

Estimación de biomasa y rendimiento utilizando métricas fenológicas de imágenes Sentinel-2 en el cultivo de arroz: caso de estudio Perú

Javier A. Quille-Mamani⁽¹⁾, Luis A. Ruiz⁽¹⁾, Lía Ramos-Fernández⁽²⁾, Paulina Raeva⁽³⁾

⁽¹⁾ Grupo de Cartografía GeoAmbiental y Teledetección, Universitat Politècnica de València, Camí de Vera s/n, 46022, Valencia, España. jaquille@upv.es, laruiz@upv.es

⁽²⁾ Departamento de Recursos Hídricos, Universidad Nacional Agraria La Molina, Perú. liarf@lamolina.edu.pe

⁽³⁾ Department of Geomatics, Faculty of Civil Engineering, Czech Technical University, Prague, Czech Republic. paulina.raeva@fsv.cvut.cz

Resumen: La fenología de los cultivos juega un papel vital en la productividad de los cultivos y es un indicador importante en las etapas de crecimiento y desarrollo. Este estudio tiene el objetivo de estimar la biomasa y el rendimiento utilizando métricas fenológicas derivadas de series temporales de imágenes Sentinel-2 en el cultivo de arroz en la región de Lambayeque (Perú). Se seleccionaron 18 fechas de imágenes Sentinel-2 extraídas de la plataforma *Google Earth Engine* (GEE), se calculó el NDVI y posteriormente métricas fenológicas, permitiendo establecer una correlación entre las variables métricas con respecto a la biomasa y el rendimiento. Finalmente, se generaron modelos mediante regresiones lineales múltiples para biomasa con un Error Absoluto Medio (MAE) de 1,71 y 0,26 t/ha, además de una R^2 de 0,51 y 0,71 para el riego por inundación y goteo, respectivamente. En los modelos generados para rendimiento se obtuvo un MAE de 0,26 y 0,21 t/ha, y una R^2 de 0,42 y 0,88 para el riego por inundación y goteo, respectivamente. Ello se plantea como un estudio preliminar para estimar la biomasa y el rendimiento de los cultivos mediante imágenes Sentinel-2.

Palabras clave: series temporales, NDVI, métricas fenológicas, índice de vegetación.

Biomass and yield estimation using phenological metrics from Sentinel-2 imagery in rice crop: a case study in Peru

Abstract: Crop phenology plays a vital role in crop productivity and is an important indicator in growth and development stages. This study aims to estimate biomass and yield using phenological metrics derived from time series of Sentinel-2 images in rice cultivation in the Lambayeque region (Peru). Eighteen dates of Sentinel-2 images extracted from the Google Earth Engine (GEE) platform were selected, the NDVI was calculated and subsequently phenological metrics, allowing to establish a correlation between the metric variables with respect to biomass and yield. Finally, models were generated using multiple linear regressions for biomass with a Mean Absolute Error (MAE) of 1.71 and 0.26 t/ha, in addition to an R^2 of 0.51 and 0.71 for flood and drip irrigation, respectively. In the models generated for yield, a MAE of 0.26 and 0.21 t/ha was obtained, and an R^2 of 0.42 and 0.88 for flood and drip irrigation, respectively. This is proposed as a preliminary study to estimate biomass and crop yield using Sentinel-2 images.

Keywords: time series, NDVI, phenological metrics, vegetation index.

1. INTRODUCCIÓN

El arroz (*Oryza sativa* L.) es uno de los alimentos más importantes de la población mundial, incluidos muchos países de África, América del Norte, América Central, América del Sur y, especialmente, en países de Asia oriental y sudoriental (Ge et al., 2021). El arroz en la región de Lambayeque (Perú), cumple un papel fundamental desde el punto de vista económico y de seguridad alimentaria para sus pobladores. La importancia de conocer el rendimiento, biomasa y otras variables biométricas siempre ha sido un desafío para la futura producción de alimentos.

