

Еje

# Ecofisiología y manejo del cultivo

8, 9 y 10 de Noviembre Pergamino, BA UNNOBA

















# PREDICCION DE RENDIMIENTO DE MAÍZ A PARTIR DE SENSORES REMOTOS

Menéndez Coccoz<sup>1</sup>, M.; Otegui<sup>2</sup>, M.E.; Rotili<sup>1</sup>, D.H.; Di Bella<sup>1</sup>, C.; Piñeiro<sup>1</sup>, G.; Martini<sup>3</sup>, G.; Paolini<sup>3</sup>, M.; Oesterheld<sup>1</sup>, M.

<sup>1</sup>IFEVA, Facultad de Agronomía, UBA-Conicet. Ciudad de Buenos Aires, Argentina. <sup>2</sup>Estación Experimental Agropecuaria INTA Pergamino, Provincia de Buenos Aires, Argentina. <sup>3</sup>AACREA, Ciudad de Buenos Aires, Argentina. Contacto: mamenendez@agro.uba.ar; otegui@agro.uba.ar; rotili@agro.uba.ar; carlos.m.dibella@gmail.com; pineiro@agro.uba.ar; gmartini@crea.org.ar; mpaolini@crea.org.ar; oesterhe@agro.uba.ar

## PREDICTION OF MAIZE GRAIN YIELD BASED ON REMOTE SENSING

## Abstract

Estimating crop grain yield is of great interest to different stakeholders in the private and public sectors. Recently, remote sensors became a promising tool for improving these estimations thanks to their increased spatial and temporal resolutions. In Argentina, maize is one the most widely planted summer crops and its total production has rapidly increased since 2010. The main objective of this work was to develop prediction models for maize grain yield at the field scale based on information provided by satellites. We used a database of 1853 maize fields (DAT-CREA) and calibrated and validated regression models between grain yield and intercepted photosynthetic radiation derived from MODIS sensor. The amount of intercepted radiation during the reproductive phase accounted for large part of grain yield variation (R<sup>2</sup> = 0.59). After adding the production region, the model increased its performance (R<sup>2</sup> = 0.73). Prediction ability increased as the crop cycle progressed and reached a maximum at physiological maturity. In addition, considering the planting date improved the prediction from the earliest stages of the cycle. Consequently, remote sensing allowed to predict maize grain yield from one to four months before harvest with MAE between 0.81 t/ha and 1.37 t/ha, depending on the region.

#### Palabras claves

Satélites, índice verde normalizado, ENSO, fecha de siembra, Monteith.

#### **Keywords**

Satellites, normalized difference vegetation index, ENSO, planting date, Monteith.



## Introducción

Si bien la producción de granos es una de las principales actividades económicas del planeta, no existen aún sistemas completamente objetivos que estimen la producción en tiempo real. Un sistema robusto de seguimiento del estado de cultivos y estimación del rendimiento podría ser muy beneficioso para diversos actores del sector agrícola. Para los productores, una estimación de rendimiento previa a la cosecha permitiría planificar el flujo de ingresos y así diseñar estrategias empresariales. A escala provincial y nacional, el sistema permitiría estimar el volumen a cosechar y establecer previsiones comerciales. Para los gobiernos y ONGs, las estimaciones de rendimiento permitirían movilizar recursos de forma anticipada y eficiente para responder ante eventuales emergencias alimenticias. En definitiva, se requiere que distintos actores, desde productores a nivel local hasta tomadores de decisiones a nivel nacional y global, dispongan de herramientas objetivas y confiables de seguimiento de los principales cultivos en tiempo real.

La mayoría de los estudios sobre predicción de rendimiento de cultivos correlacionan directamente el rendimiento con índices de vegetación provistos por sensores remotos (ej. satélites, Basso y Liu 2019). Sin embargo, el modelo contemporáneo de análisis del rendimiento se basa en la captura y eficiencia en el uso de recursos (Monteith 1972). Este modelo considera que el rendimiento es resultado del producto entre la radiación fotosintéticamente activa incidente, la fracción de ella que es interceptada, la eficiencia en el uso de la radiación (EUR), que convierte radiación interceptada en biomasa, y el índice de cosecha (IC), que representa el rendimiento por unidad de biomasa. Este modelo se utilizó junto con tecnología satelital para analizar el rendimiento de cultivos (Rattallino Edreira et al. 2019). No obstante, aún no existen trabajos de predicción de rendimiento de cultivos en los que se combine la tecnología satelital, el modelo propuesto por Monteith (1972) y una gran base de datos a escala de lote que permita validar regionalmente la metodología.

