



Facultad de Ciencias
**Astronómicas
y Geofísicas**

Universidad Nacional de La Plata
Facultad de Ciencias Astronómicas y Geofísicas

Evaluación de potenciales pronósticos de rendimiento de cultivos en la provincia de Buenos Aires basados en variabilidad climática

Tesis para obtener el grado académico de Licenciado en
Meteorología y Ciencias de la Atmósfera

Autor: Téc. Thomas Maisonnave

Director: Dr. Santiago I. Hurtado

Codirectora: Lic. Laura Aldeco

La Plata, Argentina

febrero de 2024

Agradecimientos

Quiero expresar mi más sincero agradecimiento a las personas que hicieron posible la realización de esta tesis de licenciatura. En primer lugar, mi gratitud eterna va para mi familia. El apoyo incondicional, el amor y la paciencia que me brindaron a lo largo de este arduo camino académico fueron fundamentales para mi éxito. Sus palabras de aliento y comprensión constante fueron las que me ayudaron en los momentos más difíciles. Sin su apoyo, esta tesis no habría sido posible.

Agradezco de manera especial a mis directores, Santiago Hurtado y Laura Aldeco. Su experiencia, orientación y compromiso con mi proyecto fueron esenciales en el desarrollo de esta investigación. Sus consejos sabios y su dedicación me ayudaron a dar lo mejor de mí mismo y a alcanzar un nivel de excelencia que nunca habría logrado por mí solo.

También quiero expresar mi agradecimiento a los amigos que conocí en la facultad: Jorge, Juan y Anahí. Sus amistades me han enriquecido en muchos aspectos y han sido un apoyo invaluable a lo largo de estos años. Juntos compartimos risas, desafíos y experiencias que hicieron más llevadera la carga académica. Gracias por estar siempre allí cuando los necesité.

Este logro no solo es mío, sino también de todas las personas que mencioné anteriormente. Cada uno de ustedes ha sido una pieza fundamental en este rompecabezas, y estoy profundamente agradecido por su presencia en mi vida.

Resumen

La variabilidad climática afecta a la producción agropecuaria, siendo un factor determinante en el rendimiento de los cultivos. Por este motivo, en esta tesis se desarrollaron modelos de pronóstico basados en variables del sistema acoplado océano-atmósfera con el fin de predecir el rendimiento de los principales cultivos de la provincia de Buenos Aires, Argentina. Para lograr esto, primero se caracterizaron los rendimientos de los cultivos de girasol, maíz y trigo durante el periodo 1981-2021, encontrando una tendencia lineal positiva asociada al avance tecnológico. Luego, se regionalizó el comportamiento de cada cultivo para desarrollar modelos de pronóstico para cada una de estas regiones. Posteriormente, se desarrollaron distintos tipos de modelos de pronóstico de los cultivos donde se varió el tipo de modelo, el criterio de selección de predictores y los predictores. Se consideraron como predictores por un lado índices climáticos, y por el otro, variables meteorológicas medidas por estaciones. Los modelos de pronóstico desarrollados fueron evaluados utilizando diferentes metodologías para evaluar su desempeño. Se encontraron distintos modelos de pronóstico con potencial de aplicación para los tres cultivos considerados (trigo, maíz y girasol). Los resultados de la evaluación sugieren que las regresiones lineales múltiples (MLR) y regresión logística multinomial (MULTI) son modelos con buen desempeño, que logran capturar la interacción entre las variables meteorológicas de estación y los índices climáticos con el rendimiento de cada cultivo. Por otra parte, no se destaca ningún criterio de selección sobre otro en cada uno de los modelos.

Índice general

1. Introducción	2
1.1. Motivación y antecedentes	2
1.2. Objetivos	12
2. Datos y Metodología	13
2.1. Datos	13
2.1.1. Variables meteorológicas de estación.	13
2.1.2. Índices de forzantes del sistema acoplado océano-atmósfera	15
2.1.3. Datos de rendimiento de cultivos.	16
2.2. Pre - procesamiento	17
2.2.1. Datos Faltantes	17
2.2.2. Tendencias y estandarización	17
2.2.3. Regionalización	17
2.3. Modelos de pronóstico	18
2.3.1. Regresión lineal múltiple (MLR)	19
2.3.2. Modelo de regresión logística múltiple (MULTI)	20
2.3.3. Arbol de decision (ARBOL)	20
2.3.4. Modelo medio (M.MEDIO)	20
2.4. Evaluación	21
2.4.1. Cross - Validation (Leave one out)	21

2.4.2. Métricas Evaluación	21
2.5. Software utilizado	24
3. Análisis de rendimientos de cultivos	25
3.1. Datos Faltantes	25
3.1.1. Caracterización de datos de rendimiento	26
3.1.2. Tendencias Observadas	29
3.2. Regionalización	32
4. Modelos de pronóstico de rendimiento	34
4.1. Girasol	34
4.1.1. Cluster Sur	34
4.1.2. Cluster Centro	37
4.1.3. Cluster Norte	39
4.2. Maiz	41
4.2.1. Cluster Sur	41
4.2.2. Cluster Este	43
4.2.3. Cluster Oeste	45
4.2.4. Cluster Norte	47
4.3. Trigo	50
4.3.1. Cluster Suroeste	50
4.3.2. Cluster Noreste	52
5. Evaluacion de Modelos	55
5.1. Girasol	55
5.1.1. Evaluación cluster Sur.	55
5.1.2. Evaluación cluster Centro.	56
5.1.3. Evaluación cluster Norte.	57
5.2. Maiz	58

5.2.1. Evaluación cluster Norte	58
5.2.2. Evaluación cluster Oeste	59
5.2.3. Evaluación cluster Este	60
5.2.4. Evaluación cluster Sur	61
5.3. Trigo	61
5.3.1. Evaluación cluster Noreste	61
5.3.2. Evaluación cluster Suroeste	62
6. Conclusiones	64

Capítulo 1

Introducción

1.1. Motivación y antecedentes

La agricultura es el principal pilar de la economía Argentina, lo cual es evidente si se considera que el sector aporta más del 50 % del valor total de las exportaciones y tiene una participación aproximada del 7 % sobre el producto bruto interno (PBI) total (Nogues, 2015, EXPOAGRO). En particular, Argentina está posicionada como uno de los líderes en los mercados internacionales de productos agro-alimenticios (Nogues, 2015). Por otro lado, las exportaciones asociadas al agro, tanto de forma directa como indirecta, aportan 6 de cada 10 dólares de exportaciones totales del país. El territorio argentino cuenta con 37.5 millones de hectáreas (ha) cultivables de las cuales 14.3 millones de ha corresponden a oleaginosas, 11.7 millones a cereales y 10 millones a forrajeras (Azcuay Ameghino & Fernández, 2020). Dentro de esta extensa área cultivable se encuentra la región Pampeana que representa casi la totalidad de la provincia de Buenos Aires, parte de las provincias de Córdoba, Entre Ríos, La Pampa y Santa Fe. Esta región es una de las zonas más productivas del país con su principal foco en la producción de soja, maíz, trigo y girasol (Magrin et al., 2005). La provincia de Buenos Aires (ver figura 1.1) tiene una gran importancia en la producción agropecuaria nacional, pues en sus 37 millones de hectáreas se ubica cerca del 50 % de

la superficie sembrada del país y aquí se concentra aproximadamente el 56% del total de exportaciones bovinas, el 44% de las de cereales y el 32% de las oleaginosas del país [Observatorio de Bioeconomía]

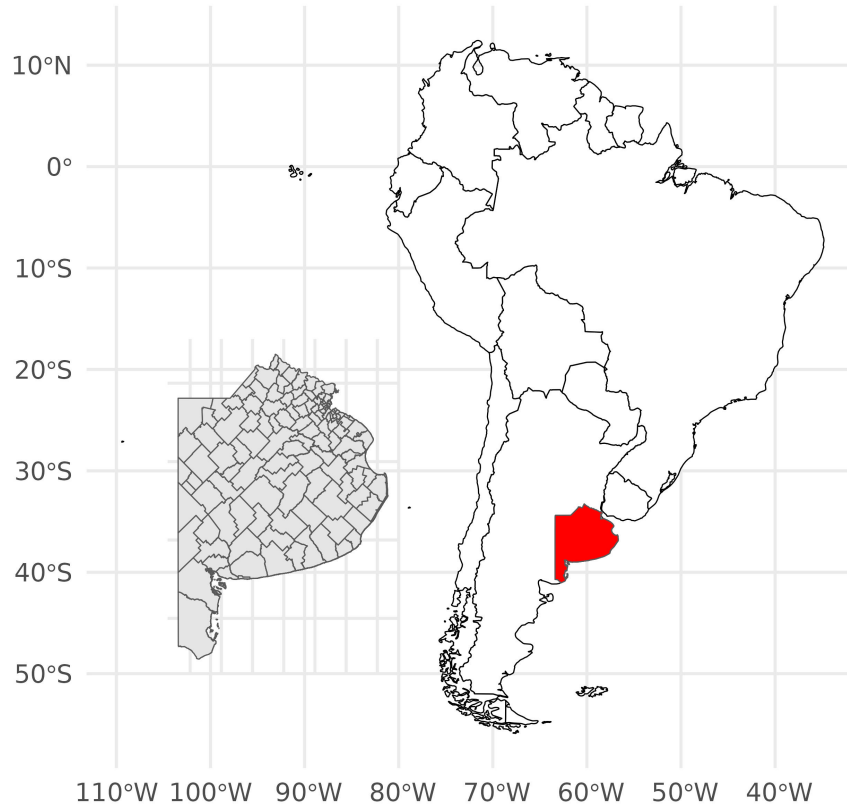


Figura 1.1: Mapa de Sudamérica, en color rojo se muestra la zona de estudio. En forma ampliada se muestra la provincia de Buenos Aires con sus respectiva división de partidos.

La producción agrícola tiene una fuerte dependencia de las condiciones meteorológicas durante la temporada de crecimiento de los cultivos, lo que puede traducirse en resultados positivos o negativos, dependiendo de si las condiciones atmosféricas son propicias o no (de la Casa et al., 2021). Las distintas etapas de crecimiento de un cultivo están definidas por las características que el mismo presenta desde su germinación hasta su muerte, las cuales no solo están fuertemente influenciadas por las condiciones climáticas sino también por las condiciones del suelo. A cada una de estas etapas se la conoce como estado fenológico del cultivo y depende exclusivamente de cada cultivo.

Las amenazas climáticas relacionadas con diversos fenómenos atmosféricos nocivos como la falta o el exceso de agua asociados a las variaciones en la precipitación tienen una correlación directa con el nivel de riesgo en el que se desarrolla la actividad agrícola de una región (Ferreyra et al., 2001). La variabilidad climática se presenta como un factor negativo en los sistemas agrarios, y la incapacidad de predecir o adaptarse a ella ha llevado al desarrollo de esquemas de gestión más conservadores en el sentido de no asumir riesgos. Al no asumir riesgos, un enfoque conservador tiene una eficacia limitada para amortiguar condiciones adversas graves, así como también se pierde la oportunidad potencial de capitalizar la variabilidad estacional cuando es favorable (Meinke & Stone, 2005). La influencia de los factores climáticos sobre el rendimiento final del cultivo no solo depende de las características de estos factores, sino también, de en qué momento del periodo de crecimiento del cultivo se produzcan estos eventos. La Figura 1.2 muestra los periodos de siembra y cosecha, junto con los periodos críticos para los cultivos de girasol, maíz y trigo. Estos periodos pueden variar algunas semanas debido a que para sembrar y cosechar se necesitan ciertas condiciones óptimas.

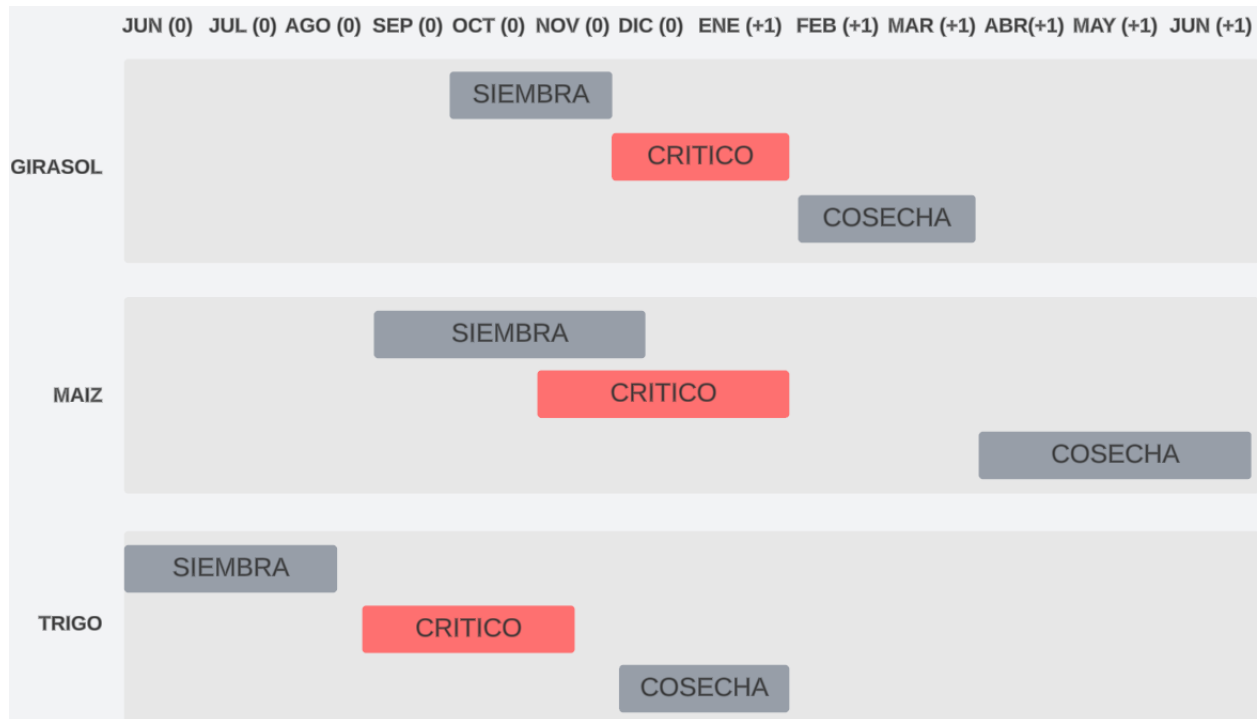


Figura 1.2: Periodos de siembra y cosecha en tonos grises de los cultivos de girasol, maíz y trigo. En color rosado los periodos críticos en el crecimiento de los cultivos de girasol, maíz y trigo. Se denota como (0): año inicial, (+1): año siguiente. Adaptado de [ORA](#)

[Anderson et al. \(2017\)](#) destacan la importancia de la memoria que presenta el suelo en conservar la disponibilidad de agua para su posterior uso por los cultivos. Asimismo, [Anastasi et al. \(2000\)](#) encuentran que es de vital importancia el estado del suelo antes de la fecha de siembra de cualquier cultivo, coincidente con lo encontrado por [Anderson et al. \(2017\)](#). Además de ser importante las condiciones del suelo previas a la siembra del cultivo, también son de suma importancia durante los periodos críticos de crecimiento. Los cultivos de maíz, girasol y trigo presentan distintas respuestas a las anomalías climáticas ocurridas en la región de estudio. [Andrade et al. \(1999\)](#) muestran que uno de los periodos más importantes en la fenología del maíz es el de floración, que es crítico a la hora de que se desarrolle el rendimiento final del maíz. [Wilhelm et al. \(1999\)](#) encuentra que durante los periodos de floración, las altas temperaturas podrían afectar el proceso de polinización y por ende el rendimiento final del maíz. [Lobell et al. \(2013\)](#) muestran cómo un aumento en la temperatura media diaria y semanal durante la etapa de floración podría afectar el

rendimiento final del maíz a causa de un aumento en la evaporación de agua del suelo, provocando un aumento de la demanda de precipitación en la etapa crítica para definir el rendimiento final del maíz. La disponibilidad de agua en el suelo es de vital importancia en el periodo crítico para el cultivo de maíz, tanto sea por las variaciones en la evaporación (moduladas por los máximos de temperatura) o por la variación en las precipitaciones. [Rebella et al. \(1984\)](#) y [AJ \(1992\)](#) encuentran asociaciones entre el rendimiento final del maíz y las precipitaciones superiores (inferiores) a lo normal entre octubre y febrero, típicas de los episodios cálidos (fríos) del El Niño-Oscilación del Sur (ENOS) en la provincia de Buenos Aires.

El cultivo de girasol se ve influenciado principalmente por las anomalías de precipitación. [Podestá et al. \(1999\)](#) encuentran una menor frecuencia de rendimientos bajos cuando se producen periodos de anomalías negativas de precipitación sobre la región de estudio. Esto podría deberse a que las condiciones húmedas favorecen la generación de enfermedades en el girasol, afectando negativamente a su rendimiento ([Travasso et al., 2009](#)). Esto es coincidente con lo estudiado por [Chapman & De la Vega \(2002\)](#) quienes encuentran que la variación en el rendimiento del girasol podría estar asociada a las variaciones estacionales de precipitación durante los meses cálidos, donde este cultivo se encuentra en su etapa crítica ([Sierra & Murphy, 1986](#)). Por otro lado, el cultivo de trigo es cosechado a partir de mediados de noviembre, y su etapa crítica comienza en septiembre ([Slafer & Rawson, 1994](#)). [Barkley et al. \(2014\)](#) muestran una correlación positiva entre las variables que afectan la disponibilidad de agua en el suelo y el rendimiento final del trigo durante los meses de otoño, es decir antes de la siembra del mismo que comienza en junio. De manera similar al girasol, los excesos de precipitación generan exceso de humedad en el suelo, que favorece el desarrollo de enfermedades en el cultivo de trigo, disminuyendo su rendimiento final ([Anderson et al., 2017](#)).

Las anomalías de temperatura también afectan el rendimiento final del trigo; [Warrick & Miller \(1999\)](#) y [Musick et al. \(1994\)](#) muestran la importancia de las temperaturas mínimas durante el periodo crítico de este cultivo, debido a que la ocurrencia de heladas afectan desfavorablemente el rendimiento final.

Las distintas variables que afectan el crecimiento óptimo de los cultivos están moduladas por diferentes forzantes de escala interanual del sistema acoplado océano - atmósfera, de los cuales el más estudiado es el fenómeno ENOS ([Hurtado, 2022](#)). El ENOS es el mayor forzante del clima en Sudamérica ([Hurtado et al., 2023](#); [Cai et al., 2020](#); [Grimm et al., 2000](#), entre otros). Este fenómeno acoplado del sistema océano-atmósfera en el océano Pacífico ecuatorial, está caracterizado por periodicidades de 3-5 años entre su fase cálida (El Niño) y fría (La Niña) ([Wang et al., 2017](#)). Las fases del ENOS tienen un ciclo de vida con anomalías positivas (negativas) que comienzan en invierno y finalizan en otoño del año siguiente ([Berrí et al., 2002](#); [Hurtado & Agosta, 2021](#)). Por otro lado, distintos autores ([Gonzalez & Vera, 2014](#); [Casarin & Kousky, 1986](#); [Nogués-Paegle & Mo, 1997](#), entre otros) muestran un comportamiento dipolar de las anomalías de precipitación en Sudamérica. Este dipolo consiste en dos fases: una fase positiva asociada a un aumento de las precipitaciones hacia el sur y sudeste de América del Sur, acompañado de un debilitamiento de la Zona de Convergencia del Atlántico Sur (SACZ, por sus siglas en inglés). En la fase negativa, se observa una intensificación de la SACZ, acompañada de una disminución de las precipitaciones en las llanuras subtropicales. La escala temporal de este dipolo oscila entre submensual e interanual, en asociación con diferentes factores como: cambios en la corriente en chorro de niveles bajos en América del sur (SALLJ; [Montini et al. \(2019\)](#)), trenes de ondas de Rossby submensuales ([Van Der Wiel et al., 2015](#)), la Oscilación Madden - Julian (MJO; [Alvarez et al. \(2016\)](#)) y el ENOS ([Hurtado et al., 2023](#)). Diversos estudios ([Hurtado et al., 2024](#); [Grimm, 2003](#); [Grimm & Zilli, 2009](#); [Tedeschi et al., 2015](#), entre otros) coinciden en que durante la primavera austral y principios de verano (octubre-diciembre) el ENOS modula la precipitación en la región de Sudamérica, alterando el comportamiento de este dipolo. Esta modulación

viene dada por la propagación de un tren de ondas de Rossby extratropicales que atraviesa el Pacífico Sur y favorece anomalías en la circulación en niveles bajos, modulando el dipolo de precipitación de Sudamérica. Los eventos El Niño se asocian con convección aumentada en el este de Argentina subtropical y convección inhibida en la región de la SACZ, mientras que durante los eventos La Niña ocurre el efecto contrario ([Hurtado & Agosta, 2021](#)).

