

**Desarrollo de un método de estimación
de rendimiento de cultivos agrícolas utilizando imágenes
satelitales ópticas en la provincia de Buenos Aires**

por Ing. Agr. Emiliano Alvarez Zanelli

Presentado ante la Facultad de Matemática, Astronomía, Física y Computación
y el Instituto de Altos Estudios Espaciales Mario Gulich
como parte de los requerimientos para la obtención del grado de

MAGISTER EN APLICACIONES DE INFORMACIÓN ESPACIAL

UNIVERSIDAD NACIONAL DE CÓRDOBA

Agosto, 2020

©IG - CONAE/UNC 2020

©FAMAF - UNC 2020

Director: Agrim. Mauro Giovanniello (ARBA)

Co-director: Lic. Mario Lanfri (CAEARTE)



Esta obra está bajo una <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Resumen:

El siguiente trabajo surge de una demanda concreta de la Agencia de Recaudación de la provincia de Buenos Aires (ARBA) con el objetivo de mejorar la calidad de los datos generados en la estimación del rendimiento de cultivos agrícolas extensivos. El alcance territorial de ARBA incluye toda la provincia de Buenos Aires (307.571 km²) y en este trabajo de investigación el área de estudio se reduce al partido de Daireaux (3.820 km²) que tiene como principal actividad productiva la agrícola-ganadera.

Se utilizaron datos de campo provenientes de monitores de rendimiento (incorporados en las máquinas cosechadoras) y se vincularon con secuencias multitemporales de tres índices de vegetación diferentes (NDVI, MSAVI, EVI) y un índice de humedad (NDMI), utilizando la serie de satélites ópticos Landsat.

Se desarrollaron modelos de regresión lineal vinculando los datos satelitales con los valores de rendimiento de cuatro cultivos estivales: soja de primera, soja de segunda, maíz y girasol; bajo cuatro maneras de vincular los datos, y en dos campañas agrícolas: una correspondiente a un ciclo hídrico de sequía (2008/2009) y otra correspondiente a un ciclo hídrico normal/húmedo (2010/2011). Las diferentes variantes de índices utilizados y los métodos de combinación de la información se analizaron con el objetivo de determinar cuál resultaba más adecuado para la estimación del rendimiento en cada tipo de cultivo. Finalmente se obtuvieron ocho modelos de estimación de rendimiento (cuatro tipos de cultivo en ciclo hídrico seco y ciclo normal/húmedo). El RRMSE (Relative Root Mean Squared Error) de los modelos en ambas campañas analizadas promediaron el 26.5% para la Soja de primera, el 19.5% para la Soja de segunda, el 23% para el Maíz y el 18% para el Girasol, cumpliendo de esta manera con los objetivos fijados.

Se aplicaron los modelos validados para el ciclo normal/húmedo, sobre la campaña agrícola 2016/2017 utilizando los sensores Landsat disponibles actualmente (Landsat 7 y 8) obteniendo el rendimiento estimado para cada tipo de cultivo analizado, y clasificado en rangos de calidad (requerimiento de ARBA), en cada parcela agrícola cultivada del partido de Daireaux.

Se propuso una metodología para la estimación del rendimiento de cultivos que se adecúa al procesamiento actual que realiza el Departamento de Tecnología de Imágenes de ARBA y en el cual se aplican los modelos desarrollados.

Summary:

The work arises from a specific demand of the Agencia de Recaudación de la Buenos Aires (ARBA) of improvement on yield estimation methods for extensive agricultural crops. The territorial scope of ARBA includes the whole province of Buenos Aires (307.571 km²) but in this work, the area studied is limited to Daireaux Department (3.820 km²), whose main productive activity is agriculture and livestock.

Field data from yield monitors (incorporated in harvesting machines) were used and linked to multi-temporal sequences from three vegetation indices (NDVI, MSAVI, EVI) and a humidity index (NDMI), using Landsat optical satellites.

Linear regression models were developed by linking the satellite data with the yield values of four summer crops: soybean 1°, soybeans 2°, corn, and sunflower; four different ways to link the data, and two agricultural campaigns: one corresponding to a drought hydrological cycle (2008/2009) and another corresponding to a normal/humid hydrological cycle (2010/2011). Variability of vegetation indexes and data combination methods were analyzed to determine which was the most suitable for yield estimation for each crop type. Finally, eight models of yield estimation were developed (one for each crop cultivate under both the hydrological cycle conditions). The RRMSE (Relative Root Mean Squared Error) of the final models for both campaigns averaged 26.5% for the first soybeans, 19.5% for the second soybean, 23% for the corn and 18% for the Sunflower, thus fulfilling the objectives previously set.

The validated models were applied to the 2016/2017 agricultural campaign using the Landsat sensors currently available (Landsat 7 and Landsat 8), obtaining the estimated yield for each type of crop analyzed and classifying them into quality ranges (ARBA requirement) within each parcel of the department of Daireaux.

The methodology proposed for crop yield estimation is based on the developed models and perfectly adapts to the current processing carried out by Departamento de Tecnología de Imágenes de ARBA.

Palabras Clave:

Estimación de Rendimiento, Cultivos Extensivos, Serie Landsat, Índices de Vegetación, Fiscalización satelital.

Agradecimientos:

Agradezco a la Agencia de Recaudación de Buenos Aires y a mis compañeros de trabajo, por haberme brindado la oportunidad de realizar la maestría, el tiempo necesario para el desarrollo de la tesis, información requerida, acompañamiento y equipos para su desarrollo.

Al Instituto Mario Gulich y a todo el plantel docente y no docente que me acompañó en esas largas jornadas de estudio que transité en las hermosas sierras de Córdoba.

A la Scuola di Ingegneria Aerospaziale de la Università la Sapienza di Roma y especialmente a Giovanni Laneve y Pablo Marzialetti que son personas excelentes que me asesoraron y acompañaron permanentemente en mi estadía en Italia.

Quiero agradecer también y enfatizar el compromiso que asumió Mauro Giovanniello para acompañarme en este proceso de aprendizaje como director de este trabajo de tesis, así como también, el que asume él diariamente por mejorar la eficiencia de los procesos que se realizan desde el estado, con una gran responsabilidad y tendiente a promover la mejora continua y la formación de recursos humanos.

A mi codirector Mario Lanfri por su dedicación, sugerencias y atención ante las consultas técnicas en cualquier momento del día.

Al tribunal evaluador que ha hecho sugerencias y correcciones significativas al trabajo.

A los doctores Germán Soracco y Luis Lozano del grupo de Física Aplicada de la Facultad de Ciencias Agrarias y Forestales (UNLP), por su acompañamiento y permanente buen humor ante mis consultas de modelización.

Agradecer a mis compañeros de maestría, que han promovido en este largo y complejo proceso un ambiente de estudio agradable con reuniones y momentos inolvidables.

A la Dra. Marta Lydia Zanelli, mi querida tía que supo acompañarme ante tantas consultas relacionadas a la estadística y la biometría.

A mi compañera de vida, Micaela Trucco que es la persona que transitó conmigo el día a día de la maestría y me acompaña siempre.

Índice General:

1. CAPITULO I	10
1. Introducción.....	10
1.1 ARBA y la actividad agrícola en la provincia	10
1.2 Estimación del rendimiento de cultivos con datos de satélite	11
1.3 Propuesta de mejora en la estimación del rendimiento	13
1.4 Objetivo general.....	14
1.5 Objetivos específicos.....	15
1.6 Resultados Esperados.....	15
2. CAPITULO II.	16
2. Materiales y métodos.	16
2.1 Sitio de estudio.....	16
2.2 Información de campo.....	17
2.2.1 Atributos de los datos de campo.	19
2.2.2 Limpieza de los datos de campo.	20
2.2.2.1 Método de Filtrado 1.	21
2.2.2.2 Método de Filtrado 2.	22
2.2.3 Cálculo del rendimiento por lote.	22
2.2.4 Análisis estadístico comparativo entre los métodos de filtrado.	23
2.2.5 Cálculo del rendimiento por pixel.	25
2.2.5.1 Generación de los polígonos de lotes cultivados.....	25
2.2.5.2 Generación de la grilla.....	25
2.2.5.3 Estimación del rendimiento a nivel pixel.	26
2.2.6 Rasterización de datos de campo.....	27
2.2.7 Control de calidad y división del set de datos (modelado y validado).	28
2.3 Información Satelital.....	29
2.3.1 Adquisición de datos.	29
2.3.2 Formación de la imagen.....	31
2.3.3 Extracción de nubes/sombras/fallas.....	33
2.3.4 Generación del pack por índice.	33
2.3.5 Reproyección.....	35
2.3.6 Control de calidad de los datos raster.....	35
3. CAPITULO III.	36
3. Desarrollo de los modelos.....	36
3.1 Conformación de la base de datos.....	36
3.2 Función de crecimiento.	39
3.2.1 Determinación del Green-Up y Green-Down.....	40
3.2.2 Nuevo filtrado de datos.....	41
3.2.3 Períodos críticos.	42
3.2.3.1 Período crítico Soja 1°.	43
3.2.3.2 Período crítico Soja 2°.	44
3.2.3.4 Período crítico Girasol.....	46
3.3 Modelo de Regresión Lineal.	47
3.3.1 Revisión del proceso de modelización.	50
3.4 Determinación de los modelos con mejor ajuste.	54
3.4.1 Modelos seleccionados para períodos de escases hídrica (2008/2009).....	58
3.4.1.1 Modelo seleccionado Soja 1.	58
3.4.1.2 Modelo seleccionado Soja 2.	59
3.4.1.3 Modelo seleccionado Maíz.....	61

3.4.1.4 Modelo seleccionado Girasol.....	62
3.4.2 Modelos seleccionados para períodos hídricos normales/húmedos (2010/2011).....	65
3.4.2.1 Modelo seleccionado Soja1.....	65
3.4.2.2 Modelo seleccionado Soja2.....	66
3.4.2.3 Modelo seleccionado Maíz.....	68
3.4.2.4 Modelo seleccionado Girasol.....	69
3.5 Validación de los modelos seleccionados.....	72
3.5.1 Soja 1 – 2008/2009.....	73
3.5.1.1 Análisis del error del modelo Soja 1 – 2008/2009.....	73
3.5.2 Soja 2 – 2008/2009.....	74
3.5.2.1 Análisis del error del modelo Soja 2 – 2008/2009.....	74
3.5.3 Maíz – 2008/2009.....	75
3.5.3.1 Análisis del error del modelo Maíz – 2008/2009.....	75
3.5.4 Girasol – 2008/2009.....	76
3.5.4.1 Análisis del error del modelo Girasol – 2008/2009.....	76
3.5.5 Validación Soja 1 – 2010/2011.....	77
3.5.5.1 Análisis del error del modelo Soja 1 – 2010/2011.....	77
3.5.6 Validación Soja 2 – 2010/2011.....	78
3.5.6.1 Análisis del error del modelo Soja 2 – 2010/2011.....	78
3.5.7 Validación Maíz – 2010/2011.....	79
3.5.7.1 Análisis del error del modelo Maíz – 2010/2011.....	79
3.5.8 Validación Girasol – 2010/2011.....	80
3.5.8.1 Análisis del error del modelo Girasol – 2010/2011.....	80
3.6 Cualificación del rendimiento estimado.....	81
4. CAPITULO IV.....	85
4. Implementación.....	85
4.1 Utilización de los modelos de ciclo hídrico normal/húmedo en 2016/2017.....	85
4.2 Desarrollo del procedimiento para la implementación.....	85
5. CAPITULO V.....	91
5. Resultados y Conclusiones.....	91
5.1 Resultados de la generación de las funciones de crecimiento.....	91
5.2 Resultados de los modelos desarrollados.....	93
5.3 Resultados de la implementación.....	94
5.4 Resultados del método propuesto.....	96
5.4.1 Instructivo de trabajo.....	98
5.5 Conclusiones.....	99
5.5.1 Conclusiones de los modelos.....	99
5.5.2 Conclusiones de la implementación.....	100
5.5.3 Conclusiones de la metodología propuesta.....	100
5.5.4 Limitaciones de los modelos.....	101
5.5.5 Propuestas de mejoramiento futuro.....	101
Anexo I.....	102
Anexo II.....	104
Anexo III.....	108
BIBLIOGRAFÍA.....	110

Índice de Figuras:

Figura 1 - Ubicación Partido de Daireaux.....	16
Figura 2 - Gráficos de precipitaciones por campaña agrícola.....	17
Figura 3 - Monitor de Rendimiento y Accesorios.....	19
Figura 4 - Datos del MR sin filtrar.....	24
Figura 5 - Datos del MR filtrados.....	24
Figura 6 - Poligonización de los datos del MR.....	25
Figura 7 - Grillas generadas sobre los 4 tipos de cultivo (verde oscuro:Soja1,verde claro:Soja2,rojo:Maíz, Amarillo: Girasol).....	26
Figura 8 - Cálculo del rendimiento a nivel pixel. 1) Grilla y polígono del lote. 2) Datos del MR luego del Kriging. 3) Intersección con la grilla. 4) Celdas de 30x30 con el valor de rendimiento asignado.....	27
Figura 9 - Rasterizado de las celdas de rendimiento.....	27
Figura 10 - Visualización conjunta de los lotes procesados para los 4 tipos de cultivo.....	27
Figura 11 - Visualización de la distribución de celdas pares e impares.....	28
Figura 12 - Serie Landsat.....	29
Figura 13 - Imagen L5-TM, P227-R085, RGB:453, Fecha 18/02/2009, con el límite del partido de Daireaux.....	30
Figura 14 - Confección de la Imagen. Izq.: archivos tiff. y txt. descargados del sitio web. Der.: Enlazado ordenado de las bandas que formarán la imagen (en este caso ya tiene aplicada la máscara de extracción de nubes). La proyección de los datos observada en "Map Info" es la original del dato.....	32
Figura 15 – Cuadrante superior izq.: imagen con nubes y sombras. Cuadrante superior der.: imagen con nubes y sombras enmascaradas automáticamente (clasificación). Cuadrante inferior izq.: clasificación de nubes y sombras mejorada a partir de la creación manual de ROI's. Cuadrante inferior der.: imagen con nubes y sombras enmascaradas.....	33
Figura 16 - Visualización de los valores de NDVI de un stack multitemporal sin equidistancia temporal (izquierda) y con equidistancia temporal (derecha).....	35
Figura 17 - Valores y atributos extraídos del stack EVI (base de datos 2008/2009) en formato ASCII.....	37
Figura 18 – Observación multitemporal de boxplots construidos con valores de índice de vegetación para el cultivo de Soja de primera. Día juliano en el eje X (05/10/2008 al 02/06/2009) e índice de vegetación EVI en el eje Y.....	38
Figura 19 – Observación multitemporal de boxplots construidos con valores de índice de vegetación para el cultivo de Soja de segunda. Día juliano en el eje X (03/10/2010 al 23/05/2011) e índice de vegetación NDVI en el eje Y.....	38
Figura 20 - Boxplots por Lote en Soja de primera. Se observa a la izquierda el lote 35 con alta variabilidad (heterogéneo) y a la derecha el lote 37 con baja variabilidad (homogéneo).....	38
Figura 21 - Curva de crecimiento estimada para un pixel de Soja de primera.....	40
Figura 22 - Curvas de crecimiento estimadas para los 39 lotes de Soja de primera. Se observan 3 lotes anómalos que descienden a mitad de ciclo. Las líneas verticales corresponden al Green-Up y el Green-Down.....	41
Figura 23 - Curvas de crecimiento estimadas para los 36 lotes de Soja de primera. Las líneas verticales corresponden al Green-Up y el Green-Down.....	41
Figura 24 - Curvas de crecimiento estimadas para Soja de primera. Las líneas verticales corresponden al Green-Up y el Green-Down y el sector verde corresponde al período crítico.....	43
Figura 25 - Curvas de crecimiento estimadas para Soja de primera en la campaña 2010/2011. Las líneas verticales corresponden al Green-Up y el Green-Down y el sector verde corresponde al período crítico.....	43
Figura 26 - Curvas de crecimiento estimadas para Soja de segunda en la campaña 2008/2009. Las líneas verticales corresponden al Green-Up y el Green-Down y el sector verde corresponde al período crítico.....	44
Figura 27 - Curvas de crecimiento estimadas para Soja de segunda en la campaña 2010/2011. Las líneas verticales corresponden al Green-Up y el Green-Down y el sector verde corresponde al período crítico.....	44
Figura 28 - Curvas de crecimiento estimadas para Maíz en la campaña 2008/2009. Las líneas verticales corresponden al Green-Up y el Green-Down y el sector verde corresponde al período crítico.....	45
Figura 29 - Curvas de crecimiento estimadas para Maíz en la campaña 2010/2011. Las líneas verticales corresponden al Green-Up y el Green-Down y el sector verde corresponde al período crítico.....	45

Figura 30 - Curvas de crecimiento estimadas para Girasol en la campaña 2008/2009. Las líneas verticales corresponden al Green-Up y el Green-Down y el sector verde corresponde al período crítico.	46
Figura 31 - Curvas de crecimiento estimadas para Girasol en la campaña 2010/2011. Las líneas verticales corresponden al Green-Up y el Green-Down y el sector verde corresponde al período crítico.	46
Figura 32 - Izq.: Ejemplo del cálculo del coeficiente de determinación R^2 durante el ciclo del cultivo de Soja de primera en la campaña 2008/2009. La línea vertical roja corresponde al día de mayor valor de R^2 . Der.: Modelo de regresión lineal que vincula los valores del índice en el día de mayor R^2 con el rendimiento en Ton/ha. Los colores de la nube de puntos definen la densidad (rojo mayor densidad, azul menor densidad).	47
Figura 33 - Izq.: Grafico de Residuos Estandarizados vs valores ajustados. Der.: Grafico de Residuos Estandarizados vs Influencia.	48
Figura 34 - Grafico de QQ-Plot (normalidad de los residuos).	48
Figura 35 - Izq.: Ejemplo del cálculo del coeficiente de determinación R^2 durante el ciclo del cultivo de Soja de primera en la campaña 2008/2009. La línea vertical roja corresponde al día de mayor valor de R^2 . Der.: Modelo de regresión lineal que vincula los valores del índice del día de mayor R^2 con el rendimiento transformado a logaritmo y en unidades de qq/ha. Los colores de la nube de puntos definen la densidad (rojo mayor densidad, azul menor densidad).	49
Figura 36 - Izq.: Grafico de Residuos Estandarizados vs valores ajustados. Der.: Grafico de Residuos Estandarizados vs Influencia.	49
Figura 37 - Grafico de QQ-Plot (normalidad de los residuos).	49
Figura 38 - Los datos corresponden al lote 2 de Soja de primera (2008/2009). Se graficó la evolución del p-valor (línea azul) derivado del análisis de normalidad de los datos de NDVI por el método de Shapiro-Wilks junto con el valor del NDVI en el ciclo del cultivo (línea verde) y la evolución de los valores del coeficiente de determinación R^2 (puntos negros) obtenido de las regresiones temporales con los datos de rendimiento. La línea vertical roja identifica el momento de máximo ajuste del modelo de regresión lineal. GU: "Green-Up" y GD: "Green-Down"	50
Figura 39 - Visualización de los 4 métodos utilizados para vincular los datos (campaña 2008/2009) utilizando como ejemplo el índice de vegetación EVI en Soja de primera. Fila 1: "a", Fila 2: "b", Fila 3: "c", Fila 4: "d"	53
Figura 40 - Visualización de los 4 métodos utilizados para vincular los datos (campaña 2010/2011) utilizando como ejemplo el índice de vegetación NDVI en Soja de primera.	54
Figura 41 – Modelo de Regresión Lineal (recta llana) y Modelo de Regresión Lineal Robusta (recta punteada) para datos de NDVI vs Rendimiento en Maíz.	56
Figura 42 - Izq.: Ejemplo del cálculo del coeficiente de determinación R^2 durante el ciclo del cultivo de Soja de primera en la campaña 2008/2009. La línea vertical roja corresponde al día de mayor valor de R^2 . La línea azul representa el desarrollo medio del cultivo (valores medios de NDVI). Der.: Modelo de regresión lineal que vincula los valores del índice NDVI del día de mayor R^2 con el rendimiento transformado a logaritmo natural. Los colores de la nube de puntos definen la densidad (rojo mayor densidad, azul menor densidad).	58
Figura 43 - Izq.: Gráfico de Residuos Estandarizados vs Valores Ajustados. Der.: Gráfico de Residuos Estandarizados vs Distancias Robustas.	59
Figura 44 - Izq.: Gráfico de QQ-Plot (normalidad de los residuos). Der.: Histograma de residuos.	59
Figura 45 - Izq.: Ejemplo del cálculo del coeficiente de determinación R^2 durante el ciclo del cultivo de Soja de segunda en la campaña 2008/2009. La línea vertical roja corresponde al día de mayor valor de R^2 . La línea azul representa el desarrollo medio del cultivo (valores medios de NDVI). Der.: Modelo de regresión lineal que vincula los valores del índice NDVI en el día posterior al Green-Up de mayor R^2 con el rendimiento transformado al logaritmo natural (qq/ha). Los colores de la nube de puntos definen la densidad (rojo mayor densidad, azul menor densidad).	60
Figura 46 - Izq.: Gráfico de Residuos Estandarizados vs Valores Ajustados. Der.: Gráfico de Residuos Estandarizados vs Distancias Robustas.	60
Figura 47 - Izq.: Gráfico de QQ-Plot (normalidad de los residuos). Der.: Histograma de residuos.	61
Figura 48 - Izq.: Ejemplo del cálculo del coeficiente de determinación R^2 durante el ciclo del cultivo de Maíz en la campaña 2008/2009. La línea vertical roja corresponde al día de mayor valor de R^2 . La línea azul representa el desarrollo medio del cultivo (valores medios de NDVI). Der.: Modelo de regresión lineal que vincula los valores del índice NDVI del día de mayor R^2 con el rendimiento transformado al logaritmo natural. Los colores de la nube de puntos definen la densidad (rojo mayor densidad, azul menor densidad).	61

Figura 49 - Izq.: Gráfico de Residuos Estandarizados vs Valores Ajustados. Der.: Gráfico de Residuos Estandarizados vs Distancias Robustas	62
Figura 50 - Izq.: Gráfico de QQ-Plot (normalidad de los residuos). Der.: Histograma de residuos	62
Figura 51 - Izq.: Ejemplo del cálculo del coeficiente de determinación R^2 durante el ciclo del cultivo de Girasol en la campaña 2008/2009. La línea vertical roja corresponde al día de mayor valor de R^2 . La línea azul representa el desarrollo medio del cultivo (valores medios de NDMI). Der.: Modelo de regresión lineal que vincula los valores del índice NDMI del día de mayor R^2 con el rendimiento transformado al logaritmo natural. Los colores de la nube de puntos definen la densidad (rojo mayor densidad, azul menor densidad)	63
Figura 52 - Izq.: Gráfico de Residuos Estandarizados vs Valores Ajustados. Der.: Gráfico de Residuos Estandarizados vs Distancias Robustas	64
Figura 53 - Izq.: Gráfico de QQ-Plot (normalidad de los residuos). Der.: Histograma de residuos	64
Figura 54 - Izq.: Ejemplo del cálculo del coeficiente de determinación R^2 durante el ciclo del cultivo de Soja de primera en la campaña 2010/2011. La línea vertical roja corresponde al día de mayor valor de R^2 . La línea verde representa el desarrollo medio del cultivo (valores medios de NDVI). Der.: Modelo de regresión lineal que vincula los valores del índice NDVI del día de mayor R^2 con el rendimiento transformado a logaritmo natural. Los colores de la nube de puntos definen la densidad (rojo mayor densidad, azul menor densidad)	65
Figura 55 - Izq.: Gráfico de Residuos Estandarizados vs Valores Ajustados. Der.: Gráfico de Residuos Estandarizados vs Distancias Robustas	66
Figura 56 - Izq.: Gráfico de QQ-Plot (normalidad de los residuos). Der.: Histograma de residuos	66
Figura 57 - Izq.: Ejemplo del cálculo del coeficiente de determinación R^2 durante el ciclo del cultivo de Soja de segunda en la campaña 2010/2011. La línea vertical roja corresponde al día de mayor valor de R^2 . La línea verde representa el desarrollo medio del cultivo (valores medios de NDVI). Der.: Modelo de regresión lineal que vincula los valores del índice NDVI del día de mayor R^2 con el rendimiento transformado a logaritmo natural en unidades de qq/ha. Los colores de la nube de puntos definen la densidad (rojo mayor densidad, azul menor densidad)	67
Figura 58 - Izq.: Gráfico de Residuos Estandarizados vs Valores Ajustados. Der.: Gráfico de Residuos Estandarizados vs Distancias Robustas	67
Figura 59 - Izq.: Gráfico de QQ-Plot (normalidad de los residuos). Der.: Histograma de residuos	68
Figura 60 - Izq.: Ejemplo del cálculo del coeficiente de determinación R^2 durante el ciclo del cultivo de Maíz en la campaña 2010/2011. La línea vertical roja corresponde al día de mayor valor de R^2 . La línea verde representa el desarrollo medio del cultivo (valores medios de NDVI). Der.: Modelo de regresión lineal que vincula los valores del índice NDVI en el día de mayor R^2 con el rendimiento transformado a logaritmo natural (qq/ha). Los colores de la nube de puntos definen la densidad (rojo mayor densidad, azul menor densidad)	68
Figura 61 - Izq.: Gráfico de Residuos Estandarizados vs Valores Ajustados. Der.: Gráfico de Residuos Estandarizados vs Distancias Robustas	69
Figura 62 - Izq.: Gráfico de QQ-Plot (normalidad de los residuos). Der.: Histograma de residuos	69
Figura 63 - Izq.: Ejemplo del cálculo del coeficiente de determinación R^2 durante el ciclo del cultivo de Soja de segunda en la campaña 2010/2011. La línea vertical roja corresponde al día de mayor valor de R^2 . La línea verde representa el desarrollo medio del cultivo (valores medios de NDVI). Der.: Modelo de regresión lineal que vincula los valores del índice NDVI del día de mayor R^2 con el rendimiento transformado a logaritmo natural. Los colores de la nube de puntos definen la densidad (rojo mayor densidad, azul menor densidad)	70
Figura 64 - Izq.: Gráfico de Residuos Estandarizados vs Valores Ajustados. Der.: Gráfico de Residuos Estandarizados vs Distancias Robustas	71
Figura 65 - Izq.: Gráfico de QQ-Plot (normalidad de los residuos). Der.: Histograma de residuos	71
Figura 66 - Observado Vs Predicho. Soja de primera campaña 2008/2009. Izq.: Diagrama de dispersión (rojo mayor densidad, azul menor densidad) con recta de regresión (línea negra) del ajuste de los datos y datos de la regresión insertos en el gráfico. Der.: Ídem Izq. con identificación de los datos con menos del 30% de error (P30) entre el valor observado y el predicho (puntos negros)	73
Figura 67 - <i>Boxplot de los residuos. Soja de primera campaña 2008/2009.</i>	73
Figura 68 - Observado Vs Predicho. Soja de primera campaña 2008/2009. Izq.: Diagrama de dispersión (rojo mayor densidad, azul menor densidad) con recta de regresión (línea negra) del ajuste de los datos y datos de la regresión insertos en el gráfico. Der.: Ídem Izq. con identificación de los datos con menos del 30% de error absoluto (P30) entre el valor observado y el predicho (puntos negros)	74

Figura 69 - Boxplot de los residuos. Soja de segunda campaña 2008/2009.....	74
Figura 70 - Observado Vs Predicho. Maíz campaña 2008/2009. Izq.: Diagrama de dispersión (rojo mayor densidad, azul menor densidad) con recta de regresión (línea negra) del ajuste de los datos y datos de la regresión insertos en el gráfico. Der.: Ídem Izq. con identificación de los datos con menos del 30% de error absoluto (P30) entre el valor observado y el predicho (puntos negros).....	75
Figura 71 - Boxplot de los residuos. Maíz campaña 2008/2009.....	75
Figura 72 - Observado Vs Predicho. Girasol campaña 2008/2009. Datos de regresión insertos en el gráfico. Izq.: Diagrama de dispersión (rojo mayor densidad, azul menor densidad) con recta de regresión (línea negra) del ajuste de los datos y datos de la regresión insertos en el gráfico. Der.: Ídem Izq. con identificación de los datos con menos del 30% de error absoluto (P30) entre el valor observado y el predicho (puntos negros).....	76
Figura 73 - Boxplot de los residuos. Girasol campaña 2008/2009.....	76
Figura 74 - Observado Vs Predicho. Soja de primera campaña 2010/2011. Datos de regresión insertos en el gráfico. Izq.: Diagrama de dispersión (rojo mayor densidad, azul menor densidad) con recta de regresión (línea negra) del ajuste de los datos y datos de la regresión insertos en el gráfico. Der.: Ídem Izq. con identificación de los datos con menos del 30% de error absoluto (P30) entre el valor observado y el predicho (puntos negros).....	77
Figura 75 - Boxplot de los residuos. Soja de primera campaña 2010/2011.....	77
Figura 76 - Observado Vs Predicho. Soja de segunda campaña 2010/2011. Datos de regresión insertos en el gráfico. Datos de regresión insertos en el gráfico. Izq.: Diagrama de dispersión (rojo mayor densidad, azul menor densidad) con recta de regresión (línea negra) del ajuste de los datos y datos de la regresión insertos en el gráfico. Der.: Ídem Izq. con identificación de los datos con menos del 30% de error absoluto (P30) entre el valor observado y el predicho (puntos negros).....	78
Figura 77 - Boxplot de los residuos. Soja de segunda campaña 2010/2011.....	78
Figura 78 - Observado Vs Predicho. Maíz campaña 2010/2011. Datos de regresión insertos en el gráfico. Izq.: Diagrama de dispersión (rojo mayor densidad, azul menor densidad) con recta de regresión (línea negra) del ajuste de los datos y datos de la regresión insertos en el gráfico. Der.: Ídem Izq. con identificación de los datos con menos del 30% de error absoluto (P30) entre el valor observado y el predicho (puntos negros).....	79
Figura 79 - Boxplot de los residuos. Maíz campaña 2010/2011.....	79
Figura 80 - Observado Vs Predicho. Girasol campaña 2010/2011. Datos de regresión insertos en el gráfico. Izq.: Diagrama de dispersión (rojo mayor densidad, azul menor densidad) con recta de regresión (línea negra) del ajuste de los datos y datos de la regresión insertos en el gráfico. Der.: Ídem Izq. con identificación de los datos con menos del 30% de error absoluto (P30) entre el valor observado y el predicho (puntos negros).....	80
Figura 81 - Boxplot de los residuos. Girasol campaña 2010/2011.....	80
Figura 82 - Tipo de errores en la cualificación del rendimiento con rangos de rendimiento para la soja.	82
Figura 83 - Porcentaje de participación de los datos observados vs predichos cualificados en cada tipo de error. Izq.: soja 1° 2008/2009. Der.: soja 1° 2010/2011.....	82
Figura 84 - Porcentaje de participación de los datos observados vs predichos cualificados en cada tipo de error. Izq.: soja 2° 2008/2009. Der.: soja 2° 2010/2011.....	83
Figura 85 - Porcentaje de participación de los datos observados vs predichos cualificados en cada tipo de error. Izq.: Maíz 2008/2009. Der.: Maíz 2010/2011.....	83
Figura 86 - Porcentaje de participación de los datos observados vs predichos cualificados en cada tipo de error. Izq.: Girasol 2008/2009. Der.: Girasol 2010/2011.....	84
Figura 87 - Curvas de crecimiento 2016/2017 con indicación de Green-Up y Green-Down (barras verticales verdes), y el día de extracción de los datos para el modelo (barra vertical roja). Izq.: Soja 1°; Der.: soja 2°.....	86
Figura 88 - Curvas de crecimiento 2016/2017 con indicación de Green-Up y Green-Down (barras verticales verdes), y el día de extracción de los datos para el modelo (barra vertical roja). Izq.: Maíz; Der.: Girasol.....	87
Figura 89 - Mapa de rendimiento obtenido para los 4 tipos de cultivo (campaña 2016/2017).....	88
Figura 90 - Lotes ubicados en un mismo sitio geográfico con su respectiva calidad (Excelente: verde, Bueno: amarillo, Regular: naranja). Izq.: Lotes de Soja 1°. Der.: Lotes de Soja 2°.....	88
Figura 91 - Lotes ubicados en un mismo sitio geográfico con su respectiva calidad (Excelente: verde, Bueno: amarillo, Regular: naranja). Izq.: Lotes de Maíz. Der.: Lotes de Girasol.....	88

Figura 92 - Imagen NDVI (Path 227, Row 085) con el shape de clasificación del cultivo de Soja de primera en campaña 2016-2017 recortado por el parcelario rural (Escala de color: "rojo" alto NDVI – "celeste" bajo NDVI).....	90
Figura 93 - Comparación de ciclos de Soja de primera. Línea continua 2008/2009 y línea punteada 2010/2011. Valores medios de NDVI de todo el set de datos.....	91
Figura 94 - Comparación de ciclos de Soja de segunda. Línea continua 2008/2009 y línea punteada 2010/2011. Valores medios de NDVI de todo el set de datos.....	92
Figura 95 - Comparación de ciclos de Maíz. Línea continua 2008/2009 y línea punteada 2010/2011. Valores medios de NDVI de todo el set de datos.....	92
Figura 96 - Comparación de ciclos de Girasol. Línea continua 2008/2009 y línea punteada 2010/2011. Valores medios de NDVI de todo el set de datos.....	92
Figura 97: ZAH VJ – Daireaux. Fuente: Zonas Agroeconómicas Homogéneas de Buenos Aires (INTA2009).....	96
Figura 98 - Flujograma del modelo de estimación de rendimientos propuesto.....	97
Figura 99 - Flujograma simplificado de procesos vigente en DTI-ARBA.....	97
Figura 100 - Flujograma de procesos modificado con la inclusión del modelo propuesto.....	98

Índice de Tablas:

Tabla 1 - Detalle del Método de Filtrado 1.....	21
Tabla 2 - Accionar del filtro sobre cada atributo.....	21
Tabla 3 - Resultados de ambos métodos de filtrado de datos.....	23
Tabla 4 - Resultados de la comparación estadística de los métodos de filtrado.....	23
Tabla 5 - Set de Imágenes 2008/2009 (izquierda) y 2010/2011 (derecha). En color amarillo se indican las fechas no utilizadas por no encontrarse disponible la imagen o contener una elevada cantidad de nubosidad.....	30
Tabla 6 - Set de Imágenes 2016/2017. En color amarillo se indican las fechas no utilizadas por no encontrarse disponible la imagen o contener una elevada cantidad de nubosidad.....	31
Tabla 7 – Ubicación de los Períodos Críticos por Cultivo y Campaña agrícola.....	42
Tabla 8 - Valores del Coeficiente de determinación R ² obtenidos al combinar: tipo de cultivo, método de análisis de datos, índice de vegetación y campaña agrícola.....	55
Tabla 9 - Modelos finalmente seleccionados.....	56
Tabla 10 - Rangos de rendimiento utilizados en DTI-ARBA y calculados a partir de los rendimientos medios históricos de cada tipo de cultivo.....	82
Tabla 11 - Precipitación en milímetros. Campaña 2016/2017. Partido de Daireaux.....	85
Tabla 12 - Modelos generados para ciclos hídricos Normales/Húmedos.....	86
Tabla 13 - Rendimientos promedios (Ton/ha) oficiales por tipo de cultivo y campaña agrícola.....	91
Tabla 14 - Resumen de modelos desarrollados y validados. "DPG-P" refiere a la cantidad de días posteriores al Green-Up Promedio en el que se halló el mejor ajuste de los datos. "n1 y n2" refieren a la cantidad de datos analizados para modelar y para validar respectivamente. (*): Método "b": Días Posteriores al Green-up de cada pixel (no es el promedio del conjunto de datos).....	93
Tabla 15 – Día de mayor ajuste y verificación del rango del Período Crítico.....	94
Tabla 16 – Estadísticas del rendimiento predicho por los modelos en 2016/2017 (qq/ha) y % de cada tipo de calidad.....	94
Tabla 17 - Datos estimados para la campaña agrícola 2016/2017 generados a partir del trabajo de investigación.....	95
Tabla 18 - Datos oficiales de producción para la Campaña agrícola 2016/2017 en Daireaux.....	95

Softwares utilizados:

- IDL-ENVI 4.8
- Q-gis 2.18
- R-Studio (R 3.5.1). Librerías (geoR, rgdal, imputeTS, car, robust, robustbase, nortest, sp, maptools, raster, rgeos, dismo, lattice, spdep).
- FastMapping 1.0

CAPITULO I - Introducción.

1.1 ARBA y la actividad agrícola en la provincia.

La Agencia de Recaudación de la provincia de Buenos Aires¹ (ARBA) [1] tiene como función recaudar los impuestos provinciales (Ingresos Brutos, Inmobiliario, Automotor, Embarcaciones y Sellos), garantizar la equidad fiscal y combatir la evasión fiscal. Su misión es “Ejecutar eficientemente la política tributaria mediante la determinación, fiscalización y percepción de los tributos y accesorios dispuestos por las normas legales, administrar el Catastro Territorial y brindar servicios a otros organismos públicos, promoviendo el cumplimiento voluntario y contribuyendo al bienestar económico y social de la ciudadanía”. Dentro de la Agencia se creó un área especializada en la utilización de tecnología satelital denominada Subgerencia de Fiscalización Satelital, dentro de la cual funciona el Departamento de Tecnología de Imágenes (DTI). DTI, realiza entre otras líneas de trabajo, la detección y cuantificación de la superficie implantada con cultivos agrícolas extensivos en todo el territorio de la provincia de Buenos Aires y en todas las campañas agrícolas desde su creación en el año 2007. El control en forma remota de la actividad agrícola, tiene como principal objetivo reducir la evasión fiscal del sector agropecuario en el impuesto de los ingresos brutos e impuesto de sellos, tanto en forma directa a través de intimaciones y/o indirectamente por percepción del riesgo.

La provincia de Buenos Aires es una de las principales provincias agrícolas de la Argentina, con aproximadamente 24 millones de hectáreas destinadas a la producción agropecuaria [2]. El sistema de producción es de tipo “secano”, los cultivos sembrados no reciben aporte externo de riego, por lo cual quedan sujetos al régimen de precipitación natural que fluctúa en períodos secos y húmedos, generando impacto en las economías regionales y nacionales debido a su extensión geográfica, intensidad y persistencia [3].

En cuanto a la actividad agrícola en la provincia, la superficie promedio² cosechada en la campaña “fina” (cultivos de invierno) es de aproximadamente 2.9 millones de hectáreas y en la campaña “gruesa” (cultivos de verano), de 10.8 millones de hectáreas [4]. Los principales cultivos extensivos son trigo, cebada y colza (campaña “fina”) y soja, maíz, girasol (campaña “gruesa”).

La producción es extraída de los lotes cosechados mediante camiones que realizan el flete corto hasta las centrales de acopio generalmente ubicadas dentro de un radio de 100 km. Posteriormente, luego de realizarse las tareas de secado, limpieza y clasificación se puede destinar al mercado interno (procesamiento industrial: molienda, aceite, etc.) o externo, enviando el grano cosechado a exportación a través de camión o tren con dirección a los puertos. Argentina exporta sus productos granarios a través de 4 puertos, 3 de los cuales se encuentran en la provincia de Buenos Aires (Puertos: Buenos Aires, Quequén y Bahía Blanca). El sector tiene gran importancia a nivel de recaudación fiscal, los montos exportados por la provincia de Buenos Aires en concepto de productos primarios para el año 2012 fueron el 22.4% del total exportado provincial y las manufacturas agropecuarias un 20.4%, lo que significa ron 6.127,3 y 5.580,2 millones de dólares respectivamente [5].

1 La Agencia de Recaudación de la Provincia de Buenos Aires es un ente autárquico cuya responsabilidad es administrar la política tributaria de la provincia de Buenos Aires. Se creó en el año 2007 por medio de la ley 13.766.

2 Promedio de las últimas 10 campañas agrícolas (2008/2009 a 2017/2018).

1.2 Estimación del rendimiento de cultivos con datos de satélite.

Las técnicas de estimación del rendimiento de cultivos agrícolas se incrementaron con el desarrollo de la tecnología satelital, dando lugar a una gama de posibilidades que a grandes rasgos comprenden [6]:

- 1_**Métodos con base en modelos matemáticos de simulación de cultivos (clasificados como empíricos, semi-empíricos y mecanicistas según Moulin S. (1998) [7]).
- 2_**Métodos que emplean información exclusivamente satelital.
- 3_**Métodos que combinan modelos matemáticos con datos provistos por imágenes satelitales.

1) Los modelos matemáticos de simulación por lo general reducen el complejo sistema planta-suelo-atmósfera, a unas pocas variables agroclimáticas que representan los factores limitantes en el rendimiento. Debido a esto, no es necesario simular todos los procesos fisiológicos del crecimiento de la planta, sino que se relacionan los factores más relevantes a través de índices que luego correlacionan con los rindes [8]. Las interacciones entre el cultivo y el medio exterior son representadas por el intercambio de energía (radiación en onda corta y onda larga, conducción y convección), intercambio de gases, agua y nutrientes. Conociendo el ritmo con el cual se intercambian estos elementos se pueden modelar los fenómenos fisiológicos que se producen y predecir el crecimiento del cultivo. Estos modelos requieren gran cantidad de datos de entrada para su funcionamiento (datos meteorológicos, características del suelo, tipo de cultivar, manejo del cultivo, disponibilidad de agua, nutrientes, etc.). Sin embargo, como contraparte, brindan gran cantidad de resultados para el análisis y permiten además predecir los requerimientos que serán necesarios para lograr un rinde óptimo [8]. Son variados los estudios que utilizan modelos para estimar y/o predecir el rendimiento de diferentes cultivos y situaciones agroclimáticas. Estos modelos generalmente definen las variables que son determinantes para la definición del rendimiento final del cultivo y a partir de mediciones directas de dichas variables, los valores (medidos en el momento indicado) se insertan en el modelo. Por ejemplo, la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO) desarrolló dos modelos empíricos basados principalmente en la disponibilidad de agua como factor limitante del rendimiento, como son el "CropWat" y el "AquaCrop" [9, 10]. Existen varios modelos de estimación de rendimiento con posibilidad de aplicación general para distintos tipos de cultivos (previa validación) como el "Simple Algorithm For Yield estimates" (SAFY) [11], el "Decision Support System for Agrotechnology Transfer" (DSSAT) [12, 13], el "Simulateur mulTIdisciplinaire pour les Cultures Standard" (STICS) [14], el "EPIC" [15] desarrollado para evaluar el rendimiento en suelos erosionados, el "Cropping Systems Simulation Model" (CropSyst) [16], entre otros; y modelos que se desarrollaron para cada tipo de cultivo en particular como lo son el "ORYZA" exclusivo para el cultivo de arroz, el "WHET" exclusivo para el trigo, etc. Una revisión de estos modelos clasificados por su potencial uso puede encontrarse en [17].

2) Los métodos de estimación de rendimiento que utilizan información exclusivamente satelital son métodos netamente empíricos [7]. Por lo general son modelos muy sencillos que correlacionan de manera directa, datos radiométricos (valores de alguna banda espectral, relaciones entre bandas y/o índices) con el rendimiento del cultivo [18].

3) Los métodos que combinan modelos matemáticos con datos provistos por imágenes satelitales contemplan un análisis cuantitativo con metodologías complejas que incluyen el acoplamiento de datos satelitales a los modelos de producción de cultivos para estimar los rendimientos a escala regional [19]. Las estimaciones agrícolas a través de sensores remotos son cada vez más considerados en todo el mundo, las variables de entrada de los modelos de simulación que son posibles de estimar de manera indirecta utilizando información satelital, han ido reemplazando las mediciones a campo [20, 18]. La medición de variables fisiológicas y del dosel del cultivo a partir de datos de teledetección está ampliamente difundido y puede encontrarse un resumen en la revisión que realiza Moulin S. en 1998 [7]. Las condiciones del ambiente se ven reflejadas en el desarrollo del cultivo de manera temporal (ciclo fenológico) y espacial (área cultivada), por lo que determinados índices de vegetación obtenidos del procesamiento de imágenes satelitales son representativos de las condiciones en la que se encuentran los cultivos como se demuestra en variados trabajos [21, 22, 20, 23]. La estrategia general de los modelos combinados, consiste en reemplazar variables o parámetros observados a campo, por observaciones radiométricas. Cada banda espectral es de gran interés: visible y NIR (Near Infrared) para la estructura del dosel y la actividad fotosintética, TIR (Thermal Infrared) para el estrés hídrico, microondas (activa o pasiva) para el contenido de agua, la humedad del suelo y la estructura del dosel. Por ejemplo, el índice de área foliar (LAI por sus siglas en inglés) puede calcularse a partir de un NDVI (Normalized Difference Vegetation Index) u algún otro índice a partir de las bandas espectrales visibles y el NIR [24]. La información satelital cuenta con la ventaja de obtener los datos con una frecuencia temporal conocida, contar con series temporales históricas y tener un dominio espacial ampliado a escala regional [7]. La aplicación de este tipo de tecnología también tiene puntos débiles y restricciones. El uso directo de datos de teledetección supone que los datos están disponibles en un paso de tiempo adecuado (de diario a semanal). Debido a la interferencia de nubosidad y las propiedades intrínsecas de los sensores y las plataformas, esto no siempre sucede. Por ejemplo, para el seguimiento de la condición de los cultivos con imágenes del satélite Landsat (resolución temporal de 16 días), la baja frecuencia de las observaciones (por su propio funcionamiento y la ocurrencia de nubosidad) representa una restricción operacional importante. Este tipo de modelo mixto, remedia el problema acoplando los datos de origen satelital con la información derivada de modelos de cultivos [6]. La utilización conjunta de datos satelitales con el modelo de simulación del crecimiento del cultivo permite resolver la discontinuidad temporal de los datos satelitales, ya que los modelos de cultivos proveen estimaciones diarias de su evolución [25]. Así es como surgen los modelos mixtos, que en su mayoría son los modelos de simulación (1) pero con sus variables medidas indirectamente con sensores remotos. Por ejemplo, el índice de área foliar que es requerido por el modelo SAFY, es estimado a partir de algún índice de vegetación calculado con medición in situ de radiómetro o a través de datos satelitales [26], [27], [28], [29].

Estos modelos mixtos pueden tener diferentes maneras de integrar los datos [30]:

- (a) el **uso directo** de una variable estimada a partir de información de teledetección;
- (b) la **actualización** de una variable del modelo, obtenida por teledetección;
- (c) la **reinicialización** del modelo, o sea, el ajuste de una condición inicial para obtener una simulación de acuerdo con las observaciones de teledetección;
- (d) la **recalibración** del modelo, es decir, el ajuste de los parámetros del modelo para obtener una simulación de acuerdo con observaciones obtenidas por teledetección.

Al nutrirse de un modelo matemático de simulación, también pueden clasificarse en empíricos, semi-empíricos y mecanicistas. Debido a que los datos de teledetección están disponibles solo en las fechas de adquisición, generalmente menos frecuentes que el paso modelo, las brechas entre las fechas deben completarse por algún procedimiento de interpolación, generalmente mediante el ajuste de una curva empírica de evolución temporal de la variable de estado con los datos observados [7]. Por lo tanto, los modelos mecanicistas, que simulan el perfil temporal de las principales variables de estado del cultivo (índice de área foliar, dimensiones y biomasa de diversos órganos, etapas de desarrollo, etc.) suelen ser más utilizados en los métodos combinados con datos de sensores remotos.

En este trabajo de tesis abordaremos el desarrollo de modelos de tipo “3” mecanicista, utilizando datos satelitales de entrada, con un análisis fenomérico que permite obtener los datos diarios que posteriormente ingresarán al modelo lineal.

1.3 Propuesta de mejora en la estimación del rendimiento.

En la Argentina existen escasos antecedentes de estimación del rendimiento agrícola utilizando información satelital por parte de las Agencias de Recaudación a nivel provincial y tampoco a nivel nacional. En general, las agencias solicitan la información a los organismos públicos relacionados a cada uno de los sectores productivos del país para realizar cálculos estimativos y determinar los sectores de la economía en los cuales se encuentra una mayor evasión presunta. Este tipo de información es global y no se encuentra asociada en forma directa a un contribuyente. En este sentido, ARBA es líder en la aplicación de los Sistemas de Información Geográfica (SIG) y la Información Satelital en el análisis espacial y temporal asociado a controlar las actividades económicas que se desarrollan en el territorio. Como efecto directo de la implementación del Régimen de Control Fiscal Agropecuario³ y su combinación con el Índice Verde⁴, el número de inscriptos en el impuesto sobre los Ingresos Brutos pasó de 5.935 en junio de 2008, a 26.016 inscriptos en junio de 2009, con un crecimiento del 338% [31]. Para realizar este seguimiento, el catastro digital geo-posicionado y actualizado es fundamental para vincular y asociar la actividad económica detectada en las imágenes, con la parcela fiscal y, por lo tanto, con el contribuyente (propietario y/o arrendatario).

La metodología empleada por DTI en la observación remota de las parcelas cultivadas ha ido evolucionando a lo largo del tiempo con el desarrollo de nuevas tecnologías, la disponibilidad de nuevos sensores remotos, y el asesoramiento de la CONAE a través del Acuerdo Marco celebrado en el año 2008 con ARBA [32]. Actualmente uno de los desafíos que enfrenta DTI, es mejorar la calidad de los productos generados a partir del procesamiento de imágenes satelitales, como es la clasificación de los cultivos agrícolas, la identificación del tipo de cultivo y la estimación de la producción por parcela. En el proceso que se desarrolla actualmente, se observan dos situaciones principales en las que se afecta la calidad del producto final. La primera se produce por la baja disponibilidad de datos satelitales (nubosidad/fallas), siendo un problema externo y aleatorio. La segunda se debe a que el método empleado para la estimación del rendimiento utiliza datos promedios por partido (datos oficiales provistos por

³ El Régimen de Control Fiscal Agropecuario (RCFA) dispuesto por Arba por resolución 32/08, obliga a propietarios de parcelas rurales superiores a 50 hectáreas a informar las actividades agropecuarias o de otra índole económica que se desarrolle en el inmueble. En caso de no ser el propietario el que realiza la actividad, debe informar el CUIT de la persona física o jurídica que la realiza.

⁴ Índice verde es como se denomina al proceso de control satelital dentro de la Agencia, ver: <http://www.arba.gov.ar/Informacion/InfoGeneral/Catastro/PropiedadesInfraccion.asp>.

Agroindustria) que carecen de una estimación del error, generando un problema interno y sistemático.

Sobre la falta de datos satelitales ha trabajado el Ing. Agr. Luciano Di Paolo perteneciente a ARBA en su tesis de maestría MAEARTE (Instituto Gulich-CONAE) en la cual analizó la viabilidad de utilizar datos SAR en forma conjunta con datos ópticos [33].

Para dar respuesta a la segunda situación, luego del proceso de clasificación de los cultivos agrícolas, con el cuál se determina el tipo de cultivo (ej.: soja, maíz, girasol, etc.) y la superficie cultivada por parcela para cada cultivo (cálculo del área de los polígonos clasificados a partir de las imágenes), es de suma importancia por los intereses de la Agencia, realizar la estimación del rendimiento de cada parcela agrícola cultivada. Este proceso se realiza relacionando un valor máximo del índice de vegetación de diferencia normalizada (NDVI) que surge del análisis multitemporal de las imágenes Landsat 7 y Landsat 8, con el valor promedio por partido/departamento otorgado por fuentes oficiales como el Ministerio de Agroindustria [4]. De esta manera se obtiene el rendimiento en el rango medio cualificado como “bueno”, luego se genera un rango superior (grupo de parcelas que tienen el rango más alto de NDVI) cualificado como “excelente” y uno inferior (grupo de parcelas que tienen el rango más bajo de NDVI) cualificado como “regular”. Este proceso requiere anualmente alimentarse de datos externos y genera dependencia permanente de datos oficiales que provienen de estimaciones realizadas a través del método de encuestas a informantes calificados con una cobertura espacial limitada en el procedimiento [6]. Teniendo en cuenta la subjetividad de la técnica empleada [20] y que el organismo oficial no acompaña el dato con el desvío estándar correspondiente, es que tomamos la decisión de innovar en otros métodos que no dependan de información externa.

La propuesta de mejora que plantea este trabajo de investigación se basa en incorporar en la cadena de producción de DTI un método de trabajo que incluya modelos de estimación del rendimiento para los principales cultivos agrícolas extensivos de la provincia de Buenos Aires utilizando únicamente información satelital como entrada de cada modelo. Esto permitiría independizarse de los datos estadísticos oficiales, comenzando a trabajar con modelos del tipo “3” que combinan modelos matemáticos con datos provistos por imágenes satelitales, estimando el rendimiento en este caso propuesto a escala de pixel⁵, con un valor central y una estimación del error que determinaría el rango de variación del rendimiento estimado para cada tipo de cultivo.

Es importante resaltar que, dado que ARBA permite presentar declaraciones juradas sobre el impuesto a los ingresos brutos hasta el cierre del ejercicio económico y con posibilidad de rectificar los montos declarados, es que no es necesario aplicar modelos de predicción del rendimiento. Es suficiente para los objetivos de ARBA, estimar el rendimiento al finalizar la campaña agrícola, lo que permite contar con el set de imágenes satelitales completo.

1.4 Objetivo general:

- Desarrollar un método que permita estimar el rendimiento de los principales cultivos agrícolas estivales de la provincia de Buenos Aires, utilizando únicamente información satelital de sensores ópticos con una confiabilidad superior al 70%.

⁵ La serie Landsat ofrece una resolución espacial de 30 m x 30 m en las bandas espectrales utilizadas para el cálculo de los índices de vegetación.

1.5 Objetivos específicos:

- Elaborar cuatro modelos de estimación del rendimiento para los cultivos estivales de soja de primera (*Glycine max* sin cultivo antecesor), soja de segunda (*Glycine max* con cultivo antecesor), maíz (*Zea mays*) y girasol (*Helianthus annuus*) que actúe a nivel de pixel (30m x 30m), para la campaña 2008/2009 (ciclo hídrico de sequía), en el partido de Daireaux provincia de Buenos Aires utilizando imágenes satelitales ópticas.
- Elaborar cuatro modelos de estimación del rendimiento para los cultivos estivales de soja de primera (*Glycine max* sin cultivo antecesor), soja de segunda (*Glycine max* con cultivo antecesor), maíz (*Zea mays*) y girasol (*Helianthus annuus*) que actúe a nivel de pixel (30m x 30m), para la campaña 2010/2011 (ciclo hídrico normal/húmedo), en el partido de Daireaux provincia de Buenos Aires utilizando imágenes satelitales ópticas.
- Implementar el método propuesto sobre la campaña agrícola 2016/2017 utilizando los sensores ópticos Landsat 7 y Landsat 8.
- Definir un método de trabajo (procedimiento) que permita incorporar los modelos desarrollados en el proceso de producción de DTI-ARBA.
- Elaborar los manuales de instrucción operativa para su puesta en funcionamiento en DTI-ARBA.

1.6 Resultados Esperados:

- Obtener ocho modelos que permitan estimar el rendimiento para los cultivos de soja de primera, soja de segunda, maíz y girasol en ciclos de clima seco y húmedo, en el partido de Daireaux provincia de Buenos Aires.
- Obtener un archivo vectorial para la campaña agrícola de implementación 2016/2017, en el partido de Daireaux, conteniendo los polígonos de cultivo con los siguientes atributos: ID, tipo de cultivo (soja1/soja2/maíz/girasol), rendimiento (qq/ha) y superficie (ha).
- Manual de instrucciones operativas para la aplicación del método desarrollado.

CAPITULO II - Materiales y métodos.

2.1 Sitio de estudio, definición espacio-temporal.

En este trabajo de investigación el área de estudio se corresponde con el partido de Daireaux (Figura 1) ubicada en el centro-oeste de la provincia de Buenos Aires, con un total de 382.000 ha que tienen como principal actividad productiva la agrícola-ganadera. La cabecera del partido es la ciudad de Daireaux ubicada en latitud $-36^{\circ}36'$ y longitud $-61^{\circ}45'$.

