



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

**Uso de productos satelitales y
meteorológicos para la estimación del
rendimiento de la caña de azúcar**

TESIS

Que para obtener el título de

Ingeniero Geomático

P R E S E N T A

Vivaldo Isaí García Perales

DIRECTORA DE TESIS

M.C.T. María Elena Osorio Tai



Ciudad Universitaria, Cd. Mx., 2025



**PROTESTA UNIVERSITARIA DE INTEGRIDAD Y
HONESTIDAD ACADÉMICA Y PROFESIONAL
(Titulación con trabajo escrito)**



De conformidad con lo dispuesto en los artículos 87, fracción V, del Estatuto General, 68, primer párrafo, del Reglamento General de Estudios Universitarios y 26, fracción I, y 35 del Reglamento General de Exámenes, me comprometo en todo tiempo a honrar a la institución y a cumplir con los principios establecidos en el Código de Ética de la Universidad Nacional Autónoma de México, especialmente con los de integridad y honestidad académica.

De acuerdo con lo anterior, manifiesto que el trabajo escrito titulado USO DE PRODUCTOS SATELITALES Y METEOROLOGICOS PARA LA ESTIMACION DEL RENDIMIENTO DE LA CAÑA DE AZUCAR que presenté para obtener el título de INGENIERO GEOMÁTICO es original, de mi autoría y lo realicé con el rigor metodológico exigido por mi Entidad Académica, citando las fuentes de ideas, textos, imágenes, gráficos u otro tipo de obras empleadas para su desarrollo.

En consecuencia, acepto que la falta de cumplimiento de las disposiciones reglamentarias y normativas de la Universidad, en particular las ya referidas en el Código de Ética, llevará a la nulidad de los actos de carácter académico administrativo del proceso de titulación.

VIVALDO ISAI GARCIA PERALES
Número de cuenta: 418048616

ÍNDICE GENERAL

1. Introducción	12
2. Antecedentes	14
2.1. Caña de azúcar	14
2.1.1. Etapas fenológicas	14
2.1.2. Requerimientos hídricos	16
2.1.3. Requerimientos térmicos	16
2.2. Percepción remota	17
2.2.1. Bases físicas	17
2.2.1.1. Fuente de radiación	17
2.2.1.2. El espectro electromagnético	18
2.2.1.3. Ventanas atmosféricas	19
2.2.2. Sistema de adquisición	20
2.2.3. Sensores	20
2.2.4. Resolución del instrumento	21
2.2.4.1. Resolución espacial	21
2.2.4.2. Resolución espectral	21
2.2.4.3. Resolución temporal	22
2.2.4.4. Resolución radiométrica	22
2.2.5. Aplicaciones en la agricultura	22
2.2.6. Estudios previos sobre estimación de rendimiento	23
3. Objetivos	25
3.1. General	25
3.2. Particulares	25
4. Área de estudio	26
4.1. Ubicación	26
4.2. Climatología	26
4.3. Edafología	27
4.4. Descripción de las parcelas	28

5. Datos y metodología	30
5.1. Imágenes de satélite	30
5.1.1. Sentinel-2	30
5.1.1.1. Adquisición	35
5.1.2. Cálculo de índices espectrales	38
5.2. Variables agrometeorológicas	43
5.2.1. Dataset AgERA5	43
6. Resultados y discusión	47
6.1. Índices espectrales	47
6.2. Variables de peso en el rendimiento de la producción de la caña de azúcar	50
6.3. Curva de desarrollo del cultivo de caña en las parcelas analizadas	57
6.4. Modelos matemáticos de rendimiento	59
6.4.1. Regresión lineal simple	59
6.4.2. Regresión lineal múltiple	64
6.5. Mapas de rendimiento basados en índices espectrales	67
7. Conclusiones	71
A. Variables meteorológicas	72
B. Rendimientos estimados	74

ÍNDICE DE FIGURAS

2.1. Caña de azúcar	14
2.2. Etapas fenológicas de la caña de azúcar	16
2.3. Representación del espectro electromagnético	19
2.4. Plataformas para la observación de la Tierra	20
2.5. Comparación entre 3 diferentes resoluciones espaciales	21
2.6. Comparación entre 3 diferentes resoluciones radiométricas	22
4.1. Mapa general de la zona de estudio	26
4.2. Mapas temáticos de la climatología en la zona de estudio	27
4.3. Mapas edafológicos de la zona de estudio	27
4.4. Mapa de la ubicación de las parcelas de estudio	28
4.5. Ejemplo de los cortes realizados en fechas diferidas	29
5.1. Esquema del satélite Sentinel-2	30
5.2. Niveles de procesamiento de los productos de Sentinel-2	31
5.3. Mosaicos	32
5.4. Ubicación de las bandas espectrales en el EM	33
5.5. Bandas espectrales del MSI	34
5.6. Cobertura y resolución temporal de la misión Sentinel-2	34
5.7. Ubicación de las parcelas de estudio en el sistema MGRS	35
5.8. Diagrama de flujo del filtrado de las geometrías	36
5.9. Enmascaramiento de los píxeles correspondientes a nubes	36
5.10. Diagrama de flujo simplificado de la metodología propuesta para la adquisición de imágenes empleando la API de GEE para Python	37
5.11. Geometría de la imagen a exportar	37
5.12. Representación de los principales índices de vegetación	39
5.13. Opciones de rasterización en el método zonal statistics de rasterstats	40
5.14. Serie de tiempo del índice NDVI durante el periodo de estudio (parcela No.5)	40
5.15. Anomalías en las series de tiempo	41
5.16. Serie de tiempo corregida	41
5.17. Modelos de regresión polinómica	42
5.18. Métricas de validación por modelo	42
5.19. Series de tiempo de las principales variables agroclimáticas (parcela No. 5)	44
5.20. Serie de tiempo de la evapotranspiración de referencia y cultivo (parcela No. 5)	45

6.1. Series de tiempo real y modelada, de las 3 temporadas que comprende el estudio	49
6.2. Correlación entre las variables con mayor relación con el rendimiento por temporada	51
6.3. Correlación entre las variables enfocadas en la humedad y el rendimiento	52
6.4. Correlación entre la temperatura promedio en las etapas clave del desarrollo del cultivo	53
6.5. Correlación entre el valor promedio de humedad relativa durante la etapa de rápido crecimiento y el rendimiento reportado	54
6.6. Correlación entre la precipitación acumulada y el rendimiento	55
6.7. Días con lluvia por temporada	56
6.8. Rangos típicos del coeficiente cultivo para las diferentes etapas del crecimiento de la caña de azúcar	57
6.9. Curva de desarrollo del cultivo a partir del coeficiente de cultivo único	58
6.10. Modelos de regresión lineal simple para la estimación del rendimiento empleando el conjunto de datos de las 3 temporadas	60
6.11. Modelos de regresión lineal simple de las mejores variables considerando el conjunto de datos de las primeras 2 temporadas	63
6.12. Representación tridimensional de los modelos de regresión lineal múltiple	66
6.13. Mapas de rendimiento generados para la parcela 5 en la temporada 2022/2023 a partir del valor máximo del índice NDRE	68
6.14. Mapas de rendimiento generados para la parcela 5 en la temporada 2022/2023 a partir del valor máximo del índice NDVI	69
6.15. Mapas de rendimiento para la parcela 5 en la temporada 2022/2023 generados a partir de los modelos de regresión lineal múltiple empleando los valores máximos de los índices NDRE y NDRE1	69
6.16. Mapa de rendimiento para la parcela 5 en la temporada 2022/2023, generados a partir de los modelos de regresión lineal múltiple empleando los valores máximos de los índices NDRE, NDRE1 y NDVI	70

ÍNDICE DE TABLAS

2.1. Ventas atmosféricas importantes	19
4.1. Datos generales de las parcelas de estudio	28
4.2. Fechas de inicio y final de temporada para cada parcela.	29
5.1. Bandas espectrales del MSI	32
5.2. Especificaciones espectrales de los satélites gemelos	33
5.3. Métricas de validación por variable	43
6.1. Fechas y valores del POS por parcela para cada una de las temporadas empleando los datos reales	48
6.2. Fechas y valores del POS por parcela para cada una de las temporadas empleando los datos modelados	48
6.3. Duración de las temporadas por parcela	49
6.4. Correlación entre los valores de los índices espectrales y el rendimiento por temporada	50
6.5. Correlación entre las variables enfocadas en la humedad y el rendimiento por temporada	53
6.6. Temperatura ideal por etapa de crecimiento	53
6.7. Correlación entre la temperatura promedio en las etapas clave del desarrollo del cultivo	54
6.8. Correlación entre la humedad relativa promedio durante la etapa de rápido crecimiento	55
6.9. Correlación entre la precipitación acumulada y el rendimiento	55
6.10. Correlación entre los días de lluvia y el rendimiento	57
6.11. Valores del coeficiente de cultivo	59
6.12. Duración de las etapas de crecimiento del cultivo	59
6.13. Métricas de validación de los modelos generados considerando el conjunto de datos de las 3 temporadas	61
6.14. Métricas de validación de los modelos generados empleando los datos individuales por temporada	62
6.15. Métricas de validación por variable empleando el conjunto de datos de las primeras 2 temporadas	63
6.16. Variables más empleadas en los mejores modelos generados	64

6.17. Comparación de las métricas de validación de los modelos generados a partir del conjunto de datos reales que comprende las 2 primeras temporadas	65
6.18. Comparación de las métricas de validación de los modelos generados a partir del conjunto de datos modelados que comprende las 2 primeras temporadas	65
6.19. Promedio de las métricas de validación por temporada	66
A.1. Descripción de las variables en el dataset AgERA5	73
B.1. Rendimientos para la zafra 2021/2022, empleando datos reales	75
B.2. Rendimientos para la zafra 2021/2022, empleando datos modelados	76
B.3. Rendimientos para la zafra 2022/2023, empleando datos reales	77
B.4. Rendimientos para la zafra 2022/2023, empleando datos modelados	78
B.5. Rendimientos para la zafra 2023/2024, empleando datos reales	79
B.6. Rendimientos para la zafra 2023/2024, empleando datos modelados	79

DEDICATORIA

Dedico este trabajo especialmente a mi abuela Otilia, que aunque físicamente nos dejó hace poco, siempre me cuidó y rezó por mi bien y por que pudiera completar este objetivo. También a mi abuelo Isaías, que me desea lo mejor y puede acompañarme en este momento importante de mi vida.

A mis padres, Vivaldo y Juana María, también a mi hermano Daniel, este trabajo es la representación de su infinito apoyo y cariño, así como su esfuerzo para que pudiera conseguir este objetivo.

A toda mi familia, amigos y personas que me acompañaron durante este proceso.

AGRADECIMIENTOS

A Dios, que me permitió cumplir con este objetivo, así como tener la oportunidad de vivir toda esta experiencia universitaria.

A mis padres, que con su esfuerzo me dieron todas las facilidades para que pudiera cumplir mis metas, apoyarme en todo momento y motivarme en todo lo que me propuse.

A mi hermano, que desde siempre y sin ser consciente, me motivó a ser un mejor hermano, mejor persona y mejor en todo lo que me propuse.

A mis tías Lupita y Katy, que desde la distancia siempre me cuidaron y apoyaron, desde mi primer día en la universidad hasta hoy.

A mis amigos, Joshua, Moises, Oswaldo, Ricardo, Samuel y Ulises, que hicieron de mi vida universitaria una experiencia que recordaré siempre y que me ayudaron a crecer como persona, no pude tener mejor compañía en esta etapa tan importante de mi vida, gracias.

A la Maestra Tai, por guiarme y ayudarme a conseguir este objetivo así como impulsarme a desarrollar mis habilidades en distintas áreas de conocimiento, siempre buscando lo mejor para mi crecimiento universitario.

A la Universidad Nacional Autónoma de México y la Facultad de Ingeniería, que me brindaron todas las herramientas para mi desarrollo académico y personal.

“Todo lo puedo en Cristo que me fortalece..”

— Filipenses 4:13

Mi abuelita solía mirarme y decírmelo, hoy y siempre la recuerdo con esto.

CAPÍTULO 1

INTRODUCCIÓN

El azúcar se produce a escala industrial en más de 133 países en todo el mundo. En la mayoría de los años, más del 70 % de la producción mundial de azúcar se consume internamente, permitiendo el desarrollo de un favorable mercado de exportación (Aguilar-Rivera et al., 2018). Aproximadamente el 80 % de la producción de azúcar a nivel mundial proviene de la caña de azúcar (Aguilar-Rivera et al., 2012; Aguilar-Rivera et al., 2018; Akbarian et al., 2022). México es el noveno mayor productor de azúcar a nivel mundial, con una participación del 2.6 % (CONADESUCA, 2024a). La industria azucarera mexicana es la mayor industria agrícola después del maíz, representa un 0.5 % del PIB nacional y 11.5 % del PIB del sector primario (Aguilar-Rivera et al., 2012; Aguilar-Rivera et al., 2018).

De acuerdo al reporte final de producción de caña y azúcar correspondiente a la zafra 2023/2024, se industrializaron un total de 743,119 hectáreas a cargo de 49 Ingenios Azucareros, a lo largo de los 15 estados en los que es cultivada, con un total de 46,093,814 toneladas de caña producida y un rendimiento en campo de 62.03 toneladas por hectárea, dando como resultado una producción total de azúcar de 4,703,547 toneladas (CONADESUCA, 2024b).

Debido a la gran importancia de esta industria agrícola, la aplicación de la percepción remota en el cultivo de caña de azúcar se ha analizado en una gran cantidad de estudios que tienen como objetivo el manejo y monitoreo para mejorar la productividad del mismo (Som-ard et al., 2021). Diversos estudios han encontrado que emplear datos derivados de series de tiempo basadas en colecciones de imágenes satelitales en conjunto con análisis de regresión permiten generar modelos para la estimación del rendimiento, siendo este enfoque el más común, esto resulta muy adecuado para datos con alta resolución espacial, espectral y temporal, como lo pueden ser los derivados de Sentinel 2 y LANDSAT 8, en este enfoque la precisión depende de la calidad de los datos multitemporales y una imagen individual de una fecha aislada no es suficiente (Som-ard et al., 2021). Los modelos basados en las series temporales de diversos índices espectrales han demostrado un buen desempeño para estimar el rendimiento (Canata et al., 2021; Cruz et al., 2017; Dimov et al., 2022; Mulianga et al., 2013; Mutanga et al., 2013; Rahman & Robson, 2020).

Para estimar el rendimiento del cultivo de caña de azúcar se desarrolló una metodología que emplea datos derivados de imágenes satelitales de alta resolución en conjunto con datos agrometeorológicos y los rendimientos previos reportados por los Ingenios Azucareros, estos datos fueron procesados y analizados para identificar las variables que influyen en el rendimiento del cultivo y se generaron modelos matemáticos para estimar el rendimiento del mismo. La metodología propuesta tiene como ventaja la facilidad para ser replicada en diferentes zonas de

estudio así como la versatilidad para adaptarse a la rigurosidad establecida por quien la replique, como desventaja tiene que depende de la calidad de los datos de entrada, principalmente de las series de tiempo generadas y de los rendimientos reportados para cada parcela.

El presente trabajo se encuentra dividido en 6 secciones principales, la primera de ellas describe los antecedentes propios del cultivo de caña de azúcar, así como fundamentos básicos de la percepción remota y un resumen de estudios previos realizados sobre el tema en cuestión. La segunda presenta los objetivos generales y particulares. La tercera ubica la zona de estudio, las características de la región y de las parcelas de estudio. La cuarta describe los datos a utilizar y la metodología a seguir para adquirir y procesar esos datos. En la quinta se analizan las variables de peso que influyen en el rendimiento del cultivo y se generan modelos matemáticos a partir de estas variables para estimar el rendimiento, dando como resultado la elaboración de mapas de rendimiento. La última sección contiene las conclusiones de este trabajo, las observaciones encontradas y las sugerencias para trabajos futuros.

CAPÍTULO 2

ANTECEDENTES

2.1. Caña de azúcar

La caña de azúcar (*Saccharum officinarum* L) (figura 2.1) es un cultivo perenne plantado en regiones tropicales y subtropicales del mundo (Lin et al., 2009). Es una gramínea tropical, un pasto gigante emparentado con el sorgo y el maíz. Su tallo contiene un jugo abundante en sacarosa, que se extrae y cristaliza en los ingenios azucareros para producir el azúcar. La sacarosa es sintetizada por la caña gracias a la energía tomada del sol durante la fotosíntesis (Comisión Veracruzana de Comercialización Agropecuaria, 2010).

Originaria de Nueva Guinea, de donde se distribuyó a toda Asia. Los árabes la trasladaron a Siria, Palestina, Arabia y Egipto, y posteriormente se propagó por África. Colón la llevó a las islas del Caribe y de ahí pasó a América tropical. A México arribó durante la conquista, instalándose las primeras industrias azucareras en las regiones cálidas del país como parte de la colonización (Comisión Veracruzana de Comercialización Agropecuaria, 2010).



Figura 2.1: Caña de azúcar. Tomada de SADER (2018)

2.1.1. Etapas fenológicas

La caña de azúcar tiene esencialmente cuatro fases de crecimiento (figura 2.2): *establecimiento*: germinación y emergencia, sea en plantillas o en rebrote/retoños (socas y resocas) de los cuales crecerán nuevos tallos (macollamiento), *ahijamiento*, *crecimiento rápido* y *maduración y cosecha* (Aguilar Rivera, 2011).

Ochoa Neira et al. (2010) mencionan que en el caso de las plantillas, de la siembra a la cosecha, el desarrollo del cultivo puede abarcar desde 14 a 17 meses, esto depende de la variedad y de la influencia del clima, en las socas la duración se reduce a un periodo de entre 11 a 13 meses.

Aguilar Rivera (2011) proporciona una descripción de las fases de crecimiento mencionadas anteriormente:

- *Establecimiento* (30 a 50 días): La germinación se refiere a la iniciación del crecimiento a partir de las yemas presentes en los tallos plantados o en los que quedan en pie después de la cosecha del cultivo anterior. Durante esta fase es necesaria la disponibilidad adecuada de agua y el control de malezas. El déficit hídrico tiene un impacto negativo importante sobre el rendimiento de azúcar ya que propicia un sistema de raíces pequeñas y poco profundas (Barbieri & Villa Nova, 1993). La germinación es el proceso que da paso de los órganos primordios latentes en la yema al estado activo de crecimiento y desarrollo. Las temperaturas óptimas para la brotación oscilan entre los 24 °C a 37 °C con disponibilidad de buena humedad en el suelo (Ochoa Neira et al., 2010).
- *Amacollamiento o ahijamiento* (50 a 70 días): Durante esta etapa el crecimiento y rendimiento son muy sensibles a cualquier déficit de agua. El ahijamiento es el proceso fisiológico de ramificación subterránea múltiple, que se origina a partir de las articulaciones nodales compactas del tallo primario. Múltiples factores, tales como la variedad, la luz, la temperatura, el riego (humedad del suelo) y las prácticas de fertilización afectan al ahijamiento. La incidencia de una iluminación adecuada en la base de la planta de caña durante el período de ahijamiento es de vital importancia. Los hijuelos o retoños que se forman primero dan origen a tallos más gruesos y pesados. Una temperatura cercana a los 30 °C es considerada como la óptima para el ahijamiento (Fauconnier & Bassereau, 1975).
- *Rápido crecimiento* (180 a 220 días): Comprende desde el cierre del dosel hasta el inicio del periodo de madurez de los tallos. Se caracteriza por el aumento de biomasa y del número de tallos por área. La humedad es un factor fundamental para que el sistema radical se desarrolle y pueda absorber los nutrimentos necesarios. Cualquier déficit de agua comenzaría el proceso de maduración y detendría la acumulación de sacarosa antes de su etapa óptima.

En el transcurso de la primera etapa de esta fase ocurre la estabilización de los retoños. De todos los retoños formados sólo el 40 - 50 % sobrevive y llega a formar cañas triturables. Esta es la fase más importante del cultivo, en la que se determinan la formación y elongación real de la caña y su rendimiento.

El riego por goteo, la fertirrigación y la presencia de condiciones climáticas de temperatura y humedad elevadas, y alta radiación favorecen una mayor elongación de la caña. El estrés hídrico reduce la longitud entre nodos. Temperaturas sobre 30 °C, con humedad cercana al 80 %, son más adecuadas para un buen crecimiento.

- *Maduración* (60 a 140 días): Comienza aproximadamente de dos a tres meses previos a la cosecha para cultivos con ciclo de 12 meses, y de los 12 a los 16 meses de edad para los que completan el ciclo en 18 a 24 meses. Durante esta fase se requiere un bajo contenido de humedad del suelo, por lo que el riego se debe reducir y luego detenerse para llevar la caña a la madurez; así se detiene el crecimiento y se propicia la acumulación

de carbohidratos y la conversión de azúcares reductores (glucosa y fructosa) a sacarosa (Leite et al., 2009).

Condiciones de abundante luminosidad, cielos despejados, noches frescas y días calurosos, junto con climas secos, favorecen notablemente la maduración. Comprender estas etapas permite al productor entender mejor el desarrollo de la planta y facilita una gestión más eficiente del riego y los nutrientes.

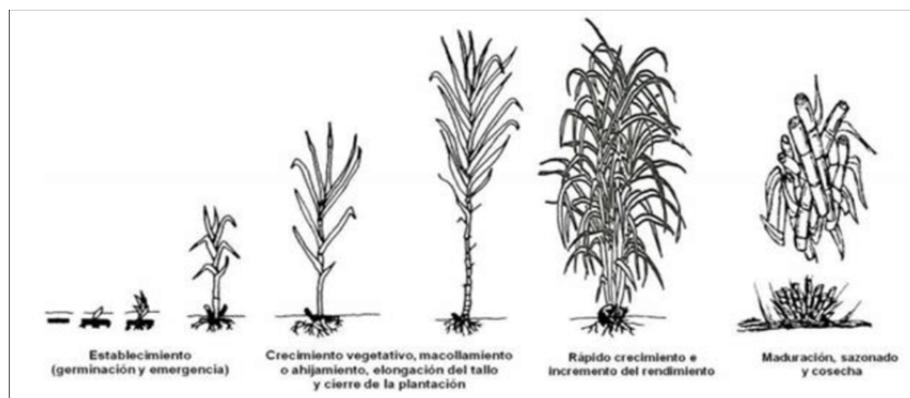


Figura 2.2: Etapas fenológicas de la caña de azúcar. Tomado de N. Aguilar et al. (2010).

El conocimiento de las fases fenológicas de la planta es de vital importancia para maximizar los rendimientos de caña y la recuperación del azúcar (Keating et al., 2003).

2.1.2. Requerimientos hídricos

Aguilar Rivera (2011) menciona que una precipitación acumulada entre 1500 y 1800 milímetros es adecuada en los meses de crecimiento vegetativo, siempre que la distribución de luz sea apropiada y abundante, similar a lo sugerido por Ochoa Neira et al. (2010) y CONADESUCA (2015).

En las condiciones ideales, el rendimiento incrementa en proporción directa con la cantidad de agua disponible y por cada 10 milímetros de agua utilizada se puede obtener alrededor de 1 tonelada por hectárea (Aguilar Rivera, 2011). Wright et al. (2017) encontraron que para producir 1 kilogramo de caña en cultivos en ciclo de plantilla se requirieron 89 litros de agua, en el caso de socas se necesitaron 118 litros de agua. La eficiencia de las plantillas se replica en la producción de azúcar, para producir 1 kilogramo se necesitaron 884 litros de agua, mientras que en las socas se necesitaron 1115 litros. Determinando que la eficiencia hídrica es menor en socas que en plantillas.

2.1.3. Requerimientos térmicos

Aguilar Rivera (2011) menciona que el crecimiento está directamente correlacionado con la temperatura. La temperatura óptima para la brotación de los esquejes se encuentra en el rango de 32 °C a 38 °C. Por debajo de 25 °C la germinación disminuye, llegando a su máximo entre los 30 °C a 34 °C y se detiene al alcanzar los 38 °C. Temperaturas superiores a 38 °C reducen la producción de materia seca por disminuir la tasa fotosintética e incrementar la respiración. Durante la maduración, las temperaturas relativamente bajas, en el rango de 12 °C a 14 °C, son preferibles ya que reducen la tasa de crecimiento vegetativo y el enriquecimiento de azúcar de la caña (Aguilar Rivera, 2011).

2.2. Percepción remota

Fagan y DeFries (2024) definen la percepción remota como “La práctica de obtener información sobre la superficie de la Tierra, utilizando imágenes adquiridas desde vehículos aerotransportados o espaciales mediante la medición de la radiación electromagnética reflejada, emitida o devuelta”. Esta definición cumple con las consideraciones impuestas por Gómez (2009), las cuales son:

- que el objeto de estudio esté sobre, o muy cercano, a la superficie de la Tierra y que el sensor se encuentre por encima del objeto.
- la información que llegue al sensor sea transportada por radiación electromagnética.

La introducción del término *radiación electromagnética* nos lleva a la descripción de los conceptos básicos de la percepción remota, las bases físicas sobre los que se sustenta y los elementos que componen el sistema.

2.2.1. Bases físicas

Gómez (2009) menciona que el estudio de las bases físicas de la percepción remota se puede representar en cuatro fases, las cuales son *fuentes de radiación, medio de transmisión, objetivo y sensor*.

A continuación se describen las fases mencionadas.

2.2.1.1. Fuente de radiación

La fuente de radiación puede ser tanto natural como artificial. En el caso de los estudios de prospección de recursos naturales, se emplea la radiación natural que proviene del Sol (Gómez, 2009).

Gomasasca (2009) señala que “El Sol y la Tierra son cuerpos físicos complejos. Los estudios sobre los fenómenos de radiación que los caracterizan son fundamentales para la aplicación de técnicas de percepción remota”(p. 130).

