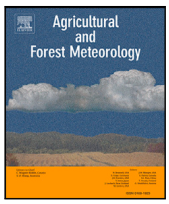
Agricultural and Forest Meteorology 353 (2024) 110055

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

journal homepage: www.elsevier.com/locate/agrformet



# Analysis and forecasting of Australian rice yield using phenology-based aggregation of satellite and weather data

James Brinkhoff [a,*], Allister Clarke [b,c], Brian W. Dunn [d], Mark Groat [b]

[a] Applied Agricultural Remote Sensing Centre, University of New England, Armidale 2351 NSW, Australia
[b] Ricegrowers Ltd. (SunRice), Leeton 2705 NSW, Australia
[c] Charles Sturt University, Wagga Wagga 2640 NSW, Australia
[d] New South Wales Department of Primary Industries, Yanco 2703 NSW, Australia

ARTICLE INFO

*Keywords:*
Rice
Phenology
Yield prediction
Shapley values
Machine learning
Time series analysis

ABSTRACT

Rice yield depends on factors including variety, weather, field management, nutrient and water availability. We analyzed important drivers of yield variability at the field scale, and developed yield forecast models for crops in the temperate irrigated rice growing region of Australia. We fused a time-series of Sentinel-1 and Sentinel-2 satellite remote sensing imagery, spatial weather data and field management information. Rice phenology was predicted using previously reported models. Higher yields were associated with early flowering, higher chlorophyll indices and higher temperatures around flowering. Successive rice cropping in the same field was associated with lower yield ($p<0.001$). After running a series of leave-one-year-out cross validation experiments, final models were trained using 2018–2022 data, and were applied to predicting the yield of 1580 fields (43,700 hectares) from an independent season with challenging conditions (2023). Models which aggregated remote sensing and weather time-series data to phenological periods provided more accurate predictions than models that aggregated these predictors to calendar periods. The accuracy of forecast models improved as the growing season progressed, reaching RMSE=1.6 t/ha and Lin's concordance correlation coefficient (LCCC) of 0.67 30 days after flowering at the field level. Explainability was provided using the SHAP method, revealing the likely drivers of yield variability overall, and of individual fields.

## 1. Introduction

Early crop yield forecasts are important for multiple levels of decision-making. For agricultural industries and processors, yield forecasts enable decision-making regarding grain storage, transport, and marketing (Basso and Liu, 2019). At governmental and intergovernmental level, they inform decision-making on food security (Meroni et al., 2021). For individual growers, yield forecasts can help with financial planning, organizing on-farm storage and transport of produce. They can also provide growers with understanding of the agrometeorological and management factors that impact yield (Delerce et al., 2016; Paudel et al., 2023), and thus aid decision making regarding management factors such as sowing date and method, nutrient and water management, potentially leading to improved productivity and profitability.

Weather has a significant impact on the yield of agricultural crops with the magnitude and direction of the effect varying relative to growth stage (Tappi et al., 2023). For rice, the effect of weather varies by variety (Sivapalan et al., 2007; Delerce et al., 2016), known as genotype by environment (G×E) interaction. Data mining approaches have been used to show how rice yield is influenced by average temperatures, solar radiation and rainfall (Delerce et al., 2016). In temperate regions, cold-temperatures during reproductive growth stages can induce sterility and thus greatly degrade yield (Godwin et al., 1994; Williams and Angus, 1994). There is some evidence that cold during vegetative stages can exacerbate this sensitivity (Shimono et al., 2007). The impact of these cold stress events on yield in temperate regions is a major cause of large yield gaps, and modeling them is not straightforward (Espe et al., 2016; Li et al., 2015). Heat stress can also impact yield, particularly when experienced during panicle initiation (PI) to flowering (Espe et al., 2017; Ali et al., 2019). The predictions from process-based rice crop models (such as APSIM-ORYZA, CERES-Rice etc.) have large uncertainties and often deviate from actual yields in the presence of such high and low temperature conditions (Li et al., 2015).

Rice yield is also influenced by biophysical and crop management factors including nitrogen status (Dunn et al., 2016), leaf area index (Hashimoto et al., 2022), water application timing (Dunn and Gaydon, 2011) and water stress (Bouman and Tuong, 2001). Many

* Corresponding author.
*E-mail address:* james.brinkhoff@une.edu.au (J. Brinkhoff).

https://doi.org/10.1016/j.agrformet.2024.110055
Received 11 October 2023; Received in revised form 2 May 2024; Accepted 8 May 2024
Available online 18 May 2024

of these factors can be detected using remote sensing (Myneni and Williams, 1994; Brinkhoff et al., 2021, 2022). Biomass depends on the interception of solar radiation by photosynthesizing leaf area, and yield is determined by the total plant biomass multiplied with the harvest index, which is typically around 50% for rice (Hashimoto et al., 2022).

The spectral reflectance of rice plants is predictive of yield. This has been demonstrated in numerous studies using ground-based sensors (Varinderpal-Singh et al., 2021; Chang et al., 2005). However, the relationships are variable with growth stage and cultivar (Varinderpal-Singh et al., 2021). Many studies have found that spectral reflectance measurements taken around the PI or booting growth stages are most related with yield (Varinderpal-Singh et al., 2021; Eugenio et al., 2023). The commonly used normalized difference vegetation index (NDVI) depends on near infrared and red reflectances. However, many works have found that indices involving green (Soriano-González et al., 2022) or red edge (Zhou et al., 2017) bands are more predictive of rice yield.

In regions where cloud frequently obstructs optical remote sensing, synthetic aperture radar (SAR) has been adopted (dela Torre et al., 2021). For example, Setiyono et al. (2019) used SAR data from the Sentinel-1 constellation to find start-of-season and leaf area index, which were then coupled with the ORYZA crop growth model to predict yield. Yu et al. (2023) found that integrating SAR and multispectral data resulted in improved rice yield predictions when the number of cloud-free multispectral observations were low.

Process-based crop models can assimilate environmental and biophysical information as well as information such as LAI derived from remote sensing to predict yield (dela Torre et al., 2021). Often, however, empirical models using statistical and machine learning techniques, can provide higher accuracies (Wang et al., 2023a), providing sufficient previous yield data is available for model training. The latter rely on algorithms to learn a predictive relationship between observed variables (remote sensing, weather, field information) and the target variable (yield) and as such, intermediate biophysical variables (LAI, fAPAR etc.) are not always required. Interpreting such models to derive useful information on the drivers of yield is challenging. However techniques such as Shapley values that summarize prediction dependencies on input variables have provided a level of explainability (Huber et al., 2022; Shendryk et al., 2021)

Yield prediction models that exploit multiple spectral and temporal features can provide more accurate predictions than single linear regression models (Kang et al., 2020; Brinkhoff and Robson, 2021; Cao et al., 2021; dela Torre et al., 2021). Finding statistics such as maximum, mean and minimum of a time-series of vegetation indices may provide good features for yield estimation (Rahman and Robson, 2020; Nguyen et al., 2022; Von Bloh et al., 2023). Using features derived from images taken at different times directly in ML models may improve accuracy relative to data taken from a single time points (Zhou et al., 2017). More advanced techniques to generate features from remote-sensing time-series include harmonic regression (Deines et al., 2020), wavelet transforms (Gu et al., 2022), or training time-series specific neural networks (such as 1DCNN and LSTM) to discover time series relationships (Paudel et al., 2023). However, some studies have found that standard ML algorithms with flattened time-series features provide as good predictions as time-series specific algorithms in general regression problems (Guijo-Rubio et al., 2023) and in yield prediction problems specifically (Kang et al., 2020; Von Bloh et al., 2023; Sabo et al., 2023). Yield prediction accuracy can be improved when remote-sensing indices are resampled according to thermal time since sowing or phenology stages (Desloires et al., 2023; Bolton and Friedl, 2013; Ji et al., 2022), particularly when there is variability in the timing of phenological events between crops, seasons and regions.

There were two overarching aims of this work. Firstly, to analyze drivers of rice yield in a temperate growing region. Secondly, to produce optimized rice yield forecasting models using time-series features. In particular, we aimed to investigate the following:

- What is the relationship between yield and factors such as variety, management, weather and remote sensing information?
- Can composing time-series variables (weather and remote sensing) according to per-field growth stages provide suitable features for yield forecast models, and do time-series models incorporating data aggregated from multiple stages across a growing season provide better predictions than those using data from a single period?
- What are the achievable yield prediction accuracies using only variables from either field management information, time-series weather data or remote sensing data, and how does accuracy improve when these variable sets are combined using linear or nonlinear machine learning algorithms?
- Can data analysis using the developed feature engineering methods and model explainability techniques be used to provide information on the drivers of crop yields, that could improve field management decisions towards greater productivity?

## 2. Methods

We collated a dataset of field-level rice yield and predictive variables. This included information on rice sowing method and date and variety. We then derived phenological and time-series information from remote sensing and weather datasets. The process is shown in Fig. 1 and is described in the following sections.

### 2.1. Field and yield data

The study area is the Murray and Murrumbidgee Valleys in New South Wales, where the majority of Australian rice is grown. The climate is temperate (Jena, 2012), which allows a single crop to be grown each year, with reproductive growth stages occurring during summer (December–February). Crops are typically sown in October and harvested in April. Throughout this paper, 'year' refers to the year of harvest rather than the year of sowing.

We obtained records of all rice crops from 2018–2023. The average field size was 34 ha. Data included geo-referenced field geometries, rice variety, sowing date and sowing method. The main sowing methods adopted in the area are drill, aerial and dry broadcast (Ward et al., 2021). Aerial and dry broadcast are quite similar, as the field is ponded (flooded) for the majority of the season. In contrast, drill sowing starts the season with intermittent irrigation, then ponding usually around the three-leaf stage. The nine varieties included in the study are the dominant ones grown in the area (Dunn and Dunn, 2023), including long grain (Doongara, Langi, Topaz) medium grain (Reiziq, Sherpa, V071), short grain (Koshihikari, Opus) and short-season medium grain (Viand) varieties.

We collected predictors (e.g. remote sensing VIs, variety, sowing dates etc.) on a per-*field* basis. However, the processor's yield data were recorded per-*crop* (unique farm × variety × year). Therefore, some yield records were aggregated from multiple fields on the same farm, which may have had diverse management (sowing date and method, fertilizer application). Since it was not clear how field-level predictors should be aggregated to match crop-level yield records, for model training, we used only *crop* records that came from a single *field*. On the other hand, during inference and model testing, we were able to predict yield for all fields, since predictors were available for all fields. These field-level predictions were then aggregated to the crop level for comparison with the processor yield data.

Data were screened to remove records with unrealistic sowing date (<September 1 or >December 31), field area errors (<0.5 ha), yield errors (<0.1 or >18 t/ha), or low normalized difference vegetation index at peak of season (<0.6, usually a result of a crop being associated with the wrong field). After this process, there were n = 1324 one-field crop records from 2018–2023. The number of datapoints per year, variety, sowing method and region is provided in Section 3.1.2.

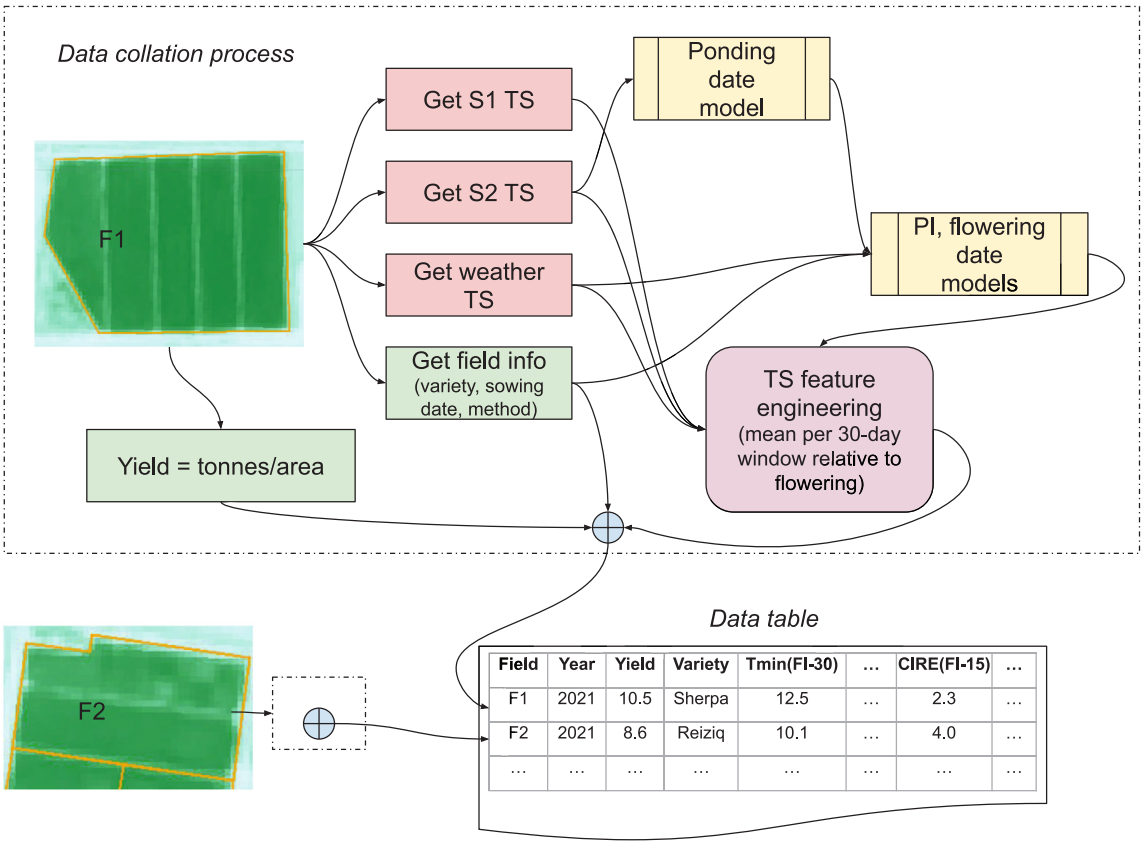



**Fig. 1.** Data collation process. Data includes time-invariant variables such as variety and sowing date, and time-series (TS) variables such as weather and remote sensing information from Sentinel-1 (S1) and Sentinel-2 (S2) satellites. Intermediate predictive models were used to estimate ponding (or flooding) date and the panicle initiation (PI) and flowering growth stages. These were then used to engineer features from the time-series data, to train and test yield prediction models.

### 2.2. Predictor variables

#### 2.2.1. Time-invariant predictor variables

The yield predictors considered are listed in Table 1. They included time-invariant variables that do not change during a season, such as rice variety, sowing method, and the dates of sowing, ponding, PI and flowering. The ponding (or flooding) date is the date when the water level is raised above the soil for the majority of the season, and was predicted using a logistic regression model based on a time-series of Sentinel-2 reflectances, as described in Brinkhoff et al. (2022). The PI and flowering dates were predicted using additional logistic regression models based on variety, sowing method, sowing date, predicted ponding date and accumulated temperatures, as described in Brinkhoff et al. (2023b).

#### 2.2.2. Time-series predictor variables

Remote sensing data from the Sentinel-1 (S1) and Sentinel-2 (S2) constellations of the European Space Agency were accessed and processed in Google Earth Engine (GEE) (Gorelick et al., 2017). The S1 synthetic aperture radar (SAR) backscatter data included vertical transmit polarization, and both vertical and horizontal receive polarization (VV and VH channels), in decibels (dB).

The S2 optical top-of-atmosphere (TOA) remote sensing data included the ten 10 m and 20 m resolution reflectance bands (Table 1). Important vegetation indices (VIs) were derived from the reflectances (Zeng et al., 2022). Though S2 surface reflectance (SR) data is also available in GEE, that dataset only starts from December 2018, so would not allow use of 2018 and 2019 yield data to train models. Further, TOA data has been found in other work to provide models with similar accuracy to those based on SR data (Brinkhoff et al., 2022; Wolters et al., 2021), and we similarly found rice yield model results (based on 2020–2023 yield records) using SR data were no better than those using TOA data. Clouds were masked in the S2 data using the s2cloudless product (Skakun et al., 2022) with threshold 40%. Cloud mask speckle was first removed using a focal minimum filter (radius 30 m), then the cloud masks were buffered by 500 m to avoid contamination of data from cloud edges and shadows (Brinkhoff et al., 2022). To ensure high-quality observations were used, only imagery with more than 99% unmasked pixels within each field was used.