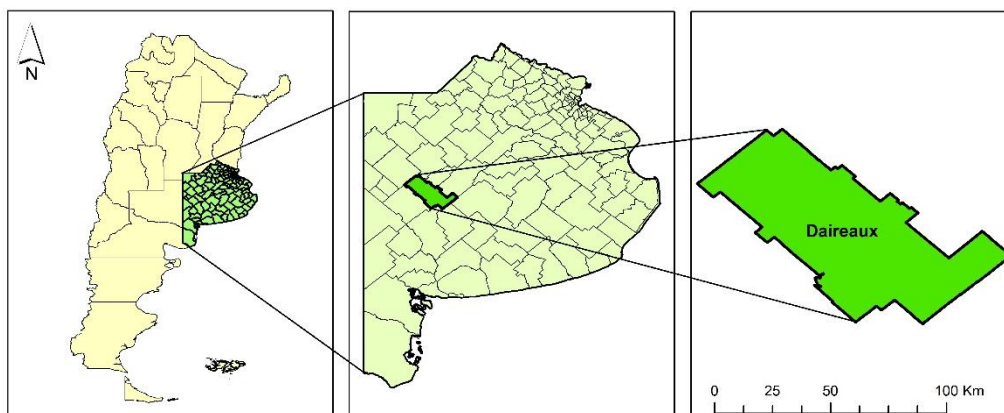


Figura 1 - Ubicación Partido de Daireaux

El partido de Daireaux se encuentra en la denominada Pampa Húmeda Argentina. El clima de la región es una variación del subtropical húmedo, también llamado templado, que se caracteriza por que la estación más cálida, es también la más lluviosa. Presenta veranos cálidos e inviernos frescos y variables, con una temperatura media anual de 17°C . El partido de Daireaux se encuentra entre las isoietas de 750 mm (separa la pampa subhúmeda de la pampa húmeda) y la de 1000 mm (separa la pampa húmeda de la pampa perhúmeda).

El suelo agrícola se compone principalmente de tres series: Bolivar (Bv), Piedritas (Pas) y Veinticinco de Mayo (VMY). La clasificación taxonómica del USDA⁶ define a las series “Bv” y “Pas” como Hapludoles Enticos, de textura Franca Gruesa, Mixta, térmica; que se corresponde con suelos de aptitud agrícola. Mientras que a la serie “VMY” la define como Udipsament Típico, Arenosa, mixta, térmica; correspondiente a suelos de uso ganadero principalmente.

La región, históricamente dominada por pastizales nativos, se transformó poco a poco a una matriz mixta de pastizales nativos, cultivos anuales y pasturas a partir de finales del siglo XIX [34]. Sin embargo, desde la década de 1980, y especialmente en la última década, el remanente de pastizales y de praderas fue casi completamente convertido a la agricultura anual, siendo la soja el cultivo dominante [35]. En cuanto a la actividad agrícola, según los últimos datos oficiales disponibles [4] en la campaña 2015/2016 se sembraron en el partido de Daireaux 188.074 has con cultivos agrícolas estivales (Soja, Maíz y Girasol) y 12.430 has con cultivos invernales (principalmente Trigo).

Dado que en la Pampa Húmeda Argentina las precipitaciones presentan un comportamiento cíclico que se manifiesta en periodos con una disminución o con un aumento de las lluvias [36], se consideró importante generar modelos para dos tipos de situación agroclimática. Para esto se seleccionó información de campo y satelital en una campaña agrícola estival 2008/2009 bajo un período seco (régimen hídrico por debajo de los valores normales)

⁶ El Departamento de Agricultura de los Estados Unidos (USDA) define el criterio de clasificación basado en el “Soil Taxonomy, 2006”.

[37, 38], y una campaña agrícola estival 2010/2011 bajo un período normal/húmedo (régimen hídrico cercano a los valores normales). La campaña 2008/2009 tuvo un período de sequía que provenía de la campaña anterior y se acentuó fuertemente, y la campaña 2010/2011 tuvo una recuperación de los niveles de precipitación a valores similares a la normal. Según [39] en su estudio en la zona sur de Buenos Aires, el evento seco de mayor duración de la región en los últimos años ocurrió durante el período 2005-2009, y la precipitación anual registrada en 2008 y 2009 fue inferior a la normal. En dicha campaña agrícola la provincia de Buenos Aires declaró Emergencia Agropecuaria para toda la región [38]. En un análisis propio (Figura 2) realizado con datos oficiales [4] sobre localidades vecinas a Daireaux (no existen datos históricos oficiales en dicha localidad) se puede observar el descenso de precipitaciones en las campañas 2006/2007 y 2007/2008, luego un incremento y estabilización para las campañas 2009/2010 y 2010/2011 seguido de un período de altas precipitaciones en 2011/2012 y 2012/2013.

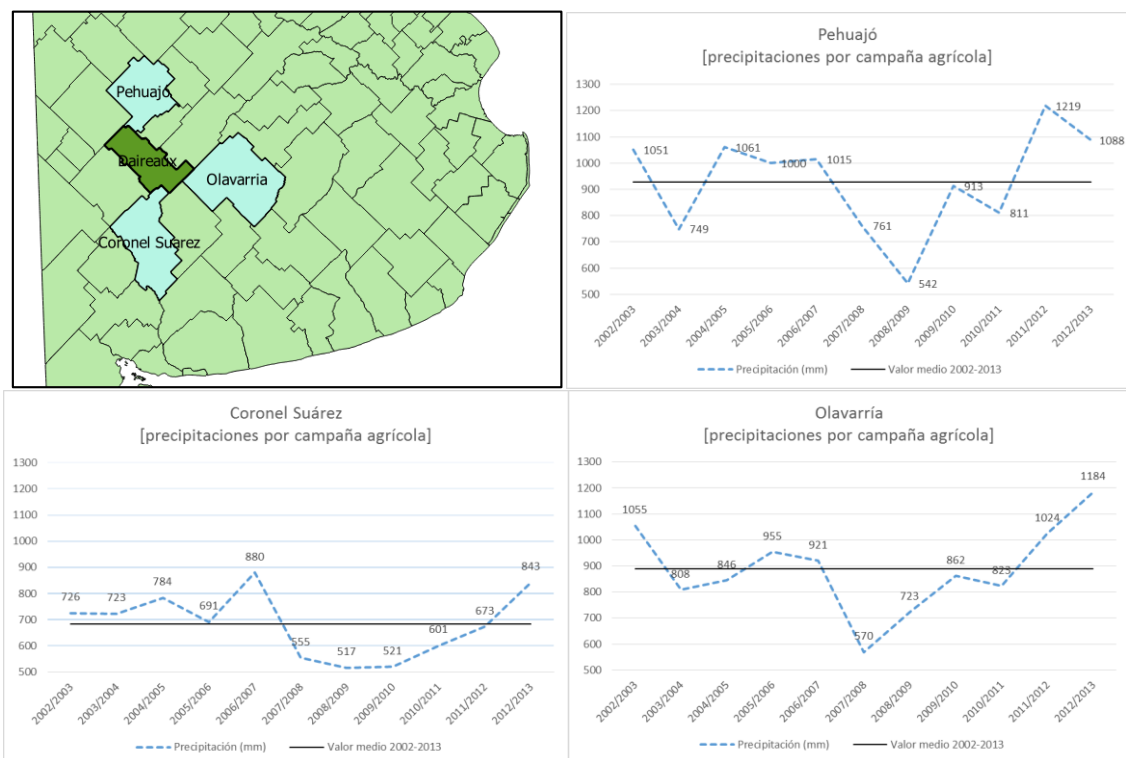


Figura 2 - Gráficos de precipitaciones por campaña agrícola.

2.2 Información de campo.

Los rendimientos de los cultivos a nivel de lote o parcela, que son imprescindibles para la realización de modelos de rendimiento, no son sencillos de obtener en general para un investigador, pero menos aún si dicha información será utilizada por una agencia de recaudación de impuestos. Este tipo de dato, por lo general es escaso y reservado por los productores, además, es muy poco frecuente generar un set con una gran cantidad de datos, ordenados por campaña agrícola y tipo de cultivo. Es por esto, que hallado el set de datos en la web de "Agrodatos" (acceso público), se procedió a interpretarla estudiándola detalladamente para aprovechar su elevado potencial. Estos datos provienen de monitores de rendimiento ofrecidos por productores de la Asociación Argentina de Consorcios Regionales de Experimentación Agropecuaria - ACREA zona Oeste de Buenos Aires. La disponibilidad de estos datos determinó la elección de la zona de estudio, el período a estudiar y, por lo tanto, la modelización utilizando

datos no actualizados. Los datos obtenidos corresponden al partido de Daireaux provincia de Buenos Aires para los cultivos de soja, maíz y girasol para las dos campañas agrícolas bajo estudio: 2008/2009 y 2010/2011.

Los datos de los Monitores de Rendimiento (MR) tienen un gran potencial de utilización como dato de campo cuando se cuenta con un amplio set de datos, tanto espacial como temporalmente. Además, los MR cuentan con una alta resolución espacial (un valor de rendimiento cada 30 m² aproximadamente), aseguran su posicionamiento en el espacio (coordenadas geográficas) y brindan el dato de tipo de cultivo [40]. Sin embargo, estos datos deben procesarse debidamente por medio de algún método de limpieza antes de ser utilizados. Esto se debe a que, en situación de trabajo, los sensores de la máquina cosechadora suelen registrar datos de manera incorrecta arrojando valores extremos incoherentes [41], [42], [43]. En el caso del cultivo de soja, los datos MR se dividieron en Soja de primera y Soja de segunda visualizando el set de imágenes de la campaña agrícola correspondiente. En el caso de visualizar un cultivo antecesor invernal, el dato del MR pasó a formar parte del set de Soja de segunda.

Los datos de rendimiento de cultivos (toneladas de grano extraídas por lote para cada cultivo) se descargaron de la web de Agrodatos [44] perteneciente a la Fundación Sadosky [45]. Estos datos en formato “shape-point” fueron relevados por monitores de rendimiento instalados en las máquinas cosechadoras de los productores o empresas contratistas que trabajan con el Grupo CREA Oeste de Buenos Aires y fueron provistos a la Fundación Sadosky por la Asociación Argentina de Consorcios Regionales de Experimentación Agropecuaria [46]. Los productores agrupados en los grupos CREA son productores por lo general tecnificados, con un alto acceso a la información, investigadores e innovadores. Son considerados productores “de punta” porque suelen estar a la cabeza de los cambios tecnológicos en el agro, con lo cual, nos permite asegurar la calidad del dato utilizado.

El MR es un equipo que se incorpora a la máquina cosechadora y consta de una serie de sensores, GPS, un monitor/consola y un sistema de registro/almacenamiento de los datos. Para estimar el rinde el MR debe contar con los siguientes componentes [47, 40] (Figura 3):

1- Sensor de flujo de grano: Permite medir el volumen de grano que se va cosechando. Existen diferentes tipos de formas de medirlo, pero generalmente es un instrumental que se coloca en el cabezal de la noria (sistema que eleva el grano limpio a la tolva de la cosechadora).

2- Sensor de humedad de grano: Mide la humedad del grano a medida que va ingresando y permite corregir el valor estimado de rinde a un valor de humedad estándar.

3- Sensor de velocidad de avance: Puede incluirse o calcularse a partir de la posición GPS. La velocidad de avance permite calcular la distancia recorrida y a partir de allí el área cosechada multiplicando por el ancho de trabajo del cabezal de cosecha, que se ingresa manualmente por el operador.

4- Switch de posición del cabezal: Define el inicio y fin del registro de datos según se encuentre o no en la posición de trabajo.

5- Consola del monitor: Es el “datalogger” en el cual se registran los datos de los sensores y se ingresan los parámetros iniciales (seteo del equipo) previo al inicio de cosecha. Las marcas comerciales más difundidas son “Ag Leader” para cosechadoras CASE y New Holland, “Green Star” para John Deere, y “Field Star” para Massey Ferguson y AGCO Allis. Los datos exportados de los monitores de rendimiento tienen diferentes formatos en función del equipo/marca instalado en la cosechadora, y esto es importante al momento de procesar el dato.

6- Receptor GPS o DGPS: Es el sensor que define la posición geográfica en la que se encuentra la máquina cosechadora. Todo dato relevado tendrá coordenadas geográficas otorgadas por el GPS.



Figura 3 - Monitor de Rendimiento y Accesorios.

Para evitar los inconvenientes de procesamiento de los datos que provienen de diferentes tipos de MR (diferentes marcas comerciales), los datos cargados en la web de Agrodatos fueron previamente pre-procesados [44], se encuentran en formato shape-point y tienen un proceso de clusterización (utilizando el atributo FIELD) que permitió separar lotes diferentes agrupados en un mismo archivo. Este último proceso, que es necesario para la utilización posterior del dato, provocó que lotes cosechados por más de una máquina o en diferentes momentos se setearan en el nombre del campo (atributo "Field" del monitor de cosecha) con denominaciones diferentes (abreviar/no abreviar el nombre, escribir en mayúscula/minúscula, escribir con algún error de tipeo, etc.), lo que generó que luego del proceso de clusterización quedaran sectores de un mismo lote agrícola separados en diferentes archivos. Por este motivo, fue necesario abrir y visualizar en el software Q-Gis todos los archivos correspondientes a un mismo cultivo para una misma campaña agrícola y posteriormente seleccionar todos los puntos correspondientes a un mismo lote y exportarlos de manera individual. A través de este procedimiento, por ejemplo, para el cultivo de soja, los inicialmente 320 shapets independientes de la campaña 2008/2009 pasaron a formar 57 lotes de Soja (se descartaron también algunos archivos inconsistentes). Este trabajo se realizó para cada uno de los cultivos y para las dos campañas agrícolas (2008/2009 y 2010/2011).

En un segundo paso se decidió separar los lotes de Soja, dividiendo en Soja de primera y Soja de segunda debido a que tienen ciclos de cultivo diferentes. Para realizar esta operación se visualizaron en forma conjunta los shapets del cultivo de soja en el software Q-Gis y las imágenes satelitales Landsat 5 y 7, path-row: 227-85, combinación RGB: 4,5,3 de ambas campañas agrícolas. Realizando un análisis visual de la serie temporal con imágenes que permitían identificar la existencia/no existencia de algún tipo de cultivo invernal para los meses de octubre y noviembre, se separaron los lotes de Soja de segunda y se exportaron con la nueva denominación. De esta manera, los 57 lotes de soja citados anteriormente se dividieron en 39 lotes de Soja de primera y 18 lotes de Soja de segunda en la campaña agrícola 2008/2009.

2.2.1 Atributos de los datos de campo.

Los shapets de los cuatro tipos de cultivos (soja1, soja2, maíz y girasol) contienen en su tabla de atributos los siguientes campos:

- LATITUD: dato provisto por el GPS que permite ubicar la posición en el espacio (grados).
- LONGITUD: dato provisto por el GPS que permite ubicar la posición en el espacio (grados).
- ID: número identificador de cada punto relevado por la máquina cosechadora.
- ELEVACIÓN: altura del cabezal de cosecha (milímetros).
- ANCHO: ancho de trabajo del cabezal de cosecha (metros).

- DISTANCIA: distancia recorrida por la cosechadora desde el punto anterior (metros).
- DURACIÓN: tiempo seteado previamente por el operador para la obtención del dato (segundos). En nuestro caso, cada 3 segundos.
- HUMEDAD: humedad del grano (%).
- FLUJO: cantidad de grano limpio recolectado (toneladas/hora).
- UTC: tiempo universal coordinado (hora).
- PRODUCTO: tipo de cultivo (seteado por el operador).
- FIELD: nombre del campo/lote a cosechar (seteado por el operador).
- CARGA: identificación que surge del pre-procesamiento realizado por Agrodatos.
- MASA: toneladas de grano calculadas en función del FLUJO.
- REND: valor de rendimiento calculado por el monitor de cosecha a partir del área cosechada (ANCHO x DISTANCIA) multiplicado por la MASA de grano corregida por la HUMEDAD.

Al individualizar los lotes por cultivo, se les agregó un nuevo campo numérico denominado LOTE (números enteros consecutivos) con el objetivo posterior de permitir realizar procesamientos a nivel lote programando ciclos automáticos.

2.2.2 Limpieza de los datos de campo.

Los datos provenientes de los MR se estima que tienen un error de entre el 2 al 5 %, y se considera que más del 5% de error requiere la re-calibración de los sensores del sistema de monitoreo [48]. Los datos obtenidos provienen de productores del grupo CREA, por lo tanto, consideramos que los datos son confiables y fueron relevados con monitores calibrados adecuadamente. Sin embargo, los datos de los MR comúnmente contienen valores que son incorrectos debido a diferentes fuentes de error: errores de registro por altura de cabezal, retardos en el inicio y fin de la cosecha, valores fuera de rango (outliers: valores que están fuera de los normales, e inliers: valores que están dentro del rango normal pero que por su posición geográfica no tienen semejanza con sus vecinos), rendimientos fuera de rango biológico (capacidad biológica del cultivo) definido por el usuario, pequeños parches o franjas angostas con rendimientos extremadamente altos o bajos que no están relacionados con sus vecinos inmediatos, pequeños segmentos y puntos de rendimientos superpuestos (grabados con iguales coordenadas), etc. [48], lo que hace indispensable la aplicación de un procedimiento de limpieza a través de la eliminación de datos erróneos (outlier/inliers) utilizando diferentes tipos de filtrado ya que no existe un consenso general de cómo deben removerse este tipo de errores, aunque sí se acuerda en que deben eliminarse debido a la gran abundancia de datos [49].

Para realizar la limpieza de los datos de campo se probaron dos tipos de filtrado:

- Método de Filtrado 1: Secuencia de (+/- 3 desvíos estándar) + Limitantes/cortes de tipo técnico.
- Método de Filtrado 2: Autocorrelación espacial con el Índice de Moran local (IMI).

Además, cabe aclarar que los datos de los MR no tienen una distribución del tipo normal (gaussiana), y que por la gran cantidad de datos relevados ya se encuentra ampliamente difundida su utilización sin ningún tipo transformación, avalando su utilización bajo el Teorema Central del Límite [49].

2.2.2.1 Método de Filtrado 1.

Utilizando los datos que ofrecen los MR (no sólo el valor de rendimiento en sí), es posible “limpiar” puntos relevados que son muy diferentes a los valores medios o que no cumplen con valores lógicos bajo un análisis técnico-agronómico. De esta manera, en la Tabla 1 se listan los 5 filtros utilizados y sus características:

Nombre del Filtro	Campo / Atributo	Descripción del Filtro	Límite Superior	Límite Inferior
Humedad	HUMEDAD	Extracción de los puntos que tengan valores de HUMEDAD de grano que sean superiores o inferiores a +/- 3 SD.	media + 3 SD	media - 3 SD
Flujo	FLUJO	Extracción de los puntos que tengan valores de FLUJO de grano que sean superiores o inferiores a +/- 3 SD.	media + 3 SD	media - 3 SD
Altura	ALTURA	Extracción de los puntos que tengan valores de ELEVACIÓN de la plataforma de cosecha que sean superiores o inferiores a +/- 3 SD.	media + 3 SD	media - 3 SD
Capacidad Biológica	REND	Extracción de los puntos que tengan valores de RENDIMIENTO de grano que sean superiores al rendimiento potencial del cultivo en la zona de estudio, y/o que sean igual a cero.	SOJA 1: < 6 ton/ha SOJA 2: < 5 ton/ha MAIZ: < 14 ton/ha GIRASOL: < 4.5 ton/ha	> 0 ton/ha
Velocidad	DISTANCIA	Extracción de los puntos que tengan valores de DISTANCIA (trayecto recorrido en 3 segundos) que sean superiores a la distancia de avance recorrida a la velocidad de trabajo máxima recomendada y/o que sean menores a la mínima velocidad de trabajo recomendada.	< 8.33 metros (velocidad menor a 10 km/h)	> 0.83 metros (velocidad mayor a 1 km/h)
Rendimiento	REND	Extracción de los puntos que tengan valores de REND de grano que sean superiores o inferiores a +/- 3 SD.	media + 3 SD	media - 3 SD

Tabla 1 - Detalle del Método de Filtrado 1

Para aplicar los filtros se desarrolló un script en el software R-Studio que permitió realizar el filtrado lote a lote en cada uno de los cultivos, generando gráficos estadísticos y posteriormente guardando la nueva base de datos “limpia” de outliers.

El límite de capacidad biológica para cada tipo de cultivo se determinó analizando los valores máximos de rendimiento que arrojan los Ensayos Comparativos de Rendimiento (ECR) que realiza el Instituto Nacional de Tecnología Agropecuaria (INTA) en la zona de estudio. En campañas agrícolas con buena disponibilidad hídrica los rendimientos de las variedades/híbridos probados no superan los valores asignados como máximos en este trabajo.

En la Tabla 2 se ejemplifica el accionar (filtrado) de la secuencia de filtros aplicada sobre el lote 1 de Soja de primera.

Filtro por Atributo	Desvío estándar (REND)	Máx. (REND)	Media (REND)	n (REND)	Cantidad de datos eliminados	% eliminado	Tipo de Filtro
DATO SIN FILTRO	5,47	301,6	1,54	10842			
Humedad	3,76	257,5	1,49	10832	10	0,09%	+/-3 sd
Flujo	3,74	257,5	1,48	10810	22	0,20%	+/-3 sd
Altura	3,74	257,5	1,48	10757	53	0,49%	+/-3 sd
Cap. Biol.	0,96	5,99	1,36	10717	40	0,37%	>0 y <6
Velocidad	0,95	5,99	1,35	10691	26	0,24%	>0,83 y <8,33
Rendimiento	0,91	4,23	1,33	10617	74	0,69%	+/-3 sd
					Σ: 225	Σ: 2,08%	

Tabla 2 - Accionar del filtro sobre cada atributo.

2.2.2.2 Método de Filtrado 2.

Se utilizó un método de autocorrelación espacial que utiliza el Índice de Moran local (IMI) y que se recomienda aplicar al análisis de la variabilidad espacial de lotes agrícolas [50]. La fórmula del índice de autocorrelación espacial local de Moran es la siguiente:

$$IMI_i = \frac{X_i - \bar{X}}{\sigma^2} \sum_{j=1, j \neq i}^n [w_{ij} (X_j - \bar{X})]$$

donde X_i es el valor de la variable X en la posición i; \bar{X} y σ^2 es la media y varianza de X, respectivamente; X_j es el valor de la variable X en todos los otros sitios (donde $j \neq i$); w_{ij} es el peso espacial entre las ubicaciones i y j.

Para implementar este filtro se utilizó un script en el software R-Studio que utiliza la librería “geoR”, “rgdal” y “spdep” (contiene las funciones localmoran y moranplot). La medición de la correlación que una variable tiene sobre el espacio geográfico (analizando las unidades espaciales contiguas), da lugar a tres posibilidades [51]:

- Autocorrelación espacial positiva: las unidades espaciales vecinas presentan valores próximos. Indica una tendencia al agrupamiento de las unidades espaciales.
- Autocorrelación espacial negativa: las unidades espaciales vecinas presentan valores muy disímiles. Indica una tendencia a la dispersión de las unidades espaciales.
- Sin autocorrelación: no ocurre ninguna de las dos situaciones anteriores. Por lo tanto, los valores de las unidades espaciales vecinas presentan valores producidos en forma aleatoria.

El método utilizado es el propuesto por [50] que realiza una permutación al azar dentro de un proceso iterativo dato por dato, calculando el valor del IMI (función localmoran) y analizándolo con respecto a los valores de su vecindario (vecinos cercanos) definidos por una grilla y un radio de búsqueda determinados previamente. En nuestro caso se utilizó un radio de análisis de 25 mts que abarca unos 2000 m² de superficie y agrupa entre 50 y 100 puntos vecinos. Al realizar este test, inicialmente hay que definir la hipótesis nula (H_0) que responde a la afirmación “la correlación de la información de un sitio con la de sus vecinos es nula”, y la alternativa (H_a) “existe correlación de la información de un sitio con la de sus vecinos”. El índice utiliza el test de comparaciones múltiples conocido como “Bonferroni” para determinar el grado de significancia.

La función moranplot construye el diagrama de dispersión, ajusta un modelo de regresión lineal y calcula una serie de estadísticos de diagnóstico⁷. Los datos que se alejen de la recta de 45° sugieren sitios que presentan un valor de autocorrelación espacial que es diferente a la de su vecindario. Esta función calcula los índices para cada observación y considera a una observación como influyente si al menos uno de los índices de diagnóstico la detecta como tal [50]. Entonces, si algún dato (punto relevado por el MR) tiene el valor del Índice de Moran local menor o igual a 0, y/o son estadísticamente significativo (alfa de significancia 0.05), y/o es una observación influyente (determinado por la función moran.plot), será considerado outlier/inlier, y luego de identificarlo será eliminado del set de datos.

2.2.3 Cálculo del rendimiento por lote.

El cálculo del rendimiento en cada lote y por cada tipo de cultivo se realizó calculando el promedio del dato de rendimiento de todos los puntos del MR que conformaban cada lote. Esta tarea se realizó con el armado de 2 scripts en el software R-Studio, uno con el método de filtrado 1 y otro con el método de filtrado 2.

Los valores medios por cada cultivo se compararon con el único dato oficial disponible de rendimiento [4] (que no posee error de estimación). Los valores de rendimiento promedio para el partido de Daireaux en la campaña 2008/2009 son: Soja=1,38 ton/ha (con 95.000 has sembradas); Maíz=4,8 ton/ha (con 16.000 has sembradas) y Girasol=2,5 ton/ha (con 11.500 has sembradas). Los lotes analizados en este trabajo y que se corresponden con productores del

⁷ Distancia de Cook, Leverage, DFFITS, DFBETAS y COVRATIO (Drapper y Smith, 1998)

grupo CREA lograron rendimientos más elevados que los valores medios oficiales (datos filtrados por el método 1): Soja (total de primera + segunda) = 2,2 ton/ha, Maíz = 8,6 ton/ha y Girasol = 2,8 ton/ha. El único cultivo que no ingresa dentro del margen de variación dado por el desvío estándar, es el maíz. Sin embargo, al ser un cultivo que requiere un alto grado de insumos (paquete tecnológico), es probable que productores pertenecientes a grupos CREA logren mejores rendimientos que el promedio de todo el partido.

Los resultados obtenidos luego de aplicar ambos tipos de filtrados (media, error estándar, intervalo de confianza, desvío estándar y rango del desvío) se comparan en la Tabla 3 para cada uno de los cultivos bajo análisis.

**Resultados de la aplicación de los métodos de filtrado sobre los datos de monitores de cosecha.
Campaña 2008/2009 - Partido de Daireaux.**

		Método 1	Método 2
Soja de primera	Cantidad de lotes	39	39
	Rinde promedio (ton/ha)	2,4	2,5
	Error estándar (ton/ha)	0,11	0,12
	Intervalo de Confianza de la Media (95%)	2,2 < 2,4 < 2,6	2,3 < 2,5 < 2,7
	Desvío estándar (ton/ha)	0,70	0,77
	Rango Desvío	1,7 < 2,4 < 3,1	1,7 < 2,5 < 3,3
Soja de segunda	Cantidad de lotes	18	18
	Rinde promedio (ton/ha)	1,4	1,4
	Error estándar (ton/ha)	0,08	0,07
	Intervalo de Confianza de la Media (95%)	1,29 < 1,4 < 1,58	1,29 < 1,4 < 1,58
	Desvío estándar (ton/ha)	0,32	0,32
	Rango Desvío	1,08 < 1,4 < 1,72	1,08 < 1,4 < 1,72
Maíz	Cantidad de lotes	21	21
	Rinde promedio (ton/ha)	8,6	9,1
	Error estándar (ton/ha)	0,29	0,30
	Intervalo de Confianza de la Media (95%)	8,02 < 8,6 < 9,14	8,47 < 9,1 < 9,65
	Desvío estándar (ton/ha)	1,31	1,38
	Rango Desvío	7,29 < 8,6 < 9,91	7,72 < 9,1 < 10,48
Girasol	Cantidad de lotes	11	11
	Rinde promedio (ton/ha)	2,8	3,2
	Error estándar (ton/ha)	0,01	0,20
	Intervalo de Confianza de la Media (95%)	2,57 < 2,8 < 3,13	2,81 < 3,2 < 3,61
	Desvío estándar (ton/ha)	0,47	0,68
	Rango Desvío	2,33 < 2,8 < 3,27	2,52 < 3,2 < 3,88

Tabla 3 - Resultados de ambos métodos de filtrado de datos.

2.2.4 Análisis estadístico comparativo entre los métodos de filtrado.

Con los valores medios por cada uno de los lotes y cada uno de los cultivos, se procedió a realizar una comparación del Método 1 y el Método 2 con el objetivo de determinar cuál método utilizar.

Se confeccionó una tabla para cada cultivo en el cual las filas corresponden a los lotes y las columnas tienen los siguientes atributos: n1 y n2 (cantidad de datos), r1 y r2 (rendimiento medio) y luego se generó una nueva columna con la diferencia entre medias ($r1 - r2$). Posteriormente, se realizó una comparación de medias apareadas que se detalla en el Anexo I y los resultados se visualizan en la Tabla 4.

	Análisis de Medias Apareadas
SOJA 1	DIFERENTES
SOJA 2	IGUALES
MAÍZ	DIFERENTES
GIRASOL	DIFERENTES

Tabla 4 - Resultados de la comparación estadística de los métodos de filtrado.

Finalmente, el análisis de la metodología a implementar realizado sobre los datos 2008/2009 arrojó resultados diferentes (excepto Soja2), demostrándolo estadísticamente con el análisis desarrollado en el Anexo I. Por lo tanto, es necesario optar por seleccionar uno u otro método de limpieza de los datos de campo.

Con el objetivo de observar espacialmente las diferencias en los datos extraídos, se realizó un análisis visual sobre los datos filtrados por uno u otro método utilizando el software Q-Gis y se encontraron algunos inconvenientes no contemplados previamente.

Seguidamente se analiza un ejemplo para un lote de Soja de primera en el cual se observan sectores en los cuales el método de filtrado 2 no actúa correctamente debido a que varios puntos (vecinos cercanos) se encuentran superando umbrales máximos permitidos por la capacidad biológica del cultivo. Como se muestra en la Figura 4, en el momento de cosecha sobre un sector del lote hubo problemas en la recolección del dato, con lo cual, se registraron valores de rendimiento para la soja de primera superiores a las 6 ton/ha. Debido a que no era un único dato elevado, sino seis (valores: 12.37, 73.93, 11.09, 9.88, 8.03, 19.08) el filtro de autocorrelación espacial eliminó únicamente los dos valores más extremos de ese pequeño vecindario: 73.93 ton/ha y 19.08 ton/ha, dejando otros cuatro datos que se encuentran fuera del rango biológico del cultivo sin eliminar (Figura 5).

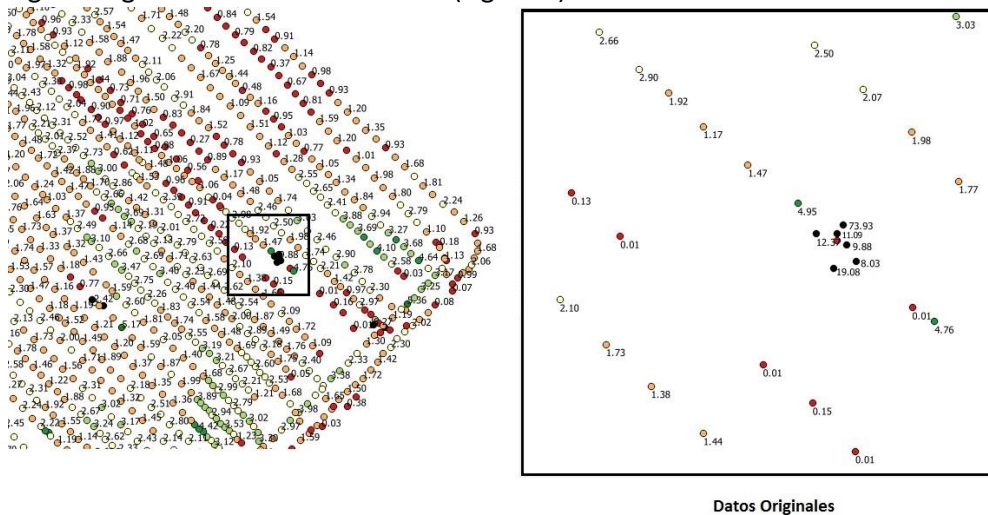


Figura 4 - Datos del MR sin filtrar.

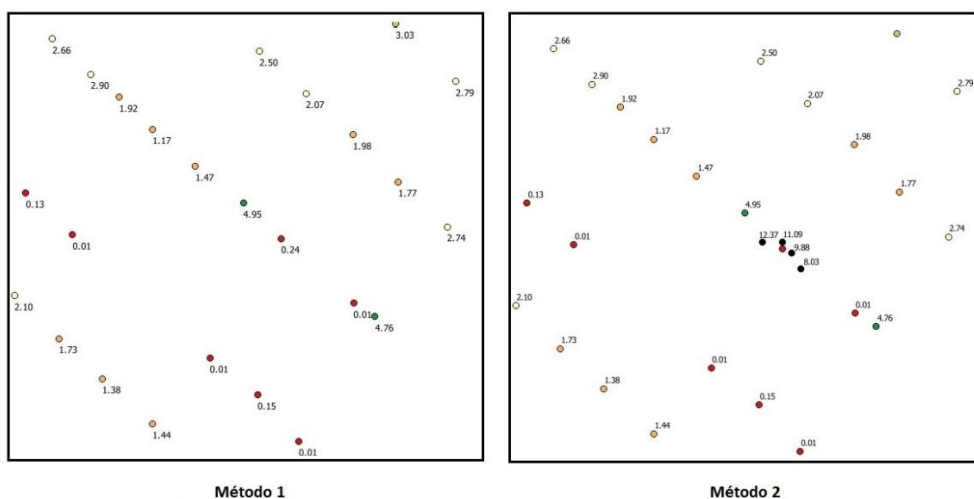


Figura 5 - Datos del MR filtrados.

Considero que debido a los posibles errores de este tipo que pueda presentar el método 2, y que se evidencia también en la sobreestimación del rendimiento (Anexo I), decido trabajar con el Método 1 de filtrado de datos en ambas campañas agrícolas, que es una metodología que

propone el INTA⁸. Cabe aclarar que el método 2 de autocorrelación espacial, fue muy satisfactorio en su resultado, aunque considero que es necesario dedicarle un mayor tiempo a definir sus parámetros internos y/o realizar previamente el corte por capacidad biológica para cada tipo de cultivo. De esta manera, se puede evitar la sobrestimación del rendimiento causada cuando el MR genera vecindarios de puntos de alto valor (atoramiento del sensor) y que se encuentran fuera del rango biológico.

2.2.5 Cálculo del rendimiento por pixel.

Es necesario para el desarrollo del trabajo de investigación, contar con el valor de rendimiento a nivel pixel. Considerando el tamaño del pixel en función de la resolución espacial de los sensores Landsat, se decidió crear una malla (grid) de 30 mts x 30 mts que permita agrupar en cada celda los puntos del MR (dato de campo). De esta manera, cada celda de 30 m x 30 m que se encuentre incluida completamente dentro de los lotes de cultivo, tendrá en su tabla de atributos el valor promedio de los puntos (datos de rendimiento) que se encuentren en su interior.

2.2.5.1 Generación de los polígonos de lotes cultivados.

Utilizando la herramienta "Concave Hull" del software Q-Gis se conformaron los polígonos de cada lote cultivado en los cuatro cultivos. Esta herramienta recibe como input la nube de puntos (datos del MR) y genera un polígono que envuelve a dichos puntos. En la Figura 6 se observan los datos del MR (nube de puntos) y luego el polígono obtenido. El algoritmo reconoce el perímetro y lo "copia" con un nivel de precisión ajustable por el usuario.

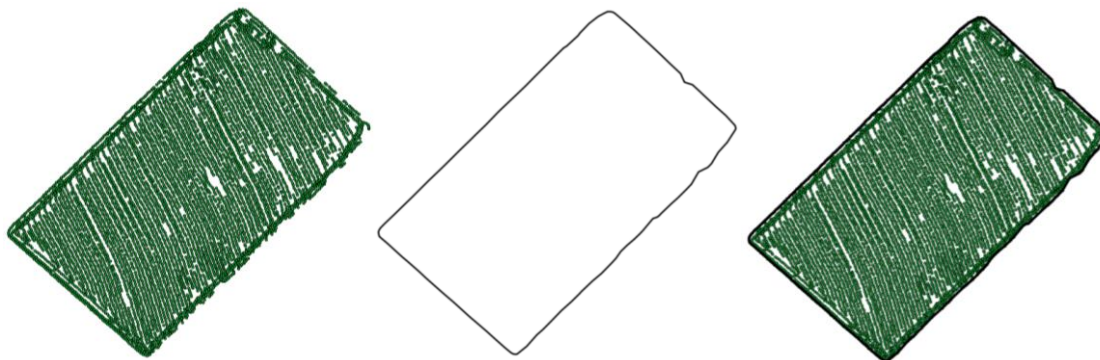


Figura 6 - Poligonización de los datos del MR.

2.2.5.2 Generación de la grilla.

Debido a que la grilla a utilizar debía corresponderse con el pixel de las imágenes satelitales para que no exista corrimiento entre lo observado por el sensor del satélite y lo medido a campo por el MR, fue necesario crear la grilla a partir de una imagen satelital (resultado del procesamiento raster: capítulo 2.3 "Información Satelital"). Utilizando el software ENVI y una imagen Landsat 5 TM libre de nubes, se procedió a realizar sucesivas multiplicaciones y divisiones de las bandas espectrales (bandas: 1,2,3,4,5,7) con el objetivo de discriminar los pixeles unos de otros (que ningún pixel quede con valor igual a alguno de sus pixeles vecinos). Una vez obtenida la nueva imagen, se exportó como tiff y utilizando la herramienta "raster to vectorial" del software Q-Gis se convirtió a polígono y cada uno de los pixeles pasó a formar una

⁸ INTA: Instituto Nacional de Tecnología Agropecuaria (<https://inta.gob.ar/>)

celda de 30 m x 30 m. A esta grilla obtenida en formato shape se le realizó un “clip” con el límite del partido, para obtener únicamente una grilla de celdas del partido de Daireaux.

Utilizando la herramienta “Selección por Localización”, se utilizaron los polígonos (generados en el ítem 2.2.5.1) con los lotes de cada tipo de cultivo para seleccionar las celdas que se encontraban incluidas completamente dentro de los polígonos que definen el lote cosechado. El resultado final es un shape con todas sus celdas de tamaño 30 m x 30 m y para cada tipo de cultivo. En la Figura 7 puede notarse que no fue necesario realizar un buffer de -30 m (como para reducir la superficie a un pixel menos y evitar posibles errores⁹) porque únicamente se incluyen las celdas que se encuentran completamente dentro del polígono del lote.

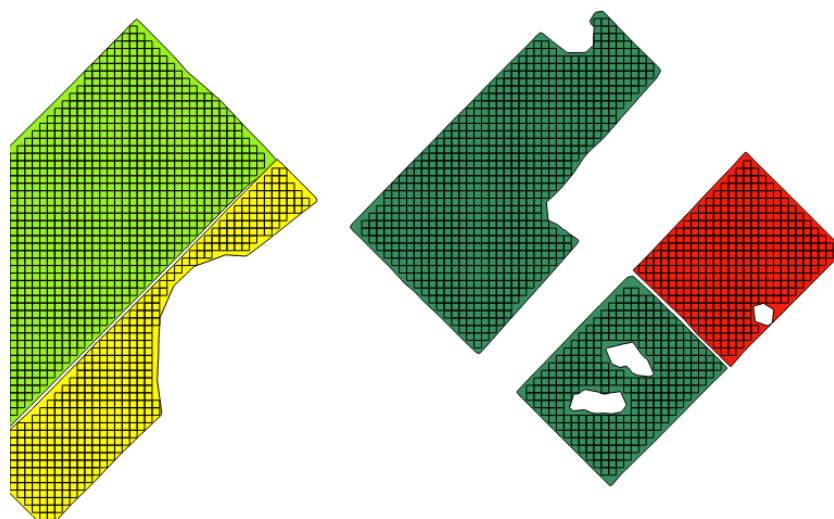


Figura 7 - Grillas generadas sobre los 4 tipos de cultivo (verde oscuro: Soja1, verde claro: Soja2, rojo: Maíz, Amarillo: Girasol).

2.2.5.3 Estimación del rendimiento a nivel pixel.

Los datos de los MR posterior a su filtrado fueron procesados por el software FastMapping [52], realizando un kriging ordinario a cada lote de manera individualizada por el atributo “REND”. Este software permite generar el semivariograma utilizando varios modelos, en nuestro caso utilizamos como opciones los modelos: esférico, exponencial y gaussiano. Luego, el mismo software procesa y realiza la interpolación kriging seleccionando el modelo con menor error (mayor ajuste). La resolución de salida de los datos interpolados se indicó en 5 m x 5 m, una resolución espacial similar a la obtenida por el MR (aproximadamente 7 m x 5 m). Los datos se guardaron de manera individualizada por lote y luego en Q-Gis se unificaron por tipo de cultivo.

Con la grilla generada anteriormente y los datos de los MR interpolados por el método kriging, se procedió a calcular el rendimiento a nivel pixel (30 m x 30 m). Utilizando la herramienta “Unir atributos por localización” del software Q-Gis e indicando a la grilla de celdas que asigne el promedio del campo rendimiento de los “objetos espaciales que interseccionen”, fue posible asignar a cada celda de la grilla el promedio de los valores de rendimiento de los puntos del MR que contenía. Este procedimiento se visualiza en la Figura 8.

⁹ Pixeles que incluyan superficie que no se corresponde con el cultivo en cuestión.

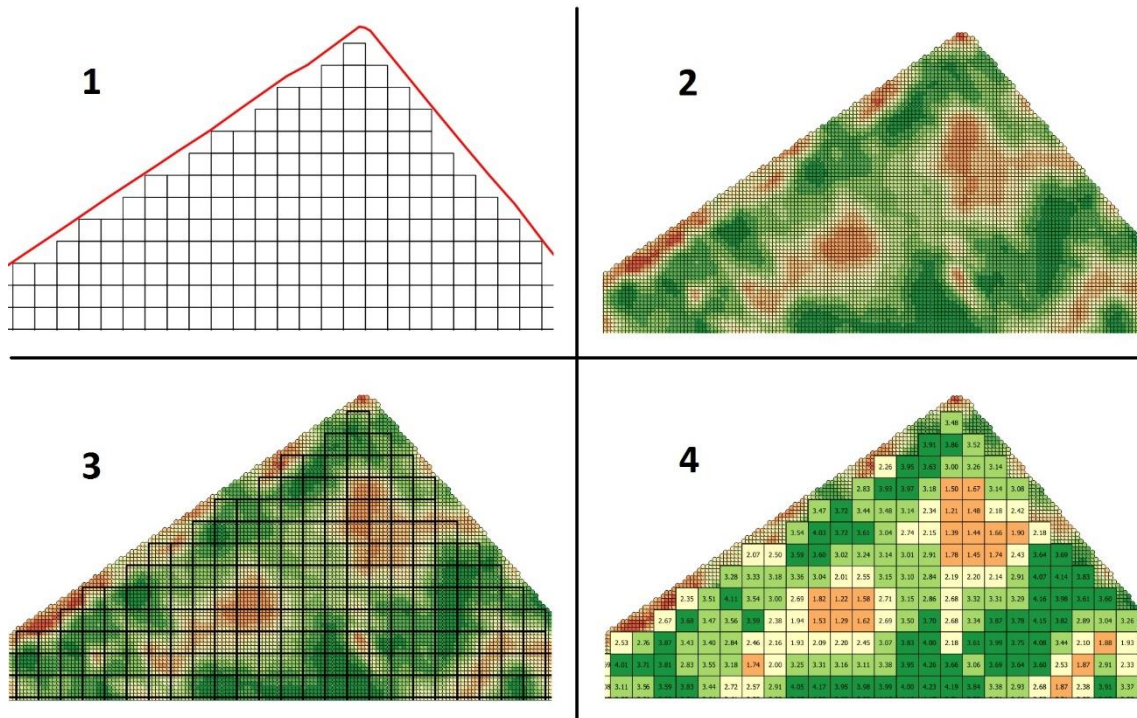


Figura 8 - Cálculo del rendimiento a nivel pixel. 1) Grilla y polígono del lote. 2) Datos del MR luego del Kriging. 3) Intersección con la grilla. 4) Celdas de 30x30 con el valor de rendimiento asignado.

2.2.6 Rasterización de datos de campo.

Los datos de campo convertidos a polígonos con el atributo del rendimiento en cada celda se unificaron en un único shape (los 4 tipos de cultivos quedaron juntos) y se convirtieron a formato raster respetando el tamaño de imagen del stack raster, para luego vincularlo con la secuencia multitemporal de imágenes de índices de vegetación. Para esto se utilizó la herramienta “v.to.rast.attribute” del software Qgis generando una imagen GEO-TIFF (Figura 9).

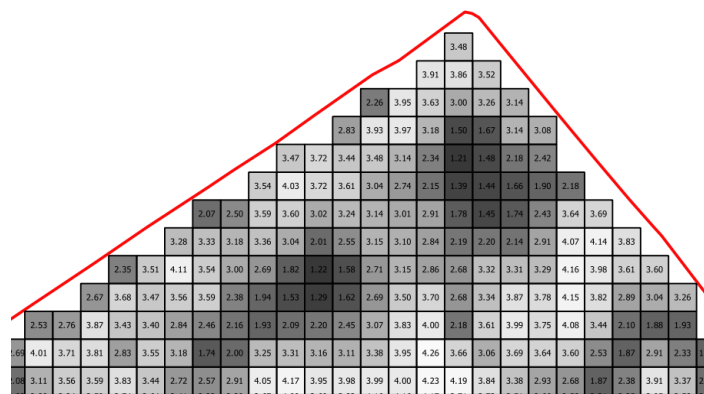


Figura 9 - Rasterizado de las celdas de rendimiento.

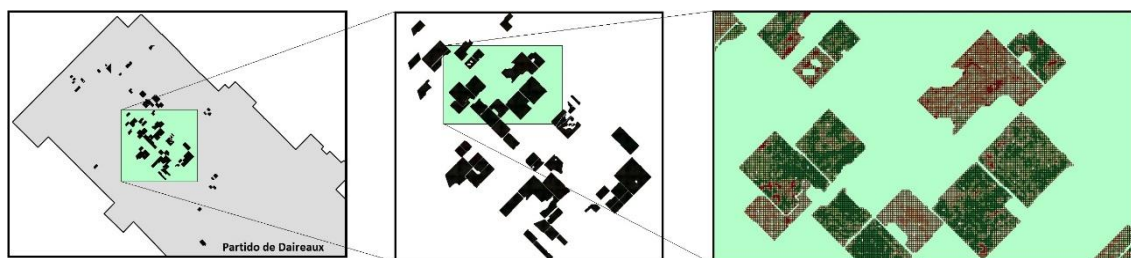


Figura 10 - Visualización conjunta de los lotes procesados para los 4 tipos de cultivo.

2.2.7 Control de calidad y división del set de datos (modelado y validado).

Se realizó un control visual sobre todas las capas generadas, chequeando el posicionamiento geográfico y la cantidad de datos para cada tipo de cultivo. También se revisaron las tablas de atributos y se realizaron cálculos manuales (3 casos por tipo de cultivo) del promedio de los puntos del MR que intersectaban con alguna celda particular. Por otro lado, se calculó la media de los valores de rendimiento asignados para la grilla (3 casos por tipo de cultivo) para obtener el rendimiento promedio del lote y compararlo con el rendimiento promedio obtenido en el total de la nube de puntos filtrada.

Luego de controlar los datos, se dividieron las celdas (los “grid” individuales para cada cultivo) generando dos sets de datos, uno para realizar el modelo y otro para su posterior validación. Se optó por realizar una selección al azar, separando los valores pares e impares del campo “ID” de las celdas (fórmula de selección: "ID" % 2 > 0)¹⁰ y exportando como “set_modelado” e invirtiendo la selección (invertir selección de objetos) y exportando como “set_validado”. Con este tipo de criterio, la base de datos fue dividida dejando aproximadamente el 50% de los datos en cada set y para cada tipo de cultivo. Siempre el set de modelado fue originado por los valores pares de la grilla obtenida y no se intercambiaron los sets de datos (modelación y validación) al generar los modelos¹¹. El campo ID de las celdas surge del proceso de generación (multiplicación de bandas espectrales), por el cual su valor no es relevante, pero permite la división de los pixeles pares e impares de manera homogénea sobre los lotes¹². Para la campaña agrícola 2008/2009 la cantidad de datos analizados fue la siguiente:

- Soja de Primera: 24.048 pixeles totales, 11.884 modelado y 12.164 validación.
- Soja de Segunda: 10.917 pixeles totales, 5.405 modelado y 5.512 validación.
- Maíz: 14.630 pixeles totales, 7.276 modelado y 7.354 validación.
- Girasol: 7.282 pixeles totales, 3.613 modelado y 3.669 validación.

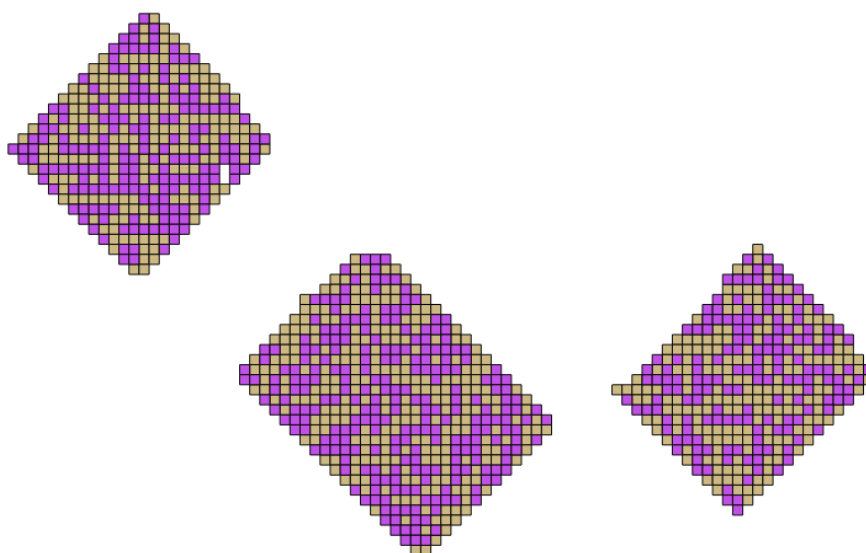


Figura 11 - Visualización de la distribución de celdas pares e impares.

¹⁰ El símbolo “%” entrega el resto de la división, en el caso de los números pares el resto es igual a cero.

¹¹ Al ser miles de celdas (pixeles) por cada tipo de cultivo, no se realizó un análisis detallado de la representatividad de cada celda para conformar una u otra base de datos.

¹² En el caso de generar una grilla con el comando “creargrilla” de Q-Gis, la misma no coincidirá con las celdas Landsat (habría que georreferenciar) y, además, las distribuciones de celdas pares e impares es por fila, con lo cual se generan “bandas” de datos para modelado y validación sin una distribución homogénea sobre el espacio.

2.3 Información Satelital.

En cuanto a la información satelital se utilizaron series temporales de imágenes de los sensores Landsat 5 y Landsat 7, calibradas (reflectancia TOA¹³), coregistradas y con extracción de: nubes, sombras y fallas. Las imágenes seleccionadas para cada campaña agrícola abarcan todo el ciclo estival iniciando a comienzos de octubre y finalizando en los últimos días de mayo. Posteriormente, se calcularon índices de vegetación (NDVI, MSAVI, EVI) y el índice de humedad NDMI (estos índices se describen en la pág. 32), para cada fecha de imagen disponible y se constituyeron en un set multitemporal de índices.

2.3.1 Adquisición de datos.

Las imágenes satelitales utilizadas en este trabajo corresponden a la serie Landsat [53] (Figura 12). Se utilizaron de manera simultánea imágenes de los satélites Landsat 5 Thematic Mapper (L5-TM) y Landsat 7 Thematic Mapper Plus (L7-TM+) para el desarrollo de los modelos de estimación de rendimiento sobre las campañas agrícolas 2008/2009 y 2010/2011, las mismas se visualizaron e identificaron utilizando el catálogo de la CONAE (catálogo web)¹⁴.

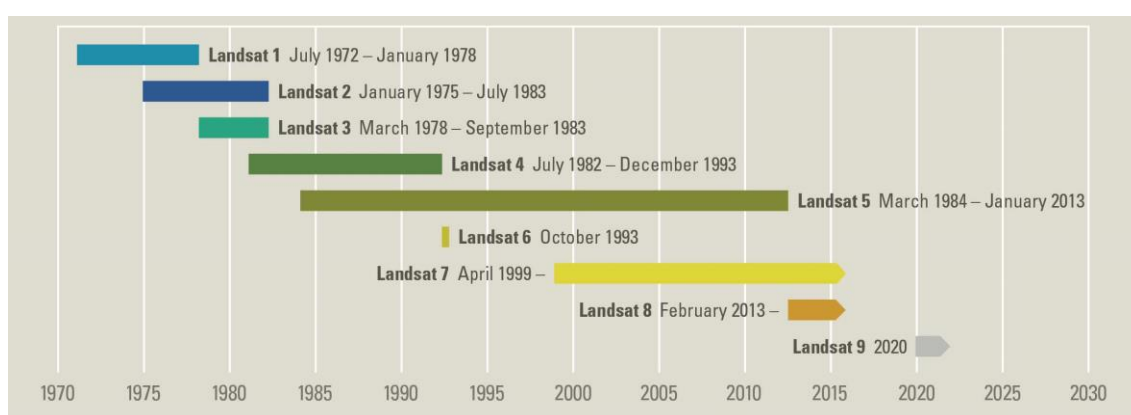


Figura 12 - Serie Landsat

La selección de las imágenes se realizó teniendo en cuenta la disponibilidad y la nubosidad sobre el área de estudio. El criterio de búsqueda utilizado para la campaña agrícola 2008/2009 fue el siguiente: Path: 227, Row: 085 (según el TGS¹⁵), Fecha inicial: 01/10/2008, Fecha final: 31/05/2009 tanto para L5-TM como para L7-TM+. En el caso de la campaña agrícola 2010/2011 se utilizó el mismo criterio de búsqueda, pero con fecha inicial: 01/10/2010 y fecha final: 31/05/2011 (Tabla 5).

El Path-Row seleccionado no cubre de manera completa todo el partido de Daireaux, sin embargo, el área no cubierta por las imágenes es principalmente ganadera con suelos bajos y generalmente anegables. Como se observa en la Figura 13 el principal sector agrícola dentro del partido se encuentra en la parte noroeste (tonalidades rojas debido a la combinación de bandas seleccionada).

¹³ Los niveles digitales (DN) se convirtieron a valores de reflectancia a tope de atmósfera (TOA) que no tiene en cuenta las posibles variaciones que genera la atmósfera terrestre.

¹⁴ Sitio web: <https://catalogos.conae.gov.ar/>

¹⁵ Tile Grid System: Es un sistema de mosaico global generado para la misión Landsat que permite presentar las escenas en cuadrículas compuestas por un Path (camino/columna) y Row (fila). Cada Path-Row cuenta con una dimensión de 5000 x 5000 píxeles, aunque en realidad las imágenes tienen una dimensión aún mayor que permite el solapamiento entre imágenes de los Path-Row vecinos, por ejemplo: L7-TM: 6166 x 5666, y L8-OLI: 6300 x 6000.

