

Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods

Elisa Kamir^{a,b}, François Waldner^{a,*}, Zvi Hochman^a

^a CSIRO Agriculture & Food, 306 Carmody Road, St Lucia, Queensland 4067, Australia

^b Supagro Montpellier, 2 Place Pierre Viala, 34060 Montpellier, France

ARTICLE INFO

Keywords:

Agriculture
Machine learning
Actual yield
Yield Gap
Remote sensing
National scale
Regression

ABSTRACT

Closing the yield gap between actual and potential wheat yields in Australia is important to meet the growing global demand for food. The identification of hotspots of the yield gap, where the potential for improvement is the greatest, is a necessary step towards this goal. While crop growth models are well suited to quantify potential yields, they lack the ability to provide accurate large-scale estimates of actual yields, owing to the sheer quantity of data they require for parameterisation. In this context, we sought to provide accurate estimates of actual wheat yields across the Australian wheat belt based on machine-learning regression methods, climate records and satellite image time series. Out of nine base learners and two ensembles, support vector regression with radial basis function emerged as the single best learner (root mean square error of 0.55 t ha^{-1} and R^2 of 0.77 at the pixel level). At national scale, this model explained 73% of the yield variability observed across statistical units. Benchmark approaches based on peak Normalised Difference Vegetation Index (NDVI) and on a harvest index were largely outperformed by the machine-learning regression models ($R^2 < 0.46$). Climate variables such as maximum temperatures and accumulated rainfall provided additional information to the 16-day NDVI time series as they significantly improved yield predictions. Variables observed up to and around the flowering period had a particularly high predictive power with additional information gained from data during grain filling. We further showed that, while all models were sensitive to a reduction of the training set size, a large majority had not reached saturation with a data set of 125 fields (2000 pixels). This indicates that additional training data are likely to further improve the skill of the models. We estimated that observations from 75 fields (1200 pixels) are required for the best single model to reach an R^2 of 0.7. We contend that machine-learning regression methods applied to climate and satellite image time series can achieve reliable crop yield monitoring across years at both the pixel and the country scale. The resulting yield estimates meet the accuracy requirements for mapping the yield gap and identifying yield gap hotspots which could be targeted for further work by agricultural researchers and advisers.

1. Introduction

The global food supply system is under mounting pressure due to sustained population growth, shifts in diets, and changes in climate and economic systems (Rounsevell et al., 2005; Searchinger et al., 2008; Godfray et al., 2010). Ensuring that food is available to all will require an increase in food production. Intensification of farming practices and cropland expansion are the two major pathways to increasing food production. Land conversion has been shown to elevate greenhouse gas emissions and to impact ecosystem services. It is responsible for 12% of the total global emissions and contributes to biodiversity erosion by fragmenting the habitat of endangered species (Nelson et al., 2010). Indeed, sharing land with wild species rather than providing them with

unfarmed and undisturbed habitats through land sparing policies is more likely to impact them (Phalan, 2018). Therefore, intensification of existing croplands, if conducted in a sustainable way that prevents ecosystem deterioration and fertility loss, appears as the most appropriate option to meet future food demand.

Considerable yield improvement opportunities exist where inefficient management of agricultural land cause deviations of actual from potential yields (Hochman and Horan, 2018). It is estimated that addressing suboptimal management practices could increase the current global crop production by 45% to 70% (Mueller et al., 2012). Wheat growers in Australia are currently achieving about 55% of their water-limited yield or potential yield for rain-fed crops (Hochman et al., 2017). As the current grain production in Australia is sufficient to feed

* Corresponding author.

E-mail address: franz.waldner@csiro.au (F. Waldner).

<https://doi.org/10.1016/j.isprsjprs.2019.11.008>

Received 23 February 2019; Received in revised form 6 November 2019; Accepted 7 November 2019

Available online 20 December 2019

0924-2716/ © 2019 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

Table 1
Characteristics of data sets used for the study.

Data type	Sample size (fields/pixels)	Spatial resolution	Temporal resolution - Years of data	Source
NDVI time series	–	250 m	16-day composites from 2009 to 2015	MODIS, collection 5
Climate time series	–	5 km	Daily records from 2009 to 2015	SILO database
Yield maps	166/3011	250 m	Yearly records from 2009 to 2015	This study
Crop type data	339/5562	Fields (vectors)	Yearly labels from 2009 to 2015	Yield Prophet* and this study
Cropland map	–	30 m	2013–2015	Teluguntla et al. (2018)
Crop area and crop yield statistics	88*	Statistical units (SA2; vectors)	2015	Australian Bureau of Statistics

* After discarding SA2s with < 21,000 ha sown to wheat (see Section 3.5.2 for details).

the projected population in 2050 and beyond, it is, therefore, more likely that global rather than national food security concerns and additional economic forces will continue to drive agricultural intensification. In fact, while Australia produces just three per cent of the world's wheat it accounts for 10–15% of the world's 100 million tonnes annual global wheat trade. Identifying yield gap hotspots is thus a prerequisite to target areas that need “intervention” to help close the gap. This requires detailed maps of water-limited and actual yields over large areas and across scales.

Water-limited yields are a function of biophysical characteristics of the environment, the genetic material, and the best management practices (van Ittersum et al., 2013). Hence, they are commonly estimated with locally-validated deterministic crop growth models such as the Agricultural Production System sIMulator (APSIM; Holzworth et al., 2014) using the optimal management decisions (see Hochman et al., 2016 for an Australian example). Crop growth models are also appropriate to estimate actual yield at field scale (Hochman et al., 2009), but far less at larger scales, because they require intensive knowledge about the management practices and the environment (e.g., soil type, starting soil moisture), leading to complex point-based parametrisation (Bazgeer et al., 2007). With limited information (reflecting the typical use of models for large-scale applications), most crop growth models poorly estimate actual yield (Palosuo et al., 2011).

As early as 1975, there has been a strong interest in estimating actual yields from space using optical remote sensing (Hammond, 1975; Morain and Williams, 1975). Efforts can be classified into three main categories: (1) empirical models that relate remotely-sensed quantities such as vegetation indices to instantaneous values of various canopy state variables which themselves correlate with yield, (2) semi-empirical models based on light use efficiency theory, and (3) data assimilation into crop growth models. Given their simplicity, empirical models have been widely utilised in the remote sensing literature. However, the main criticism is that these models are specific to the crop cultivars, the crop growth stages, and the geographical regions they were calibrated on (Doraiswamy et al., 2004; Fang et al., 2011) and, therefore, do not generalise well. Advanced nonparametric alternative methods in the field of machine learning have been recently tested to estimate crop variables such as leaf area index (Rivera-Caicedo et al., 2017; Houborg and McCabe 2018), leaf chlorophyll content (Verrelst et al., 2013), and yield (Jiang et al., 2004; Johnson et al., 2016; Kuwata and Shibasaki 2016; Mishra et al., 2016; Wolanin et al., 2019; Khaki and Wang 2019) for a range of crop types and environments. These methods learn to predict a target value, yield in this case, without prior assumptions about the underlying data distribution (Behmann et al., 2015). They are also particularly suitable for the interpretation of data from optical sensors because the impact of noise may be compensated for, provided availability of sufficient representative training data. Finally, machine learning methods also bolster the capacity to efficiently process big Earth Observation data (Verrelst et al., 2015).

The objective of this study was to design a machine learning approach to accurately map actual end-of-season wheat yields across Australia using climate and satellite image time series. More specifically, we sought (1) to compare and identify the most suitable model to estimate wheat yield, determine its accuracy at the pixel and statistical-

unit levels and compare their performance against two benchmark models (an empirical model based on peak NDVI and a light-use efficiency model based on a harvest index), (2) to evaluate if ensembles of base learners can further reduce prediction errors, (3) to identify important features in the yield estimate process and evaluate the relative contribution of climate data, and finally (4) to determine the sensitivity of the considered algorithms to the size of the training set.

2. Materials

2.1. Study area

Approximately 6% of the agricultural land in Australia is used for growing crops (26.1 M ha), of which over 99% is rain-fed. Wheat is the main crop in Australia, as it represents 43% of the cropland area and yields an average (1997–2016) of 1.79 t ha⁻¹ (www.abs.gov.au). Wheat production is dedicated to both human and livestock consumption. Most varieties are spring wheat varieties planted in late autumn to early winter (April–mid June) and generally harvested early in the summer (November–December). Rainfall in the cropping areas ranges from 325 mm to 700 mm and is winter-dominant rainfall in the west and south to summer-dominant rainfall in the north-east, where stored soil moisture is a critical supplement to in-crop rainfall.

2.2. Data

We collected data from different sources, providing a set of yield, satellite image time series and climate data at 250 m resolution for six years between 2009 and 2015 (Table 1).

First, we sourced fortnightly composites of Normalised Vegetation Difference Index (NDVI) at 250 m resolution from the MODIS “MOD13Q1” data set, collection 5. All tiles covering the region of interest were downloaded and mosaicked. Pixels contaminated by clouds or affected by sensor malfunctions were masked based on the image quality flags, missing values were then gap-filled using linear temporal interpolation.

Second, we sourced daily rainfall and maximum temperature estimates from SILO (Scientific Information for Land Owners), a database of historical climate records for Australia. The spatially and temporally complete SILO data sets are constructed from point-based observational records provided by the Bureau of Meteorology which are then gap-filled and interpolated to produce a raster with a 5-km resolution (Jeffrey et al., 2001). It was hypothesised that higher resolution was not required for our study due to low local variability of the features. We resampled the gridded estimates to 250 m using a nearest neighbour algorithm and produced 16-day composites so that the climate data matched the resolution of MODIS in space and time. We converted the 16-day rainfall estimates to cumulative rainfall starting from the first of January each year.

