



Proceeding Paper

# Rice Crop Yield Prediction from Sentinel-2 Imagery Using Phenological Metric <sup>†</sup>

Javier A. Quille-Mamani <sup>1,\*</sup> , Luis A. Ruiz <sup>1</sup> and Lía Ramos-Fernández <sup>2</sup>

<sup>1</sup> Grupo de Cartografía GeoAmbiental y Teledetección, Universitat Politècnica de València, Camí de Vera s/n, 46022 Valencia, Spain; laruiz@upv.es

<sup>2</sup> Departamento de Recursos Hídricos, Universidad Nacional Agraria La Molina, Lima 15024, Peru; liarf@lamolina.edu.pe

\* Correspondence: jaquille@upv.es

<sup>†</sup> Presented at the IV Conference on Geomatics Engineering, Madrid, Spain, 6–7 July 2023.

**Abstract:** Crop yield prediction at plot scale is a vitally important magnitude for farmers at the socio-economic level. This study aims to quantify rice yield using phenological metrics from a normalized difference vegetation index (NDVI) time series derived from Sentinel-2 imagery, with yield data collected from 32 plots with an area of 36 ha in the Ferreñafe District of the Lambayeque region, Peru. Three different rice yield models were obtained, the best linear regression models were obtained for the SVM classification, with  $R^2$  of 0.69, MAE = 1.01 and RMSE = 1.23 t ha<sup>-1</sup>; and MRL with  $R^2$  of 0.61, MAE = 1.10 and RMSE = 1.38 t ha<sup>-1</sup>; RF with  $R^2$  of 0.44, MAE = 1.23 and RMSE = 1.66 t ha<sup>-1</sup>. The models obtained open the possibility to generate more robust models using a larger number of samples, which would be useful for farmers as well as for management and planning decisions for food and economic security.

**Keywords:** NDVI; multispectral imaging; machine learning; time series; precision agriculture



**Citation:** Quille-Mamani, J.A.; Ruiz, L.A.; Ramos-Fernández, L. Rice Crop Yield Prediction from Sentinel-2 Imagery Using Phenological Metric. *Environ. Sci. Proc.* **2023**, *28*, 16. <https://doi.org/10.3390/environsciproc2023028016>

Academic Editors: María Belén Benito Oterino, José Fernández Torres, Rosa María García Blanco, Jorge Miguel Gaspar Escribano, Miguel Ángel Manso Callejo and Antonio Vázquez Hoehne

Published: 2 January 2024



**Copyright:** © 2024 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

One of the world's major cereal crops is rice [1], which meets the nutritional demand and food security of 50% of the world's population [2]. The importance of yield knowledge has been a challenge for future food production due to climatic factors and water scarcity. Remote sensing tools allow yield prediction using satellite imagery (MODIS, LandSat, Sentinel) [3]. However, being Sentinel-2 optical imagery, with higher spatial and temporal resolution it allows to decrease cloud-induced noise. Therefore, it allows the use of different methodologies, including Machine Learning for yield prediction [4].

Northern Peru's contribution to national rice production is distributed in three regions: San Martín with 24%, Piura with 15% and Lambayeque with 13% of the national production. In the Lambayeque region, with a per capita consumption of 63.5 kg, approximately 417,597 ha are planted, representing 40% of the total sown area [5,6], playing a fundamental role from an economic and food security point of view for its inhabitants.

Therefore, and in line with similar studies [3,4,7], this study aims to quantify rice yield using phenological metrics from a Normalized Difference Vegetation Index (NDVI) time series derived from Sentinel-2 images.

## 2. Materials and Methods

### 2.1. Study Area

The study area of this work is the district of Ferreñafe, located in the region of Lambayeque, Peru (79°47'09.73" W; 6°35'36.68" S; 46 m above sea level); with a mean annual rainfall of 22 mm and mean annual temperatures with a minimum average of 15.4 °C and a maximum average of 28.8 °C, with clay soils. The tests were carried out on a total area

of 36 ha in 5 different zones of approximately 5–12 ha each, with a total of 32 plots of rice (Tinajones variety) (Figure 1). Yield samples were taken from each study plot at the end of the growing season. See Table 1.