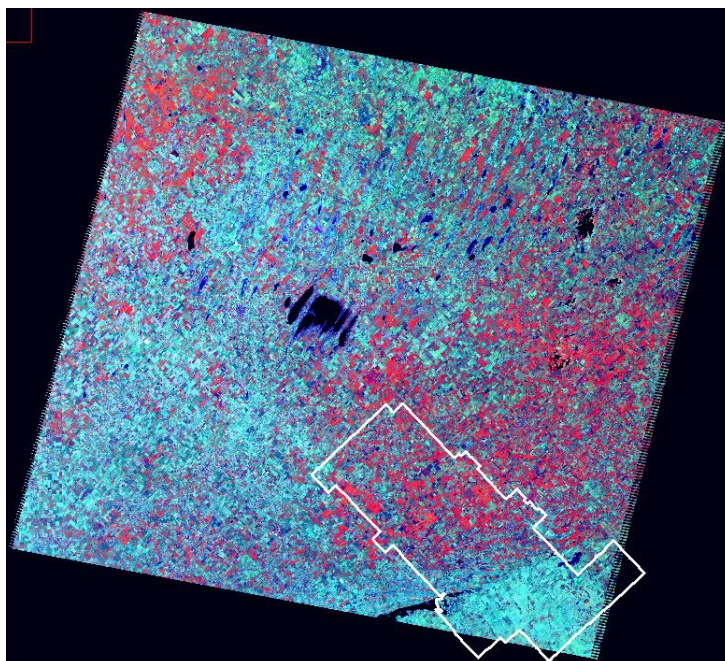


Figura 13 - Imagen L5-TM, P227-R085, RGB:453, Fecha 18/02/2009, con el límite del partido de Daireaux.

De la misma manera que se seleccionaron las imágenes de las campañas 2008/2009 y 2010/2011, se procedió sobre las imágenes L7-TM+ y Landsat 8 Operational Land Imager (L8-OLI) de la campaña agrícola 2016/2017, que son las imágenes para la implementación del modelo (Tabla 5).

PATH: 227 - ROW: 85		CAMPAÑA 2008/2009	PATH: 227 - ROW: 85		CAMPAÑA 2010/2011
FECHA	SENSOR	OBSERVACIONES	FECHA	SENSOR	OBSERVACIONES
20081005	L7	ALGUNAS NUBES	20101003	L5	SIN NUBES
20081013	L5	CON ELEVADA CANTIDAD DE NUBES	20101011	L7	SIN NUBES
20081021	L7	CON ELEVADA CANTIDAD DE NUBES	20101019	L5	SIN NUBES
20081029	L5	SIN NUBES	20101027	L7	NO DISPONIBLE
20081106	L7	SIN NUBES	20101104	L5	SIN NUBES
20081114	L5	ALGUNAS NUBES	20101112	L7	SIN NUBES
20081122	L7	SIN NUBES	20101120	L5	CON ELEVADA CANTIDAD DE NUBES
20081130	L5	SIN NUBES	20101128	L7	ALGUNAS NUBES
20081208	L7	CON ELEVADA CANTIDAD DE NUBES	20101206	L5	ALGUNAS NUBES
20081216	L5	ALGUNAS NUBES	20101214	L7	SIN NUBES
20081224	L7	SIN NUBES	20101222	L5	SIN NUBES
20090101	L5	CON ELEVADA CANTIDAD DE NUBES	20101230	L7	SIN NUBES
20090109	L7	SIN NUBES	20110107	L5	SIN NUBES
20090117	L5	CON ELEVADA CANTIDAD DE NUBES	20110115	L7	SIN NUBES
20090125	L7	CON ELEVADA CANTIDAD DE NUBES	20110123	L5	SIN NUBES
20090202	L5	CON ELEVADA CANTIDAD DE NUBES	20110131	L7	SIN NUBES
20090210	L7	CON ELEVADA CANTIDAD DE NUBES	20110208	L5	CON ELEVADA CANTIDAD DE NUBES
20090218	L5	SIN NUBES	20110216	L7	SIN NUBES
20090226	L7	SIN NUBES	20110224	L5	CON ELEVADA CANTIDAD DE NUBES
20090306	L5	ALGUNAS NUBES	20110304	L7	ALGUNAS NUBES
20090314	L7	CON ELEVADA CANTIDAD DE NUBES	20110312	L5	CON ELEVADA CANTIDAD DE NUBES
20090322	L5	NO DISPONIBLE	20110320	L7	SIN NUBES
20090330	L7	CON ELEVADA CANTIDAD DE NUBES	20110328	L5	SIN NUBES
20090407	L5	SIN NUBES	20110405	L7	CON ELEVADA CANTIDAD DE NUBES
20090415	L7	SIN NUBES	20110413	L5	NO DISPONIBLE
20090423	L5	ALGUNAS NUBES	20110421	L7	CON ELEVADA CANTIDAD DE NUBES
20090501	L7	CON ELEVADA CANTIDAD DE NUBES	20110429	L5	NO DISPONIBLE
20090509	L5	NO DISPONIBLE	20110507	L7	SIN NUBES
20090517	L7	CON ELEVADA CANTIDAD DE NUBES	20110515	L5	NO DISPONIBLE
20090525	L5	NO DISPONIBLE	20110523	L7	SIN NUBES
20090602	L7	SIN NUBES			

Tabla 5 - Set de Imágenes 2008/2009 (izquierda) y 2010/2011 (derecha). En color amarillo se indican las fechas no utilizadas por no encontrarse disponible la imagen o contener una elevada cantidad de nubosidad.

PATH: 227 - ROW: 85		CAMPAÑA 2016/2017
FECHA	SENSOR	OBSERVACIONES
20161003	L8	CON NUBES
20161011	L7	CON ELEVADA CANTIDAD DE NUBES
20161019	L8	CON ELEVADA CANTIDAD DE NUBES
20161027	L7	CON ELEVADA CANTIDAD DE NUBES
20161104	L8	SIN NUBES
20161112	L7	SIN NUBES
20161120	L8	CON ELEVADA CANTIDAD DE NUBES
20161128	L7	CON ELEVADA CANTIDAD DE NUBES
20161206	L8	SIN NUBES
20161214	L7	SIN NUBES
20161222	L8	CON ELEVADA CANTIDAD DE NUBES
20161230	L7	ALGUNAS NUBES
20170107	L8	ALGUNAS NUBES
20170115	L7	CON ELEVADA CANTIDAD DE NUBES
20170123	L8	ALGUNAS NUBES
20170131	L7	CON ELEVADA CANTIDAD DE NUBES
20170208	L8	CON NUBES
20170216	L7	CON ELEVADA CANTIDAD DE NUBES
20170224	L8	CON NUBES
20170304	L7	CON ELEVADA CANTIDAD DE NUBES
20170312	L8	SIN NUBES
20170320	L7	CON NUBES
20170328	L8	CON ELEVADA CANTIDAD DE NUBES
20170405	L7	SIN NUBES
20170413	L8	CON ELEVADA CANTIDAD DE NUBES
20170421	L7	SIN NUBES
20170429	L8	SIN NUBES
20170507	L7	CON ELEVADA CANTIDAD DE NUBES
20170515	L8	SIN NUBES

Tabla 6 - Set de Imágenes 2016/2017. En color amarillo se indican las fechas no utilizadas por no encontrarse disponible la imagen o contener una elevada cantidad de nubosidad.

2.3.2 Formación de la imagen.

Las imágenes se descargaron seleccionando el producto: “Producto en GeoTiff compactado .TAR.GZ” y se calibraron en reflectancia TOA cada banda espectral manualmente para Landsat 5 y Landsat 7 (en el caso de imágenes Landsat 8 se utilizó el procedimiento automático de DTI). Esta información incluye la banda de nubes (b9) y la imagen de calidad (QA)¹⁶ que se utilizarán para extraer la nubosidad. También se descargó el producto que contiene los índices espectrales ya generados: “Índices Espectrales en GeoTiff compactados como .RAR”. De este producto se utilizarán los siguientes índices: NDVI, MSAVI, EVI y NDMI¹⁷. Las bandas espectrales y los 4 índices se agruparon en un paquete (formación de la imagen) de manera ordenada: B1, B2, B3, B4, B5, B6, B7, NDVI, MSAVI, EVI, NDMI utilizando el software ENVI. El marco de referencia y el sistema de proyección se mantuvieron como se disponen en el sitio de descarga: WGS-84¹⁸ UTM Zona 20 Sur¹⁹. En la Figura 14 se observan los archivos descargados (imágenes TIFF y datos TXT) a la izquierda y la generación de la imagen de 11 bandas a la derecha (7 espectrales + 4 índices).

¹⁶ QA: Pixel Quality Assessment (Es una banda dentro de la imagen, que contiene información para identificar los píxeles que tienen condiciones adversas dadas por el instrumento, por la atmósfera o por la superficie).

¹⁷ NDVI: Índice de Vegetación de Diferencia Normalizada; MSAVI: Índice de Vegetación Ajustado al Suelo y Modificado; EVI: Índice de Vegetación Mejorado; NDMI: Índice Normalizado de Diferencia de Humedad.

¹⁸ WGS-84: Son las siglas en inglés de “World Geodetic System 84”, que significa Sistema Geodésico Mundial de 1984.

¹⁹ UTM: “Universal Transverse Mercator”. Sistema de coordenada transversal al Mercator.

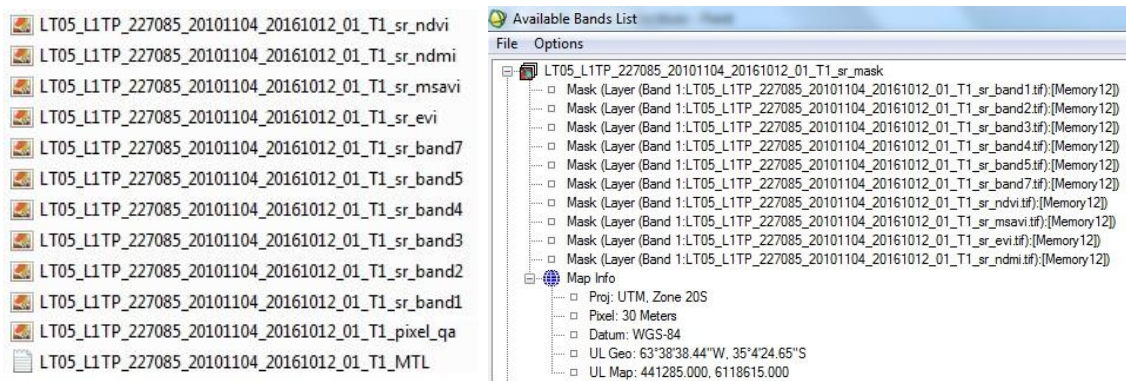


Figura 14 - Confección de la Imagen. Izq.: archivos tiff. y txt. descargados del sitio web. Der.: Enlazado ordenado de las bandas que formarán la imagen (en este caso ya tiene aplicada la máscara de extracción de nubes). La proyección de los datos observada en "Map Info" es la original del dato.

En DTI se trabaja desde los inicios con el Índice de Vegetación de Diferencia Normalizada conocido por sus siglas como "NDVI". Este índice estaba previsto ser utilizado desde el comienzo del trabajo de investigación, pero luego, se decidió incorporar otros índices con la idea de probar su desempeño en cada tipo de cultivo y en cada situación hídrica. La siguiente información sobre los índices fue extraída de [54] que cuenta con un listado que describe los productos que ofrece CONAE en su catálogo web. En este trabajo denominaremos a todos los índices como Índices de Vegetación, aunque el NDMI es un Índice de humedad.

- 1) Índice de Vegetación de Diferencia Normalizada (NDVI): Es un índice que permite identificar la presencia de vegetación verde en la superficie y caracterizar su distribución espacial. Se calcula como la diferencia normalizada entre la banda del rojo (R) y del infrarrojo cercano (NIR).

$$\text{NDVI} = (\text{NIR} - \text{R}) / (\text{NIR} + \text{R})$$

Dónde: NIR= Infrarrojo Cercano, R= Rojo

- 2) Índice de Vegetación Mejorado (EVI): Este índice incorpora un valor "L" para realizar un ajuste en relación al canopy, un valor "C" como coeficiente de resistencia de la atmósfera, y un valor correspondiente a la banda azul (B). Teniendo en cuenta estos coeficientes se realiza una corrección del índice normal reduciendo los ruidos producto de la interferencia de la atmósfera, el canopy y la saturación.

$$\text{EVI} = G * (\text{NIR} - \text{R}) / (\text{NIR} + \text{C1} * \text{R} - \text{C2} * \text{B} + \text{L})$$

Dónde: G=2.5, NIR= Infrarrojo Cercano, R= Rojo, B= Azul, L= 1, C1=6, C2=7.5

- 3) Índice de Vegetación ajustado modificado (MSAVI): Este índice se calcula como la diferencia normalizada entre la banda roja (R) y la banda del infrarrojo cercano (NIR) donde se aplica una función que determina el valor "L" para reducir el efecto del suelo en la señal de la vegetación.

$$\text{MSAVI} = (2 * \text{NIR} + 1 - \sqrt{((2 * \text{NIR} + 1)^2 - 8 * (\text{NIR} - \text{R}))}) / 2$$

Dónde: NIR= Infrarrojo Cercano, R= Rojo

- 4) Índice Normalizado de la Diferencia de Humedad (NDMI): Este índice se calcula como la diferencia normalizada entre el Infrarrojo cercano (NIR) y el infrarrojo medio (SWIR1, banda 6).

$$\text{NDMI} = (\text{NIR} - \text{SWIR}) / (\text{NIR} + \text{SWIR})$$

Dónde: NIR= Infrarrojo Cercano, SWIR= Infrarrojo Medio

2.3.3 Extracción de nubes/sombras/fallas.

Para la extracción de las nubes y sombras se utilizaron los algoritmos de extracción que se utilizan en DTI-ARBA (relación matemática entre bandas espectrales que incluye en un árbol de decisión la imagen QA) que genera como producto una máscara de nubes, sombras y fallas (falla del SLC²⁰). De esta manera los sectores con nubes, sombras y fallas son reemplazados para que no generen error en el análisis posterior. A través de la apertura de la imagen en RGB:453 se controló el nivel de extracción de la máscara aplicada. Posteriormente, en un análisis visual detallado, se optó por mejorar (en el caso que la imagen lo requiriese) utilizando la herramienta de construcción de áreas de interés (ROI) la digitalización de áreas no incluidas automáticamente. Las nuevas áreas se convirtieron a máscara y se volvió a aplicar sobre la imagen (ver Figura 15). Este nuevo producto se guardó con el nombre original de la imagen agregando el sufijo “_mask” haciendo referencia a la máscara aplicada sobre el stack.

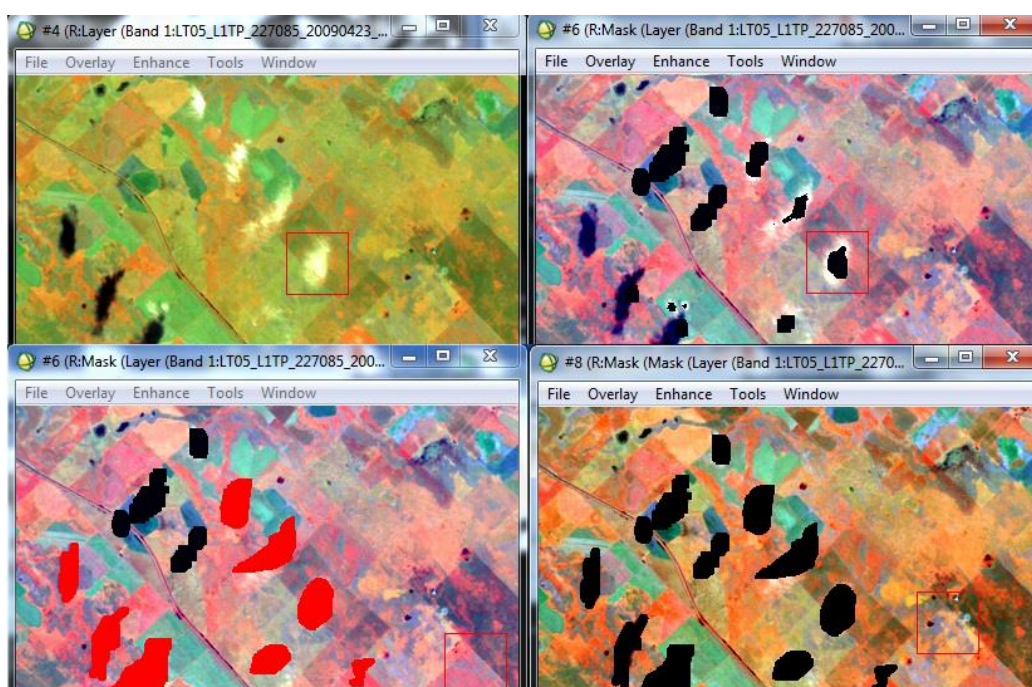


Figura 15 – Cuadrante superior izq.: imagen con nubes y sombras. Cuadrante superior der.: imagen con nubes y sombras enmascaradas automáticamente (clasificación). Cuadrante inferior izq.: clasificación de nubes y sombras mejorada a partir de la creación manual de ROI's. Cuadrante inferior der.: imagen con nubes y sombras enmascaradas.

2.3.4 Generación del pack por índice.

De cada una de las imágenes que componen el set para cada campaña (2008/2009; 2010/2011 y 2016/2017), se extrajeron los índices de vegetación de manera individualizada y se los dividió por 10.000 debido a que los datos descargados no se encuentran en el rango normal de los índices (-1 a 1) sino que se encuentran multiplicados por 10.000 para obtener un menor tamaño de archivo (peso). Estas nuevas imágenes se guardaron con el nombre anterior, pero agregando el sufijo correspondiente al índice, por ejemplo: “_NDVI”.

Estas imágenes tienen un posicionamiento espacial denominado “orbital” (define la georreferencia de la imagen según el posicionamiento del satélite sobre la órbita al tomar la

²⁰ Scan Line Corrector: Es un sistema corrector de la imagen que tiene el satélite Landsat7, y que a partir de mayo del año 2003 se averió. Esto produce la falta de datos en forma de cuñas a ambos laterales de la imagen.

imagen), que no siempre se encuentra bien definido. Es por este motivo que es necesario co-registrar las imágenes entre ellas. Con la herramienta “registration -> image to image” del software ENVI se realizó el proceso de co-registración. Se utilizó una imagen “base” disponible en DTI-ARBA (Imagen Landsat 8 libre de nubes del path row correspondiente) que tiene su posicionamiento verificado con respecto a un mosaico de imágenes SPOT²¹ de 2,5 m de resolución espacial²². Los corrimientos en la serie de imágenes Landsat, por lo general, correspondieron a 1 y 2 píxeles, salvo en algunas excepciones en las cuales se encontraron corrimientos de hasta 12 píxeles. En el proceso de co-registración se disminuyó el error de posicionamiento a menos de medio píxel entre imágenes.

Para obtener la equidistancia temporal entre fechas (imágenes de índices de vegetación) se agregaron imágenes con valor cero intercaladas²³ entre las imágenes con dato según correspondía (ver: Tablas 5 y 6) y luego, todas las imágenes previamente geoposicionadas se agruparon en un stack multitemporal (de octubre a mayo) ordenadas cronológicamente. Como se puede apreciar en la Figura 16, en la izquierda se observan los valores del índice de vegetación de un píxel (utilizando herramienta Z-Profile²⁴) sobre un lote agrícola, en un stack conformado con las fechas con orden temporal pero sin equidistancia temporal, y a la derecha el mismo píxel con la correcta separación temporal entre fechas.

Este proceso se realizó para los ocho meses de análisis (octubre a mayo) y cada campaña agrícola:

- 16 imágenes de la campaña agrícola 2008/2009 agregando 15 imágenes en valor cero. Stack NDVI de 31 fechas.
- 20 imágenes de la campaña agrícola 2010/2011 agregando 10 imágenes en valor cero. Stack NDVI de 30 fechas.
- 16 imágenes de la campaña agrícola 2016/2017 agregando 13 imágenes en valor cero. Stack NDVI de 29 fechas.

Luego se agregaron al stack dos nuevas imágenes, una correspondiente a la rasterización de los rendimientos medidos en campo (el valor del píxel corresponde al rendimiento en ton/ha) y otra correspondiente a la rasterización de los lotes (el valor del píxel corresponde al número del lote según el tipo de cultivo).

²¹ SPOT (Satellite Pour l'Observation de la Terre), son una serie de satélites de teledetección civiles para la observación terrestre desarrollados por el CNES (Centro Nacional de Estudios Espaciales francés).

²² Este mosaico SPOT fue realizado por la CONAE y compartido a ARBA a través del convenio marco de colaboración.

²³ Este método de intercalar bandas con valor cero en el set de imágenes fue una solución simple hallada previamente a mejorar los conocimientos en programación.

²⁴ Z-Profile: Herramienta del software ENVI que permite visualizar en un gráfico los valores de un determinado píxel (espectro) dentro de una imagen compuesta por más de una banda (stack).

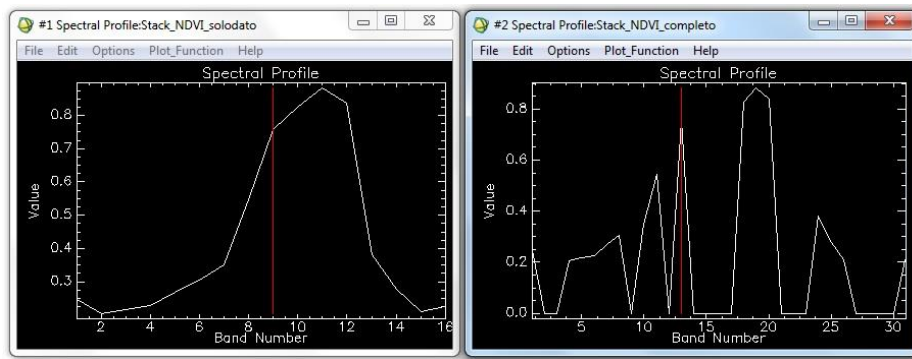


Figura 16 - Visualización de los valores de NDVI de un stack multitemporal sin equidistancia temporal (izquierda) y con equidistancia temporal (derecha)

2.3.5 Reproyección.

Como se observa en la Figura 14 la proyección del stack de imágenes se corresponde con la original (WGS-84, UTM zona 20 Sur), por lo que se re proyectó de manera completa todo el stack²⁵ de cada campaña agrícola al marco de referencia y proyección que utiliza la Agencia (Posgar 94, Gauss-Kruger Faja 5).

Posteriormente se corroboró la georreferenciación de los stack's re proyectados nuevamente con el mosaico SPOT. Este mosaico de imágenes se encuentra geoposicionado con la base de puntos GPS relevados por DTI-ARBA en toda la provincia y esto garantiza la correcta vinculación de los datos satelitales con la cartografía de ARBA. Cabe aclarar que es de suma importancia en los productos que desarrolla DTI, el correcto posicionamiento del set de datos satelitales, el set de datos de campo y la cartografía.

2.3.6 Control de calidad de los datos raster.

El control de la calidad de cada uno de los stacks de imágenes NDVI procesado se realizó haciendo la apertura de alguna de sus bandas (fechas) y analizando con la herramienta Z-Profile el comportamiento temporal de los lotes agrícolas. De esta manera es posible detectar errores en el ordenamiento cronológico de las imágenes, corrimientos de georreferenciación entre imágenes, chequear los valores de los índices, verificar la calidad de la extracción de nubes/fallas/sombras y chequear el valor de los reemplazos realizados.

²⁵ La opción de re proyectar las imágenes de manera independiente y luego unificarlas en un stack, no permitió una buena vinculación de las imágenes y se originaron corrimientos de 1 pixel entre algunas imágenes.

CAPITULO III - Desarrollo de los modelos.

La información para el desarrollo de los modelos de estimación del rendimiento consiste en valores de índices de vegetación que se relacionan con los valores de rendimiento de los cultivos. Los valores de los índices de vegetación obtenidos por cada fecha de imagen disponible se interpolaron para cubrir los datos faltantes (por imagen no disponible o extracción de nubes/fallas). La interpolación se realizó de manera lineal obteniendo un dato de índice de vegetación cada 8 días, y luego, se ajustó una función matemática de suavizado (interpolación de tipo “spline”) de manera de obtener datos diarios. Esto permitió conocer el comportamiento del crecimiento de cada tipo de cultivo a través del tiempo (modelado del crecimiento a nivel pixel) y analizar la “fenometría” de los cultivos.

Luego se generaron dos series de modelos para cada tipo de cultivo (Soja de primera, Soja de segunda, Maíz y Girasol), los modelos bajo régimen hídrico de sequía (por debajo de los valores normales) y los modelos bajo un régimen hídrico normal/húmedo. Los modelos se generaron relacionando valores de rendimiento (escalados a tamaño del pixel Landsat), con valores de los índices de vegetación (NDVI, MSAVI, EVI y NDMI). El momento óptimo de vincular los datos de rendimiento con el valor del índice fue definido a través del análisis multitemporal (curva de crecimiento del cultivo). Con las curvas de crecimiento para cada uno de los pixeles, se definió el Green-Up (momento en el que comienza a desarrollarse el cultivo, “enverdecimiento”) y el Green-Down (momento en que el cultivo concluyó su senescencia y fue finalmente cosechado) utilizando el cálculo de la derivada segunda de la función de crecimiento. Se realizó una serie consecutiva de regresiones lineales diarias a partir de la fecha establecida como inicio (fecha de inicio del estudio, Green-Up ó porcentaje del ciclo del cultivo definido a partir del Green-up y el Green-Down). El valor del coeficiente de determinación (R^2) de cada regresión lineal fue el indicador del ajuste del modelo. Una vez determinado la fecha de mayor ajuste del modelo de regresión (día del máximo R^2) se extrajeron los coeficientes de la función de regresión. Luego, fueron validados los modelos utilizando los datos de campo reservados para tal fin, realizando el cálculo de estimación del rendimiento con la función lineal hallada aplicada a los valores de los índices de vegetación. Por último, se compararon los resultados (rendimiento estimado) con el rendimiento observado (dato de campo reservado para validación), se realizaron los estudios de los residuos y se calcularon los errores para cada modelo.

3.1 Conformación de la base de datos.

Las celdas individualizadas por tipo de cultivo e identificadas como “Modelación” y “Validación” se cargaron como vector en el software ENVI y se transformaron en Región de Interés (ROI) para cada uno de los stacks de imágenes con los índices de vegetación. Luego fueron extraídos los valores de los índices de cada celda (pixeles) en forma conjunta con el valor de rendimiento y el número de lote en formato ASCII de forma individualizada por tipo de cultivo e índice de vegetación. Para la obtención de los modelos se utilizó la base de datos guardada como “modelación” y la base de datos “validación” se conservó para utilizar en la validación de los modelos obtenidos.

Cada pixel exportado contiene los siguientes atributos (Figura 17): “ID” (identificador: N° de pixel), “MapX” y “MapY” (coordenadas en la imagen), “Lat” y “Lon” (coordenadas

geográficas), 31 datos de índice de vegetación (de B1 a B31), “B32” (Rendimiento en Ton/ha) y “B33” (N° de Lote).

ID	MapX	MapY	Lat	Lon	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14
1	5309832.54	5980234.37	-36.306571	-62.117216	0.0000	0.0000	0.0000	0.2587	0.2863	0.2624	0.3050	0.3405	0.0000	0.5434	0.7161	0.0000	0.7734	0.0000
2	5309832.54	5980234.37	-36.306848	-62.116889	0.0000	0.0000	0.0000	0.2885	0.3032	0.2639	0.2915	0.3488	0.0000	0.4675	0.6659	0.0000	0.6488	0.0000
3	5309832.54	5980234.37	-36.306842	-62.117223	0.0000	0.0000	0.0000	0.3329	0.3634	0.3063	0.3133	0.3731	0.0000	0.4546	0.5535	0.0000	0.6340	0.0000
4	5309772.54	5980264.37	-36.307100	-62.117898	0.0000	0.0000	0.0000	0.3694	0.0000	0.3408	0.3263	0.3337	0.0000	0.3229	0.4556	0.0000	0.5077	0.0000
5	5309802.54	5980264.37	-36.307106	-62.117564	0.0000	0.0000	0.0000	0.3628	0.0000	0.3265	0.3263	0.3141	0.0000	0.3484	0.4419	0.0000	0.4775	0.0000
6	5309832.54	5980264.37	-36.307112	-62.117231	0.0000	0.0000	0.0000	0.3685	0.0000	0.3410	0.3161	0.3596	0.0000	0.4005	0.5009	0.0000	0.5272	0.0000
7	5309892.54	5980264.37	-36.307124	-62.116563	0.0000	0.0000	0.0000	0.2497	0.3015	0.2484	0.2537	0.3036	0.0000	0.4506	0.6306	0.0000	0.6596	0.0000
8	5309892.54	5980234.37	-36.307394	-62.116570	0.0000	0.0000	0.0000	0.2339	0.0000	0.2515	0.2452	0.2968	0.0000	0.4623	0.6236	0.0000	0.7673	0.0000
9	5309802.54	5980234.37	-36.307376	-62.117572	0.0000	0.0000	0.0000	0.3884	0.0000	0.3158	0.3240	0.3596	0.0000	0.3687	0.5009	0.0000	0.5395	0.0000
10	5309922.54	5980234.37	-36.307400	-62.116236	0.0000	0.0000	0.0000	0.2497	0.0000	0.2299	0.2523	0.2997	0.0000	0.5066	0.7296	0.0000	0.8233	0.0000

Figura 17 - Valores y atributos extraídos del stack EVI (base de datos 2008/2009) en formato ASCII.

Con el objetivo de realizar un análisis exploratorio de la información extraída del set obtenido para modelar, se realizaron boxplots multitemporales para cada tipo de cultivo, cada campaña agrícola y cada índice de vegetación (a continuación, Figura 18 y 19 se exponen ejemplos).

En el gráfico de la Figura 18 pueden observarse las fechas de imágenes satelitales disponibles, la variación de los valores medios (línea central del gráfico de caja) a lo largo del tiempo y la variación de los valores del índice EVI en el cultivo de soja de primera en cada fecha observada. Como puede apreciarse, inicia la campaña agrícola con valores bajos entre 0.1 y 0.2 y a partir del día juliano 2008335 (30 de noviembre) comienza a ascender su vigor vegetal alcanzando valores cercanos a 0.8 a comienzos del año 2009009 (9 de enero), mantiene valores más o menos constantes (plateau) hasta 2009065 (6 de marzo) y luego comienza la senescencia del cultivo bajando nuevamente al final del ciclo a los valores iniciales. También es posible observar en forma clara la distribución de las fechas (16 imágenes) de los datos satelitales a lo largo del tiempo y que el principal faltante de 4 fechas se evidencia (en la campaña 2008/2009) en el período de plateau. Otra información que se expresa en el gráfico es la variabilidad de los datos, inicia el análisis multitemporal con cajas pequeñas que luego se agrandan significativamente durante el período de crecimiento del cultivo y finaliza con cajas nuevamente pequeñas. Esto se debe a la gran diversidad que tiene el sistema productivo, con cultivares, largo de ciclos, fechas de siembras, manejos culturales, etc. muy diferentes, además de la variabilidad que ocurre a nivel intralote que se puede interpretar en la Figura 20. En la Figura 19 se puede apreciar el comportamiento de la soja de segunda, o sea, el cultivo de soja en segunda ocupación del lote para la misma campaña agrícola. En general se siembra sobre un rastrojo de un cultivo invernal recientemente cosechado que en la zona suele ser cebada o trigo (la cebada se cosecha unos días antes que el trigo, con lo cual se gana tiempo y mejora el desempeño del cultivo de segunda ocupación, en este caso la soja). En el período analizado (octubre-mayo, campaña 2010/2011) se puede observar el vigor vegetal obtenido por el cultivo antecesor que supera los valores medios de 0.8 de NDVI, luego se produce la senescencia hacia el fin del año calendario y reinicia el vigor vegetal del nuevo cultivo que en este caso es la soja. El máximo valor de NDVI lo logra en el mes de marzo, y en general, con niveles más bajos que un cultivo de soja de primera.

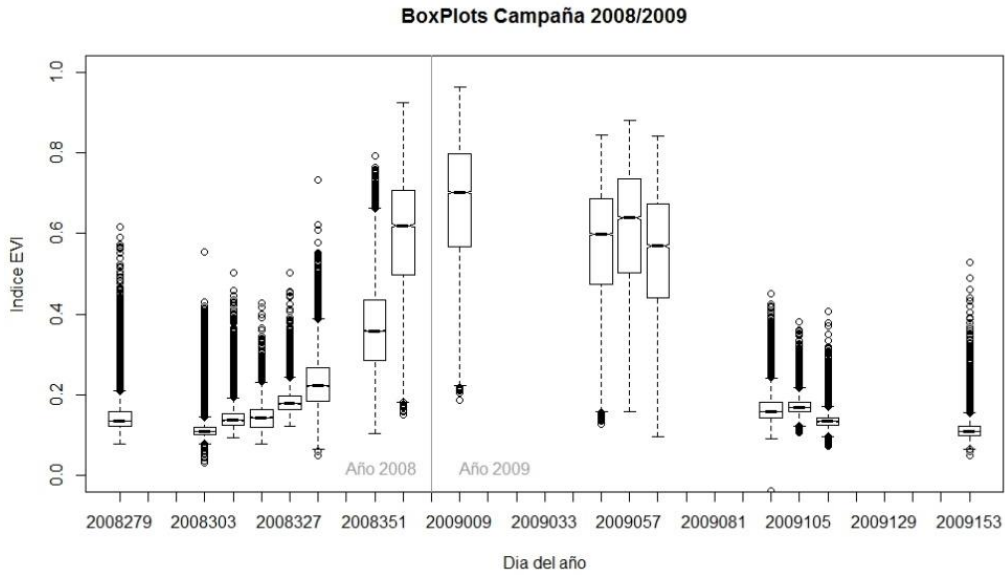


Figura 18 – Observación multitemporal de boxplots construidos con valores de índice de vegetación para el cultivo de Soja de primera. Día juliano en el eje X (05/10/2008 al 02/06/2009) e índice de vegetación EVI en el eje Y.

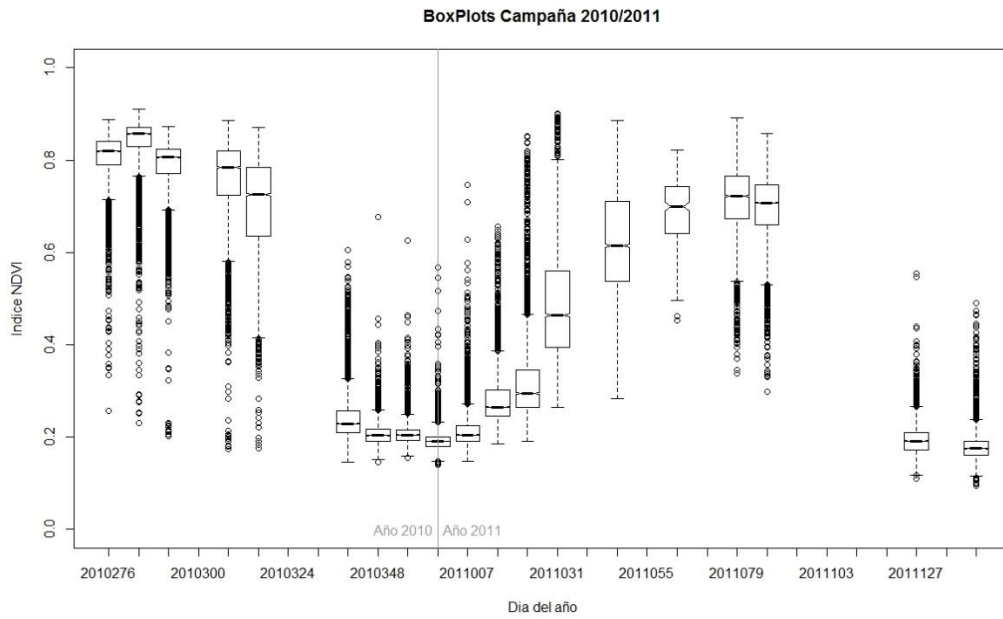


Figura 19 – Observación multitemporal de boxplots construidos con valores de índice de vegetación para el cultivo de Soja de segunda. Día juliano en el eje X (03/10/2010 al 23/05/2011) e índice de vegetación NDVI en el eje Y.

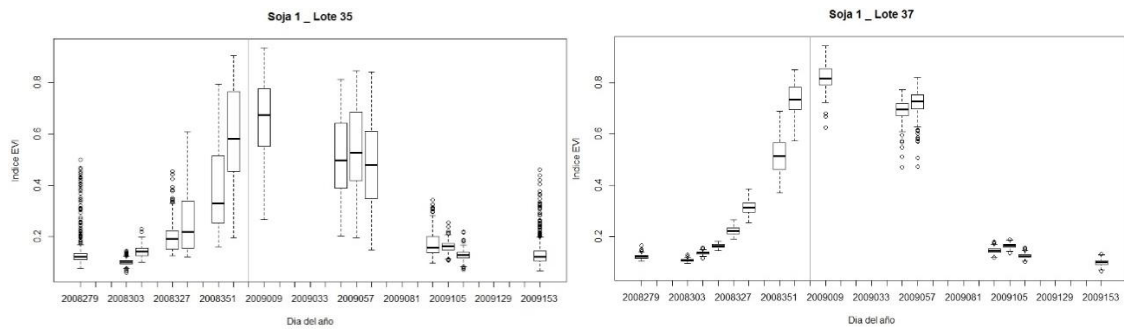


Figura 20 - Boxplots por Lote en Soja de primera. Se observa a la izquierda el lote 35 con alta variabilidad (heterogéneo) y a la derecha el lote 37 con baja variabilidad (homogéneo).

Se elaboraron 8 scripts en R-Studio para el análisis y desarrollo de los modelos (4 cultivos en 2 campañas agrícolas) en los cuales se define ingresar como inputs la información correspondiente al índice de vegetación que se desea procesar. En los mismos se realizó un análisis exploratorio de los datos (ejemplos en Figuras 18, 19 y 20) y se aplicó un filtrado de los datos de los índices de vegetación en cada fecha y cada lote de manera individualizada, eliminando los datos que se encontraban fuera del rango establecido por ± 3 desvíos estándar (eliminación de datos outliers).

3.2 Función de crecimiento.

Con el objetivo de lograr independizarse de la alta variabilidad en la disponibilidad de los datos satelitales ópticos generada principalmente por la nubosidad y las fallas (en el caso de Landsat 7), se procedió a realizar una interpolación lineal sobre la información de cada pixel para estimar los datos faltantes (imágenes con datos en valor cero). El total de pixeles con información de los 4 tipos de cultivos para la campaña 2008/2009 fue de 24764 y para 2010/2011 de 27978. De los 16 datos iniciales (fechas) para la campaña 2008/2009 (20081005, 20081029, 20081106, 20081114, 20081122, 20081130, 20081216, 20081224, 20090109, 20090218, 20090226, 20090306, 20090407, 20090415, 20090423, 20090602) y los 20 disponibles para la campaña 2010/2011 (20101003, 20101011, 20101019, 20101104, 20101112, 20101128, 20101206, 20101214, 20101222, 20101230, 20110107, 20110115, 20110123, 20110131, 20110216, 20110304, 20110320, 20110328, 20110507, 20110523), se obtuvieron los set completos con datos cada 8 días (resolución temporal de la constelación Landsat), o sea, una matriz de datos de 31 columnas y 24764 filas en 2008/2009 y de 30 columnas y 27978 filas en 2010/2011.

Una vez realizada la interpolación lineal (datos cada 8 días) se procedió a aplicar sobre ese nuevo set de datos una nueva interpolación de tipo “cubic spline²⁶” también aplicada por pixel. La misma se realizó con un paso de tiempo diario con el objetivo de suavizar las curvas fenológicas de crecimiento. Inicialmente, se aplicó el cubic spline directamente sin la interpolación lineal de 8 días, pero debido a faltantes de datos de hasta 4 fechas consecutivas, la respuesta no era óptima y los valores interpolados quedaban fuera del rango permitido por los valores de los índices de vegetación en cultivos (0 a 1).

Finalizada la interpolación se obtuvieron para los scripts 2008/2009 un set de datos con 240 columnas que es la cantidad de días entre la fecha inicial del estudio 05/10/2008 = 2008279, y la fecha final 02/06/2009 = 2009153. En cambio, para los scripts 2010/2011 se generó un set de datos con 232 columnas con fecha inicial 03/10/2010 = 2010276 y fecha final 23/05/2011 = 2011143.

Un proceso similar de interpolado pero con un suavizado más pronunciado (smooth.spline) se aplicó sobre el resultado obtenido por cubic spline y se guardó como una nueva variable. De este suavizado de los datos obtenido por cada pixel se extrajo la derivada primera y la derivada segunda con el objetivo de utilizarlas para hallar los puntos de comienzo del enverdecimiento conocido como “Green-Up”²⁷ y el final del ciclo del cultivo que

²⁶ Las interpolaciones de tipo “spline” son interpolaciones realizadas por segmentos (trozos) en los que se aplica una interpolación mediante polinomios. En el caso de cubic spline utiliza un polinomio de grado 3 para interpolar de manera segmentada.

²⁷ Como no es posible determinar con el uso de imágenes satelitales el momento de siembra del cultivo, se considera como punto de inicio el denominado enverdecimiento o Green-Up, que es posible de detectar con los índices de vegetación. La fecha de siembra según las condiciones del suelo y clima se encuentra aproximadamente un mes antes del Green-Up.

denominaremos “Green-Down” en contraposición con el primero. El proceso de suavizado de la función de crecimiento fue fundamental para lograr obtener valores derivados (derivada primera y segunda) con curvas que permitiesen extraer la información requerida. Si no se realiza el suavizado de la función de crecimiento, los sucesivos cambios de la función de crecimiento a lo largo del tiempo (cambios de concavidad, valles y crestas con sus mínimos y máximos locales) producen un efecto “serrucho” sobre los valores de la función derivada que no permite hallar el valor estimado de Green-Up y Green-Down (ver Figura 21). Para la determinación del inicio y fin del ciclo del cultivo, luego de una gran cantidad de pruebas realizadas, se decidió utilizar como indicador del momento de Green-Up y Green-Down el valor máximo de la derivada segunda de la función suavizada del crecimiento del cultivo [55] [56]. Para determinar el valor máximo de la derivada segunda en el inicio y en el final del ciclo, se utilizan rangos de tiempo definidos para cada tipo de cultivo.

3.2.1 Determinación del Green-Up y Green-Down.

Para cada tipo de cultivo se definió una zona temporal de búsqueda del Green-Up y el Green-Down que se corresponde con los días del año en el cual comienza y finaliza el cultivo. La misma se definió a partir de las fechas de siembra y cosecha en la zona agrícola de Daireaux. Por ejemplo, para la soja de primera, el período de siembra es del 15/10 al 15/11, y es muy variable debido a la gran plasticidad del cultivo, pero también porque depende del grupo de madurez (se siembra en la zona el Grupo III, IV y V) y del ambiente de siembra (Bajo, Media Loma y Loma). La cosecha del cultivo se realiza a partir de abril cuando el cultivo ya se encuentra en senescencia y la humedad del grano baja al orden del 10-13%. En la Figura 21 se grafica el comportamiento de un pixel de soja de primera y las fechas estimadas de Green-Up y Green-Down.

Posteriormente, utilizando el valor de la derivada segunda de la función de crecimiento suavizada, se hallaron el Green-Up y el Green-Down equivalentes al día del año en el que la derivada segunda obtiene el mayor valor (máxima aceleración que se produce tanto en el aumento como en el descenso del valor del índice) en los rangos de tiempo establecidos.

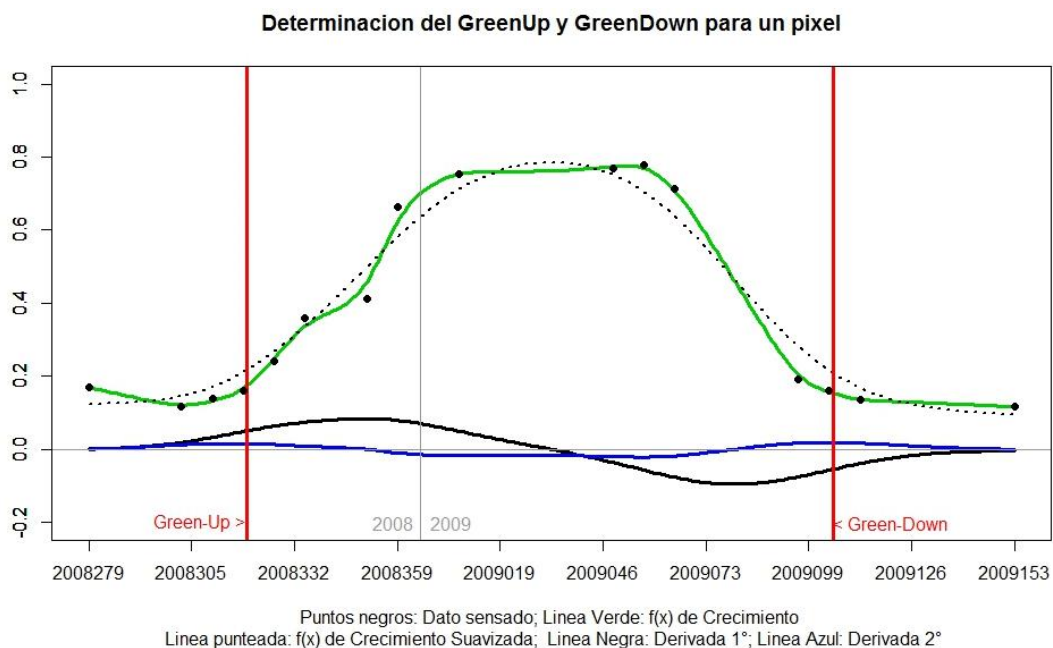


Figura 21 - Curva de crecimiento estimada para un pixel de Soja de primera.

El ciclo promedio del cultivo de soja de primera que surge del análisis es de 138 días, iniciando el Green-Up aproximadamente del 15/11 al 15/12, elevando los valores de los índices de vegetación a los primeros días del año y manteniendo el “plateau” (meseta) hasta inicios del mes de marzo (aproximadamente: 2009065 = 06/03/2009), luego comienza a disminuir el vigor vegetal por la traslocación de los fotoasimilados hacia los granos ingresando a la fase de senescencia. Cumple la madurez fisiológica del grano aproximadamente en el cambio de concavidad (mínimo de la derivada primera, 2009080= 21/03/2009), y continúa la senescencia del cultivo hasta cosecha, momento identificado como Green-Down (2009112= 21/04/2009), aunque puede realizarse la cosecha a partir de la madurez fisiológica del grano.

Esta determinación de Green-Up y Green-Down se realizó en cada pixel para cada tipo de cultivo, para cada tipo de índice de vegetación y para las dos campañas agrícolas bajo estudio.

3.2.2 Nuevo filtrado de datos.

Al comienzo del proceso (conformación de la base de datos) se aplicó un filtrado de los datos de los índices de vegetación en cada fecha y cada lote de manera individualizada, eliminando los datos que se encontraban fuera del rango establecido por +/-3 desvíos estándar (eliminación de datos outliers). Luego, al graficar el comportamiento de cada pixel temporalmente se encontraron casos fuera del patrón de crecimiento típico del cultivo. Por este motivo, se graficaron las curvas con los valores medios por tipo de cultivo y lote de producción (Figura 22) y se pudo detectar para cada campaña agrícola los lotes que tienen un comportamiento anómalo. Por ejemplo, para la campaña 2008/2009 fueron eliminados del análisis el conjunto de pixeles que comprendían 3 lotes anómalos de soja de primera (Figura 23), 1 lote en la soja de segunda, y 1 lote en el maíz.

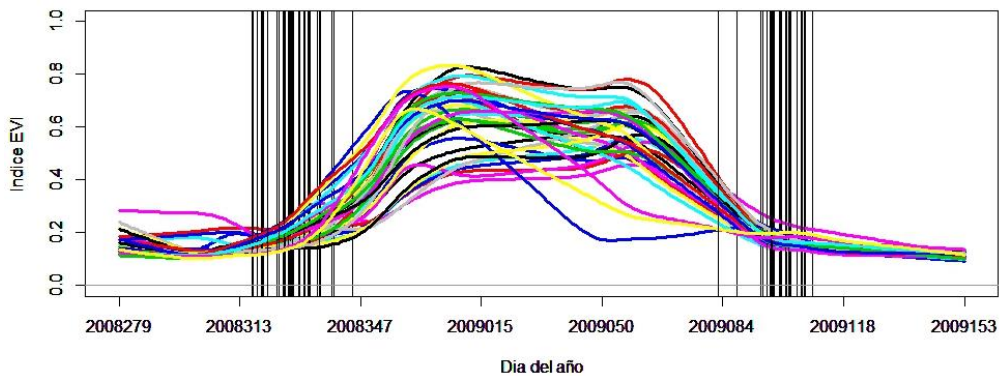


Figura 22 - Curvas de crecimiento estimadas para los 39 lotes de Soja de primera. Se observan 3 lotes anómalos que descienden a mitad de ciclo. Las líneas verticales corresponden al Green-Up y el Green-Down.

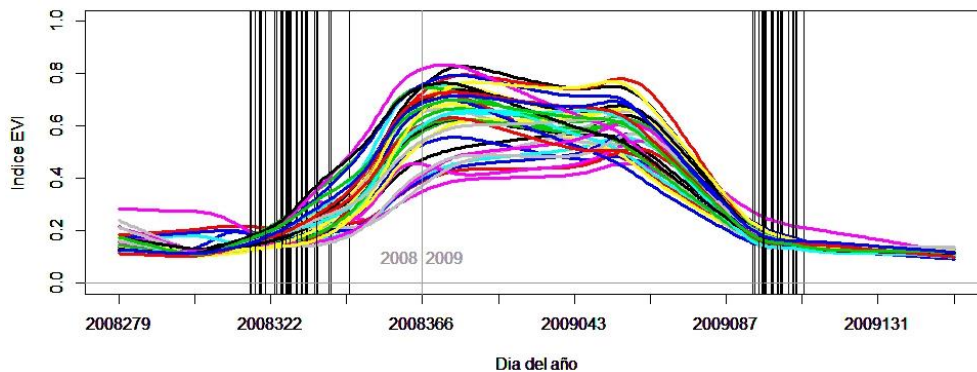


Figura 23 - Curvas de crecimiento estimadas para los 36 lotes de Soja de primera. Las líneas verticales corresponden al Green-Up y el Green-Down.

3.2.3 Períodos críticos.

Las especies agrícolas anuales cultivadas tienen un momento determinado de su ciclo en el cual las deficiencias hídricas, nutricionales y de sanidad se manifiestan de manera muy notable en los rindes obtenidos. En general existe una brecha entre el rendimiento de los cultivos bajo condiciones de producción (a campo) y el potencial de rendimiento bajo condiciones controladas sin limitantes (ensayos). Los factores limitantes del rendimiento no revisten la misma importancia en todo el ciclo, sino que toman mayor relevancia en determinados momentos en los que el cultivo se encuentra con mayor necesidad para la definición de los componentes del rendimiento. En general, la cantidad de asimilados producidos a nivel del cultivo está directamente relacionada con la tasa de crecimiento lograda (materia seca producida por unidad de tiempo) en dicho periodo. La tasa de crecimiento depende de la cantidad de energía solar capturada, y que a su vez depende de la habilidad del cultivo para captarla y de la intensidad de la misma. En definitiva, el rendimiento de un cultivo depende de la cantidad de asimilados que son asignados a las estructuras reproductivas para la formación de granos durante un periodo denominado crítico [57], y cada tipo de cultivo tiene un período crítico que abarca determinados estadios fenológicos y estos ocurren en momentos del año que varían de acuerdo a la zona de producción.

Los períodos críticos (variables en cada tipo de cultivo) son considerados como una ventana temporal en la cual se analizan detenidamente los datos del cultivo. Los índices de vegetación representan la condición ecofisiológica en la que se encuentra el cultivo y, por lo tanto, una buena condición del cultivo en el período crítico debería ser coincidente con un buen valor de rendimiento y viceversa. Teniendo en cuenta esto, se espera que el momento de mayor ajuste de los modelos en cada tipo de cultivo ocurra dentro del período crítico del mismo.

Con los datos obtenidos a partir de las curvas de crecimiento promedio de los cultivos y la determinación del momento del Green-Up en cada pixel, se calculó la fecha promedio del mismo, que denominamos "Green-Up Promedio", y se atribuyó como el momento de emergencia²⁸ del cultivo, que es un dato que por lo general no se encuentra disponible. Definido el momento de emergencia de los cultivos, los períodos críticos se ubicaron en función de la ocurrencia de los estadios fenológicos determinados por la cantidad de Días Desde la Emergencia (DDE) (Tabla 7).

	Tipo de Cultivo	Green-Up Promedio	Variación Green-Up	Ubicación del Período Crítico	Duración Período Crítico	Período Crítico
2008/2009	Soja 1°	19/11	+/- 12 días	A partir de los 80 DDE	60 días	28/01 – 29/03
	Soja 2°	14/12	+/- 10 días	A partir de los 60 DDE	50 días	02/02 – 24/03
	Maíz	05/11	+/- 15 días	Floración (+/- 20 días)	40 días	16/12 – 25/01
	Girasol	16/11	+/- 16 días	Floración (-30 días y +20 días)	50 días	08/12 – 27/01
2010/2011	Soja 1°	01/12	+/- 20 días	A partir de los 80 DDE	60 días	31/01 – 01/04
	Soja 2°	22/12	+/- 10 días	A partir de los 60 DDE	50 días	10/02 – 01/04
	Maíz	07/11	+/- 15 días	Floración (+/- 20 días)	40 días	21/12 – 30/01
	Girasol	12/11	+/- 11 días	Floración (-30 días y +20 días)	50 días	04/12 – 23/01

Tabla 7 – Ubicación de los Períodos Críticos por Cultivo y Campaña agrícola.

²⁸ Emergencia es un concepto agronómico que indica la aparición de la plántula por sobre la superficie del suelo.

3.2.3.1 Período crítico Soja 1°.

Los estados fenológicos reproductivos desde “plena fructificación” (R4) a “tamaño máximo del grano” (R6) corresponden al período crítico para la soja. En la zona agrícola bajo estudio coinciden con el fin de la etapa de “plateau”, con una duración de alrededor de 60 días (variabilidad dada por las diversas fechas de siembra y grupos de madurez) en los cuales se determinan 2 de los 4 componentes del rendimiento (N° de vainas por nudo y N° de granos por vaina). Se puede observar la ubicación del período crítico en soja de primera para la campaña agrícola 2008/2009 en la Figura 24 iniciando el 23/01, y un corrimiento de 8 días en el ciclo para la campaña 2010/2011 (Figura 25).

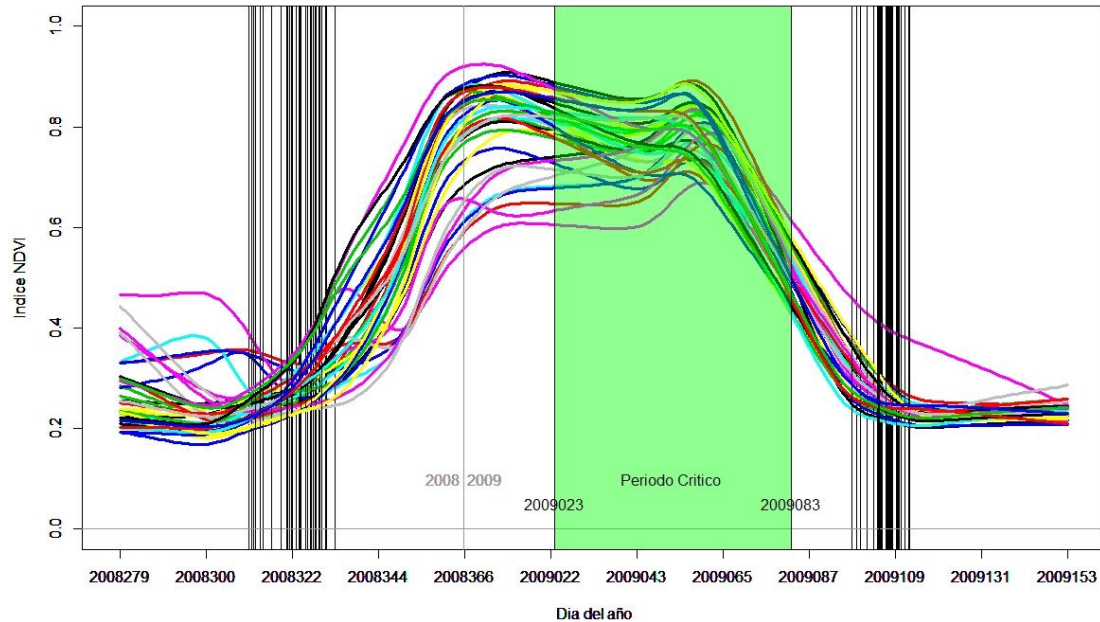


Figura 24 - Curvas de crecimiento estimadas para Soja de primera. Las líneas verticales corresponden al Green-Up y el Green-Down y el sector verde corresponde al período crítico.