Las imágenes satelitales constituyen una importante fuente de información para el seguimiento de los cambios temporales en los cultivos, particularmente en áreas extensas. Las trayectorias temporales de índices de vegetación (IV) obtenidas a partir de imágenes de satélite

se relacionan con la densidad y el estado fitosanitario de los cultivos. Se han realizado estudios donde los IV están directamente correlacionados con características como el índice de área foliar, biomasa y la productividad (Djurfeldt et al., 2018; Yang et al., 2022).

En particular, las series temporales del Índice de Vegetación de Diferencia Normalizada (NDVI) se han utilizado para analizar el historial de crecimiento de la vegetación, monitorear las condiciones de crecimiento y predecir la dinámica futura de la vegetación.

La estimación de la biomasa y el rendimiento del arroz han sido de gran importancia para la gestión agrícola, por ello se han realizado diferentes investigaciones sobre modelos de estimación (regresión simple, regresión lineal múltiple, redes neuronales artificiales, *Random Forest*, etc.) en distintas zonas geográficas como Taiwán

(Son et al., 2022), España (Soriano-González et al., 2022) o China (Ge et al., 2021), entre otras.

El objetivo de este estudio es obtener resultados preliminares para la estimación de la biomasa y el rendimiento en el cultivo de arroz en el valle de Lambayeque (Perú) utilizando métricas fenológicas derivadas de imágenes Sentinel-2.

2. MATERIAL Y MÉTODOS

2.1. Zona de estudio

El ámbito de estudio de la investigación es el fundo Zapote Figueroa, provincia de Ferreñafe, Lambayeque-Perú. Ubicado en las coordenadas geográficas de la 6° 35' 47,23" S y 79° 46' 57,52" O, se caracteriza por ser una zona semiárida, con una precipitación media anual de 22 mm y temperaturas medias anual con promedio mínimo de 15,4°C y promedio máximo de 28,8°C, con suelos arcillosos.

Se consideraron dos parcelas de 800 m², con dos sistemas de riego distintos: inundación y goteo. El trasplante del cultivo de arroz se realizó el 16 de enero de 2019 (16 DoY, Day of Year), teniendo un periodo vegetativo de 145 días después del trasplante (DDT).

2.2. Datos de campo y procesado

Se seleccionaron imágenes Sentinel-2 de 18 fechas mediante la plataforma *Google Earth Engine* (GEE), con una resolución espacial de 10 m, a lo largo del periodo vegetativo del cultivo, considerando un filtrado <30% de nubosidad. Además, se realizó el cálculo del NDVI para cada una de las fechas seleccionadas. Posteriormente, se utilizó el software QGIS para delimitar las parcelas. Finalmente, se extrajeron siete píxeles de las imágenes para cada tipo de riego y fecha seleccionada, según se representa en la Figura 1.

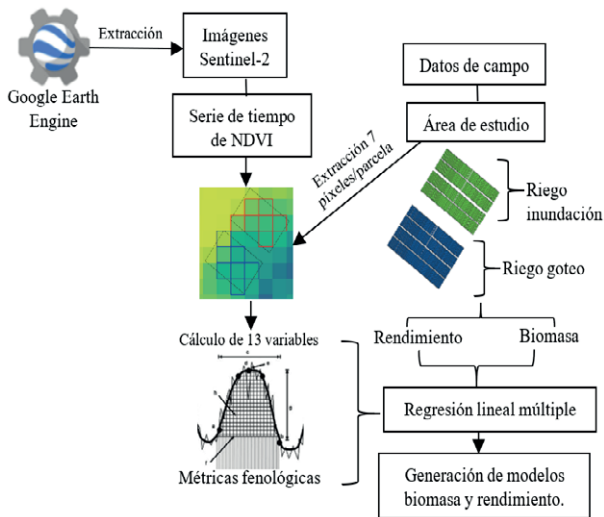


Figura 1. Procedimiento de la metodología de trabajo.

2.3. Extracción de las métricas fenológicas

Las métricas fenológicas se extrajeron considerando estudios realizados por Araya et al. (2017). En la

Figura 2 se representan las variables extraídas de la serie temporal del NDVI y en la Tabla 1 se definen según su utilización en este estudio.

