

Article

Model for Predicting Rice Yield from Reflectance Index and Weather Variables in Lowland Rice Fields

Chinaza B. Onwuchekwa-Henry^{1,2,3,*}, Floris Van Ogtrop¹, Rose Roche² and Daniel K. Y. Tan¹

¹ Sydney Institute of Agriculture, School of Life and Environmental Sciences, Faculty of Science, The University of Sydney, Sydney, NSW 2006, Australia; floris.vanogtrop@sydney.edu.au (F.V.O.); daniel.tan@sydney.edu.au (D.K.Y.T.)

² Agriculture and Food/DATA61, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Canberra, ACT 2601, Australia; Rose.Roche@csiro.au

³ Department of Agronomy, College of Crop and Soil Sciences, Michael Okpara University of Agriculture, Umudike P.O. Box 7627, Nigeria

* Correspondence: chinaza.onwuchekwa-henry@sydney.edu.au

Abstract: Smallholder rice farmers need a multi-purpose model to forecast yield and manage limited resources such as fertiliser, irrigation water supply in-season, thus optimising inputs and increasing rice yield. Active sensing tools like Canopeo and GreenSeeker-NDVI have provided the opportunity to monitor crop health and development at different growth stages. In this study, we assessed the effectiveness of in-season estimation of rice yield in lowland fields of northwest Cambodia using weather data and vegetation cover information measured with; (1) the mobile app-Canopeo, and (2) the conventional GreenSeeker hand-held device that measures the normalised difference vegetative index (NDVI). We collected data from a series of on-farm field experiments in the rice-growing regions in 2018 and 2019. Average temperature and cumulative rainfall were calculated at panicle initiation and pre-heading stages when the crop cover index was measured. A generalised additive model (GAM) was generated using log-transformed data for grain yield, with the combined predictors of canopy cover and weather data during panicle initiation and pre-heading stages. The pre-heading stage was the best stage for grain yield prediction with the Canopeo-derived vegetation index and weather data. Overall, the Canopeo index model explained 65% of the variability in rice yield and Canopeo index, average temperature and cumulative rainfall explained 5, 65 and 56% of the yield variability in rice yield, respectively, at the pre-heading stage. The model (Canopeo index and weather data) evaluation for the training set between the observed and the predicted yield indicated an R^2 value of 0.53 and root mean square error (RMSE) was 0.116 kg ha⁻¹ at the pre-heading stage. When the model was tested on a validation set, the R^2 value was 0.51 (RMSE = 925.533 kg ha⁻¹) between the observed and the predicted yield. The NDVI-weather model explained 62% of the variability in yield, NDVI, average temperature and cumulative rainfall explained 3, 62 and 54%, respectively, of the variability in yield for the training set. The NDVI-weather model evaluation for the training set showed a slightly lower fit with R^2 value of 0.51 (RMSE = 0.119 kg ha⁻¹) between the observed and the predicted yield at pre-heading stage. The accuracy performance of the model indicated an R^2 value of 0.46 (RMSE = 979.283 kg ha⁻¹) at the same growth stage for validation set. The vegetation-derived information from Canopeo index-weather data increasingly correlated with rice yield than NDVI-weather data. Therefore, the Canopeo index-weather model is a flexible and effective tool for the prediction of rice yield in smallholder fields and can potentially be used to identify and manage fertiliser and water supply to maximise productivity in rice production systems. Data availability from more field experiments are needed to test the model's accuracy and improve its robustness.

Keywords: rice; yield prediction; generalised additive model; active remote sensors



Citation: Onwuchekwa-Henry, C.B.; Ogtrop, F.V.; Roche, R.; Tan, D.K.Y. Model for Predicting Rice Yield from Reflectance Index and Weather Variables in Lowland Rice Fields. *Agriculture* **2022**, *12*, 130. <https://doi.org/10.3390/agriculture12020130>

Academic Editor: Francisco Gutiérrez

Received: 15 November 2021

Accepted: 11 January 2022

Published: 18 January 2022

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



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Rice production is a significant contributor to the Cambodian economy for poverty reduction, income growth and national and household level food security. However, the yield productivity is 1.6 t ha^{-1} below the potential yields of between 3.3 and 5.5 t ha^{-1} for the local varieties in Cambodia [1]. As the predominant crop and means of livelihood for most farmers in the southeast Asia, projections indicated the need to increase rice production by 1.1% p.a. over the next decades to meet the demand of a rising population [2]. Cambodia has progressed from a rice importer after the Khmer Rouge war to a rice exporting country. At present, rice yield remains stagnant due to variability in climatic conditions, irrigation water supply and traditional practices. A past survey reported that rice yield varied from 0.6 to 5.5 t ha^{-1} across villages and regions in Cambodia due to variability in management practices [3]. The inability to predict yield may result in low yield given that crop yield may vary from year to year, depending on the local weather conditions, soil types and management inputs [4]. To maximise grain yield, rice farmers need scalable decision support tools to enhance resource use efficiency and intensify rice production. Recently, the discourse to increase rice productivity in smallholder fields has extended to sustainable and practices and precision agriculture [5]. Achieving an increase in rice productivity in smallholder fields will require a locally calibrated and validated model to predict yield at different rice growth stages.

Predicting rice yield is a diagnostic technique that can enhance resource inputs, minimise the risk of crop failure, decrease yield productivity and ultimately promote food security. Many studies have reported significant increases in rice yield using different yield prediction models in rice crops [6,7]. For example, the Agricultural Production Systems sIMulator (APSIM) model has been heralded for its ability to estimate yield and offers opportunities to improve yield management in Cambodia [8,9]. However, using the APSIM model to forecast rice yield may not be practical for smallholder rice growers because of its complexity in using extensive information on cultivar type and soil information that are not readily available. Hence, over the decades, prediction models using remote sensing data from advanced platforms such as Unmanned aerial vehicles and Sentinel (satellite) have been successfully used to make crop management decisions and estimate rice yield [10–12]. Conventionally, many reports focus on using normalised difference vegetation indices (NDVI) to estimate rice biomass at critical growth stages on small-scale rice fields [13,14]. A past study reported a positive relationship at the panicle initiation ($R^2 = 0.36$) and panicle differentiation ($R^2 = 0.42$) stages between NDVI and grain yield [15]. A recent study predicted rice yield at panicle initiation using GreenSeeker-NDVI in rice fields [16]. Nevertheless, the usefulness of its application to predict rice yield is limited in smallholder farmer fields, particularly in developing countries. Apart from the challenge of spectral saturation when biomass is high, the GreenSeeker sensor is expensive and not easily accessible to smallholder farmers. An alternative approach that is cost-effective to estimate rice yield is using freely available mobile application such as the Canopeo app.