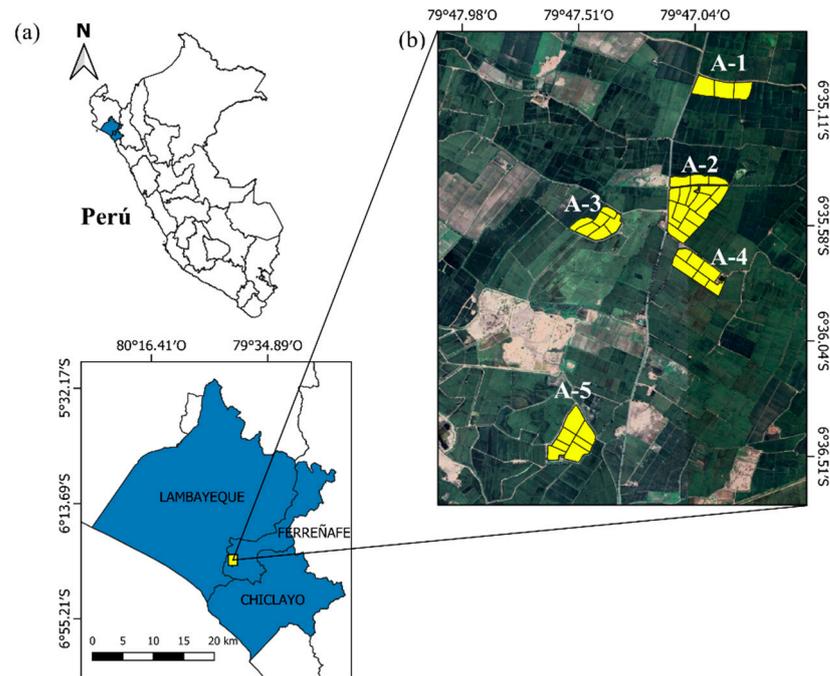


Figure 1. Rice cultivation map of the study region (a) and study rice growing areas (b).

Table 1. Information rice fields.

A-1		A-2		A-3		A-4		A-5	
Area (Ha)	Yield (t ha <sup>-1</sup> )	Area (Ha)	Yield (t ha <sup>-1</sup> )	Area (Ha)	Yield (t ha <sup>-1</sup> )	Area (Ha)	Yield (t ha <sup>-1</sup> )	Area (Ha)	Yield (t ha <sup>-1</sup> )
1.98	10.41	1.11	10.45	0.64	11.63	0.77	11.91	1.99	8.81
1.74	9.66	0.96	10.51	1.14	8.11	0.99	13.26	1.42	7.36
1.58	9.72	0.76	10.90	1.11	10.00	1.38	14.31	1.11	7.87
		0.75	10.75	1.03	12.04	1.49	12.29	0.80	6.28
		1.16	9.91	0.85	10.21	0.95	13.88	1.58	4.98
		1.13	11.06					0.66	5.06
		0.90	10.18					0.99	7.57
		0.78	11.53						
		1.13	10.71						
		1.03	10.81						
		1.02	10.61						
		1.15	8.69						

## 2.2. Data Processing

The Google Earth Engine (GEE) platform was used to extract the NDVI time series from ten Sentinel-2 images, with a spatial resolution of 10 m, considering a filtering of <30% cloud cover. In addition, the mean and maximum NDVI value at a 15-day interval was extracted for the entire rice growing season (December 2021 to June 2022). Subsequently, the extracted values of the NDVI time series were input to generate the phenological metrics in the R programming language with the “CropPhenology” library [8]. In addition, “Weka” software [9] was used for variable selection, model generation and cross-validation, as shown in Figure 2.

