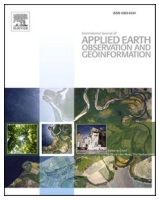
International Journal of Applied Earth Observations and Geoinformation 105 (2021) 102627

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

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



Rice nitrogen status detection using commercial-scale imagery

James Brinkhoff [a,*], Brian W. Dunn [b], Andrew J. Robson [a]

[a] Applied Agricultural Remote Sensing Centre, University of New England, Armidale, 2350 NSW, Australia
[b] New South Wales Department of Primary Industries, Yanco, 2198 NSW, Australia

ARTICLE INFO

Keywords:
Remote sensing
Precision Agriculture
Statistical Modeling
Machine Learning
Rice
Nitrogen

ABSTRACT

Determining the mid-season nitrogen status of rice is important for precision application of fertilizer to optimize productivity. While there has been much research aimed at developing remote-sensing-based models to predict the nitrogen status of rice, this has been predominantly limited to scientific small plot trials, relying on experts performing radiometric calibrations, encompassing limited cultivars, seasons and locations, and uniform management practices. As such, there has been little testing of models at commercial scale, against the range of conditions encountered across entire growing regions. To fill this gap, this work brings together four years of data, from both experimental replicated plot trials (38 datasets with 1734 observations) and commercial farms (12 datasets with 106 observations). Using commercial scale imagery acquired from airplanes, a number of nitrogen uptake modeling methodologies were evaluated. Universal single vegetation index based linear regression models had prediction root mean squared error (RMSE) of more than 45 kg/ha when tested at the 12 commercial sites. Machine learning models using multiple remote sensing features were able to improve predictions somewhat (RMSE $>$ 30 kg/ha). Practically useful accuracies were achieved after using three local field samples to calibrate models to each field image. The prediction RMSE using this methodology was 22.9 kg/ha, or 19.4%. This approach enables provision of optimal variable-rate mid-season rice fertilizer prescriptions to growers, while motivating continued research towards development of methods that reduce requirement of local sampling.

1. Introduction

Precision agriculture aims to apply the correct amount of crop inputs at the correct time in the correct location to optimize productivity (Mulla, 2013). One of the most important inputs is nitrogen (N) fertilizer, which is a key driver of rice yield and quality (Lee and Lee, 2012). If the N supplied is insufficient to meet crop demand, rice yield potential is not achieved. Too much N can lead to reduction in yield due to lodging and cold-induced sterility (Dunn et al., 2016b), and cause pollution through volatization and leaching (Bacenetti et al., 2020). Typically, a base N rate is applied at the start of the growing season. Mid season deficiencies in the plant uptake of N become evident, often with spatially variable N uptake within fields. Therefore, methodologies that support optimal mid-season variable-rate N application promise to improve yields, reduce spatial variability (Nutini et al., 2021; Simmonds et al., 2013) and minimize negative environmental impacts resulting from over-application (Peng et al., 2010).

Precision application of mid-season N requires two pieces of information: the actual plant N uptake, and the amount of additional N to apply in order to achieve optimal production. The second piece of information can be guided by knowledge of the critical N levels required to achieve yield potential, and N usage efficiency (Ata-Ul-Karim et al., 2017). Alternatively, tables of N recommendation versus mid-season N-uptake can be used (Dunn, 2008). Such tables are developed by conducting field trials with varying levels of base N, applying a number of mid-season N rates, and selecting the optimal application rate versus N uptake that maximizes yield (Dunn et al., 2016b). These methods can be used to transform spatial N uptake estimation maps into variable-rate N prescription maps for farmers to use in precision fertilizer application equipment.

The first piece of information (actual mid-season plant N uptake) can be obtained using various methods. The most accurate way is through destructive sampling of rice plants followed by laboratory analysis (Ata-Ul-Karim et al., 2017). This however is labour intensive, costly, and provides limited information on the spatial variability of N uptake (Berger et al., 2020). Numerous studies have shown relationships

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

https://doi.org/10.1016/j.jag.2021.102627
Received 19 September 2021; Received in revised form 11 November 2021; Accepted 12 November 2021
Available online 27 November 2021

between plant chlorophyll concentration, nitrogen content and optical reflectance (Gitelson et al., 2003; Schlemmer et al., 2013; Dunn et al., 2016a). Therefore, another way is using hand-held spectral sensors, with models predicting the relationship between sensor data and actual plant N uptake (Yao et al., 2012). This can be performed much more quickly than plant sampling, however still only provides information at a limited number of point source locations. Remote sensing promises an alternative method, allowing complete spatial coverage of fields at fine resolution (Inoue et al., 2012). Models are needed to predict actual N uptake from remote sensing reflectance data, which has been the subject of much research (Berger et al., 2020).

These remote sensing based approaches have utilised a variety of data sources, including hyperspectral (Inoue et al., 2012), multispectral (Brinkhoff et al., 2019), and visible (Shi et al., 2021) imagery. Images have been acquired from satellite (Huang et al., 2017), aerial (Ryu et al., 2011) and unmanned aerial vehicle (UAV) (Brinkhoff et al., 2020) platforms. The vast majority have considered experimental field trials with replicated N treatments across multiple plots. These have shown that remote sensing can describe a high degree of the variability of N uptake (R^2 is often above 0.9), perhaps leading to an optimistic view of the status of current techniques and technology to deliver reliable prediction accuracy ready for industry adoption.

However, a limitation of past research is a general lack of validation and extension at commercial farming operations, with few exceptions (Nutini et al., 2021). The use of UAVs for scientific research trials using replicated small plots to develop fundamental remote sensing techniques is appropriate. However the ability to deploy UAVs over large areas of broad acre production is limited due to flight time, speed and altitude constraints, and line-of-sight requirements. Models developed these trials often specify precisely controlled and calibrated image acquisition campaigns. Thus, they have limited exposure to variable: image acquisition time (Brinkhoff et al., 2020), solar-sensor geometry interaction with plant row orientation (Li et al., 2020), growth phase (Huang et al., 2017), variety (Ata-Ul-Karim et al., 2016), management strategies, seasonal differences (Ryu et al., 2011), soil characteristics (Wang et al., 2021), quality of atmospheric correction (Houborg and Boegh, 2008) and surface anisotropy (Li et al., 2018). These confounding factors limit success of implementing the developed tools and techniques at the commercial production level (Colaço and Bramley, 2019).

Whilst the ultimate goal is to develop a universal model, that does not need any local calibration (Houborg and Boegh, 2008), there is still much work needed to achieve this for operational rice N management. In the meantime, adequately accurate and practical methods (that may require some field sampling for local site calibration) are required to achieve acceptable accuracies at commercial scales, and thereby support industry adoption.

Currently, Australian rice grower standard practice is to apply mid-season N at the panicle initiation (PI) growth phase to address deficiencies (Dunn et al., 2016b). To determine the appropriate application rate, growers are recommended to gather nine biomass samples per field, three in each of low, medium and high areas based on imagery or soil zones (Dunn, 2012). These samples are analyzed in a laboratory to determine average field N uptake in kg/ha. From this, a uniform field N application is recommended (Batten et al., 1991). The limitations of this are (i) the three samples may not give an accurate representation of the average field N uptake and (ii) there is limited information about spatial variability, so precision variable application recommendations are not provided.

We aim to develop a process for providing timely and accurate N uptake maps at the PI growth stage, which can provide coverage of the majority of the Australian rice growing industry. The maps should allow assessment of in-field variability and facilitate variable rate nitrogen prescriptions. We used commercially provided multispectral aerial imagery collected at scale (10,000 ha in 2021). Our methodology investigates effects of imperfect image calibration, suggestion ways of providing practically useful prediction accuracy.