The Canopeo app uses spectral values of red-green-blue (RGB) colour systems to diagnose plant health conditions and estimate crop biomass [17]. It can be used to develop a crop model that includes a comprehensive range of crop–canopy attributes. For example, Canopeo showed a positive and significant relationship in biomass estimation in sorghum varieties [18], and other yield-related traits in sorghum [19,20]. However, the disadvantage is that variations in soil conditions and variability in agronomic management practices in farmers' fields can affect yield estimation and management decision in rice fields. Most yield prediction models are improved by long-time series weather data to provide a quantitative estimation of expected yield in advance because of their influence on vegetative growth and development at different growth stages [21,22]. A past study reported a significant influence of accumulated temperature and precipitation in predicting rice yield phenological development stages [23]. However, our goal is using easy-to-measure weather variables that are localised to support smallholder farmers to predict rice yield within rice fields. While the weather variables are great predictors of crop yield across experiments that

are characterised with spatial and temporal variations, integrating weather components can offset the limitations associated with the yield prediction from each sensor, which is a useful strategy to support smallholder farmers to improve rice productivity. The use of statistical-based model such as multiple linear regression (MLR) provides potential to determine the relationships between environmental and/or vegetation factors and crop yield and improve yield predictions [24]. Gnyp et al. [25] reported an R^2 value of 0.50 between vegetation reflectance indices and rice aboveground biomass at booting stage with using multiple linear regressions. However, the MLR does not allow flexible specification of the dependence of the response variable on covariates to capture hidden patterns emanating from various field management conditions.

The generalised additive model (GAM) can explore the non-linear relationships between the dependence variables and independent variables for it depends linearly on unknown smooth functions of some predictor variables, and interest focuses on inference from these smooth functions to determine the relationship between the predictors and the response variable. GAMs can account for the potential non-linearity in the relationship between rice yield and potential predictors of yield, and the linear predictors include a sum of smooth functions of covariates. The use of the GAMs in forecasting rice yield are easy to interpret by plotting fitted smooths against yield, which potentially can improve our understanding of crop-vegetation indices-weather interactions. For example, Evans et al. [26] used seasonal water availability, soil type and location for predicting wheat yield with GAM. In addition, GAM increased the accuracy of yield ($R^2 = 0.65$) estimation with seasonal climate and phenological metrics in wheat [27]. Hence, this study aimed to determine if Canopeo index and weather data could predict yield under various field conditions with the generalised additive model (GAM) compared with only the NDVI and weather data. The rice yield data from various management and environmental conditions were examined using vegetation indices and local weather variables across Cambodian rice fields at specific growth stages to make rice management decisions. The purpose was to develop a model to forecast yield with a few variables that rice farmers can access locally to support and improve strategic management practices such as fertilisation, irrigation scheduling applications and weed control for rice under dynamic field conditions. Therefore, we hypothesised that developing a Canopeo index and weather model can predict rice yield in-season using a generalised additive model. The specific objective was to assess the effectiveness of a model combining Canopeo index and weather data to predict rice yield at different rice growth stages and to compare its performance with the conventional NDVI model with weather data.

2. Materials and Methods

2.1. Experiments and Data Collection

Data used in this study were collected from different on-farm experiments with common measurements undertaken as part of the CamSID project (ACIAR project-CSE/2015/044). The project is targeted to promote sustainable intensification and diversification in the lowland rice system in Cambodia. During these experiments, Canopeo index and NDVI data were collected at panicle initiation and at pre-heading stages to provide good coverage of varying field conditions in rice growing regions of northwest Cambodia. These experiments were conducted in the farmers' field across two provinces during the 2018 and 2019 growing seasons (Figure 1). These were categorised as Experiments 1, 2, 3, 4 and 5.

Experiments 1 and 2 were conducted in the early wet season (EWS) and main wet season (MWS), respectively, in the 2018 cropping season at Kork Ton Loab (KTL) village in Banteay Meanchey (BMC) province. Experiment 3 was conducted at Svay Cheat (SVC) village in Battambang (BTB) province in the early wet season. Experiments 1, 2 and 3 involved evaluations of study on seeding rates that included 20, 40, 60, 80 and 180 kg ha⁻¹ laid out in a randomised complete block design (RCBD) with three replications using the same management practices. Experiments 4 and 5 were conducted at Don Bosco School (DBS) farm in BTB province and KTL village, respectively, during the early wet season

(EWS) in 2019 under different management conditions and were laid out in a split-plot design with four replications.

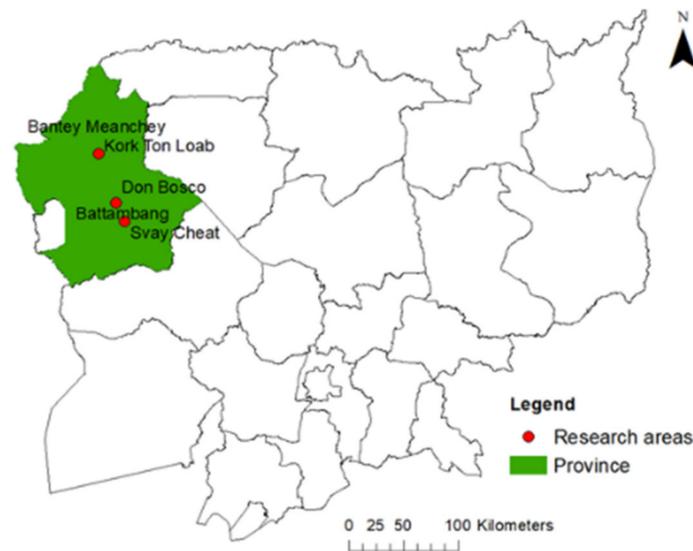


Figure 1. Location of the study area for on-farm field experiments.

For Experiment 4, the subplots consisted of five herbicide options (pendimethalin, pretilachlor, butachlor, oxadiazon and control) applied at 1.5 a.i. kg ha⁻¹; while the main plot had 1.0 a.i. kg ha⁻¹ of bysipyribac-sodium (post-emergence herbicide) and a control plot. The subplot for Experiment 5 consisted of different seeding rates, 20, 40, 60, 80 and 180 kg ha⁻¹, while the main plots comprised different fertiliser rates as low, medium and high rates; N-50, P-40 and K-30, N-75, P-60 and K-40 and N-100, P-80 and K-50 kg ha⁻¹, respectively. Rice fields were divided into smaller plots based on the experimental layout of each experiment treatment, as stated above. The total field size for each experiment was not less than 0.8 ha, and rice field owners managed the experiments using the protocols in each rice field. Sen Kra Oub was variety was sown at different dates for each experiment across the experiments in both cropping seasons.

2.2. Canopeo, NDVI and Weather Data Collection

The spectral measurements were collected using a freely available mobile app-Canopeo and the commercial GreenSeeker-NDVI (Trimble[®], Sunnyvale, CA, USA) (Figure 2). Canopeo is a smartphone image processing tool that estimates canopy cover using colour values in the red–green–blue (RGB) bands. It classifies the green pixels from rice canopy within the spectral domain of 500–570 nm to produce white pixels according to ratios of red/green, blue/green and the excess green index. The vegetation information collected with the Canopeo app reproduces the white canopy pixels as well as the value of the percentage of the canopy cover (Figure 2b–d). Fractional green canopy cover from the app ranges from 0 to 1, where the value of 0 means no green canopy cover, and value of 1 means 100% of green canopy cover (Equation (1)).

$$R/G < P_1 \text{ and } B/G < P_2 \text{ and } 2G - R - B > P_3 \quad (1)$$

P_1 and P_2 are parameters that have values close to 1 and classify pixels mainly in the green bands of ~500–570 nm, and P_3 is a parameter that sets the minimum excess green index at a constant value of 20 to select green vegetation of rice canopy. Default parameter values for Canopeo are $P_1 = 0.95$, $P_2 = 0.95$ and $P_3 = 20$. GreenSeeker-NDVI takes canopy measurements between ~650 and 780 nm at red and near-infrared wavelength bands, respectively (Equation (2)). The vegetation information collected with the GreenSeeker-NDVI sensor generates an output of NDVI values with no image (Figure 2a).

