

Article

Long-Term Hindcasts of Wheat Yield in Fields Using Remotely Sensed Phenology, Climate Data and Machine Learning

Fiona H. Evans ^{1,2,*}  and Jianxiu Shen ¹ ¹ Centre for Crop and Food Innovation, Food Futures Institute, Murdoch University, 90 South Street, Murdoch, WA 6150, Australia; Jenny.Shen@murdoch.edu.au² Centre for Digital Agriculture, Centre for Crop and Disease Management, Curtin University, Kent Street, Bentley, WA 6102, Australia

* Correspondence: Fiona.Evans@murdoch.edu.au

Abstract: Satellite remote sensing offers a cost-effective means of generating long-term hindcasts of yield that can be used to understand how yield varies in time and space. This study investigated the use of remotely sensed phenology, climate data and machine learning for estimating yield at a resolution suitable for optimising crop management in fields. We used spatially weighted growth curve estimation to identify the timing of phenological events from sequences of Landsat NDVI and derive phenological and seasonal climate metrics. Using data from a 17,000 ha study area, we investigated the relationships between the metrics and yield over 17 years from 2003 to 2019. We compared six statistical and machine learning models for estimating yield: multiple linear regression, mixed effects models, generalised additive models, random forests, support vector regression using radial basis functions and deep learning neural networks. We used a 50-50 train-test split on paddock-years where 50% of paddock-year combinations were randomly selected and used to train each model and the remaining 50% of paddock-years were used to assess the model accuracy. Using only phenological metrics, accuracy was highest using a linear mixed model with a random effect that allowed the relationship between integrated NDVI and yield to vary by year ($R^2 = 0.67$, MAE = 0.25 t ha⁻¹, RMSE = 0.33 t ha⁻¹, NRMSE = 0.25). We quantified the improvements in accuracy when seasonal climate metrics were also used as predictors. We identified two optimal models using the combined phenological and seasonal climate metrics: support vector regression and deep learning models ($R^2 = 0.68$, MAE = 0.25 t ha⁻¹, RMSE = 0.32 t ha⁻¹, NRMSE = 0.25). While the linear mixed model using only phenological metrics performed similarly to the nonlinear models that are also seasonal climate metrics, the nonlinear models can be more easily generalised to estimate yield in years for which training data are unavailable. We conclude that long-term hindcasts of wheat yield in fields, at 30 m spatial resolution, can be produced using remotely sensed phenology from Landsat NDVI, climate data and machine learning.

Keywords: Landsat; NDVI; crop phenology; yield estimation; long-term; hindcasts; seasonal climate metrics; machine learning



Citation: Evans, F.H.; Shen, J. Long-Term Hindcasts of Wheat Yield in Fields Using Remotely Sensed Phenology, Climate Data and Machine Learning. *Remote Sens.* **2021**, *13*, 2435. <https://doi.org/10.3390/rs13132435>

Academic Editors: Roger Lawes and Aniruddha Ghosh

Received: 20 May 2021

Accepted: 17 June 2021

Published: 22 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Crop management decisions depend on our understanding of how changes to management will affect yield, but management is only one of many determinants of yield. Crop yield varies in space and time according to soil types, local weather conditions and seasonal climate variability as well as management. Moreover, spatial and temporal variation in yield is often much larger than effects of management. Precision agriculture (PA) aims to optimise crop management across farms and fields to sustainably improve yield and profit [1,2]. While many sources of information inform PA decision-making, the primary source is geographically referenced yield data recorded by yield monitors mounted on harvesting machinery [3]. While yield monitors are standard in most modern harvesting

equipment, many farmers do not have data for a sufficiently long period to understand how seasonal climate conditions interact with spatially -varying influences on yield. Use of yield monitor data is also hampered by lack of data inter-operability and standards [4], and because poor Internet connectivity in rural areas means that transferring yield maps from farms cannot be easily automated [5].

Satellite remote sensing offers a cost-effective means of generating comprehensive, long-term yield maps that can be used instead of yield monitor data to understand spatial-temporal variation in yield. Crop growth can be monitored using vegetation indices (VIs) that combined spectral reflectance measurements into a single index that reflects biophysical characteristics of the crop canopy, such as greenness, biomass and leaf area index (LAI) [6]. For example, the normalised difference vegetation index (NDVI) measures the ratio of the difference between the near-infrared and red reflectances and their sum [7].

Land surface phenology considers the development of crops using time-sequences of remotely sensed VIs [8–11]. Yield estimation from land surface phenology then aims to identify spatial and interannual variation in phenology and use it to map or predict yield. Early work simply used the time within the detected growing season for which a VI had maximum correlation with yield [12,13]. However, this neglected general knowledge that yield is influenced by the timing of ‘true’ phenological events, such as germination, flowering and senescence. Higher yields are associated with earlier emergence, longer growing season and longer green leaf area duration [14–18]. Using this knowledge, various metrics have been derived from sequences of VIs to better estimate yield. These include the timing and duration of growth stages identified from the VI sequence [19,20], peak VI [21] and time-integrated VIs [22–27]. Use of phenological metrics for yield estimation is further supported by Waldner, et al. [28], who showed that linear regression of LAI metrics derived from simulated phenology can explain between 30 and 78% of simulated grain yield variability in the ‘crop-model’ space.

Recent approaches for improving on yield estimates made directly from phenological metrics include assimilation of remotely sensed data into simple crop models [29,30] and use of machine learning models to combine phenological metrics with climate data [24,31,32]. These studies aimed to estimate yield over large, national-scale areas and used data from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite with spatial resolution of 250 m or 500 m. This is adequate for regional-scale estimation but is too coarse to support PA decision-making about how to optimise crop management in fields. In contrast, the Landsat satellite series is ideal for yield estimation to support PA. It has a long-term record of archived images dating back to 1982 [33], with a 16-day revisit period and global coverage at 30 m resolution. Because Landsat data have been freely available since 2008 [34], they can be used to generate inexpensive information for farmers. Landsat data are used operationally to map and monitor land condition and land cover change over large areas in Australia [35,36], North America [37,38] and South America [39]. In the United States, the Landsat-derived Cropland Data Layer (CDL) released in 2009 [40] is still in operational use. Recently, Landsat data were used to retrospectively map crops prior to the release of the CDL, from as early as 1984 [41].

With the goal of producing long-term hindcasts of yield at within-field resolution to support PA decision-making, we investigate the use of sequences of Landsat NDVI for estimating wheat yields in a study area in Western Australia (WA). While many VIs are used for remotely sensed phenology detection, including the enhanced vegetation index [42], the wide dynamic range vegetation index [43] and more [9], we use the NDVI because it is well-understood by farmers and is frequently used by farm managers for crop monitoring. However, use of Landsat NDVI for yield estimation is limited by the presence of cloud and cloud shadows, which occur most frequently during wet periods that drive crop growth [44–47]. Because MODIS has higher temporal frequency and more cloud-free images available, methods developed for yield estimation using MODIS-derived phenology may not work for Landsat data. Spatially weighted growth curve (SWG) estimation is a new method designed specifically for Landsat NDVI to fill spatial and

temporal gaps caused by cloud contamination during the growing season [48]. SWGC estimates crop growth curves using data from a local neighbourhood around each cell, where the data are weighted according to geographical distance from the central cell. It combines spatial smoothing by use of spatial weights with temporal smoothing by growth curve estimation to improve estimation of land surface phenology from Landsat NDVI. Because it is not dependent on individual cells having sufficient cloud-free images within the growing season, SWGC enables phenology detection at more cells than non-spatial approaches which typically exclude cells with insufficient observations (e.g., [44,49]).

To support decision-making about how to optimise crop management within fields, this study aims to determine whether phenological and seasonal climate metrics obtained from SWGC estimation have utility for estimating wheat yield. We use the SWGC estimated growth curves to identify the timing of phenological events and derive phenological metrics that describe the timing and degree of crop growth stages occurring at each cell. Detected phenology is then combined with daily weather data to produce seasonal climate metrics that describe the water availability and growing degree days during different growth stages of the crop. We aim to use the metrics as predictors of wheat yield. Previous studies that combined phenological metrics with climate data to predict yield reported large differences in accuracy using different statistical and machine learning models [24,31,32]. We therefore compare six of these models for our purpose: multiple linear regression, linear mixed models, generalised additive models, random forests, support vector regression using radial basis functions and deep learning.

The specific objectives of this study are to: (1) Investigate relationships between phenological and seasonal climate metrics derived from Landsat NDVI using SWGC estimation with wheat yield; (2) assess and compare statistical and machine learning models for estimating wheat yield using phenological metrics as predictors; (3) quantify improvements in accuracy of yield estimation when seasonal climate metrics are also used as predictors; and (4) identify an optimal model and produce long-term hindcasts of wheat yield at 30 m resolution.

2. Materials and Methods

2.1. Data

2.1.1. Study Area

The study area is located in the Western Australian grainbelt at around 31.28 S and 118.16 E (Figure 1a). It is approximately 17,000 ha in size. Grain crops are grown in a dryland system that is heavily reliant on winter rainfall during the May to October growing season. Average growing season rainfall is between 200 and 300 mm with large inter-annual variability (Table 1). Wheat is the main crop grown, with barley, lupins, canola and pasture included in cropping rotations. Besides crop fields, the study area includes areas of remnant vegetation (native trees and shrubs) and salt scalds caused by dryland salinity.

2.1.2. Landsat Data

All bands of Landsat-7 (path/row: 111/082) data from 2003 to 2019 were obtained from the United States Geological Survey (USGS) web site (<https://earthexplorer.usgs.gov> accessed on 2 February 2021). There were 366 Landsat-7 images available during 2003 to 2019. Data pre-processing including geometric correction, top of atmosphere reflectance correction and surface reflectance correction were completed using the USGS Earth Resources Observation and Science (EROS) Centre Science Processing Architecture (ESPA) online interface (<https://espa.cr.usgs.gov> accessed on 3 February 2021). Cells with cloud contamination or cloud shadow were removed using the cell quality control band. The corrected red and near-infrared reflectances were used to calculate NDVI for all available image dates.

