An automatic method based on daily in situ images and deep learning to date wheat heading stage

Kaaviya Velumani, Simon Madec, Benoit de Solan, Raul Lopez-Lozano, Jocelyn Gillet, Jeremy Labrosse, Stephane Jezequel, Alexis Comar, Frédéric Baret

An automatic method based on daily *in situ* images and deep learning to date wheat heading stage

Kaaviya Velumani[1,2], Simon Madec[2], Benoit de Solan[3], Raul Lopez-Lozano[2], Jocelyn Gillet[1], Jeremy Labrosse[1], Stephane Jezequel[3], Alexis Comar[1], Frédéric Baret[2]*

**Corresponding author: kvelumani@hiphen-plant.com; kaaviya.velumani@inra.fr*

[1]Hiphen SAS, 228, route de l'aérodrome – CS 40509, 84914 Avignon Cedex 9 – France

[2]INRAE, Avignon Université, UMR EMMAH, UMT CAPTE, 228, route de l'aérodrome – CS 40509, 84914 Avignon Cedex 9 – France

[3]Arvalis, 228, route de l'aérodrome – CS 40509, 84914 Avignon Cedex 9 – France

Abstract

Accurate and timely observations of wheat phenology and, particularly, of heading date are instrumental for many scientific and technical domains such as wheat ecophysiology, crop breeding, crop management or precision agriculture. Visual annotation of the heading date *in situ* is a labour-intensive task that may become prohibitive in scientific and technical activities where high-throughput is needed. This study presents an automatic method to estimate wheat heading date from a series of daily images acquired by a fixed RGB camera in the field. A convolutional neural network (CNN) is trained to identify

the presence of spikes in small patches. The heading date is then estimated from the dynamics of the spike presence in the patches over time. The method is applied and validated over a large set of 47 experimental sites located in different regions in France, covering three years with nine wheat cultivars. Results show that our method provides good estimates of the heading dates with a root mean square error close to 2 days when compared to the visual scoring from experts. It outperforms the predictions of a phenological model based on the ARCWHEAT crop model calibrated for our local conditions. The potentials and limits of the proposed methodology towards a possible operational implementation in agronomic applications and decision support systems are finally further discussed.

Keywords

Phenology, Internet of things for Agriculture, Convolutional Neural Networks, field sensors, phenology modelling

1 Introduction

Phenological observations are essential in agronomy, as crop management strategies (irrigation, fertilizing or crop protection) are planned considering plant development (Brown et al., 2005; Chmielewski, 2013). In wheat, the heading stage is one of the critical developmental phases as the plant becomes highly sensitive to abiotic stress – heat stress, frost, water constraints – with a strong impact on yield components (Slafer and Rawson, 1994). Several studies have quantified the effect of post-heading abiotic stress on yield (Ferris et al., 1998; Gooding et al., 2003; Wheeler et al., 1996) due to a significant decrease in the grain weight,, along with a fall in the number of grains per plant. More recently, Balla et al., (2019) analysed the possible impact of temperature on yield components at different development stages over a large wheat genotype panel, highlighting the high sensitivity of grain number to heat stress episodes around heading. Also, García et al., (2015) and Lobell and Ortiz-Monasterio, (2007) have shown

that the increase in minimum temperature during nights owing to changing climatic conditions accelerates the rate of crop development and reduces yield.

The timing and duration of stress through wheat developmental phases are thus essential to understand the impact of environmental factors on yield (Sadras and Slafer, 2012). Wheat phenology is driven by several eco-physiological mechanisms involving the response to temperature, photoperiod and vernalization (Gate, 1995) that are regulated by complex genetic pathways (Guedira et al., 2016; Whittal et al., 2018). Phenology has been traditionally one of the most important traits used to genetically improve wheat adaptation, matching crop development – particularly reproductive and grain-filling phases – to the optimal growing conditions of a target environment (Foulkes et al., 2011; Slafer, 2012). Indeed, phenology constitutes, in wheat breeding programs, one of the main levers enabling to optimize assimilates partitioning while reducing the impact of adverse weather events such as heat stress and frost during the grain-filling stage (Camargo et al., 2016; Chapman et al., 2012; Reynolds et al., 2009).

Accurate and timely observations of wheat phenology and, particularly, of heading date are, therefore, instrumental for many scientific and technical domains such as wheat ecophysiology, crop breeding, crop management or precision agriculture. Heading date is commonly scored visually *in situ* by operators that are frequently surveying the crops. It constitutes a labour-intensive task that requires skilled experts. For certain applications like high-throughput phenotyping in crop breeding programs on large genotype panels (Araus and Cairns, 2014; Cabrera-Bosquet et al., 2012) an accurate visual annotation of the heading date –with experts frequently visiting the field– can be difficult to achieve as the number of microplots increase geometrically. In other applications oriented towards crop management, phenological models may represent a valid alternative to *in situ* phenology observations (Bogard et al., 2014; White et al., 2008; Zheng et al., 2012), provided that those models are calibrated for each specific genotypes. However, such genotype-specific calibration necessitates extensive phenology observations from field experiments (Wallach et al., 2019).

The recent development of field sensors and unmanned platforms with imaging capabilities –including aerial (UAVs) and ground vehicles (UGVs)– have opened new avenues to monitor automatically crops in near-real time (Baret et al., 2018; Comar et al., 2012; Jay et al., 2017; Madec et al., 2017; White and Conley, 2013; Yang et al., 2017). In parallel, the advances in computational resources and data science achieved in the last years have fostered a significant breakthrough in computer vision and has paved the way to implement advanced algorithms to extract relevant information from high spatial resolution imagery. Deep learning algorithms including convolutional neural network (CNN), have shown excellent performances for object recognition (LeCun et al., 2015). These capabilities have favoured their progressive adoption in the fields of agronomy and phenomics (Kamilaris and Prenafeta-Boldú, 2018; Singh et al., 2018). For instance, they have been successfully used to detect and count individual cereal heads from RGB images (Hasan et al., 2018; Lu et al., 2017; Madec et al., 2019) and LiDAR data (Malambo et al., 2019). Nevertheless, the potential of such algorithms to provide accurate, automatic *in situ* estimations of crop phenology remains, up to now, underexploited. The use of well-known capabilities of CNNs to detect plant organs on individual images to derive crop phenology from image series needs to be further explored. At the time of writing this article, only very few studies have attempted to do that. The work of Yalcin (2018) proposing a CNN-based method for the discrimination of phenological stages for several crops, including wheat, seems promising, but no results on the absolute accuracy of the method have been provided. More recently Desai et al., (2019) have developed a deep learning approach to estimate the heading date on rice, but it was only tested over a small number of situations, preventing from drawing general conclusions about its performance under operational conditions.

The development of operational methods for automatic heading date detection using *in situ* images would enable to increase enormously the acquisition throughput at a reasonable cost. This would

represent an important contribution to the afore-mentioned scientific and technical domains, where frequent phenology observations are critical.

In that context, the objective of this study is to present an automatic method to estimate wheat heading date from daily high-resolution images taken in the field from fixed cameras. The method developed rely on deep learning techniques: daily images are interpreted using a CNN classifier that detects the presence of wheat spikes in the image, and the dynamics of the presence of spikes along the season permits to determine the heading date. A major aspect that differentiates this study against the existing works on this subject is the use of an extensive dataset of observations used to validate the method proposed, which enables to discuss its possible implementation under operational conditions. The dataset comprises 47 field plots sown with several soft and durum wheat cultivars in different regions of France, where fixed cameras were installed, and actual heading dates were annotated by experts. The performance of our CNN-based method to estimate the actual heading date are also concurrently compared with those of a phenology model based on ARCWHEAT (Weir et al., 1984) and calibrated for local conditions. The robustness of our method and its potential operational implementation are discussed with emphasis on possible limitations–e.g. image quality issues, environmental conditions during image acquisition– that may impact the performances.

2 Materials and Methods

2.1 Study sites

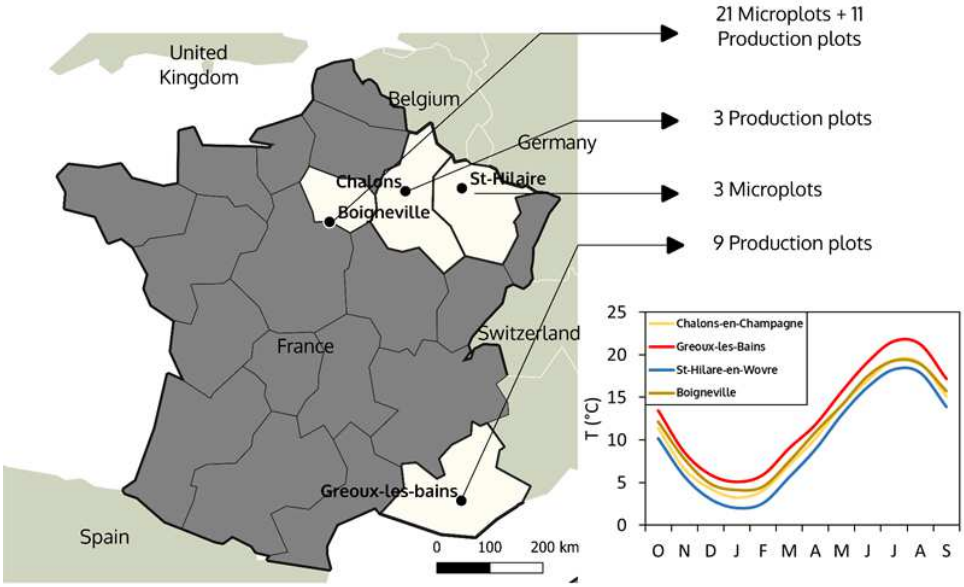


Figure 1. Location of the 4 regions where field sensors installed during the 2017, 2018 and 2019 growing seasons. The graph indicates the average monthly temperatures in the period 1986-2018 observed in 4 near weather stations from INRA and Meteo france: St. Gilles (Gréoux-les-Bains), Gif-sur-Yvette (Boigneville, only since 1999), Erneville (St-Hilare-en-Wovre), and Fagnieres (Chalons-en-Champagne).