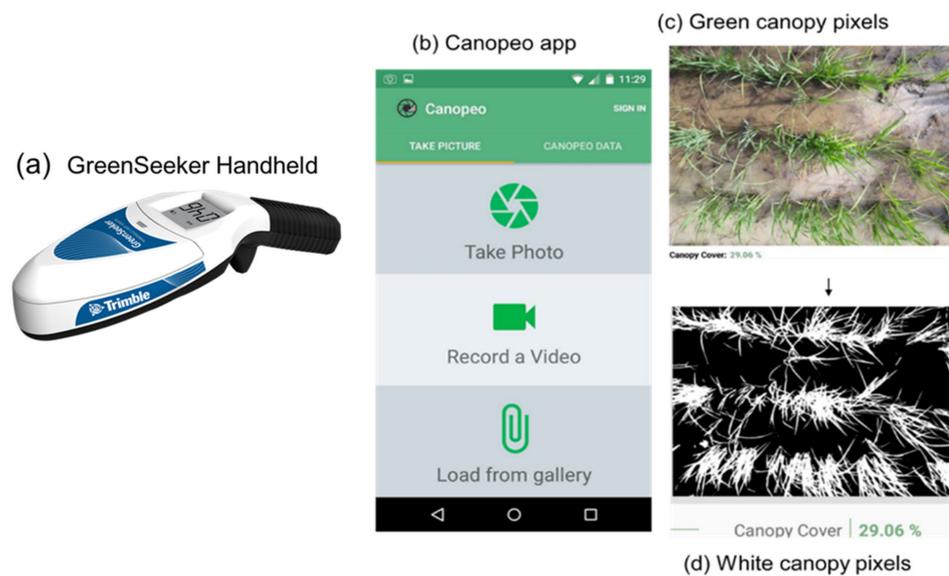


Figure 2. Ground-based sensors for assessing vegetation information in the rice field.

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (2)$$

NIR is a reflection in the near-infrared spectrum, while red reflects the red range of the spectrum. Spectral measurements were taken above a canopy height of between 60–80 cm using a standard sampling method and position to reduce sampling errors that might result from both sensors. The scanning method was done in transects by walking eight steps within plant rows in each plot. Two transects were sampled from each rice plot during panicle initiation (PI) and pre-heading (PH) stages. Reflectance data for each sensor were averaged for each plant plot where yield data was collected with a total number of 169 data points for NDVI and Canopeo data, respectively, in all five experiments (Table 1). Plants were harvested for grain yield at maturity, dried at 70 °C and weighed. The weather data used in this study were obtained from the local weather stations. In Cambodia, climate is tropical and fairly uniform across the province. The average temperatures (°C) and cumulative precipitations were estimated from the time of sowing to the panicle initiation and pre-heading stages, respectively, for each experiment (Table 2).

Table 1. Agronomic information, sensing dates and spectral reflectance data from the on-farm experiments at panicle initiation (PI) and pre-heading (PH) for the early wet season (EWS) and main wet season (MWS) rice growing seasons in 2018 and 2019 at Kork Ton Loab (KTL) village, Banteay Meanchey (BMC) province, and Svay Cheat (SVC) village and Don Bosco School (DBS), Battambang province.

Experiment	Site/Season	Planting	PI Stage	PH Stage	Harvesting	Year	Reflectance Data Total ($n = 169$)
1	KTL/EWS	26 April	26 June	10 July	9 July	2018	30
2	KTL/MWS	24 September	29 November	13 December	8 January	2018	30
3	SVC/EWS	17 May	17 July	31 July	4 September	2018	24
4	DBS/EWS	20 July	24 August	9 September	30 September	2019	40
5	KTL/EWS	3 May	1 July	10 July	31 August	2019	45

Table 2. Climate data for each on-farm field experiment during rice growing seasons in 2018 and 2019.

Experiment	Mean Temperature PI (°C)	Mean Temperature PH (°C)	Mean Cumulative Rainfall PI (mm)	Mean Cumulative Rainfall PH (mm)
1	26.11	26.5	341.3	407.8
2	28.1	28.3	589.1	626.0
3	28.3	25.9	369.3	381.1
4	26.8	27.0	346.2	407.8
5	29.6	26.0	748.4	829.3

Where PI and PH represent panicle initiation and pre-heading, respectively.

2.3. Data Processing and Statistical Analysis

2.3.1. Simple Linear Regression and Generalised Additive Models

The relationships of individual datasets with each vegetative index and yield at both growth stages were initially examined for initial exploration using simple linear regression (SLR) model in the form of:

$$Y_t = \beta_{0,t} + \beta_1 V_t + \varepsilon_t \quad (3)$$

where $\beta_{0,t}$, Y_t , V_t and ε_t represent an intercept, yield, the reflectance index of Canopeo or NDVI and error term for each site 't'. The generalised additive model was developed for yield predictive and validation purposes using combined predicting components of individual spectral index and weather data: average temperature and cumulative rainfall. Grain yield data was log-transformed to meet the assumptions of the regression. A combination of growing cumulative rainfall and growing average temperature were selected as variables to avoid collinearity. Overall, we developed two models from GAM, which were the Canopeo-weather model and NDVI-weather model. Data on the growing average temperature and cumulative rainfall, which coincided with the sensing dates at panicle initiation and pre-heading stages, were collected across the experiments (Table 1). To increase the model's application to be used extensively under any environmental conditions in Cambodia, each on-farm field experiment dataset was pooled together before model calibration irrespective of the differences in their management conditions.

We randomly split the pooled dataset into approximately 70% for training and 30% for testing. The model was calibrated using 70% of the dataset, while the remaining 30% was used to validate the model's accuracy (GAM). Two different spectral index-weather data were developed with a generalised additive model (GAM) that combines the specified sum of smooth terms as 'f' of the predictors and the dimension of the basis that represent the smooth term as 'k' [28], and the residuals satisfied the criterial for normality.

$$\text{Model} = \text{gam}(\text{Log10}(\text{Yield}) \sim f_1(V_1, k = 3) + f_2(T_2, k = 3) + f_3(R_3, k = 3) \quad (4)$$

where gam, V_1 , T_2 and R_3 represent the function for general additive modelling, the spectral index (Canopeo or NDVI), growing average temperature and cumulative precipitation data, respectively, while 'f' is the smooth function and 'k' is the dimension of the basis that represents the smooth term for all field experiments. The descriptive statistics were estimated for the observed and predicted yield in the subsets of calibration and validation for mean, minimum (min), maximum (max), standard deviation (SD) and coefficient of variation (CV) (Table 3).

Table 3. Descriptive statistics of yield samples from the pooled dataset.

Yield (kg ha ⁻¹)	Total (n)	Max (kg ha ⁻¹)	Min (kg ha ⁻¹)	Mean (kg ha ⁻¹)	SD (kg ha ⁻¹)	CV (%)
Calibration	169	9118.86	669.72	3051.58	1460.92	47.87
Validation	60	8445.03	662.73	2807.32	1324.75	52.01

Where Max, Min, SD and CV represent maximum, minimum, standard deviation and coefficient of variation, respectively.