El ENOS además de ser el principal forzante de la precipitación en la región de estudio también tiene efectos sobre la temperatura en dicha región. [Rusticucci & Vargas \(2002\)](#) encuentran que durante los eventos La Niña todos los periodos bimensuales (excepto agosto-septiembre) presentan una mayor probabilidad de ocurrencia de valores de temperatura fuera de lo normal, con temperaturas más bajas en invierno y más altas en verano. Por el contrario, los diferentes episodios de El Niño son más heterogéneos y por lo tanto, sus efectos sobre la temperatura son más variados.

Otro forzante del clima en la región es el modo dipolar del océano Índico (IOD, por sus siglas en Inglés; [Saji & Yamagata \(2003\)](#)). El IOD es un gradiente anómalo de temperatura de la superficie del océano entre el océano Índico ecuatorial occidental y el océano Índico ecuatorial sudoriental. Durante un IOD positivo el tren de ondas inducido desde el océano Índico provoca un aumento del transporte de humedad hacia la Cuenca del Plata, lo que da lugar a anomalías de humedad en la región y posibles aumentos en las precipitaciones ([Sena & Magnusdottir, 2021](#)). Por otro lado, se ha encontrado que durante la primavera austral (periodo de mayor influencia del índice IOD) existen correlaciones positivas entre la temperatura y el IOD para la región de estudio ([Saji et al., 2005](#)). Estudios recientes han demostrado que el ENOS y el IOD pueden coexistir de forma tal que el IOD positivo tiende a ocurrir junto a la fase cálida del ENOS, reforzando sus impactos en la precipitación, y el IOD negativo junto a la fase fría del ENOS ([Sun et al., 2015](#); [Wang & Wang, 2014](#)).

De todos modos hay eventos IOD que ocurren de manera independiente a alguna de las fases del ENOS y que representan un 32 % de los casos según [Stuecker et al. \(2017\)](#), aunque su impacto en Sudamérica subtropical se ve disminuido cuando ocurren de forma independiente ([Andrian L. G & M., 2022](#)).

Respecto del océano Atlántico, diferentes estudios han demostrado que existen dos forzantes con influencia en el clima de Sudamérica. Uno de ellos está definido por las variaciones de la temperatura superficial en una región del mismo ([Taschetto & Wainer, 2008](#); [Jorgetti et al., 2014](#); [Barrucand et al., 2008](#)). Sus efectos se han estudiado mediante la definición de un índice que se basa en la anomalía de temperatura de la superficie del mar (SST) en el océano Atlántico tropical (TSA, por sus siglas en inglés). Según [Melice & Servain \(2003\)](#), este es un modo de variabilidad climática conocido por estar asociado a distintos regímenes de anomalías de precipitación sobre Sudamérica. El otro forzante que afecta Argentina es el Dipolo Oceánico del Atlántico Sur (SAODI, por sus siglas en inglés). Este dipolo consiste en anomalías de SST positivas en el noreste y negativas en el sudoeste del océano Atlántico tropical Sur. [Nnamchi et al. \(2011\)](#) encontraron que los episodios activos del SAODI mostraron un acople con las anomalías de circulación de la atmósfera y precipitación durante los meses de mayo a agosto. [Garbarini & Gonzalez \(2014\)](#) encontraron correlaciones negativas entre el índice SAODI y la precipitación en la región central de Argentina durante el invierno y primavera austral.

El Modo Anular Austral (SAM) es el principal modo de variabilidad de la circulación extratropical en el hemisferio sur y se caracteriza por estructuras zonalmente simétricas o anulares, con perturbaciones de la altura geopotencial de signos opuestos entre la Antártida y latitudes medias ([Thompson & Wallace, 2000](#)). Diferentes estudios muestran que este forzante climático presenta correlaciones positivas con la precipitación en la región central de Argentina durante el invierno austral ([Marshall, 2003](#); [Thompson & Wallace, 2000](#); [Silvestri & Vera, 2003](#)). Por otro lado, [Reboita et al. \(2009\)](#) evaluaron las anomalías de precipitación estacional sobre Sudamérica para diferentes fases de SAM durante el periodo

de 1980 - 1999. En la fase negativa del SAM, se observaron anomalías de precipitación positivas sobre el sur de Brasil, Uruguay y el centro y norte de Argentina, principalmente en verano y otoño. Las precipitaciones máximas se produjeron durante el otoño, extendiéndose a todo el sur de Sudamérica. El SAM también actúa como forzante para las anomalías de temperatura sobre la región central de Argentina. [Vasconcellos et al. \(2019\)](#) encontraron que para los meses de febrero a mayo se producen anomalías negativas significativas de temperatura en la fase positiva de la SAM. Este comportamiento se vuelve a repetir entre septiembre y noviembre, mientras que en los meses de junio a agosto no se ve una anomalía clara de temperatura asociada a la SAM.

Los tomadores de decisión del sector agrícola suelen tener en cuenta las variaciones del clima y su posible evolución para definir la fecha de siembra o la variedad del cultivo a sembrar para una determinada campaña ([Bert et al., 2006](#)). La respuesta del suelo frente a la precipitación y la temperatura condiciona el rendimiento de los cultivos ([Pántano et al., 2014](#)), con lo cual no sólo es importante comprender los forzantes climáticos que favorecen la variabilidad de estas variables, sino que también es necesario poder anticiparse de manera tal que los usuarios puedan tomar decisiones que les permitan maximizar el rendimiento de los cultivos ([Oram, 1989](#)). Por estos motivos, es claro que utilizar un modelo de pronóstico de rendimiento de cultivos que no sólo se base en las variables atmosféricas como la precipitación y la temperatura, sino también en los índices climáticos mencionados, puede ser de gran ayuda en ese sentido. Es por ello que esta tesis se orientó al desarrollo de modelos de pronósticos de rendimientos de cultivos basados en estos modos de variabilidad climática. Los principales modelos utilizados para estimar el rendimiento de los cultivos bajo diferentes inferencias son: empíricos, mecanísticos, funcionales o modelos orientados en procesos ([Link, 2005](#)). Los modelos empíricos funcionan identificando términos significativos desde el punto de vista físico del problema y realizan un proceso estadístico-matemático hasta encontrar relaciones que representen adecuadamente una característica del cultivo con los parámetros escogidos ([De Wit & Van Keulen, 1987](#)). En los modelos mecanísticos,

los procesos que ocurren en el sistema forman la base del modelo. El crecimiento del cultivo expresado en este sentido está construido sobre la base del conocimiento de los procesos fisiológicos, físicos y químicos inherentes y el efecto de los factores ambientales sobre estos (Guarín Giraldo, 2011). Los modelos funcionales son definidos como modelos que incorporan un acercamiento simplificado para describir los procesos más complejos. Estos modelos no dependen de muchos parámetros y por lo tanto simplifican los procesos descritos mediante un acercamiento mecanístico (Link, 2005; Brisson et al., 1998; Stöckle et al., 2003). Los modelos basados en procesos fueron desarrollados con la finalidad de simular sistemas complejos, como son los sistemas de cultivos donde existen relaciones con el clima, las prácticas agrícolas, las características del suelo, etc. Por lo que este tipo de modelo contiene una combinación de diferentes tipos de modelos (Link, 2005; Stöckle et al., 2003). En este sentido los modelos empíricos permiten al desarrollador introducir parámetros significativos desde el punto de vista físico del problema hasta encontrar la mejor predicción posible, así como también, su facilidad y su bajo costo de aplicación permiten desarrollar varios modelos con distintos enfoques y evaluar su desempeño. Por estos motivos, se decidió utilizar modelo empíricos para el desarrollo de esta tesis

1.2. Objetivos

Por lo anteriormente expuesto el objetivo general de esta tesis es estudiar posibles modelos estadísticos de pronóstico de rendimiento de los principales cultivos de la provincia de Buenos Aires basados en variabilidad climática. Para esto se plantearon los siguientes objetivos específicos:

- Objetivo Específico 1: (Capítulo 3) Caracterizar estadísticamente el rendimiento de los principales cultivos y agrupar los partidos en subregiones de común variabilidad inter-anual.
- Objetivo Específico 2: (Capítulo 4) Diagramar potenciales modelos de pronóstico basados en variables meteorológicas observadas a nivel de estación, en función de su relación con las variaciones del rendimiento de cultivos a nivel de subregiones.
- Objetivo Específico 3: (Capítulo 4) Diagramar potenciales modelos de pronósticos basados en índices de variabilidad del sistema acoplado océano-atmósfera.
- Objetivo Específico 4: (Capítulo 5) Evaluar y comparar los modelos de pronóstico confeccionados a partir de los resultados obtenidos en el OE 2 y 3, a fin de determinar los mejores modelos.

Capítulo 2

Datos y Metodología

En el presente capítulo se detallan los datos utilizados y las metodologías que se han implementado para el desarrollo de la presente tesis.

2.1. Datos

2.1.1. Variables meteorológicas de estación.

Para el desarrollo de esta tesis se utilizaron datos del Servicio Meteorológico Nacional (SMN) de 18 estaciones meteorológicas distribuidas en la provincia de Buenos Aires, desde 1981 hasta 2021. Las variables utilizadas fueron la precipitación acumulada mensual (PP) y las temperaturas máxima media, mínima media y media mensuales en el período mencionado. Por otro lado, del SMN también se obtuvieron datos diarios de evapotranspiración real (ETR) y evapotranspiración potencial (ETP) calculada a partir del modelo BHOA (Long et al., 2013). En la Figura 2.1 se observa la distribución de estaciones con los datos anteriormente mencionados.

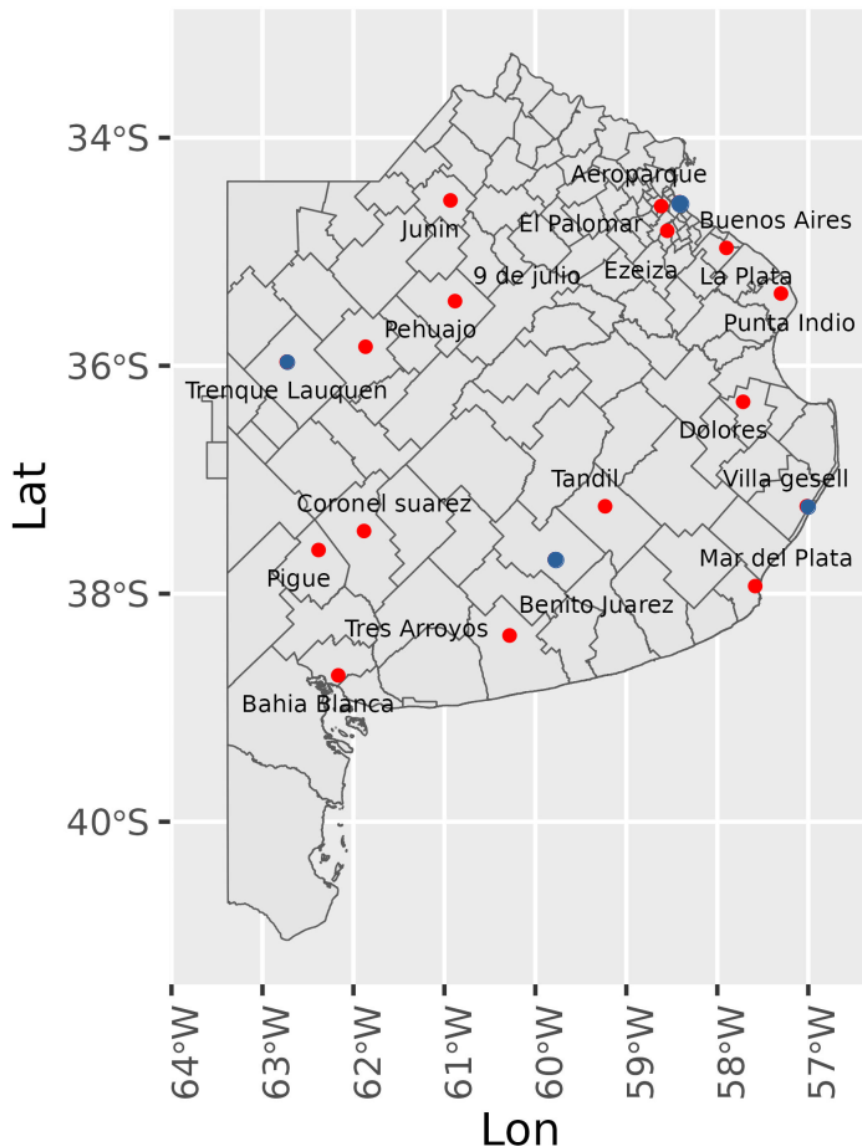


Figura 2.1: Distribución de estaciones meteorológicas utilizadas en la provincia de buenos aires. En colores rojos las estaciones que presentan datos de temperatura máxima, mínima y media junto con datos de precipitación y evapotranspiración. En color azul estaciones que solo presentan datos de evapotranspiración

2.1.2. Índices de forzantes del sistema acoplado océano-atmósfera

Para esta tesis se utilizaron datos de temperatura de la superficie del océano de ERSSTv5 (Huang et al., 2017) de la NOAA (National Oceanic and Atmospheric Administration). El conjunto de datos ERSST es un análisis mensual de la temperatura global de la superficie del mar. La producción de ERSST se realiza en una cuadrícula de 2° - 2° Latitud-Longitud, con resolución espacial mejorada mediante métodos estadísticos. Este conjunto de datos mensuales comienza en enero de 1854 y continúa hasta el presente, con actualizaciones mensuales. Huang et al. (2018) realizaron una validación de este análisis, encontrando buena consistencia con las observaciones. Además, esta base de datos es ampliamente utilizada para investigaciones climáticas y diferentes aplicaciones (Seager et al., 2019; Simpson et al., 2018; Cowtan et al., 2018). Por otro lado, al actualizarse mensualmente estos datos pueden ser usados en herramientas operativas.

A través de los datos de la superficie del océano de ERSSTv5 se calcularon los siguientes índices climáticos :

- Dipolo Oceánico del Atlántico Sur (SAODI): consiste en el calentamiento de las aguas superficiales frente a las costas de África ecuatorial occidental/central acompañado de un enfriamiento de magnitud similar frente a las costas de Argentina - Uruguay - Brasil (Nnamchi et al., 2011). Una medida simple del SAODI se define diferenciando la anomalía normalizada de temperatura superficial del mar promediada en el dominio de los dos centros de calentamiento y enfriamiento. Estos dominios se describen por su ubicación en el océano Atlántico Sur como el polo noreste (10°E - 20°W , 0° - 15°S) y el polo suroeste (10°W - 40°W , 25°S - 40°S).
- Índice del Atlántico Sur Tropical (TSA): es la anomalía media mensual de temperatura de la superficie del océano promediada en la región 30°W - 10°E , 20°S - 0° .
- Índice Niño 1+2: es la anomalía de la TSM del pacífico tropical extremo oriental,

promediada en la región $0^{\circ} - 10^{\circ}\text{S}$, $90^{\circ}\text{W} - 80^{\circ}\text{W}$.

- Índice Niño 4: es la anomalía de la TSM del pacifico tropical central, promediada en la región $5^{\circ}\text{N} - 5^{\circ}\text{S}$, $160^{\circ}\text{E} - 150^{\circ}\text{W}$.
- Índice Niño 3: es la anomalía de la TSM del pacifico tropical oriental, promediada en la región $5^{\circ}\text{N} - 5^{\circ}\text{S}$, $150^{\circ}\text{W} - 90^{\circ}\text{W}$.
- Índice del modo dipolar del océano Índico (IOD; Saji & Yamagata (2003)): es un gradiente anómalo de temperatura de la superficie del océano entre el océano Índico ecuatorial occidental ($50^{\circ}\text{E} - 70^{\circ}\text{E}$, $10^{\circ}\text{S} - 10^{\circ}\text{N}$) y el océano Índico ecuatorial sudoriental ($90^{\circ}\text{E} - 110^{\circ}\text{E}$, $10^{\circ}\text{S} - 0^{\circ}\text{N}$).
- Modo anular del sur (SAM): este índice es calculado a partir de diferenciar el promedio de la presión a nivel medio del mar entre seis estaciones en torno a los 40°S y seis estaciones al sur de los 60°S , utilizando la definición propuesta por Marshall (2003). El índice calculado con esta metodología es provisto por el "British Antarctic Survey" y está disponible online (<https://legacy.bas.ac.uk/met/gjma/sam.html>).

2.1.3. Datos de rendimiento de cultivos.

En el Ministerio de Agricultura, Ganadería y Pesca (MAGYP) se encuentran disponibles los datos de rendimientos de cultivos para cada uno de los partidos de la provincia de Buenos Aires. En ese sentido, en esta tesis se utilizaron también datos de rendimientos de los cultivos de trigo, maíz y girasol desde 1981 hasta 2021. Estos datos resultan del promedio de rendimiento en kilogramos por hectárea de cada partido.

2.2. Pre - procesamiento

2.2.1. Datos Faltantes

El periodo de estudio es entre 1981 y 2021, por lo tanto, como primer paso se realizó un control de calidad para evaluar los datos faltantes de cada una de las estaciones observacionales. Las estaciones que presentaban más de un 5 % de datos faltantes en el período de estudio fueron desestimadas, mientras que las estaciones que presentaban menos del 5 % de datos faltantes fueron rellenadas utilizando regresión lineal múltiple para obtener series continuas ([Hurtado et al., 2021](#)).

2.2.2. Tendencias y estandarización

La tendencia es un cambio gradual de un conjunto de datos medidos en el tiempo, haciendo que la muestra pierda estacionariedad, lo cual es indeseable para una correcta aplicación de análisis estadísticos y dinámicos. Por lo tanto, en esta tesis se aplicó el test estadístico de Mann-Kendall ([Gocic & Trajkovic, 2013](#)) con un nivel de significancia estadística de 0.1, este test se utiliza para determinar si nuestra serie de datos presenta un aumento o una disminución sistemática del valor del dato. Aquellas tendencias significativas no nulas fueron posteriormente filtradas utilizando residuos de la regresión lineal de los datos, para asegurar la estacionariedad de los datos. Por otro lado, las variables fueron estandarizadas para poder ser comparadas entre sí, este proceso se realizó restando la media y dividiendo por la varianza de la serie.

2.2.3. Regionalización

Basándonos en que las variaciones climáticas son regionales, y por lo tanto la inferencia que podría llegar a tener el clima en el rendimiento de los cultivos es regional también, en esta tesis se decidió regionalizar la región de estudio alineándose con este principio. Para

ello, se realizó una clasificación espacial de partidos que presentan características similares en su comportamiento, con el fin de sub-regionalizar la región de estudio en grupos espaciales que tengan un comportamiento similar en sus series de tiempo. Esta metodología se realizó para los tres cultivos disponibles agrupándolos mediante el método K-means (Hartigan & Wong, 1979). Por otro lado, la cantidad de clusters o grupos óptima se calculó a partir del método "silhouette", ya que este método mide tanto la cohesión como la separación de clusters (Dinh et al., 2019). La cohesión es una medida de cuán similares son los rendimientos de una región, mientras que la separación mide qué tan diferentes son los datos de rendimientos en las distintas regiones (Kaufman & Rousseeuw, 2009).

Una vez definidos los cluster para cada cultivo se procedió a calcular un índice I de cultivo representativo para cada cluster.

$$I(t) = \frac{\sum_{i=1}^n P_i(t)}{N} \quad (2.1)$$

Donde P_i representa el valor estandarizado de rendimiento de cada partido dentro del cluster mientras que N representa la cantidad total de partidos del cluster. Los índices para cada cluster fueron calculados utilizando la media de todos los partidos contenidos en el cluster como se muestra en la ecuación 2.1, finalmente, para cada cultivo tendremos índices representativos para cada cluster.

2.3. Modelos de pronóstico

Un modelo es un sistema que se desarrolla para intentar predecir la realidad de la forma más exacta posible. En esta tesis se van a desarrollar modelos de predicción de rendimiento de los cultivos de girasol, maíz y trigo donde las variables utilizadas como "input" (entrada) varían entre datos de estaciones meteorológicas e índices climáticos, seleccionados por las metodologías que se detallan a continuación. Las estaciones meteorológicas utilizadas como "input" fueron las cercanas a cada región obtenida. La cantidad máxima de predictores se

definió en 5 para evitar posible sobreajuste de los modelos. La predicción de rendimientos se va a realizar a partir de predecir tres categorías acordes al rendimiento en kilogramos por hectárea cosechada:

- Rendimiento Alto: cuando el rendimiento es mayor al percentil 75
- Rendimiento Normal: cuando el rendimiento es entre el percentil 75 y 25
- Rendimiento Bajo: cuando el rendimiento es menor al percentil 25.

Estas tres clasificaciones van a ser las que devuelva el modelo ('output') una vez ingresado el input mencionado.

2.3.1. Regresión lineal múltiple (MLR)

El primer modelo empleado fue el modelo de regresión lineal múltiple ([Hurtado et al., 2021](#)), donde el predictando es el valor de rendimiento, y los múltiples predictores son las variables meteorológicas obtenidas a partir de los datos de estación. Para determinar cuáles son las variables meteorológicas que pueden ser posibles predictores, se utilizó el criterio de información bayesiano (BIC, por sus siglas en inglés), este criterio es una de las herramientas más conocidas y utilizadas en la selección de modelos estadísticos ([Neath & Cavanaugh, 2012](#)). Además del criterio de información bayesiano, se seleccionaron los predictores realizando un análisis de sensibilidad (S). El procedimiento de este análisis consiste en tomar las variables meteorológicas en tiempos mensuales, bimensuales y trimestrales que mayor correlación tienen con el rendimiento de los cultivos. Para el cálculo de correlación se utilizó la correlación del primer momento de Pearson ([Wilks, 2011](#)).