This study was conducted during the years 2017, 2018 and 2019 in different commercial and experimental fields belonging to four contrasted agro-climatic regions around the following cities: Gréoux-les-Bains (43.8° N, 5.9° E) in the south-east of France, Boigneville (48.3° N 2.4° E) in the center of France; Chalons-en-Champagne (49.0° N, 4.4° E) and Saint-Hilaire-en-Woëvre (49.1° N, 5.7° E) in the north-east of France (Figure 1).

The climate in Greoux-les-Bains is Mediterranean (Kottek et al., 2006), with a maximum average temperature of 20°C, 690 mm of rainfall. In Boigneville, the climate is temperate and humid, with a maximum average temperature of 15.3^{o}C over the year and rainfall of 677 mm (Meteo France). The

climate in Chalons-en-Champagne is similar to that of Boigneville, whereas in Saint Hilaire-en-Wovre conditions are slightly colder, especially during winter (Figure 1), and more humid (average precipitation close to 1000 mm/year).

Among these four sites, 24 field sensors equipped with RGB cameras (see 2.2) were installed in microplots with a size of 10 x 2 meters belonging to larger experimental fields. The remaining 23 sensors were installed in production plots, with a size similar to a commercial field (around 800 x 200 meters). Often, the production plots are subdivided in homogeneous units with different cultivars or agro management. In the sites at the north of France, production plots and microplots are sown with winter soft wheat (*Triticum aestivum*) cultivars Descartes, Oregrain, Fructidor, RGT Sacramento, Matheo and Rebelde. In *Greoux-les-Bains*, the winter durum wheat (*Triticum durum*) cultivars RGT Voilur, Anvergur and Toscadou were grown. A summary of the 47 sites considered is given in Table 1. Please refer to Table A in Appendix for a detailed description of the sites, their location and average temperature over the growing season. The field sensors were installed in relatively homogenous areas of the fields which provided daily information over a footprint of about 10m^2 (see 2.2). During the installation, special attention was paid to orientate the camera field-of-view towards the centre of the microplot to prevent possible border effects.

Table 1 Description of the distribution of the 47 sites across years and regions. The number of sites available each year per region and year is indicated along with the corresponding number of cultivars.

Locations	**2017**		**2018**		**2019**	
	Sites	**Cultivars**	**Sites**	**Cultivars**	**Sites**	**Cultivars**
Gréoux-les-Bains	-	-	7	3	2	2
Boigneville	8	1	12	3	12	3
Chalons en Champagne	-	-	-	-	3	1

Saint_Hilaire en Woëvre	-	-	-	-	3	1

2.2 Acquisition of daily images of the canopy with IoTA systems

The field observation systems installed in the 47 sites were developed by Bosch and Hiphen (www.hiphen-plant.com/our-solutions/iot-field-sensor) and are named IoTA (Internet of Things for Agriculture). They consist of a telescopic pole placed vertically and equipped with an RGB camera as well as meteorological sensors (Figure 2a). The RGB camera takes one image (Figure 2b) each day at solar noon, and automatically uploads it to a cloud storage system through a GSM network. The image dimensions are 1024 x 768 pixels and is recorded in PNG format. The camera was set up at a height of approximately 1m above the top of the crop (installed after emergence), with a 45° inclination angle oriented in a compass direction perpendicular to the row. In some of the fields of 2017 and 2018, the length of the pole was adjusted mid-growing campaign to ensure that the camera was always well above the plant canopy. In the 2019 campaign, the height of the pole was fixed at 1.5 m to avoid this mid-campaign intervention. It has a field of view (FOV) of 55° x 41° providing a footprint of 10.8 m^2. Because of the relatively large FOV, the ground resolution is non-uniform throughout the image, particularly in the vertical direction (Figure 2b).

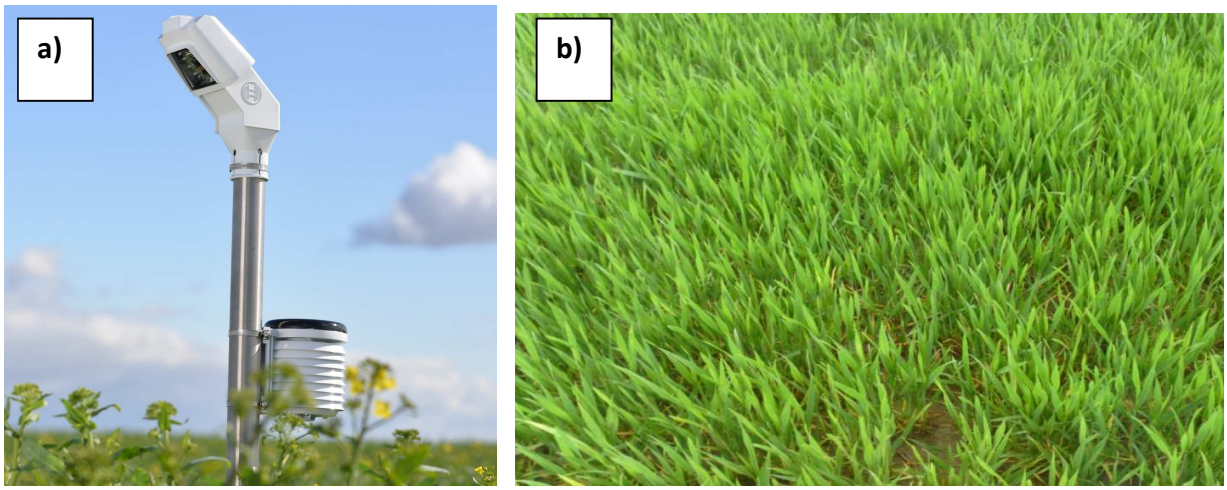


Figure 2 a) The IoTA systems installed in a field. The top part inclined at 45° hosts the RGB camera. The cylinder attached to the vertical pole includes sensors to monitor the local air temperature and moisture. b) An example of a raw image.

2.3 Heading date determined by experts

The heading stage corresponds to the emergence of the developing spike from the flag leaf sheath. (Bonnett, 1936; Zadoks et al., 1974). In the field, the heading stage is identified according to the definition given by Zadoks et al., (1974): 50% stems with spikes, at least, half-emerged, corresponding to the phase code 54 of the Zadoks scale. For the 27 sites monitored in 2017 and 2018 (Table 1), an online questionnaire was prepared with IoTA images covering a 9 to 13 consecutive days period approximately centered on the heading date. A panel of 14 experts was asked to identify the reference heading date for all the sites by applying the definition of Zadoks *et al.*, (1974) for heading stage: 50% stems with spikes at least half-emerged. The experts could view the images at their full resolution. Eight of the experts had more than 10 years of experience in working with wheat phenology and only four of them had less than two years of experience. The reference heading date for each of the 27 sites was considered as the average date of all the 14 experts. The standard deviation of the heading dates for each site was also calculated to quantify the variability of the expert replies.

In 2019, the actual heading dates were determined *in situ* by experts on all the 20 available sites (Table 1). They followed the definition by Zadoks *et al.* (1974) and visited the fields every two or three days. Note that this reference heading date might be different from the one derived from the 2017-2018 questionnaire since experts were scoring the heading date from images, not from direct observation of the crop in the field.

2.4 Heading date estimation from IoTA images and CNN

2.4.1 Image preparation

To get a more uniform ground pixel size, each image was first cropped into 1024 x 384 pixels by removing the top region of the image where the crop-sensor distance is too large, resulting in blurred objects (Figure 2b). Then 14 overlapping patches of size 256 x 256 pixels were extracted from the cropped image. The overlap between patches was 50% in either vertical or horizontal directions to minimize possible border effects. Working with patches permits to benefit from the full resolution of the IoTA image while avoiding memory issues.

2.4.2 Spikes Labelling

The patches from the 27 sites observed during the 2017 and 2018 growing seasons were labelled into two classes: “spikes present” or “spikes absent”. This represents a total of 40,500 patches out of which 17,000 were labelled as “spikes present” and 23,500 as “spikes absent”. All the patches belonging to images acquired until five days before the actual heading date determined by experts were automatically assigned to the “spikes absent” class. Similarly, the patches from the images acquired from five days after the actual heading date onwards were assigned to the “spikes present” class. Therefore, only those patches within a window of ±5 days around the actual heading date were visually attributed to their respective classes. Few patches with unclear assignation, such as emerging and sparse spikes were

excluded from the training dataset. A few examples belonging to the two classes can be found in Figure 3.

