

ESTIMACIÓN DE RENDIMIENTO DE SOJA A TRAVÉS DE IMÁGENES MODIS

*Trabajo final presentado para la Especialización en Teledetección y Sistemas de
Información Geográfica Aplicados al Estudio de los Recursos Naturales y la
Producción Agropecuaria*

Alfonso de Lara

Ingeniero Agrónomo – Universidad Nacional de Mar del Plata - 2011



Escuela para graduados Ing. Agr. Alberto Soriano
Facultad de Agronomía – Universidad de Buenos Aires

COMITÉ CONSEJERO

Director de tesis

Martín Oesterheld

Ingeniero Agrónomo (Universidad de Buenos Aires)

PhD. en Biología (Syracuse University - EEUU)

JURADO DE TESIS

Gervasio Piñeiro

Ingeniero Agrónomo (Universidad de la República Oriental del Uruguay)

Doctor en Ciencias Agropecuarias (Universidad de Buenos Aires - Argentina)

Mariano Oyarzabal

Ingeniero Agrónomo (Universidad de Buenos Aires)

Doctor en Ciencias Agropecuarias (Universidad de Buenos Aires)

Agradecimientos

En primer lugar, quiero agradecer a todo el equipo de la Especialización en teledetección y sistemas de información geográfica aplicados al estudio de los recursos naturales y la producción agropecuaria, por su excelencia académica y su apoyo constante durante la cursada, e inclusive una vez finalizadas ésta.

Particularmente un agradecimiento a Martín Oesterheld, por aceptarme como alumno y guiarme durante la realización la presente tesis.

A Carlos Di Bella, por responder e instruirme ante cada inquietud que se me presentase, inclusive ajenas a la especialización.

Finalmente, a Julieta Straschnoy por ayudarme de distintas maneras.

"Declaro que el material incluido en esta tesis es, a mi mejor saber y entender, original, producto de mi propio trabajo (salvo en la medida en que se identifique explícitamente las contribuciones de otros), y que este material no ha sido presentado, en forma parcial o total, como una tesis en ésta u otra institución."

Resumen

La predicción de rendimientos agrícolas es demandada por distintos usuarios y por ello es motivo de numerosos estudios. Con el gran crecimiento del cultivo de soja, crece la importancia relativa de definir métodos más precisos de estimación y/o predicción de su rendimiento. Tomando como base trabajos hechos con herramientas de teledetección sobre diferentes cultivos y recursos forrajeros, se propone obtener modelos de estimación basados en el Índice Mejorado de Vegetación (EVI) y el Índice de agua superficial terrestre (LSWI) a nivel de lote productivo. La zona específica estudiada fue la del sudeste bonaerense, Argentina, de la cual se contó con numerosa información de lotes de distintos establecimientos en distintos años. Se calculó la radiación absorbida acumulada y se estimó el contenido hídrico foliar en segmentos de 16 días, y se la correlacionó con los rendimientos de campo. Los modelos obtenidos demostraron un buen ajuste con el rendimiento ($r^2 > 0,84$), pudiendo ser estimado con similar precisión ($\pm 300 \text{ kg ha}^{-1}$) tanto a partir de información recolectada hasta solo fines de marzo como durante todo el ciclo de cultivo.

Palabras clave:

Teledetección, EVI, LSWI, soja, predicción, rendimiento.

INDICE GENERAL

Agradecimientos.....	iii
Declaración.....	iv
Resumen.....	v
Índice general.....	vi
Índice de tablas.....	vii
Índice de figuras.....	viii
Estimación de rendimiento de soja a través de imágenes MODIS.....	1
Introducción.....	1
Materiales y métodos.....	3
Datos de rendimientos de soja.....	3
Radiación incidente.....	3
Información satelital.....	4
Selección de lotes por criterio de representatividad de la información satelital.....	4
Radiación absorbida por lote.....	4
Estado hídrico del cultivo.....	5
Sistema de información geográfica.....	5
Análisis estadístico.....	5
Resultados.....	6
Estimación del rendimiento.....	9
Predicción del rendimiento.....	11
Discusión.....	12
Conclusión.....	14
Bibliografía.....	15

INDICE DE TABLAS

Tabla 1. Coeficientes de regresión entre los valores de RFAA acumulados en cada período y el rendimiento observado a cosecha. NA significa que el valor de correlación no fue significativo con un nivel de significancia del 5%. Los colores en los coeficientes van de azul a rojo, de menor a mayor valor	8
Tabla 2. Coeficientes de regresión entre los valores de LSWI de cada período y el rendimiento observado a cosecha. NA significa que el valor de correlación no fue significativo con un nivel de significancia del 5%. Los colores en los coeficientes van de azul a rojo, de menor a mayor valor	8
Tabla 3. Resultados estadísticos de la selección del mejor modelo estimativo del rendimiento de soja según distintos métodos. Descripción de los modelos indicando los segmentos de RFAA e índices de estrés hídrico incluidos, y resultados de validación cruzada del método “LOOCV”	11
Tabla 4. Resultados estadísticos de la selección del mejor modelo predictivo del rendimiento de soja según distintos métodos. Descripción de los modelos indicando los segmentos de RFAA e índices de estrés hídrico incluidos, y resultados de validación cruzada del método “LOOCV”	12