Third, maps of observed grain yield were collected using growers' yield monitoring equipment attached to harvesters. Observed yields were pre-processed according to Bramley and Williams (2001) and Whelan et al., (2002); and were then aggregated to the resolution of MODIS using an average resampling algorithm. In total, 166 yield maps

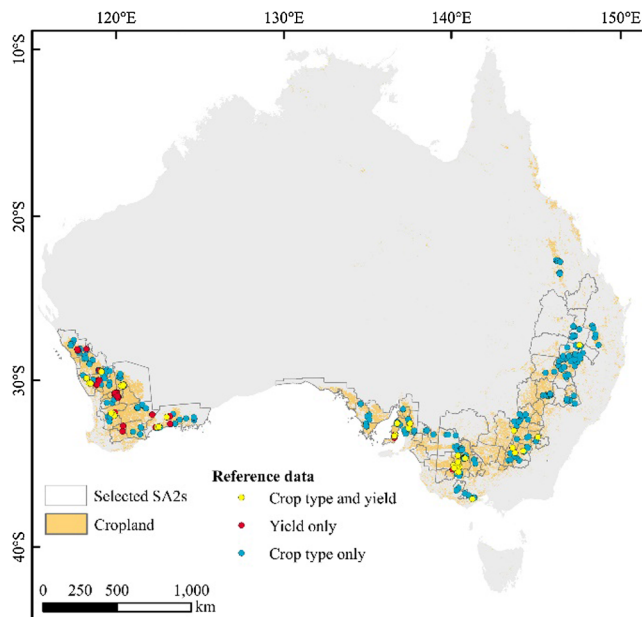


Fig. 1. Location of the reference data available to estimate wheat yields and to map crop types. Red and yellow dots were used to build yield prediction models whereas yellow and blue dots served as training data to classify different crop types. Areas in orange illustrate the extent of the Australian cropland. The black lines outline the statistical units (SA2s) for which aggregate yield estimates were available from the Australian Bureau of Statistics for 2015.

covering 3011 pixels were available (Fig. 1).

Fourth, we generated the three input data sets needed to implement a semi-empirical model based on a harvest index as benchmark. The fraction of absorbed photosynthetically active radiation was derived from the MODIS NDVI data by linearly rescaling NDVI values using pixel-specific thresholds of bare soil and of full vegetation cover (Donohue et al., 2014). The average daily incoming short-wave surface irradiance was calculated at 250 m according to Donohue et al. (2010). The diffuse fraction, which is the ratio of diffuse to total solar irradiance and varies from 0.2 under clear skies to 1.0 under overcast skies, was calculated following Roderick (1999). All data sets were resampled to 250 m.

Fifth, 295 additional fields with known crop type and geolocation in 2015 were collected from the Yield Prophet® database (Hochman et al., 2009) to increase the number of training data available for crop mapping. Yield Prophet® is an online agricultural decision-support tool that helps growers match crop inputs to potential yield to maximise profit and minimise climate and soil water risk. Crop type information from

Yield Prophet® is high quality because it has been collected from growers who pay a subscription fee to use it. The full data set totalled 339 fields (5562 pixels) with an average of 16 pixels per field.

Sixth, we sourced a 30-m cropland extent product for Australia from (Teluguntla et al., 2018) which we resampled to 250 m and used to mask out non cropping areas. This map was produced using Landsat-8 images acquired in the years 2013–2015 and a pixel-based random forest algorithm (Breiman 2001). Training (958 pixels) and validation (900 pixels) data were compiled from field surveys, photo-interpretation of very high spatial resolution imagery, and other ancillary sources. The map has an overall accuracy of 97% but the users' accuracy of the cropland class only reached 79%. Some of these errors might be reduced due to resampling from 30 m to 250 m.

Finally, we collected wheat area and yield statistics from the Australian Bureau of Statistics for each SA2 in the country for 2015. SA2s are the finest statistical units, roughly equivalent to shires, which are built up by the Australian Bureau of Statistics and updated every five years. They are designed to reflect functional areas regarding the statistics they are built for (Fig. 1).

3. Methods

We first extracted a set of temporal NDVI metrics to complement the 16-day composites to predict actual wheat yields with different machine learning techniques (Section 3.1). Secondly, we trained nine base algorithms and relied on cross-validation to optimise their hyperparameters (Section 3.2). We also tested two ensembles to consolidate the estimates of the base learners (Section 3.3). The respective performances of all algorithms (base learners and ensembles) were assessed using cross-validation and compared against the two benchmark approaches (Section 3.4). Using the most accurate algorithm, we identified the most informative predictor variables and predicted wheat yields spatially across Australia for the 2015 growing season (Sections 3.4.5 and 3.5). For additional assessment, we compared the average yield predicted by the selected algorithm for each SA2 against the estimates provided by the Bureau of Statistics (Section 3.5). This last step required the production of a wheat map to mask out areas that were not sown to wheat.

3.1. Metric extraction

To remove outliers, MODIS NDVI time series were smoothed using a double logistic function which accurately models their specific shape (Beck et al., 2006; Battude et al., 2016):

$$NDVI(t) = (Peak - min) \times \left(\frac{1}{1 + e^{-Smax(t-TSmax)}} + \frac{1}{1 + e^{-Smin(t-TSmin)}} - 1 \right) \quad (1)$$

Table 2

Metrics derived from NDVI series and used as predictor variables of yield (12 in total).

Acronym	Metric	Definition	Ref.
Peak	Maximum	Maximum value within NDVI time series	(Lopresti et al., 2015)
min	Minimum	Minimum value within NDVI time series	(Battude et al., 2016)
amp	Amplitude	Amplitude of NDVI time series	(Battude et al., 2016)
Mean	Mean	Within-season mean of NDVI time series	(Bernardes et al., 2012)
Smax	Maximum slope	Maximum slope of the NDVI time series	This study
TSmax	Day maximum slope	Day corresponding to Smax	(Battude et al., 2016)
Smin	Minimum slope	Minimum slope of the NDVI time series	(Battude et al., 2016)
TSmin	Day minimum slope	Day corresponding to Smin	(Battude et al., 2016)
gsl	Growing season length	Number of days for which NDVI values are above the NDVI threshold	(Duveiller et al., 2013)
INT	Total integral	Total integral under curve of NDVI	(Jiang et al., 2004)
INTb	Integral before Peak	Integral before Peak under curve of NDVI for which NDVI values are above the minimum	(Labus et al., 2002)
INTa	Integral after Peak	Integral after Peak under curve of NDVI for which NDVI values are above the minimum	(Lai et al., 2018)
			This study
			(Benedetti and Rossini 1993)

where *Peak*, *min*, *Smax* and *Smin* correspond to the maximum, the minimum, the maximum slope, and the minimum slope of the NDVI time series respectively. *TSmax* and *TSmin* are the dates of maximum and minimum slopes. Fig. 2 illustrates the smoothing for a wheat pixel in Western Australia.

Based on the literature, we supplemented the smoothed NDVI values with a set of 12 temporal metrics (Table 2 and Fig. 2). Each of them highlights specific characteristics of the time series that were expected to correlate with yield. Time series were resampled on a daily step for calculating the growing season length and the integrals under the curve. The start and end of the season were defined as the first/last day when NDVI values were above a threshold defined as $\text{min} + 0.1\text{min}$.

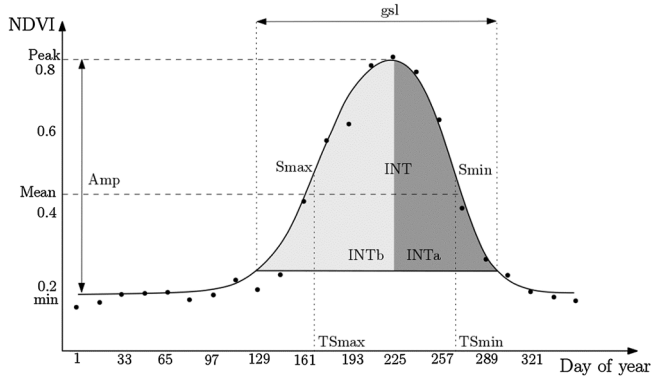


Fig. 2. Smoothed NDVI values and the corresponding metrics derived from an NDVI time series from an illustrative pixel located in Western Australia in 2015 with an average yield of 2.5 t ha^{-1} .

3.2. Calibration of the selected algorithms

The input to any machine learning algorithm is a set of predictor variables and the corresponding target values. Here, there were 81 predictor variables: 23 smoothed NDVI observations, 12 NDVI metrics, and 46 climate observations (biweekly cumulative rainfall and maximum temperature) as defined in Section 3.1, and the target value was the observed yield. The algorithm then learns to predict the target value by minimising the training loss or error. In this study, nine diverse learners were selected and compared: Random Forest (Breiman, 2001), XGBoost (Chen and Guestrin, 2016), Cubist (Quinlan, 1992), Multi-Layer Perceptron (Murtagh, 1991), Support Vector Regression (Vapnik et al., 1997), Gaussian Process (Williams, 1998), k-Nearest Neighbours (Györfi et al., 2002), and Multivariate Adaptive Regression Splines (Friedman, 1991). A brief description of the nine base learners is provided in Table 3; readers are referred to the Appendix and the seminal literature for more details.

Each algorithm has a set of hyper-parameters, properties of the model such as its complexity, structure or learning rate, that is defined before the learning process and must be optimised. The best hyper-parameters, those providing the most accurate predictions, were identified for each algorithm by grid search using a 10-fold cross-validation on the full data set (Table 3). Cross-validation consists of dividing the data into k independent folds of the same size, training the algorithm on $(k-1)$ folds and testing its accuracy on the remaining fold based on the error between predicted and target values.

3.3. Ensembles of base learners

We sought to consolidate the predictions of the individual base learners by combining them in ensembles. The rationale was that ensembles gather strengths and compensates for weaknesses of individual algorithms (Tumer and Ghosh, 1996). Specifically, we implemented an *average ensemble* (ens.mean), which simply averaged the predictions of all base learners, and a *Bayesian Data Fusion* (ens.BDF), which weighted

predictions according to the uncertainty associated with each learner and data point. Weights were defined based on the prediction uncertainty of the base learners as measured by their prediction variance (Bériaux et al., 2015):

$$w_i^k = \frac{1/\sigma_i^k}{\sum_{j=1}^n 1/\sigma_i^j} \quad (2)$$

where w_i^k is the weight of the k^{th} learner for the i^{th} data point, and where σ_i^k and σ_i^j are the prediction variances of learners k and j , and n the total number of algorithms used (9 in this case). Because all base learners did not directly provide estimates of the prediction variance, we generated the prediction uncertainty using bootstrapping (Efron, 1979). Bootstrapping generates subsets of the same dimension as the original data set by random sampling with replacement. Each sub-data set is a new sample, typically of the same size as the original. Thus, in each bootstrap sample, some of the original individuals may not be represented and others may be represented more than once. The variance corresponds to the prediction variance of the bootstrapped models. Here, we generated 25 bootstrapped calibration sets per base learner.

3.4. Model performance, sensitivity, and informative predictor variables

3.4.1. Pixel-based model accuracy

The performance of each algorithm was assessed using k -fold cross-validation; 100 folds were randomly generated using an 80%-20% split at the field-year level – or ca. 2400 and 603 pixels for each training and validation set, respectively. This approach ensured independence between the training and test set as all pixels belonging to the same field-year were assigned to the same fold, which prevented the models from being overtrained. All algorithms and ensembles were then trained with the training folds using the hyper-parameters optimised for the full data set and then tested on the validation subsets from which several performance metrics were derived. The Root Mean Square Error (RMSE) was used as the objective function during parameterisation and to assess the algorithm performance:

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

where y_i is the i^{th} observed yield and \hat{y}_i is the i^{th} prediction, and n is the total number of pixel observations. The Root Mean Square Percentage Error (RMSPE) and the coefficient of variation (R^2) were computed to further evaluate the models' performance:

$$RMSPE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(100 \left| \frac{y_i - \hat{y}_i}{y_i} \right| \right)^2} \quad (4)$$

$$R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} \quad (5)$$

where \bar{y} is the average observed yield.

