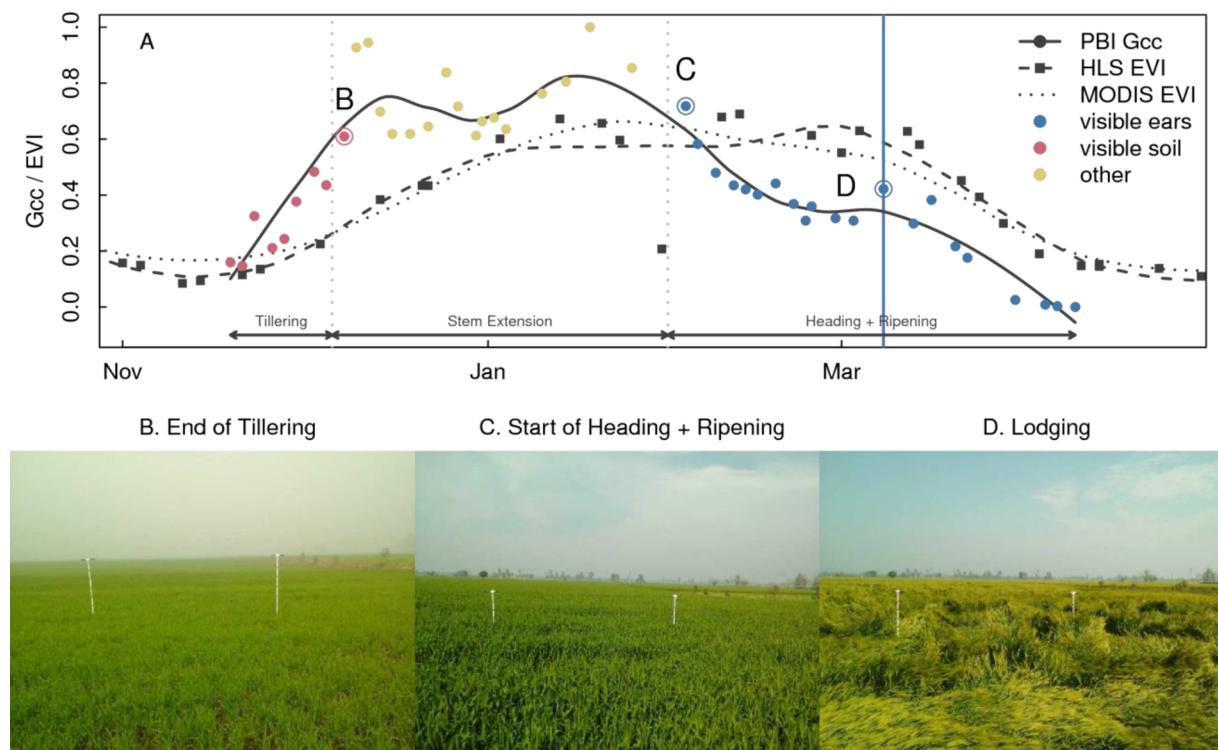


Fig. 5. Panel plot showing Gcc, Harmonized Landsat Sentinel (HLS) and MODIS EVI time series in combination with various phenological phases and illustration images for one selected site **a**. Normalized Gcc (solid circles, solid line), HLS (black squares and dashed line), and MODIS EVI (dotted line) time series throughout the growing season. Gcc points are color-coded according to the labels assigned by visual classification of the image. Vertical grey dashed lines indicate the estimated threshold-based phenological phases, i.e. the day at which the derived phenological thresholds for canopy closure and development of ears are exceeded. The period until the canopy closes approximates the tillering phase, the period of canopy closure until the emergence of ears approximates the stem extension phase, and the period after the first ears appear corresponds with the heading and ripening phase. These respective periods are marked with horizontal arrows. A vertical blue line indicates a lodging event. **b**. Last image marked as having no closed canopy. **c**. First image marked as having visible ears. **d**. Image showing an instance of lodging in the field of view (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

thresholds for both. For this, we interpreted the phenological stages from the in-situ crop imagery as our ground-truth data.

Next, we used visible observations of the end-of-tillering (EOT) and start-of-heading (SOH) to derive common growth-stage vegetation index threshold values rather than arbitrarily defined values. A robust common threshold would provide a way to estimate crop phenology without a visual assessment. We compared dates at which individual Gcc time series reach the common thresholds to dates at which the vegetation index for the same plot reaches a threshold as derived from two remote sensing products, with differing spatial and temporal scales.

First, we used Harmonized Landsat Sentinel (HLS) version 1.3 data, which fuses both Landsat 8 and Sentinel-2 A data into a consistent data product with higher temporal frequency (2–4 days) and a standardized 30 m resolution (Claverie and Masek, 2016). For all 52 sites, we extracted time series of the Enhanced Vegetation Index (EVI, Huete et al., 2002) in an 11×11 pixel window surrounding the site location. The EVI index is preferred over a Normalized Difference Vegetation Index (NDVI) as it accounts for vegetation background signals and reduces atmosphere influences (Huete et al., 2002). The image location is defined as the location 15 m to the north, or towards the direction covered by the field-of-view of the smartphone, of the smartphone GPS location. Two sites, 363636-1 and 105803-1, were removed from the HLS analysis as the time series were too noisy or showed no seasonal changes in their EVI values (Appendix Fig. A4).

Second, in addition to HLS imagery, we extracted matching EVI time series from the Moderate Resolution Imaging Spectroradiometer (MODIS) MCD43A4 (Collection 6) product at a daily interval and 500 m resolution for all sites. To assess the landscape-wide land cover class at the site locations we extracted the latest available Collection 5 MODIS Land Cover IGBP classes. All MODIS data was accessed through Google

Earth Engine® and a custom subset script (GEE; Hufkens, 2017), while HLS data was downloaded through the project's data portal and subset using a custom R script. Similar to the processing of the near-surface remote sensing data, we used smoothed and normalized time series using a fitted loess model with a fixed span of 0.4 for both remote sensing products (Appendix Fig. A5).

Phenological thresholds used in satellite remote sensing products or analysis are rarely directly related to observed plant physiological development (Zhang et al., 2003; Hufkens et al., 2012). We derive a set of generalized phenological dates for the start of season corresponding to the end-of-tillering and start-of-heading based on a range of potential arbitrary VI amplitude thresholds (10–90% with 20% increments) for both remote sensing datasets and Gcc time series. Comparison with Gcc-derived phenological dates based on common thresholds described in Section 2.4 is then used to evaluate possible errors in estimation of crop phenology based on use of arbitrary amplitude thresholds commonly used in analysis of satellite remote sensing imagery.