INDICE DE FIGURAS

- Figura 1. Evolución de la radiación fotosintéticamente activa absorbida (MJ m^{-2}) de cada lote en los 13 segmentos calculados dentro de cada ciclo de cultivo. Cada año representa el inicio de dichas campañas, y se visualiza con distintos colores 7
- Figura 2. Dispersión entre los valores de rendimiento y RFAA para tres períodos de acumulación (a partir del 10 de febrero, 26 de febrero y 14 de marzo, por 16 días). Los colores de cada punto dependen de la campaña de análisis. La línea roja grafica la línea de regresión lineal que minimiza la suma de cuadrados 9
- Figura 3. Dispersión entre los valores de rendimiento y LSWI para tres fechas de adquisición (10 de febrero, 26 de febrero y 14 de marzo). Los colores de cada punto dependen de la campaña de análisis. La línea roja grafica la línea de regresión lineal que minimiza la suma de cuadrados 9

Estimación de rendimiento de soja a través de imágenes MODIS

Introducción

La soja (*Glycine max* L.Merrill) ha significado una gran revolución productiva-comercial en la cadena agroalimentaria mundial a partir de los años setenta, ya que es la principal fuente de proteína vegetal para la alimentación animal en sistemas de cría intensiva y a la vez provee aceites para alimentación humana. Adicionalmente, en estos últimos años su demanda continúa creciendo debido a que es parte de la formulación de múltiples productos alimenticios e industriales. Estos abarcan un amplio espectro de alimentos a base de soja, biocombustibles, sustituto proteico de carnes y productos lácteos. La soja es la oleaginosa dominante en el comercio mundial, con un 56% de la producción global (Wilson, 2008). Estados Unidos, Brasil, Argentina, China e India, en ese orden, son los cinco mayores productores, con el 93% de la producción mundial (Steduto et al., 2012). Además de ser un importante ingrediente en las comidas, es una de las principales bases económicas de países agro-exportadores, como es el caso de Brasil y Argentina. El valor de la producción argentina de soja de 2013 rondaría los U\$S 26.000 millones, el máximo de los últimos 12 años (Daniela and Bulgarelli, 2013).

El monitoreo de la condición de los cultivos y la estimación de su producción es importante para la agricultura y los departamentos económicos a nivel municipal, provincial y nacional. Información temprana en el suministro de cultivos es crucial para los responsables políticos y personas involucradas en tomar decisiones sobre mercado e inversión (Esquerdo et al., 2011). La predicción de rendimiento con una anticipación de meses a cosecha podría ser de extrema importancia para una oportuna comercialización, asegurar la demanda nacional y organizar la logística entre países (Bastiaanssen and Ali, 2003). El desarrollo de metodologías para la predicción de rendimiento no es tarea sencilla. El rendimiento de cultivos agrícolas a nivel de lote se caracteriza por poseer una gran variabilidad, resultado de cambios en los factores que afectan su producción (por ejemplo, precipitaciones, temperaturas, y manejo del cultivo). Es por eso que, para obtener una eficaz estimación de rendimientos, es necesario el seguimiento y monitoreo de las variables que tienen mayor impacto sobre el crecimiento y desarrollo de los cultivos.

La productividad primaria neta aérea (PPNA) puede ser estimada a partir del modelo de eficiencias propuesto por Kumar and Monteith (1981). En este modelo, la PPNA se relaciona linealmente con la radiación fotosintéticamente activa (RFA) absorbida por la vegetación (RFAA) y la eficiencia de uso de la radiación (EUR), la cual relaciona la conversión de la RFAA en materia seca. La RFAA es el producto de la RFA incidente (RFAi) y la fracción de RFAi interceptada por la vegetación (fRFAA) (Cristiano et al., 2007). Las imágenes satelitales proveen índices espectrales como el Índice Mejorado de Vegetación (EVI), el cual puede ser convertido linealmente a fRFAA en cultivos de maíz (*Zea mayz* L.) y soja (Bandaru et al., 2013). Por otro lado, las estaciones meteorológicas tienen sensores que miden distintos aspectos de la radiación incidente, entre ellos la RFAi diaria. Por lo tanto, la RFAA puede ser calculada a partir de la combinación de sensores remotos, e información de estaciones meteorológicas. El cálculo de la EUR de los cultivos es uno de los desafíos más importantes en la predicción de rendimientos. Dong et al. (2015) estudiaron cuatro modelos basados en EVI y otras variables para estimar el rendimiento en maíz y soja en condiciones de sequía severa. El modelo que incluía variables relacionadas al estrés hídrico (Índice de agua superficial terrestre, LSWI por sus siglas en inglés) fue el modelo con mejor desempeño. Del mismo modo, Chandrasekar et al. (2010) encontró que el LSWI es útil para estimar el contenido de agua líquida en la cobertura vegetal y detectar sequías.

Finalmente, el rendimiento del cultivo se define como el producto entre la PPNA y el índice de cosecha (IC). En otras palabras, la porción de la materia seca acumulada que se destina a grano. Cabe destacar que el IC de los cultivos es una de las variables más difíciles de estimar, y varios métodos han sido estudiados (Kemanian et al., 2007). Sin embargo, a diferencia de la mayoría de los cultivos, el IC en soja podría suponerse constante para distintos niveles de producción (Andrade and Sadras, 2000). Es decir, ante distintos niveles de acumulación de materia seca, el porcentaje destinado a la producción de granos no variaría en forma significativa. Esto indicaría que el IC podría suponerse como constante sin inducir una fuente de error importante a fines de estimar el rendimiento a cosecha.

Objetivos

El primer objetivo de este estudio fue la calibración de modelos de estimación de rendimiento de soja en la zona sudeste de la Provincia de Buenos Aires, a partir de índices de vegetación y del estado hídrico de la superficie terrestre. Como segundo objetivo se indagó sobre la búsqueda de un modelo de estimación de rendimiento con capacidad predictiva con la mayor anticipación a cosecha posible.