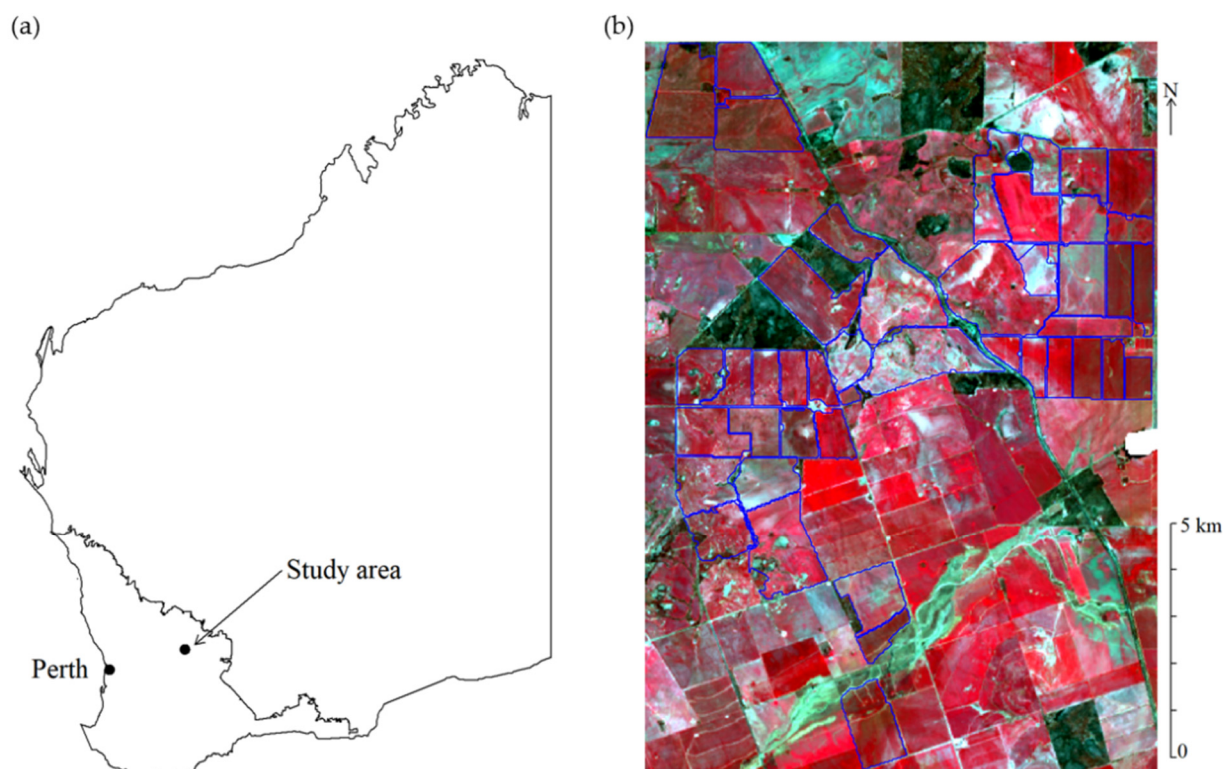


Figure 1. The study area: (a) Study area location in the eastern grainbelt of Western Australia; and (b) false-colour composites of the near-infrared, red and green bands of the Landsat-8 OLI image acquired 29 August 2019 with 44 paddocks with wheat yield data overlaid in blue.

Table 1. Number of wheat paddocks, area of wheat grown and growing season (May to October) rainfall for each year from 2003 to 2019.

Year	Number of Wheat Paddocks	Area of Wheat Grown (ha)	Growing Season Rainfall (mm)
2003	7	629	327
2004	18	1812	265
2005	23	2028	264
2006	17	1572	270
2007	9	932	178
2008	10	722	275
2009	43	4284	225
2010	31	2549	141
2011	28	2364	353
2012	35	3168	178
2013	32	3176	275
2014	31	2695	215
2015	36	2674	224
2016	29	2554	284
2017	30	2868	218
2018	30	2849	235
2019	17	1385	171

2.1.3. Wheat Yield Data

Yield monitor data were obtained for 44 paddocks (5281 ha) across the study area (Figure 1b) during 2003 to 2019. The yield data were cleaned to remove high (above the 99th percentile) and low (below the 1st percentile) outliers and kriged to produce 30-m resolution raster maps using the R ‘gstat’ package [50,51]. The number of wheat paddocks and the total area of wheat grown vary from year to year (Table 1). The total number of paddock-year combinations is 426. Yields ranged from 0 to 4 tonnes per hectare ($t\ ha^{-1}$)

with substantial spatial and year to year variation. The total number of 30-m resolution cells containing yield data during the 17-year period was 421,766.

2.1.4. Climate Data

Point-source and gridded weather data, at 5-km resolution, are available for Australia via the Long Paddock SILO data base (<https://www.longpaddock.qld.gov.au/silo> accessed on 3 February 2021). Because wheat yield in dryland cropping systems can be highly influenced by intense winter rainfall events, we use point-source weather data from an actively recording weather station instead of gridded data. Daily point-source weather data from 2002 to 2019 were obtained from the Long Paddock ‘Patched Point’ Database for the nearest recording weather station to the study area, Nungarin (ID = 10112, 31.18 S and 118.10 E). ‘Patched Point’ data have had temporal data gaps filled with an estimate obtained by interpolating data from surrounding weather stations. Consequently, they form a complete daily data record with no missing data.

2.2. Methods

2.2.1. Spatially Weighted Growth Curve Estimation

The spatially weighted growth curve (SWGC) estimation method estimates growth curves for each year using data from a spatial neighbourhood around each cell, where the data are weighted according to distance from the central cell using a distance-decay kernel [48]. Spatial weights are applied using a truncated or ‘moving window’ Gaussian kernel, where the distance weights are set to zero for all cells (x, y) outside of a rectangular region centred on the location for which the growth curve is being estimated. The moving window size is specified by $MAXD$ and has width and length equal to $(2 * MAXD + 1)$ metres (m). Within the window, distance weights are given by:

$$w_{x,y} = \exp\left(-\frac{1}{2}\left(\frac{d_{x,y}}{b}\right)^2\right) \quad (1)$$

where $w_{x,y}$ is the weight for cell (x, y) , $d_{x,y}$ is the distance of (x, y) from the location for which the growth curve is to be estimated and b is the Gaussian kernel bandwidth. To ensure filling of spatial gaps in Landsat-7 data caused by the scanline corrector failure in 2003, we use bandwidth $b = 60$ m and $MAXD = 200$ m. This bandwidth ensures sufficient data with non-zero weights in the estimation of SLC failure gaps. $MAXD$ is chosen to be as small as possible while retaining all data with weights within three decimal places from zero.

Crop growth is modelled using an asymmetric double Lorentz function for the two main phases of crop development as measured by NDVI, which we refer to as the vegetative phase (from germination to maximum vegetative growth) and the grain-fill phase (from maximum growth to senescence). The curve takes the form:

$$y = \begin{cases} c + (d - c) / (1 + b(x - e)^2), & x \leq e \\ c + (d - c) / (1 + b(x - f)^2), & x > e \end{cases} \quad (2)$$

where x is the day of year, $0 \leq c \leq 0.9$ is the minimum NDVI, $0.1 \leq d \leq 1$ is the maximum NDVI, $0 \leq e \leq 260$ is the day of year at which maximum NDVI is observed and b and f are the shape parameters for the curve before and after maximum NDVI is observed.

Figure 2a demonstrates SWGC estimation using spatial weights and an asymmetric double Lorentz function.

2.2.2. Phenological Metrics

Phenological metrics (PM) are derived from the estimated SWGC growth curves as follows. We define the start of growing season (SOS) and end of growing season (EOS) as the days of the year when NDVI is 20% higher than minimum NDVI (c in

Equation (2)). The peak of season (POS) is the day of the year that NDVI is at its maximum (e in Equation (2)). Peak of season vegetative growth (POSV) is identified by maximum NDVI (d in Equation (2)). Three integrated NDVI measures are created that sum NDVI in three stages of the crop growth cycle: iNDVI for the whole growing season (SOS to EOS), VLAD for the vegetative stage (SOS to POS) and GLAD for the grainfill phase (POS to EOS). Growth periods and metrics are shown in Figure 2b.

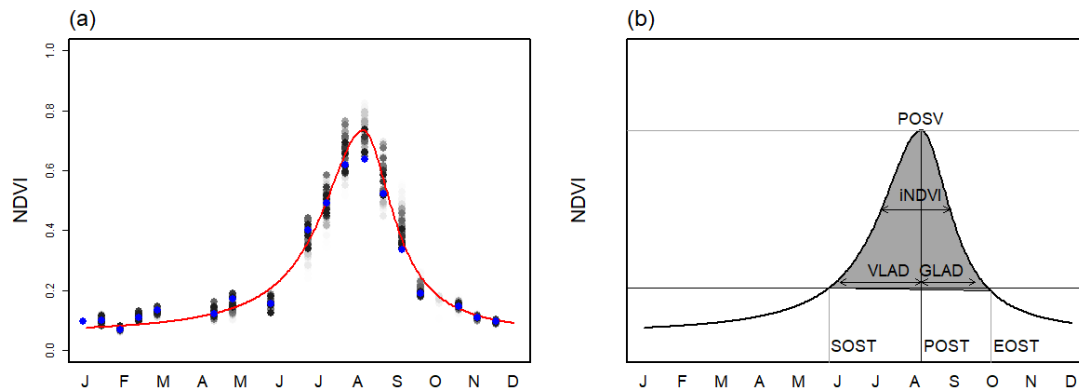


Figure 2. Demonstration of: (a) Spatially weighted growth curve estimation, where the estimated growth curve is shown in red, NDVI data used in the estimation are shown in shades of grey according to their spatial weight (white lowest and black highest) and NDVI data for the cell are shown in blue and (b) calculation of phenological metrics.

2.2.3. Seasonal Climate Metrics

Seasonal climate metrics (SCM) are produced by combining detected phenology with daily weather data. Water availability has been shown to be the main driver of wheat yield in dryland cropping systems of Western Australia [52]. It is defined as the sum of growing season (May to October) rainfall and one-third of summer (November to April) rainfall. We modify this using estimated phenological dates to define annual water availability (AWavail) as the sum of rainfall falling between SOS and EOS and one-third of rainfall falling during the 180-day period preceding SOS. We also consider total rainfall falling during the growing season (GSR), vegetative phase (VR) and grainfill phase (GR). In addition, growing degree days, defined as the sum of one half of daily maximum temperature minus minimum temperature, are calculated for the three periods: growing season (GSDD), vegetative phase (VDD) and grainfill phase (GDD).