Figure 3 Samples extracted from patches belonging to the 9 wheat varieties monitored in our study. The patches were classified as 'spikes present' (top row) and 'spikes absent' (bottom row).

2.4.3 Identifying the presence of wheat spikes with the ResNet50

Wheat spikes were identified in the images using the ResNet50 (He et al., 2016), which obtained the best results in object detection at the ImageNet Large Scale Visual Recognition Challenge 2015 (Russakovsky et al., 2015). This network has a depth of 50 layers and uses residual blocks with identity mappings. The ResNet50 pre-trained on the ImageNet dataset which is available in the Keras Python deep learning library (Chollet, 2015) was used.

We replaced the original top layers of the pre-trained ResNet50 by two fully connected layers with dimensions, respectively of 512 and 1 to build a binary classifier (spikes present/absent). The network was re-trained with the labelled image patches by fine-tuning the weights of the entire network with a low learning rate to identify only the high-level features which were relevant to detect spikes and classify the patches as "spikes present" or "spikes absent". This strategy, called transfer learning, performs generally better than training the full network from scratch (Lee et al., 2015; Tajbakhsh et al., 2016). The

206 training dataset comprises the 27 sites available in 2017 and 2018. First-order data augmentation was
207 applied to the patches: translation, rotation, zoom, flip and changes in brightness levels. This improves
208 the generalization capacity of the neural network and increases the size of the training set at marginal
209 cost. Two different training and validation schemes were followed:

210 • **Twofold-cross validation on 2017 and 2018 sites**: the 27 sites available in 2017 and 2018 were
211 divided randomly into folds of 13 and 14 sites with approximately 19500 and 21000 labelled
212 patches in each fold. ResNet50 was re-trained on each of the two folds independently and
213 validated with the other one.
214 • **Independent validation on 2019 sites**: In this second scheme, a unique set of the 27 sites
215 corresponding to 40500 patches from the 2017 and 2018 sites is used for fine-tuning ResNet50.
216 The purpose of this scheme is to mimic the operational conditions when CNNs are trained with
217 data from the previous years: the CNN is trained on data acquired in years different from those
218 used for the validation.

219 In both schemes, 20% of the data were held back for performance testing at the end of each epoch of
220 the ResNet50 re-training. The “binary cross entropy loss” available in the Keras library was used as the
221 loss function. To avoid over-fitting, we reduced the learning rate by a factor of 0.5 when the validation
222 mean absolute error did not improve after three consecutive epochs and stopped the training when the
223 validation mean absolute error did not improve after five consecutive epochs. The pertinence of the
224 ResNet50 network was further evaluated using Gradient-weighted Class Activation Maps (Grad-CAM,
225 developed by Selvaraju *et al.* , 2016). These maps highlight the regions that contribute to the output
226 score using the gradient values input to the final convolutional layer (shown in Figure 6 and Figure 8).

2.4.4 Heading date estimation

The heading date is determined from the dynamics of the presence of spikes in the images. For each day, d, the fraction of patches per image classified by the CNN as 'spikes present' , $f_{head}(d)$, is calculated along the growing season (Figure 4). It provides a fair approximation of the Zadoks definition followed by the experts: the proportion of patches with emerged spikes in an image is a reasonable proxy to the proportion of stems with spikes emerged in the image. Then, a three-parameter logistic function is fitted to the time-series of $f_{head}(d)$ for every site:

$$f_{head}(d) = \frac{L}{1+e^{-k(d-d_0)}} \qquad \text{Eq. 1}$$

where L is the maximum value of $f_{head}(d)$, fixed by construction to $L = 1.0$ (Figure 4). The maximum growth rate, k, and d_0 are estimated using the Scipy Python package (Jones et al., 2001). Parameter d_0 represents the date when 50% of the patches have spikes.

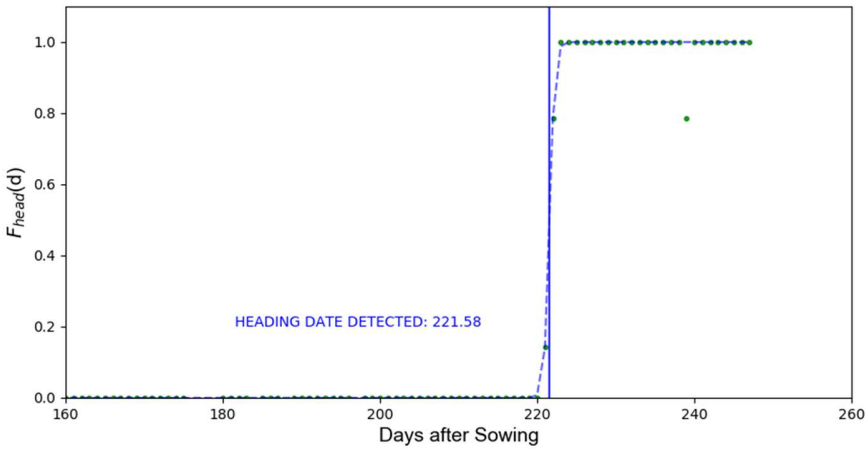


Figure 4 Typical dynamics of the fraction of patches with spikes (f_{head} shown as green dots). The logistic curve fitted to the dynamics of f_{head} is shown using dashed blue line. The heading date estimated from adjusted parameter d_0 is represented by the vertical blue line.

2.5 Heading date from ARCWHEAT model

A version of the phenology module of ARCWHEAT (Weir et al., 1984) adapted to the French local conditions by Gate (1995) was used. It is based on cumulated temperature with the effects of vernalization and photoperiod (Gouache et al., 2012). The model was run with actual daily temperatures collected by the Meteo France weather stations that are the closest to the sites in the period 2016-2019. Cultivar-specific parameters for vernalization and photoperiod corresponding to the nine cultivars observed in this study were adjusted using optimization algorithms based on independent field experiments (Thépot, 2014).

2.6 Performance metrics

The accuracy of the CNN classification was evaluated based on the overall accuracy (OA):

$$OA = \frac{T_p + T_n}{N} \quad \text{Eq. 2}$$

where T_p and T_n are, respectively, the number of patches correctly classified as "spikes present" (true positive) and "spikes absent" (true negative); and N is the total number of patches in the test dataset.

The root mean squared error (RMSE) was computed to quantify the errors between estimated and observed heading dates:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (HD_r - HD_e)^2}{N}} \quad \text{Eq. 3}$$

where HD_r is the reference heading date obtained from experts, HD_e is the heading date estimated by the indirect method (CNNs or phenology model) and N is the number of sites used.

3 Results

3.1 Accuracy of heading date estimates from the CNN model

The automatic method proposed estimated the heading date with a RMSE close to 2.0 days (Figure 5), as compared to the reference dates given by the experts. Moreover, the errors when the CNN is trained and validated with images from years 2017 and 2018 using a twofold cross-validation (Figure 5a) are similar to those obtained when validating the method against a completely independent set of images from year 2019 (Figure 5b). Moreover, the coefficient of determination R^2 is very close to 1 especially in 2019, where the variability in the time to heading among the plots monitored is large. In that year, heading was observed about 180 days after sowing for two durum wheat plots (varieties Anvergur and RGT Voilur) at Greoux-les-Bains, more than 20 days earlier than both varieties in 2018. This is explained by higher seasonal temperatures in the 2019 season compared to the 2018 season (seasonal average of 13.3 °C in 2019, against 10.1°C in 2018, see details in the Appendix Table A).

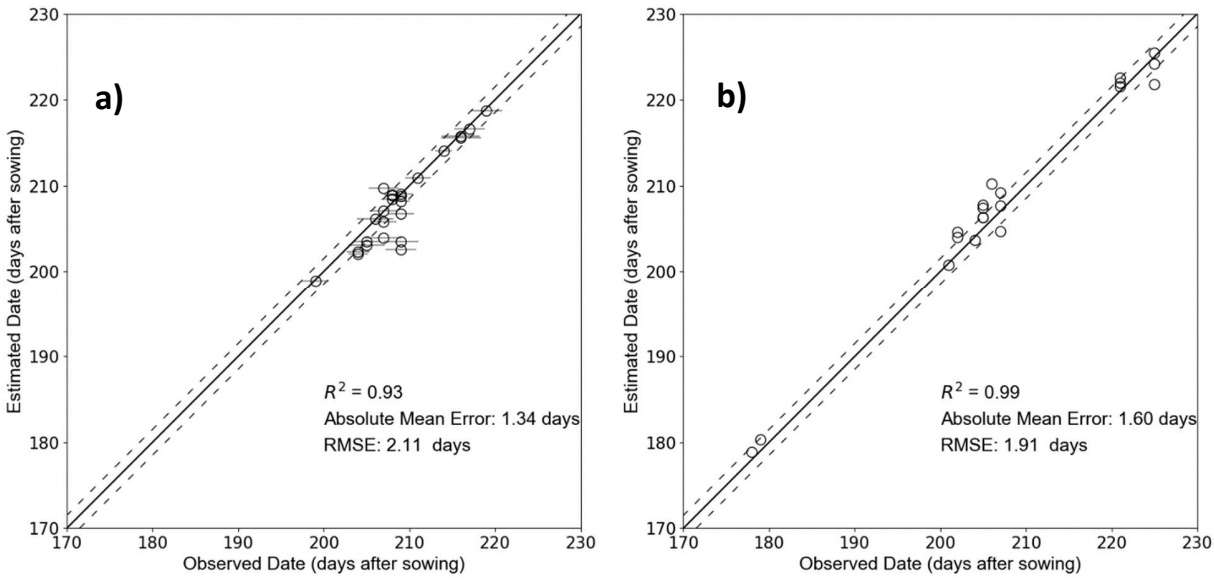


Figure 5 Comparison of reference heading dates and those estimated by our CNN-based method. a) 2017 and 2018, using a twofold cross-validation for years 2017 and 2018; b) independent validation over year 2019. Horizontal error bars represent the

standard deviation of the expert annotations for each site in 2017 and 2018. The dotted black lines represent the standard deviation (1.5 days) between experts when identifying the heading dates from photographs.