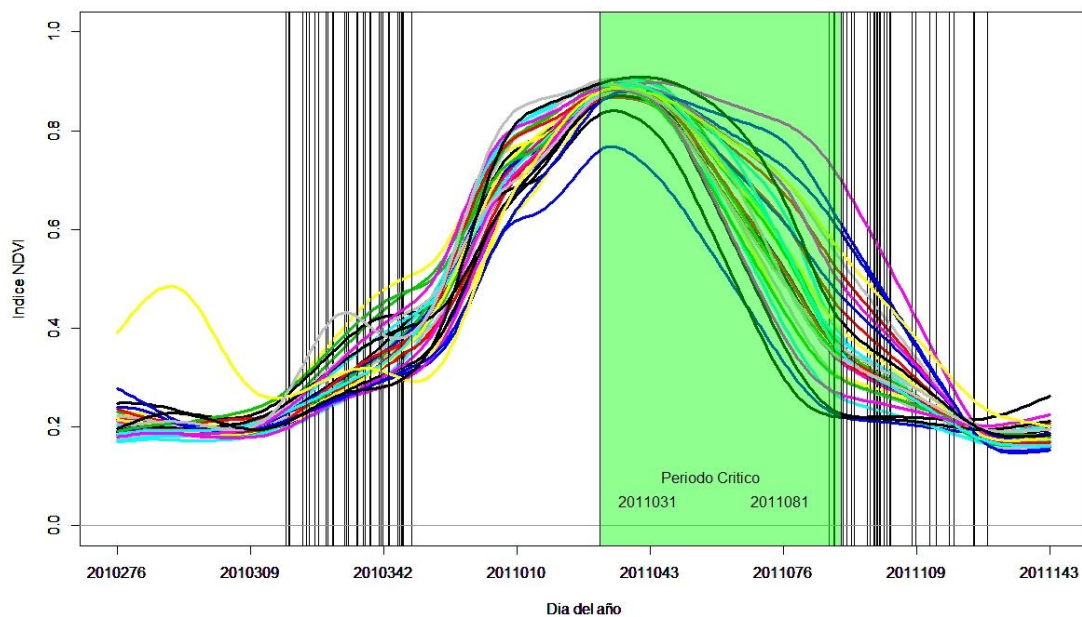


Figura 25 - Curvas de crecimiento estimadas para Soja de primera en la campaña 2010/2011. Las líneas verticales corresponden al Green-Up y el Green-Down y el sector verde corresponde al período crítico.

3.2.3.2 Período crítico Soja 2°.

El desarrollo fenológico del cultivo de Soja de segunda sufre un acortamiento de las fases de desarrollo, principalmente en la etapa vegetativa. Se denomina “de segunda” porque en la misma campaña agrícola se realizan dos cultivos en el mismo sitio geográfico, por lo tanto, se siembra la soja sobre el rastrojo de un cultivo de invierno recientemente cosechado. En la zona de estudio esto sucede a fines de noviembre comienzos de diciembre, por lo que, al contar con menor tiempo para el desarrollo, y a pesar del acortamiento de las fases, hay un corrimiento de aproximadamente tres semanas entre el período crítico de la Soja de primera y la Soja de segunda (Figura 26 y 27)

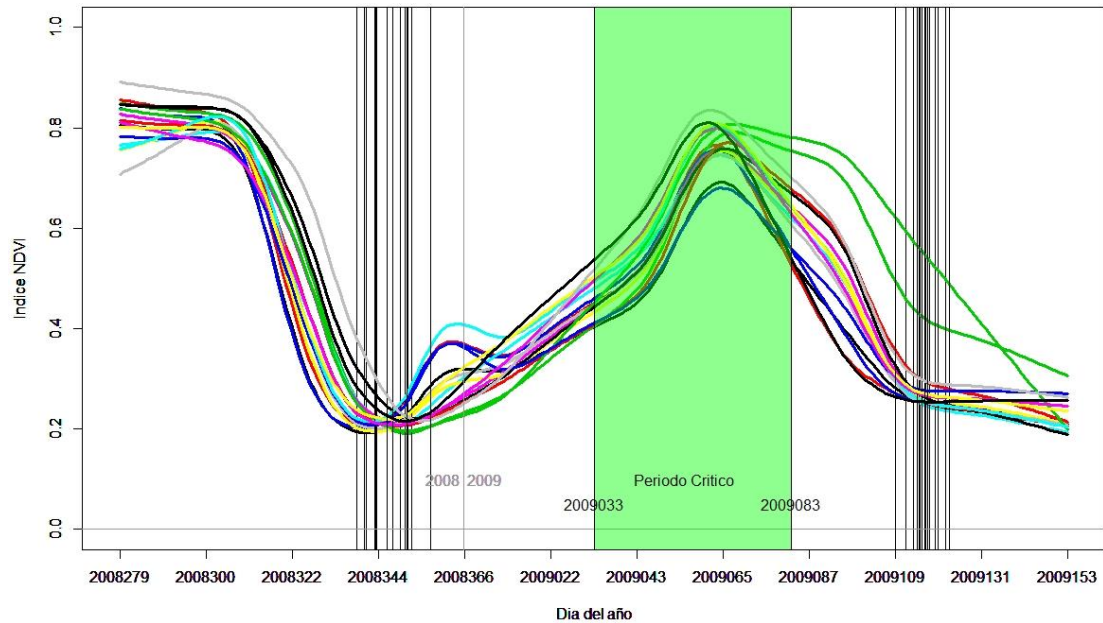


Figura 26 - Curvas de crecimiento estimadas para Soja de segunda en la campaña 2008/2009. Las líneas verticales corresponden al Green-Up y el Green-Down y el sector verde corresponde al período crítico.

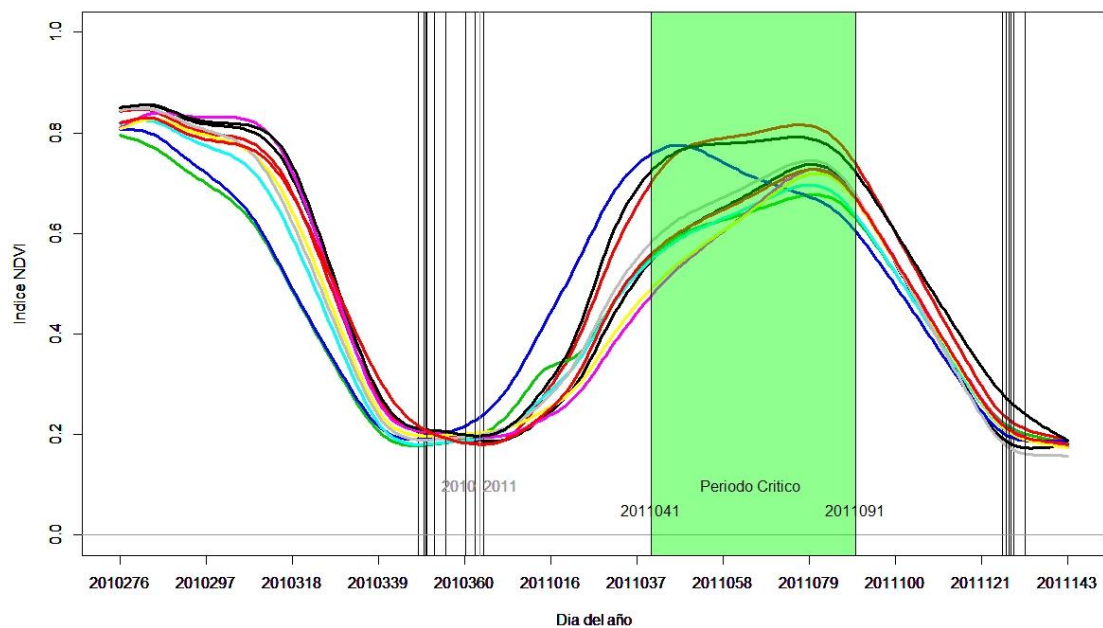


Figura 27 - Curvas de crecimiento estimadas para Soja de segunda en la campaña 2010/2011. Las líneas verticales corresponden al Green-Up y el Green-Down y el sector verde corresponde al período crítico.

3.2.3.3 Período crítico Maíz.

La siembra de este cultivo ocurre apenas iniciado el período libre de heladas, y esto sucede en la zona de estudio en la segunda quincena del mes de octubre. El período crítico del maíz se encuentra determinado por la floración, debido a que es un período muy sensible a las condiciones climáticas y es clave en la definición del rendimiento. La floración de los maíces tradicionales de ciclo largo se ubica a comienzos de enero y abarca un período de +/- 20 días. En este trabajo no tenemos en cuenta los maíces de ciclo corto que en los últimos años han tomado gran importancia por el “escape” que le hacen al efecto de la alta evapotranspiración que ocurre en la época de floración tradicional, y que, debido a esto, florecen tardíamente en febrero/marzo. Como se observa en las figuras, los lotes de maíz en situación de sequía (Figura 28) tuvieron un comportamiento más heterogéneo que en la situación hídrica normal (Figura 29) en la que el rango de fechas de Green-Up y Green-Down quedaron fuertemente definidas.

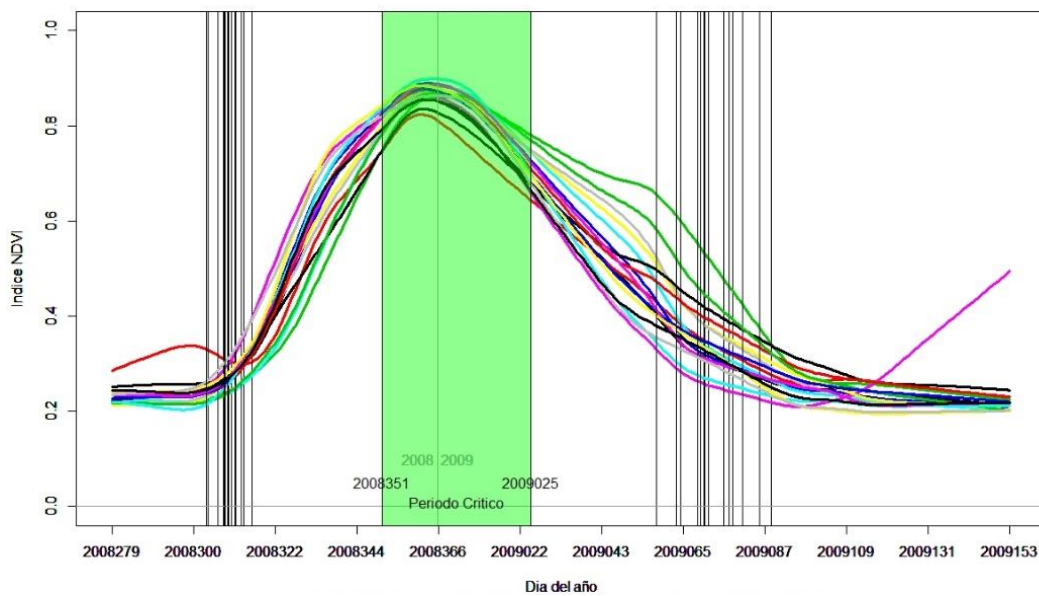


Figura 28 - Curvas de crecimiento estimadas para Maíz en la campaña 2008/2009. Las líneas verticales corresponden al Green-Up y el Green-Down y el sector verde corresponde al período crítico.

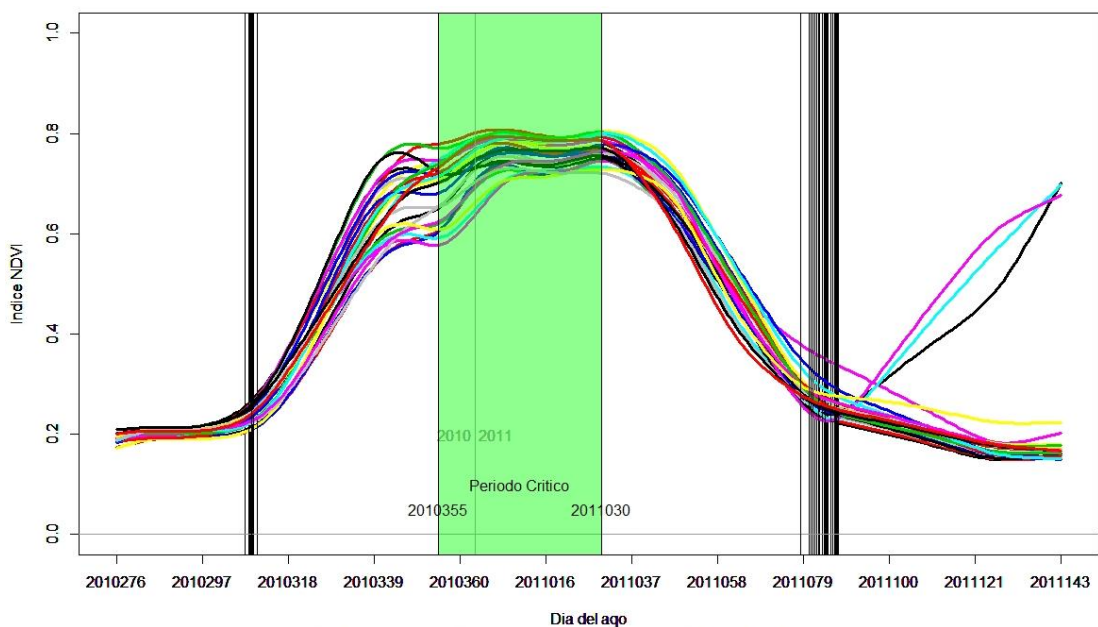


Figura 29 - Curvas de crecimiento estimadas para Maíz en la campaña 2010/2011. Las líneas verticales corresponden al Green-Up y el Green-Down y el sector verde corresponde al período crítico.

3.2.3.4 Período crítico Girasol.

La siembra de este cultivo, al igual que con los otros cultivos de verano, ocurre apenas iniciado el período libre de heladas (en la zona de estudio en la segunda quincena del mes de octubre). El período crítico del girasol se encuentra determinado por la floración al igual que el maíz, salvo que, en este caso, se estima 30 días antes y 20 días después de la floración. La floración del girasol en la zona de estudio (fase fenológica R5) se registra generalmente en la primera semana del mes de enero, por lo tanto, el período crítico estimado para la zona de estudio abarca por lo general los meses de diciembre y enero (Figuras 30 y 31).

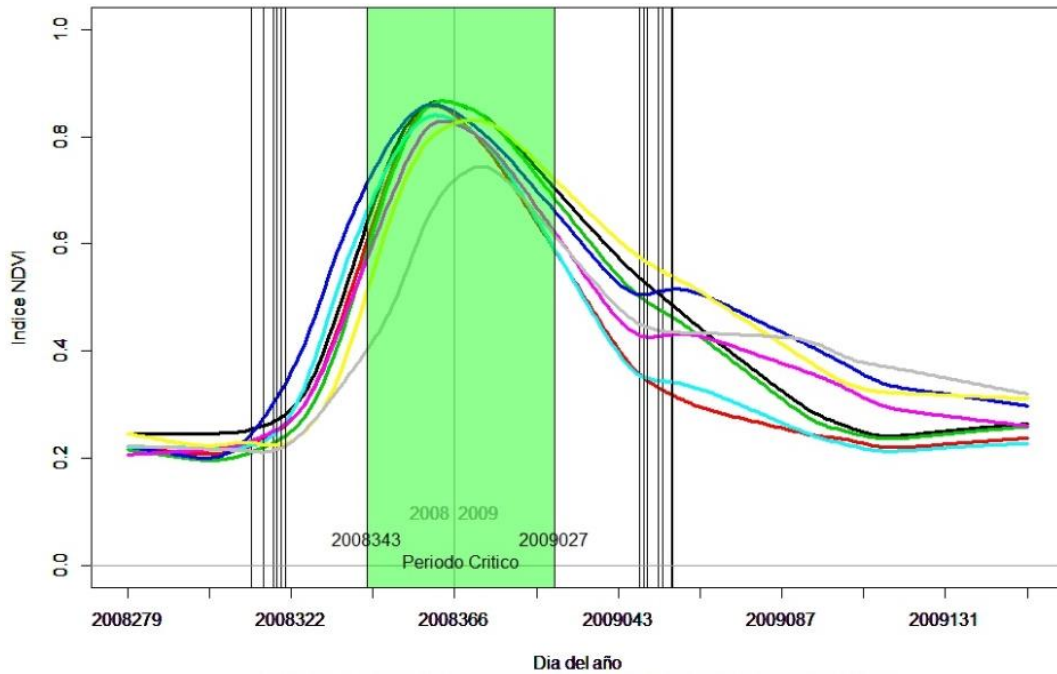


Figura 30 - Curvas de crecimiento estimadas para Girasol en la campaña 2008/2009. Las líneas verticales corresponden al Green-Up y el Green-Down y el sector verde corresponde al período crítico.

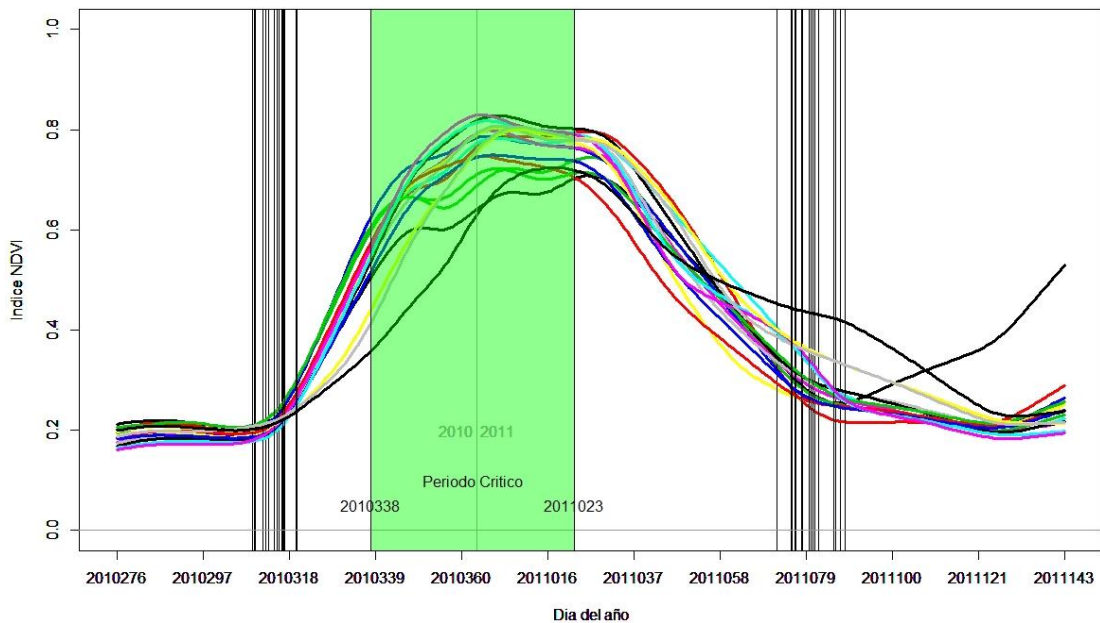


Figura 31 - Curvas de crecimiento estimadas para Girasol en la campaña 2010/2011. Las líneas verticales corresponden al Green-Up y el Green-Down y el sector verde corresponde al período crítico.

3.3 Modelo de Regresión Lineal.

Se generaron los modelos de regresión lineal con los datos previamente procesados de rendimiento e índices. La regresión lineal es un modelo matemático usado para aproximar la relación de dependencia entre una variable dependiente, y una (regresión lineal simple) o varias (regresión lineal múltiple) variables independientes. Este modelo de regresión lineal simple puede ser expresado como:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Donde, Y: variable dependiente (respuesta), X: variable independiente (explicativa), β_0 : ordenada al origen, β_1 : pendiente de la recta, ε : error residual (incertidumbre del modelo).

Las regresiones lineales requieren que se cumplan determinados supuestos para su correcta implementación, estos, se examinarán en los modelos propuestos.

- linealidad: el gráfico de dispersión indicaría una distribución lineal en la relación entre la variable dependiente y la independiente.
- normalidad: los errores del modelo tienen una distribución normal con media igual a cero.
- homocedasticidad: los errores del modelo tienen una varianza constante.
- independencia: los errores del modelo son variables aleatorias independientes.

Para cuantificar la bondad del ajuste de los modelos utilizaremos el coeficiente de determinación lineal R^2 , que es la proporción de variabilidad total presente explicada por el modelo. Este coeficiente toma valores entre 0 y 1, y cuanto más se aproxime a 1 mejor será el ajuste y, por lo tanto, mejores serán las predicciones del modelo.

$$R^2 = 1 - \frac{SCR}{SCT}$$

Donde, R^2 : Coeficiente de Determinación, SCR: Suma de Cuadrados de los Residuos, SCT: Suma de los Cuadrados Total.

Utilizando scripts desarrollados en R-Studio se vincularon, a través de regresiones lineales simples, el rendimiento del cultivo con el valor del índice de vegetación en cada día del ciclo del cultivo. El proceso para determinar el momento del año en el que se obtiene el mejor ajuste de la regresión lineal se realizó vinculando los valores diarios del índice de vegetación durante todo el ciclo del cultivo con el rendimiento, y guardando en una nueva variable (vector) el coeficiente de determinación R^2 de cada una de las regresiones. El día en el que se obtuvo el valor máximo de R^2 definió el momento óptimo para el modelo (Figura 32).

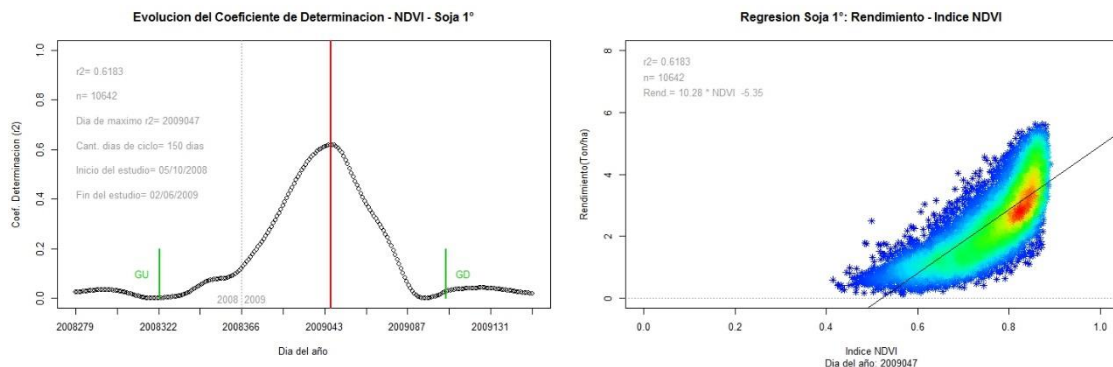


Figura 32 - Izq.: Ejemplo del cálculo del coeficiente de determinación R^2 durante el ciclo del cultivo de Soja de primera en la campaña 2008/2009. La línea vertical roja corresponde al día de mayor valor de R^2 . Der.: Modelo de regresión lineal que vincula los valores del índice en el día de mayor R^2 con el rendimiento en Ton/ha. Los colores de la nube de puntos definen la densidad (rojo mayor densidad, azul menor densidad).

Se observa una respuesta de tipo exponencial (no lineal) (Figura 32) en los datos en el momento de mayor ajuste del modelo de regresión lineal. Al analizar gráficamente los residuos

del modelo de regresión (Figura 33) se observa una distribución de residuos de tipo “trompeta” (respuesta no lineal), con acumulación de residuos hacia la derecha, por lo que no parece cumplirse el supuesto de homocedasticidad de la varianza. En cuanto a la normalidad de los residuos (Figura 34), se observa una leve dispersión en las colas del gráfico.

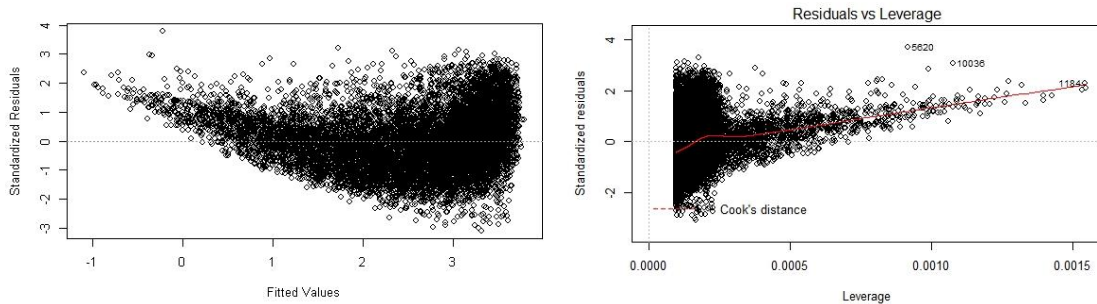


Figura 33 - Izq.: Gráfico de Residuos Estandarizados vs valores ajustados. Der.: Gráfico de Residuos Estandarizados vs Influencia

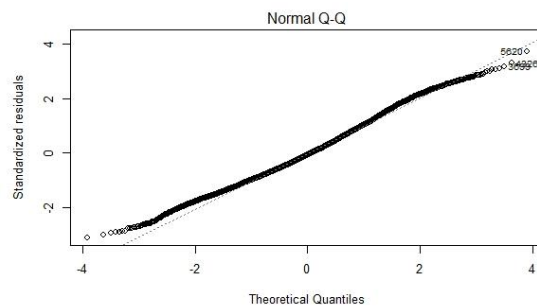


Figura 34 - Gráfico de QQ-Plot (normalidad de los residuos).

En base al problema observado, se decidió transformar la variable rendimiento (respuesta) al logaritmo natural con base “e” (arrojó el mejor ajuste en el modelo²⁹). Para evitar números negativos (por los valores decimales de la variable), los valores de rendimiento inicialmente obtenidos en ton/ha se multiplicaron por 10 obteniendo el valor en una variable de rendimiento denominada en el ámbito agropecuario como “quintales por hectárea” (qq/ha).

Con la variable transformada (logaritmo del rendimiento), la regresión analizada como ejemplo (Figura 32) obtiene un mayor valor del coeficiente de determinación R^2 (Figura 35), se obtiene una notable mejora en la linealidad de los datos en el gráfico de dispersión, la distribución de las varianzas de los residuos también mejora (Observar que a la izquierda de la Figura 36, el gráfico de residuos estandarizados vs valores ajustados no es más de tipo “trompeta”, aunque pasa a tomar una leve forma “cónica” con acumulación de datos hacia la derecha, en coincidencia con el sector de mayor densidad del gráfico de regresión de la Figura 35). El análisis de los residuos mejora notablemente al transformar la variable dependiente al logaritmo, sin embargo, igualmente parece no cumplirse gráficamente el supuesto de homocedasticidad de la varianza (Figura 36) y el de normalidad de los residuos (Figura 37) debido a que se incrementa el desvío en la cola negativa del gráfico QQ-Plot.

²⁹ Se probaron tres transformaciones en los datos: el logaritmo, la raíz cuadrada, y la inversa.

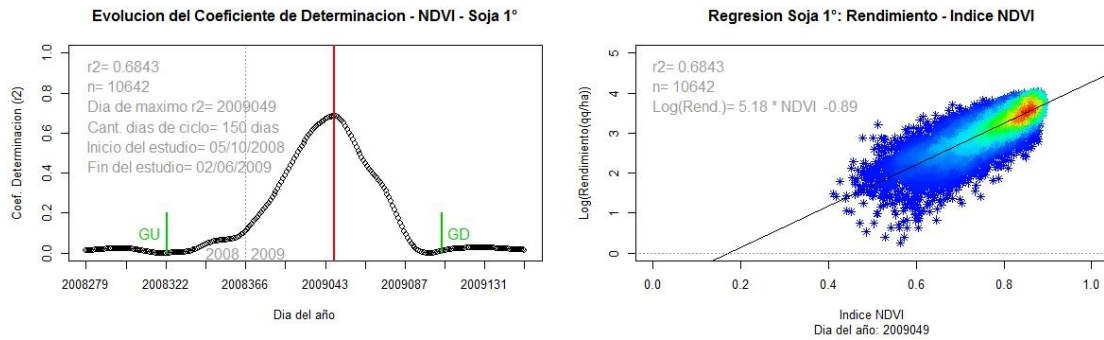


Figura 35 - Izq.: Ejemplo del cálculo del coeficiente de determinación R^2 durante el ciclo del cultivo de Soja de primera en la campaña 2008/2009. La línea vertical roja corresponde al día de mayor valor de R^2 . Der.: Modelo de regresión lineal que vincula los valores del índice del día de mayor R^2 con el rendimiento transformado a logaritmo y en unidades de qq/ha. Los colores de la nube de puntos definen la densidad (rojo mayor densidad, azul menor densidad)

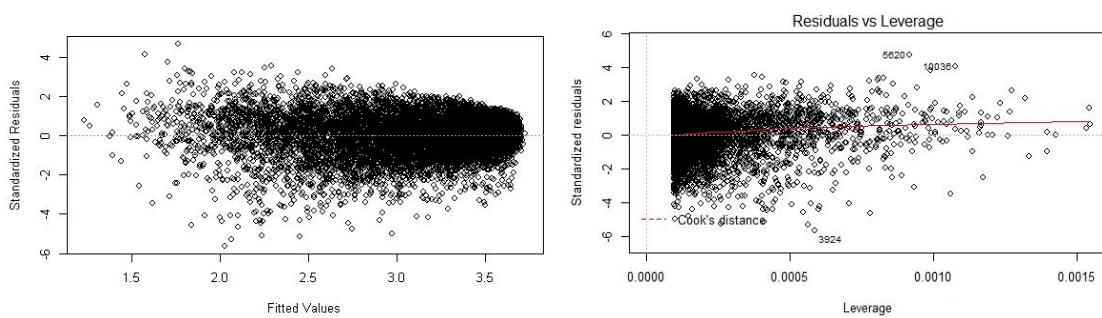


Figura 36 - Izq.: Grafico de Residuos Estandarizados vs valores ajustados. Der.: Grafico de Residuos Estandarizados vs Influencia

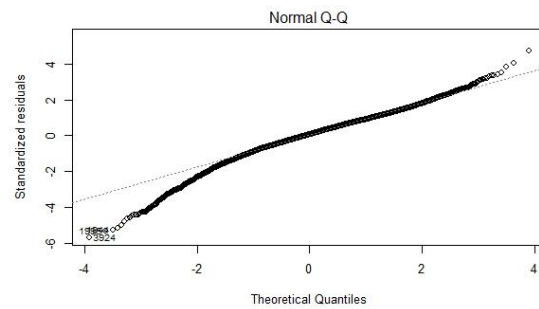


Figura 37 - Grafico de QQ-Plot (normalidad de los residuos).

3.3.1 Revisión del proceso de modelización.

La transformación al logaritmo natural en base “e” de los valores de rendimiento permitió obtener un mejor ajuste en los modelos, en general para los 4 tipos de cultivos, utilizando como indicador el valor de R^2 . Los residuos de los modelos lograron mejorar en algunos casos su distribución (analizando visualmente los gráficos), aunque analíticamente, y en general, no permitieron cumplir los supuestos de los modelos de regresión lineal.

Inicialmente surgió la propuesta de transformar únicamente la variable dependiente, o sea, el rendimiento. Esto se propuso debido a que la variable regresora, al cambiar de base de datos a medida que avanza temporalmente en el ciclo del cultivo (paso temporal diario), cambia la distribución (sólo en determinados momentos es de tipo “normal”) en los datos y resulta complejo determinar un tipo de transformación a realizar para todos los datos en su conjunto. Se analizó la evolución temporal de los datos de los índices de vegetación para determinar su comportamiento e intentar transformar adecuadamente los datos. Para esto, se realizó un estudio temporal de la distribución de los datos del índice de vegetación utilizando como ejemplo el NDVI en el cultivo de Soja de primera.

Los test de normalidad permiten determinar si una serie de datos tiene o no distribución normal. Se aplicó el test de Shapiro-Wilks³⁰ sobre los datos correspondientes a un lote de producción de soja (este tipo de test no permite trabajar con elevada cantidad de datos) para observar su comportamiento. En la Figura 38 se muestra la evolución del p-value del test de normalidad (Shapiro-Wilks) y se observan dos momentos en los que la distribución de los datos encuentra un comportamiento del tipo normal. Estos dos momentos son coincidentes con el inicio del crecimiento del cultivo (incremento de los valores del índice de vegetación) y el fin o senescencia del cultivo (disminución de los valores del índice de vegetación) respectivamente.

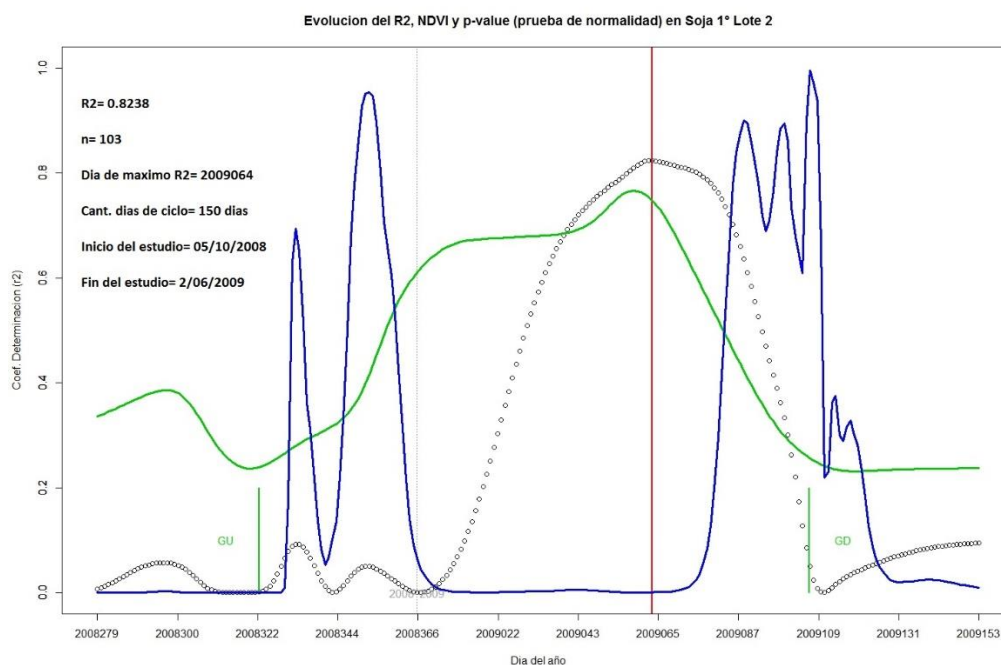


Figura 38 - Los datos corresponden al lote 2 de Soja de primera (2008/2009). Se graficó la evolución del p-valor (línea azul) derivado del análisis de normalidad de los datos de NDVI por el método de Shapiro-Wilks junto con el valor del NDVI en el ciclo del cultivo (línea verde) y la evolución de los valores del coeficiente de determinación R^2 (puntos negros) obtenido de las regresiones temporales con los datos de rendimiento. La línea vertical roja identifica el momento de máximo ajuste del modelo de regresión lineal. GU: “Green-Up” y GD: “Green-Down”.

³⁰ Test de Shapiro-Wilks: Plantea como hipótesis nula que los datos están distribuidos normalmente, y si el p-valor es menor a alfa (nivel de significancia considerado como 0,05) entonces la hipótesis nula es rechazada, y se concluye que los datos no cuentan con una distribución normal. Si el p-valor es mayor a alfa, no se rechaza la hipótesis nula y se concluye que los datos cuentan con distribución normal.

La diversidad de estadios de crecimiento y de senescencia (analizado en cantidad de píxeles) que ocurren en un determinado momento en la zona de estudio, son los que contribuyen a obtener una distribución de los datos del tipo normal. En el resto de los momentos del ciclo del cultivo lidera la homogeneidad de los datos y con lo cual los valores del índice de vegetación se encuentran concentrados (datos leptocurticos) hacia la izquierda (asimetría positiva) o hacia la derecha (asimetría negativa). Esto se relaciona con lo que ocurre en el campo, por ejemplo, un lote en barbecho (suelo desnudo sin cultivo) arroja una asimetría positiva (valores del índice de vegetación concentrados y cercanos a cero), en cambio cuando el cultivo se encuentra en su máximo vigor vegetal el canopeo se encuentra homogéneo arrojando una asimetría negativa (valores del índice de vegetación cercanos a uno).

Con el objetivo de intentar mejorar el análisis de los residuos de los modelos, y conociendo que la variable regresora mantiene asimetría negativa en los momentos de máximo valor de R^2 , se probó utilizar una transformación del tipo Box-Cox [58] [59], utilizando la función "powerTransform" (package car) la cual determina un valor de "lambda" (exponente) al cual se eleva la variable para lograr mejorar la distribución de los datos tendiente a la normalidad. Esta transformación se aplicó al par de datos (grupo de valores del índice para cada día de análisis y el rendimiento), conformando una nueva base de datos transformada. El resultado obtenido no fue superior a la aplicación de la transformación al logaritmo natural en base "e" de la variable dependiente, por lo que se desestimó este análisis.

Finalmente, el procedimiento de cambiar los valores iniciales de ton/ha a qq/ha y transformar la variable rendimiento al logaritmo natural en base "e", permitió obtener en general, el mayor ajuste de los modelos. Las pruebas se realizaron sobre todas las combinaciones de datos bajo estudio. Se realizó nuevamente el proceso de regresión sucesivo, día a día, determinando el día del año en el que se halla la mayor correlación entre el rendimiento (variable respuesta - dependiente) y los valores del Índice de Vegetación (variable regresora - independiente). Los parámetros extraídos del análisis de las curvas de crecimiento para realizar las regresiones lineales fueron los siguientes: Valores diarios de cada Índice de Vegetación, Fecha del Green-Up, Fecha del Green-Down, y Máximo valor alcanzado del Índice de Vegetación en el Período Crítico. Con esta información se probaron 4 métodos diferentes de vinculación de los datos.

- Todos los datos (información por píxel) vinculándose en forma conjunta por día con el valor de rendimiento:
 - a) Correlacionando todo el set de datos de cada índice de vegetación por tipo de cultivo desde el inicio del estudio y en forma consecutiva día por día hasta el fin del estudio. La cantidad de regresiones es igual a la cantidad de días de análisis 2008/2009= 240 días y 2010/2011= 232 días.
- Cada dato (información por píxel) vinculándose independientemente y en días variables con el valor de rendimiento:
 - b) Correlacionando el valor de cada píxel de cada índice de vegetación por tipo de cultivo desde el inicio del enverdecimiento ó Green-Up (fecha variable en cada píxel, porque cada píxel tiene su dato propio de Green-up). Se reordena la información de cada píxel en una nueva tabla en el cual el inicio de los datos queda determinado por el día del Green-Up, y se correlacionan los datos consecutivamente hasta la

cantidad de días promedio de ciclo de cultivo. La cantidad de regresiones es igual a la cantidad de días promedio de ciclo para cada tipo cultivo (por ejemplo, 140 días para la soja de primera). Incorpora la variable Green-Up al análisis de cada pixel.

- c) Correlacionando el valor de cada pixel de cada índice de vegetación por tipo de cultivo según el porcentaje de avance en la duración del ciclo (fecha variable en cada pixel, porque cada pixel tiene su dato de Green-Up y Green-Down). Se reordena la información de cada pixel en una nueva tabla en el cual el inicio de los datos queda determinado por el día del Green-Up y el fin del análisis por el valor del Green-Down. El tiempo transcurrido en días, desde Green-Up y Green-Down se toma como el 100%. Luego, se calcula cada día para realizar la regresión de los datos en función de un avance de un 1% en el ciclo del cultivo. La cantidad de regresiones es igual a 100. Incorpora la variable Green-Up y Green-Down al análisis de cada pixel.
- d) Valor máximo alcanzado por el índice de vegetación durante el Período Crítico de cada tipo de cultivo. En este caso, NO se identifica el valor máximo de R^2 en algún momento del ciclo del cultivo, sino que se vinculan directamente los valores máximos del índice de vegetación obtenidos durante el período crítico del cultivo directamente con el rendimiento.

En los tres primeros casos (a, b, c), el coeficiente de determinación R^2 fue el indicador del ajuste del modelo (Figura 39, sector izquierdo) y el momento en el cual éste ocurre es el día del año con mayor grado de ajuste (mayor valor de R^2). Para el caso del valor máximo del índice de vegetación, no se determinó el mejor ajuste del R^2 , sino que se identificaron los máximos valores de cada índice dentro del período crítico de cada cultivo.

En las siguientes páginas se graficaron en forma de ejemplo los datos de Soja de primera en 2008/2009 y 2010/2011. En el sector izquierdo de la Figura 39 se presenta la evolución temporal del coeficiente R^2 , y a la derecha se observa la función de regresión.

Los 4 métodos de vinculación de los datos arrojaron valores similares de ajuste ($a=0.665$, $b=0.644$, $c=0.663$, $d=0.652$). Por la simplicidad de procesar los datos, se selecciona como método el "a". Para el ejemplo de la Soja de primera por el método "a" se evidencia en el gráfico de la izquierda el momento de máximo valor de ajuste ($>R^2$) en la campaña 2008/2009 el día 2009049 (18/02), o 86 días luego del Green-Up promedio. La fecha de Green-Up promedio que surge del análisis para la soja de primera es fines de noviembre, por lo que 86 días después (2^{da} quincena febrero), el cultivo se encontraba finalizando su etapa fenológica de R6, en la cual define el tamaño de los granos (a mitad del período crítico para la soja).

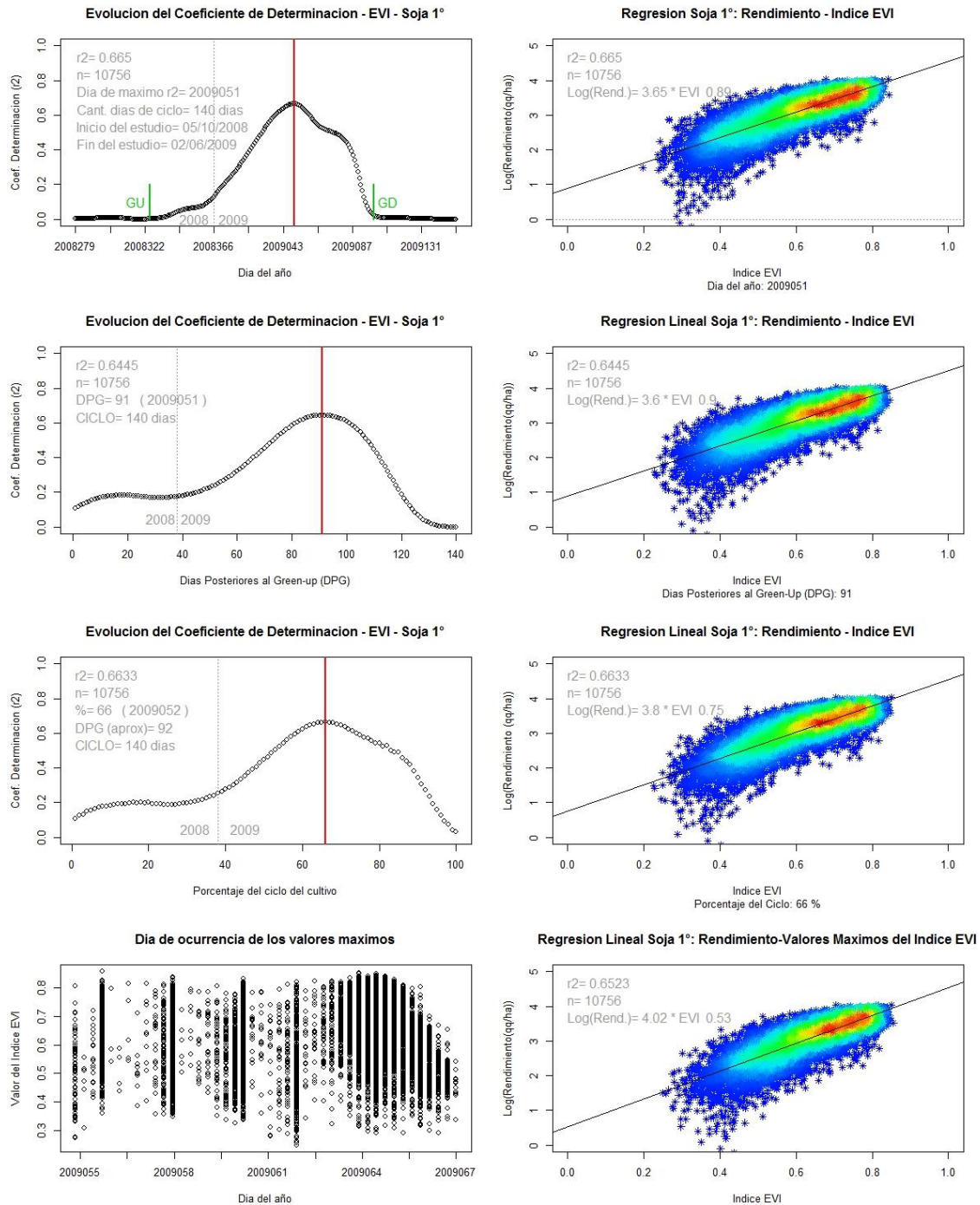


Figura 39 - Visualización de los 4 métodos utilizados para vincular los datos (campana 2008/2009) utilizando como ejemplo el índice de vegetación EVI en Soja de primera. Fila 1: "a", Fila 2: "b", Fila 3: "c", Fila 4: "d".

Continuando con el ejemplo de la Soja de primera, en la campaña agrícola de ciclo hídrico normal/húmedo (2010/2011), el fenómeno observado es significativamente diferente (Figura 40). Es posible que haya existido un cambio en la elección de los grupos de madurez del cultivo de soja, ya que se observa un ciclo más corto que encuentra su máximo valor de R^2 en la primera quincena de enero (13/01). El método "a" logra el mejor ajuste de los datos (a=0.535, b=0.330, c=0.279, d=0.252).

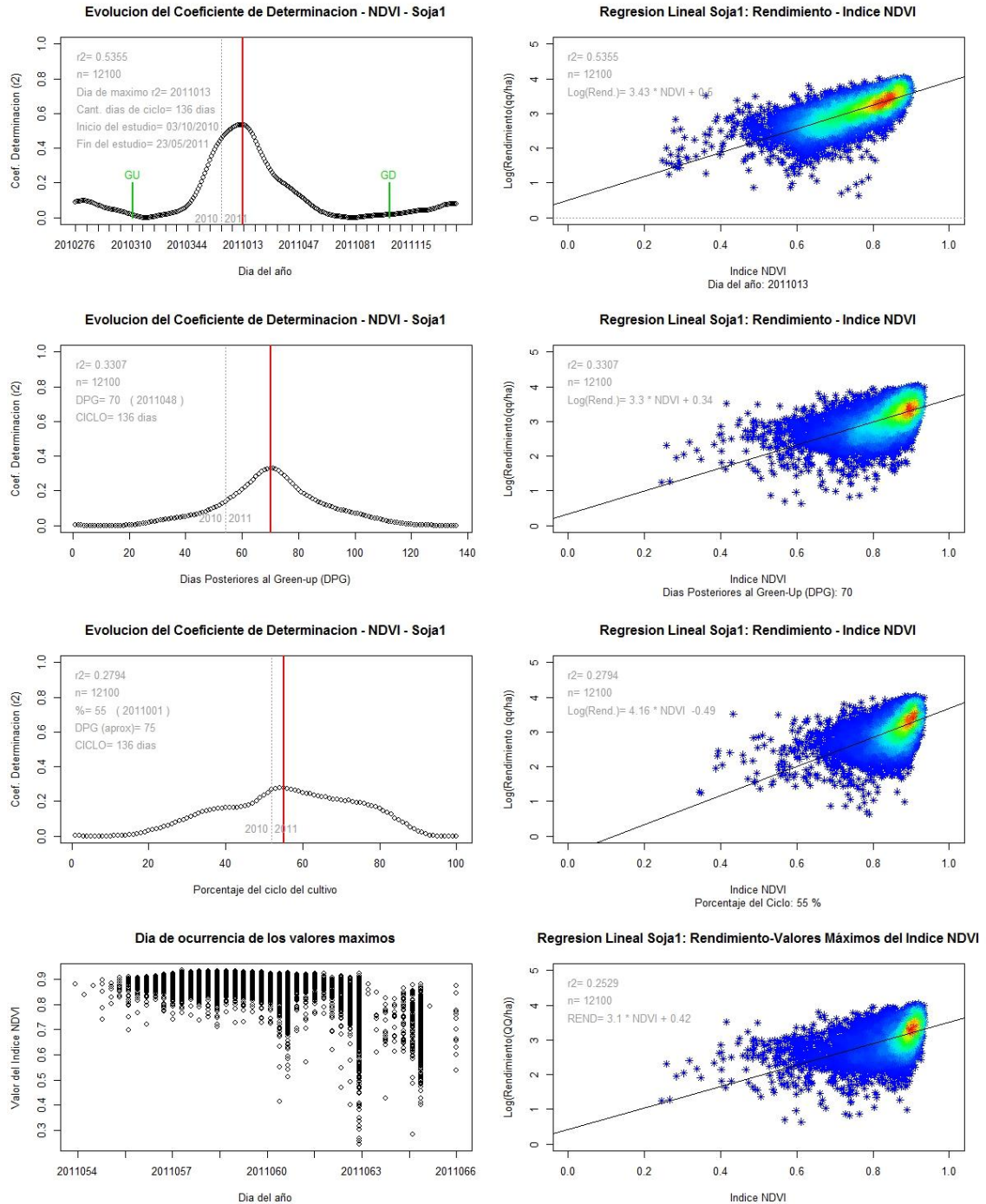


Figura 40 - Visualización de los 4 métodos utilizados para vincular los datos (campana 2010/2011) utilizando como ejemplo el índice de vegetación NDVI en Soja de primera.

3.4 Determinación de los modelos con mejor ajuste.

Durante el proceso de combinación de las variables analizadas, se generaron 128 posibles modelos. Esto se debe a la combinación de las siguientes variables: 4 cultivos, 4 índices de vegetación, 4 formas de vincular la información, y 2 campañas agrícolas.

Las regresiones lineales se realizaron vinculando el rendimiento (variable respuesta) con el dato de rendimiento transformado al logaritmo natural con base "e", por lo tanto, se debe re-transformar el resultado para obtener el rendimiento en quintales por hectárea (qq/ha).

En la Tabla 8 se comparan los resultados de las combinaciones de las regresiones de los 4 tipos de cultivo, las 4 maneras de combinar los datos, los 4 índices de vegetación utilizados y las 2 campañas agrícolas. El objetivo fue obtener un panorama general del comportamiento de todos los modelos utilizando como primer indicador de ajuste el valor del coeficiente de determinación R^2 . Los modelos seleccionados para cada tipo de cultivo y período hídrico fueron elegidos a partir del máximo valor del coeficiente de determinación R^2 (indicados con color amarillo en la Tabla 8). Debido a que la obtención de los métodos “b”, “c” y “d” tienen una mayor complejidad de cálculo, y considerando que las diferencias obtenidas entre los modelos con respecto al método “a” en algunos casos son mínimas, es que se definió dar prioridad al método “a” en los casos que fuera posible.

CULTIVO	MÉTODO	Campaña 2008/2009				Campaña 2010/2011					
		n	NDVI	EVI	MSAVI	NDMI	n	NDVI	EVI	MSAVI	NDMI
SOJA 1	a	10977	0,6843	0,665	0,6709	0,6827	13181	0,5287	0,5141	0,5182	0,4595
	b		0,5888	0,6349	0,6367	0,6689		0,3307	0,3479	0,3517	0,279
	c		0,6289	0,6628	0,6338	0,6863		0,2788	0,3336	0,3365	0,2944
	d		0,5928	0,6504	0,6579	0,683		0,255	0,2988	0,3099	0,2662
SOJA 2	a	4943	0,4855	0,4718	0,4871	0,5092	2866	0,4154	0,3047	0,3096	0,329
	b		0,5215	0,4909	0,4697	0,4837		0,4151	0,3052	0,313	0,3211
	c		0,5322	0,4749	0,489	0,5056		0,3399	0,2204	0,2205	0,2505
	d		0,4733	0,4523	0,4645	0,4721		0,1847	0,1308	0,1286	0,143
MAÍZ	a	6100	0,4684	0,3513	0,3836	0,4044	7233	0,4102	0,2728	0,2715	0,2526
	b		0,4514	0,3641	0,3921	0,4057		0,4211	0,2719	0,244	0,2529
	c		0,4316	0,3258	0,3439	0,3404		0,4121	0,2678	0,264	0,2466
	d		0,3286	0,3178	0,3439	0,3184		0,3436	0,2025	0,1996	0,1729
GIRASOL	a	2744	0,4097	0,425	0,4359	0,4749	4698	0,5826	0,5233	0,5205	0,5565
	b		0,4095	0,4034	0,4149	0,4634		0,4497	0,4604	0,4523	0,4724
	c		0,3965	0,4018	0,4071	0,4602		0,437	0,4197	0,4082	0,3491
	d		0,3816	0,3975	0,3979	0,4566		0,4443	0,4069	0,3954	0,3505

Tabla 8 - Valores del Coeficiente de determinación R^2 obtenidos al combinar: tipo de cultivo, método de análisis de datos, índice de vegetación y campaña agrícola.

De la Tabla 8 se puede concluir en forma general que de los diferentes métodos de combinar la información (a, b, c, d), el más sencillo: método “a”, es el que mejores resultados obtuvo en los diferentes cultivos. La propuesta de modelar teniendo en cuenta el momento de Green-Up en cada pixel o la proporción de avance dentro del ciclo del cultivo no aportaron las diferencias significativas esperadas. Sólo para el cultivo de Soja de segunda en la campaña 2008/2009 el método “b” obtuvo un mejor desempeño.

En cuanto a los índices de vegetación, se puede observar un mayor ajuste de los datos para el NDVI en ambas campañas agrícolas, con excepción del NDMI (índice de humedad de la vegetación) que logró un mejor desempeño para el cultivo de Girasol en la campaña seca.

En el caso de los cultivos, el que mayor relación encuentra con los valores de rendimiento es la Soja de primera, que puede deberse a la gran “plasticidad”³¹ del cultivo.

³¹ Se denomina “plasticidad de un cultivo” a la capacidad de superar posibles condiciones adversas mediante la acción de diferentes mecanismos internos.

Además, es notable la mayor cantidad de datos disponibles para este cultivo. En el caso de la Soja de segunda, los menores valores del R^2 podrían deberse a las condiciones del cultivo antecesor, que por lo general tienen gran incidencia sobre el cultivo de segunda ocupación y no se encuentra ese efecto analizado en estos modelos. El Maíz es uno de los cultivos más difíciles de modelar a través de datos satelitales debido a que es muy sensible al estrés hídrico en su período crítico, pudiendo tener plantas de muy buen vigor vegetal, pero sin granos cuajados en la espiga. Para ambas campañas agrícolas los valores del coeficiente de determinación R^2 para el Maíz son menores a 0,5. En el caso del Girasol, se obtienen valores similares al Maíz, aunque en la campaña de ciclo hídrico normal/húmedo se logran mejores resultados.