3.4.2. Benchmarking against standard yield models

We implemented two benchmark models widely tested in the remote sensing literature. The first benchmark model (maxNDVI) establishes an empirical relationship between Peak NDVI and wheat yield through a linear regression model, where the dependent variable was represented by wheat yield and the independent variable by maximum NDVI ($\text{Yield} = a \text{ Peak} + b$; Lopresti et al., 2015).

The second benchmark model is a harvest index (HI) approach (see Lobell et al., 2003; Lobell 2013; Yuan et al., 2016; Löw et al., 2018, for instance). It relies on the light-use efficiency theory (Monteith, 1972) which states that the total biomass production is directly proportional to the total absorption of photosynthetically active radiation (PAR) over the course of the growing season. The HI approach derives grain yield

Table 3
Algorithm hyper-parameters tuned, and the range of values tested for each of them. N is the total number of combinations tested for each algorithm and p is the number of features.

Acronym	Algorithm	Description	Hyper-parameter	Values tested	N
RF	Random Forest	Ensemble learning model where the output averages the result of multiple regression trees.	Number of variables available at each node	2, 3, 5, 10, 20, 30, $p/3$, \sqrt{p}	8
CUB	Cubist	Rule-based model tree where terminal leaves contain linear regression models	Number of trees	500	
XB	XgBoost	Gradient-boosting ensemble model where regression trees are trained iteratively to predict residual errors of previous trees and are then added together.	Number of boosting iterations	From 3 to 6 by step of 1	28
MLP	Multi-Layer Perceptron	Artificial neural network which predicts the output variable using a nonlinear weighted combination of the predictor variables.	Maximum number of nodes per tree	From 2 to 8 by step of 1	18
			Learning rate	0.005, 0.01, 0.05	
			Number of neurons in hidden layers	Layer1: from 6 to 12 by step of 2	64
				Layer2: from 8 to 14 by step of 2	
				Layer3: from 2 to 8 by step of 2	
SVMl	Support Vector Regression -linear	Regressor which finds the optimal regression hyperplane so that most training samples lie within a certain margin around it.	Cost function parameter c	From 1 to 10 by step of 1	10
SVMr	Support Vector Regression – radial basis function	The optimal regression hyperplane is identified by mapping the data into a high dimensional feature space using non-linear kernel functions, here, radial basis functions [*] .	Cost function parameter c	From 1 to 5 by step of 0.1	50
			Margin of error tolerance ϵ	From 0.005 to 0.05 by step of 0.005	
GP	Gaussian Process regression	Bayesian non-parametric regression model	Parameter of the covariance function v	From 0.0025 to 0.03 by step of 0.0025	12
KNN	k-Nearest Neighbours	Instance-based learning algorithm where the output is the average of the values of its k nearest neighbours	Number of neighbours k	From 3 to 30 by step of 3	10
MARS	Multivariate Adaptive Regression Splines	Non-parametric method based on the divide-and-conquer strategy where complex non-linear relationships are estimated by piece-wise linear segments	Maximum number of parameters to be kept in the model	From 2 to 12 by step of 2	30
			Degree of freedom in the model	From 2 to 10 by step of 2	

* The inverse kernel width for the Radial Basis kernel function was optimised using heuristics per [Caputo et al. \(2002\)](#).

from total biomass using a scaling factor, *i.e.*, the harvest index (see SI2 for more details about our specific implementation). In comparison to maxNDVI, the HI model can capture variations in yield due to changes in PAR, and the interpolation of fraction of absorbed photosynthetically active radiation allows more flexibility in the timing of images relative to crop growth stage (Lobell, 2013). As for the machine learning models, the coefficients of the maxNDVI model and the harvest index were optimised during cross-validation to minimise the prediction residuals of the training sets.

3.4.3. Statistical analysis

We relied on non-parametric statistical tests to select the best model following recommendations by Demšar (2006). First, we carried out Friedman tests to determine whether there was a significant difference in mean ranks among the eleven machine learning models and the two benchmark models. The Friedman test detects statistical differences between treatments (different models) by comparing their ranks (Friedman 1937). Where the Friedman test rejected the null hypothesis of equal performance between methods, Nemenyi *post hoc* tests that compare pairwise performance differences were conducted (Nemenyi, 1963). Two algorithms were declared significantly different when the actual difference between their averaged rank was greater than the critical distance at the 95% confidence level. As these two statistical tests assess performance based on ranks, they were computed using 1-RMSPE so that low ranks indicate high performance, and conversely. Results from the Nemenyi *post hoc* test were presented on a critical distance diagram (Demšar, 2006), which shows the average rank of each algorithm and the critical distance, *i.e.*, the distance for which the difference between two algorithms is considered statistically different. The groups of algorithms that are not significantly different are also connected with a line.

To analyse the impact of climate data on the accuracy of the models, we repeated the cross-validation without climate variables as training input. We then compared the RMSPE values obtained with and without climate features by means of Wilcoxon signed-rank tests, which assess the existence of a significant difference between two groups (Wilcoxon, 1945).

3.4.4. Impact of the training set size on accuracy

To assess the sensitivity of the algorithms but also to identify the required number of fields, we retrained all base learners with reduced sample sizes ranging from 125 to 25 fields (*ca.* 2000–400 pixels). For consistency, the splits between the training and validation sets were the same as in the full cross-validation experiment (80%-20% split at the field-year level), and climate variables were included in the feature set. For each sample size and training split, the required number of field-years were randomly selected 25 times. All pixels belonging to the selected field-years were then used for training. To quantify the accuracy of the new models, we predicted the wheat yields for each validation set and computed the R^2 by comparing predictions against observations (see above). R^2 values were then averaged across all splits with the same sample size.

3.4.5. Predictor variables of wheat yield

We identified informative predictor variables for the best algorithm using a filter approach (Kuhn and Johnson, 2013). The relationships between predictor variables and yield were fit independently using local regression smoothers (Cleveland, 1979) which can accommodate non-linear relationships. The relative measure of importance was given by the R^2 of these models against the intercept only null model.

3.5. Country-scale wheat yield prediction and validation

3.5.1. Crop masking

To generate the national wheat map of 2015, we computed the phase and amplitude of the gap-filled NDVI time series as they can be

used to differentiate land cover and crop types (Wilson et al., 2018; Ghazaryan et al., 2018). Similar to Fourier transform, harmonic regression decomposes a time series into its constituent frequencies, thereby providing representation with fewer dimensions:

$$f(t) = a_0 + \sum_{i=1}^m \left(a_i \cos \frac{i2\pi t}{n} + b_i \sin \frac{i2\pi t}{n} \right) \quad (6)$$

where t is the NDVI timestamp value, n is the length of the cycle, and m is the number of harmonics, a_0 , a_i and b_i respectively represent the intercept, the cosine and the sine coefficients of the polynomials. The main difference compared with the Fourier transform is that harmonic regression requires defining the number of harmonics to keep before fitting the function. Geerken (2009) suggested to retain between three and five harmonics for time series imagery because a 3rd order Fourier series explains between 81 and 99% of the variance of MODIS time series. After masking non-cropland pixels, we applied ordinary least squares regression to fit a 3rd order harmonic to the NDVI time series for each cropland pixel in the study area. Non-cropland pixels were masked out using the Landsat-based GFSAD cropland map of 2015 (Teluguntla et al., 2018). After harmonic regression, seven features were available for classification (a_0 , a_1 , a_2 , a_3 , b_1 , b_2 , and b_3).

We then trained a random forest classification model on the coefficients extracted from the harmonic regression and the crop label data set. Random forests have been used with high accuracy in land cover, cropland, and crop type classification (Gislason et al., 2006; Waldner et al., 2015; Inglada et al., 2015). Instead of generating a hard output, *i.e.*, one class per pixel, we created a map of the wheat probability. Per-pixel wheat probabilities can be derived by counting the number of trees in the model that voted for the wheat class.

We finally generated a binary wheat/no-wheat map by defining probability cut-off values per SA2 so that the area covered by pixels above the threshold matched the wheat area reported in the official statistics.

3.5.2. Yield estimates by statistical units and validation

The best algorithm was re-calibrated on the whole yield data set (166 fields) and then applied all pixels identified as wheat in the crop map using the satellite and climate data from 2015. Using wheat area reported by the Australian Bureau of Statistics we masked the statistical units containing less than 21,000 ha sown to wheat (equivalent to approximately 15 farms). Eighty-eight statistical units covering more than 93% of the Australian wheat area remained for further analysis. We then averaged the pixel-level yield estimates of all wheat pixels within each of the selected statistical units. Finally, we compared the average predicted wheat yields with the reported yields and reported the RMSE, R^2 , and the bias (b).

4. Results

4.1. Performance of the algorithms

We trained each base learner with and without climate variables and computed their cross-validated performance (Table 4). Focusing on models that include climate variables, SVMr appeared as the best base learner regardless of the accuracy measure. It explained 77% of the wheat yield variability with an RMSE of 0.55 t ha⁻¹. Interestingly, its linear counter-part (SVML) displayed the lowest performance of all; it explained 55% of wheat yield variability (RMSE of 0.87 t ha⁻¹). The inclusion of climate variables (accumulated rainfalls and maximum temperatures) had a positive impact on prediction accuracy. Across all models, the RMSPE was on average 24% higher when climate variables were included. As indicated by the asterisks in Table 4, these differences were significant for most algorithms (Wilcoxon pairwise test; p -value < 0.05). Note that those base learners that did not show significant differences are those with poor performance, *i.e.*, MARS and

Table 4

Summary of the cross-validated performances of the base learners. Asterisks (*) indicate a significant difference between the accuracy obtained with and without climate variables according to a Wilcoxon test at the 95% confidence level. The best learner for each accuracy measure is indicated in bold. SVMr yielded the best performance across models trained with climate variables.