2.3.2. Modelo de regresión logística múltiple (MULTI)

Por otro lado, se empleó el modelo de regresión logística multinomial ([LaValley, 2008](#)). Un modelo de regresión logística se basa en las probabilidades de ocurrencia de un resultado de interés o evento categórico, es decir la predicción de una variable no continua. El criterio de selección utilizado es el mismo que en el modelo de regresión lineal múltiple, es decir el criterio de información bayesiano (BIC).

2.3.3. Arbol de decision (ARBOL)

Otro de los modelos empleados es el de árbol de decisión ([Griva & Basualdo, 2018](#)). Los árboles de decisión se basan en definir umbrales para determinar la categoría de un dado registro. Cada umbral definido es un nodo donde se abren dos ramas, una si supera el umbral y otra si no lo supera. En este sentido, cada árbol de decisión tiene un nodo inicial de partida de donde surgen dos ramas, de cada rama puede haber un nuevo nodo del cual surgen nuevamente dos ramas. El algoritmo concluye cuando se llega a una rama que termina en una decisión (por ejemplo, “es categoría 3”) en vez de en un nuevo nodo. Para definir el nodo (qué variable utilizar y con qué umbral), se utilizó el índice de “GINI”. Este índice busca que la separación generada por el nodo forme los dos grupos más homogéneos posibles en términos de la categoría a determinar (predictando). Se determinó usar este criterio debido a que minimiza la clasificación errónea ([Gulati et al., 2016](#)).

2.3.4. Modelo medio (M.MEDIO)

Para complementar los modelos mencionados en las secciones anteriores se decidió utilizar el modelo “Medio”. Este modelo se basa en utilizar las predicciones estimadas por todos los modelos anteriormente mencionados y calcular su valor medio, de manera similar a un ensamble, y teniendo así un modelo más para su posterior evaluación.

2.4. Evaluación

2.4.1. Cross - Validation (Leave one out)

A la hora de entrenar y posteriormente evaluar nuestro modelo de predicción se decidió utilizar el método de “leave one out” de validación cruzada (LOOCV, por sus siglas en inglés). Esta metodología considera una observación de toda la muestra para la evaluación del modelo y los $n - 1$ valores restantes los utiliza para el entrenamiento del modelo. Luego se cambia la observación a considerar para la evaluación y se realiza este proceso n veces, donde n es igual al largo de toda la muestra (Shao & Er, 2016).

2.4.2. Métricas Evaluación

Para la comparación entre el dato modelado y el dato observado se utilizaron métricas calculadas a partir de una matriz de confusión que es una tabla de contingencia (Tabla 2.1) entre categorías predichas y observadas (Irizarry, 2019).

		Observaciones			Total
		Bajo	Normal	Alto	
Valor Predicho	Bajo	Acierto	No Acierto	Error_sev_1	N Bajo Predicho
	Normal	No Acierto	Acierto	No Acierto	N Normal Predicho
	Alto	Error_sev_3	No Acierto	Acierto	N Alto Predicho
Total		N Bajo Observado	N Normal Obsevado	N Alto Observado	N TOTAL

Tabla 2.1: En tonos Naranjas se representa la matriz de confusión A. Las columnas representan la cantidad de años observados con rendimientos bajos, normales y altos mientras que las filas representan la cantidad de años pronosticados con la misma clasificación.

La matriz de confusión es una matriz que permite comparar la cantidad de años que el modelo predijo de manera correcta una categoría o no, respecto de las observaciones en las mismas categorías. Las matrices de confusión se calcularon para cada modelo y para cada cluster de cada uno de los cultivos. Una matriz de confusión de un modelo perfecto solo tendría los valores para cada categoría sobre la diagonal de la matriz mientras que en las posiciones restantes tendría ceros. El caso en el cual un modelo no logra predecir ningún resultado sería el de una matriz con ceros en su diagonal.

A la hora de identificar qué modelos predicen mejor el rendimientos de los cultivos, se decidió utilizar diferentes métricas de evaluación que se calculan a partir de la matriz de confusión (A) de cada modelo, para cada cluster y cultivo.

Métricas de evaluacion	
Sensibilidad	$\frac{A[i, i]}{sum(A[, i])}$
Error-sev-1	$\frac{A[1, 3]}{sum(A[1, i])}$
Error-sev-3	$\frac{A[3, 1]}{sum(A[3, i])}$
Precision	$\frac{sum(A[i, i])}{sum(A)}$

Tabla 2.2: Métricas de evaluación calculadas. A representa la matriz de confusión y el subíndice i representa cada una de las posiciones de la matriz. Los dos puntos en la dimensión implican todos los valores de fila/columna

La Tabla 2.2 muestra las distintas métricas de evaluación utilizadas para determinar la fiabilidad de los modelos desarrollados en esta tesis. La primera métrica se define como la Precisión y es la suma de aciertos en la predicción dividido por la cantidad de observaciones totales. Otra de las métricas utilizadas es la **sensibilidad** (Sen (i)) la cual se define como la cantidad de predicciones correctas de cada categoría, dividido el total de observaciones de esa categoría. El subíndice i indica a qué categoría está asociada dicha sensibilidad, donde el 1 indica rendimientos bajos, el 2 indica rendimientos normales y 3 indica rendimientos altos. Las dos métricas restantes utilizadas se definieron para determinar los errores severos (**Error_sev_1/3**) en los modelos. Estos errores están definidos como 1: la cantidad de veces que se predijeron rendimientos bajos y las observaciones mostraron rendimientos altos y 3: la cantidad de veces donde se predijeron rendimientos altos y las observaciones mostraron rendimientos bajos. Estas últimas dos métricas representan la posibilidad de que un modelo tenga errores severos en sus predicciones, indicando en ese caso que tiene un desempeño bajo y por lo tanto que sea poco adecuado para su uso. A la hora de discretizar qué modelo representa de mejor manera el rendimiento de los cultivos, se definió un umbral para la precisión y la sensibilidad de 0.50. Es decir, que los modelos cuyas métricas estén por encima de este valor se consideran buenos. Los modelos que no cumplan dicho criterio se consideran de bajo desempeño y por lo tanto no serían útiles para obtener un pronóstico de calidad. Por otro lado, para el caso de los errores severos, se espera que el modelo más apto en la predicción de rendimiento sea el que tenga valores cercanos o iguales a cero en ambas de estas métricas.

2.5. Software utilizado

Todos los algoritmos implementados, cálculos, y análisis estadísticos fueron computados en R, lenguaje de programación de software libre ([Team, 2010](#)). Dentro de este software se utilizaron los siguientes paquetes para distintos cálculos específicos: Tidyverse ([Wickham et al., 2019](#)), Lubridate ([Grolemund & Wickham, 2011](#)), Factoextra ([Kassambara, 2016](#)), Cluster ([Maechler et al., 2013](#)), Rpart([Therneau et al., 2015](#)), Nnet ([Ripley et al., 2016](#))). A la hora de realizar gráficos se utilizó el paquete ggplot2 ([Wickham, 2011](#))

Capítulo 3

Análisis de rendimientos de cultivos

3.1. Datos Faltantes

En la Figura 3.1 se muestra la provincia de Buenos Aires con todos sus partidos en colores rojos y verdes. Los partidos pintados en color rojo fueron desestimados por tener un total de datos faltantes mayor al 5 %, mientras que los partidos que se encuentran en tonos de verde fueron los retenidos para utilizar en esta tesis. En este último grupo se encuentran tanto los partidos sin datos faltantes en el período de estudio como los que fueron completados mediante la metodología de rellenado mencionada en el Capítulo 2. Se puede observar que en general los partidos que presentan mayor cantidad de datos faltantes son los que se encuentran cerca de los grandes conglomerados urbanos. El girasol es uno de los cultivos que presenta mayor cantidad de partidos desestimados, mientras que el trigo es el que presenta mayor cantidad de partidos disponibles.

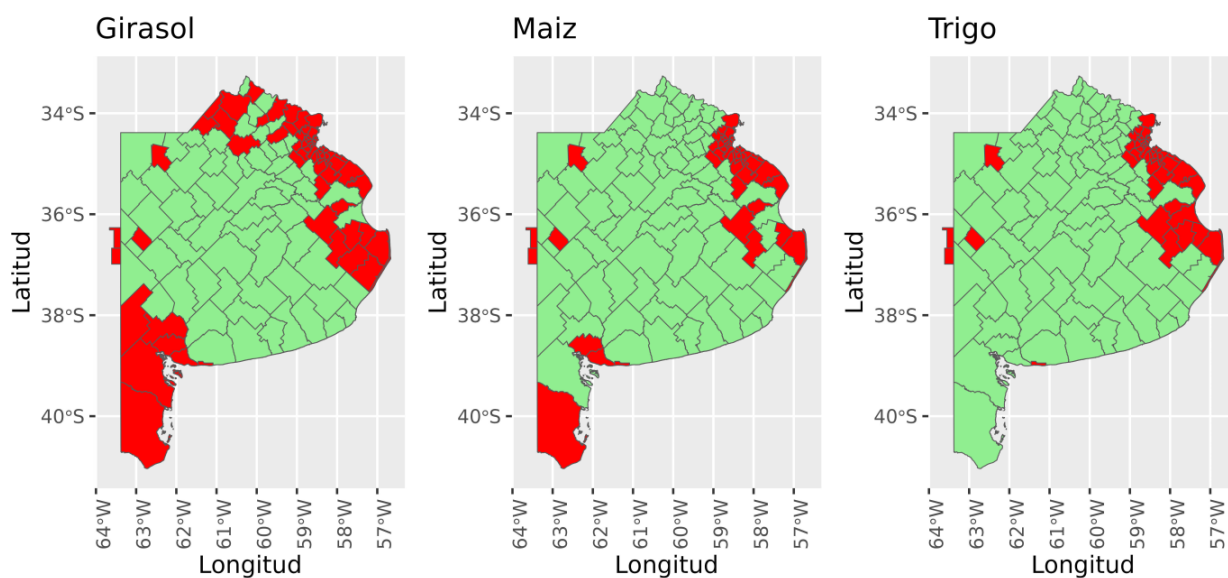


Figura 3.1: En tonos verdes los partidos que tienen menos del 5 % de datos faltantes o que su serie es continua. En tonos rojos los partidos que presentan mas del 5 % de datos faltantes.

3.1.1. Caracterización de datos de rendimiento

En la Figura 3.2 se muestran los valores de rendimiento máximo, mínimo y el rendimiento medio anual en kg/ha (kilogramos por hectárea) del período 1981-2021 para todos los partidos de la provincia de Buenos Aires, y para los tres cultivos. Se puede observar que el rendimiento de trigo tiene valores mínimos por debajo de los 2000 kg/ha. Las distribuciones, tanto para los mínimos, medios y máximos son similares, teniendo los valores más altos en el norte y en el sureste de la provincia. Los valores medios de este cultivo se encuentran en torno a los 4000 kg/ha mientras que los máximos superan, en algunos partidos, los 6000 kg/ha. El girasol muestra máximos en el sector norte de la provincia de Buenos Aires para el rendimiento medio y máximo, mientras que para el rendimiento mínimo se observa una distribución espacial más uniforme. Los rendimientos mínimos se encuentran por debajo de los 1000 kg/ha en toda la provincia, los medios están en torno a los 2000 kg/ha y los máximos se encuentran por encima de los 3000 kg/ha en el norte de la provincia. El

rendimiento de maíz tiene los valores mínimos por debajo de los 3000 kg/ha para todos los partidos. Por otro lado, se observa que los rendimientos más altos se encuentran en los partidos del norte de la provincia, mientras que los más bajos se encuentran en el extremo sur de la provincia. Por su parte, los valores medios para la provincia de Buenos Aires muestran que los rendimientos siguen una distribución similar a la de los valores mínimos. Los valores medios tienen rendimientos del orden de los 7000 kg/ha, mientras que los máximos de maíz siguen una distribución similar, teniendo valores de rendimiento aproximados a 12000 kg/ha. Observando el comportamiento de los tres cultivos para la provincia de Buenos Aires, podemos inferir que esta distribución podría estar asociada a las capacidades productivas de la superficie que fueron presentadas por la oficina de riesgo agropecuario. (ORA).

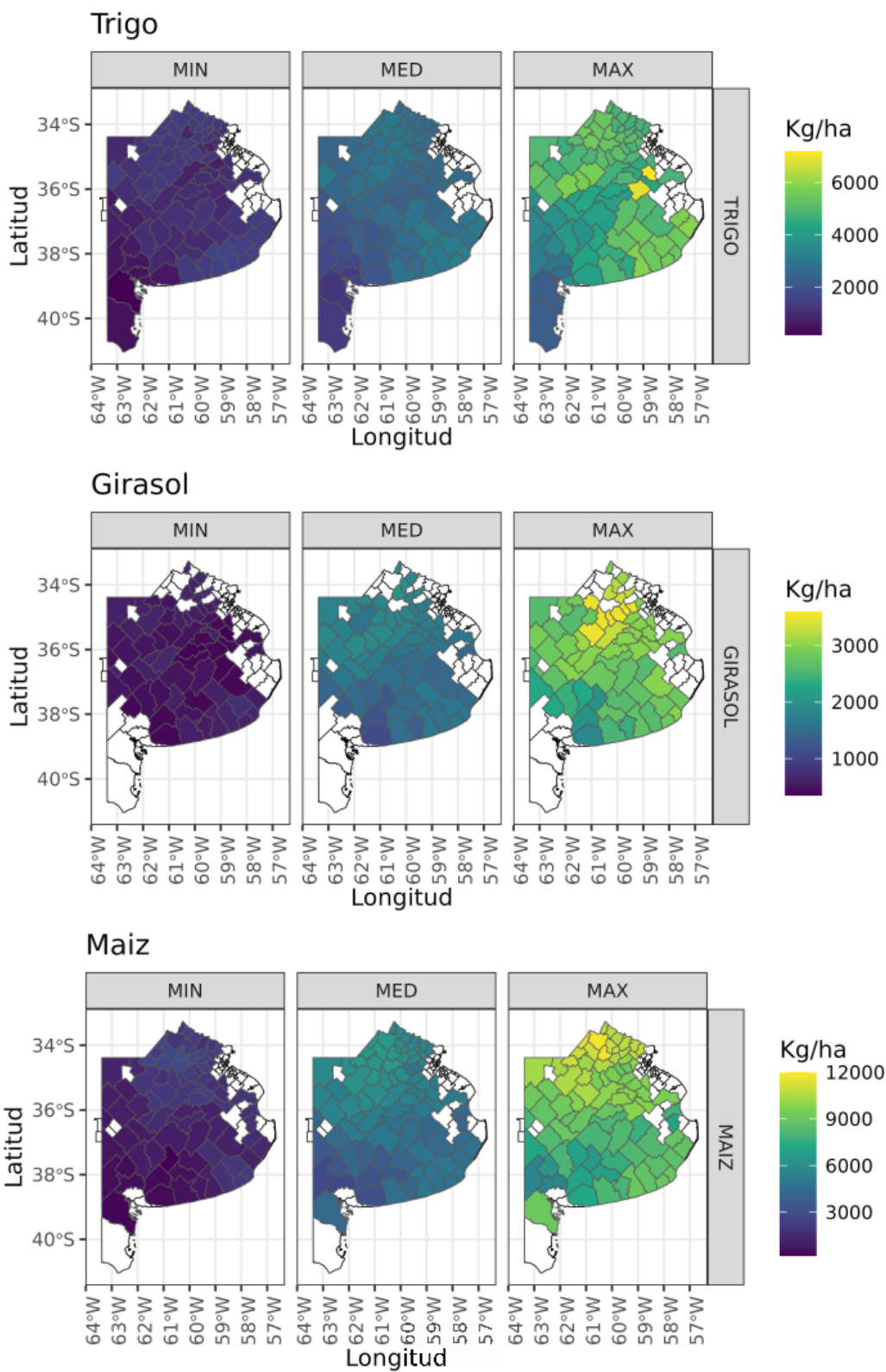


Figura 3.2: Distribucion de rendimientos por partidos de valores minimos, medios y maximos para el periodo desde 1981 hasta 2021 sobre los cultivos de trigo, maiz y girasol. Los partidos que se encuentran en color blanco fueron desestimados por tener muchos datos faltantes.

3.1.2. Tendencias Observadas

Para poder analizar la existencia de tendencias en los rendimientos de cultivos, se calcularon los valores medios de rendimiento en períodos de 13 años para los 3 cultivos. Mediante la comparación entre períodos es posible detectar si existen cambios en los valores medios de los rendimientos a lo largo de la serie completa. Para ello se seleccionaron tres períodos: 1981-1994 (período 1), 1995-2008 (período 2) y 2009-2021 (período 3). En la Figura 3.3 se muestran los rendimientos medios para los tres períodos mencionados, para los tres cultivos. En dicha figura se encuentra que para el trigo existe un aumento del valor medio de rendimiento a través de los periodos. En el periodo 1 el trigo tiene un rendimiento promedio del orden de los 2000 kg/ha, con una distribución que muestra valores más altos de rendimiento en el norte y en el sureste, dejando la zona central con mínimos relativos durante este periodo. Para el periodo 2 se encuentra una distribución espacial similar al período 1, con máximos relativos en el sector norte y en la franja costera del sureste de la provincia, y no se observa un aumento en los valores medios en el extremo sur de la provincia de Buenos Aires. En este período los valores medios se encuentran en torno a los 3000 kg/ha. En el periodo 3 los valores máximos están distribuidos más uniformemente en la zona central de la provincia comparado con los periodos anteriores. Estos máximos están por debajo de los 5000 kg/ha y el extremo sur de la provincia presenta valores de rendimiento medio similares a los del periodo 1. Los partidos disponibles para el cultivo de girasol presentan, para los tres periodos un comportamiento similar al del trigo, donde desde la franja central hacia el noroeste se encuentran rendimientos más altos que en el sureste de la provincia. Para el periodo 1 los valores de rendimiento para este cultivo rondan los 1000 kg/ha mientras que en el periodo 2 el rendimiento es del orden de los 2000 kg/ha. Este aumento en los valores de rendimiento también se observa en el periodo 3 donde los valores de rendimiento en algunos partidos son cercanos a los 3000 kg/ha. Por último, el maíz presenta máximos en el extremo norte de la provincia para el periodo 1, con rendimientos que rondan los 3000

kg por hectárea. Para el periodo 2 se observan variaciones en la distribución de los rendimientos, donde se encuentran mayores rendimientos en toda la franja norte de la provincia y menores rendimientos en el extremo suroeste. En este período los valores de rendimiento se encuentran en torno a los 6000 kg/ha. El periodo 3 por su parte también muestra los máximos rendimientos en la franja norte, alcanzando 10000 kg/ha mientras que la distribución de los rendimientos es similar al periodo 2, salvo por un aumento en el rendimiento en el extremo suroeste. A partir del análisis gráfico de los rendimientos en los tres períodos seleccionados, se observa una clara tendencia positiva en los mismos para los tres cultivos. Para determinar si la tendencia observada en el rendimiento de los cultivos a lo largo de los años es significativa o no, se realizó un test de Mann-Kendall. Los resultados mostraron que, efectivamente todas las series de rendimientos de cultivo tanto para el girasol, el maíz y el trigo presentan una tendencia significativa a un nivel de confianza 0.1.