2.2.4. Data Exploration

We explore the relationships between the phenological and seasonal climate metrics and yield using the Pearson correlation coefficient (R). We consider both the correlation between phenological and seasonal climate metrics and across all years and the mean of their annual correlations with yield. To visualise differences in relationships between the metrics and yield and how the relationships vary with seasonality, we plot the metrics against yield for each year.

2.2.5. Statistical and Machine Learning Models

We test six predictive models for yield estimation: multiple linear regression (MLR), linear mixed models (LMM), generalised additive models (GAM), random forests (RF), support vector regression using radial basis functions (SVR) and deep learning using multilayer perceptrons (DL). We perform all analyses in the R environment [53].

Multiple linear regression (MLR) is used as a baseline against which to compare other predictive models. Feature selection for MLR is performed using a stepwise forward selection process to find the optimal set of explanatory variables to use in each model. The forward selection process starts with the intercept only model and adds predictors to the

model one at a time. At each step, the single variable that improves the goodness-of-fit of the model as measured by the Akaike Information criterion (AIC) [54] is added until a minimum AIC is attained. The ordering in which predictors are added provides a measure of the importance of each predictor.

Linear mixed models (LMMs) are an extension of MLR that allows both fixed and random effects [55,56]. Fixed effects have a common linear relationship for all the data, as is the case for predictors in MLR. Random effects can be used to account for structure in the data. We estimate LMMs using the optimal set of predictors identified by stepwise forward selection for MLR plus one random effect that allows the relationship between a single important predictor and yield to vary by year. The predictor with maximum mean annual correlation with yield is used as the random effect. We use the 'lmer' function from R package 'lme4' [57] to perform linear mixed modelling.

Generalised additive models (GAMs) are generalised linear models with a linear predictor composed of smooth functions of the predictors [58,59]. Practical variable selection for GAMs can be performed by adding penalty terms that can shrink components of the linear term to zero, thereby eliminating predictors from the model [60]. We adopt this approach, using thin plate regression splines [61] for the smooths. We use the 'gam' function from R package 'mgcv' [62] for estimation of GAMs. To prevent overfitting and improve generalisation, we also impose an additional constraint to enforce smoother models by setting the 'gamma' parameter equal to six.

A regression tree is a hierarchic structure, where a test is applied at each level to either one predictor or a linear combination of predictors that may have one of two outcomes. The result is a partitioning of the predictor space into disjoint regions that each correspond to a single prediction. Random forests (RF) perform nonlinear regression by model averaging of many regression trees where each tree uses a random number of predictors sampled with replacement according to a uniform probability distribution [63]. We use the 'ranger' function with default parameters from R package 'ranger' for random forests [64].

Support vector regression (SVR) identifies the optimal regression hyperplanes by minimising the linear regression coefficients within specified error tolerances. Nonlinear regression is performed by first transforming the data into a high-dimensional feature space using a nonlinear kernel function [65]. In this study, we use Gaussian radial basis kernels to transform the data. We use the 'svm' function from R package 'kernlab' to estimate the kernel scale parameter and perform epsilon-sensitive loss regression [66].

Deep learning neural networks are artificial neural networks or multilayer perceptrons with multiple inner layers [67]. We train a deep learning network with three inner layers with 64, 128 and 64 nodes respectively. Rectified linear unit (ReLU) activation functions are used for the first two inner layers, and a linear activation function is used for the third. To prevent overfitting and improve generalisation, we perform regularisation by dropping out 20% of the nodes in each layer during training [68]. Estimation is performed using root mean square propagation (RMSProp) [69]. We use the Tensorflow interface [70] for training and prediction of deep learning models, accessed via R package 'keras' [71].

2.2.6. Yield Predictors

Using cells in all paddocks with wheat data, we test two sets of predictors for yield prediction: (1) Phenological metrics derived from the SWGC estimated growth curves; and (2) phenological metrics combined with seasonal climate metrics (Table 2). To avoid collinearity, for all models except linear mixed models, we do not use predictors that are sums of other predictors. These include iNDVI (equal to VLAD + GLAD), GSR (VR + GR) and GSDD (VDD + GDD). For linear mixed models, we include one random effect that is chosen to be the predictor that has maximum mean annual correlation with yield. That predictor may be one of the summed predictors.

Table 2. Sets of yield predictors tested: phenological (PM) and phenological plus seasonal climate metrics (PM + SCM).

Predictor Set	Metrics
PM	SOST, POST, EOST, POSV, VLAD, GLAD
PM + SCM	SOST, POST, EOST, POSV, VLAD, GLAD, AWavail, VR, GR, VDD, GDD

2.2.7. Model Comparison

Model comparison aims to determine the optimal machine learning model for producing long-term spatial hindcasts of yield to improve our understanding of crop response to environmental conditions and management. We use a 50–50 train-test split on paddock-years where 50% of paddock-year combinations are randomly selected and used to train each model. The remaining 50% of paddock-years are used as test data to assess model accuracy. Accuracy is measured using the coefficient of determination (R^2) between observed (y_i) and predicted (\hat{y}_i) yields, mean absolute error (MAE), root mean square error (RMSE) and normalised root mean square error (NRMSE), defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

$$NRMSE = \frac{RMSE}{\bar{y}} \quad (6)$$

The coefficient of determination provides the proportion of variance in the observed data explained by a model, relative to observed mean. The MAE and RMSE provide measures of the model error. NRMSE is the ratio of the model error and mean observed value. The optimal model is identified by having highest R^2 and lowest MAE, RMSE and NRMSE.

2.2.8. Optimal Model Validation

After the optimal model is identified we perform further validation of the results using the test data. To understand how the model performs in different years, we consider scatterplots of observed versus predicted yield for each year.

2.2.9. Yield Hindcasts

The machine learning model with highest R^2 and lowest MAE, RMSE and NRMSE is selected as the optimal yield prediction model. The model is then trained using all the yield data, and the estimated model parameters are used to predict yield in each year to form a set of long-term yield hindcasts.

3. Results

3.1. Spatially Weighted Growth Curve Estimation

The SWGC method for spatially weighted growth curve estimation was applied to each cell in the study area ($n = 189,280$) for 2003–2019. Figure 3 shows an example of the estimated growth curves for a cell in a field where wheat was grown in 10 years out of 17 years. It shows that there are different numbers of cloud-free Landsat images in different years. In 2003, there were only two cloud-free images during the May–October growing season. Years 2005 and 2009 have few cloud-free images during the vegetative growth phase, while 2006, 2011, 2014 and 2018 have few images during the grainfill phase. In some

years, when there were no cloud-free images available around the peak of season, such as 2011 and 2012, SWGC estimates higher peak NDVI values than observed.

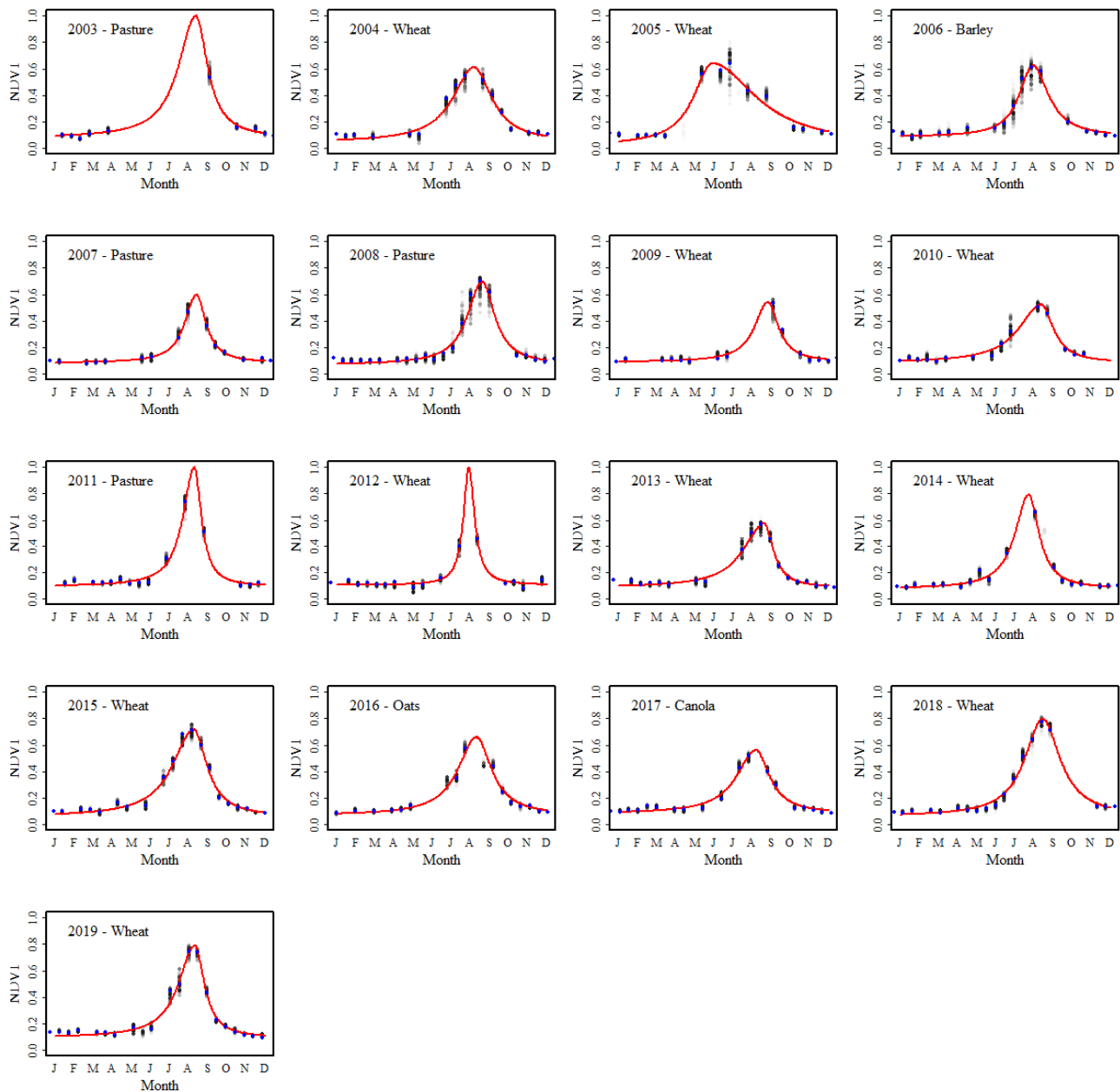


Figure 3. Example of spatially weighted growth curve estimation for a typical cell. The estimated growth curve is shown in red, NDVI data used in the estimation are shown in shades of grey according to their spatial weight (white lowest and black highest) and NDVI data for the cell are shown in blue.