La emisión de energía o radiación solar en el espacio se puede describir utilizando el modelo del *cuerpo negro*. En términos simples, un cuerpo negro es un objeto que absorbe y emite radiación de manera perfecta. La distribución de la energía en función de la longitud de onda o frecuencia no es uniforme, sino que sigue un patrón que está relacionado con la temperatura. Este patrón funcional fue descubierto por el físico alemán M. Planck en 1900. La relación entre la temperatura y la radiación implica que a medida que la temperatura de un objeto incandescente aumenta, su color cambia desde el rojo hasta el azul-blanco, según lo establecido por la Ley de desplazamiento de Wien (Colwell, 1983). En consecuencia, se puede considerar que el Sol se comporta de manera aproximada como un cuerpo negro con una temperatura de alrededor de 6000 ° K. A esta temperatura, la radiación máxima se encuentra en una longitud de onda de aproximadamente 500 nanómetros, lo que corresponde al color verde-amarillo (Gómez, 2009).

La radiación o energía electromagnética se transmite de un lugar a otro siguiendo un patrón de onda armónico y continuo, a la velocidad de la luz y que contiene dos campos de fuerzas que son ortogonales entre sí: eléctrico y magnético. La longitud de onda (λ) y la frecuencia (ν), son los dos elementos que describen las características de este flujo de energía. La longitud de onda es la distancia entre dos picos sucesivos de una onda y la frecuencia se refiere a la velocidad de oscilación o el número de ciclos que pasan por un punto fijo en una unidad de tiempo

(Chuvienco S., 2008). Ambos elementos están relacionados por una relación de proporcionalidad inversa:

$$c = \lambda v \quad (2.1)$$

donde c : la velocidad de la luz en el vacío (3×10^8 m/s); λ : longitud de onda (expresada habitualmente en micrómetros, $1 \mu\text{m} = 10^{-6}$ m); v : frecuencia (Hertz, $1 \text{ Hz} = 1$ ciclo por segundo). Debido a la relación inversa entre ambos elementos, cuanto mayor sea la longitud de onda, menor será la frecuencia y viceversa, siendo posible caracterizar el flujo de energía indicando solo un término (Chuvienco S., 2008).

La ley general de la emisión electromagnética fue enunciada por Planck en Diciembre de 1900 en forma sintética, permitiendo calcular la cantidad de energía transportada por un fotón, si se conoce su frecuencia (Gomasasca, 2009):

$$e = hv \quad (2.2)$$

donde e : la energía radiante de un fotón (en Julios); h : la constante de Planck (6.6×10^{-34} J/s); v : la frecuencia. Sustituyendo en la ecuación (2.1) se tiene

$$e = h \left(\frac{c}{\lambda} \right) \quad (2.3)$$

, resultando que, a mayor longitud de onda, la cantidad de energía será menor y viceversa. Como consecuencia, la radiación en longitudes de onda largas será más difícil de detectar, a diferencia de aquellas con una longitud de onda corta (Chuvienco S., 2008).

2.2.1.2. El espectro electromagnético

A partir de las ecuaciones anteriores se pueden definir todas las formas de energía electromagnética en función de su longitud de onda o frecuencia. El arreglo de radiación de todas las longitudes de onda comprende desde las longitudes de onda más cortas (rayos Gamma) hasta las ondas de mayor tamaño (ondas de radio), se conoce como espectro electromagnético (Gómez, 2009) (figura 2.3).

Para el uso práctico, el espectro electromagnético está dividido en regiones, en función de la longitud de onda y las unidades de medida se relacionan con su respectiva longitud. Los micrómetros (μm) se utilizan para las longitudes más cortas y las más largas se miden en centímetros o metros. En el caso de las ondas de radio, se miden por valores de frecuencia (Gigahertz, $1 \text{ GHz} = 10^9$ Hz) (Chuvienco S., 2008; Gomasasca, 2009).

Chuvienco S. (2008) destaca una serie de regiones del espectro electromagnético (figura 2.3), las cuales son las más frecuentemente empleadas desde el punto de vista de la percepción remota:

- *Espectro visible* ($0.4 \mu\text{m}$ a $0.7 \mu\text{m}$): Recibe este nombre por ser la única región del espectro en la que nuestros ojos pueden percibir la radiación electromagnética. Dentro de esta región se distinguen las bandas: azul (**A**: $0.4 \mu\text{m}$ a $0.5 \mu\text{m}$), verde (**V**: $0.5 \mu\text{m}$ a $0.6 \mu\text{m}$) y rojo (**R**: $0.6 \mu\text{m}$ a $0.7 \mu\text{m}$).
- *Infrarrojo cercano* (IRC: $0.7 \mu\text{m}$ a $1.3 \mu\text{m}$): También denominado infrarrojo próximo, reflejado o fotográfico, parte de él se puede detectar a partir de películas dotadas de emulsiones especiales. Tiene una especial importancia debido a su capacidad para discriminar masas vegetales y concentraciones de humedad.

- *Infrarrojo medio (IRM: 1.3 μm a 8 μm):* En esta región del espectro se entremezclan los procesos de reflexión de la luz solar y de emisión de la superficie terrestre. Para el primer proceso, se habla del Infrarrojo de onda corta (Short Wave Infrared, SWIR), el cual está situado entre 1.3 μm y 2.5 μm , esta región es ideal para estimar el contenido de humedad en la vegetación y en los suelos. El segundo caso se denomina como infrarrojo medio (IRM), está comprendido entre los 3 μm y 5 μm , teniendo una gran utilidad para la detección de focos de alta temperatura.
- *Infrarrojo lejano o térmico (IRT: 8 μm a 14 μm):* corresponde a la porción emisiva del espectro terrestre, donde se detecta el calor que proviene de la mayor parte de las cubiertas terrestres.
- *Micro-ondas (M: >1 mm):* es un tipo de radiación transparente a la cubierta nubosa.

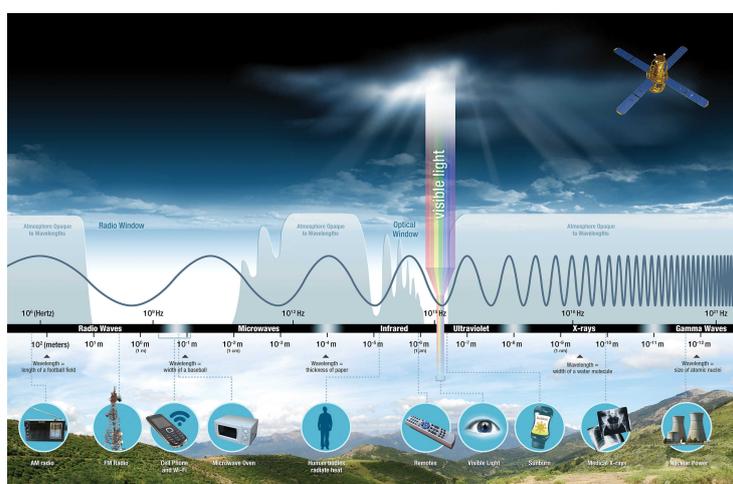


Figura 2.3: Espectro electromagnético. Tomada de National Aeronautics and Space Administration, Science Mission Directorate (2010)

2.2.1.3. Ventanas atmosféricas

La radiación electromagnética que proviene del Sol tarda aproximadamente ocho minutos en alcanzar la Tierra. En ese punto la radiación solar interactúa con la atmósfera de la Tierra. Los componentes de la atmósfera (nitrógeno, oxígeno, CO₂, vapor de agua, O₃, etc.) absorben, reflejan y dispersan la radiación, en forma selectiva, en función de la longitud de onda. En consecuencia, la transmisibilidad atmosférica posee características espectrales. Las regiones donde la radiación electromagnética pasa sin interactuar significativamente con los elementos de la atmósfera se conocen como “ventanas atmosféricas” (Gómez, 2009).

Visible e IR (λ en μm)	Radio (ν en GHz)
0.3 a 0.9	0.1 a 15
1 a 1.1	25 a 35
1.2 a 1.3	80 a 100
1.5 a 1.8	140 a 160
2 a 2.4	230 a 250
3.5 a 4	260 a 290
4.6 a 4.9 y 8 a 13	

Tabla 2.1: Ventanas atmosféricas importantes. Adaptado de Gómez (2009).

2.2.2. Sistema de adquisición

El sistema de adquisición consiste principalmente en el sensor y la plataforma en la que reside. La plataforma puede estar sobre la superficie, el aire, o el espacio (Liang et al., 2020).

Las plataformas espaciales son principalmente satélites y transbordadores espaciales (Liang et al., 2020).

La observación de la Tierra (EO) desde distancias orbitales es la principal característica de la percepción remota desde el espacio. El rango de altitudes desde los 200 km, para transbordadores espaciales, hasta los 35 800 km, para órbitas geoestacionarias de satélites meteorológicos (figura 2.4). En el caso de los empleados para observar y estudiar los recursos naturales, la altitud varía desde los 450 km a 900 km, en órbitas heliosíncronas, casi polares y circulares (Gomasca, 2009).

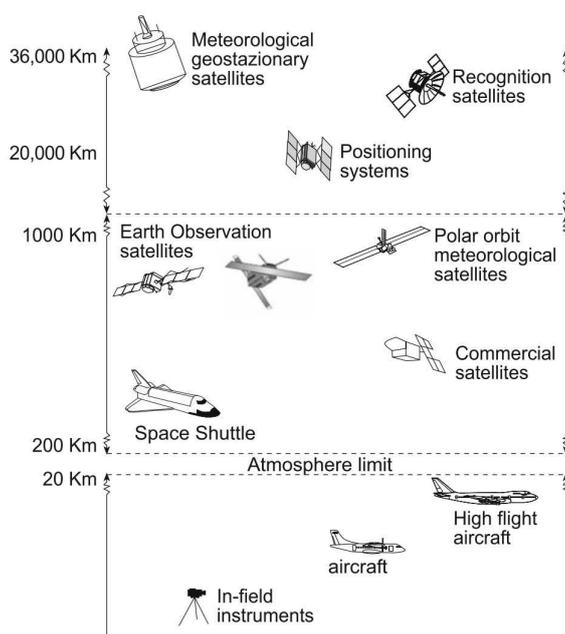


Figura 2.4: Plataformas para la observación de la Tierra y sistemas satelitales de posicionamiento global con referencia a la altitud orbital operacional. Tomada de Gomasca, 2009.

Liang et al. (2020) proporcionan información adicional acerca de los satélites con órbitas geoestacionarias y polares. Mencionan que un satélite geoestacionario está en una órbita que solo puede ser alcanzada a una altitud muy cercana a los 35 786 km y que mantiene el satélite fijo en una longitud del ecuador. Los satélites con órbita polar pueden proporcionar una plataforma de observación para la superficie completa de la Tierra, son capaces de rodear el globo aproximadamente una vez cada 100 mín. También mencionan que la mayoría de estos satélites tienen una altitud cercana a los 800 km, están en una órbita heliosíncrona y pasan directamente sobre una ubicación específica en la superficie a la misma hora en tiempo local.

2.2.3. Sensores

Existen dos tipos de sensores: pasivos y activos. Los sensores pasivos detectan la radiación natural, ya sea de forma mecánica o electrónica, a través de una gama de detectores. Un ejemplo de estos detectores es el espectroradiómetro, que es capaz de medir la radiancia en múltiples bandas espectrales. Los sensores activos proveen su propia fuente de radiación electromagnética

para iluminar la escena en cuestión, mandando un pulso de energía desde el sensor a la escena y recibiendo la radiación que es reflejada o dispersada de regreso de la escena, ejemplos de estos son el Radar, SAR, InSAR, LiDAR, etc (Liang et al., 2020).

2.2.4. Resolución del instrumento

Las especificaciones de la plataforma y el sensor determina la resolución de los datos. Cada instrumento (sensor) se caracteriza por tener diferentes resoluciones (Gomasasca, 2009; Liang et al., 2020).

Las resoluciones espaciales, espectrales, radiométricas y temporales se utilizan para comparar la resolución de datos analógicos y digitales en el campo de la percepción remota, en este contexto, la resolución es la capacidad de separar y distinguir objetos o elementos adyacentes en una escena (Shankar, 2017).

2.2.4.1. Resolución espacial

La resolución espacial es una medida de la nitidez o finura de los detalles espaciales en una imagen, comúnmente se expresa en términos de distancia y describe la capacidad de un sensor para identificar el detalle de menor tamaño en una escena, ya sean patrones u objetos (Shankar, 2017) (figura 2.5).

Zhang y Li (2023) se refieren a la resolución espacial como “la escala o tamaño de la unidad más pequeña de una imagen capaz de distinguir objetos, o es una medida de la distancia angular o lineal más pequeña para identificar objetos adyacentes en una imagen”.

En el contexto de la percepción remota, una imagen está compuesta por elementos básicos llamados píxeles. Cada pixel está caracterizado por dos coordenadas (x , y), que indican su posición en la imagen. El tamaño del pixel influye directamente en la resolución de la imagen. Esta resolución puede variar desde centímetros hasta cientos de kilómetros (Gomasasca, 2009).



Figura 2.5: Comparación entre tres diferentes resoluciones espaciales sobre la misma escena. De izquierda a derecha se puede apreciar como una mayor resolución espacial permite distinguir objetos presentes en la escena hasta el punto de solo poder distinguir la forma aproximada de la escena en cuestión. Tomada de National Aeronautics and Space Administration, Earth Observatory (s.f.).

2.2.4.2. Resolución espectral

Zhang y Li (2023) se refieren a la resolución espectral como “las características espectrales del sensor remoto y a los parámetros espectrales seleccionados por el sensor remoto, como el rango de detección espectral, el número de bandas espectrales y el ancho de banda espectral”.

La resolución espectral describe el número y ancho de las bandas espectrales en el sensor. En el caso de las imágenes digitales, corresponde al número y ubicación de las bandas espectrales (respecto al espectro electromagnético), su ancho, y el rango de sensibilidad de cada una de las bandas (Jensen, 2007; Liang et al., 2020).

2.2.4.3. Resolución temporal

Liang et al. (2020) define la resolución temporal como una medida del ciclo de repetición o la frecuencia en la que el sensor revisita la misma parte de la superficie de la Tierra.

Zhang y Li (2023) proporcionan una definición más simple, siendo la resolución temporal una medida de que tan seguido la misma área es visitada por el sensor.

Gomarasca (2009) la define como el período entre 2 adquisiciones sucesivas de la misma área.

2.2.4.4. Resolución radiométrica

Describe la capacidad del sensor para medir la intensidad de la señal de brillo de los objetos. Es una medida de la capacidad del sensor para distinguir entre dos objetos con una reflectancia similar (figura 2.6). Entre más sensible es un sensor a la reflectancia de un objeto comparado a los que están a su alrededor, menor será el objeto que puede ser detectado e identificado (Shankar, 2017).

En el caso de imágenes digitales, la resolución radiométrica se refiere al rango dinámico, o el número de diferentes números de salida en cada banda de información, y está determinada por el número de *bits* en los que la radiación almacenada es dividida (Liang et al., 2020).



Figura 2.6: Comparación de distintas resoluciones radiométricas sobre la misma escena. De izquierda a derecha, 16 bit, 8 bit y 4 bit. Entre mayor sea esta resolución se podrá distinguir de mejor forma objetos con una reflectancia similar, como en la imagen con resolución de 16 bit, la cual es más sensible a los valores del agua permitiendo distinguir la diferencia de reflectancias, por el contrario en la imagen de 4 bit, los valores de reflectancia en el agua se agrupan. Tomada de USGS (s.f.).

2.2.5. Aplicaciones en la agricultura

Weiss et al. (2020) mencionan que la Percepción Remota tiene la capacidad de ayudar a la evolución adaptativa de las prácticas agrícolas para seguir cumpliendo su papel en un contexto de sostenibilidad ambiental y cambio climático, para continuar abasteciendo a la humanidad de alimento, combustibles y materias primas que son fundamentales para el sustento humano.

Weiss et al. (2020) destacan la cualidad de la Percepción Remota de ser una forma no-destruccionista de monitorear la vegetación desde una perspectiva espacial y temporal, permitiendo

a los actores agrícolas (agricultores, cooperativas agrícolas, autoridades locales, nacionales o internacionales) cumplir con sus objetivos. Puede contribuir en la identificación de nuevas variedades que se ajusten mejor en contextos desafiantes (fenotipado), monitoreo de uso de suelo agrícola, la estimación de la producción del cultivo durante la temporada, la optimización de la producción a corto plazo y la prestación de servicios ecosistémicos relacionados con los recursos del suelo o el agua, la biodiversidad animal o vegetal.

2.2.6. Estudios previos sobre estimación de rendimiento

Bégué et al. (2010) mencionan que “la estimación del rendimiento de los cultivos ha sido estudiada por la comunidad de sensores remotos durante muchos años, especialmente para los cereales, que son cultivos principales”.

Para el cultivo de caña de azúcar, un gran número de estudios se han publicado dirigiéndose al manejo y monitoreo de esta para incrementar la productividad. En su revisión, Som-ard et al. (2021) resumieron las aplicaciones de la Percepción Remota desde 1981 hasta el 2020. 30 de esas aplicaciones fueron dirigidas a la estimación de rendimiento del cultivo. A continuación se describen algunas de esas aplicaciones:

Rahman y Robson (2020) emplearon series de tiempo del índice de vegetación GNDVI, generadas a partir de imágenes derivadas de S2 MSI y L8 OLI, en conjunto con un algoritmo de aprendizaje automático, para estimar el rendimiento de la caña de azúcar. Encontrando que el valor máximo del GNDVI está altamente correlacionado con el rendimiento reportado.

Picoli et al. (2014) combinaron datos agrometeorológicas y del sensor ALOS/PALSAR para estimar el rendimiento usando un modelo de regresión lineal múltiple. Encontraron que modelos predictivos de rendimiento basados en datos SAR pueden asistir a los agricultores e ingenios azucareros.

Mutanga et al. (2013) determinaron el momento óptimo para predecir el rendimiento de la caña de azúcar usando el índice NDVI derivado de imágenes de vegetación de SPOT. Encontrando que el mejor periodo para adquirir las imágenes es cerca de los 2 meses antes de la cosecha.

Mulianga et al. (2013) estudiaron la conveniencia del NDVI de MODIS para predecir el rendimiento de la caña de azúcar. Considerando un manejo de cultivo de 15 meses para un ciclo de soca. Las series temporales del NDVI se normalizaron mediante un método original de pesos que considera el periodo de crecimiento del cultivo y lo relaciona con el conjunto de datos históricos de rendimiento. Concluyendo que el uso de datos provenientes de satélites de baja resolución dificulta la predicción de rendimiento.

Adicional a los estudios analizados en la obra de Som-ard et al. (2021), se presentan los siguientes:

Canata et al. (2021) desarrollaron modelos predictivos de rendimiento integrando series de tiempo y técnicas de aprendizaje automático. Emplearon imágenes de una parcela comercial desde el principio del brote del retoño hasta la cosecha de 2 temporadas consecutivas. Determinando que el modelo de Bosque Aleatorio (RF) tuvo un mejor desempeño en comparación con el modelo de regresión lineal múltiple (MLR) para predecir el rendimiento de la caña de azúcar.

Dimov et al. (2022) compararon la utilización y el uso combinatorio de tres diferentes conjuntos de datos predictores para la estimación de rendimiento de caña de azúcar, empleando datos de los satélites de la misión Sentinel-2. Las variables empleadas fueron: 1) series de tiempo por parcela basadas en índices de vegetación, 2) descriptores de series de tiempo (marcadores y métricas fenológicas) y 3) variables espacio-temporales. Obteniendo como resultado que los marcadores fenológicos consiguen buenos resultados al estimar el rendimiento.

Akbarian et al. (2022) emplearon imágenes multiespectrales de alta resolución derivadas de vehículos aéreos no tripulados (UAV) para predecir el rendimiento de la caña de azúcar a nivel de surco. Estudiaron el desempeño de múltiples índices de vegetación en tres etapas principales de desarrollo del cultivo durante 3 temporadas de crecimiento. Encontrando que la etapa de crecimiento óptimo para los índices empleado fue de mediados de Marzo a principios de Mayo. Determinando que el mejor mes para realizar las predicciones fue Marzo, aproximadamente un mes antes de lo reportado anteriormente para la zona de estudio. Alcanzando el mejor desempeño en cuanto a la estimación al combinar los índices NDRE y GNDVI.

En el panorama nacional, Cruz et al. (2017) estimaron el rendimiento de la caña de azúcar en 23 parcelas de control en la zona de abasto del Ingenio Emiliano Zapata, en el estado de Morelos. Empleando los valores promedio del NDVI y MSI, así como la evapotranspiración del cultivo acumulada para desarrollar modelos de regresión lineal simples y así obtener el valor del rendimiento estimado. Obteniendo resultados favorables ($R^2 = 0.92$ para el NDVI) con una diferencia del 5.3 % respecto a lo reportado por el Ingenio.

CAPÍTULO 3

OBJETIVOS

3.1. General

Diseñar una metodología para estimar el rendimiento del cultivo de caña de azúcar en parcelas ubicadas en la región del sur de Tamaulipas, mediante datos derivados de productos satelitales de la misión espacial Sentinel 2, herramientas de scripting y datos meteorológicos derivados de modelos numéricos.

3.2. Particulares

- Adquirir los productos satelitales y datos derivados de modelos meteorológicos numéricos mediante scripts de Python.
- Identificar el momento de máximo desarrollo del cultivo para cada parcela.
- Determinar las variables de peso en el rendimiento de la producción de caña de azúcar.
- Calcular modelos matemáticos para la estimación del rendimiento.
- Generar mapas de rendimiento basados en los modelos matemáticos.

CAPÍTULO 4

ÁREA DE ESTUDIO

4.1. Ubicación

El presente trabajo se llevó a cabo en el sur del estado de Tamaulipas, perteneciente a la denominada región cañera Noreste, de acuerdo a la clasificación establecida por la CONADE-SUCA. Los municipios asociados a la producción de caña de azúcar en esta región son, Antiguo Morelos, El Mante, Gómez Farías, González, Llera, Nuevo Morelos, Ocampo y Xicoténcatl, siendo El Mante y Gómez Farías los municipios en donde se ubican las parcelas elegidas para este estudio (figura 4.1).



Figura 4.1: Localización de la zona de estudio.

4.2. Climatología

De acuerdo al INEGI (2023), la precipitación pluvial anual en el estado de Tamaulipas en los años 2021, 2022 y 2023 fue de 740.2 mm, 487.3 mm y 535.6 mm, respectivamente. En el

caso de los municipios de El Mante y Gómez Farías, la precipitación media anual se encuentra en el intervalo de 800 mm a 1200 mm y 1200 mm a 1500 mm, respectivamente (figura 4.2a), clasificando ambos en el régimen de precipitación subhúmedo (CONABIO, 1998).

El grupo de climas existentes en el estado son, templado, seco y cálido, con un intervalo en la temperatura media anual de 10 °C a 26 °C (INEGI, 2023). Los municipios de El Mante y Gómez Farías entran en la clasificación de clima cálido y semicálido, con una temperatura media anual que oscila entre 24 °C y 26 °C para EL Mante y 16 °C a 26 °C para Gómez Farías (Cuervo-Robayo et al., 2015).

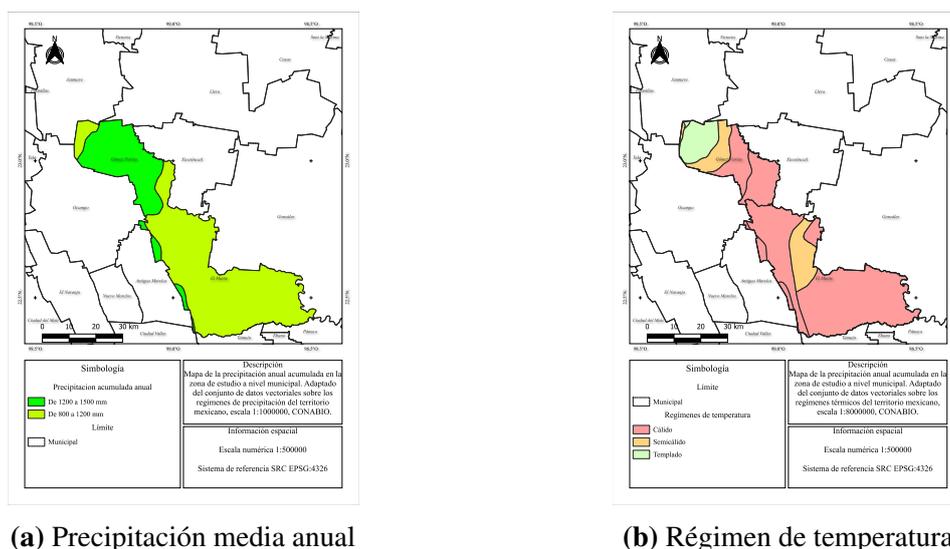


Figura 4.2: Climatología de la zona de estudio. Generados a partir de los datos de CONABIO (1998)

4.3. Edafología

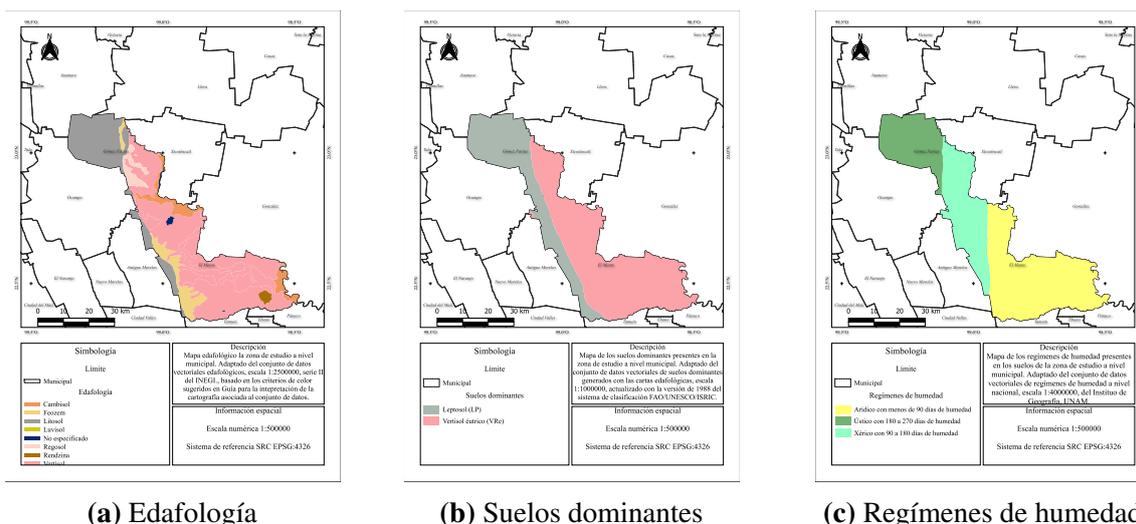


Figura 4.3: Mapas edafológicos de la zona de estudio

De acuerdo a lo reportado por CONABIO (2001), en la zona de estudio se encuentran diferentes tipos de suelo, desde Cambisol a Vertisol (figura 4.3a), siendo los dominantes el Leptosol y Vertisol (figura 4.3b), con 3 principales regímenes de humedad (figura 4.3c).