Accuracy measure	Base Learner								
	RF	CUB	XB	SVMI	SVMr	MLP	MARS	GP	kNN
<i>With climate variables</i>									
RMSE (t ha ⁻¹)	0.615*	0.787*	0.635*	0.872	0.545*	0.714*	0.787	0.554*	0.608*
RMSPE	0.455	0.567	0.441	0.622	0.400	0.480	0.594	0.404	0.427
R ²	0.713	0.575	0.691	0.551	0.770	0.668	0.557	0.764	0.725
<i>Without climate variables</i>									
RMSE (t ha ⁻¹)	0.735	0.837	0.767	0.775	0.739	0.848	0.782	0.728	0.730
RMSPE	0.581	0.616	0.515	0.603	0.592	0.613	0.601	0.582	0.558
R ²	0.592	0.521	0.579	0.562	0.593	0.534	0.551	0.599	0.603

SVMI.

The best ensemble was Ens.BDF and it explained 76% of the yield variability with an RMSE and RMSPE of 0.57 t ha⁻¹ and 40%, respectively (Table 5). Its performance remained below the performance of the single best learner when climate variables were not included. Adding climate variables led to significant improvements (> 0.14 t ha⁻¹) of the prediction accuracy of the ensemble models which then outperformed SVMr. Both benchmarks performed poorly in comparison with best learner and ensemble as they explained between 42 and 46% of the yield variability (RMSE > 0.735 t ha⁻¹), a drop of 30% compared to the single best machine learning method (SVMr). Removing climate variables from the feature set reduced these differences. In that case, HI exhibited a similar of slightly better RMSE than SVMr but still explained less variance.

There was a significant difference among regression models (Friedman test; $\chi^2 = 870$; p -value < 0.001). A *post hoc* analysis with Nemenyi tests was thus conducted to identify groups of algorithms that significantly differed from one another (Fig. 3). Algorithms belonging to the first group (highest average rank) included ens.BDF, SVMr, and GP. The last group consisted of SVMI, CUB, MARS, and the benchmarks HI and maxNDVI. As SVMr had the highest cross-validated accuracy regardless of the metrics and was not significantly different from the method with the highest average rank, it was selected as the single best yield prediction algorithm in all subsequent analyses.

The weights allocated to each algorithm for building ens.BDF, which were based on prediction uncertainty, were consistent with the estimated performance of the algorithms. The largest weights were allocated to GP, RF, SVMr and kNN which were assessed as the best single learners (Figs. 3 and 4). MARS, SVMI had generally weights below the average.

To further understand why ensembles were not rewarded with

Table 5

Summary of the cross-validated performances of ensembles and benchmarks. Asterisks (*) indicate a significant difference between the accuracy obtained with and without climate variables according to a Wilcoxon test at the 95% confidence level. The best ensemble was ens.BDF but it remained less accurate than SVMr. Benchmarks were largely outperformed by machine learning approaches.

Accuracy measures	Single best	Ensembles		Benchmarks	
	SVMr	ens.mean	ens.BDF	maxNDVI	HI
<i>With climate variables</i>					
RMSE (t ha ⁻¹)	0.545*	0.581*	0.559*	–	–
RMSPE	0.400	0.438	0.410	–	–
R ²	0.770	0.742	0.760	–	–
<i>Without climate variables</i>					
RMSE (t ha ⁻¹)	0.739	0.755	0.730	0.888	0.735
RMSPE	0.592	0.593	0.579	0.648	0.583
R ²	0.593	0.610	0.601	0.429	0.463

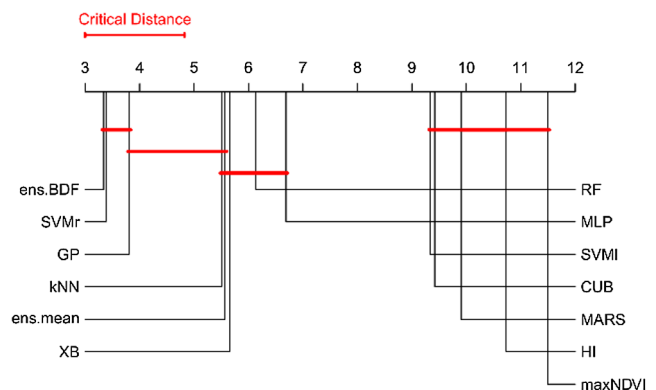


Fig. 3. Critical distance diagram of the algorithm performance with climate variable input. Groups of algorithms that are not significantly different are connected (p -value > 0.05). Here, ens.BDF, SVMr, and GP were the top three methods. These three methods are not significantly different from one another.

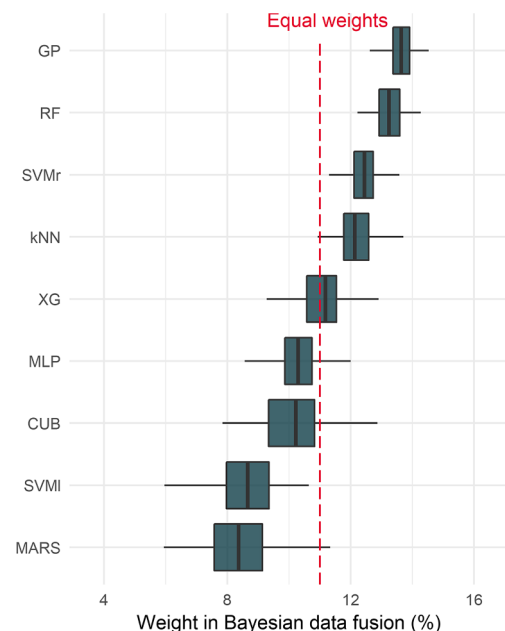


Fig. 4. Distribution of the obtained weights of each base learner in the Bayesian data fusion ensemble. The red vertical line corresponds to the weighting of the average ensemble, that is, equal weights.

higher accuracies, we analysed the correlations between the predictions of the base learners (Fig. 5). Correlations between the most accurate algorithms (e.g., SVMr or GP) were higher than the ones with lower accuracy (e.g., SVMI or MARS): the correlations exceeding 0.8 were

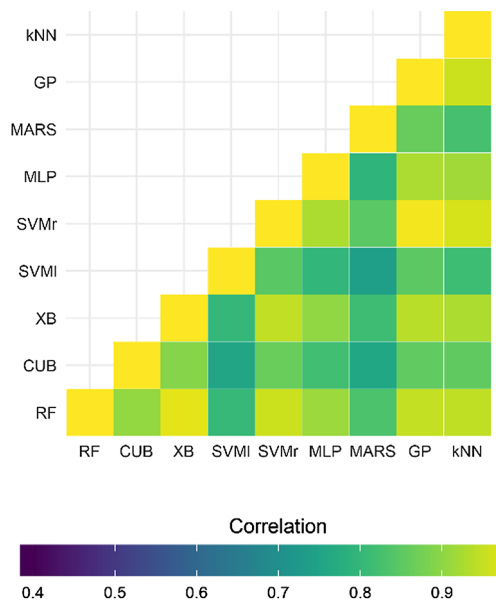


Fig. 5. Correlation between the predictions of base learners. High correlations explain the limited improvement of ensembles compared to base learners.

observed for GP, SVMr, RF, XB, and MLP while MARS, CUB and SVMl showed only low correlations with other algorithms. As ensembles rely on the assumption that prediction accuracy is improved by combining predictions of uncorrelated base learners, the high correlations of accurate algorithms shed some light on the limited improvement.

4.2. Effect of the training set size

The performance of base learners was found to improve as more data were added to the training set while the rank of algorithms did not change once 50 fields (ca. 800 pixels) were in the training set (Fig. 6). Algorithms with the highest response to added data included SVMr, GP, RF, kNN, and XB. Only SVMl did not show a steady increase in performance, which further underlined the poor suitability of this algorithm for wheat yield estimation. Although it seemed that most learners were close to saturation, they could probably still transform larger quantities of data into higher accuracy.

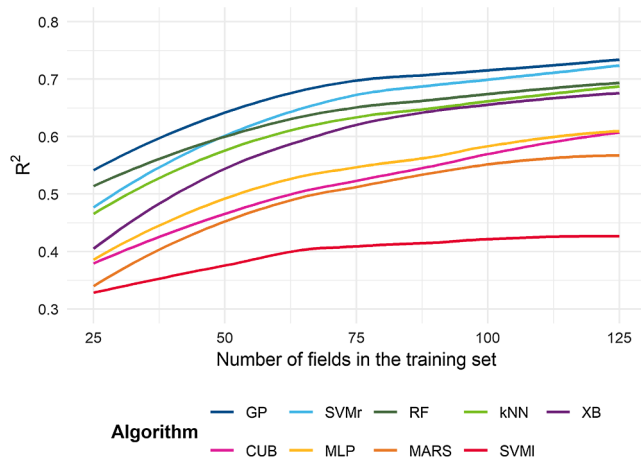


Fig. 6. Effect of training set size on the performance of machine-learning regression methods. Each field contained 16 pixels on average. The learners in the legend were ranked by decreasing accuracy when 125 fields were used.

4.3. Informative predictor variables for wheat yield prediction

Out of the top 14 predictor variables, seven were NDVI values, five were NDVI metrics and two were climate variables (Fig. 7). Four of the top five variables (NDVI 14/09; NDVI 29/08; NDVI 30/09; and Peak) occurred during the optimal flowering period throughout much of the cropping zone (Flohr et al., 2017). Mean, INT and cumulative rainfall on 16/10 are representative of the whole growing season, while NDVI 16/10, TSmin and INTa represent conditions during grain filling. NDVI 13/08 and NDVI 28/07, which stand among the lesser half of the predictor variables, relate to the pre-anthesis period.

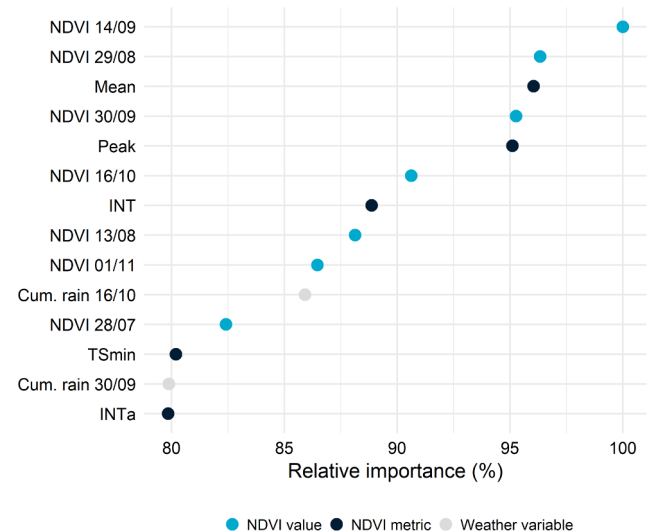


Fig. 7. Important variables for wheat yield prediction. Four of the top five variables (NDVI 14/09; NDVI 29/08; NDVI 30/09; and Peak) occur during optimal flowering period throughout much of the cropping zone.