2. Methods

2.1. Sites and experiments

As a temperate growing region, rice in Australia is grown once per year (McDonald, 1994). Representative soil and weather summaries are described in (Dunn et al., 2020). Rice is typically sown in October. Basal nitrogen fertilizer is applied before ponding water, usually in late November. However, some growers have adopted delayed permanent water (DPW) (Dunn and Gaydon, 2011), which delays ponding until late December, and has the potential to improve water use efficiency (WUE). The panicle initiation (PI) growth phase occurs late December to mid January (Darbyshire et al., 2019), and additional nitrogen is top dressed at this time in order to maximize yield potential and minimize spatial variability (Dunn et al., 2016b). Crops are harvested in April–May.

We used data from four years, three experimental sites and twelve commercial fields (Fig. 1), covering the Murrumbidgee, Coleambally and Murray Valley irrigation areas, where the majority of rice is grown in Australia (McDonald, 1994).

2.1.1. Nitrogen rate experiments

The nitrogen rate experiments at sites E1, E2 and E3 (Fig. 1) were run over four years (harvested in 2018, 2019, 2020 and 2021). All experiments are listed in Table 1. The experiments consisted of replicated plots laid out in randomized complete block designs, each including three rice varieties and multiple base nitrogen rates. Individual plot areas ranged from 50–80 m^2. Experiments included conventional, delayed permanent water (DPW) and late sown management practices. The 2021 experiments at E1 included 54 conventional management plots, and 36 DPW plots.

For each experiment, all plots were sampled at least once, close to PI. Samples included all above ground biomass within 6 rows by 1 meter. Biomass and plant nitrogen concentration were extracted using methods described in (Dunn et al., 2016b). Multispectral imagery was acquired from airplanes close to the respective sampling dates (dates provided in Table 1). In some cases, the experiments were sampled and imaged on multiple dates to characterize temporal variability, especially in the 2021 season.

2.1.2. Commercial fields

Twelve commercial fields were sampled and imaged close to their respective PI dates (Fig. 1), 4 in 2019 (sites 01–04), 3 in 2020 (05–07) and 5 in 2021 (08–12). Nine samples were taken from each of these fields. These were distributed among low, medium and high vigour zones (which were derived from NDVI, which is related to leaf area index (Houborg and Boegh, 2008), acquired from Sentinel-2 or aerial imagery). Biomass and N concentration determined for each sample (using the methods described above). Each sample included biomass within a 0.2 m^2 ring. Two of the twelve sites had missing data for one of their samples (due to location recording error and abnormal N concentration lab test result), so those two samples were discarded (leaving eight samples for those two sites).

In the 2021 harvest season, approximately 10,000 hectares of commercial rice fields were imaged in New South Wales, Australia, close to the predicted PI date of each field. PI dates were predicted using the growth degree accumulation models described in (Darbyshire et al., 2019). The data was processed to provide spatial N uptake and N topdressing recommendation maps to growers using methods described in following sections.

2.2. Remote sensing data

A commercial provider (Ceres Imaging) acquired imagery of the experiment and commercial sites from airplanes on the image dates

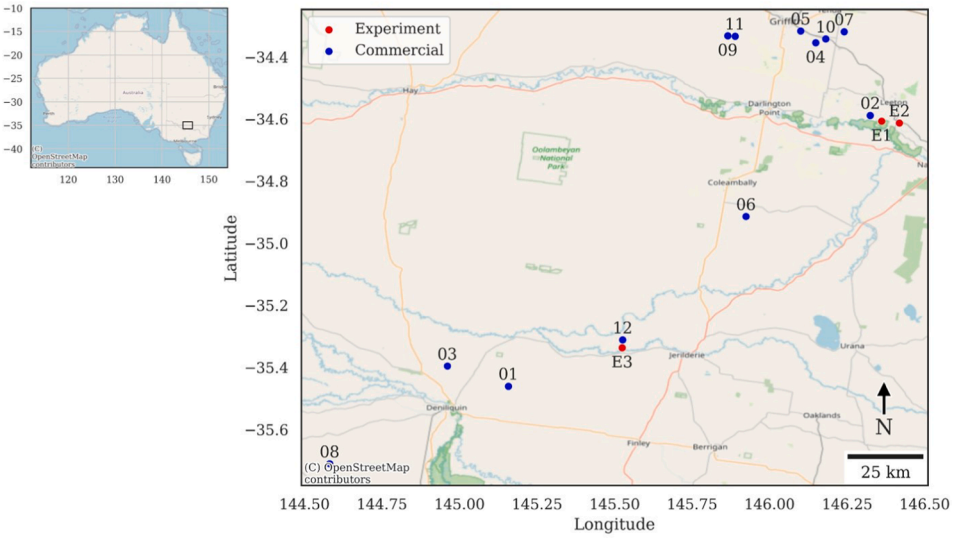


Fig. 1. Map showing the location of the study area within Australia (left) and experiment (E1-E3) and commercial (01–12) sites (right).

Table 1
Details of nitrogen rate experiments.

Harvest year	Site	Experiment	Sowing date	Ponding date	Nitrogen rates (kg/ha)	Plots	Image date	Sample date
2018	E1	Conventional	2017-10-30	2017-11-30	0, 60, 120, 180, 240, 300	54	2018-01-10	2018-01-12
		Late sown	2017-11-15	2017-12-21	0, 60, 120, 180, 240, 300	54	2018-01-10	2018-01-15
	E2	Conventional	2017-10-27	2017-12-01	0, 60, 120, 180	36	2018-01-10	2018-01-12
2019	E3	DPW	2018-10-10	2018-12-21	0, 60, 120, 180	36	2018-12-27 2019-01-06	2018-12-28 2019-01-07
		Conventional	2018-10-29	2018-12-08	0, 60, 120, 180, 240	45	2018-12-27 2019-01-06	2018-12-28 2019-01-07
		Late sown	2018-11-05	2018-12-21	0, 60, 120, 180	24	2019-01-06	2019-01-07
2020	E1	DPW	2019-10-11	2019-12-20	0, 60, 120, 180	36	2019-12-27	2020-01-01
		Conventional	2019-10-23	2019-11-21	0, 60, 120, 180, 240, 300	54	2019-12-27	2019-12-27
2021	E1	DPW	2020-10-14	2020-12-23	0, 60, 120, 180	36	2021-01-05 10:12 2021-01-05 11:36 2021-01-06 10:49 2021-01-06 11:53 2021-01-07 11:11 2021-01-07 14:15 2021-01-08 11:23 2021-01-08 11:55 2021-01-09 13:03	2021-01-07
							2021-01-11 11:38 2021-01-13 11:08	2021-01-14
							2020-12-19 11:08 2020-12-19 15:57	2020-12-18
							2020-12-23 11:31	2020-12-23
							2020-12-28 11:41 2020-12-28 14:49 2020-12-29 14:12	2020-12-28
		Conventional	2020-10-24	2020-11-25	0, 60, 120, 180, 240, 300	54	2021-01-05 10:12 2021-01-05 11:36 2021-01-06 10:49 2021-01-06 11:53 2021-01-07 11:11 2021-01-07 14:15 2021-01-08 11:23 2021-01-08 11:55 2021-01-09 13:03	2021-01-07
							2021-01-11 11:38 2021-01-13 11:08	2021-01-13

shown in Table 1. Resolution ranged from 0.2–0.8 m. The sensor bands are given in Table 2. The band wavelengths were consistent over years, except for the red edge (RE) band as shown. Methods to deal with the inconsistent RE band are described below. The imagery was provided as surface reflectance multispectral GeoTIFFs.