2.3.2. Model Evaluation

The coefficient of determination (R^2) and the root mean square error (RMSE) was used to evaluate the model performances.

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

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (6)$$

where y_i represents the yield data, \hat{y}_i is the predicted data and \bar{y} is the mean yield of the model, n is the number of samples. All data were analysed using R scripts using these models [29].

3. Results

3.1. Relationships between Yield and Sensors

There was a positive relationship between yield and Canopeo with simple linear regression (SLR). However, the relationships between yield and Canopeo was poor at different growth stages across the locations for each experiment (Table 4). The coefficients of determination (R^2) were 0.24 and 0.20 at panicle initiation (PI) and pre-heading (PH) stages, respectively, in Svay Cheat (SVC). For experiment 3 at Kork Ton Lab (KTL), R^2 was 0.24 at the PI stage with the Canopeo index. Similarly, NDVI showed a weak relationship with yield using the SLR model at both growth stages in each location regardless of different management practices used in the on-farm field experiments (Table 5). With NDVI, R^2 was 0.23 at KTL in experiment 1 and 0.22 at SVC at pre-heading in experiment 2 (Table 5).

Table 4. Relationship between yield and Canopeo showing coefficients of determination at panicle initiation (PI) and pre-heading stages across locations at Kork Ton Loab (KTL) village, Banteay Meanchey (BMC) province, and Svay Cheat (SVC) village and Don Bosco School (DBS), Battambang province.

Experiment	Location	Growth Stage	Intercept (β_0)	Predictor ($x = \text{Canopeo}$)	R^2	p -Value
1	KTL	PI	1157.48	11.718	0.14	0.04 *
		PH	1640.75	4.717	0.03	0.37
2	SVC	PI	426.88	16.719	0.24	0.02 *
		PH	216.49	18.29	0.20	0.06 *
3	KTL	PI	−8108.39	156.53	0.10	0.11
		PH	−221.19	71.5	0.24	0.01 **
4	DBS	PI	4280.34	−10.15	0.02	0.35
		PH	4784.33	−20.95	0.10	0.09
5	KTL	PI	2492.65	8.069	0.03	0.22
		PH	2375.26	9.159	0.03	0.09

* and ** are significant at $p < 0.05$ and 0.01 .

3.2. Yield Estimation and Model Evaluation with Explanatory Variables Using a Generalised Additive Models (GAM)

With GAM, the analysis was done on the pooled dataset and ($n = 169$) across different experiments. Generally, the GAM model appeared to be more robust in predicting rice yield using Canopeo and weather variables combinations despite the different management practices and various field conditions used in the experiments. The Canopeo-weather model fitted well with the observed yield data (deviance = 65%) and the 5, 56 and 65% were the proportion explained by Canopeo index, cumulative rainfall and the average temperature explained, respectively, at the pre-heading stage (Table 6). More so, the explanatory smooth terms for average temperature contributed more to the model and yield increased in a non-linear way with increased average temperature at pre-heading stage (Figure 3).

Table 5. Relationship between yield and Canopeo showing coefficients of determination at panicle initiation (PI) and pre-heading (PH) stages across each location at Kork Ton Loab (KTL) village, Banteay Meanchey (BMC) province, and Svay Cheat (SVC) village and Don Bosco School (DBS), Battambang province.

Experiment	Location	Growth Stage	Intercept (y)	Predictor (x = NDVI)	R ²	p-Value
1	KTL	PI	−14.25	2739.63	0.14	0.042 *
		PH	−1845	6046	0.23	0.007 **
2	SVC	PI	−1675	4570	0.23	0.019 *
		PH	−1655	4751	0.22	0.020 *
3	KTL	PI	9710	−7490	0.0033	0.762
		PH	850.7	5294.8	0.002	0.813
4	DBSC	PI	4900	−1778	0.010	0.610
		PH	6760	−5157	0.10	0.116
5	KTL	PI	1272	2834.3	0.20	0.004 **
		PH	2105.3	1305.1	0.04	0.187

* and ** are significant at $p < 0.05$ and 0.01 , respectively.

Table 6. Deviance of the models and individual model component at panicle initiation (PI) and pre-heading (PH) stages with generalised additive model. ‘gam’ in the model is the function for general additive modelling ‘f’ represents the smooth function and ‘k’ is the dimension of the basis that represents the smooth term.

Models	Predictors/Model Components	Deviance (%)	
		PI	PH
gam(yield ~ f(Canopeo, k = 3) + f(average temperature, k = 3) + f(cumulative rainfall, k = 3))	-	62	65
	f(Canopeo)	5	5
	f(average temperature)	62	65
	f(cumulative rainfall)	35	56
gam(yield ~ f(NDVI, k = 3) + f(average temperature, k = 3) + f(cumulative rainfall, k = 3))	-	62	62
	f(NDVI)	<0.1	3
	f(average temperature)	62	62
	f(cumulative rainfall)	30	54

Evaluating the Canopeo-weather model on the calibration (training) set showed an R² of 0.51 (RMSE = 0.118 kg ha^{−1}) between observed and predicted yield at the panicle initiation (Figure 4a). For the validation set, the model showed an accuracy performance with R² value of 0.46 (RMSE = 977.016 kg ha^{−1}) between observed and predicted yield at a similar growth stage (Figure 4b). On the other hand, the Canopeo-weather model showed a higher R² of 0.53 (RMSE = 0.116 kg ha^{−1}) when the model was evaluated on the calibrated set between observed and predicted at the pre-heading stage (Figure 5a). The R² was 0.51 (RMSE = 925.533 kg ha^{−1}) when the model was tested on the validation set at the pre-heading stage (Figure 5b).

The model developed from the variable inputs of NDVI and weather variables showed slightly lower fit model at the pre-heading stage with deviance value of 65% compared with Canopeo-weather model (Table 6). The proportion of the variability in yield explained by the NDVI was 3%, cumulative rainfall explained 54% whereas 62% of the variability in yield was explained by average temperature. In addition, yield increased in a non-linear pattern with increased average temperature, while cumulative rain decreased with increased yield and NDVI showed a flat response (Figure 6.)