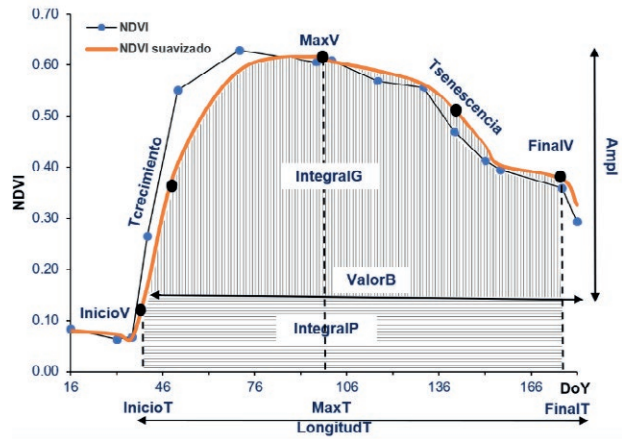


Figura 2. Extracción de las métricas fenológicas a partir de la serie temporal de NDVI.

Tabla 1. Definición de las métricas fenológicas utilizadas.

Métricas	Definición de las variables y descripción fisiológica
InicioT	Momento en que se logra el inicio (borde izquierdo incrementa 10% de la amplitud de la temporada medida desde el nivel mínimo izquierdo).
InicioV	Valor NDVI al inicio del crecimiento, plántula.
MaxV	Valor máximo de NDVI en la temporada.
MaxT	Momento en que MaxV alcanza la etapa de desarrollo antes.
FinalT	Momento en que alcanza el final, la etapa de desarrollo de la senescencia (borde derecho decrece el 10% de la amplitud de la temporada medida desde el nivel mínimo derecho).
FinalV	Valor NDVI al final de la senescencia.
ValorB	Promedio de los valores mínimos del InicioV y FinalV.
Ampl	Diferencia entre MaxV y el ValorB.
LongitudT	Duración de la temporada.
Trecimiento	Tasa de aumento del valor NDVI entre InicioV y MaxV.
Tsenescencia	Tasa de disminución del valor NDVI entre MaxV y FinalV.
IntegralG	Área bajo la curva suavizada entre InicioT y FinalT.
IntegralP	Área por debajo del ValorB desde el InicioT hasta el FinalT.

2.4. Análisis estadístico

El análisis de los datos se realizó determinando la media, desviación estándar (SD), error absoluto medio (MAE) y error absoluto medio normalizado (NMAE). Los modelos de regresión lineal múltiple se generaron con el *software* estadístico Statgraphics (Statgraphics Technologies, Inc., Virginia, EE.UU.).

Los modelos se generaron a partir de regresiones múltiples utilizando las métricas fenológicas como

variables independientes y los datos de biomasa y rendimiento medidos en campo como referencia. Debido a la reducida extensión del área de trabajo y a lo preliminar del estudio, no se pudo contar con muestras suficientes para la validación, que se prevé realizar en trabajos futuros.

3. RESULTADOS Y DISCUSIÓN

En la Tabla 2, se muestran los coeficientes de correlación de Pearson entre las variables fenológicas extraídas de la serie temporal NDVI del cultivo de arroz. Se observa una alta colinealidad entre algunas variables, obteniendo correlaciones >0,95, entre ellas ValorB vs. (IntegralP y FinalV), FinalV vs. IntegralP, Trecimiento vs. MaxV, para el riego por inundación.

Por otro lado, en el riego por goteo las siguientes variables presentaron alta correlación: MaxV vs. Ampl, FinalV vs. ValorB, IntegralG vs. (MaxV y Ampl) y Trecimiento vs. Tsenescencia, permitiendo preseleccionar algunas de ellas con baja correlación para la obtención final de los modelos.

La Tabla 3, muestra una correlación de Pearson de 0,38, 0,32 y 0,39 entre ValorB, IntegralG y FinalV y la biomasa, respectivamente, y una correlación negativa entre el rendimiento y MaxV, Trecimiento, Tsenescencia, InicioV y Ampl de -0,47; -0,49; -0,50; -0,43 y -0,38 en el sistema de riego por inundación. En el riego por goteo se obtuvieron correlaciones positivas de 0,54 y 0,25 para IntegralG y IntegralP, y negativa de -0,22 entre InicioV y biomasa.