For two sites at Boigneville monitored in 2018 (Figure 5a), the CNN estimations show discrepancies reaching up to 6 days with the reference dates from the experts. The dynamics of spike appearance for these two sites are shown in Figure 6 along with the GradCAM (Gradient Class Activation Maps) which highlights the regions that influence the CNN prediction. In one of the sites (Figure 6a) the misclassification was due to the poor quality of the images: leaf blades appear bluish because of a poor white balance camera setup that was fixed only after day 203 (Figure 6a). These quality issues introduced substantial artifacts in the time-series that impacted the logistic curve fit. In the second site (Figure 6b), the water droplets on leaves observed on days 195 and 196 after sowing were wrongly identified by the CNN as spikes. Although the CNN seems to slightly over-detect the presence of spikes in this site (Figure 6b) even when no droplets were observed, the errors due to presence of droplets contributed to increase substantially the discrepancies against observed dates. Besides these two specific cases, the issues of CNN misclassification were marginal over the whole dataset. The cross-validation conducted with images from 2017 and 2018 revealed an overall accuracy of 98.45% for classifying individual patches as spikes present/absent. Moreover, the classification errors observed did not exhibit any systematic bias.

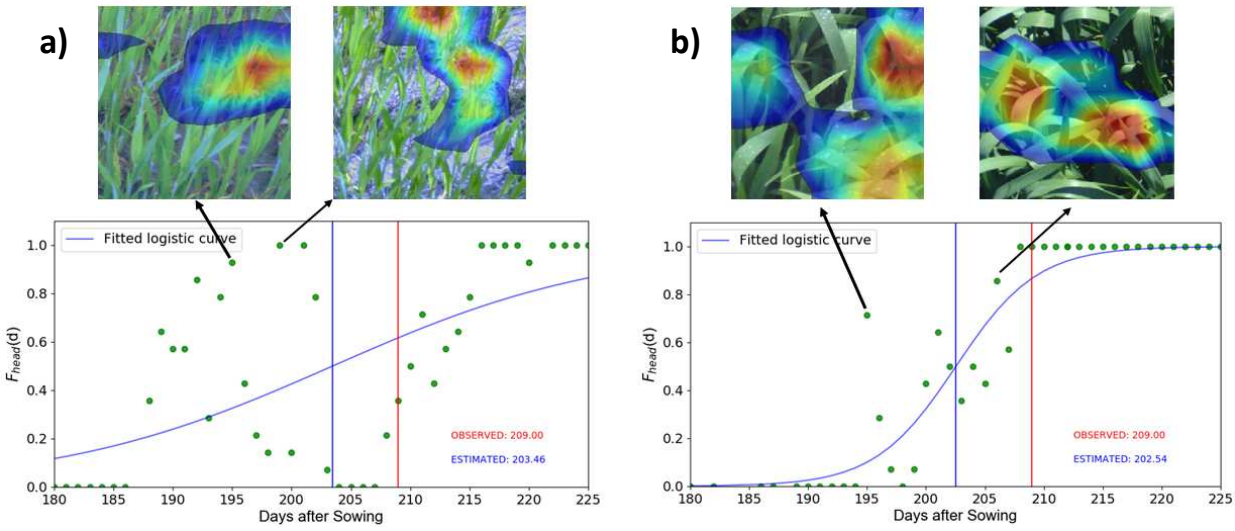


Figure 6 Dynamics of f_{head} for two sites in Boigneville in 2018 where large errors on heading date estimations by the CNN method were observed due to non-optimal image quality (a); image misclassification due to the presence of water droplets (b).

3.2 Accuracy of heading date prediction by the ARCWHEAT phenology model

The ARCWHEAT crop model, adapted to French conditions had an RMSE of 4.1 days to predict the heading date for the 47 sites monitored from 2017 to 2019 (Figure 7). This represents twice the error yielded by our CNN based method. A similar study conducted by Bogard *et al.* (2014), which evaluated the use of phenology models specifically calibrated for different cultivars to predict heading dates, reported errors comparable to those given by ARCWHEAT in this study. Other works where phenology models are applied at the regional or continental scale with no cultivar-specific calibration show even larger discrepancies reaching up to 20 days (McMaster and Smika, 1988; Ceglar *et al.*, 2019).

Although the ARCWHEAT model is capable to simulate the variability of the heading date among sites and years (R^2=0.92), it clearly underperformed when compared to the automatic method based on CNN. Only 37% of the sites were within the ± 1.5 days interval that represent the variability of the heading dates determined by the experts. Moreover the discrepancies between the model predictions and the

observed heading dates were above 10 days for two sites (Figure 7), which is not acceptable for precision agriculture or phenotyping applications.

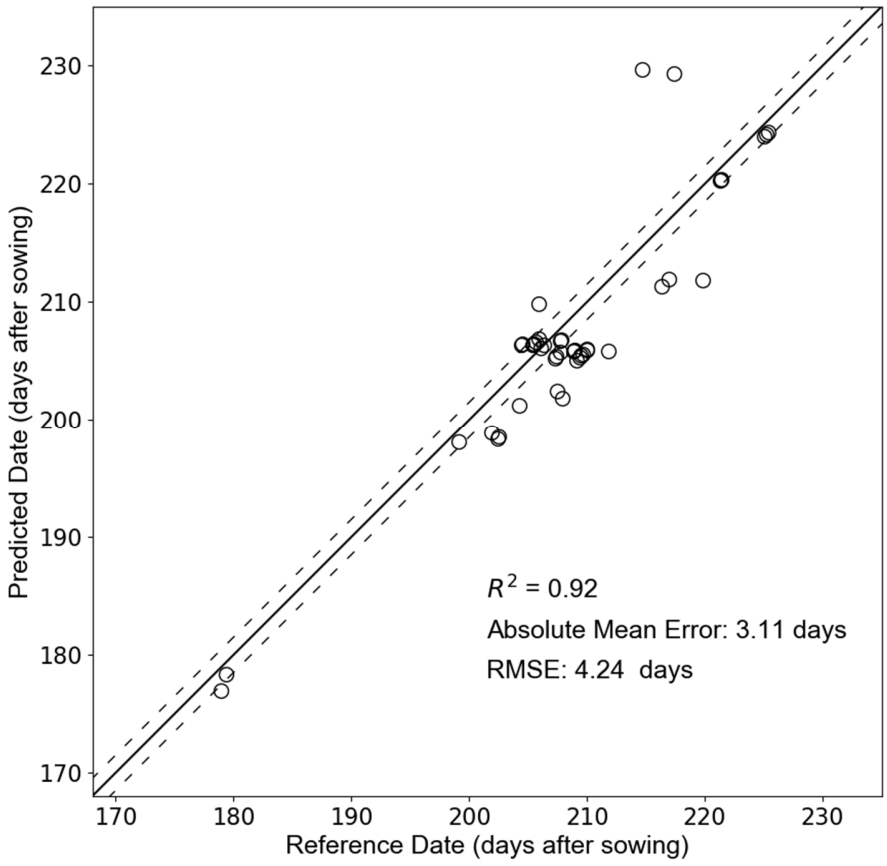


Figure 7 A comparison between the heading date predictions from the crop phenology model with the reference heading date for all the 47 sites in our study from the 3 growing seasons. The dashed lines around the 1:1 line corresponds to the 1.5 days interval. To ensure that all the 47 points are visible on the graph despite the overlap, a negligible random noise was added while plotting the points.

4 Discussion

4.1 Suitability of the automatic CNN-based methodology to estimate wheat heading date in operational applications

Our proposed CNN based method estimated heading dates with about 2 days uncertainties as evaluated over the 47 sites spanning across several regions, years and cultivars. The performances of the method proposed are considered satisfactory since the errors are close to the standard deviation of the expert panel replies provided for years 2017 and 2018. Moreover, this is close to the expected accuracy of an expert visiting the fields every two or three days to annotate crop development stage. The recent study of Desai et al., (2019) reported estimation errors between one and two days using also a CNN-based approach for paddy rice heading date, but the number of observations and sites were substantially smaller than in our study. Further, among the 20 sites monitored in the 2019 validation dataset, 6 of them were sown with cultivars that were not included in the training set of the CNN, indicating that the method is resistant to possible morphological differences among spikes between cultivars. The robustness of our method is essentially due to three factors: the reliability of the ResNet50 CNN to identify the presence of spikes in the patches; the statistics computed across patches within each image that smooth out individual errors; and the use of a logistic fit to determine the heading date from the daily statistics.