3.2. Phenological and Seasonal Climate Metrics

We calculated the phenological and seasonal climate metrics for all cells where wheat was grown. Maps of all metrics are included in the Supplementary Material to this article (Supplementary Material 1).

3.3. Data Exploration

Table 3 shows the correlation between phenological and seasonal climate metrics and yield across all years and the mean of their annual correlations. Figure 4 shows scatter

plots of the phenological and seasonal climate metrics against yield for each year, with data coloured to show their correlation. Annual water availability (AWavail) has maximum correlation with yield across all years ($R = 0.64$), but its annual correlation with yield is lower and varies from year to year. This indicates that AWavail has potential usefulness for explaining temporal yield variability. Integrated NDVI over the entire growing season (iNDVI) has the next highest overall correlation with yield ($R = 0.56$) with strong positive annual correlations with yield that range from 0.10 in 2014 to 0.72 in 2015, with a mean annual correlation of 0.45. This indicates that iNDVI has usefulness for explaining both temporal and spatial variability. As seen in Figure 4, yield tends to increase with increasing iNDVI until it reaches a maximum, sometimes linearly (e.g., 2003, 2008), but it can appear that the rate of increase slows and the amount of yield variability increases with increasing values of iNDVI. The relationship between the peak of season NDVI (POSV) and yield is similar to the relationship between iNDV and yield but is less strong. Vegetative phase (early season)-integrated phenological metric (VLAD) has higher correlation with yield than grainfill phase (late season)-integrated metric (GLAD) across all years, but lower mean annual correlation with yield. Similarly, vegetative phase rainfall (VR) has higher correlation with yield than grainfill phase rainfall (GR) across all years, but VR has lower mean annual correlation with yield. However, grainfill phase degree days (GDD) has higher correlation with yield across all years and annually than vegetative phase degree days (VDD), so the relative importance of vegetative phase and grainfill phase metrics varies depending on the metric. Start of season timing (SOST) is negatively correlated with yield and end of season timing (EOST) is positively correlated with yield, indicating that early sowing and late harvest result in increased yield, but this is not the case in all years.

Table 3. Overall and annual relationships between phenological (PM) and seasonal climate metrics (SCM) and yield.

Metric	Metric Type	R	Mean Annual R
SOST	PM	−0.23	−0.10
POST	PM	−0.03	−0.12
EOST	PM	0.32	0.08
POSV	PM	0.34	0.36
iNDVI	PM	0.56	0.45
VLAD	PM	0.49	0.33
GLAD	PM	0.41	0.36
AWavail	SCM	0.64	0.12
GSR	SCM	0.50	0.12
VR	SCM	0.39	0.08
GR	SCM	0.31	0.10
GSDD	SCM	0.29	0.08
VDD	SCM	0.19	0.02
GDD	SCM	0.24	0.13

3.4. Model Comparison

Table 4 shows the accuracy statistics for estimating yield, calculated for the independent test data, for each of the machine learning models. The time taken to estimate each model is also shown. The nonlinear models (GAM, RF, SVR and DL) had higher accuracy than MLR. The LMM included a random effect that estimated different linear iNDVI-yield relationships for each year. This resulted in the highest accuracy of all models that used only phenological metrics as predictors ($R^2 = 0.67$, MAE = 0.25 t ha^{-1} , RMSE = 0.33 t ha^{-1} , NRMSE = 0.25).

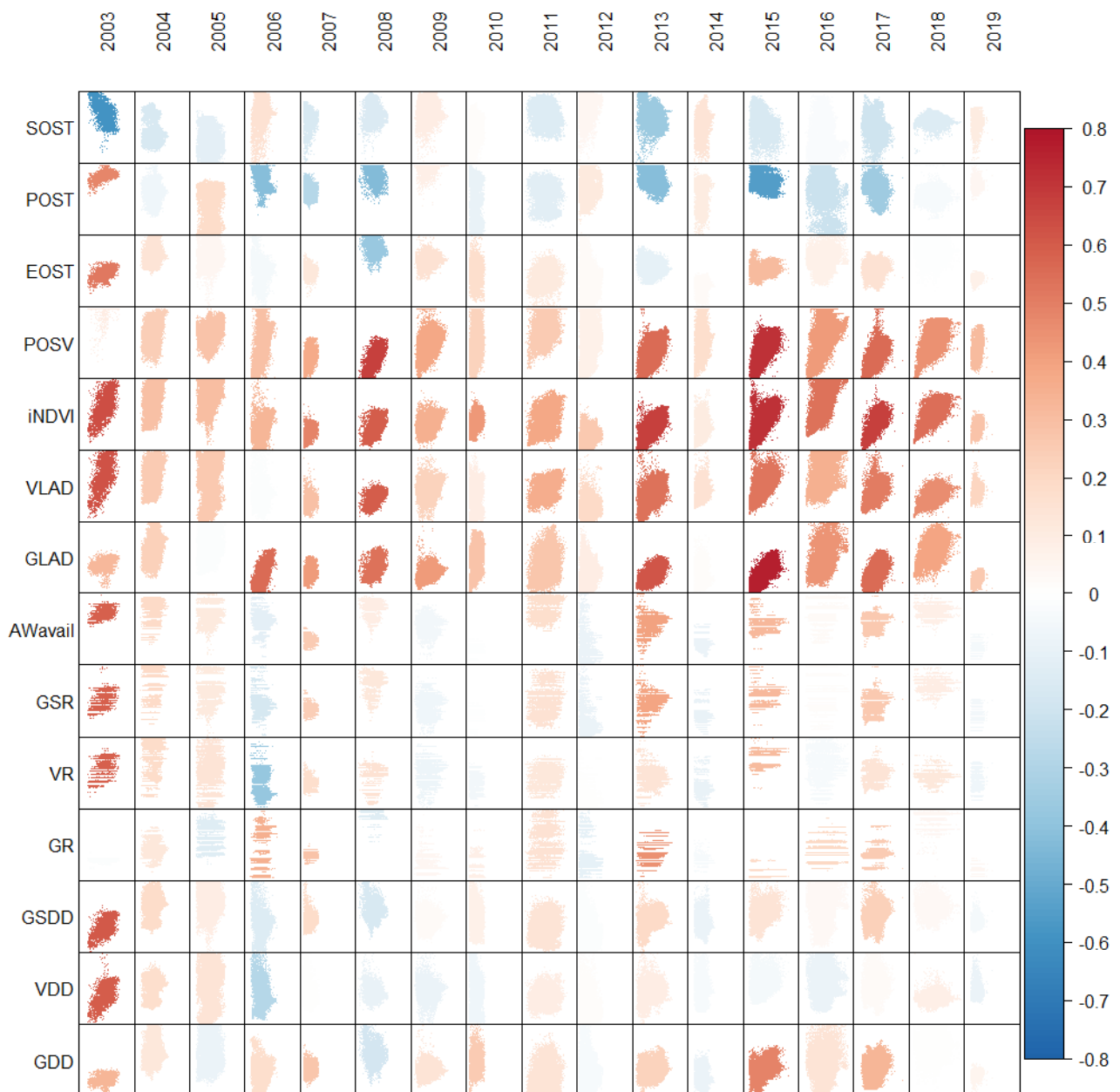


Figure 4. Scatterplots of yield (x-axes) and phenological and seasonal climate metrics (y-axes) for each year from 2003 to 2019. Data are coloured according to their correlation.

Use of seasonal climate metrics improved accuracy for all models except LMM. By estimating different iNDVI–yield relationships for each year, the LMM using only phenological metrics appeared to capture much of the interannual information provided by the seasonal climate metrics to the other models. Adding seasonal climate metrics to the model provided little improvement.

Two equally optimal models were identified: support vector regression with radial basis functions (SVR) and deep learning (DL) models ($R^2 = 0.68$, $MAE = 0.25 \text{ t ha}^{-1}$, $RMSE = 0.32 \text{ t ha}^{-1}$, $NRMSE = 0.25$) using combined phenological and seasonal climate metrics. There is little difference between the predictions of the two models across all years ($R = 0.98$). Figure 5 shows that the annual error (observed–predicted) distributions using the SVR and DL models are also similar. However, model estimation for SVM was more computationally intensive, taking almost 20 times as long to run than DL (Table 4). Because

the SVM and DL models perform similarly but DL is faster, we select DL as the optimal machine learning model.

Table 4. Machine learning model comparison for hindcasting wheat yields in paddocks (using a 50–50 train-test split on paddock-years). Models tested include multiple linear regression (MLR), generalised additive model (GAM), linear mixed model (LMM), random forest (RF), support vector regression with radial basis functions (SVR) and deep learning (DL). Predictors used include phenological metrics (PM) and seasonal climate metrics (SCM). Timing is the computational time required to fit the model to the training data ($n = 205,889$) on a Dell Latitude laptop with 2.8 GHz CPU and 32 GB RAM.