Materiales y métodos

Datos de rendimientos de soja

Se utilizaron datos de rendimiento de lotes agrícolas localizados en la zona sudeste de la provincia de Buenos Aires, República Argentina (partidos de San Cayetano, Necochea y Lobería). Por lo tanto, los lotes analizados tenían características agronómicas variables: distintos tipos de suelos, condiciones climáticas y técnicas de producción (densidades de siembra, fertilización, etc.). Las campañas de cultivo de las cuales se tenían datos de rendimiento a campo fueron: 2007-2008, 2008-2009, 2009-2010, 2010-2011 y 2011-2012. Combinando lotes y campañas, se contó con 109 datos de rendimiento, que luego del proceso de selección que se describe más adelante resultó en un número final de 97.

Radiación incidente

En base a la ubicación “promedio” de los lotes se obtuvieron los datos de radiación provenientes de la estación meteorológica más conveniente, i.e. Estación Experimental Agropecuaria Balcarce (EEAB). Los datos de radiación incidente diarios ($\text{MJ m}^{-2} \text{ día}^{-1}$) fueron obtenidos a partir de la página web oficial del Instituto Nacional de Tecnología Agropecuaria (INTA), en la sección de registro de datos agrometeorológicos (<http://siga2.inta.gov.ar/en/datoshistoricos/>). La radiación fotosintéticamente activa (RFA) se estimó como 47,3% de la radiación incidente total (Papaioannou et al., 1993).

Información satelital

Se digitalizaron los límites de cada lote, dentro de los cuales se establecieron los píxeles puros y semi-puros (mínimo de 95% dentro del lote) del sensor MODIS (sigla en inglés de “Moderate Resolution Imaging Spectroradiometer”), a bordo del satélite AQUA. Los índices utilizados para el cálculo de los predictores de rendimiento fueron el EVI y el LSWI, ambos obtenidos a partir del producto MYD13Q1 del sensor MODIS (Huete et al., 1999). Estos productos satelitales fueron provistos por el laboratorio ORNL DAAC (siglas en inglés de “Oak Ridge National Laboratory Distributed Active Archive Center”). Utilizando sólo píxeles puros y semi-puros, se extrajeron los valores EVI y LSWI (ambos píxeles de 5,37 hectáreas de extensión) a lo largo de cada temporada de cultivo para cada uno de los lotes. Las fechas consideradas abarcaron desde el 24 de octubre de un año hasta el 1 de mayo del siguiente. Dado que la resolución temporal fue de 16 días, cada campaña incluyó un total de 13 índices consecutivos (fechas iniciales correspondientes al 24 de octubre, 9 de noviembre, 25 de noviembre, 11 de diciembre, 27 de diciembre, 9 de enero, 25 de enero, 10 de febrero, 26 de febrero, 14 de marzo, 30 de marzo, 15 de abril y 1 de mayo).

Selección de lotes por criterio de representatividad de la información satelital

Con el fin de comprobar la medida en que con los datos satelitales se representaba una alta proporción del lote, un lote fue incluido cuando al menos la mitad de su superficie era cubierta por píxeles puros y semi-puros. Para esto, se realizó una exploración visual de imágenes Landsat 5, 7 y 8 del mes de febrero (enero en su defecto) de cada campaña de soja, sobreponiendo los píxeles MODIS. De los 109 lotes originales, 97 cumplieron con este criterio de inclusión.

Radiación absorbida por lote

A partir de los 13 índices EVI disponibles en cada campaña, se calcularon 13 segmentos de acumulación de RFAA por cada lote y campaña. A su vez, cada segmento fue calculado como la suma producto de la RFA diaria y el valor único de rRFAA de dicho segmento (relación lineal con el EVI). Se utilizó como fecha de inicio de cada segmento, la fecha que figura en la nomenclatura de cada producto MYD13Q1. Sin embargo, la fecha de adquisición real del valor EVI no es exactamente la descripta

en su nomenclatura. Cada píxel posee una fecha específica de adquisición en base a las mejores condiciones atmosféricas existentes en el periodo comprendido entre un producto y el siguiente. Esto es posible ya que el sensor MODIS en realidad realiza tomas diarias, pero como se mencionó anteriormente, el cálculo de los índices es cada 16 días. Es por ello que para obtener mayor precisión en nuestros análisis se utilizó la fecha real del píxel a la hora de asignar la fRFAA diaria.

Finalmente, combinando los valores de RFAi provistos por la estación experimental con los valores de fRFAA, se calculó la RFAA diaria, y con éstas, la acumulación de RFAA de cada segmento. En caso de contar con más de un píxel puro dentro de los límites de cada lote, se promediaron los valores de RFAA, ya que los datos de cosecha fueron de un solo valor por lote.

Estado hídrico del cultivo

Los valores de LSWI para los mismos periodos de 16 días fueron obtenidos a partir las bandas 5 (NIR) y 7 (SWIR) del compuesto de 16 días del satélite AQUA (MYD13Q1). El cálculo del LSWI se realiza mediante la siguiente fórmula (Chandrasekar et al., 2010):

$$LSWI = [(NIR_{858 \text{ nm}}) - (SWIR_{2130 \text{ nm}})] / [(NIR_{858 \text{ nm}}) + (SWIR_{2130 \text{ nm}})]$$

Procediendo de igual manera que con el EVI, en caso de existir más de un píxel puro por lote, se extrajo la media aritmética.