The ResNet50 CNN was used here to classify the patches with spikes. This approach was preferred to directly identifying spikes in the image and counting them (Madec et al., 2019). Using CNN for image classification increases the efficiency with which the training dataset is generated. Indeed, assigning individual images to the spikes present/absent classes is relatively straightforward: it permitted to generate a training dataset of more than 40,000 patches that contributed to improve the robustness of ResNet50. By contrast, annotating images for spikes identification or spikes counting is time-consuming.

335 Although CNN-based classification provides much less information than object counting/detection
336 algorithms, the results indicate that this approach is sufficient for heading date estimation. In Figure 8,
337 the first image (top row) spikes are not yet emerged and the Grad-CAM heat map shows low confidence,
338 whereas in the second one (bottom row) the presence of emerged spikes is obvious, and the probability
339 increases up to 0.82. ResNet50 may only detect the presence of a proportion of the spikes present in the
340 patch, but this appears sufficient to correctly classify it. This makes the classification approach more
341 robust for phenology identification as compared to the approaches based on object detection or
342 counting by regression, where the variability of spike size and shape among cultivars or environmental
343 conditions during image acquisition may severely affect the performances (Park et al., 2010). The
344 presence of water droplets induced only a moderate bias in heading date estimation in two of the sites
345 (Figure 6b). Images with droplets represent less than 1% of the training dataset cases. That was probably
346 not sufficient to teach the CNN to distinguish between droplets and spikes. These issues can only be
347 solved by increasing the variability by including more images taken under diverse environmental
348 conditions.

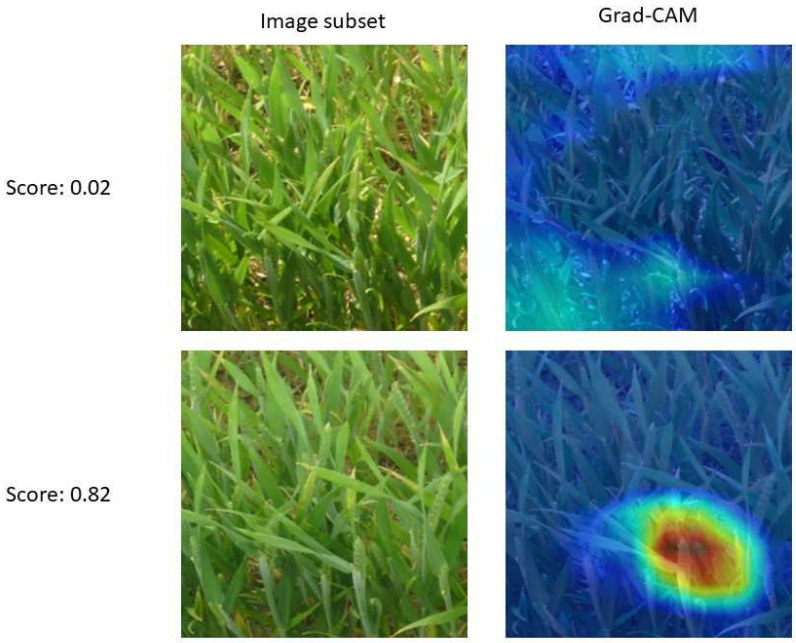


Figure 8 Patches extracted from two images of the same IoTA system and acquired on the reference heading date (top) and one day after (bottom) with their corresponding class activation maps. The confidence score provided by the CNN towards the presence of spikes is shown along with the Grad-CAM heat map which shows the areas which influence the CNN output

The logistic function smooths the daily values of $f_{head}(d)$, and reduces the influence of possible unsystematic classification errors on the heading date estimation. Furthermore, to minimize the possible impact of classification errors, in operational conditions it would be possible to prevent unrealistic estimations by imposing some constraints on the logistic model based on prior knowledge of the heading date:

$$\begin{aligned} f_{head}(d) &= 0.0 \quad if \quad d < d_{prior} - \delta \\ f_{head}(d) &= 1.0 \quad if \quad d > d_{prior} + \delta \end{aligned} \qquad \text{Eq. 4}$$

With d_{prior} being the prior value of the heading date derived from previous years observations or from phenological models, and δ being the associated maximum error. Anyhow, the use of a logistic function fit makes difficult to estimate the heading date in real time. The correct function fit is only possible *a posteriori*, i.e. when some images where $f_{head}(d) = 1.0$ have been already observed, which may only

363 happen 5 to 10 days after the actual heading date. This may limit the use of the method proposed on
364 operational applications of crop management requiring a rapid assessment of the heading date.

365 Compared to a phenological model calibrated under local conditions, our CNN-based method provides
366 better accuracy while not requiring ancillary information (e.g. sowing dates, variety-specific model
367 parameters) or daily weather data. This is an important advantage, since the accuracy achieved by
368 ARCWHEAT over the 47 sites monitored in this study (RMSE= 4 days) was mainly due to the cultivar-
369 specific calibration of model parameters. However, such a cultivar-specific calibration is time-consuming
370 and requires frequent observations throughout the crop cycle, repeated over several locations and years
371 (Cabelguenne et al., 1990; Jégo et al., 2010). Moreover, phenological models largely rely on the quality of
372 meteorological variables, which sometimes are interpolated from weather stations that are far from the
373 sites to be monitored without accounting for possible microclimatic effects (Joly et al., 2011; Monestiez
374 et al., 2001). In any case, mechanistic modelling of plant phenology remains always necessary on
375 prospective studies and to predict cultivar performance under a range of climatic scenarios. It also
376 constitutes an attractive alternative for heading date estimation for some applications where the
377 acquisition of high-throughput canopy images is not feasible. Phenological models can also largely
378 benefit from the method developed in this paper: the use of IoTA field systems and deep learning
379 approaches would substantially reduce the cost of calibration experiments and would permit also to
380 increase the environmental variability of the field trials giving access to frequent observations from
381 remote locations.

382 Our CNN based method constitutes a robust and cost-efficient approach for heading date estimation for
383 operational applications when daily images of the canopy are available, as it is the case for some high-
384 throughput phenotyping platforms. In those platforms, vectors such as unmanned ground vehicles and
385 hand-pushed carts are often used to frequently monitor the plant development and characterize
386 biophysical traits of the different cultivars using optical images (Deery et al., 2014; Mueller-Sim et al.,

2017; White and Conley, 2013). In such applications, the proposed automatic method showcased in this study could be directly integrated into other data processing pipelines at almost no cost to estimate heading date from the existing RGB images.

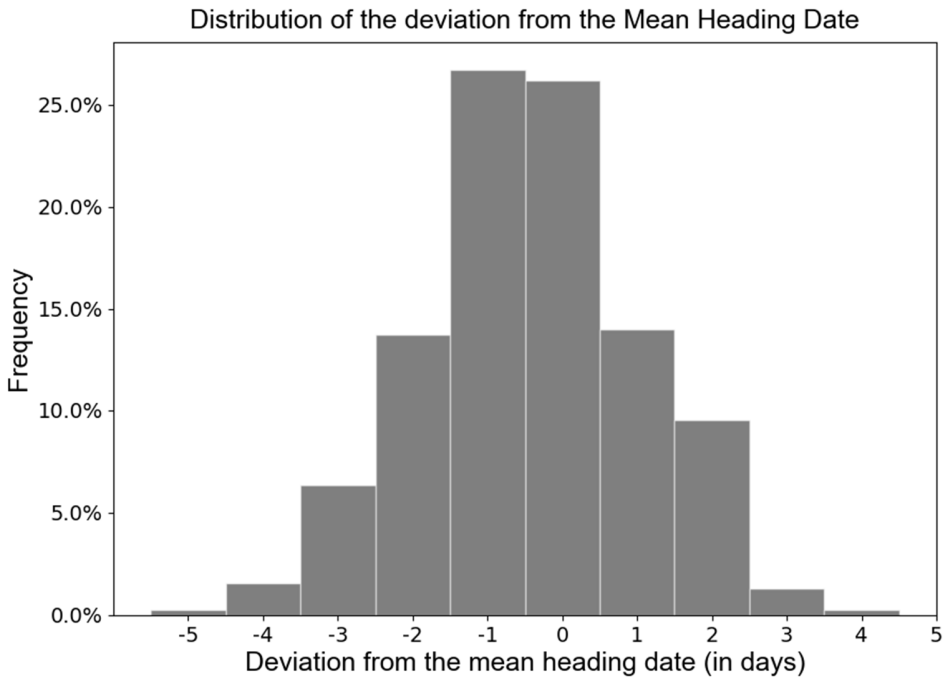
In agronomic or breeding applications where the phenology on distant fields need to be monitored –e.g. regional or national networks of agricultural fields– the whole system presented, including the fixed camera and the CNN-based method to estimate heading dates, may present important advantages in terms of costs compared to in situ visual annotation. The initial investment in each camera (including the pole, batteries and the hardware for data transmission) raises up to, approximately, 650 €; and the yearly costs of system maintenance and real-time data transmission by GSM is about 150 € per camera. The cost-efficiency –and the benefits in terms of environmental impact– of a system based on network of cameras compared to expert visits has to be determined case by case, and will largely depend on the distance between the fields to be monitored: the larger the network is, the more efficient remote observations are compared to field visits.