En Argentina, el cultivo de maíz se produce en ambientes muy contrastantes en cuanto a oferta de recursos. Por esta razón, las tecnologías que mejor se adaptan a cada ambiente también lo son. Por ejemplo, la fecha de siembra es una de las variables de manejo agronómico más importante. Con su elección se puede posicionar el período crítico del cultivo (Fischer y Palmer, 1984) en momentos del año con mayor radiación incidente o un balance hídrico favorable y así aumentar la productividad. En regiones con buena provisión hídrica durante el mes de enero, las siembras tradicionales de septiembre u octubre aumentan los rendimientos. En cambio, en regiones donde la alta demanda atmosférica del mes de enero no es suplida por las precipitaciones o en temporadas en las que la fase "Niña" del fenómeno del ENSO se acentúa, la siembra de maíz "tardío" constituye la mejor tecnología de manejo (Otegui et al., 2021). Debido a estas diferencias, una estimación de rendimiento propia para cada región que tenga en cuenta la fecha de siembra y las fases del ENSO podría ser una buena estrategia para disminuir el error de los modelos de predicción.

El objetivo de este trabajo fue desarrollar modelos de predicción del rendimiento de maíz a escala de lote a partir de información provista por satélites. Los objetivos específicos fueron conocer la medida en que (1) el rendimiento a nivel de lote es predicho por modelos basados en la radiación interceptada inferida a partir de información satelital, (2) los modelos difieren según la región, la fase del ENSO o la fecha de siembra del cultivo y (3) los modelos predicen el rendimiento con distinta anticipación al fin del ciclo del cultivo.



## Materiales y Métodos

#### Región de estudio y base de datos

La región de estudio abarcó las principales zonas productivas de maíz de la Argentina (26° S - 39° S y 57° O - 66° O, Fig. 1). Las condiciones ambientales y las prácticas de manejo del cultivo varían ampliamente dentro de la región (ReTAA, 2021).

Se compaginó una base de datos de lotes georeferenciados de maíz para grano sembrados entre 2017 V 2021 (Base de datos agrícolas trazados de CREA, DAT CREA, https://www.crea.org.ar/dat-crea/). Se conformó una base de datos que para cada lote incluyó la radiación solar incidente, la temperatura media y el índice verde normalizado (IVN). La radiación solar incidente diaria promedio se obtuvo del Instituto Nacional de Pesquisas Espaciais (INPE), mientras que la temperatura media diaria se obtuvo de "NASA POWER". Las series temporales de IVN se extrajeron a partir del producto "MODIS NDVI 16-Day", con una resolución espacial de 250 metros y un dato de IVN cada 16 días.

En definitiva, se trabajó con 1853 registros únicos de lote/campaña que contaron con información sobre año, establecimiento, lote, superficie sembrada, fecha de siembra y rendimiento, fueron delimitables a partir de coordenadas geográficas y no presentaron huecos de más de 40 días en su serie de datos de IVN.



Fig. 1: Región de estudio y ejemplo (recuadro) de delimitación de los lotes de producción. Los puntos rojos refieren a los registros de lote/campaña en toda la región de estudio. En el ejemplo, los puntos blancos refieren a la coordenada geográfica informada por el productor y los polígonos amarillos a límites de los lotes de producción definidos mediante fotointerpretación.

## Estimación del ciclo del cultivo, fases fenológicas y radiación interceptada

A partir de la fecha de siembra y un modelo de acumulación de tiempo térmico (Maddonni 2012), se estimaron tres fases fenológicas para cada lote de producción: siembra - emergencia (S - Ve), emergencia – floración femenina (Ve - R1) y floración femenina – madurez fisiológica (R1 - R6). Además, se determinó el período crítico del cultivo (PC) y el período efectivo de llenado de grano (LLG). El ciclo del cultivo abarcó el período entre emergencia y madurez fisiológica (Ve – R6).