Sistema de información geográfica

El análisis, procesamiento y extracción de datos de toda la información georreferenciada antes mencionada (imágenes satelitales, límite de lotes, etc.) fue llevado a cabo en el software ArcGIS 10.2.2 (Environmental System Research Institute, Redlands, CA).

Análisis estadístico

Los valores de rendimiento observados a campo fueron la variable a explicar (dependiente), en base a 26 predictores (13 segmentos de RFAA + 13 valores de

LSWI). El software estadístico R (R Core Team, 2016) fue utilizado para realizar análisis de correlaciones, exploración y visualización de la información. Tres métodos de selección de modelos fueron analizados dentro del mismo software, aunque utilizando distintos paquetes de extensión (es decir librerías externas o “plugins”). Dentro de la función “step” del paquete “stats” (R Core Team, 2016), se usaron los métodos “forward” y “stepwise”, ambos partiendo de una situación inicial diferente. El método “forward” elige el primer predictor en base a la mayor correlación con la variable dependiente. Luego se adiciona el segundo predictor que más adicione poder explicativo al modelo, hasta que ya no se incremente la regresión significativamente. Para el caso de “stepwise”, el modelo inicial es con la mitad de los predictores disponibles, y se va agregando o restando de acuerdo a cómo incrementa el poder explicativo. El tercer y más novedoso método de selección fue llevado a cabo utilizando la función “dredge” del paquete “MuMIn” (Barton, 2016). El modelo inicial incluye todos los predictores, por lo que analiza todos los modelos posibles haciendo de éste el método más computacionalmente exigente de los tres. Esta metodología brinda como resultado un ranking de modelos ordenados en base a varios criterios disponibles. En nuestro caso, el criterio seleccionado fue el de “corrected Akaike Information Criterion” (AICc). Este parámetro estadístico relaciona el poder explicativo del modelo (r^2), penalizando por la utilización de cada predictor extra. La versión corregida (“Corrected”) representa la versión corregida, el más apropiado en este caso debido a que la relación entre observaciones y predictores es menor a 40 (Burnham and Anderson, 2003).

Resultados

La evolución temporal de los valores de RFAA de cada segmento, en los lotes finalmente incluidos en el modelo completo, pueden ser observados en la Figura 1. Se aprecia una clara tendencia a partir del 27 de diciembre (día juliano 361), en la cual se incrementan los valores de RFAA en casi la totalidad de los años. Esto es esperable para el cultivo de soja, en cual se observa un marcado crecimiento del cultivo a fines de año. Los máximos valores de absorción se observaron desde el 25 de enero al 14 de marzo (días julianos 25 y 73 respectivamente), lo cual demuestra la plasticidad del cultivo de soja en su fase reproductiva para adaptarse a la oferta hídrica disponible.

Esto fue muy claro para el año 2008, en el cual, comparado con los promedios históricos, las lluvias fueron escasas y tardías.

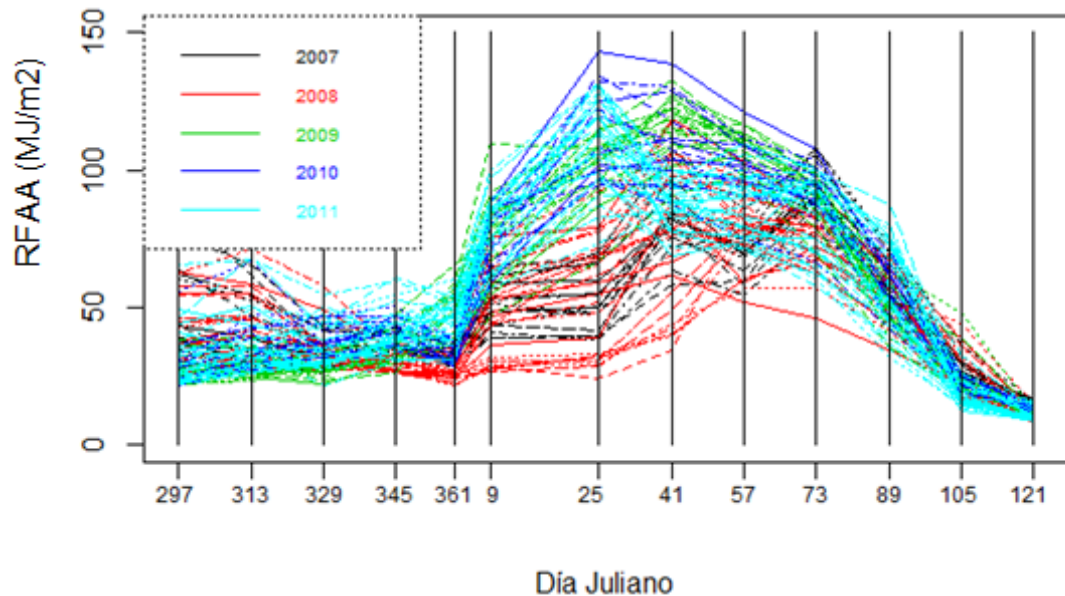


Figura 1. Evolución de la radiación fotosintéticamente activa absorbida (MJ m^{-2}) de cada lote en los 13 segmentos calculados dentro de cada ciclo de cultivo. Cada año representa el inicio de dichas campañas, y se visualiza con distintos colores.

Con el objetivo de explorar y mejorar el entendimiento de la información disponible, se calcularon tablas de correlaciones entre los valores acumulados de RFAA de cada segmento y el rendimiento (Tabla 1), y entre los valores del índice LSWI y rendimiento (Tabla 2). En ambas tablas puede verse cómo las fechas alrededor del período crítico del cultivo son las que más explican el rendimiento, es decir son las fechas más importantes a la hora de la definición del rinde del cultivo.