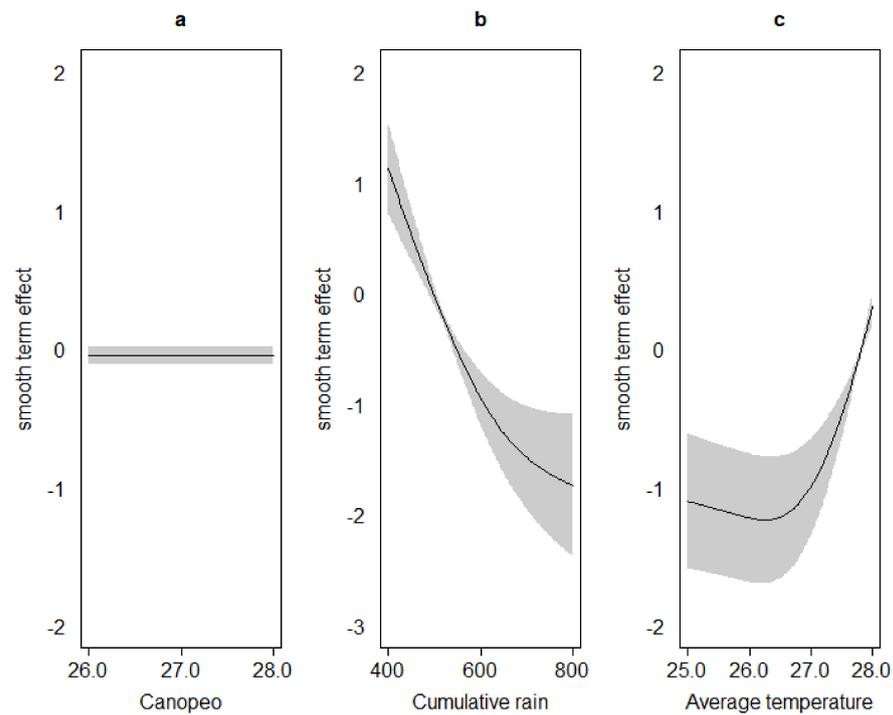


Figure 3. Smooth term effect of each predictor (a) Canopeo, (b) cumulative rain, (c) average temperature, on log-transformed yield at pre-heading stage Canopeo-weather model.

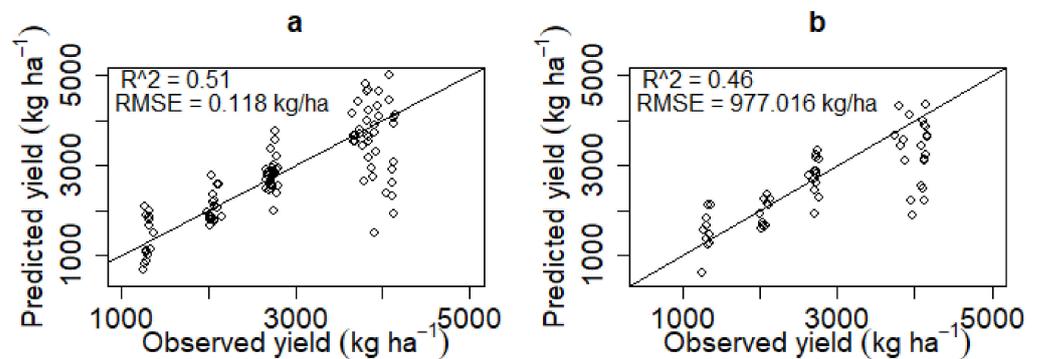


Figure 4. Evaluation of model (Canopeo and weather variables; average temperature and cumulative rainfall) for calibration set (a) and model performance for the validation set (b) showing the model predictive power between the predicted and the observed yield at panicle initiation stage.

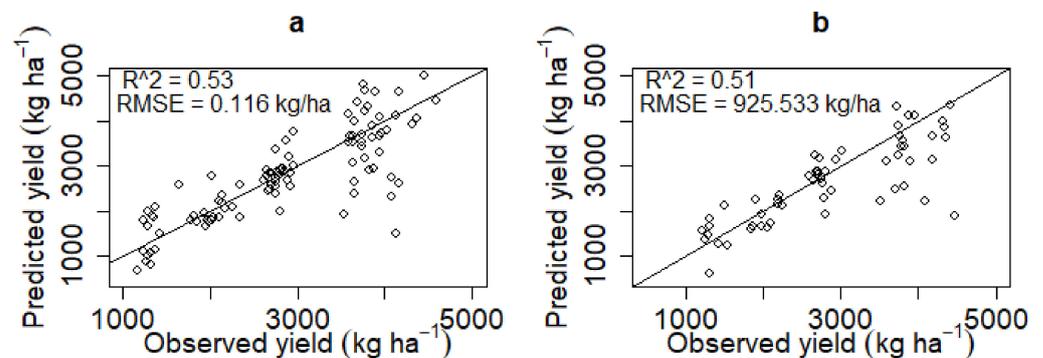


Figure 5. Evaluation of model (Canopeo and weather variables; average temperature and cumulative rainfall) for the calibration set (a) and model performance for the validation set (b) showing the model predictive power between the predicted and the observed yield at pre-heading stage.

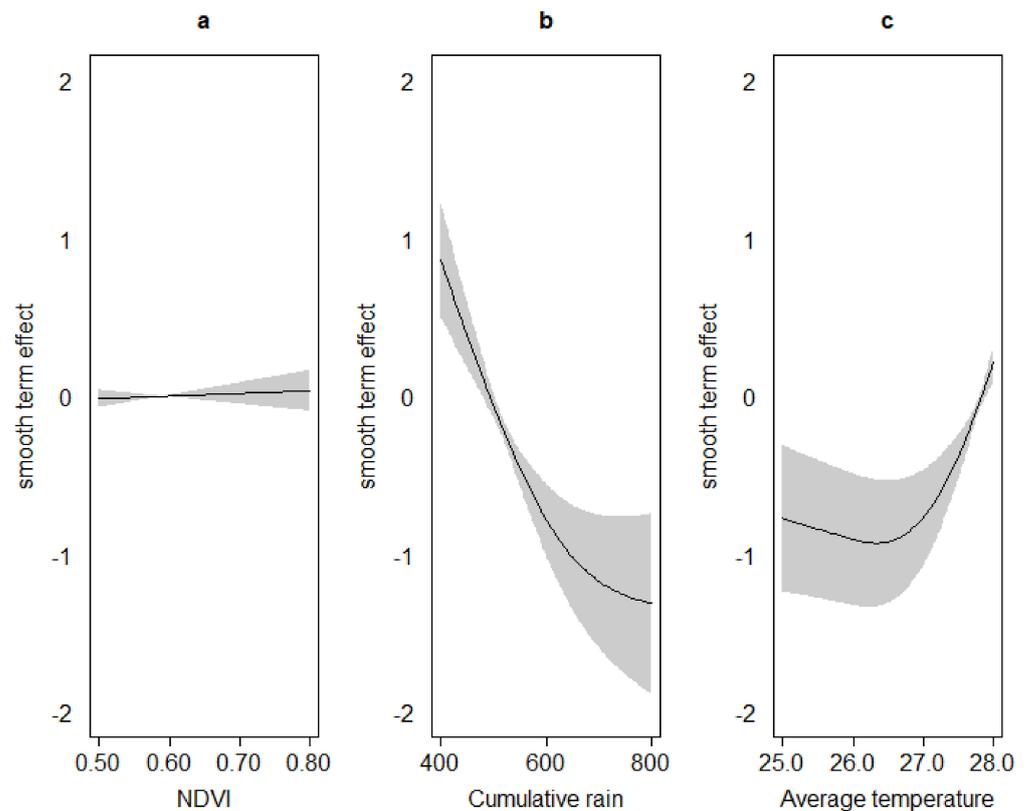


Figure 6. Smooth term effect of each predictor (a) NDVI, (b) cumulative rain, (c) average temperature, on log-transformed yield at pre-heading stage NDVI-weather model.

The model performance was low (R^2 value = 0.51 RMSE = 0.119 kg ha⁻¹) for the calibration set between the observed and the predicted yield at the panicle initiation stage (Figure 7a). There was similar a similar fit of an R^2 value of 0.47 when the model was tested on the validation set with an accuracy model performance with RMSE = 968.688 kg ha⁻¹ in predicting yield at the panicle initiation stage and (Figure 7b).