Para cada lote, se calculó la radiación interceptada durante todo el ciclo del cultivo y durante cada fase fenológica o período siguiendo tres pasos. El primer paso consistió en transformar los datos de IVN a eficiencia de intercepción (e;). Esto se realizó siguiendo los modelos propuestos por Gitelson et al. (2014), según la fase fenológica del cultivo. Para la fase Ve – R1 el modelo fue ei = 1,35 IVN – 0,32 (R<sup>2</sup> = 0,95),



mientras que para la fase R1 – R6 el modelo fue  $e_i = 1,89 IVN - 0,78 (R^2 = 0.93)$ . Cuando fue necesario, los valores de el negativos se fijaron en 0, mientras que los mayores a 0,95 se fijaron en 0,95. El segundo paso consistió en calcular la radiación interceptada diaria como el producto entre la eficiencia de intercepción y la radiación incidente diaria. El tercer y último paso consistió en calcular la radiación interceptada durante todo el ciclo y para cada período como la suma de la radiación interceptada diaria correspondiente.

## Análisis de datos

Para cumplir con los dos primeros objetivos específicos se calibraron y validaron regresiones simples y múltiples entre el rendimiento, la radiación interceptada en distintos períodos del ciclo del cultivo y otras variables potencialmente explicativas. La variación entre regiones se evaluó mediante el agregado de la variable categórica "Región CREA". La Región CREA corresponde a una agrupación espacial según condiciones edafoclimáticas y de tecnología comunes entre productores. La fase del ENSO se analizó mediante la clasificación de las cuatro campañas agrícolas en tres categorías: "Niña", "Neutro" y "Niño" (Tokyo Climate Center 2022). Por último, la fecha de siembra se incorporó de forma categórica con dos niveles ("Temprano" o "Tardío"), según la fecha de siembra respecto al 01/12 en las regiones ubicadas hacia el norte del país y respecto al 20/11 en las regiones ubicadas hacia el sur.

Para cumplir con el tercer objetivo específico, se utilizó el modelo de regresión de mayor ajuste del paso anterior para predecir el rendimiento en diferentes ventanas de tiempo posteriores a la siembra del cultivo. Se ocultó toda la información posterior a la fecha evaluada desde los 40 hasta los 150 días posteriores a la siembra, a intervalos crecientes de 10 días. Además, se evalúo si el agregado de la fecha de siembra como "Temprano" o "Tardío" modificaba el poder predictivo de los modelos en las diferentes ventanas de tiempo.

La calibración y validación de los modelos se realizó mediante el procedimiento "Repeated K-fold Cross Validation". Se dividió la base de datos en diez porciones iguales y se calibró y validó cada modelo en diez oportunidades, cada una con una base de validación diferente. Este procedimiento se repitió tres veces, lo que resultó en un total de treinta validaciones independientes. Todo el análisis estadístico se realizó con el software estadístico "R" y el paquete "caret".

## Resultados

## 1. Rendimiento en función de la radiación interceptada

La fase reproductiva fue la que más explicó la variación del rendimiento a partir de la radiación interceptada. La radiación interceptada durante la fase reproductiva (R1 – R6) explicó el 59% de la variación del rendimiento entre lotes y campañas, mientras que la de la fase vegetativa (Ve – R1) solo explicó el 11% (Fig. 2A). A su vez, la radiación interceptada del ciclo completo (Ve – R6) explicó el 47% de la variación del rendimiento, mientras que la radiación interceptada durante el llenado efectivo de grano (LLG) explicó el 54%. La radiación interceptada durante el período crítico (PC) explicó solo el 29% de la variación del rendimiento.



Fig. 2: A) Rendimiento (t/ha) de cada lote/campaña en función de la radiación interceptada durante la fase reproductiva (MJ/m²). La línea roja representa los valores predichos por el modelo. B) Rendimiento observado y rendimiento predicho por el modelo que considera la radiación interceptada durante las fases vegetativa y reproductiva para cada lote/campaña. La línea negra representa la recta y = x. La línea roja representa el ajuste entre rendimiento observado y predicho.