De acuerdo a estos resultados, el período determinante del rendimiento utilizando sensores remotos comprendería aproximadamente entre el 10 de febrero al 14 de marzo. Cabe destacar que, para el caso de la RFAA, este período se extendería hasta el 29 de marzo, ya que la acumulación de radiación solar es hasta el día anterior del inicio del siguiente segmento. En las Figuras 2 y 3 se puede visualizar los gráficos de dispersión entre los valores de rendimiento y RFAA, y LSWI y rendimiento respectivamente, para el período antes descrito. Efectivamente, se observa una clara correlación positiva entre ambos predictores y el rendimiento. Es decir, ante una mayor cantidad de radiación activa absorbida hay un mayor rendimiento. De la misma manera,

Tabla 1. Coeficientes de regresión entre los valores de RFAA acumulados en cada período y el rendimiento observado a cosecha. NA significa que el valor de correlación no fue significativo con un nivel de significancia del 5%. Los colores en los coeficientes van de azul a rojo, de menor a mayor valor.

RFAA		
Día Juliano	Fecha	Correlación Pearson
297	24-Oct	NA
313	9-Nov	NA
329	25-Nov	NA
345	11-Dec	0,29
361	27-Dec	0,31
9	9-Jan	0,35
25	25-Jan	0,39
41	10-Feb	0,72
57	26-Feb	0,62
73	14-Mar	0,79
89	30-Mar	0,32
105	15-Apr	NA
121	1-May	0,44

Tabla 2. Coeficientes de regresión entre los valores de LSWI de cada período y el rendimiento observado a cosecha. NA significa que el valor de correlación no fue significativo con un nivel de significancia del 5%. Los colores en los coeficientes van de azul a rojo, de menor a mayor valor.

LSWI		
Día Juliano	Fecha	Correlación Pearson
297	24-Oct	NA
313	9-Nov	NA
329	25-Nov	NA
345	11-Dec	NA
361	27-Dec	NA
9	9-Jan	0,34
25	25-Jan	0,53
41	10-Feb	0,64
57	26-Feb	0,64
73	14-Mar	0,59
89	30-Mar	0,50
105	15-Apr	0,54
121	1-May	0,33

ante mayores valores de LSWI, menor estrés hídrico sufrido por el cultivo, y una mayor producción de granos.

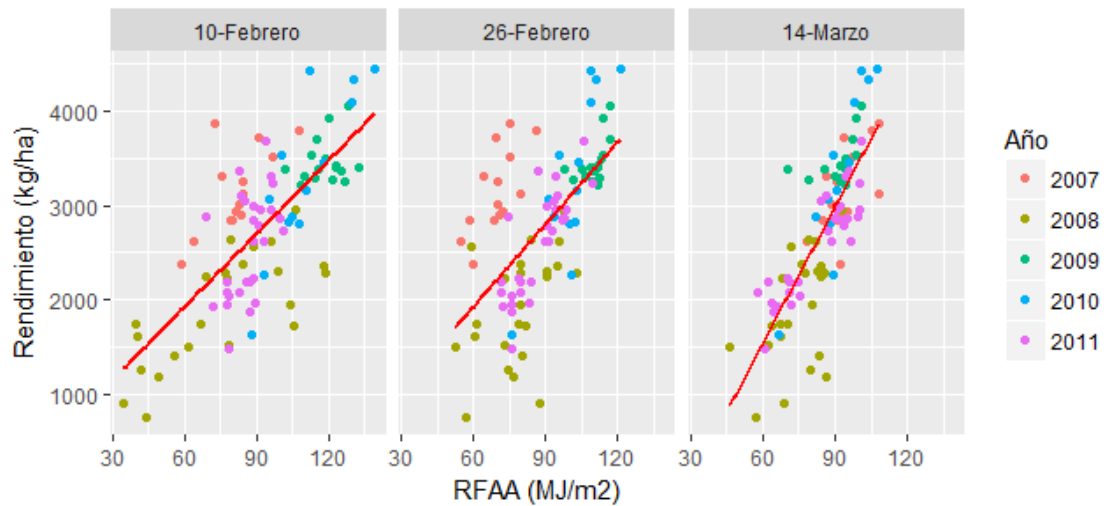


Figura 2. Dispersión entre los valores de rendimiento y RFAA para tres períodos de acumulación (a partir del 10 de febrero, 26 de febrero y 14 de marzo, por 16 días). Los colores de cada punto dependen de la campaña de análisis. La línea roja grafica la línea de regresión lineal que minimiza la suma de cuadrados.

