

Estimation of corn yield thro ugh image treatments obtained by Sentinel 2: case of Las Arenas, Acambay

Mónica Ivonne García-Martínez¹ Xanat Antonio-Némiga^{1,§} José Francisco Ramírez-Dávila² Adán Salazar-Garibay³

1 Facultad de Geografía-Universidad Autónoma del Estado de México. Ciudad Universitaria s/n. Toluca, Estado de México. México. CP. 50110.

2 Facultad de Agronomía-Universidad Autónoma del Estado de México. Carretera Toluca-Ixtlahuaca km 15.5, El Cerrillo Piedras Blancas, Toluca, Estado de México. México. CP. 50200.

3 Agencia Espacial Mexicana. Insurgentes sur 1685, Álvaro Obregón, Ciudad de México, México. CP. 01020.

Autor para correspondencia: xantonion@uaemex.mx

Abstract

Corn (*Zea mays* L.) is the basis of food and culture in the state of Mexico, so estimating its production to sustain a growing population is a current need. Therefore, images obtained by ESA Copernicus' Sentinel 2 were used to estimate corn yield in plots in the localities of Las Arenas, Acambay, state of Mexico. The efficiency of various indices and biophysical indicators calculated with information from these images was tested to establish their correlation against the harvest measured in the field. The indices calculated in Sentinel 2 were: NDVI and EVI and the LAI and FAPAR indicators. In this region and under conditions of intense drought, the NDVI calculated in Sentinel 2 had the best predictive ability for corn yield (model fit $r^2 = 0.79$). Based on the correlation, the production of 10 randomly selected plots was estimated, demonstrating that, in the range of values between 0.4 and 0.5, NDVI is an excellent predictor of corn harvest under drought conditions, whereas higher NDVI values tend to overestimate yield by up to 1 t ha⁻¹. This information is useful for estimating the harvest and insurance of agricultural production.

Keywords:

biophysical indicators, corn production, vegetation indices.



License (open-access): Este es un artículo publicado en acceso abierto bajo una licencia Creative Commons

Introduction

Estimating the yield of grain corn, in order to guarantee the production required to feed a constantly growing population, is of utmost relevance. Crop yield prediction is a priority to increase knowledge of climate/crop relationships and to generate information that can be used in timely planning and production management (Hernández and Caballero, 2009). Yield estimation techniques diversified with the design of satellite technology, giving rise to a range of possibilities that include, in addition to procedures based on mathematical crop simulation models, those that use satellite information exclusively and others that integrate both (Weiss *et al.*, 2001; Doraiswamy *et al.*, 2003).

The yield of a plot is a multifactorial phenomenon, in which soil and climatic components intervene, interacting with management practices. The characteristic spectral response of plants during their production and growth process is very specific and has been widely documented. In essence, healthy vegetation has a low response in the red region of the visible spectrum and very high in the infrared region of the spectrum (Chuvieco, 2002; Manzo and Meave, 2003). The difference in response between these two regions has led to the development of various vegetation indices, the value of which can be considered an indicator of the state of vegetation health.

Since plant reflectance has a high correlation with crop health status, there is a significant relationship between vegetation spectral indices and their yield (Cruz-Durán *et al.*, 2011). Vegetation indices have been documented to be sensitive to vegetation changes in terms of physiological development (Asrar *et al.*, 1984; Kolotii *et al.*, 2015). Each vegetation index has its own limitations, hence the importance of combining indices and biophysical indicators.

Biophysical indicators are variables that can be used to evaluate vegetation stress and forecast agricultural yields. The leaf area index (LAI) calculates the area of leaves per unit of soil area, whereas the fraction of absorbed photosynthetically active radiation (FAPAR) quantifies the solar radiation absorbed by plants within the photosynthetically active spectral region (Ovando, 2021). LAI is closely related to plant productivity and biomass, a higher value indicates better canopy development and therefore higher potential yield (Hu *et al.*, 2014). For its part, FAPAR is an indicator of photosynthetic activity and vegetation productivity. By being directly related to the energy absorbed by plants for their growth, it can complement other indices (Qin *et al.*, 2018).