Debido a que fue la variable que más correlacionó con el rendimiento en los modelos simples, la radiación interceptada durante la fase reproductiva (R1 – R6) se utilizó en todos los modelos múltiples. El modelo que más valor explicativo agregó fue el que incluyó la radiación interceptada durante la fase vegetativa (Ve – R1), que explicó el 62% de la variación del rendimiento (Fig. 2B). Este modelo se basó en el mismo período que el anterior descripto (Ve - R6) que explicó el 47% de la variación de rendimiento. El mayor poder explicativo se debió a que permitió pesar diferencialmente a los estados vegetativo y reproductivo.

## 2. Efecto de la región, la fase del ENSO y la fecha de siembra

Para evaluar el efecto marginal del agregado de cada variable se utilizó el modelo de mejor ajuste en el paso anterior (fase reproductiva y fase vegetativa,  $R^2 = 0,62$ ).

## 2.1. Efecto de la región

Con la incorporación de la variable "Región CREA", el R<sup>2</sup> del modelo incrementó de 0,62 a 0,73 (Fig. 3). Por ejemplo, la región "Centro" (CEN) presentó un rendimiento similar que la región "Litoral Sur" (LIS) a valores bajos de radiación interceptada durante la fase reproductiva (200 MJ/m<sup>2</sup>), pero 2,2 t/ha mayor a valores altos (500 MJ/m<sup>2</sup>), cuando la radiación interceptada durante la fase vegetativa fue la misma.

Entre regiones CREA, el R<sup>2</sup> de los modelos varió entre 0,35 y 0,87 (Fig. 4). Además, la variación de rendimiento también difirió entre regiones. Por ejemplo, la región "Noroeste Argentino" (NOA) presentó el menor ajuste (0,35) y solo 25% de variación de rendimiento entre los percentiles 25 y 75. En cambio, la región "Sur de Santa Fe" (SFC) presentó el mayor ajuste (0,87) y una variación de rendimiento entre los percentiles 25 y 75 de 86%. Sin embargo, al comparar el MAE, la región NOA fue la que menor error presentó.





Fig. 3: Rendimiento observado y rendimiento predicho por el modelo para cada lote/campaña según el modelo que incorpora la Región CREA. La línea negra representa la recta y = x. La línea roja representa el ajuste entre rendimiento observado y predicho.



Fig. 4: A) Regiones CREA. B) R2 del modelo que incorpora la radiación interceptada por el cultivo en la fase vegetativa y reproductiva según la región CREA. C) Igual a B pero con MAE.

## 2.2. Efecto de la fase del ENSO

El agregado de la fase del ENSO como variable incrementó marginalmente el R<sup>2</sup> del modelo basado en la radiación interceptada de las fases vegetativa y reproductiva (p < 0,05, Fig. 5). Ante la misma radiación interceptada durante la fase vegetativa y valores bajos de radiación interceptada durante la fase reproductiva (200 MJ/m<sup>2</sup>), los maíces sembrados en años "Niño" rindieron 1,3 t/ha más que los sembrados en años "Niña". En cambio, a valores altos de radiación interceptada durante la fase reproductiva (500 MJ/m<sup>2</sup>), tanto los maíces sembrados en años "Niño" como años "Niña" presentaron rendimientos similares.



Fig. 5: Rendimiento observado y rendimiento predicho por el modelo para cada lote/campaña según el modelo que incorpora la fase del ENSO. La línea negra representa la recta y = x. La línea roja representa el ajuste entre rendimiento observado y predicho.



#### 2.3. Efecto de la fecha de siembra

Al incorporar la fecha de siembra como variable en el modelo, el R<sup>2</sup> incrementó de 0,62 a 0,65 (valor p < 0,05, Fig. 6). En general, los cultivos de maíz "tardíos" rindieron más que los cultivos de maíz "tempranos" con la misma cantidad de radiación interceptada durante la fase vegetativa y la fase reproductiva. Sin embargo, a medida que los cultivos interceptaron mayor cantidad de radiación durante la fase reproductiva, las diferencias de rendimiento se redujeron.