Model	Predictors	R^2	MAE (t ha^{-1})	RMSE (t ha^{-1})	NRMSE	Timing
MLR	PM	0.39	0.37	0.45	0.35	1 s
	PM + SCM	0.56	0.30	0.38	0.30	5 s
LMM	PM	0.67	0.25	0.33	0.26	10 s
	PM + SCM	0.68	0.25	0.33	0.25	17 s
GAM	PM	0.44	0.34	0.43	0.33	34 s
	PM + SCM	0.65	0.26	0.34	0.27	80 s
RF	PM	0.48	0.32	0.41	0.32	3 m 5 s
	PM + SCM	0.66	0.26	0.34	0.26	3 m 42 s
SVR	PM	0.49	0.31	0.41	0.32	3 h 58 m 30 s
	PM + SCM	0.68	0.25	0.32	0.25	4 h 18 m 6 s
DL	PM	0.51	0.31	0.40	0.31	11 m 20 s
	PM + SCM	0.68	0.25	0.32	0.25	13 m 13 s

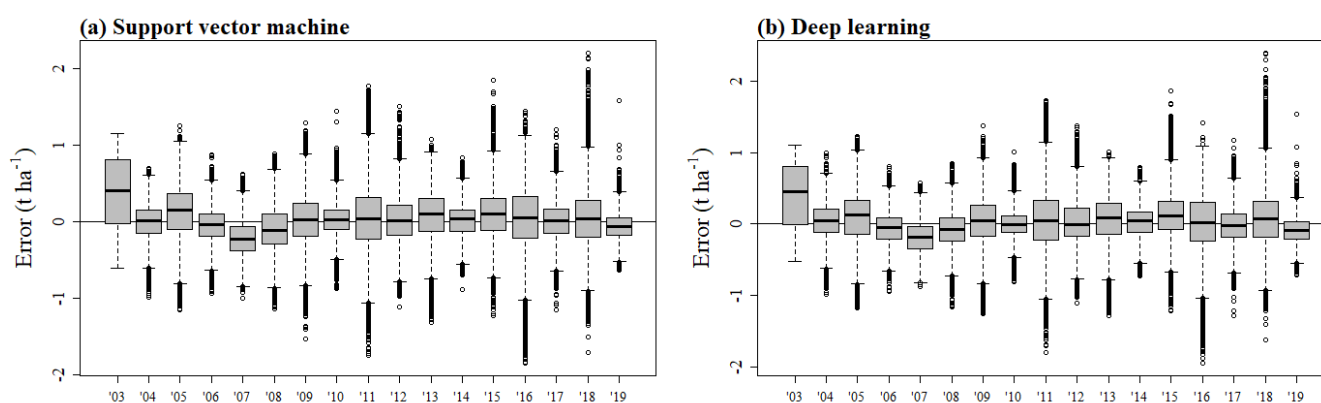


Figure 5. Boxplots showing error distributions (observed–predicted yield) for each year using the two machine learning models identified as equally optimal: (a) Support vector machines with radial basis functions and (b) deep learning.

3.5. Optimal Model Validation

Figure 6 shows the scatterplots of observed versus predicted yield using the deep learning model. The accuracy of the DL model varies across different years. It explained nearly 70% of the yield variability in 2015 (as it does overall, see Table 4), but much less in many years. R^2 was lower in low-yielding years (e.g., 2007, 2010, 2012, 2019) than in high-yielding years (e.g., 2015, 2017, 2018), but there are exceptions: 2005 and 2016 had high yields and low R^2 . It might be expected that the error statistics MAE and RMSE would be lower in lower-yielding years, and this was generally the case. In general, the predicted yields had a lower range than observed and the model over-predicts low yields and under-predicts high yields.

3.6. Yield Hindcasts

Yield hindcasts for all wheat fields from 2003 to 2019 were produced using the deep learning (DL) model trained on all available data. Figure 7 shows the maps of the observed and predicted yield for each year. High-resolution maps can be viewed in the Supplementary Material to this article (Supplementary Material 2). There is evidence of reduced range

of predicted versus observed yields, as suggested by Figure 6. However, the predicted yield maps show spatial consistency with the observed yield maps, with each showing similar spatial patterns of yield variation. There is a degree of spatial smoothing evident in the predicted maps, when viewed closely. This is due to the spatial smoothing component of spatially weighted growth curve estimation [48], which is propagated into the predicted yields.

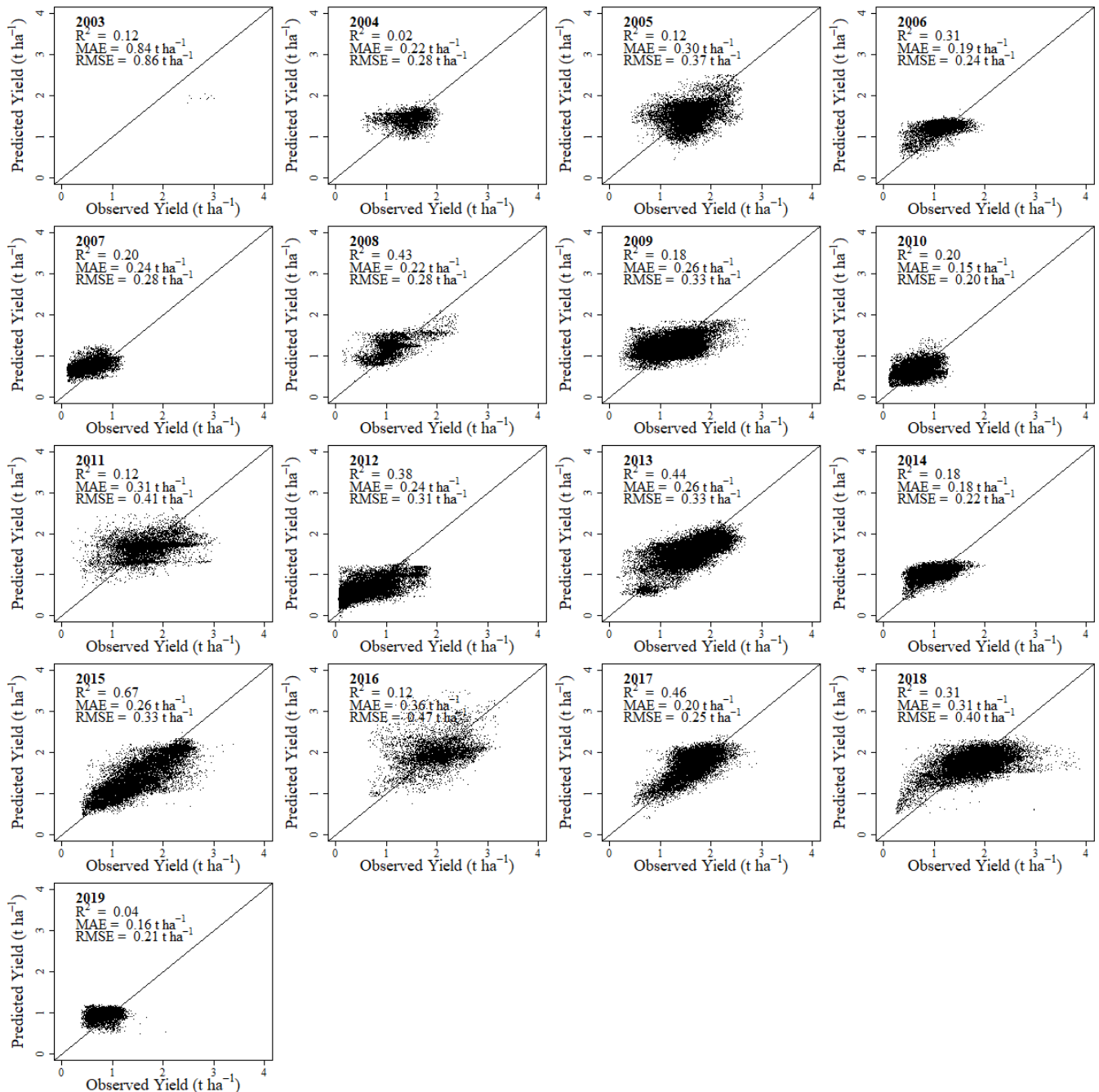


Figure 6. Scatterplots of observed versus predicted yield for 2004 to 2019 using the optimal deep learning (DL) model for the test data.

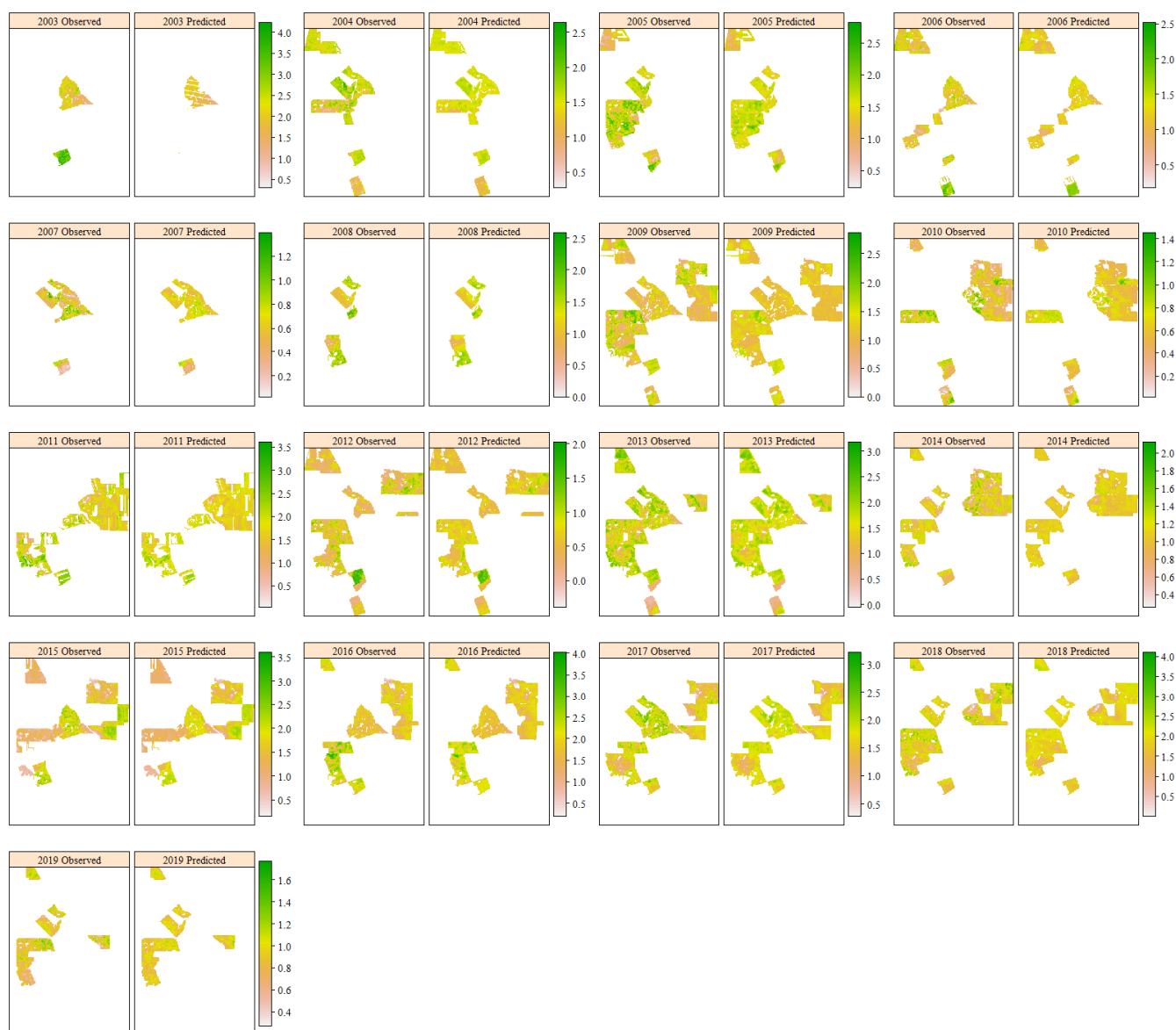


Figure 7. Hindcasts of wheat yield: Maps of observed and predicted wheat yield (t ha^{-1}) for 2004 to 2019.