Several studies around the world have tested the performance of vegetation indices calculated in Sentinel-2 (Huang *et al.*, 2014; Xu *et al.*, 2022). For example, Xiuliang *et al.* (2020) report that the three-band water index (TBWI) was ideal for estimating corn biomass, establishing the relationship between biomass and leaf area index at various stages of growth. For their part, Chi *et al.* (2022) estimated corn biomass in the province of Jilin, China, by using Sentinel-1 and Sentinel-2 indices and biophysical variables, and they found that the difference of cross-polarizations in Sentinel-1 provides a more accurate estimate of biomass (R^2 = 0.81-0.83, RMSE= 0.4-0.41 kg m⁻²) than models based on simple polarizations, and they combine predictors (optical indices, radar indices, and biophysical indicators) to improve accuracy.

Authors such as Chen *et al.* (2021) combine information from the Sentinel-1 radar with the multispectral images from Sentinel-2 for multitemporal mapping of corn in highly complex and heterogeneous landscapes. Bolton and Friedl (2013) found high correlations between vegetation indices and yield when the crop was fully developed. For their part, Lewis *et al.* (1998) evaluated corn production in Kenya and found that the maximum values of the NDVI index are a sensitive indicator of corn production in the agricultural district. Their results from a simple regression model with NDVI as an independent variable to predict corn production are encouraging (r^2 = 0.75, p< 0.05).

Various remote sensing techniques have been applied to estimate the yield of corn plots in advance (Soria *et al.*, 2004). The present study used data obtained in the field to obtain a more robust model for predicting yield in the municipality of Acambay, state of Mexico. This is because, although satellite images and the indices derived from them provide valuable information, measurements of plant biomass, density, and physiological parameters measured in the field are essential to validate the models, improving their accuracy.



Materials and methods

Study area. The plots monitored were 40 plots of native corn cultivation in well 2 of the agricultural irrigation system of the locality of Las Arenas, municipality of Acambay, state of Mexico. Together, these cover a total area of 63 ha (Figure 1). The area is characterized by a temperate climate with rainfall from June to September, with a maximum rainfall of 160 mm and an average temperature of 15 °C. The dry season includes December, January, and February, whereas the rainy season covers the months from June to October (Figure 2).



Sentinel 2 image processing

Feb

Mar

Abr

May

Jun

Jul

Precipitación 🛛 — Temperatura media 🚽 Temperatura máxima 🚽 Temperatura mínima

Ago

Sep

Oct

Nov

15 12° 9°

> 6° 3° 0

> > Ene

CONAGUA

Sentinel 2B is a multispectral optical sensor from the European Space Agency (ESA) that collects information on the reflectance of the Earth's crust in the visible and infrared regions. The image used was an image obtained on October 13, 2022, the date that corresponds to the maturity of the ears of corn and in which cloud-free information was acquired, a criterion of utmost importance for the calculation of the indices. The scene was downloaded from the Open Access Hub (https:// scihub.copernicus.eu/dhus/#/home) server with an L2 level, which includes radiometric calibration. The image was co-registered with vector mapping and the pixels of the image bands were

80

0

Dic

resampled to a size of 10 m pixel⁻¹ taking advantage of the finest spatial resolution available from the source in the optical bands and using the nearest neighbor method and the scene cut to the study area. These used SNAP's thematic processing functions for optical sensors to calculate the indices and indicators listed below (Table 1). The minimum, maximum, and average values of the indices indicated above in the pixels located at the boundaries of the plots were obtained by means of zonal statistics operations in the QGIS software (QGIS development team, 2024).