4.2 Subjectivity of visual annotations of the heading date from experts using RGB images

The visual determination of heading date is a task that includes some degree of subjectivity. For the 27 sites monitored in 2017 and 2018, the panel of 14 experts provided different estimates of the heading dates after inspecting visually the IoTA images. In most of the cases, the panel members selected 5 to 7 different dates per site out of those proposed in the questionnaire, although they were asked to follow the same definition of the heading date (Zadoks *et al.* 1974). The distribution of the deviations between individual replies and the average date for each site was roughly Gaussian with a standard deviation of 1.5 days (Figure 9), very close to the RMSE of our CNN based method proposed in this paper.

The subjectivity when determining visually heading date from images is obviously higher compared to *in situ* scoring. Issues regarding the image quality exacerbated the discrepancies among experts. According to the feedback provided by the panel, image saturation and blur due to a suboptimal camera setup made difficult to see the emergence of spikes in some of the images. Further, spikes were harder to identify in images taken under direct illumination conditions due to low image contrast, as well as for cultivars with awns.

The variability of expert replies evaluated in the 27 sites constitutes a good benchmark for our CNN based method, as image quality issues may also affect the identification of spikes by ResNet50 when classifying patches. The similarity between the 2 days RMSE associated to our CNN based method with the 1.5 days confidence interval of the expert date demonstrates that the performances of our CNN based method can be considered comparable to the expert reply when observing the same images.



420

421 *Figure 9 Distribution of the difference in days between the individual heading date identified by the experts and the mean*
422 *heading date value.*

423 5 Conclusion

424 In this study, we propose a CNN based method to estimate the heading date from daily images acquired
425 over wheat crops using an RGB camera fixed in the field. Images are processed per patch based on a
426 binary CNN-classifier to construct the dynamics of spikes appearance. Our method is easy to implement
427 since the labelling of patches is not time-consuming as compared to individual object annotation
428 required for other CNN models used for object identification or counting. The reliability of the CNN-
429 classifier to identify the presence of emerged spikes was marginally affected by the illumination
430 conditions and cultivar diversity, since the training dataset included images acquired under diverse
431 environmental conditions. Our method achieved satisfactory performances with RMSE≈2.0 days, which is
432 close to the uncertainties of expert annotations, and substantially better than phenological models
433 specifically calibrated for the cultivars monitored.

The proven robustness of the proposed method suggests a strong potential for cost-efficient operational applications in the field of phenotyping and agronomic applications. However, our method is limited by the image footprint close to 5 m² if only the bottom half of the image is used, as done in our study. However, the good consistency with expert observations taken over a larger sampling area demonstrates that our restricted sampling was sufficient. Nevertheless, the representativeness of such small footprint estimations to characterize phenology over large and heterogenous fields remains an open question for future works.

This method has been developed using a camera looking to the crop in a fixed position, but a similar approach could be transposed to time series of images from other vectors used in phenotyping experiments, such as unmanned ground and aerial vehicles, providing that the revisit time and resolution are sufficient. Further, the method could be adapted to identify other crop development stages associated with the identification of certain organs, such as the appearance of anthers for wheat to date flowering, or the appearance of tassels for the male flowering in maize.

Acknowledgements

The authors would like to thank the Association Nationale de Recherche Technologique (ANRT) for their grant CIFRE. The authors acknowledge Guy Deshayes (Arvalis) for his kind support with consolidating the reference data from the experts in the field.

6 References

Araus, J.L., Cairns, J.E., 2014. Field high-throughput phenotyping: the new crop breeding frontier. Trends Plant Sci. 19, 52–61. https://doi.org/10.1016/j.tplants.2013.09.008

Balla, K., Karsai, I., Bónis, P., Kiss, T., Berki, Z., Horváth, Á., Mayer, M., Bencze, S., Veisz, O., 2019. Heat stress responses in a large set of winter wheat cultivars (Triticum aestivum L.) depend on the timing

456 and duration of stress. PLoS One 14, e0222639. https://doi.org/10.1371/journal.pone.0222639

457 Baret, F., Madec, S., Irfan, K., Lopez, J., Comar, A., Hemmerlé, M., Dutartre, D., Praud, S., Tixier, M.H.,
458 2018. Leaf-rolling in maize crops: from leaf scoring to canopy-level measurements for phenotyping.
459 J. Exp. Bot. 69, 2705–2716. https://doi.org/10.1093/jxb/ery071

460 Bogard, M., Ravel, C., Paux, E., Bordes, J., Balfourier, F., Chapman, S.C., Le Gouis, J., Allard, V., 2014.
461 Predictions of heading date in bread wheat (Triticum aestivum L.) using QTL-based parameters of
462 an ecophysiological model. J. Exp. Bot. 65, 5849–5865. https://doi.org/10.1093/jxb/eru328

463 Bonnett, O.T., 1936. The development of the wheat spike. J. Agric. Res. 53, 445–451.

464 Brown, B., Westcott, M., Christensen, N., Pan, B., Stark, J., 2005. Nitrogen management for hard wheat
465 protien enhancement. Pacific Northwest Ext. Publ. 578, 1–14.

466 Cabelguenne, M., Jones, C.A., Marty, J.R., Dyke, P.T., Williams, J.R., 1990. Calibration and validation of
467 EPIC for crop rotations in southern France. Agric. Syst. 33, 153–171. https://doi.org/10.1016/0308-
468 521X(90)90078-5

469 Cabrera-Bosquet, L., Crossa, J., von Zitzewitz, J., Serret, M.D., Luis Araus, J., 2012. High-throughput
470 Phenotyping and Genomic Selection: The Frontiers of Crop Breeding ConvergeF. J. Integr. Plant Biol.
471 54, 312–320. https://doi.org/10.1111/j.1744-7909.2012.01116.x

472 Camargo, A. V., Mott, R., Gardner, K.A., Mackay, I.J., Corke, F., Doonan, J.H., Kim, J.T., Bentley, A.R., 2016.
473 Determining Phenological Patterns Associated with the Onset of Senescence in a Wheat MAGIC
474 Mapping Population. Front. Plant Sci. 7, 1540. https://doi.org/10.3389/fpls.2016.01540

475 Ceglar, A., van der Wijngaart, R., de Wit, A., Lecerf, R., Boogaard, H., Seguini, L., van den Berg, M., Toreti,
476 A., Zampieri, M., Fumagalli, D., Baruth, B., 2019. Improving WOFOST model to simulate winter
477 wheat phenology in Europe: Evaluation and effects on yield. Agric. Syst. 168, 168–180.

https://doi.org/10.1016/j.agsy.2018.05.002

Chapman, S.C., Chakraborty, S., Dreccer, M.F., Howden, S.M., 2012. Plant adaptation to climate change—opportunities and priorities in breeding. Crop Pasture Sci. 63, 251. https://doi.org/10.1071/CP11303

Chmielewski, F.-M., 2013. Phenology in Agriculture and Horticulture, in: Schwartz, M.D. (Ed.), Phenology: An Integrative Environmental Science. Springer Netherlands, Dordrecht, pp. 539–561. https://doi.org/10.1007/978-94-007-6925-0_29

Chollet, F., 2015. Keras. https://keras.io.

Comar, A., Burger, P., de Solan, B., Baret, F., Daumard, F., Hanocq, J.-F., 2012. A semi-automatic system for high throughput phenotyping wheat cultivars in-field conditions: description and first results. Funct. Plant Biol. 39, 914. https://doi.org/10.1071/FP12065

Deery, D., Jimenez-Berni, J., Jones, H., Sirault, X., Furbank, R., 2014. Proximal Remote Sensing Buggies and Potential Applications for Field-Based Phenotyping. Agronomy 4, 349–379. https://doi.org/10.3390/agronomy4030349

Desai, S.V., Balasubramanian, V.N., Fukatsu, T., Ninomiya, S., Guo, W., 2019. Automatic estimation of heading date of paddy rice using deep learning. Plant Methods 15, 76. https://doi.org/10.1186/s13007-019-0457-1

Ferris, R., Ellis, R.H., Wheeler, T.R., Hadley, P., 1998. Effect of High Temperature Stress at Anthesis on Grain Yield and Biomass of Field-grown Crops of Wheat. Ann. Bot. 82, 631–639. https://doi.org/10.1006/anbo.1998.0740

Foulkes, M.J., Slafer, G.A., Davies, W.J., Berry, P.M., Sylvester-Bradley, R., Martre, P., Calderini, D.F., Griffiths, S., Reynolds, M.P., 2011. Raising yield potential of wheat. III. Optimizing partitioning to

grain while maintaining lodging resistance. J. Exp. Bot. 62, 469–486. https://doi.org/10.1093/jxb/erq300

García, G.A., Dreccer, M.F., Miralles, D.J., Serrago, R.A., 2015. High night temperatures during grain number determination reduce wheat and barley grain yield: a field study. Glob. Chang. Biol. 21, 4153–4164. https://doi.org/10.1111/gcb.13009

Gate, P., 1995. Ecophysiologie du blé. Tec & Doc-Lavoisier, France.