For each field and image date, the spatial mean of all S1 and S2 pixels were calculated. This resulted in time-series tabular data with columns being the image bands listed in Table 1. These time-series were linearly interpolated to a daily time step. A Savitzky–Golay filter was then applied to smooth each time series.

The weather data was gathered at a daily time step for each farm from the spatially interpolated SILO dataset (Jeffrey et al., 2001). To attempt to model cold-temperature induced sterility, a feature was engineered from the daily minimum temperature, quantifying how far below a critical temperature the daily minimum temperature was, similar to Shimono et al. (2005). This was defined as:

$$Tmin < Tcrit = min(Tmin - Tcrit, 0) \quad (1)$$

After testing $Tcrit$ = 12,15,18,21, it was found that $Tcrit$ = 15 offered the greatest improvement in model prediction accuracy, so only that critical temperature was retained, with the variable named Tmin < 15 in the following. We note that this threshold was chosen empirically based on improvement of model predictions, while other works (Godwin et al., 1994; Espe et al., 2016) have discussed critical night temperatures from 12–20 °C depending on variety and region. While we did analyze the correlation between solar radiation and yield, we found it added no accuracy to ML yield model predictions, so it was omitted from the models.

**Table 1**
Time-invariant and time-series predictor variables. Time-series variables were aggregated both to calendar months, and to 30-day windows relative to the predicted flowering date.

| Type | Feature set name | Name | Details |
|---|---|---|---|
| Time-invariant | Variety | Variety | Categorical [Reiziq, V071, Langi, Sherpa...] |
| | Dates+Method (D+M) | Sowing date | Supplied by growers |
| | | Ponding date | Predicted (Brinkhoff et al., 2022) |
| | | PI date | Predicted (Brinkhoff et al., 2023b) |
| | | Flowering date | Predicted (Brinkhoff et al., 2023b) |
| | | Sowing method | Categorical [Drill, Aerial/Dry broadcast] |
| Time series | Reflectances (Refs) | B | Blue (490 nm, 10 m) |
| | | G | Green (560 nm, 10 m) |
| | | R | Red (665 nm, 10 m) |
| | | RE1 | Red edge (705 nm, 20 m) |
| | | RE2 | Red edge (740 nm, 20 m) |
| | | RE3 | Red edge (780 nm, 20 m) |
| | | NIR1 | Near infrared (835 nm, 10 m) |
| | | NIR2 | Near infrared (865 nm, 20 m) |
| | | SWIR1 | Shortwave infrared (1610 nm, 20 m) |
| | | SWIR2 | Shortwave infrared (2200 nm, 20 m) |
| | Vegetation indices (VIs) | GRVI | $(G-R)/(G+R)$ |
| | | GNDVI | $(NIR-G)/(NIR+G)$ |
| | | NDVI | $(NIR-R)/(NIR+R)$ |
| | | NDRE | $(NIR-RE1)/(NIR+RE1)$ |
| | | CIG | $NIR/G-1$ |
| | | CIRE | $NIR/RE1-1$ |
| | | LSWI | $(NIR-SWIR1)/(NIR+SWIR1)$ |
| | | MNDWI | $(G-SWIR1)/(G+SWIR1)$ |
| | SAR (S1) | VV | Vertical-vertical backscatter (dB) |
| | | VH | Vertical-horizontal backscatter (dB) |
| | Weather (Temps+Srad) | Tmin | Minimum daily temperature (°C) |
| | | Tmin < 15 | min(Tmin-15,0) (°C) |
| | | Tmax | Maximum daily temperature (°C) |
| | | Srad | Solar radiation (MJ/m$^2$) |

#### 2.2.3. Time-series feature engineering

Temporal aggregations of the time-series variables was performed firstly on a monthly basis. All observations within a calendar month were averaged. Secondly, they were also aggregated on a phenological basis. An example of the feature engineering process is illustrated in Fig. 2, showing how the time-series variables were aggregated to 30-day windows relative to each field's predicted flowering date. There were 15-day overlaps between the windows. For example, a window centered on 30 days before flowering was denoted Fl-30 (where the '–' indicates before flowering, while '+' is used for windows after flowering, Fig. 2), and Fl-30 window averages data from 45 days before flowering to 15 days before flowering. These aggregated time-series features were used in correlation analyses and as inputs for the yield prediction models. The nominal models used all windows from Fl-90 to Fl+30, though we also tested inclusion of later windows in Section 3.2.3.

#### 2.2.4. Field cropping history

In order to investigate the influence of field cropping history on rice yield, we found which fields from each year intersected fields from the previous one and two years. A variable (rice-rice) was then added to each field's record, to indicate if the rice crop was following a rice crop planted in the same field from the previous year. Another variable (rice-rice-rice) was added to indicate if the field had been planted with rice in both of the previous years. As there were very few crops planted in 2019 and 2020 due to lack of availability of water during drought, we performed this analysis for 2022 and 2023 crops.

### 2.3. Model algorithms and explanations

We compared two contrasting ML algorithms to build yield forecast models using scikit-learn (Pedregosa et al., 2011). Firstly, the Ridge regression algorithm, which applies L2 regularization to the multiple linear regression problem, so that the coefficients of less important variables are shrunk towards zero (Hastie et al., 2009). Secondly, the LightGBM (LGB) algorithm, which is a very efficient gradient boosting decision tree framework (Ke et al., 2017). It is able to quickly solve problems with a large number of features and instances, while achieving high accuracy. Gradient boosting decision tree models such as this are still often outperform other algorithms such as neural networks for tabular data problems (Grinsztajn et al., 2022). Ridge is a linear algorithm, while LGB allows fitting of nonlinear relationships.

For the Ridge algorithm, all input features were normalized based on the distribution of the training data, using a standard scaler. This step was not necessary for LGB, as it is a tree-based method. Categorical features (variety and sowing method) were one-hot encoded for Ridge, which again was not necessary for LGB as it can deal with categorical features directly.

We explored the impact of different features (field information, weather, remote sensing) on yield using the TreeExplainer method of the SHAP library, which is based on Shapley values (Lundberg et al., 2020). This was used firstly to provide global model explanations, showing the most important variables in predicting yield, and whether they had a positive or negative impact. Secondly, yield predictions for individual fields were explained, demonstrating the likely causes of low and high yields in specific cases.

### 2.4. Model selection

Models were assessed using standard error metrics. Bias is the average error between actual and predicted yields:

$$Bias = \frac{1}{N}\sum_{n=1}^{N}(Y_n - \hat{Y}_n), \tag{2}$$

where $Y_n$ are the actual yields and $\hat{Y}_n$ are the predicted yields.

The root mean squared error (RMSE) is:

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(Y_n - \hat{Y}_n)^2}. \tag{3}$$

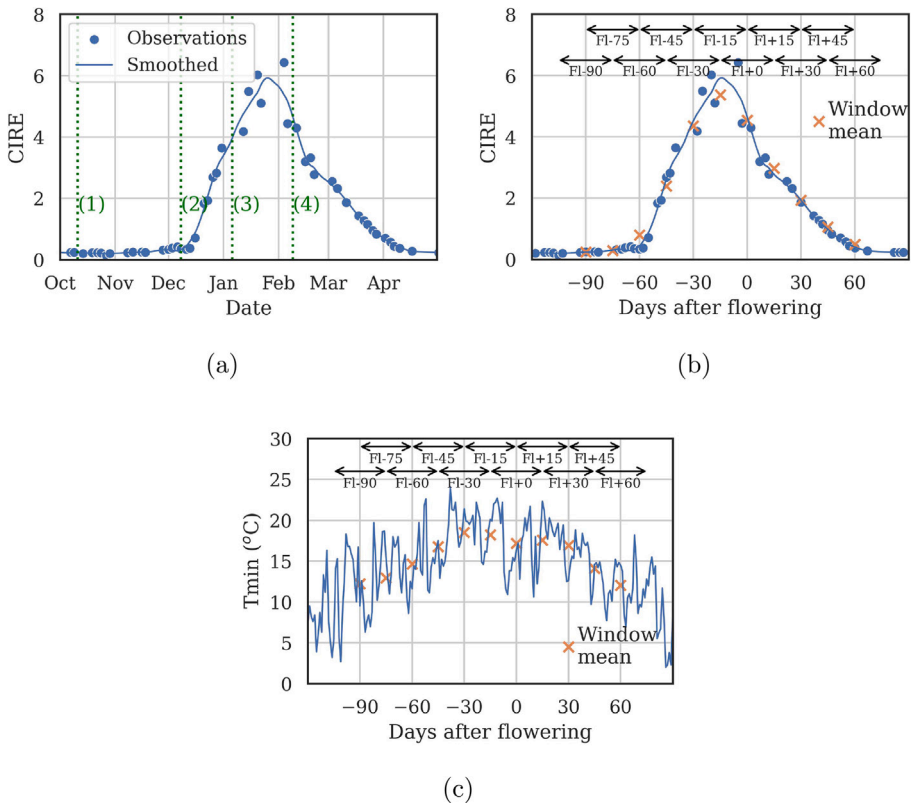



**Fig. 2.** Example of deriving time-series predictors for a drill-sown Reiziq crop from the 2022 season, illustrating the feature engineering process. The original chlorophyll index red edge (CIRE) time series is shown in (a), with vertical lines indicating sowing (1), predicted ponding (2), PI (3) and flowering (4) dates. Then (b–c) plot CIRE and Tmin relative to the flowering date, and the mean of the observations within the 30-day windows.

The relative RMSE (RRMSE) is the RMSE divided by the average actual yield, in %:

$$RRMSE(\%) = RMSE/\overline{Y} \times 100 \quad (4)$$

where $\overline{Y}$ is the mean of the actual yields.

Lin's concordance correlation coefficient (LCCC, Lin (1989)), ranges from 0–1. It measures both how much of the variability in actual yields are predicted, and how closely the predicted yields are to the actual yields:

$$LCCC = \frac{2s_{\hat{Y}Y}}{s^2_{\hat{Y}} + s^2_{Y} + (\overline{\hat{Y}} - \overline{Y})^2} \quad (5)$$

where $s_{\hat{Y}Y}$ is the covariance between predicted and measured yields, $s^2_{\hat{Y}}$ and $s^2_{Y}$ are the variances of the predicted and measured yields respectively, and $\overline{\hat{Y}}$ is the mean of the predicted yields.

We held out the 2023 data for final model testing. To rigorously test various combinations of algorithms and predictor variables, we used the 2018–2022 data from one-field crops (n = 1072, see Fig. 4). To ensure models were not overfit to in-season data, and to obtain realistic expectation of model accuracy, we adopted a leave-one-year-out (LOYO) cross validation (CV) procedure, instead of using standard random training-validation splits such as in K-fold CV (Brinkhoff et al., 2019; Meroni et al., 2021). K-fold CV often produces lower errors than LOYO CV (Yu et al., 2023), and these lower errors do not represent the realistic operational scenario of forecasting yield in an independent season where no training data is yet available.

Each model was trained on four "training" years of data and then tested on the remaining "test" year, and this process was repeated five times, once for each year in the dataset. For example, the first experiment was trained using [2018,2019,2020,2021] data, then tested on 2022; the second was trained using [2018,2019,2020,2022] and tested on 2021; and so on.

We tested all 64 combinations of the feature sets, [Variety, Dates+Method, Refs, VIs, S1, Temps] (Table 1). We also compared against a baseline null model with no predictors, which generated all predictions as simply the average of all yield records in the training data. Including such baseline models enables testing if more complex ML models are providing any real benefit in prediction accuracy. Other simple models included those that used only variety, or only sowing date and sowing method information. The total number of experiments was 2 algorithms (Ridge and LGB) × 64 feature sets × 5 LOYO CV combinations = 640 experiments. The performance of the feature sets and algorithms was ranked using the mean and maximum RMSE from the 5 LOYO CV experiments of each feature combination. These experiments were performed for both the monthly and phenology-based time-series feature aggregations.

### 2.5. Model validation

Three feature sets were chosen based on the results of the above experiments (Variety+Temps, Variety+Temps+CIRE, Variety+Temps+Refs+VIs, see Table 1). We then trained models using these feature sets and the 2018–2022 data.

These models were used to predict rice yields from the independent 2023 season for *all* fields (n = 1580, not just the 252 one-field crops, see Section 2.1). Where multiple fields came from the same farm × variety, the per-field yield predictions were combined to match the crop-level data from the processor. The 1580 fields were thus reduced to 762 crops. Finally, these crop-level predictions were compared with the processor's data.

To evaluate how prediction accuracy improved as the season progressed, we also trained and tested models with the time-series data truncated at various points through the season.

We also tested the model's ability to predict inter-annual yield variability, using independent 2006–2017 yield data for one variety (Reiziq). As the data did not include precise field locations, and Sentinel-2 remote sensing data was not available for most of the years, we used a model that only included the Variety+Temps feature sets. We estimated the ponding date to be coincident with the sowing data for aerial and dry broadcast sowing methods, and to be 44 days later than the sowing date for the drill sowing method (which is the average delay from sowing to ponding from the 2018–2023 data). These sowing and ponding dates were then used, together with the weather data, to predict the flowering date per field following (Brinkhoff et al., 2023b). The weather data was then aggregated to windows relative to flowering date as described above. Finally, the variety and windowed weather data was used to predict the yield per field, which was compared with the actual yields. 2007 data was excluded, as there was no sowing date or sowing method information, and 2008–2009 data was excluded as there were a very small number of fields due to drought.

## 3. Results

We first report results of analysis on factors affecting yield (Section 3.1), including variety, sowing, phenology, and relationship with weather and remote sensing variables. Secondly, we develop machine learning models to provide in-season yield forecasts (Section 3.2), including comparison of feature sets and algorithms, validation of per-field predictions in an independent year, prediction explainability, and testing model ability to predict inter-annual yield variation using a historical dataset.

### 3.1. Yield factor analysis

#### 3.1.1. Season characteristics

The fields in the study experienced considerable variation in weather conditions and exhibited diverse characteristics in remote sensing indices. Fig. 3 shows the average Chlorophyll Index Red Edge (CIRE, Gitelson et al. (2005)) and temperatures per year and aggregation period. CIRE and Tmin were some of the most important indices for predicting yield, as will be shown in following sections.

When CIRE was aggregated to monthly periods (Fig. 3(a)), there was significant spread in the timing of green-up, peak and senescence from year-to-year. In particular, 2021 and 2023 exhibited delayed progression, attributed to a combination of colder conditions, and later sowing in 2023 due to early-season flooding. In contrast, when CIRE was aggregated to windows relative to the per-field flowering dates (Fig. 3(b)), the inter-annual spread was much reduced. There were still differences in the length of season (e.g. 2021 was longer due to slower maturing). Additionally, the peak CIRE varied, with 2020 and 2023 seeing lower values and 2021 higher. In general, CIRE peaked at, or just before, flowering.

There were differences in the temperatures from year-to-year (see Figs. 3(c)–3(d)). 2019 was a drought year and had significantly higher temperatures. 2021 experienced warm temperatures at the start of the season, but became cooler starting around Fl-60. In 2023, temperatures remained consistently cooler throughout the season.

#### 3.1.2. Categorical factors (season, sowing method, variety, region)

Yield differed between categorical factors (Fig. 4). First, yields were different between seasons (Fig. 4(a)). The lowest average yields were observed in 2021, and this year also had the largest yield variability with a negative skew. Many crops were affected by cold-temperature induced sterility in this year. The highest yields were observed in 2022, and there was also relatively less variability in yield. There were very few crops planted in 2019 and 2020 due to lack of water availability during a severe drought.