4.4. National-scale wheat yield prediction

The accuracy of the crop classification assessed using the out-of-bag confusion matrix reached 86% (Table S1). Most classification errors were commission errors (17%), i.e., the classifier identified wheat too frequently. This issue was mitigated by constraining the final wheat map to the most likely wheat pixels so that the wheat area per statistical unit matched the area reported in by the Australian Bureau of Statistics. Omission errors for the wheat class did not exceed 5%.

The SVMr model was applied to all wheat pixels and the corresponding yield estimates were then averaged by statistical unit (see Fig. 1 for their locations) and compared against the reported yields (Fig. 8). Predicted and reported yields strongly agreed, with predictions explaining more than 70% of the yield variability. The small negative bias indicated that the model tends to over-predict yield, especially in those statistical units with low wheat areas. Estimates for statistical units with a large wheat area (> 300,000 ha) were closer to the 1:1 line. The spatial distribution of wheat yields across Australia is shown in Fig. 9.

5. Discussion

The main goal of this study was to develop a methodology based on machine learning to predict the actual wheat yields across Australia so that they could later be combined with estimates of potential yields in order to identify yield gap hotspots. Such information, together with a good understanding of the causes of the yield gap, is required to facilitate action towards closing the yield gap. The performance of the best learning algorithm (SVMr) showed high accuracy: it reached an RMSE of 0.545 t ha⁻¹ and an R² of 0.77 at the pixel level, and an R² of

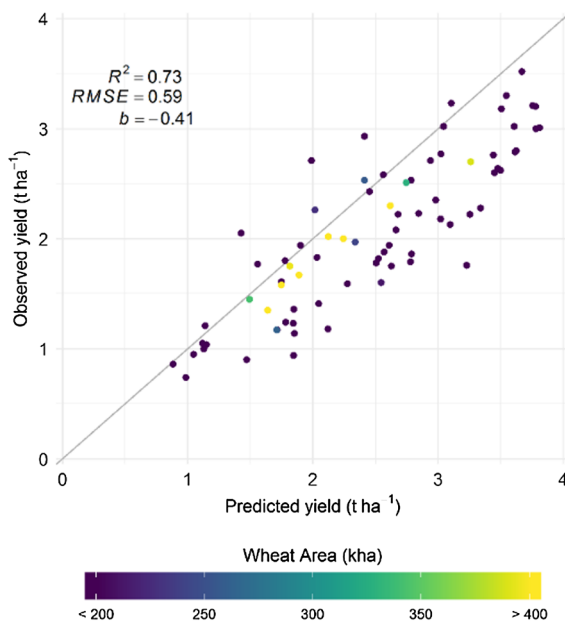


Fig. 8. Agreement between observed and predicted yields for Australian wheat in 2015. The predicted yields explain 70% of the observed yield variability and have an RMSE of 0.59 t ha^{-1} .

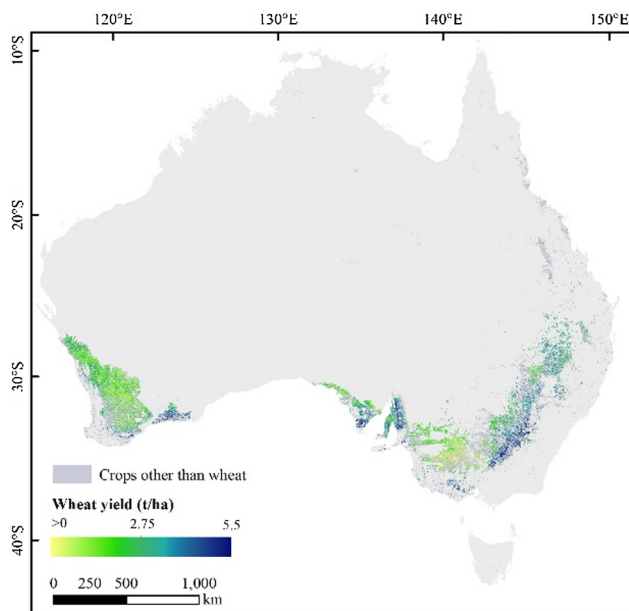


Fig. 9. Spatial distribution of wheat yields across Australia for 2015.

0.66 at the level of statistical units. We contend that this level of accuracy is fit for purpose because it is less than the size of the yield gap. As an example, Hochman et al., 2012 estimated that the yield gap over the whole Wimmera region of Victoria varied annually from 0.63 to 4.12 t ha^{-1} with an average of 2.00 t ha^{-1} .

The simplicity of the maximum NDVI and harvest index benchmarks came at the cost of a considerable deterioration of accuracy. This is probably related to the wide range of harvest indices encountered in Australia (Unkovich et al., 2010) which such simple models cannot accommodate. While adding climate predictors to these approaches might improve their prediction skills, it is unlikely that these simple models would challenge machine-learning regression models. The performance of some machine learning methods represents a clear improvement compared to other approaches tested in Australia,

although the validation protocols differ. For instance, Lai et al., 2018 predicted wheat yield at the pixel level with an average RMSE of 0.79 t ha^{-1} using an empirical model based on a time-integrated Landsat NDVI metric while C-Crop, an assimilation model, explained 68% of field-scale yield variability (Donohue et al., 2018). The model also outperformed other regional empirical models based on NDVI and meteorological data such as the one proposed for wheat yield modelling in Ukraine by Kogan et al. (2013) who reported RMSEs ranging from 0.565 t ha^{-1} to 0.794 t ha^{-1} . A deterministic crop growth model (APSIM) can predict grain yields with an R^2 of 0.77 across a range of varieties, management practices and growing conditions (Asseng et al., 1998). Therefore, machine learning of climate and NDVI time series appears to be an efficient way of building large-scale yield prediction models, as its performance was comparable or better than other approaches. Advantage of machine learning approaches over conventional empirical models has already been reported in Jiang et al. (2004) and Uno et al. (2005) and might be attributed to their ability to absorb large quantities of data, and the absence of assumptions about the underlying data distribution. Furthermore, these methods seem to overcome the issue of local relevance undermining most other empirical approaches: a single model achieved high accuracy at the pixel level across years and provided consistent yield estimates nationally. SVM with radial basis functions emerged as the single best model for wheat yield modelling in Australia whereas its linear counter-part consistently ranked last. Its high performance may be due to its use of the risk minimisation principle which minimises an upper bound of the generalisation error rather than minimizing the training error, eventually leading to better generalisation performance. Moreover, significant differences in accuracy were observed between non-linear and linear models. However, the fact that previous research reported that SVM performed poorly compared to RF and GP (Aghighi et al., 2018) or that non-linear models did not show an advantage over linear models (Johnson, 2016) underlines the difficulty of making assumptions about the performance of an algorithm. It might not only depend on the crop of interest, but also on the environment, the predictor variables, and the training set. Ranking might also differ according to the performance metric selected.

The most useful NDVI values were those measured around the anthesis period (NDVI 14/09; NDVI 29/08; NDVI 30/09; and Peak) when viable grain number is determined. Next in importance were NDVI and climate variables that integrate over the whole wheat growth season (Mean, INT and Cum. Rain 16/10) and thus reflect overall seasonal conditions which in turn determines total above-ground dry matter production. Other important values reflect conditions during the grain-filling period (NDVI 16/10, TSmin and INTa). These observations are consistent with crop physiological analysis findings that viable grain number and total above-ground dry matter are much more highly correlated to grain yield than grain weight (Rathey et al., 2009; Lawes et al., 2016). The absence of informative yield predictors before late July reflects the high uncertainty about final wheat yields before this date. Despite their coarser spatial resolution, including climate variables as predictor variables led to significant improvement of the model accuracy, which suggests that they provide complementary information to 16-day NDVI time series. Cumulative rainfall in mid-October (16/10) emerged as the most important climate variable for wheat yield prediction which was also picked up by other empirical estimation models (Waldner et al., 2019). The importance of within-season rainfall to predict yield is consistent with other studies that demonstrated a high correlation between seasonal rainfall and grain yield such as French and Schultz (1984).

There are multiple pathways to further improve the model accuracy such as increasing the size of the training set, including more predictor variables, or using time series with shorter revisit cycles. All models tested in this study were sensitive to reduced training data sets. This analysis also highlighted that saturation was not yet achieved, i.e., additional calibration data are likely to improve accuracy as more

diverse growing conditions become available for learning. For instance, Johnson et al. (2016) reported that additional training data from years with high or low yields improved yield forecast accuracy, especially for extreme values. This would certainly be observed with SVM as extreme or rare observations are likely to become support vectors. Learning from a larger set of predictor variables and thus from a more complete description of plant growth is another option to improve accuracy. NDVI is affected by soil colour and tends to saturate for dense canopies so synergistic combination with other indices is likely to bring about improvement. Complementing NDVI with Enhanced Vegetation Index, which does not saturate as rapidly as NDVI in response to over dense canopy cover, led to a significant improvement in skill score compared to models using NDVI solely (Johnson et al., 2016). Besides, using data from various sources significantly improves regional crop yield predictions (Guan et al., 2017). In particular, QuickSCAT Ku-band backscatter, thermal-based ALEXI Evapotranspiration and AMSR-E X-band passive microwave Vegetation Optical Depth provided them unique information on environmental stresses and improved overall crop yield predictive skill. Finally, biweekly compositing might not adequately characterise crop productivity at the pixel level if crop critical periods are smoothed by the compositing algorithm (Labus et al., 2002). Noise reduction can indeed limit the ability in detecting detailed variations during periods near the seasonal peak (Cai et al., 2017; Zhao and Zhang, 2018; Shao et al., 2016). Increasing the temporal resolution will lead to improved yield estimates without changing the estimation method (Waldner et al., 2019). Indeed, the higher temporal resolution provides a finer description of temporal patterns and, in turn, more informative temporal metrics. Interestingly, the relative importance of climate data also decreases as the temporal resolution increases (Waldner et al., 2019). As adding predictors by increasing the sources or the temporal frequency of the data, a feature selection step should be considered to reduce the training time and the risk of overfitting.

It is also worth noting that the accuracy of yield estimates using MODIS time series was also affected by pixel purity. Low pixel purity typically occurs at the edges of large fields or where the landscape fragmentation is high, i.e. small fields and large crop diversity (Waldner et al., 2018). These effects are particularly relevant for off-nadir pixels with large viewing angles (Duveiller et al., 2011). Therefore, the multi-angularity of MODIS is likely to have impacted the results in two ways. First, all boundary pixels were retained in the analyses, regardless of their purity. The average field size in our data set was larger than 100 ha (16 MODIS pixels) so that MODIS still largely provides pure observations. An important next step would certainly be to relate yield prediction errors to pixel purity as proposed by Duveiller et al. (2015). Secondly, the point spread function of the sensor, which depicts the response of the satellite to a point source, was neglected when observed yields were resampled to 250 m. As a substantial portion of the spectral signal of each pixel comes also from surrounding of the original footprint, there is an intrinsic mismatch between the target and the observed spatial response.