Index	Characteristics	Formula			
NDVI	It correlates the visible red and infrared	NDVI=(NIR-Rvis)/(NIR+Rvis)			
	regions and is associated with crop				
	vigorousness (Rouse et al., 1974).				
EVI	In addition to the visible red infrared region, it	EVI=Gvis*($\frac{NIR-Rvis}{NIR+c1*Rvis-C2*Avis+L}$)			
	incorporates the blue band to correct the effect				
	of the atmosphere and soil (Huete et al., 2002).				
LAI	This is generated from EVI by means	LAI=(3.618*EVI-0.118)			
	of correction factors (Chen, 1992).				
FAPAR	Radiation that plants absorb in the range	FAPAR= ND * scale factor + offset ND			
	of the electromagnetic spectrum from 0.4	the reflectance value of the vegetation			
	to 0.7 μm (Vega and Alvarado, 2019).				

Calculation of corn production with data obtained in the field

Field sampling was carried out following the methodology of the International Maize and Wheat Improvement Center (CIMMYT, for its acronym in Spanish) in its corn yield estimation manual (CIMMYT, 2019). Before harvest (from December 01 to 15, 2022), plants were cut and collected from three samples chosen randomly by using random selection methods. The number of rows in each plot was counted and based on a table of random numbers, the number immediately below was chosen to indicate which was the row to collect the sample.

To estimate the distance between furrows to be measured, the average length of the furrow was calculated and divided by 10. Once the furrow to be measured was identified, all the ears found in an area of 5 m were harvested and these were shelled and dried to later be weighed. Dry yield was calculated according to the following formula:

$$Yield\left(\frac{Kg}{Ha}\right) = \left(\frac{\text{total grain weight-moisture}}{\text{area}}\right) * 10$$

For its part, the amount of moisture was estimated by multiplying the total grain weight by the percentage of moisture in the grain. This percentage of moisture was measured with a moisture meter and its value was 14%.

Yield at $14\% \left(\frac{\text{Kg}}{\text{Ha}}\right) = \frac{\text{Dry yield}}{0.86}$

The yield value is extrapolated for the evaluated plots

In this study, the plot is the territorial unit of analysis, that is, the pixel values of the indices obtained from the satellite images were averaged per plot and the yield values sampled in the field are also summed per plot. This is because the spatial resolution of the image prevents a direct relationship between the pixel of the image and the individual (plant) in the field.



Results and discussion

Figure 3 shows the cartography of the indices and indicators calculated in the study area.



The NDVI showed minimum values of 0.17 and maximum values of 0.89, with the values most frequently found between 0.4 and 0.75, which, according to the scale proposed by Alarcón-Rozo (2021), indicates that the plots in the study area have abundant vegetation cover (Figure 4).







The enhanced vegetation index, EVI, showed its average ranges between 0.1 and 0.25, obtaining a minimum value of 0.025 and a maximum of 0.34 (Figure 5).







The leaf area index, LAI, showed minimum values of 0.26 and maximum of 4.7 (Figure 6).







The FAPAR indicator showed minimum values of 0.16 and maximum of 0.89 (Figure 7). This indicator stands out for being the one with the greatest variability between plots.







Relationship of the indices with the yield obtained in the field

As can be seen in Figure 8, the yield in the 40 plots evaluated was highly variable, with yields ranging from less than 1 t ha⁻¹ to 6.2 t ha⁻¹. The size of the plot does not necessarily determine production as there are some small plots with high production and other large ones with low productions, which highlighted the importance of crop management as a decisive factor in production.





A producer from the Pozo 2 ejido applied irrigation in March. However, given the drought conditions reported by CONAGUA, it is not considered a significant variation compared to the other plots. The Table 2 reported indicators of the regression model relating the productivity measured in the field and the indices calculated on the Sentinel 2 image.