There was variability in the yields between varieties (Fig. 4(b)). The recently developed V071 variety produced the highest yields. It has similar grain characteristics to the previously dominant Reiziq variety, but V071 has higher yields and is less susceptible to cold-induced sterility. This is evident in the reduced negative skew of V071 yields compared with Reiziq (Fig. 4(b)). The high-value Koshihikari variety had lower yields, as did the short-season Viand variety.

The median yield of drill sown crops was 0.9 t/ha lower than dry broadcast and aerial (DB+Aerial) crops, which have longer ponding duration (Fig. 4(c)). However, following sections will show that this is likely not due to the sowing methods themselves, but due to drill sowing being associated with later PI and flowering dates.

Yields varied between growing regions (Fig. 4(d)). The Murrumbidgee Irrigation Area (MIA) had the highest yields, and the lowest came from the Coleambally Irrigation Area (CIA). The MIA is further north and warmer than the Eastern and Western Murray Valley (EMV and WMV) regions. The lower yields in the CIA could be at least partly explained by the predominance of the lower yielding Viand variety (24% of fields) and widespread practice of growing rice-after-rice on the same fields (40% of fields in 2022 and 2023).

Fields that had rice grown in the preceding year ($n = 161$) had 1.1 t/ha lower yield than those that did not ($p < 0.001$), see Fig. S1. Those that had rice grown in both of the preceding two years (i.e. 3 consecutive years of rice, $n = 26$) had 2.1 t/ha lower yield ($p < 0.001$).

#### 3.1.3. Management and phenology dates

Next, the relationships between yield and continuous variables were explored. The distribution of sowing, ponding, PI and flowering dates per year are shown in Fig. 5. Ponding date was predicted using remote sensing methods described in Brinkhoff et al. (2022) and PI and flowering using (Brinkhoff et al., 2023b). The sowing dates were highly variable. They were generally later in 2023, due to intense rainfall and flooding at the start of the season, which hindered field access. Sowing dates were earlier for drill-sown fields, as recommended in the growing guides (Ward et al., 2021). The predicted ponding dates were later for drill sowing compared with aerial and dry broadcast as expected, as drill sowing involves intermittent irrigations before ponding later in the season.

The PI dates were predicted to be later for drill-sown fields than aerial and dry broadcast fields by an average of 6 days, and flowering by 9 days . The average PI and flowering dates were the 8th of January and 8th of February respectively. The phenology dates in 2023 were also much later, attributed to both later sowing, and cold temperatures as discussed above.

The trend of yield against management and phenology dates are shown in Fig. 6. The yield decreased for crops subjected to later sowing, ponding, PI and flowering dates ($p < 0.001$). The relationship between yield and sowing date, and between yield and ponding date, were different according to the sowing method employed. There was more negative impact on yield with delayed ponding dates for drill than for aerial and dry broadcast.

Conversely, the yield vs. PI and flowering date relationship among the different sowing methods was similar (Figs. 6(c)–6(d)). On average, yield declined by 0.09 t/ha per day PI and flowering were delayed. This suggests the key driver of yield decline with date is PI and flowering timing rather than sowing and ponding dates or sowing method themselves. Later PI and flowering dates expose the crop to higher risk of encountering low temperatures during the sensitive reproductive growth stages (Farrell et al., 2006; Brinkhoff et al., 2023b). Thus, the later phenology dates of drill sown crops compared with aerial and dry broadcast crops could explain the lower yields of drill sown crops noted previously (Fig. 4(c)).

#### 3.1.4. Correlation of yield with weather and remote sensing indices

We investigated the relationship of yield with the time-series variables (remote sensing and weather). The correlation coefficients between the variables aggregated to calendar months is shown in

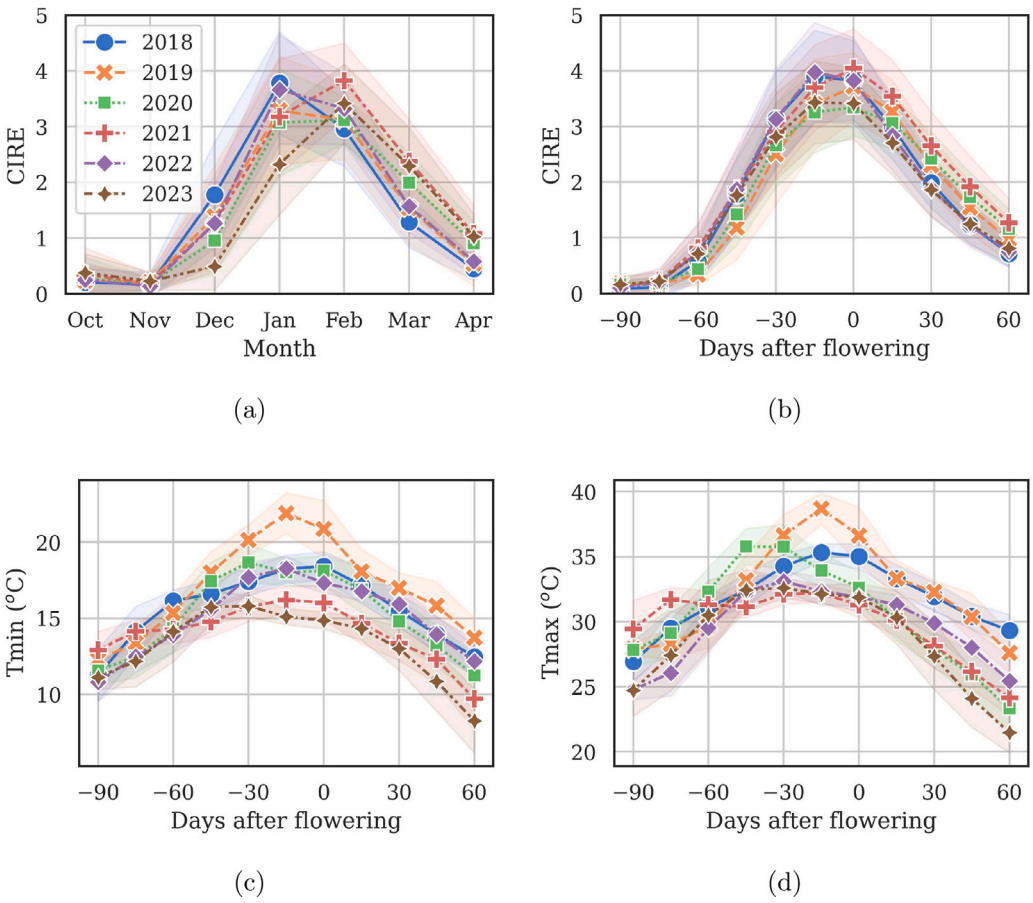



**Fig. 3.** Average chlorophyll index red edge (CIRE) and temperatures for the 1324 fields, aggregated per month (a) and phenology period (b–d). The shaded areas show the standard deviations.

Fig. 7(a), and aggregated to 30-day windows relative to each field's flowering date in Fig. 7(b).

The monthly aggregations showed higher correlations between yield and remote sensing variables, reaching r = 0.61 for LSWI and 0.58 for CIRE in January. Higher VIs in January are associated with higher yields, while higher VIs in March onwards (closer to harvest) are associated with lower yields. The correlations between monthly aggregations of weather variables and yield were much lower ($r \leq 0.3$), as were those of the SAR coefficients.

Temperatures were more related with yield when aggregated to phenology windows (Fig. 7(b)) than when aggregated to months, while the opposite was observed for remote sensing variables. Higher temperatures were associated with higher yield after Fl-30, and correlations increased after flowering ($r > 0.45$). Tmax was less correlated with yield than Tmin or the critical temperature feature Tmin < 15. Solar radiation (Srad) was less correlated with yield then temperatures. Later investigations revealed that including Srad in the yield prediction models did not improve accuracy and so this variable was not used further.

Most VIs had the highest correlations with yield 15 days before flowering (Fl-15). The most important VIs were LSWI (r = 0.45) and CIRE (r = 0.44). NDVI was less related to yield (r = 0.33). The SAR predictors VV and VH were relatively less correlated with yield ($r < 0.25$) than temperature or multispectral variables.

Examples of the relationship between yield and two of the important predictors, CIRE(Fl-15) and Tmin(Fl+15), are shown in Figs. 7(c)–7(d).

These correlations are global, including all varieties, sowing methods and sowing dates. Higher correlations are sometimes obtained when, for example, only a single variety is included. This motivates the use of ML methods that can exploit categorical variables (such as variety) and learn the possibly distinct relationship between predictors and yield per variety as well as use many of the predictors simultaneously. The correlations in Fig. 7 are thus not necessarily indicative of our ML model performance (results below), rather they indicate which predictors are most related with rice yield overall.

Two contrasting seasons were compared, by examining the yield, phenology, Tmin and CIRE of low, medium and high yielding fields (Fig. S2). Many fields in 2021 were affected by cold-induced sterility, and this may be related to the late phenology and corresponding minimum temperatures experienced by the low yielding fields. In contrast, 2022 was a typical year, and the difference in yield between the groups was less extreme, and was more related with the chlorophyll index CIRE.

### 3.2. Machine learning yield prediction models

#### 3.2.1. 2018–2022 Leave-one-year-out model selection

The results of the LOYO-CV model selection experiments using 2018–2022 data are summarized in Fig. 8 for models built both using calendar month-based and phenology-based aggregations of the time series variables. The RMSE was calculated for each test year and results were ranked by the mean RMSE (corresponding LCCC results are shown in Fig. S3). The worst RMSE over the 5 years is also shown, which generally came from 2021, the year significantly affected by cold-induced sterility.

The month-based aggregation models (Fig. 8(a)) generally produced lower errors than phenology-based models (Fig. 8(b)) over the 2018–2022 dataset. For example the lowest mean RMSE for month-based aggregation was 1.72 t/ha, compared with 1.86 t/ha for phenology-based aggregation. On the other hand, the month-based results were

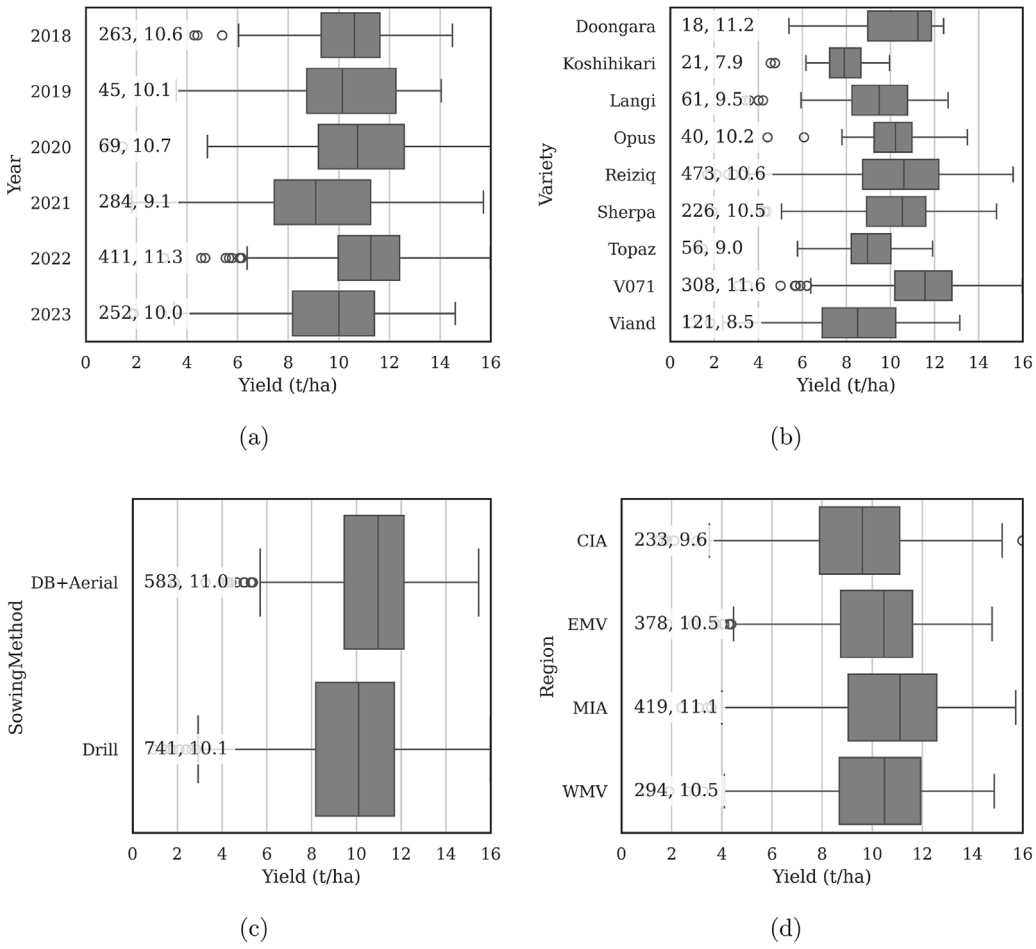



**Fig. 4.** Yield vs. categorical factors: year (a), variety (b), sowing method (c) and growing region (d). The numbers to the left of the boxes indicate the number of data points and median yield respectively.

worse than phenology-based results for the Variety+Temps feature set (compare line #59 in Fig. 8(a) to #33 in Fig. 8(b), indicating that phenology is important to make good use of the weather variables. Following results in this paper will show the phenology-based models gave better results in the challenging 2023 test season, which had significantly later PI and flowering dates than other years. Therefore, in the following, we focus on the phenology-based LOYO CV experiments (Fig. 8(b)).

Generally the Ridge models had slightly lower average RMSE than the LGB models, particularly for models including many features. However, the maximum RMSE for LGB models was often lower when less features were available, potentially indicating better generalization to challenging years.

The null model (#64, with no predictors, simply predicting the average of training set yields) had average RMSE across the five test years of 2.5 t/ha, and worst RMSE of 3.3 t/ha (in 2021). The best models gave much better performance than models using no predictors (#64). Very little improvement from the null model is obtained when only including either S1, Dates+Method or variety (#60-63). These results indicate the value of forecasting yield using our methodology over simpler approaches.

A model built around weather and variety (#33) was similar to the model using only remote sensing and variety (#34), and both are worse than models that combine weather and remote sensing (#1–10).

Variety and Temps are included in all the 10 best feature sets (#1–10), and S2 reflectances (Refs) in all of the top 5. Reflectances may be slightly better than VIs (comparing #5 and #7), but including both of them (#2) is better than including either one. The complete set of features (#9) had near to best performance (mean RMSE within 0.08 and 0.01 t/ha of the best LGB and Ridge models respectively). However, there was no real penalty in performance when the S1 and Dates+Method predictor sets were omitted (#2, Variety+Temps+Refs+VIs).

### 3.2.2. 2023 Yield prediction validation

Models were trained using 2018–2022 data (1072 single-field crops), and then used to predict 2023 yield. As analyses reported above showed, 2023 was an uncharacteristic year, with many late-sown crops, cold temperatures and a corresponding large spread in phenological dates. It thus offers a challenging scenario to assess the generalizability of the yield models. We predicted the yield for all fields (n = 1580, 43,700 ha), then aggregated predictions from multiple fields belonging to each of the 762 crops (unique farm × variety combinations).