2.6. Statistical analysis

2.6.1. Quantifying spatial and temporal synchrony in growth phases

The seasonal development of plants and crops is largely driven by temperature, and, as a result, phenology is more accurately defined as a function of thermal time (growing degree days, GDD) rather than calendar days. However, due to uncertainty regarding sowing dates, it is not possible to use a simple GDD-based approach to assess spatial and temporal variability in crop phenology in smallholder regions such as our study site. We therefore rely on the high degree of synchronization in vegetation phenology across most landscapes (Defriez and Reuman, 2017), due to the slow varying nature of temperature in regions with

flat or mildly sloping topography. In order to test if phenological phases are geographically correlated across our sample of individual fields, we assess spatial autocorrelation among phenological indicator values (all threshold values for Gcc and remote sensing products) using a global Moran's I index. The Moran's I index summarizes both the spatial location and value of a measurement into an inferential statistic of spatial autocorrelation, under a null hypothesis that the measurements being analyzed are spatially randomly distributed. When a Moran's I index is not statistically significant, we cannot reject the null hypothesis, and we assume therefore that the spatial distribution of the measurements is the result of random spatial processes, for example differences in individual farmer management practices and production behavior.

2.6.2. Differences in growth stage as a function of image data and resolution

We assess the spatial heterogeneity (or lack thereof) in growth stage timing within and among sites to quantify the degree of spatial and temporal consistency in crop phenology between the in-situ crop imagery and remote sensing products. We estimate local spatial heterogeneity in the HLS data across a region of both 3×3 and 5×5 pixels surrounding each site location by describing the mean difference in days between the EOT (or SOH) of the pixel containing the in-situ crop imaging site location, and EOT (or SOH), of the surrounding pixels. For both the EOT and SOH we used two common amplitude thresholds for vegetation indices derived through remote sensing (10 and 50%, respectively) on the time series' ascending and descending sections. We compare the derived EOT and SOH from Gcc (based on the date at which a site-specific Gcc reaches the common thresholds defined in Fig. 4) with those from both HLS and MODIS (based on the date at which a site-specific EVI reaches the 10 or 50% thresholds) using ordinary least squares (OLS) regressions to assess the correspondence between near-surface remote sensing (Gcc) and satellite remote sensing time series.

2.6.3. Quantifying a lodging response in Gcc values

We quantified the influence of lodging on the Gcc values using a comparison of Gcc (means) between weeks surrounding a lodging event. We calculated the difference between the week containing the lodging event (at time t) and the preceding and following week ($t-1$, $t+1$ respectively, see Appendix Fig. A6). The first difference of these three weeks are indicative of the rate of change due to the normal seasonal progression rate during the growing season. We hypothesize that if consistently large differences occur in the sign and magnitude of the Gcc signal due to lodging events a comparison of mean progression rate should be significantly different between weeks during, preceding and potentially following lodging (i.e. lodging will lead to an accelerated rate of Gcc decline relative to the rate of Gcc decline before lodging). The responses in the weekly mean Gcc differences (Gcc rate of change) surrounding the timing of lodging (t) were compared using an analysis of variance (ANOVA, as $\Delta\text{Gcc} \sim \text{lodging} | \text{week}$), with lodging as an explanatory variable (factor), and week as a random effect.

3. Results

3.1. Visually-determined winter wheat phenology, Gcc thresholds, and lodging

As described in Section 2, the visual classification of images divides the seasonal trajectory of a Gcc curve into three distinct sections, breaking at approximately the inflection points of the step-shaped curve (Fig. 4, Appendix Fig. A3 for each site individually). The two defined labels, "visible soil" and "visible wheat ears", respectively, approximate the end of "tillering", which is marked by an increasing lengthening of the plant after canopy closure, and the start of heading, characterized by occurrence of visible flower structures (Fig. 3). We find that across all sites a threshold for normalized Gcc time series (0–1) of $0.73\% \pm 0.13$ of the overall amplitude best characterizes the end of tillering (the ascending part of the curve), while a slightly higher value

of $0.83\% \pm 0.10$ in amplitude is estimated for the start of heading (or the descending part of the curve). Our leave-p-out cross-validation supports these numbers with mean values of $0.73\% \pm 0.01$ and $0.83\% \pm 0.01$ across the 1000 iterations for the end of tillering and start of heading phases, respectively. These common thresholds explain 54% of the variation in the observed end of tillering dates, and 46% of the variation in the start of heading dates across sites, regardless of incomplete data due to inconsistent image acquisition times by farmers. Examples of both actual and estimated threshold-based phenological phases are illustrated in Fig. 4 and Appendix Fig. A6.

Lodging events were common throughout our study sample, with 31 out of 52 sites affected (~60%). However, the severity and timing of these events varied considerably from site to site (visually assessed but not explicitly quantified). Almost all lodging events (30 out of 31 sites) were limited to the heading and ripening phase of the growth cycle (15th of January - 11th of April), consistent with previous research showing that most lodging typically occurs during 2–3 months preceding a harvest (Berry et al., 2003). As expected for a process driven by stochastic events such as excessive rain and high wind conditions, no well-defined Gcc threshold can be derived for lodging events with a threshold of 0.21 ± 0.25 , spanning a range from 0.002 to 0.93. When we compare the rate of change in Gcc values in the week before, during and after the lodging event no significant differences are found between the 3 means (ANOVA, $F = 0.56$, $p > 0.05$), indicating that lodging has little influence on the rate of change between successive weeks. Indeed, visual inspection confirms no abrupt changes in Gcc values during or after lodging events. We thus conclude that Gcc values recorded, at this particular view angle, are insensitive to lodging events. Similarly, satellite remote sensing EVI time series did not show consistent VI values for lodging events, varying from 0.14 to 0.99 (0.64 ± 0.22) and 0.11 to 0.98 (0.70 ± 0.23) for MODIS and HLS data, respectively. Furthermore, these time series do not contain visible evidence of the crop damage that can be detected from in-situ imagery (e.g. Figs. 3d and 5d). Section 4.2 discusses alternative approaches to exploit information contained in the in-situ imagery beyond Gcc to support the detection and classification of lodging events.