4.4. Descripción de las parcelas

Se cuenta con 15 parcelas de estudio, propiedad de productores particulares, distribuidas en los municipios mencionados, 7 se encuentran en el municipio de El Mante y 8 en Gómez Farías (figura 4.4), cuyas áreas van desde 2.48 ha hasta 35.83 ha. Se recopilaron los datos de interés de cada una de las parcelas correspondientes a las temporadas 2021/2022, 2022/2023 y 2023/2024 (tabla 4.1), comenzando por su ubicación y respectiva geometría, área, variedad, tipo de riego y rendimiento ($t\ ha^{-1}$) reportado por el Ingenio Azucarero por cada temporada.

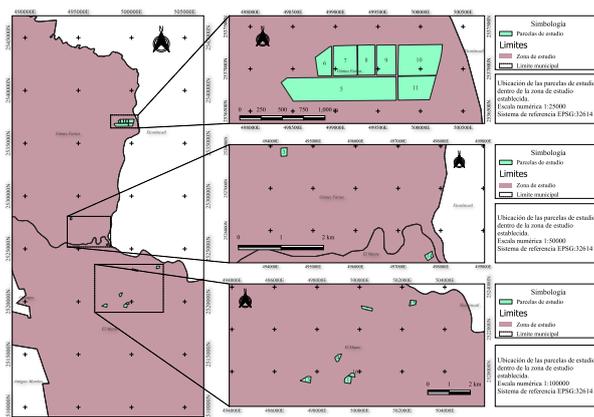


Figura 4.4: Ubicación de las parcelas de estudio sobre la zona de estudio, enumeradas de acuerdo a su identificador.

Parcela	Área (ha)	Variedad	Riego	Rendimiento		
				21/22	22/23	23/24
1	9.206	CP:722086	Gravedad	70	60	40
2	6.247	CP:722086	Gravedad	75	65	50
3	2.783	CP:722086	Gravedad	45	50	65
4	6.387	CP:722086	Bombeo	50	65	65
5	35.833	CP:722086	Gravedad	80	75	65
6	5.207	CP:722086	Gravedad	80	75	65
7	9.976	CP:722086	Gravedad	80	75	65
8	7.227	CP:722086	Gravedad	80	75	65
9	8.159	CP:722086	Gravedad	80	75	65
10	17.871	CP:722086	Gravedad	75	75	65
11	11.500	CP:722086	Gravedad	75	75	40
13	9.185	CP:722086	Gravedad	65	50	40
14	2.480	CP:722086	Gravedad	75	65	45
15	6.551	CP:722086	Gravedad	75	50	45
16	3.748	CP:722086	Gravedad	40	30	25

Tabla 4.1: Datos generales de la parcela de estudio y rendimientos de las últimas 3 temporadas.

Los datos presentados en tabla 4.1 son los proporcionados por los productores, en el caso del rendimiento, es el que les reportó el Ingenio Azucarero. Las parcelas mostradas en la parte superior derecha de la figura 4.4 pertenecen al mismo productor, por lo que el Ingenio les asignó un valor de rendimiento por *lote*, aún cuando no todas fueron cosechadas en la misma fecha, por lo que su análisis se realizó individualmente asignándoles el rendimiento disponible, por esa razón las parcelas con identificador del 5 al 11 tienen el mismo rendimiento. Las fechas de corte

proporcionadas por los productores se verificaron con la revisión de una imagen satelital de la parcela en la fecha dada, empleando la herramienta *Planet Explorer* (Planet Labs PBC, 2018), propiedad de Planet Labs, encontrando diferencias respecto a la información proporcionada por los productores, por lo que las fechas de corte se establecieron de acuerdo al momento inmediato anterior en la que no hay presencia de vegetación en la parcela. En algunos casos especiales los cortes se realizaron de forma diferida, en días distintos y la fecha de algunas de esos días coincidió con la adquisición de la imagen por parte del sensor, dando como resultado imágenes en las que la parcela se encuentra parcialmente cubierta por vegetación y la otra parte de suelo desnudo.

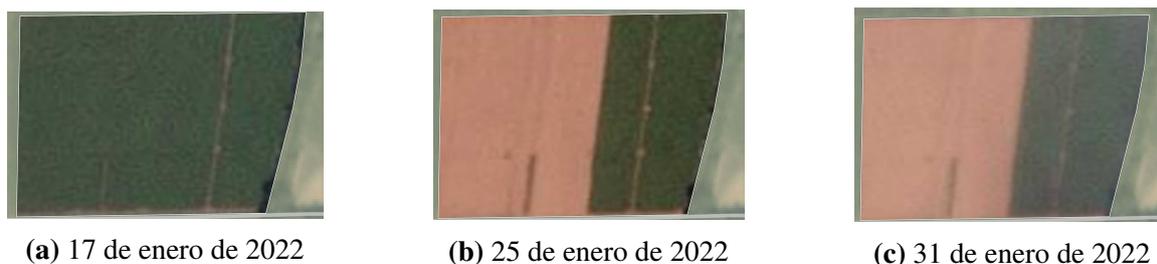


Figura 4.5: Cortes en fechas diferidas. **a)** la imagen fue capturada cuando la parcela estaba totalmente cubierta por el cultivo, **b)** la captura se realiza durante un corte parcial, **c)** el corte parcial se extiende a la siguiente captura por parte del sensor.

En la figura 4.5 se ejemplifica el caso mencionado anteriormente, en el que se puede apreciar que el 17 de enero de 2022, la superficie de la parcela estaba completamente cubierta por vegetación (el cultivo), el 25 de enero de 2022 el sensor capturó el comienzo de la cosecha pero esta no fue terminada el mismo día, prolongándose por varios días y el 31 de enero de 2022 coincidió con la adquisición de la escena. En este caso se procede a considerar como fecha de corte cuando se comienza el corte, al abarcar 2 temporadas consecutivas esto afecta a la siguiente, por lo que se omitirán estos días en los que la parcela tiene partes con vegetación y suelo desnudo, comenzando a considerar a partir de que la superficie total no presente vegetación. Después de estas aclaraciones se presentan las fechas de corte para las temporadas 2021/2022, 2022/2023 y 2023/2024, así como su delimitación temporal, inicio y final (tabla 4.2).

Parcela	2021/2022		2022/2023		2023/2024	
	Inicio	Final	Inicio	Final	Inicio	Final
1	2020-11-22	2022-03-10	2022-03-10	2023-04-29	2023-04-29	2024-04-15
2	2020-11-16	2022-04-23	2022-04-23	2023-03-06	2023-03-06	2024-03-26
3	2020-09-27	2022-03-02	2022-03-02	2023-02-16	2023-02-16	2024-04-05
4	2020-04-13	2022-03-10	2022-03-10	2023-02-07	2023-02-07	2024-04-11
5	2021-03-23	2022-03-21	2022-03-21	2023-05-01	2023-05-01	2024-04-17
6	2021-05-02	2022-05-16	2022-05-16	2023-05-02	2023-05-02	2024-07-29
7	2020-12-15	2022-01-25	2022-01-25	2023-02-23	2023-02-23	2024-04-11
8	2021-03-19	2022-05-17	2022-05-17	2023-05-02	2023-05-02	2024-04-11
9	2021-04-05	2022-05-17	2022-05-17	2023-05-02	2023-05-02	2024-07-29
10	2020-12-12	2022-01-25	2022-01-25	2023-02-23	2023-02-23	2024-03-25
11	2020-11-23	2022-01-25	2022-01-25	2023-02-23	2023-02-23	2024-03-25
13	2020-12-22	2022-03-28	2022-03-28	2023-02-21	2023-02-21	2024-04-23
14	2020-11-14	2022-02-18	2022-02-18	2023-03-16	2023-03-16	2024-07-29
15	2021-01-13	2022-04-25	2022-04-25	2023-03-11	2023-03-11	2024-04-15
16	2021-02-05	2022-05-01	2022-05-01	2023-03-30	2023-03-30	2024-04-23

Tabla 4.2: Fechas de inicio y final de temporada

CAPÍTULO 5

DATOS Y METODOLOGÍA

En esta sección se describen los datos empleados para la realización de este estudio, así como la metodología propuesta para cumplir con los objetivos descritos anteriormente y se introducen los datos resultantes.

5.1. Imágenes de satélite

Una imagen satelital es el producto generado por un sensor instalado a bordo de un satélite artificial, el cual capta la radiación electromagnética emitida o reflejada por un cuerpo celeste; producto que posteriormente se transmite a estaciones terrestres para su visualización, procesamiento y análisis (C. G. Aguilar, 2014).

5.1.1. Sentinel-2

La misión Sentinel-2 ofrece una combinación sin precedentes de una sistemática cobertura global de las superficies terrestres, a una alta frecuencia de revisita, alta resolución espacial y un amplio campo de visión (Jutz & Milagro-Pérez, 2018). Es una misión de imágenes multispectrales de alta resolución desarrollada por la ESA, dentro del programa Copernicus. Consta de una constelación de 2 satélites gemelos, Sentinel-2A y Sentinel-2B (figura 5.1), que operan simultáneamente, desfasados 180° entre sí, en una órbita heliosíncrona a una altitud media de 786 km. La posición de cada satélite en su órbita es medida por un receptor GNSS de frecuencia doble.



Figura 5.1: Vista esquemática del satélite Sentinel-2 desplegado. Tomada de European Spatial Agency (s.f.-b).

los productos contienen tiles/mosaicos de una sola toma de datos. Los productos son partes elementales de tamaño fijo, dentro de una única orbita. Un mosaico es la partición mínima indivisible de un producto y contiene todas las posibles bandas espectrales. Para los mosaicos de nivel-1C y nivel-2A, son ortoimágenes de $110 \text{ km} \times 110 \text{ km}$ en proyección UTM/WGS84 (figura 5.3). La Tierra se subdivide en un conjunto predefinido de mosaicos, definidos en proyección UTM/WGS84 y utilizando un paso de 100 km. Sin embargo, cada mosaico tiene una superficie de 110 km^2 para proporcionar un gran solapamiento con los vecinos. El sistema UTM (Universal Transverse Mercator) divide la superficie terrestre en 60 zonas. Cada zona UTM tiene un ancho vertical de 6° de longitud y un ancho horizontal de 8° de latitud (European Spatial Agency, s.f.-b).

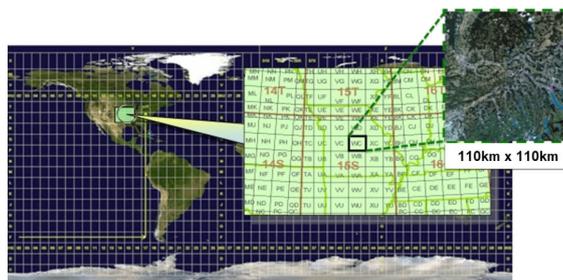


Figura 5.3: Ubicación de los mosaicos en el sistema UTM. Los mosaicos pueden estar parcial o completamente cubiertos por una imagen. Los mosaicos cubiertos parcialmente corresponden a aquellos que están en las orillas o en la parte superior/inferior de la tira de datos. Tomada de European Spatial Agency (s.f.-a)

Los productos tienen características heredadas del MSI, las cuales son las resoluciones, temporal, espacial, espectral y radiométrica. Las 13 bandas abarcan desde la porción visible (VIS) e infrarrojo cercano (NIR) hasta el infrarrojo de onda corta (SWIR) a diferentes resoluciones espaciales, de 10 m a 60 m. Las cuatro bandas con resolución de 10 m aseguran la compatibilidad con SPOT-4 y SPOT-5. Las 6 bandas a 20 m son requisitos para otros parámetros de los datos de nivel 2. Las bandas a 60 m son principalmente dedicadas para las correcciones atmosféricas (Jutz & Milagro-Pérez, 2018). El generador de imágenes multiespectrales es la primer misión de observación terrestre en su tipo al incluir tres bandas en la parte “red edge” del espectro electromagnético (a una resolución espacial de 20 m), la cual proporciona información clave del estado de la vegetación (Jutz & Milagro-Pérez, 2018).

Banda	Longitud de onda central	Resolución	Descripción
1	~ 443	60	Aerosoles
2	~ 493	10	Blue
3	~ 560	10	Green
4	~ 665	10	Red
5	~ 704	20	Red Edge 1
6	~ 740	20	Red Edge 2
7	~ 783	20	Red Edge 3
8	~ 883	10	NIR
8A	~ 865	20	Red Edge 4
9	~ 945	60	Water Vapor
10	~ 1375	60	Cirrus
11	~ 1610	20	SWIR 1
12	~ 2190	20	SWIR 2

Tabla 5.1: Bandas espectrales del instrumento MSI a bordo en los satélites gemelos 2A y 2B

Las longitudes de onda central presentadas en la tabla 5.1 corresponden al aproximado por banda, ya que para cada satélite este valor difiere en función de ciertos parámetros. La determinación de la longitud de onda central se define por

$$\lambda_c = \frac{\int \lambda \times S(\lambda) .d\lambda}{\int S(\lambda) .d\lambda}$$

, donde λ_c es la longitud de onda central de una banda dada, λ la longitud de onda y S es la función de respuesta espectral del instrumento.

Con base a lo anterior, se presentan las longitudes de onda central y ancho de banda por banda espectral, diferenciando entre los satélites 2A y 2B.

Banda	S2A		S2B	
	Lon. de onda central	Ancho de banda	Lon. de onda central	Ancho de banda
1	442.7	20	442.3	20
2	492.7	65	492.3	65
3	559.8	35	558.9	35
4	664.6	30	664.9	31
5	704.1	14	703.8	15
6	740.5	14	739.1	13
7	782.8	19	779.7	19
8	832.8	105	832.9	104
8A	864.7	21	864.0	21
9	945.1	19	943.2	20
10	1373.5	29	1376.9	29
11	1613.7	90	1610.4	94
12	2202.4	174	2185.7	184

Tabla 5.2: Longitudes de onda central y anchos de banda por cada banda generada por los satélites 2A y 2B

Con los datos presentados en las tablas 5.1 y 5.2 se puede ubicar cada una de las bandas dentro del espectro electromagnético y visualizar su ancho de banda, tal como se presenta en la figura 5.4.

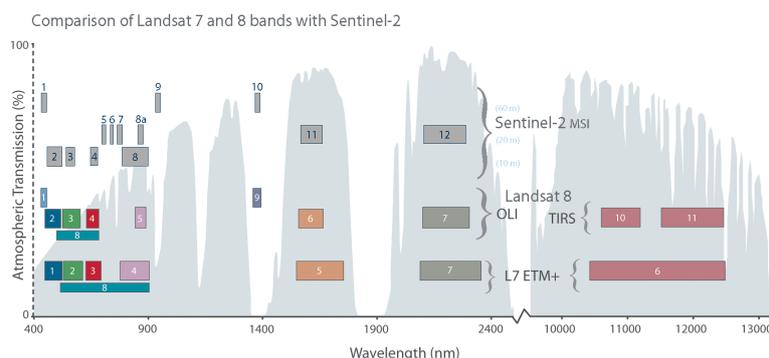


Figura 5.4: Ubicación de las bandas espectrales en el espectro electromagnético en comparación con las misiones Landsat 8 y 7. Tomada de Earth Resources Observation and Science (EROS) Center (2015)

Si se quiere visualizar la ubicación de cada banda de forma mas precisa, se presenta la figura 5.5, en la que se presenta un arreglo de las ubicaciones de las bandas divididas por su resolución espacial.

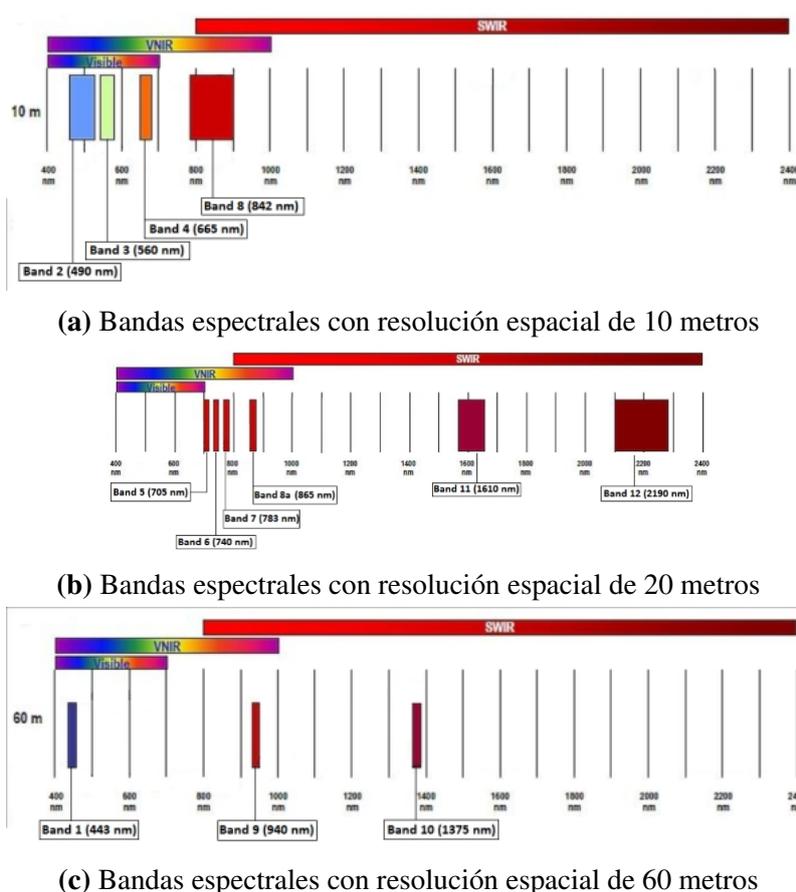
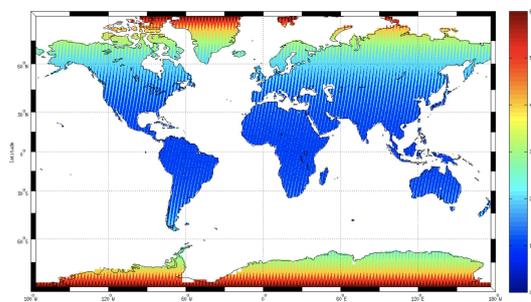
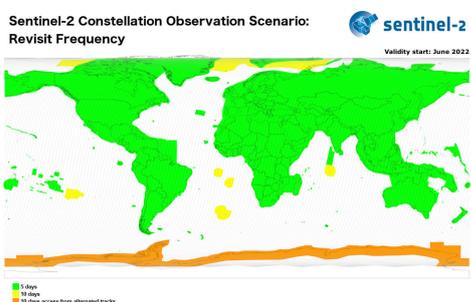


Figura 5.5: Bandas espectrales generadas por el instrumento MSI ubicadas en el espectro electromagnético. **a)** corresponde a las bandas con una resolución de 10 metros y ubicadas en la porción visible y de infrarrojo cercano del EM, **b)** bandas espectrales con resolución de 20 metros, asociadas a la parte ‘red edge’ e infrarrojo de onda corta, **c)** bandas con resolución de 60 metros. Fuente European Spatial Agency (s.f.-b).

La resolución radiométrica del MSI es de 12 bit, lo que otorga una alta sensibilidad al instrumento para distinguir diferentes niveles de brillo o tono, desde 0 a 4095. En el caso de la resolución temporal, considerando la constelación completa (2 satélites), el tiempo de revisita en condiciones libres de nubes es de 5 días (figura 5.6).



(a) Cobertura y tiempo previsto de revisita para las adquisiciones de la misión Sentinel-2.

(b) Frecuencia de revisita geométrica debido a la superposición entre orbitas adyacentes.

Figura 5.6: Representación de la resolución temporal en función de la cobertura. Debido a la superposición o traslape entre las franjas generado por las orbitas adyacentes, la frecuencia de revisita incrementará con diferentes condiciones de visibilidad. Fuente European Spatial Agency (s.f.-b).

5.1.1.1. Adquisición

Se presenta una metodología o “framework” para adquirir las imágenes satelitales deseadas, priorizando la individualidad de las mismas, en la que cada imagen corresponde a una única parcela y permite ajustar las dimensiones en unidades de píxeles a la geometría de la parcela en cuestión, dando como resultado una imagen de tipo raster compuesta solamente de los píxeles necesarios para representar la escena de interés. Para la adquisición de las imágenes de satélite se hizo uso de la plataforma de *Google Earth Engine* (Gorelick et al., 2017), aprovechando la API desarrollada para el lenguaje de programación *Python*. Esta API permite acceder a las funciones, métodos y clases disponibles para manipular y procesar los conjuntos de datos que ofrece la plataforma, a través de la librería destinada para el cliente (usuario), la cual tiene el nombre de *ee*. Villegas y Suárez (2022) describen de forma muy detallada las funcionalidades de esta API, enfocados en el lenguaje de programación *JavaScript* pero con la posibilidad de adaptarlo al lenguaje *Python*. Dentro de la plataforma *Google Earth Engine* las misiones satelitales reciben el nombre de *datasets*, dentro de estos se encuentran los productos correspondientes como una colección de imágenes a las que se puede acceder a través de la librería *ee*.

Se desarrolló un Script en *Python* para identificar, delimitar y adquirir las imágenes deseadas, requiere datos de entrada proporcionados por el usuario para su correcto funcionamiento y posible adaptación a diferentes necesidades, las entradas necesarias son la colección de imágenes a estudiar, la geometría o geometrías de las parcelas de estudio, el intervalo temporal que abarca el análisis (fecha de inicio y final) y el código de identificación único de la zona asociada a la imagen de acuerdo con el sistema de referencia de la red militar, por su siglas en inglés (MGRS), el cual es un derivado del sistema UTM, el cual permite ubicar y referirse de forma precisa a ubicaciones específicas en la Tierra. Adicionalmente es necesario corroborar una serie de condiciones previo a seleccionar y exportar las imágenes, la primera condición es asegurar que las geometrías de las parcelas y las imágenes intersecten, ya que como se mencionó anteriormente, las imágenes pueden cubrir parcial o completamente los mosaicos, por lo que si la geometría de la parcela se ubica en las zonas de traslape entre los mosaicos es probable que la geometría se encuentre en la línea divisoria entre mosaicos con *tile id's* (MGRS) distintos, provocando que 2 imágenes se consideren para una única geometría, violando el principio de unicidad establecido en la introducción de este marco de trabajo. Este proceso se realiza ya que al momento de generar una colección de imágenes se filtran los elementos mediante una de sus propiedades, el *MGRS_TILE*. Para este caso en particular, las parcelas se encuentran en estos traslapes por lo que se pueden encontrar imágenes con un *MGRS_TILE* diferente y esto resultaría en un problema de duplicación en la colección de imágenes.

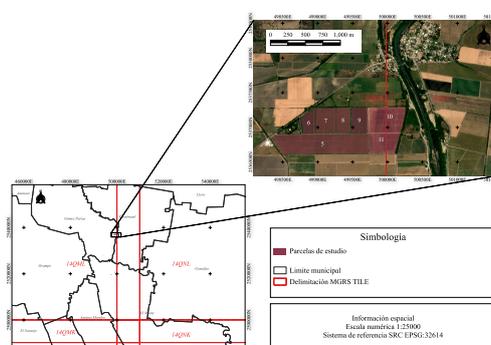


Figura 5.7: Ubicación de las parcelas sobre las cuadrículas que representan la grilla MGRS TILE. Encontrando que algunas parcelas se ubican en la zona de traslape entre los mosaicos y por lo tanto se pueden ubicar con 2 MGRS TILE distintos.

En la figura 5.7 se puede apreciar como algunas parcelas se pueden encontrar tanto en el mosaico con el identificador 14QML y 14QNL, pero otras solo se encuentran en el 14QML, por lo que este mosaico se selecciona para filtrar la colección de imágenes. Una vez que se tiene la colección imágenes con ese identificador se procede a verificar que la geometría de cada parcela interseca con la de cada imagen, para esto se desarrolló una función que verifique la intersección con un margen de error igual a 0.001 y devuelva un valor booleano dependiendo del resultado, si la condición se cumple, la imagen se considera apta para pasar al segundo filtro (figura 5.8).

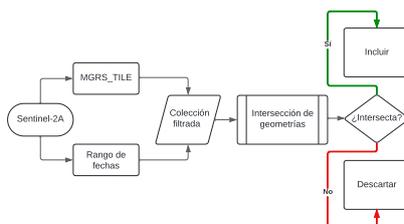


Figura 5.8: Diagrama de flujo del proceso de filtrado de la colección de imágenes y la verificación de la intersección entre geometrías

La segunda condición es minimizar/eliminar la presencia de nubes en las escenas, esta cuestión se puede afrontar desde diferentes enfoques, dependiendo el autor y los objetivos, Rahman y Robson (2016) consideraron como libre de nubes a las imágenes con un porcentaje menor al 40 % de píxeles asociados a nubes, Rahman y Robson (2020) emplearon imágenes con un porcentaje de nubosidad menor al 20 %, Canata et al. (2021) seleccionaron imágenes con una presencia de nubes menor al 1 %. Estos enfoques se basan en la adquisición y empleo de imágenes con el menor porcentaje de nubosidad posible dependiendo del propósito y disponibilidad, pero con la probabilidad de que en la zona de estudio exista la influencia de nubes aún cuando el porcentaje de nubosidad en la imagen sea muy bajo. En este trabajo se presenta un enfoque distinto respecto a la selección y adquisición de imágenes, que puede modificarse de acuerdo a la rigurosidad de quien lo replique. Similar a como se afrontó el problema de la intersección de las geometrías, en este enfoque se emplea la geometría de la parcela para identificar la presencia de píxeles considerados como nubes dentro de la escena. Para llevar a cabo esto es necesario emplear alguna de las bandas dedicadas a la identificación de nubes en la imagen, en este caso se seleccionó la banda MSK_CLDPRB, esta banda representa un mapa con la probabilidad que tiene cada píxel de ser o no una nube, dicha probabilidad va desde el 0 % hasta 100 %, con una resolución espacial de 20 metros. Se establece una condición a cumplir para considerar un píxel como nube (tener un valor mayor al 50 %, permitiendo identificar (figura 5.9) y cuantificar los píxeles considerados como nubes dentro de la geometría de las parcelas de estudio).