There are several data quality issues that might have impacted our results. First, reported yields are based on information obtained from agricultural businesses that responded to the 2015 Agricultural Census. These estimates are subject to sampling error depending on the response rate and it was estimated that the reported yield estimates had a relative standard error of < 10%. Second, the yield maps from which training data were derived were likely affected by miscalibration of the harvester equipment. Such errors are frequent and occur for instance when the flow sensor or the harvest width are not properly set up. A sample selection bias might also have occurred while collecting these yield maps as they were sourced from cooperative producers and consultants which tend to reach above-average yields. This selection bias might partly explain the prediction bias observed across statistical units. Third, yield estimates were also likely affected by the quality of the wheat mask. To minimize classification errors, we used imagery with resolutions substantially finer than the scale of crop fields. The

wheat map generally displayed consistent spatial patterns rather than isolated pixels or over-sized patches. Nonetheless, the more accurate yield estimations in statistical units with large areas sown to wheat are probably related to a reduced likelihood of classification errors in the wheat map. As yield prediction models are crop-specific, classification errors would propagate through to the yield estimation process. Collecting a larger training set, increasing the feature set size with new spectral indices or data sources, and introducing stratification (pre-classification based on ancillary data) are some of the options to further improve classification accuracy (Lu and Weng, 2007). Higher-quality data are thus expected to improve the skill of the regression models as well as to provide more reliable validation data.

The overarching goal of this work was to advance towards more accurate and precise estimates of the wheat yield gap at scale which explains why we focused on end-of-season yield estimates. The high performance of the yield model indicates directions for future work. Foremost among these is extending the proposed approach to in-season yield forecasting. This would entail slight modifications of how the smoothing is performed as the current use of double logistic functions requires annual time series and, therefore, prevents dynamical computing of temporal metrics. This could easily be overcome by changing the smoothing method and defining new ways of extracting temporal metrics. The main constraint to operationalisation resides in the lack of in-season crop type data to produce a crop mask. While access to sufficient ground truth data (both yield maps and crop type data) and official statistics has limited our analysis to wheat and to the selected years, it is expected that the approach could be extended to additional crops when more yield maps become available.

6. Conclusion

One of the prerequisites to closing the wheat yield gap in Australia is to identify hotspots of yield gap. In this study, we capitalised on a large training data set of very high-resolution yield maps obtained from harvesters to predict actual wheat yields in Australia using machine learning based on climate records (rainfall and maximum temperature) and NDVI time series data. Out of the nine base learners and two ensembles that were compared, support vector regression with radial basis functions emerged as the single best algorithm and predicted pixel-level yield estimates with an R^2 of 0.77 and an RMSE of 0.55 t ha^{-1} . This corresponds to an improvement of $> 0.2 \text{ t ha}^{-1}$ when compared to standard yield models. At the national scale, it successfully explained 73% of yield variability across statistical areas. Climate variables such as maximum temperature and accumulated rainfall were found to synergistically complement the 16-day NDVI time series and metrics as they significantly improved the model accuracy. NDVI and climate features that were aligned with conditions around anthesis, overall seasonal conditions and conditions during the grain-filling period were among the more important predictors of grain yield. Results showed that non-linear models clearly outperformed linear models and that ensembles, despite their larger computational cost, did not show a significant advantage over the single best model. All models were sensitive to a reduction of the training set size and most had not yet reached saturation. This indicates that additional training data are likely to further improve the skill of the models. We conclude that machine learning of crop yield is a promising approach for delivering accurate actual yield estimates at the pixel level across large areas and years. This work provides the basis for mapping the yield gap at field scale across Australia to identify yield gap hotspots which can be targeted for further work by researchers and advisers.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors received funding from “GrainCast”, a project of the CSIRO Digiscape Future Science Platform.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.isprsjprs.2019.11.008>.