Fig. 6: Rendimiento observado y rendimiento predicho por el modelo para cada lote/campaña según el modelo que incorpora la fecha de siembra. La línea negra representa la recta y = x. La línea roja representa el ajuste entre rendimiento observado y predicho.

#### 3. Predicción de rendimiento en diferentes ventanas de tiempo

Para predecir el rendimiento en diferentes ventanas de tiempo se utilizó el modelo de mejor ajuste en el paso anterior. Este modelo fue el que relacionó el rendimiento con la radiación interceptada durante la fase vegetativa y la fase reproductiva según la región del cultivo ( $R^2 = 0,73$ ). A medida que avanzó el ciclo del cultivo aumentó el poder predictivo de los modelos. Además, el agregado de la fecha de siembra como variable aumentó la precisión de la predicción en las etapas más tempranas del ciclo (Fig. 7). De esta forma, fue posible predecir el rendimiento a los ciento treinta días posteriores a la siembra con una precisión similar a la del momento de fin de ciclo ( $R^2 = 0,66$ ). A medida que se acercó el estado R6, el agregado de la fecha de siembra como variable repercutió menos en la predicción.



Fig. 7: R2 de los modelos de predicción de rendimiento en función del ciclo del cultivo de maíz.



## Discusión y conclusiones

La radiación interceptada durante las fases reproductiva y vegetativa explicó un 62% de la variación del rendimiento a escala de lote en todo el país durante cuatro campañas. Entre estas dos fases, la que más explicó fue la reproductiva (59% de la variación total). Durante esta fase el cultivo define los componentes numéricos del rendimiento: número y peso de los granos. La magnitud de estos componentes varía según la tasa de crecimiento del cultivo, que depende de la radiación interceptada y de la eficiencia en el uso de la radiación y, en definitiva, afecta al índice de cosecha, último término de la ecuación de Monteith (1972). Los satélites permitieron estimar la tasa de crecimiento del cultivo a través de la radiación interceptada y así predecir el rendimiento.

El poder predictivo de la radiación interceptada aumentó cuando los modelos se diferenciaron por regiones. En cambio, no varió según la fase del ENSO de la campaña. Las regiones difieren en condiciones edafoclimáticas y de tecnologías de manejo (ej. fecha de siembra) que serían las causantes de la diferencia regional entre los modelos. Por ejemplo, las diferencias de temperatura entre regiones durante el ciclo del cultivo significan diferencias en la EUR, lo que impacta en la conversión de radiación interceptada en rendimiento. En contraste, la incorporación de la fase del ENSO no mejoró sustancialmente el ajuste de los modelos. Si bien el rendimiento medio fue menor en las fases "Niña", este efecto fue capturado por el sensor a través del IVN y no se reflejó en los residuos del modelo. Las condiciones de estrés hídrico prolongadas de un evento "Niña" reducen la eficiencia de intercepción por acartuchamiento del cultivo durante el período crítico o senescencia anticipada durante el llenado de grano. En menor medida, el estrés hídrico reduce la fotosíntesis y, por ende, la EUR. Además, el índice de cosecha se reduce solo ante episodios puntuales de estrés hídrico durante el período cercano a floración y/o en el llenado de los granos. Dado que la base de datos incluía campañas "Niña", "Neutro" y "Niño", los modelos podrían ser utilizados para predecir el rendimiento de maíz en campañas futuras con cualquier fase del ENSO. En resumen, resulta beneficioso utilizar diferentes modelos de predicción de rendimiento entre regiones y no así entre campañas o fases del ENSO.