(a) Representación en color natural de la escena



(b) Píxeles enmascarados

Figura 5.9: Enmascaramiento de los píxeles asociados a nubes. **a)** se visualiza una representación a color natural lo cual permite distinguir claramente la presencias de nubes en la escena, **b)** presenta los píxeles que cumplieron la condición establecida (color rojo) y permite identificar de forma cuantitativa la presencia de nubes en la escena.

Resultado del enmascaramiento, se generan píxeles con valores de 0 o 1, siendo 0 un píxel que no cumplió con la condición y 1 el caso en el que el píxel si cumplió la condición establecida. Si dentro de la geometría no se encuentra ningún píxel con valor igual a 1 se asume que la escena está completamente libre de nubes, en el caso de que dentro de la región comprendida por la geometría de la parcela se encuentren píxeles con ambos valores la imagen en cuestión se descarta por la presencia de nubes en la escena. Al implementar este enfoque junto a la verificación de geometrías se genera el flujo de trabajo completo (figura 5.10).

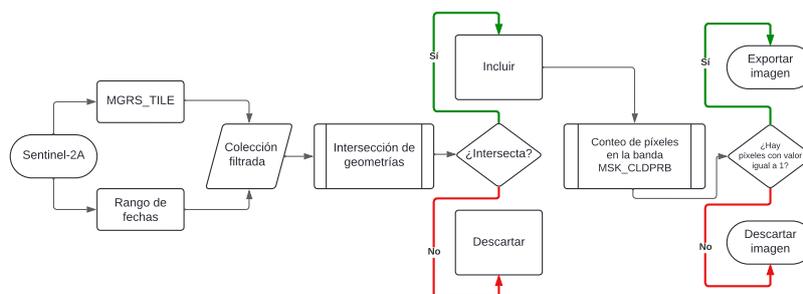
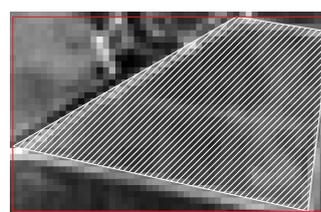


Figura 5.10: Diagrama de flujo simplificado de la metodología propuesta para la adquisición de imágenes empleando la API de GEE para Python

Una vez que se tienen identificadas y seleccionadas las imágenes aptas para exportar, se especifican las bandas, para este caso se exportan las 13 bandas que incluye el sensor MSI, en el caso de las bandas con una resolución diferente a 10 metros se remuestran mediante un método bilineal (esto lo realiza internamente Earth Engine al momento de exportar una imagen) y la extensión o geometría de la imagen resultante es la geometría generada a partir del método *bounds* provisto por GEE, el cual genera el recuadro mínimo que delimita la totalidad de la geometría de entrada (figura 5.11).



(a) Recuadro delimitador generado con el método *bounds* a partir de la geometría de la parcela



(b) Imagen resultante de recortar la imagen completa con el recuadro generado

Figura 5.11: Geometría de la imagen exportada y del objeto empleado para el recorte. **a)** sobre una imagen base en color natural se sobrepone la geometría de la parcela (color blanco) y la geometría del recuadro mínimo delimitador (color rojo), la cual se emplea para recortar la imagen raster, **b)** representa la imagen raster recortada a la extensión mínima que delimita la geometría y así la imagen se compone únicamente de los píxeles necesarios para contener la geometría de la parcela.

En la figura 5.11b se visualiza como las dimensiones de la imagen se recortaron a la geometría de la parcela, con un margen de ± 1 píxel. Siguiendo la metodología propuesta y representada en la figura 5.10, se inicia una colección de imágenes filtrada al rango de fechas establecido, que en este caso es desde 2020-01-01 al 2024-08-01 y con el MGRS_TILE 14QML, teniendo un total de 692 imágenes que cumplen con estos filtros, el siguiente paso es comprobar para cada parcela que efectivamente su geometría se encuentra dentro de la geometría de la imagen en turno,

dando un total de 9030 imágenes por verificar, de las cuales el 48.35 % cumplen con este criterio y son válidas para revisar la presencia de nubes en la escena, de las 4366 imágenes, el 62.35 % se consideran como libre de nubes y se exportan, el restante 37.65 % presenta nubosidad en la escena, por lo que se descartan. Una vez exportadas las imágenes se agrupan por identificador de parcela para tener un conjunto de datos ordenado y listo para su posterior manipulación y análisis.

5.1.2. Cálculo de índices espectrales

A partir de las imágenes adquiridas en la sección anterior se calcularon diferentes índices espectrales que sirven como datos de entrada para análisis realizados en sección posteriores. Los índices espectrales se calculan típicamente como la razón entre dos bandas espectrales para contrastar una característica absorbente con una característica de referencia no absorbente, y existen muchos ratios de banda estrecha y banda ancha que cuantifican los procesos de absorción del dosel asociados con pigmentos, agua y compuestos de ligno-celulosa de la materia orgánica y material leñoso (Ustin et al., 2004). Los índices espectrales a utilizar en este estudio son los denominados como Índices de Vegetación (VI's) y los Índices de agua en la vegetación (WI's), los VI's son medidas espectrales del verdor del dosel, expresadas mediante la cantidad y contenido de clorofila del follaje del dosel, mientras que los WI's miden la cantidad de follaje y el contenido de humedad del dosel (Huete, 2012).

Estudios previos han considerado emplear diferentes tipos de VI's y WI's para estimar el rendimiento de la caña de azúcar, tales como el *NDVI* (Cruz et al., 2017; Dengia et al., 2023; Dimov et al., 2022; Wang et al., 2019), propuesto por Tucker (1979), con un rango de -1 a 1 y adaptado a las bandas del instrumento MSI como,

$$NDVI = \frac{SR - 1}{SR + 1} = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red}} = \frac{b8 - b4}{b8 + b4} \quad (5.1)$$

GNDVI (Akbarian, Jamnani et al., 2023; Rahman & Robson, 2020; Rahman & Robson, 2016), propuesto por A. Gitelson et al. (1996) como una adaptación del *NDVI* al emplear la banda correspondiente a la porción del verde en lugar del rojo en el espectro electromagnético, el "Green *NDVI*" o *GNDVI*, con valores en el rango de -1 a 1 , definido tal que,

$$GNDVI = \frac{\rho_{NIR} - \rho_{green}}{\rho_{NIR} + \rho_{green}} = \frac{b8 - b3}{b8 + b3} \quad (5.2)$$

el índice *MSI* (Cruz et al., 2017; Palacios-Vélez et al., 2011), propuesto por Rock et al. (1986), con un rango de 0 a 3 y definido como,

$$MSI = \frac{\rho_{NIR}}{\rho_{SWIR}} = \frac{b8}{b11} \quad (5.3)$$

NDWI, propuesto por Gao (1996), con valores en el intervalo de -1 a 1 y definido como,

$$NDWI = \frac{\rho_{NIR} - \rho_{1240\text{ nm}}}{\rho_{NIR} + \rho_{1240\text{ nm}}} = \frac{b8 - b11}{b8 + b11} \quad (5.4)$$

, *NDRE1* (Dimov et al., 2022), propuesto por A. Gitelson y Merzlyak (1994), definido como la razón,

$$NDRE1 = \frac{\rho_{750\text{ nm}} - \rho_{705\text{ nm}}}{\rho_{750\text{ nm}} + \rho_{705\text{ nm}}} = \frac{b6 - b5}{b6 + b5} \quad (5.5)$$

, *NDRE* (Akbarian, Xu et al., 2023; Akbarian et al., 2022; Canata et al., 2021), propuesto por Barnes et al. (2000) como,

$$NDRE = \frac{\rho_{790\text{ nm}} - \rho_{720\text{ nm}}}{\rho_{790\text{ nm}} + \rho_{720\text{ nm}}} = \frac{b8 - b5}{b8 + b5} \quad (5.6)$$

, *CIRE* (Akbarian et al., 2022; Dimov et al., 2022), propuesto por A. A. Gitelson et al. (2005) y definido como,

$$CIRE = \left(\frac{\rho_{NIR}}{\rho_{red\ edge}} \right) - 1 = \left(\frac{b7}{b5} \right) - 1$$

Para la manipulación de las imágenes y calcular los índices mencionados anteriormente, se utilizó la biblioteca de Python *rasterio* (Gillies et al., 2016), la cual permite manipular datos de tipo raster, así como leer y escribir nuevos datos, esta acción se realiza seleccionando las bandas deseadas y operarlas aritméticamente. Con la biblioteca *rasterio*, se procede a calcular los índices de vegetación y almacenarlos en memoria de virtual (figura 5.12). El siguiente paso es extraer las estadísticas de los datos raster generados para cada parcela, Rahman y Robson (2016) propusieron obtener el valor promedio de los píxeles de la imagen, para asegurarse que los píxeles correspondieran a los valores dentro de la geometría de la parcela, implementaron un buffer interno de 10 metros, por su parte Canata et al. (2021) emplearon un buffer interno de 5 metros. Dimov et al. (2022) calcularon los valores mínimo, máximo, promedio y la desviación estándar por parcela. Para obtener los valores estadísticos de los píxeles se empleó la biblioteca *rasterstats* (Perry, 2015a) y como geometría de entrada un buffer interno de 10 metros (este buffer se realiza para cada parcela).

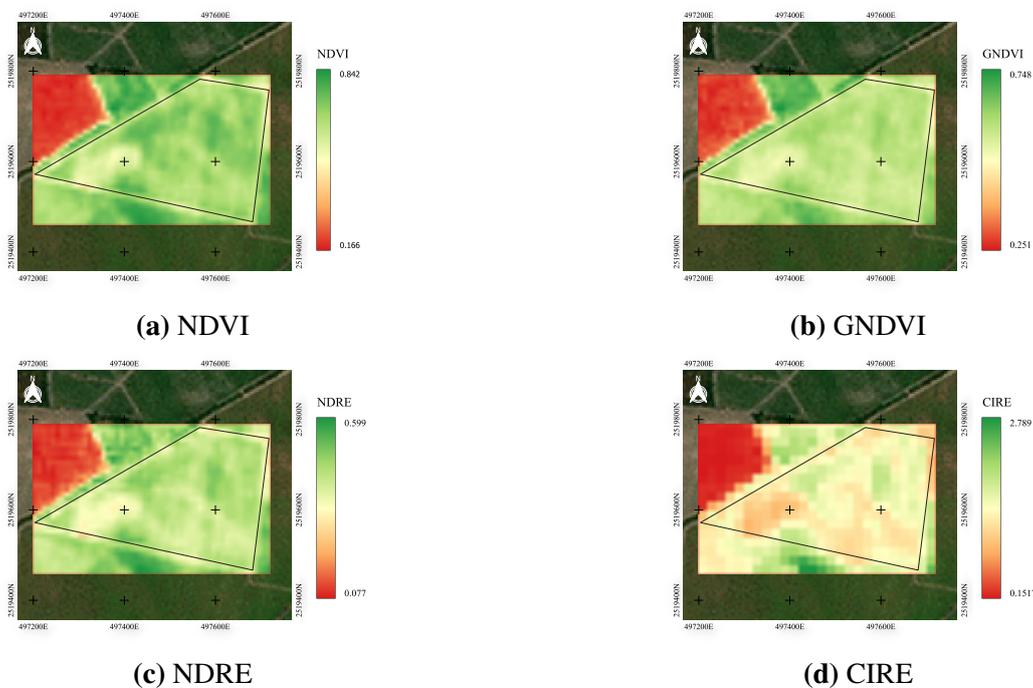


Figura 5.12: Representación de los principales índices de vegetación. Se puede apreciar como los índices GNDVI (b) y CIRE (d) representan las zonas con menor desarrollo vegetativo pero los índices NDVI y NDRE1 no, siendo que estos últimos dos comparten la banda 8 correspondiente al infrarrojo cercano por lo que se enfocan en una zona del espectro electromagnético en común.

Este buffer se guarda en memoria virtual y permite modificar el tamaño del buffer dependiendo de las necesidades sin la restricción de guardar los archivos en el disco por cada modificación. El objetivo de *rasterstats* es extraer información de datos de tipo raster basados en datos de entrada de tipo vectorial (ShapeFile, GeoJSON, etc.), aplicando el concepto de estadísticas por zona. El método *zonal statistics* devuelve por defecto 4 valores básicos, el valor mínimo, máximo, promedio y el conteo de píxeles dentro de la región o que intersectan con la misma.

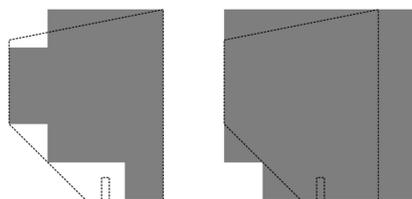


Figura 5.13: Opciones de rasterización en el método *zonal statistics*. En la representación del lado izquierdo, la opción *all_touched* es falsa, por lo que no incluye todos los píxeles con los que intersecta la geometría de entrada, en la imagen de la derecha, la opción es verdadera, por lo que incluye todos los píxeles que intersectan con la geometría de entrada. Tomada de Perry (2015b)

Se selecciona la opción *false* en conjunto con el buffer interno de 10 metros para asegurar que los píxeles correspondan únicamente al interior de la parcela y no a las orillas u otras zonas no deseadas. Los datos resultantes se asocian a la fecha correspondiente a la imagen de la cual se calcula las estadísticas y se almacenan dentro de un DataFrame de *pandas* (McKinney, 2010), dando como resultado una estructura de datos que se exporta al formato csv para su posterior manipulación. Este conjunto de datos recibirán el nombre de *serie de tiempo*, en la cual se puede visualizar el desarrollo vegetativo del cultivo en función del tiempo transcurrido.

La figura 5.14 muestra la serie de tiempo del *NDVI* para una de las parcelas de estudio (parcela No.5), esta serie conformada por observaciones secuenciales del índice puede mostrar la progresión del dosel del cultivo desde la emergencia hasta la senescencia (Rahman & Robson, 2016).

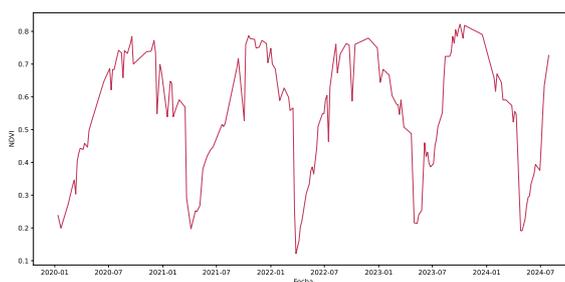
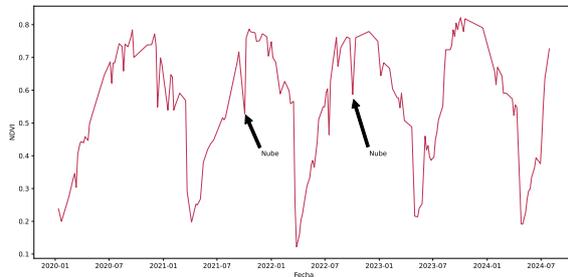


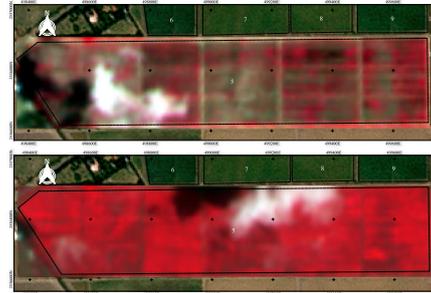
Figura 5.14: Serie de tiempo del índice *NDVI* de la parcela No. 5 durante el periodo de estudio. Se puede apreciar un patrón que representa los ciclos de desarrollo durante múltiples temporadas de cultivo.

En la serie de tiempo presentada en la figura 5.14, se pueden visualizar valores atípicos al comportamiento del desarrollo vegetativo, esto se puede deber a anomalías en los sistemas de adquisición o fenómenos meteorológicos, siendo principalmente la presencia de nubes la causa de estos valores. En la figura 5.15a se identifican los valores atípicos como el resultado de la presencia de nubes en la escena en el momento de la adquisición de la imagen, para asegurar que estos valores correspondan a la presencia de nubes es necesario visualizar la escena (figura 5.15b), la presencia de nubes cubre gran parte de la escena, al igual que la sombra de la

misma, por lo que al momento de obtener los valores estadísticos, los píxeles que corresponden a las nubes causan una anomalía con respecto a los que corresponden a la vegetación, en este tipo de situaciones en las que nubes, sombras o datos en mal estado, generan inconsistencias, Rahman y Robson (2020) optaron por descartar estas imágenes.



(a) Presencia de nube en la serie de tiempo



(b) Presencia de nube dentro de una escena

Figura 5.15: Anomalías en la serie de tiempo asociadas a la presencia de nubes en la escena. **a)** se puede identificar valores atípicos en la serie de tiempo, ya que estos contradicen por completo el comportamiento del cultivo, **b)** al revisar la escena correspondiente en una composición de falso color, esos valores atípicos se verifica la presencia de nubes en la escena, representadas de color blanco y su respectiva sombra en color negro.

Siguiendo el enfoque propuesto por Rahman y Robson (2020), se procede a revisar individualmente las imágenes correspondientes a estos picos en los que el valor del índice para esa fecha difiere con lo esperado o como debería ser para esa etapa del desarrollo vegetativo, una vez que se tienen las fechas identificadas en las que existe presencia de nubes, sombra, niebla u otros posibles efectos atmosféricos, se procede a descartar los datos correspondientes a esas fechas en la serie de tiempo. En la figura 5.16 se puede visualizar como el perfil de la serie de tiempo tiene un comportamiento más uniforme, sin cambios drásticos en los valores del índice que no corresponden al desarrollo del cultivo y corresponden a condiciones atmosféricas.

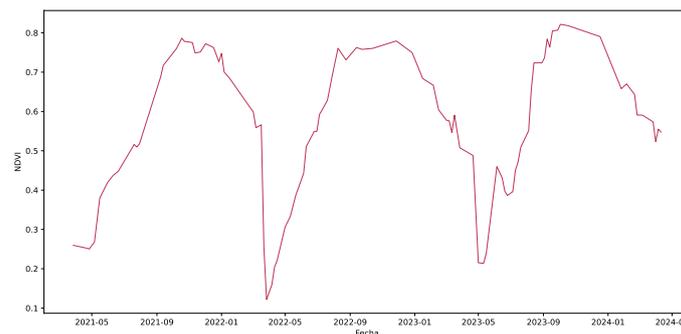


Figura 5.16: Serie de tiempo corregida. La presencia de nubes fue eliminada del conjunto de datos, obteniendo una serie de tiempo que contiene el comportamiento propio del cultivo.

Como se mencionó anteriormente, en algunas circunstancias es necesario descartar imágenes por la presencia de nubes en la escena u otras situaciones que merman la calidad de los datos, por lo que el conjunto de datos disminuye y no permite tener un panorama general del desarrollo del cultivo durante todo el periodo de crecimiento, con el propósito de tener una vista de la progresión del crecimiento del cultivo, Rahman y Robson (2016) ajustaron los valores promedio del índice GNDVI con un modelo cuadrático, con el que reconstruyeron la serie de tiempo del

crecimiento del cultivo durante su periodo de desarrollo. De forma similar, se ajustaron los datos de las series de tiempo por cada índice espectral, dividida por temporada, a polinomios de distintos grados para determinar cuales ajustan mejor los datos sin perder en cuenta la complejidad que implica aumentar grados de ajuste, para esto se comparó el AIC, el cual mide la complejidad de un modelo. A partir de un grado 2, el valor del AIC se comporta de una forma muy similar hasta llegar a un grado en el que comienza a aumentar, indicando que la complejidad del modelo aumenta. A medida que se aplica un grado mayor, los datos modelados tienden a asimilarse más a los datos reales o de entrada, pero a mayor grado del polinomio se tiende a sobre ajustar el modelo, por lo que los datos resultantes no describirán adecuadamente la característica deseada, si no que tenderán a replicar o imitar los datos de entrada, esto se puede visualizar en la figura 5.17b, en la que a partir de un grado 4 el modelo tiende a imitar los datos de entrada, independientemente de si los datos resultantes siguen el comportamiento esperado. Para la elección del grado adecuado se priorizó evitar el sobre ajuste de los datos, en la figura 5.17a se observa como el coeficiente de determinación y el AIC, para los grados 2, 3 y 4, son muy similares (para el caso de la parcela No.1), por lo que se optó por un grado 2, un modelo cuadrático, similar a lo realizado por Rahman y Robson (2016).

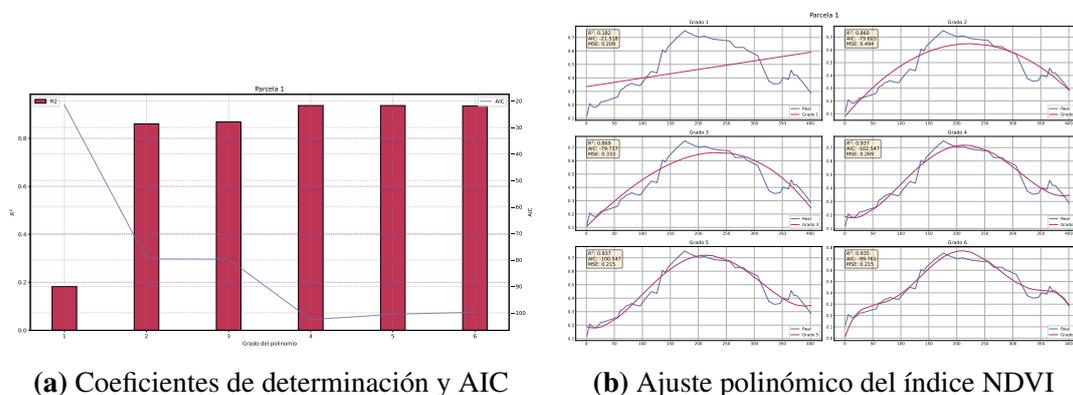


Figura 5.17: Modelos de regresión polinómica para diferentes grados de ajuste. **a)** conforme se aumentan los grados de ajuste la complejidad del modelo comienza a aumentar sin que esto implique una mejor sustancial en el coeficiente de determinación, por lo que la elección del grado adecuado es importante para modelar el comportamiento de los datos de forma adecuada, **b)** al aumentar los grados de ajuste llega un punto en el que los datos tienden a imitar las variaciones sin que estas correspondan al comportamiento natural de los datos.

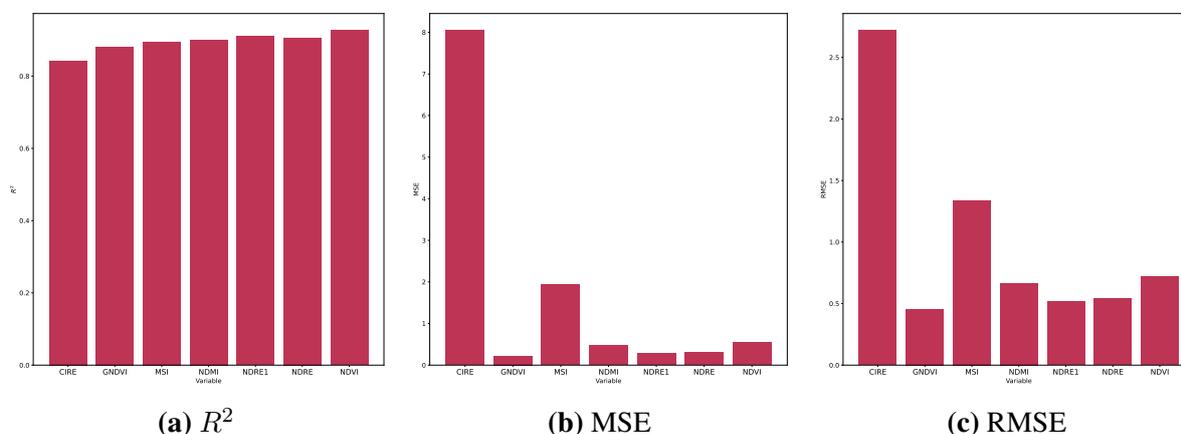


Figura 5.18: Métricas de validación por modelo considerando las 3 temporadas.

Al ajustar los datos con este grado, se obtuvieron los modelos y métricas estadísticas por cada índice, para cada parcela (15 en total) y por cada temporada (3 en total), encontrando que el coeficiente de determinación (R^2) tiene un comportamiento similar en la mayoría de los índices, siendo el índice *NDVI* el que obtiene mejor resultado, por su parte el índice *GNDVI* obtuvo los mejores *MSE* (5.18b) y *RMSE* (5.18c), siendo la variable que menores errores tiene en su modelo. Los resultados de las métricas se presentan en la tabla 5.3.

A partir de estos modelos se obtuvieron las series de tiempo modeladas de cada uno de los índices espectrales, por parcela y de forma individual para cada una de las 3 temporadas que comprende este estudio, haciendo una distinción en la terminología empleada hasta el momento, teniendo series de tiempo *reales* y *modeladas*.

Índice	R^2	<i>MSE</i>	<i>RMSE</i>
<i>NDVI</i>	0.927	0.542	0.720
<i>GNDVI</i>	0.881	0.220	0.456
<i>NDMI</i>	0.901	0.476	0.667
<i>MSI</i>	0.895	1.935	1.337
<i>NDREI</i>	0.911	0.281	0.516
<i>NDRE</i>	0.906	0.310	0.541
<i>CIRE</i>	0.843	8.049	2.721

Tabla 5.3: Métricas de validación por variable considerando la totalidad de temporadas.