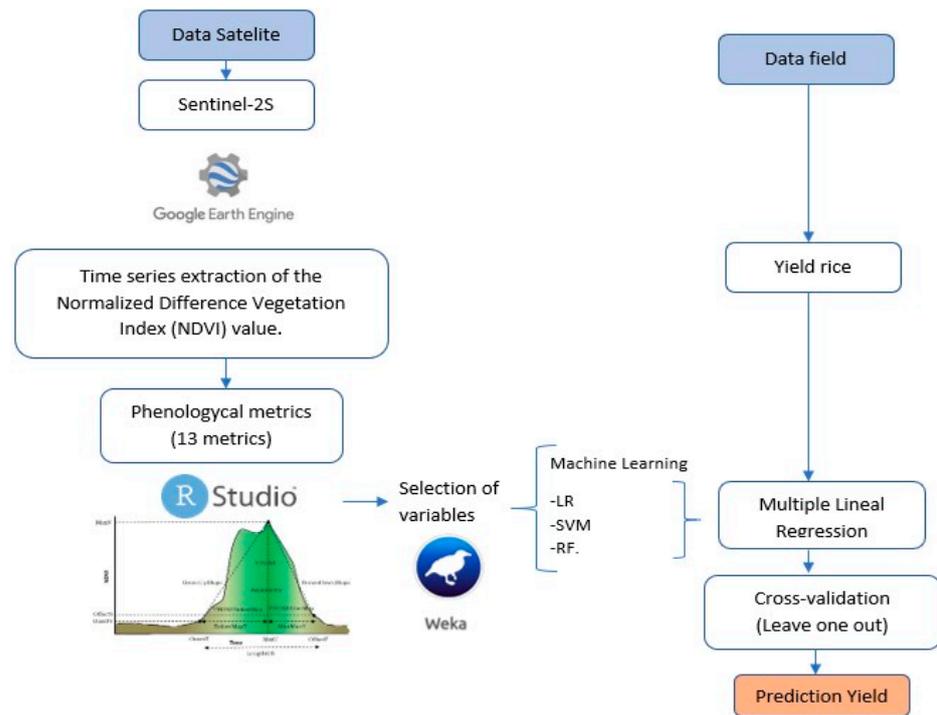


Figure 2. Methodological Working Procedure.

### 2.3. Phenological Metrics

The phenological metrics were extracted considering studies conducted by Araya et al. [8]. They are defined as shown in Figure 3.

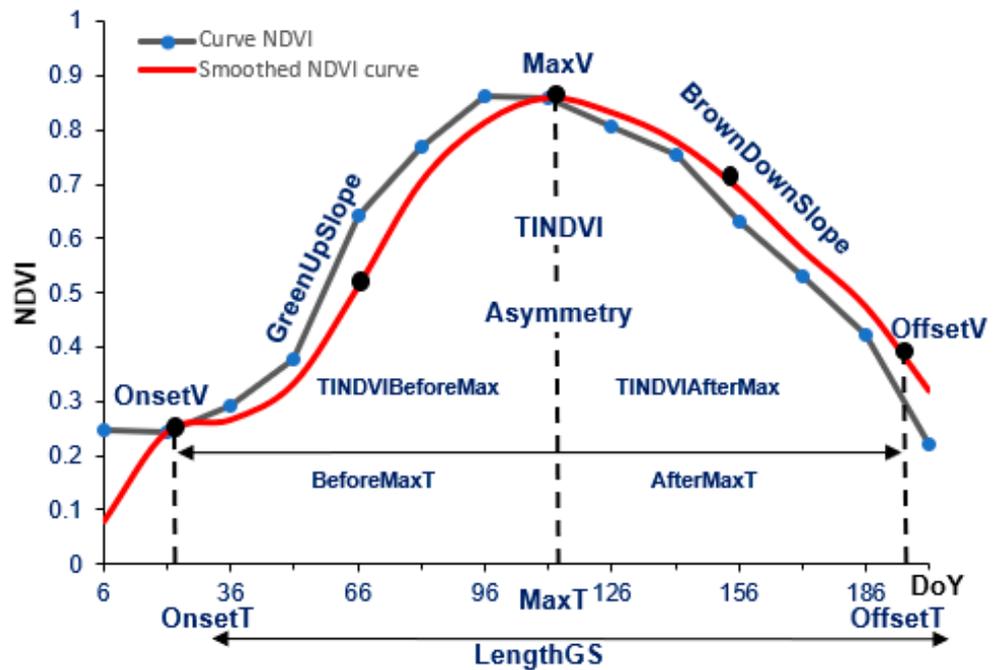


Figure 3. NDVI curve with representation of phenological metrics of rice.

### 2.4. Statistical Analysis

Data analysis was performed by applying Support Vector Machine (SVM), Linear Regression (LR) and Random Forest (RF) for the selection of the metrics, and Multiple Linear Regression (MLR) to obtain rice yield models, which were evaluated by Leave-One-Out Cross-Validation (LOOCV) [10], obtaining the coefficients of determination ( $R^2$ ), the root mean square error (RMSE) and the Mean Absolute Error (MAE).

### 3. Results and Discussion

Figure 4 shows the r-Pearson correlation coefficients between the phenological metric variables extracted from the NDVI time series of the rice crop. Presenting collinearity between the variables “GreenUpSolpe” and “MaxV”, and in the other hand, between “TINDVI” and “TINDVIAfterMax”, with an r-Pearson correlation  $\geq 0.90$ .