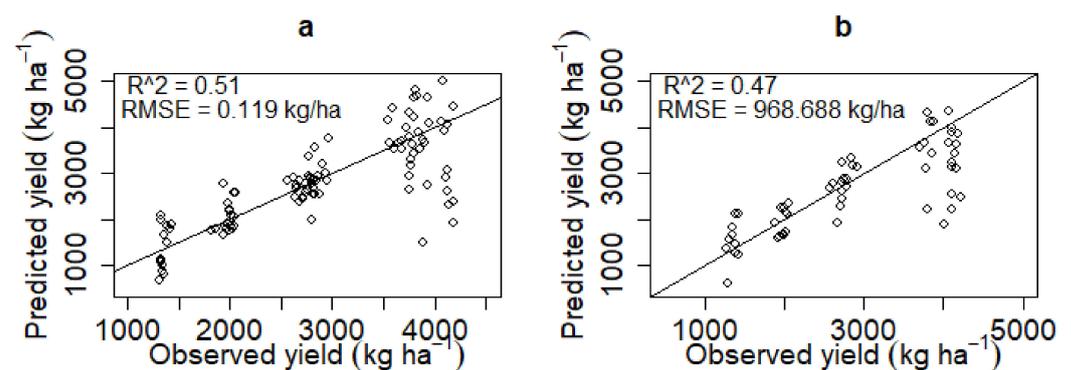


Figure 7. Evaluation of model (NDVI and weather variables; average temperature and cumulative rainfall) for the calibration set (a) and model performance for the validation set (b) showing the model predictive power between the predicted and the observe at panicle initiation stage.

At the pre-heading stage, the model evaluation on training set showed an R^2 value of 0.51 between the observed and predicted yields with RMSE value of 0.119 kg ha⁻¹ (Figure 8a). For the validations set, the R^2 was 0.46 (RMSE = 979.283 kg ha⁻¹) between the observed and the predicted yield at the same growth stage (Figure 8b).

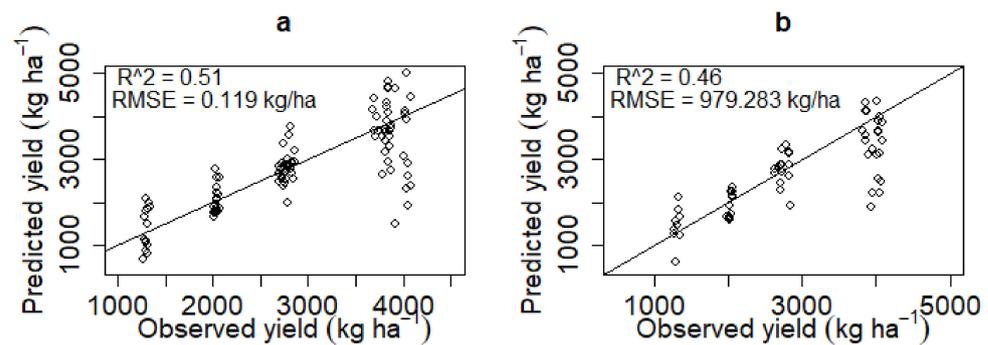


Figure 8. Evaluation of model (NDVI and weather variables; average temperature and cumulative rainfall) for the calibration set (a) and model performance for the validation set (b) showing model predictive power between the predicted and the observed yield at the pre-heading stage.

4. Discussion

4.1. NDVI Yield Model

This study demonstrates that individual NDVI showed poor relationships with yield under different management practices using the SLR. The possible reasons are the noise from standing water in the rice fields might have reduced the correlation between NDVI and yield. The reflectance index with a red-light band within a narrow range from ~650–670 nm may not accurately differentiate agronomic parameters and specific physiological characteristics in rice crops resulting in the saturation of the spectral sensors at all the growth stages for each experiment (Tables 4 and 5) [30,31]. In addition to this, GreenSeeker-NDVI does not have the in-built feature to remove noise associated with background effect in the rice fields. Overall, the low coefficient of determination could be attributed to poor canopy development because of no standing water in the rice fields and high incidence of blast incidence during the panicle initiation stage. In other words, NDVI can saturate with excessive biomass or standing water. NDVI is the most widely spectral index used to estimate rice yield in the fields [32,33], before this study, neither NDVI nor Canopeo has been used to obtain vegetation information during mid-growth stages under Cambodian rice ecosystems. The poor sensor–yield relationship implies that it is not practical for growers to rely solely on individual spectral indices with red light bands to estimate rice yield at mid-growth stages. However, having a universal model is more practical to estimate yield at different rice growth stages in a wide range of environment [34,35]. In Cambodian rice fields, weather variables have been identified as important predictor variables that can estimate rice yield [36]. Therefore, NDVI did not explain much deviance with the NDVI-Weather model (Table 6). Other explanatory variables such as the average temperature in each model had non-linear relationship with the yield as the higher yield, the higher average temperature and the higher cumulative rainfall, the lower the yield. The low model performance at pre-heading was due to the saturation effect frequently associated with NDVI in the visible wavelength at high biomass conditions [35,36].

4.2. Canopeo-Index Yield Model

As demonstrated in this study, combining weather data and Canopeo index improved the model's performance marginally at the pre-heading stage. However, the model was biased and under-predicted yield higher than 4000 kg ha⁻¹ between panicle initiation and pre-heading stages using weather variables and spectral indices likewise NDVI-weather model. This result is consistent with other findings [37–39] that reported difficulty to predict yield in-season at 4000 kg ha⁻¹ with a different spectral index from the rice canopies. However, the possible reason for the scattered plots with the Canopeo-weather model might be that Canopeo extracts vegetation information from a single spectral and rice vegetation acquired with a single band as well as its dynamics reduced its effectiveness to analyse images in the red light compared with NDVI that uses two bands of visible light and near-infrared light. To the best of our knowledge, this study represents the first attempt to predict

rice yield in-season the combination of Canopeo and cumulative rainfall and average temperature. The SLR model increased our understanding of yield-sensor relationships in each location under the variable growing conditions in Cambodian rice farms. At the same time, the Canopeo-weather model explained more of the variability in yield; 53% and 51% for calibration and validation set, respectively, at the pre-heading stage with a degree of improvement in rice prediction accuracy from weather integration with GAM than using the NDVI-weather model. This agrees with the accuracy reported with GAM for yield estimation in crops by reducing the limitations associated with using only the sensors to predict rice yield [40]. This suggests that yield could be estimated with higher accuracy if more weather data are available and accessible to farmers in the decision support tools. (Figures 4 and 5). Therefore, it is more advantageous to combine Canopeo and weather data to predict rice yield because of the higher accuracies between expected and observed yield than using NDVI as revealed in this study (Figures 3 and 4).