Both LGB and Ridge models were trained for three different feature sets, chosen from the LOYO-CV experiments (Fig. 8(b)):

1. The Variety+Temps set was ranked #33, and is useful in cases where remote sensing data is not available.
2. The Variety+Temps+Refs+VIs set was ranked #2, offering near to best performance with a reduced set of predictors (omitting S1 and Dates+Management features).
3. An additional feature set was included, Variety+Temps+CIRE. This used a single VI (CIRE). It was anticipated this would provide better predictions than the Variety+Temps model, but would be more interpretable than the full Variety+Temps+Refs+VIs model, due to having less remote sensing variables. Thus, this model would be useful to provide explainability

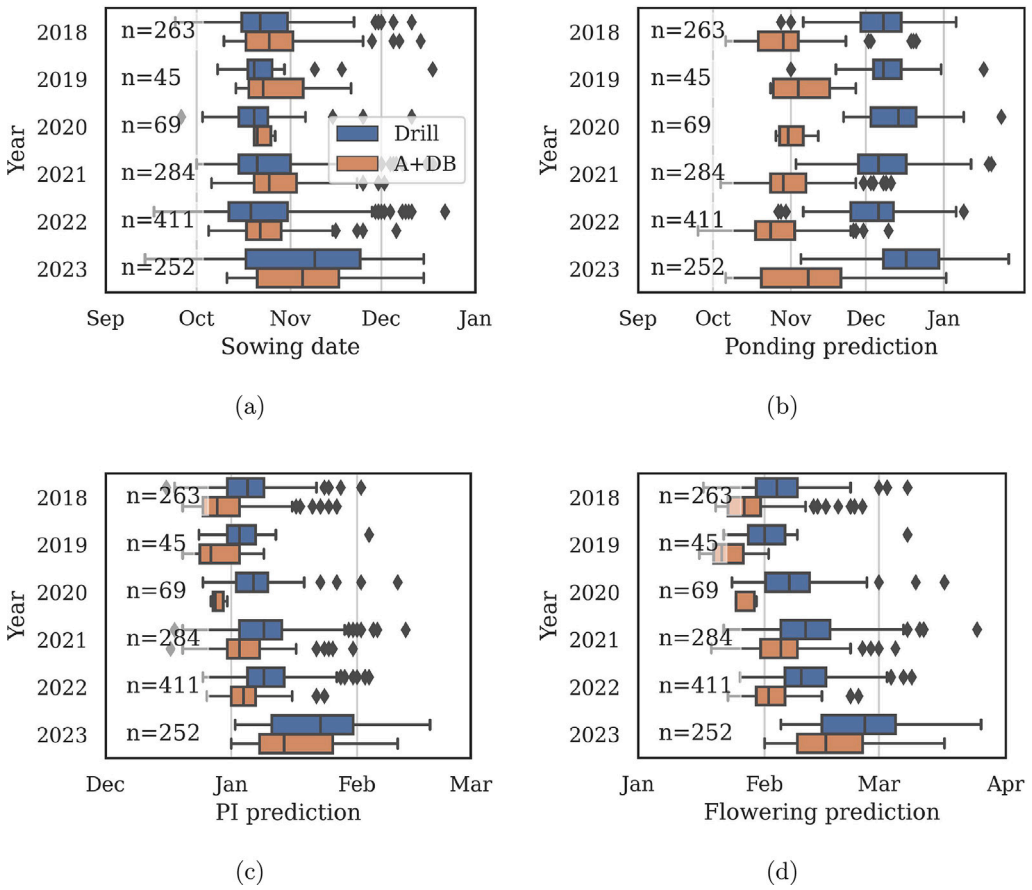



**Fig. 5.** Distribution of sowing dates (a), predicted ponding dates (b), PI dates (c) and flowering dates (d). The relationships for different sowing methods (drill, aerial and dry broadcast (A+DB)) are separated by color.

and insights on yield variability, potentially aiding future field management decisions to improve productivity.

There was not a large difference between Ridge and LGB models for each of the 3 feature sets (<0.06 t/ha difference). Model predictions are plotted against actual yields in Fig. 9. The model that only included Variety+Temps had higher errors (2 t/ha RMSE) and lower capability to predict the variability in yield between fields (LCCC = 0.44) than the models including remote sensing variables. The full model achieved 1.6 t/ha RMSE and LCCC of 0.67. The average error (bias) was less than 0.6 t/ha for the three models. The models were not able to predict the full extent of yield loss for some fields that had very low yields (<4 t/ha), which was probably due to low-temperature sterility.

We also trained a model using Variety+Temps+Refs+VI feature sets, but with calendar month-based aggregation of time-series variables instead of using phenology-based aggregation, also using the Ridge algorithm. The results (Fig. 9(d)) were worse than the model that used aggregation windows relative to phenology (RMSE 1.8 vs. 1.6 t/ha). The month-based model had higher bias, and more scatter. Interestingly, the month-based model had less tendency to under-predict high yield and over-predict low yields compared with the phenology-based model. These observations may indicate the month-based model is slightly overfitting to the training data, while the aggregation-based model is regularizing more, thereby giving less variance. Similar results were obtained using the LGB algorithm (Fig. S4). Though lower errors were obtained with the phenology-based models in 2023, these results (and the LOYO-CV results in Fig. 8) indicate the month-based models may be useful in seasons where sowing and phenology dates and weather are more typical and reflective of those in the training data.

Spatial prediction and error maps for all 2023 crops using the Ridge Variety+Temps+Refs+VIs model were produced (Fig. 10). There was a tendency to slightly under-predict yields generally, although in the Coleambally Irrigation Area, the yields were generally slightly over-predicted (by 0.7 t/ha on average).

When field-level predictions and actual yields were aggregated at region and variety levels, the errors reduced. For the Variety+Temps+Refs+VIs Ridge model, the mean yield prediction errors per region were all less than 9%, and the mean errors per variety were all less than 10%. The whole industry prediction error was 4.1% (average predicted yield for all crops was 9.3 t/ha, and the average actual yield was 9.7 t/ha).

### 3.2.3. *Prediction errors vs. forecast timing*

Fig. 11 shows how the yield forecast accuracy changed as the season progressed, using both time-series models (including all windows from Fl-90) and single-window models. The single window model predictions were worse than the time-series models, particularly after Fl-15. The RMSE for the full model was less than 1.8 t/ha by the Fl-15 window (approximately 50 days before harvest (Brinkhoff et al., 2023a)). RMSE continued to reduce to 1.6 t/ha by Fl+30. The model with only Variety+Temps features had higher errors than the model including remote sensing, but reached RMSE close to 2 t/ha by the Fl+0 window. There was no advantage in any case to including data beyond Fl+30.

### 3.2.4. *Model and prediction explanations*

After training models using 2018–2023 data, the TreeExplainer method of the SHAP library (Lundberg et al., 2020) was used with the LGB algorithm to further investigate the relationship between important features and yield. The contribution of the 12 most important features to yield predictions for the each of the three feature sets discussed in the previous section are shown in Fig. 12. Variety, Tmax at 30 days after flowering and Tmin at 15 days after flowering feature

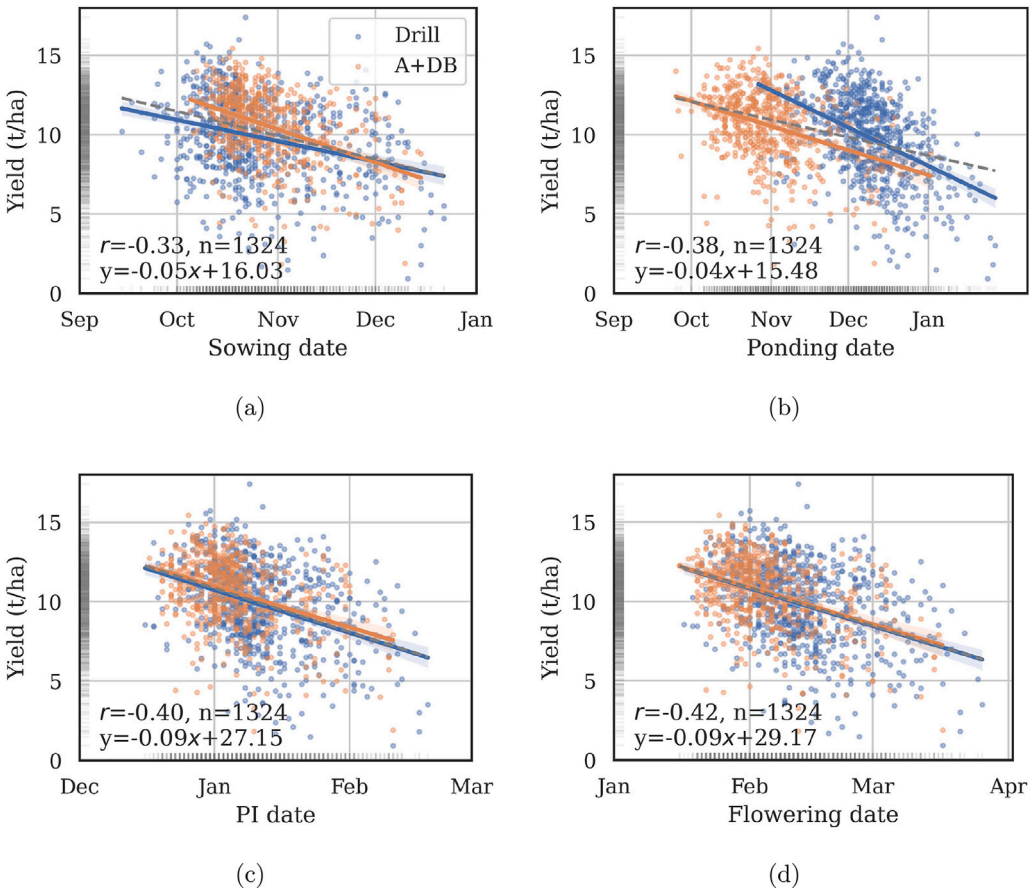



**Fig. 6.** The relationship of yield of all 1324 fields with continuous factors: sowing date (a), ponding date (b), PI date (c) and flowering date (d). The relationships for different sowing methods are separated by color.

strongly in all the models. Higher temperatures at and after flowering were generally predictive of higher yields. Lower temperatures early in the season (before Fl-30) were predictive of higher yields. This may be indicative of a secondary relationship, as earlier sowing dates encounter lower temperatures early in the season, and earlier sowing dates are associated with higher yields (Fig. 6(a)).

For the model that added CIRE to the features (Fig. 12(b)), CIRE just before flowering (Fl-15) was the most important predictor, and was positively related with yield, possibly because it is indicative of higher N uptake (Schlemmer et al., 2013). Conversely, it is not advantageous to have CIRE remaining high after flowering, as high CIRE(Fl+30) is predictive of lower yield. Slow decline in CIRE indicates slow maturing as occurred in 2021 (Fig. 3(b)), possibly caused by sterility.

The complete Variety+Temps+Refs+VIs model explanations (Fig. 12(c)) show the importance of reflectance bands, and particularly the red edge bands RE and RE74. Higher RE reflectances before flowering (e.g. RE(Fl-15)) were associated with lower yields. Conversely higher RE reflectances after flowering (e.g. RE74(Fl+30)) were predictive of higher yields. This can be explained by noting that lower RE means higher VIs (such as CIRE = NIR/RE-1), and as noted above, higher VIs are advantageous before flowering, and disadvantageous after flowering. VIs were also important, particularly CIRE and LSWI, though not as important as reflectances.

Examples of SHAP local explanations of high and low yield crops are provided in Fig. S5.

*3.2.5. Historical seasonal variability validation using weather-based model*

The historical validation test involved Reiziq yields from 2006–2017 using the LGB(Variety+Temps) model. As expected, the model could not predict much of the field-to-field variability (LCCC = 0.2), as no remote sensing information to characterize such spatial variability was available. However, the model predicted much of the inter-annual variability (LCCC = 0.6). This demonstrated the ability of our method to describe inter-annual rice yield fluctuations due to weather variability. Interested readers are referred to Fig. S6 for more detail.

## 4. Discussion

In this study, we have analyzed factors influencing rice yield using a multi-year dataset of field-level data including information on variety, sowing date and sowing method. We leveraged previously developed models that predicted water management (Brinkhoff et al., 2022) and growth stages (Brinkhoff et al., 2023b), and used these to engineer phenology-specific time-series features from weather and remote sensing sources for each field. These phenology-based aggregations were very important in generalizing yield predictions to a year (2023) with atypical weather and very different phenology dates to those in the training data (2018–2022). In more typical years, calendar-month based aggregation of at least the remote sensing variables may offer better predictions as our LOYO-CV experiments showed.

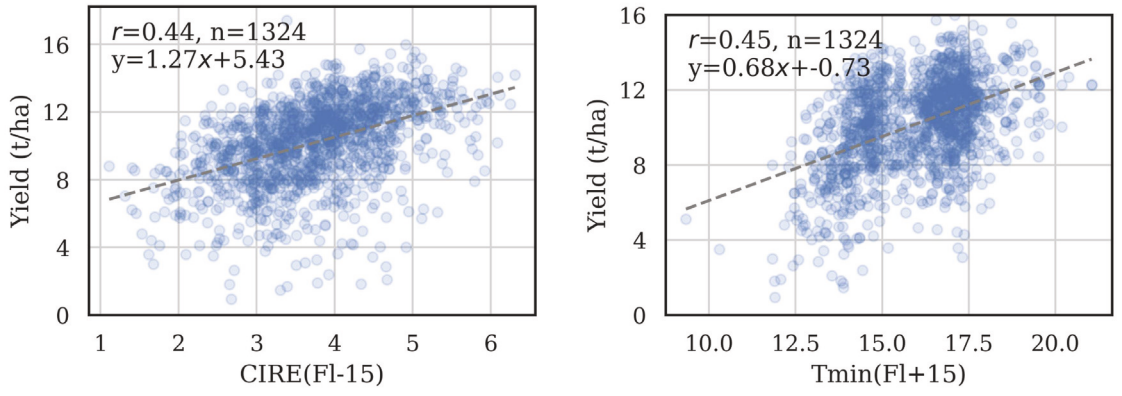
Our data fusion methodology and analyses facilitated several significant findings. We observed strong positive association between yield and remote sensing indices between PI and flowering, and between yield and minimum temperatures around flowering. It is important to reach PI and flowering early, before cooler seasonal temperatures become likely. This can be somewhat controlled by sowing early. We found significantly decreasing yields for fields that had consecutive rice crops. The reasons for this warrants further investigation in our study area, although work in other areas and growing systems suggested one possible cause is declining nutrient availability (Saito et al., 2006).

| | Aug | Sep | Oct | Nov | Dec | Jan | Feb | Mar | Apr | May |
|---|---|---|---|---|---|---|---|---|---|---|
| Tmax | 0.25 *** | 0.1 *** | 0.0 | -0.13 *** | 0.18 *** | 0.12 *** | 0.09 ** | 0.2 *** | 0.06 * | -0.1 *** |
| Tmin | -0.01 | -0.15 *** | -0.17 *** | -0.01 | 0.16 *** | 0.28 *** | 0.23 *** | 0.27 *** | 0.3 *** | 0.17 *** |
| Tmin<15 | -0.01 | -0.16 *** | -0.18 *** | 0.03 | 0.24 *** | 0.3 *** | 0.24 *** | 0.28 *** | 0.3 *** | 0.17 *** |
| Srad | 0.22 *** | 0.2 *** | 0.16 *** | -0.19 *** | 0.06 * | -0.19 *** | -0.15 *** | 0.07 ** | -0.14 *** | -0.1 *** |
| NDVI | -0.17 *** | -0.23 *** | -0.3 *** | 0.04 | 0.48 *** | 0.52 *** | 0.27 *** | -0.21 *** | -0.37 *** | -0.39 *** |
| CIG | -0.17 *** | -0.2 *** | -0.27 *** | -0.15 *** | 0.44 *** | 0.52 *** | 0.22 *** | -0.26 *** | -0.4 *** | -0.41 *** |
| LSWI | -0.22 *** | -0.25 *** | 0.03 | 0.25 *** | 0.43 *** | 0.61 *** | 0.52 *** | -0.11 *** | -0.32 *** | -0.38 *** |
| GRVI | -0.16 *** | -0.21 *** | -0.08 ** | 0.29 *** | 0.49 *** | 0.38 *** | 0.11 *** | -0.26 *** | -0.38 *** | -0.09 *** |
| NDRE | -0.18 *** | -0.23 *** | -0.31 *** | -0.08 ** | 0.48 *** | 0.56 *** | 0.12 *** | -0.29 *** | -0.41 *** | -0.42 *** |
| GNDVI | -0.17 *** | -0.21 *** | -0.31 *** | -0.15 *** | 0.43 *** | 0.52 *** | 0.29 *** | -0.2 *** | -0.35 *** | -0.37 *** |
| MNDWI | -0.22 *** | -0.19 *** | 0.18 *** | 0.24 *** | -0.05 * | -0.26 *** | 0.17 *** | 0.19 *** | -0.18 *** | -0.17 *** |
| CIRE | -0.18 *** | -0.21 *** | -0.26 *** | -0.03 | 0.47 *** | 0.58 *** | 0.06 * | -0.35 *** | -0.45 *** | -0.43 *** |
| VV | -0.03 | 0.04 | -0.2 *** | -0.28 *** | 0.02 | -0.15 *** | -0.1 *** | -0.03 | 0.0 | 0.18 *** |
| VH | -0.17 *** | -0.15 *** | -0.3 *** | -0.32 *** | 0.14 *** | 0.12 *** | 0.16 *** | 0.13 *** | 0.0 | 0.07 * |