5.2. Variables agrometeorológicas

Se ha sugerido contrastar los datos generados a partir de imágenes de satélite con datos climáticos y agronómicos de la zona de estudio para estimar la producción del cultivo de caña de azúcar (Cruz et al., 2017; Morel et al., 2014). Por otro lado, autores como Dimov et al. (2022) descartaron incluir los datos meteorológicos debido al tamaño de su zona de estudio, ya que relativamente era pequeña y la resolución espacial de esos datos no permitió observar una variabilidad espacial, además de un bajo nivel de asociación entre los datos del suelo y el rendimiento del cultivo. Por lo que en este estudio se analizará la relación y posible influencia de variables agrometeorológicas como, precipitación, temperatura del aire, humedad relativa, radiación solar, evapotranspiración de referencia (ET_o), evapotranspiración del cultivo (ET_c) y el coeficiente de cultivo, y se analizará su influencia en el rendimiento del cultivo.

5.2.1. Dataset AgERA5

La adquisición de estos datos proviene comúnmente de bases de datos climáticas que recopilan los datos a distintos niveles, desde global a local (Aguilar-Rivera et al., 2018; Baez-Gonzalez et al., 2017), en México estos datos provienen principalmente de estaciones meteorológicas pertenecientes al SMN o al INIFAP (Baez-Gonzalez et al., 2017; Cruz et al., 2017). En este estudio no fue posible emplear los datos de las estaciones mas cercanas a la zona de estudio debido a que estas dejaron de recopilar los datos hasta el 2011, por lo que fue necesario buscar otra fuente de datos. Una alternativa es el dataset *Agrometeorological indicators from 1979 to present derived from reanalysis* (Boogaard et al., 2020), el cual provee datos meteorológicos diarios de la superficie desde 1979 a la fecha presente que se consulte para aplicaciones agroecológicas (Araghi et al., 2022; Karaman & Akyürek, 2023; Roffe & van der Walt, 2023), se encuentra disponible a una resolución espacial de 0.1° y está basado en la quinta generación del

reanálisis atmosférico del clima global de ECMWF (Hersbach et al., 2020). El dataset AgERA5 se compone de 12 variables principales, las cuales se distribuyen en un formato NetCDF-4 con una resolución espacial de 0.1° y temporal diaria, se actualiza cada mes y tiene una cobertura global. En la tabla A.1 se describen las 12 variables que provee el dataset. Adicionalmente y como se mencionó anteriormente, una de las variables a estudiar es la Evapotranspiración, esta variable no se encuentra dentro de las provistas por el dataset AgERA5, para esto se emplea el dataset *Reference evapotranspiration - AgERA5 derived (Global - Daily - ~ 10 km)*, que se generó a partir de las variables del dataset AgERA5 y acorde al método FAO Penman - Monteith que se describe en FAO: Drenaje y Riego 56 (Allen et al., 2006). Ambos dataset están disponibles en la Climate Data Store (CDS) y se pueden descargar a través de ordenes individuales por dataset o por medio de las API's que proporcionan en distintos lenguajes de programación, también se puede acceder a ellos desde el catalogo de dataset de *Earth Engine* y a través de herramientas como la desarrollada por Brown et al. (2023). Se descargaron los datos desde la CDS y se procesaron siguiendo de forma similar a lo propuesto anteriormente para las imágenes satelitales. A partir de los centroides de cada una de las parcelas se extrajo el valor de la celda por cada dataset y se asoció a la fecha correspondiente, obteniendo un conjunto de datos compuesto por fecha y valor, conformando una serie de tiempo por cada variable y cada parcela.

A partir de estos datos se generaron series de tiempo por variable (figura 5.19), con la finalidad de tener un mayor contexto histórico y distinguir efectos atípicos o de temporada, la extensión temporal de estas series de tiempo abarcan el periodo de inicios de 2020 a agosto de 2024.

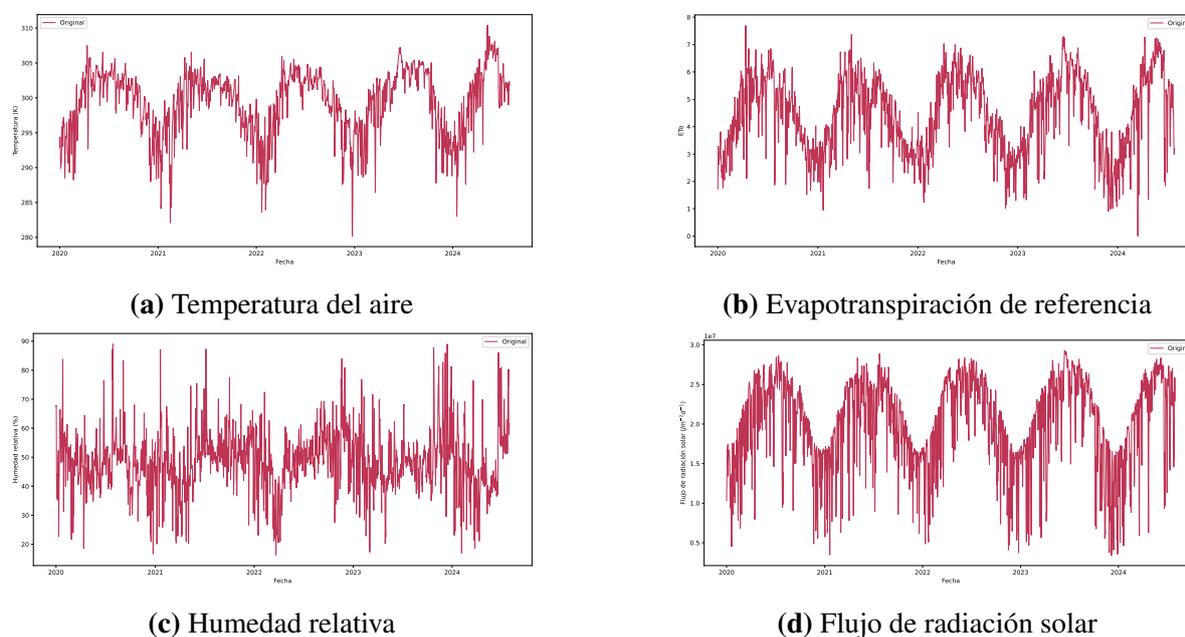


Figura 5.19: Series de tiempo de las principales variables agroclimáticas en la ubicación de la parcela 5. Se puede apreciar un patrón cíclico, esto se debe a que el estudio abarca un periodo de 4 años y por consecuencia se identifican las variaciones propias de cada variable a lo largo de las temporadas que lo componen.

Las dos variables restantes, evapotranspiración del cultivo y coeficiente del cultivo, se pueden calcular a partir de las ecuaciones descritas por Allen et al. (2006), si se cuenta con los datos de las variables de entrada para el modelo sugerido, una alternativa es emplear el índice *NDVI* para calcular dichos valores, Dingre et al. (2021) desarrollaron una metodológica para estimar

ambas variables con base a este índice, obteniendo un coeficiente de cultivo (K_c) 16.6 % menor a lo sugerido en FAO-56 (Allen et al., 2006) para el cultivo de caña de azúcar. Por su parte Palacios-Vélez et al. (2011), como parte de su metodología, obtuvieron una fórmula para el coeficiente de cultivo (K_c) casi idéntica a la sugerida por Calera y González-Piqueras (2007). Cruz et al. (2017) adaptaron el estudio de Palacios-Vélez et al. (2011) que originalmente se desarrolló para trigo, al cultivo de la caña de azúcar y con base a la ecuación propuesta por Calera y González-Piqueras (2007), calcularon las variables K_c y ET_c empleando los datos obtenidos de estaciones meteorológicas para estimar el rendimiento.

La evapotranspiración del cultivo ET_c se calcula como el producto de la evapotranspiración de referencia ET_o y el coeficiente de cultivo K_c (Allen et al., 2006):

$$ET_c = K_c ET_o \quad (5.7)$$

Los datos de ET_o se descargaron desde el dataset AgERA5, por lo que queda calcular el coeficiente calculado K_c a partir de los enfoques mencionados:

$$K_c = 1.15 * NDVI + 0.17 \quad (5.8)$$

propuesta por Calera y González-Piqueras (2007).

$$K_c = 1.1478 * NDVI + 0.1716 \quad (5.9)$$

obtenida por Palacios-Vélez et al. (2011).

Para calcular las variables objetivo se emplearon las ecuaciones (5.7) y (5.8), los valores resultantes se recortaron al periodo de tiempo que comprende el ciclo productivo del cultivo (por cada parcela) para visualizar su comportamiento a lo largo de este. En la figura 5.20 se puede distinguir la diferencia entre evapotranspiración de referencia (ET_o) y de cultivo (ET_c), esto se debe a la inclusión de las características propias del cultivo (Allen et al., 2006).

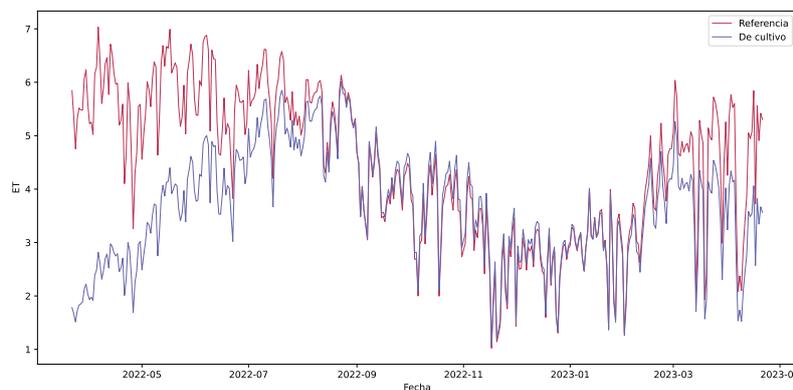


Figura 5.20: Serie de tiempo de la evapotranspiración de referencia y cultivo en la parcela No.5. Al incluir las características propias del cultivo (NDVI) el comportamiento de la evapotranspiración cambia, se puede apreciar como al comienzo de la serie de tiempo hay un desfase negativo entre el valor de la ET_o y ET_c , esto se debe al bajo nivel de desarrollo vegetativo del cultivo, que conforme avanza el tiempo casi iguala el comportamiento de la ET_o y en la parte final disminuye ligeramente, propio de esa parte de la etapa de crecimiento.

El coeficiente de cultivo único K_c incorpora las características del cultivo y los efectos promedios de la evaporación en el suelo. Durante el período de crecimiento del cultivo, la variación del coeficiente del cultivo K_c expresa los cambios en la vegetación ($NDVI$) y en el

grado de cobertura del suelo. Esta variación del coeficiente K_c a lo largo del crecimiento del cultivo está representada por la curva del coeficiente del cultivo.

Por cada parcela se generó un dataset conformado por los valores promedios modelados por VI's y WI's, así como las variables agrometeorológicas en una escala temporal diaria, delimitados a las fechas de cosecha proporcionadas para cada una de las temporadas, con la finalidad de relacionar estos datos con el rendimiento reportado y analizar su posible efecto en su desempeño.

Los datos generados en esta sección (series de tiempo) se analizarán en la siguiente sección y serán los datos de entrada para la generación de modelos que permitan estimar el rendimiento del cultivo de la caña de azúcar en la zona de estudio.

CAPÍTULO 6

RESULTADOS Y DISCUSIÓN

6.1. Índices espectrales

Resultado de lo desarrollado en la sección anterior, se obtuvieron las series de tiempo de los VI's, WI's y las variables agrometeorológicas para cada una de las parcelas analizadas en las 3 temporadas que comprende el estudio. A partir de estas series de tiempo fue posible calcular características propias del desarrollo del cultivo, las cuales reciben del nombre de métricas fenológicas. Dimov et al. (2022) calcularon distintas métricas con base a las propuestas por el USGS (Reed et al., 1994). Típicamente estas métricas así como los llamados marcadores fenológicos se extraen de las series de tiempo derivadas del índice *NDVI*. Para cada una de las parcelas se calcularon los marcadores y las métricas fenológicas, considerando las series de tiempo reales y modeladas, con base a lo sugerido por Dimov et al. (2022). A continuación se describen cada uno de ellos:

- Inicio de la temporada (SOS): la fecha en la que inicia el crecimiento y el cultivo emerge
- Cumbre de la temporada (POS): la fecha correspondiente al máximo valor presente en la serie de tiempo
- Final de la temporada (EOS): la fecha de cosecha
- Valor del POS: valor máximo del *NDVI* en la serie de tiempo
- Duración de la temporada: Numero de días transcurridos entre el SOS y POS
- Primera mitad: duración entre el SOS y POS
- Segunda mitad: duración entre el POS y EOS

El inicio de la temporada (SOS) se considera como la primera fecha en la serie de tiempo en la que el valor del *NDVI* rebasa el umbral de 0.2 y continua con su correspondiente desarrollo (Reed et al., 1994). Se calculó la duración promedio de cada temporada y el valor máximo promedio que alcanza el índice *NDVI*, en la temporada 2021/2022, la duración promedio de la temporada fue de 424 días con un valor promedio máximo de 0.771, para la temporada 2022/2023, la duración promedio de la temporada es de 333 días con un valor promedio máximo de 0.769, finalmente en el caso de la temporada 2023/2024, tuvo una duración de 353 días con

un valor máximo promedio de 0.765. De acuerdo a los resultados, se encontró que para las 3 temporadas, el valor máximo del *NDVI* se alcanzó en el periodo de Junio a Enero (tabla 6.1).

Parcela	21/22		22/23		23/24	
	Fecha	POS	Fecha	POS	Fecha	POS
1	2021-07-20	0.722	2022-09-08	0.751	2024-02-20	0.733
2	2021-10-18	0.812	2022-11-27	0.796	2023-10-18	0.745
3	2021-06-05	0.710	2022-11-27	0.714	2023-10-28	0.762
4	2021-10-18	0.805	2022-11-27	0.783	2023-10-18	0.748
5	2021-10-18	0.787	2022-11-27	0.779	2023-10-03	0.822
6	2021-10-18	0.761	2022-11-27	0.795	2024-07-14	0.757
7	2021-10-18	0.819	2022-10-13	0.811	2023-12-17	0.796
8	2021-12-27	0.735	2022-11-27	0.812	2023-10-03	0.829
9	2021-10-18	0.785	2022-11-27	0.788	2024-04-10	0.746
10	2021-11-07	0.789	2022-08-09	0.793	2023-12-17	0.787
11	2021-10-18	0.784	2022-08-09	0.802	2023-10-18	0.802
13	2021-10-18	0.768	2022-11-27	0.668	2023-10-18	0.722
14	2021-07-20	0.770	2022-11-27	0.811	2024-07-14	0.820
15	2021-10-03	0.808	2022-11-27	0.785	2023-12-27	0.771
16	2021-06-10	0.712	2022-11-27	0.653	2024-02-20	0.640

Tabla 6.1: Fechas y valores del valor máximo alcanzado por el índice *NDVI* extraído de las serie de tiempo reales.

Al igual que con las series de tiempo reales, se calcularon los marcadores y las métricas fenológicas de las series modeladas (tabla 6.2), se encontró que la duración promedio en el caso de la temporada 2021/2022 fue de 425.86 días, con un valor promedio máximo de 0.730, para la 2022/2023 fue de 331.13 días y 0.735, para la 2023/2024, la temporada duró en promedio 356.53 días y alcanzó un valor máximo promedio de 0.773.

Parcela	21/22		22/23		23/24	
	Fecha	POS	Fecha	POS	Fecha	POS
1	2021-09-08	0.638	2022-10-10	0.651	2024-04-10	0.911
2	2021-11-14	0.725	2022-11-21	0.781	2023-11-08	0.727
3	2021-09-08	0.655	2022-10-12	0.676	2023-09-27	0.718
4	2021-11-09	0.729	2022-11-27	0.743	2023-10-26	0.689
5	2021-11-16	0.742	2022-11-07	0.776	2023-11-08	0.809
6	2021-11-27	0.735	2022-12-11	0.760	2024-07-14	0.818
7	2022-01-16	0.789	2022-10-16	0.796	2023-11-06	0.791
8	2021-11-12	0.703	2022-12-07	0.750	2023-11-23	0.835
9	2021-11-21	0.748	2022-12-08	0.765	2024-03-10	0.706
10	2021-10-19	0.779	2022-10-05	0.797	2023-10-22	0.794
11	2021-10-15	0.775	2022-10-02	0.811	2023-10-19	0.808
13	2021-10-08	0.749	2022-11-11	0.667	2023-10-19	0.706
14	2021-10-24	0.738	2022-11-22	0.702	2024-07-14	0.910
15	2021-10-26	0.769	2022-11-28	0.753	2023-11-08	0.769
16	2021-09-28	0.687	2022-11-18	0.608	2024-04-15	0.610

Tabla 6.2: Fechas y valores del valor máximo alcanzado por el índice *NDVI* extraído de las serie de tiempo modeladas.

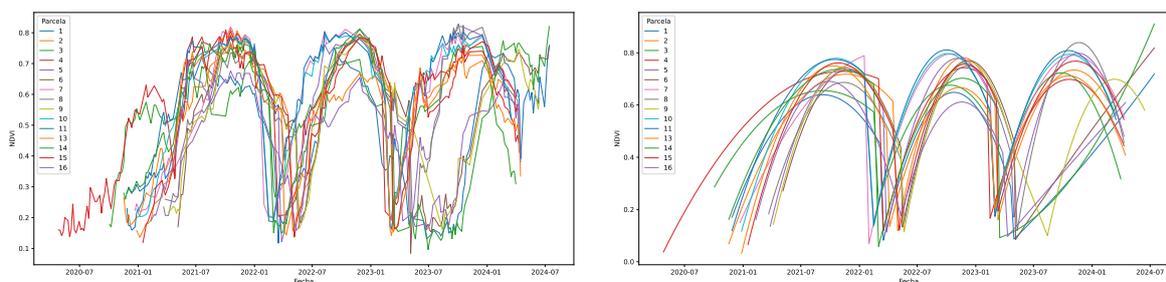
Se comparó la duración de temporada entre las parcelas de estudio para cada una de las 3 temporadas que se analizan (tabla 6.3), encontrando que en el caso de la parcela No.4 en la

temporada 2021/2022 tiene una duración de 655 días, fuera del rango esperado que como se mencionó en la sección de antecedentes del cultivo, puede durar hasta 17 meses en casos de ciclo de plantilla, que de acuerdo a la proporcionado por el propietario, no es el caso, por lo que se procedió a descartar los datos asociados a esta parcela durante el desarrollo de este estudio.

Parcela	21/22	22/23	23/24
1	465	345	335
2	475	290	315
3	485	345	390
4	655	325	405
5	325	395	335
6	360	345	365
7	375	365	395
8	400	340	320
9	375	330	290
10	375	365	370
11	385	345	365
13	400	290	395
14	450	330	435
15	420	295	365
16	415	305	220

Tabla 6.3: Duración de la temporada en días. Desde el SOS al EOS.

En la figura 6.1 se presentan las series de tiempo, en las que se puede visualizar el inicio y final de cada temporada, la duración de la misma, así como el desarrollo del cultivo a través de la evolución del *NDVI*.



(a) Series de tiempo del índice NDVI

(b) Series de tiempo del índice NDVI modelado

Figura 6.1: Series de tiempo del índice NDVI por parcela, abarca el periodo de tiempo establecido para este estudio, 3 temporadas. **a)** en la serie de tiempo ‘real’ es posible apreciar la evolución en el desarrollo del cultivo para cada una de las parcelas, y como las aparentes en la temporada 2022/2023, **b)** las series de tiempo modeladas permiten visualizar de una forma más sencilla el comportamiento del desarrollo del cultivo, identificando visualmente su momento de máximo desarrollo vegetativo así como el patrón de su ciclo productivo.

Aguilar-Rivera et al. (2018) mencionan que la temporada de cosecha o zafra comprende el periodo de Noviembre a Junio para cultivos de 12 meses de desarrollo, socas y resocas, esto se confirma con el comportamiento de las series de tiempo presentadas en la figura 6.1 y las tablas 6.1 y 6.2.

6.2. Variables de peso en el rendimiento de la producción de la caña de azúcar

Se analizaron las variables asociadas al rendimiento en la producción de la caña de azúcar, así como su correlación para identificar y determinar las que mayor influencia tienen en el proceso de producción. Tomando como punto de partida que el rendimiento está condicionado al nivel de desarrollo vegetativo durante el ciclo de desarrollo, se tomarán en consideración los VI's y WI's, esto corroborado por los estudios previos desarrollados (Cruz et al., 2017; Dimov et al., 2022; Rahman & Robson, 2020; Rahman & Robson, 2016). Som-ard et al. (2021) menciona que factores climáticos, tales como precipitación, temperatura, radiación solar y humedad relativa, afectan significativamente la producción de la caña de azúcar. Otro factor a considerar es la evapotranspiración del cultivo (ETc), empleada por Cruz et al. (2017) para caña de azúcar y validada por Palacios-Vélez et al. (2011) para trigo.

El primer conjunto de variables a comparar fueron los valores promedios y máximos por cada índice espectral calculado previamente, extraídos y calculados a partir de sus respectivas series de tiempo (reales y modeladas). Se compararon las muestras de las 3 zafras (tabla 6.4), 2021/2022, 2022/2023 y 2023/2024, dando como resultado una muestra de 39 datos (los valores correspondientes a la parcela 3 y 4 de la zafra 2021/2022, parcela 1, 6, 14 y 6 de la zafra 2023/2024 se descartaron debido a inconsistencias con el comportamiento de las otras parcelas) cuando se analizan los datos de las 3 zafras en conjunto.

Datos	Variable	21/22	22/23	23/24	Conjunto
Reales	max <i>GNDVI</i>	0.692	0.823	0.441	0.543
	mean <i>GNDVI</i>	0.678	0.761	0.019	0.422
	max <i>CIRE</i>	0.512	0.841	0.403	0.550
	mean <i>CIRE</i>	0.673	0.706	0.077	0.418
	max <i>NDREI</i>	0.500	0.827	0.436	0.578
	mean <i>NDREI</i>	0.680	0.741	-0.093	0.360
	max <i>NDRE</i>	0.567	0.845	0.430	0.549
	mean <i>NDRE</i>	0.706	0.771	-0.040	0.405
	max <i>NDVI</i>	0.573	0.845	0.317	0.578
	mean <i>NDVI</i>	0.665	0.711	-0.190	0.316
Modelados	max <i>GNDVI</i>	0.506	0.875	0.216	0.441
	mean <i>GNDVI</i>	0.625	0.796	0.146	0.390
	max <i>CIRE</i>	0.598	0.849	0.272	0.474
	mean <i>CIRE</i>	0.627	0.795	0.204	0.410
	max <i>NDREI</i>	0.524	0.834	0.203	0.446
	mean <i>NDREI</i>	0.585	0.743	0.081	0.334
	max <i>NDRE</i>	0.590	0.843	0.216	0.472
	mean <i>NDRE</i>	0.628	0.777	0.100	0.377
	max <i>NDVI</i>	0.463	0.828	0.072	0.402
	mean <i>NDVI</i>	0.526	0.721	-0.015	0.269

Tabla 6.4: Coeficiente de correlación (pearson) entre valores de los VI's empleados y el rendimiento reportado, por zafra y en conjunto. Para las primeras 2 temporadas la correlación entre las variables estudiadas mostraron un comportamiento similar, caso opuesto a las correspondientes a la temporada 2023/2024, que disminuyeron su correlación o en algunos casos se anularon o cambiaron de sentido y magnitud.

En el caso de la zafra 2021/2022, se encontró que la variable real que obtuvo la mejor correlación fue el valor promedio del índice *NDRE* (figura 6.2a), ligeramente superior al máximo

6.2. VARIABLES DE PESO EN EL RENDIMIENTO DE LA PRODUCCIÓN DE LA CAÑA DE AZÚCAR

del *GNDVI*, por el lado de los datos modelados, el valor promedio del *NDRE* (figura 6.2b) fue el que mayor correlación obtuvo con el rendimiento reportado para esa zafra, con un valor casi idéntico al obtenido por el promedio del *GNDVI*. Para la zafra 2022/2023, las variables reales tuvieron un comportamiento muy similar respecto al rendimiento, destacando ligeramente sobre las demás, los valores máximos de los VI's, en especial los valores máximos del *NDVI* y *NDRE* (figura 6.2c), mismo caso para los datos modelados pero destacando el máximo del *GNDVI* (figura 6.2d). En la temporada 2023/2024, el valor máximo del *GNDVI* (figura 6.2e) obtuvo la mayor correlación en el caso de los datos reales, por el lado de los modelados, el máximo del *CIRE* (figura 6.2f) destacó de entre las otras variables.