3.2. Comparison of near-surface and satellite derived phenology metric

The SOS and EOS phenological dates derived through satellite remote sensing with the discrete phenological dates of the Gcc time series (EOT and SOH) explain a limited, but significant, fraction of the variance in phenological timing/dates by near-surface remote sensing data (Table 1). The SOS derived from MODIS data show a positive relationship with Gcc end-of-tillering dates, explaining ~4 to 14% of the variance. The HLS data only shows a significant relation with the Gcc-based tillering estimates for thresholds of 50 and 70% amplitude, explaining ~16% and 9% of the variance, respectively. Start of heading dates show weaker relationships with the Gcc-derived estimates for the SOH. Here, only the MODIS EOS value for a 50% threshold shows a significant relationship with the Gcc-based start of heading, explaining only 11% of the variance. All other threshold values for both HLS and MODIS data were non-significant. Overall we note a consistent root mean squared error of ~12 days across all comparisons, due to the monotonic increase of both curves and their relatively constant rate of change over the period considered.

When comparing arbitrary amplitude thresholds between SOS/EOS dates derived from Gcc time series and satellite remote sensing we find that there are some significant relationships (Table 2). Agreement between HLS / MODIS and Gcc SOS and EOS values is highest for thresholds of 50 / 30%, 50 / 50% and 50 / 50%, 50 / 90% respectively. Although in this cross-comparison relationships are at times significant, they only explain a limited amount of the variance, with values generally < 20% and with a maximum of 38%. Summarizing, across both analysis and irrespective of the choice of threshold or Gcc metric, only a limited proportion of the variance in Gcc derived phenology can be explained by satellite imagery.

Table 1

Summary statistics of the comparison of near-surface remote sensing derived wheat phenology metrics and phenological metrics for various amplitude thresholds for two remote sensing products at different spatial and temporal scales. We report the absolute difference and R^2 , RMSE in days of a linear fit between the Gcc derived values and those derived from satellite remote sensing using Harmonized Landsat Sentinel (HLS) and Moderate resolution radiometer (MODIS) based EVI time series at two different scales of 30 and 500 m, respectively. All significant values are marked with * ($p < 0.05$), non-significant values are marked with NS. Statistics are reported for start of season and end of season thresholds. Thresholds used for the Gcc derived phenological metrics were determined by visually assessing images for canopy closure (end of tillering) or a visible flowering and fruiting body (heading).

Start of Season (SOS) ~ end of tillering (EOT)						
Threshold	HLS EVI (30 m)			MODIS EVI (500 m)		
	R^2	RMSE	Mean Bias	R^2	RMSE	Mean Bias
10	0 ^{NS}	12.54	25.14	0.04 ^{NS}	12.36	23.98
30	0 ^{NS}	12.59	12.84	0.14*	11.68	7.27
50	0.16*	11.55	0.22	0.14*	11.66	-3.44
70	0.09*	11.98	-10.94	0.14*	11.67	-13.81
90	0 ^{NS}	12.56	-33.34	0.14*	11.66	-26.37

End of Season (EOS) ~ start of heading (SOH)						
Threshold	R^2	RMSE	Mean Bias	R^2	RMSE	Mean Bias
10	0.06 ^{NS}	11.21	-60.72	0.01 ^{NS}	11.47	-60.02
30	0.04 ^{NS}	11.31	-48.58	0.02 ^{NS}	11.4	-47.40
50	0.04 ^{NS}	11.31	-40.04	0.11*	10.92	-37.69
70	0.02 ^{NS}	11.43	-31.06	0.05 ^{NS}	11.25	-26.65
90	0.04 ^{NS}	11.34	18	0.02 ^{NS}	11.42	-7.15

3.3. Between-site heterogeneity in land cover and crop phenology

For our sample of 52 fields, we compare time series of canopy development over widely varying scales: from field-based Gcc values, to remote sensing-based EVI indices at a stand (HLS) and landscape (MODIS) scale. At the landscape level, the MODIS land cover product (MCD12Q1) indicates a large homogeneity in land cover type, with all sites labeled as croplands (IGBP class 12).

Irrespective of this homogeneity in landscape-level land cover, close visual inspection using high resolution remote sensing imagery shows predominantly smaller field sizes in this region that are interspersed with natural vegetation and man-made features (e.g. buildings, roads;

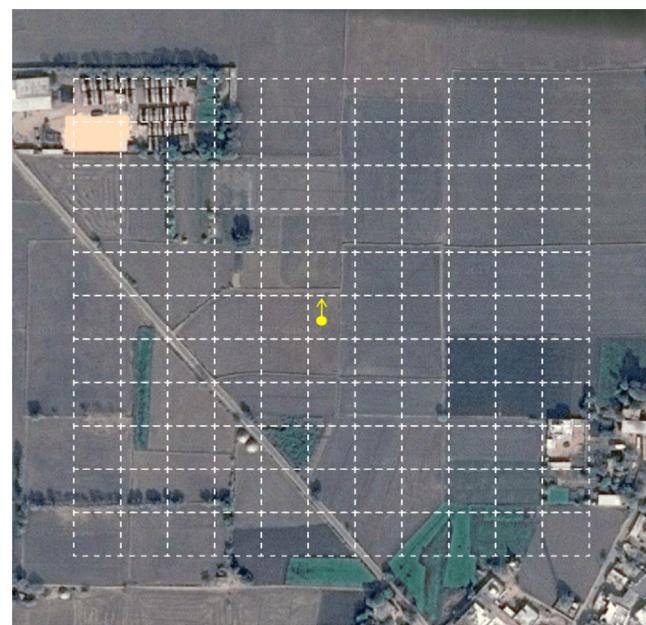


Fig. 6. Illustration of a field sizes relative to the harmonized Landsat sentinel (HLS) grid cells for a randomly selected site in the study region. The potential view angle of the camera, had this farmer been included in the study, is indicated by the yellow dot, and yellow arrow respectively. White dashed grid cells display the relative size of the HLS product (~30 m). Note the spatial heterogeneity of the landscape, with very few HLS grid cells not intersected by roads, man-made structures, other vegetation types or dirt barriers, marking field boundaries. (image: Google Maps®) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Fig. 6. Consequently, for any given site, the surroundings show only a limited number of HLS 30 m grid cells not intersected by roads, man-made structures, different vegetation types, or dirt barriers defining field margins (e.g., Fig. 6, white dashed polygons). This fine scale heterogeneity has a marked influence on the local derived phenology surrounding each site. For example, the start of season across a 5 × 5 (3 × 3) cell window surrounding our sites shows dispersion (standard deviation) between the central pixel and the surrounding values of up to 22 days (14 days), with a mean dispersion across sites of 7 ± 4 days (5 ± 4 days) (Table 2).