4. Discussion

Motivated by the need for long-term sequences of yield data to support precision agriculture, this study investigated the use of Landsat NDVI for estimating wheat yields over a 17,000 ha study area in WA for 17 years from 2003 to 2019. For this purpose, we tested the use of a new method for estimating crop growth curves from sequences of Landsat NDVI that may contain spatial and temporal gaps caused by cloud contamination during the growing season: spatially weighted growth curve (SWGC) estimation. We used the estimated growth curves to identify the timing of phenological events and derive phenological metrics to describe the timing and degree of crop growth stages occurring at each cell. We then combined the detected phenology with climate data to produce seasonal climate metrics that summarise water availability and growing degree days during different growth stages of the crop. We investigated the relationships between the phenological and seasonal climate metrics and yield, and found that in general, the remotely sensed phenological metrics tend to correlate more highly with annual yields than the seasonal climate metrics. While annual water availability (AWavail) had maximum correlation with yield across all years, its correlation with yield in any one year was lower and varied

from year to year. In contrast, integrated NDVI over the growing season (iNDVI) had high correlation with yield across all years and annually, but the relationship between iNDVI and yield varied from year to year. We interpret this as suggesting that iNDVI has use for explaining both spatial and temporal variability and AWavail may be useful for explaining temporal yield variability or the interannual variability in the iNDVI–yield relationship. In general, the remotely sensed phenological metrics tend to correlate more highly with annual yields than the seasonal climate metrics. Across all years, SOST is negatively correlated with yield and EOST is positively correlated with yield, which reflects agronomic knowledge about the impact of early sowing and growing season length on yield [14–18]. However, this is not the case in all years, suggesting that other unmeasured factors affect these relationships.

We used the phenological metrics as predictors of yield in six statistical and machine learning models: multiple linear regression (MLR), linear mixed models (LMM), generalised additive models (GAM), random forests (RF), support vector regression using radial basis functions (SVR) and deep learning (DL). The nonlinear models (GAM, RF, SVR and DL) all had higher accuracy than MLR. The LMM included a random effect that estimated different iNDVI–yield relationships for each year. This resulted in the highest accuracy of all models using phenological metrics only ($R^2 = 0.67$, $MAE = 0.25 \text{ t ha}^{-1}$, $RMSE = 0.33 \text{ t ha}^{-1}$, $NRMSE = 0.25$). We then added the seasonal climate metrics to each of the models and quantified improvements in accuracy using the combined set of predictors. Use of seasonal climate metrics improved the accuracy for all models except LMM. By estimating different iNDVI–yield relationships for each year, the LMM using only phenological metrics appeared to capture much of the interannual information provided by the seasonal climate metrics to the other models. The nonlinear models all performed similarly. This contrasts with the range of accuracies reported across machine learning models by other studies that have compared machine learning models for yield estimation [24,31,32]. This may be because we have used a smaller study area than used in those national-scale studies, and the range of yields recorded for our study area is likely smaller than the range of yields across nations.

Two equally optimal models were identified: SVR and DL models ($R^2 = 0.68$, $MAE = 0.25 \text{ t ha}^{-1}$, $RMSE = 0.32 \text{ t ha}^{-1}$, $NRMSE = 0.25$) using combined phenological and seasonal climate metrics. They offer only marginally higher accuracy than the LMM using phenological metrics only. This means that long-term hindcasts of yield can be performed using only phenological metrics derived from Landsat NDVI. However, the nonlinear models that also use seasonal climate metrics have an additional advantage over the LMM. The LMM relies on having sufficient data from each year so that linear iNDVI–yield relationships can be estimated for each year. This means that the model cannot be used to estimate yield in years for which there is no training data. It cannot be used to extrapolate farther into the past, or to predict yield in future years without obtaining additional training data and re-fitting the model. Because they do not explicitly encode annual iNDVI–yield relationships but instead use the seasonal climate metrics to distinguish between different seasonal patterns, the nonlinear SVM and DL models can be used to make predictions for other years. This is advantageous for operational yield hindcasting and while we have focused on generating hindcasts in this study, it also has implications for yield forecasting. Yield forecasts are generally produced prior to the end of the growing season, when actual yield is unknown. Early forecasts, made before the crop is sown, can help farmers make decision about crop management, such as which crop and cultivar to plant, and how much fertiliser to apply at sowing. However, early forecasts cannot make use of in-season remotely sensed information and generally rely on the use of seasonal forecasts and crop models [72,73]. In contrast, remotely sensed phenology can contribute to forecasts made during the growing season that are useful for management decision about fertiliser, weed and disease management and also for industry-wide decision-making to support logistics, such as scheduling of grain transportation by road and rail, marketing and policy-making [10,74–76]. Use of phenological metrics,

climate data and machine learning has shown promise for forecasting wheat yield with 2-months lead time in Australia [31]. Our work supports this, and to enable the possibility of extending our work from considering hindcasts only to also considering forecasts, we recommend using climate metrics in nonlinear models, such SVM and DL found optimal by our study, rather than LMMs.

Of the two models identified as optimal, we prefer DL because it is considerably faster to implement. Validation of the DL model showed that it performed better in some years than others (Figure 5). Lower R^2 was observed for low-yielding years, suggesting that the phenological and seasonal climate metrics may not provide sufficient information to capture the complex relationship between crop growth and yield in dry years. For some years, poorer accuracy may be due to fewer cloud-free Landsat images being available during the growing season (Figure 2). For example, in 2003 there was only one cloud-free image between May and October. In 2005, 2009 and 2014 there were a few cloud-free images during the vegetative or early part of the growing season. In 2011 and 2012, there were a few cloud-free images available during the peak of season. Accuracy was highest in years with a reasonable number of cloud-free images regularly distributed throughout the growing season: 2008, 2013, 2015, 2017 and 2018. This suggests that yield could be mapped with higher accuracy if more cloud-free images were available. This could be achieved by combining data from multiple sensors [12,77,78], or by fusing Landsat data with MODIS data which has a more frequent revisit capacity [79–81].

We trained a DL model using all the available data and produced hindcast maps of yield at 30 m resolution for each year from 2003 to 2019. The maps showed spatial and temporal consistency with observed yield. However, the range of predicted yields in some years was lower than observed, indicating that the model tends to under-predict high yields and over-predict low yields. This is also demonstrated in plots of observed versus predicted yields for the test data. This is a common result of using statistical and machine learning models that fit observations in an average sense: higher errors are observed for extreme observations. Moreover, it is not uncommon for yield estimation [29,30]. There is no evidence of systematic error or bias evident in some other studies, such as under-prediction of high yields [32], or over-prediction of yields [24].

There is a degree of spatial smoothing in the yield hindcasts caused by the use of SWGC for growth curve estimation (see Evans and Shen [48] for discussion of the smoothing effect of SWGC). This affects the precision of the hindcasts, but not their value. A model has value if it helps the user make a better decision. Our goal of producing long-term hindcasts of yield at within-field resolution is to support precision agriculture decisions about how to vary crop management within fields. Currently, most decisions about in-field crop management are based on interpolated yield maps that have been smoothed [2], thereby reducing the yield range and the yield variability. Moreover, farmers may not have a sufficiently long record to quantify effects of seasonal variability and climate change. While our hindcasts may not be as spatially precise as yield monitor data, they do form a longer temporal record that can be used to understand impacts of seasonal climate conditions on yield.

Although we are primarily interested in estimating yield within fields, estimation of yield at field-scale is important to agricultural planning and policy-making. We therefore averaged estimated and predicted yields over each field-year combination so that field-average results could be considered. The accuracy statistics for field-average yield prediction using the deep learning model, calculated for the test data, was $R^2 = 0.85$, $MAE = 0.14 \text{ t ha}^{-1}$ and $RMSE = 0.19 \text{ t ha}^{-1}$. This suggests an added utility of yield estimation using this approach beyond our original goal of supporting on-farm decision-making: accurate estimates of field-average yields can be produced over large areas to support off-farm decision-making.

5. Conclusions

To support decision-making about how to optimise crop management within fields, this study aimed to determine whether phenological and seasonal climate metrics obtained from SWGC estimation have utility for estimating wheat yield.

We investigated the relationships between phenological and seasonal climate metrics and yield and found that integrated NDVI over the growing season (iNDVI) could explain the spatial and temporal variability in yield, but the relationship between iNDVI and yield varied from year to year. Annual water availability had the highest correlation with yield across all years, suggesting its potential for explaining temporal variability in yield or for explaining the interannual variability in the iNDVI–yield relationship.

We assessed and compared six statistical and machine learning models for estimating wheat yield using two sets of predictors: phenological metrics only and combined phenological and seasonal climate metrics. Using only phenological metrics, accuracy was highest using a linear mixed model with a random effect that allowed the relationship between iNDVI and yield to vary by year ($R^2 = 0.67$, MAE = 0.25 t ha⁻¹, RMSE = 0.33 t ha⁻¹, NRMSE = 0.25). For all other models, accuracy was higher when seasonal climate metrics were also used as predictors. We identified two equally optimal models using the combined phenological and seasonal climate metrics: support vector regression and deep learning models ($R^2 = 0.68$, MAE = 0.25 t ha⁻¹, RMSE = 0.32 t ha⁻¹, RMSE = 0.25). While the linear mixed model using only phenological metrics performed similarly to the nonlinear models that also used seasonal climate metrics, the nonlinear models can be more easily generalised to estimate yield in years for which training data are unavailable. We selected the deep learning model as optimal because it is faster to implement than the support vector regression. We performed further validation of the deep learning model by comparing observed and predicted yields for each year and found that it performed better in higher-yielding years and in years with a reasonable number of cloud-free images regularly distributed throughout the growing season. We used the model to produce yield hindcasts for all years from 2003 to 2019.