All image processing was done in Google Earth Engine (Gorelick et al., 2017). Vegetation indices (VIs) that have been shown to have strong relationships with plant N were computed from the reflectance bands. These included the normalized difference red edge (NDRE) = (NIR-RE)/(NIR + RE) (Brinkhoff et al., 2019), chlorophyll index red edge CIre = NIR/RE and chlorophyll index green CIg = NIR/G (Schlemmer et al., 2013; Gitelson et al., 2003). Note that Gitelson et al. (2003) proposed subtracting 1 from CIre and CIg, as they found this zeroed the y-intercept for the relationship between chlorophyll content and the respective VIs. This was not the case for the relationship between N uptake and the respective VIs in our datasets, so we omitted the –1.

In order to make the red edge data consistent through the years (Table 2), we applied calibrations to the RE 700 nm data. To do this, we used 2020 data, regressing the coincident 700 nm against the 717 nm results, for all 90 experiment plots. NDRE and CIre using 700 nm vs 717 nm were highly correlated, so we applied a transformation directly to these VIs:

$$\begin{aligned} \mathrm{NDRE}_{717nm} &= 0.84 \times \mathrm{NDRE}_{700nm} - 0.02 \quad (R^2 = 0.98) \\ \mathrm{CIre}_{717nm} &= 0.64 \times \mathrm{CIre}_{700nm} + 0.43 \quad (R^2 = 0.98) \end{aligned}$$

The transformations were applied to NDRE and CIre from 2018 and 2019, resulting in 717 nm equivalent data in all years.

All image dates for the experiment sites are provided in Table 1. Each of these flights also acquired imagery of many commercial fields. Multiple image acquisitions occurred on some days in 2021, because multiple flights were required to cover the area of the commercial fields. Image acquisition time was only supplied in 2021, but generally images were acquired between 10am and 3 pm, avoiding solar noon where specular reflection off the ponded water surface can saturate the multispectral sensor.

For each sample location, the mean pixel values of the VIs were calculated. The remote sensing data was merged with the plant sample data from the closest date (Table 1). This yields 38 datasets (one dataset per experiment image), with a combined total of 1734 observations, allowing detailed investigation of the variability of the models extracted from each of these datasets.

2.3. Model calibration

We assessed the correlation between N uptake and all possible 2-band normalized difference spectral indices and ratio spectral indices, finding that N uptake is most strongly related to indices combining NIR and G bands, and NIR and RE bands (Inoue et al., 2012; Brinkhoff et al., 2019). Therefore, we assessed four formulae that have been used to model N uptake (NU) in previous studies:

$$\mathrm{NU} = \alpha \times e^{\beta \times \mathrm{NDRE}} \ (\text{Brinkhoff et al., 2020}) \tag{1}$$

$$\ln(\mathrm{NU}) = \ln(\alpha) + \beta \times \mathrm{NDRE}$$

Table 2
Aerial imagery band wavelengths.

	Wavelength (nm)		
Year	**2018–2019**	**2020**	**2021**
Green (G)	500	500	500
Red (R)	670	670	670
Red edge (RE)	700	700 & 717	717
Near infrared (NIR)	800	800	800

$$\mathrm{NU} = \alpha + \beta \times \mathrm{NDRE}^2 \ (\text{Brinkhoff et al., 2019}) \tag{2}$$

$$\mathrm{NU} = \alpha + \beta \times \mathrm{CIre} \ (\text{Schlemmer et al., 2013}) \tag{3}$$

$$\mathrm{NU} = \alpha + \beta \times \mathrm{CIg} \ (\text{Schlemmer et al., 2013}) \tag{4}$$

The α and β coefficients were extracted for each dataset in Table 1 using ordinary least squares (OLS). The α coefficient is the intercept and β the slope wrt the VI for each model. Eq. (1) was transformed into a linear formulation as shown, so OLS could be used.

We assessed how well each calibrated model described the variability of plant N uptake in the respective dataset using the coefficient of determination:

$$R^2 = 1 - (\text{residual sum of squares})/(\text{total sum of squares}) \tag{5}$$

Note that this definition of R^2 (also known as the Nash–Sutcliffe model efficiency coefficient) is negative if predictions are worse than simply predicting the mean of the observations. Prediction error was evaluated using root mean squared error:

$$\mathrm{RMSE} = \sqrt{(\text{residual sum of squares}/\mathrm{N})} \tag{6}$$

2.4. Model variability assessment

A key question of this study is how remote sensing based N uptake predictions can be generated accurately in operational contexts, over large areas, multiple times, dates and diverse agronomic practices. We assessed the variability of the coefficients of the N uptake models between the 38 datasets from the experimental sites by examining the distributions of α and β for each formula (1)–(4) over all the datasets (Table 1). The normalized inter-quartile range of the model coefficients were calculated and expressed as a percentage:

$$\mathrm{nIQR} = (75\text{th percentile} - 25\text{th percentile})/\text{median} \times 100\% \tag{7}$$

2.5. Predictive model training and testing methodologies

We tested three methods of extracting models to generate N uptake predictions across experiment and commercial site (Fig. 2).

Method A uses a single experiment dataset to calibrate models, then tests predictions on remaining experiment datasets and commercial sites, repeating for all training datasets. This simulates the situation where an N uptake prediction model is calibrated from an image of a replicated N rate trial, which is then applied to other sites and images.

Method B trains on data from all experiment datasets except one, then tests on this remaining dataset (or trains on all experiment datasets when testing on commercial datasets). Machine learning algorithms including ridge, lasso, random forest and support vector regression (Pedregosa et al., 2011) are used to make use of the extensive training data. The 20 features included all reflectance bands (4: G, R, RE, NIR), all reflectance ratios (12: G/R, R/G, G/RE, RE/G etc.) and normalized difference vegetation indices (4: NDRE, NDVI, GNDVI, GRVI). Algorithm hyperparameters were tuned using leave one group out cross validation procedures (Brinkhoff et al., 2019). This method assesses achievable accuracy when a very comprehensive training dataset is used, encompassing multiple sites, seasons and images.

Method C takes a different approach. Instead of generalizing predictions using models generated from other sites/images, it uses local site-specific calibration. A limited number of samples from each site were used to train models, then predictions are tested against the remaining samples from that site.