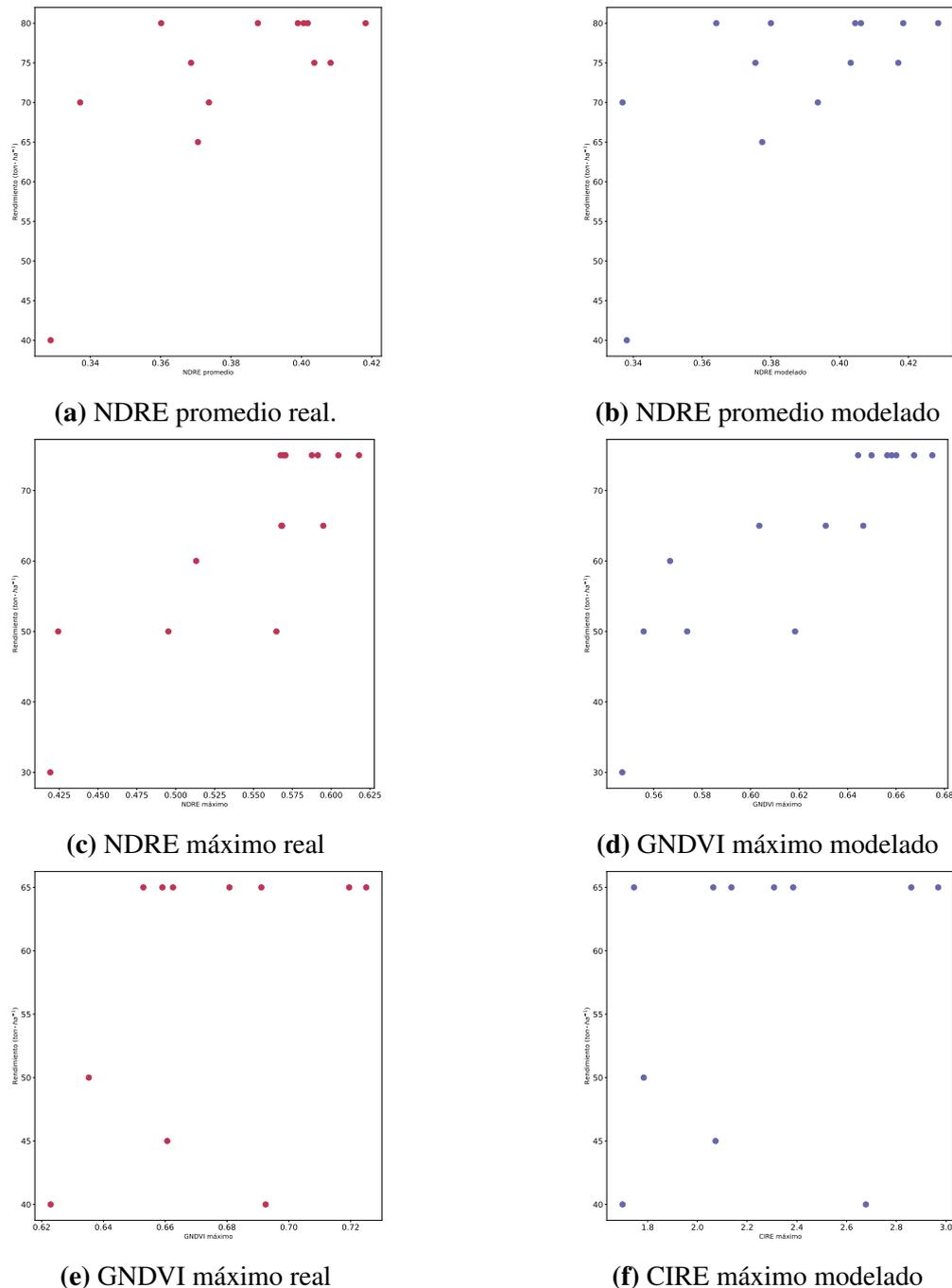


Figura 6.2: Variables con la mayor correlación con el rendimiento reportado por temporada y tipo de dato. **a)** y **b)** corresponden a la temporada 2021/2022, **c)** y **d)** a la 2022/2023, **e)** y **f)** a la 2023/2024.

La figura 6.2 permite visualizar las diferencias entre la correlación presente en las distintas variables analizadas, encontrando que para la temporada 2021/2022, las variables empleadas obtienen una correlación positiva que va desde moderada a fuerte, en la siguiente temporada, 2022/2023, la correlación aumenta, su magnitud y dirección se hace mas evidente, en el caso de la 2023/2024, la correlación disminuye drásticamente y se ve representada en las gráficas de dispersión presentadas.

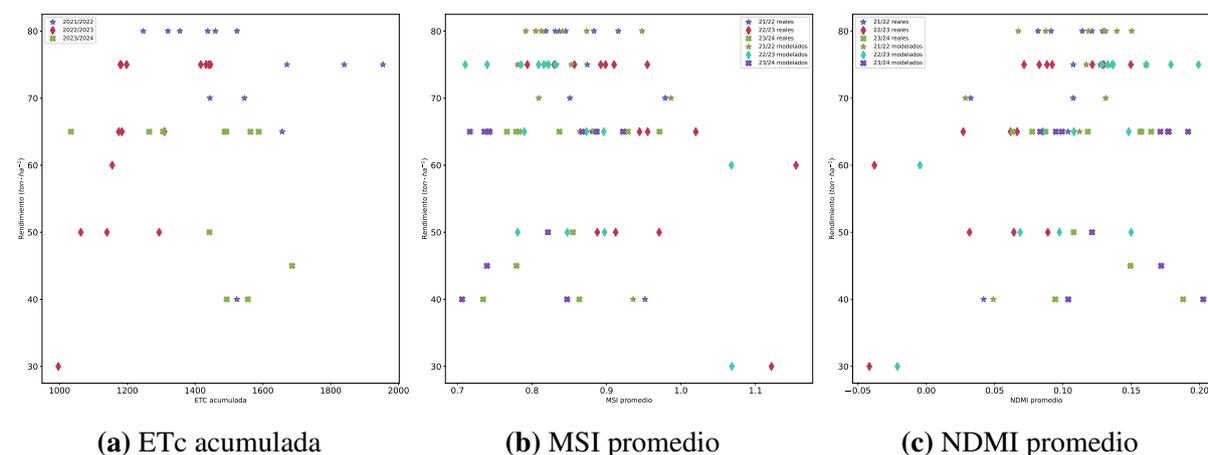


Figura 6.3: Comparación entre el valor de la evapotranspiración del cultivo acumulada, el promedio de los índices *MSI* y *NDMI* y su correlación con el rendimiento reportado. **a)** al analizarse de forma individual se aprecia un desfase entre los valores acumulados entre las temporadas 2021/2022 y 2022/2023, esto se debe a la mayor extensión promedio de la temporada y por consecuencia un valor acumulado mayor, sin que esto influya directamente en el rendimiento, en el caso de la ultima temporada la correlación cambia su comportamiento drásticamente, **b)** el promedio del *MSI* proporciona una correlación fuerte con el rendimiento, esto debido a que ejemplifica correctamente la influencia del estrés hídrico en el rendimiento, a excepción de nuevamente la temporada 2023/2024, que presenta una correlación positiva, lo que va en contra de lo mencionado, **c)** el valor promedio del *NDMI* indica que conforme aumenta la humedad en el cultivo el rendimiento aumentará, esto aplica para las 2 primeras temporadas, ya que para el caso de la ultima los resultados indican lo contrario.

De la figura 6.3a se puede determinar que para el caso de las parcelas de estudio, la evapotranspiración del cultivo acumulada durante la temporada tiene una correlación débil frente al rendimiento anual declarado cuando se evalúan las 3 temporadas en conjunto, al desglosarse de manera individual por temporada como en la tabla 6.5, se puede apreciar como para la zafra 2022/2023 a medida que la ETc acumulada es mayor el rendimiento también aumenta, esto confirma lo dicho por Cruz et al. (2017), a medida que aumenta la evapotranspiración, se espera más fotosíntesis y, en consecuencia, más biomasa y rendimiento., mientras que para el índice MSI se observa una correlación negativa fuerte, ya que conforme aumenta el estrés hídrico, el rendimiento disminuye, similar a lo que ocurre con el índice *NDMI* (Cruz et al., 2017; Palacios-Vélez et al., 2011).

A partir de lo establecido en las tablas 6.4 y 6.5, tanto los VI's, WI's y la ETc acumulada tienen un correlación considerable respecto al rendimiento anual que fue reportado por cada parcela en el estudio, se encontraron excepciones durante el análisis como lo señalado para los datos obtenidos a partir del conjunto correspondiente a la temporada 2023/2024, los cuales presentan inconsistencias en comparación con el comportamiento de las temporadas previas.

6.2. VARIABLES DE PESO EN EL RENDIMIENTO DE LA PRODUCCIÓN DE LA CAÑA DE AZÚCAR5

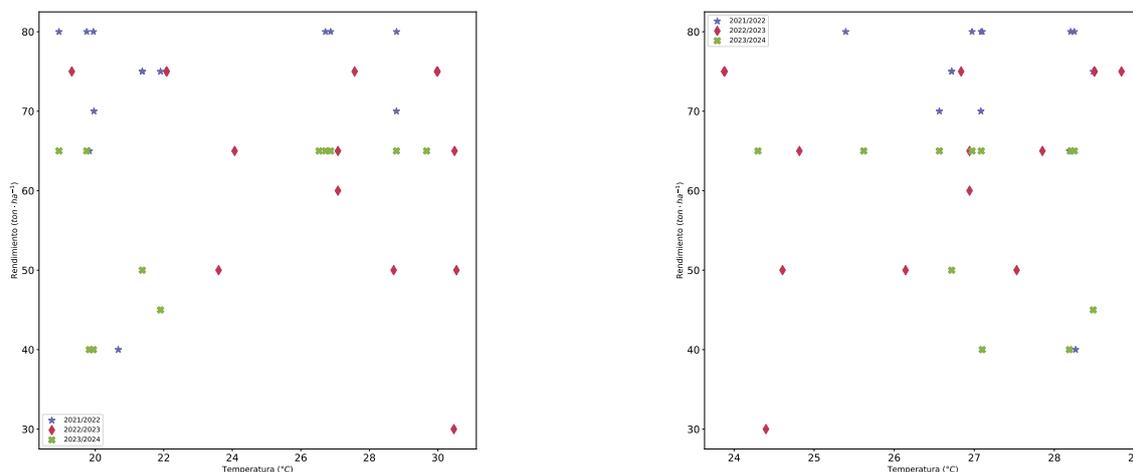
Datos	Variable	21/22	22/23	23/24	Conjunto
Reales	mean MSI	-0.534	-0.591	0.316	-0.223
	mean NDMI	0.577	0.696	-0.247	-0.302
Modelados	ETc acum	-0.175	0.679	-0.429	0.164
	mean MSI	-0.422	-0.648	0.176	-0.213

Tabla 6.5: Correlación de los WI's y ETc acumulada con el rendimiento reportado, por temporada y en conjunto. Las primeras 2 temporadas siguen un comportamiento similar, donde las variables logran explicar el aumento o disminución en el rendimiento en función de estas, los datos de la temporada 2023/2024 disminuyen el valor total de la correlación al incluirlos en el análisis.

En conjunto con las variables generadas a partir de los índices espectrales calculados, se midió la correlación entre las variables climatologías y el rendimiento reportado, de acuerdo con Som-ard et al. (2021), las variables que influyen el rendimiento son la temperatura, precipitación, radiación solar y humedad relativa. La primer variable a analizar es la temperatura, conforme a lo establecido en la definición de las etapas fenológicas y los requerimientos térmicos (tabla 6.6), se calculó el valor promedio de la temperatura durante los intervalos de tiempo que comprenden las etapas fenológicas críticas para la evolución y desarrollo del cultivo y se midió la correlación con el rendimiento reportado (figura 6.4 y tabla 6.7).

Etapa	Periodo (días)	Temperatura
Germinación	30 a 50	30 °C
Rápido crecimiento	180 a 220	32 °C

Tabla 6.6: Temperatura óptima por etapa de crecimiento.



(a) Temperatura promedio durante la etapa de germinación

(b) Temperatura promedio durante la etapa de rápido crecimiento

Figura 6.4: Correlación entre la temperatura promedio durante las etapas clave del desarrollo de cultivo. **a)** durante la etapa de germinación, en 2 de las 3 temporadas se encontró que conforme la temperatura promedio era mayor, el rendimiento asociado también, mientras que en la 2022/2023 la correlación fue opuesta, sugiriendo que conforme la temperatura promedio es mayor el rendimiento disminuirá, **b)** los datos obtenidos de la temporada 2022/2023 sugieren que conforme la temperatura se acerca a los 30 °C durante la etapa de crecimiento rápido el rendimiento es mayor.

6.2. VARIABLES DE PESO EN EL RENDIMIENTO DE LA PRODUCCIÓN DE LA CAÑA DE AZÚCAR

Se ha demostrado que la precipitación está directamente relacionada con el rendimiento, diversos autores han estudiado su efecto y medido su correlación, Santillán-Fernández et al. (2016) estudiaron la influencia de la sequía y la irrigación en la producción de caña de azúcar en diferentes regiones de México, mencionaron que la sequía es el mayor problema para la producción de caña de azúcar, concluyendo que la disminución de precipitación y los cambios en la distribución anual de la misma ha reducido la producción de caña de azúcar en México.

Etapa	21/22	22/23	23/24	Conjunto
Rápido crecimiento	-0.142	-0.218	0.366	-0.018

Tabla 6.8: Correlación entre la humedad relativa promedio durante la etapa de rápido crecimiento. Al medir la correlación en conjunto el valor resultante es muy cercano a 0, lo que sugiere una correlación nula, cuando se analiza de manera individual el comportamiento cambia dependiendo la temporada que se analiza, encontrando que las primeras 2 tienen un comportamiento similar, mientras que la última se comporta en dirección opuesta.

Variable	21/22	22/23	23/24	Conjunto
Precipitación acumulada	0.107	0.610	-0.349	0.471

Tabla 6.9: Correlación entre la precipitación acumulada y el rendimiento. Al medir la correlación en conjunto, el valor resultante indica una correlación moderada y positiva, lo que sugiere que a medida que la precipitación acumulada es mayor el rendimiento aumenta. Cuando se analiza de manera individual se puede apreciar la diferencia entre temporadas, donde la temporada 2022/2023 presenta la mayor correlación, seguida de la 2021/2022, que como se comentó, presenta una extensión mayor en la duración de la temporada, afectando directamente en los valores acumulados y la última temporada presenta un comportamiento opuesto.

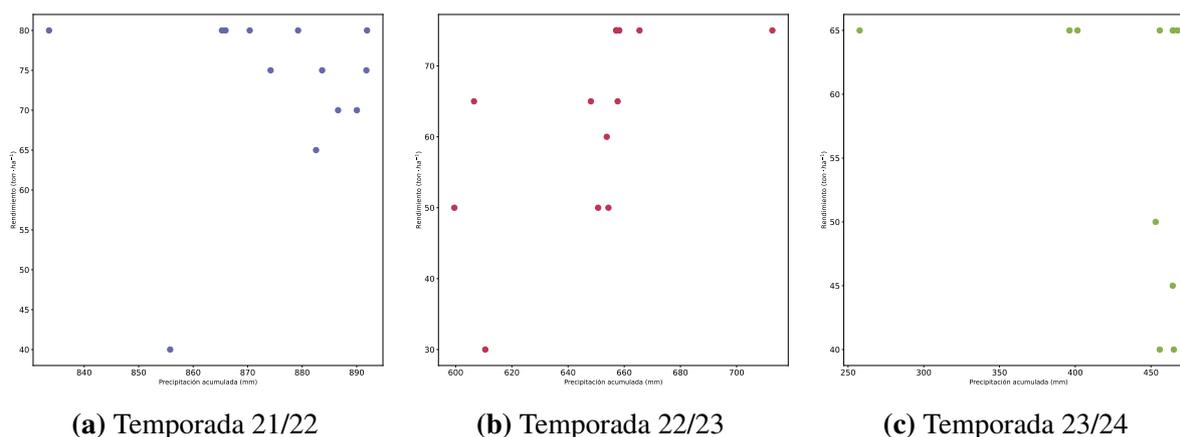


Figura 6.6: Correlación entre la precipitación acumulada y el rendimiento. Se puede apreciar que la correlación más marcada se encuentra en el conjunto de datos correspondientes a la temporada 2022/2023, también se puede distinguir la variación entre los valores acumulados entre las 3 temporadas, siendo la 2021/2022 las que presentan un mayor valor debido a la duración de su temporada, seguida de la 2022/2023 y por último la 2023/2024, con el menor valor de precipitación acumulada.

Tomando en cuenta lo anterior, se midió la correlación entre la precipitación acumulada (abarca desde el inicio de la temporada hasta la fecha de cosecha) y el rendimiento reportado

(tabla 6.9), se obtuvo que para la zafra 2021/2022 la correlación fue positiva y muy débil, con un valor de 0.107, lo que sugiere que no hay un efecto en el rendimiento, esto se puede deber a las inconsistencias con la duración de la temporada (tabla 6.3) y su efecto en la acumulación de la precipitación, en la 2022/2023 la correlación medida fue de 0.610 (figura 6.6), esta correlación fuerte y positiva sugiere que a medida que la precipitación acumulada aumenta el rendimiento también, corroborando lo mencionado por Santillán-Fernández et al. (2016), por el otro lado, para la zafra 2023/2024, la correlación fue similar pero en sentido opuesto, con un valor de -0.349. Al considerar las 3 temporadas la correlación fue de 0.472, positiva y fuerte. Es importante aclarar que únicamente se está considerando la precipitación, no las laminas de agua producto de los riegos realizados en cada parcela, de los cuales no se tienen datos y eso puede generar las variaciones mencionadas. Aunque la correlación al considerar las 3 temporadas sea buena es necesario analizar la razón de la significativa variación entre la correlación obtenida por temporada, una de las posibles razones puede ser la diferencia de agua irrigada por parcela y la distribución temporal de esta, la cantidad de días con lluvia durante la temporada, este último caso fue el que se analizó. La clasificación se realizó conforme a lo establecido por Servicio Meteorológico Nacional, el cual establece que se considera como lluvia cuando hay una precipitación acumulada de 0.1 a 25 milímetros en un periodo de 24 horas (la resolución temporal de los datos de precipitación con los que se cuentan), no se consideran el resto de clasificaciones ya que solo se busca distinguir entre días con o sin lluvia.

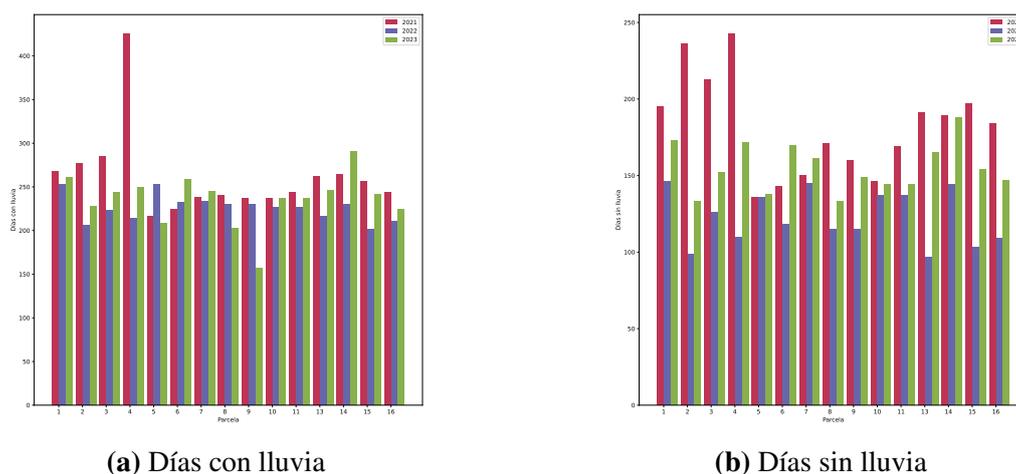


Figura 6.7: Días con lluvia por parcela. Los días con lluvia se consideran a partir de una precipitación acumulada mayor a 0.1 milímetros, menor a eso se considera como día sin lluvia, **a)** se puede apreciar como los días con lluvia correspondientes a la parcela numero 4 difiere radicalmente, esto debido a la duración excesiva de su temporada, que al no poder justificarse se optó por descartar para análisis posteriores, en el caso del resto de las parcelas, el comportamiento es similar con ligeras excepciones como la parcela 9 en la temporada 2023/2024, en general estuvieron expuestas a una cantidad de días con lluvia muy similares, **b)** igualmente destacan los datos correspondientes a la temporada 2021/2022, pero al no considerar esto debido a lo antes mencionado se puede apreciar como los días sin lluvia son mayores en la temporada 2023/2024, sugiriendo un déficit en la cantidad de precipitación durante esa temporada.

En la figura 6.7 se puede distinguir la diferencia entre días con y sin lluvia entre las parcelas de estudio durante cada una de las temporadas, encontrando variaciones llamativas resultado de anomalías mencionadas anteriormente. Se midió la correlación entre los días con y sin lluvia (tabla 6.10), encontrando que al considerar el conjunto de las 3 temporadas, la correlación entre los días de lluvia y sin lluvia tienen un comportamiento similar y en el mismo sentido

($\rho = 0.142$ y $\rho = 0.175$, respectivamente.), lo que sugiere una ligera influencia en el rendimiento. Al medir la correlación individualmente se encontró que varía en función de la temporada que se analiza, cambiando su comportamiento ligeramente o de forma totalmente opuesta, por lo que al momento de emplearlas en los modelos matemáticos futuros se debe tener en cuenta esto para seleccionar la variable que representa el correcto comportamiento de los datos.

Dato	21/22	22/23	23/24	Conjunto
Con lluvia	-0.338	0.510	-0.341	0.142
Sin lluvia	-0.394	0.434	-0.06	0.175

Tabla 6.10: Correlación entre los días de lluvia y el rendimiento.

6.3. Curva de desarrollo del cultivo de caña en las parcelas analizadas

A medida que el cultivo se desarrolla, varía progresivamente tanto la cobertura del suelo por la vegetación como la altura del cultivo y el área foliar. Estas variaciones afectan la evapotranspiración en las distintas fases del desarrollo, lo que provoca que el valor de K_c para un cultivo específico también se modifique a lo largo de su ciclo de crecimiento. Este período de crecimiento puede ser dividido en cuatro etapas: inicial, de desarrollo del cultivo, de mediados de temporada y de final de temporada (figura 6.8) (Allen et al., 2006). La etapa inicial abarca desde la siembra hasta el momento en que el cultivo alcanza aproximadamente el 10% de cobertura del suelo. La duración de la etapa inicial depende en gran medida del tipo de cultivo, la variedad del mismo, la fecha de siembra y del clima. El final de la etapa inicial ocurre cuando la vegetación verde cubre aproximadamente un 10% de la superficie del suelo. En el caso de cultivos permanentes, la fecha de siembra es reemplazada por el momento en que aparecen las primeras hojas. La etapa de desarrollo del cultivo va desde el momento en que la cobertura vegetal alcanza un 10% hasta el momento de alcanzar la cobertura efectiva total. La etapa de mediados de temporada comprende el período de tiempo entre la cobertura completa hasta el comienzo de la madurez. Durante la etapa de mediados de temporada, el coeficiente K_c alcanza su valor máximo.

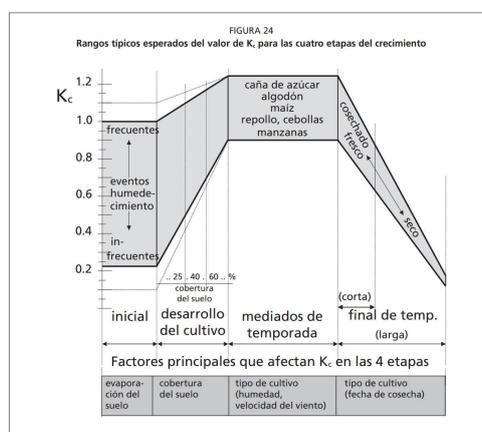
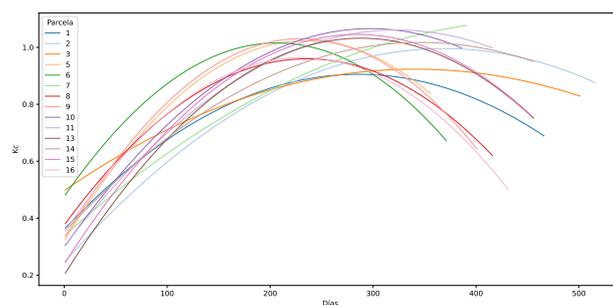


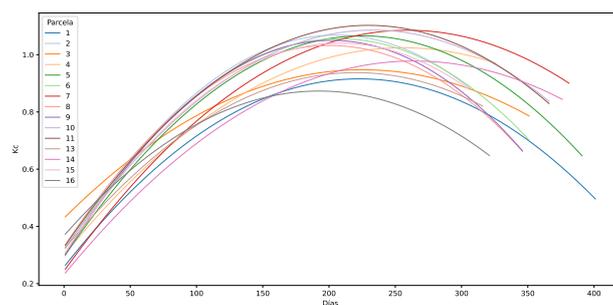
Figura 6.8: Rangos típicos esperados del valor de K_c para las cuatro etapas del crecimiento. Tomado de Allen et al. (2006).

La etapa final o tardía de crecimiento abarca desde el inicio de la madurez hasta el momento de la cosecha o la completa senescencia. Se asume que el cálculo de los valores de K_c y ET_c finaliza cuando el cultivo es cosechado, secado al natural, alcanza la completa senescencia o experimenta la caída de las hojas. Para algunos tipos de vegetación perenne en climas libres de heladas, los cultivos pueden desarrollarse durante todo el año, por lo que podría tomarse la fecha de término de la etapa final como la misma fecha de ‘siembra’ (Allen et al., 2006).

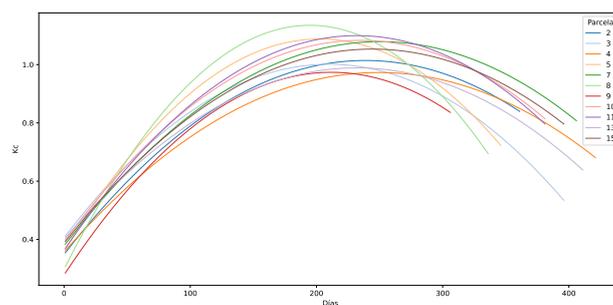
En la figura 6.9 se muestra la curva de desarrollo del cultivo, esta es muy similar a las series de tiempo del índice $NDVI$ ya que el cálculo del K_c se realiza a partir de este índice, como se indicó en la ecuación propuesta por Calera y González-Piqueras (2007) (ecuación).



(a) Temporada 2021/2022



(b) Temporada 2022/2023



(c) Temporada 2023/2024

Figura 6.9: Curva de desarrollo del cultivo a partir del coeficiente de cultivo único (K_c). **a)** se puede apreciar la mayor duración (en días) de la temporada, así como una mayor dispersión en las curvas de desarrollo y en la fase final de la misma, **b)** en este caso las curvas siguen un comportamiento más homogéneo y la duración de la misma es similar en la mayoría de las parcelas, **c)** se aprecia un inicio homogéneo y ligeras variaciones en el comportamiento de la curva, con una duración similar a la temporada anterior.

De acuerdo a la clasificación (tabla 6.11) establecida por Allen et al. (2006) para los distintas etapas del periodo de crecimiento del cultivo, el K_c toma 3 valores: la etapa inicial representada por ($K_{c\ ini}$), la etapa de mediados de temporada ($K_{c\ med}$) y la etapa final ($K_{c\ fin}$).