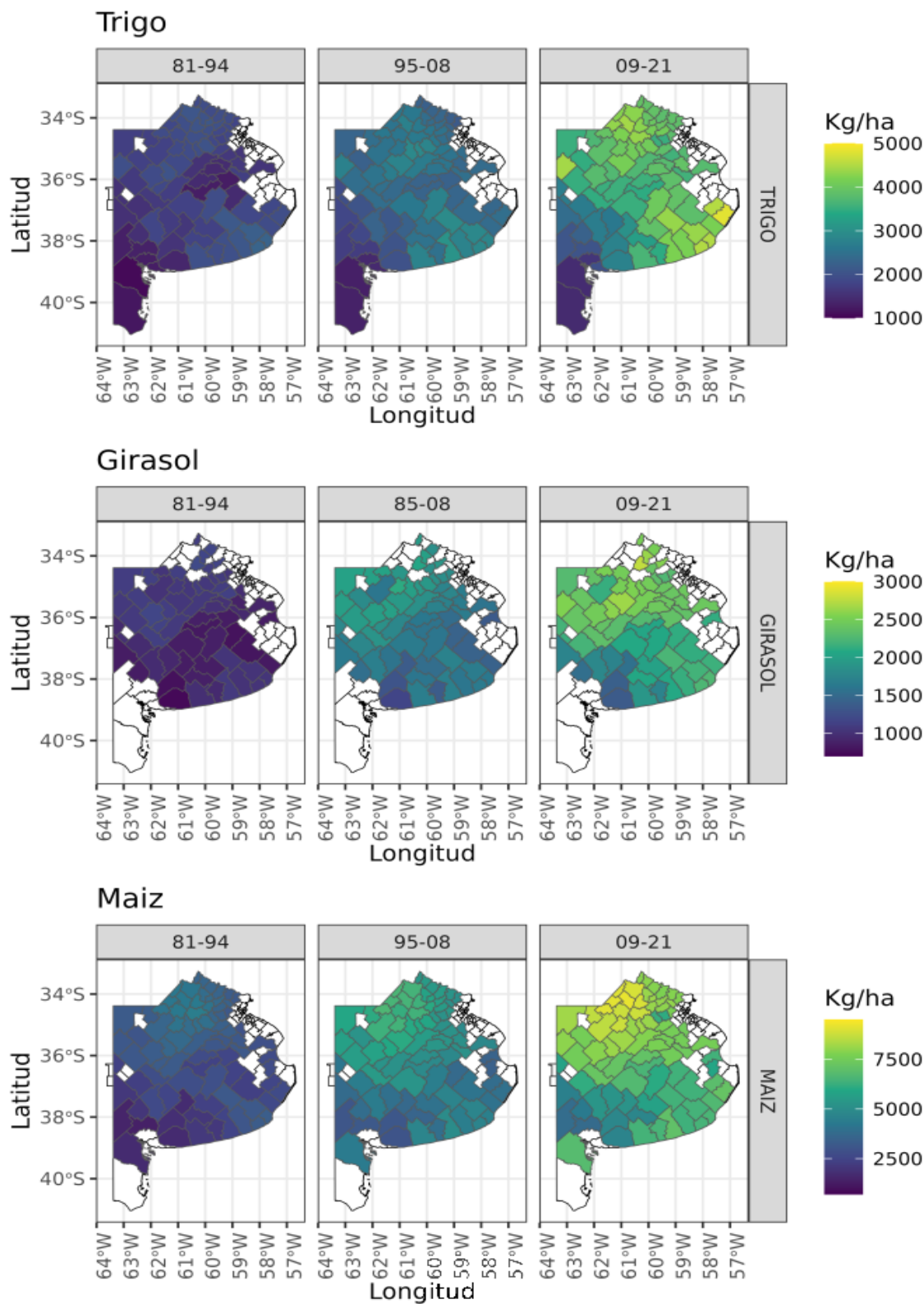


Figura 3.3: Distribucion de valores medios de rendimiento para los partidos de la provincia de Buenos Aires, los valores medios estan calculados para tres periodos, estos son: 1981-1994 (período 1), 1995-2008 (período 2) y 2009-2021 (período 3).

3.2. Regionalización

Para la división de la región de estudio en subgrupos que tengan un comportamiento de rendimiento similar, primero se determinó la cantidad de grupos que se debe dividir para cada cultivo. Para ello, como se mencionó en el capítulo anterior, se aplicó el método K-means, con el cual se determinó que para el cultivo de girasol es necesaria una división de la región en 3 grupos, para el cultivo de maíz una división de la región en 4 grupos y para el cultivo de trigo la división óptima es de 2 grupos para toda la provincia de Buenos Aires. Introduciendo la cantidad de divisiones óptimas se encontraron los grupos para cada cultivo, como se muestra en la Figura 3.4. En el cultivo de girasol los 3 grupos que quedaron se distribuyen de sur a norte, mientras que para el cultivo de maíz se observan los 4 grupos distribuidos de la siguiente manera: dos en la zona de la costa Atlántica y otros dos grupos en la región central y norte de la provincia. Por último, el cultivo de trigo, que quedó con dos grupos, muestra uno que abarca una región que se extiende hacia el suroeste y el otro que abarca una región más extensa en toda la región central y noreste.

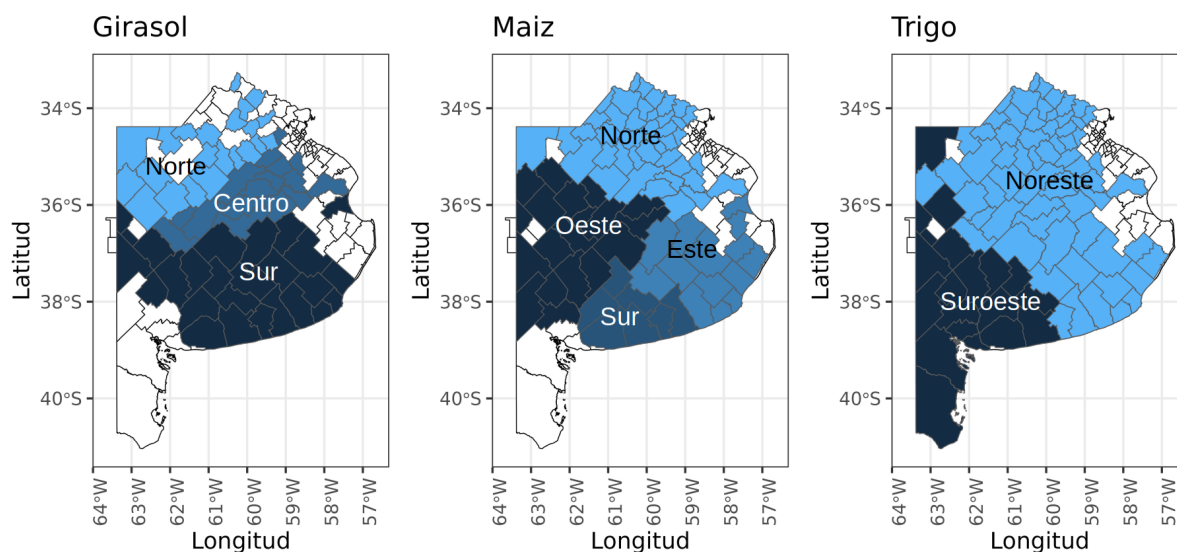


Figura 3.4: Se muestran los grupos para cada cultivo, el girasol presenta 3 grupos, el maíz presenta 4 grupos y el trigo presenta 2 grupos.

Estos resultados muestran que los clusters del rendimiento de los cultivos presentan una distribución similar, principalmente en el cultivo de girasol, al que presentan las regiones de estadios fenológicos propuestos por la oficina de riesgo agropecuario [ORA](#). Los clusters de los cultivos de maíz y trigo no presentan similitudes tan marcadas con los estadios fenológicos, esto podría deberse a que la variabilidad de los mismos no está únicamente determinada por su estadio fenológico.

Capítulo 4

Modelos de pronóstico de rendimiento

En este capítulo se describen los modelos de predicción de rendimientos de cultivos comenzando por el girasol seguido por el maíz y por último el trigo. En cada Tabla se muestran los modelos utilizados con sus respectivos predictores aplicados a cada cluster.

4.1. Girasol

4.1.1. Cluster Sur

Modelos basados en estaciones meteorológicas

La Tabla 4.1 muestra qué predictores utiliza cada modelo de la región sur de girasol. **MLR_BIC** presenta tres predictores que utilizan la variable **ETR**, un predictor con **PP** y otro con **ETP**. La variable **ETR** utiliza los períodos bimestrales abril-mayo y agosto-septiembre, y el mes de diciembre. Para la variable de **ETP** se utiliza el período bimestral octubre-noviembre. Por último, para la variable de **PP** se utiliza el mes de mayo. En el caso del modelo **MLR_S**, el mismo presenta tres predictores con **ETP**, uno con **ETR** y por último presenta un predictor con **PP**. La variable **ETP** utiliza el mes de abril y el bimestre octubre-noviembre., mientras que el predictor con **PP** utiliza el mes de mayo. El predictor de **ETR**

Estacion-Variable-Temporada	MLR_BIC	MLR_S	MULTI_BIC	ÁRBOL
Coronel Suárez - ETP - abr		X		
Tres Arroyos - ETP - abr		X		
Dolores - ETP - oct/nov	X	X		
Dolores - ETR - abr/may	X			
Benito Juarez - ETR - ago/sep	X			
Villa Gesell - ETR - ago/sep			X	
Benito Juarez - ETR - ago/sep/oct		X		
Mar del Plata - ETR - jul			X	
Tandil - ETR - dic			X	
Tres Arroyos - ETR - dic	X			X
Tres Arroyos - PP - may	X	X		
Mar Del Plata - PP - jul/ago			X	
Punta Indio - PP - ago				X
Mar Del Plata - PP - ago				X
Tandil - Tmáxima - jun			X	
Dolores - Tmedia - abr/may				X

Tabla 4.1: Predictores de rendimientos para el cluster sur del cultivo de girasol. En la primera columna se muestra la estación junto con la variable meteorológica y el periodo que se utiliza como predictor para cada modelo. ETP: Evapotranspiracion potencial ; ETR: Evapotranspiracion real; PP: precipitacion ; Tmaxima: temperatura maxima ; Tmedia: temperatura media ; Tminima: temperatura minima. Los modelos son: MLR_BIC: Modelo de regresión lineal múltiple con criterio Bayesiano, MLR_S: Modelo de regresión lineal múltiple con análisis de sensibilidad, MULTI_BIC: Modelo de regresión logística múltiple y ÁRBOL: Modelo de ÁRBOL de decisión.

utiliza el periodo trimestral que va de agosto a octubre Podemos detectar que en estos dos modelos se repiten dos predictores: estos son **ETP** en Dolores para el bimestre que va de octubre a noviembre y **PP** en Tres Arroyos para el mes de mayo. El modelo **MULTI_BIC** presenta tres predictores con la variable **ETR** y otros dos predictores con la variable **PP** y **Tmáxima**. Cabe destacar que 3 de los 5 predictores seleccionados para el modelo **MULTI_BIC** presentan periodos mensuales mientras que dos presentan periodos bimestrales. Este modelo utiliza el predictor **ETR** en diciembre y en julio y en el periodo bimestral agosto-septiembre. Otro de los predictores para **MULTI_BIC** utiliza PP en el bimestre julio-agosto, y el quinto predictor para este modelo utiliza la variable **Tmáxima** en el mes de junio. El modelo **ÁRBOL** presenta cuatro predictores, a diferencia de los demás modelos utilizados que presentan 5 predictores. Dos de estos cuatro predictores utilizan la variable

PP mientras que los demás predictores utilizan las variables **Tmedia** y **ETR**. Dos predictores con **PP** utilizan el mes de agosto, mientras que el predictor **ETR** utiliza el mes de diciembre. Por último, el predictor **Tmedia** utiliza el bimestre de abril-mayo.

Modelos basados en Índices climáticos

La Tabla 4.2 muestra que para el modelo **MLR_BIC** se determinaron dos predictores que corresponden a los índices **SAM** e **IOD**. El **IOD** utiliza el mes de septiembre mientras que el índice **SAM** toma el trimestre de abril a junio. En el caso del modelo **MLR_S**, tres de sus cuatro predictores utilizan el índice **TSA** en periodos invernales que van de junio a agosto.

El modelo **MULTI_BIC** presenta dos predictores: uno asociado al índice **IOD** en el mes de octubre y otro asociado a la **TSA** en el bimestre mayo-junio. Por último, el modelo **ÁRBOL** contiene tres predictores, de los cuales uno corresponde al índice **TSA** en el mes de julio, mientras que los otros dos son el índice **IOD** en el bimestre agosto-septiembre y el índice **NIÑO1+2** en el mes de agosto.

Índice - Temporada	MLR_BIC	MLR_S	MULTI_BIC	ÁRBOL
TSA - may/jun			X	
TSA - jun/jul		X		
TSA - jul		X		X
TSA - jun/jul/ago		X		
SAODI - oct		X		
SAM - abr/may/jun	X			
IOD - sep	X			
IOD - oct			X	
IOD - ago/sep				X
NINO1+2 - ago				X

Tabla 4.2: Predictores de rendimientos para el cluster sur del cultivo de girasol. En la primera columna se muestra el índice junto con el periodo que se utiliza como predictor para cada modelo. Los modelos son: **MLR_BIC**: Modelo de regresión lineal múltiple con criterio Bayesiano, **MLR_S**: Modelo de regresión lineal múltiple con análisis de sensibilidad, **MULTI_BIC**: Modelo de regresión logística múltiple y **ÁRBOL**: Modelo de **ÁRBOL** de decisión.

4.1.2. Cluster Centro

Modelos basados en estaciones meteorológicas

Estacion-Variable-Temporada	MLR_BIC	MLR_S	MULTI_BIC	ÁRBOL
Coronel Suárez - ETP - ago/sep/oct	X			
Ezeiza - ETP - sept/oct				X
Punta indio - ETP - ago/sep/oct		X		
Punta Indio - ETP - may			X	
Tandil - ETR - jun/jul		X		
Tandil - ETR - jun/jul/ago		X		
Pehuajó - ETR - jul/ago/sep			X	
Nueve de julio - ETR - ago/sep/oct	X			
Coronel Suarez - ETR - ago/sep/oct		X		X
Coronel Suarez - ETR - sep/oct			X	
Ezeiza - ETR - oct	X			
Tandil - PP - jun	X			
Pehuajó - PP - jun			X	
Tandil - PP - jul		X		
Tandil - PP - oct	X		X	
La Plata - Tmáxima - abr				X

Tabla 4.3: Idem a Tabla 4.1 pero para cluster centro

La Tabla 4.3 muestra que el modelo **MLR_BIC** presenta un predictor con la variable **ETP**, dos predictores con la variable **ETR** y otros dos predictores con **PP**. Para la variable **ETP** el predictor utiliza el trimestre que va de agosto a octubre, mientras que la variable **ETR** presenta un predictor en el trimestre que va de agosto a octubre y otro predictor en el mes de octubre. Por último, los predictores de **MLR_BIC** que utilizan la **PP** como variable, lo hacen en los meses de junio y octubre. El modelo **MLR_S** utiliza tres de sus cinco predictores con la variable **ETR** en los siguientes períodos: junio a julio, junio a agosto y agosto a octubre. Los dos predictores restantes en este modelo utilizan la variable **ETP** en el trimestre de agosto a octubre por un lado, y por otro utiliza la variable **PP** durante julio. El modelo **MULTI_BIC** considera dos predictores utilizando la variable **ETR**, uno en el trimestre que va de julio a septiembre y otro en el bimestre septiembre-octubre. Por otro lado considera

otros dos predictores: uno es la **PP** para los meses de junio y octubre, y el otro es la **ETP** en el mes de mayo. El modelo **ÁRBOL** basado en estaciones para el cluster centro sobre el cultivo de girasol presenta tres predictores. El primero utiliza la variable **ETP** para el bimestre septiembre-octubre, el segundo utiliza la variable **ETR** durante el trimestre de agosto a octubre y el último de los predictores es la **Tmáxima** en el mes de abril.

Modelos basados en Índices climáticos

Indice-Temporada	MLR_BIC	MLR_S	MULTI_BIC	ÁRBOL
SAODI - ago	X			
SAODI - ago/sep/oct				X
SAODI - nov/dic		X		
SAM - ago/sep/oct	X			
SAM - ago				X
IOD - nov/dic		X		
IOD - nov		X		
NIÑO1+2 - abr			X	
NIÑO1+2 - may				X
NIÑO4 - may			X	
NIÑO4 - may/jun				X

Tabla 4.4: Idem a Tabla 4.2 pero para cluster centro

De la Tabla 4.4 se desprende que **MLR_BIC** presenta dos predictores, uno utiliza el índice **SAODI** para el mes de agosto y otro utiliza el índice **SAM** para el trimestre que va de agosto a octubre. Por otro lado, el modelo **MLR_S** presenta tres predictores, uno de los cuales utiliza el índice **SAODI** en el bimestre noviembre-diciembre. Los dos predictores restantes utilizan el índice **IOD** en el bimestre noviembre-diciembre y en el mes de noviembre. El modelo **MULTI_BIC** utiliza dos predictores a través de los índices **NIÑO1+2** y **NIÑO4**, el primero para el mes de abril y el segundo para el mes de mayo. El modelo **ÁRBOL** para el cluster centro del cultivo girasol presenta cuatro predictores. El primer predictor utiliza el índice **SAODI** en el trimestre que va de agosto a octubre, mientras que otro de los predictores utiliza el índice **SAM** en el mes de agosto. Los últimos dos predictores usan el **NIÑO1+2** y **NIÑO4** en el mes de mayo y en el bimestre mayo-junio, respectivamente.

4.1.3. Cluster Norte

Modelos basados en estaciones meteorológicas

Estacion-Variable-Temporada	MLR_BIC	MLR_S	MULTI_BIC	ÁRBOL
Nueve de Julio - ETP - jul/ago/sep			X	
Trenque Lauquen - ETR - nov/dic	X	X		X
Trenque Lauquen - ETR - abr			X	
Trenque Lauquen - ETR - sep/oct	X			
Pehuajo - ETR - sep/oct/nov			X	
Trenque Lauquen - ETR - jul/ago/sep		X		
Ezeiza - ETR - abr		X		
Ezeiza - ETR - abr/may		X	X	
Junin - ETR - jul	X			X
Ezeiza - PP - jul			X	
Pehuajo - PP - sep	X			
Pehuajo - Tmáxima - abr/may	X	X		X
Nueve de Julio - Tmáxima - jul				X

Tabla 4.5: Idem a Tabla 4.1 pero para cluster norte

De la Tabla 4.5 se desprende que el modelo **MLR_BIC** contiene tres de sus 5 predictores con la variable **ETR**, para los bimestres noviembre-diciembre y septiembre-octubre. El otro de los predictores que utiliza la variable **ETR** se centra en el mes de junio. Los dos predictores restantes se encuentran sobre las variables de **Tmáxima** y **PP**, en el bimestre abril-mayo y para el mes de septiembre, respectivamente. El modelo **MLR_S** presenta cuatro de sus cinco predictores con la variable **ETR**. Los periodos utilizados para los primeros tres predictores de esta variable son: el bimestre abril-mayo, bimestre noviembre-diciembre y el trimestre que va de julio a septiembre. El cuarto predictor que utiliza la variable **ETR** es en el mes de abril, mientras que el último de los predictores utiliza la variable **Tmáxima** en el bimestre abril-mayo. El modelo **MULTI_BIC** cuenta con tres predictores con la variable **ETR**, de los cuales dos utilizan el mes de abril y el predictor restante utiliza el trimestre que va de septiembre a noviembre. Otro de los predictores de este modelo utiliza la variable **ETP** en el trimestre que va de julio a septiembre, mientras que el predictor que completa

el modelo utiliza la variable **PP** en el mes de julio. Por último, el modelo basado en **ÁRBOL** presenta cuatro predictores, de los cuales dos utilizan la variable **ETR**, uno en el bimestre noviembre-diciembre y el otro en el mes de junio. Los dos predictores restantes utilizan la **Tmáxima** en el bimestre abril-mayo y en el mes de junio.

Modelos basados en Índices climáticos

Indice-Temporada	MLR_BIC	MLR_S	MULTI_BIC	ÁRBOL
SAODI - ago/sep/oct	X			
SAODI - sep/oct/nov		X	X	
SAODI - mar/abr	X			
SAODI - ene	X			
SAM - ago				X
SAM - may/jun/jul	X			
IOD - dic		X	X	
IOD - jun				X
NIÑO1+2 - ago	X	X	X	
NIÑO1+2 - abr				X
NIÑO1+2 - jun/jul/ago		X		
NIÑO1+2 - jul/ago/sep			X	
NIÑO3 - jun/jul/ago		X	X	

Tabla 4.6: Idem a Tabla 4.2 pero para cluster norte

En la Tabla 4.6 podemos identificar que el modelo **MLR_BIC** presenta cinco predictores, entre ellos encontramos que tres corresponden al índice **SAODI** en los siguientes períodos: el trimestre que va de agosto a octubre, el bimestre marzo-abril y el mes de enero. Otro de los predictores utiliza el índice **SAM** en el trimestre que va de mayo a julio, mientras que el predictor restante utiliza el índice **NIÑO1+2** en el mes de agosto. El modelo **MLR_S** utiliza tres índices basados en los índices NIÑO: los primeros dos utilizan el **NIÑO1+2** en el mes de agosto y en el trimestre que va de junio a agosto, y el restante utiliza el índice **NIÑO3** en el mismo trimestre. Los dos predictores restantes utilizan los índices **SAODI** e **IOD**, en el trimestre que va de septiembre a noviembre y en el mes de diciembre, respectivamente. El modelo **MULTI_BIC** utiliza dos predictores a partir del índice **NIÑO1+2**, uno en el mes de

agosto y otro en el trimestre que va de julio a septiembre. Otro de los predictores utiliza el **NIÑO3** en el trimestre que va de julio a agosto. Los predictores restantes para este modelo se basan en el índice en el trimestre que va de septiembre a noviembre, y en el índice **IOD** en el mes de diciembre. El último modelo para el cluster norte es **ÁRBOL**, este presenta 3 predictores que se basan en los índices **SAODI**, **IOD** y **NIÑO1+2** para los meses de agosto, junio y abril, respectivamente.