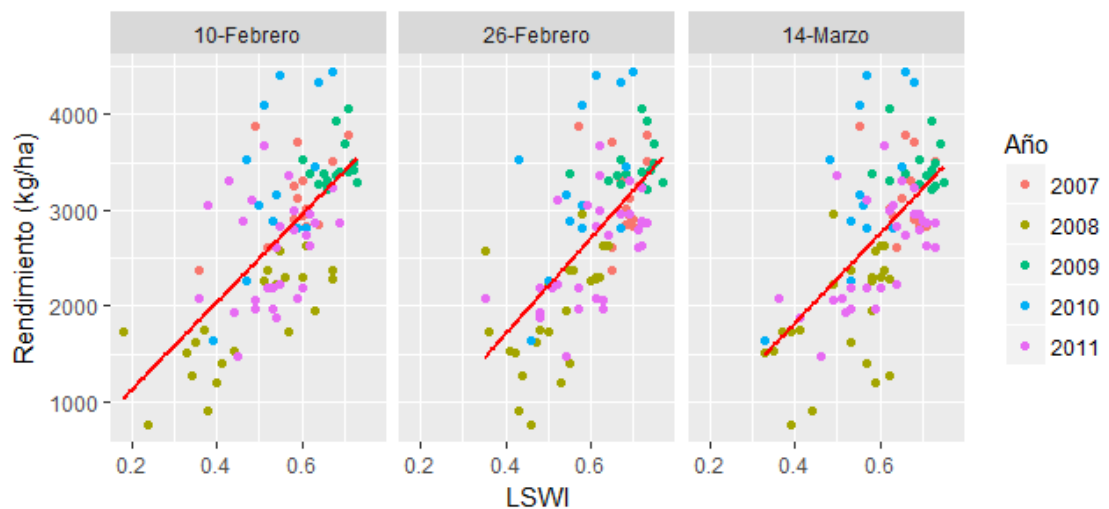


Figura 3. Dispersión entre los valores de rendimiento y LSWI para tres fechas de adquisición (10 de febrero, 26 de febrero y 14 de marzo). Los colores de cada punto dependen de la campaña de análisis. La línea roja grafica la línea de regresión lineal que minimiza la suma de cuadrados.

Estimación del rendimiento

Como primer objetivo, se planteó la estimación del rendimiento usando toda la extensión temporal como información de entrada para los tres métodos de selección del

mejor modelo. En otras palabras, los segmentos de radiación absorbida y valores de LSWI que comprendían desde el 24 de octubre, hasta el 1 de mayo. Debido a restricciones computacionales, la función más completa es decir “dredge”, no fue posible de correr a la hora de hacer la selección automática de modelos. Esto se debe a que la cantidad total a analizar fue de 26, es decir aproximadamente 403 cuatrillones (403×10^{24}) combinaciones. Los resultados estadísticos obtenidos de los métodos de selección “forward” y “stepwise” se detallan en la Tabla 3. En ambos modelos se alcanzó un coeficiente de regresión (r^2) significativamente alto (0,88). El error residual estándar fue de 292 y 290 kg ha^{-1} para los modelos “forward” y “step”, respectivamente. Estos valores significan que, en promedio, los desvíos de las estimaciones de rendimiento fueron alrededor de unas 290-292 unidades por encima o por debajo de lo observado a campo. Por último, las dos últimas columnas de la Tabla 3 describen los estadísticos del desempeño de los modelos selectos en cuanto a la repetitividad de los resultados utilizando una base de datos externa. Para analizar esto se utilizó el método de Validación cruzada dejando-uno-fuera por vez (LOOCV por sus siglas en inglés). Este método consiste en dejar una observación fuera de la modelación, para luego compararla con la estimada y obtener así la diferencia. Este proceso se repite con cada una de las observaciones, obteniendo así el promedio de todas las diferencias. Como puede visualizarse en ambos casos (Tabla 3), el ERS se incrementa y el r^2 disminuye. Esto es lógico ya que en esta ocasión se examina el poder predictivo de los modelos ante hipotéticos nuevos datos. En general podría decirse que ambos métodos tuvieron un buen ajuste, con valores de ERS de 310 y 313 kg ha^{-1} con el método “forward” y “step” respectivamente, y un r^2 de 0,85 y 0,84 respectivamente. En cuanto a los predictores utilizados, se seleccionaron a los del último segmento de acumulación de radiación (R13) entre sus parámetros en ambos métodos. Esto implicaría la menor capacidad de estimación de rendimientos anticipadamente en el tiempo. Importante destacar que ambos métodos solo seleccionaron predictores LSWI hasta el 27 de diciembre. Esto podría explicarse por una mejor lectura de la realidad del lote por parte del LSWI en etapas tempranas (mayor superficie cubierta con suelo en comparación con el cultivo) y colinealidad con el EVI en etapas más avanzadas. Esto último significa que ambos índices explicarían el rendimiento de manera similar, por lo que la información es redundante y solo el más importante es seleccionado.

Tabla 3. Resultados estadísticos de la selección del mejor modelo estimativo del rendimiento de soja según distintos métodos. Descripción de los modelos indicando los segmentos de RFAA e índices de estrés hídrico incluidos, y resultados de validación cruzada del método “LOOCV”.

Método	Estadísticos		Descripción del modelo estimativo‡	LOOCV§	
	ERS†	r ²		ERS	r ²
step.forward	291,80	0,88	~ R5 + R7 + R8 + R9 + R10 + R13 + L3 + L4 + L5	310,43	0,85
step.wise	290,20	0,88	~ R1 + R2 + R5 + R7 + R8 + R9 + R10 + R13 + L3 + L5	312,76	0,84
dredge	-	-	-	-	-

† ERS: Error residual estándar, en kg ha⁻¹.