Overall, Canopeo-weather model has great potential to support rice growers in taking management decisions on resource input such as irrigation water supply, fertiliser and insecticide in rice fields to minimise any future risk in the rice production system. Technically, the Canopeo-weather model is more flexible by their lesser demand for inputs for ease of adaptation to other region because rice growers can easily use their mobile phones to assess information on the model components. The Canopeo app is freely available online, and weather data can locally be sourced from their weather stations and online (<https://www.weather-forecast.com/maps/Cambodia>). Notwithstanding, the models were biased because they always under-predicted the yield of Sen Kra Oub variety when the predicted yields were compared with the observed yields at different rice stages because it is relatively a small dataset. This is a common result of using statistical and models that fit observations in an average sense for yield estimation [41,42]. Therefore, the Canopeo-weather model developed in this study can be used as a base model for further effects of crop management for fertiliser application, irrigation schedule and weed management to predict yield. It is also essential to determine the degree of uncertainty associated with the model to increase its reliability in predicting rice yield using other management conditions that are not included in this study. However, models may improve as data availability increases with more field experiments to test the model's accuracy and improve its robustness.

5. Conclusions

Prediction of rice yield improved using the Canopeo and weather data with a generalised additive model compared with the traditional NDVI algorithm and weather data. Average temperature and cumulative rainfall, when combined with NDVI, significantly influenced yield predictions. The combination of Canopeo and weather data further improved the model's prediction capability at the pre-heading stage than at panicle initiation between the observed and predicted yield. The model indicated a higher accuracy for model calibration at $R^2 = 0.53$, $RMSE = 0.116 \text{ kg ha}^{-1}$ and the validation of $R^2 = 0.51$ ($RMSE = 925.533 \text{ kg ha}^{-1}$) at pre-heading stage. The model is promising and can be useful for rice yield prediction to optimise management practices such as fertiliser and irrigation water supply. Still, proper modification and more testing are required with additional field experiments to improve the robustness and reliability of the model. It will be beneficial to adopt this intervention for managing low yield in smallholder rice fields in lowland rice fields.

Author Contributions: Conceptualisation, C.B.O.-H., F.V.O. and D.K.Y.T.; methodology, C.B.O.-H., F.V.O. and D.K.Y.T.; statistical analysis, C.B.O.-H., F.V.O. and D.K.Y.T.; investigation and data processing, C.B.O.-H., F.V.O. and D.K.Y.T.; writing of the original manuscript, C.B.O.-H.; review and editing, D.K.Y.T., F.V.O. and R.R.; supervision, D.K.Y.T., F.V.O. and R.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Australian Centre for International Agricultural Research (ACIAR) Project number CSE/2015/044. The Norman Matheson Scholarship was funded by the University of Sydney, Australia.

Data Availability Statement: Data are available on request.

Acknowledgments: We thank the Australian Centre for International Agricultural Research (ACIAR) Project number CSE/2015/044 for funding this research and Tertiary Education Trust Fund (TETFund), Norman Matheson Scholarship and CSIRO/DATA 61 for the financial assistance on the field expenses.

Conflicts of Interest: The authors declare no conflict of interest.

References

- MAFF. *Ministry of Agriculture, Forestry and Fisheries Agricultural Statistics*; Department of Planning, Statistics and International Cooperation: Phnom Penh, Cambodia, 2018.
- Gathala, M.K.; Kumar, V.; Sharma, P.C.; Saharawat, Y.S.; Jat, H.S.; Singh, M.; Kumar, A.; Jat, M.L.; Humphreys, E.; Sharma, D.K.; et al. Reprint of “Optimizing intensive cereal-based cropping systems addressing current and future drivers of agricultural change in the Northwestern Indo-Gangetic Plains of India”. *Agric. Ecosyst. Environ.* **2014**, *187*, 33–46. [[CrossRef](#)]
- Li, P.; Zhang, X.; Wang, W.; Zheng, H.; Yao, X.; Tian, Y.; Zhu, Y.; Cao, W.; Chen, Q.; Cheng, T. Estimating aboveground and organ biomass of plant canopies across the entire season of rice growth with terrestrial laser scanning. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *91*, 102132. [[CrossRef](#)]
- Iizumi, T.; Ramankutty, N. How do weather and climate influence cropping area and intensity? *Glob. Food Sec.* **2015**, *4*, 46–50. [[CrossRef](#)]
- Campbell, B.M.; Thornton, P.; Zougmore, R.; van Asten, P.; Lipper, L. Sustainable intensification: What is its role in climate smart agriculture? *Curr. Opin. Environ. Sustain.* **2014**, *8*, 39–43. [[CrossRef](#)]
- Biswas, R.; Bhattacharyya, B.; Banerjee, S. Predicting Rice Yield from Weather Variable through Detrended Production Index. *Contemp. Res. India* **2017**, *7*, 462–466.
- Zhang, J.; Liu, X.; Liang, Y.; Cao, Q.; Tian, Y.; Zhu, Y.; Cao, W.; Liu, X.; Zhang, J.; Liu, X.; et al. Using a Portable Active Sensor to Monitor Growth Parameters and Predict Grain Yield of Winter Wheat. *Sensors* **2019**, *19*, 1108. [[CrossRef](#)] [[PubMed](#)]
- Poulton, P.L.; Vesna, T.; Dalgliesh, N.P.; Seng, V. Applying simulation to improve rice varieties in reducing the on-farm yield gap in Cambodian lowland rice ecosystems. *Exp. Agric.* **2015**, *51*, 264–284. [[CrossRef](#)]
- Gaydon, D.S.; Balwinder-Singh; Wang, E.; Poulton, P.L.; Ahmad, B.; Ahmed, F.; Akhter, S.; Ali, I.; Amarasingha, R.; Chaki, A.K.; et al. Evaluation of the APSIM model in cropping systems of Asia. *Field Crops Res.* **2017**, *204*, 52–75. [[CrossRef](#)]
- Babar, M.A.; Reynolds, M.P.; Van Ginkel, M.; Klatt, A.R.; Raun, W.R.; Stone, M.L. Spectral reflectance to estimate genetic variation for in-season biomass, leaf chlorophyll, and canopy temperature in wheat. *Crop Sci.* **2006**, *46*, 1046–1057. [[CrossRef](#)]
- Prasad, B.; Carver, B.F.; Stone, M.L.; Babar, M.A.; Raun, W.R.; Klatt, A.R. Genetic analysis of indirect selection for winter wheat grain yield using spectral reflectance indices. *Crop Sci.* **2007**, *47*, 1416–1425. [[CrossRef](#)]
- Rudiyanto; Minasny, B.; Shah, R.M.; Soh, N.C.; Arif, C.; Setiawan, B.I. Automated near-real-time mapping and monitoring of rice extent, cropping patterns, and growth stages in Southeast Asia using Sentinel-1 time series on a Google Earth Engine platform. *Remote Sens.* **2019**, *11*, 1666. [[CrossRef](#)]
- Cheng, T.; Song, R.; Li, D.; Zhou, K.; Zheng, H.; Yao, X.; Tian, Y.; Cao, W.; Zhu, Y. Spectroscopic Estimation of Biomass in Canopy Components of Paddy Rice Using Dry Matter and Chlorophyll Indices. *Remote Sens.* **2017**, *9*, 319. [[CrossRef](#)]
- Gnyp, M.L.; Yu, K.; Miao, Y.; Bareth, G. Estimating biomass of rice in farmers’ fields by red-edge indices. In Proceedings of the Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Lausanne, Switzerland, 24–27 June 2014; pp. 1–4.
- Harrell, D.L.; Tubaña, B.S.; Walker, T.W.; Phillips, S.B. Estimating rice grain yield potential using normalized difference vegetation index. *Agron. J.* **2011**, *103*, 1717–1723. [[CrossRef](#)]
- Guo, Y.; Xiang, H.; Li, Z.; Ma, F.; Du, C. Prediction of Rice Yield in East China Based on Climate and Agronomic Traits Data Using Artificial Neural Networks and Partial Least Squares Regression. *Agronomy* **2021**, *11*, 282. [[CrossRef](#)]
- Patrignani, A.; Ochsner, T.E. Canopeo: A powerful new tool for measuring fractional green canopy cover. *Agron. J.* **2015**, *107*, 2312–2320. [[CrossRef](#)]
- Chung, Y.S.; Choi, S.C.; Silva, R.R.; Kang, J.W.; Eom, J.H.; Kim, C. Case study: Estimation of sorghum biomass using digital image analysis with Canopeo. *Biomass Bioenergy* **2017**, *105*, 207–210. [[CrossRef](#)]
- Yellareddygari, S.K.R.; Gudmestad, N.C. Bland-Altman comparison of two methods for assessing severity of Verticillium wilt of potato. *Crop Prot.* **2017**, *101*, 68–75. [[CrossRef](#)]
- Da Silva, L.A.; Bressan, P.O.; Gonçalves, D.N.; Freitas, D.M.; Machado, B.B.; Gonçalves, W.N. Estimating soybean leaf defoliation using convolutional neural networks and synthetic images. *Comput. Electron. Agric.* **2019**, *156*, 360–368. [[CrossRef](#)]
- Yin, X.; Kropff, M.J.; McLaren, G.; Visperas, R.M. A nonlinear model for crop development as a function of temperature. *Agric. For. Meteorol.* **1995**, *77*, 1–16. [[CrossRef](#)]