(a) Monthly aggregations

| | Fl-120 | Fl-105 | Fl-90 | Fl-75 | Fl-60 | Fl-45 | Fl-30 | Fl-15 | Fl+0 | Fl+15 | Fl+30 | Fl+45 | Fl+60 | Fl+75 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tmax | -0.32 *** | -0.3 *** | -0.4 *** | -0.36 *** | -0.26 *** | -0.07 * | 0.14 *** | 0.16 *** | 0.27 *** | 0.41 *** | 0.41 *** | 0.37 *** | 0.32 *** | 0.27 *** |
| Tmin | -0.4 *** | -0.36 *** | -0.37 *** | -0.27 *** | -0.16 *** | -0.08 ** | 0.17 *** | 0.26 *** | 0.33 *** | 0.45 *** | 0.46 *** | 0.46 *** | 0.47 *** | 0.39 *** |
| Tmin<15 | -0.4 *** | -0.35 *** | -0.35 *** | -0.26 *** | -0.14 *** | -0.12 *** | 0.22 *** | 0.3 *** | 0.38 *** | 0.46 *** | 0.46 *** | 0.47 *** | 0.48 *** | 0.39 *** |
| Srad | -0.27 *** | -0.36 *** | -0.37 *** | -0.33 *** | -0.27 *** | 0.04 | 0.17 *** | 0.21 *** | 0.3 *** | 0.3 *** | 0.3 *** | 0.28 *** | 0.26 *** | 0.29 *** |
| NDVI | -0.15 *** | -0.14 *** | -0.09 ** | 0.02 | 0.17 *** | 0.26 *** | 0.31 *** | 0.33 *** | 0.31 *** | 0.24 *** | 0.12 *** | 0.01 | -0.07 ** | -0.15 *** |
| CIG | -0.12 *** | -0.15 *** | -0.15 *** | -0.09 ** | 0.12 *** | 0.25 *** | 0.32 *** | 0.36 *** | 0.33 *** | 0.2 *** | 0.09 *** | 0.01 | -0.08 ** | -0.16 *** |
| LSWI | -0.08 ** | 0.09 *** | 0.17 *** | 0.21 *** | 0.25 *** | 0.31 *** | 0.41 *** | 0.45 *** | 0.42 *** | 0.33 *** | 0.2 *** | 0.08 ** | -0.03 | -0.13 *** |
| GRVI | -0.03 | 0.12 *** | 0.17 *** | 0.21 *** | 0.24 *** | 0.24 *** | 0.21 *** | 0.18 *** | 0.14 *** | 0.14 *** | 0.02 | -0.08 ** | -0.12 *** | -0.14 *** |
| NDRE | -0.14 *** | -0.15 *** | -0.14 *** | -0.06 * | 0.14 *** | 0.27 *** | 0.36 *** | 0.43 *** | 0.37 *** | 0.18 *** | 0.05 | -0.03 | -0.1 *** | -0.18 *** |
| GNDVI | -0.16 *** | -0.17 *** | -0.15 *** | -0.1 *** | 0.08 ** | 0.24 *** | 0.32 *** | 0.36 *** | 0.34 *** | 0.25 *** | 0.15 *** | 0.06 * | -0.05 | -0.14 *** |
| MNDWI | 0.02 | 0.13 *** | 0.17 *** | 0.19 *** | 0.13 *** | -0.08 ** | -0.15 *** | -0.12 *** | -0.06 * | 0.01 | 0.06 * | 0.09 ** | 0.02 | -0.06 * |
| CIRE | -0.1 *** | -0.13 *** | -0.14 *** | -0.03 | 0.16 *** | 0.28 *** | 0.38 *** | 0.44 *** | 0.36 *** | 0.13 *** | -0.02 | -0.09 ** | -0.14 *** | -0.2 *** |
| VV | -0.11 *** | -0.19 *** | -0.22 *** | -0.2 *** | -0.11 *** | -0.06 * | -0.12 *** | -0.18 *** | -0.2 *** | -0.11 *** | -0.03 | 0.02 | 0.14 *** | 0.17 *** |
| VH | -0.21 *** | -0.23 *** | -0.22 *** | -0.2 *** | -0.1 *** | -0.0 | 0.01 | -0.08 ** | -0.06 * | 0.05 | 0.1 *** | 0.1 *** | 0.18 *** | 0.19 *** |

(b) Phenology aggregations

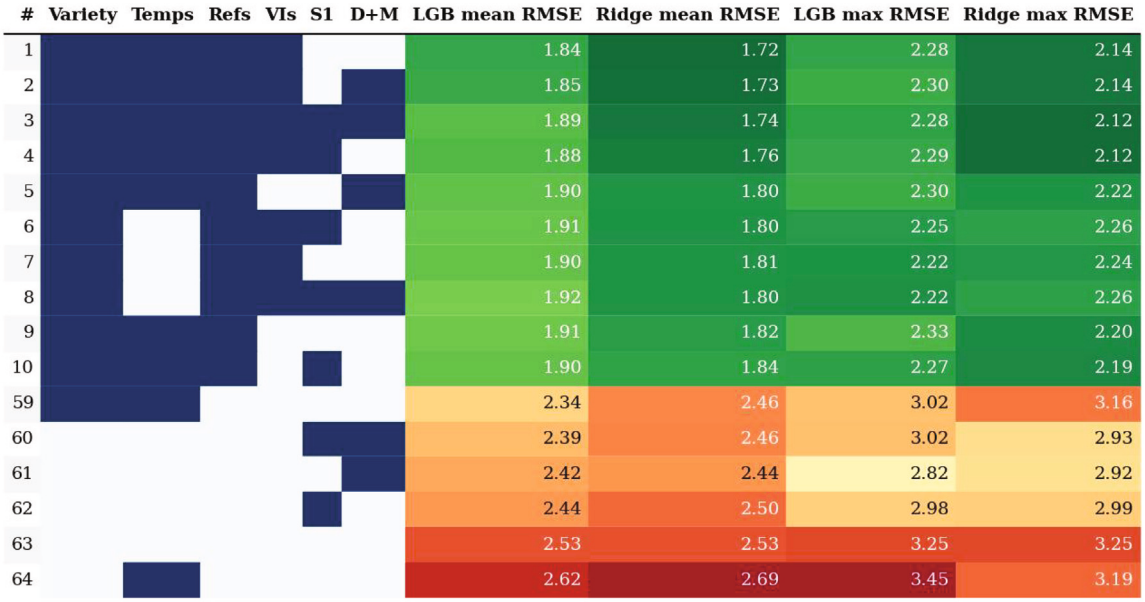



(c) (d)

**Fig. 7.** Correlation between yield and windowed aggregations of the time-series predictors for the 1324 2018–2023 fields. Time-series aggregations are the mean per month in (a), and per 30-day window relative to flowering date in (b). P-values are indicated with <0.001 = ***, <0.01 = ** and <0.05 = *. The individual relationships between yield and CIRE(Fl-15) and Tmin(Fl+15) are shown in (c) and (d).

Our yield forecast machine learning models were able to provide predictions well in advance of harvest in an independent season (2023) with challenging conditions (later sowing and colder temperatures than normal). The RMSE was 1.6 t/ha, the RRMSE was 16.3% and LCCC was 0.67. Errors reduced as yield was aggregated to coarser scales (region, variety or whole industry). Furthermore, we developed models based on a subset of variables to provide explainability of yield predictions for individual fields and all fields. These showed the negative effects both of low temperatures around flowering, and of low vegetation indices related with nitrogen status just before flowering.

These findings have implications for both industry stakeholders and farmers, as they enable yield forecasts well before harvest, offer insights



| # | Variety | Temps | Refs | VIs | S1 | D+M | LGB mean RMSE | Ridge mean RMSE | LGB max RMSE | Ridge max RMSE |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ■ | ■ | ■ | ■ | | | 1.84 | 1.72 | 2.28 | 2.14 |
| 2 | ■ | ■ | ■ | ■ | | ■ | 1.85 | 1.73 | 2.30 | 2.14 |
| 3 | ■ | ■ | ■ | ■ | ■ | ■ | 1.89 | 1.74 | 2.28 | 2.12 |
| 4 | ■ | ■ | ■ | ■ | ■ | | 1.88 | 1.76 | 2.29 | 2.12 |
| 5 | ■ | ■ | ■ | | | ■ | 1.90 | 1.80 | 2.30 | 2.22 |
| 6 | ■ | | ■ | ■ | ■ | | 1.91 | 1.80 | 2.25 | 2.26 |
| 7 | ■ | | ■ | ■ | | | 1.90 | 1.81 | 2.22 | 2.24 |
| 8 | ■ | | ■ | ■ | ■ | ■ | 1.92 | 1.80 | 2.22 | 2.26 |
| 9 | ■ | ■ | ■ | | | | 1.91 | 1.82 | 2.33 | 2.20 |
| 10 | ■ | ■ | ■ | | ■ | | 1.90 | 1.84 | 2.27 | 2.19 |
| 59 | ■ | ■ | | | | | 2.34 | 2.46 | 3.02 | 3.16 |
| 60 | | | | | ■ | ■ | 2.39 | 2.46 | 3.02 | 2.93 |
| 61 | | | | | | ■ | 2.42 | 2.44 | 2.82 | 2.92 |
| 62 | | | | | ■ | | 2.44 | 2.50 | 2.98 | 2.99 |
| 63 | | | | | | | 2.53 | 2.53 | 3.25 | 3.25 |
| 64 | | ■ | | | | | 2.62 | 2.69 | 3.45 | 3.19 |

(a) Monthly time-series aggregations

| # | Variety | Temps | Refs | VIs | S1 | D+M | LGB mean RMSE | Ridge mean RMSE | LGB max RMSE | Ridge max RMSE |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ■ | ■ | ■ | ■ | ■ | | 1.91 | 1.86 | 2.55 | 2.33 |
| 2 | ■ | ■ | ■ | ■ | | | 1.90 | 1.90 | 2.39 | 2.41 |
| 3 | ■ | ■ | ■ | ■ | | ■ | 1.93 | 1.88 | 2.50 | 2.38 |
| 4 | ■ | ■ | ■ | ■ | ■ | ■ | 1.92 | 1.90 | 2.44 | 2.46 |
| 5 | ■ | ■ | ■ | ■ | | | 1.92 | 1.91 | 2.41 | 2.57 |
| 6 | ■ | ■ | ■ | | ■ | | 1.92 | 1.92 | 2.40 | 2.48 |
| 7 | ■ | ■ | ■ | | ■ | | 1.89 | 1.95 | 2.45 | 2.61 |
| 8 | ■ | ■ | ■ | ■ | | ■ | 1.94 | 1.91 | 2.47 | 2.55 |
| 9 | ■ | ■ | ■ | ■ | ■ | ■ | 1.99 | 1.87 | 2.63 | 2.30 |
| 10 | ■ | ■ | ■ | ■ | ■ | | 1.96 | 1.92 | 2.62 | 2.43 |
| 33 | ■ | ■ | ■ | | | | 2.06 | 2.12 | 2.40 | 2.68 |
| 34 | ■ | | ■ | ■ | | | 2.07 | 2.11 | 2.81 | 2.84 |
| 60 | ■ | | | | ■ | | 2.42 | 2.34 | 3.09 | 3.13 |
| 61 | ■ | | | | | | 2.39 | 2.39 | 3.18 | 3.18 |
| 62 | | | | | | ■ | 2.42 | 2.44 | 2.82 | 2.92 |
| 63 | | | | | ■ | | 2.48 | 2.49 | 3.25 | 3.15 |
| 64 | | | | | | | 2.53 | 2.53 | 3.25 | 3.25 |

(b) Phenology time-series aggregations

**Fig. 8.** 2018-2022 LOYO CV experiments (n = 1072), ranked by average RMSE (t/ha). The results with time-series aggregated by month (a) and phenology (b) are shown. Of the 64 possible combinations of feature sets (rank indicated by the # column), only the top 10, and bottom 5, as well as the Variety+Temps and Variety+Refs+VIs sets are shown. D+M denotes the dates and sowing method feature sets.

into the drivers of yield variability, and suggest actionable strategies to improve productivity. Such strategies include optimizing sowing date, crop rotation and nitrogen management.

### *4.1. Time-series features for enhanced yield forecasting*

The integration of time-series data with machine learning techniques has emerged as a dynamic field with substantial potential, and in tandem with developments in high-cadence remote sensing data, offers promise for more powerful predictive capabilities (Foumani et al., 2023). In the context of rice yield forecasting, Zhou et al. (2017) found an improvement in accuracy when UAV images from two dates were used instead of a single image and Ha et al. (2023) showed the importance of multi-temporal weather features. Marshall et al. (2022) found better predictions when all images from a season were used as predictors. We also demonstrated that models based on a time-series of features provided better predictions than those based on features from a single point in time.

Other research has demonstrated the benefit of resampling time-series data using thermal time from sowing (Desloires et al., 2023), or aligning time-series features with crop phenology (Bolton and Friedl, 2013; Ji et al., 2022). These methods can provide improved predictions relative to sampling time-series according to fixed calendar periods, particularly in situations where phenology is temporally diverse (Bregaglio et al., 2023). This was also the case in our work. In 2023, which had later sowing due to flooding, delayed phenological progression