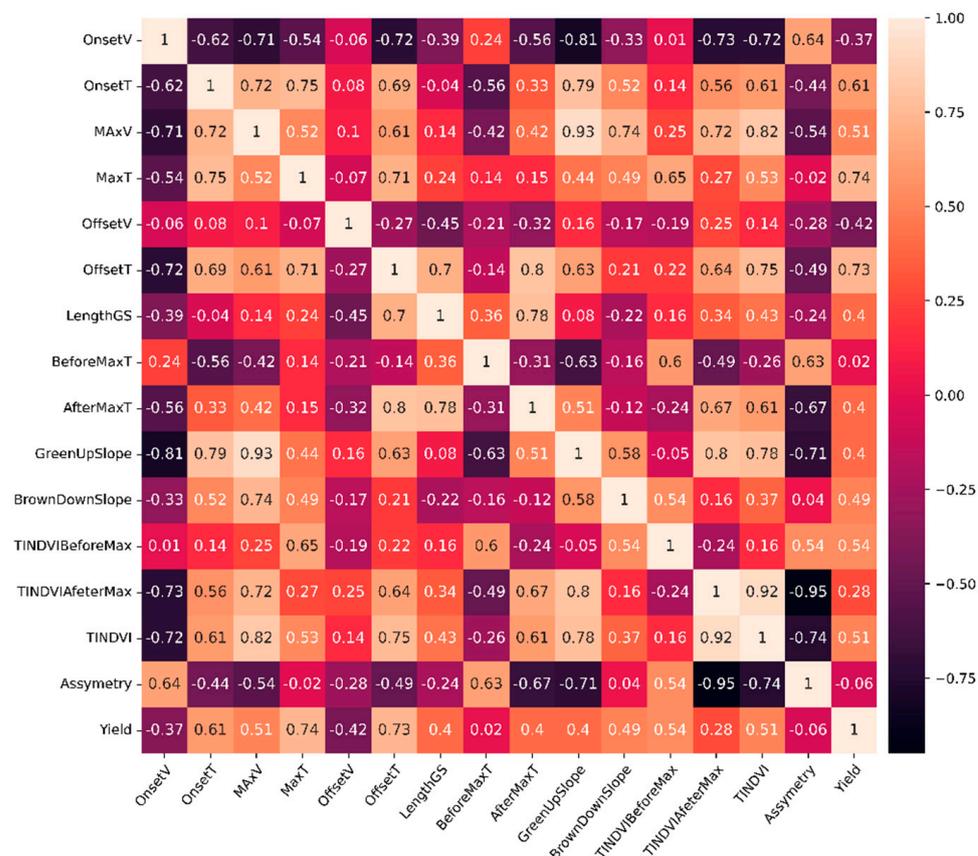


Figure 4. r-Pearson correlation of rice yield and phenological metrics.

Normally, rice yield models are based on NDVI analysis during different phenological stages of the crop, which are reflected in the phenological metrics according to [8]. This allows to generate models by automatic learning, thus obtaining reasonable errors even with few images available due to weather conditions. Therefore, a higher frequency of images could potentially allow a better accuracy in determining the key date for model creation.

The machine learning models selected the following metrics: LR (OffsetT and TINDVIBeforeMax), SVM (OnsetV, MaxT, OffsetT and BrownDownSolpe) and RF (MaxT, TINDVIAfterMax and TINDVI). Table 2 shows the summary statistics of the models generated by the multiple linear regression of the variables selected for each automatic selection method. The best result was obtained using SVM for metrics selection, with  $R^2$  of 0.69, MAE = 1.01 and RMSE = 1.23 t ha<sup>-1</sup>; and LR with  $R^2$  of 0.61, MAE = 1.10 and RMSE = 1.38 t ha<sup>-1</sup>; RF with  $R^2$  of 0.44, MAE = 1.23 and RMSE = 1.66 t ha<sup>-1</sup>.

**Table 2.** Statistical summary of yield prediction models evaluated using cross-validation (leave-one-out), coefficient of determination, root mean square error and mean absolute error.