‡ R, segmento de acumulación de RFAA; L, índice de agua superficial terrestre

§ LOOCV (Leave-one-out at a time cross-validation): Validación cruzada dejando uno fuera por vez

Predicción del rendimiento

Como segundo objetivo, se planteó la predicción del rendimiento filtrando una porción de los predictores como información de entrada para los métodos de selección automática del modelo. Descartar predictores adquiridos en las etapas más próximas al final del cultivo, otorgaría capacidad predictiva a los modelos de estimación. Específicamente, el criterio de filtrado se basó en un balance entre la correlación de los predictores con el rendimiento, y con su anticipación temporal con respecto al ciclo del cultivo. En otras palabras, los predictores más tardíos y menos importantes fueron los primeros en ser removidos. Como se observa en los gráficos de dispersión y en las tablas de correlaciones, el mayor potencial explicativo del rendimiento estaría a partir del 11 de diciembre, hasta el 30 de marzo aproximadamente. Es por ello que en los análisis siguientes se descartaron los índices y segmentos de absorción posteriores al 14 de marzo, es decir al día juliano 73 (R11, R12, R13, L11, L12 y L13). Los resultados de los tres diferentes métodos de selección automática de modelos se detallan en la Tabla 4. En un principio, el método “forward” y “dredge” arrojaron el mismo modelo. De igual manera, los resultados de la función “stepwise” fueron muy similares al resto. En cuanto al ERS, estuvo 7,7 kg ha⁻¹ por encima de los modelos anteriores, aunque con un r² menor. Sin embargo, cuando se realizó el análisis de desempeño de los modelos ante situaciones de nueva información (LOOCV), todos los r² fueron iguales pero el

ERS de “forward” y “dredge” fue 2,2 kg ha⁻¹ mayor. Si bien todos los modelos fueron similares y con buen poder predictivo, estadísticamente el mejor modelo sería el elegido por el método “dredge”. Esto se debe a que, entre todos los modelos posibles, el mejor es el que da como resultado el mayor coeficiente de regresión con la menor cantidad de predictores. Similar a lo observado anteriormente en los modelos de “estimación”, los índices LSWI seleccionados por “forward” o “dredge” no fueron posteriores al 27 de diciembre (9 de enero para “step.wise”).

Tabla 4. Resultados estadísticos de la selección del mejor modelo predictivo del rendimiento de soja según distintos métodos. Descripción de los modelos indicando los segmentos de RFAA e índices de estrés hídrico incluidos, y resultados de validación cruzada del método “LOOCV”.

Método	Mejor modelo		Descripción del modelo predictivo‡	LOOCV§	
	ERS†	r ²		ERS	r ²
step.forward	299,50	0,87	~ R5 + R8 + R9 + R10 + L3 + L5	312,38	0,84
step.wise	294,00	0,86	~ R1 + R2 + R3 + R5 + R8 + R9 + R10 + L3 + L5 + L6	314,60	0,84
dredge	299,50	0,87	~ R5 + R8 + R9 + R10 + L3 + L5	312,38	0,84

† ERS: Error residual estándar, en kg ha⁻¹.

‡ R, segmento de acumulación de RFAA; L, índice de agua superficial terrestre

§ LOOCV (Leave-one-out at a time cross-validation): Validación cruzada dejando uno fuera por vez

Discusión

Similares tendencias fueron encontradas tanto en las tablas de correlaciones entre la RFAA y el LSWI con el rendimiento, como en los coeficientes de los predictores de los modelos de estimación de rinde de soja. Un incremento de éste ante mayores niveles de RFAA, como así también ante mayores valores de LSWI (mayor disponibilidad hídrica del cultivo). Esto concuerda con lo encontrado por Chandrasekar et al. (2010), quien observó mayores valores de LSWI para los píxeles ubicados en zonas de mayor registro de lluvias. Para nuestro estudio, esto quiere decir que el LSWI fue un buen índice a la hora de estimar el contenido hídrico del cultivo, ya que los valores de SWIR_{2130nm} disminuyeron ante mayores contenidos de humedad, por lo tanto, incrementando los valores de LSWI. También siguiendo con lo observado por

Chandrasekar et al. (2010), el índice LSWI tiene un retraso temporal en cuanto a su manifestación con respecto al EVI. Esto se explica debido a que el contenido hídrico del cultivo condiciona el crecimiento futuro de biomasa del cultivo, mientras que el EVI y los índices calculados a partir de éste estiman las condiciones de biomasa actual o ‘instantánea’ del cultivo.

En general, podría concluirse que los distintos métodos de selección de modelos para la estimación y predicción de rinde produjeron resultados aceptables a buenos. Por limitaciones computacionales, el método “dredge” sólo pudo ejecutarse para la “predicción” del rendimiento, devolviendo el mismo resultado que el método “forward”. Esto sugiere que, por lo menos para casos de análisis similares (desarrollo temporal de un cultivo), el método “forward” sería efectivo a la hora de delinear el mejor modelo cuando la función “dredge” no pueda ejecutarse. A pesar de que los valores r^2 obtenidos fueron relativamente altos, todavía faltaría exactitud en la estimación o predicción del rendimiento de soja a nivel de lote de producción. Un desvío promedio aproximado de 300 kg ha^{-1} significa que, si bien en muchos casos se obtendría una estimación o predicción de rinde con un error menor a esa cifra, va a haber casos en el que los que el error se acerque e inclusive sobrepase los 600 kg ha^{-1} . Para mejorar esto se deberían estudiar nuevos parámetros más representativos de las características del lote. Es por ello que, a partir de los resultados obtenidos en este trabajo, se proponen nuevas ideas para incorporar esta información. Como se observa en la Figura 1, aunque puede haber menores efectos intra-año debido a diferentes fechas de siembra en cada lote, los valores máximos de EVI se dan en distintas fechas dependiendo principalmente del año de campaña. Si a esto le sumamos que en las Tablas 1 y 2 las mayores correlaciones de los índices con el rendimiento se dieron justamente en estos picos, todo nos indicaría que para mejorar los modelos se debería tener en cuenta justamente estas curvas de crecimiento variables según años y lotes. Jönsson and Eklundh (2004) sugirieron el uso de un programa de computadora en el cual se analizan las curvas de crecimiento y se puede determinar los momentos de transición entre las distintas etapas fenológicas de los cultivos. Con esta información, podrían estudiarse distintas ponderaciones a los predictores utilizados en este trabajo (e.g. menor ponderación a la RFAA absorbida en estados vegetativos, y/o mayor ponderación al LSWI durante períodos críticos). Otra posibilidad es la mejora en la integración de las condiciones climáticas en los modelos de estimación/predicción. El