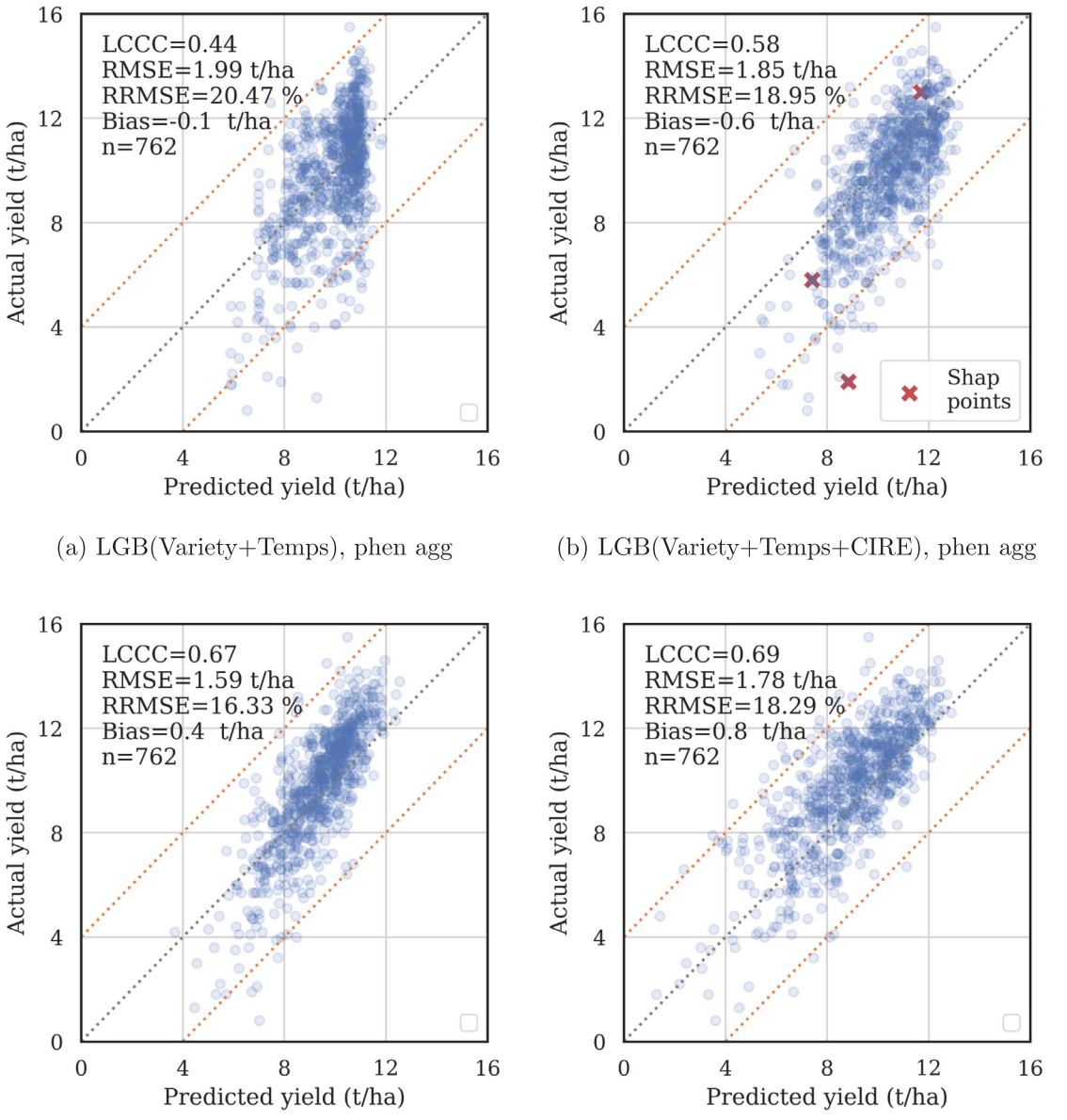



**Fig. 9.** Independent test year (2023) actual vs. predicted yields for 3 phenology-based aggregation models, LGB with Variety+Temps features (a), LGB with Variety+Temps+CIRE features (b) and Ridge with Variety+Temps+Refs+VIs features (c). For comparison, (d) shows a model with the same base features as (c), but with monthly instead of phenological time-series aggregations. The yield points used in SHAP explanations in Fig. S5 are indicated in (b).

and colder temperatures, our method of feature engineering time-series using per-field flowering dates yielded superior predictions than models that generated time-series features using fixed calendar months. Therefore, phenology-based aggregations are important for generalizing to environments not adequately captured in training data. We also showed that if remote sensing variables are not available, phenology-based aggregation is very important to make use of temperature variables, as models with temperatures aggregated to monthly periods gave very poor predictions compared to those with temperatures aggregated to phenological periods.

In contrast to some previous methods (Bolton and Friedl, 2013; Ji et al., 2022) which use generic land-surface phenology 'greenup' to reference time-series against, we used a rice-specific growth stage model that predicts and was validated against actual physiological growth stages, such as panicle initiation and flowering (Brinkhoff et al., 2023b). The advantage of land-surface phenology-based methods is that they are inherently scalable, as they rely only on satellite data. In contrast, our method relies on rice-specific phenology models and sowing date, which may not be available in other rice-growing areas, potentially limiting scalability. However, when we tested including phenology dates and sowing method as model variables, these were not very important to obtain accurate predictions (Fig. 8). Also, in years where phenology dates were typical, the monthly aggregations provided as good or better predictions than phenology-based aggregations, which may therefore enable scalability of the monthly aggregation models. However, as discussed above, phenology aggregations were very important in a season with abnormal sowing and phenology dates (2023).

We note that our results also showed the importance of including rice variety as a model input, with the top 10 models in the LOYO-CV experiments all using variety (Fig. 8). However, variety may not be an

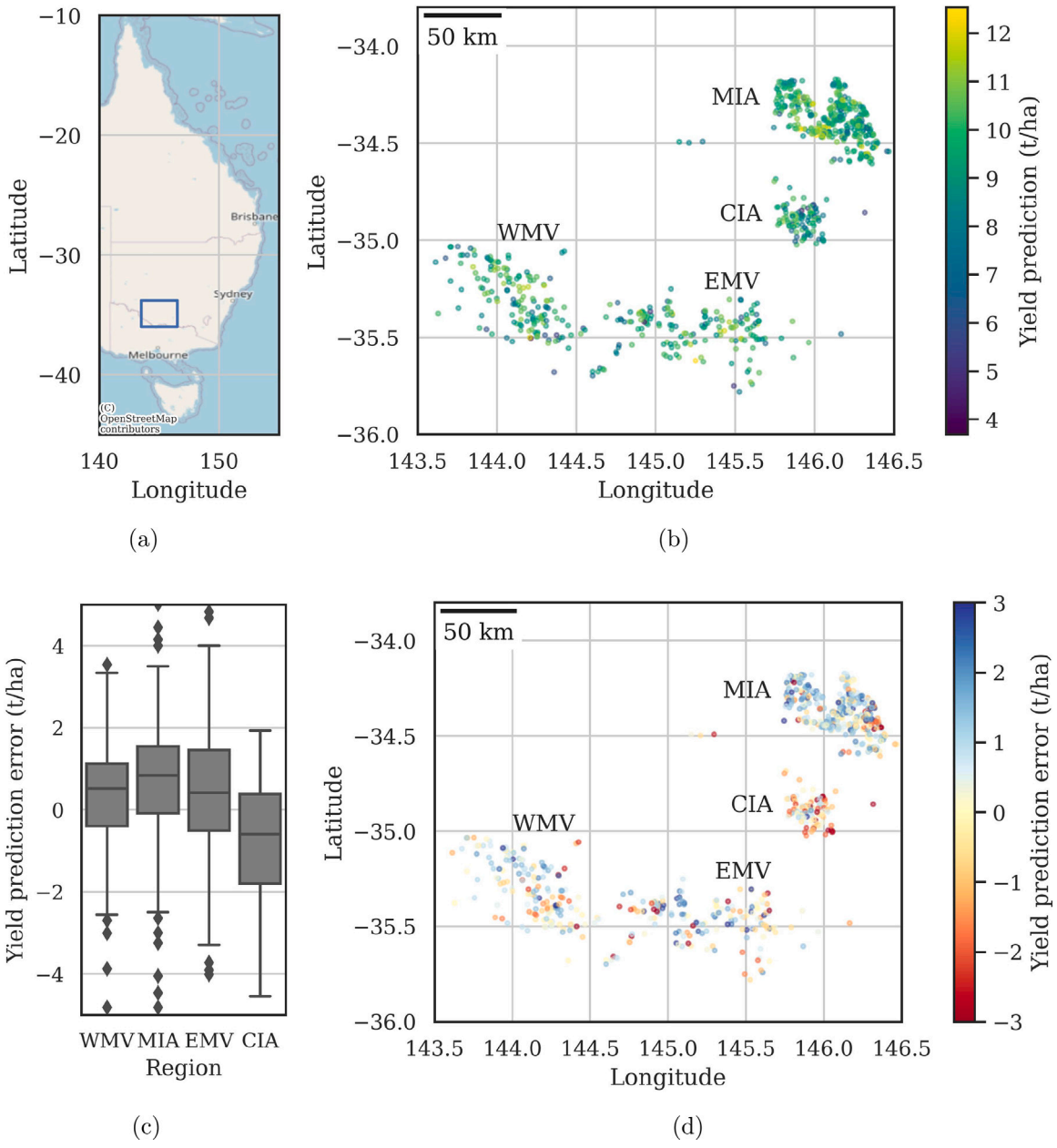



**Fig. 10.** Box showing the extent of the study area in New South Wales, Australia (a), 2023 yield predictions (b), distribution of errors per region (c), and map of prediction errors (d) for the 762 crops using the Variety+Temps+Refs+VIs Ridge model.

important model input in regions where there is less variance in the yield and phenology characteristics of the varieties grown.

Time-series yield forecasting methods allow a sequence of predictions to be generated, making use of the latest remote sensing and weather data available as a growing season progresses. As other work has demonstrated, this technique provides both early forecasts with reduced accuracy, and improving forecast accuracy as harvest approaches (Von Bloh et al., 2023; Potgieter et al., 2022). We showed continuously improving accuracy from forecasts provided 90 days before flowering to 30 days after flowering. The later is typically about 20–25 days before harvest (Brinkhoff et al., 2023a).

### 4.2. Important spectral reflectance bands

In common with other work, we found that the red edge bands were very important in yield prediction (Zhang et al., 2019). These wavelengths are strong indicators of chlorophyll and thus nitrogen status (Schlemmer et al., 2013; Wang et al., 2023b; Brinkhoff et al., 2021; Inoue et al., 2012), which may explain their importance. We also found indices based on short wave infrared bands (LSWI) were important (Marshall et al., 2022). This highlights the importance of inclusion of the red edge and shortwave infrared bands in remote sensing products targeting agricultural applications.

There were positive relationships between yield and CIRE and LSWI between PI and flowering. Conversely, CIRE 30 days after flowering was negatively related with yield, possibly because higher CIRE after flowering is indicative of slower maturing. Anecdotally, 2021 crops that suffered low-temperature induced sterility tended to mature slower. This may be due to a similar effect as that noted by Nakano et al. (1995), where removal of rice panicles retarded the decrease in photosynthesis of the rice flag leaf during senescence.

Most yield prediction studies using remote sensing consider only VIs derived from reflectances. However, we found that using the raw reflectances themselves were more powerful than VIs, as others have

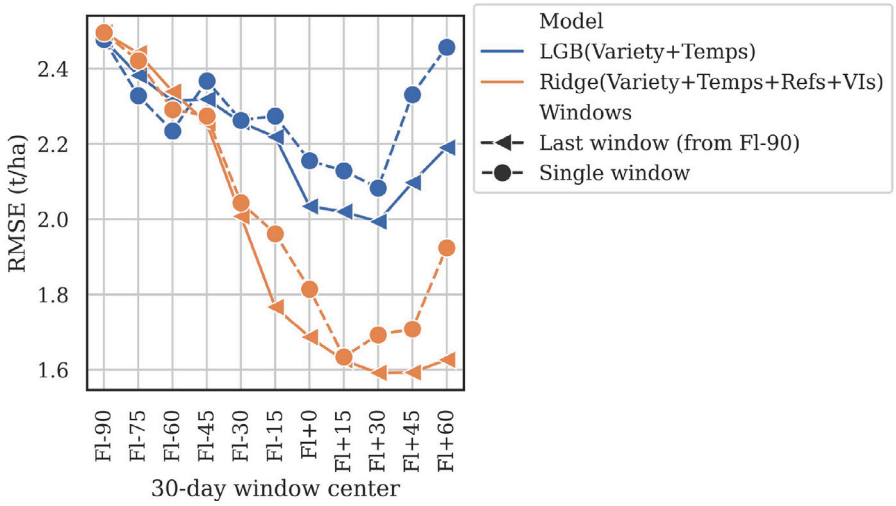



**Fig. 11.** 2023 yield forecast accuracy evolution as models incorporated later data as the season progressed. The graph compares models using only a single time window, to models using all time windows from Fl-90, for two models.

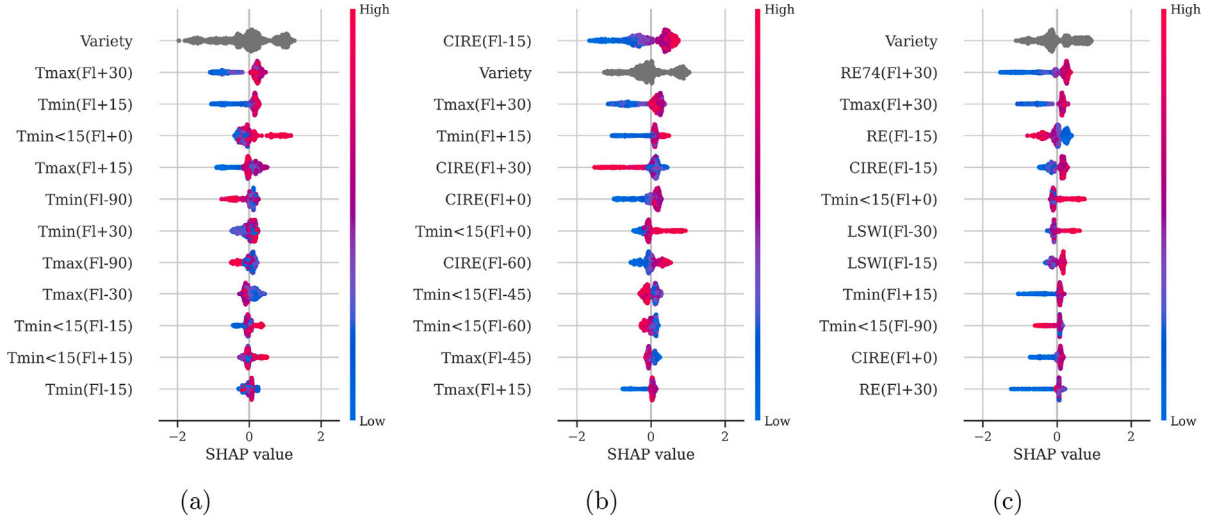



**Fig. 12.** Global model explanations, indicating contribution of features to yield predictions using 2018–2023 data for Variety+Temps (a), Variety+Temps+CIRE (b), and Variety+Temps+Refs+VIs (c) feature sets.

also found (Marszalek et al., 2022; Franch et al., 2021; Skakun et al., 2021). But combining VIs and reflectances gave even better results.

### 4.3. Scale of predictions

We predicted yield at the individual field level. This is in contrast to much previous work that has either predicted at the regional or county level (Zhou et al., 2023; Bolton and Friedl, 2013; Meroni et al., 2021; Ha et al., 2023; Liu et al., 2022; Islam et al., 2023) or those that focus on in-field variability or plot-level predictions using relatively expensive high resolution imagery from UAV and aerial platforms (Zhou et al., 2017; Luo et al., 2022; Wan et al., 2020). Both regional-level yield statistics and plot experiments are sources of more readily available yield data, compared with large-scale field-level datasets such as the one used in this work. However, field-level yield predictions arguably provide the most value, as they provide insights for individual farmers and fields, and are also scalable to provide accurate regional forecasts.

Previous works have shown that as fine-scale yield predictions are aggregated to coarser scales (regional or country), accuracies improve (Deines et al., 2020; Filippi et al., 2020; Yu et al., 2023), which is another benefit of providing field-level, rather than region yield predictions. We similarly found when field predictions were aggregated to variety and region levels, the errors were lower (<10% over all regions/varieties, and 4% for all rice crops).

Field-level yield predictions will have benefits for individual growers, supporting pre-harvest finance and logistic decisions (Filippi et al., 2020) Additionally, the SHAP model explanations demonstrated important factors that can lead to low yield, including low chlorophyll content (related to nitrogen application, Schlemmer et al. (2013) and Dunn et al. (2016), and late planting, which increases the chance of encountering low temperatures during the sensitive reproductive stages (Brinkhoff et al., 2023b). These explanations could be provided to growers to promote understanding of the causes of low or high yields of individual fields in previous seasons, thus supporting optimized decision-making to improve productivity in future seasons.

### 4.4. The challenge of forecasting the effect of cold-induced sterility

One of the biggest challenges to predicting rice yield in temperate growing regions is the impact of cold temperature shocks during sensitive reproductive stages, which can result in sterility and thus very

low yields (Espe et al., 2017; Sivapalan et al., 2007; Alemayehu et al., 2021). Though the trends are clear, it is very difficult to characterize the magnitude of the effect of such cold shocks on yield, as seen in the low correlation coefficients between temperatures and spikelet sterility in experiments (Farrell et al., 2006). There are complex interactions of the degree of cold-induced sterility and factors such as nitrogen (Gunawardena and Fukai, 2005; Heenan, 1984) and interaction between the effects of temperatures experienced at different times during a growing season (Godwin et al., 1994; Shimono et al., 2007). Crop models still struggle to describe yield loss due to cold temperatures adequately (Espe et al., 2016).