We conclude that long-term hindcasts of wheat yield in fields, at 30 m spatial resolution, can be produced by using SWGC estimation to detect remotely sensed phenology from Landsat NDVI and creating phenological and seasonal climate metrics to use as predictors of yield in machine learning models. However, better accuracy could be obtained if more regular time-sequences of NDVI were available. This might be achieved by combining remotely sensed data from multiple sources or by fusing Landsat data with more frequent MODIS observations.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/rs13132435/s1>, Supplementary Materials 1 (Maps of phenological and seasonal climate metrics) and Supplementary Materials 2 (High resolution maps of observed and hindcast yields).

Author Contributions: F.H.E.: conceptualisation, methodology, software, validation, formal analysis, data curation, writing (original draft, review and editing), funding acquisition; J.S.: data curation and pre-processing, writing (review and editing). All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the WA Government ‘Royalties for Regions’ program administered by the Department of Primary Industries and Regional Development.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Landsat data are available from <https://earthexplorer.usgs.gov> accessed on 2 February 2021. Climate data are available from <https://www.longpaddock.qld.gov.au/silo> accessed on 3 February 2021.

Acknowledgments: The authors thank Neil Smith, Ellanna Farm L-4-S Pastoral Co., for his generosity in providing data for this research.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Robert, P.C. Precision agriculture: A challenge for crop nutrition management. *Plant Soil* **2002**, *247*, 143–149. [[CrossRef](#)]
- Whelan, B.; Taylor, J. *Precision Agriculture for Grain Production Systems*; CSIRO Publishing: Melbourne, Australia, 2013. [[CrossRef](#)]
- Plant, R.E. Site-specific management: The application of information technology to crop production. *Comput. Electron. Agric.* **2001**, *30*, 9–29. [[CrossRef](#)]
- White, E.L.; Thomasson, J.A.; Auvermann, B.; Kitchen, N.R.; Pierson, L.S.; Porter, D.; Baillie, C.; Hamann, H.; Hoogenboom, G.; Janzen, T.; et al. Report from the conference, ‘identifying obstacles to applying big data in agriculture’. *Precis. Agric.* **2020**, *22*, 306–315. [[CrossRef](#)]
- Leonard, E.; Rainbow, R.; Baker, I.; Barry, S.; Darragh, L.; Darnell, R.; George, A.; Heath, R.; Jakku, E.; Laurie, A.; et al. *Accelerating Precision Agriculture to Decision Agriculture: Enabling Digital Agriculture in Australia*; Cotton Research and Development Corporation: Narrabri, Australia, 2017.
- Xue, J.; Su, B. Significant remote sensing vegetation indices: A review of developments and applications. *J. Sens.* **2017**, *2017*, 1353691. [[CrossRef](#)]
- Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring Vegetation Systems in the Great Plains with ERTS. In Proceedings of the Goddard Space Flight Center 3rd ERTS-1 Symposium, Washington, DC, USA, 10–14 December 1973; pp. 309–317.
- De Beurs, K.M.; Henebry, G.M. Spatio-Temporal Statistical Methods for Modelling Land Surface Phenology. In *Phenological Research*; Hudson, I.L., Keatley, M.R., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 177–208. [[CrossRef](#)]
- Zeng, L.; Wardlow, B.D.; Xiang, D.; Hu, S.; Li, D. A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. *Remote Sens. Environ.* **2020**, *237*, 111511. [[CrossRef](#)]
- Mkhabela, M.S.; Bullock, P.; Raj, S.; Wang, S.; Yang, Y. Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. *Agric. For. Meteorol.* **2011**, *151*, 385–393. [[CrossRef](#)]
- Salazar, L.; Kogan, F.; Roytman, L. Use of remote sensing data for estimation of winter wheat yield in the United States. *Int. J. Remote Sens.* **2007**, *28*, 3795–3811. [[CrossRef](#)]
- Bolton, D.K.; Friedl, M.A. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agric. For. Meteorol.* **2013**, *173*, 74–84. [[CrossRef](#)]
- Sakamoto, T.; Gitelson, A.A.; Arkebauer, T.J. MODIS-based corn grain yield estimation model incorporating crop phenology information. *Remote Sens. Environ.* **2013**, *131*, 215–231. [[CrossRef](#)]
- Gregersen, P.L.; Culetic, A.; Boschian, L.; Krupinska, K. Plant senescence and crop productivity. *Plant Mol. Biol.* **2013**, *82*, 603–622. [[CrossRef](#)]
- Zeleeke, K.T.; Nendel, C. Analysis of options for increasing wheat (*Triticum aestivum* L.) yield in south-eastern Australia: The role of irrigation, cultivar choice and time of sowing. *Agric. Water Manag.* **2016**, *166*, 139–148. [[CrossRef](#)]
- Flohr, B.M.; Hunt, J.R.; Kirkegaard, J.A.; Evans, J.R.; Trevaskis, B.; Zwart, A.; Swan, A.; Fletcher, A.L.; Rheinheimer, B. Fast winter wheat phenology can stabilise flowering date and maximise grain yield in semi-arid Mediterranean and temperate environments. *Field Crop. Res.* **2018**, *223*, 12–25. [[CrossRef](#)]
- Hunt, J.R.; Lilley, J.M.; Trevaskis, B.; Flohr, B.M.; Peake, A.; Fletcher, A.; Zwart, A.B.; Gobbett, D.; Kirkegaard, J.A. Early sowing systems can boost Australian wheat yields despite recent climate change. *Nat. Clim. Chang.* **2019**, *9*, 244–247. [[CrossRef](#)]
- Fischer, R.A.; Kohn, G.D. The relationship of grain yield to vegetative growth and post-flowering leaf area in the wheat crop under conditions of limited soil moisture. *Aust. J. Agric. Res.* **1966**, *17*, 281–295. [[CrossRef](#)]
- Rezaei, E.E.; Ghazaryan, G.; Gonzalez, J.; Cornish, N.; Dubovyk, O.; Siebert, S. The use of remote sensing to derive maize sowing dates for large-scale crop yield simulations. *Int. J. Biometeorol.* **2021**, *65*, 565–576. [[CrossRef](#)] [[PubMed](#)]
- Ji, Z.; Pan, Y.; Zhu, X.; Wang, J.; Li, Q. Prediction of Crop Yield Using Phenological Information Extracted from Remote Sensing Vegetation Index. *Sensors* **2021**, *21*, 1406. [[CrossRef](#)] [[PubMed](#)]
- Becker-Reshef, I.; Vermote, E.; Lindeman, M.; Justice, C. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. *Remote Sens. Environ.* **2010**, *114*, 1312–1323. [[CrossRef](#)]
- Lai, Y.R.; Pringle, M.J.; Kopittke, P.M.; Menzies, N.W.; Orton, T.G.; Dang, Y.P. An empirical model for prediction of wheat yield, using time-integrated Landsat NDVI. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *72*, 99–108. [[CrossRef](#)]
- Labus, M.P.; Nielsen, G.A.; Lawrence, R.L.; Engel, R.; Long, D.S. Wheat yield estimates using multi-temporal NDVI satellite imagery. *Int. J. Remote Sens.* **2002**, *23*, 4169–4180. [[CrossRef](#)]
- Kamir, E.; Waldner, F.; Hochman, Z. Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. *ISPRS J. Photogramm. Remote Sens.* **2020**, *160*, 124–135. [[CrossRef](#)]
- Benedetti, R.; Rossini, P. On the use of NDVI profiles as a tool for agricultural statistics: The case study of wheat yield estimate and forecast in Emilia Romagna. *Remote Sens. Environ.* **1993**, *45*, 311–326. [[CrossRef](#)]
- Araya, S.; Ostendorf, B.; Lyle, G.; Lewis, M. Remote sensing derived phenological metrics to assess the spatio-temporal growth variability in cropping fields. *Adv. Remote Sens.* **2017**, *6*, 212–228. [[CrossRef](#)]