4.2. Maiz

4.2.1. Cluster Sur

Modelos basados en estaciones meteorológicas

La Tabla 4.7 presenta cinco predictores para el modelo **MLR_BIC**, el primer predictor utiliza la variable **ETP** durante el bimestre octubre-noviembre. Dos predictores de este modelo utilizan la variable **PP**, uno durante el trimestre que va de noviembre a enero y el otro durante el bimestre agosto-septiembre. Los dos predictores restantes utilizan la **Tmínima** y la **Tmedia**, y están determinados en el bimestre que va de noviembre a diciembre para la **Tmínima**, y el trimestre que va de octubre a noviembre para la **Tmedia**. El modelo **MLR_S** utiliza dos de sus cinco predictores con la variable **PP**: uno durante el bimestre diciembre-enero y otro en el trimestre que va de noviembre a enero. Otros dos predictores de este modelo son a partir de la variable **ETR** para el mes de enero y para el bimestre octubre-noviembre. El último predictor de este modelo está basado en la variable **ETP** en el mes de octubre. En el caso del modelo **MULTI_BIC**, éste presenta dos de sus cinco predictores asociados a la variable temperatura, y uno de ellos utiliza la **Tmedia** y **Tmáxima** para la misma estación. Los periodos utilizados por estos dos predictores son los trimestres que van de julio a septiembre y de junio a agosto. Otros dos predictores utilizan la variable **ETP** y **ETR**: la variable **ETP** utiliza como periodo el mes de enero mientras que la variable **ETR**

Estacion-Variable-Temporada	MLR_BIC	MLR_S	MULTI_BIC	ÁRBOL
Bahia Blanca - ETP - Oct/Nov	X			
Bahia Blanca - ETP - Ene			X	
Coronel Suarez - ETP - Oct		X		
Tres Arroyos - ETR - Ene		X		
Tres Arroyos - ETR - Jul/Ago				X
Bahia Blanca - ETR - Jul/Ago/Sep			X	
Coronel Suarez - ETR - Oct/Nov		X		
Tres Arroyos - PP - Nov/Dic/Ene	X	X		
Tres Arroyos - PP - Dic/Ene		X	X	
Coronel Suarez - PP - Ago/Sep	X			
Coronel Suarez - Tmínima - Nov/Dic	X			
Tres Arroyos - Tmedia - Oct/Nov/Dic	X			
Bahia Blanca - Tmedia - Jul/Ago/Sep			X	
Bahia Blanca - Tmedia - Oct/Nov/Sep				X
Bahia Blanca - Tmáxima - Jun/Jul/Ago			X	
Tres Arroyos - Tmáxima - Sep/Oct/Nov				X
Bahia Blanca - Tmáxima - Jul				X

Tabla 4.7: Predictores de rendimientos para el cluster sur del cultivo de maíz. En la primera columna se muestra la estación junto con la variable meteorológica y el periodo que se utiliza como predictor para cada modelo. ETP: Evapotranspiracion potencial ; ETR: Evapotranspiracion real; PP: precipitacion ; Tmáxima: temperatura maxima ; Tmedia: temperatura media ; Tmínima: temperatura minima. Los modelos son: MLR_BIC: Modelo de regresión lineal múltiple con criterio Bayesiano, MLR_S: Modelo de regresión lineal múltiple con análisis de sensibilidad, MULTI_BIC: Modelo de regresión logística múltiple y ÁRBOL: Modelo de ÁRBOL de decisión.

utiliza el trimestre de julio a septiembre. El predictor que completa este modelo está basado en la variable PP para el bimestre que va de diciembre a enero. Por último, el modelo **ÁRBOL** presenta cuatro predictores, de los cuales 3 corresponden a variables asociadas a la temperatura. Por un lado utiliza la **Tmáxima** para el trimestre que va de septiembre a noviembre y el mes de julio. Por otro lado, utiliza la **Tmedia** para el trimestre que va de octubre a diciembre. El último predictor utiliza la variable **ETR** para la estación de Tres Arroyos durante el bimestre julio-agosto

Modelos basados en Índices climáticos

La Tabla 4.8 muestra que el modelo **MLR_BIC** presenta dos predictores que utilizan el índice **SAM**, uno durante julio y el otro durante el bimestre noviembre-diciembre. El

modelo **MLR_S** presenta cuatro predictores, de los cuales tres de ellos utilizan los índices **IOD**, **NIÑO3** y **NIÑO1+2**, todos en el bimestre diciembre-enero. El último de los predictores utiliza el índice **SAM** en el bimestre noviembre-diciembre. El modelo **MULTI_BIC** presenta dos predictores, uno basado en el índice **SAM** durante el bimestre enero-febrero, y el otro basado en la **TSA** durante el mes de julio. El modelo **ÁRBOL** presenta dos predictores, uno en base al índice **NIÑO1+2** durante el bimestre agosto-septiembre y el otro a partir de la **TSA** durante el mes de julio.

Indice-Temporada	MLR_BIC	MLR_S	MULTI_BIC	ÁRBOL
TSA - Jul			X	X
SAM - Nov/Dic	X	X		
SAM - Jul	X			
SAM - Ene/Feb			X	
IOD - Dic/Ene		X		
NIÑO1+2 - Ago/Sep				X
NIÑO3 - Dic/Ene		X		
NINO1+2 - Dic/Ene		X		

Tabla 4.8: Predictores de rendimientos para el cluster sur del cultivo de maíz. En la primera columna se muestra el índice junto con el periodo que se utiliza como predictor para cada modelo. Los modelos son: **MLR_BIC**: Modelo de regresión lineal múltiple con criterio Bayesiano, **MLR_S**: Modelo de regresión lineal múltiple con análisis de sensibilidad, **MULTI_BIC**: Modelo de regresión logística múltiple y **ÁRBOL**: Modelo de **ÁRBOL** de decisión.

4.2.2. Cluster Este

Modelos basados en estaciones meteorológicas

La Tabla 4.9 muestra que en el cluster este del cultivo de maíz el modelo **MLR_BIC** presenta cinco predictores, de los cuales dos están asociados a la variable **PP**, uno en el bimestre noviembre-diciembre y otro en el mes de julio. También, utiliza otro predictor con la variable **ETP** para el trimestre que va de octubre a diciembre. Por otro lado, otro de los predictores utiliza la variable **ETR** en el mes de enero, mientras que el último predictor de **MLR_BIC** utiliza la variable **Tmedia** durante el trimestre que va de agosto a octubre.

Estacion-Variable-Temporada	MLR_BIC	MLR_S	MULTI_BIC	ÁRBOL
Dolores - ETP - Oct/Nov/Dic	X			X
Dolores - ETP - Sep/Oct			X	
Dolores - ETP - Dic			X	
Dolores - ETP - Dic/Ene/Feb			X	
Mar del Plata - ETP - Jul/Ago				X
Punta Indio - ETP - Oct			X	
Tandil - ETR - Ene	X	X		
Tandil - ETR - Sep/Oct		X		
Dolores - PP - Nov/Dic	X	X		
Punta Indio - PP - Nov/Dic				X
Tandil - PP - Dic		X		
Dolores - PP - Jul	X			
Mar del Plata - PP - Ago/Sep/Oct			X	
Punta Indio - Tmínima - Ene		X		
Punta Indio - Tmedia - Ago/Sep/Oct	X			

Tabla 4.9: Idem a Tabla 4.7 pero para cluster este

El modelo **MLR_S** presenta dos predictores utilizando la variable **ETR**: uno en el mes de enero y otro en el bimestre septiembre-octubre. Este modelo presenta otros dos predictores utilizando la variable **PP** en el bimestre noviembre-diciembre y en el mes de diciembre. El último predictor de este modelo utiliza la variable **Tmínima** en el mes de enero. El modelo **MULTI_BIC** presenta cuatro predictores utilizando la variable **ETP**. El conjunto de estos cuatro predictores utilizan periodos mensuales, bimestrales y trimestrales que van desde septiembre hasta febrero. El último predictor utiliza la variable **PP** para el trimestre que va de agosto a octubre. El último modelo para esta región y este cultivo es **ÁRBOL**, el cual presenta tres predictores, de los cuales dos utilizan la variable **ETP** en el trimestre que va de octubre a diciembre y en el bimestre julio-agosto. El tercer predictor utiliza **PP** en el bimestre noviembre-diciembre.

Modelos basados en Índices climáticos

De la Tabla 4.10 se desprende que el modelo **MLR_BIC** para el cluster este del cultivo de maíz presenta tres predictores que utilizan los índices **NIÑO4**, **SAM** y **SAODI**. Los períodos

Indice-Temporada	MLR_BIC	MLR_S	MULTI_BIC	ÁRBOL
TSA - Jul/Ago				X
SAODI - Oct	X			
SAODI - Ago/Sep		X		
SAM - Jul/Ago/Sep	X		X	X
IOD - Dic/Ene		X		
NIÑO1+2 - Oct				X
NIÑO3 - Sep/Oct				X
NIÑO4 - Ago	X			

Tabla 4.10: Idem a Tabla 4.8 pero para el cluster este

utilizados son el mes de agosto, el trimestre que va de julio a septiembre y el mes de octubre, respectivamente. El modelo **MLR_S** presenta dos predictores, uno de ellos utiliza el índice **IOD** en el bimestre diciembre-enero, mientras que el otro utiliza el índice **SAODI** en el bimestre agosto-septiembre. El modelo **MULTI_BIC** presenta un solo predictor a través del índice **SAM** en el trimestre que va de julio a septiembre. Por último, el modelo **ÁRBOL** presenta cuatro predictores: el primero utiliza el índice **SAM** en el trimestre que va de julio a septiembre, el segundo utiliza el índice **TSA** en el bimestre julio-agosto, el tercero utiliza el índice **NIÑO1+2** en el mes de octubre y el cuarto utiliza el índice **NIÑO3** en el bimestre septiembre-octubre.

4.2.3. Cluster Oeste

Modelos basados en estaciones meteorológicas

La Tabla 4.11 muestra que el modelo **MLR_BIC** presenta dos predictores utilizando la variable **ETR**, uno durante el trimestre que va de diciembre a febrero y el otro en el bimestre noviembre-diciembre. El modelo **MLR_BIC** también utiliza dos predictores a partir de la variable **PP** durante los meses de diciembre y agosto. El último predictor de este modelo utiliza la **Tmínima** como variable durante el trimestre que va de agosto a octubre. El modelo **MLR_S** utiliza **ETP** durante el bimestre enero-febrero. La variable **ETR** utiliza también el bimestre enero-febrero y el trimestre que va de noviembre a enero. Otro de los predictores

Estacion-Variable-Temporada	MLR_BIC	MLR_S	MULTI_BIC	ÁRBOL
Pehuajo - ETP - Ene/Feb		X		
Pehuajo - ETP - Sep			X	
Coronel Suarez - ETP - Dic				X
Pehuajo - ETP - Oct			X	
Pehuajo - ETR - Ene/Feb		X		
Coronel Suarez - ETR - Jul			X	
Pehuajo - ETR - Dic/Ene/Feb	X			
Pehuajo - ETR - Nov/Dic	X			
Pehuajo - ETR - Nov/Dic/Ene		X		
Pehuajo - PP - Dic	X		X	
Pehuajo - PP - Sep				X
Pehuajo - PP - Nov/Dic		X		
Coronel Suarez - PP - Ago	X			
Pehuajo - Tmáxima - Dic/Ene/Feb		X		
Pehuajo - Tmedia - Dic			X	
Pigue - Tminima - Ago/Sep/Oct	X			
Coronel Suarez - Tmínima - Sep				X

Tabla 4.11: Idem a Tabla 4.7 pero para cluster oeste

utiliza **PP** para el bimestre noviembre-diciembre y por último, el modelo **MLR_S** utiliza **Tmáxima** para el trimestre que va de diciembre a febrero. El modelo **MULTI_BIC** presenta dos predictores utilizando la variable **ETP**, uno para el mes de septiembre y otro para el mes de octubre. Otro de los predictores utiliza la variable **ETR** durante el mes de julio. Los últimos dos predictores de este modelo utilizaron variables de **PP** y **Tmedia**, ambos durante el mes de diciembre. El modelo **ÁRBOL** presenta tres predictores para este cluster, utiliza dos predictores basados en **Tmínima** y **ETP** durante los meses de septiembre y diciembre, respectivamente. El tercer predictor utiliza **PP** en septiembre.

Modelos basados en Índices climáticos

La Tabla 4.12 muestra que el modelo **MLR_BIC** presenta un único predictor que utiliza como variable el índice **NIÑO4** para el mes de septiembre. Por otro lado el modelo **MLR_S** presenta cinco predictores, de los cuales tres utilizan el índice **NIÑO1+2**, **NIÑO3** y **NIÑO4**. Los períodos utilizados son el bimestre noviembre-diciembre, el bimestre agosto-septiembre

Indice-Temporada	MLR_BIC	MLR_S	MULTI_BIC	ÁRBOL
SAODI- Dic/Ene/Feb		X		
SAODI - Sep			X	
SAODI - Oct			X	
SAODI - Jul/Ago/Sep				X
SAM - Dic/Ene		X		
SAM - Ago				X
NIÑO1+2 - Nov/Dic		X		
NIÑO3 - Ago/Sep		X		
NIÑO3 - Sep				X
NIÑO4 - Sep	X			
NIÑO4 - Ago/Sep/Oct		X		

Tabla 4.12: Idem a Tabla 4.8 pero para el cluster oeste

y el trimestre que va de agosto a octubre, respectivamente. Este modelo presenta otros dos predictores utilizando los índices **SAM** y **SAODI**, en el bimestre diciembre-enero y en el trimestre que va de diciembre a febrero. Por su parte, el modelo **MULTI_BIC** muestra dos predictores que utilizan el índice **SAODI**, en los meses de septiembre y de octubre. Por último, el modelo **ÁRBOL** presenta cuatro predictores, de los cuales uno utiliza el índice **SAM** en el mes de agosto. Los predictores restantes utilizan el índice **SAODI** en el trimestre que va de julio a septiembre y el índice **NIÑO3** en el mes de septiembre.

4.2.4. Cluster Norte

Modelos basados en estaciones meteorológicas

La Tabla 4.13 muestra que el modelo **MLR_BIC** presenta cuatro predictores, de los cuales dos utilizan la variable **ETR**, uno para el bimestre de enero-febrero y el otro en el mes de enero. Completando el modelo los dos predictores restantes utilizan las variables de **ETP** y **Tmáxima** en el trimestre que va de septiembre a noviembre y en el bimestre octubre-noviembre, respectivamente. El modelo **MLR_S** presenta dos predictores que utilizan temperatura: uno con **Tmáxima** durante el bimestre enero-febrero y otro con **Tmedia** durante el trimestre que va de diciembre a febrero. Este modelo también presenta dos predictores

Estacion-Variable-Temporada	MLR_BIC	MLR_S	MULTI_BIC	ÁRBOL
Nueve de julio - ETP - Ene		X		
Nueve de julio - ETP - Ene/Feb		X		
Nueve de julio - ETP - Sep/Oct/Nov			X	
Junín - ETP - Oct/Nov				X
Junín - ETP - Sep/Oct/Nov	X			
Nueve de julio - ETR - Oct			X	
Junín - ETR - Ene/Feb	X		X	
Junín - ETR - Ene	X	X		
Junín - PP - Oct/Nov			X	X
Ezeiza - PP - Jul/Ago				X
Nueve de julio - Tmáxima - Oct/Nov	X			
Junín - Tmáxima - Ene/Febr		X		
Junín - Tmáxima - Jul/Ago/Sep			X	
Junín - Tmedia - Dic/Ene/Feb		X		

Tabla 4.13: Idem a Tabla 4.7 pero para cluster norte

a través de la variable **ETP**, en el mes de enero y en el bimestre enero-febrero. El último predictor de este modelo utiliza la variable **ETR** para el mes de enero. Otro de los modelos es **MULTI_BIC** que utiliza tiene dos predictores con la variable **ETR**, en el mes de octubre y en el bimestre enero-febrero. Otro de los predictores utiliza la variable **ETP** en el trimestre que va de septiembre a noviembre. Los últimos predictores del modelo **MULTI_BIC** utilizan la variable **PP** en el bimestre octubre-noviembre y la variable **Tmáxima** durante el trimestre que va de julio a septiembre. El último modelo de este cluster, **ÁRBOL**, presenta tres predictores de los cuales dos utilizan la variable **PP**: uno durante el bimestre octubre-noviembre y otro en el bimestre julio-agosto. El último predictor de este modelo utiliza la variable **ETP** durante el bimestre octubre-noviembre.

Modelos basados en Índices climáticos

De la Tabla 4.14 se desprende que el modelo **MLR_BIC** presenta tres predictores. Dos de ellos utilizan los índices **SAODI** y **NIÑO4**, ambos en el bimestre septiembre-octubre, mientras que el tercer predictor utiliza el índice **SAM** en el mes de julio. El modelo **MLR_S** presenta cinco predictores, de los cuales tres utilizan índices Niño: **NIÑO1+2** en el bi-

Indice-Temporada	MLR_BIC	MLR_S	MULTI_BIC	ARBOL
TSA - Jul/Ago/Sep				X
SAODI - Sep/Oct	X			
SAODI - Jul			X	
SAM - Jul	X		X	
SAM - Oct/Nov/Dic		X		
SAM - Dic/Ene/Feb			X	
SAM - Sep			X	
IOD - Nov/Dic		X		
NIÑO1+2 - Ago				X
NIÑO1+2 - Nov/Dic		X		
NIÑO1+2 - Sep/Oct				X
NIÑO3 - Oct/Nov/Dic		X		
NIÑO4 - Sep/Oct	X			
NIÑO4 - Oct/Nov/Dic		X		

Tabla 4.14: Idem a Tabla 4.8 pero para el cluster norte

mestre noviembre-diciembre, **NIÑO3** y **NIÑO4**, ambos en el trimestre que va de octubre a diciembre. Otro de los predictores utiliza el índice **SAM** en el trimestre mencionado, mientras que el último predictor utiliza el índice **IOD** para el bimestre noviembre-diciembre. El modelo **MULTI_BIC** utiliza tres predictores a través del índice **SAM** en los meses de julio y septiembre, y en el trimestre que va de diciembre a febrero. El predictor restante de este modelo utiliza el índice **SAODI** en el mes de julio. El modelo **ÁRBOL** presenta tres predictores, de los cuales dos utilizan el índice **NIÑO1+2**, en el bimestre septiembre-octubre y en el mes de agosto. El último de los predictores utiliza la **TSA** en el trimestre que va de julio a septiembre.

4.3. Trigo

4.3.1. Cluster Suroeste

Modelos basados en estaciones meteorológicas

Estacion-Variable-Temporada	MLR_BIC	MLR_S	MULTI_BIC	ÁRBOL
Benito Juarez - ETP - Ago			X	
Pehuajo - ETR - Oct	X			
Coronel Suarez - ETR - Ago/Sep/Oct	X			
Bahia Blanca - ETR - Sep				X
Coronel Suarez - ETR - Ago	X	X		
Coronel Suarez - ETR - Jul/Ago/Sep		X	X	
Coronel Suarez - ETR - Ago/Sep		X	X	
Coronel Suarez - PP - Jun/Jul		X	X	
Coronel Suarez - PP - Jul/Ago/Sep			X	
Pigue - Tmáxima - Ago/Sep/Oct		X		
Pigue - Tmedia - Abril	X			
Tres Arroyos - Tmedia - Mayo				X
Tres Arroyos - Tmáxima - Ago	X			
Tres Arroyos - Tmínima - Abril				X

Tabla 4.15: Predictores de rendimientos para el cluster suroeste del cultivo de Trigo. En la primera columna se muestra la estación junto con la variable meteorológica y el periodo que se utiliza como predictor para cada modelo. ETP: Evapotranspiracion potencial ; ETR: Evapotranspiracion real; PP: precipitacion ; Tmáxima: temperatura maxima ; Tmedia: temperatura media ; Tmínima: temperatura minima. Los modelos son: MLR_BIC: Modelo de regresión lineal múltiple con criterio Bayesiano, MLR_S: Modelo de regresión lineal múltiple con análisis de sensibilidad, MULTI_BIC: Modelo de regresión logística múltiple y ÁRBOL: Modelo de ÁRBOL de decisión.

La Tabla 4.15 muestra que el modelo **MLR_BIC** presenta cinco predictores, de los cuales tres utilizan la variable ETR: uno durante el mes de octubre, otro en el trimestre que va de agosto a octubre y el tercero en el mes de agosto. Los dos predictores restantes utilizan las variables de **Tmedia** y **Tmáxima** en el mes de abril y en el mes de agosto, respectivamente. El modelo **MLR_S** presenta tres predictores utilizando la variable **ETR**: uno durante el mes de agosto, otro en el trimestre que va de julio a septiembre y otro en el bimestre agosto-septiembre. Otro de los predictores utiliza la variable **PP** durante el bimestre junio-julio, y el

último de los predictores utiliza la variable **T**máxima durante el trimestre que va de agosto a octubre. El modelo **MULTI_BIC** presenta cuatro predictores que utilizan las variables de **ETR** y **PP**, ambas en el trimestre que va de julio a septiembre, mientras que **ETR** también utiliza el bimestre agosto-septiembre y **PP** el bimestre junio-julio. El modelo **ÁRBOL** presenta tres predictores de los cuales dos utilizan las variables **T**media y **T**mínima en el mes de mayo y abril, respectivamente. El último predictor utiliza la variable **ETR** durante el mes de septiembre.