References

- Aghighi, H., Azadbakht, M., Ashourloo, D., Shahrabadi, H.S., Radiom, S., 2018. Machine learning regression techniques for the silage maize yield prediction using time-series images of Landsat 8 OLI. *IEEE J. Sel. Top. Appl. Earth Observ. Rem. Sens.* 1–15. <https://doi.org/10.1109/JSTARS.2018.2823361>.
- Asseng, S., Keatinge, B.A., Fillery, I.R.P., Gregory, P.J., Bowden, J.W., Turner, N.C., Palta, J.A., Abrecht, D.G., 1998. Performance of the APSIM-wheat model in Western Australia. *Field Crops Res.* 57 (2), 163–179. [https://doi.org/10.1016/S0378-4290\(97\)00117-2](https://doi.org/10.1016/S0378-4290(97)00117-2).
- Battude, Marjorie, Bitar, Ahmad Al, Morin, David, Cros, Jérôme, Huc, Mireille, Sicre, Claire Marais, Le Dantec, Valérie, Demarez, Valérie, 2016. Estimating maize biomass and yield over large areas using high spatial and temporal resolution sentinel-2 like remote sensing data. *Rem. Sens. Environ.* 184 (October), 668–681. <https://doi.org/10.1016/j.rse.2016.07.030>.
- Bazgeer, S., Kamali, Gh., Mortazavi, A., 2007. “Wheat Yield Prediction through Agrometeorological Indices for Hamedan, Iran,” pp. 7.
- Beck, Pieter S.A., Atzberger, Clement, Högda, Kjell Arild, Johansen, Bernt, Skidmore, Andrew K., 2006. Improved monitoring of vegetation dynamics at very high latitudes: a new method using MODIS NDVI. *Rem. Sens. Environ.* 100 (3), 321–334. <https://doi.org/10.1016/j.rse.2005.10.021>.
- Behmann, Jan, Mahlein, Anne-Katrin, Rumpf, Till, Römer, Christoph, Plümer, Lutz, 2015. A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. *Precision Agriculture* 16 (3), 239–260. <https://doi.org/10.1007/s11119-014-9372-7>.
- Benedetti, R., Rossini, P., 1993. On the use of NDVI profiles as a tool for agricultural statistics: the case study of wheat yield estimate and forecast in emilia romagna. *Rem. Sens. Environ. (USA)*. <http://agris.fao.org/agris-search/search.do?recordID=US9421008>.
- Bériaux, Emilie, Waldner, François, Collienne, François, Bogaert, Patrick, Defourny, Pierre, 2015. Maize Leaf area index retrieval from synthetic quad pol SAR time series using the water cloud model. *Remote Sensing* 7 (12), 16204–16225. <https://doi.org/10.3390/rs71215818>.
- Bernardes, T., Alves Moreira, M., Adami, M., Friedrich Theodor Rudolf, B., 2012. Monitoring biennial bearing effect on coffee yield using modis remote sensing imagery. In: 2012 IEEE International Geoscience and Remote Sensing Symposium, pp. 3760–3763. <https://doi.org/10.1109/IGARSS.2012.6350499>.
- Bramley, R.G.V., Williams, S.K., 2001. “A Protocol for the Construction of Yield Maps from Data Collected Using Commercially Available Grape Yield Monitors.” <https://publications.csiro.au/rpr/pub?list=BRO&pid=changement:2061&sb=RECENT&n=14&rpp=25&page=181&tr=4579&dr=all&dc4.browseYear=2001>.
- Breiman, Leo, 2001. Random Forests. *Machine Learning* 45 (1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Cai, Zhanzhang, Jönsson, Per, Jin, Hongxiao, Eklundh, Lars, 2017. Performance of smoothing methods for reconstructing NDVI time-series and estimating vegetation phenology from MODIS data. *Remote Sensing* 9 (12), 1271. <https://doi.org/10.3390/rs9121271>.
- Caputo, Barbara, Kim Lan Sim, Furesjo, F., Alexander J. Smola, 2002. “Appearance-Based Object Recognition Using SVMs: Which Kernel Should I Use.” In.
- Chen, Tianqi, and Carlos Guestrin. 2016. “XGBoost: a scalable tree boosting system.” In, 785–94. ACM Press. <https://doi.org/10.1145/2939672.2939785>.
- Cleveland, William S., 1979. Robust locally weighted regression and smoothing scatterplots. *J. Am. Stat. Assoc.* 74 (368), 829–836. <https://doi.org/10.1080/01621459.1979.10481038>.
- Demšar, Janez, 2006. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* 7 (Jan), 1–30.
- Donohue, R.J., Hume, I.H., Roderick, M.L., McVicar, T.R., Beringer, J., Hutley, L.B., Gallant, J.C., et al., 2014. Evaluation of the remote-sensing-based DIFFUSE model for estimating photosynthesis of vegetation. *Rem. Sens. Environ.* 155 (December), 349–365. <https://doi.org/10.1016/j.rse.2014.09.007>.
- Donohue, Randall J., Lawes, Roger A., Mata, Gonzalo, Gobbett, David, Ouzman, Jackie, 2018. Towards a national, remote-sensing-based model for predicting field-scale crop yield. *Field Crops Res.* 227 (October), 79–90. <https://doi.org/10.1016/j.fcr.2018.08.005>.
- Donohue, Randall J., McVicar, Tim R., Roderick, Michael L., 2010. Assessing the ability of potential evaporation formulations to capture the dynamics in evaporative demand within a changing climate. *J. Hydrol.* 386 (1), 186–197. <https://doi.org/10.1016/j.jhydrol.2010.03.020>.
- Doraiswamy, P.C., Hatfield, J.L., Jackson, T.J., Akhmedov, B., Prueger, J., Stern, A., 2004. Crop condition and yield simulations using Landsat and MODIS. *Rem. Sens. Environ.* 92 (4), 548–559. <https://doi.org/10.1016/j.rse.2004.05.017>.
- Duveiller, Grégory, Baret, Frédéric, Defourny, Pierre, 2011. Crop specific green area index retrieval from MODIS data at regional scale by controlling pixel-target adequacy. *Rem. Sens. Environ.* 115 (10), 2686–2701. <https://doi.org/10.1016/j.rse.2011.05.026>.
- Duveiller, Grégory, López-Lozano, Raúl, Baruth, Bettina, 2013. Enhanced processing of 1-Km spatial resolution FAPAR time series for sugarcane yield forecasting and monitoring. *Remote Sensing* 5 (3), 1091–1116. <https://doi.org/10.3390/rs5031091>.
- Duveiller, Grégory, Lopez-Lozano, Raul, Cescatti, Alessandro, 2015. Exploiting the multi-angularity of the MODIS temporal signal to identify spatially homogeneous vegetation cover: a demonstration for agricultural monitoring applications. *Rem. Sens. Environ.* 166 (September), 61–77. <https://doi.org/10.1016/j.rse.2015.06.001>.
- Efron, B., 1979. “Bootstrap Methods: Another Look at the Jackknife.” 1979. <https://projecteuclid.org/euclid.aos/1176344552>.
- Fang, Hongliang, Liang, Shunlin, Hoogenboom, Gerrit, 2011. Integration of MODIS LAI and vegetation index products with the CSM-CERES-maize model for corn yield estimation. *Int. J. Rem. Sens.* 32 (4), 1039–1065. <https://doi.org/10.1080/0143160903505310>.
- Floh, B.M., Hunt, J.R., Kirkegaard, J.A., Evans, J.R., 2017. Water and temperature stress define the optimal flowering period for wheat in South-Eastern Australia. *Field Crops Res.* 209 (August), 108–119. <https://doi.org/10.1016/j.fcr.2017.04.012>.
- French, R.J., Schultz, J.E., 1984. Water Use Efficiency of wheat in a Mediterranean-type environment. I. The relation between yield, Water Use and climate. *Aust. J. Agric. Res.* 35 (6), 743–764. <https://doi.org/10.1071/ar9840743>.
- Friedman, Jerome H., 1991. Multivariate adaptive regression splines. *Ann. Stat.* 19 (1), 1–67. <https://doi.org/10.1214/aos/1176347963>.
- Friedman, Milton, 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Am. Stat. Assoc.* 32 (200), 675–701. <https://doi.org/10.1080/01621459.1937.10503522>.
- Geerken, Roland A., 2009. An algorithm to classify and monitor seasonal variations in vegetation phenologies and their inter-annual change. *ISPRS J. Photogramm. Rem. Sens.* 64 (4), 422–431. <https://doi.org/10.1016/j.isprsjprs.2009.03.001>.
- Ghazaryan, Gohar, Duboviyk, Olena, Löw, Fabian, Lavreniuk, Mykola, Kolotii, Andrii, Schellberg, Jürgen, Kussul, Nataliia, 2018. A rule-based approach for crop identification using multi-temporal and multi-sensor phenological metrics. *Eur. J. Rem. Sens.* 51 (January), 511–524. <https://doi.org/10.1080/22797254.2018.1455540>.
- Gislason, Pall Oskar, Benediktsson, Jon Atli, Sveinsson, Johannes R., 2006. Random forests for land cover classification. *Patt. Recog. Lett.* 27 (4), 294–300. <https://doi.org/10.1016/j.patrec.2005.08.011>.
- Godfray, H. Charles J., Beddington, John R., Crute, Ian R., Haddad, Lawrence, Lawrence, David, Muir, James F., Pretty, Jules, Robinson, Sherman, Thomas, Sandy M., Toulmin, Camilla, 2010. Food security: the challenge of feeding 9 billion people. *Science* 327 (5967), 812–818. <https://doi.org/10.1126/science.1185383>.
- Guan, Kaiyu, Wu, Jin, Kimball, John S., Anderson, Martha C., Frolking, Steve, Li, Bo, Hain, Christopher R., Lobell, David B., 2017. The shared and unique values of optical, fluorescence, thermal and microwave satellite data for estimating large-scale crop yields. *Rem. Sens. Environ.* <https://doi.org/10.1016/j.rse.2017.06.043>.
- Györfi, László, Kohler, Michael, Krzyżak, Adam, Walk, Harro, 2002. A Distribution-Free Theory of Nonparametric Regression. Springer Series in Statistics. New York, NY: Springer New York. <https://doi.org/10.1007/b97848>.
- Hammond, Allen L., 1975. Crop forecasting from space: toward a global food watch. *Science* 188 (4187), 434–436.
- Hochman, Z., van Rees, H., Carberry, P.S., Hunt, J.R., McCown, R.L., Gartmann, A., Holzworth, D., et al., 2009. Re-inventing model-based decision support with Australian Dryland farmers. 4. Yield prophet helps farmers monitor and manage crops in a variable climate. *Crop Past. Sci.* 60 (11), 1057. <https://doi.org/10.1071/CP09020>.
- Hochman, Zvi, Gobbett, David, Holzworth, Dean, McClelland, Tim, van Rees, Harm, Marinoni, Oswald, Garcia, Javier Navarro, Horan, Heidi, 2012. Quantifying yield gaps in rainfed cropping systems: a case study of wheat in Australia. *Field Crops Res.* 136 (September), 85–96. <https://doi.org/10.1016/j.fcr.2012.07.008>.
- Hochman, Zvi, Gobbett, David, Horan, Heidi, Garcia, Javier Navarro, 2016. Data rich yield gap analysis of wheat in Australia. *Field Crops Res.* 197 (October), 97–106. <https://doi.org/10.1016/j.fcr.2016.08.017>.
- Hochman, Z., Gobbett, D.L., Horan, H., 2017. Climate trends account for stalled wheat yields in Australia since 1990. *Global Change Biol.* 23 (5), 2071–2081. <https://doi.org/10.1111/gcb.13604>.
- Hochman, Zvi, Horan, Heidi, 2018. Causes of wheat yield gaps and opportunities to advance the water-limited yield frontier in Australia. *Field Crops Res.*
- Holzworth, D.P., Huth, N.I., deVoil, P.G., Zurcher, E.J., Herrmann, N.I., McLean, G., Chenu, K., et al., 2014. APSIM—evolution towards a new generation of agricultural systems simulation. *Environ. Model. Soft.* 62 (December), 327–350. <https://doi.org/10.1016/j.envsoft.2014.07.009>.
- Houborg, Rasmus, McCabe, Matthew F., 2018. A hybrid training approach for leaf area index estimation via cubist and random forests machine-learning. *ISPRS J. Photogramm. Rem. Sens.* 135, 173–188. <https://doi.org/10.1016/j.isprsjprs.2017.10.004>.
- Inglada, Jordi, Arias, Marcela, Tardy, Benjamin, Hagolle, Olivier, Valero, Silvia, Morin, David, Dedieu, Gérard, et al., 2015. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. *Remote Sensing* 7 (9), 12356–12379. <https://doi.org/10.3390/rs70912356>.
- Ittersum, Martin K. Van, Cassman, Kenneth G., Grassini, Patricio, Wolf, Joost, Titttonell, Pablo, Hochman, Zvi, 2013. Yield Gap Analysis with Local to Global Relevance—A Review. *Field Crops Res.* 143 (March), 4–17. <https://doi.org/10.1016/j.fcr.2012.09.009>.
- Jeffrey, Stephen J., Carter, John O., Moodie, Keith B., Beswick, Alan R., 2001. Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environ. Model. Soft.* 16 (4), 309–330. [https://doi.org/10.1016/S1364-8152\(01\)134](https://doi.org/10.1016/S1364-8152(01)134)