Table 2. Statistical properties of corn crop yield regression models.						
Statistic	NDVI	EVI	LAI	FAPAR		
R ²	0.797	0.556	0.591	0.603		
F _{stat}	149	47.5	55	57.7		
p-val	<0.001	<0.001	<0.001	<0.001		
RMSE	0.672	0.993	0.953	0.939		

The index that presented the greatest correlation with the production measured in the field was the NDVI ($R^2 = 0.79$), with an RMSE of 0.672, which indicates a better fit compared to the different indices evaluated (Figure 9); the result is similar to that obtained by Lewis *et al.* (1998), who found a correlation value of $R^2 = 0.75$, concluding that NDVI can be a sensitive indicator of corn production, especially in areas of high spatial variability, as was the case of the study area, where, due to the agricultural management of each plot, the NDVI value presented values ranging from 0.2 to 0.9.



This study found a direct linear relationship between NDVI and the production measured in the field for the plots studied, with a strong correlation (R^2 = 0.79). This allows us to be optimistic about the possibilities of predicting the harvest of grain corn, it is necessary to have more replications that allow this relationship to be sustained (Figure 10).



Figure 10. Estimated corn crop yield in Las Arenas, in October 2022 based on the regression with the NDVI index, derived from ESA's Sentinel-2B image.



The EVI index had a coefficient of $R^2 = 0.55$ in the fit to actual production. This is used in early or late stages where NDVI tends to become saturated, because its correction coefficients reduce the distortion of the atmosphere and soil in reflectances; however, it showed a low relationship with yield.

The same occurred with the biophysical indicators LAI (R^2 = 0.59), FAPAR (R^2 = 0.6). This is explained by the temporality of the satellite image since it corresponds to October, the season in which corn begins its senescence. There are studies that have reported the efficiency of the LAI and FAPAR indicators, they are developed in forested areas with abundant vegetation and dense canopy cover, such as the Monarch Butterfly Biosphere Reserve (Champo *et al.*, 2014). This would explain why, in herbaceous crops such as corn, these indicators are not as efficient in estimating productivity.

Given the high correlation found between the values of NDVI and production measured in the field, a predictive model was run on the image to estimate production considering the equation:

Yield=(18.512*NDVI-7.4722)

To validate the data predicted by the model, estimated yield values were chosen in 10 randomly chosen plots and their values were contrasted against the yield reported by the owners of the plots. Table 3 contrasts the estimated and actual values of corn yield. It was observed that there were plots where the model estimated the yield very accurately, such as number 2 (3.6/3.5) and 4 (3.3/3). Nonetheless, there were others where the model overestimated production by up to 2 t such as plot 5 (7.5/5.5).

of corn yield.				
Plot	NDVI	Estimated yield (t ha ⁻¹)	Actual yield (t ha ⁻¹)	Difference
1	0.7	5.5	4.5	1
2	0.6	3.6	3.5	0.1
3	0.59	3.4	2.5	0.9
4	0.58	3.3	3	0.3
5	0.79	7.2	5.5	1.7
6	0.73	6	5.5	0.5
7	0.76	6.6	5	1.6
8	0.44	0.7	1.5	0.8
9	0.68	5.1	3	2.1
10	0.7	5.5	4.5	1
Average	0.7	4.7	3.9	1
		Overestimation= 26%		

The model tends to overestimate yield when there are high NDVI values (0.7); calculated with average NDVI values (0.5-0.6), its performance is better. On average, the difference between yields is up to 1.5 t, that is, the model varies by up to 26% compared to actual production. Given the strong correlation between the average NDVI value and the estimated harvest, yield could be estimated at least two months before harvest, in agreement with Prasad *et al.* (2006). Replications in other agricultural areas and other climatic conditions are required to confirm this predictive ability of the NDVI index in Sentinel 2. This is because crop yields are driven by a variety of factors, including soil properties, weather conditions, pest and disease pressures, management practices, and crop genetic traits.

Conclusions

In the plots of well number 2 in the locality of Las Arenas, Acambay, state of Mexico and under conditions of intense drought, the NDVI calculated with spectral information from images from the Sentinel 2 sensor had the best predictive capacity for corn production (model fit r^2 = 0.79) so it turned out to be an adequate predictor of the corn harvest under drought conditions, whereas higher NDVI values tend to overestimate yield by up to 1 t ha⁻¹. This information can be useful for the estimation of harvest.