Table 2

Linear regressions between various arbitrary thresholds for the near-surface greenness time series (Gcc) and satellite remote sensing products, in particular the Harmonized Landsat Sentinel (HLS) and MODIS EVI time series. (indicated by the row and column headers), for both the start-of-season (SOS) and end-of-season (EOS). We report the R^2 values of the linear regressions. Significant relationships are marked with an asterisk (*, $p < 0.05$).

Start of Season (SOS)										End of Season (EOS)				
		10	30	50	70	90			10	30	50	70	90	
Gcc	10	0.038	0.057	0.232*	0.201*	0.038	0.092*	0.094*	0.09*	0.091*	0.071			
	30	0.002	0.002	0.378*	0.295*	0.06	0.127*	0.148*	0.162*	0.178*	0.161*			
	50	0.004	0.001	0.364*	0.263*	0.041	0.157*	0.189*	0.186*	0.183*	0.164*			
	70	0.003	0.002	0.216*	0.145*	0.016	0.103*	0.149*	0.152*	0.144*	0.134*			
	90	0.005	0.003	0.136*	0.087*	0.003	0.062	0.089*	0.108*	0.104*	0.089*			
Start of Season (SOS)										End of Season (EOS)				
		10	30	50	70	90			10	30	50	70	90	
Gcc	10	0.028	0.115*	0.116*	0.122*	0.111*	0.004	0.002	0.013	0	0.003			
	30	0.061	0.228*	0.236*	0.245*	0.228*	0	0	0.018	0.001	0			
	50	0.065	0.239*	0.245*	0.251*	0.24*	0.018	0.017	0.069	0.016	0.007			
	70	0.033	0.145*	0.145*	0.143*	0.144*	0.015	0.024	0.099*	0.045	0.031			
	90	0.044	0.098*	0.118*	0.115*	0.121*	0.016	0.029	0.119*	0.059	0.022			

Despite the fairly consistent use of the same wheat variety (HD2967) across most sites (78%), a large variability in derived crop phenology was noted in all but the MODIS data. For example, the start of the growing season, as defined by the 10% Gcc amplitude threshold at the start of the (Rabi) growing season (Hufkens et al., 2012) (which on average started on the 23th of November) showed a difference of 47 days between the first and the last emerging site. The subsequent wheat tillering phase, as observed from images, ended on average 25 ± 10 days later (18th of December) (Fig. 4). The end of wheat tillering is only weakly coupled to the start of the growing season, suggesting high variability between fields in the rate of crop development. As such, the start of the growing season explains only ~44% of the variability ($R^2 = 0.44$, $p < 0.05$) in the end of the tillering phase. Consequently, there are up to 70 days between the first and last site reaching the end of tillering. On average two months after the end of tillering, on the 12th of February 2017, the first wheat ears appeared, marking the state of the heading and ripening phase. Across sites, 48 days separated the first and last occurrence of the start of heading. However, the date of tillering only explains a small amount ($R^2 = 0.16$, $p < 0.05$) of the variability in the date of heading again showing a large variability in crop development rates, arguably due to the influence of management practices and local weather (Table 3).

When evaluating spatial autocorrelation in phenology timing using a Moran's I index for both Gcc- and satellite remote sensing-derived values, we find no significant autocorrelation for either the Gcc- and HLS-based phenology metrics (Table 4). We therefore conclude that the start of the growing season, the end of tillering, and the start of heading were to a large extent governed by a spatially-random process. These differences are likely driven by non-linear responses in crop growth, classification uncertainty, and the idiosyncratic nature of farmer planting dates. In contrast, all MODIS phenological metrics show a positive and significant Moran's I index, an indication that the MODIS time series shows landscape wide-synchronicity. In other words, the derived phenological metrics are not spatially random in their location and value due to the smoothing out of field-level heterogeneity in farmer management practices and timing.

4. Discussion

Smallholder farmers play a critical role in supporting food security in developing countries, but often have limited capacity to manage production risks or mitigate income losses resulting from extreme weather events. The timing of weather extremes relative to in-traseasonal crop development is known to have a significant influence on crop yields (Cui et al., 2018; Tariq et al., 2017). We discuss key insights gained from our analysis about the potential of near-surface imagery for crop phenological monitoring in heterogeneous smallholder systems, highlighting in particular the advantages of this approach over traditional satellite-based data products.

Table 3

Variability in estimated start of season (SOS) and end of season (EOS) as derived from Harmonized Landsat Sentinel (HLS) time series for 3x3 and 5x5 windows surrounding site locations. We report mean (and maximum) and their deviation as differences in number of days between the pixel containing the site location and the surrounding window for two commonly used phenological threshold values (10 and 50%).

Window Size	SOS		EOS	
	10%	50%	10%	50%
3 x 3	2 ± 2 (12)	5 ± 4 (14)	5 ± 4 (14)	5 ± 7 (34)
5 x 5	3 ± 2 (12)	7 ± 4 (22)	6 ± 3 (13)	8 ± 8 (34)

Table 4

Summary statistics of the comparison of near-surface remote sensing derived wheat phenology metrics and phenological metrics for various amplitude thresholds for two remote sensing products at different spatial and temporal scales. We report spatial-autocorrelation of the phenological metrics using a global Moran's I value. All significant values are marked with * ($p < 0.05$), non-significant values are marked with NS. Statistics are reported for start of season and end of season thresholds. Thresholds used for the Gcc derived phenological metrics were determined by visually assessing images for canopy closure (end of tillering) or a visible flowering and fruiting body (heading). Thresholds for HLS and MODIS remote sensing were arbitrary.