27. Kouadio, L.; Duveiller, G.; Djaby, B.; El Jarroudi, M.; Defourny, P.; Tychon, B. Estimating regional wheat yield from the shape of decreasing curves of green area index temporal profiles retrieved from MODIS data. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *18*, 111–118. [[CrossRef](#)]
28. Waldner, F.; Horan, H.; Chen, Y.; Hochman, Z. High temporal resolution of leaf area data improves empirical estimation of grain yield. *Sci. Rep.* **2019**, *9*, 15714. [[CrossRef](#)] [[PubMed](#)]
29. Chen, Y.; Donohue, R.J.; McVicar, T.R.; Waldner, F.; Mata, G.; Ota, N.; Houshmandfar, A.; Dayal, K.; Lawes, R.A. Nationwide crop yield estimation based on photosynthesis and meteorological stress indices. *Agric. For. Meteorol.* **2020**, *284*, 107872. [[CrossRef](#)]
30. Donohue, R.J.; Lawes, R.A.; Mata, G.; Gobbett, D.; Ouzman, J. Towards a national, remote-sensing-based model for predicting field-scale crop yield. *Field Crop. Res.* **2018**, *227*, 79–90. [[CrossRef](#)]
31. Cai, Y.; Guan, K.; Lobell, D.; Potgieter, A.B.; Wang, S.; Peng, J.; Xu, T.; Asseng, S.; Zhang, Y.; You, L.; et al. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agric. For. Meteorol.* **2019**, *274*, 144–159. [[CrossRef](#)]
32. Wang, Y.; Zhang, Z.; Feng, L.; Du, Q.; Runge, T. Combining Multi-Source Data and Machine Learning Approaches to Predict Winter Wheat Yield in the Conterminous United States. *Remote Sens.* **2020**, *12*, 1232. [[CrossRef](#)]
33. Belward, A.S.; Skøien, J.O. Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. *ISPRS J. Photogramm. Remote Sens.* **2015**, *103*, 115–128. [[CrossRef](#)]
34. Wulder, M.A.; Masek, J.G.; Cohen, W.B.; Loveland, T.R.; Woodcock, C.E. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sens. Environ.* **2012**, *122*, 2–10. [[CrossRef](#)]
35. Caccetta, P.; Furby, S.; Richards, G.; Wallace, J.; Waterworth, R.; Wu, X. Long-term monitoring of Australian land cover change using Landsat data: Development, implementation, and operation. In *Global Forest Monitoring from Earth Observation*; Achard, F., Hansen, M.C., Eds.; CRC Press: Boca Raton, FL, USA, 2002; pp. 243–258. [[CrossRef](#)]
36. Furby, S.L.; Caccetta, P.A.; Wallace, J.F.; Lehmann, E.A.; Zdunic, K. Recent development in vegetation monitoring products from Australia’s National Carbon Accounting System. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Cape Town, South Africa, 12–17 July 2009.
37. Masek, J.G.; Vermote, E.F.; Saleous, N.E.; Wolfe, R.; Hall, F.G.; Huemmrich, K.F.; Gao, F.; Kutler, J.; Lim, T.K. A Landsat Surface Reflectance Dataset for North America, 1990–2000. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 68–72. [[CrossRef](#)]
38. Jin, S.; Yang, L.; Danielson, P.; Homer, C.; Fry, J.; Xian, G. A comprehensive change detection method for updating the National Land Cover Database to circa 2011. *Remote Sens. Environ.* **2013**, *132*, 159–175. [[CrossRef](#)]
39. Achard, F.; Stibig, H.-J.; Eva, H.D.; Lindquist, E.J.; Bouvet, A.; Arino, O.; Mayaux, P. Estimating tropical deforestation from Earth observation data. *Carbon Manag.* **2014**, *1*, 271–287. [[CrossRef](#)]
40. Boryan, C.; Yang, Z.; Mueller, R.; Craig, M. Monitoring US agriculture: The US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto Int.* **2011**, *26*, 341–358. [[CrossRef](#)]
41. Johnson, D.M. Using the Landsat archive to map crop cover history across the United States. *Remote Sens. Environ.* **2019**, *232*, 111286. [[CrossRef](#)]
42. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [[CrossRef](#)]
43. Sakamoto, T.; Wardlow, B.D.; Gitelson, A.A.; Verma, S.B.; Suyker, A.E.; Arkebauer, T.J. A Two-Step Filtering approach for detecting maize and soybean phenology with time-series MODIS data. *Remote Sens. Environ.* **2010**, *114*, 2146–2159. [[CrossRef](#)]
44. Roy, D.P.; Yan, L. Robust Landsat-based crop time series modelling. *Remote Sens. Environ.* **2020**, *238*, 110810. [[CrossRef](#)]
45. Whitcraft, A.K.; Vermote, E.F.; Becker-Reshef, I.; Justice, C.O. Cloud cover throughout the agricultural growing season: Impacts on passive optical earth observations. *Remote Sens. Environ.* **2015**, *156*, 438–447. [[CrossRef](#)]
46. Younes, N.; Joyce, K.E.; Maier, S.W. All models of satellite-derived phenology are wrong, but some are useful: A case study from northern Australia. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *97*, 102285. [[CrossRef](#)]
47. Weiss, M.; Jacob, F.; Duveiller, G. Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* **2020**, *236*, 111402. [[CrossRef](#)]
48. Evans, F.H.; Shen, J. Spatially weighted estimation of broadacre crop growth improves gap-filling of Landsat NDVI. *Remote Sens.* **2021**, *13*, 2128. [[CrossRef](#)]
49. Shen, J.; Evans, F.H. The Potential of Landsat NDVI Sequences to Explain Wheat Yield Variation in Fields in Western Australia. *Remote Sens.* **2021**, *13*, 2202. [[CrossRef](#)]
50. Gräler, B.; Pebesma, E.; Heuvelink, G. Spatio-temporal interpolation using gstat. *R J.* **2016**, *8*, 204–218. [[CrossRef](#)]
51. Pebesma, E.J. Multivariable geostatistics in S: The gstat package. *Comput. Geosci.* **2004**, *30*, 683–691. [[CrossRef](#)]
52. Chen, K.; O’Leary, R.A.; Evans, F.H. A simple and parsimonious generalised additive model for predicting wheat yield in a decision support tool. *Agric. Syst.* **2019**, *173*, 140–150. [[CrossRef](#)]
53. R Core Team. R: A Language and Environment for Statistical Computing. Available online: <https://www.R-project.org/> (accessed on 5 July 2019).
54. Akaike, H. Information theory and an extension of the maximum likelihood principle. In Proceedings of the Second International Symposium on Information Theory, Tsahkadsor, Armenia, 2–8 September 1971; pp. 267–281.
55. Pinheiro, J.; Bates, B. *Mixed-Effects Models in S and S-PLUS*; Springer Science Business Media: New York, NY, USA, 2004.

56. Zuur, A.F.; Ieno, E.N.; Walker, N.J.; Saveliev, A.A.; Smith, G.M. *Mixed Effects Models and Extensions in Ecology with R*; Springer: New York, NY, USA, 2009.
57. Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **2015**, *67*. [[CrossRef](#)]
58. Hastie, T.; Tibshirani, R. *Generalized Additive Models*; Chapman and Hall: London, UK, 1990. [[CrossRef](#)]
59. Wood, S.N. *Generalized Additive Models: An Introduction with R*; Chapman and Hall: London, UK, 2006. [[CrossRef](#)]
60. Marra, G.; Wood, S.N. Practical variable selection for generalized additive models. *Comput. Stat. Data Anal.* **2011**, *55*, 2372–2387. [[CrossRef](#)]
61. Wood, S.N. Thin plate regression splines. *J. R. Stat. Soc. Ser. B* **2003**, *65*, 95–114. [[CrossRef](#)]
62. Wood, S.N. Stable and efficient multiple smoothing parameter estimation for generalized additive models. *J. Am. Stat. Assoc.* **2004**, *99*, 673–686. [[CrossRef](#)]
63. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
64. Wright, M.N.; Ziegler, A. Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *J. Stat. Softw.* **2017**, *77*. [[CrossRef](#)]
65. Vapnik, V.; Golowich, S.E.; Smola, A.J. Support vector method for function approximation, regression estimation, and signal processing. *Adv. Neural Inf. Process. Syst.* **1997**, *9*, 281–287.
66. Karatzoglou, A.; Smola, A.; Hornik, K.; Zeileis, A. Kernlab—An S4 Package for Kernel Methods in R. *J. Stat. Softw.* **2004**, *11*. [[CrossRef](#)]
67. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
68. Srivastava, N.; Hinton, G.; Krizhevsky, I.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
69. Tieleman, T.; Hinton, G. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA Neural Netw. Mach. Learn.* **2012**, *4*, 26–31.
70. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*; Software available from tensorflow.org; TensorFlow: Mountain View, CA, USA, 2015.
71. Allaire, J.J.; Chollet, F. Keras: R Interface to ‘Keras’. Available online: <https://keras.rstudio.com> (accessed on 5 February 2021).
72. Evans, F.H.; Guthrie, M.M.; Foster, I. Accuracy of six years of operational statistical seasonal forecasts of rainfall in Western Australia (2013 to 2018). *Atmos. Res.* **2020**, *233*, 104697. [[CrossRef](#)]
73. Brown, J.N.; Hochman, Z.; Holzworth, D.; Horan, H. Seasonal climate forecasts provide more definitive and accurate crop yield predictions. *Agric. For. Meteorol.* **2018**, *260*, 247–254. [[CrossRef](#)]
74. Kouadio, L.; Newlands, N.; Davidson, A.; Zhang, Y.; Chipanshi, A. Assessing the Performance of MODIS NDVI and EVI for Seasonal Crop Yield Forecasting at the Ecodistrict Scale. *Remote Sens.* **2014**, *6*, 10193–10214. [[CrossRef](#)]
75. Franch, B.; Vermote, E.F.; Becker-Reshef, I.; Claverie, M.; Huang, J.; Zhang, J.; Justice, C.; Sobrino, J.A. Improving the timeliness of winter wheat production forecast in the United States of America, Ukraine and China using MODIS data and NCAR Growing Degree Day information. *Remote Sens. Environ.* **2015**, *161*, 131–148. [[CrossRef](#)]
76. Huang, J.; Gómez-Dans, J.L.; Huang, H.; Ma, H.; Wu, Q.; Lewis, P.E.; Liang, S.; Chen, Z.; Xue, J.-H.; Wu, Y.; et al. Assimilation of remote sensing into crop growth models: Current status and perspectives. *Agric. For. Meteorol.* **2019**, 276–277, 107609. [[CrossRef](#)]
77. Skakun, S.; Vermote, E.; Franch, B.; Roger, J.-C.; Kussul, N.; Ju, J.; Masek, J. Winter Wheat Yield Assessment from Landsat 8 and Sentinel-2 Data: Incorporating Surface Reflectance, through Phenological Fitting, into Regression Yield Models. *Remote Sens.* **2019**, *11*, 1768. [[CrossRef](#)]
78. Bolton, D.K.; Gray, J.M.; Melaas, E.K.; Moon, M.; Eklundh, L.; Friedl, M.A. Continental-scale land surface phenology from harmonized Landsat 8 and Sentinel-2 imagery. *Remote Sens. Environ.* **2020**, *240*, 111685. [[CrossRef](#)]
79. Gao, F.; Anderson, M.C.; Zhang, X.; Yang, Z.; Alfieri, J.G.; Kustas, W.P.; Mueller, R.; Johnson, D.M.; Prueger, J.H. Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. *Remote Sens. Environ.* **2017**, *188*, 9–25. [[CrossRef](#)]
80. Walker, J.J.; de Beurs, K.M.; Wynne, R.H. Dryland vegetation phenology across an elevation gradient in Arizona, USA, investigated with fused MODIS and Landsat data. *Remote Sens. Environ.* **2014**, *144*, 85–97. [[CrossRef](#)]
81. Zhou, F.; Zhong, D. Kalman filter method for generating time-series synthetic Landsat images and their uncertainty from Landsat and MODIS observations. *Remote Sens. Environ.* **2020**, *239*, 111628. [[CrossRef](#)]