Modelos basados en Índices climáticos

Indice-Temporada	MLR_BIC	MLR_S	MULTI_BIC	ÁRBOL
TSA - Abr/May/Jun		X		
SAM - Mar	X			
SAM - Ago/Sep		X	X	
SAM - Jun/Jul/Ago		X		X
SAM - Jul			X	
SAM - Abr/May/Jun			X	
SAM - Abri	X			
IOD - Ene/Feb/Mar	X			
IOD - Mar				X
IOD - Mar/Abr			X	
IOD - Abr				X
NIÑO1+2 - May				X
NIÑO1+2 - Jun/Jul/Ago			X	

Tabla 4.16: Predictores de rendimientos para el cluster suroeste del cultivo de trigo. En la primera columna se muestra el índice junto con el periodo que se utiliza como predictor para cada modelo. Los modelos son: **MLR_BIC**: Modelo de regresión lineal múltiple con criterio Bayesiano, **MLR_S**: Modelo de regresión lineal múltiple con análisis de sensibilidad, **MULTI_BIC**: Modelo de regresión logística múltiple y **ÁRBOL**: Modelo de **ÁRBOL** de decisión.

La Tabla 4.16 muestra que el modelo **MLR_BIC** presenta tres predictores, de los cuales dos utilizan el índice **SAM** en el mes de marzo y en el mes de abril, mientras que el último de los predictores utiliza el índice **IOD** en el trimestre que va de enero a marzo. El modelo **MLR_S** presenta también tres predictores, de los cuales dos de ellos utilizan el índice **SAM** en el bimestre agosto-septiembre y en el trimestre que va de junio a agosto. El

predictor restante utiliza el índice **TSA** en el trimestre que va de abril a junio. El modelo **MULTI_BIC** utiliza cinco predictores donde tres de ellos utilizan el índice **SAM**: uno en el mes de julio, otro en el bimestre agosto-septiembre y el último en el trimestre que va de abril a junio. Los dos predictores restantes para completar el modelo utilizan los índices **IOD** y **NIÑO1+2** en el bimestre marzo-abril y en el trimestre que va de junio a agosto, respectivamente. El modelo **ÁRBOL** utiliza cuatro predictores, de los cuales dos utilizan el índice **IOD**, uno en marzo y el otro en abril. Los dos predictores que completan el modelo utilizan el índice **SAM** y el **NIÑO1+2** en el trimestre que va de junio a agosto y en el mes de mayo, respectivamente.

4.3.2. Cluster Noreste

Modelos basados en estaciones meteorológicas

Estacion-Variable-Temporada	MLR_BIC	MLR_S	MULTI_BIC	ÁRBOL
Villa Gesell - ETP - Mar		X		
Punta Indio - ETP - Sep	X			
Tandil - ETP - Feb				X
Nueve de Julio - ETP - May/Jun/Jul	X			
El Palomar - ETP - Feb			X	
Mar del Plata - ETR - Mar/Abr/May		X		
El Palomar - ETR - Jun/Jul/Ago		X		
Tandil - ETR - Jun	X			
Dolores - ETR - May/Jul		X		
Nueve de Julio - PP - Mar/Abr/May		X		
El palomar- PP - Mar			X	
El palomar - PP - Ago/Sep/Oct			X	
Punta Indio - PP - Marzo				X
Tandil - PP - Jul/Ago/Sep			X	
Mar del Plata - Tmáxima - Jun/Jul/Ago			X	
Ezeiza - Tmedia - Feb/Mar/Abr	X			
Punta Indio - Tmedia - Abr/May/Jul				X
Nueve de Julio - Tmínima - Sep	X			

Tabla 4.17: Idem a 4.15 pero para el cluster Noreste

La Tabla 4.17 ilustra que el modelo **MLR_BIC** presenta cinco predictores, de los cuales

dos utilizan la variable **ETP**: uno en el mes de septiembre y otro en el trimestre que va de mayo a julio. Otros dos predictores utilizan las variables **Tmedia** y **Tmínima**, durante el trimestre que va de febrero a abril y durante el mes de septiembre, respectivamente. Por último, para completar el modelo, el predictor restante utiliza la variable **ETR** durante el mes de junio. El modelo **MLR_S** utiliza tres predictores con la variable **ETR**, en los trimestres que van de marzo a mayo, junio a agosto y en el bimestre mayo-junio. Otro de los predictores utiliza la variable **ETP** durante el mes de marzo. El último predictor de este modelo utiliza la variable **PP** durante el trimestre que va de marzo a mayo. El modelo **MULTI_BIC** utiliza tres predictores con la variable **PP**, durante el mes de marzo y en los trimestres que van de agosto a octubre y de julio a septiembre. Otro de los predictores utiliza la variable **ETP** durante el mes de febrero y el último predictor de este modelo utiliza la variable **Tmáxima** durante el trimestre que va de junio a agosto. El último modelo **ÁRBOL** presenta tres predictores, el primero utiliza la variable **ETP** durante el mes de febrero, el segundo predictor utiliza la variable **PP** durante el mes de marzo. El predictor restante utiliza la variable **Tmedia** durante el trimestre que va de abril a junio.

Modelos basados en Índices climáticos

Indice-Temporada	MLR_BIC	MLR_S	MULTI_BIC	ÁRBOL
TSA - May/Jun/Jul		X		
SAODI - Mar/Abr				X
SAM - Abr/May/Jun	X		X	X
SAM - Feb/Mar/Abr		X		
SAM - Ago				X
SAM - Oct		X		
SAM - Sep/Oct			X	
SAM - May/Jun/Jul	X			
IOD - Abr	X			
IOD - Ago				X
NINO3 - Abr				X
NINO4 - Ago/Sep/Oct		X		

Tabla 4.18: Idem a 4.16 pero para el cluster Noreste

La Tabla 4.18 muestra que el modelo **MLR_BIC** presenta tres predictores, de los cuales dos utilizan el índice **SAM** en los trimestres que van de abril a junio y de mayo a julio. El predictor que completa este modelo utiliza el índice **IOD** en el mes de abril. El modelo **MLR_S** utiliza también dos predictores a partir del índice **SAM**, uno en el trimestre que va de febrero a abril y el otro en el mes de octubre. Otro de los predictores de este modelo utiliza el índice **TSA** en el trimestre que va de mayo a julio. El modelo **MULTI_BIC** presenta dos predictores utilizando el índice **SAM**: uno en el trimestre que va de abril a junio y otro en el bimestre septiembre-octubre. El modelo **ÁRBOL** presenta cinco predictores, de los cuales dos utilizan el índice **SAM** en el trimestre que va de abril a junio y en el mes de agosto. Otros dos predictores utilizan los índices **IOD** y **NIÑO3**, en los meses de agosto y abril, respectivamente. El último de los predictores de este modelo utiliza el índice **SAODI** en el bimestre marzo-abril.

Capítulo 5

Evaluacion de Modelos

En este capítulo se evalúan las métricas descritas en el Capítulo 2, sección 2.4.2: precisión, sensibilidad y los errores severos 1 y 3, provenientes de la matriz de confusión. La precisión y los errores severos se evaluaron para cada cultivo y para cada cluster, mientras que la sensibilidad se evaluó también para cada categoría de rendimiento: bajo, normal y alto. Además de los cuatro modelos utilizados, también se evaluó el Modelo medio o media del ensamble de modelos (ver Cap.2, sección. 2.3.4).

5.1. Girasol

5.1.1. Evaluación cluster Sur.

La Tabla 5.1 muestra que los modelos basados en estaciones presentan una precisión mayor que los modelos basados en índices climáticos. Por otro lado, la sen(2) muestra un valor mayor respecto a la sen(1) y sen(3), lo cual está asociado a una mayor predictibilidad de rendimientos de cultivos en torno a los valores normales de los mismos. El modelo **ÁRBOL** basado en estaciones presenta Error_sev_1 y Error_sev_3, tanto para estaciones como para índices. Esto indica que este modelo predijo rendimientos de cultivos inferiores (superiores) a lo normal cuando se observaron rendimientos superiores (inferiores) a lo normal

Variable	Modelo	Precisión	Sen(1)	Sen(2)	Sen(3)	Error_sev_1	Error_sev_3
Estaciones	MULTI_BIC	0.63	0.54**	0.63	0.72**	0	0.090
	ÁRBOL	0.68**	0.27	0.89**	0.72**	0.4	0.27
	MLR_BIC	0.60*	0.54**	0.78*	0.36**	0	0
	MLR_S	0.48	0.18	0.73	0.36**	0	0
	M.Medio	0.60*	0.45**	0.89**	0.27*	0	0
I.Climatico	MULTI_BIC	0.51	0.54**	0.78	0	0.25	0
	ÁRBOL	0.36	0.27	0.57	0.09	0.27	0.33
	MLR_BIC	0.41	0.27	0.78*	0.18	0	0
	MLR_S	0.39	0	0.84**	0	0	0
	M.Medio	0.48	0.36*	0.84**	0	0.12	0

Tabla 5.1: Valores de las métricas de evaluación para modelos basados en estaciones meteorológicas e índices climáticos para el cluster Sur de girasol. En la primera columna se muestra la variable, en la segunda el modelo y luego cada una de las métricas. En el caso de sensibilidad, sen(1) es la sensibilidad para el rendimiento bajo, sen(2) para el rendimiento normal y sen(3) para el rendimiento alto. Un asterisco indica el tercer mejor valor de cada métrica, doble asterisco indica el segundo mejor valor de cada métrica y doble asterisco y en negrita indica el mejor valor de cada métrica.

dentro del período evaluado. El modelo **MLR_BIC** basado en estaciones y el modelo **MULTI_BIC** presentan los valores de precisión, sen(1) y sen(2) de los más altos del grupo. Sin embargo el modelo **MLR_BIC** presenta un valor de sen(3) por debajo del umbral de 0.50, por lo tanto queda descartado para su uso para este cluster y para este cultivo. El modelo **MULTI_BIC** basado en estaciones presenta un error severo del tipo 3, es decir, que existe una probabilidad de que se pronostiquen rendimientos altos y se observen rendimientos bajos. A pesar de esto, al tratarse de una probabilidad menor a 0.1, también podría llegar a ser un modelo adecuado para pronosticar el rendimientos del cluster sur de girasol.

5.1.2. Evaluación cluster Centro.

De la Tabla 5.2 se desprende que el modelo **MLR_BIC** basado en estaciones presenta los mayores valores de precisión y sensibilidad para las categorías 1 y 3. En la sensibilidad de la categoría 2 observamos que los valores superiores se encuentran en los modelos basados en índices climáticos mientras que en la categoría 3 los valores más altos los presenta el modelo **MLR_BIC** basado en estaciones y el modelo **MULTI_BIC** basado en índices. A la hora

Variable	Modelo	Precisión	Sen(1)	Sen(2)	Sen(3)	Error_sev_1	Error_sev_3
Estaciones	MULTI_BIC	0.46	0.18	0.57	0.54**	0	0
	ÁRBOL	0.60	0.81**	0.63	0.36	0.07	0.1
	MLR_BIC	0.70**	0.81**	0.68	0.63**	0	0
	MLR_S	0.48	0.18	0.73	0.36	0	0
	M.Medio	0.63*	0.72**	0.68	0.45*	0	0
I.Climatico	MULTI_BIC	0.58	0.27	0.78	0.63**	0.25	0.25
	ÁRBOL	0.68**	0.63**	0.84*	0.45*	0.25	0.14
	MLR_BIC	0.43	0.27	0.73	0.09	0.14	0.25
	MLR_S	0.51	0.18	0.94**	0.09	0.33	0
	M.Medio	0.48	0.27	0.89**	0	0.28	0

Tabla 5.2: idem Tabla 5.1 pero para modelos aplicados al cluster centro de girasol.

de observar los errores severos podemos ver que una gran cantidad de modelos basados en índices presentan errores severos, tanto de tipo 1 como de tipo 3. El modelo **MLR_BIC** basado en estaciones presenta valores de precisión y sensibilidad por encima del umbral de 0.5 y además son los más altos del cluster con respecto a los demás modelos. Por otro lado, la sensibilidad del tipo 2 presenta valores aceptables y se destaca que este modelo no presenta errores severos de ninguno de los dos tipos posibles.

5.1.3. Evaluación cluster Norte.

Variable	Modelo	Precisión	Sen(1)	Sen(2)	Sen(3)	Error_sev_1	Error_sev_3
Estaciones	MULTI_BIC	0.53	0.83	0.68	0.36	0.22	0
	ÁRBOL	0.70*	0.72**	0.78	0.54	0.18	0.1
	MLR_BIC	0.60	0.54	0.73	0.83**	0	0
	MLR_S	0.51	0.45	0.68	0.27	0	0
	M.Medio	0.60	0.54	0.84*	0.27	0.1	0
I.Climatico	MULTI_BIC	0.51	0.18	0.68	0.72*	0	0.23
	ÁRBOL	0.70*	0.72**	0.84**	0.45	0.09	0
	MLR_BIC	0.63	0.54	0.89**	0.27	0	0.2
	MLR_S	0.97**	0.90**	1**	1**	0	0
	M.Medio	0.75**	0.63**	1**	0.45	0	0

Tabla 5.3: idem Tabla 5.1 pero para modelos aplicados al cluster norte de girasol

Observando la Tabla 5.3, se puede ver que en el caso de los modelos basados en índices,

el valor más alto de precisión se encuentra en el modelo **MLR_S** con 0.97. Este mismo modelo basado en índices también presenta los valores más altos de las sensibilidades en las tres categorías posibles de rendimiento. Asimismo, este modelo no tiene errores severos y sus métricas de sensibilidad y precisión se encuentran muy por encima del umbral de 0.5, por lo tanto podría ser uno de los modelos que mejor predice el rendimiento de girasol en el cluster norte. Dentro de los modelos basados en estaciones el modelo **ÁRBOL** y **MLR_BIC** son los que presentan métricas más elevadas pero siguen siendo inferiores a los modelos basados en índices. El modelo **ÁRBOL** además presenta errores severos de ambos tipos.

5.2. Maiz

5.2.1. Evaluación cluster Norte.

Variable	Modelo	Precisión	Sen(1)	Sen(2)	Sen(3)	Error_sev_1	Error_sev_3
Estaciones	MULTI_BIC	0.63	0.72*	0.68	0.45	0.9	0.5
	ÁRBOL	0.41	0.54	0.52	0.09	0	0.18
	MLR_BIC	0.60	0.54	0.73	0.83*	0.1	0.1
	MLR_S	0.48	0.45	0.68	0.18	0	0
	M.Medio	0.58	0.45	0.94**	0.09	0	0
I.Climatico	MULTI_BIC	0.90**	0.90**	0.89*	0.90**	0	0
	ÁRBOL	1**	1**	1**	1**	0	0
	MLR_BIC	0.53	0.54	0.57	0.45	0.09	0
	MLR_S	0.39	0.27	0.57	0.18	0.1	0.25
	M.Medio	0.75*	0.90**	0.84	0.45	0	0

Tabla 5.4: ídem Tabla 5.1 pero para modelos aplicados al cluster norte de maíz.

En la Tabla 5.4 se observan valores levemente más elevados en las métricas de modelos basados en índices que en los basados en estaciones. Respecto al modelo **MLR_S**, tanto el basado en índices como el basado en estaciones, en la mayoría de las métricas no supera el umbral de 0.50, y por lo tanto queda descartado por su bajo desempeño. Se destacan los modelos **MULTI_BIC** y **ÁRBOL** basados en índices, que presentan los valores más altos en sus métricas, entre 0.89 y 1, y no presentan errores severos. Por estos motivos, se consideran

los más adecuados del cluster para realizar pronósticos. El Modelo Medio basado en índices presenta valores elevados de precisión y de sensibilidad en las categorías 1 y 2, pero no son tan elevados como en los modelos **MULTI_BIC** y **ÁRBOL** basados en índices. Esto sugiere que el alto desempeño del Modelo Medio está explicado principalmente por estos dos modelos. Los modelos **MLR_BIC**, **MULTI_BIC** y **ÁRBOL** basado en estaciones presentan diferentes grados de errores severos.

5.2.2. Evaluación cluster Oeste

Variable	Modelo	Precisión	Sen(1)	Sen(2)	Sen(3)	Error_sev_1	Error_sev_3
Estaciones	MULTI_BIC	0.56**	0.54*	0.68*	0.36	0.9	0
	ÁRBOL	0.17	0.18	0.15	0.18	0.2	0.3
	MLR_BIC	0.63**	0.54*	0.73**	0.54**	0.12	0
	MLR_S	0.51*	0.63**	0.63	0.36	0	0.09
	M.Medio	0.56**	0.72**	0.73**	0.72**	0	0
I.Climatico	MULTI_BIC	0.39	0.09	0.63	0.45	0.28	0.33
	ÁRBOL	0.34	0.27	0.26	0.54**	0.13	0.09
	MLR_BIC	0.41	0.09	0.84**	0	0.3	0
	MLR_S	0.46	0.18	0.73**	0.27	0	0.1
	M.Medio	0.41	0.09	0.78	0.09	0.16	0

Tabla 5.5: ídem Tabla 5.1 pero para modelos aplicados al cluster oeste de maíz

De la Tabla 5.5 se desprende que **MLR_BIC** basado en estaciones es el modelo con mayor precisión para el cultivo de maíz en el cluster oeste. El valor más alto en la sensibilidad para la categoría 1 se encuentra en el Modelo Medio basado en estaciones, mientras que para la sensibilidad en la categoría 2 este modelo presenta el segundo mejor valor. La métrica más alta en esta categoría se encuentra en el modelo **MLR_BIC** basado en índices climáticos, pero también presenta sensibilidades que no superan el umbral de 0.50. La sensibilidad en la categoría 3 presenta su valor más elevado en el modelo Modelo Medio basado en estaciones. Además, este modelo no presenta errores severos, lo cual lo posiciona como uno de los modelos que mejor predice el rendimiento de maíz en el cluster oeste.

5.2.3. Evaluación cluster Este

Variable	Modelo	Precisión	Sen(1)	Sen(2)	Sen(3)	Error_sev_1	Error_sev_3
Estaciones	MULTI_BIC	0.68**	0.72**	0.78*	0.45*	0.1	0
	ÁRBOL	0.80**	0.36	0.94**	1**	0	0.3
	MLR_BIC	0.65*	0.54*	0.68	0.63**	0	0
	MLR_S	0.38	0.45*	0.47	0.18	0.16	0.12
	M.Medio	0.68**	0.45*	1	0.36	0	0
I.Climatico	MULTI_BIC	0.43	0.36	0.68	0.90**	0	0
	ÁRBOL	0.51	0.45*	0.68	0.27	0.1	0
	MLR_BIC	0.53	0.36	0.84**	0.18	0	0.33
	MLR_S	0.43	0.09	0.78*	0.18	0	0
	M.Medio	0.48	0.45	0.73	0.09	0	0

Tabla 5.6: ídem Tabla 5.1 pero para modelos aplicados al cluster este de maíz.

La Tabla 5.6 muestra las métricas de los modelos para el cluster este de maíz. Este cluster presenta valores en general más bajos en sus métricas que otros cluster, descartando los modelos que presentan precisión y sensibilidad por debajo del umbral de 0.5. Teniendo en cuenta esto, el modelo **ÁRBOL** basado en estaciones presentan valores máximos de precisión, mientras que el segundo mejor valor se encuentra en los modelos **MULTI_BIC** y Modelo Medio basados en estaciones. Por otro lado en las métricas de sensibilidad el modelo **ÁRBOL** basado en estaciones tiene valores de sensibilidad bajos para una categoría y altos para la otra, esto perjudica las predicciones de rendimiento. Lo mismo ocurre con esta métrica en los modelos basados en índices. Los errores severos se observan en modelos particulares como **MLR_S** y **ÁRBOL** basado en estaciones, y **ÁRBOL** y **MLR_BIC** basados en índices. Uno de los modelos más adecuados en este cluster parece ser **MLR_BIC** basado en estaciones, ya que si bien no presenta los valores más altos en todas las métricas, es el único en el cual todas superan el umbral de 0.5. Sumado a que no presenta errores severos, este modelo es el que mejor podría predecir el rendimiento del maíz en el cluster Este.

Variable	Modelo	Precisión	Sen(1)	Sen(2)	Sen(3)	Error_sev_1	Error_sev_3
Estaciones	MULTI_BIC	0.70**	0.63**	0.73	0.72**	0.16	0.09
	ÁRBOL	0.65*	0.45*	0.84*	0.54*	0.44	0.18
	MLR_BIC	0.48	0.45*	0.52	0.45	0.1	0
	MLR_S	0.43	0.27	0.73	0.09	0.37	0.2
	M.Medio	0.56	0.45*	0.84*	0.18	0	0
I.Climatico	MULTI_BIC	0.46	0	0.73	0.45	0.25	0.36
	ÁRBOL	0.46	0.27	0.78	0.090	0.25	0.16
	MLR_BIC	0.43	0.27	0.68	0.18	0	0
	MLR_S	0.97**	1**	1**	0.90**	0	0
	M.Medio	0.53	0.36	0.89**	0.09	0	0

Tabla 5.7: ídem Tabla 5.1 pero para modelos aplicados al cluster sur de maíz.