Threshold	Start of Season (EOT)		
	Gcc Moran's I	HLS EVI (30m) Moran's I	MODIS EVI (500m) Moran's I
10	/	-0.07 ^{NS}	0.17*
30	/	0.06 ^{NS}	0.15*
50	/	0.11 ^{NS}	0.17*
70	/	0.08 ^{NS}	0.17*
90	/	-0.13 ^{NS}	0.15*
73 (reference)	0.02 ^{NS}	/	/

Threshold	End of Season (SOH)		
	Moran's I	Moran's I	Moran's I
10	/	0.05 ^{NS}	0.21*
30	/	0.04 ^{NS}	0.13*
50	/	0 ^{NS}	0.27*
70	/	0 ^{NS}	0.37*
90	/	0.03 ^{NS}	0.34*
83 (reference)	0.12 ^{NS}	/	/

4.1. Improving phenological monitoring with near-surface and satellite imagery

Although the use of smartphones to monitor seasonal changes in vegetation is not novel (Graham et al., 2011, Fang et al. 2018), our study presents the first large-scale implementation of a flexible protocol to monitor seasonal crop development and damage that can be carried out by farmers with limited technical skills. Our analysis shows that smartphone based near-surface remote sensing can capture the progression of the crop growth season in a way that improves upon the level of information captured by alternative methods. Similar to Kosmala et al. (2016), we used visual assessments of the near-surface remote sensing imagery and found that well-constrained amplitude thresholds on vegetation greenness correlated with the timing of key phenological phases during wheat development.

In contrast, unlike previous studies on deciduous broadleaf canopies (Hufkens et al., 2012; Klosterman et al., 2014), we show a weak relationship between Gcc- and satellite remote sensing-derived phenological metrics when considering visually-determined common thresholds for the Gcc-derived metrics. Satellite remote sensing-based phenology metrics for a range of amplitude thresholds explain at most 16% of the overall variance in the timing of the end of tillering or the start of heading. Similarly, cross-comparison using arbitrary thresholds between Gcc- and satellite- remote sensing derived phenological metrics show significant relationships but with only a limited amount of variance explained.

Critically, satellite remote-sensing techniques are unable to capture the significant degree of spatial heterogeneity in crop phenology that is observed both within and between smallholder fields in our sample. Observed non-synchronicity in crop phenology contrasts strongly with the highly synchronous nature of phenology in both natural and managed non-agricultural ecosystems (Defriez and Reuman, 2017), and suggests that crop phenology, while still driven by climate, is also influenced heavily by management practices such as sowing dates, irrigation, and fertilizer application. In our study sample, we discount

choice of crop variety as a major driver of heterogeneity in phenology as ~76% of surveyed farmers report having sowed the same HD2967 wheat variety, a value which remains roughly constant across our broader sample of survey sites. In addition, near-surface remote sensing differentiates itself from most satellite remote sensing phenology products by the ability to link threshold values, which approximate inflection points on the Gcc curves, with physiologically well characterized phenological phases. Importantly, we show that such phases can't be quantified through ad-hoc thresholds of high-resolution satellite remote sensing vegetation indices.

Several factors may explain the improved performance of phenological estimation from near-surface imagery compared with satellite-based data. Smallholder agricultural systems, in particular in the tropics, are characterized by small plot sizes, intercropping, and cloud cover during the main agricultural production season. Despite advances in both the processing and the resolution of satellite imagery, use of earth observation datasets for assessing smallholder crop production remains a challenge due to the small size of fields (e.g. < 1 ha) and the heterogeneous nature of cropping patterns, with multiple vegetation/crop types and fallow areas interspersed (Duncan et al., 2015; Mann and Warner, 2017). In addition, existing assessments of crop yields using satellite remote sensing (Gao et al., 2017; Johnson et al., 2016; Kouadio et al., 2014) typically do not control for effects of field-level crop phenology (Dalhaus et al., 2018; Tariq et al., 2017). In this context, phenocam-type data using smartphone cameras on the ground provide high-resolution plot-level data, with consistent coverage in cloudy environments that cannot be achieved using existing satellite datasets. Additionally, improved performance from near-surface imagery may also reflect fundamental differences in the properties of vegetation indices derived from satellite and near-surface remote sensing. In our study, near-surface remote sensing registers the greenness of vegetation while the EVI derived from satellite data registers both photosynthetic and structural canopy properties. Estimates of phenological timing from satellite imagery are biased by the fact that structural changes in crop growth often lag behind changes in vegetation pigment (Huete et al., 2002; Keenan et al., 2014). Furthermore, the relative proximity of near-surface remote sensing to vegetation and its view angle might play an important role in registering more subtle changes in vegetation color needed to capture the onset of crop development phases (Vrieling et al., 2018).

4.2. Implications for monitoring and managing agricultural production risk

Near-surface imagery from phenocam-type data could provide valuable additional information about field-scale variability in crop phenology, which may help explain some of the heterogeneity in crop yield estimates observed in satellite-based assessments of smallholder crop yields in our study region (Jain et al., 2017) and elsewhere (Azzari et al., 2017). In contrast to satellite data, use of near-surface crop imagery also enables detection of yield loss events that do not appreciably change vegetation greenness, irrespective of the resolution of the earth observation dataset. For example, lodging impacts on wheat yields result from reductions in net photosynthesis as the more photosynthetically active upper leaves are unable to function effectively, along with physical difficulties in harvesting grain in lodged crops. These events are not captured by greenness time series, but can be readily detected using a visual archive of the standing state of the vegetation through smartphone imagery (Fig. 3d). Importantly, while in this study we classify lodging manually, additional computer vision-based texture indices or deep learning approaches may also provide a way to characterize various image features and hence identify wheat lodging (Cimpoi et al., 2015). Future developments such as the Planet Labs micro satellites and other sub-10 m scale image products might provide sufficient resolution if made freely available to researchers. However, such new high temporal and spatial resolution imagery often is only available at significant cost or for restricted geographic extents,

rendering these data financially prohibitive for use in research or policy analysis while also failing to address the limited correlation between remotely sensed vegetation indices and wheat phenology observed in our study.