2.6. Nitrogen application recommendation maps

N uptake prediction maps were generated by applying the respective

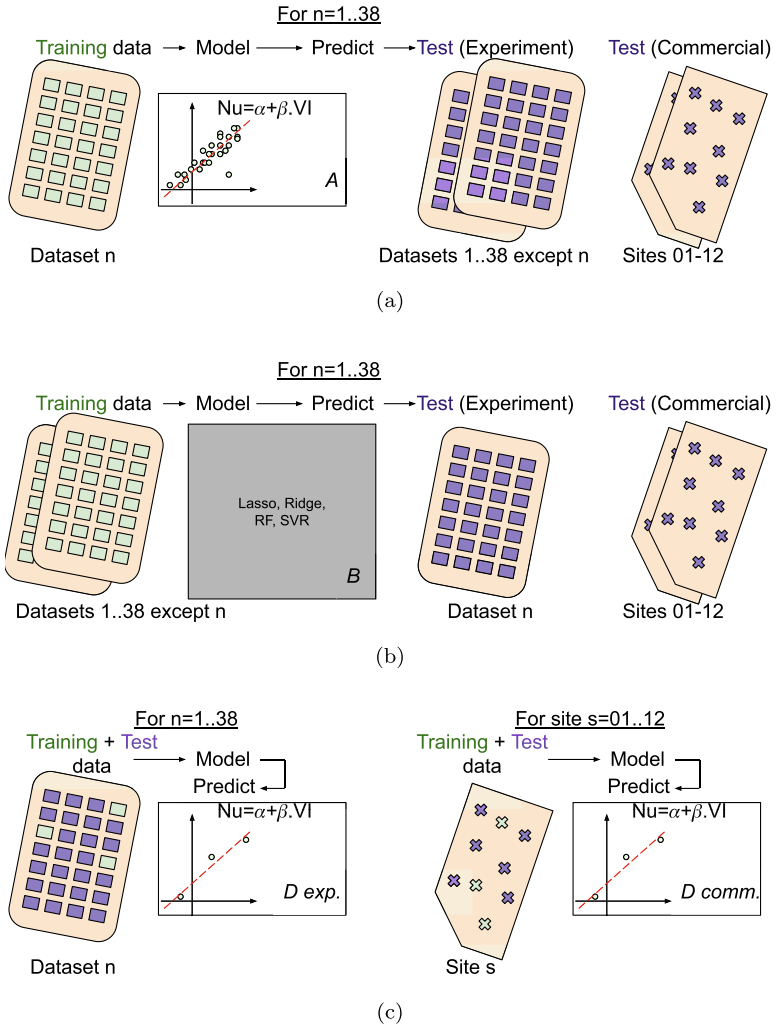


Fig. 2. Three methods of training and testing models. Method A (a) trains on one experiment dataset, and tests on all other experiment and commercial datasets. Method B (b) trains on all experiment datasets (except the test dataset), using machine learning models with 20 input features. Method C (c), trains on 3 local samples, and predictions are tested on remaining local samples.

model formulae (1)–(4) to every pixel of the experiment and commercial images. Variable-rate N recommendation maps were generated by applying tables that map N uptake to optimal N application rates. These tables are described in Dunn (2008). This process resulted in high resolution N recommendation maps. In order to provide variable rate maps that are practical to deliver with current fertilizer spreading technology,

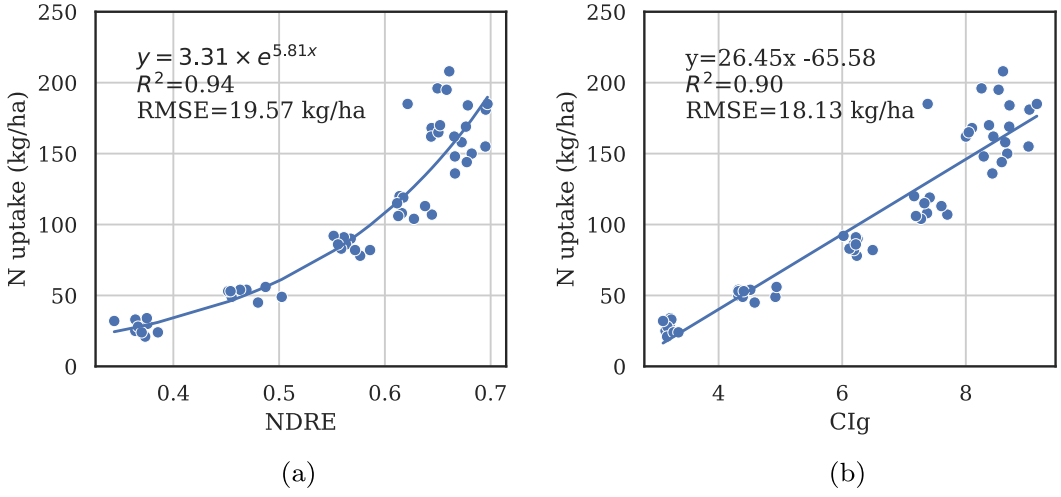


Fig. 3. Regression results for the 2021 E1 dataset (54 samples, sample date 2020–12-28 and image date 2020–12-28 11:41) for formulae (1) and (4).

a neighborhood mode filter was applied, with a radius of 30 m and 3 iterations.

3. Results

We first examine N-uptake model variability from the experiment datasets. Then prediction accuracies are assessed for the three modeling strategies at both experiment and commercial sites.

The four model formulae (1)–(4) were fit to each of the 38 datasets (Table 1). Fig. 3 shows an example fit for NDRE (1) and CIg (4), for the 2021 E1 conventional water management experiment, for a 2020–12-28 dataset. The formulae fit the characteristics of the N uptake well. There is some heteroskedasticity, with more scatter at very high N uptake values. However, we note that for practical applications, accuracy above 150 kg/ha is of lesser importance, as at that level the plants have sufficient N, so the fertilizer prescription in those areas would be 0 kg/ha (Dunn et al., 2016b).

For all 38 datasets, and 4 formulae, the β slope coefficient was significant at $p < 0.001$. The intercept parameter α was significant at $p < 0.05$ for formulae (3) and (4) for all datasets. For formula (1), α was significant at $p < 0.05$ for all but 7 of the datasets, and for formula (2) α was significant for all but 2 of the datasets.

A summary for the model fit results over the 38 experiment datasets (Table 3) shows all models have median R^2 better than 0.83, with the exponential NDRE model (1) having the highest value of 0.92. The variability of R^2 is reasonably constrained, with nIQR between 6.3–11.2 % (i.e. the inter-quartile range is less than 11.2 % of the median R^2). The median RMSE for the 4 formulae is between 19.9–23.8 kg/ha.

For practical applications, variability between models is important, particularly if the goal is to extract a model from one or more images of experiments and then apply it to commercial farms. The first two formulae (1–2) have very high variability in the intercept parameter α (nIQR>60%, Table 3), while the slope β is somewhat constrained (nIQR<20%). The CIre and CIg formulae (3–4) produce models with neither coefficient having low variability. The scatter of coefficients for the 38 datasets is shown in Fig. 4. These results suggest models from one image will not generalize to other images well.

3.1. Method A: Applying regression models extracted from single experiment datasets

3.1.1. Testing on experiment datasets

The first test of model scalability was to use the model from one experiment dataset (training), and test accuracy predicting N uptake for all other datasets (Fig. 2a). This simulates the situation where a single N-rate trial is used to develop a model, which is then used to supply prediction for other sites and/or dates. This assessment was performed 38 times (once for each dataset in Table 1 as model training data, testing model predictions on the 37 remaining datasets). The median RMSEs are very large, from 43.3–49.1 kg/ha for the four formulae (Fig. 5a). The CIg model (4) gave the best results both in terms of median and variability of RMSEs. If datasets from 2021 only are used, the median RMSE is between 31.4–34.9 kg/ha. However these values are still too high for practical use.

3.1.2. Testing on commercial datasets

To test this method on the commercial site data, the experiment model from the dataset closest in time to the image from each respective commercial site was used to predict the 8 or 9 N uptake samples for each of the 12 sites (total of 106 samples). The best results were obtained with formulae (2), with RMSE = 48 kg/ha (Fig. 5b).

3.2. Method B: Applying machine learning models extracted from many experiment images

Four ML algorithms were trained on the experiment datasets, using 20 remote-sensing based features (Fig. 2b).

3.2.1. Testing on experiment datasets

Models were trained using 37 of the experiment datasets, with predictions tested on the remaining dataset, repeating for the 38 test datasets. The best results were achieved with ridge regression (median RMSE = 27.6 kg/ha, Fig. 6a), followed by support vector regression. These results are significantly better than the simple linear regression models, although variability in prediction accuracy over the datasets is still large.

3.2.2. Testing on commercial datasets

The ML models were trained on all 38 experiment datasets. The models were then applied to predict N uptake of the 106 samples of the 12 commercial sites. The best RMSE of 32.1 kg/ha was obtained with support vector regression (Fig. 6b), followed by ridge (34.1 kg/ha), lasso (34.2 kg/ha) and random forest (37.8 kg/ha). If only 2021 training and testing data was used, the best test RMSE of 21.5 kg/ha was obtained using lasso.

The relative importance of reflectance, VI and ratio features was assessed. For SVR and ridge, the best results were obtained when only the 4 reflectance features were used (31.0 kg/ha for SVR). RF and lasso results were better when all 20 features were used.