Method of Selection	Models	R <sup>2</sup>	RMSE (t ha <sup>-1</sup> )	MAE	N-Value
LR	$2.3369 \times \text{OffsetT} + 8.6715 \times \text{TINDVIBeforeMax} - 21.0089$	0.61	1.38	1.10	32
SVM	Multiple Linear Regression $36.4304 \times \text{OnsetV} + 1.4759 \times \text{MaxT} + 3.0196 \times \text{OffsetT} + 74.65 \times \text{BrownDownSlope} - 34.6315$	0.69	1.23	1.01	32
RF	$2.8828 \times \text{MaxT} - 4.9856 \times \text{TINDVIAfeterMax} + 7.2392 \times \text{TINDVI} - 15.0098$	0.44	1.66	1.23	32

#### 4. Conclusions

The present study evaluated three rice yield models, based on the analysis of sentinel-2 NDVI time series, with the extraction of 15 phenological metrics. Three machine learning methods (LR, SVM and RF) were used for variable selection, then multiple linear regressions were generated for each machine learning model. The best models were generated using the following phenological metrics of OnsetV, MaxT, OffsetT, brownDownSlopw, TINDVIBeforeMax, TINDVIAfterMax and TINDVI. It is expected to obtain data from future seasons at the same points in order to develop a more robust model applicable to plots without yield information.

**Author Contributions:** Conceptualization, J.A.Q.-M., L.R.-F. and L.A.R.; methodology, J.A.Q.-M., L.R.-F. and L.A.R.; software, J.A.Q.-M. and L.A.R.; validation, J.A.Q.-M., L.R.-F. and L.A.R.; writing—original draft preparation, J.A.Q.-M., L.R.-F. and L.A.R.; writing—review and editing, J.A.Q.-M., L.R.-F. and L.A.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research has been funded by the projects AGENCIA ESPAÑOLA DE COOPERACION INTERNACIONAL/2020/ACDE/000307 and the scholarship (Generación del Bicentenario) of the Peruvian government for the completion of the thesis (Programa Nacional de Becas y Crédito Educativo (Pronabec) of the Peruvian Ministry of Education (Minedu)).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All data generated during this study are included in this published article.

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

#### References

- Rahman, M.M.; Hasan, S.; Ahmed, R.; Adham, A.K.M. Recycling deep percolated water in continuously flooding irrigated rice fields to mitigate water scarcity. *Paddy Water Environ.* **2022**, *20*, 449–466. [[CrossRef](#)]
- Jiang, Y.; Carrijo, D.; Huang, S.; Chen, J.; Balaine, N.; Zhang, W.; van Groenigen, K.J.; Linquist, B. Water management to mitigate the global warming potential of rice systems: A global meta-analysis. *Field Crop. Res.* **2019**, *234*, 47–54. [[CrossRef](#)]
- Tornos, L.; Domínguez, J.A.; Moyano, M.C.; Recuero, L.; Cicuéndez, V.; García-García, M.J.; Palacios-Orueta, A. Assessment of the SASI Spectral Shape Index Time Series for Mapping Rice Ecosystems in the Mediterranean Region. *Agronomy* **2021**, *11*, 1365. [[CrossRef](#)]
- López-Andreu, F.J.; Erena, M.; Dominguez-Gómez, J.A.; López-Morales, J.A. Sentinel-2 Images and Machine Learning as Tool for Monitoring of the Common Agricultural Policy: Calasparra Rice as a Case Study. *Agronomy* **2021**, *11*, 621. [[CrossRef](#)]
- Aguilar, E.H.; Pando, L.G.; Sosa, G. Utilización de los índices de selección en la identificación de genotipos de arroz (*Oryza sativa* L.) tolerantes a sequía. *Producción Agropecuaria y Desarrollo Sostenible* **2017**, *2*, 11–31. [[CrossRef](#)]
- Porrás-Jorge, R.; Ramos-Fernández, L.; Ojeda-Bustamante, W.; Ontiveros-Capurata, R. Performance assessment of the AquaCrop model to estimate rice yields under alternate wetting and drying irrigation in the coast of Peru. *Sci. Agropecu.* **2020**, *11*, 309–321. [[CrossRef](#)]

7. Kucuk, C.; Taskin, G.; Erten, E. Paddy-Rice Phenology Classification Based on Machine-Learning Methods Using Multitemporal Co-Polar X-Band SAR Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **2016**, *9*, 2509–2519. [[CrossRef](#)]
8. Araya, S.; Ostendorf, B.; Lyle, G.; Lewis, M. CropPhenology: An R package for extracting crop phenology from time series remotely sensed vegetation index imagery. *Ecol. Informatics* **2018**, *46*, 45–56. [[CrossRef](#)]
9. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. The WEKA data mining software: An update. *SIGKDD Explor. Newsl.* **2009**, *11*, 10–18. [[CrossRef](#)]
10. Stone, M. Cross-Validatory Choice and Assessment of Statistical Predictions. *J. R. Stat. Soc. Ser. B (Methodol.)* **1974**, *36*, 111–133. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.