5.2.4. Evaluación cluster Sur

Las métricas que observamos en la Tabla 5.7 muestran que el modelo de mayor precisión es **MLR_S** basado en índices, mientras que el segundo y el tercer mejor valor se encuentran en los modelos **MULTI_BIC** y **ÁRBOL** basado en estaciones. En cuanto a métricas de sensibilidad el modelo **MLR_S** asociado a índices climáticos presenta los máximos valores en cada categoría, y sumado a que no presenta errores severos, es uno de los modelos que mejor predice el rendimiento de maíz en el cluster sur. Los modelos asociados a estaciones presentan métricas por debajo del umbral de 0.50, salvo el modelo **MULTI_BIC**, pero este presenta errores severos. Todos los modelos basados en estaciones presentan errores severos salvo el modelo Modelo Medio pero sus métricas no son tan elevadas.

5.3. Trigo

5.3.1. Evaluación cluster Noreste

Observando la Tabla 5.8 vemos que la precisión tiene sus tres valores más altos dentro de los modelos basados en estaciones, siendo la más alta la del modelo **MULTI_BIC**. Este modelo también presenta los valores más altos de sensibilidad para las categorías 1

Variable	Modelo	Precisión	Sen(1)	Sen(2)	Sen(3)	Error_sev_1	Error_sev_3
Estaciones	MULTI_BIC	078**	0.90**	0.73*	0.72**	0	0
	ÁRBOL	0.68*	0.63*	0.73*	0.63**	0.18	0.1
	MLR_BIC	0.75**	0.81**	0.78**	0.63**	0.18	0.1
	MLR_S	0.39	0.09	0.68	0.18	0.42	0
	M.Medio	0.75**	0.81**	1**	0.37*	0	0
I.Climatico	MULTI_BIC	0.46	0.63*	0.57	0.09	0.08	0.1
	ÁRBOL	0.56	0.36	0.57	0.72**	0	0.13
	MLR_BIC	0.53	0.63*	0.57	0.36	0	0
	MLR_S	0.41	0.18	0.78**	0	0	0
	M.Medio	0.51	0.54	0.78**	0	0	0

Tabla 5.8: ídem Tabla 5.1 pero para modelos aplicados al cluster noreste de trigo.

y 3 mientras que para la categoría dos tiene el tercer valor más alto. El modelo **MLR_BIC** basado en estaciones presenta los segundos valores más altos de sensibilidad para las tres categorías y tiene la segunda precisión más alta, pero a diferencia del modelo **MULTI_BIC** este modelo presenta errores severos del tipo 1. Los modelos basados en índices climáticos presentan valores de sensibilidad elevados, pero también presentan valores de esta métrica que no superan el umbral de 0.50 por este motivo son descartados para el pronóstico de rendimientos de cultivos. Por el análisis previo, se considera que el modelo **MULTI_BIC** basado en estaciones es uno de los más adecuados para predecir el rendimiento de trigo para el cluster noreste.

5.3.2. Evaluación cluster Suroeste

La Tabla 5.9 muestra que la precisión máxima se encuentra en el modelo **MULTI_BIC** basado en índices climáticos, y el segundo y tercer valor más alto se encuentran en este mismo modelo y en **MLR_BIC** basados en estaciones. La sensibilidad para la categoría 1 es máxima también en el modelo **MULTI_BIC** basado en índices climáticos, y los segundos valores más elevados se encuentran en los modelos basados en estaciones. La sensibilidad en la categoría 2 presenta valores más elevados en los modelos basados en índices climáticos, siendo máxima en el modelo Modelo Medio. La sensibilidad para la categoría 3 presenta su valor

Variable	Modelo	Precisión	Sen(1)	Sen(2)	Sen(3)	Error_sev_1	Error_sev_3
Estaciones	MULTI_BIC	0.70	0.72**	0.73	0.63*	0.09	0
	ÁRBOL	0.46	0.36*	0.68	0.18	0.12	0.28
	MLR_BIC	0.73**	0.72**	0.78	0.90**	0	0
	MLR_S	0.48	0.72**	0.42	0.36	0	0
	M.Medio	0.41	0.72**	0.73	0.45	0	0
I.Climatico	MULTI_BIC	0.80**	0.81**	0.78	0.81**	0	0
	ÁRBOL	0.56	0.36	0.68	0.54	0.1	0.2
	MLR_BIC	0.39	0	0.83*	0.09	0	0
	MLR_S	0.46	0.27	0.84**	0	0.16	0.5
	M.Medio	0.51	0.27	0.89**	0.09	0	0

Tabla 5.9: ídem Tabla 5.1 pero para modelos aplicados al cluster suroeste de trigo.

más elevado en el modelo **MLR_BIC** basado en estaciones. Varios de los modelos basados en índices climáticos y estaciones presentan errores severos en diferentes grados, mientras que el modelo **MULTI_BIC** basado en índices climáticos no presenta errores severos, pudiendo ser este último uno de los modelos más adecuados para predecir el rendimiento del trigo en el cluster suroeste. El modelo **MLR_BIC** basado en estaciones también presenta métricas que podrían predecir de manera adecuada el rendimiento de este cultivo en esta región, pero con un desempeño levemente inferior al del modelo **MULTI_BIC** basado en índices climáticos.

Capítulo 6

Conclusiones

En esta tesis se implementaron y evaluaron modelos de pronóstico de rendimientos de los cultivos de girasol, maíz y trigo para la provincia de Buenos Aires, Argentina. El objetivo principal fue obtener modelos de pronóstico que puedan aplicarse operativamente para uno de los sectores económicos y productivos más importantes de la región.

Los modelos utilizados fueron regresión lineal múltiple utilizando análisis de sensibilidad y criterio de información bayesiano para la selección de predictores. También se utilizaron un modelo multinomial mediante el criterio de selección bayesiano y un modelo basado en árbol de decisión. Por último, se definió un modelo medio que es el promedio de los valores predichos por los modelos anteriormente mencionados. Como predictores disponibles para la selección se utilizaron, por un lado variables meteorológicas observadas en estaciones meteorológicas de superficie de la provincia de Buenos Aires, y por otro se utilizaron índices de diferentes componentes del sistema acoplado océano-atmósfera que afectan a dichas variables.

Para el aprendizaje de los modelos anteriormente mencionados se utilizaron datos de rendimientos de cultivos de trigo, maíz y girasol para la provincia de Buenos Aires. Estos datos presentaban una tendencia positiva significativa a un nivel de confianza de 0.1, por lo que fue filtrada y removida. Esta tendencia está asociada a las tecnologías implementadas

en el sector agropecuario para mejorar el rendimiento de los cultivos (García Bernado, 2018).

Luego de realizar una caracterización del rendimiento de cada cultivo, se dividió la región de estudio en clusters mediante el método K-means. Primero, se determinó la cantidad de grupos o clusters y luego se determinaron las distintas regiones que ocupa cada grupo. La regionalización mostró que el cultivo de girasol queda dividido en tres clusters que se distribuyen de norte a sur mientras que el maíz está dividido en cuatro grupos sobre la región de estudio. Por su parte, el cultivo de trigo presenta dos divisiones, una comprende el noreste y la otra el suroeste de la provincia. Los clusters para el cultivo de girasol presentan una distribución similar a la que se observan en los estadios fenológicos de este cultivo propuestos por la oficina de riesgo agropecuario (ver capítulo 3, sección 3.2). En cuanto a las regionalizaciones realizadas para los cultivos de trigo y maíz, éstas no presentan una distribución similar a la de los estadios fenológicos. Esto podría deberse a la influencia de otros factores en la separación de regiones de estos dos cultivos.

Para evaluar el desempeño de los modelos se calculó la matriz de confusión para cada modelo y cada cluster. A partir de esta matriz se calcularon matrices de evaluación para determinar el desempeño de cada modelo. Se determinó que los modelos que no cumplan con una precisión y sensibilidad mayor al 0.50 para cada una de sus categorías serían descartados, debido a que esos valores pueden llevar a errores no aceptables en las predicciones. Hecho esto, se pasó a seleccionar el modelo que mejor predice el rendimiento de cada cluster para cada cultivo. Los resultados obtenidos de este análisis se resumen en la Tabla 6.1.

Cultivo	Cluster	Base Del Modelo	Mejor/es Modelo/s	Variable
Girasol	Norte	Índices Climáticos	MLR_S	SAODI_SON IOD_DIC NIÑO1+2_AGO NIÑO1+2_JAS NIÑO3_JJA
Girasol	Centro	Estaciones Meteorológicas	MLR_BIC	ETP_ASO ETR_ASO ETR_OCT PP_JUN PP_OCT
Girasol	Sur	Estaciones Meteorológicas	MULTI_BIC	ETR_AS ETR_JUL ETR_DIC PP_JA Tmax_JUN
Maiz	Norte	Índices Climáticos	ÁRBOL	TSA_JAS NIÑO1+2_AGO NIÑO1+2_SO
Maiz	Oeste	Estaciones Meteorológicas	Modelo Medio	
Maiz	Este	Estaciones Meteorológicas	MLR_BIC	ETP_OND ETR_ENE PP_ND PP_JUL Tmed_ASO
Maiz	Sur	Índices Climáticos	MLR_S	SAM_ND IOD_DE NIÑO3_DE NIÑO1+2_DE
Trigo	Noreste	Estaciones Meteorológicas	MULTI_BIC	ETP_FEB PP_MAR PP_ASO PP_JAS Tmax_JJA
Trigo	Suroeste	Índices Climáticos	MULTI_BIC	SAM_JUL SAM_AS SAM_AMJ IOD_MA NIÑO1+2_JJA

Tabla 6.1: Primera columna: cultivo, segunda columna: cluster, tercera columna: base de cada modelo, cuarta columna: mejores modelos y quinta columna: predictores de los mejores modelos.

Para el cluster sur y centro del cultivo de girasol, los modelos basados en estaciones utilizan como variables predictoras ETR, ETP, PP, y Temperatura máxima y media sobre los meses que van de abril a diciembre, principalmente en invierno y primavera. Por otro lado, los modelos basados en índices climáticos utilizaron los índices TSA, SAODI, IOD, SAM, NINO4 y NIÑO1+2, también principalmente en invierno y primavera.

Para ambos clusters los modelos basados en estaciones presentaron una mejor performance respecto a los basados en índices climáticos, siendo el mejor modelo para el cluster sur el MUTLI_BIC basado en estaciones y para el cluster centro el MLR_BIC, también basado en estaciones. Ambos modelos utilizan ETR y PP como variables predictores entre los meses de julio y octubre principalmente. El modelo MLR_BIC utiliza, además, la variable ETP. Los predictores de ambos modelos aparecen con mayor frecuencia en períodos invernales y primaverales. [Anderson et al. \(2017\)](#) destacan la importancia de la memoria que presenta el suelo en conservar la disponibilidad de agua para su posterior uso por los cultivos. Teniendo en cuenta que este cultivo se siembra a finales de noviembre, esto podría indicar una gran importancia de las condiciones del suelo antes del periodo de siembra. Esta inferencia es consistente con [Anastasi et al. \(2000\)](#) que encuentran que son de vital importancia las condiciones del suelo antes de la fecha de siembra. No obstante, estos resultados no se condicen con [Chapman & De la Vega \(2002\)](#), quienes encuentran que la variación del rendimiento de girasol podría estar asociada a las variaciones estacionales de PP durante los meses cálidos, durante la etapa de floración del cultivo ([Sierra & Murphy, 1986](#)).

El cluster norte del cultivo de girasol presentó una mayor cantidad de predictores utilizando la variable ETR para los períodos comprendidos en invierno y primavera, similar a lo encontrado en los clusters centro y sur. Los modelos basados en índices climáticos para este cluster mostraron una mayor cantidad de predictores utilizando el índice SAODI principalmente en otoño e invierno. Estos modelos también utilizan predictores basados en los índices SAM, IOD, NIÑO1+2 y NIÑO3 con predominio de los periodos invernales.

A diferencia de los clusters sur y centro, los modelos basados en índices climáticos para el cluster norte de girasol mostraron un mejor desempeño respecto a los basados en estaciones. El modelo que mejor predice el rendimiento del girasol en este cluster es el MLR_S basado en índices climáticos. Éste presenta predictores utilizando los índices SAODI, IOD, NIÑO1+2 y NIÑO3, principalmente, durante periodos invernales y primaverales. Uno de los predictores que utiliza el modelo es el índice IOD durante el mes de diciembre. Esto podría estar asociado a lo encontrado por [Chan et al. \(2008\)](#) quienes mostraron que la precipitación en la región de la cuenca del Plata durante la primavera y verano austral aumenta cuando el dipolo del océano Índico (IOD) se encuentra en su fase positiva. Como se mencionó previamente, estas variaciones de PP durante los meses cálidos podrían estar asociadas al rendimiento del girasol, cuya etapa de floración ocurre en esos meses ([Chapman & De la Vega, 2002](#)). La fase positiva del índice SAODI durante los meses invernales y primaverales se asocia a menores precipitaciones en la zona contenida por el cluster norte ([Garbarini & Gonzalez, 2014](#)). La selección de los predictores utilizando los índices NIÑO 1+2 y NIÑO 3 en periodos invernales podría asociarse al comienzo de eventos El niño o La niña, ya que los mismos suelen comenzar en el invierno austral y llegar a su máximo de anomalías durante los meses de verano ([Hurtado & Agosta, 2021](#); [Ropelewski & Halpert, 1996](#); [Grimm et al., 2000](#)). Los eventos El Niño (La Niña) están asociados a anomalías positivas (negativas) de precipitación en la región de estudio durante la primavera y verano austral ([Berri et al., 2002](#); [Hurtado, 2022](#)).

Por otro lado, el fenómeno ENOS también modula la temperatura sobre gran parte del Sudeste de Sudamérica (SESA). [Rusticucci et al. \(2017\)](#) encontraron que durante eventos cálidos del ENOS, se favorecen los días más cálidos durante el invierno y condiciones más frías en verano, donde diciembre presenta una disminución significativa de su amplitud térmica. El caso contrario ocurre durante eventos fríos del ENOS; desde noviembre a febrero se observa una mayor cantidad de días cálidos, mientras que durante el invierno se observan temperaturas más frías respecto a la normal. La modulación en las precipitaciones y en las

temperaturas que generan estos forzantes, tanto en el periodo previo a la siembra (invierno - primavera) como en el periodo de floración (verano) podrían ser las causantes de una variación en el rendimiento final del girasol.

El modelo que mejor predice el rendimiento del cultivo de maíz en el cluster sur fue el modelo MLR_S basado en índices climáticos. Este modelo utiliza predictores conformados por los índices SAM, IOD, NIÑO3 y NIÑO1+2, en su mayoría durante los meses de noviembre a enero. El uso del índice SAM durante periodos que van de noviembre a diciembre, podría estar asociado a las anomalías negativas en la precipitación durante fases positivas del índice SAM (Reboita et al., 2021). Los predictores que utilizan índices NIÑO1+2 y NIÑO3 lo hacen en periodos entre fines de primavera y principios del verano, coincidente con lo encontrado por AJ (1992). Ellos encontraron que el rendimiento del maíz probablemente está regulado por las precipitaciones superiores (inferiores) a lo normal entre octubre y febrero, típicas de los episodios cálidos (fríos) del ENOS. Otro de los predictores utiliza el índice IOD durante fines de primavera y verano. Esto podría estar asociado a distintos factores, uno de ellos podría ser que durante la fase positiva del IOD se observa un aumento de la precipitación en la región de estudio (Chan et al., 2008). Además, Sun et al. (2015), y Wang & Wang (2014) han demostrado que el ENOS y el IOD presentan un acople en sus distintas fases. Por otro lado, durante la primavera austral existen correlaciones positivas entre el IOD y la temperatura en la zona central de Argentina afectando la región de estudio (Saji et al., 2005). Lobell et al. (2011) muestran cómo un aumento en la temperatura media diaria y semanal durante la etapa de floración podría afectar el rendimiento del maíz. Un aumento de temperaturas favorece un aumento de la evaporación de agua del suelo, lo cual a su vez provoca un aumento de la demanda de precipitaciones en la etapa crítica para definir el rendimiento del maíz. Otro de los factores a considerar es el aumento de humedad que se produce en la fase positiva del IOD durante el periodo noviembre-diciembre sobre la región de estudio (Sena & Magnusdottir, 2021).

El modelo que mejor predice el rendimiento del maíz en el cluster Este es MLR_BIC

basado en estaciones meteorológicas. El mismo utiliza las variables de ETP, ETR, PP y Temperatura media, principalmente en periodos que van de fines del invierno a principios del verano austral. Por otro lado, el modelo que mejor predice el rendimiento del cultivo de maíz en el cluster Oeste es el Modelo Medio basado en estaciones meteorológicas, donde también se utilizan predictores a través de las variables ETP, ETR, PP y Temperatura en periodos que van del invierno al verano austral. La selección de periodos invernales, primaverales y de verano en variables de ETP, ETR y PP podría estar asociado a la importancia de la disponibilidad de agua en el suelo antes y durante el periodo de siembra (Anderson et al., 2017). Uno de los periodos más importantes en la fenología del maíz es el de floración, que es de vital importancia en el rendimiento final del maíz (Andrade et al., 1999). En nuestra región este periodo es aproximadamente a finales de la primavera y comienzos del verano austral. Durante este periodo es importante que el cultivo cuente con disponibilidad de agua en el suelo asociada a la precipitación caída en los meses pasados, así como también a las precipitaciones caídas durante la etapa de floración. Otro de los predictores utiliza la temperatura media en el trimestre que va de agosto a octubre. Esto se podría asociar al proceso de polinización, ya que se sabe que las temperaturas extremadamente altas afectan el proceso de polinización (Wilhelm et al., 1999), y esto influye en el rendimiento final del maíz. No obstante no se ha encontrado bibliografía que estudie la influencia en el maíz de la temperatura en los meses que van de agosto a octubre.

El modelo que mejor predice el rendimiento de maíz en el cluster Norte es el ÁRBOL basado en índices. Este utiliza como predictores a los índices TSA durante el trimestre que va de julio a septiembre y NIÑO1+2 en el trimestre que va de agosto a octubre. Barros et al. (2000) encontraron que las anomalías de la temperatura de la superficie del mar en el Atlántico tropical influyen en la posición de la SACZ durante el verano austral. Estos desplazamientos tienden a inhibir (aumentar) las precipitaciones en la zona de estudio debido al aumento (disminución) de la subsidencia. Por lo tanto, el uso de la TSA en invierno como predictor podría estar asociado al inicio de una anomalía de TSA para finales de primavera

y verano, la cual podría influir en la disponibilidad de agua en el suelo a través de la precipitación. Esto podría afectar el rendimiento final del cultivo de maíz debido a que durante estos meses está en su etapa crítica ([Andrade et al., 1999](#)). Como para el caso del cluster Norte de girasol, la presencia del índice NIÑO1+2 durante fines de invierno y principios de primavera austral pueden estar asociados al comienzo de un evento El Niño o La Niña.

El modelo que mejor predice el rendimiento del trigo en el cluster Suroeste es el MULTI_BIC basado en índices, el cual está formado por predictores que utilizan el índice SAM desde abril hasta septiembre, el índice IOD durante marzo y abril y el índice NIÑO1+2 en el trimestre invernal. Diversos autores han encontrado evidencias de que existen correlaciones positivas entre el índice SAM y la precipitación en el sudeste de sudamérica (SESA), en los periodos entre abril y septiembre ([Marshall, 2003](#); [Thompson & Wallace, 2000](#); [Silvestri & Vera, 2003](#), entre otros). Teniendo en cuenta que el trigo es un cultivo que se siembra en junio, el uso de este predictor sugiere una influencia de la precipitación desde los meses de abril y mayo, antes de que el cultivo se siembre. Posterior a esto, el uso de este predictor también puede tener influencia en las condiciones del suelo antes de la etapa crítica del trigo que comienza en septiembre ([Slafer & Rawson, 1994](#)). No se han encontrado trabajos en la bibliografía que asocien el IOD en otoño con variaciones de la precipitación o temperatura en la región de estudio. Debido a esto, no se puede encontrar una relación directa entre este índice y el rendimiento del trigo.

El modelo que mejor predice el rendimiento del trigo para el cluster Noreste es el MULTI_BIC basado en estaciones. Utiliza ETP, PP y temperatura máxima y sus periodos van desde otoño hasta principios de primavera. El uso de PP en marzo es consistente con lo encontrado por [Barkley et al. \(2014\)](#), quienes muestran una correlación positiva entre las variables que afectan la disponibilidad de agua en el suelo y el cultivo de trigo durante el otoño. La selección de estos periodos en los predictores infiere en la importancia de las condiciones del suelo antes del periodo de floración del trigo ([Nalley et al., 2009](#)). [Musick et al. \(1994\)](#) encuentra una relación entre el estrés por temperatura y el rendimiento del

trigo durante la primavera. Diversos estudios muestran la importancia de la temperatura en primavera para el cultivo de trigo ([Musick et al., 1994](#); [Warrick & Miller, 1999](#)), no obstante ningún predictor utiliza la temperatura mínima como variable.