3.3. Method C: Extracting a local regression model for each image

Next, we extracted model coefficients using a limited number of samples from each dataset, and predicted the remaining samples from that dataset (Fig. 2c). This is the situation where a local model is extracted from each site-image, requiring at least 2 local physical samples to determine α and β.

3.3.1. Testing on experiment datasets

The training samples from experiment plots were randomly selected 10 times for each of the 38 datasets, done separately with 2, 3 and 6 training samples. The resulting models were used to predict remaining plot samples in each dataset. The median RMSE and variance in RMSE declines as the number of training samples increases (Fig. 7a). For 3 training samples, the best results are obtained with the NDRE formulae (1), with median RMSE of 29.5 kg/ha, and 24.9 kg/ha with 6 training samples.

Because the training samples were randomly selected, some models were parameterized using samples with very small variability in sample VI values, others with large variability. The prediction errors are generally smaller if the model training samples encompass a range of VI

Table 3
Summary of model coefficients and accuracy metrics for the 38 experiment datasets.

	Median				nIQR (%)			
Formula	α	β	R^2	RMSE (kg/ha)	α	β	R^2	RMSE
(1)	2.21	6.48	0.92	21.10	96.52	19.19	7.82	24.60
(2)	−70.56	567.47	0.83	23.80	68.74	14.22	11.21	21.75
(3)	−104.64	54.86	0.85	20.99	27.31	31.89	10.81	24.50
(4)	−91.66	33.95	0.88	19.92	25.29	35.70	6.32	27.38

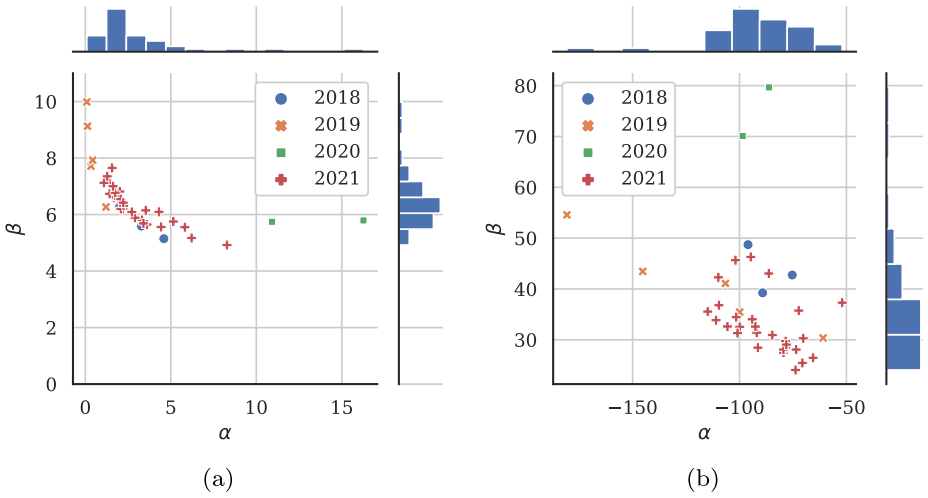


Fig. 4. Model coefficients for the exponential NDRE (1) and CIg (4) models extracted from all experiment datasets.

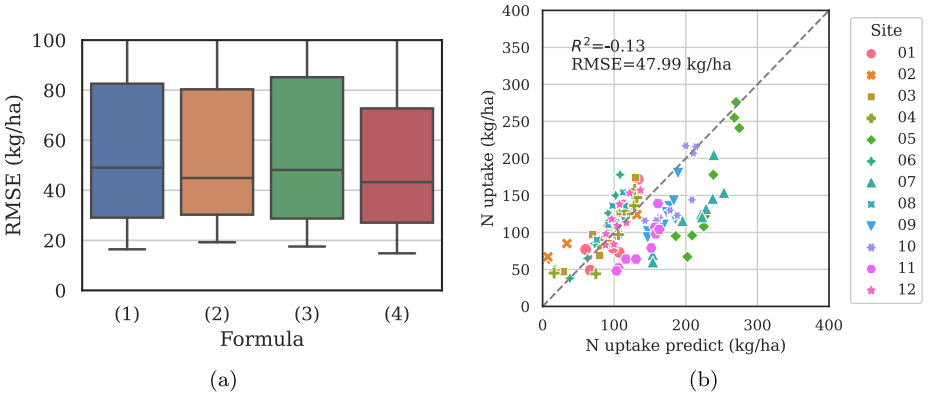


Fig. 5. Results of extracting a model from one experiment image (method A), and applying it to (a) all other experiment sites, and (b) commercial sites using formula (2).

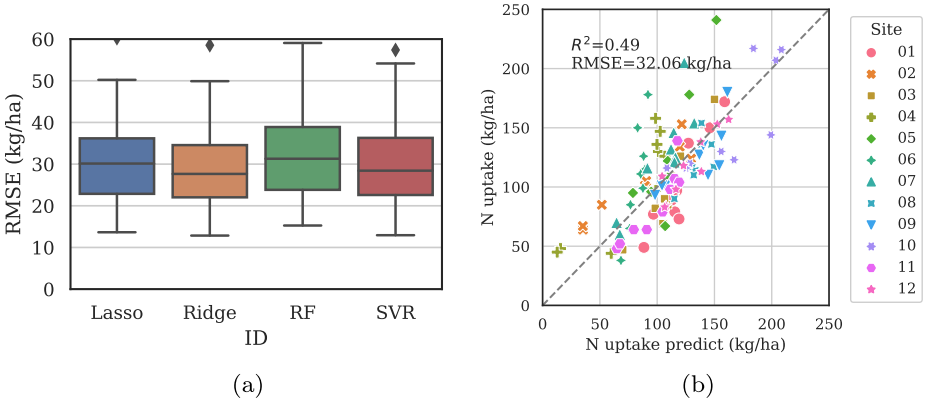


Fig. 6. Results of extracting ML models from many training experiment images (method B), and applying it to (a) test experiment images, and (b) commercial site images (support vector regression results shown).

values (Fig. 7b). This suggests using a model-based sample selection strategy similar to latin hypercube sampling, so that the model is trained on data that reflects the distribution of the data (Minasny and McBratney, 2006). This is important in practical cases, where very limited field samples are available at commercial sites.

Therefore, we ran another test, selecting the samples with minimum, median and maximum VI to train the model. The median RMSEs for the four formulae are lower using this sample selection strategy (Fig. 7c), between 22.9 kg/ha for (4) and 25.1 kg/ha for (2).

3.3.2. *Testing on commercial datasets*

The minimum-median-maximum sampling strategy was applied to each commercial site, and predictions were tested on the 5 or 6 remaining site samples. The best results were obtained using formulae (4) with RMSE = 22.9 kg/ha (Fig. 7d). This method was used to produce spatial N uptake and recommendation maps for the commercial sites