- 00008-1.
- Jiang, D., Yang, X., Clinton, N., Wang, N., 2004. An Artificial neural network model for estimating crop yields using remotely sensed information. *Int. J. Rem. Sen.* 25 (9), 1723–1732. <https://doi.org/10.1080/0143116031000150068>.
- Johnson, D.M., 2016. A comprehensive assessment of the correlations between field crop yields and commonly used MODIS products. *Int. J. Appl. Earth Observ. Geoinf.* 52, 65–81.
- Johnson, Michael D., Hsieh, William W., Cannon, Alex J., Davidson, Andrew, Bédard, Frédéric, 2016. Crop Yield forecasting on the canadian prairies by remotely sensed vegetation indices and machine learning methods. *Agric. Forest Meteorol.* 218–219 (March), 74–84. <https://doi.org/10.1016/j.agrformet.2015.11.003>.
- Khaki, Saeed, Wang, Lizhi, 2019. Crop Yield Prediction using deep neural networks. 621–621. *Frontiers Plant Sci.* 10 (May). <https://doi.org/10.3389/fpls.2019.00621>.
- Kogan, F., Kussul, N., Adamenko, T., Skakun, S., Kravchenko, O., Kryvobok, O., Shelestov, A., Kolotii, A., Kussul, O., Lavrenyuk, A., 2013. Winter wheat yield forecasting in Ukraine based on earth observation, meteorological data and biophysical models. *Int. J. Appl. Earth Observ. Geoinf.* 23 (August), 192–203. <https://doi.org/10.1016/j.jag.2013.01.002>.
- Kuhn, Max, Johnson, Kjell, 2013. Applied Predictive Modeling. In: Kuhn, Max, Johnson, Kjell, (Eds.). *Applied Predictive Modeling*. New York, NY: Springer New York, pp. 1–16. https://doi.org/10.1007/978-1-4614-6849-3_1.
- Kuwata, K., Shibasaki, R., 2016. Estimating corn yield in the united states with modis evi and machine learning methods. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, III–8: 131–36. Copernicus GmbH. <https://doi.org/10.5194/isprs-annals-III-8-131-2016>.
- Labus, M.P., Nielsen, G.A., Lawrence, R.L., Engel, R., Long, D.S., 2002. Wheat yield estimates using multi-temporal NDVI satellite imagery. *Int. J. Rem. Sen.* 23 (20), 4169–4180. <https://doi.org/10.1080/01431160110107653>.
- Lai, Y.R., Pringle, M.J., Kopittke, P.M., Menzies, N.W., Orton, T.G., Dang, Y.P., 2018. An empirical model for prediction of wheat yield, using time-integrated landsat NDVI. *Int. J. Appl. Earth Observ. Geoinf.* 72 (October), 99–108. <https://doi.org/10.1016/j.jag.2018.07.013>.
- Lawes, R.A., Huth, N.D., Hochman, Z., 2016. Commercially available wheat cultivars are broadly adapted to location and time of sowing in Australia's grain zone. *Eur. J. Agronomy* 77 (July), 38–46. <https://doi.org/10.1016/j.eja.2016.03.009>.
- Lobell, David B., 2013. The use of satellite data for crop yield gap analysis. *Field Crops Res.* 143 (March), 56–64. <https://doi.org/10.1016/j.fcr.2012.08.008>.
- Lobell, David B., Asner, Gregory P., Ivan Ortiz-Monasterio, J., Benning, Tracy L., 2003. Remote sensing of regional crop production in the Yaqui Valley, Mexico: estimates and uncertainties. *Agric. Ecosyst. Environ.* 94 (2), 205–220. [https://doi.org/10.1016/S0167-8809\(02\)00021-X](https://doi.org/10.1016/S0167-8809(02)00021-X).
- Lopresti, Mariano F., Di Bella, Carlos M., Degioanni, Américo J., 2015. Relationship between MODIS-NDVI data and wheat yield: a case study in Northern Buenos Aires Province, Argentina. *Inform. Proc. Agric.* 2 (2), 73–84. <https://doi.org/10.1016/j.inpa.2015.06.001>.
- Löw, Fabian, Biradar, Chandrashekar, Dubovyk, Olena, Fliemann, Elisabeth, Akramkhanov, Akmal, Vallejo, Alejandra Narvaez, Waldner, Francois, 2018. Regional-scale monitoring of cropland intensity and productivity with multi-source satellite image time series. *GI Sci. Rem. Sen.* 55 (4), 539–567. <https://doi.org/10.1080/15481603.2017.1414010>.
- Lu, D., Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. *Int. J. Rem. Sen.* 28 (5), 823–870. <https://doi.org/10.1080/01431160600746456>.
- Mishra, Subhadra, Mishra, Debahuti, Santra, Gour Hari, 2016. Applications of machine learning techniques in agricultural crop production: a review paper. *Indian J. Sci. Technol.* 9 (38).
- Monteith, J.L., 1972. Solar radiation and productivity in tropical ecosystems. *J. Appl. Ecol.* 9 (3), 747–766. <https://doi.org/10.2307/2401901>.
- Morain, S.A., Williams, D.L., 1975. Wheat production estimates using satellite images 1. *Agronomy J.* 67 (3), 361–364. <https://doi.org/10.2134/agronj1975.00021962006700030020x>.
- Mueller, Nathaniel D., Gerber, James S., Johnston, Matt, Ray, Deepak K., Ramankutty, Navin, Foley, Jonathan A., 2012. Closing yield gaps through nutrient and water management. *Nature* 490 (7419), 254–257. <https://doi.org/10.1038/nature11420>.
- Murtagh, Fionn, 1991. Multilayer perceptrons for classification and regression. *Neurocomputing* 2 (5), 183–197. [https://doi.org/10.1016/0925-2312\(91\)90023-5](https://doi.org/10.1016/0925-2312(91)90023-5).
- Nelson, Erik, Sander, Heather, Hawthorne, Peter, Conte, Marc, Ennaanay, Driss, Wolny, Stacie, Manson, Steven, Polasky, Stephen, 2010. Projecting global land-use change and its effect on ecosystem service provision and biodiversity with simple models. Edited by Adina Maya Merenlender. *PLoS ONE* 5 (12), e14327. <https://doi.org/10.1371/journal.pone.0014327>.
- Nemenyi, Peter, 1963. Distribution-Free Multiple Comparisons.
- Palosuo, Taru, Kersebaum, Kurt Christian, Angulo, Carlos, Hlavinka, Petr, Moriondo, Marco, Olesen, Jørgen E., Patil, Ravi H., et al., 2011. Simulation of winter wheat yield and its variability in different climates of Europe: a comparison of eight crop growth models. *Eur. J. Agron.* 35 (3), 103–114. <https://doi.org/10.1016/j.eja.2011.05.001>.
- Phalan, Benjamin, 2018. What have we learned from the land sparing-sharing model? *Sustainability* 10 (6), 1760. <https://doi.org/10.3390/su10061760>.
- Quinlan, J. R. 1992. "Learning With Continuous Classes." *World Scientific*, pp. 343–348.
- Rathey, Allan, Shorter, Ray, Chapman, Scott, Drecker, Fernanda, van Herwaarden, Anthony, 2009. Variation for and relationships among biomass and grain yield component traits conferring improved yield and grain weight in an elite wheat population grown in variable yield environments. *Crop Pasture Sci.* 60 (8), 717–729. <https://doi.org/10.1071/CP08460>.
- Rivera-Caicedo, Juan Pablo, Verrelst, Jochem, Muñoz-Marí, Jordi, Camps-Valls, Gustau, Moreno, José, 2017. Hyperspectral dimensionality reduction for biophysical variable statistical retrieval. *ISPRS J. Photogramm. Rem. Sen.* 132 (October), 88–101. <https://doi.org/10.1016/j.isprsjprs.2017.08.012>.
- Roderick, Michael L., 1999. Estimating the diffuse component from daily and monthly measurements of global radiation. *Agric. Forest Meteorol.* 95 (3), 169–185. [https://doi.org/10.1016/S0168-1923\(99\)00028-3](https://doi.org/10.1016/S0168-1923(99)00028-3).
- Rounsevell, M.D.A., Ewert, F., Reginster, I., Leemans, R., Carter, T.R., 2005. Future scenarios of european agricultural land use: II. Projecting changes in cropland and grassland. *Agric. Ecosyst. Environ.* 107 (2), 117–135. <https://doi.org/10.1016/j.agee.2004.12.002>.
- Searchinger, T., Heimlich, R., Houghton, R.A., Dong, F., Elobeid, A., Fabiosa, J., Tokgoz, S., Hayes, D., Yu, T.-H., 2008. Use of U.S. Croplands for biofuels increases greenhouse gases through emissions from land-use change. *Science* 319 (5867), 1238–1240. <https://doi.org/10.1126/science.1151861>.
- Shao, Yang, Lunetta, Ross S., Wheeler, Brandon, Iames, John S., Campbell, James B., 2016. An evaluation of time-series smoothing algorithms for land-cover classifications using MODIS-NDVI Multi-temporal data. *Rem. Sen. Environ.* 174 (March), 258–265. <https://doi.org/10.1016/j.rse.2015.12.023>.
- Teluguntla, P., Thenkabail, P.S., Oliphant, A., Xiong, J., Gumma, M.K., Congalton, R.G., Yadav, K., Huete, A., 2018. A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. *ISPRS J. Photogramm. Rem. Sen.* 144 (October), 325–340. <https://doi.org/10.1016/j.isprsjprs.2018.07.017>.
- Tumer, Kagan, Ghosh, Joydeep, 1996. Error Correlation and Error Reduction in Ensemble Classifiers. *Connection Science* 8 (3–4), 385–404. <https://doi.org/10.1080/095400996116839>.
- Unkovich, Murray, Baldock, Jeff, Forbes, Matthew, 2010. Chapter 5 - variability in harvest index of grain crops and potential significance for carbon accounting: examples from Australian agriculture. In: *Advances in Agronomy. Advances in Agronomy*. Academic Press, pp. 173–219. [https://doi.org/10.1016/S0065-2113\(10\)05005-4](https://doi.org/10.1016/S0065-2113(10)05005-4).
- Uno, Y., Prasher, S.O., Lacroix, R., Goel, P.K., Karimi, Y., Viau, A., Patel, R.M., 2005. Artificial neural networks to predict corn yield from Compact Airborne Spectrographic Imager data. *Comput. Electron. Agric.* 47 (2), 149–161.
- Vapnik, Vladimir, Golowich, Steven E., Smola, Alex J., 1997. Support vector method for function. *Approx. Regres. Estim. Signal Proc.* 7.
- Verrelst, Jochem, Camps-Valls, Gustau, Muñoz-Marí, Jordi, Rivera, Juan Pablo, Veroustraete, Frank, Clevers, Jan G.P.W., Moreno, José, 2015. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – A review. *ISPRS J. Photogramm. Rem. Sen.* 108 (October), 273–290. <https://doi.org/10.1016/j.isprsjprs.2015.05.005>.
- Verrelst, J., Rivera, J.P., Moreno, J., Camps-Valls, G., 2013. Gaussian processes uncertainty estimates in experimental Sentinel-2 LAI and Leaf chlorophyll content retrieval. *ISPRS J. Photogramm. Rem. Sen.* 86 (December), 157–167. <https://doi.org/10.1016/j.isprsjprs.2013.09.012>.
- Waldner, François, Duveiller, Gregory, Defourny, Pierre, 2018. Local adjustments of image spatial resolution to optimize large-area mapping in the era of big data. *Int. J. of Appl. Earth Observ. Geoinf.* 73 (December), 374–385. <https://doi.org/10.1016/j.jag.2018.07.009>.
- Waldner, François, Horan, Heidi, Chen, Yang, Hochman, Zvi, 2019. High temporal resolution of leaf area data improves empirical estimation of grain yield. *Sci. Rep.* 9 (1), 1–14.
- Waldner, François, Lambert, Marie-Julie, Li, Wenjuan, Weiss, Marie, Demarez, Valérie, Morin, David, Marais-Sicre, Claire, et al., 2015. Land cover and crop type classification along the season based on biophysical variables retrieved from multi-sensor high-resolution time series. *Remote Sensing* 7 (8), 10400–10424. <https://doi.org/10.3390/rs70810400>.
- Whelan, B.M., McBratney, A.B., Minasny, B., 2002. "Vesper 1.5 – spatial prediction software for precision agriculture," 14.
- Wilcoxon, Frank, 1945. Individual comparisons by ranking methods. *Biometrics Bulletin* 1 (6), 80. <https://doi.org/10.2307/3001968>.
- Williams, C.K.I., 1998. Prediction with Gaussian Processes: From Linear Regression to Linear Prediction and Beyond. In: *Learning in Graphical Models*, pp. 599–621. NATO ASI Series. Springer, Dordrecht. https://doi.org/10.1007/978-94-011-5014-9_23.
- Wilson, Barry T., Knight, Joseph F., McRoberts, Ronald E., 2018. Harmonic regression of landsat time series for modeling attributes from national forest inventory data. *ISPRS J. Photogramm. Rem. Sen.* 137 (March), 29–46. <https://doi.org/10.1016/j.isprsjprs.2018.01.006>.
- Wolanin, Aleksandra, Camps-Valls, Gustau, Gómez-Chova, Luis, Mateo-García, Gonzalo, van der Tol, Christiaan, Zhang, Yongguang, Gunter, Luis, 2019. Estimating crop primary productivity with sentinel-2 and landsat 8 using machine learning methods trained with radiative transfer simulations. *Rem. Sen. Environ.* 225 (May), 441–457. <https://doi.org/10.1016/j.rse.2019.03.002>.
- Yuan, Wenping, Chen, Yang, Xia, Jiangzhou, Dong, Wenjie, Magliulo, Vincenzo, Moors, Eddy, Olesen, Jørgen Eivind, Zhang, Haicheng, 2016. Estimating crop yield using a satellite-based light use efficiency model. *Ecological Indicators* 60 (January), 702–709. <https://doi.org/10.1016/j.ecolind.2015.08.013>.
- Zhao, Jianyun, Zhang, Xiaohua, 2018. An Adaptive noise reduction method for NDVI time series data based on S-G filtering and wavelet analysis. *J. Indian Soc. Rem. Sen.* 46 (12), 1975–1982. <https://doi.org/10.1007/s12524-018-0855-2>.