Remote sensing alone is unlikely to predict sterility, as sterility may not significantly affect spectral reflectance. Therefore, we fused weather and remote sensing data. Our model was able to predict a large degree of the variability in yields at the field level in a year where low temperatures were encountered (2023). However, the model had limited ability to predict the magnitude of yield loss in some particularly hard-hit fields. Developing machine learning models that can accurately predict the impact of extreme weather events on yield is a generally challenging problem (Paudel et al., 2023). One management factor growers use to minimize risk of cold-induced sterility is to ensure deep water during microspore, which provides a buffer against cold night time temperatures (Williams and Angus, 1994). Shimono et al. (2005) showed that it is the temperature of the panicle which largely determines spikelet sterility. Therefore predicting the extent of sterility requires knowledge of multiple variables including water depth and temperature, air temperature, and the height of the panicle above the water. However, models based on public remote sensing and weather data, such as ours, cannot describe most of these factors. This may explain some of the difficulty in predicting the amount of yield loss due to cold on a per-field basis. Incorporation of in-field water level and temperature sensors may improve predictions (Shimono et al., 2005). However only a limited number of fields across a region are likely to have such sensors, and the data may be hard to obtain. Additionally, water levels are not consistent between fields, so measuring one field does not guarantee good predictions in an adjacent field. Therefore, solutions based on in-field sensors are not likely to be scalable industry-wide. Satellite sensors that measure land surface temperature (Wei et al., 2023) or SAR data that is able to quantify water levels, such as polarimetric L-band data (Arai et al., 2022), may provide solutions to these problems in the future. However, the revisit times of such satellites is currently limited so may not detect the dynamics of temperatures and water levels with sufficient temporal detail to characterize field status at critical timings such as microspore and flowering.

## 5. Conclusion

We firstly analyzed factors driving rice yield variability in a temperate growing region using a multi-year dataset. Significant factors included variety, field cropping history (lower yields when a field was planted with rice in consecutive years) and phenology date (later panicle initiation and flowering were associated with lower yields). We then leveraged previously developed rice ponding date and phenological date prediction models to develop phenology-specific windows to aggregate time-series data from remote sensing and weather datasets. Vegetation indices related to nitrogen status (e.g. CIRE) and temperatures around flowering were both positively correlated with yield.

We then developed rice yield forecast models, building on these analyses and using ML algorithms. The fusion of remote sensing and weather variables aggregated to periods relative to field-specific flowering dates provided powerful features for these models. They were able to forecast yield in an independent season with challenging weather conditions (2023), achieving RRMSE of 16.3% on a per-crop basis, and the error between average predicted and actual yield for all crops was 4%. In a further validation experiment, models based only on weather and estimated phenology dates were able to predict a large degree of the inter-annual variability of yield (LCCC = 0.6) using an independent historical dataset. We developed a model using a subset of predictors, which, though less accurate than the full models, provided enhanced interpretability of yield predictions. This, combined with Shapley additive explanations (SHAP) explained the reasons for variability in yield performance of individual fields, identifying factors such as sowing date, impact on temperatures encountered during flowering, and chlorophyll content.

The yield analyses and the ML models both showed the importance of sowing on time, mitigating the potentially harmful impacts of cold temperatures around flowering, and of achieving sufficient nitrogen uptake before flowering (indicated by high CIRE). The challenge of accurately predicting the full impact of cold-temperature induced sterility remains, and we suggested it is likely that additional in-field information such as water depth may be needed for this. This work provides a framework for analyzing yield variability, and a method to develop models that are suitable to operationally forecast in-season yield, thus providing value to both individual farmers and to rice industry bodies.

## CRediT authorship contribution statement

**James Brinkhoff:** Writing – original draft, Validation, Software, Methodology, Investigation, Funding acquisition, Formal analysis. **Allister Clarke:** Writing – review & editing, Methodology, Data curation. **Brian W. Dunn:** Writing – review & editing, Supervision, Conceptualization. **Mark Groat:** Writing – review & editing, Supervision, Methodology, Conceptualization.

## 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.

## Data availability

The data that has been used is confidential.

## Acknowledgments

We appreciate the detailed suggestions provided by the reviewers and editor, which helped improve our paper. This work was funded by AgriFutures Australia, grant PRO-013078 (Real-time remote-sensing based monitoring for the rice industry). Rice planting and productivity data was supplied by SunRice. Many discussions with members of the Rice Extension, the New South Wales Department of Primary Industries, and the University of New England Applied Agricultural Remote Sensing Centre teams have provided inspiration and suggestions for this research. We are particularly grateful for the input from Chris Quirk, Troy Mauger, Peter McDonnell, Christopher Proud, Peter Snell, Alex Schultz and Sharon McGavin.

## Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.agrformet.2024.110055.

## References


Alemayehu, H.A., Dumbuya, G., Hasan, M., Tadesse, T., Nakajyo, S., Fujioka, T., Abe, A., Matsunami, M., Shimono, H., 2021. Genotypic variation in cold tolerance of 18 Ethiopian rice cultivars in relation to their reproductive morphology. Field Crops Res. 262, 108042. http://dx.doi.org/10.1016/j.fcr.2020.108042.

Ali, F., Waters, D.L., Ovenden, B., Bundock, P., Raymond, C.A., Rose, T.J., 2019. Australian rice varieties vary in grain yield response to heat stress during reproductive and grain filling stages. J. Agron. Crop Sci. 205 (2), 179–187. http://dx.doi.org/10.1111/jac.12312.

Arai, H., Le Toan, T., Takeuchi, W., Oyoshi, K., Fumoto, T., Inubushi, K., 2022. Evaluating irrigation status in the Mekong Delta through polarimetric L-band SAR data assimilation. Remote Sens. Environ. 279, 113139. http://dx.doi.org/10.1016/j.rse.2022.113139.

Basso, B., Liu, L., 2019. Seasonal crop yield forecast: Methods, applications, and accuracies. In: Advances in Agronomy. Vol. 154, Elsevier, pp. 201–255. http://dx.doi.org/10.1016/bs.agron.2018.11.002, URL https://linkinghub.elsevier.com/retrieve/pii/S0065211318300944.

Bolton, D.K., Friedl, M.A., 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agricult. Forest Meteorol. 173, 74–84. http://dx.doi.org/10.1016/j.agrformet.2013.01.007.

Bouman, B.A.M., Tuong, T.P., 2001. Field water management to save water and increase its productivity in irrigated lowland rice. Agricult. Water Manag. 49 (1), 11–30. http://dx.doi.org/10.1016/S0378-3774(00)00128-1.

Bregaglio, S., Ginaldi, F., Raparelli, E., Fila, G., Bajocco, S., 2023. Improving crop yield prediction accuracy by embedding phenological heterogeneity into model parameter sets. Agric. Syst. 209, 103666. http://dx.doi.org/10.1016/j.agsy.2023.103666.

Brinkhoff, J., Dunn, B.W., Dunn, T., 2023a. The influence of nitrogen and variety on rice grain moisture content dry-down. Field Crops Res. 302, 109044. http://dx.doi.org/10.1016/j.fcr.2023.109044.

Brinkhoff, J., Dunn, B.W., Robson, A.J., 2021. Rice nitrogen status detection using commercial-scale imagery. Int. J. Appl. Earth Obs. Geoinf. 105, 102627. http://dx.doi.org/10.1016/j.jag.2021.102627.

Brinkhoff, J., Dunn, B.W., Robson, A.J., Dunn, T.S., Dehaan, R.L., 2019. Modeling mid-season rice nitrogen uptake using multispectral satellite data. Remote Sens. 11 (15), 1837. http://dx.doi.org/10.3390/rs11151837.

Brinkhoff, J., Houborg, R., Dunn, B.W., 2022. Rice ponding date detection in Australia using Sentinel-2 and Planet Fusion imagery. Agricult. Water Manag. 273, 107907. http://dx.doi.org/10.1016/j.agwat.2022.107907.

Brinkhoff, J., McGavin, S.L., Dunn, T., Dunn, B.W., 2023b. Predicting rice phenology and optimal sowing dates in temperate regions using machine learning. Agron. J. http://dx.doi.org/10.1002/agj2.21398.

Brinkhoff, J., Robson, A.J., 2021. Block-level macadamia yield forecasting using spatio-temporal datasets. Agricult. Forest Meteorol. 303, 108369. http://dx.doi.org/10.1016/j.agrformet.2021.108369.

Cao, J., Zhang, Z., Tao, F., Zhang, L., Luo, Y., Zhang, J., Han, J., Xie, J., 2021. Integrating multi-source data for rice yield prediction across China using machine learning and deep learning approaches. Agricult. Forest Meteorol. 297, 108275. http://dx.doi.org/10.1016/j.agrformet.2020.108275.

Chang, K.-W., Shen, Y., Lo, J.-C., 2005. Predicting rice yield using canopy reflectance measured at booting stage. Agron. J. 97 (3), 872–878. http://dx.doi.org/10.2134/agronj2004.0162.

Deines, J.M., Patel, R., Liang, S.-Z., Dado, W., Lobell, D.B., 2020. A million kernels of truth: Insights into scalable satellite maize yield mapping and yield gap analysis from an extensive ground dataset in the US Corn Belt. Remote Sens. Environ. 112174. http://dx.doi.org/10.1016/j.rse.2020.112174.

dela Torre, D.M.G., Gao, J., Macinnis-Ng, C., 2021. Remote sensing-based estimation of rice yields using various models: A critical review. Geo-Spat. Inf. Sci. 24 (4), 580–603. http://dx.doi.org/10.1080/10095020.2021.1936656, Publisher: Taylor & Francis.

Delerce, S., Dorado, H., Grillon, A., Rebolledo, M.C., Prager, S.D., Patiño, V.H., Varón, G.G., Jiménez, D., 2016. Assessing weather-yield relationships in rice at local scale using data mining approaches. PLoS One 11 (8), e0161620. http://dx.doi.org/10.1371/journal.pone.0161620, Publisher: Public Library of Science.

Desloires, J., Ienco, D., Botrel, A., 2023. Out-of-year corn yield prediction at field-scale using Sentinel-2 satellite imagery and machine learning methods. Comput. Electron. Agric. 209, 107807. http://dx.doi.org/10.1016/j.compag.2023.107807.

Dunn, B., Dunn, T., 2023. Rice variety guides. URL https://www.dpi.nsw.gov.au/_data/assets/pdf_file/0007/1472380/Rice-variety-guides.pdf.

Dunn, B.W., Dunn, T.S., Orchard, B.A., 2016. Nitrogen rate and timing effects on growth and yield of drill-sown rice. Crop Pasture Sci. 67 (11), 1149–1157. http://dx.doi.org/10.1071/CP16331.

Dunn, B.W., Gaydon, D.S., 2011. Rice growth, yield and water productivity responses to irrigation scheduling prior to the delayed application of continuous flooding in south-east Australia. Agricult. Water Manag. 98 (12), 1799–1807. http://dx.doi.org/10.1016/j.agwat.2011.07.004.

Espe, M.B., Hill, J.E., Hijmans, R.J., McKenzie, K., Mutters, R., Espino, L.A., Leinfelder-Miles, M., van Kessel, C., Linquist, B.A., 2017. Point stresses during reproductive stage rather than warming seasonal temperature determine yield in temperate rice. Global Change Biol. 23 (10), 4386–4395. http://dx.doi.org/10.1111/gcb.13719.

Espe, M.B., Yang, H., Cassman, K.G., Guilpart, N., Sharifi, H., Linquist, B.A., 2016. Estimating yield potential in temperate high-yielding, direct-seeded US rice production systems. Field Crops Res. 193, 123–132. http://dx.doi.org/10.1016/j.fcr.2016.04.003.

Eugenio, F.C., Grohs, M., Schuh, M., Venancio, L.P., Schons, C., Badin, T.L., Mallmann, C.L., Fernandes, P., Pereira da Silva, S.D., Fantinel, R.A., 2023. Estimated flooded rice grain yield and nitrogen content in leaves based on RPAS images and machine learning. Field Crops Res. 292, 108823. http://dx.doi.org/10.1016/j.fcr.2023.108823.

Farrell, T.C., Fox, K.M., Williams, R.L., Fukai, S., Lewin, L.G., Farrell, T.C., Fox, K.M., Williams, R.L., Fukai, S., Lewin, L.G., 2006. Minimising cold damage during reproductive development among temperate rice genotypes. II. Genotypic variation and flowering traits related to cold tolerance screening. Aust. J. Agric. Res. 57 (1), 89–100. http://dx.doi.org/10.1071/AR05186, Publisher: CSIRO PUBLISHING.

Filippi, P., Whelan, B.M., Vervoort, R.W., Bishop, T.F.A., 2020. Mid-season empirical cotton yield forecasts at fine resolutions using large yield mapping datasets and diverse spatial covariates. Agric. Syst. 184, 102894. http://dx.doi.org/10.1016/j.agsy.2020.102894.

Foumani, N.M., Miller, L., Tan, C.W., Webb, G.I., Forestier, G., Salehi, M., 2023. Deep learning for time series classification and extrinsic regression: A current survey. URL http://arxiv.org/abs/2302.02515. arXiv:2302.02515 [cs].

Franch, B., Bautista, A.S., Fita, D., Rubio, C., Tarrazó-Serrano, D., Sánchez, A., Skakun, S., Vermote, E., Becker-Reshef, I., Uris, A., 2021. Within-field rice yield estimation based on Sentinel-2 satellite data. Remote Sens. 13 (20), 4095. http://dx.doi.org/10.3390/rs13204095, Number: 20 Publisher: Multidisciplinary Digital Publishing Institute.

Gitelson, A.A., Viña, A., Ciganda, V., Rundquist, D.C., Arkebauer, T.J., 2005. Remote estimation of canopy chlorophyll content in crops. Geophys. Res. Lett. 32 (8), http://dx.doi.org/10.1029/2005GL022688.

Godwin, D.C., Meyer, W.S., Singh, U., 1994. Simulation of the effect of chilling injury and nitrogen supply on floret fertility and yield in rice. Aust. J. Exp. Agric. 34 (7), 921–926. http://dx.doi.org/10.1071/ea9940921, Publisher: CSIRO PUBLISHING.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: Planetary-scale geospatial analysis for everyone. Big Remotely Sensed Data: tools, applications and experiences, Remote Sens. Environ. Big Remotely Sensed Data: tools, applications and experiences, 202, 18–27.http://dx.doi.org/10.1016/j.rse.2017.06.031,

Grinsztajn, L., Oyallon, E., Varoquaux, G., 2022. Why do tree-based models still outperform deep learning on tabular data? http://dx.doi.org/10.48550/arXiv.2207.08815, URL http://arxiv.org/abs/2207.08815.

Gu, C., Ji, S., Xi, X., Zhang, Z., Hong, Q., Huo, Z., Li, W., Mao, W., Zhao, H., Zhang, R., Li, B., Tan, C., 2022. Rice yield estimation based on continuous wavelet transform with multiple growth periods. Front. Plant Sci. 13.

Guijo-Rubio, D., Middlehurst, M., Arcencio, G., Silva, D.F., Bagnall, A., 2023. Unsupervised feature based algorithms for time series extrinsic regression. URL http://arxiv.org/abs/2305.01429. arXiv:2305.01429 [cs, stat].

Gunawardena, T.A., Fukai, S., 2005. The interaction of nitrogen application and temperature during reproductive stage on spikelet sterility in field-grown rice. Aust. J. Agric. Res. 56 (6), 625. http://dx.doi.org/10.1071/AR04099.

Ha, S., Kim, Y.-T., Im, E.-S., Hur, J., Jo, S., Kim, Y.-S., Shim, K.-M., 2023. Impacts of meteorological variables and machine learning algorithms on rice yield prediction in Korea. Int. J. Biometeorol. http://dx.doi.org/10.1007/s00484-023-02544-x.