Los modelos presentados permitieron estimar con cierta precisión el rendimiento de maíz con anticipación a la cosecha. El mejor modelo recolectó información hasta la madurez fisiológica del cultivo (R6). Sin embargo, el cultivo permanece en el lote hasta alcanzar la madurez comercial. Este período puede extenderse desde uno hasta cuatro meses según la tasa de secado de los granos. De esta forma, los modelos que integraron hasta R6 permitieron predecir el rendimiento de maíz con una gran anticipación a la cosecha y MAE entre 0.81 y 1,37 t/ha según la región. Además, permitieron predecir antes de R6 con poca pérdida de precisión. Por ejemplo, ciento treinta días posteriores a la siembra (aproximadamente veinte días previo a R6) el R<sup>2</sup> del modelo fue bastante similar al del modelo que incluyó toda la información (0,66 vs 0,73). Las estimaciones de rendimiento previas a la finalización del ciclo del cultivo mejoraron sustancialmente cuando se incorporó como variable la fecha de siembra. Esta mejora se debe a que su inclusión es una forma indirecta de adicionar información del ambiente que explora el cultivo como la radiación incidente, la temperatura y la condición hídrica, que podrían afectar tanto la EUR como también el índice de cosecha.

El resultado de este trabajo beneficia a distintos actores del sector agrícola. Los productores podrían estimar fácilmente el rendimiento de sus cultivos. Al inicio de la campaña deberían digitalizar los lotes destinados al cultivo de maíz e informar la fecha de siembra. A partir de ese momento, el



modelo simularía los sucesivos estados fenológicos y recolectaría información satelital de forma periódica. Simultáneamente, se realizarían estimaciones de rendimiento basadas en la fecha de siembra, la región y la radiación interceptada informada por el satélite. Así, el productor contaría con estimaciones periódicas de rendimiento que le permitirían pronosticar ingresos, mejorar la toma de decisiones, etc. En base a estimaciones de lotes puntuales, se podrían escalar los rendimientos y obtener una estimación a nivel regional. La estimación a nivel regional, junto con datos de área sembrada de maíz, otorgarían una estimación del volumen a cosechar. De esta forma, tanto empresas exportadoras como gobiernos y ONGs podrían planificar sus acciones a partir de estimaciones objetivas de rendimiento.

## Apoyo financiero

Este proyecto se realizó en el marco de un convenio entre la Facultad de Agronomía de la Universidad de Buenos Aires y AACREA. Se financió con recursos del PICT "Modelo de Estimación y Seguimiento Regional de la Producción Agrícola" y la beca UBA200 "Predicción de rendimiento de maíz a partir de sensores remotos".



## Referencias bibliográficas

Basso, B.; Liu, L. 2019. Chapter Four - Seasonal crop yield forecast: Methods, applications, and accuracies. Advances in Agronomy 154, 201-255.

Fischer, K.S.; Palmer, F.E. 1984. Tropical Maize. In: Goldsworthy, P.R.; Fisher, N.M. (ed). The Physiology of Tropical Field Crops. Wiley, Chichestor, pp 213- 248.

Gitelson A.A.; Peng Y.; Huemmrich K.F. 2014. Relationship between fraction of radiation absorbed by photosynthesizing maize and soybean canopies and NDVI from remotely sensed data taken at close range and from MODIS 250m resolution data. Remote Sensing of Environment 147, 108 – 120.

Maddonni, G. A. 2012. Analysis of the climatic constraints to maize production in the current agricultural region of Argentina, a probabilistic approach. Theoretical and Applied Climatology 107, 325 -345.

Monteith, J. 1972. Solar radiation and productivity in tropical ecosystems. Journal of Applied Ecology 9, 747–766.

Otegui, M.E.; Riglos, M.; Mercau, J.L. 2021. Genetically modified maize hybrids and delayed sowing reduced drought effects across a rainfall gradient in temperate Argentina. Journal of Experimental Botany 72, 5180 – 5188.

Rattalino Edreira, J. I.; Mourtzinis, S.; Azzari, G.; Andrade, J.F.; Conley, S.P.; Lobell, D.; Specht, J.E.; Grassini, P. 2020. From sunlight to seed: Assessing limits to solar radiation capture and conversion in agro-ecosystems. Agricultural and Forest Meteorology 280, 107775.

ReTAA, Relevamiento de Tecnología Agrícola Aplicada, Bolsa de Cereales. Maíz2020/21. Tokyo Climate Center, Climate Prediction Division, Japan Meteorological Agency. https://ds.data.jma.go.jp/tcc/tcc/