A modo de resumen de la Tabla 8 los modelos seleccionados se presentan en la Tabla 9:

	Período seco	Período normal/húmedo
Soja de primera	NDVI método "a"	NDVI método "a"
Soja de segunda	NDVI método "b"	NDVI método "a"
Maíz	NDVI método "a"	NDVI método "a"
Girasol	NDMI método "a"	NDVI método "a"

Tabla 9 - Modelos finalmente seleccionados.

Al graficar las regresiones (Figuras 39 y 40) se observan datos muy alejados del ajuste de la recta de regresión lineal, que pueden considerarse outliers. Estos datos pueden desestimarse del análisis para evitar su influencia sobre la definición de la recta de regresión. Para evitar la posible influencia de estos outliers, se propuso realizar las regresiones a través de algún método de estadística robusta. La regresión robusta se puede usar en cualquier situación en la que use la regresión por mínimos cuadrados y es una solución alternativa entre excluir datos de alto apalancamiento (alto leverage) y tomar todos los datos para tratarlos por igual en la regresión por mínimos cuadrados. La regresión robusta pondera las observaciones de forma diferente en función de su comportamiento, minimizando el efecto de aquellas observaciones que cambian de manera significativa la estimación de los coeficientes de regresión.

Los cambios generados a partir de la utilización de las regresiones robustas en el coeficiente de determinación R^2 no fueron significativos (debido posiblemente a la alta cantidad de datos utilizados en este trabajo), pero consideramos conveniente aplicarlo para robustecer los modelos a desarrollar y evitar realizar la extracción de outliers de cada set de datos. Por lo tanto, se decidió aplicar modelos lineales robustos del tipo "MM" estimadores, utilizando la librería "robustbase" del software R-studio y su función "lmrob" que estima un valor de R^2 a pesar de no realizar el cálculo por mínimos cuadrados.

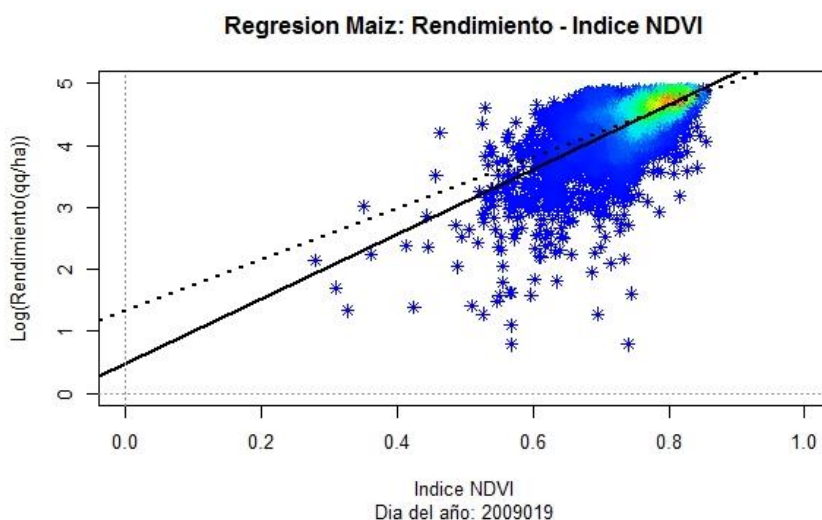


Figura 41 – Modelo de Regresión Lineal (recta llena) y Modelo de Regresión Lineal Robusta (recta punteada) para datos de NDVI vs Rendimiento en Maíz.

En la Figura 41 se puede observar el comportamiento de los modelos lineales sobre la base de datos del cultivo de maíz (índice NDVI vs Rendimiento). La línea llena corresponde al modelo lineal (función “lm”) y la línea punteada corresponde al modelo lineal robusto (función “lmrob”). Se evidencia como el modelo robusto desestima los outliers (datos que se desprenden de la nube de puntos) presentes y que alcanzan los 32 datos según los resultados de la regresión.

Como se observa en los siguientes resúmenes de los dos tipos de regresión (extracción del programa R) los coeficientes e indicadores de la regresión cambian al pasar del modelo lineal al modelo lineal robusto de la siguiente manera: $b_0= 0.48$ vs 1.33 ; $b_1= 5.20$ vs 4.12 ; $RSE= 0.33$ vs 0.19 ; $R^2= 0.4636$ vs 0.4793 .

MODELO LINEAL (lm):

Call:

```
lm(formula = log(rendimiento) ~ NDVI(día de max R²)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.5535	-0.1154	0.0290	0.1680	1.3611

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.48898	0.05455	8.964	<2e-16 ***
NDVI(día de max R²)	5.20878	0.07175	72.592	<2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: **0.3309** on 6098 degrees of freedom

Multiple R-squared: **0.4636**, Adjusted R-squared: **0.4635**

F-statistic: 5270 on 1 and 6098 DF, p-value: < 2.2e-16

MODELO LINEAL ROBUSTO (lmrob):

Call:

```
lmrob(formula = log(rendimiento) ~ NDVI(día de max R²)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.617564	-0.160608	0.006538	0.126214	1.095479

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.33415	0.04455	29.95	<2e-16 ***
NDVI(día de max R²)	4.12165	0.05744	71.76	<2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Robust residual standard error: **0.1995**

Multiple R-squared: **0.4793**, Adjusted R-squared: **0.4792**

Convergence in 28 IRWLS iterations

Robustness weights:

32 observations c(351,1904,1911,2169,2195,2370,2546,2596,2615,2798,2981,3758,3854,3937,3942,3943,3982,4011,4012,4045,4046,4090,4092,4159,4315,4567,4767,4943,5021,5518,5591,5705) are outliers with |weight| = 0 (< 1.6e-05);

3896 weights are ~1. The remaining 2172 ones are summarized as

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0000238	0.4736000	0.8426000	0.7093000	0.9643000	0.9990000

3.4.1 Modelos seleccionados para períodos de escases hídrica (2008/2009):

Los modelos seleccionados surgen de la comparación realizada en la Tabla 8, y sólo se agregó la aplicación del método de regresión robusta.

3.4.1.1 Modelo seleccionado Soja 1:

Utilizando la base de datos de modelación (datos guardados al inicio del estudio), se procesaron los datos del índice NDVI para la Soja de primera obteniendo valores diarios y se obtuvo la función lineal en el día del año que se logra el coeficiente de determinación R^2 más elevado. Para este caso, el modelo de mejor ajuste resultó ser el "a", que vincula los valores por día del año con el rendimiento sin ajustar por Green-Up. Se utilizaron para el desarrollo del modelo un total de 10642 datos de pixeles de Soja de primera.

En el siguiente gráfico de la izquierda (Figura 42) se incluyen de manera conjunta, la evolución del coeficiente de determinación R^2 a lo largo del tiempo del estudio (puntos negros) y el valor medio del índice de vegetación NDVI (línea azul). Como puede observarse, el máximo valor de R^2 se obtiene el día juliano 2009048, y en momentos en los que ocurre un valle dentro del "plateau". La fecha corresponde al 17 de febrero, momento en que en la zona de estudio el ciclo del cultivo de soja de primera se encuentra en etapas avanzadas. En el mismo gráfico se agrega el valor medio de Green-Up y de Green-Down indicados en color verde. En el gráfico de la derecha se visualiza la regresión de la variable NDVI (regresora - independiente) que varía de 0 a 1 sobre el eje de coordenadas X, y la variable RENDIMIENTO transformada al logaritmo natural (respuesta - dependiente) sobre el eje Y. Los datos fueron coloreados de acuerdo a la densidad, siendo el color rojo la zona de mayor densidad y la azul la zona de menor densidad de puntos.

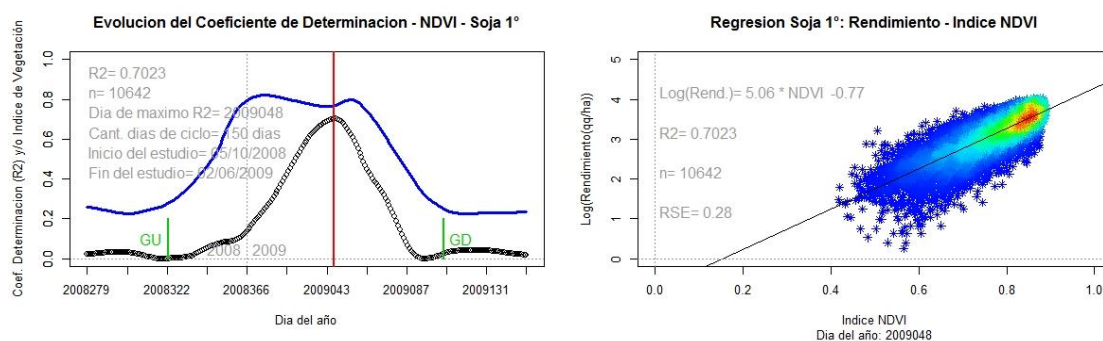


Figura 42 - Izq.: Ejemplo del cálculo del coeficiente de determinación R^2 durante el ciclo del cultivo de Soja de primera en la campaña 2008/2009. La línea vertical roja corresponde al día de mayor valor de R^2 . La línea azul representa el desarrollo medio del cultivo (valores medios de NDVI). Der.: Modelo de regresión lineal que vincula los valores del índice NDVI del día de mayor R^2 con el rendimiento transformado a logaritmo natural. Los colores de la nube de puntos definen la densidad (rojo mayor densidad, azul menor densidad).

Modelo "Sequía-Soja1":

$$YS1 = \frac{e^{(5.06 * IVS1)}}{2.159}$$

Donde:

YS1: Valor de rendimiento estimado por el modelo para la Soja de primera en períodos de escases hídrica.

IVS1: Valor del índice de vegetación NDVI para datos de Soja de primera a 86 DPG-Promedio que se corresponde con el día juliano 48 en el modelo.

En el análisis de la regresión, la significancia de la pendiente indica la existencia de relación lineal entre las variables (se rechaza la hipótesis nula de pendiente=0). Error Estándar Residual del modelo es de 0.28 (Anexo II). En cuanto a los residuos del modelo, el supuesto de

homocedasticidad de la varianza de los residuos gráficamente parece no cumplirse porque se observa un leve comportamiento cónico que evidencia una mayor varianza inicial de los residuos, y analizado con el test Breusch-Pagan, analíticamente no logra cumplirse. En cuanto a la normalidad de los residuos, la mayor densidad de datos se mantiene dentro de un rango fijo (Figura 43), pero observando el gráfico de QQ-plot y el histograma de residuos, se evidencia el alejamiento de la cola de residuos negativos (Figura 44), y analizado con el test de Kolmogorov-Smirnov modificado por lilliefors, analíticamente no se cumple la normalidad (Anexo II).

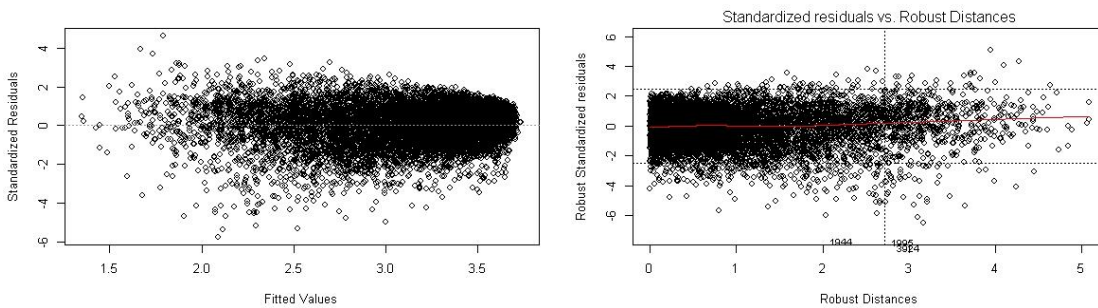


Figura 43 - Izq.: Gráfico de Residuos Estandarizados vs Valores Ajustados. Der.: Gráfico de Residuos Estandarizados vs Distancias Robustas.

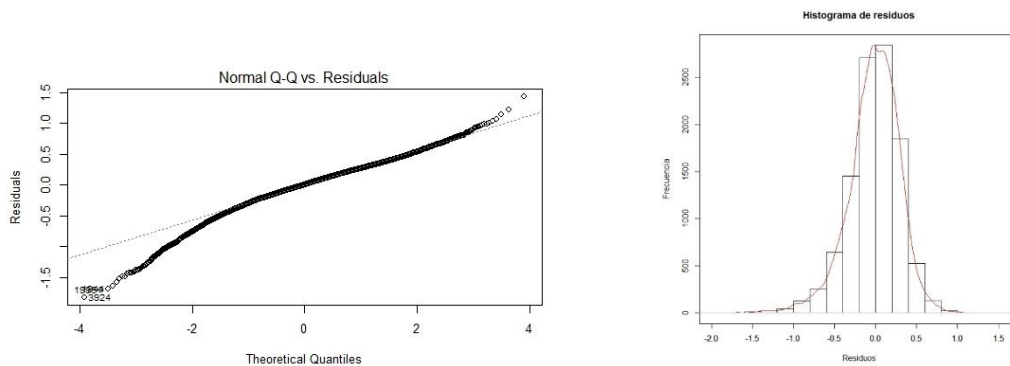


Figura 44 - Izq.: Gráfico de QQ-Plot (normalidad de los residuos). Der.: Histograma de residuos.

3.4.1.2 Modelo seleccionado Soja 2:

Al igual que en la Soja de primera, se procesaron los datos del índice NDVI para la Soja de segunda obteniendo valores diarios y se obtuvo la función lineal en el día del año que se logra el coeficiente de determinación R^2 más elevado. Para este caso, el modelo de mejor ajuste resultó ser el tipo “b”, que ordena el inicio del crecimiento del cultivo en base al Green-Up. Se utilizaron para el desarrollo del modelo un total de 4913 datos de pixeles de Soja de segunda.

En el siguiente gráfico de la izquierda (Figura 45) se incluyen de manera conjunta, la evolución del coeficiente de determinación R^2 desde el día promedio de Green-Up hasta el día promedio del largo del ciclo del cultivo, y el valor medio del índice de vegetación NDVI (línea azul) en ese mismo intervalo de tiempo. Como puede observarse, el máximo valor de R^2 se obtiene en el DPG 82 que se corresponde al día juliano 2009063. La fecha corresponde al 4 de marzo, momento en que en la zona de estudio el ciclo del cultivo de soja de segunda se encuentra en una etapa avanzada (formación del grano). En el gráfico de la derecha se visualiza la regresión de la variable NDVI (regresora - independiente) que varía de 0 a 1 sobre el eje de coordenadas X, y la variable RENDIMIENTO transformada al logaritmo natural (respuesta -

dependiente) sobre el eje Y. Los datos fueron coloreados de acuerdo a la densidad, siendo el color rojo la zona de mayor densidad y la azul la zona de menor densidad de puntos.

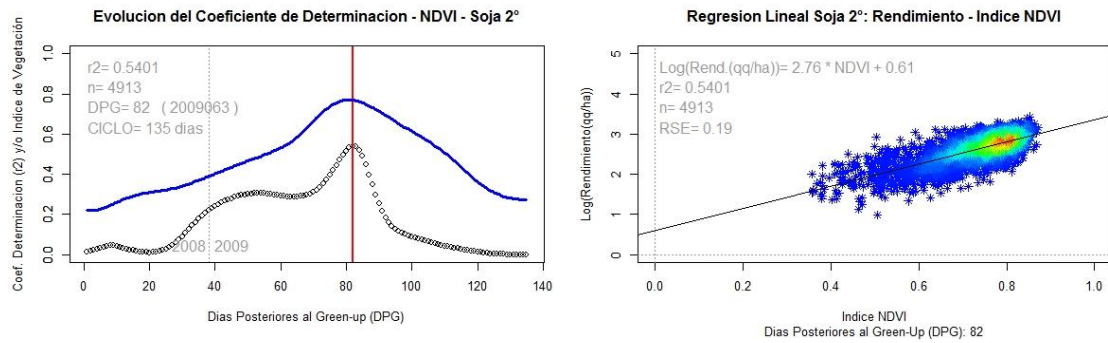


Figura 45 - Izq.: Ejemplo del cálculo del coeficiente de determinación R^2 durante el ciclo del cultivo de Soja de segunda en la campaña 2008/2009. La línea vertical roja corresponde al día de mayor valor de R^2 . La línea azul representa el desarrollo medio del cultivo (valores medios de NDVI). Der.: Modelo de regresión lineal que vincula los valores del índice NDVI en el día posterior al Green-Up de mayor R^2 con el rendimiento transformado al logaritmo natural (qq/ha). Los colores de la nube de puntos definen la densidad (rojo mayor densidad, azul menor densidad)

Modelo “Sequía-Soja2”:

$$YS2 = 1.84 e^{(2.76 * IVS2)}$$

Donde:

- YS2:** Valor de rendimiento estimado por el modelo para la Soja de segunda en períodos de escasez hídrica.
- IVS2:** Valor del índice de vegetación NDVI sobre datos de Soja de segunda, 82 días posteriores al Green-Up.

En el análisis de la regresión, la significancia de la pendiente indica la existencia de relación lineal entre las variables (se rechaza la hipótesis nula de pendiente=0). Error Estándar Residual del modelo es de 0.19 (Anexo II). En cuanto a los residuos del modelo, el supuesto de homocedasticidad de la varianza de los residuos gráficamente parece no cumplirse porque se observa un leve comportamiento cónico que evidencia una mayor varianza inicial de los residuos, y analizado con el test Breusch-Pagan, analíticamente no logra cumplirse. En cuanto a la normalidad de los residuos, la mayor densidad de datos se mantiene dentro de un rango fijo (Figura 46), observando el gráfico de QQ-plot y el histograma de residuos, se evidencia un leve alejamiento de la cola de residuos negativos, pero la distribución parece normal (Figura 47), y analizado con el test de Kolmogorov-Smirnov modificado por lilliefors, analíticamente no cumple la normalidad (Anexo II).

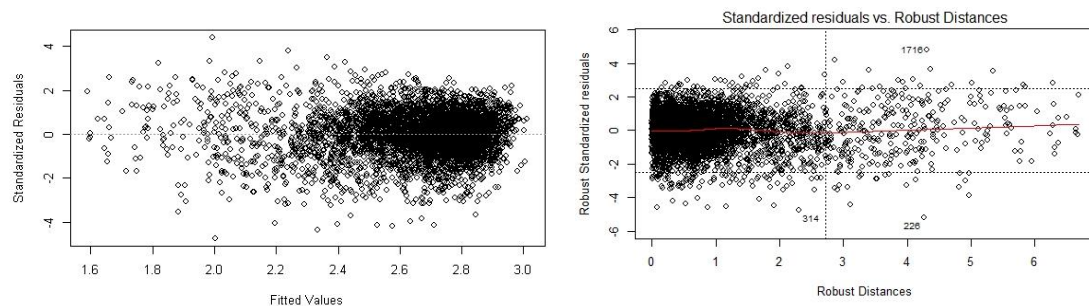


Figura 46 - Izq.: Gráfico de Residuos Estandarizados vs Valores Ajustados. Der.: Gráfico de Residuos Estandarizados vs Distancias Robustas.

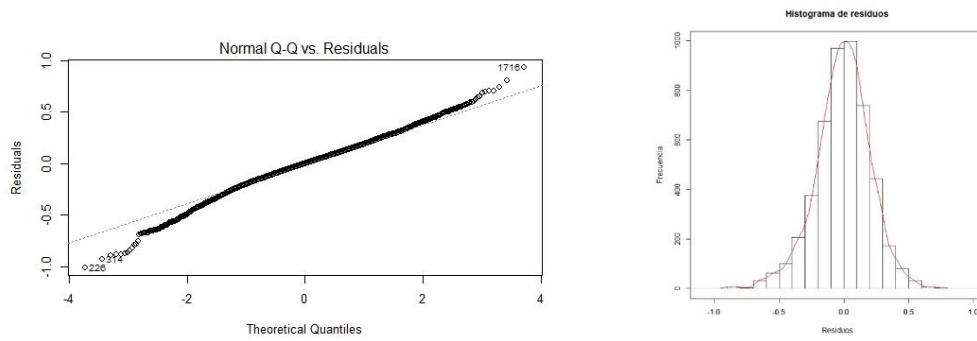


Figura 47 - Izq.: Gráfico de QQ-Plot (normalidad de los residuos). Der.: Histograma de residuos.

3.4.1.3 Modelo seleccionado Maíz:

Utilizando la base de datos de modelación (datos guardados al inicio del estudio), se procesaron los datos del índice NDVI para el maíz obteniendo valores diarios y se obtuvo la función lineal en el día del año que se logra el coeficiente de determinación R^2 más elevado. Para este caso, el modelo de mejor ajuste resultó ser el “a”, que vincula los valores por día del año con el rendimiento sin ajustar por Green-Up. Se utilizaron para el desarrollo del modelo un total de 6100 datos de pixeles de Maíz.

En el siguiente gráfico de la izquierda (Figura 48) se incluyen de manera conjunta, la evolución del coeficiente de determinación R^2 desde el día promedio de Green-Up hasta el día promedio del Green-Down, y el valor medio del índice de vegetación NDVI (línea azul) en ese mismo intervalo de tiempo. Como puede observarse, el máximo valor de R^2 se obtiene el día juliano 2009018. La fecha corresponde al 18 de enero, momento en que en la zona de estudio el ciclo del cultivo de maíz (maíz común de ciclo tradicional) se encuentra finalizando su etapa de floración. Esta campaña agrícola es muy particular para el maíz por el efecto del estrés hídrico sobre el cultivo, y pasado el momento de máximo valor del índice de vegetación (fin del año calendario, aproximadamente a los 60 días de la emergencia), comienza un marcado descenso del vigor vegetal de manera temprana. En el gráfico de la derecha se visualiza la regresión de la variable NDVI (regresora - independiente) que varía de 0 a 1 sobre el eje de coordenadas X, y la variable RENDIMIENTO transformada al logaritmo natural (respuesta - dependiente) sobre el eje Y. Los datos fueron coloreados de acuerdo a la densidad, siendo el color rojo la zona de mayor densidad y la azul la zona de menor densidad de puntos.

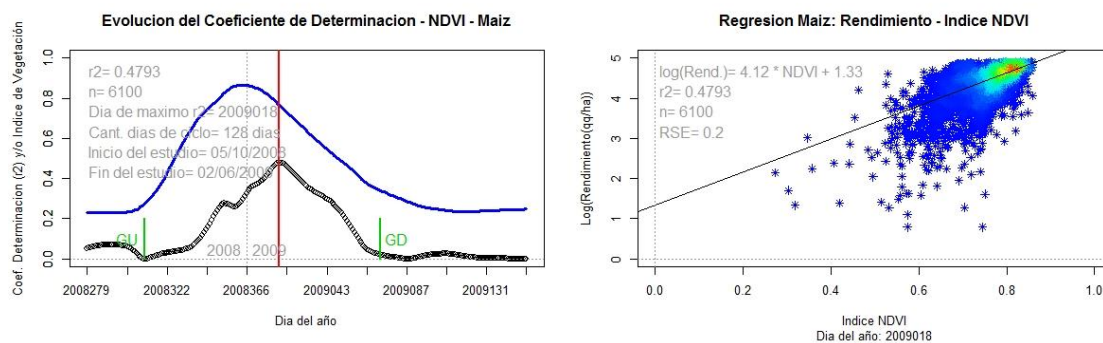


Figura 48 - Izq.: Ejemplo del cálculo del coeficiente de determinación R^2 durante el ciclo del cultivo de Maíz en la campaña 2008/2009. La línea vertical roja corresponde al día de mayor valor de R^2 . La línea azul representa el desarrollo medio del cultivo (valores medios de NDVI). Der.: Modelo de regresión lineal que vincula los valores del índice NDVI del día de mayor R^2 con el rendimiento transformado al logaritmo natural. Los colores de la nube de puntos definen la densidad (rojo mayor densidad, azul menor densidad)

Modelo "Sequía-Maíz":

$$YM = 3,781 e^{(4,12 * IVM)}$$

Donde:

YM: Valor de rendimiento estimado por el modelo para la Maíz en períodos de escasez hídrica.

IVM: Valor del índice de vegetación NDVI sobre datos de Maíz a 80 DPG-Promedio que se corresponde con el día juliano 18 en el modelo.

En el análisis de la regresión, la significancia de la pendiente indica la existencia de relación lineal entre las variables (se rechaza la hipótesis nula de pendiente=0). Error Estándar Residual del modelo es de 0.19 (Anexo II). En cuanto a los residuos del modelo, el supuesto de homocedasticidad de la varianza de los residuos gráficamente parece no cumplirse porque se observa un leve comportamiento trapezoidal que evidencia una mayor varianza hacia el valor medio de los valores ajustados, y analizado con el test Breusch-Pagan, analíticamente no logra cumplirse. En cuanto a la normalidad de los residuos, se evidencia en la Figura 49 que el valor medio de los residuos no se encuentra centrado en cero, y observando el gráfico de QQ-plot junto con el histograma de residuos, se evidencia el alejamiento de la cola de residuos negativos (Figura 50). Analizado con el test de Kolmogorov-Smirnov modificado por lilliefors, analíticamente no se cumple la normalidad (Anexo II).

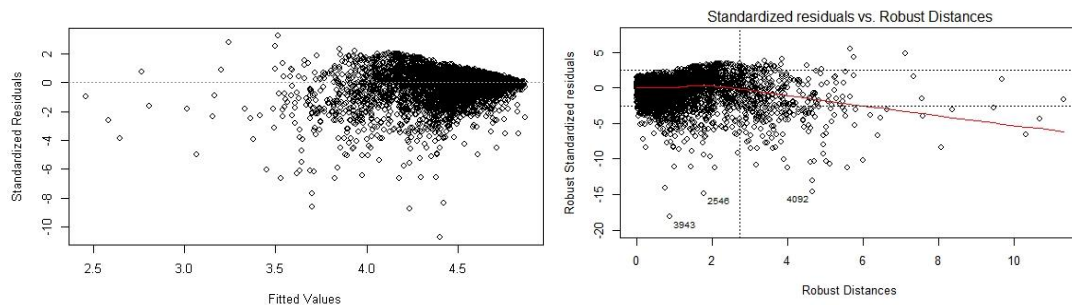


Figura 49 - Izq.: Gráfico de Residuos Estandarizados vs Valores Ajustados. Der.: Gráfico de Residuos Estandarizados vs Distancias Robustas.

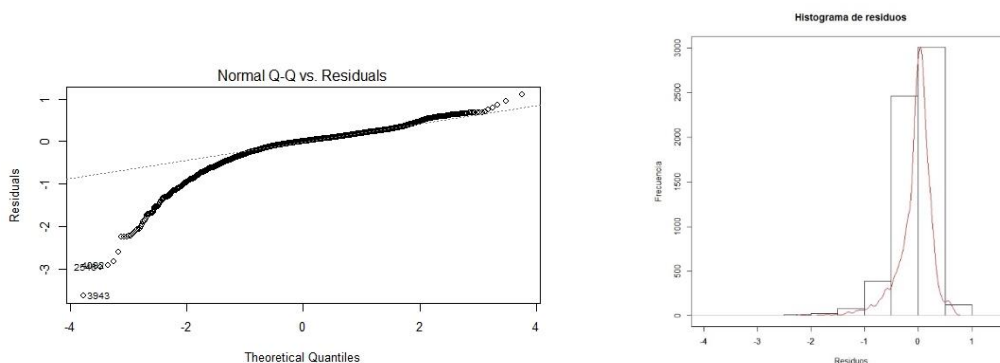


Figura 50 - Izq.: Gráfico de QQ-Plot (normalidad de los residuos). Der.: Histograma de residuos.

3.4.1.4 Modelo seleccionado Girasol:

Utilizando la base de datos de "modelación", se procesaron los datos del índice NDMI para el Girasol obteniendo valores diarios y se obtuvo la función lineal en el día del año que se logra el coeficiente de determinación R^2 más elevado. Para este caso, el modelo de mejor ajuste resultó ser el "a", que vincula los valores por día del año con el rendimiento sin ajustar por Green-Up. Se utilizaron para el desarrollo del modelo un total de 2744 datos de pixeles de Girasol.

En el siguiente gráfico de la izquierda (Figura 51) se incluyen de manera conjunta, la evolución del coeficiente de determinación R^2 a lo largo del tiempo del estudio (puntos negros) y el valor medio del índice NDMI (línea azul). Como puede observarse, el máximo valor de R^2 se obtiene el día juliano 2009006, coincidente con el mayor valor alcanzado por el índice de humedad. Al utilizarse en este caso, el índice NDMI, los valores del índice para el ciclo del cultivo varían entre -0.5 y 0.5, y es por este motivo que la curva azul no se encuentra completa en el gráfico. La fecha corresponde al 6 de enero, momento en que en la zona de estudio el ciclo del cultivo de girasol se encuentra en su etapa de floración. En el mismo gráfico se agrega el valor medio de Green-Up y de Green-Down indicados en color verde. En el gráfico de la derecha se visualiza la regresión de la variable NDMI³² (regresora - independiente) que varía de 0 a 1 sobre el eje de coordenadas X, y la variable RENDIMIENTO transformada al logaritmo natural (respuesta - dependiente) sobre el eje Y. Los datos fueron coloreados de acuerdo a la densidad, siendo el color rojo la zona de mayor densidad y la azul la zona de menor densidad de puntos.

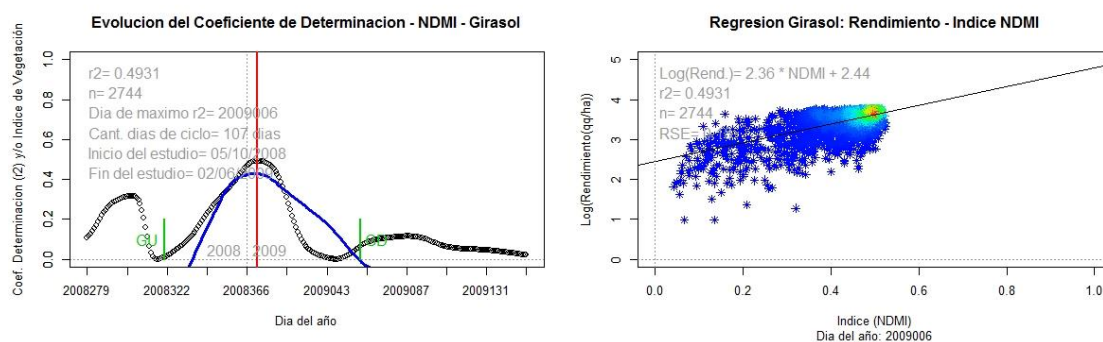


Figura 51 - Izq.: Ejemplo del cálculo del coeficiente de determinación R^2 durante el ciclo del cultivo de Girasol en la campaña 2008/2009. La línea vertical roja corresponde al día de mayor valor de R^2 . La línea azul representa el desarrollo medio del cultivo (valores medios de NDMI). Der.: Modelo de regresión lineal que vincula los valores del índice NDMI del día de mayor R^2 con el rendimiento transformado al logaritmo natural. Los colores de la nube de puntos definen la densidad (rojo mayor densidad, azul menor densidad)

Modelo "Sequía-Girasol":

$$YG = 11,473 e^{(2,36 * IVG)}$$

Donde:

YG: Valor de rendimiento estimado por el modelo para la Girasol en períodos de escasas hídrica.

IVG: Valor del índice de vegetación NDMI sobre datos de Girasol a los 49 DPG-Promedio que se corresponde con el día juliano 6 en el modelo.

En el análisis de la regresión, la significancia de la pendiente indica la existencia de relación lineal entre las variables (se rechaza la hipótesis nula de pendiente=0). Error Estándar Residual del modelo es de 0.18 (Anexo II). En cuanto a los residuos del modelo, el supuesto de homocedasticidad de la varianza de los residuos gráficamente parece no cumplirse porque se observa una mayor varianza hacia los valores bajos del eje x (valores ajustados) (Figura 52), y analizado con el test Breusch-Pagan, analíticamente no logra cumplirse. En cuanto a la normalidad de los residuos, se evidencia en la Figura 52 que el valor medio de los residuos no se encuentra centrado en cero, y observando el gráfico de QQ-plot junto con el histograma de residuos, se evidencia el alejamiento de la cola de residuos negativos (Figura 53). Analizado con

³² A los valores de NDMI se le sumo el valor 1 para evitar números negativos y posteriormente se transformó.

el test de Kolmogorov-Smirnov modificado por lilliefors, analíticamente no se cumple la normalidad (Anexo II).

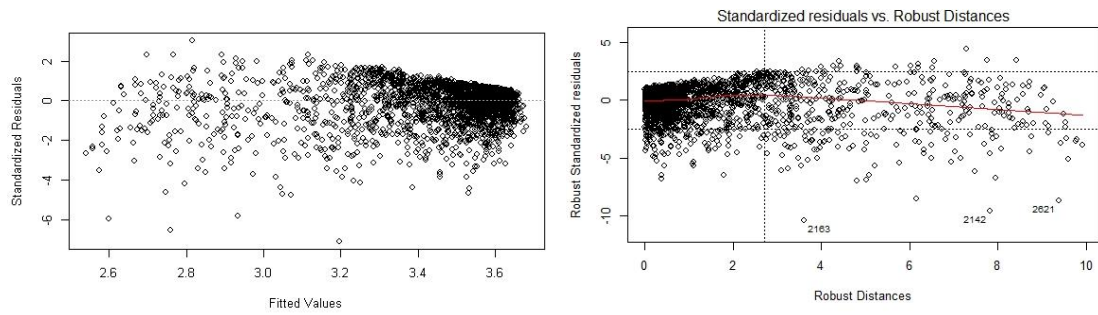


Figura 52 - Izq.: Gráfico de Residuos Estandarizados vs Valores Ajustados. Der.: Gráfico de Residuos Estandarizados vs Distancias Robustas.

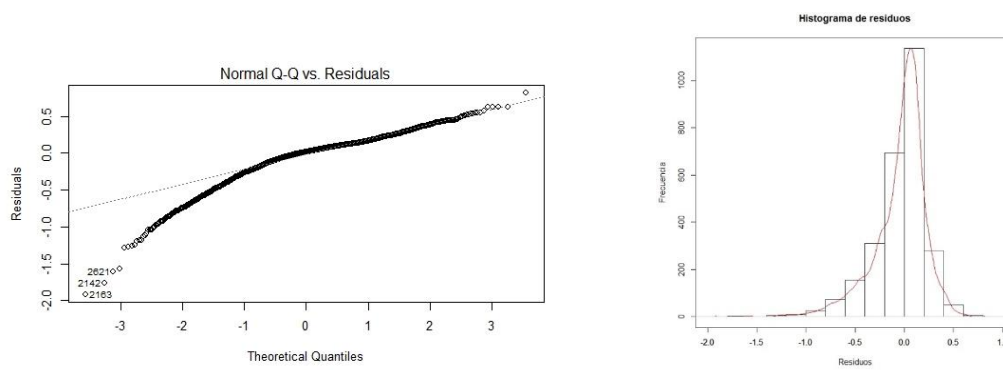


Figura 53 - Izq.: Gráfico de QQ-Plot (normalidad de los residuos). Der.: Histograma de residuos.

3.4.2 Modelos seleccionados para períodos hídricos normales/húmedos (2010/2011):

3.4.2.1 Modelo seleccionado Soja1:

Utilizando la base de datos de modelación (datos guardados al inicio del estudio), se procesaron los datos del índice NDVI para la Soja de primera obteniendo valores diarios, y se obtuvo la función lineal en el día del año que se logra el coeficiente de determinación R^2 más elevado. Para este caso, el modelo de mejor ajuste resultó ser el “a”, que vincula los valores por día del año con el rendimiento sin ajustar por Green-Up. Se utilizaron para el desarrollo del modelo un total de 12100 datos de pixeles de Soja de primera.

En el siguiente gráfico de la izquierda (Figura 54) se incluyen de manera conjunta, la evolución del coeficiente de determinación R^2 a lo largo del tiempo del estudio (puntos negros) y el valor medio del índice de vegetación NDVI (línea verde). Como puede observarse, el máximo valor de R^2 se obtiene el día juliano 2011013. La fecha corresponde al 13 de enero, momento en que en la zona de estudio el ciclo del cultivo de soja de primera se encuentra en pleno crecimiento. En el mismo gráfico se agrega el valor medio de Green-Up y de Green-Down indicados en color verde. En el gráfico de la derecha se visualiza la regresión de la variable NDVI (regresora - independiente) que varía de 0 a 1 sobre el eje de coordenadas X, y la variable RENDIMIENTO transformada al logaritmo (respuesta - dependiente) sobre el eje Y. Los datos fueron coloreados de acuerdo a la densidad, siendo el color rojo la zona de mayor densidad y la azul la zona de menor densidad de puntos.

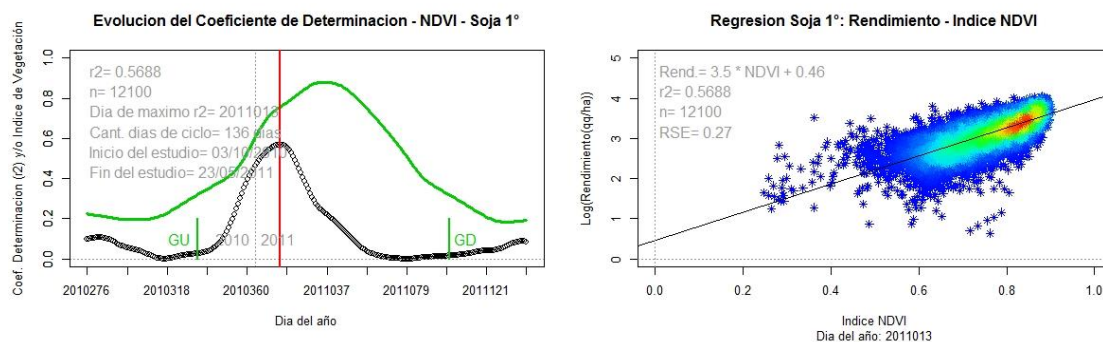


Figura 54 - Izq.: Ejemplo del cálculo del coeficiente de determinación R^2 durante el ciclo del cultivo de Soja de primera en la campaña 2010/2011. La línea vertical roja corresponde al día de mayor valor de R^2 . La línea verde representa el desarrollo medio del cultivo (valores medios de NDVI). Der.: Modelo de regresión lineal que vincula los valores del índice NDVI del día de mayor R^2 con el rendimiento transformado a logaritmo natural. Los colores de la nube de puntos definen la densidad (rojo mayor densidad, azul menor densidad).

Modelo “Normal-Soja1”:

$$YS1 = 1,584 e^{(3,5 * IVS1)}$$

Donde:

YS1: Valor de rendimiento estimado por el modelo para la Soja de primera en períodos hídricos normales.

IVS1: Valor del índice de vegetación NDVI sobre datos de Soja de primera a los 46 DPG-Promedio que se corresponde con el día juliano 13 en el modelo.

En el análisis de la regresión, la significancia de la pendiente indica la existencia de relación lineal entre las variables (se rechaza la hipótesis nula de pendiente=0). Error Estándar Residual del modelo es de 0.27 (Anexo II). En cuanto a los residuos del modelo, el supuesto de homocedasticidad de la varianza de los residuos gráficamente parece no cumplirse porque se observa una mayor varianza hacia los valores bajos del eje x (valores ajustados) (Figura 55), y

analizado con el test Breusch-Pagan, analíticamente no logra cumplirse. En cuanto a la normalidad de los residuos, se observa que el valor medio de los residuos no se encuentra centrado en cero, y visualizando el gráfico de QQ-plot junto con el histograma de residuos, se evidencia el alejamiento de la cola de residuos negativos (Figura 56). Analizado con el test de Kolmogorov-Smirnov modificado por lilliefors, analíticamente no se cumple la normalidad (Anexo II).

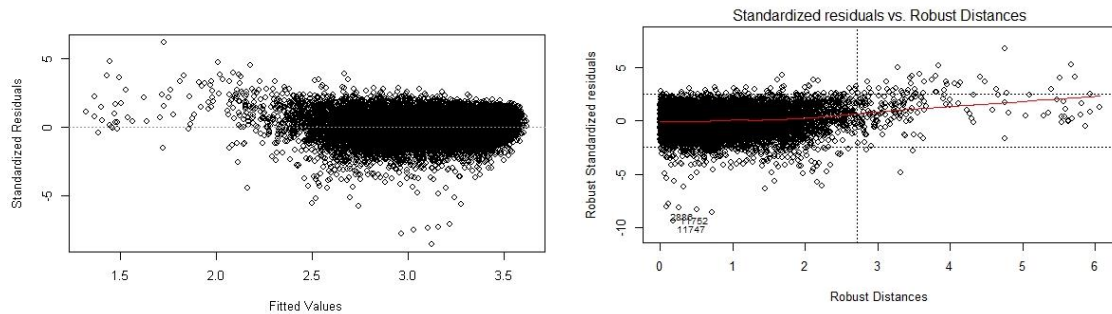


Figura 55 - Izq.: Gráfico de Residuos Estandarizados vs Valores Ajustados. Der.: Gráfico de Residuos Estandarizados vs Distancias Robustas.

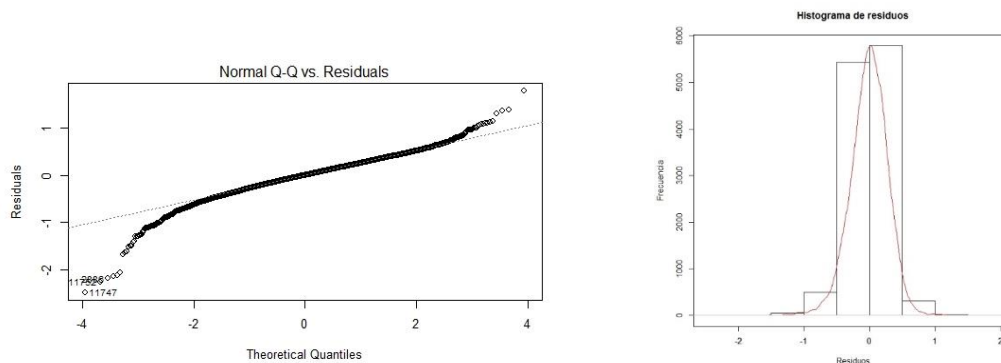


Figura 56 - Izq.: Gráfico de QQ-Plot (normalidad de los residuos). Der.: Histograma de residuos.

3.4.2.2 Modelo seleccionado Soja2:

Al igual que en la Soja de primera, se procesaron los datos del índice NDVI para la Soja de segunda obteniendo valores diarios y se obtuvo la función lineal en el día del año que se logra el coeficiente de determinación R^2 más elevado. Para este caso, el modelo de mejor ajuste resultó ser el “a”, que vincula los valores por día del año con el rendimiento sin ajustar por Green-Up. Se utilizaron para el desarrollo del modelo un total de 3435 datos de pixeles de Soja de segunda.

En el siguiente gráfico de la izquierda (Figura 57) se incluyen de manera conjunta, la evolución del coeficiente de determinación R^2 desde el día promedio de Green-Up hasta el día promedio del Green-Down, y el valor medio del índice de vegetación NDVI (línea verde) en ese mismo intervalo de tiempo. Como puede observarse, el máximo valor de R^2 se obtiene el día juliano 2009084. La fecha corresponde al 25 de marzo, momento en que en la zona de estudio el ciclo del cultivo de soja de segunda se encuentra en una etapa avanzada (formación del grano). En el gráfico de la derecha se visualiza la regresión de la variable NDVI (regresora - independiente) que varía de 0 a 1 sobre el eje de coordenadas X, y la variable RENDIMIENTO transformada al logaritmo natural (respuesta - dependiente) sobre el eje Y. Los datos fueron coloreados de acuerdo a la densidad, siendo el color rojo la zona de mayor densidad y la azul la zona de menor densidad de puntos.

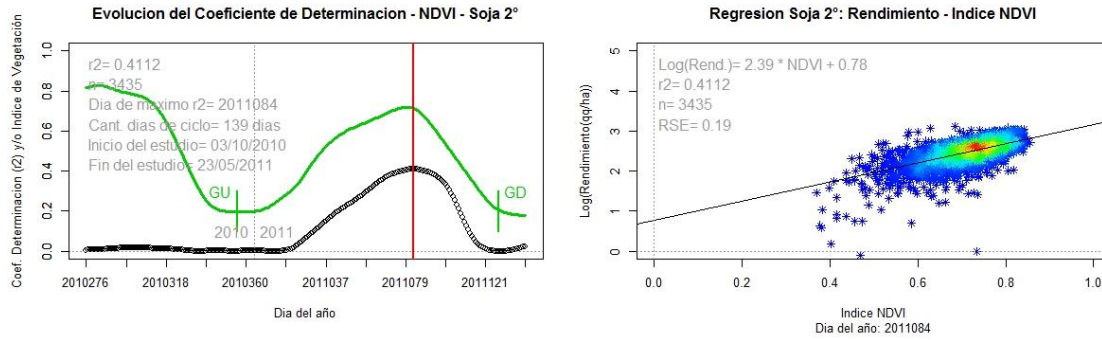


Figura 57 - Izq.: Ejemplo del cálculo del coeficiente de determinación R^2 durante el ciclo del cultivo de Soja de segunda en la campaña 2010/2011. La línea vertical roja corresponde al día de mayor valor de R^2 . La línea verde representa el desarrollo medio del cultivo (valores medios de NDVI). Der.: Modelo de regresión lineal que vincula los valores del índice NDVI del día de mayor R^2 con el rendimiento transformado a logaritmo natural en unidades de qq/ha. Los colores de la nube de puntos definen la densidad (rojo mayor densidad, azul menor densidad)

Modelo "Normal-Soja2":

$$YS2 = 2,181 e^{(2,39 * IVS2)}$$

Donde:

YS2: Valor de rendimiento estimado por el modelo para la Soja de segunda en períodos hídricos normales.

IVS2: Valor del índice de vegetación NDVI sobre datos de Soja de segunda a los 82 DPG-Promedio que se corresponde con el día juliano 84 en el modelo.

En el análisis de la regresión, la significancia de la pendiente indica la existencia de relación lineal entre las variables (se rechaza la hipótesis nula de pendiente=0). Error Estándar Residual del modelo es de 0.19 (Anexo II). En cuanto a los residuos del modelo, el supuesto de homocedasticidad de la varianza de los residuos gráficamente parece no cumplirse porque se observa una mayor varianza hacia los valores bajos del eje x (valores ajustados) (Figura 58), y analizado con el test Breusch-Pagan, analíticamente no logra cumplirse. En cuanto a la normalidad de los residuos, se observa que el valor medio de los residuos no se encuentra centrado en cero, y visualizando el gráfico de QQ-plot junto con el histograma de residuos, se evidencia el alejamiento de la cola de residuos negativos (Figura 59). Analizado con el test de Kolmogorov-Smirnov modificado por lilliefors, analíticamente no se cumple la normalidad (Anexo II).

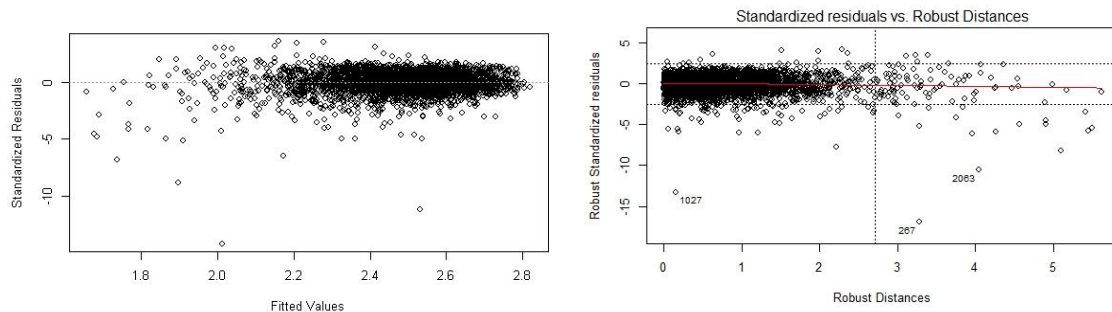


Figura 58 - Izq.: Gráfico de Residuos Estandarizados vs Valores Ajustados. Der.: Gráfico de Residuos Estandarizados vs Distancias Robustas.

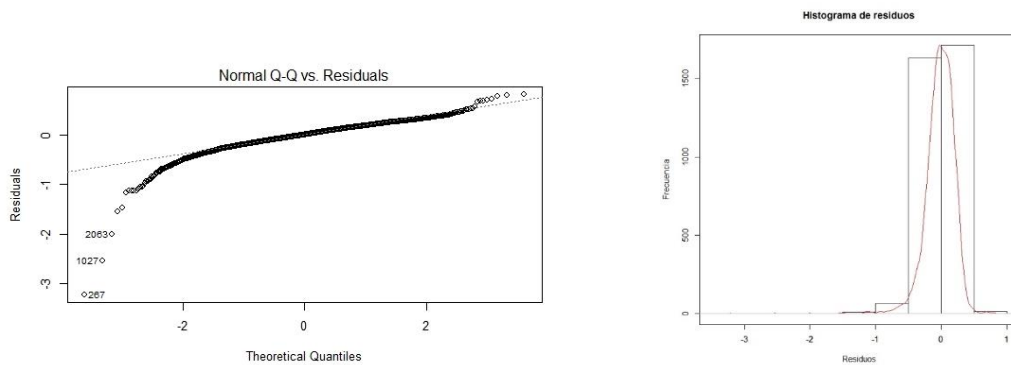


Figura 59 - Izq.: Gráfico de QQ-Plot (normalidad de los residuos). Der.: Histograma de residuos.

3.4.2.3 Modelo seleccionado Maíz

Utilizando la base de datos de modelación (datos guardados al inicio del estudio), se procesaron los datos del índice NDVI para Maíz obteniendo valores diarios y se obtuvo la función lineal en el día del año que se logra el coeficiente de determinación R^2 más elevado. Para este caso, el modelo de mejor ajuste resultó ser el “a”, que vincula los valores por día del año con el rendimiento sin ajustar por Green-Up. Se utilizaron para el desarrollo del modelo un total de 7233 datos de pixeles de Maíz.

En el siguiente gráfico de la izquierda (Figura 60) se incluyen de manera conjunta, la evolución del coeficiente de determinación R^2 desde el día promedio de Green-Up hasta el día promedio del Green-Down, y el valor medio del índice de vegetación NDVI (línea verde) en ese mismo intervalo de tiempo. Como puede observarse, el máximo valor de R^2 se obtiene el día juliano 2011017. La fecha corresponde al 17 de enero, momento en que en la zona de estudio el ciclo del cultivo de Maíz se encuentra aún en su período crítico (finalizando la floración). En el gráfico de la derecha se visualiza la regresión de la variable NDVI (regresora - independiente) que varía de 0 a 1 sobre el eje de coordenadas X, y la variable RENDIMIENTO transformada al logaritmo natural (respuesta - dependiente) sobre el eje Y. Los datos fueron coloreados de acuerdo a la densidad, siendo el color rojo la zona de mayor densidad y la azul la zona de menor densidad de puntos.

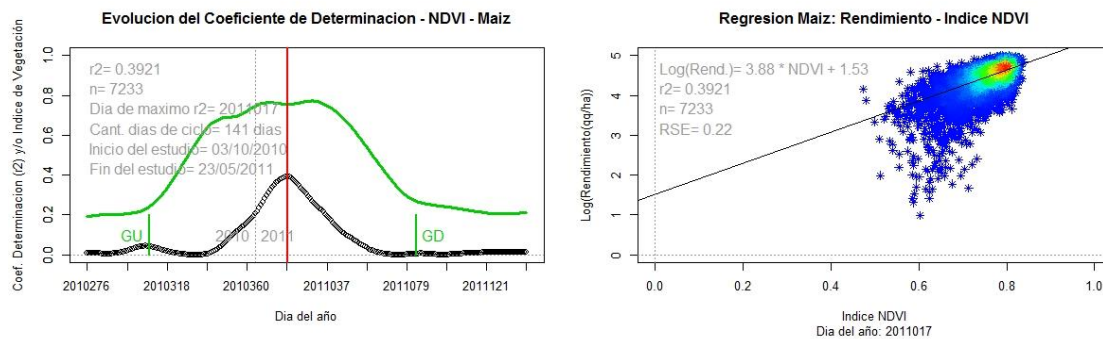


Figura 60 - Izq.: Ejemplo del cálculo del coeficiente de determinación R^2 durante el ciclo del cultivo de Maíz en la campaña 2010/2011. La línea vertical roja corresponde al día de mayor valor de R^2 . La línea verde representa el desarrollo medio del cultivo (valores medios de NDVI). Der.: Modelo de regresión lineal que vincula los valores del índice NDVI en el día de mayor R^2 con el rendimiento transformado a logaritmo natural (qq/ha). Los colores de la nube de puntos definen la densidad (rojo mayor densidad, azul menor densidad)

Modelo "Normal-Maíz":

$$YM = 4,618 e^{(3.88 * IVM)}$$

Donde:

YM: Valor de rendimiento estimado por el modelo para la Maíz en períodos hídricos normales.

IVM: Valor del índice de vegetación NDVI sobre datos de Maíz a los 52 DPG-Promedio que se corresponde con el día juliano 17 en el modelo.

En el análisis de la regresión, la significancia de la pendiente indica la existencia de relación lineal entre las variables (se rechaza la hipótesis nula de pendiente=0). Error Estándar Residual del modelo es de 0.22 (Anexo II). En cuanto a los residuos del modelo, el supuesto de homocedasticidad de la varianza de los residuos gráficamente parece no cumplirse porque se observa una mayor varianza en los valores medios del eje x (valores ajustados) (Figura 61), y analizado con el test Breusch-Pagan, analíticamente no logra cumplirse. En cuanto a la normalidad de los residuos, se observa que el valor medio de los residuos no se encuentra centrado en cero, y visualizando el gráfico de QQ-plot junto con el histograma de residuos, se evidencia el alejamiento de la cola de residuos negativos (Figura 62). Analizado con el test de Kolmogorov-Smirnov modificado por lilliefors, analíticamente no se cumple la normalidad (Anexo II).

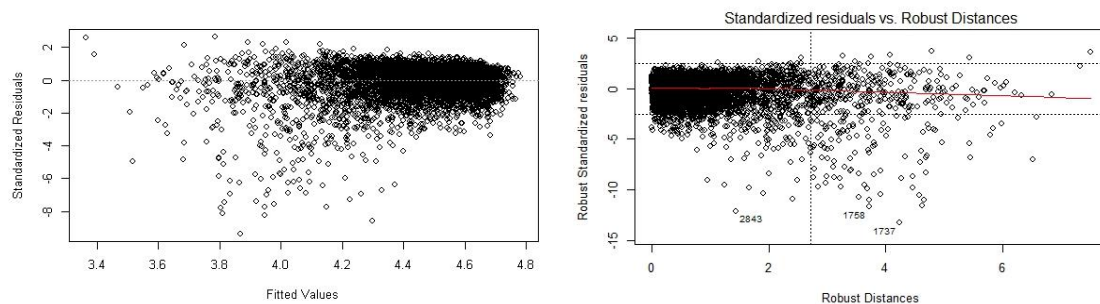


Figura 61 - Izq.: Gráfico de Residuos Estandarizados vs Valores Ajustados. Der.: Gráfico de Residuos Estandarizados vs Distancias Robustas.

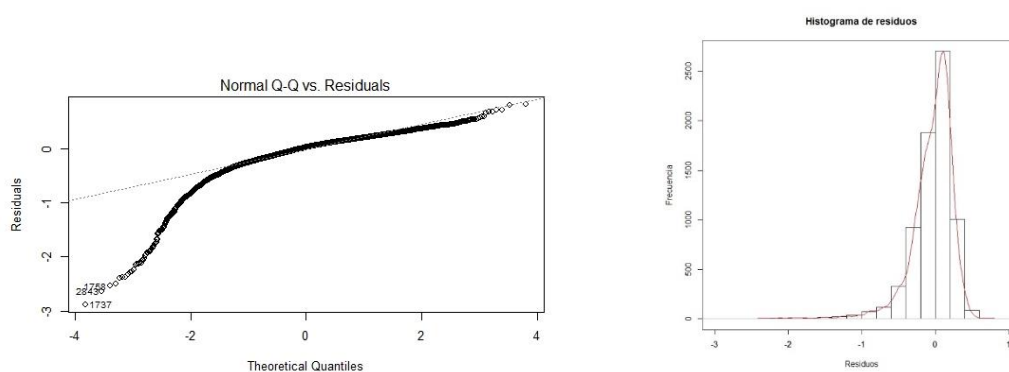


Figura 62 - Izq.: Gráfico de QQ-Plot (normalidad de los residuos). Der.: Histograma de residuos.