Bibliography

- Asrar, G.; Fuchs, M.; Kanemasu, E. T. y Hatfield, J. L. 1984. Estimación de la radicación fotosintéticamente absorbida y el índice de área foliar a partir de la reflectancia espectral del trigo. Agronomy Journal. 76(2):300-306. https://acsess.onlinelibrary.wiley.com/doi/10.2134/agronj1984.00021962007600020029x.
- Bolton, D. K. and Friedl, M. A. 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agricultural and Forest Meteorology. 173:74-84. 10.1016/j. agrformet.2013.01.007.
- 3 Chen, J. M. and Black, T. A. 1992. Defining leaf area index for non-flat leaves. Plant, Cell & Environment. 15(4):421-429.
- 4 CONAGUA, 2022. Comisión Nacional del Agua. Estaciones meteorológicas automáticas (EMAS). https://smn.conagua.gob.mx/es/observandoeltiempo/estacionesmeteorologicasautomaticasemas.
- Champo-Jiménez, O.; España-Boquera, M. L.; Sánchez-Vargas, N.; Cruz-de-León, J.; Lobit,
 P. y López-Pérez, L. 2014. Construcción de mapas de LAI y fAPAR de la Reserva



de la Biósfera Mariposa Monarca y su comparación con mapas globales. CienciaUAT. 8(2):22-31. https://www.redalyc.org/articulo.oa?id=441942931003.

- 6 Chi, X.; Yangling, D. and Dai, Z. 2022. Estimation of maize biomass components from Sentinel-1 SAR data using multi-target regressors. IEEE International Geoscience and Remote Sensing Symposium. 1392-1395 pp. Doi:10.1109/IGARSS46834.2022.9884054.
- 7 Chuvieco-Salinero, E. 2002. Teledetección ambiental. La observación de la tierra desde el espacio. 1^a. Ed. Editorial Ariel. Barcelona, España. 586 p.
- 8 CIMMYT. 2012. Centro Internacional de Mejoramiento de Maíz y Trigo. Manual de estimación de rendimiento de maíz. Mexico. El Batán, Estado de México. 16-27 pp.
- 9 Cruz-Durán, J. P.; Sánchez-García, A.; Galvis-Spínola, P. y Carrillo-Salazar, J. A. 2011. Índices espectrales en pimiento para el diagnóstico nutrimental de nitrógeno. Terra Latinoamericana. 29(3):259-265.
- 10 Doraiswamy, P. C.; Moulin, S.; Cook, W. P. and Stern, A. 2003. Crop yield assessment from remote sensing. Photogramm. Eng. Remote Sens. 69(6):665-674.
- Hernández, N.; Soto, F. y Caballero, A. 2009. Modelos de simulación de cultivos: Características y usos. Cultivos Tropicales. 30(1):73-82. http://scielo.sld.cu/scielo.php?script=sciarttext&pid=S025859362009000100014&Ing=es&tIng=es.
- Hu, R.; Yan, G.; Mu, X. and Luo, J. 2014. Indirect measurement of leaf area index on the basis of path length distribution. Remote Sensing of Environment. 155:239-247. Doi: 10.1016/ j.rse.2014.08.032.
- Huang, J.; Wang, H.; Dai, Q. y Han, D. 2014. Analysis of NDVI data for crop identification and yield estimation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 7(11):4374-4384.
- Kolotii, A.; Kussul, N.; Shelestov, A.; Skakun, S.; Yailymov, B.; Basarab, R. y Ostapenko, V. 2015. comparación de predictores biofísicos y satélites para predicciones de rendimiento de trigo en Ucrania. Archivos internacionales de fotogrametría, teledetección y ciencias de la información espacial. 39-44 pp.
- 15 Lewis, J. E.; Rowland, J. y Nadeau, A. 1998. Estimating maize production in Kenya using NDVI: some statistical considerations. International Journal of Remote Sensing, 19(13):2609-2617.
- Manzo-Delgado, L. y Meave, J. A. 2003. La vegetación vista desde el espacio. La fenología foliar a través de la percepción remota. Ciencia. 54(3):18-28. https://www.amc.edu.mx/ revistaciencia/images/revista/54-3/vegetacion-vista-espacio.pdf.
- 17 Ovando, G.; de la Casa, A.; Díaz, G.; Díaz, P.; Bressanini, L. y Miranda, C. 2021. Desempeño de diferentes índices de vegetación de Sentinel-2A para estimar el rendimiento de soja en agricultura de precisión. Agriscientia. 38(2):1-12.
- Prasad, A. K.; Chai, L.; Sigh, R. P. and Kafatos, M. 2006. Crop yield estimation model for lowa using remote sensing surface parameters. International Journal of Applied Earth Observation and Geoinformation. 8(1):26-33.
- 19 QGIS 2024. Geographic Information System. Open-Source Geospatial Foundation Project. https://qgis.org.
- 20 Qin, H.; Wang, C.; Zhao, K. and Xi, X. 2018. Estimación de la fracción de radiación fotosintéticamente activa absorbida (fPAR) en los canopies de maíz utilizando datos LiDAR e imágenes hiperespectrales. Plos one. 13(5):e0197510. 10.1371/journal.pone.0197510.
- 21 Soria-Ruiz, J.; Fernández-Ordóñez, Y. y Granados-Ramirez, R. 2004. Metodología para la predicción del rendimiento del maíz utilizando datos satelitales de teledetección en el Centro de México. Investigaciones Geográficas. 55:61-78.