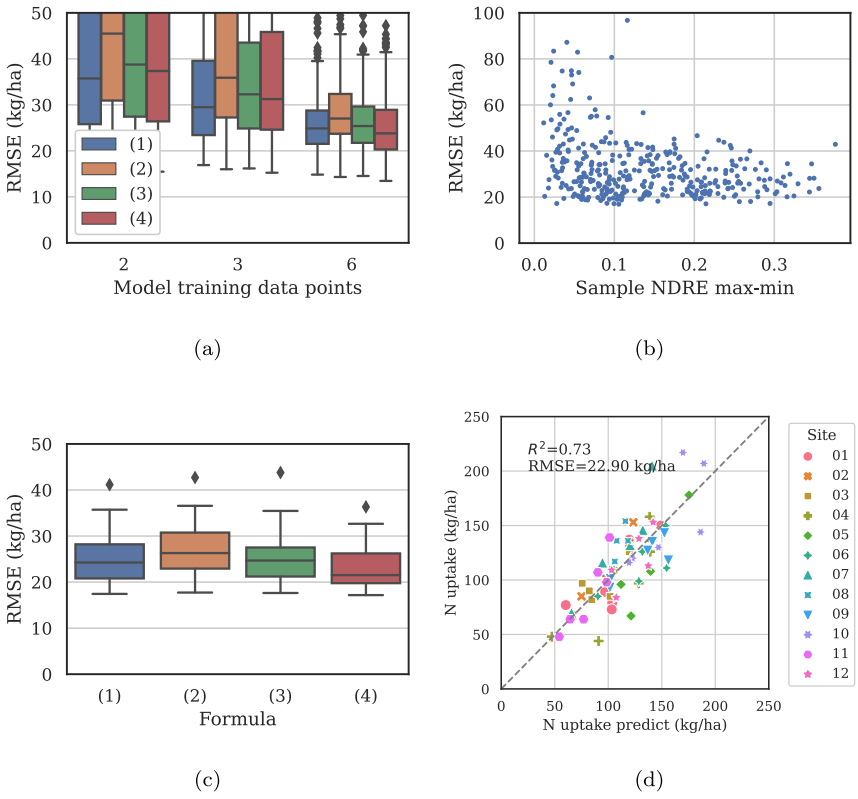


Fig. 7. Results of extracting a model from the local site and testing on remaining samples from the site (method C). For the experiments, (a) shows results with random selection of training samples, (b) shows how the distribution of the sample VI values affects prediction errors, (c) shows predictions when the 3 selected training samples are those with minimum, median and maximum VI. (d) shows the same sample selection as (c) for commercial sites, using model formula (4).

(examples shown in Fig. 8).

4. Discussion

We tested a variety of model calibration strategies to enable prediction of rice N uptake using remote sensing imagery. To generate PI N application maps, we used tables of recommended application of N (at the panicle initiation growth stage) relative to N uptake (Dunn, 2008). N uptake prediction errors are expected to be less than 30 kg/ha (near the practical limits of current aerial variable rate spreaders). However, the N uptake models were variable over multiple image acquisition times, dates, seasons and sites, rendering generalized model predictions inaccurate (though ML models trained on much data were better than regression models trained on one image). In the presence of this variability, three targeted local samples are able to provide predictions with sufficient accuracy.

We examined four model formulae and four ML algorithms (summary in Table 4). The $NDRE^2$ formula (2) was less accurate when fit to single experiment datasets (Table 3), but sometimes gave better predictions when scaling to new sites and images. The formulae using an exponential transformation of NDRE (1) gave reasonably consistent prediction accuracy, but has the drawback of requiring nonlinear transformation to fit linear models. The CIg (4) and CIre (3) formulae were linear with respect to N uptake (Schlemmer et al., 2013; Zheng et al., 2018). The CIg formula will be particularly useful for using satellite data where the red-edge band is not available, or is at a lower resolution (Cai et al., 2019). The CIg formula consistently gave the best results when models were trained with a 3 local samples (method C, Table 4). Model performance was better when only data from 2021 was used, possibly due to improvements in image calibration and elimination of seasonal differences. The 2020 model coefficients were quite different to those from other years (Fig. 4, which may be at least partly due to the presence of bushfire smoke causing errors in surface reflectance calibration.

The linear regression coefficients varied greatly across the 38 experiment site datasets, even for images taken of the same site on the same day. A probable cause is atmospheric correction and reflectance calibration imperfections, which is likely to be a factor in any large-scale image acquisition from aerial or UAV platforms given cost and logistic limitations. Atmospheric correction imperfections and view angle effects also affect retrieval of vegetation indices from satellite data (Zhang et al., 2018). Model coefficients depend on a variety of additional factors: image acquisition time (Brinkhoff et al., 2020), regional differences (Inoue et al., 2012), season and variety (Brinkhoff et al., 2019), growth stages (Zheng et al., 2018; Zhou et al., 2018), different view angles and surface anisotropy (Li et al., 2018), and the relationship between solar angle and plant row orientation (Li et al., 2020).

Given the variability in N prediction models, it seems unlikely that a universal model based on simple linear regression can be applied across varieties, management practices, images, sites and seasons. Machine learning (ML) methods using multiple variables produced more consistent models. Other authors have found similar benefits for ML models (Shi et al., 2021), and have also found that band reflectances are as good as vegetation indices for ML-based nitrogen prediction models (Peng et al., 2021). However, these methods did not provide predictions with sufficient accuracy over all sites and years for our data. Adding additional features to those derived from remote sensing data for ML models to be trained on (such as time of day, days after sowing, environmental conditions) may provide improved predictions. However, data from sufficient sites and seasons would be needed to provide sufficient coverage of the possible range of these variables to make them useful predictors. Our work motivates further studies on generalized prediction models using such data.

While generalizable models are still an area of research, we proposed using three plant samples from each field, from low, medium and high growth areas, to calibrate models to local sites. This provided useful

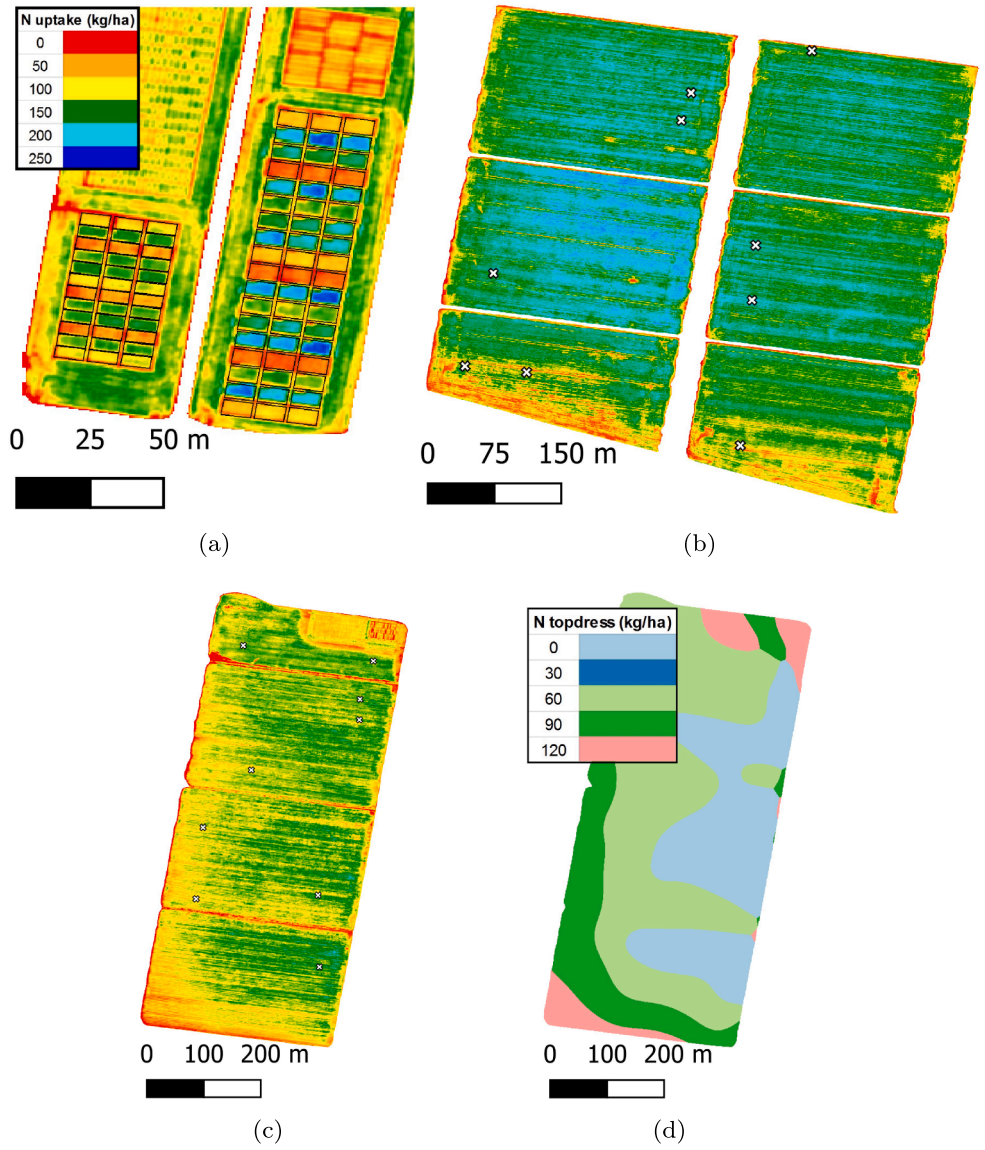


Fig. 8. Predicted N uptake for: (a) 2021 experiments at site E1, (b) commercial site 10, and (c) commercial site 09. (d) shows the PI N topdressing recommendation for commercial site 09, smoothed using a 30 m filter.