3.4.2.4 Modelo seleccionado Girasol:

Utilizando la base de datos de modelación (datos guardados al inicio del estudio), se procesaron los datos del índice NDVI para el Girasol obteniendo valores diarios y se obtuvo la función lineal en el día del año que se logra el coeficiente de determinación R^2 más elevado. Se utilizaron para el desarrollo del modelo un total de 4698 datos de píxeles de Girasol.

En el siguiente gráfico de la izquierda (Figura 63) se incluyen de manera conjunta, la evolución del coeficiente de determinación R^2 a lo largo del tiempo del estudio (puntos negros)

y el valor medio del índice de vegetación NDVI (línea verde). Como puede observarse, el máximo valor de R^2 se obtiene el día juliano 2011001 (inicio del año calendario), y en momentos en los que el índice de vegetación alcanza el valor del “plateau” (meseta). La fecha corresponde al 1° de enero, y en la zona de estudio el cultivo de Girasol se encuentra iniciando su etapa de floración. En el mismo gráfico se agrega el valor medio de Green-Up y de Green-Down indicados en color verde. En el gráfico de la derecha se visualiza la regresión de la variable NDVI (regresora - independiente) que varía de 0 a 1 sobre el eje de coordenadas X, y la variable RENDIMIENTO transformada a logaritmo natural (respuesta - dependiente) sobre el eje Y. Los datos fueron coloreados de acuerdo a la densidad, siendo el color rojo la zona de mayor densidad y la azul la zona de menor densidad de puntos.

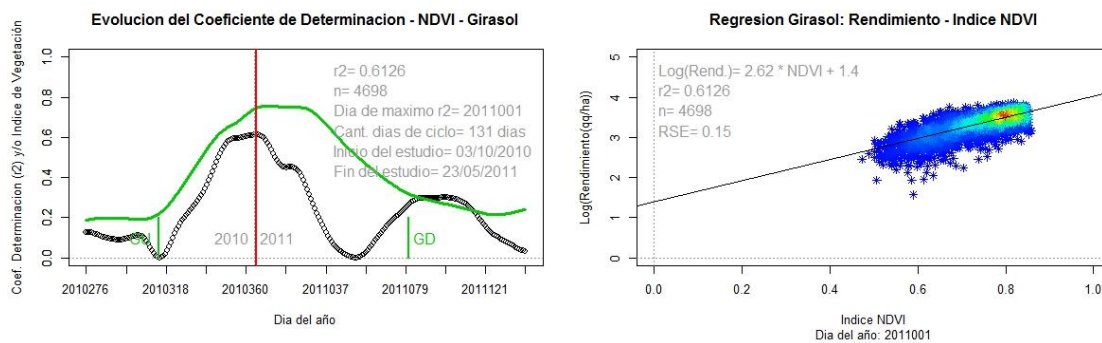


Figura 63 - Izq.: Ejemplo del cálculo del coeficiente de determinación R^2 durante el ciclo del cultivo de Soja de segunda en la campaña 2010/2011. La línea vertical roja corresponde al día de mayor valor de R^2 . La línea verde representa el desarrollo medio del cultivo (valores medios de NDVI). Der.: Modelo de regresión lineal que vincula los valores del índice NDVI del día de mayor R^2 con el rendimiento transformado a logaritmo natural. Los colores de la nube de puntos definen la densidad (rojo mayor densidad, azul menor densidad)

Modelo “Normal-Girasol”:

$$YG = 4,05 e^{(2.62 * IVG)}$$

Donde:

YG: Valor de rendimiento estimado por el modelo para la Girasol en períodos hídricos normales.

IVG: Valor del índice de vegetación NDVI sobre datos de Girasol a los 51 DPG-Promedio que se corresponde con el día juliano 1 en el modelo.

En el análisis de la regresión, la significancia de la pendiente indica la existencia de relación lineal entre las variables (se rechaza la hipótesis nula de pendiente=0). Error Estándar Residual del modelo es de 0.15 (Anexo II). En cuanto a los residuos del modelo, el supuesto de homocedasticidad de la varianza de los residuos gráficamente parece no cumplirse porque se observa una mayor varianza en los valores medios del eje x (valores ajustados) (Figura 64), y analizado con el test Breusch-Pagan, analíticamente no logra cumplirse. En cuanto a la normalidad de los residuos, se observa que el valor medio de los residuos se encuentra bastante centrado en cero, visualizando el gráfico de QQ-plot se evidencia el alejamiento de la cola de residuos negativos, y en el histograma de residuos simula una distribución normal (Figura 65). Analizado con el test de Kolmogorov-Smirnov modificado por lilliefors, analíticamente se cumple la normalidad con alfa de significancia de 0.01 (Anexo II).

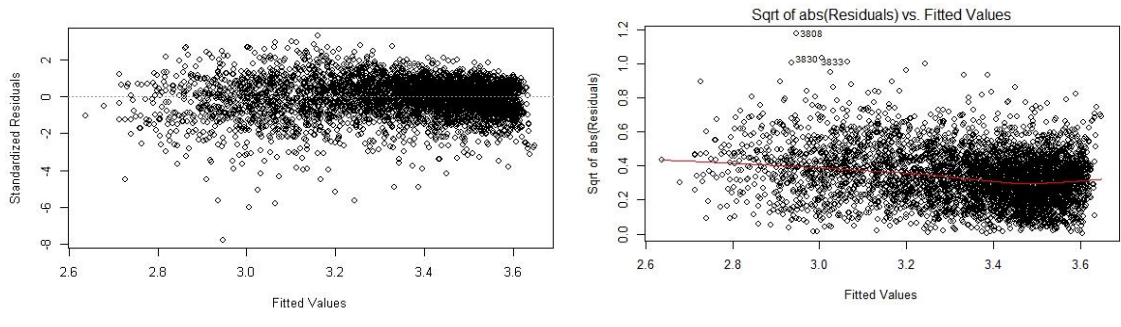


Figura 64 - Izq.: Gráfico de Residuos Estandarizados vs Valores Ajustados. Der.: Gráfico de Residuos Estandarizados vs Distancias Robustas.

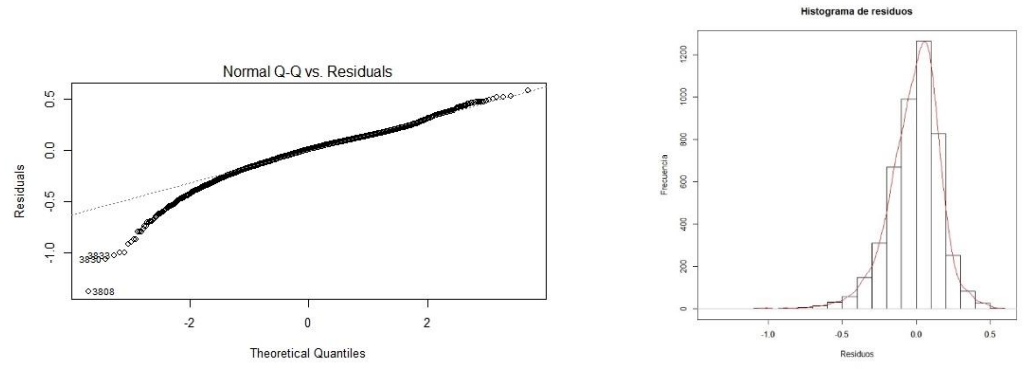


Figura 65 - Izq.: Gráfico de QQ-Plot (normalidad de los residuos). Der.: Histograma de residuos.

3.5 Validación de los modelos seleccionados.

Se utilizó la base de datos reservada para la validación de los modelos. La misma fue procesada e interpolada (utilizando el mismo script generado en R-Studio para procesar los datos de la base “modelación”). Finalizado el proceso, se identificó la columna de datos del índice de vegetación correspondiente al día del año en el que se deben extraer los datos para ingresar al modelo y se le aplicó el modelo lineal (ecuación de rendimiento). Este proceso se realizó para cada tipo de cultivo y cada campaña agrícola.

El resultado del proceso es el rendimiento predicho por cada modelo, y que, en relación al rendimiento observado, nos permitirá evaluar el desempeño de cada uno de ellos. La evaluación se realizará por medio de estos 5 indicadores:

P30 (% de datos menores a 30% de error)

$$P_{30} = \frac{N_{\varepsilon < 0,3}}{N} \times 100$$

MBE (Mean Bias Error)

$$MBE = \frac{\sum_{i=1}^N (O_i - E_i)}{N}$$

RMSE (Root Mean Squared Error)

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (O_i - E_i)^2}{N}}$$

RE (Relative Error)

$$RE = \frac{RMSE}{O}$$

d (Índice de concordancia de Willmott)

$$d = 1 - \frac{\sum_{i=1}^N (O_i - E_i)^2}{\sum_{i=1}^N (|E_i - \bar{O}| + |O_i - \bar{O}|)^2}$$

E: Dato Estimado; *O*: Dato Observado; *N*: Cantidad de observaciones. $N_{\varepsilon < 0,3}$: Cantidad de observaciones con error menor a 30%

El P30 nos permite conocer el porcentaje del total de los datos que logran errores entre los valores observado y los predichos menores al 30%.

El Bias ó MBE es el valor medio de los residuos del modelo (valor medio de las diferencias positivas y negativas). Si las diferencias medias entre los observado y lo predicho arroja un valor negativo, el modelo sobrestima (los valores estimados son mayores a los observados), por el contrario, si el valor es positivo, el modelo subestima. El valor del Bias se encuentra en las unidades de medida del modelo, por lo cual, en este caso permite determinar el valor medio del error de sobrestimación o subestimación en quintales por hectárea (qq/ha).

El “RMSE” es la raíz cuadrada de la media de los residuos elevados al cuadrado. Este indicador del error de un modelo permite determinar la capacidad de predicción y su resultado mantiene las unidades de medida de origen. Valores de RMSE cercanos a cero indican un bajo error del modelo con una alta capacidad de predicción. Este valor también se conoce como “desviación cuadrática media” y arroja valores similares a la desviación estándar de los residuos (si la distribución de los residuos es de tipo normal) aunque el RMSE es sensible a la presencia de valores atípicos (debido a que eleva los residuos al cuadrado). Se podría asumir (como para comprender mejor este indicador), que el 68% de los datos (1 desvío estándar), tienen un error medio de estimación igual al RMSE.

El error relativo “RE” se calcula como el RMSE dividido el valor medio de los datos observados, y es por esta razón que se lo conoce también como RRMSE (Relative Root Mean Squared Error). Este indicador permite obtener un valor porcentual del error y por lo tanto lo hace más intuitivo para entender el desempeño del modelo.

El valor “d” conocido como Índice de Concordancia de Willmott permite determinar la alta o baja relación entre los valores estimados y los observados. Los valores de “d” varían de 0 a 1, siendo el valor de 1 una relación perfecta.

3.5.1 Validación Soja 1 – 2008/2009.

El modelo desarrollado para la Soja de primera en períodos de escasas hídricas logró un buen ajuste entre los valores observados (rendimiento de datos de campo guardados para la validación) y los predichos (rendimiento estimado por el modelo) utilizando una base de control de 10897 datos (píxeles).

Se observa en el diagrama de dispersión (Figura 66) que la distribución de los datos ajustan muy bien, aunque al superar los 40 qq/ha se ingresa en una meseta en la cual valores observados arrojan valores predichos no superiores a los 40 qq/ha. Esto determina que la función de ajuste (observado/predicho) indique una subestimación del rendimiento al superar los 40 qq/ha y una sobre estimación en valores inferiores a 10 qq/ha.

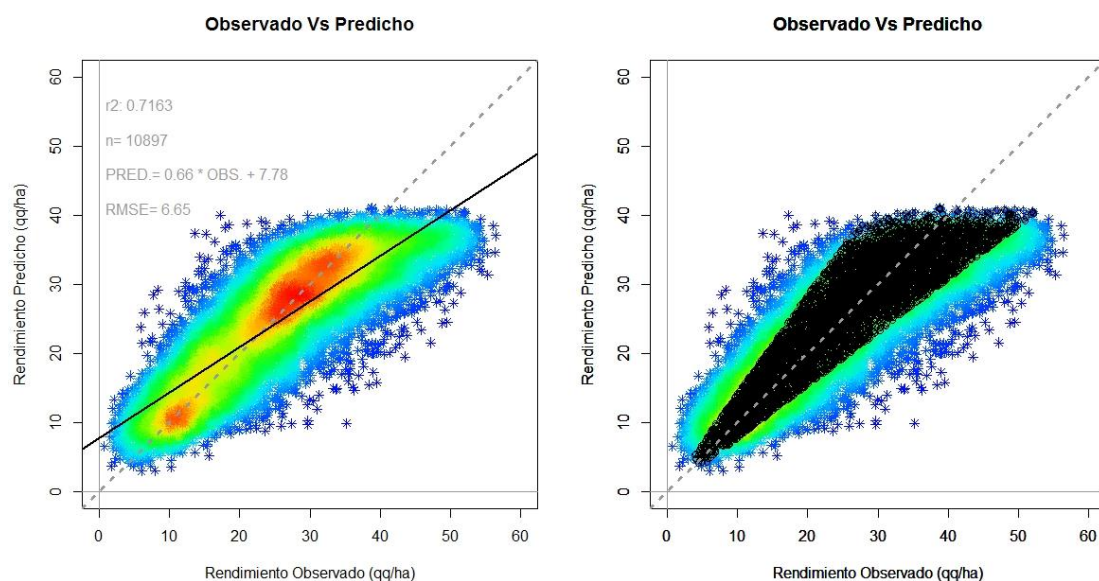


Figura 66 - Observado Vs Predicho. Soja de primera campaña 2008/2009. Izq.: Diagrama de dispersión (rojo mayor densidad, azul menor densidad) con recta de regresión (línea negra) del ajuste de los datos y datos de la regresión insertos en el gráfico. Der.: Ídem Izq. con identificación de los datos con menos del 30% de error (P30) entre el valor observado y el predicho (puntos negros).

3.5.1.1 Análisis del error del modelo Soja 1 – 2008/2009.

- P30 = 71.02 %
- MBE = 0.90 qq/ha
- RMSE = 6.65 qq/ha
- RE = 26.11 %
- d = 0.9999864

Valores de los residuos expresados en quintales:

- Mínimo = -23.78 qq/ha
- Máximo = 25.64 qq/ha
- Desvío estándar = 6.58 qq/ha

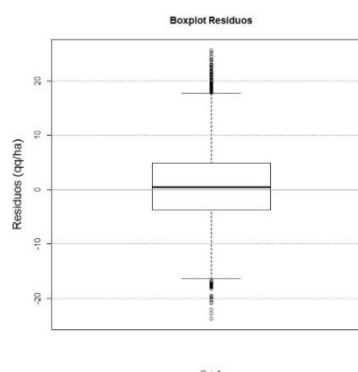


Figura 67 - Boxplot de los residuos. Soja de primera campaña 2008/2009.

Este modelo logró un buen ajuste entre los datos observados y los predichos obteniendo un valor de R^2 elevado (0.71). Los datos que obtienen un porcentaje de error absoluto entre los datos observados y los estimados por el modelo menores al 30% acumulan un total del 71%. El MBE indica una subestimación promedio del rendimiento del orden de los 90 kg/ha. El RMSE arroja un error de predicción promedio del modelo de 665 kg/ha que representa un error relativo del 26.11%. El valor del índice “d” indica una perfecta relación entre los datos observados y los predichos.

3.5.2 Validación Soja 2 – 2008/2009.

El modelo desarrollado para la Soja de segunda en períodos de escasas hídricas logró un ajuste medio entre los valores observados y los predichos por el modelo utilizando una base de control de 5043 datos (píxeles).

Se observa en el diagrama de dispersión (Figura 68) que la distribución de los datos ajusta muy bien, aunque al superar los 20 qq/ha se ingresa en una meseta en la cual valores observados arrojan valores predichos no superiores a los 20 qq/ha.

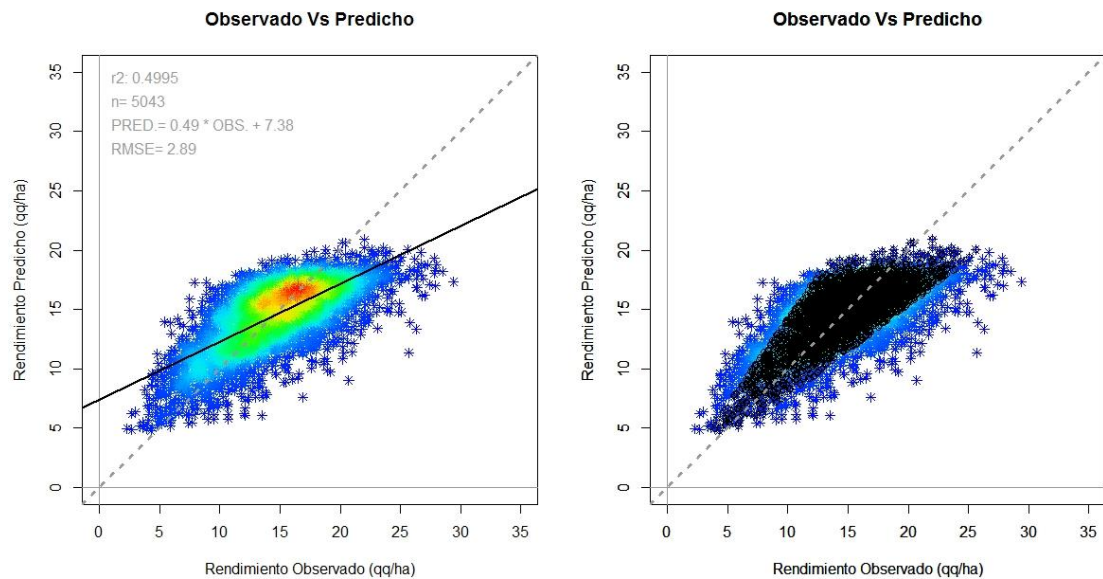


Figura 68 - Observado Vs Predicho. Soja de primera campaña 2008/2009. Izq.: Diagrama de dispersión (rojo mayor densidad, azul menor densidad) con recta de regresión (línea negra) del ajuste de los datos y datos de la regresión insertos en el gráfico. Der.: Ídem Izq. con identificación de los datos con menos del 30% de error absoluto (P30) entre el valor observado y el predicho (puntos negros).

3.5.2.1 Análisis del error del modelo Soja 2 – 2008/2009.

- P30 = 86.29 %
- MBE = 0.22 qq/ha
- RMSE = 2.89 qq/ha
- RE = 19.69 %
- d = 0.9999435

Valores de los residuos expresados en quintales:

- Mínimo = -9.38 qq/ha
- Máximo = 14.34 qq/ha
- Desvío estándar = 2.88 qq/ha

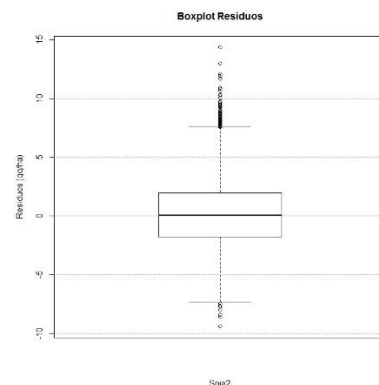


Figura 69 - Boxplot de los residuos. Soja de segunda campaña 2008/2009.

Este modelo logró un ajuste medio entre los datos observados y los predichos obteniendo un valor de R^2 de 0.49. El MBE indica una leve subestimación promedio del rendimiento del orden de los 22 kg/ha. El RMSE arroja un error de predicción promedio del modelo de 289 kg/ha que representa un error relativo del 19.69%. El valor del índice “d” indica una perfecta relación entre los datos observados y los predichos.

3.5.3 Validación Maíz – 2008/2009.

El modelo desarrollado para Maíz en períodos de escasas hídricas logró un ajuste medio entre los valores observados y los predichos por el modelo utilizando una base de control de 6186 datos (píxeles).

Se observa en la nube de puntos del gráfico de regresión (Figura 70) que los datos no ajustan bien por como se encuentran distribuidos espacialmente. Los valores más concentrados (coloración roja) y que se encuentran en el rango de los 80 a 130 qq/ha, tienen un buen ajuste a la respuesta lineal 1 a 1 (45°), pero el modelo lineal se ve influenciado por una larga cola de datos hacia los valores bajos de rendimiento.

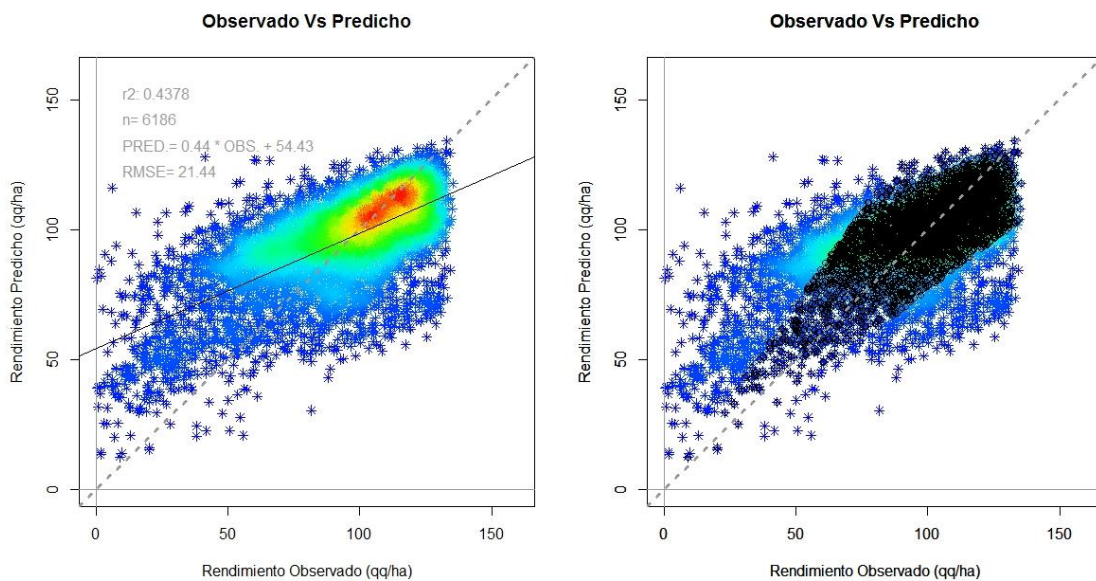


Figura 70 - Observado Vs Predicho. Maíz campaña 2008/2009. Izq.: Diagrama de dispersión (rojo mayor densidad, azul menor densidad) con recta de regresión (línea negra) del ajuste de los datos y datos de la regresión insertos en el gráfico. Der.: Ídem Izq. con identificación de los datos con menos del 30% de error absoluto (P30) entre el valor observado y el predicho (puntos negros).

3.5.3.1 Análisis del error del modelo Maíz – 2008/2009.

- P30 = 78.8 %
- MBE = -3.3 qq/ha
- RMSE = 21.44 qq/ha
- RE = 23.4 %
- d = 0.9999487

Valores de los residuos expresados en quintales:

- Mínimo = -110.21 qq/ha
- Máximo = 66.81 qq/ha
- Desvío estándar = 21.18 qq/ha

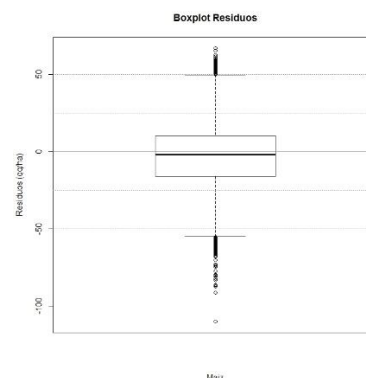


Figura 71 - Boxplot de los residuos. Maíz campaña 2008/2009.

Este modelo logró un ajuste medio entre los datos observados y los predichos obteniendo un valor de R^2 de 0.43. El MBE indica una sobrestimación promedio del rendimiento del orden de los 330 kg/ha. El RMSE arroja un error de predicción promedio del modelo de 2144 kg/ha que representa un error relativo del 23.4%. El valor del índice “d” indica una perfecta relación entre los datos observados y los predichos.

3.5.4 Validación Girasol – 2008/2009.

El modelo desarrollado para el Girasol en períodos de escasas hídrica logró un buen ajuste entre los valores observados y los predichos por el modelo utilizando una base de control de 2768 datos (píxeles).

Se observa en la nube de puntos del gráfico de regresión (Figura 72) que los datos ajustan con una distribución irregular, aunque se mantienen dentro del rango de error del 30%. Se evidencia una sobre estimación (observaciones desperdigas hacia la izquierda) pero de baja densidad (rendimientos muy bajos no habituales para la zona de estudio).

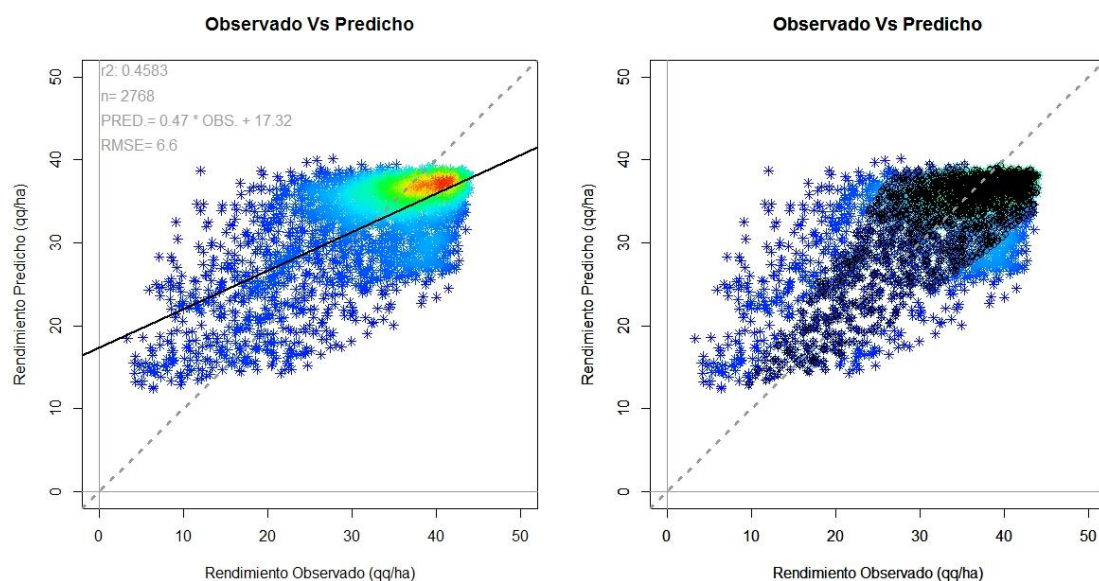


Figura 72 - Observado Vs Predicho. Girasol campaña 2008/2009. Datos de regresión insertos en el gráfico. Izq.: Diagrama de dispersión (rojo mayor densidad, azul menor densidad) con recta de regresión (línea negra) del ajuste de los datos y datos de la regresión insertos en el gráfico. Der.: Ídem Izq. con identificación de los datos con menos del 30% de error absoluto (P30) entre el valor observado y el predicho (puntos negros).

3.5.4.1 Análisis del error del modelo Girasol – 2008/2009.

- P30 = 80.4 %
- MBE = -0.24 qq/ha
- RMSE = 6.60 qq/ha
- RE = 20.68 %
- d = 0.9998957

Valores de los residuos expresados en quintales:

- Mínimo = -26.7 qq/ha
- Máximo = 16.9 qq/ha
- Desvío estándar = 6.59 qq/ha

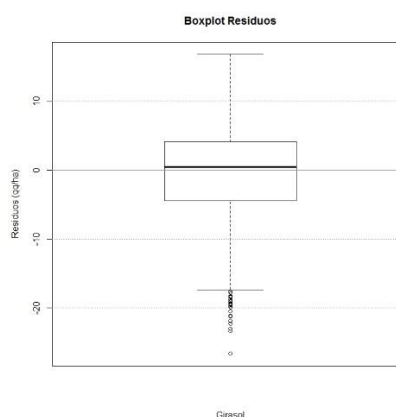


Figura 73 - Boxplot de los residuos. Girasol campaña 2008/2009.

Este modelo logró un ajuste medio entre los datos observados y los predichos obteniendo un valor de R^2 de 0.45. El MBE indica una sobrestimación promedio del rendimiento del orden de los 24 kg/ha. El RMSE arroja un error de predicción promedio del modelo de 660 kg/ha que representa un error relativo del 20.68%. El valor del índice "d" indica una perfecta relación entre los datos observados y los predichos.

3.5.5 Validación Soja 1 – 2010/2011.

El modelo desarrollado para la Soja de primera en períodos hídricos normales/húmedos logró un buen ajuste entre los valores observados y los predichos por el modelo utilizando una base de control de 12338 datos (píxeles).

Se observa en la distribución de puntos del gráfico de regresión (Figura 74) que los datos ajustan muy bien hasta valores de rendimiento de 35 qq/ha, luego ingresa en una meseta en la cual valores observados entre 35 y 55 qq/ha arrojan valores predichos no superiores a los 40 qq/ha. Esto determina que la función de ajuste (observado/predicho) indique una subestimación del rendimiento al superar los 35 qq/ha y una sobre estimación en valores inferiores a 15 qq/ha.

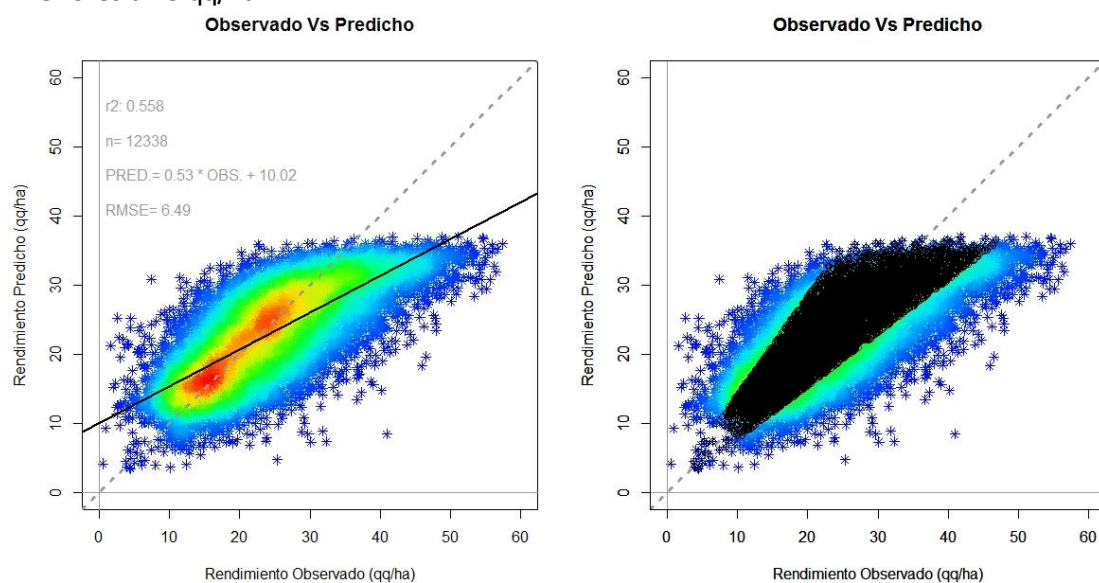


Figura 74 - Observado Vs Predicho. Soja de primera campaña 2010/2011. Datos de regresión insertos en el gráfico. Izq.: Diagrama de dispersión (rojo mayor densidad, azul menor densidad) con recta de regresión (línea negra) del ajuste de los datos y datos de la regresión insertos en el gráfico. Der.: Idem Izq. con identificación de los datos con menos del 30% de error absoluto (P30) entre el valor observado y el predicho (puntos negros).

3.5.5.1 Análisis del error del modelo Soja 1 – 2010/2011.

- P30 = 71.9 %
- MBE = 1.30 qq/ha
- RMSE = 6.49 qq/ha
- RE = 26.84 %
- d = 0.9999806

Valores de los residuos expresados en quintales:

- Mínimo = -23.43 qq/ha
- Máximo = 32.43 qq/ha
- Desvío estándar = 6.36 qq/ha

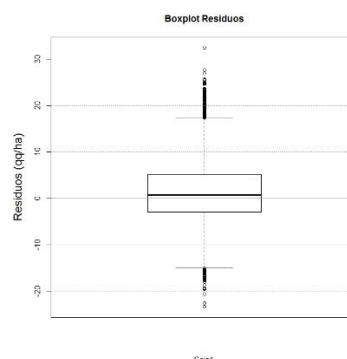


Figura 75 - Boxplot de los residuos. Soja de primera campaña 2010/2011.

Este modelo logra un ajuste adecuado entre los datos observados y los predichos obteniendo un valor de R^2 de 0.55. El MBE indica una subestimación promedio del rendimiento del orden de los 130 kg/ha. El RMSE arroja un error de predicción promedio del modelo de 649 kg/ha que representa un error relativo del 26.84%. El valor del índice “d” indica una perfecta relación entre los datos observados y los predichos.

3.5.6 Validación Soja 2 – 2010/2011.

El modelo desarrollado para la Soja de segunda en períodos normales/húmedos no logró un elevado ajuste entre los valores observados y los predichos por el modelo utilizando una base de control de 3378 datos (píxeles).

Se observa en la distribución de puntos del gráfico de regresión (Figura 76) que los datos ajustan muy bien entre valores de rendimiento de 8 a 18 qq/ha, por fuera del rango el modelo subestima o sobreestima.

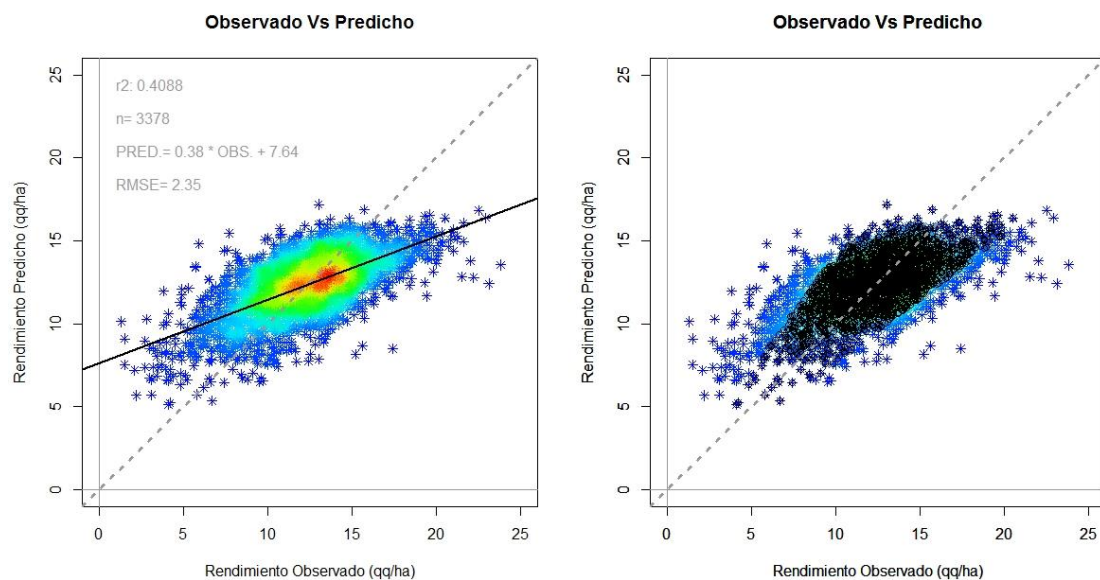


Figura 76 - Observado Vs Predicho. Soja de segunda campaña 2010/2011. Datos de regresión insertos en el gráfico. Datos de regresión insertos en el gráfico. Izq.: Diagrama de dispersión (rojo mayor densidad, azul menor densidad) con recta de regresión (línea negra) del ajuste de los datos y datos de la regresión insertos en el gráfico. Der.: Ídem Izq. con identificación de los datos con menos del 30% de error absoluto (P30) entre el valor observado y el predicho (puntos negros).

3.5.6.1 Análisis del error del modelo Soja 2 – 2010/2011.

- P30 = 88. 90 %
- MBE = -0.06 qq/ha
- RMSE = 2.34 qq/ha
- RE = 19.13 %
- d = 0.9998875

Valores de los residuos expresados en quintales:

- Mínimo = -8.91 qq/ha
- Máximo = 10.69 qq/ha
- Desvío estándar = 2.34 qq/ha

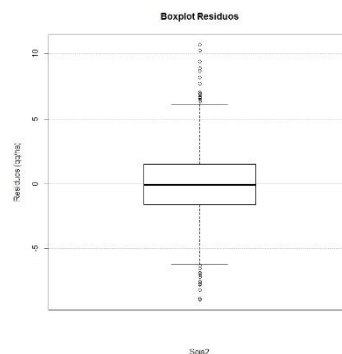


Figura 77 - Boxplot de los residuos. Soja de segunda campaña 2010/2011.

Este modelo no logra un buen ajuste entre los datos observados y los predichos obteniendo un valor de R^2 de 0.4, sin embargo, el MBE cercano a cero indica que en el promedio de los datos el rendimiento no sobrestima ni subestima, y el RMSE arroja un error de predicción promedio del modelo de 234 kg/ha que representa un error relativo del 19.13%. O sea, que a pesar de no lograr un alto valor de R^2 , el error del modelo es aceptable. El valor del índice “d” indica una perfecta relación entre los datos observados y los predichos.

3.5.7 Validación Maíz – 2010/2011.

El modelo desarrollado para el Maíz en períodos normales/húmedos no logró un buen ajuste entre los valores observados y los predichos por el modelo utilizando una base de control de 7323 datos (píxeles).

Se observa en la distribución de puntos del gráfico de regresión (Figura 78) que los datos ajustan muy bien entre valores de rendimiento de 50 a 120 qq/ha (rindes normales para la zona de estudio) pero, por fuera del rango, el modelo subestima o sobreestima.

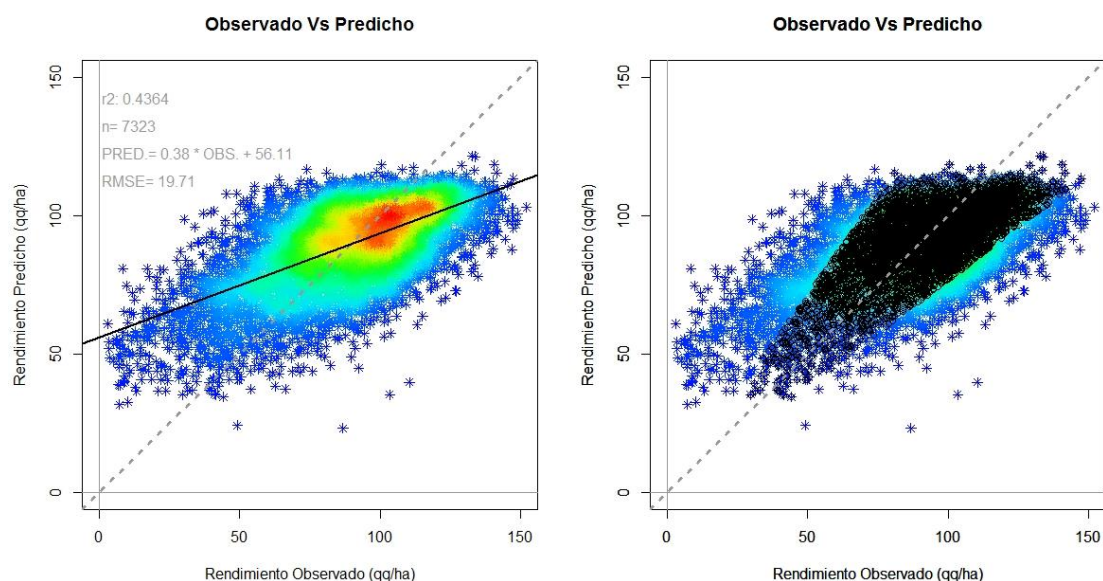


Figura 78 - Observado Vs Predicho. Maíz campaña 2010/2011. Datos de regresión insertos en el gráfico. Izq.: Diagrama de dispersión (rojo mayor densidad, azul menor densidad) con recta de regresión (línea negra) del ajuste de los datos y datos de la regresión insertos en el gráfico. Der.: Ídem Izq. con identificación de los datos con menos del 30% de error absoluto (P30) entre el valor observado y el predicho (puntos negros).

3.5.7.1 Análisis del error del modelo Maíz – 2010/2011.

- P30 = 80.40 %
- MBE = -0.34 qq/ha
- RMSE = 19.71 qq/ha
- RE = 22.5 %
- d = 0.9999501

Valores de los residuos expresados en quintales:

- Mínimo = -72.59 qq/ha
- Máximo = 70.49 qq/ha
- Desvío estándar = 19.71 qq/ha

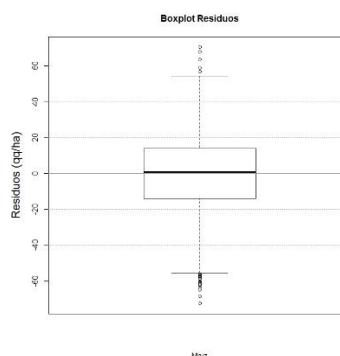


Figura 79 - Boxplot de los residuos. Maíz campaña 2010/2011.

Este modelo no logra un buen ajuste entre los datos observados y los predichos obteniendo un valor de R^2 de 0.43, sin embargo, el MBE indica una subestimación promedio del rendimiento del orden de los 34 kg/ha (muy bajo para los rindes de maíz), y el RMSE arroja un error de predicción promedio del modelo de 1971 kg/ha que representa un error relativo del 22.5%. O sea, que a pesar de no lograr un alto valor de R^2 , el error del modelo es aceptable. El valor del índice “d” indica una perfecta relación entre los datos observados y los predichos.

3.5.8 Validación Girasol – 2010/2011.

El modelo desarrollado para el Girasol en períodos normales/húmedos logró un ajuste aceptable entre los valores observados y los predichos por el modelo utilizando una base de control de 4848 datos (píxeles).

Se observa en la distribución de puntos del gráfico de regresión (Figura 80) que los datos ajustan muy bien en todo el rango de datos analizado, sin denotar en una primera instancia, sobreestimación o subestimación de rinde.

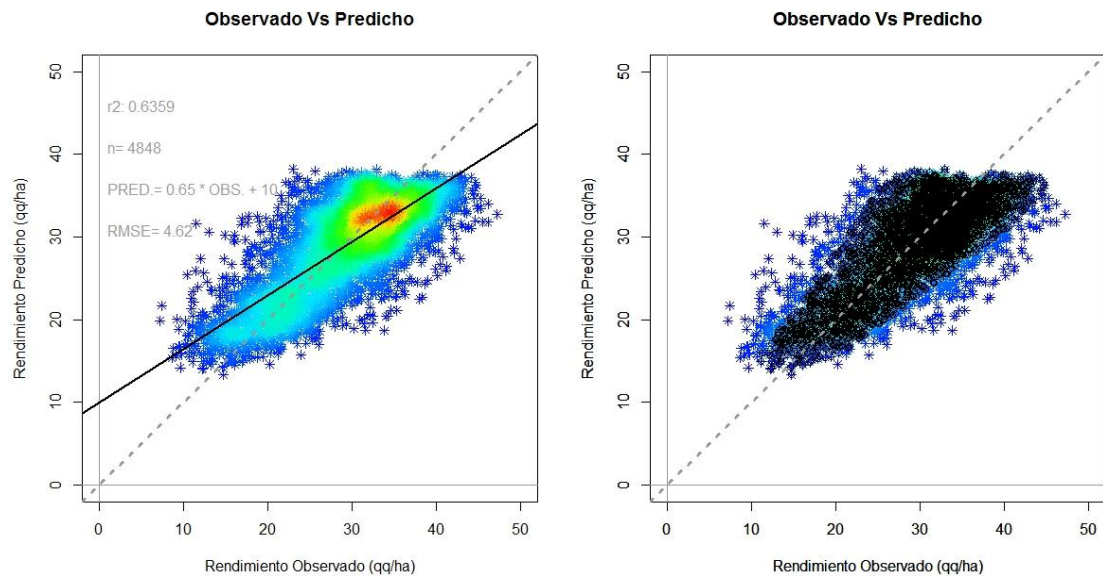


Figura 80 - Observado Vs Predicho. Girasol campaña 2010/2011. Datos de regresión insertos en el gráfico. Izq.: Diagrama de dispersión (rojo mayor densidad, azul menor densidad) con recta de regresión (línea negra) del ajuste de los datos y datos de la regresión insertos en el gráfico. Der.: Ídem Izq. con identificación de los datos con menos del 30% de error absoluto (P30) entre el valor observado y el predicho (puntos negros).

3.5.8.1 Análisis del error del modelo Girasol – 2010/2011.

- P30 = 92.3 %
- MBE = 0.19 qq/ha
- RMSE = 4.62 qq/ha
- RE = 15.87 %
- d = 0.9999586

Valores de los residuos expresados en quintales:

- Mínimo = -20.12 qq/ha
- Máximo = 15.56 qq/ha
- Desvío estándar = 4.61 qq/ha

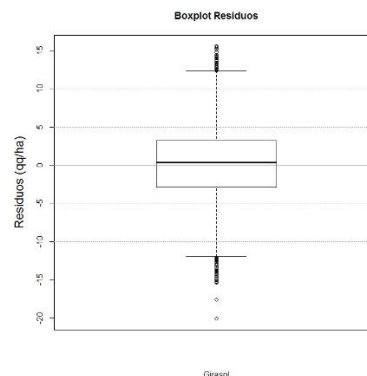


Figura 81 - Boxplot de los residuos. Girasol campaña 2010/2011.

Este modelo logra un buen ajuste entre los datos observados y los predichos obteniendo un valor de R^2 de 0.63. El MBE indica una sobrestimación promedio del rendimiento del orden de los 20 kg/ha. El RMSE arroja un error de predicción promedio del modelo de 462 kg/ha que representa un error relativo del 15.87 %. El valor del índice “d” indica una perfecta relación entre los datos observados y los predichos.

3.6 Cualificación del rendimiento estimado.

Por requerimiento de la Agencia de Recaudación (ARBA) y a través de su normativa interna, los datos de rendimiento se presentan de manera cualitativa en tres rangos de calidad: Regular, Bueno, Excelente. La estimación final (ingreso en valor monetario) que realiza otra Gerencia dentro de ARBA, se obtiene a partir de datos de rendimiento medio asignados a cada una de las 3 categorías de calidad y un análisis del valor monetario del tipo de grano. La nueva metodología que se presenta en este trabajo de investigación obtiene el dato de rendimiento por tipo de cultivo y a nivel de pixel (mínima unidad de trabajo a nivel procesamiento/obtención del dato), luego este dato es promediado a nivel de parcela (mínima unidad de trabajo a nivel fiscalización). Por lo tanto, los rendimientos a nivel pixel obtenidos en quintales por hectárea (qq/ha) serán asignados a las categorías de calidad que ya cuentan con un valor de rendimiento medio. Luego, a todos los datos (píxeles) incluidos dentro de una determinada parcela se les realizará el promedio ponderado (por calidad) para obtener el rendimiento medio en la parcela y el valor de la producción (rinde medio x superficie) por cada tipo de cultivo implantando en la parcela. Este paso final en el que interviene la intersección con el parcelario rural y los cálculos respectivos en forma masiva, se realiza en la Subgerencia de Tecnología Geoespacial de ARBA.

Es importante remarcar que la metodología propuesta en este trabajo permite estimar el rendimiento del cultivo a nivel pixel (30 m x 30 m) y con datos para la obtención de los modelos que provienen de monitores de cosecha (datos relevados aproximadamente en una celda de 7 m x 5 m). Esta resolución espacial definida para el trabajo tuvo como objetivo hallar modelos de estimación de rendimiento lo más precisos posible, a pesar de que finalmente los datos se agrupan en rangos de calidad. En el caso de que existan reclamos de los contribuyentes en relación a las estimaciones realizadas por la Agencia, será posible informar el rendimiento estimado en la parcela por tipo de cultivo y a escala de pixel.

Los rangos de rendimiento para estratificar los rangos de calidad que utiliza DTI-ARBA se definieron en función de las estimaciones agrícolas oficiales del Ministerio de Agricultura de la Nación³³. Para esto se calculó el valor medio de rendimiento histórico³⁴ para cada tipo de cultivo y el desvío estándar para luego estratificar en 3 rangos (Tabla 10). El valor de rinde promedio del tipo de cultivo en la provincia se asigna al rango de calidad “Bueno”, la clase superior e inferior de calidad se define por la suma y resta del valor de un desvío estándar respectivamente. De esta manera se definieron los rindes regular, bueno y excelente, y a partir de éstos se definió el rango de valores que abarca cada clase a partir del valor medio entre cada rinde de calidad.

³³ <http://datos.estimaciones.magyp.gob.ar/reportes.php?reporte=Estimaciones>

³⁴ Análisis desde la campaña agrícola 2000/2001 hasta la campaña 2016/2017 (datos oficiales).

		REGULAR		BUENO		EXCELENTE	
SOJA	RINDE	1550		2500		3450	
	RANGO	1100	2000	2000	3000	3000	3900
MAIZ	RINDE	4750		6850		8950	
	RANGO	3700	5800	5800	7900	7900	10000
GIRASOL	RINDE	1900		2350		2850	
	RANGO	1700	2100	2100	2600	2600	3100

Tabla 10 - Rangos de rendimiento utilizados en DTI-ARBA y calculados a partir de los rendimientos medios históricos de cada tipo de cultivo.

Los datos estimados a través de los modelos se clasificaron en las tres clases de calidad de rinde por tipo de cultivo. De esta manera, al clasificar el dato observado y el dato predicho al ejecutar los modelos desarrollados, es esperable que surjan 6 tipos de errores:

- Error de tipo 1: Se observa un rinde Regular y se estima un rinde Bueno.
- Error de tipo 2: Se observa un rinde Bueno y se estima un rinde Excelente.
- Error de tipo 3: Se observa un rinde Regular y se estima un rinde Excelente.
- Error de tipo 4: Se observa un rinde Bueno y se estima un rinde Regular.
- Error de tipo 5: Se observa un rinde Excelente y se estima un rinde Bueno.
- Error de tipo 6: Se observa un rinde Excelente y se estima un rinde Regular.

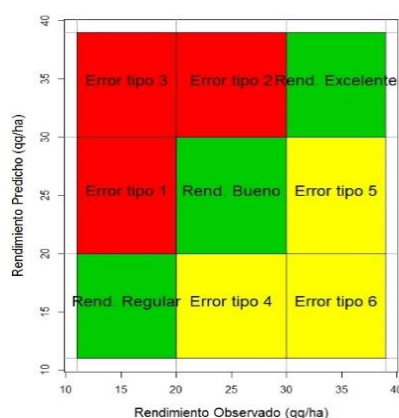


Figura 82 - Tipo de errores en la cualificación del rendimiento con rangos de rendimiento para la soja.

Los errores 1, 2 y 3 corresponden a los casos de sobreestimación del rendimiento (color rojo), mientras que los errores del tipo 4, 5 y 6 se corresponden con errores de subestimación (color amarillo) (Figura 82).

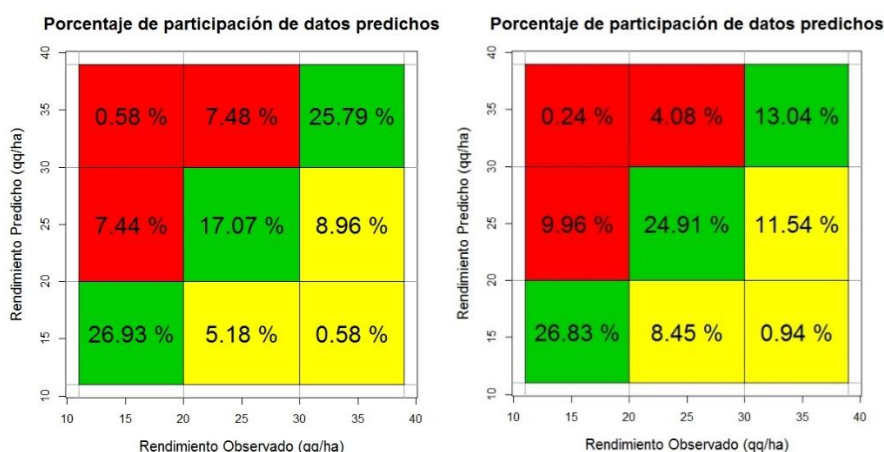


Figura 83 - Porcentaje de participación de los datos observados vs predichos cualificados en cada tipo de error. Izq.: soja 1° 2008/2009. Der.: soja 1° 2010/2011.

En la Figura 83 se observa el porcentaje de participación de los datos de Soja de primera para la campaña agrícola 2008/2009 (izquierda). Los datos categorizados de manera correcta logran el 69,79%, los errores de subestimación un 14,72% y los de sobreestimación llegan al

15,5%. En la campaña agrícola 2010/2011 (derecha) los datos categorizados de manera correcta logran el 64,78%, los errores de subestimación un 20,93% y los de sobreestimación llegan a 14,28%.

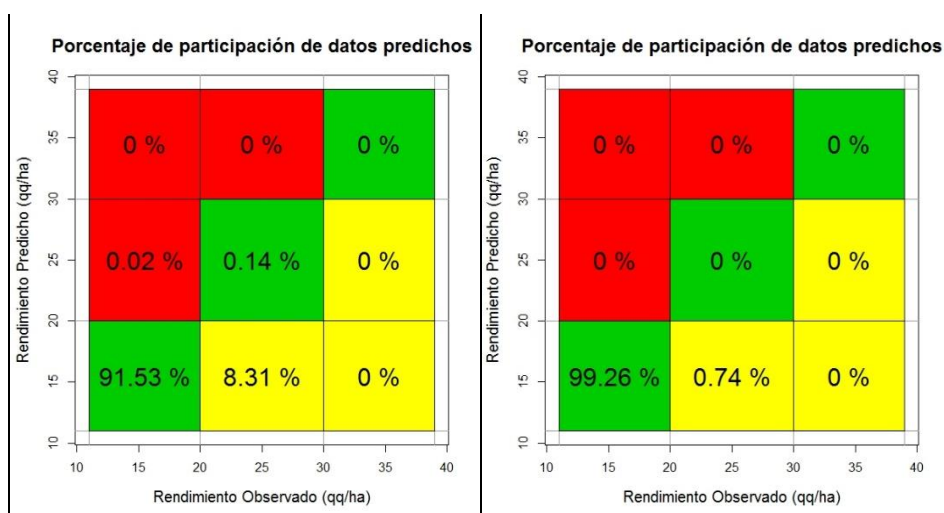


Figura 84 - Porcentaje de participación de los datos observados vs predichos cualificados en cada tipo de error. Izq.: soja 2° 2008/2009. Der.: soja 2° 2010/2011.

En la Figura 84 se observa el porcentaje de participación de los datos de Soja de segunda para la campaña agrícola 2008/2009 (izquierda). Los rangos de calidad fueron establecidos para la Soja en general (soja 1 y soja 2), por lo tanto, la Soja de segunda a obtener menores rindes, concentra sus datos en el rango de calidad inferior denominado regular. Los datos categorizados de manera correcta logran el 91,67%, los errores de subestimación un 8,13% y los de sobreestimación 0,02%. En la campaña agrícola 2010/2011 (derecha) los datos categorizados de manera correcta logran el 99,26%, los errores de subestimación un 0,74% y no hay presencia de datos sobreestimados.