- REMEXCA
- Vega-Araya, M. y Alvarado-Barrantes, R. 2019. Análisis de las series de tiempo de variables biofísicas para cuatro ecorregiones de Guanacaste, Costa Rica. Revista de Ciencias Ambientales. 53(2):60-96. Doi: https://doi.org/10.15359/rca.53-2.
- 23 Weiss, M.; Troufleau, D.; Baret, F.; Chauki, H.; Prévot, L.; Olioso, A.; Bruguier, N. y Brisson, N. 2001. Coupling canopy functioning and radiative transfer models for remote sensing data assimilation. Agric. For. Meteorol. 108(2):113-128.
- Xiuliang, J.; Zhenhai, L.; Haikuan, F.; Zhibin, R. y Shaokun, L. 2019. Deep neural network algorithm for estimating maize biomass based on simulated Sentinel 2A vegetation indices and leaf area index. KeAi Chinese Roots Global Impact. 8(1):87-97.
- Xu, C.; Ding, Y.; Zheng, X.; Wang, Y.; Zhang, R.; Zhang, H.; Dai, Z. and Xie, Q. A. 2022. Comprehensive comparison of machine learning and feature selection methods for maize biomass estimation using Sentinel-1 SAR, Sentinel-2 Vegetation Indices, and Biophysical Variables. Remote Sensing. 14:4083. 10.3390/rs14164083.



Estimation of corn yield thro ugh image treatments obtained by Sentinel 2: case of Las Arenas, Acambay

Journal Information

Journal ID (publisher-id): remexca

Title: Revista mexicana de ciencias agrícolas

Abbreviated Title: Rev. Mex. Cienc. Agríc

ISSN (print): 2007-0934

Publisher: Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias

Article/Issue Information

Date received: 01 January 2025

Date accepted: 01 March 2025

Publication date: 12 May 2025

Publication date: Apr-May 2025

Volume: 16

Issue: 3

Electronic Location Identifier: e3636

DOI: 10.29312/remexca.v16i3.3636

Categories

Subject: Articles

Keywords:

Keywords: biophysical indicators corn production vegetation indices

Counts

Figures: 10 Tables: 3 Equations: 12 References: 25 Pages: 0