Gooding, M.J., Ellis, R.H., Shewry, P.R., Schofield, J.D., 2003. Effects of Restricted Water Availability and Increased Temperature on the Grain Filling, Drying and Quality of Winter Wheat. J. Cereal Sci. 37, 295–309. https://doi.org/10.1006/jcrs.2002.0501

Gouache, D., Le Bris, X., Bogard, M., Deudon, O., Pagé, C., Gate, P., 2012. Evaluating agronomic adaptation options to increasing heat stress under climate change during wheat grain filling in France. Eur. J. Agron. 39, 62–70. https://doi.org/10.1016/j.eja.2012.01.009

Guedira, M., Xiong, M., Hao, Y.F., Johnson, J., Harrison, S., Marshall, D., Brown-Guedira, G., 2016. Heading Date QTL in Winter Wheat (Triticum aestivum L.) Coincide with Major Developmental Genes VERNALIZATION1 and PHOTOPERIOD1. PLoS One 11, e0154242. https://doi.org/10.1371/journal.pone.0154242

Hasan, M.M., Chopin, J.P., Laga, H., Miklavcic, S.J., 2018. Detection and analysis of wheat spikes using Convolutional Neural Networks. Plant Methods 14, 100. https://doi.org/10.1186/s13007-018-0366-8

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Jay, S., Maupas, F., Bendoula, R., Gorretta, N., 2017. Retrieving LAI, chlorophyll and nitrogen contents in

522 sugar beet crops from multi-angular optical remote sensing: Comparison of vegetation indices and
523 PROSAIL inversion for field phenotyping. F. Crop. Res. 210, 33–46.
524 https://doi.org/10.1016/j.fcr.2017.05.005

525 Jégo, G., Pattey, E., Bourgeois, G., Morrison, M.J., Drury, C.F., Tremblay, N., Tremblay, G., 2010.
526 Calibration and performance evaluation of soybean and spring wheat cultivars using the STICS crop
527 model in Eastern Canada. F. Crop. Res. 117, 183–196. https://doi.org/10.1016/j.fcr.2010.03.008

528 Joly, D., Brossard, T., Cardot, H., Cavailhes, J., Hilal, M., Wavresky, P., 2011. Temperature interpolation
529 based on local information: the example of France. Int. J. Climatol. 31, 2141–2153.
530 https://doi.org/10.1002/joc.2220

531 Jones, E., Oliphant, T., Peterson, P., others, 2001. SciPy: Open source scientific tools for Python.
532 https://www.scipy.org.

533 Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: A survey. Comput. Electron. Agric.
534 147, 70–90. https://doi.org/10.1016/j.compag.2018.02.016

535 Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World Map of the Köppen-Geiger climate
536 classification updated. Meteorol. Zeitschrift 15, 259–263. https://doi.org/10.1127/0941-
537 2948/2006/0130

538 LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.
539 https://doi.org/10.1038/nature14539

540 Lee, S.H., Chan, C.S., Wilkin, P., Remagnino, P., 2015. Deep-plant: Plant identification with convolutional
541 neural networks, in: 2015 IEEE International Conference on Image Processing (ICIP). IEEE, pp. 452–
542 456. https://doi.org/10.1109/ICIP.2015.7350839

543 Lobell, D.B., Ortiz-Monasterio, J.I., 2007. Impacts of day versus night temperatures on spring wheat

544 yields: A comparison of empirical and CERES model predictions in three locations, in: Agronomy
545 Journal. American Society of Agronomy, pp. 469–477. https://doi.org/10.2134/agronj2006.0209

546 Lu, H., Cao, Z., Xiao, Y., Zhuang, B., Shen, C., 2017. TasselNet: counting maize tassels in the wild via local
547 counts regression network. Plant Methods 13, 79. https://doi.org/10.1186/s13007-017-0224-0

548 Madec, S., Baret, F., de Solan, B., Thomas, S., Dutartre, D., Jezequel, S., Hemmerlé, M., Colombeau, G.,
549 Comar, A., 2017. High-Throughput Phenotyping of Plant Height: Comparing Unmanned Aerial
550 Vehicles and Ground LiDAR Estimates. Front. Plant Sci. 8, 1–14.
551 https://doi.org/10.3389/fpls.2017.02002

552 Madec, S., Jin, X., Lu, H., De Solan, B., Liu, S., Duyme, F., Heritier, E., Baret, F., 2019. Ear density
553 estimation from high resolution RGB imagery using deep learning technique. Agric. For. Meteorol.
554 264, 225–234. https://doi.org/10.1016/j.agrformet.2018.10.013

555 Malambo, L., Popescu, S.C., Horne, D.W., Pugh, N.A., Rooney, W.L., 2019. Automated detection and
556 measurement of individual sorghum panicles using density-based clustering of terrestrial lidar data.
557 ISPRS J. Photogramm. Remote Sens. 149, 1–13. https://doi.org/10.1016/j.isprsjprs.2018.12.015

558 McMaster, G.S., Smika, D.E., 1988. Estimation and evaluation of winter wheat phenology in the central
559 Great Plains. Agric. For. Meteorol. 43, 1–18. https://doi.org/10.1016/0168-1923(88)90002-0

560 Monestiez, P., Courault, D., Allard, D., Ruget, F., 2001. Spatial interpolation of air temperature using
561 environmental context: Application to a crop model. Environ. Ecol. Stat. 8, 297–309.
562 https://doi.org/10.1023/A:1012726317935

563 Mueller-Sim, T., Jenkins, M., Abel, J., Kantor, G., 2017. The Robotanist: A ground-based agricultural robot
564 for high-throughput crop phenotyping, in: Proceedings - IEEE International Conference on Robotics
565 and Automation. Institute of Electrical and Electronics Engineers Inc., pp. 3634–3639.

https://doi.org/10.1109/ICRA.2017.7989418

Park, D., Ramanan, D., Fowlkes, C., 2010. Multiresolution Models for Object Detection, in: Daniilidis, K., Maragos, P., Paragios, N. (Eds.), Computer Vision – ECCV 2010. Springer, Berlin, Heidelberg, pp. 241–254. https://doi.org/10.1007/978-3-642-15561-1_18

Reynolds, M., Foulkes, M.J., Slafer, G.A., Berry, P., Parry, M.A.J., Snape, J.W., Angus, W.J., 2009. Raising yield potential in wheat. J. Exp. Bot. 60, 1899–1918. https://doi.org/10.1093/jxb/erp016

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet Large Scale Visual Recognition Challenge. Int. J. Comput. Vis. 115, 211–252. https://doi.org/10.1007/s11263-015-0816-y

Sadras, V.O., Slafer, G.A., 2012. Environmental modulation of yield components in cereals: Heritabilities reveal a hierarchy of phenotypic plasticities. F. Crop. Res. 127, 215–224. https://doi.org/10.1016/j.fcr.2011.11.014

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization, in: The IEEE International Conference on Computer Vision (ICCV). pp. 618–626.

Singh, A.K., Ganapathysubramanian, B., Sarkar, S., Singh, A., 2018. Deep Learning for Plant Stress Phenotyping: Trends and Future Perspectives. Trends Plant Sci. 23, 883–898. https://doi.org/10.1016/j.tplants.2018.07.004

Slafer, G., 2012. Wheat development: its role in phenotyping and improving crop adaptation, in: Reynolds, M., Pask, A., Mullan, D. (Eds.), WHEAT Physiological Breeding I : Interdisciplinary Approaches to Improve Crop Adaptation. CIMMYT, Mexico, pp. 107–121.

Slafer, G., Rawson, H., 1994. Sensitivity of Wheat Phasic Development to Major Environmental Factors: a

588 Re-Examination of Some Assumptions Made by Physiologists and Modellers. Funct. Plant Biol. 21,
589 393. https://doi.org/10.1071/PP9940393

590 Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., Liang, J., 2016.
591 Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? IEEE Trans.
592 Med. Imaging 35, 1299–1312. https://doi.org/10.1109/TMI.2016.2535302

593 Thépot, S., 2014. Utilisation d'une population multi-parentale et hautement recombinante de blé tendre
594 pour l'étude de l'architecture génétique de la précocité de floraison. Util. d'une Popul. multi-
595 parentale hautement Recomb. blé Tendr. pour l'étude l'architecture génétique la précocité
596 floraison, Univ. Paris Sud (Paris 11)(2014). Université Paris Sud 11.