Finally, the ability of near-surface smartphone imagery to capture spatial and temporal heterogeneity in crop phenology for individual smallholder fields has important implications for efforts to target policies to address low agricultural productivity and resultant impacts on farmer livelihoods. In many parts of the developing world, there is growing interest in the use of crop insurance as a tool to protect farmers against financial losses caused by extreme weather events. For example, a recent study by Dalhaus et al. (2018) for winter wheat production in Germany has shown that phenology models based on crop observation networks can help to increase the reliability of yield loss estimates and financial payouts from weather index-based parametric insurance, reducing cost of insurance provision and increasing farmer demand for such products. Government-funded crop observation networks are rarely present in smallholder agricultural systems in the developing world due to their significant operation costs. However, phenological monitoring using near-surface imagery from farmer-reported smartphone imagery—as proposed in this study—offers a promising low-cost opportunity for scaling phenological data collection to support rapid assessment of distributed yield losses in smallholder systems (Corbeels et al., 2018).

4.3. Future directions

Our study suggests that vegetation indices derived from either low or high resolution satellite imagery are unable to reliably monitor crop phenological stages in smallholder agricultural systems such as northwest India. Although high resolution satellite remote sensing data may help overcome limitations of the spatial resolution on current data products, it remains subject to other shortcomings that we were able to overcome by relying on smartphone images to monitor crop phenology, growth disturbances and consequently the effect on crop yields. In particular, it should be noted that there is a critical lack of high resolution ground-truth data. Plot-level data from nationally-representative crop cut surveys are almost never made public, but instead aggregated over large administrative areas. This aggregation of data not only greatly obscures critical variation in yields, and its causes, but is also aggregated across administrative regions that do not correspond to agricultural or climatic regions.

Although our study addresses practical issues related to post-processing of acquired images, socio-economic considerations around the picture-taking protocol cannot be ignored ahead of a potential wider scale deployment. For example, retention of farmers during the entire growing season has proven challenging. In total, approximately 28% of all farmers took sufficient images during the latter half of the growing season (> 15 images after January 15th) with only ~10% adhering to the strict selection requirements for our analysis. Although partial seasonal data limits its uses, for instance within a crop modelling framework, the derived phenological statistics would still remain valid and provide useful information. It is important to note that in our study we did not rely on sending farmers reminders to take pictures during the growing season, so retention rates could potentially be higher when automatically scheduling picture taking reminders. A valuable area for future research therefore might be to determine the required number and timing of images for both monitoring phenology and derivative applications, for example for yield prediction. Notwithstanding the overall success of this first field trial, incentive programs, co-benefits to participate in field trials and reminders should be carefully considered to maximize farmer engagement within both an insurance and research context (Ceballos et al., 2018).

Further work should also explore the potential to exploit growing capabilities in computer vision and machine learning techniques to support automation of crop feature extraction from smartphone images.

Previous research has shown that machine learning in combination with large image databases can successfully detect plant stress (Ghosal et al., 2018) or complex leaf venation patterns (Wilf et al., 2016). Preliminary machine learning classification experiments of winter wheat growth stages by the authors are promising (data not shown) and will be important for the potential to scale our proposed approach across smallholder landscapes such as the Indo-Gangetic Plain and others.

5. Conclusion

The ability to monitor and model crop development and phenology is essential to support monitoring and management of agricultural production and food security globally. In this study, we show that near-surface remote sensing can be used to quantify physiologically important phenological stages in agricultural crops, where satellite remote sensing based vegetation indices fail to capture the start of wheat heading and the end of the tillering phases in highly heterogeneous settings, such as smallholder farming. Additionally, we demonstrate that near-surface imagery also provides a valuable record of agronomically significant crop damage events (e.g. wheat lodging) that cannot be detected readily using moderate or coarse spatial resolution satellite imagery. Information about crop phenology and damage has significant potential to support improved estimation of heterogeneous agricultural productivity within smallholder systems, and also offers opportunities to support delivery of interventions, such as crop insurance, to increase farmers' resilience to financial risks posed by extreme weather events. Additionally, these image records could fill a critical information gap by providing high resolution ground-truth data, not currently available from national crop cut surveys.

Acknowledgements

This work was undertaken as part of the CGIAR Research Program on Policies, Institutions, and Markets (PIM) led by the International Food Policy Research Institute (IFPRI). Funding support for this study was provided by the CGIAR Research Programs on Climate Change and Agricultural Food Security (CCAFS) and Policies, Institutions, and Markets (PIM); the CGIAR Platform for Big Data in Agriculture; 3ie Award No. TW13-1038; and NERC-ESRC-DFID Award No. NE/R014094/1. KH acknowledges support from the National Science Foundation's Macro-system Biology Program (award EF-1065029) and the Belgian Science Policy Office (contract BR/175/A3/COBECORE). EKM acknowledges support from NASA grant number NNN14ZDA001N-LCUC. This paper has not gone through IFPRI's standard peer-review procedure. The opinions expressed here belong to the authors, and do not necessarily reflect those of CCAFS, PIM, 3ie, NERC, IFPRI, or CGIAR.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.agrformet.2018.11.002>.