Table 4
Summary of results from the four methodologies in Fig. 2. The results shown are RMSE (kg/ha) with best formulae/algorithm in brackets.

Method	Years	Experiment	Commercial
A. Train on one experiment dataset	All	43.3 (4)	48.0 (2)
	2021	31.4 (4)	41.0 (2)
B. ML train on many experiment datasets	All	27.6 (Ridge)	34.1 (SVR)
	2021	26.1 (SVR)	21.5 (Lasso)
C. Train on 3 local samples	All	21.5 (4)	22.9 (4)
	2021	20.8 (4)	21.2 (4)

spatial N uptake prediction accuracy. However, this method has disadvantages. Sampling introduces delay in the process of generating prescription maps due to laboratory test time. Obtaining very accurate geolocation for grower supplied samples is difficult. Ultimately, growers desire variable-rate N recommendation maps that don't require any sampling.

Most previous studies on remote sensing for rice N management used data from scientific plot-scale experiments to assess predictions. A significant source of uncertainty in applying these methodologies to commercial sites is the relatively larger spatial variability in plant N at small scales observed at growers sites. To train and validate models in such situations, work will require (i) pixel sizes smaller than the spatial variability, (ii) precise sample and image co-location and/or (iii) selection of sampling areas with minimal spatial variability. Given these results, we recommend more studies should include assessment of N uptake prediction models on commercial scale fields to give realistic expectations of accuracy when these models are adopted by industry.

We used aerial imagery, as it provided the most practical trade-off between ability to cover the large area (10,000 ha in 2021) and sufficient spatial resolution to enable correlation of imagery with samples. UAVs could not currently cover such an area in a constrained timeframe (Weiss et al., 2020), and satellite data at sufficient resolution (< 2m) is currently too expensive. Due to the large rice area and the logistical challenges to acquire the imagery close to the predicted PI date of each field (Darbyshire et al., 2019), image acquisition times and atmospheric conditions could not be tightly controlled (as they can be in small-scale field trials). Scaling to medium-resolution satellite data is an interesting prospect, which will require consideration of the different resolution of pixels and plant samples (Cai et al., 2019) and reflectance calibration

(Zhang et al., 2018).

5. Conclusion

Using an extensive dataset, we have shown that commercial-scale multispectral imagery may not be sufficiently consistent to provide rice N uptake predictions with universal linear regression models based on remote sensing data. Predictions were not only tested on plot-scale field trials, but also at commercial rice operations, with a dataset encompassing four years, eight varieties and various sowing and water management strategies. We demonstrated that three targeted field samples (encompassing the range of field vegetation index variation) is able to improve N prediction accuracy to required levels (RMSE<25 kg/ha). This provides a practical way for remote-sensing based precision N applications to be adopted in practice, while motivating the continued development of generalizable models, which require ongoing research to encapsulate sources of variability and commercial scale validation.

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.

Acknowledgment

This research was funded by AgriFutures Australia, project number PRJ-011058, 'Improving mid-season nitrogen management of rice', and project number PRJ-009772, 'Moving forward with NIR and remote sensing'. The authors are grateful to growers who allowed trial work in their fields, and to for field work performed by Tina Dunn, Chris Dawe and Craig Hodges. The imagery was purchased by SunRice, who are facilitating industry adoption.

References

Ata-Ul-Karim, S.T., Cao, Q., Zhu, Y., Tang, L., Rehmani, M.I.A., Cao, W., 2016. Non-destructive Assessment of Plant Nitrogen Parameters Using Leaf Chlorophyll Measurements in Rice. Frontiers in Plant Science 7. https://doi.org/10.3389/fpls.2016.01829.

Ata-Ul-Karim, S.T., Liu, X., Lu, Z., Zheng, H., Cao, W., Zhu, Y., 2017. Estimation of nitrogen fertilizer requirement for rice crop using critical nitrogen dilution curve. Field Crops Research 201, 32–40. https://doi.org/10.1016/j.fcr.2016.10.009.

Bacenetti, J., Paleari, L., Tartarini, S., Vesely, F.M., Foi, M., Movedi, E., Ravasi, R.A., Bellopede, V., Durello, S., Ceravolo, C., Amicizia, F., Confalonieri, R., 2020. May smart technologies reduce the environmental impact of nitrogen fertilization? A case study for paddy rice. Science of The Total Environment 715, 136956. https://doi.org/10.1016/j.scitotenv.2020.136956.

Batten, G.D., Blakeney, A.B., Glennie-Holmes, M., Henry, R.J., McCaffery, A.C., Bacon, P. E., Heenan, D.P., 1991. Rapid determination of shoot nitrogen status in rice using near infrared reflectance spectroscopy. J. Sci. Food Agric. 54, 191–197. https://doi.org/10.1002/jsfa.2740540204.

Berger, K., Verrelst, J., Féret, J.B., Wang, Z., Wocher, M., Strathmann, M., Danner, M., Mauser, W., Hank, T., 2020. Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. Remote Sens. Environ. 242, 111758. https://doi.org/10.1016/j.rse.2020.111758.

Brinkhoff, J., Dunn, B.W., Hart, J., Dunn, T., 2020. Impact of UAV Time-of-Flight on Rice Nitrogen Uptake Models. In: IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium, pp. 4355–4358. https://doi.org/10.1109/IGARSS39084.2020.9323619.

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 Sensing 11, 1837. https://doi.org/10.3390/rs11151837.

Cai, Y., Guan, K., Nafziger, E., Chowdhary, G., Peng, B., Jin, Z., Wang, S., Wang, S., 2019. Detecting In-Season Crop Nitrogen Stress of Corn for Field Trials Using UAV- and CubeSat-Based Multispectral Sensing. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12, 5153–5166. https://doi.org/10.1109/JSTARS.2019.2953489.

Colaço, A.F., Bramley, R.G.V., 2019. Site-Year Characteristics Have a Critical Impact on Crop Sensor Calibrations for Nitrogen Recommendations. Agronomy Journal 111, 2047–2059. https://doi.org/10.2134/agronj2018.11.0726.

Darbyshire, R., Crean, E., Dunn, T., Dunn, B., 2019. Predicting panicle initiation timing in rice grown using water efficient systems. Field Crops Research 239, 159–164. https://doi.org/10.1016/j.fcr.2019.05.018.

Dunn, B., 2008. Improving topdressing recommendations for rice. URL https://agrifuturesrice.squarespace.com/s/Improving-topdressing-recommendations-for-rice.pdf.

Dunn, B., 2012. Nitrogen tissue test for rice panicle initiation. URL https://www.agrifutures.com.au/wp-content/uploads/publications/12-047.pdf.