Por último, de los modelos seleccionados como más adecuados para predecir el rendimiento de los cultivos estudiados en esta tesis, tres utilizan el modelo MULTI con el criterio de selección bayesiano. El modelo MLR es utilizado en cuatro ocasiones pero este se divide en dos que utilizan el criterio de selección bayesiano y otros dos que utilizan el criterio de sensibilidad. En un sólo caso se utiliza el modelo ÁRBOL y el Modelo Medio. Estos resultados sugieren que MLR y MULTI son modelos con buen desempeño que logran capturar la interacción entre las variables. La buena performance de MLR por sobre otros modelos puede deberse a su capacidad de modelar las conexiones entre distintas variable meteorológicas ([Khan & Naeem, 2024](#)). No obstante, el criterio de selección de variables predictoras de MLR no es trivial y ambos criterios utilizados dieron buenos resultados. La base de los modelos que se mencionaron anteriormente presenta cinco cluster que utilizan variables medidas en estaciones meteorológicas, mientras que los cuatro restantes utilizan índices climáticos. Cabe mencionar que no se observa una relación clara entre la base de los predictores y cluster.

Esta tesis aportó en el conocimiento sobre la predictibilidad de los rendimientos de tres cultivos en la provincia de Buenos Aires. Se encontraron modelos que presentaron buen desempeño, con potencialidad de uso operativo. No obstante, aún se requieren más estudios para tener mejores modelos en algunas regiones, y también para poder extender el estudio al resto de las regiones productivas del país. Por último, sería de suma utilidad para los usuarios poder ampliar el estudio al cultivo de la soja, de vital importancia económica en la región, pero que al no disponer de datos históricos de rendimiento, no se pudo incluir en el marco de la presente tesis.

Bibliografía

AJ, H. (1992). Field-crop systems of the pampas. *Field crop ecosystems*, (pp. 413–449).

Alvarez, M. S., Vera, C. S., Kiladis, G. N., & Liebmann, B. (2016). Influence of the madden julian oscillation on precipitation and surface air temperature in south america. *Climate Dynamics*, 46, 245–262.

Anastasi, U., Cammarata, M., & Abbate, V. (2000). Yield potential and oil quality of sunflower (oleic and standard) grown. *Ital. J. Agron*, 4(1), 23–36.

Anderson, W., Seager, R., Baethgen, W., & Cane, M. (2017). Life cycles of agriculturally relevant enso teleconnections in north and south america. *International Journal of Climatology*, 37(8), 3297–3318.

Andrade, F. H., Vega, C., Uhart, S., Cirilo, A., Cantarero, M., & Valentinuz, O. (1999). Kernel number determination in maize. *Crop Science*, 39(2), 453–459.

Andrian L. G, V. C. S. & M., O. (2022). Impactos del dipolo de océano indico en sesa y su acción combinada con el niño oscilación del sur sobre el clima de sudamérica durante la primavera. In *Congreso Argentino de Meteorología XIV. Actas de Congreso*.

Azcuy Ameghino, E. & Fernández, D. A. (2020). El censo nacional agropecuario 2018: visión general y aproximación a la región pampeana. *Revista Interdisciplinaria de Estudios Agrarios*.

- Barkley, A., Tack, J., Nalley, L. L., Bergtold, J., Bowden, R., & Fritz, A. (2014). Weather, disease, and wheat breeding effects on kansas wheat varietal yields, 1985 to 2011. *Agronomy Journal*, 106(1), 227–235.
- Barros, V., Gonzalez, M., Liebmann, B., & Camilloni, I. (2000). Influence of the south atlantic convergence zone and southatlantic sea surface temperature on interannual summer-rainfall variability in southeastern south america. *Theoretical and Applied Climatology*, 67, 123–133.
- Barrucand, M., Rusticucci, M., & Vargas, W. (2008). Temperature extremes in the south of south america in relation to atlantic ocean surface temperature and southern hemisphere circulation. *Journal of Geophysical Research: Atmospheres*, 113(D20).
- Berri, G. J., Ghietto, M. A., & García, N. O. (2002). The influence of enso in the flows of the upper paraná river of south america over the past 100 years. *Journal of Hydrometeorology*, 3(1), 57–65.
- Bert, F. E., Satorre, E. H., Toranzo, F. R., & Podestá, G. P. (2006). Climatic information and decision-making in maize crop production systems of the argentinean pampas. *Agricultural systems*, 88(2-3), 180–204.
- Brisson, N., Mary, B., Ripoche, D., Jeuffroy, M. H., Ruget, F., Nicoullaud, B., Gate, P., Devienne-Barret, F., Antonioletti, R., Durr, C., et al. (1998). Stics: a generic model for the simulation of crops and their water and nitrogen balances. i. theory and parameterization applied to wheat and corn. *Agronomie*, 18(5-6), 311–346.
- Cai, W., McPhaden, M. J., Grimm, A. M., Rodrigues, R. R., Taschetto, A. S., Garreaud, R. D., Dewitte, B., Poveda, G., Ham, Y.-G., Santoso, A., et al. (2020). Climate impacts of the el niño–southern oscillation on south america. *Nature Reviews Earth & Environment*, 1(4), 215–231.

-
- Casarin, D. P. & Kousky, V. E. (1986). Anomalias de precipitação no sul do brasil e variações na circulação atmosférica. *Revista Brasileira de Meteorologia*, 1(2), 83–90.
- Chan, S. C., Behera, S. K., & Yamagata, T. (2008). Indian ocean dipole influence on south american rainfall. *Geophysical Research Letters*, 35(14).
- Chapman, S. C. & De la Vega, A. J. (2002). Spatial and seasonal effects confounding interpretation of sunflower yields in argentina. *Field Crops Research*, 73(2-3), 107–120.
- Cowtan, K., Rohde, R., & Hausfather, Z. (2018). Evaluating biases in sea surface temperature records using coastal weather stations. *Quarterly Journal of the Royal Meteorological Society*, 144(712), 670–681.
- de la Casa, A. C., Ovando, G. G., & Díaz, G. J. (2021). Enso influence on corn and soybean yields as a base of an early warning system for agriculture in córdoba, argentina. *European Journal of Agronomy*, 129, 126340.
- De Wit, C. & Van Keulen, H. (1987). Modelling production of field crops and its requirements. *Geoderma*, 40(3-4), 253–265.
- Dinh, H. T., Bonner, A., Ramsbotham, J., & Clark, R. (2019). Cluster randomized controlled trial testing the effectiveness of a self-management intervention using the teach-back method for people with heart failure. *Nursing & Health Sciences*, 21(4), 436–444.
- Ferreira, R. A., Podestá, G. P., Messina, C. D., Letson, D., Dardanelli, J., Guevara, E., & Meira, S. (2001). A linked-modeling framework to estimate maize production risk associated with enso-related climate variability in argentina. *Agricultural and Forest Meteorology*, 107(3), 177–192.
- Garbarini, E. & Gonzalez, M. (2014). Variabilidad de la precipitación en argentina: factores de influencia. In *the Proceedings of 2do Congreso Internacional de Hidrología de Llanuras, Santa Fe, Argentina* (pp. 23–26).

- García Bernado, R. (2018). Cambio tecnológico en la producción agraria argentina: algunas teorías en disputa. *Mundo agrario*, 19(41), 90–e1.
- Gocic, M. & Trajkovic, S. (2013). Analysis of changes in meteorological variables using mann-kendall and sen's slope estimator statistical tests in serbia. *Global and planetary change*, 100, 172–182.
- Gonzalez, P. L. & Vera, C. S. (2014). Summer precipitation variability over south america on long and short intraseasonal timescales. *Climate dynamics*, 43, 1993–2007.
- Grimm, A. M. (2003). The el niño impact on the summer monsoon in brazil: regional processes versus remote influences. *Journal of Climate*, 16(2), 263–280.
- Grimm, A. M., Barros, V. R., & Doyle, M. E. (2000). Climate variability in southern south america associated with el niño and la niña events. *Journal of climate*, 13(1), 35–58.
- Grimm, A. M. & Zilli, M. T. (2009). Interannual variability and seasonal evolution of summer monsoon rainfall in south america. *Journal of Climate*, 22(9), 2257–2275.
- Griva, L. & Basualdo, M. (2018). Clasificación de pacientes con diabetes mellitus tipo 1 mediante técnicas de árbol de decisión. In *IX Congreso Argentino de Informática y Salud (CAIS)-JAIIO 47 (CABA, 2018)*.
- Grolemund, G. & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of statistical software*, 40, 1–25.
- Guarín Giraldo, G. W. (2011). *Impacto de la variabilidad climática en la producción de banano en el Urabá antioqueño*. PhD thesis, Universidad Nacional de Colombia.
- Gulati, P., Sharma, A., & Gupta, M. (2016). Theoretical study of decision tree algorithms to identify pivotal factors for performance improvement: A review. *International Journal of Computer Applications*, 141(14), 19–25.

- Hartigan, J. A. & Wong, M. A. (1979). Algorithm as 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 28(1), 100–108.
- Huang, B., Angel, W., Boyer, T., Cheng, L., Chepurin, G., Freeman, E., Liu, C., & Zhang, H.-M. (2018). Evaluating sst analyses with independent ocean profile observations. *Journal of Climate*, 31(13), 5015–5030.
- Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., Menne, M. J., Smith, T. M., Vose, R. S., & Zhang, H.-M. (2017). Extended reconstructed sea surface temperature, version 5 (ersstv5): upgrades, validations, and intercomparisons. *Journal of Climate*, 30(20), 8179–8205.
- Hurtado, S. I. (2022). *Precipitación observada en la temporada húmeda sobre Argentina subtropical: calidad de datos, variabilidad y forzantes asociados de gran escala*. PhD thesis, Universidad Nacional de La Plata.
- Hurtado, S. I. & Agosta, E. A. (2021). El niño southern oscillation-related precipitation anomaly variability over eastern subtropical south america: Atypical precipitation seasons. *International Journal of Climatology*, 41(7), 3793–3812.
- Hurtado, S. I., Agosta, E. A., & Zaninelli, P. G. (2023). Monthly variations of forcing mechanisms of austral summer precipitation in subtropical argentina. *Atmospheric Research*, 285, 106609.
- Hurtado, S. I., Agosta, E. A., & Zaninelli, P. G. (2024). Tropical-subtropical south american midsummer precipitation under enso events. *METEOROLOGICA*. *In press*.
- Hurtado, S. I., Zaninelli, P. G., Agosta, E. A., & Ricetti, L. (2021). Infilling methods for monthly precipitation records with poor station network density in subtropical argentina. *Atmospheric Research*, 254, 105482.

- Irizarry, R. A. (2019). *Introduction to data science: Data analysis and prediction algorithms with R*. CRC Press.
- Jorgetti, T., da Silva Dias, P. L., & de Freitas, E. D. (2014). The relationship between south atlantic sst and sacz intensity and positioning. *Climate dynamics*, 42, 3077–3086.
- Kassambara, A. (2016). Factoextra: extract and visualize the results of multivariate data analyses. *R package version*, 1.
- Kaufman, L. & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons.
- Khan, S. F. & Naeem, U. A. (2024). Performance evaluation of various techniques in estimating precipitation record of a sparsely gauged mountainous watershed. *Environmental Monitoring and Assessment*, 196(2), 112.
- LaValley, M. P. (2008). Logistic regression. *Circulation*, 117(18), 2395–2399.
- Link, E. J. (2005). *Investigation and modeling of the optimization potential of adapted nitrogen fertilization strategies in corn cropping systems with regard to minimize nitrogen losses*. PhD thesis, Universität Hohenheim.
- Lobell, D. B., Hammer, G. L., McLean, G., Messina, C., Roberts, M. J., & Schlenker, W. (2013). The critical role of extreme heat for maize production in the united states. *Nature climate change*, 3(5), 497–501.
- Lobell, D. B., Schlenker, W., & Costa-Roberts, J. (2011). Climate trends and global crop production since 1980. *Science*, 333(6042), 616–620.
- Long, M. E. F., Spescha, L., Barnatán, I., & Murphy, G. (2013). Modelo de balance hidrológico operativo para el agro (bhoa). *Agronomía & Ambiente*, 32(1-2).
- Maechler, M. et al. (2013). Cluster analysis extended rousseeuw et al. *R CRAN*.

- Magrin, G. O., Travasso, M. I., & Rodríguez, G. R. (2005). Changes in climate and crop production during the 20th century in argentina. *Climatic change*, 72, 229–249.
- Marshall, G. J. (2003). Trends in the southern annular mode from observations and reanalyses. *Journal of climate*, 16(24), 4134–4143.
- Meinke, H. & Stone, R. C. (2005). Seasonal and inter-annual climate forecasting: the new tool for increasing preparedness to climate variability and change in agricultural planning and operations. *Climatic change*, 70(1-2), 221–253.
- Melice, J.-L. & Servain, J. (2003). The tropical atlantic meridional sst gradient index and its relationships with the soi, nao and southern ocean. *Climate Dynamics*, 20, 447–464.
- Montini, T. L., Jones, C., & Carvalho, L. M. (2019). The south american low-level jet: A new climatology, variability, and changes. *Journal of Geophysical Research: Atmospheres*, 124(3), 1200–1218.
- Musick, J. T., Jones, O. R., Stewart, B. A., & Dusek, D. A. (1994). Water-yield relationships for irrigated and dryland wheat in the us southern plains. *Agronomy Journal*, 86(6), 980–986.
- Nalley, L. L., Barkley, A. P., & Sayre, K. (2009). Photothermal quotient specifications to improve wheat cultivar yield component models. *Agronomy Journal*, 101(3), 556–563.
- Neath, A. A. & Cavanaugh, J. E. (2012). The bayesian information criterion: background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(2), 199–203.
- Nnamchi, H. C., Li, J., & Anyadike, R. N. (2011). Does a dipole mode really exist in the south atlantic ocean? *Journal of Geophysical Research: Atmospheres*, 116(D15).
- Nogues, J. (2015). Barreras sobre las exportaciones agropecuarias: impactos económicos

- y sociales de su eliminación. *Serie de informes técnicos del Banco Mundial en Argentina, Paraguay y Uruguay*, 3.
- Nogués-Paegle, J. & Mo, K. C. (1997). Alternating wet and dry conditions over south america during summer. *Monthly Weather Review*, 125(2), 279–291.
- Oram, P. A. (1989). Sensitivity of agricultural production to climatic change, an update.
- Pántano, V. C., Spescha, L. B., Penalba, O. C., & Murphy, G. M. (2014). Influencia de la variabilidad de temperatura y precipitación en la situación hídrica del suelo, en la región oriental de secano de la argentina. *Meteorológica*, 39(2), 21–36.
- Podestá, G. P., Messina, C. D., Grondona, M. O., & Magrin, G. O. (1999). Associations between grain crop yields in central-eastern argentina and el niño–southern oscillation. *Journal of Applied Meteorology and Climatology*, 38(10), 1488–1498.
- Rebella, C., Díaz, R., & Funston, L. (1984). Tendencias de lluvias y rendimientos de maíz en la región maicera típica argentina (precipitation trends and maize yields in argentina’s corn belt). In *Proc. III Congreso Nacional de Maíz* (pp. 129–133).
- Reboita, M. S., Ambrizzi, T., Crespo, N. M., Dutra, L. M. M., Ferreira, G. W. d. S., Rehbein, A., Drumond, A., da Rocha, R. P., & Souza, C. A. d. (2021). Impacts of teleconnection patterns on south america climate. *Annals of the New York Academy of Sciences*, 1504(1), 116–153.
- Reboita, M. S., Ambrizzi, T., & Rocha, R. P. d. (2009). Relationship between the southern annular mode and southern hemisphere atmospheric systems. *Revista Brasileira de Meteorologia*, 24, 48–55.
- Ripley, B., Venables, W., & Ripley, M. B. (2016). Package ‘nnet’. *R package version*, 7(3-12), 700.

- Ropelewski, C. F. & Halpert, M. S. (1996). Quantifying southern oscillation-precipitation relationships. *Journal of climate*, 9(5), 1043–1059.
- Rusticucci, M., Barrucand, M., & Collazo, S. (2017). Temperature extremes in the argentina central region and their monthly relationship with the mean circulation and enso phases. *International Journal of Climatology*, 37(6), 3003–3017.
- Rusticucci, M. & Vargas, W. (2002). Cold and warm events over argentina and their relationship with the enso phases: risk evaluation analysis. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 22(4), 467–483.
- Saji, N., Ambrizzi, T., & Ferraz, S. E. T. (2005). Indian ocean dipole mode events and austral surface air temperature anomalies. *Dynamics of atmospheres and oceans*, 39(1-2), 87–101.
- Saji, N. & Yamagata, T. (2003). Possible impacts of indian ocean dipole mode events on global climate. *Climate research*, 25(2), 151–169.
- Seager, R., Cane, M., Henderson, N., Lee, D.-E., Abernathey, R., & Zhang, H. (2019). Strengthening tropical pacific zonal sea surface temperature gradient consistent with rising greenhouse gases. *Nature Climate Change*, 9(7), 517–522.
- Sena, A. C. & Magnusdottir, G. (2021). Influence of the indian ocean dipole on the large-scale circulation in south america. *Journal of Climate*, 34(15), 6057–6068.
- Shao, Z. & Er, M. J. (2016). Efficient leave-one-out cross-validation-based regularized extreme learning machine. *Neurocomputing*, 194, 260–270.
- Sierra, E. M. & Murphy, G. M. (1986). Agroclimas del cultivo de girasol en la argentina. *Revista de la Facultad de Agronomía, UBA*, 7, 25–44.
- Silvestri, G. E. & Vera, C. S. (2003). Antarctic oscillation signal on precipitation anomalies over southeastern south america. *Geophysical Research Letters*, 30(21).

-
- Simpson, I. R., Deser, C., McKinnon, K. A., & Barnes, E. A. (2018). Modeled and observed multidecadal variability in the north atlantic jet stream and its connection to sea surface temperatures. *Journal of Climate*, 31(20), 8313–8338.
- Slafer, G. A. & Rawson, H. (1994). Sensitivity of wheat phasic development to major environmental factors: a re-examination of some assumptions made by physiologists and modellers. *Functional Plant Biology*, 21(4), 393–426.
- Stöckle, C. O., Donatelli, M., & Nelson, R. (2003). Cropsyst, a cropping systems simulation model. *European journal of agronomy*, 18(3-4), 289–307.
- Stuecker, M. F., Timmermann, A., Jin, F.-F., Chikamoto, Y., Zhang, W., Wittenberg, A. T., Widiasih, E., & Zhao, S. (2017). Revisiting enso/indian ocean dipole phase relationships. *Geophysical Research Letters*, 44(5), 2481–2492.
- Sun, S., Lan, J., Fang, Y., Gao, X., et al. (2015). A triggering mechanism for the indian ocean dipoles independent of enso. *Journal of Climate*, 28(13), 5063–5076.
- Taschetto, A. S. & Wainer, I. (2008). The impact of the subtropical south atlantic sst on south american precipitation. In *Annales Geophysicae*, volume 26/N°11 (pp. 3457–3476).: Copernicus Publications Göttingen, Germany.
- Team, R. D. C. (2010). R: A language and environment for statistical computing. (*No Title*).
- Tedeschi, R. G., Grimm, A. M., & Cavalcanti, I. F. (2015). Influence of central and east enso on extreme events of precipitation in south america during austral spring and summer. *International Journal of Climatology*, 35(8), 2045–2064.
- Therneau, T., Atkinson, B., & Ripley, B. (2015). rpart: Recursive partitioning and regression trees. *R package version*, 4, 1–9.
- Thompson, D. W. & Wallace, J. M. (2000). Annular modes in the extratropical circulation. part i: Month-to-month variability. *Journal of climate*, 13(5), 1000–1016.

- Travasso, M. I., Magrin, G. O., Grondona, M. O., & Rodriguez, G. R. (2009). The use of sst and soi anomalies as indicators of crop yield variability. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 29(1), 23–29.
- Van Der Wiel, K., Matthews, A. J., Stevens, D. P., & Joshi, M. M. (2015). A dynamical framework for the origin of the diagonal south pacific and south atlantic convergence zones. *Quarterly Journal of the Royal Meteorological Society*, 141(691), 1997–2010.
- Vasconcellos, F. C., Pizzochero, R. M., & de Albuquerque Cavalcanti, I. F. (2019). Month-to-month impacts of southern annular mode over south america climate. *Anuário do Instituto de Geociências*, 42(1), 783–792.
- Wang, C., Deser, C., Yu, J.-Y., DiNezio, P., & Clement, A. (2017). El niño and southern oscillation (enso): a review. *Coral reefs of the eastern tropical Pacific*, (pp. 85–106).
- Wang, X. & Wang, C. (2014). Different impacts of various el niño events on the indian ocean dipole. *Climate dynamics*, 42, 991–1005.
- Warrick, B. E. & Miller, T. D. (1999). Freeze injury on wheat. *Texas Agricultural Extension Service The Texas A and M University System, San Angelo, Texas*.
- Wickham, H. (2011). ggplot2. *Wiley interdisciplinary reviews: computational statistics*, 3(2), 180–185.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the tidyverse. *Journal of open source software*, 4(43), 1686.
- Wilhelm, E., Mullen, R., Keeling, P., & Singletary, G. (1999). Heat stress during grain filling in maize: effects on kernel growth and metabolism. *Crop science*, 39(6), 1733–1741.
- Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences*, volume 100. Academic press.