Cultivo	$K_c\ ini$	$K_c\ med$	$K_c\ fin$
Caña de azúcar	0.4	1.25	0.75

Tabla 6.11: Valores del coeficiente único del cultivo (K_c). Adaptado de Allen et al., 2006

A partir de lo representado en las figura 6.9 y tabla 6.11, es posible corroborar que la etapa inicial comienza con valores del K_c cercanos al 0.4, aumentando hasta acercarse al máximo establecido de 1.25 y comenzando a disminuir hasta bajar a 0.75 o menos, en esta gráfica se puede apreciar la diferencia de desarrollo entre las parcelas, específicamente el $K_c\ mid$, equivalente al valor máximo del $NDVI$, relacionado directamente con el rendimiento. Las etapas de crecimiento mencionadas anteriormente tienen una duración sugerida por FAO No. 24 y presentada en FAO No. 56 dependiendo el tipo de clima y de la ubicación en la que se encuentre el cultivo (tabla 6.12).

Cultivo	Inicial	Desarrollo	Medio	Final	Total	Región
Plantilla	35	60	190	120	405	Latitudes bajas
	50	70	220	140	480	Trópicos
	75	105	330	210	720	Hawai, EU
Soca	25	70	135	50	280	Latitudes bajas
	30	50	180	60	320	Trópicos
	35	105	210	70	420	Hawai, EU

Tabla 6.12: Duración de las etapas de crecimiento del cultivo. Adaptado de Allen et al., 2006

Se calculó la duración promedio de las etapas de crecimiento del cultivo a partir de las curvas de desarrollo, tomando como el inicio cuando el valor del coeficiente K_c supera el umbral de 0.4 hasta la fecha de cosecha, se obtuvo que para la zafra 2021/2022, la duración promedio fue de 425.86 días, para la zafra 2022/2023, 331 días y para la 2023/2024, 363.73. Con base a lo establecido en la tabla 6.12, la duración total corresponde a lo sugerido para caña de azúcar en su ciclo de soca, para la región de trópicos (Allen et al., 2006).

6.4. Modelos matemáticos de rendimiento

La mayoría de los acercamientos realizados para la estimación de producción de caña de azúcar utiliza análisis de regresión entre una variable predictora ($NDVI$, $GNDVI$, MSI , ETc acumulada) y la producción real (Abdel-Rahman & Ahmed, 2008). Se emplean regresiones lineales simples o múltiples para producir los modelos y métodos de validación, usan la raíz del error cuadrático medio (RMSE) y otras estadísticas (R^2) para evaluar la mejor variable y el modelo óptimo para la estimación de la producción (Som-ard et al., 2021). Tomando como guía lo realizado por los autores mencionados en la recopilación de estudios previos, se generaron modelos de regresión lineal simple y múltiple, empleando como variables predictoras las analizadas en secciones anteriores (tablas 6.4 y 6.5).

6.4.1. Regresión lineal simple

El objetivo del análisis lineal es encontrar un conjunto de variables predictoras (X_1 a X_n), en el que los cambios en cada variable predictora provoquen un cambio en la variable de respuesta

(Y) como un múltiplo del cambio en la variable predictora (Nisbet et al., 2009).

Los objetivos principales de la regresión lineal son

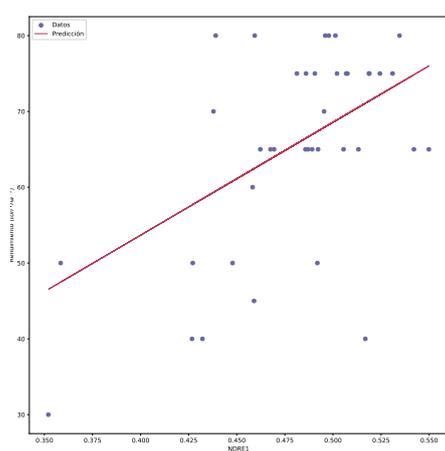
- Determinar si existe la relación entre una variable y otra
- Describir la naturaleza de esa relación, si existe
- Cuantificar la exactitud de esa relación
- Evaluar la contribución relativa de cada variable, si se usan múltiples variables

El proceso básico de la regresión lineal de un conjunto de datos es estimar los parámetros (coeficientes) por cada variable predictora (X) para representar el efecto que tiene en la variable de respuesta (Y) (Nisbet et al., 2009). Se define como

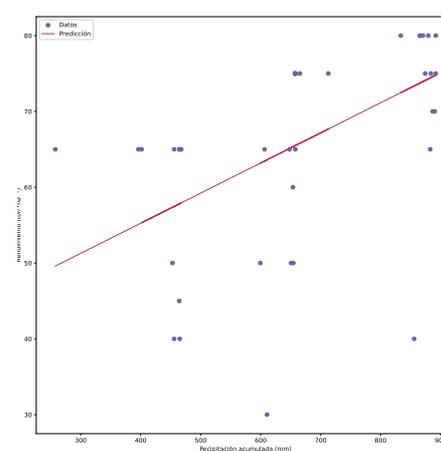
$$Y = a + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n$$

donde a es el valor de Y cuando $X = 0$ (intercepto), β_1 es el coeficiente para el factor X_1 , β_2 para el factor X_2 y así hasta la última variable X_n .

Los modelos de regresión lineal se generaron con el método de mínimos cuadrados ordinarios, por sus siglas en inglés, OLS, empleando la clase proporcionada por la biblioteca *statsmodels* (Seabold & Perktold, 2010) y dividiendo los datos en conjuntos de entrenamiento y prueba, en una proporción 80 y 20, respectivamente. Para determinar la variable adecuada para estimar el rendimiento, se comparó el resultado de los métodos de validación empleados anteriormente (R^2 , MSE y $RMSE$). Las variables a estudiar son las presentadas en las tablas 6.4 y 6.5, se analizó su desempeño de forma individual por temporada y en conjunto para visualizar las diferencias en la precisión de las predicciones. Los resultados obtenidos muestran que para el conjunto de datos reales que comprende las 3 temporadas (tabla 6.13), el valor máximo del $NDREI$ logra explicar mejor la variabilidad de los datos (figura 6.10a), en el caso de los modelados, el valor acumulado de la precipitación obtuvo el mejor coeficiente de determinación (R^2) (figura 6.10b).



(a) $NDREI$ máximo vs rendimiento



(b) Precipitación acumulada vs rendimiento

Figura 6.10: Modelos de regresión lineal simple por variable para la estimación del rendimiento. **a)** el valor máximo del índice $NDREI$ presente una relación lineal directa con el rendimiento, conforme una aumenta la otra también lo hace en proporción, **b)** indica que conforme aumenta la cantidad de precipitación acumulada el rendimiento será mayor.

Al analizarlo por temporada (tabla 6.14), para la 2021/2022, en el caso de los reales, el valor promedio del $CIRE$ obtuvo el mayor coeficiente de determinación (R^2), seguido del máximo

del *GNDVI*, destacando del resto de variables, en el caso de los modelados, el máximo del *CIRE* destaca significativamente del resto con un $R^2 = 0.9$. Para la temporada 2022/2023, el valor máximo del índice *CIRE* obtuvo el mejor R^2 junto al máximo del *NDRE*, para el conjunto de los datos reales, mientras que para los modelados, el valor máximo del *GNDVI* fue el que obtuvo el mejor coeficiente de determinación, siguiendo lo establecido por Rahman y Robson (2016) y Rahman y Robson (2020). En la temporada 2023/2024, el *MSI* promedio explicó mejor la variabilidad de los datos al considerar el conjunto de datos reales, para los modelados, los valores acumulados de la evapotranspiración del cultivo y precipitación obtuvieron los mejores coeficientes de determinación.

Datos	Variable	R^2	<i>MSE</i>	<i>RMSE</i>
Reales	max <i>NDREI</i>	0.272	1418.989	37.669
	max <i>NDVI</i>	0.257	1218.989	37.659
	max <i>NDRE</i>	0.234	1221.957	34.957
	max <i>CIRE</i>	0.229	1193.380	34.545
	max <i>GNDVI</i>	0.191	994.904	31.542
	mean <i>GNDVI</i>	0.165	859.977	29.325
	mean <i>NDRE</i>	0.144	750.012	27.386
	mean <i>CIRE</i>	0.136	712.575	27.386
	mean <i>NDREI</i>	0.111	579.646	24.076
	mean <i>NDVI</i>	0.096	499.965	22.360
	mean <i>NDMI</i>	0.079	413.761	20.341
	mean <i>MSI</i>	0.056	292.022	17.089
	Modelados	acum <i>LLUVIA</i>	0.309	1614.946
max <i>NDRE</i>		0.161	838.909	28.964
max <i>CIRE</i>		0.160	834.488	28.888
max <i>NDREI</i>		0.149	775.387	27.846
max <i>GNDVI</i>		0.137	717.842	26.793
max <i>NDVI</i>		0.125	653.248	25.559
mean <i>CIRE</i>		0.102	534.740	23.124
mean <i>GNDVI</i>		0.088	457.911	21.399
acum <i>ETC</i>		0.087	455.898	21.352
mean <i>NDRE</i>		0.085	441.798	21.019
mean <i>NDREI</i>		0.060	314.657	17.739
mean <i>NDMI</i>		0.040	206.806	14.381
mean <i>NDVI</i>		0.034	178.540	13.362
mean <i>MSI</i>	0.02	104.428	10.219	

Tabla 6.13: Métodos de validación por variable, considerando el conjunto de datos de las 3 temporadas, ordenadas de forma descendente. Se puede apreciar que conforme disminuye el coeficiente de determinación (R^2) el valor el *RMSE* también lo hace.

La tabla 6.13 presenta la diferencia de precisión entre los conjuntos de variables empleadas para estimar el rendimiento, al emplear el conjunto de las 3 temporadas se encontraron resultados esperados y otros inconsistentes, por lo que analizar con mayor atención el comportamiento por temporada permitió identificar las variables que mejor se adaptan al objetivo. En la tabla 6.14 se presentan las métricas de desempeño desglosadas por temporada con la finalidad de distinguir los cambios que hay de una temporada a otra. En la 2021/2022 las variables en general tienen un buen desempeño cuando se considera el conjunto de datos reales ($R^2 > 0.5$), destacando el valor promedio del *CIRE* y el máximo del *GNDVI*, en el caso de los modelados, algunas variables aumentan o disminuyen su desempeño ligeramente pero siguen el comportamiento de su contraparte real, a excepción de las variables acumuladas (*LLuvia* y *ETc*), estas se acercan a

un $R^2 = 0$ debido a lo mencionado anteriormente, la duración de la temporada se prolongó y su comportamiento no influye directamente al rendimiento. La temporada siguiente, 2022/2023, los valores máximos de los variables tuvieron mejores desempeños, tanto si se considera el conjunto de datos reales o modelados. La última temporada, 2023/2024, casi la totalidad de las variables no pueden explicar la variabilidad de los datos en ningún sentido ni magnitud a excepción de las que se enfocan en el estrés hídrico y los valores acumulados de precipitación y evapotranspiración del cultivo.

Datos	Variable	21/22		22/23		23/24	
		R^2	RMSE	R^2	RMSE	R^2	RMSE
Reales	max <i>NDVI</i>	0.729	31.496	0.710	41.861	0.053	7.179
	mean <i>NDVI</i>	0.610	28.794	0.554	36.958	0.143	11.783
	max <i>GNDVI</i>	0.813	33.253	0.671	40.668	0.168	12.759
	mean <i>GNDVI</i>	0.503	26.155	0.598	38.392	0.017	4.063
	max <i>NDREI</i>	0.736	20.866	0.690	41.255	0.136	11.480
	mean <i>NDREI</i>	0.775	27.620	0.609	38.754	0.074	8.485
	max <i>CIRE</i>	0.848	33.963	0.740	42.728	0.120	10.802
	mean <i>CIRE</i>	0.733	31.566	0.685	41.115	0.006	2.481
	max <i>NDRE</i>	0.763	32.222	0.727	42.336	0.169	12.810
	mean <i>NDRE</i>	0.759	32.132	0.637	39.634	0.039	6.129
	mean <i>MSI</i>	0.414	23.734	0.483	34.532	0.267	16.112
	Modelados	max <i>NDVI</i>	0.635	29.393	0.760	43.290	0.006
mean <i>NDVI</i>		0.411	23.645	0.627	39.332	0.022	4.669
max <i>GNDVI</i>		0.628	29.218	0.826	45.125	0.008	2.737
mean <i>GNDVI</i>		0.455	24.866	0.711	41.883	0.002	1.350
max <i>NDREI</i>		0.795	32.884	0.757	43.219	0.008	2.872
mean <i>NDREI</i>		0.644	29.590	0.644	39.858	0.001	1.132
max <i>CIRE</i>		0.900	34.980	0.783	43.949	0.036	5.894
mean <i>CIRE</i>		0.797	32.923	0.705	41.705	0.012	3.347
max <i>NDRE</i>		0.841	33.813	0.777	43.789	0.016	3.947
mean <i>NDRE</i>		0.651	29.751	0.686	41.122	0.000	0.144
mean <i>MSI</i>		0.244	18.221	0.580	37.840	0.107	10.174
acum <i>LLUVIA</i>		0.002	1.654	0.380	30.607	0.199	13.890
acum <i>ETC</i>	0.011	3.921	0.533	36.268	0.203	14.033	

Tabla 6.14: Métricas de validación por temporada. Permite apreciar el cambio en como las variables analizadas pueden explicar la variabilidad de los datos y estimar el rendimiento de una temporada a otra

Los datos correspondientes a la temporada 2023/2024 presentaron irregularidades en comparación con los de las temporadas anteriores, provocando que el desempeño de los modelos de regresión lineal disminuya al considerar el conjunto de datos de las 3 temporadas. Por lo que para tener una perspectiva distinta se optó por realizar el análisis considerando únicamente los datos de la temporadas 2021/2022 y 2022/2023 (figura 6.11), con la finalidad de encontrar las diferencias en el desempeño de los modelos cuando se usan datos de mejor calidad. De los datos reales se encontró un desempeño similar en casi la totalidad de las variables, destacando ligeramente el valor promedio del *CIRE* y el máximo del *NDRE*, en los modelados, los valores máximos de los VI's obtuvieron mejores R^2 , destacando el máximo del *CIRE* y *NDRE* (tabla 6.15). Se encontró que al incluir solo los datos de las 2 primeras temporadas los modelos resultantes pueden describir mejor la variabilidad de los datos sin que estos estén condicionados o sean válidos solo para una temporada particular, al no incluir los datos de la 2023/2024 las inconsistencias de esta no influyen en los modelos generados y permiten estimar el rendimiento en las condiciones normales de una temporada.

Datos	Variable	R^2	MSE	RMSE
Reales	mean <i>CIRE</i>	0.622	2481.72	49.817
	max <i>NDRE</i>	0.605	2412.569	49.118
	mean <i>NDRE</i>	0.594	2369.656	48.679
	max <i>CIRE</i>	0.587	2339.822	48.372
	max <i>NDVI</i>	0.567	2262.787	47.569
	max <i>NDREI</i>	0.560	2234.697	47.273
	mean <i>NDREI</i>	0.557	2221.821	47.136
	mean <i>GNDVI</i>	0.552	2200.772	46.912
	max <i>GNDVI</i>	0.536	2137.373	46.232
	mean <i>NDVI</i>	0.511	2036.587	45.129
	mean <i>NDMI</i>	0.509	2031.286	45.070
	mean <i>MSI</i>	0.407	811.249	28.482
	Modelados	max <i>CIRE</i>	0.613	2444.538
max <i>NDRE</i>		0.598	2384.557	48.832
max <i>NDREI</i>		0.570	2273.241	47.679
max <i>GNDVI</i>		0.554	2209.642	47.007
max <i>NDVI</i>		0.540	2154.575	46.417
mean <i>CIRE</i>		0.523	2083.548	45.646
mean <i>NDRE</i>		0.482	1922.400	43.845
mean <i>GNDVI</i>		0.466	1860.125	43.129
mean <i>NDREI</i>		0.429	1709.401	41.345
mean <i>NDMI</i>		0.403	1607.922	40.099
mean <i>NDVI</i>		0.373	1485.400	38.541
mean <i>MSI</i>		0.362	721.384	26.859
acum <i>ETC</i>		0.194	772.307	27.790
acum <i>LLUVIA</i>		0.156	621.603	24.932

Tabla 6.15: Métodos de validación por variable, considerando el conjunto de datos de las 2 primeras temporadas. Ordenados por coeficiente de determinación de forma descendente.

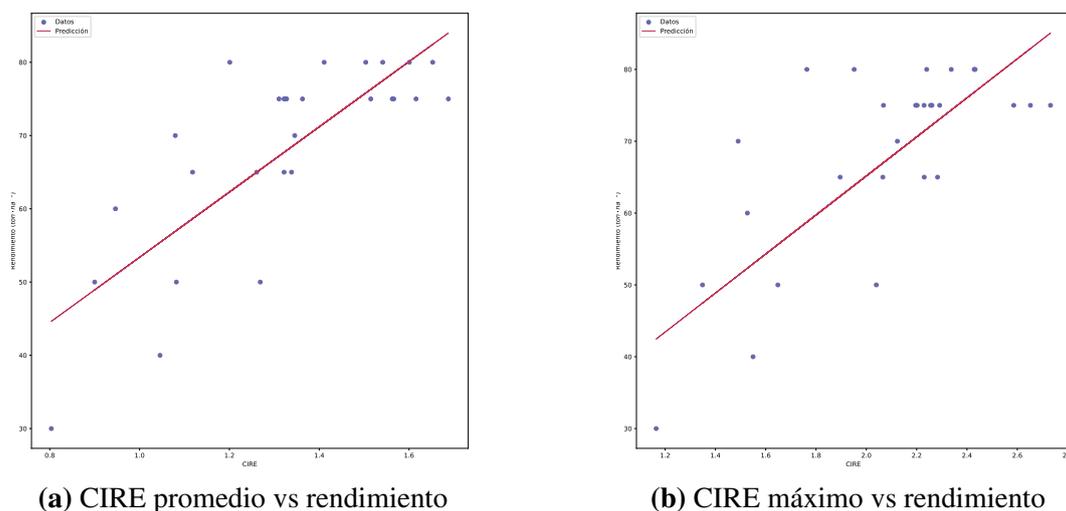


Figura 6.11: Modelos de regresión lineal simple de las mejores variables considerando el conjunto de datos de las primeras 2 temporadas. **a)** el valor promedio del índice *CIRE* real presenta una relación lineal directa con el rendimiento, conforme una aumenta la otra también lo hace en proporción, **b)** mismo caso para el valor máximo del *CIRE* modelado.

Como se pudo apreciar en las métricas de validación de los modelos generados, además de la relación de la variable a analizar, los datos empleados tienen un papel crucial en la precisión

con la que pueden estimar el rendimiento, ya que si estos corresponden a una temporada que presenta anomalías como las presentes en la 2023/2024, estos pueden mermar el desempeño de los modelos y no explicar correctamente la influencia de la variable predictora con la de respuesta, de igual forma usar datos correspondientes a una única temporada pueden no explicar el comportamiento real de los datos aunque la precisión sea buena, por lo que ampliar los datos a mas de una temporada como se hizo anteriormente permite eliminar la especificidad de una sola temporada y no contemplar las anomalías propias de temporadas como la mencionada, esto permite obtener modelos con una buena precisión y que explican correctamente el comportamiento de los datos.

6.4.2. Regresión lineal múltiple

A diferencia de lo realizado en la sección anterior, en esta se analizará la relación entre 2 o mas variables predictoras con la variable de respuesta. Akbarian et al. (2022) analizaron la correlación (ρ) entre 23 VI's y el rendimiento de la caña de azúcar, reduciendo a 11 los VI's a incluir para la construcción de modelos predictivos, calculando el *AIC* para elegir las mejores variables predictoras, finalmente calcularon el *MSE* resultante de los modelos generados a partir de las variables empleadas para validar sus desempeños. En este estudio se generaron modelos de regresión lineal múltiple empleando como datos de entrada los conjuntos usados en la sección anterior, que de la misma forma fueron divididos en reales y modelados, también se distinguió entre el valor máximo que alcanza la variable y su promedio. La cantidad de variables incluidas por modelo se determinó de acuerdo a la cantidad de variables disponibles, en el caso de las máximas se cuenta con 5 y de las promedio con 7 (tablas 6.4 y 6.5), por lo que al realizar todas las combinaciones posible entre variables, se generaron modelos que van desde las 2 hasta las 7 variables, aplicado a los conjuntos de datos da como resultado 1460 modelos distintos, debido a la gran cantidad de opciones se optó por solo considerar los que mejores métricas de validación obtuvieron, esto se llevó a cabo comparando el coeficiente de determinación (R^2) por cada conjunto de datos, tipos (reales o modelados) y la cantidad de variables, ej. para el conjunto correspondiente a la temporada 2022/2023, en su tipo de variable máxima, se compararon todas las combinaciones conformadas por 2 variables y se eligió la que mejor R^2 obtiene, esto permitió reducir a 100 los modelos a considerar en el análisis. Para medir el grado de presencia o efecto de las variables en este conjunto de los 100 mejores modelos (tabla 6.16), se realizó el conteo de las ocasiones que se emplean una variable en alguno de los 100 modelos, encontrando que el valor promedio del *NDRE* se encuentra en 46 de los 100 mejores modelos.

Variabes	Contador
ndvi max	22
gndvi max	33
cire max	23
ndre max	31
ndre1 max	31
ndvi mean	31
gndvi mean	36
cire mean	38
ndre mean	46
ndre1 mean	36
ndmi mean	40
msi mean	43

Tabla 6.16: Conteo de combinaciones en las que está presente una variable.

Se encontró que conforme se incluyan más variables en los modelos la precisión de estos al estimar el rendimiento aumentará (menor $RMSE$), mientras que los coeficientes de determinación mantienen un comportamiento similar y una complejidad AIC muy parecida entre ellos. Se presenta como ejemplo los mejores modelos para el conjunto de datos que comprenden las temporadas 2021/2022 y 2022/2023, en las tablas 6.17 y 6.18 se puede visualizar como los resultados al emplear los datos reales y modelados son muy similares, también el AIC presenta un comportamiento adecuado con una desviación estándar $\sigma = 0.005$ promedio para el caso de los modelos que incluyen las variables máximas y $\sigma = 0.031$ para las variables promedio.

	Combinación	Variables	R^2	AIC	MSE	$RMSE$
Máximos	1	5	0.633	166.776	504.837	22.469
	3	4	0.630	164.936	628.379	25.067
	6	3	0.627	163.136	833.365	28.868
	10	2	0.622	161.451	1239.311	35.204
Promedio	1	7	0.711	165.535	404.913	20.122
	7	6	0.711	163.554	472.236	21.731
	19	5	0.710	161.612	566.067	23.792
	29	4	0.705	159.967	702.878	26.512
	25	3	0.684	159.456	909.721	30.162
	21	2	0.655	159.415	1306.010	36.139

Tabla 6.17: Comparación de las métricas de validación de los modelos de regresión lineal múltiple generados a partir de los datos reales, usando el conjunto de datos correspondientes a las primeras 2 temporadas. Conforme se emplean más variables la capacidad de explicar la variabilidad de los datos aumenta, sin que su complejidad aumente drásticamente, teniendo una diferencia en el AIC pequeña.

	Combinación	Variables	R^2	AIC	MSE	$RMSE$
Máximos	1	5	0.635	166.683	506.079	22.496
	5	4	0.634	164.739	631.661	25.133
	9	3	0.631	162.911	838.406	28.955
	8	2	0.623	161.356	1242.568	35.250
Promedios	1	7	0.697	166.551	397.130	19.928
	7	6	0.697	164.551	463.317	21.525
	21	5	0.697	162.564	555.836	23.576
	35	4	0.680	161.751	678.046	26.039
	27	3	0.657	161.279	873.495	29.555
	12	2	0.596	162.906	1187.724	34.463

Tabla 6.18: Comparación de las métricas de validación de los modelos de regresión lineal múltiple generados a partir de los datos modelados, usando el conjunto de datos correspondientes a las primeras 2 temporadas. Conforme se emplean más variables la capacidad de explicar la variabilidad de los datos aumenta, sin que su complejidad aumente drásticamente, teniendo una diferencia en el AIC pequeña.

Cuando se analiza el desempeño de los modelos generados a partir de los datos de temporadas individuales, las métricas de validación tienden a mejorar sustancialmente en comparación con los presentados en las tablas 6.17 y 6.18, caso contrario cuando se analizan los correspondientes al conjunto de las 3 temporadas, donde estos valores disminuyen abruptamente (tabla 6.19), al igual que como se presentó en la sección anterior.

Datos		21/22		22/23		23/24		Conjunto	
		R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE
Reales	max	0.919	19.639	0.900	26.142	0.758	14.827	0.333	23.181
	mean	0.878	17.373	0.885	23.388	0.946	15.132	0.303	19.945
Modelados	max	0.965	20.203	0.840	25.378	0.705	14.388	0.199	17.932
	mean	0.926	17.783	0.820	22.478	0.952	15.181	0.287	19.335

Tabla 6.19: Promedios de las métricas de validación por temporada. Se calcularon tomando en cuenta todos los modelos generados por cada conjunto de datos, por esta razón la precisión se incrementa significativamente debido a la especificidad de los datos y a su propio comportamiento, presentando buenos valores incluso para la temporada 2023/2024, que como se mencionó antes presenta anomalías pero al emplear hasta 7 variables en algunos modelos permite obtener precisiones altas, pero al considerar los 3 conjuntos el desempeño disminuye.