References

- Auffhammer, M., Ramanathan, V., Vincent, J.R., 2012. Climate change, the monsoon, and rice yield in India. *Clim. Change* 111, 411–424. <https://doi.org/10.1007/s10584-011-0208-4>.
- Azzari, G., Jain, M., Lobell, D.B., 2017. Towards fine resolution global maps of crop yields: testing multiple methods and satellites in three countries. *Remote Sens. Environ.* 202, 129–141. <https://doi.org/10.1016/j.rse.2017.04.014>.
- Barnett, B.J., Barrett, C.B., Skees, J.R., 2008. Poverty traps and index-based risk transfer products. *World Dev.* 36, 1766–1785. <https://doi.org/10.1016/j.worlddev.2007.10.016>.
- Below, T.B., Mutabazi, K.D., Kirschke, D., Franke, C., Sieber, S., Siebert, R., Tscherning, K., 2012. Can farmers' adaptation to climate change be explained by socio-economic household-level variables? *Glob. Environ. Change* 22, 223–235. <https://doi.org/10.1016/j.gloenvcha.2011.11.012>.
- Berry, P.M., Sterling, M., Baker, C.J., Spink, J., Sparkes, D.L., 2003. A calibrated model of wheat lodging compared with field measurements. *Agri. For. Meteorol.* 119, 167–180.
- Burke, M., Lobell, D.B., 2017. Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proc. Natl. Acad. Sci.* 114, 2189–2194. <https://doi.org/10.1073/pnas.1616919114>.
- Carletto, C., Jolliffe, D., Banerjee, R., 2015. From tragedy to renaissance: improving agricultural data for better policies. *J. Dev. Stud.* 51, 133–148. <https://doi.org/10.1080/00220388.2014.968140>.
- Ceballos, F., Kramer, B., Robles, M., 2018. The Feasibility of Picture-Based Crop Insurance (PBI). *Smartphone Pictures for Affordable Crop Insurance*.
- Chen, M., Melaas, E.K., Gray, J.M., Friedl, M.A., Richardson, A.D., 2016. A new seasonal-deciduous spring phenology submodel in the Community Land Model 4.5: impacts on carbon and water cycling under future climate scenarios. *Glob. Change Biol.* 22, 3675–3688. <https://doi.org/10.1111/gcb.13326>.
- Cimpoi, M., Maji, S., Vedaldi, A., 2015. 2015 IEEE Conf. Deep Filter Banks for Texture Recognition and Segmentation. *Comput. Vis. Pattern Recognit. (CVPR) 2015. Deep Filter Banks for Texture Recognition and Segmentation. Comput. Vis. Pattern Recognit. (CVPR) 2015*, 3828–3836. <https://doi.org/10.1109/CVPR.2015.7299007>.
- Claverie, M., Masek, J.G., 2016. Harmonized Landsat-8 Sentinel-2 (HLS) Product User's Guide Website. pp. 1–16. 2. <https://hls.gsfc.nasa.gov/>.
- Corbeels, M., Berre, D., Rusinamudzi, L., Lopez-Ridaura, S., 2018. Can we use crop modelling for identifying climate change adaptation options? *Agric. For. Meteorol.* 256–257, 46–52. <https://doi.org/10.1016/j.agrformet.2018.02.026>.
- Cui, Z., Zhang, H., Chen, X., Zhang, C., Ma, W., Huang, C., Zhang, W., Mi, G., Miao, Y., Li, X., Gao, Q., Yang, J., Wang, Z., Ye, Y., Guo, S., Lu, J., Huang, J., Lv, S., Sun, Y., Liu, Y., Peng, X., Ren, J., Li, S., Deng, X., Shi, X., Zhang, Q., Yang, Z., Tang, L., Wei, C., Jia, L., Zhang, J., He, M., Tong, Y., Tang, Q., Zhong, X., Liu, Z., Cao, N., Kou, C., Ying, H., Yin, Y., Jiao, X., Zhang, Q., Fan, M., Jiang, R., Zhang, F., Dou, Z., 2018. Pursuing sustainable productivity with millions of smallholder farmers. *Nature*. <https://doi.org/10.1038/nature25785>.
- Dalhaus, T., Musshoff, O., Finger, R., 2018. Phenology information contributes to reduce temporal basis risk in agricultural weather index insurance. *Sci. Rep.* 8, 1–10. <https://doi.org/10.1038/s41598-017-18656-5>.
- Defriez, E.J., Reuman, D.C., 2017. A global geography of synchrony for terrestrial vegetation. *Glob. Ecol. Biogeogr.* 26, 867–877. <https://doi.org/10.1111/geb.12594>.
- Duncan, J.M.A., Dash, J., Atkinson, P.M., 2015. The potential of satellite-observed crop phenology to enhance yield gap assessments in smallholder landscapes. *Front. Environ. Sci.* 3, 1–16. <https://doi.org/10.3389/fenvs.2015.00056>.
- Farooq, M., Bramley, H., Palta, J.A., Siddique, K.H.M., 2011. Heat stress in wheat during reproductive and grain-filling phases. *CRC. Crit. Rev. Plant Sci.* 30, 491–507. <https://doi.org/10.1080/07352689.2011.615687>.
- Fishman, R., 2016. More uneven distributions overturn benefits of higher precipitation for crop yields. *Environ. Res. Lett.* 11, 24004.
- Gao, F., Anderson, M.C., Zhang, X., Yang, Z., Alfieri, J.G., Kustas, W.P., Mueller, R., Johnson, D.M., Prueger, J.H., 2017. Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. *Remote Sens. Environ.* 188, 9–25. <https://doi.org/10.1016/j.rse.2016.11.004>.
- Gent, M., Kiyomoto, R., 1998. Physiological and agronomic consequences of Rht Genes in wheat. *J. Crop. Prod. Process.* 1, 27–46. <https://doi.org/10.1300/J144v01n01>.
- Ghosal, S., Blystone, D., Singh, A.K., Ganapathysubramanian, B., Singh, A., Sarkar, S., 2018. An explainable deep machine vision framework for plant stress phenotyping. *Proc. Natl. Acad. Sci.* 201716999.
- Harvey, C.A., Rakotobe, Z.L., Rao, N.S., Dave, R., Razafimahatratra, H., Rabarijohn, H., Rajafarara, H., Mackinnon, J.L., B, P.T.R.S., Razafimahatratra, H., Rabarijohn, R.H., Rajafarara, H., Harvey, C.A., 2014. Extreme vulnerability of smallholder farmers to agricultural risks and climate change in Madagascar. *Extreme vulnerability of smallholder farmers to agricultural risks and climate change in Madagascar*. Author for correspondence. *Philos. Trans. R. Soc. B Biol. Sci.* 369, 20130089.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978. <https://doi.org/10.1002/joc.1276>.
- Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* 83, 195–213.
- Hufkens, K., 2017. A Google Earth Engine Subset Script & Library. <https://doi.org/10.5281/zenodo.833789>.
- Hufkens, K., Friedl, M., Sonnentag, O., Braswell, B.H., Milliman, T., Richardson, A.D., 2012. Linking near-surface and satellite remote sensing measurements of deciduous broadleaf forest phenology. *Remote Sens. Environ.* 117, 307–321. <https://doi.org/10.1016/j.rse.2011.10.006>.
- 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. *Nat. Clim. Change*. <https://doi.org/10.1038/nclimate2942>.
- IFAD, UNEP, 2013. Smallholders, food security and the environment. *Rome Int. Fund Agric. Dev.* 17, 54.
- Jain, M., Naeem, S., Orlove, B., Modi, V., DeFries, R.S., 2015. Understanding the causes and consequences of differential decision-making in adaptation research: adapting to a delayed monsoon onset in Gujarat, India. *Glob. Environ. Change* 31, 98–109. <https://doi.org/10.1016/j.gloenvcha.2014.12.008>.
- Jain, M., Singh, B., Srivastava, A.A.K., Lobell, D.B., Azzari, G., 2017. Using satellite data to identify the causes of and potential solutions for yield gaps in India's Wheat Belt. *Satellite detection of rising maize yield heterogeneity in the U.S. Midwest*. *Environ. Res. Lett.* 12. <https://doi.org/10.1088/1748-9326/aa5371>.