Hashimoto, N., Saito, Y., Yamamoto, S., Ishibashi, T., Ito, R., Maki, M., Homma, K., 2022. Feasibility of yield estimation based on leaf area dynamics measurements in rice paddy fields of farmers. Field Crops Res. 286, 108609. http://dx.doi.org/10.1016/j.fcr.2022.108609.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, second ed. Springer Series in Statistics, Springer-Verlag, New York, URL https://www.springer.com/gp/book/9780387848570.

Heenan, D., 1984. Low-temperature induced floret sterility in the rice cultivars Calrose and Inga as influenced by nitrogen supply. Aust. J. Exp. Agric. 24 (125), 255. http://dx.doi.org/10.1071/EA9840255.

Huber, F., Yushchenko, A., Stratmann, B., Steinhage, V., 2022. Extreme gradient boosting for yield estimation compared with deep learning approaches. Comput. Electron. Agric. 202, 107346. http://dx.doi.org/10.1016/j.compag.2022.107346.

Inoue, Y., Sakaiya, E., Zhu, Y., Takahashi, W., 2012. Diagnostic mapping of canopy nitrogen content in rice based on hyperspectral measurements. Remote Sens. Environ. 126, 210–221. http://dx.doi.org/10.1016/j.rse.2012.08.026.

Islam, M.D., Di, L., Qamer, F.M., Shrestha, S., Guo, L., Lin, L., Mayer, T.J., Phalke, A.R., 2023. Rapid rice yield estimation using integrated remote sensing and meteorological data and machine learning. Remote Sens. 15 (9), 2374. http://dx.doi.org/10.3390/rs15092374, Number: 9 Publisher: Multidisciplinary Digital Publishing Institute.

Jeffrey, S.J., Carter, J.O., Moodie, K.B., Beswick, A.R., 2001. Using spatial interpolation to construct a comprehensive archive of Australian climate data. Environ. Model. Softw. 16 (4), 309–330. http://dx.doi.org/10.1016/S1364-8152(01)00008-1.

Jena, K.K. (Ed.), 2012. Advances in Temperate Rice Research. IRRI, Metro Manila, Philippines.

Ji, Z., Pan, Y., Zhu, X., Zhang, D., Wang, J., 2022. A generalized model to predict large-scale crop yields integrating satellite-based vegetation index time series and phenology metrics. Ecol. Indic. 137, 108759. http://dx.doi.org/10.1016/j.ecolind.2022.108759.

Kang, Y., Ozdogan, M., Zhu, X., Ye, Z., Hain, C., Anderson, M., 2020. Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest. Environ. Res. Lett. 15 (6), 064005. http://dx.doi.org/10.1088/1748-9326/ab7df9, Publisher: IOP Publishing.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. LightGBM: A highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems. Vol. 30, Curran Associates, Inc., URL https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html.

Li, T., Hasegawa, T., Yin, X., Zhu, Y., Boote, K., Adam, M., Bregaglio, S., Buis, S., Confalonieri, R., Fumoto, T., Gaydon, D., Marcaida, M., Nakawaga, H., Oriol, P., Ruane, A., Ruget, F., Singh, B., Singh, U., Tang, L., Bouman, B., 2015. Uncertainties in predicting rice yield by current crop models under contrasting climatic environments, 2014..

Lin, L.I.-K., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics 45 (1), 255–268. http://dx.doi.org/10.2307/2532051, Publisher: [Wiley, International Biometric Society].

Liu, Y., Wang, S., Chen, J., Chen, B., Wang, X., Hao, D., Sun, L., 2022. Rice yield prediction and model interpretation based on satellite and climatic indicators using a transformer method. Remote Sens. 14 (19), 5045. http://dx.doi.org/10.3390/rs14195045, Number: 19 Publisher: Multidisciplinary Digital Publishing Institute.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2 (1), 56–67. http://dx.doi.org/10.1038/s42256-019-0138-9, Number: 1 Publisher: Nature Publishing Group.

Luo, S., Jiang, X., Jiao, W., Yang, K., Li, Y., Fang, S., 2022. Remotely sensed prediction of rice yield at different growth durations using UAV multispectral imagery. Agriculture 12 (9), 1447. http://dx.doi.org/10.3390/agriculture12091447, Number: 9 Publisher: Multidisciplinary Digital Publishing Institute.

Marshall, M., Belgiu, M., Boschetti, M., Pepe, M., Stein, A., Nelson, A., 2022. Field-level crop yield estimation with PRISMA and Sentinel-2. ISPRS J. Photogramm. Remote Sens. 187, 191–210. http://dx.doi.org/10.1016/j.isprsjprs.2022.03.008.

Marszalek, M., Körner, M., Schmidhalter, U., 2022. Prediction of multi-year winter wheat yields at the field level with satellite and climatological data. Comput. Electron. Agric. 194, 106777. http://dx.doi.org/10.1016/j.compag.2022.106777.

Meroni, M., Waldner, F., Seguini, L., Kerdiles, H., Rembold, F., 2021. Yield forecasting with machine learning and small data: What gains for grains? Agricult. Forest Meteorol. 308–309, http://dx.doi.org/10.1016/j.agrformet.2021.108555.

Myneni, R.B., Williams, D.L., 1994. On the relationship between FAPAR and NDVI. Remote Sens. Environ. 49 (3), 200–211. http://dx.doi.org/10.1016/0034-4257(94)90016-7.

Nakano, H., Makino, A., Mae, T., 1995. Effects of panicle removal on the photosynthetic characteristics of the flag leaf of rice plants during the ripening stage. Plant Cell Physiol. 36 (4), 653–659. http://dx.doi.org/10.1093/oxfordjournals.pcp.a078806.

Nguyen, L.H., Robinson, S., Galpern, P., 2022. Medium-resolution multispectral satellite imagery in precision agriculture: mapping precision canola (Brassica napus L.) yield using Sentinel-2 time series. Precis. Agric. 23 (3), 1051–1071. http://dx.doi.org/10.1007/s11119-022-09874-7.

Paudel, D., de Wit, A., Boogaard, H., Marcos, D., Osinga, S., Athanasiadis, I.N., 2023. Interpretability of deep learning models for crop yield forecasting. Comput. Electron. Agric. 206, 107663. http://dx.doi.org/10.1016/j.compag.2023.107663.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12 (Oct), 2825–2830.

Potgieter, A.B., Schepen, A., Brider, J., Hammer, G.L., 2022. Lead time and skill of Australian wheat yield forecasts based on ENSO-analogue or GCM-derived seasonal climate forecasts – A comparative analysis. Agricult. Forest Meteorol. 324, 109116. http://dx.doi.org/10.1016/j.agrformet.2022.109116.

Rahman, M.M., Robson, A., 2020. Integrating Landsat-8 and Sentinel-2 time series data for yield prediction of sugarcane crops at the block level. Remote Sens. 12 (8), 1313. http://dx.doi.org/10.3390/rs12081313, Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.

Sabo, F., Meroni, M., Waldner, F., Rembold, F., 2023. Is deeper always better? Evaluating deep learning models for yield forecasting with small data. Environ. Monit. Assess. 195 (10), 1153. http://dx.doi.org/10.1007/s10661-023-11609-8.

Saito, K., Linquist, B., Keobualapha, B., Phanthaboon, K., Shiraiwa, T., Horie, T., 2006. Cropping intensity and rainfall effects on upland rice yields in northern Laos. Plant Soil 284 (1), 175–185. http://dx.doi.org/10.1007/s11104-006-0049-5.

Schlemmer, M., Gitelson, A., Schepers, J., Ferguson, R., Peng, Y., Shanahan, J., Rundquist, D., 2013. Remote estimation of nitrogen and chlorophyll contents in maize at leaf and canopy levels. Int. J. Appl. Earth Obs. Geoinf. 25, 47–54. http://dx.doi.org/10.1016/j.jag.2013.04.003.

Setiyono, T.D., Quicho, E.D., Holecz, F.H., Khan, N.I., Romuga, G., Maunahan, A., Garcia, C., Rala, A., Raviz, J., Collivignarelli, F., Gatti, L., Barbieri, M., Phuong, D.M., Minh, V.Q., Vo, Q.T., Intrman, A., Rakwatin, P., Sothy, M., Veasna, T., Pazhanivelan, S., Mabalay, M.R.O., 2019. Rice yield estimation using synthetic aperture radar (SAR) and the ORYZA crop growth model: development and application of the system in South and South-east Asian countries. Int. J. Remote Sens. 40 (21), 8093–8124. http://dx.doi.org/10.1080/01431161.2018.1547457, Publisher: Taylor & Francis.

Shendryk, Y., Davy, R., Thorburn, P., 2021. Integrating satellite imagery and environmental data to predict field-level cane and sugar yields in Australia using machine learning. Field Crops Res. 260, 107984. http://dx.doi.org/10.1016/j.fcr.2020.107984.

Shimono, H., Hasegawa, T., Moriyama, M., Fujimura, S., Nagata, T., 2005. Modeling spikelet sterility induced by low temperature in rice. Agron. J. 97 (6), 1524–1536. http://dx.doi.org/10.2134/agronj2005.0043.

Shimono, H., Okada, M., Kanda, E., Arakawa, I., 2007. Low temperature-induced sterility in rice: Evidence for the effects of temperature before panicle initiation. Field Crops Res. 101 (2), 221–231. http://dx.doi.org/10.1016/j.fcr.2006.11.010.

Sivapalan, S., Batten, G., Goonetilleke, A., Kokot, S., 2007. Yield performance and adaptation of some Australian-grown rice varieties through multivariate analysis. Aust. J. Agric. Res. 58 (9), 874. http://dx.doi.org/10.1071/AR06357.

Skakun, S., Kalecinski, N.I., Brown, M.G.L., Johnson, D.M., Vermote, E.F., Roger, J.-C., Franch, B., 2021. Assessing within-field corn and soybean yield variability from WorldView-3, Planet, Sentinel-2, and Landsat 8 satellite imagery. Remote Sens. 13 (5), 872. http://dx.doi.org/10.3390/rs13050872, Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.

Skakun, S., Wevers, J., Brockmann, C., Doxani, G., Aleksandrov, M., Batič, M., Frantz, D., Gascon, F., Gómez-Chova, L., Hagolle, O., López-Puigdollers, D., Louis, J., Lubej, M., Mateo-García, G., Osman, J., Peressutti, D., Pflug, B., Puc, J., Richter, R., Roger, J.-C., Scaramuzza, P., Vermote, E., Vesel, N., Zupanc, A., Žust, L., 2022. Cloud Mask Intercomparison eXercise (CMIX): An evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2. Remote Sens. Environ. 274, 112990. http://dx.doi.org/10.1016/j.rse.2022.112990.

Soriano-González, J., Angelats, E., Martínez-Eixarch, M., Alcaraz, C., 2022. Monitoring rice crop and yield estimation with Sentinel-2 data. Field Crops Res. 281, 108507. http://dx.doi.org/10.1016/j.fcr.2022.108507.

Tappi, M., Carucci, F., Gagliardi, A., Gatta, G., Giuliani, M.M., Santeramo, F.G., 2023. Crop varieties, phenological phases and the yield-weather relationship: evidence from the Italian durum wheat production. Bio-Based Appl. Econ. http://dx.doi.org/10.36253/bae-13745.

Varinderpal-Singh, Kunal, Kaur, R., Mehtab-Singh, Mohkam-Singh, Harpreet-Singh, Bijay-Singh, 2021. Prediction of grain yield and nitrogen uptake by basmati rice through in-season proximal sensing with a canopy reflectance sensor. Precis. Agric. http://dx.doi.org/10.1007/s11119-021-09857-0.

Von Bloh, M., Nóia Júnior, R.D.S., Wangerpohl, X., Saltık, A.O., Haller, V., Kaiser, L., Asseng, S., 2023. Machine learning for soybean yield forecasting in Brazil. Agricult. Forest Meteorol. 341, 109670. http://dx.doi.org/10.1016/j.agrformet.2023.109670.

Wan, L., Cen, H., Zhu, J., Zhang, J., Zhu, Y., Sun, D., Du, X., Zhai, L., Weng, H., Li, Y., Li, X., Bao, Y., Shou, J., He, Y., 2020. Grain yield prediction of rice using multi-temporal UAV-based RGB and multispectral images and model transfer – A case study of small farmlands in the South of China. Agricult. Forest Meteorol. 291, 108096. http://dx.doi.org/10.1016/j.agrformet.2020.108096.

Wang, J., Chen, J., Zhang, J., Yang, S., Zhang, S., Bai, Y., Xu, R., 2023a. Consistency and uncertainty of remote sensing-based approaches for regional yield gap estimation: A comprehensive assessment of process-based and data-driven models. Field Crops Res. 302, 109088. http://dx.doi.org/10.1016/j.fcr.2023.109088.

Wang, S., Guan, K., Zhang, C., Jiang, C., Zhou, Q., Li, K., Qin, Z., Ainsworth, E.A., He, J., Wu, J., Schaefer, D., Gentry, L.E., Margenot, A.J., Herzberger, L., 2023b. Airborne hyperspectral imaging of cover crops through radiative transfer process-guided machine learning. Remote Sens. Environ. 285, 113386. http://dx.doi.org/10.1016/j.rse.2022.113386.

Ward, R., Brickhill, H., Bull, N., Dunn, B., Dunn, T., Fowler, J., Hart, J., Mauger, T., 2021. Rice Growing Guide 2021, 2nd ed. NSW Department of Primary Industries, URL https://www.dpi.nsw.gov.au/_data/assets/pdf_file/0004/1361173/RGG-2021-web-final-26Oct2021.pdf.

Wei, G., Chen, H., Lin, E., Hu, X., Xie, H., Cui, Y., Luo, Y., 2023. Identification of water layer presence in paddy fields using UAV-based visible and thermal infrared imagery. Agronomy 13 (7), 1932. http://dx.doi.org/10.3390/agronomy13071932, Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.

Williams, R., Angus, J., 1994. Deep floodwater protects high-nitrogen rice crops from low-temperature damage. Aust. J. Exp. Agric. 34 (7), 927. http://dx.doi.org/10.1071/EA9940927.

Wolters, S., Söderström, M., Piikki, K., Reese, H., Stenberg, M., 2021. Upscaling proximal sensor N-uptake predictions in winter wheat (Triticum aestivum L.) with Sentinel-2 satellite data for use in a decision support system. Precis. Agric. 22 (4), 1263–1283. http://dx.doi.org/10.1007/s11119-020-09783-7.

Yu, W., Yang, G., Li, D., Zheng, H., Yao, X., Zhu, Y., Cao, W., Qiu, L., Cheng, T., 2023. Improved prediction of rice yield at field and county levels by synergistic use of SAR, optical and meteorological data. Agricult. Forest Meteorol. 342, 109729. http://dx.doi.org/10.1016/j.agrformet.2023.109729.

Zeng, Y., Hao, D., Huete, A., Dechant, B., Berry, J., Chen, J.M., Joiner, J., Frankenberg, C., Bond-Lamberty, B., Ryu, Y., Xiao, J., Asrar, G.R., Chen, M., 2022. Optical vegetation indices for monitoring terrestrial ecosystems globally. Nat. Rev. Earth Environ. 3 (7), 477–493. http://dx.doi.org/10.1038/s43017-022-00298-5.

Zhang, K., Ge, X., Shen, P., Li, W., Liu, X., Cao, Q., Zhu, Y., Cao, W., Tian, Y., 2019. Predicting rice grain yield based on dynamic changes in vegetation indexes during early to mid-growth stages. Remote Sens. 11 (4), 387. http://dx.doi.org/10.3390/rs11040387.

Zhou, S., Xu, L., Chen, N., 2023. Rice yield prediction in hubei province based on deep learning and the effect of spatial heterogeneity. Remote Sens. 15 (5), 1361. http://dx.doi.org/10.3390/rs15051361, Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.

Zhou, X., Zheng, H., Xu, X., He, J., Ge, X., Yao, X., Cheng, T., Zhu, Y., Cao, W., Tian, Y., 2017. Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. ISPRS J. Photogramm. Remote Sens. 130, 246–255. http://dx.doi.org/10.1016/j.isprsjprs.2017.05.003.