597 Wallach, D., Palosuo, T., Thorburn, P., Seidel, S.J., Gourdain, E., Asseng, S., Basso, B., Buis, S., Crout, N.,
598 Dibari, C., Dumont, B., Ferrise, R., Gaiser, T., Garcia, C., Gayler, S., Ghahramani, A., Hochman, Z.,
599 Hoek, S., Horan, H., Hoogenboom, G., Huang, M., Jabloun, M., Jing, Q., Justes, E., Kersebaum, K.C.,
600 Klosterhalfen, A., Launay, M., Luo, Q., Maestrini, B., Moriondo, M., Zadeh, H.N., Olesen, J.E., Poyda,
601 A., Priesack, E., Pullens, J.W.M., Qian, B., Schütze, N., Shelia, V., Souissi, A., Specka, X., Srivastava,
602 A.K., Stella, T., Streck, T., Trombi, G., Wallor, E., Wang, J., Weber, T.K.D., Weihermüller, L., Wit, A.
603 de, Wöhling, T., Xiao, L., Zhao, C., Zhu, Y., 2019. How well do crop models predict phenology, with
604 emphasis on the effect of calibration? bioRxiv 708578.
605 https://doi.org/https://doi.org/10.1101/708578

606 Weir, A.H., Bragg, P.L., Porter, J.R., Rayner, J.H., 1984. A winter wheat crop simulation model without
607 water or nutrient limitations. J. Agric. Sci. 102, 371–382.
608 https://doi.org/10.1017/S0021859600042702

609 Wheeler, T.R., Hong, T.D., Ellis, R.H., Batts, G.R., Morison, J.I.L., Hadley, P., 1996. The duration and rate of
610 grain growth, and harvest index, of wheat (Triticum aestivum L.) in response to temperature and

611 CO 2. J. Exp. Bot. 47, 623–630. https://doi.org/10.1093/jxb/47.5.623

612 White, J.W., Conley, M.M., 2013. A Flexible, Low-Cost Cart for Proximal Sensing. Crop Sci. 53, 1646.
613 https://doi.org/10.2135/cropsci2013.01.0054

614 White, J.W., Herndl, M., Hunt, L.A., Payne, T.S., Hoogenboom, G., 2008. Simulation-Based Analysis of
615 Effects of and Loci on Flowering in Wheat. Crop Sci. 48, 678.
616 https://doi.org/10.2135/cropsci2007.06.0318

617 Whittal, A., Kaviani, M., Graf, R., Humphreys, G., Navabi, A., 2018. Allelic variation of vernalization and
618 photoperiod response genes in a diverse set of North American high latitude winter wheat
619 genotypes. PLoS One 13, e0203068. https://doi.org/10.1371/journal.pone.0203068

620 Yalcin, H., 2018. Phenology recognition using deep learning: DeepPheno, in: 2018 26th Signal Processing
621 and Communications Applications Conference (SIU). IEEE, pp. 1–4.
622 https://doi.org/10.1109/SIU.2018.8404165

623 Yang, G., Liu, J., Zhao, C., Li, Zhenhong, Huang, Y., Yu, H., Xu, B., Yang, X., Zhu, D., Zhang, X., Zhang, R.,
624 Feng, H., Zhao, X., Li, Zhenhai, Li, H., Yang, H., 2017. Unmanned Aerial Vehicle Remote Sensing for
625 Field-Based Crop Phenotyping: Current Status and Perspectives. Front. Plant Sci. 8, 1111.
626 https://doi.org/10.3389/fpls.2017.01111

627 Zadoks, J.C., Chang, T.T., Konzak, C.F., 1974. A decimal code for the growth stages of cereals. Weed Res.
628 14, 415–421. https://doi.org/10.1111/j.1365-3180.1974.tb01084.x

629 Zheng, B., Chenu, K., Fernanda Dreccer, M., Chapman, S.C., 2012. Breeding for the future: what are the
630 potential impacts of future frost and heat events on sowing and flowering time requirements for
631 Australian bread wheat (Triticum aestivium) varieties? Glob. Chang. Biol. 18, 2899–2914.
632 https://doi.org/10.1111/j.1365-2486.2012.02724.x

633

634 Appendix

635 *Table A Description of the 47 sites considered in our study over 3 growing seasons and 9 different varieties of wheat*

Field Id	Coordinates	Year	Mean Seasonal Temperature Oct to July (°C)	Growing Degree Days °C (sowing to heading)	Heading date	Variety	Sowing date	Harvest date	Comments
D-D1	48.322452, 2.383738	2017	11.43	1518.15	17-05-2017	Descartes	20-10-2016	-	Experimental plots
D-D2	48.322483, 2.383812	2017	11.43	1518.15	16-05-2017	Descartes	20-10-2016	-	
D-D3	48.321520, 2.382034	2017	11.43	1518.15	17-05-2017	Descartes	20-10-2016	-	
D-I1	48.321570, 2.382026	2017	11.43	1518.15	17-05-2017	Descartes	20-10-2016	-	
D-I2	48.321530, 2.383126	2017	11.43	1518.15	16-05-2017	Descartes	20-10-2016	-	
D-I3	48.322499, 2.383748	2017	11.43	1518.15	17-05-2017	Descartes	20-10-2016	-	
D-S1	48.320248, 2.379639	2017	11.43	1518.15	18-05-2017	Descartes	20-10-2016	-	
D-S2	48.320221, 2.379579	2017	11.43	1518.15	19-05-2017	Descartes	20-10-2016	-	

E-A	44.238959, 5.928383	2018	7.45	1612.904	11-05-2018	RGT Voilur	25-10-2017	-	Production plots
E-ME	44.261415, 5.870530	2018	8.61	1841.848	06-05-2018	Anvergur	04-10-2017	-	
E-MO	44.262088, 5.869253	2018	8.61	1894.37	09-05-2018	Anvergur	04-10-2017	-	
E-MN	43.766323, 6.098798	2018	10.08	1780.29	17-05-2018	Toscadou	10-10-2017	-	
E-MS	43.765653, 6.100184	2018	10.15	1754.06	14-05-2018	Toscadou	10-10-2017	-	
E-S	43.812474, 5.772345	2018	11.22	1620.32	09-05-2018	RGT Voilur	22-10-2017	-	
E-V	43.791706, 6.120393	2018	10.08	1878.21	14-05-2018	Toscadou	10-10-2017	-	
A-201	48.323864, 2.378989	2018	10.06	1586.40	11-05-2018	Oregrain	16-10-2017	10-07-2018	Experimental plots with different nitrogen management
A-207	48.348859, 2.432189	2018	10.06	1575.05	10-05-2018	Oregrain	16-10-2017	10-07-2018	
A-210	48.345231, 2.432945	2018	10.06	1613.05	13-05-2018	Oregrain	16-10-2017	10-07-2018	
A-301	48.349030, 2.432504	2018	10.06	1586.40	11-05-2018	Oregrain	16-10-2017	10-07-2018	
A-305	48.323589, 2.378951	2018	10.06	1601.60	12-05-2018	Oregrain	16-10-2017	10-07-2018	

A-308	48.347419, 2.43389	2018	10.06	1586.40	11-05-2018	Oregrain	16-10-2017	10-07-2018	
A-310	48.344959, 2.432586	2018	10.06	1613.05	13-05-2018	Oregrain	16-10-2017	10-07-2018	
22-DS	48.325583, 2.386405	2018	10.06	1624.20	14-05-2018	Fructidor	20-10-2017	09-07-2018	Production plots
22-SS	48.324231, 2.385779	2018	10.06	1639.85	15-05-2018	Fructidor	20-10-2017	09-07-2018	
331-MS	48.321029, 2.384333	2018	10.06	1613.05	13-05-2018	RGT Sacramento	20-10-2017	09-07-2018	
331-SS1	48.319074, 2.380753	2018	10.06	1601.60	12-05-2018	RGT Sacramento	20-10-2017	09-07-2018	
331-SS2	48.319046, 2.380693	2018	10.06	1601.60	12-05-2018	RGT Sacramento	20-10-2017	09-07-2018	
SI-1	49.073911, 5.704641	2019	8.80	1582.76	27-05-2019	Fructidor	14-10-2018	-	
SI-2	49.073735, 5.704535	2019	8.80	1582.76	27-05-2019	Fructidor	14-10-2018	-	
SI-3	49.074086, 5.704128	2019	8.80	1582.76	27-05-2019	Fructidor	14-10-2018	-	
C-B1	48.958538, 4.253838	2019	9.48	1671.19	23-05-2019	Matheo	14-10-2018	-	
C-B2	48.958576,	2019	9.48	1671.19	23-05-2019	Matheo	14-10-2018	-	

	4.253507								
C-B3	48.95826, 4.253713	2019	9.48	1671.19	23-05-2019	Matheo	14-10-2018	-	
T-217	48.323273, 2.379553	2019	9.62	1529.25	16-05-2019	Oregrain	23-10-2018	-	Experimental plots with different nitrogen management
T-205	48.323624, 2.379038	2019	9.62	1529.25	16-05-2019	Oregrain	23-10-2018	-	
T-202	48.323761, 2.378927	2019	9.62	1543.0	17-05-2019	Oregrain	23-10-2018	-	
T-305	48.323589, 2.378951	2019	9.62	1529.25	16-05-2019	Oregrain	23-10-2018	-	
T-201	48.323864, 2.378989	2019	9.62	1557.70	18-05-2019	Oregrain	23-10-2018	-	
T-110	48.323586, 2.379289	2019	9.62	1529.25	16-05-2019	Oregrain	23-10-2018	-	
332-DS	48.32201, 2.383564	2019	9.62	1518.65	16-05-2019	Rebelde	24-10-2018	-	Production plots
332-SS	48.319752, 2.37894	2019	9.62	1558.70	19-05-2019	Rebelde	24-10-2018	-	
332-MS	48.320652, 2.381057	2019	9.62	1558.70	19-05-2019	Rebelde	24-10-2018	-	
411-SS	48.317192, 2.387024	2019	9.62	1570.45	27-05-2019	Fructidor	06-11-2018	-	
411-DS	48.318287, 2.388875	2019	9.62	1553.5	26-05-2019	Fructidor	06-11-2018	-	
411-	48.316154,	2019	9.62	1570.45	27-05-2019	Fructidor	06-11-2018	-	

MS	2.384733								
E-F1	43.710176, 4.534401	2019	13.31	1854.88	21-04-2019	RGT Voilur	25-10-2018	-	
E-M1	43.843360, 4.442453	2019	13.12	1874.5	22-04-2019	Anvergur	25-10-2018	-	

636