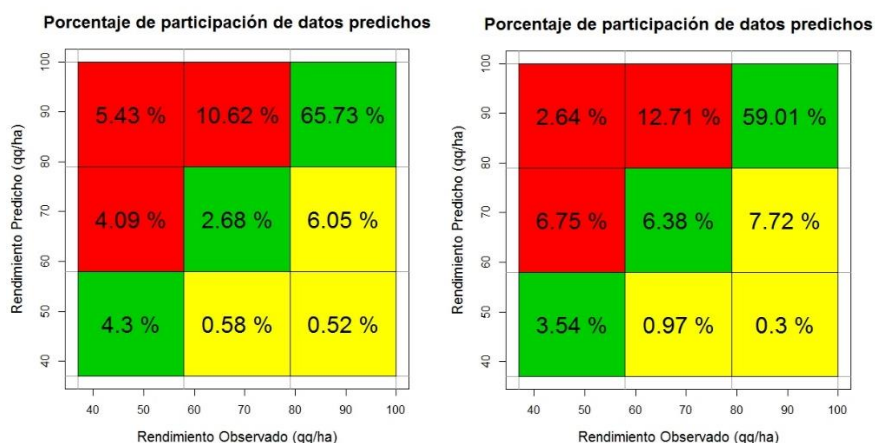


Figura 85 - Porcentaje de participación de los datos observados vs predichos cualificados en cada tipo de error. Izq.: Maíz 2008/2009. Der.: Maíz 2010/2011.

En la Figura 85 se observa el porcentaje de participación de los datos de Maíz para la campaña agrícola 2008/2009 (izquierda). Los datos categorizados de manera correcta logran el 72,71%, los errores de subestimación un 7,15% y los de sobreestimación alcanzan el 20,14%. En la campaña agrícola 2010/2011 (derecha) los datos categorizados de manera correcta logran el

68,93%, los errores de subestimación un 8,99% y los de sobreestimación alcanzan el 22,1%. Los modelos desarrollados para Maíz tienden a la sobreestimación de los resultados como ya se expuso en el capítulo de validación del modelo.

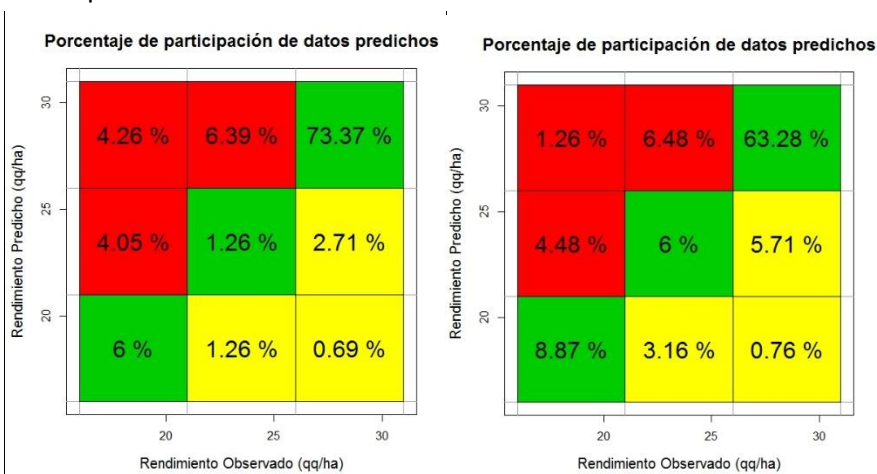


Figura 86 - Porcentaje de participación de los datos observados vs predichos cualificados en cada tipo de error. Izq.: Girasol 2008/2009. Der.: Girasol 2010/2011.

En la Figura 86 se observa el porcentaje de participación de los datos de Girasol para la campaña agrícola 2008/2009 (izquierda). Los datos categorizados de manera correcta logran el 80,63%, los errores de subestimación un 4,66% y los de sobreestimación alcanzan el 14,7%. En la campaña agrícola 2010/2011 (derecha) los datos categorizados de manera correcta logran el 78,15%, los errores de subestimación un 9.63% y los de sobreestimación llegan al 12,22%.

CAPITULO IV - Implementación.

4.1 Utilización de los modelos de ciclo hídrico normal/húmedo en 2016/2017.

Para determinar qué tipos de modelos aplicar en la campaña de implementación 2016/2017 se analizó el registro de lluvias correspondiente a la campaña agrícola (01/07/2016 al 30/06/2017) en la zona de Daireaux. Debido a que las series temporales de registros pluviométricos no se encuentran completas para el año 2016 y 2017 en las localidades seleccionadas en el capítulo 1 (Pehuajó, Coronel Suarez y Olavarría), se utilizó información no oficial proveniente de una estación meteorológica automática del partido de Daireaux [60]. En los datos (Tabla 11) se evidencia un régimen húmedo de 935.7 mm que supera los valores medios (800 mm) con 187 mm de precipitación previo a la siembra de los cultivos de verano y un acumulado de 508 mm de noviembre 2016 a marzo 2017. Los altos valores de precipitaciones acumulados durante el año 2017 sumado a altos valores de precipitación en julio, agosto y septiembre del año 2017 produjeron una gran inundación durante el invierno en la zona de Daireaux. Los modelos a utilizar en la campaña 2016/2017 serán los propuestos para ciclo hídrico normal/húmedo elaborados con datos del 2010/2011.

Año	2016						2017						Total
	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.	Ene.	Feb.	Mar.	Abr.	May.	Jun.	
Precipitación	33,6	0	36	151,2	67,2	45	81	213,3	102,3	111,6	43,8	50,7	935,7

Tabla 11 - Precipitación en milímetros. Campaña 2016/2017. Partido de Daireaux.

4.2 Desarrollo del procedimiento para la implementación.

Una vez obtenidos y validados los modelos de rendimiento, se procedió a documentar el método empleado (procedimiento) en formato de flujograma. Luego, utilizando los sensores Landsat disponibles actualmente (L7 y L8), se aplicaron los modelos en datos provenientes de series temporales de imágenes satelitales de la campaña agrícola 2016-2017. La campaña agrícola en cuestión, se corresponde con un régimen hídrico húmedo por lo que se aplicaron los modelos desarrollados en la campaña 2010/2011.

Los datos para alimentar el modelo surgen de la secuencia multitemporal del índice de vegetación NDVI desde inicios de octubre a fines de mayo, utilizando un total de 16 imágenes (ver Tabla 6 “set de imágenes”) entre Landsat 7 y Landsat 8. Las imágenes se descargaron y procesaron de la misma manera de que se procesaron las imágenes para el desarrollo de los modelos (ver capítulo 2.3).

La identificación del tipo de cultivo se corresponde con la clasificación de cultivos que realizó DTI-ARBA para la campaña “Verano 2016-2017”. La clasificación de cultivos se realiza de modo supervisado utilizando el método de máxima verosimilitud. El set de datos de campo se recolecta en cada campaña agrícola para los cultivos agrícolas extensivos (Soja1, Soja2, Maíz y Girasol) y para las clases pastura y pastizal; y luego se completa con un set de datos generados en gabinete que contiene las clases urbano, agua y monte. Cada categoría se divide al 50% para conservar un set de datos para el control de la clasificación. Los datos de la clasificación de cultivos de la campaña gruesa 2016/2017 se detallan en el Anexo III.

Utilizando el software ENVI se vincularon los polígonos (shape) por tipo de cultivo al set de imágenes de los índices de vegetación acotada únicamente al partido de Daireaux, y se exportaron los datos multitemporales en formato ASCII para luego trabajarlos en el software R-Studio. En este proceso de validación, los datos exportados en cantidad de píxeles por tipo de cultivo fueron: 679.031 píxeles para Soja de primera, 275.828 píxeles para Soja de segunda,

173.881 píxeles para Maíz y 152.582 para Girasol. A pesar de la gran cantidad de datos a procesar, el procedimiento de generación de la interpolación de los datos y la obtención de los parámetros para la aplicación de los modelos se realizaron de manera normal con el script desarrollado.

Las curvas de crecimiento de cada uno de los cultivos con los datos 2016-2017 se presentan en las Figuras 87 y 88, e indican el momento (día del año) en el que se extrajeron los valores para aplicar en el modelo (línea vertical roja). A cada conjunto de datos discriminado por tipo de cultivo (datos multitemporales interpolados) se le aplicó la función (modelo) correspondiente al cultivo (ver Tabla 12) para la estimación del rendimiento predicho a nivel de píxel. El rendimiento se obtuvo en quintales por hectárea (qq/ha), y posteriormente se clasificó en los rangos de calidad propuestos (Excelente, Bueno, Regular). Luego se visualizaron los resultados en software Q-Gis con el objetivo de corroborar y controlar los resultados obtenidos con las variaciones espaciales observadas en las imágenes satelitales.

Ciclo	Cultivo	Modelo	DPG-P
Normal/Húmedo	Soja 1°	$YS1 = 1,584 e^{(3,5 * IVS1)}$	46
	Soja 2°	$YS2 = 2,181 e^{(2,39 * IVS2)}$	82
	Maíz	$YM = 4,618 e^{(3,88 * IVM)}$	52
	Girasol	$YG = 4,05 e^{(2,62 * IVG)}$	51

Tabla 12 - Modelos generados para ciclos hídricos Normales/Húmedos.

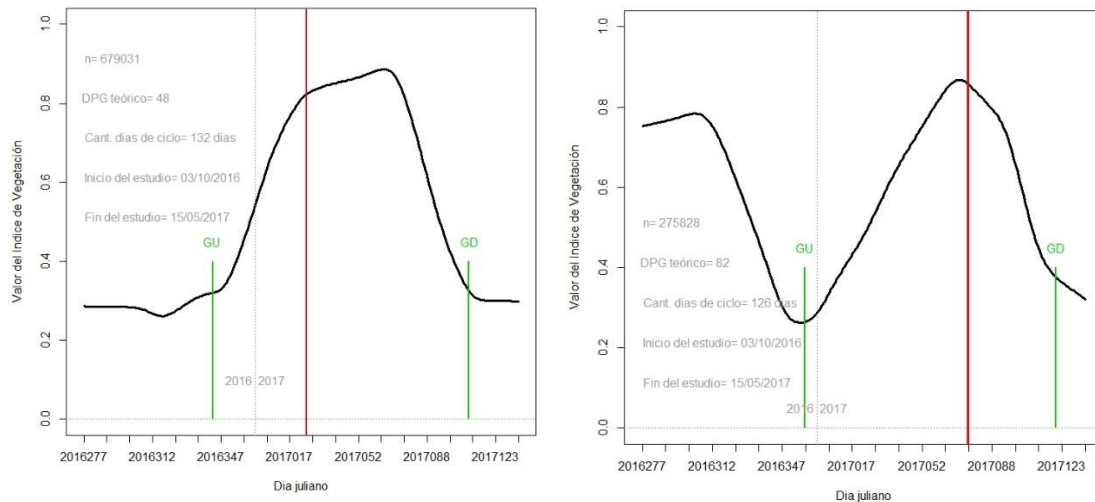


Figura 87 - Curvas de crecimiento 2016/2017 con indicación de Green-Up y Green-Down (barras verticales verdes), y el día de extracción de los datos para el modelo (barra vertical roja). Izq.: Soja 1°; Der.: soja 2°.

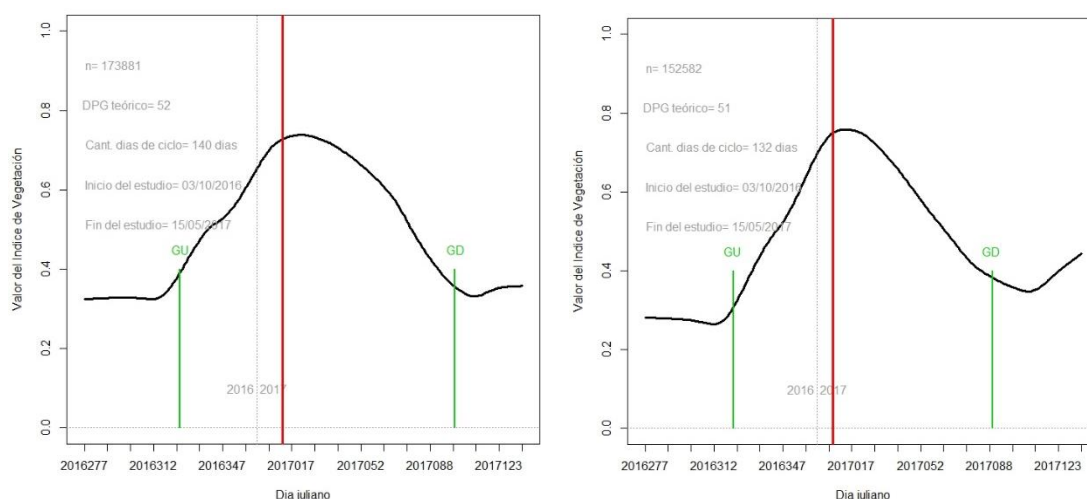


Figura 88 - Curvas de crecimiento 2016/2017 con indicación de Green-Up y Green-Down (barras verticales verdes), y el día de extracción de los datos para el modelo (barra vertical roja). Izq.: Maíz; Der.: Girasol.

El resultado de todos los modelos de estimación de rendimiento absoluto (qq/ha)³⁵ y el resultado de la clasificación en rangos de calidad (excelente, bueno, regular) se guardaron junto a la clasificación de coberturas (sólo los cultivos³⁶), en formato raster (imagen de 3 bandas). Es importante aclarar que estas capas de información no deben desvincularse, debido a que el rendimiento y la calidad fueron calculados en base al tipo de cultivo que se encuentra en la banda de clasificación de coberturas.

La imagen generada, posteriormente se vectorizó³⁷ y se obtuvieron 3 capas vectoriales: A_tipo de cultivo, B_rendimiento y C_calidad. Luego se unieron A y C creando la capa “cultivos x calidad” (unión vectorial de la clase de cultivo con su correspondiente calidad). Esta capa vectorial generada “cultivos x calidad” es el resultado final del proceso de producción de DTI, y posteriormente, la subgerencia de Tecnología Geoespacial realiza la intersección de estos datos con el parcelario rural y realiza los cálculos necesarios para obtener los valores de producción en pesos por parcela.

Una vez obtenido el archivo vectorial final se visualizó en el software Q-Gis para observar su comportamiento espacial y realizar exportaciones de mapas. En la Figura 89 se observan los rendimientos en sus tres tipos de calidad (Excelente: verde, Bueno: amarillo, Regular: naranja) de los 4 cultivos simultáneamente, y en las Figuras 90 y 91 se observa un detalle (zoom) que discrimina las calidades del rendimiento predicho por tipo de cultivo.

³⁵ El rendimiento absoluto se conserva con objetivos de control interno ante algún descargo que puedan realizar los contribuyentes a la Agencia.

³⁶ La clasificación incluye todo tipo de coberturas, pero se enmascara y se exportan únicamente los cultivos (Soja1, Soja2 Maíz y Girasol).

³⁷ En situación de producción el paquete raster generado, que incluye todo el path-row de Landsat, ingresa al procesamiento vectorial final que actualmente realiza el Departamento GIS de ARBA



Figura 89 – Mapa de rendimiento obtenido para los 4 tipos de cultivo (campaña 2016/2017).



Figura 90 – Lotes ubicados en un mismo sitio geográfico con su respectiva calidad (Excelente: verde, Bueno: amarillo, Regular: naranja). Izk.: Lotes de Soja 1°. Der.: Lotes de Soja 2°.



Figura 91 – Lotes ubicados en un mismo sitio geográfico con su respectiva calidad (Excelente: verde, Bueno: amarillo, Regular: naranja). Izk.: Lotes de Maíz. Der.: Lotes de Girasol.

Los resultados predichos en valores de rendimiento (qq/ha), en rangos de calidad y a escala pixel (datos por fila) se presentan a modo de ejemplo en las siguientes líneas (extracción de los primeros datos “cabeza de tabla” por cada tipo de cultivo):

> head(soja1.predicho)

	X	Y	Lat	Lon	REND.PREDICHO	CAL.PREDICHA
1	5307235	5987008	-36.24563	-62.14446	19.42	15.5
2	5307205	5987008	-36.24563	-62.14480	25.25	25.0
3	5307205	5986978	-36.24590	-62.14480	37.07	34.5
4	5307235	5986978	-36.24590	-62.14447	24.82	25.0
5	5307175	5986978	-36.24589	-62.14514	36.81	34.5
6	5307145	5986948	-36.24616	-62.14548	36.93	34.5

> head(Soja2.predicho)

	X	Y	Lat	Lon	REND.PREDICHO	CAL.PREDICHA
1	5308645	5985538	-36.25915	-62.12915	9.68	15.5
2	5308615	5985508	-36.25942	-62.12949	11.90	15.5
3	5308645	5985508	-36.25942	-62.12915	11.03	15.5
4	5308675	5985508	-36.25943	-62.12882	10.87	15.5
5	5308705	5985508	-36.25943	-62.12849	8.23	15.5
6	5308585	5985508	-36.25941	-62.12982	12.81	15.5

> head(Maiz.predicho)

	X	Y	Lat	Lon	REND.PREDICHO	CAL.PREDICHA
1	5308465	5985778	-36.25695	-62.13109	106.46	89.5
2	5308285	5985778	-36.25692	-62.13309	133.38	89.5
3	5308435	5985778	-36.25695	-62.13142	65.19	68.5
4	5308315	5985778	-36.25693	-62.13276	122.61	89.5
5	5308345	5985778	-36.25693	-62.13242	117.21	89.5
6	5308255	5985748	-36.25718	-62.13343	138.49	89.5

> head(Girasol.predicho)

	X	Y	Lat	Lon	REND.PREDICHO	CAL.PREDICHA
1	5301805	5986438	-36.24967	-62.20498	34.76	28.5
2	5301775	5986438	-36.24966	-62.20531	36.12	28.5
3	5301775	5986408	-36.24993	-62.20532	36.69	28.5
4	5301805	5986408	-36.24994	-62.20498	36.60	28.5
5	5301835	5986408	-36.24994	-62.20465	34.49	28.5
6	5301865	5986408	-36.24995	-62.20432	32.97	28.5

Con el fin de presentar la información generada en la implementación de los modelos desarrollados, se realizó la intersección del vectorial “Cultivos x Calidad” con el parcelario, recortando el vectorial generado (descartando sitios de rutas y calles, parcelas inconsistentes, parcelas menores a 5 has y parcelas urbanas) y conservando los lotes cultivados (de los diferentes cultivos) dentro de cada parcela rural consistente, generando la capa D. En la tabla de atributos de la capa D se agregó una columna de superficie y se calculó la superficie de cada polígono de cultivo y de calidad. Los valores de producción estimada (Toneladas) por parcela para cada tipo de cultivo se calculó multiplicando la superficie total de cada tipo de calidad por el rendimiento medio de cada calidad (ver Tabla 10).

El rendimiento absoluto (sin cualificar) se obtuvo vinculando la capa de cultivos (A) intersectada por el parcelario rural y la capa de rendimiento (B), esta información queda archivada en los servidores internos (no se publica).

Como ya se aclaró anteriormente, el resultado final de DTI y consecuentemente del trabajo de investigación es la capa “Cultivos x Calidad”, sin embargo, se realizó el proceso sucesivo correspondiente a la subgerencia de Tecnología Geoespacial para obtener el rendimiento por parcela en el partido de Daireaux (Figura 92). Este producto que tiene como base el parcelario rural, contiene en los atributos los siguientes campos: tipo de cultivo,

rendimiento, calidad, superficie y producción de cada parcela. La superficie es calculada a partir del área vectorial del cultivo en la parcela y el dato de producción se calcula como:

$$PC = \sum_{[E,B,R]} (RC_{[E,B,R]} \times SC_{[E,B,R]})$$

PC: Producción por tipo de cultivo.

RC: Rendimiento discriminado en calidades.

SC: Superficie de cada tipo de calidad.

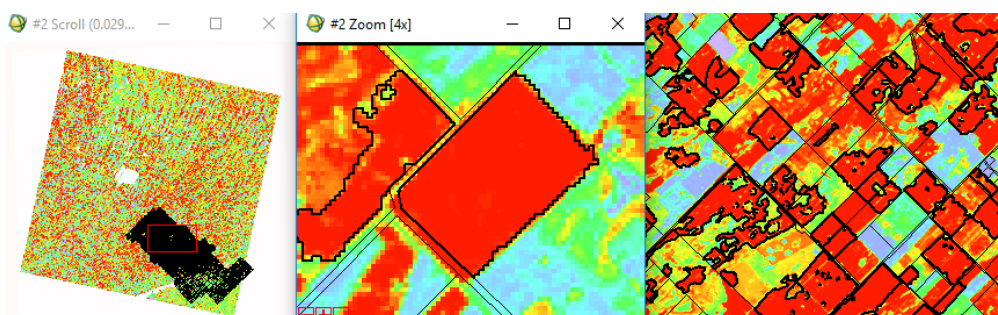


Figura 92 - Imagen NDVI (Path 227, Row 085) con el shape de clasificación del cultivo de Soja de primera en campaña 2016-2017 recortado por el parcelario rural (Escala de color: "rojo" alto NDVI – "celeste" bajo NDVI).

El partido de Daireaux tiene 2714 parcelas rurales registradas sin inconsistencias en el parcelario actualizado de ARBA. De este total, se detectó cultivo agrícola en 1796 (66,17 %), discriminados en 555 parcelas de Girasol, 894 parcelas de Maíz, 1213 parcelas de Soja de primera, y 868 parcelas de Soja de segunda. La sumatoria de las parcelas discriminadas por cultivo suma un total de 3530, con lo cual, se puede presumir que la gran mayoría de las parcelas agrícolas cultivadas produce al menos 2 tipos diferentes de cultivo estival en sus lotes cultivados.

CAPITULO V - Resultados y Conclusiones.

5.1 Resultados de la generación de las funciones de crecimiento.

Es este estudio se utilizaron diferentes variantes de los índices espectrales: NDVI, EVI, MSAVI y NDMI, y el comportamiento de los mismos fue similar, destacándose el NDVI a pesar de que a este índice se le atribuye el efecto de “saturación” [61] [62]. El efecto de saturación sucede en momentos en los que la biomasa del cultivo es abundante y con alto vigor, pero en nuestro caso, el NDVI fue el índice de vegetación que mejor desempeño tuvo en este estudio (a excepción del Girasol en ciclo seco donde el que el NDMI al ser un índice de humedad, logra un mejor ajuste). Por otro lado, se observó que los índices tuvieron un comportamiento temporal diferente en el ciclo seco 2008/2009 y en el ciclo normal/húmedo 2010/2011, y posiblemente lo tenga en cada campaña agrícola, ya que las diferentes variables ambientales y técnico-productivas que hacen al sector agrícola y su toma de decisiones, cambian en cada ciclo productivo.

El vigor vegetal de los cultivos en un año normal/húmedo debería ser superior al de una campaña con efecto de estrés hídrico, por lo tanto, los índices obtienen valores máximos más elevados acompañando el aumento de la biomasa vegetal que cubre completamente el suelo, incrementa su volumen y el verdor de las hojas. Este incremento del vigor vegetal (leve aumento del valor del índice de vegetación) se reflejó únicamente en el cultivo de Soja de primera (Figura 93), y obtuvo un leve incremento del rendimiento (Tabla 13). Para el caso de la Soja de segunda, Maíz y Girasol (Figuras 94, 95 y 96), los máximos valores de NDVI registrados en 2010/2011 fueron levemente inferiores a 2008/2009. A pesar de esto, las regresiones realizadas sobre los respectivos sets de datos, indican que los índices de vegetación y el rendimiento tuvieron desempeños similares en ambas campañas en cuanto a los ajustes (R^2) y los errores observados.

Cultivo	Campaña agrícola											
	06/07	07/08	08/09	09/10	10/11	11/12	12/13	13/14	14/15	15/16	16/17	17/18
Soja	2.259	2.586	1.389	2.714	2.205	2.820	2.311	2.356	2.556	3.327	3.210	2.284
Maíz	7.900	7.000	4.800	8.000	6.210	8.000	8.000	6.000	7.000	7.700	8.000	5.000
Girasol	2.400	2.300	2.500	2.200	2.600	2.200	2.400	2.300	2.700	3.000	2.500	2.000

Tabla 13 - Rendimientos promedios (Ton/ha) oficiales por tipo de cultivo y campaña agrícola.

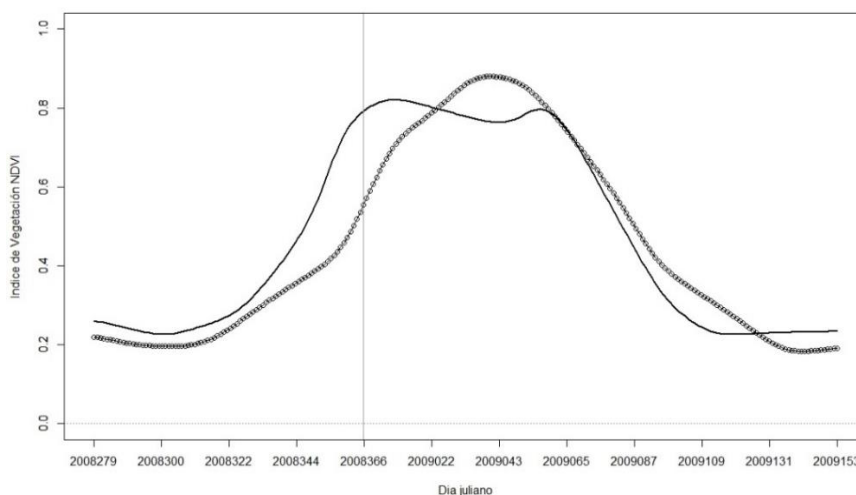


Figura 93 - Comparación de ciclos de Soja de primera. Línea continua 2008/2009 y línea punteada 2010/2011. Valores medios de NDVI de todo el set de datos.

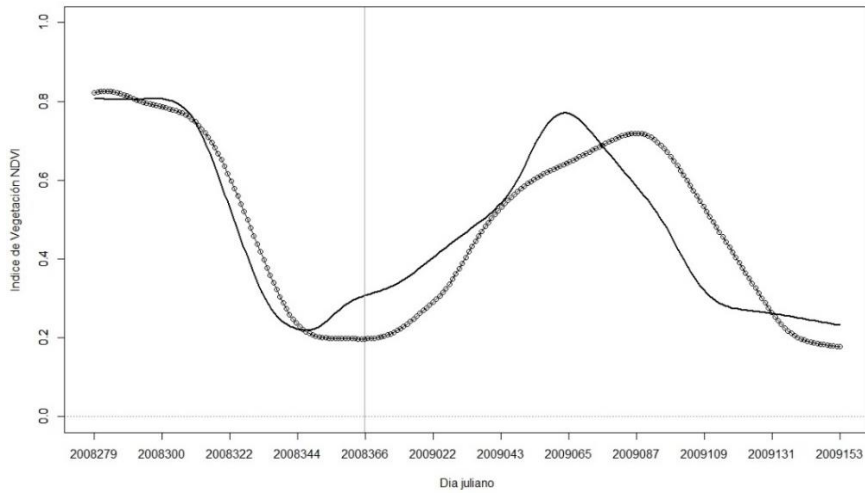


Figura 94 - Comparación de ciclos de Soja de segunda. Línea continua 2008/2009 y línea punteada 2010/2011. Valores medios de NDVI de todo el set de datos.

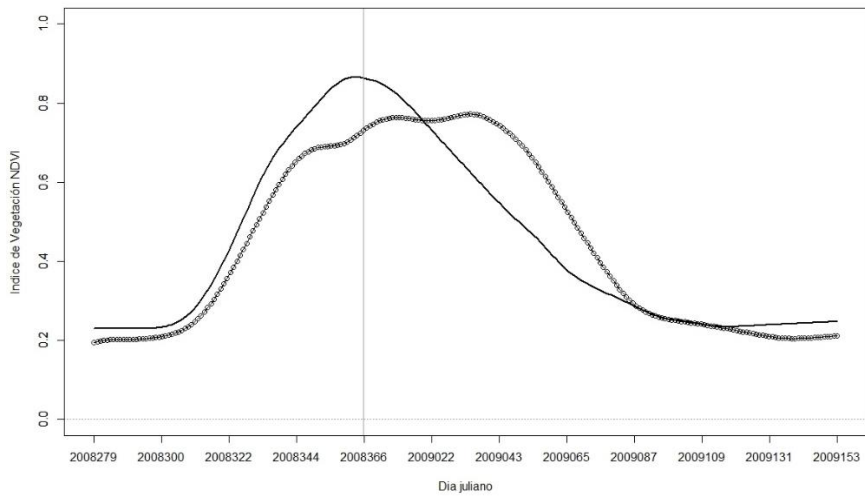


Figura 95 - Comparación de ciclos de Maíz. Línea continua 2008/2009 y línea punteada 2010/2011. Valores medios de NDVI de todo el set de datos.

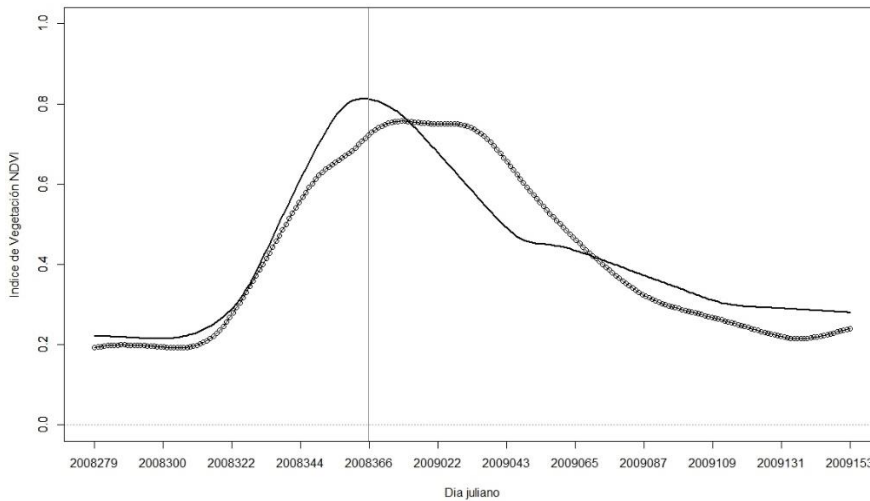


Figura 96 - Comparación de ciclos de Girasol. Línea continua 2008/2009 y línea punteada 2010/2011. Valores medios de NDVI de todo el set de datos.

5.2 Resultados de los modelos desarrollados.

El desarrollo de los modelos de estimación del rendimiento fue una tarea sumamente compleja. La diversidad de opciones que se probaron (2 campañas agrícolas, 4 tipos de cultivos, 4 índices de vegetación, 4 maneras de vincular los datos) arrojaron 128 resultados que debían ser analizados detalladamente y con criterio agronómico. Por otro lado, el hecho de tener que insertarse en una cadena de procesamiento existente (proceso de producción de DTI), condiciona el método propuesto para que pueda ser aplicado sistemáticamente en el área.

Los modelos de estimación del rendimiento para los cultivos de soja1, soja2, maíz y girasol validados en el partido de Daireaux provincia de Buenos Aires se presentan en la siguiente tabla resumen (Tabla 14):

Ciclo	Cultivo	Modelo	r ²	DPG-P	n ₁	n ₂	RMSE (qq/ha)	RE (%)	BIAS (qq/ha)	d
Seco	Soja 1°	$YS1 = e^{(5.06 * IVS1)} / 2,159$	0.7023	86	10642	10897	6,64	26,11	0,9	0,9999864
	Soja 2°	$YS2 = 1.84 e^{(2.76 * IVS2)}$	0.5401	82 DPG*	4913	5043	2,89	19,69	0,22	0,9999435
	Maíz	$YM = 3,781 e^{(4,12 * IVM)}$	0.4793	80	6100	6186	21,44	23,4	-3,3	0,9999485
	Girasol	$YG = 11,473 e^{(2,36 * IVG)}$	0.4931	49	2744	2768	6,60	20,68	-0,24	0,9998957
Normal / Húmedo	Soja 1°	$YS1 = 1,584 e^{(3,5 * IVS1)}$	0.5688	46	12100	12338	6,49	26,84	1,3	0,9999806
	Soja 2°	$YS2 = 2,181 e^{(2,39 * IVS2)}$	0.4112	82	3435	3378	2,34	19,13	-0,06	0,9998875
	Maíz	$YM = 4,618 e^{(3,88 * IVM)}$	0.3921	52	7233	7323	19,71	22,5	-0,34	0,9999501
	Girasol	$YG = 4,05 e^{(2,62 * IVG)}$	0.6126	51	4698	4848	4,62	15,87	0,19	0,9999586

Tabla 14 - Resumen de modelos desarrollados y validados. "DPG-P" refiere a la cantidad de días posteriores al Green-Up Promedio en el que se halló el mejor ajuste de los datos. "n1 y n2" refieren a la cantidad de datos analizados para modelar y para validar respectivamente. (*): Método "b": Días Posteriores al Green-up de cada pixel (no es el promedio del conjunto de datos).

En general los modelos tienen como input datos del índice de vegetación NDVI, con lo cual se simplifica el proceso posterior, sin embargo, el modelo de Girasol en ciclo hídrico seco ajustó mejor con el índice de humedad NDMI. Esto requiere que el proceso automático además de generar el cálculo del NDVI en cada una de las imágenes, se deba incluir también el cálculo del NDMI en las campañas de ciclo seco, únicamente para ser aplicado al cultivo de girasol. Se podría revisar a futuro este modelo, y modificarlo de alguna forma para que logre un mayor desempeño a través del NDVI y de esta manera simplificar los procesos posteriores.

Los errores porcentuales de los modelos desarrollados en ambas campañas promedian el 22%. Los modelos desarrollados para la Soja de primera en general subestiman levemente el rendimiento estimado, los modelos de Soja de segunda y Girasol son relativamente estables, y en los modelos de Maíz se sobreestima el rendimiento. (Capítulo 3.5: Validación de los modelos seleccionados)

En cuanto a los momentos en los cuales se obtiene el mayor ajuste de los datos del índice de vegetación con el rendimiento de los cultivos, en la campaña de ciclo de sequía se promedian 82.5 días desde la fecha de Green-Up Promedio salvo para el girasol que sucede a los 49 días. En la campaña normal/húmeda sucede en promedio a los 50 días desde la fecha estimada de Green-Up Promedio a excepción de la Soja de segunda que se extiende hasta los 82 días. Estas fechas en general coincidieron con los rangos estimados de período crítico para cada tipo de cultivo, con excepción de la Soja de primera en el ciclo normal/húmedo que tuvo su mejor ajuste del modelo 18 días previo al inicio del período crítico (Tabla 15).

	Modelo Sequía		¿Dentro del período Crítico?	Modelo Normal/húmedo		¿Dentro del período Crítico?
	Día de > r ²	Rango PC		Día de > r ²	Rango PC	
Soja 1°	17/02	28/01 – 29/03	SI	13/01	31/01 – 01/04	NO
Soja 2°	4/03	02/02 – 24/03	SI	25/03	10/02 – 01/04	SI
Maíz	18/01	16/12 – 25/01	SI	17/01	21/12 – 30/01	SI
Girasol	06/01	08/12 – 27/01	SI	01/01	04/12 – 23/01	SI

Tabla 15 – Día de mayor ajuste y verificación del rango del Período Crítico.

Los ajustes (R²) en general obtienen valores medios (promedio 0.5) con lo cual, se puede inferir que los modelos desarrollados predicen con una única variable de detección remota aproximadamente el 50% de la variabilidad total que determina el rendimiento de un cultivo. Sin embargo, teniendo en cuenta la complejidad del sistema agrícola, lograr predicciones con errores dentro de niveles aceptables, es un muy buen resultado obtenido.

5.3 Resultados de la implementación.

Luego de procesar los datos satelitales para la campaña agrícola de implementación 2016/2017, en el partido de Daireaux, se aplicó el script de procesamiento asignando a cada tipo de cultivo el modelo de rendimiento desarrollado. El resultado obtenido son 3 capas raster con los siguientes datos: tipo de cultivo, rendimiento (qq/ha) y calidad (excelente, bueno, regular). Posteriormente, estos datos se vectorizaron obteniendo en la tabla de atributos de cada polígono los siguientes campos: ID, tipo de cultivo (soja1, soja2, maíz, girasol), rendimiento (qq/ha) y superficie (ha).

Predicción del rendimiento:

Cultivo	Min.	1°Q.	Mediana	Media	3°Q.	Max.	sd	%Reg.	%Bue.	%Exc.
Soja 1°	0.59	27.68	34.42	31.83	37.66	50.26	7.36	9.53	22.23	68.23
Soja 2°	2.70	12.02	15.32	14.62	17.44	20.67	3.29	99.94	0.05	0.00
Maíz	8.68	61.91	82.70	81.41	102.17	197.14	28.30	21.60	23.05	55.33
Girasol	8.06	26.19	32.85	30.44	36.26	45.24	7.45	14.81	9.74	75.44

Tabla 16 – Estadísticas del rendimiento predicho por los modelos en 2016/2017 (qq/ha) y % de cada tipo de calidad.

La predicción del rendimiento logró resultados esperados y adecuados para cada tipo de cultivo. Es notable un valor máximo atípico para el cultivo de maíz para la zona de estudio (19.714 kg/ha) que puede surgir de ingresar al modelo el valor de un pixel que no se corresponde con el cultivo. Sin embargo, los valores medios, mediana y desvío son coherentes con lo esperado para dicho cultivo en la zona. Es importante resaltar la importancia de la cualificación del rendimiento, ya que, de esta manera, este tipo de errores son rectificadas al ingresar al rango de calidad que le corresponde.

Los rangos de calidad establecidos por tipo de cultivo (Regular, Bueno, Excelente) son generales para toda la provincia, por lo tanto, para los datos obtenidos en Daireaux hay rangos que no se obtienen (bueno y excelente en Soja de segunda). Esto podría mejorarse incluyendo un mayor número de rangos de calidad.

Los datos de producción obtenidos a partir de este procesamiento en la campaña 2016/2017 (Tabla 17) se compararon con los datos publicados oficialmente [4] en la misma campaña agrícola (Tabla 18), que es la única fuente de información disponible, aunque de baja veracidad (no cuenta con la estimación del desvío), debido a que no disponemos de datos de campo. Para el Girasol se observa una subestimación del área cosechada de un 20% (dato que se calcula a partir de la clasificación de coberturas) y un valor de rendimiento estimado por el modelo es levemente superior al promedio oficial en alrededor del 6% (160 kg/ha). En cuanto al Maíz, la superficie relevada a partir de la clasificación es prácticamente la mitad de la informada oficialmente, pero esto se debe principalmente a una tendencia de los últimos años de migrar hacia cultivares de ciclo corto y siembras tardías (para este tipo de maíz no se desarrolló el

modelo por la falta de datos de este tipo en las campañas en las cuales se crearon los modelos). El rendimiento estimado para el maíz tradicional de ciclo largo es levemente inferior al rendimiento informado oficialmente en un 5.6% (442 kg/ha). Para la Soja (no se distingue la información entre Soja de primera y Soja de segunda), la superficie detectada en la clasificación es un 30% inferior y el rendimiento estimado para la Soja de primera es levemente inferior en un 5% (153 kg/ha) al informado por el organismo oficial.

CULTIVO	SUP. COSECHADA (ha)	PRODUCCIÓN (ton)	RENDIMIENTO (kg/ha)
Girasol	13.732	36.535	2.660
Maíz	15.649	118.280	7.558
Soja1 + Soja2	61.112 + 24.824	186.860 + 38.489	3.057 y 1.550

Tabla 17 - Datos estimados para la campaña agrícola 2016/2017 generados a partir del trabajo de investigación.

CULTIVO	SUP. COSECHADA (ha)	PRODUCCIÓN (ton)	RENDIMIENTO (kg/ha)
Girasol	17.000	42.500	2.500
Maíz	34.200	273.600	8.000
Soja	111.600	358.200	3.210

Tabla 18 - Datos oficiales de producción para la Campaña agrícola 2016/2017 en el partido de Daireaux.

5.4 Resultados del método propuesto.

Luego de desarrollar los modelos, validarlos e implementarlos fue posible definir el proceso de producción para incorporar la estimación del rendimiento por cada tipo de cultivo en el partido de Daireaux. Los modelos fueron creados y validados con datos ubicados físicamente dentro del partido de Daireaux, por lo tanto, estos modelos se podrán aplicar en próximas campañas agrícolas dentro de la Zona Agroeconómica Homogénea (ZAH) que incluye al partido “ZAH V J – Daireaux” y que agrega algunas circunscripciones del sur del partido de Trenque Lauquen (Figura 97). Se entiende por ZAH a áreas que por sus condiciones agronómicas y económicas pueden ser consideradas relativamente homogéneas [63]. Sin embargo, la propuesta metodológica está orientada a diseñar un procedimiento que permita aplicar modelos de estimación de rendimiento por ZAH con alcance en toda la provincia de Buenos Aires, por lo tanto, se ejemplificará el proceso completo sin hacer énfasis en el partido de Daireaux.

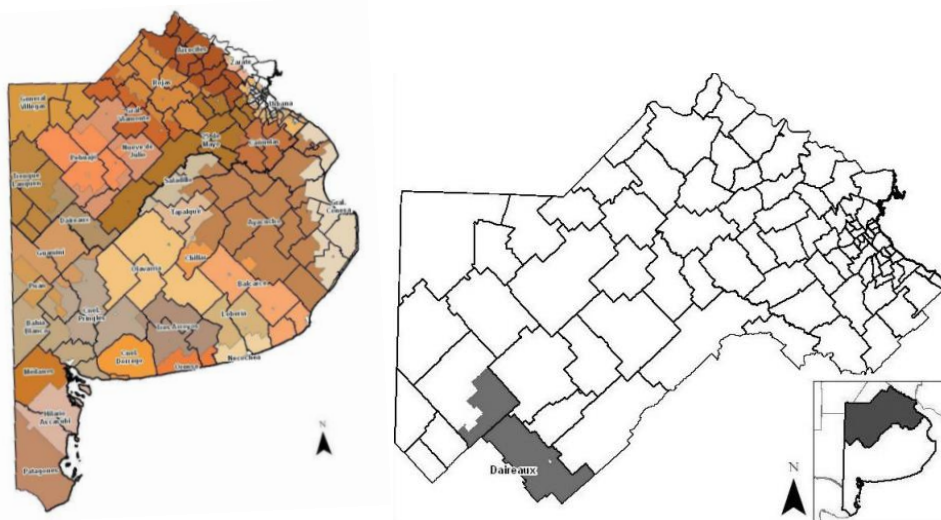


Figura 97: ZAH V J – Daireaux. Fuente: Zonas Agroeconómicas Homogéneas de Buenos Aires (INTA 2009)

El método propuesto consiste en incorporar los modelos de estimación de rendimiento en la secuencia de procesamiento de imágenes existente en DTI-ARBA. La entrada (input) al modelo es un paquete multitemporal de imágenes de NDVI creados a partir de imágenes calibradas a reflectancia, con extracción de nubes y sombras, con su correspondiente georeferenciación y reproyección e intercalación de bandas “cero”³⁸; al que se le agrega la banda de ZAH (recorte de dimensiones equivalente al Path-Row de Landsat) y la banda de Clasificación de Coberturas (CC). La salida (output) es un paquete raster de 3 bandas: B1_Banda de tipo de cultivo (Soja1, Soja2, Maíz, Girasol); B2_Banda de rendimiento estimado (qq/ha); B3_Banda de rangos de calidad (Excelente, Bueno, Regular).

Una vez confeccionado el paquete multitemporal (NDVI's + ZAH + CC), se procede a ejecutar el script de procesamiento. En primera instancia actúan dos árboles de decisión, el primero identifica la ZAH para definir qué conjunto de modelos utilizar, y el segundo discrimina por tipo de cultivo (Soja1, Soja2, Maíz, Girasol). Los datos correspondientes a cada tipo de cultivo se interpolan diariamente y se calcula la fecha promedio de Green-Up. Teniendo en cuenta el tipo de cultivo, se extraen los datos de NDVI en el día asignado por el modelo (DPG ó DPG-P). A la base de datos extraída se le aplica la fórmula correspondiente del modelo por tipo de cultivo obteniendo el rendimiento en cada pixel y en unidades de quintales por hectárea. Posteriormente se clasifica el dato de rendimiento en los 3 rangos de calidad preestablecidos

³⁸ En caso de no disponer de imagen en una determinada fecha dentro de la resolución temporal de 8 días (constelación L7 y L8), se adiciona una banda con valores cero.

(Excelente, Bueno, Regular). Finalmente se guardan los datos en formato raster generando un paquete de 3 bandas (Figura 98).

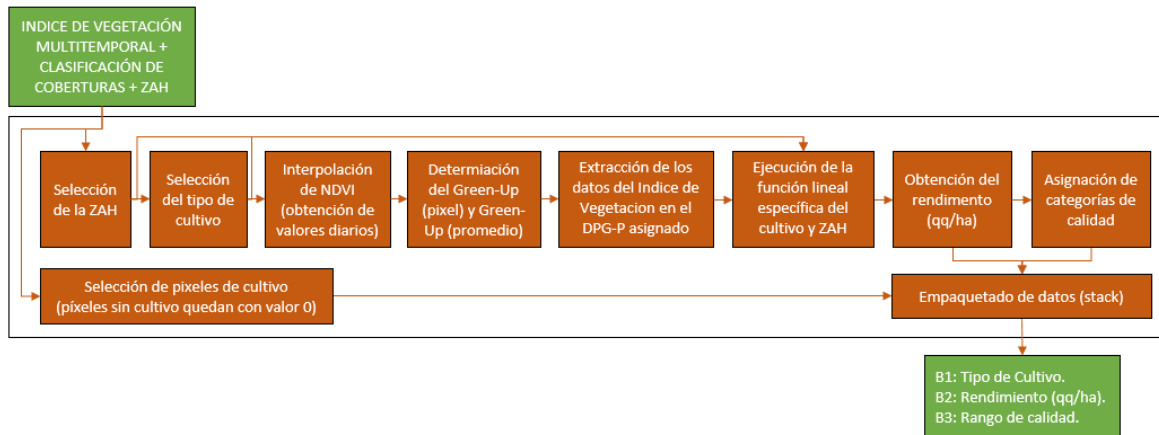


Figura 98 - Flujoograma del modelo de estimación de rendimientos propuesto.

El proceso de producción vigente se muestra en la Figura 99 discriminando en bloques de “pre-procesamiento de imágenes”, “procesamiento de imágenes”, “procesamiento vectorial inicial” y “procesamiento vectorial final”. Los cambios propuestos se insertan en el bloque de “procesamiento de imágenes” (Figura 100), y consisten básicamente en incluir los procesos indicados en la Figura 98.

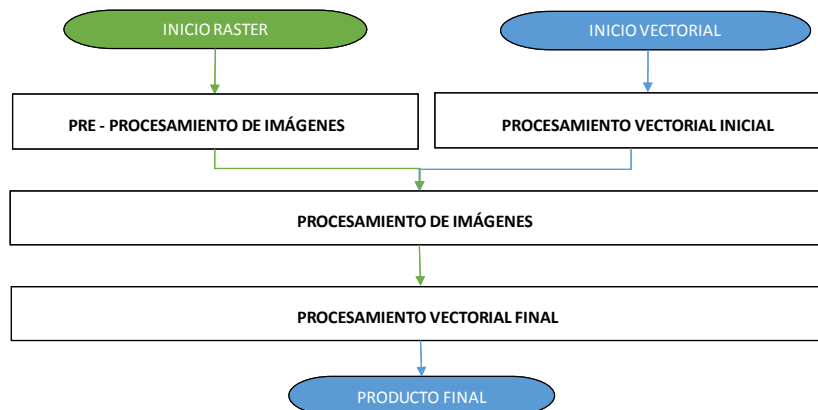


Figura 99 - Flujoograma simplificado de procesos vigente en DTI-ARBA.

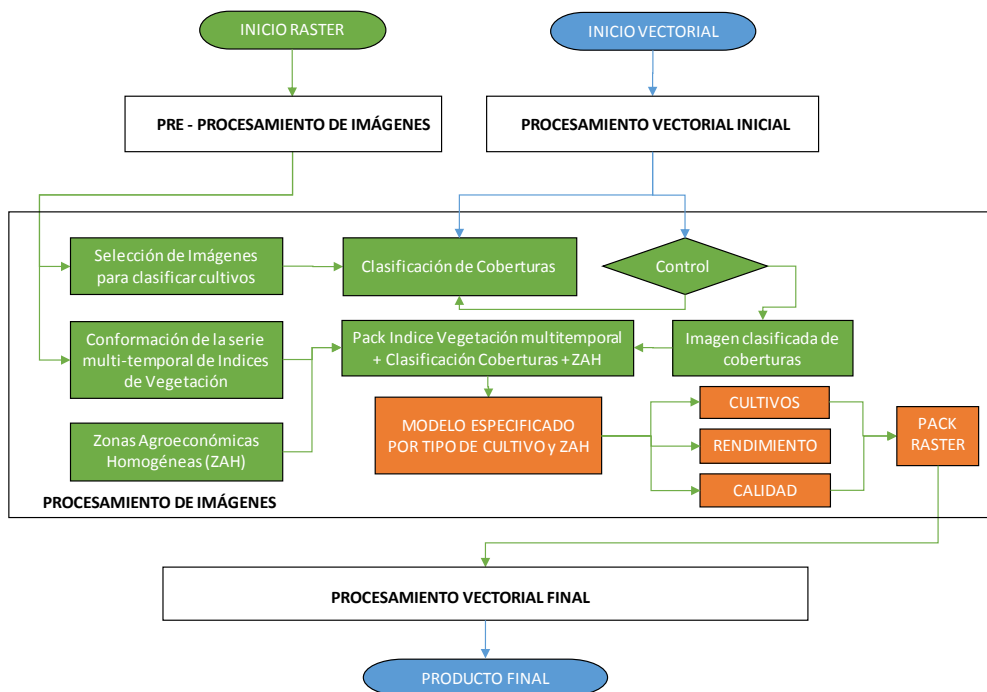


Figura 100 - Flujoograma de procesos modificado con la inclusión del modelo propuesto.

5.4.1 Instructivo de trabajo.

Uno de los objetivos específicos del trabajo de investigación consistía en la confección de un “Manual de instrucciones operativas” que permita la puesta en funcionamiento del método desarrollado en la cadena de procesamiento de DTI-ARBA. Debido a que la propuesta diseñada se encuentra en formato de “prototipo”, que es necesario aún continuar avanzando en el desarrollo y mejoramiento de los modelos, que aún falta ingresar a la propuesta de mejora a “producción”, evaluar sus resultados y posteriormente presentar las modificaciones a la Gerencia, no es posible comenzar a elaborar los Instructivos de Trabajo en el tiempo estimado para la concreción de la tesis.

Por otro lado, los protocolos de producción vigentes (Manuales de Procedimientos y Manuales de Instrucciones Operativas) fueron elaborados en forma conjunta con el Departamento de Calidad y Procesos perteneciente a la Gerencia de Planificación y Control General de ARBA, y este proceso de trabajo conjunto tiene un alto costo en cantidad de tiempo invertido por ambas partes.

Por estos motivos, se tomó la decisión de finalizar la tesis con los resultados obtenidos de estimaciones de rendimiento realizadas en la campaña de implementación 2016/2017 y la metodología propuesta a través de los flujogramas presentados (Cap. 5.4).

5.5 Conclusiones.

Los modelos de estimación de rendimiento, como todo modelo, son una simplificación de la realidad para facilitar la comprensión de algún proceso. Es indudable que modelar procesos biológicos complejos y prolongados en el tiempo, como el crecimiento de los cultivos, e intentar predecir el rendimiento de los mismos, es una tarea sumamente compleja.

Es variada y numerosa la bibliografía y la comunidad científica dedicada a la elaboración y/o validación de modelos de estimación del rendimiento de cultivos. A pesar de los notables avances en la temática, los modelos disponibles no se encuentran validados en las amplias zonas que en general se demandan, y además requieren contar con un gran set de datos agrometeorológicos, información sobre el manejo cultural de cada cultivo, información sobre los cultivos antecesores, etc. En Argentina no es sencillo contar con todas las variables que requieren los modelos, y si se encuentran disponibles, generalmente no cumplen con la resolución espacial adecuada, la cobertura espacial y/o temporal que requieren los modelos para operar correctamente y cumplir con los objetivos propuestos. Por otro lado, la información del manejo del cultivo como: fecha de siembra, tipo de siembra, tipo de cultivar, aplicaciones de fertilización, momento de la fertilización, tipo de nutriente incorporado, por nombrar algunos ejemplos, son datos a escala de lote que son imposibles de reunir para trabajar a escala regional.

Los resultados obtenidos luego de la validación de los modelos y su aplicación, cumplieron con las expectativas previstas y resultados esperados al iniciar el trabajo de investigación. Este desarrollo es muy importante para el Departamento de Tecnología de Imágenes (DTI-ARBA) en el cual me desempeño laboralmente debido a que es el primer antecedente con base científica que se realiza para determinar los rendimientos de los cultivos agrícolas extensivos de manera autónoma, sin depender de datos provistos por otros organismos oficiales.

Este método propuesto y aplicado en el partido de Daireaux es el comienzo de un proceso de desarrollo que debe tener como objetivo principal “escalarse” a toda la zona agrícola de la provincia de Buenos Aires. Los mayores inconvenientes para realizar esta tarea se deben a la baja disponibilidad de datos de campo con información de rendimientos de los diferentes tipos de cultivo, y distribuidos espacial y temporalmente.

5.5.1 Conclusiones de los modelos.

Fue posible desarrollar modelos de estimación del rendimiento de cultivos agrícolas extensivos utilizando únicamente información satelital de libre disponibilidad como input de los modelos.

Se probó que es posible estimar el rendimiento con errores aceptables utilizando información provista por la serie Landsat con resolución temporal de 16 días y constelación de 8 días (Landsat 7 y 8), interpolando los valores temporales de aproximadamente 15 imágenes disponibles (en promedio de las 3 campañas analizadas).

Se determinó que es imprescindible para la aplicación de los modelos en otras campañas agrícolas, definir el Día Posterior al Green-UP (DPG) por cada pixel y el DPG-Promedio que es el valor medio de Green-UP para cada tipo de cultivo en el partido. De esta manera es posible identificar el día en el que se deben extraer los datos de la serie multitemporal interpolada del índice de vegetación que es parte fundamental del modelo propuesto.

Surge del análisis de los modelos que el mejor predictor del rendimiento resultó ser el NDVI en comparación con el EVI, el MSAVI y el NDMI, obteniendo los mayores valores del coeficiente de determinación R^2 y mayor homogeneidad en el comportamiento de los 4 cultivos analizados.

La manera de vincular los datos para realizar las regresiones lineales del modo “a” (correlacionando todo el set de datos de cada índice de vegetación por tipo de cultivo desde el inicio del estudio y en forma consecutiva día por día hasta el fin del estudio) fue la que mejor desempeño tuvo para los 4 tipos de cultivo y en ambas campañas agrícolas.

Los modelos de soja de primera (tanto en 2008/2009 como en 2010/2011) resultaron ser los que mejor ajuste lograron en la regresión lineal, sin embargo, estos modelos también fueron los que mayor error porcentual expresaron debido a que el rango de variación del rendimiento de la soja es acotado. Los errores porcentuales (Raíz del Error Medio Cuadrático Relativo, RRMSE) de los modelos desarrollados en ambas campañas promediaron el 26.4% para la Soja de primera, el 19.4% para la Soja de segunda, el 22.9% para el Maíz y el 18,2% para el Girasol, cumpliendo de esta manera con los objetivos fijados.

Las fechas para extraer la información de los índices de vegetación y vincular con el rendimiento se concentraron principalmente en 2 momentos: 50 días posteriores al Green-Up Promedio y 80 días posteriores al Green-Up Promedio. Particularmente, para la Soja de segunda en ambos modelos la extracción de la información es aproximadamente a los 80 DPG-P, y para el Girasol en ambos modelos es aproximadamente a los 50 DPG-P. En el caso de los cultivos de Soja de primera y Maíz, en los modelos de ciclo seco la extracción de la información se realiza a los 86 y 80 DPG-P respectivamente, y para el ciclo húmedo la extracción de la información baja aproximadamente a los 50 DPG-P.