Tabla 2. Coeficiente de correlación de Pearson para variables métricas fenológicas, riego por inundación y riego por goteo.

Riego por inundación	ValorB	MaxV	Ampl	IntegralG	IntegralP	Trecimiento	Tsenescencia	InicioV	FinalV
ValorB	1,00	0,24	-0,13	0,34	0,98	0,18	0,69	-0,20	0,98
MaxV		1,00	0,93	0,86	0,18	0,99	0,86	0,63	0,32
Ampl			1,00	0,75	-0,19	0,94	0,62	0,72	-0,04
IntegralG				1,00	0,36	0,81	0,84	0,27	0,46
IntegralP					1,00	0,11	0,64	-0,27	0,96
Trecimiento						1,00	0,82	0,72	0,25
Tsenescencia							1,00	0,30	0,76
InicioV								1,00	-0,25
FinalV									1,00

Riego por goteo	ValorB	MaxV	Ampl	IntegralG	IntegralP	Trecimiento	Tsenescencia	InicioV	FinalV
ValorB	1,00	-0,53	-0,70	-0,59	0,89	-0,25	0,00	-0,52	0,99
MaxV		1,00	0,98	0,96	-0,22	0,84	0,72	0,91	-0,43
Ampl			1,00	0,96	-0,41	0,77	0,61	0,90	-0,61
IntegralG				1,00	-0,22	0,73	0,60	0,85	-0,49
IntegralP					1,00	-0,05	0,17	-0,26	0,92
Trecimiento						1,00	0,97	0,72	-0,15
Tsenescencia							1,00	0,58	0,11
InicioV								1,00	-0,49
FinalV									1,00

Tabla 3. Valores de media, desviación estándar (SD) y correlación estadística de Pearson para las métricas fenológicas con respecto a la biomasa y el rendimiento.

Métrica fenológica	Riego Inundación				Riego Goteo			
	Media	SD	r		Media	SD	r	
			Biomasa	Rendimiento			Biomasa	Rendimiento
ValorB	0,18	0,02	0,38	-0,04	0,17	0,02	0,10	-0,14
MaxV	0,63	0,05	0,02	-0,47	0,53	0,06	0,13	0,11
Ampl	0,44	0,06	-0,08	-0,38	0,36	0,06	0,09	0,17
IntegralG	43,26	7,56	0,03	-0,33	34,24	3,56	0,54	-0,29
IntegralP	24,14	2,15	0,32	-0,24	22,23	3,75	0,25	-0,23
Trecimiento	0,36	0,06	0,17	-0,49	0,33	0,03	0,07	0,23
Tsenescencia	0,47	0,06	0,27	-0,50	0,41	0,04	0,19	-0,07
InicioV	0,12	0,01	0,05	-0,43	0,12	0,01	-0,22	0,76
FinalV	0,34	0,03	0,39	-0,09	0,29	0,05	0,18	-0,27

Tabla 4. Modelos de regresión lineal múltiple de biomasa(B) y rendimiento(R) obtenidos con variables fenológicas.

Tipo de riego	Modelos de estimación de Biomasa(B) y Rendimiento(R)	N° píxeles	MAE (t/ha)	NMAE (%)	R ²	p-valor
Riego Inundación	B= 571,32xValorB + 0,54xIntegralG - 2,88xIntegralP - 45,09	7	1,71	12,36	0,51	0,49 NS
	R= 0,10xIntegralG - 17,96xMaxV + 13,4	7	0,26	4,02	0,42	0,34 NS
Riego goteo	B = 0,96xIntegralG - 45,21xMaxV + 3,22	7	0,77	6,34	0,71	0,09 NS
	R = 87,60xInicioV - 9,88xAmpl -2,67	7	0,21	4,66	0,88	0,01 *

Significancia: NS (No significativo), * (significativo) y ** (altamente significativo).

Para el rendimiento se obtuvieron correlaciones positivas de 0,23 y 0,76 para el Trecimiento e InicioV, además de negativas de -0,29; -0,23 y -0,27 con IntegralG, IntegralP y FinalV.

En la Tabla 4 se recogen los modelos de regresión lineal múltiple para la biomasa y el rendimiento, en cada tipo de riego.