- Jain, M., Srivastava, A.K., Balwinder-Singh Joon, R.K., McDonald, A., Royal, K., Lisaius, M.C., Lobell, D.B., 2016. Mapping smallholder wheat yields and sowing dates using micro-satellite data. *Remote Sens.* 8, 1–18. <https://doi.org/10.3390/rs8100860>.
- Johnson, M.D., Hsieh, W.W., Cannon, A.J., Davidson, A., Bédard, F., 2016. Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. *Agric. For. Meteorol.* 218–219, 74–84. <https://doi.org/10.1016/j.agrformet.2015.11.003>.
- Jönsson, P., Eklundh, L., 2004. TIMESAT - a program for analyzing time-series of satellite sensor data. *Comput. Geosci.* 30, 833–845. <https://doi.org/10.1016/j.cageo.2004.05.006>.
- 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. *Ecol. Appl.* 24, 1478–1489. <https://doi.org/10.1890/13-0652.1>.
- Killlick, R., Fearnhead, P., Eckley, I.A., 2011. Optimal detection of changepoints with a linear computational cost. *J. Am. Stat. Assoc.* 1459. <https://doi.org/10.1080/01621459.2012.737745>.
- 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. *Biogeosci. Discuss.* 11, 2305–2342. <https://doi.org/10.5194/bgd-11-2305-2014>.
- Kosmala, M., Crall, A., Cheng, R., Hufkens, K., Henderson, S., Richardson, A.D., 2016. Season spotter: using citizen science to validate and scale plant phenology from near-surface remote sensing. *Remote Sens.* 8. <https://doi.org/10.3390/rs8090726>.
- Kouadio, L., Newlands, N.K., Davidson, A., Zhang, Y., Chipanshi, A., 2014. Assessing the performance of MODIS NDVI and EVI for seasonal crop yield forecasting at the ecodistrict scale. *Remote Sens.* 6, 10193–10214. <https://doi.org/10.3390/rs61010193>.
- Lobell, D.B., Schlenker, W., Costa-Roberts, J., 2014. Climate trends and global crop production since 1980. *Science* (80-) 343, 1017–1021. <https://doi.org/10.1126/science.1244510>. [dryad.5t110.Supplementary](https://doi.org/10.5061/dryad.5t110.Supplementary).
- Lowder, S.K., Skoet, J., Raney, T., 2016. The number, size, and distribution of farms, smallholder farms, and family farms worldwide. *World Dev.* 87, 16–29. <https://doi.org/10.1016/j.worlddev.2015.10.041>.
- Mann, M.L., Warner, J.M., 2017. Ethiopian wheat yield and yield gap estimation: a spatially explicit small area integrated data approach. *F. Crop. Res.* 201, 60–74. <https://doi.org/10.1016/j.fcr.2016.10.014>.
- Melaas, E.K., Friedl, M.A., Richardson, A.D., 2016. Multiscale modeling of spring phenology across Deciduous Forests in the Eastern United States. *Glob. Chang. Biol.* 22, 792–805. <https://doi.org/10.1111/gcb.13122>.
- Morton, J.F., 2007. The impact of climate change on smallholder and subsistence agriculture. *Proc. Natl. Acad. Sci.* 104, 19680–19685. <https://doi.org/10.1073/pnas.0701855104>.
- Ricciardi, V., Ramankutty, N., Mehrabi, Z., Jarvis, L., Chookolingo, B., 2018. How much of the world's food do smallholders produce? *Global Food Sec.* 2018 (17), 64–72. <https://doi.org/10.1016/j.gfs.2018.05.002>.
- Rodell, M., Velicogna, I., Famiglietti, J.S., 2009. Satellite-based estimates of groundwater depletion in India. *Nature* 460, 999–1002.
- Saini, H.S., Aspinall, D., 1982. Abnormal sporogenesis in wheat (*Triticum-Aestivum L*) induced by short periods of high-temperature. *Ann. Bot.* 49, 835–846. <https://doi.org/10.1093/oxfordjournals.aob.a086310>.
- Shah, T., Singh, O.P., Mukherji, A., 2006. Some aspects of South Asia's groundwater irrigation economy: analyses from a survey in India, Pakistan, Nepal Terai and Bangladesh. *Hydrogeol. J.* 14 (3), 286–309.
- 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. *Agric. For. Meteorol.* 152, 159–177. <https://doi.org/10.1016/j.agrformet.2011.09.009>.
- Tariq, M., Ahmad, S., Fahad, S., Abbas, G., Hussain, S., Fatima, Z., Nasim, W., Mubeen, M., Habib ur Rehman, M., Azam Khan, M., Adnan, M., Wilkerson, C.J., Hoogenboom, G., 2017. The impact of climate warming and crop management on phenology of sunflower-based cropping systems in Punjab, Pakistan. *Agric. For. Meteorol.* 256, 270–282. <https://doi.org/10.1016/j.agrformet.2017.07.012>.
- Vera, C.L., Duguid, S.D., Fox, S.L., Rashid, K.Y., Dribenki, J.C.P., Clarke, F.R., 2012. Short Communication: comparative effect of lodging on seed yield of flax and wheat. *Can. J. Plant Sci.* 92, 39–43. <https://doi.org/10.4141/cjps2011-031>.
- Vrieling, A., Meroni, M., Darvishzadeh, R., Skidmore, A.K., Wang, T., Zurita-Milla, R., Oosterbeek, K., O'Connor, B., Paganini, M., 2018. Vegetation phenology from Sentinel-2 and field cameras for a Dutch barrier island. *Remote Sens. Environ.* 1–13. <https://doi.org/10.1016/j.rse.2018.03.014>.
- Wilf, P., Zhang, S., Chikkerur, S., Little, S.A., Wing, S.L., Serre, T., 2016. Computer vision cracks the leaf code. *Proc. Natl. Acad. Sci.* 113, 3305–3310. <https://doi.org/10.1073/pnas.1524473113>.
- Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C.F., Gao, F., Reed, B.C., Huete, A., 2003. Monitoring vegetation phenology using MODIS. *Remote Sens. Environ.* 84, 471–475.