índice LSWI fue incluido a modo de estimar las condiciones hídricas del cultivo a lo largo del ciclo. Sin embargo, la dinámica de la evolución del EVI se debe principalmente a las condiciones de humedad y temperatura (Holzman et al., 2014). Es por ello que se sugiere la adición de la temperatura como predictor, o la combinación de esta variable con el LSWI. Lógicamente se podría mejorar la caracterización de las condiciones ambientales del ciclo del cultivo, y por ende una menor dependencia del modelo de estimación de rendimiento del año de análisis en particular.

Conclusión

En resumen, los resultados de nuestra investigación indicarían la posibilidad de combinar información de adquisición remota con información de radiación incidente local, con fines de estimación y predicción del rendimiento de soja a nivel de lote. Una adecuada estimación de los niveles finales de producción de soja a nivel zonal puede ser resuelta a medida que el ciclo de cultivo progresa y la información se vuelve disponible. Sin embargo, todavía no se alcanzó la precisión deseada en la estimación de rendimientos a nivel de lote. Nuevos predictores deberían estudiarse con este fin. Debido a los patrones de evolución de la biomasa observados en los distintos años, se sugiere trabajar en una mejor caracterización de éstos a fines de ganar precisión. Una buena alternativa para ello, podría ser la incorporación de información de temperaturas ya sea de estaciones meteorológicas o fuentes de adquisición remota.

Bibliografía

- Andrade, F.H. and V.O. Sadras. 2000. Bases para el manejo del maíz, el girasol y la sojaINTA, Buenos Aires (Argentina). EEA Balcarce.
- Bandaru, V., T.O. West, D.M. Ricciuto and R.C. Izaurralde. 2013. Estimating crop net primary production using national inventory data and MODIS-derived parameters. *ISPRS Journal of Photogrammetry and Remote Sensing* 80: 61-71.
- Barton, K. 2016. MuMIn: Multi-Model Inference.
- Bastiaanssen, W.G. and S. Ali. 2003. A new crop yield forecasting model based on satellite measurements applied across the Indus Basin, Pakistan. *Agriculture, ecosystems & environment* 94: 321-340.
- Burnham, K.P. and D.R. Anderson. 2003. Model selection and multimodel inference: a practical information-theoretic approach Springer Science & Business Media.
- Chandrasekar, K., M. Sessa Sai, P. Roy and R. Dwevedi. 2010. Land Surface Water Index (LSWI) response to rainfall and NDVI using the MODIS Vegetation Index product. *International Journal of Remote Sensing* 31: 3987-4005.
- Cristiano, P.M., M.E. Beget, C. Di Bella, G. Posse and T. Hartmann. 2007. Estimación de la productividad primaria neta aérea en pastizales de la Región Pampeana utilizando información MODIS y GOES.
- Daniela, C. and V. Bulgarelli. 2013. INFORME ECONÓMICO N° 213.
- Dong, J., X. Xiao, P. Wagle, G. Zhang, Y. Zhou, C. Jin, et al. 2015. Comparison of four EVI-based models for estimating gross primary production of maize and soybean croplands and tallgrass prairie under severe drought. *Remote Sensing of Environment* 162: 154-168.
- Esquerdo, J., J. Zullo Júnior and J. Antunes. 2011. Use of NDVI/AVHRR time-series profiles for soybean crop monitoring in Brazil. *International Journal of Remote Sensing* 32: 3711-3727.
2014. Spatial crop yield estimation based on remotely sensed stress index. XLIII Jornadas Argentinas de Informática e Investigación Operativa (43JAIIO)-VI Congreso Argentino de AgroInformática (CAI)(Buenos Aires, 2014).
- Huete, A., C. Justice and W. van Leeuwen. 1999. MODIS vegetation index (MOD13) algorithm theoretical basis document. Greenbelt: NASA Goddard Space Flight Centre, <http://modarch.gsfc.nasa.gov/MODIS/LAND/#vegetation-indices>.
- Jönsson, P. and L. Eklundh. 2004. TIMESAT—a program for analyzing time-series of satellite sensor data. *Computers & Geosciences* 30: 833-845.
- Kemarian, A.R., C.O. Stöckle, D.R. Huggins and L.M. Viega. 2007. A simple method to estimate harvest index in grain crops. *Field Crops Research* 103: 208-216.
- Kumar, M. and J. Monteith. 1981. Remote sensing of crop growth. *Plants and the daylight spectrum*: 133-144.
- Papaioannou, G., N. Papanikolaou and D. Retalis. 1993. Relationships of photosynthetically active radiation and shortwave irradiance. *Theoretical and Applied Climatology* 48: 23-27.
- R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2016. ISBN 3-900051-07-0.
- Steduto, P., T.C. Hsiao, D. Raes and E. Fereres. 2012. Crop yield response to water Food and Agriculture Organization of the United Nations Rome.
- Wilson, R.F. 2008. Soybean: market driven research needs. *Genetics and genomics of soybean*. Springer. p. 3-15.