El promedio y la desviación estándar de los valores biomasa y rendimiento para el riego por inundación fueron de 13,84±3,08 t/ha y 6,47±0,47 t/ha, respectivamente. Para el riego por goteo fueron de 12,15±2,10 t/ha y 4,50±0,71 t/ha. La disminución tanto de la biomasa como del rendimiento en el sistema de riego por goteo puede deberse a que el arroz es una planta semiacuática y no tolera una estrés hídrico ocasionado por un sistema de riego controlado.

El error medio absoluto (MAE) obtenido en la estimación de biomasa y rendimiento para el riego por inundación fue de 1,71 t/ha y 0,26 t/ha, respectivamente, con coeficientes de determinación (R²) de 0,51 y 0,42. En el riego por goteo el MAE fue de 0,77 t/ha para biomasa y 0,21 t/ha para rendimiento, con R² de 0,71 y 0,88, respectivamente. Ge et al. (2021), Son et al. (2022), obtuvieron resultados similares en estimaciones de rendimiento, con R² de 0,51 a 0,80 y MAE de 0,38 t/ha a 0,42 t/ha con p-valor de 0,69. Sin embargo, tanto las condiciones abióticas locales como el número de muestras eran distintas.

4. CONCLUSIONES

En este estudio se utilizaron métricas fenológicas obtenidas de series temporales de NDVI para predecir la biomasa y el rendimiento del cultivo de arroz en dos sistemas de riego diferentes (inundación y goteo). Los resultados, aunque preliminares debido a la escasez de muestras de referencia, muestran que el NDVI integrado en el tiempo y las métricas fenológicas derivadas reflejan el efecto estacional en el crecimiento y desarrollo del cultivo. Las ecuaciones obtenidas mediante regresión lineal múltiple muestran una mayor capacidad de las variables ValorB, IntegralG, IntegralP y MaxV para la predicción de la biomasa. En el caso del rendimiento, las variables IntegralG, MaxV, InicioV y Ampl son las que mejor se comportan. Más allá de este estudio preliminar se prevé desarrollar trabajos más exhaustivos que permitan aprovechar al máximo el potencial de los indicadores fenológicos para la estimación del rendimiento y biomasa en áreas más extensas y para diferentes cultivos.

5. AGRADECIMIENTOS

Al grupo de investigación del Área Experimental de Riego (EAR) de la Universidad Nacional Agraria La Molina (UNALM) y a la beca (Generación del Bicentenario) del gobierno peruano para la realización de la tesis (Programa Nacional de Becas y Crédito Educativo (Pronabec) del Ministerio de Educación (Minedu) del Perú).

6. BIBLIOGRAFÍA

Araya, S., Ostendorf, B., Lyle, G., Lewis, M. 2017. Remote Sensing Derived Phenological Metrics to Assess the Spatio-Temporal Growth Variability in Cropping Fields. *Advances in Remote Sensing*, 6, 212-228.

Djurfeldt, G., Hall, O., Jirstrom, M., Archila Bustos, M., Holmquist, B., Nasrin, S. 2018. Using panel survey and remote sensing data to explain yield gaps for maize in sub-Saharan Africa. *Journal of Land Use Science*, 13, 344-357.

Ge, H., Ma, F., Li, Z., Du, C. 2021. Grain yield estimation in rice breeding using phenological data and vegetation indices derived from UAV images. *Agronomy*, 11, 2439.

Son, N.T., Chen, C.F., Cheng, Y.S., Toscano, P., Chen, C.R., Chen, S.L., Tseng, K.H. et al. 2022. Field-scale rice yield prediction from Sentinel-2 monthly image composites using machine learning algorithms. *Ecological Informatics*, 69, 101618.

Soriano-González, J., Angelats, E., Martínez-Eixarch, M., Alcaraz, C. 2022. Monitoring rice crop and yield estimation with Sentinel-2 data. *Field Crops Research*, 281, 108507.

Yang, H., Li, H., Wang, W., Li, N., Zhao, J., Pan, B. 2022. Spatio-Temporal Estimation of Rice Height Using Time Series Sentinel-1 Images. *Remote Sensing*, 14.