22. Shi, P.; Tang, L.; Lin, C.; Liu, L.; Wang, H.; Cao, W.; Zhu, Y. Modeling the effects of post-anthesis heat stress on rice phenology. *Field Crops Res.* **2015**, *177*, 26–36. [[CrossRef](#)]
23. Samsul Huda, A.K.; Ghildyal, B.P.; Tomar, V.S.; Jain, R.C. Contribution of climatic variables in predicting rice yield. *Agric. Meteorol.* **1975**, *15*, 71–86. [[CrossRef](#)]
24. Lobell, D.B.; Burke, M.B. On the use of statistical models to predict crop yield responses to climate change. *Agric. For. Meteorol.* **2010**, *150*, 1443–1452. [[CrossRef](#)]
25. Gnyp, M.L.; Miao, Y.; Yuan, F.; Ustin, S.L.; Yu, K.; Yao, Y.; Huang, S.; Bareth, G. Hyperspectral canopy sensing of paddy rice aboveground biomass at different growth stages. *Field Crops Res.* **2014**, *155*, 42–55. [[CrossRef](#)]
26. Chen, K.; O’Leary, R.A.; Evans, F.H. A simple and parsimonious generalised additive model for predicting wheat yield in a decision support tool. *Agric. Syst.* **2019**, *173*, 140–150. [[CrossRef](#)]
27. Evans, F.H.; Shen, J. Long-term hindcasts of wheat yield in fields using remotely sensed phenology, climate data and machine learning. *Remote Sens.* **2021**, *13*, 2435. [[CrossRef](#)]
28. Wood, S.N. *Generalized Additive Models: An Introduction with R*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2017; ISBN 9781498728348.
29. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021; Available online: <https://www.R-project.org/> (accessed on 26 October 2021).
30. Liu, K.; Li, Y.; Hu, H.; Zhou, L.; Xiao, X.; Yu, P. Estimating Rice Yield Based on Normalized Difference Vegetation Index at Heading Stage of Different Nitrogen Application Rates in Southeast of China. *J. Environ. Agric. Sci.* **2015**, *2*, 13.
31. Zhou, X.; Zheng, H.B.; Xu, X.Q.; He, J.Y.; Ge, X.K.; Yao, X.; Cheng, T.; Zhu, Y.; Cao, W.X.; Tian, Y.C. Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 246–255. [[CrossRef](#)]
32. Chang, K.W.; Shen, Y.; Lo, J.C. Predicting rice yield using canopy reflectance measured at booting stage. *Agron. J.* **2005**, *97*, 872–878. [[CrossRef](#)]
33. Ali, A.M.; Thind, H.S.; Sharma, S.; Singh, V. Prediction of dry direct-seeded rice yields using chlorophyll meter, leaf color chart and GreenSeeker optical sensor in northwestern India. *Field Crops Res.* **2014**, *161*, 11–15. [[CrossRef](#)]
34. Pandey, K.K.; Rai, V.N.; Sisodia, B.V.S.; Singh, S.K. Effect of Weather Variables on Rice Crop in Eastern Uttar Pradesh, India. *Plant Arch.* **2015**, *15*, 575–579.
35. Das, B.; Nair, B.; Reddy, V.K.; Venkatesh, P. Evaluation of multiple linear, neural network and penalised regression models for prediction of rice yield based on weather parameters for west coast of India. *Int. J. Biometeorol.* **2018**, *62*, 1809–1822. [[CrossRef](#)] [[PubMed](#)]
36. Poulton, P.L.; Laing, A.; Poulton, P.L.; Dalglish, N.P.; Vang, S.; Charlesworth, P. Resilience of smallholder farmers in Cambodian lowland rice ecosystems in managing for future climate uncertainty. In Proceedings of the 17th ASA Conference, Hobart, Australia, 20–24 September 2015; pp. 1–4.
37. Yao, Y.; Miao, Y.; Cao, Q.; Wang, H.; Gnyp, M.L.; Bareth, G.; Khosla, R.; Yang, W.; Liu, F.; Liu, C. In-season estimation of rice nitrogen status with an active crop canopy sensor. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 4403–4413. [[CrossRef](#)]
38. Stevens, G.; Wrather, A.; Rhine, M.; Vories, E.; Dunn, D. Predicting rice yield response to midseason nitrogen with plant area measurements. *Agron. J.* **2008**, *100*, 387–392. [[CrossRef](#)]
39. Liu, X.; Ferguson, R.B.; Zheng, H.; Cao, Q.; Tian, Y.; Cao, W.; Zhu, Y. Using an active-optical sensor to develop an optimal NDVI dynamic model for high-yield rice production (Yangtze, China). *Sensors* **2017**, *17*, 672. [[CrossRef](#)] [[PubMed](#)]
40. Marcillo, G.S.; Martin, N.F.; Diers, B.W.; Da Fonseca Santos, M.; Leles, E.P.; Chigeza, G.; Francischini, J.H. Implementation of a generalized additive model (Gam) for soybean maturity prediction in african environments. *Agronomy* **2021**, *11*, 1043. [[CrossRef](#)]
41. Chen, Y.; Donohue, R.J.; McVicar, T.R.; Waldner, F.; Mata, G.; Ota, N.; Houshmandfar, A.; Dayal, K.; Lawes, R.A. Nationwide crop yield estimation based on photosynthesis and meteorological stress indices. *Agric. For. Meteorol.* **2020**, *284*, 107872. [[CrossRef](#)]
42. Donohue, R.J.; Lawes, R.A.; Mata, G.; Gobbett, D.; Ouzman, J. Towards a national, remote-sensing-based model for predicting field-scale crop yield. *Field Crops Res.* **2018**, *227*, 79–90. [[CrossRef](#)]