Dunn, B.W., Dehaan, R., Schmidtke, L.M., Dunn, T.S., Meder, R., 2016a. Using Field-Derived Hyperspectral Reflectance Measurement to Identify the Essential Wavelengths for Predicting Nitrogen Uptake of Rice at Panicle Initiation. J. Near Infrared Spectrosc. 24, 473–483. https://doi.org/10.1255/jnirs.1246.

Dunn, B.W., Dunn, T.S., Mitchell, J.H., Brinkhoff, J., 2020. Effects of plant population and row spacing on grain yield of aerial-sown and drill-sown rice. Crop and Pasture Science 71, 219–228. https://doi.org/10.1071/CP19421.

Dunn, B.W., Dunn, T.S., Orchard, B.A., 2016b. Nitrogen rate and timing effects on growth and yield of drill-sown rice. Crop and Pasture Science 67, 1149–1157. https://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. Agric. Water Manag. 98, 1799–1807. https://doi.org/10.1016/j.agwat.2011.07.004.

Gitelson, A.A., Gritz, Y., Merzlyak, M.N., 2003. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. J. Plant Physiol. 160, 271–282. https://doi.org/10.1078/0176-1617-00887.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031.

Houborg, R., Boegh, E., 2008. Mapping leaf chlorophyll and leaf area index using inverse and forward canopy reflectance modeling and SPOT reflectance data. Remote Sens. Environ. 112, 186–202. https://doi.org/10.1016/j.rse.2007.04.012.

Huang, S., Miao, Y., Yuan, F., Gnyp, M.L., Yao, Y., Cao, Q., Wang, H., Lenz-Wiedemann, V.I.S., Bareth, G., 2017. Potential of RapidEye and WorldView-2 Satellite Data for Improving Rice Nitrogen Status Monitoring at Different Growth Stages. Remote Sensing 9, 227. https://doi.org/10.3390/rs9030227.

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. https://doi.org/10.1016/j.rse.2012.08.026.

Lee, K.J., Lee, B.W., 2012. Modeling for recommending panicle nitrogen topdressing rates for yield and milled-rice protein content. Journal of Crop Science and Biotechnology 15, 335–343. https://doi.org/10.1007/s12892-012-0117-8.

Li, D., Chen, J.M., Zhang, X., Yan, Y., Zhu, J., Zheng, H., Zhou, K., Yao, X., Tian, Y., Zhu, Y., Cheng, T., Cao, W., 2020. Improved estimation of leaf chlorophyll content of row crops from canopy reflectance spectra through minimizing canopy structural effects and optimizing off-noon observation time. Remote Sens. Environ. 248, 111985. https://doi.org/10.1016/j.rse.2020.111985.

Li, D., Zheng, H., Xu, X., Lu, N., Yao, X., Jiang, J., Wang, X., Tian, Y., Zhu, Y., Cao, W., Cheng, T., 2018. BRDF Effect on the Estimation of Canopy Chlorophyll Content in Paddy Rice from UAV-Based Hyperspectral Imagery. In: IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 6464–6467. https://doi.org/10.1109/IGARSS.2018.8517684.

McDonald, D.J., 1994. Temperate rice technology for the 21st century: an Australian example. Aust. J. Exp. Agric. 34, 877–888. https://doi.org/10.1071/ea9940877.

Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Computers & Geosciences 32, 1378–1388. https://doi.org/10.1016/j.cageo.2005.12.009.

Mulla, D.J., 2013. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. Biosyst. Eng. 114, 358–371. https://doi.org/10.1016/j.biosystemseng.2012.08.009.

Nutini, F., Confalonieri, R., Paleari, L., Pepe, M., Criscuolo, L., Porta, F., Ranghetti, L., Busetto, L., Boschetti, M., 2021. Supporting operational site-specific fertilization in rice cropping systems with infield smartphone measurements and Sentinel-2 observations. Precision Agric. https://doi.org/10.1007/s11119-021-09784-0.

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. Journal of Machine Learning Research 12, 2825–2830.

Peng, J., Manevski, K., Kørup, K., Larsen, R., Andersen, M.N., 2021. Random forest regression results in accurate assessment of potato nitrogen status based on multispectral data from different platforms and the critical concentration approach. Field Crops Research 268, 108158. https://doi.org/10.1016/j.fcr.2021.108158.

Peng, S., Buresh, R.J., Huang, J., Zhong, X., Zou, Y., Yang, J., Wang, G., Liu, Y., Hu, R., Tang, Q., Cui, K., Zhang, F., Dobermann, A., 2010. Improving nitrogen fertilization in rice by sitespecific N management. A review. Agronomy for Sustainable Development 30, 649–656. https://doi.org/10.1051/agro/2010002.

Ryu, C., Suguri, M., Umeda, M., 2011. Multivariate analysis of nitrogen content for rice at the heading stage using reflectance of airborne hyperspectral remote sensing. Field Crops Research 122, 214–224. https://doi.org/10.1016/j.fcr.2011.03.013.

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. https://doi.org/10.1016/j.jag.2013.04.003.

Shi, P., Wang, Y., Xu, J., Zhao, Y., Yang, B., Yuan, Z., Sun, Q., 2021. Rice nitrogen nutrition estimation with RGB images and machine learning methods. Computers and Electronics in Agriculture 180, 105860. https://doi.org/10.1016/j.compag.2020.105860.

Simmonds, M.B., Plant, R.E., Peña-Barragán, J.M., van Kessel, C., Hill, J., Linquist, B.A., 2013. Underlying causes of yield spatial variability and potential for precision

management in rice systems. Precision Agric. 14, 512–540. https://doi.org/10.1007/s11119-013-9313-x.

Wang, X., Miao, Y., Dong, R., Zha, H., Xia, T., Chen, Z., Kusnierek, K., Mi, G., Sun, H., Li, M., 2021. Machine learning-based in-season nitrogen status diagnosis and sidedress nitrogen recommendation for corn. Eur. J. Agron. 123, 126193. https://doi.org/10.1016/j.eja.2020.126193.

Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. Remote Sens. Environ. 236, 111402. https://doi.org/10.1016/j.rse.2019.111402.

Yao, Y., Miao, Y., Huang, S., Gao, L., Ma, X., Zhao, G., Jiang, R., Chen, X., Zhang, F., Yu, K., Gnyp, M.L., Bareth, G., Liu, C., Zhao, L., Yang, W., Zhu, H., 2012. Active canopy sensor-based precision N management strategy for rice. Agronomy for Sustainable Development 32, 925–933. https://doi.org/10.1007/s13593-012-0094-9.

Zhang, H.K., Roy, D.P., Yan, L., Li, Z., Huang, H., Vermote, E., Skakun, S., Roger, J.C., 2018. Characterization of Sentinel-2A and Landsat-8 top of atmosphere, surface, and nadir BRDF adjusted reflectance and NDVI differences. Remote Sens. Environ. 215, 482–494. https://doi.org/10.1016/j.rse.2018.04.031.

Zheng, H., Cheng, T., Li, D., Zhou, X., Yao, X., Tian, Y., Cao, W., Zhu, Y., 2018. Evaluation of RGB, Color-Infrared and Multispectral Images Acquired from Unmanned Aerial Systems for the Estimation of Nitrogen Accumulation in Rice. Remote Sensing 10, 824. https://doi.org/10.3390/rs10060824.

Zhou, K., Cheng, T., Zhu, Y., Cao, W., Ustin, S.L., Zheng, H., Yao, X., Tian, Y., 2018. Assessing the Impact of Spatial Resolution on the Estimation of Leaf Nitrogen Concentration Over the Full Season of Paddy Rice Using Near-Surface Imaging Spectroscopy Data. Frontiers in Plant Science 9. https://doi.org/10.3389/fpls.2018.00964.