Con respecto al análisis comparativo de las dos campañas agrícolas con diferente situación de disponibilidad hídrica (seca y normal/húmeda), se evidencian cambios en los ciclos de los cultivos, en los valores máximos de los índices de vegetación, y en los rendimientos. Por lo tanto, existen diferencias entre los modelos desarrollados para una u otra situación climática. Esto supone una correcta decisión en utilizar diferentes modelos para diferentes ciclos hídricos, aunque es necesario aumentar el número de campañas agrícolas a analizar para lograr afirmarlo.

Los modelos creados permiten predecir rendimientos con antelación a la cosecha debido a que técnicamente no requieren completar el set de datos hasta la cosecha, sino que, si se cuenta con una imagen libre de nubes posterior a la fecha de extracción de la información ya es posible realizar la interpolación, calcular el Green-UP Promedio y realizar el conteo de los DPG-P temporal en los datos para aplicar los modelos.

5.5.2 Conclusiones de la implementación.

Los modelos desarrollados para campañas de ciclos hídricos normales/húmedos se aplicaron en toda la extensión del partido de Daireaux utilizando los sensores Landsat 7 y Landsat 8 en la campaña agrícola 2016-2017. El cambio en la escala de trabajo, con el procesamiento de todo el set de datos de Daireaux (1.280.000 píxeles) operó de forma adecuada, obteniendo las funciones de crecimiento con valores diarios, el Green-Up, y el DPG a nivel pixel. El resultado obtenido del proceso fue el esperado: imagen raster con 3 bandas: una correspondiente al tipo de cultivo, otra indicando el valor del rendimiento estimado (qq/ha), y finalmente otra banda (capa) indicando la calidad asignada. Los datos de rendimiento obtenidos se encontraron dentro de los valores normales para la zona de estudio (análisis cuantitativo) y la distribución espacial de los mismos se consideró adecuada, respetando la topografía del sitio y los efectos "borde" (análisis cualitativo).

5.5.3 Conclusiones de la metodología propuesta.

Se definió una metodología para aplicar los modelos desarrollados por cada tipo de cultivo sobre el área de influencia correspondiente (Daireaux y Circunscripciones al sur de Trenque Lauquen) dentro del proceso de producción actual de DTI-ARBA. La propuesta de aplicación se desarrolló en formato de flujogramas en el proceso de producción de forma sencilla y factible de aplicar. Los manuales de procedimiento no se realizaron debido a que es una tarea que requiere un tiempo muy prolongado de desarrollo e incluye la intervención de otras áreas de la institución.

5.5.4 Limitaciones de los modelos.

- Requerimiento de datos mínimos: Los modelos no podrían aplicarse en el caso de no disponer imágenes libres de nubosidad por un prolongado período de tiempo (mayor a 4 fechas = 32 días) y principalmente si escasean en los períodos críticos de los cultivos.
- Dependencia en la calidad de las clasificaciones de cobertura: En el caso de que la clasificación de los cultivos no tenga un buen desempeño y existan altos porcentajes de omisión/comisión dentro de las clases de cultivos, ingresarán al modelo datos con curvas de crecimiento que no se corresponden con el modelo a aplicar según el tipo de cultivo. La interpolación se realizará de igual manera y los datos se extraerán en un DPG-P que no se corresponde con el tipo de cultivo, por lo tanto, se calculará un valor de rendimiento erróneo (por ejemplo, al girasol le aplico el modelo de maíz).
- Variaciones ambientales y de sanidad posteriores a la estimación: Si sucede un evento posterior a la fecha del dato del índice de vegetación extraído (granizo, inundación, sequía, plaga/enfermedad, etc.) que reduzca significativamente el rendimiento, los modelos realizarán una estimación adecuada pero que no coincidirá con el resultado cosechado al finalizar el ciclo.
- Cambios en la utilización de híbridos y cultivares: Por ejemplo, el modelo desarrollado para el cultivo de maíz fue construido con datos de maíz de ciclo tradicional (siembra temprana) y en la zona de estudio en los últimos años ha cambiado la forma de producir, pasando a siembra de híbridos de ciclo corto que permiten realizar siembras tardías con el objetivo de “escapar” a las frecuentes sequías que ocurren a fines de diciembre y comienzos de enero.

5.5.5 Propuestas de mejoramiento futuro.

- El procedimiento definido puede automatizarse para aplicarse a cada uno de los path row que cubren la provincia en el caso de que se logren generar los modelos correspondientes. Se encuentra escrito en lenguaje R, pero es posible implementarse en IDL para utilizar con ENVI o en Java Script para aplicar en Google Earth Engine, por ejemplo.
- Es posible incorporar índices de vegetación provenientes de otros sensores ópticos como Sentinel 2, aumentando la cantidad de datos para mejorar las interpolaciones temporales.
- Crear modelos de estimación del rendimiento en otras áreas de la provincia de Buenos Aires para lograr escalar la metodología, y modificarlos para que los modelos trabajen por tipos de ambientes, o sea, por sitios de características homogéneas de textura y estructura del suelo, profundidad del perfil, posición en el paisaje, drenaje, altura de la napa, etc.
- Evaluar el comportamiento de los modelos desarrollados para sequía y también continuar aplicando los modelos en otras campañas agrícolas, aumentando el número de campañas analizadas.
- Surge la necesidad de ampliar el número de clases de calidad (excelente, bueno, regular) para mejorar la precisión de las estimaciones. Las categorías se deben ajustar mejor a la nueva precisión con la cual se estima el dato de rendimiento, ampliando por lo menos, de 3 a 5 el número de clases.
- Los modelos podrían ajustarse para mejorar las subestimaciones y las sobreestimaciones, utilizando otro tipo de funciones no lineales, o regresiones lineales segmentadas denominadas “piecewise”.

Anexo I. Comparación de los métodos de filtrado de datos de campo.

Soja de Primera.

- Comparación de medias apareadas

“T - Test” para la diferencia $r_1 - r_2$

Hipótesis Nula: Diferencia = 0

Hipótesis Alternativa: Diferencia $\neq 0$

Media	-0,0755
Error estándar	0,0262
IC inferior (95%)	-0,1286
Media - H ₀	-0,0755
IC superior (95%)	-0,0224
Valor T	-2,88
Grados de libertad	38
P-valor	0,0066

El p-value es menor al valor alfa de significancia 0.05, por lo tanto, SI se rechaza la hipótesis nula y los valores medios estimados por el método 1 y por el método 2 tienen diferencias significativas. El valor de T negativo se debe a que las diferencias ($r_1 - r_2$) es en promedio negativa (-0,0755), o sea, que r_2 tiene valores de rinde mayores que r_1 (el método 2 sobrestima el rendimiento).

Soja de Segunda.

- Comparación de medias apareadas

“T - Test” para la diferencia $r_1 - r_2$

Hipótesis Nula: Diferencia = 0

Hipótesis Alternativa: Diferencia $\neq 0$

Media	0,0016
Error estándar	0,0103
IC inferior (95%)	-0,0200
Media - H ₀	0,0016
IC superior (95%)	0,0233
Valor T	0,16
Grados de libertad	17
P-valor	0,8757

El p-valor es mayor al valor alfa de significancia 0.05, por lo tanto, NO se rechaza la hipótesis nula y los valores medios estimados por el método 1 y por el método 2 no tienen diferencias significativas. Es posible utilizar indistintamente, uno u otro método de filtrado de datos.

Maíz.

- Comparación de medias apareadas

“T - Test” para la diferencia $r_1 - r_2$

Hipótesis Nula: Diferencia = 0

Hipótesis Alternativa: Diferencia $\neq 0$

Media	-0,4824
Error estándar	0,0938
IC inferior (95%)	-0,6781
Media - H ₀	-0,4824
IC superior (95%)	-0,2868
Valor T	-5,14
Grados de libertad	20
P-valor	0,0000

El p-valor es menor al valor alfa de significancia 0.05, por lo tanto, SI se rechaza la

hipótesis nula y los valores medios estimados por el método 1 y por el método 2 tienen diferencias significativas. NO es posible utilizar indistintamente, uno u otro método de filtrado de datos.

Girasol.

Nota: Para analizar la base de datos de Girasol se descartó uno de los lotes que tiene valores de rendimiento muy bajos y distorsionaba el análisis de correlación (al alejarse significativamente del resto de la nube de puntos, generaba que las variables parecieran estar correlacionadas), no así el de comparación de las medias apareadas que obtiene el mismo resultado.

- Comparación de medias apareadas

“T - Test” para la diferencia $r_1 - r_2$

Hipótesis Nula: Diferencia = 0

Hipótesis Alternativa: Diferencia $\neq 0$

Media	-0,4056
Error estándar	0,1371
IC inferior (95%)	-0,7159
Media - H_0	-0,4056
IC superior (95%)	-0,0954
Valor T	-2,96
Grados de libertad	9
P-valor	0,0160

El p-valor es menor al valor alfa de significancia 0.05, por lo tanto, SI se rechaza la hipótesis nula y los valores medios estimados por el método 1 y por el método 2 tienen diferencias significativas. NO es posible utilizar indistintamente, uno u otro método de filtrado de datos.

Anexo II. Resultados de los residuos de los modelos seleccionados.

MODELO SOJA 1° SEQUIA

```
> summary(regremax)
```

```
Call:
```

```
lmrob(formula = log(dato.interpol[, 242]) ~ dato.interpol[, which.max(rcuadrado)],  
      data = as.data.frame(dato.interpol[, ]), method = "MM", setting = "KS2011")  
  \--> method = "MM"
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-1.825858	-0.191111	0.003687	0.190697	1.433562

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.76943	0.02527	-30.44	<2e-16 ***
dato.interpol[, which.max(rcuadrado)]	5.05689	0.03252	155.49	<2e-16 ***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Robust residual standard error: 0.2818
```

```
Multiple R-squared:  0.7023, Adjusted R-squared:  0.7022
```

```
Convergence in 11 IRWLS iterations
```

```
Robustness weights:
```

```
7381 weights are ~ = 1. The remaining 3261 ones are summarized as
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.01358	0.76160	0.91430	0.82690	0.97620	0.99900

```
> lillie.test(regremax$residuals)
```

```
Lilliefors (Kolmogorov-Smirnov) normality test
```

```
data:  regremax$residuals
```

```
D = 0.043828, p-value < 2.2e-16
```

```
> lmtest::bptest(regremax)
```

```
studentized Breusch-Pagan test
```

```
data:  regremax
```

```
BP = 803.44, df = 1, p-value < 2.2e-16
```

MODELO SOJA 2° SEQUIA

```
> summary(regremax)
```

```
Call:
```

```
lmrob(formula = log(dato.interpol[, 242]) ~ tabla[, which.max(tabla2)], data = as.data.f  
rame(dato.interpol[,]), method = "MM", setting = "KS2011")\--> method = "MM"
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-1.009160	-0.130874	0.002236	0.126813	0.933560

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.60825	0.02787	21.82	<2e-16 ***
tabla[, which.max(tabla2)]	2.75854	0.03730	73.96	<2e-16 ***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Robust residual standard error: 0.194
```

```
Multiple R-squared:  0.5401, Adjusted R-squared:  0.54
```

```
Convergence in 11 IRWLS iterations
```

```
Robustness weights:
```

```
3377 weights are ~ = 1. The remaining 1536 ones are summarized as
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.06271	0.75010	0.91020	0.82930	0.97570	0.99900

```
> lmtest::bptest(regremax)
```

```
studentized Breusch-Pagan test
```

```
data:  regremax
```

```
BP = 249.84, df = 1, p-value < 2.2e-16
```

```
> lillie.test(regremax$residuals)
```

```
Lilliefors (Kolmogorov-Smirnov) normality test
```

```
data:  regremax$residuals
```

```
D = 0.033514, p-value = 9.957e-14
```

MODELO MAÍZ SEQUIA

```
> summary(regremax)
Call:
lmrob(formula = log(dato.interpol[, 242]) ~ dato.interpol[, which.max(rcuadrado)],
      data = as.data.frame(dato.interpol[, ]), method = "MM", setting = "KS2011")
\--> method = "MM"
Residuals:
    Min       1Q   Median       3Q      Max
-3.617564 -0.160608  0.006538  0.126214  1.095479
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.33415    0.04455   29.95 <2e-16 ***
dato.interpol[, which.max(rcuadrado)]  4.12165    0.05744   71.76 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Robust residual standard error: 0.1995
Multiple R-squared:  0.4793, Adjusted R-squared:  0.4792
Convergence in 28 IRWLS iterations
Robustness weights:
 32 observations c(351,1904,1911,2169,2195,2370,2546,2596,2615,2798,2981,3758,3854,3937,
3942,3943,3982,4011,4012,4045,4046,4090,4092,4159,4315,4567,4767,4943,5021,5518,5591,570
5) are outliers with |weight| = 0 (< 1.6e-05);
 3896 weights are ~ = 1. The remaining 2172 ones are summarized as
    Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
0.0000238 0.4736000 0.8426000 0.7093000 0.9643000 0.9990000
> lmtest::bptest(regremax)
studentized Breusch-Pagan test
data:  regremax
BP = 565.96, df = 1, p-value < 2.2e-16
> lillie.test(regremax$residuals)
Lilliefors (Kolmogorov-Smirnov) normality test
data:  regremax$residuals
D = 0.14308, p-value < 2.2e-16
```

MODELO GIRASOL SEQUIA

```
> summary(regremax)
Call:
lmrob(formula = log(dato.interpol[, 242]) ~ dato.interpol[, which.max(rcuadrado)],
      data = as.data.frame(dato.interpol[, ]), method = "MM", setting = "KS2011")
\--> method = "MM"
Residuals:
    Min       1Q   Median       3Q      Max
-1.91582 -0.15184  0.01494  0.12036  0.81417
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      2.44318    0.02143  114.01 <2e-16 ***
dato.interpol[, which.max(rcuadrado)]  2.35786    0.04805   49.07 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Robust residual standard error: 0.1838
Multiple R-squared:  0.4931, Adjusted R-squared:  0.4929
Convergence in 23 IRWLS iterations
Robustness weights:
 4 observations c(1522,2142,2163,2621) are outliers with |weight| = 0 (< 3.6e-05);
 1772 weights are ~ = 1. The remaining 968 ones are summarized as
    Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
0.005288 0.536900 0.842100 0.723700 0.957900 0.999000
> lmtest::bptest(regremax)
studentized Breusch-Pagan test
data:  regremax
BP = 279.88, df = 1, p-value < 2.2e-16
> lillie.test(regremax$residuals)
Lilliefors (Kolmogorov-Smirnov) normality test
data:  regremax$residuals
D = 0.11919, p-value < 2.2e-16
```

MODELO DE SOJA 1 NORMAL:

```
> summary(regremax)
Call:
lmrob(formula = log(dato.interpol[, 234]) ~ dato.interpol[, which.max(rcuadrado)],
      data = as.data.frame(dato.interpol[, ]), method = "MM", setting = "KS2011")
  \--> method = "MM"
Residuals:
      Min       1Q   Median       3Q      Max
-2.484377 -0.177492  0.003151  0.179169  1.787832
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.46046    0.02168   21.23  <2e-16 ***
dato.interpol[, which.max(rcuadrado)]  3.49884    0.02838  123.29  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Robust residual standard error: 0.2655
Multiple R-squared:  0.5688, Adjusted R-squared:  0.5687
Convergence in 11 IRWLS iterations
Robustness weights:
 5 observations c(2886,11741,11745,11750,11883) are outliers with |weight| = 0 (< 8.3e-06);
 8352 weights are ~ 1. The remaining 3743 ones are summarized as
      Min.  1st Qu.  Median    Mean 3rd Qu.    Max.
0.0000838 0.7973000 0.9184000 0.8461000 0.9767000 0.9990000

> lmtest::bptest(regremax)
studentized Breusch-Pagan test
data:  regremax
BP = 464.82, df = 1, p-value < 2.2e-16

> lillie.test(regremax$residuals)
Lilliefors (Kolmogorov-Smirnov) normality test
data:  regremax$residuals
D = 0.031597, p-value < 2.2e-16
```

MODELO DE SOJA 2 NORMAL:

```
> summary(regremax)
Call:
lmrob(formula = log(dato.interpol[, 234]) ~ dato.interpol[, which.max(rcuadrado)],
      data = as.data.frame(dato.interpol[, ]), method = "MM", setting = "KS2011")
  \--> method = "MM"
Residuals:
      Min       1Q   Median       3Q      Max
-3.2163581 -0.1295615  0.0005785  0.1295167  0.8037147
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.77549    0.03548   21.86  <2e-16 ***
dato.interpol[, which.max(rcuadrado)]  2.39060    0.04939   48.40  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Robust residual standard error: 0.1899
Multiple R-squared:  0.4112, Adjusted R-squared:  0.4111
Convergence in 12 IRWLS iterations
Robustness weights:
 5 observations c(267,1027,2050,2063,2064) are outliers with |weight| <= 2.3e-05 (< 2.9e-05);
 2394 weights are ~ 1. The remaining 1036 ones are summarized as
      Min.  1st Qu.  Median    Mean 3rd Qu.    Max.
0.02207 0.80220 0.93280 0.84240 0.97850 0.99900

> lmtest::bptest(regremax)
studentized Breusch-Pagan test
data:  regremax
BP = 89.514, df = 1, p-value < 2.2e-16

> lillie.test(regremax$residuals)
Lilliefors (Kolmogorov-Smirnov) normality test
data:  regremax$residuals
D = 0.060866, p-value < 2.2e-16
```

MODELO DE MAIZ NORMAL:

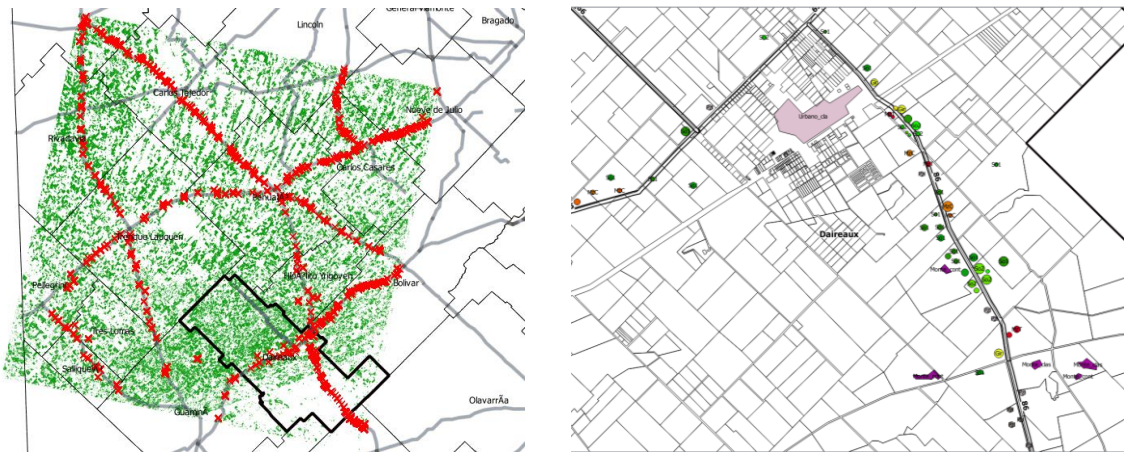
```
> summary(regremax)
Call:
lmrob(formula = log(dato.interpol[, 234]) ~ dato.interpol[, which.max(rcuadrado)],
      data = as.data.frame(dato.interpol[, ]), method = "MM", setting = "KS2011")
\--> method = "MM"
Residuals:
      Min       1Q   Median       3Q      Max
-2.87483 -0.16671  0.01844  0.14380  0.80447
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.53099    0.04374   35.00  <2e-16 ***
dato.interpol[, which.max(rcuadrado)]  3.87771    0.05754   67.39  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Robust residual standard error: 0.2168
Multiple R-squared:  0.3921, Adjusted R-squared:  0.392
Convergence in 14 IRWLS iterations
Robustness weights:
 35 observations c(1005,1473,1737,1745,1751,1757,1758,1764,1765,1774,1793,1794,1799,1803
,1804,1806,1819,1821,1826,1852,1901,2785,2843,2846,3022,3672,3701,4859,4892,4945,4946,49
47,5433,5739,5922) are outliers with |weight| = 0 ( < 1.4e-05); 4976 weights are ~ = 1. T
he remaining 2222 ones are summarized as
      Min.  1st Qu.  Median    Mean  3rd Qu.    Max.
0.0000599 0.7296000 0.9181000 0.7973000 0.9770000 0.9990000
> lmtest::bptest(regremax)
studentized Breusch-Pagan test
data:  regremax
BP = 643.08, df = 1, p-value < 2.2e-16
> lillie.test(regremax$residuals)
Lilliefors (Kolmogorov-Smirnov) normality test
data:  regremax$residuals
D = 0.097445, p-value < 2.2e-16
```

MODELO DE GIRASOL NORMAL:

```
> summary(regremax)
Call:
lmrob(formula = log(dato.interpol[, 234]) ~ dato.interpol[, which.max(rcuadrado)],
      data = as.data.frame(dato.interpol[, ]), method = "MM", setting = "KS2011")
\--> method = "MM"
Residuals:
      Min       1Q   Median       3Q      Max
-1.37765 -0.10882  0.00921  0.10176  0.58235
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.40033    0.02333   60.01  <2e-16 ***
dato.interpol[, which.max(rcuadrado)]  2.62023    0.03102   84.46  <2e-16 ***
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Robust residual standard error: 0.1524
Multiple R-squared:  0.6126, Adjusted R-squared:  0.6125
Convergence in 12 IRWLS iterations
Robustness weights:
 observation 3808 is an outlier with |weight| = 0 ( < 2.1e-05);
 3254 weights are ~ = 1. The remaining 1443 ones are summarized as
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.00543 0.72540 0.90930 0.81340 0.97570 0.99900
> lmtest::bptest(regremax)
studentized Breusch-Pagan test
data:  regremax
BP = 204.35, df = 1, p-value < 2.2e-16
> lillie.test(regremax$residuals)
Lilliefors (Kolmogorov-Smirnov) normality test
data:  regremax$residuals
D = 0.053823, p-value < 2.2e-16
```

Anexo III. Clasificación de cultivos de DTI-ARBA para la campaña verano 2016-2017

Relevamiento de datos de campo: La recolección de los datos se realiza con tablets PC con GPS incorporado que permite asignar polígonos sobre los lotes seleccionados por el operador, indicando el tipo de cobertura. Los relevamientos se realizan sobre las vías asfaltadas y/o caminos mejorados abarcando toda la provincia y posteriormente se exportan los datos en las unidades de trabajo (Path-Row), en este caso el Path-Row 227-85. Los datos de campo que componen la extracción para la clasificación del Path-Row fueron recolectados en 3 recorridos distintos: 08 y 09 de marzo 2017, 15 y 16 de marzo, y 28 y 29 de marzo de 2017. Los datos de campo contienen por lo general para la campaña de verano (gruesa) los siguientes tipos de cobertura: soja1 (soja de primera), soja2 (soja de segunda), maízC (maíz de ciclo tradicional o común), maízT (maíz de siembra tardía), pastura (forraje implantado) y pastizal (forraje espontáneo). A estos datos se le agregan otros polígonos creados en gabinete (visualización de imágenes): monte (cobertura arbórea), agua (cuerpos de agua) y urbano (zonas residenciales). Cada una de estas coberturas se divide en dos sets de datos: clasificación y control.



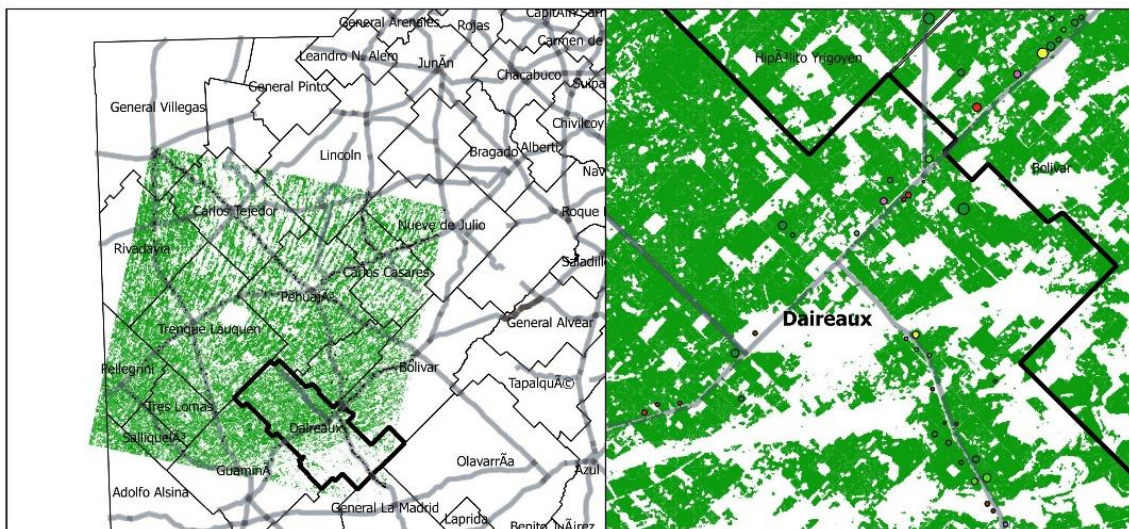
Izq: Clasificación de cultivos (verde) y datos de cultivos (rojo) en el Path row 227-085. Der: Visualización de los datos de campo recolectados sobre los caminos transitados.

Overall Accuracy:	95,74%	(340884/356017)											
Kappa Coefficient:	0.8819												
Valores en pixeles													
Valores en pixeles	Agua_control	Girasol_control	MaiZ_control	MaiZT_control	Monte_control	Pastizal_control	Pastura_control	Soja1_control	Soja2_control	Urbano_control	Total		
Agua_clas	281727	0	0	0	0	9	0	0	0	0	281736		
Girasol_clas	0	1407	236	118	0	2	0	0	0	0	1763		
MaiZ_clas	1	125	5994	562	0	139	0	32	0	5	6858		
MaiZT_clas	8	132	1557	4881	2	271	19	31	0	5	6906		
Monte_clas	631	350	103	99	2138	640	25	192	0	182	4360		
Pastizal_clas	1910	83	204	15	36	3886	36	39	0	857	7066		
Pastura_clas	146	18	3	391	52	3298	2353	98	234	534	7127		
Soja1_clas	0	0	0	4	3	0	0	14913	31	4	14955		
Soja2_clas	0	0	0	58	2	0	21	781	1868	0	2730		
Urbano_clas	162	0	191	286	45	51	1	63	0	21717	22516		
Total	284585	2115	8288	6414	2278	8296	2455	16149	2133	23304	356017		
Valores en porcentaje													
Valores en porcentaje	Agua_control	Girasol_control	MaiZ_control	MaiZT_control	Monte_control	Pastizal_control	Pastura_control	Soja1_control	Soja2_control	Urbano_control	Total		
Agua_clas	99.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	79.14		
Girasol_clas	0.00	66.52	2.85	1.84	0.00	0.02	0.00	0.00	0.00	0.00	0.50		
MaiZ_clas	0.00	5.91	72.32	8.76	0.00	1.68	0.00	0.20	0.00	0.02	1.93		
MaiZT_clas	0.00	6.24	18.79	76.10	0.09	3.27	0.77	0.19	0.00	0.02	1.94		
Monte_clas	0.22	16.55	1.24	1.54	93.85	7.71	1.02	1.19	0.00	0.78	1.22		
Pastizal_clas	0.67	3.92	2.46	0.23	1.58	46.84	1.47	0.24	0.00	3.68	1.98		
Pastura_clas	0.05	0.85	0.04	6.10	2.28	39.75	95.85	0.61	10.97	2.29	2.00		
Soja1_clas	0.00	0.00	0.00	0.06	0.13	0.00	0.00	92.35	1.45	0.02	4.20		
Soja2_clas	0.00	0.00	0.00	0.90	0.09	0.00	0.86	4.84	87.58	0.00	0.77		
Urbano_clas	0.06	0.00	2.30	4.46	1.98	0.61	0.04	0.39	0.00	93.19	6.32		
Total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00		

	Commission (Percent)	Omission (Percent)	Commission (Pixels)	Omission (Pixels)
Agua	0.00	1.00	9/281736	2858/284585
Girasol	20.19	33.48	356/1763	708/2115
MaiZ	12.60	27.68	864/6858	2294/8288
MaiZT	29.32	23.90	2025/6906	1533/6414
Monte	50.96	6.15	2222/4360	140/2278
Pastizal	45.00	53.16	3180/7066	4410/8296
Pastura	66.98	4.15	4774/7127	102/2455
Soja1	0.28	7.65	42/14955	1236/16149
Soja2	31.58	12.42	862/2730	265/2133
Urbano	3.55	6.81	799/22516	1587/23304

	Prod. Acc. (Percent)	User Acc. (Percent)	Prod. Acc. (Pixels)	User Acc. (Pixels)
Agua	99.00	100.00	281727/284585	281727/281736
Girasol	66.52	79.81	1407/2115	1407/1763
MaiZ	72.32	87.40	5994/8288	5994/6858
MaiZT	76.10	70.68	4881/6414	4881/6906
Monte	93.85	49.04	2138/2278	2138/4360
Pastizal	46.84	55.00	3886/8296	3886/7066
Pastura	95.85	33.02	2353/2455	2353/7127
Soja1	92.35	99.72	14913/16149	14913/14955
Soja2	87.58	68.42	1868/2133	1868/2730
Urbano	93.19	96.45	21717/23304	21717/22516

Resultados de la matriz de confusión de la clasificación de cultivos de verano 2016-2017



Clasificación de cultivos campaña 2016-2017 y ubicación de datos de campo

BIBLIOGRAFÍA.

- [1] ARBA, «www.arba.gov.ar,» [En línea]. Available: www.arba.gov.ar.
- [2] INDEC, «Censo Nacional Agropecuario 2008,» 2009.
- [3] J. L. Minetti, W. M. Vargas, B. Vega y M. C. Costa, «LAS SEQUÍAS EN LA PAMPA HÚMEDA: IMPACTO EN LA PRODUCTIVIDAD DEL MAÍZ,» vol. 22, nº 2, pp. 218-232, 2007.
- [4] M. d. Agroindustria, «Portal de Datos Abiertos Agroindustria,» [En línea]. Available: <http://www.agroindustria.gov.ar/datosabiertos/>. [Último acceso: 1 12 2017].
- [5] DINREP, «DINREP-Buenos Aires,» Dirección Nacional de Relaciones Económicas con las Provincias, 2012.
- [6] A. De La Casa y G. Ovando, «Integración del índice de vegetación de la diferencia normalizada (ndvi) y del ciclo fenológico de maíz para estimar el rendimiento a escala departamental en Córdoba, Argentina,» vol. 67, nº 4, pp. 362-371, 2007.
- [7] S. Moulin, A. Bondeau y R. Delecalle, *Combining agricultural crop models and satellite observations: From field to regional scales*, 1998.
- [8] UNER. Cátedra de Climatología y Fenología Agrícola, «Unidad 10: Modelos de cultivos,» 2009.
- [9] T. R. ACIAR, «A comparison of three empirical models for assessing cropping options in a data-sparse environment, with reference to Laos and Cambodia,» 2015.
- [10] FAO, Respuesta del rendimiento de los cultivos al agua, vol. 66, 2012.
- [11] B. Duchemin, «A simple algorithm for yield estimates: Evaluation for semi-arid irrigated winter wheat monitored with green leaf area index,» *Science Direct*, p. 16, 2008.
- [12] J. W. Jones, G. Hoogenboom, C. H. Porter, K. J. Boote, W. D. Batchelor, L. A. Hunt, P. W. Wilkens, U. Singh, A. J. Gijssman y J. T. Ritchie, «The DSSAT cropping system model,» de *European Journal of Agronomy*, 2003.
- [13] K. R. Thorp, K. C. DeJonge, A. L. Kaleita, W. D. Batchelor y J. O. Paz, «Methodology for the use of DSSAT models for precision agriculture decision support,» *Computers and Electronics in Agriculture*, 2008.
- [14] Nadine Brisson, Bruno Mary, Dominique Ripoche, Marie Hélène Jeuffroy, Françoise Ruget, et al., «STICS: a generic model for the simulation of crops and their water and nitrogen balances. I. Theory,» *Agronomie, EDP Sciences*, pp. 311-346, 1998.
- [15] Williams, J.R. Jones, C.A. Kiniry, J.R. Spanel, D.A., «The EPIC crop growth model,» *American Society of Agricultural and Biological Engineers*, 1989.
- [16] C. O. Stockle, M. Donatelli y R. Nelson, «CropSyst, a cropping systems simulation model,» *European Journal of Agronomy*, nº 18, p. 289/307, 2003.
- [17] M. Van Ittersum, P. Leffelaar, H. v. Keulen, M. Kropff, L. Bastiaans y J. Goudriaan, «On approaches and applications of the Wageningen crop models,» *European Journal of Agronomy*, nº 18, p. 201/234, 2003.
- [18] M. Bocco, S. Sayago, S. Violini y E. Willing, «Modelos simples para estimar rendimiento de cultivos agrícolas a partir de imágenes satelitales: una herramienta para la planificación,» de *44 JAIIO*, Rosario, 2015.
- [19] C. L. Wiegand, A. J. Richardson, R. D. Jackson, P. J. Pinter, J. Kris Aase, D. E. Smika, L. F. Lautenschlager y J. E. McMurtrey, «Development of Agrometeorological Crop Model Inputs from Remotely Sensed Information,» *IEEE Transactions on Geoscience and Remote Sensing*, 1986.
- [20] M. F. Lopresti, C. M. Di Bella y A. J. Degioanni, «Relationship between MODIS-NDVI data and wheat yield: A case study in Northern Buenos Aires province, Argentina,» *Information Processing in Agriculture*, 2015.
- [21] A. Chipanshi, Y. Zhang, L. Kouadio, N. Newlands, A. Davidson, H. Hill, R. Warren, B. Qian, B. Daneshfar, F. Bedard y G. Reichert, «Evaluation of the Integrated Canadian Crop Yield Forecaster (ICCYF) model for in-season prediction of crop yield across the Canadian agricultural landscape,» *Agricultural and Forest Meteorology*, 2015.

- [22] M. P. Labus, G. A. Nielsen, R. L. Lawrence, R. Engel y D. S. Long, «International Journal of Remote Sensing Wheat yield estimates using multi-temporal NDVI satellite imagery Wheat yield estimates using multi-temporal NDVI satellite imagery,» *int. j. remote sensing*, vol. 23, nº 20, pp. 4169-4180, 2002.
- [23] R. Shrestha, L. Di, E. G. Yu, L. Kang, L. Li, M. S. Rahman, M. Deng y Z. Yang, «Regression based corn yield assessment using MODIS based daily NDVI in Iowa state,» de *2016 5th International Conference on Agro-Geoinformatics, Agro-Geoinformatics 2016*, 2016.
- [24] J. M. Chen y J. Cihlar, «Retrieving leaf area index of boreal conifer forests using landsat TM images,» *Remote Sensing of Environment*, 1996.
- [25] S. J. Mass, «Using Satellite Data to Improve Model Estimates of Crop Yield,» *Agronomy Journal*, nº 80:655-662, 1988.
- [26] R. Casa, E. A. Zanelli, S. Pascucci, G. Laneve, H. Yang, G. Yang, W. Huang y S. Pignatti, «Exploitation of Multitemporal and Multisensor Earth Observation Data for Arable Crop Classification and Yield Assessment at the Farm Scale,» de *Dragon 4 Symposium, Copenhagen - Denmark*, 2017.
- [27] D. Gaso y A. G. Berger, «Early estimated wheat yield with a simple crop growth model through the assimilation of Landsat images,» 2015.
- [28] A. Chahbi, M. Zribi, Z. Lili-Chabaane, B. Duchemin, M. Shabou, B. Mougnot y G. Boulet, «Estimation of the dynamics and yields of cereals in a semi-arid area using remote sensing and the SAFY growth model,» *International Journal of Remote Sensing*, 2014.
- [29] M. Battude, A. Al Bitar, D. Morin, J. Cros, M. Huc, C. Marais Sicre, V. Le Dantec y V. Demarez, «Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2 like remote sensing data,» *Remote Sensing of Environment*, 2016.
- [30] R. Delécolle, S. J. Maas, M. Guérif y F. Baret, «Remote sensing and crop production models: present trends,» *ISPRS Journal of Photogrammetry and Remote Sensing*, 1992.
- [31] ARBA, «http://www.arba.gov.ar/NoticiasHome/MasInfo.asp?id_contenido=1178,» [En línea]. Available: www.arba.gov.ar. [Último acceso: 13 12 2017].
- [32] ARBA, «Acuerdo Marco entre ARBA y CONAE,» de *Expediente N° 043/08*, La Plata, 2008.
- [33] D. P. Luciano Elias, «Clasificación de cultivos en la provincia de Buenos Aires mediante la utilización de imágenes SAR e imágenes ópticas,» 2016.
- [34] Hall, Rebella, Ghersa, Culot, «Field-crop systems of the Pampas. Ecosystems of the World,» Netherlands, 1992, pp. 413-451.
- [35] S. I. Ballesteros, «Inundaciones y su relación con el clima y la hidrología subterránea en el Noroeste de Buenos Aires (1980-2010). Aplicación de percepción remota 2014,» 2005.
- [36] O. E. Scarpati y A. D. Capriolo, «Sequías e inundaciones en la provincia de Buenos Aires (Argentina) y su distribución espacio-temporal,» *Investigaciones Geográficas: Boletín del Instituto de Geografía*, vol. 2013, pp. 38-51, 2013.
- [37] F. Ferrelli, «La sequía 2008-2009 en el Sudoeste de la provincia de Buenos Aires (Argentina),» *Ecosistemas*, vol. 21, pp. 235-238, 2012.
- [38] P. E. d. I. p. d. B. Aires, *Declaración del estado de Emergencia y/o Desastre Agropecuario por sequía*, La Plata, 2009.
- [39] F. Ferrelli, V. Yael Bohn y M. C. Piccolo, «Variabilidad de la precipitación y ocurrencia de eventos secos en el sur de la provincia de Buenos Aires (Argentina),» *IX JORNADAS NACIONALES DE GEOGRAFIA FISICA BAHIA BLANCA*, pp. 15-28, 2012.
- [40] M. Braga chini, «Monitor de Rendimiento y Conocimientos de Calibración,» 2006.
- [41] G. Lyle, B. A. Bryan y B. Ostendorf, «Post-processing methods to eliminate erroneous grain yield measurements: Review and directions for future development,» *Precision Agriculture*, 2014.
- [42] J. Kleinjan, J. Chang, J. Wilson, D. Humburg, G. Carlson, D. Clay y D. Long, «Cleaning Yield Data,» 2002.
- [43] T. W. Griffin, J. P. Brown, J. Lowenberg-Deboer y G. R. Assistants, «YIELD MONITOR DATA ANALYSIS: DATA ACQUISITION, MANAGEMENT, AND ANALYSIS PROTOCOL,» 2005.

- [44] F. Sadosky, «Agrodatos,» [En línea]. Available: agrodatos.info/dataset?groups=aacrea2. [Último acceso: 05 Agosto 2017].
- [45] Sadosky, «Fundación Sadosky,» [En línea]. Available: www.fundacionsadosky.org.ar. [Último acceso: 05 Agosto 2017].
- [46] AACREA, «www.aacrea.org.ar,» [En línea]. Available: www.aacrea.org.ar. [Último acceso: 05 Agosto 2017].
- [47] PROCISUR, Agricultura de Precisión, Montevideo, Uruguay, 2006.
- [48] M. S. Albarenque y J. P. Vélez, «Técnicas para el procesamiento de mapas de rendimiento,» 2011.
- [49] G. Lyle, M. Lewis y B. Ostendorf, «Testing the temporal ability of Landsat imagery and precision agriculture technology to provide high resolution historical estimates of wheat yield at the farm scale,» *Remote Sensing*, 2013.
- [50] Balzarini Mónica [et.al.], Análisis de la variabilidad espacial en lotes agrícolas. Manual de Buenas Prácticas Agrícolas, Córdoba: 1a ed. Eudecor. ISBN 978-987-1536-66-5, 2014.
- [51] J. P. CELEMIN, «Autocorrelación espacial e indicadores locales de asociación espacial: Importancia, estructura y aplicación,» *Revista Universitaria de Geografía*, vol. vol.18, nº n.1, pp. 11-31, 2009.
- [52] Córdoba M., Balzarini M., Bruno C., Aguade F., Paccioretti P., *FastMapping v.1.0*, Córdoba, Argentina., 2015.
- [53] USGS, «<https://landsat.usgs.gov/>,» [En línea]. Available: <https://landsat.usgs.gov/landsat-missions-timeline>. [Último acceso: 2009 2017].
- [54] CONAE, «catalogos.conae.gov.ar,» [En línea]. Available: <https://catalogos.conae.gov.ar/landsat8/Docs/IndicesEspectralesDerivadosDeLandsat8-v2.pdf>. [Último acceso: 2018].
- [55] Wang, Cong et. al., «Analysis of Differences in Phenology Extracted from the Enhanced Vegetation Index and the Leaf Area Index,» *Sensors*, 2017.
- [56] Yu, Fangfang et. al., «Response of seasonal vegetation development to climatic variations in eastern central Asia,» *Remote Sensing of Environment*, 2003.
- [57] F. H. Andrade, Bases para el manejo del maíz, el girasol y la soja., Argentina : Medica Panamericana, 2000.
- [58] C. D. R. Box G. E. P., «An Analysis of Transformations,» *Journal of the Royal Statistical Society*, vol. Vol. 26, nº No. 2, pp. 211-252, 1964.
- [59] E. Castaño, «Una estimación no paramétrica y robusta de la transformación Box-Cox para el modelo de regresión,» *Lecturas de Economía - No. 75*, vol. 75, pp. 89-106, julio-diciembre 2011.
- [60] A. M. Iturri, «[Deroweb.com.ar](http://deroweb.com.ar),» [En línea]. Available: <http://meteo.deroweb.com.ar/lluvias.php>. [Último acceso: Octubre 2018].
- [61] A. K. S. O. Mutanga, «Narrow band vegetation indices overcome the saturation problem in biomass estimation,» *International Journal of Remote Sensing*, vol. VOL. 25, nº NO. 19, p. 3999–4014, 2004.
- [62] Yingxin Gua, Bruce K. Wylieb, Daniel M. Howardc, Khem P. Phuyalc, Lei Ji a, «NDVI saturation adjustment: A new approach for improving cropland performance estimates in the Greater Platte River Basin, USA,» *Ecological Indicators*, nº 30, pp. 1-6, 2013.
- [63] Rubén Álvarez, Sebastián Leavy, Magdalena Marino, «Zonas Agroeconómicas Homogéneas Bs As Norte,» INTA, CABA, 2009.
- [64] Y. Zhao, S. Chen y S. Shen, «Assimilating remote sensing information with crop model using Ensemble Kalman Filter for improving LAI monitoring and yield estimation,» *Ecological Modelling*, 2013.
- [65] Q. Xie, W. Huang, B. Zhang, P. Chen, X. Song, S. Pascucci, S. Pignatti, G. Laneve y Y. Dong, «Estimating Winter Wheat Leaf Area Index from Ground and Hyperspectral Observations Using Vegetation Indices,» *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2016.

- [66] M. Weiss, D. Troufleau, F. Baret, H. Chauki, L. Prévot, A. Olioso, N. Bruguier y N. Brisson, «Coupling canopy functioning and radiative transfer models for remote sensing data assimilation,» *Agricultural and Forest Meteorology*, vol. 108, pp. 113-128, 2001.
- [67] J. Van Wart, K. C. Kersebaum, S. Peng, M. Milner y K. G. Cassman, «Estimating crop yield potential at regional to national scales,» *Field Crops Research*, 2013.
- [68] J. Sergio, S. Efraín y M. Farfán, «El arte de programar en R: un lenguaje para la estadística,» 2014.
- [69] G. Sepulcre-Cantó, F. Gellens-Meulenberghs, A. Arboleda, G. Duveiller, A. de Wit, H. Eerens, B. Djab y P. Defourny, «Estimating crop-specific evapotranspiration using remote-sensing imagery at various spatial resolutions for improving crop growth modelling,» *International Journal of Remote Sensing*, 2013.
- [70] J. Salmon, M. A. Friedl, S. Froking, D. Wisser y E. M. Douglas, «Global rain-fed, irrigated, and paddy croplands: A new high resolution map derived from remote sensing, crop inventories and climate data,» *International Journal of Applied Earth Observation and Geoinformation*, 2015.
- [71] D. Raes, P. Steduto, T. C. Hsiao y E. Fereres, «AquaCrop – The FAO crop model to simulate yield response to water AquaCrop Reference Manual,» *Reference Manual*, 2009.
- [72] E. Nana, C. Corbari y D. Bocchiola, «A model for crop yield and water footprint assessment: Study of maize in the Po valley,» *Agricultural Systems*, 2014.
- [73] W. Mora, «Matemática, Educación e Internet Cómo utilizar R en métodos numéricos,» *Setiembre – Febrero*, vol. 16, nº 1, 2016.
- [74] W. Lück y A. van Niekerk, «Evaluation of a rule-based compositing technique for Landsat-5 TM and Landsat-7 ETM+ images,» *International Journal of Applied Earth Observation and Geoinformation*, 2016.
- [75] R. Luciani, G. Laneve, M. Jahjah y C. Mito, «Developing a satellite based automatic system for crop monitoring: Kenya's Great Rift valley, a case study,» de *European Space Agency, (Special Publication) ESA SP*, 2016.
- [76] L. Kouadio, N. K. Newlands, A. Davidson, Y. Zhang y A. Chipanshi, «Assessing the performance of MODIS NDVI and EVI for seasonal crop yield forecasting at the ecodistrict scale,» *Remote Sensing*, 2014.
- [77] N. Kim y Y.-W. Lee, «Machine Learning Approaches to Corn Yield Estimation Using Satellite Images and Climate Data: A Case of Iowa State,» *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, 2016.
- [78] D. M. Johnson, «An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States,» *Remote Sensing of Environment*, 2014.
- [79] Ing. Agr. Mariano Augusto Córdoba, «Herramientas estadísticas para el monitoreo y uso de la variabilidad espacial del rendimiento y propiedades de suelo intralote,» 2014.
- [80] M. E. Holzman y R. E. Rivas, «Early Maize Yield Forecasting from Remotely Sensed Temperature/Vegetation Index Measurements,» *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2016.
- [81] M. E. Holzman, R. Rivas y M. C. Piccolo, «Estimating soil moisture and the relationship with crop yield using surface temperature and vegetation index,» *International Journal of Applied Earth Observation and Geoinformation*, 2014.
- [82] Global Yield Gap Atlas, «List of crop modelling studies as collected for the Global yield gap atlas project,» p. 10, 2012.
- [83] P. C. Doraiswamy, J. L. Hatfield, T. J. Jackson, B. Akhmedov, J. Prueger y A. Stern, «Crop condition and yield simulations using Landsat and MODIS,» de *Remote Sensing of Environment*, 2004.
- [84] D. De Abelleira, S. Verón y C. Di Bella, «GEO Joint Experiment for Crop Assessment and Monitoring (JECAM): 2014 Site Progress Report».
- [85] R. C. Chicana, «Introducción al uso de R y R Commander para el análisis estadístico de datos en ciencias sociales,» 2014.
- [86] N. Chandel, K. Agrawal, P. Tiwari, K. Golhani y H. Tripathi, «IDW Interpolation of Soybean Yield Data acquired by Automated Yield Monitor,» vol. 13, nº 1, pp. 36-45, 2013.

- [87] D. K. Bolton y M. A. Friedl, «Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics,» *Agricultural and Forest Meteorology*, 2013.
- [88] R. Balaghi, B. Tychon, H. Eerens y M. Jlibene, «Empirical regression models using NDVI, rainfall and temperature data for the early prediction of wheat grain yields in Morocco,» *International Journal of Applied Earth Observation and Geoinformation*, 2008.
- [89] Agrim Juan Pablo Silva, «Elaboración de un protocolo para la generación de Información Geográfica, que permita una mejor gestión de las inundaciones, en la Provincia De Buenos Aires,» 2017.
- [90] X. Zhang y Q. Zhang, «Monitoring interannual variation in global crop yield using long-term AVHRR and MODIS observations,» *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016.
- [91] M. K. Van Ittersum, K. G. Cassman, P. Grassini, J. Wolf, P. Tittone y Z. Hochman, «Yield gap analysis with local to global relevance-A review,» *Field Crops Research*, 2013.
- [92] Thomson Reuters, «THE IMPOSSIBLE REALIZED,» 2014.
- [93] F. Tao, M. Yokozaawa, Z. Zhang, Y. Xu y Y. Hayashi, «Remote sensing of crop production in China by production efficiency models: Models comparisons, estimates and uncertainties,» *Ecological Modelling*, 2005.
- [94] R. A. Seiler, F. Kogan, G. Wei y M. V. Vinocur, «Seasonal and interannual responses of the vegetation and production of crops in Cordoba - Argentina assessed by AVHRR derived vegetation indices,» *Advances in Space Research*, 2007.
- [95] T. Sakamoto, A. A. Gitelson y T. J. Arkebauer, «Near real-time prediction of U.S. corn yields based on time-series MODIS data,» *Remote Sensing of Environment*, 2014.
- [96] T. Sakamoto, A. A. Gitelson y T. J. Arkebauer, «MODIS-based corn grain yield estimation model incorporating crop phenology information,» *Remote Sensing of Environment*, 2013.
- [97] J. Ren, Z. Chen, Q. Zhou y H. Tang, «Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China,» *International Journal of Applied Earth Observation and Geoinformation*, 2008.
- [98] N. A. Quarmby, M. Milnes, T. L. Hindle y N. Silleos, «The use of multi-temporal NDVI measurements from AVHRR data for crop yield estimation and prediction,» *International Journal of Remote Sensing*, 1993.
- [99] J. D. Luck, «Yield Monitoring Systems: Understanding how we Estimate Yield,» 2013.
- [100] D. B. Lobell, «The use of satellite data for crop yield gap analysis,» *Field Crops Research*, 2013.
- [101] W. Kowalik, K. Dabrowska-Zielinska, M. Meroni, T. U. Raczka y A. de Wit, «Yield estimation using SPOT-VEGETATION products: A case study of wheat in European countries,» *International Journal of Applied Earth Observation and Geoinformation*, 2014.
- [102] A. D. Kleshchenko, T. A. Goncharova y T. A. Naidina, «Using the satellite data in dynamic models of crop yield forecasting,» *Russian Meteorology and Hydrology*, 2012.
- [103] L. Ji y A. J. Peters, «Performance evaluation of spectral vegetation indices using a statistical sensitivity function,» *Remote Sensing of Environment*, 2007.
- [104] A. Gusso, J. R. Ducati, M. R. Veronez, D. Arvor, L. Gonzaga y D. Silveira, «Spectral Model for Soybean Yield Estimate Using MODIS/EVI Data *,» *International Journal of Geosciences*, vol. 4, pp. 1233-1241, 2000.
- [105] T. W. Griffin, C. L. Dobbins, T. J. Vyn, R. J. G. M. Florax y J. M. Lowenberg-Deboer, «Spatial analysis of yield monitor data: Case studies of on-farm trials and farm management decision making,» de *Precision Agriculture*, 2008.
- [106] P. Bose, N. K. Kasabov, L. Bruzzone y R. N. Hartono, «Spiking Neural Networks for Crop Yield Estimation Based on Spatiotemporal Analysis of Image Time Series,» *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, vol. 54, nº 11, 2016.
- [107] «THE YIELD FORECASTING PROGRAM OF NASS The Statistical Methods Branch,» 2012.
- [108] Maas S. J., «Using Satellite Data to Improve Model Estimates of Crop Yield,» *Agronomy Journal*, vol. 80, pp. 655-662, 1988.

- [109] M. I. ANDRADE, P. LAPORTA y L. IEZZI, «Sequías en el sudoeste bonaerense : Vulnerabilidad e incertidumbre,» *Revista de Estudios Geográficos*, vol. 5, nº 55, pp. 213-231, 2009.
- [110] C. J. Tucker, B. N. Holben, J. H. Elgin y J. E. McMurtrey Iii, «Relationship of Spectral Data to Grain Yield Variation,» *PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING*, vol. 46, nº 5, pp. 657-666, 1980.
- [111] C. J. Tucker, «Red and photographic infrared linear combinations for monitoring vegetation,» *Remote Sensing of Environment*, 1979.
- [112] K. J. Boote, J. W. Jones y N. B. Pickering «Potential Uses and Limitations of Crop Models (AJ),» *Agronomy Journal*, vol. 88, pp. 704-716, 1996.
- [113] de WIT C. T., «ON COMPETITION,» 1960.
- [114] D. B. Lobell y C. B. Field, «Global scale climate-crop yield relationships and the impacts of recent warming,» *Environmental Research Letters*, 2007.
- [115] D. Makowski, «A simple Bayesian method for adjusting ensemble of crop model outputs to yield observations,» *European Journal of Agronomy*, 2017.
- [116] B. Duchemin, P. Maisongrande, G. Boulet y I. Benhadj, «A simple algorithm for yield estimates: Evaluation for semi-arid irrigated winter wheat monitored with green leaf area index,» *Environmental Modelling and Software*, 2008.
- [117] Zajac CC, Vallejos AG, Galantini JA y Zajac EE, «Análisis del rendimiento de trigo mediante mapas de rendimiento y el NDVI en el SO bonaerense,» 2013.
- [118] UNFCCC, «United Nations Framework Convention on Climate Change. Capitulo_7_Agricultura,» 2006.
- [119] A. Li, S. Liang, A. Wang y J. Qin, «Estimating Crop Yield from Multi-temporal Satellite Data Using Multivariate Regression and Neural Network Techniques,» 2007.
- [120] P. C. Doraiswamy, S. Moulin, P. W. Cook y A. Stern, «Crop Yield Assessment from Remote Sensing,» 2003.
- [121] V. K. Dadhwal, «CROP GROWTH AND PRODUCTIVITY MONITORING AND SIMULATION USING REMOTE SENSING AND GIS,» 2003.
- [122] J. Andrey Zamora Araya y L. Rosibel Tatiana Vallejos Brenes Liceo Mauro Fernández Acuña, «Aprendiendo estadística con R,» 2012.
- [123] E. Paradis, J. A. Ahumada, Q. Agraderle, J. Claude, C. Declercq, E. Gazave, F. Leisch y M. Ros, «R para Principiantes,» 2003.
- [124] D. M. Lambert, J. Lowenberg-Deboer y R. Bongiovanni, «SPATIAL REGRESSION MODELS FOR YIELD MONITOR DATA: A CASE STUDY FROM ARGENTINA,» 2003.
- [125] Ing. Agr. Andrés Méndez, Técnico Juan Pablo Vélez y Técnico Diego Villarreal, «PUNTOS CLAVES PARA LOGRAR UN MAPA DE RENDIMIENTO CON DATOS CONFIABLES,» *Proyecto Maquinas y Agrocomponentes Precisos – INTA Manfredi*, 2012.
- [126] D. Boretto, «RELACIÓN ENTRE LOS ATRIBUTOS DEL TERRENO, LA HUMEDAD DEL SUELO Y LA PRODUCTIVIDAD DE LOS CULTIVOS,» 2016.
- [127] J. G. Bernal, P. Eliecer, D. Díaz, A. Javier, F. M. Monroy, P. Mejías, D. Raes, K. Leuven, E. Fereres y J. F. Boshell, «Uso del modelo AquaCrop para estimar rendimientos para el cultivo de maíz en los departamentos de Córdoba, Meta, Tolima y Valle del Cauca,» 2013.
- [128] J. Aparicio, A. M. Mayoral y J. Morales, «Modelos Lineales Aplicados en R,» 2004.
- [129] I. Ahmad, A. Ghafoor, M. Iftikhar Bhatti, I.-u. Hassan Akhtar y M. Ibrahim, «Satellite Remote Sensing and GIS based Crops Forecasting & Estimation System in Pakistan,» 2014.