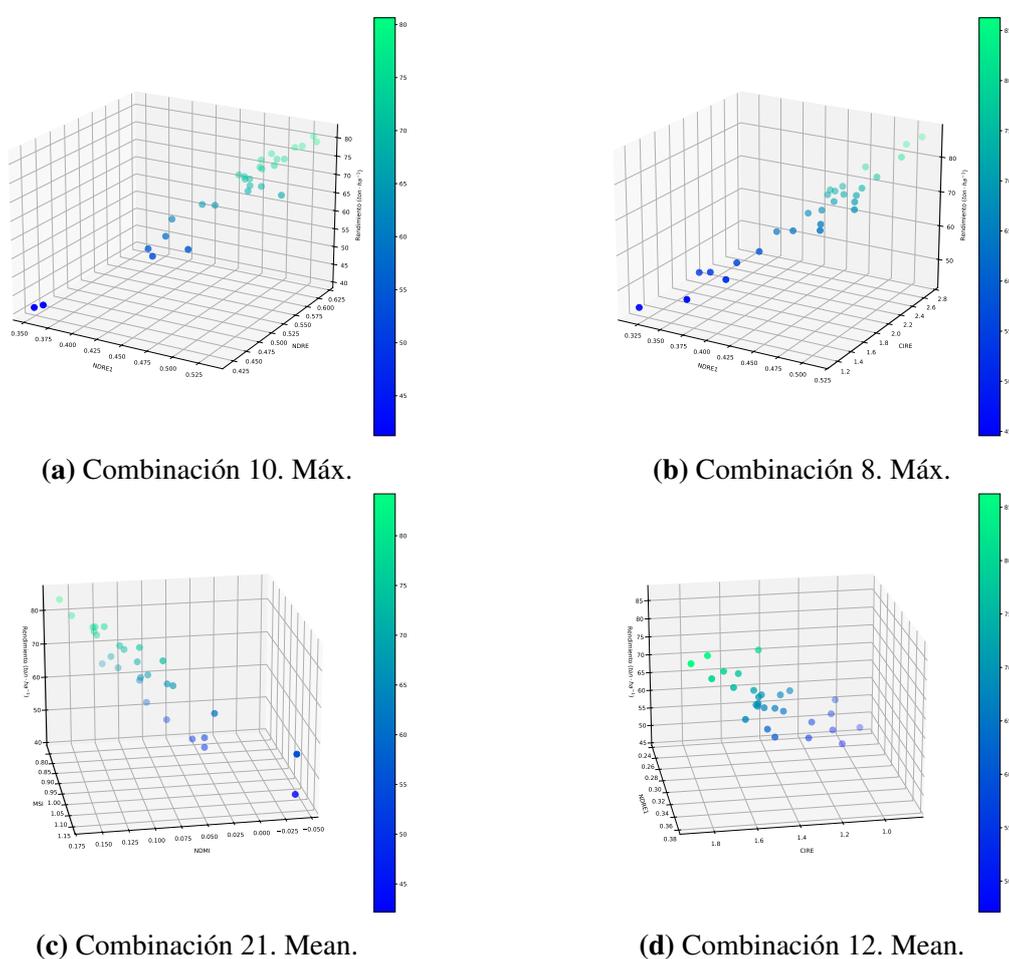


Figura 6.12: Dispersión en 3D de los rendimientos estimados por los modelos con las mejores combinaciones de 2 variables por tipo de datos, reales y modelados, máximos y promedios, para el conjunto de datos que contempla las primeras 2 temporadas. **a)** compuesto por los valores máximos del $NDRE$ y $NDREI$ reales, indica correctamente que conforme el valor de estas variables sea mayor el rendimiento también lo será, **b)** conforme el valor máximo modelado de los índices $NDREI$ y $CIRE$ aumenta, el rendimiento también, **c)** emplea los promedios reales del MSI y $NDMI$, cuando el valor del MSI disminuye y $NDMI$ aumenta, el rendimiento es mayor, lo que se puede describir que a menor estrés hídrico y mayor cantidad de agua en el cultivo, el rendimiento aumenta, **d)** $CIRE$ y $NDREI$, conforme aumentan el rendimiento también.

Debido a la cantidad de variables que un modelo puede emplear (máximo 7 en el caso de los datos promedio), la representación visual se ve limitada únicamente a los modelos que involucran 2 variables independientes para estimar una variable dependiente Y . Por lo que gráficos de dispersión en 3 dimensiones que representen el comportamiento de las variables independientes son los adecuados para una correcta visualización. En la figura 6.12 se presenta el comportamiento de los datos generados por los modelos que emplean 2 variables predictoras como entrada para el conjunto de datos de las temporadas 2021/2022 y 2022/2023 de acuerdo a la combinación de variables que obtiene las mejores métricas de validación.

Una de las ventajas de distinguir el conjunto de variables entre máximas y promedios es la generación de mapas de rendimiento a partir de los modelos que emplean los valores máximos de las variables como datos de entrada, estos modelos (figuras 6.12a y 6.12b) se representan mediante sus ecuaciones y se definen como:

$$Y = 414.844 (ndre_{max}) - 240.258 (ndre1_{max}) - 48.111 \quad (6.1)$$

$$Y = 47.801 (cire_{max}) - 171.545 (ndre1_{max}) - 43.395 \quad (6.2)$$

Adicionalmente se obtuvieron los modelos que emplean el valor promedio de las variables para estimar el rendimiento, útil cuando se tienen las series de tiempo que abarcan el ciclo de crecimiento del cultivo, estos modelos son:

$$Y = 757.389 (ndmi_{mean}) + 327.423 (msi_{mean}) - 293.399 \quad (6.3)$$

$$Y = 110.377 (cire_{mean}) - 538.806 (ndre1_{mean}) + 81.099 \quad (6.4)$$

6.5. Mapas de rendimiento basados en índices espectrales

La predicción, modelado y el mapeo del rendimiento de los cultivos basados en percepción remota tiene un gran potencial para una multitud de aplicaciones y las partes interesadas en los procesos agrícolas (Perich et al., 2023). Los mapas de rendimiento proveen información esencial para guiar las prácticas de la agricultura de precisión (Canata et al., 2021). Para los agricultores, los responsables de la formulación de políticas, los seguros de cosechas y las organizaciones no gubernamentales, es de gran interés anticipar el rendimiento de los cultivos (Weiss et al., 2020).

En este estudio se ha presentado una metodología para estimar el rendimiento del cultivo a través de la regresión lineal simple/múltiple de variables predictoras y el rendimiento reportado, siendo el valor máximo de algunos VI's opciones a considerar debido a su buena correlación y capacidad de explicar la variabilidad de los datos, tal como lo sugirieron Rahman y Robson (2020) y Rahman y Robson (2016), otro enfoque fue el desarrollado por Canata et al. (2021), que encontraron la etapa de desarrollo que mejor correlación obtenía con el rendimiento reportado y aplicaron técnicas de aprendizaje automático como lo es el algoritmo de Bosque Aleatorio y regresión múltiple, emplearon la imagen satelital asociada a esa etapa para representar el mapa de rendimiento.

La elaboración de los mapas de rendimiento se llevó a cabo utilizando los modelos de regresión lineal generados en los apartados 6.4.1 y 6.4.2, empleando los valores máximos de los índices presentados en la tabla 6.1 y la imagen asociada a esa fecha, del conjunto de datos de las primeras 2 temporadas, como se presentó anteriormente, existe una variación entre el máximo real y el modelado, debido a que la resolución temporal de la misión Sentinel-2 es

de 5 días en las condiciones ideales y en muchas ocasiones no es posible obtener la imagen correspondiente a la fecha óptima para estimar el rendimiento, una alternativa para generar mapas de rendimiento con una precisión mayor es emplear las imágenes de otra misión espacial que correspondan al valor máximo modelado, en este caso se pudo contar con las imágenes de la constelación *PlanetScope*, distribuida por la empresa Planet Labs. Las imágenes empleadas fueron las adquiridas por los satélites que cuentan con el sensor PSD.SD (también conocido como *Super Dove*), el cual genera imágenes diarias con una resolución espacial de 3 metros y espectral de 8 bandas que van desde los aerosoles hasta el infrarrojo cercano, las bandas del Super Dove son interoperables con las de Sentinel 2, por lo que los modelos generados a partir de los datos de Sentinel son válidos para aplicarlos en las imágenes generadas por el Super Dove y generar los mapas de rendimiento, permitiendo obtener mapas correspondientes a los datos reales (resolución temporal de Sentinel) y a los modelados, con una mayor resolución espacial.

Partiendo de la diversidad de modelos y conjuntos de datos, se comenzó por comparar los modelos generados a partir de la regresión lineal simple (tabla 6.15), donde la variable máxima independiente que destacó fue el *NDRE*, el modelo generado por el conjunto de datos reales de esa variable se aplicó a la imagen de Sentinel 2 asociada a esa fecha (figura 6.13a) y el modelo generado a partir de los datos modelados se aplicó a la imagen de Planet (figura 6.13b), la precisión de ambos modelos es casi igual ($R^2 = 0.605$ y $R^2 = 0.598$, respectivamente) y permiten identificar las zonas de variación en el rendimiento y la calidad del cultivo, con la diferencia que en las imágenes de Planet se puede apreciar con más detalle esas variaciones debido a su gran resolución espacial.

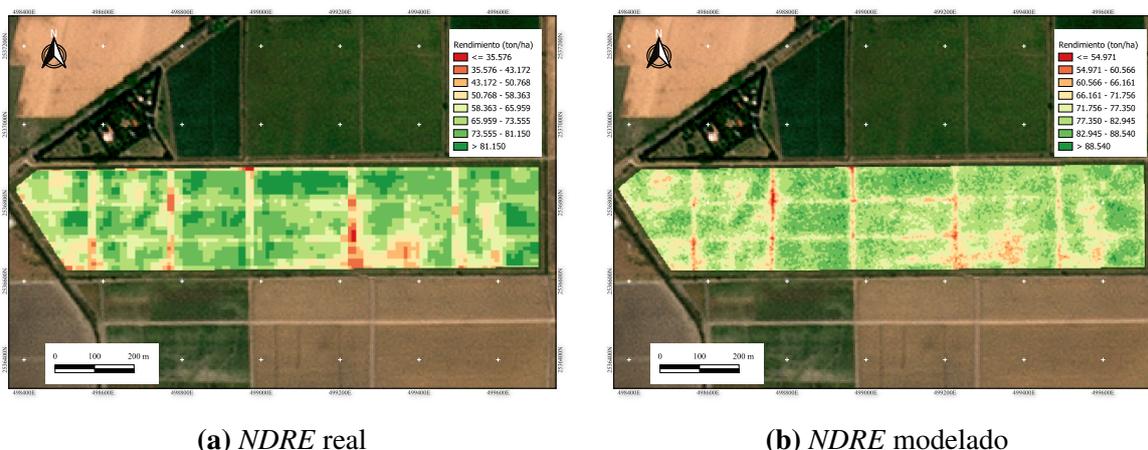


Figura 6.13: Mapas de rendimiento de la parcela No. 5 en la temporada 2022/2023 generados a partir de los modelos de regresión lineal simple empleando el valor máximo del índice *NDRE*, real y modelado, con el conjunto de datos de las primeras dos temporadas. **a)** corresponde a la imagen de Sentinel 2 asociada a la fecha donde el *NDRE* alcanza su valor máximo en la serie de tiempo real, **b)** corresponde a la imagen del Super Dove asociada a la fecha en la que el *NDRE* alcanza su valor máximo en la serie de tiempo modelada.

Continuando con los modelos de regresión simple, se compararon los mapas de rendimiento generados a partir de los modelos que emplean la variable más usada para esta aplicación, el *NDVI*, en su valor máximo para el conjunto de datos reales y modelados (figura 6.14). La figura 6.14a logra representar la variación de rendimiento en las distintas zonas de la parcela de forma adecuada, similar a lo generado a partir del modelo que emplean los datos modelados (figura 6.14c), se comparó el mapa de rendimiento aplicado a la imagen de Planet en la misma fecha del valor máximo real registrado por Sentinel y se encontraron ligeras diferencias en el

rendimiento estimado (figura 6.14b).

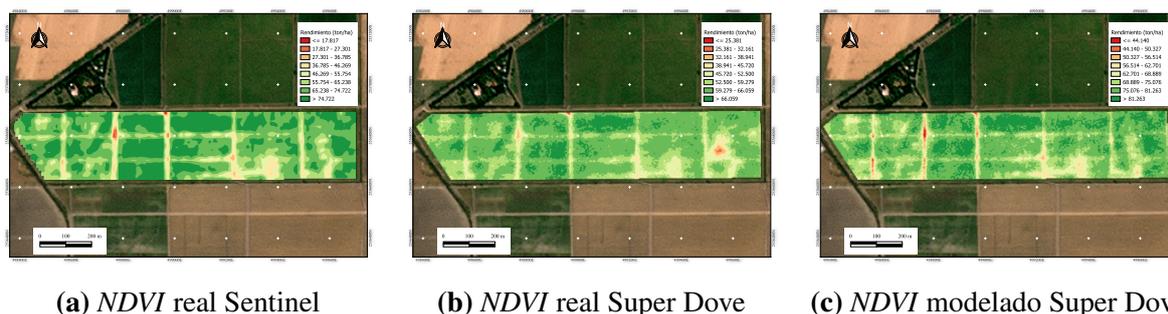


Figura 6.14: Mapas de rendimiento para la parcela 5 en la temporada 2022/2023 generados a partir de los modelos de regresión lineal simple empleando el valor máximo del índice *NDVI*, real y modelado, con el conjunto de datos de las primeras dos temporadas. **a)** y **b)** corresponden a la fecha donde el valor real del *NDVI* alcanza su valor máximo y **c)** el valor modelado en la fecha estimada. Se pueden apreciar ligeras diferencias entre las estimaciones pero en general comparten las zonas con rendimiento medio, permitiendo distinguir entre la zona cultivada y los pasillos entre bloques.

Para emplear los modelos de regresión lineal múltiple generados en la sección anterior es necesario contar con las 2 imágenes correspondientes a la fecha en la que el índice llega a su máximo, el primer modelo empleado fue el representado por la ecuación (6.1), el mapa generado (figura 6.15) presenta similitudes con sus contrapartes generadas por medio de modelos de regresión simple (figuras 6.13a y 6.13b) y con una precisión en la identificación de la variación en el rendimiento parecida al mostrar zonas de mayor calidad de forma acertada.

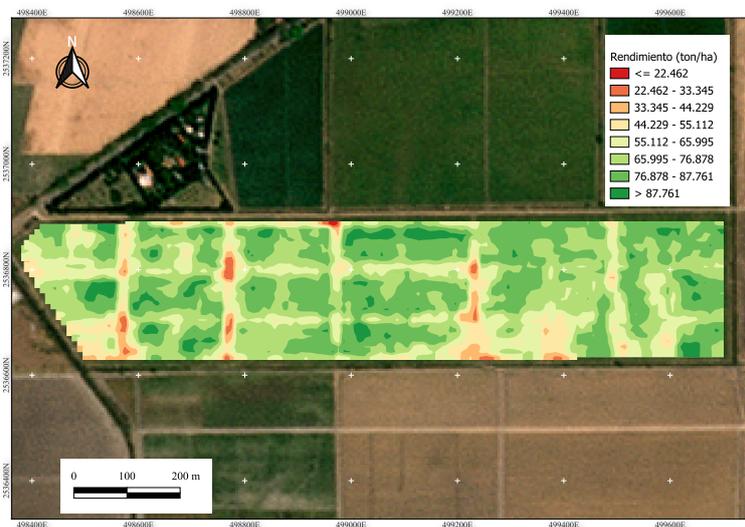


Figura 6.15: Mapa de rendimiento para la parcela 5 en la temporada 2022/2023 generado a partir del modelo de regresión múltiple que emplea los valores reales máximos de los índices *NDRE* y *NDRE1*. Se aprecia una mayor distinción entre los distintos valores de rendimiento estimados, detallando las zonas de media/alta productividad y resaltando las partes de los pasillos como zonas de muy bajo rendimiento

Con la finalidad de ampliar la representación de las ventajas de la generación de mapas de rendimiento a partir de modelos de regresión múltiple, se aplicó el modelo y se generó el mapa de rendimiento (figura 6.16) basado en la combinación de variables No. 6 (tabla 6.17), el cual

cuenta con 3 variables independientes y aunque su capacidad de explicar la variabilidad de los datos es similar al modelo anterior, su precisión (*RMSE*) al estimar el rendimiento es superior (22 % mayor), el modelo se define como

$$Y = 401.145 (ndre_{max}) - 327.1379 (ndre1_{max}) + 111.324 (ndvi_{max}) - 84.882 \quad (6.5)$$

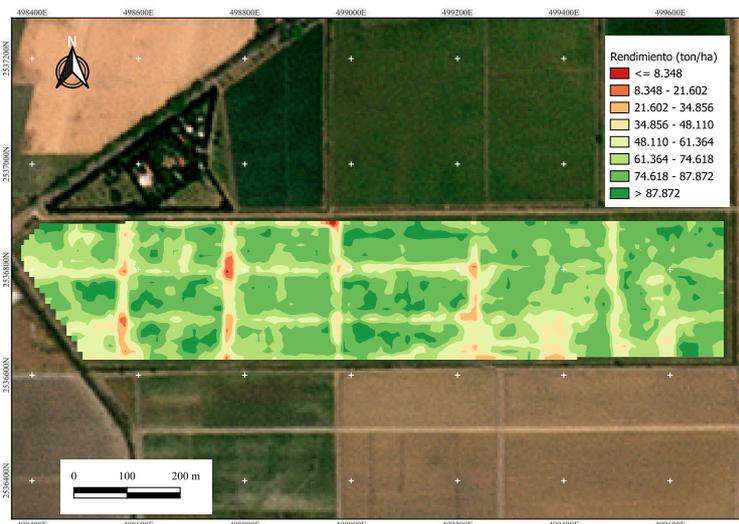


Figura 6.16: Mapa de rendimiento de la parcela 5 en la temporada 2022/2023 generado a partir del modelo de regresión múltiple que emplea los valores reales máximos de los índices *NDRE*, *NDRE1* y *NDVI*.

Para comparar la precisión de los modelos al estimar el rendimiento se obtuvo una muestra de 344 de puntos aleatorios contenidos dentro de un buffer de 10 metros de la geometría de la parcela, el equivalente aproximado al 10 % del total de píxeles de las imágenes y se calculó el promedio de los valores de pixel asociados a esos puntos, se encontró que para el modelo que emplea 3 variables máximas reales el rendimiento promedio fue de 72.266 t ha^{-1} (figura 6.16), en el caso del modelo de regresión simple basado en el *NDVI* el promedio es de 68.981 t ha^{-1} (figura 6.14a), el rendimiento promedio resultante del modelo de regresión múltiple que emplea las variables *NDRE* y *NDRE1* (figura 6.15) es de 71.885 t ha^{-1} , rendimientos cercanos a los declarados para esa de esa parcela para la zafra 2022/2023 que fue de 75 t ha^{-1} .

Como se mencionó al inicio de la sección, los mapas de rendimiento son una herramienta útil para visualizar la productividad de la parcela y proveen información espacial para la toma de decisiones, al emplear el valor máximo de un índice espectral permite generar estos mapas meses previos a la cosecha, lo que proporciona una pronóstico que puede tener múltiples aplicaciones y permite a las partes involucradas tomar acciones a partir de esta información.

CAPÍTULO 7

CONCLUSIONES

La metodología desarrollada en este estudio permite estimar el rendimiento del cultivo de la caña de azúcar con una precisión que depende de la calidad y cantidad de los datos que se emplean, ya que al contar con únicamente 15 parcelas y que 7 de esas estuvieran concentradas en la misma zona no permitió analizar la variabilidad del cultivo a lo largo de toda la zona de estudio, por lo que en un futuro trabajo, incluir un mayor número de parcelas con una mayor distribución espacial y que se tenga mejor control de sus datos será de gran utilidad para que las estimaciones sean más cercanas al rendimiento reportado. La generación de series de tiempo de buena calidad a partir en los índices de vegetación empleados en el estudio permitieron identificar apropiadamente las distintas métricas fenológicas propias del cultivo y por consecuencia distinguir variables como el momento de máximo desarrollo del mismo, el cual tiene una correlación directa con el rendimiento y sirve como variable de entrada en los modelos matemáticos dedicados a estimar el rendimiento y generar los mapas de rendimiento que permiten al productor o cualquier otro ente involucrado en el proceso productivo del cultivo, tener una perspectiva de la variación en la productividad de la parcela. Las inconsistencias señaladas durante el desarrollo de este estudio son la principal limitante encontrada, la discrepancia entre los datos de correlación encontrados de una temporada a otra, la diferencia en la duración de cada temporada entre parcelas, la baja resolución espacial de los datos agrometeorológicos no permitió distinguir con precisión la variación de sus efectos puntuales en el rendimiento de cada parcela, a pesar de que se encontró una correlación entre variables como la temperatura y humedad relativa durante etapas cruciales en el desarrollo del cultivo, estas no tienen la capacidad de explicar por sí mismas la variabilidad en el rendimiento ni la suficiente influencia para ser consideradas en los modelos de rendimiento, similar a lo sugerido por los autores citados. Por lo que se puede concluir que se cumplieron la totalidad de los objetivos propuestos al desarrollar una metodología capaz de estimar el rendimiento del cultivo de la caña de azúcar, desarrollada para la región sur de Tamaulipas y con la capacidad de ser replicada en otra zona.

APÉNDICE A

VARIABLES METEOROLÓGICAS

Descripción de las variables meteorológicas presentes en el dataset AgERA5.

Variable	Unidad	Descripción
10m wind speed	m s^{-1}	Velocidad promedio del viento a una altura de 10 metros sobre la superficie durante un periodo de 24 horas
2m dewpoint temperature	K	La temperatura media de punto de rocío a una altura de 2 metros sobre la superficie durante el período de tiempo local de 00h a 24h.
2m relative humidity	%	Humedad relativa a una altura de 2 metros sobre la superficie.
2m temperature	K	Temperatura del aire a una altura de 2 metros sobre la superficie.
Cloud cover	Adimensional	Numero de horas con presencia de nubes sobre el periodo de 24 horas dividido entre 24 horas.
Liquid precipitation duration fraction	Adimensional	Numero de horas con precipitación durante el periodo de 24 horas dividido entre 24 horas y por unidad de área.
Precipitation flux	mm d^{-1}	Volumen total de líquido (mm^3) precipitado durante el periodo de 24 horas por unidad de área (mm^2).
Snow thickness	cm	Profundidad promedio de la nieve durante el periodo de 24 horas medido como el volumen de nieve (cm^3) por unidad de área (cm^2).
Snow thickness LWE	cm	Profundidad promedio de nieve liquida (equivalente agua) sobre el periodo de 24 horas medida como volumen de nieve (cm^3) por unidad de área (cm^2)
Solar radiation flux	$\text{J m}^{-2} \text{d}^{-1}$	Cantidad total de energía provista por la radiación solar en la superficie en un periodo de 24 horas por unidad de área y tiempo.
Solid precipitation duration fraction	Adimensional	Numero de horas con precipitación sólida (nieve, mezcla de nieve y agua, etc) en un periodo de 24 horas dividido entre 24 horas y por unidad de área.
Vapour pressure	hPa	Contribución a la presión atmosférica total provista por el vapor de agua sobre el periodo de 24 horas por unidad de tiempo.

Tabla A.1: Variables del dataset AgERA5 con sus unidades y descripción

APÉNDICE B

RENDIMIENTOS ESTIMADOS

Rendimiento reportado por parcela durante las 3 temporadas que comprende este estudio y su correspondiente estimación en función de las variables utilizadas, en su versión real y modelada.

Parcela	Yield	max <i>NDVI</i>	mean <i>NDVI</i>	max <i>GNDVI</i>	mean <i>GNDVI</i>	max <i>CIRE</i>	mean <i>CIRE</i>	max <i>NDREI</i>	mean <i>NDREI</i>	mean <i>MSI</i>
1	70	61.125	69.368	59.373	65.187	63.588	56.929	62.910	54.683	24.885
2	75	80.868	69	81.548	62.595	79.554	69.27	79.673	67.051	73.058
3	60	58.430	67.972	55.098	61.310	61.557	63.487	61.994	67.111	58.056
5	80	75.205	81.628	77.449	79.511	74.026	76.248	74.534	76.477	77.345
6	80	69.615	79.353	71.169	77.159	66.548	71.927	67.061	77.024	78.114
7	80	82.334	59.536	83.835	76.908	81.624	78.011	81.651	71.280	58.410
8	80	63.985	65.721	66.579	69.665	63.664	61.533	63.132	60.969	70.071
9	70	74.965	72.762	76.237	71.302	73.722	68.874	74.050	67.326	76.719
10	80	75.744	73.049	76.611	85.067	73.202	83.198	74.165	81.677	77.968
11	80	74.663	71.669	75.251	72.516	76.928	80.778	75.192	78.609	78.174
13	65	71.203	63.558	68.213	64.332	68.879	69.461	69.009	70.332	73.297
14	75	71.600	86.904	73.549	81.431	78.471	79.607	75.352	78.152	78.217
15	75	79.995	74.043	80.643	74.365	75.853	76.375	78.571	78.240	73.975
16	40	58.932	55.970	55.102	60.259	61.620	55.488	61.800	53.509	42.351

Tabla B.1: Rendimiento reportado y estimados por variable para la zafra 2021/2022, empleando los datos reales

No.	Yield	ndvi_max	ndvi_mean	gndvi_max	gndvi_mean	cire_max	cire_mean	ndrel_max	ndrel_mean	msi_mean	lluvia	etc
1	70	56.008	51.642	58.421	63.043	57.247	56.042	56.486	52.766	-4.475	71.666	71.311
2	75	68.879	63.345	68.680	63.235	70.523	68.948	69.837	66.813	73.917	72.415	67.988
3	60	58.571	66.531	57.778	62.202	59.311	64.084	59.303	67.429	65.258	75.027	69.626
5	80	72.807	75.837	74.447	77.609	76.610	76.380	74.316	75.383	76.924	64.018	73.803
6	80	71.669	80.836	73.578	76.959	69.152	72.200	69.524	76.430	77.625	70.608	72.909
7	80	80.455	60.425	84.965	71.187	81.594	74.093	78.068	66.893	29.702	68.689	73.200
8	80	63.735	61.685	63.410	67.452	62.350	60.737	62.971	59.837	65.075	72.432	72.182
9	70	73.824	73.873	74.664	74.496	73.597	71.391	73.146	71.469	77.936	72.162	72.175
10	80	78.846	80.669	79.348	84.935	81.501	83.370	80.449	81.807	78.123	68.595	72.228
11	80	78.147	79.514	77.893	75.372	79.137	81.339	78.570	81.363	75.949	69.329	71.512
13	65	74.054	68.414	70.332	64.256	72.143	69.070	74.115	69.129	70.408	71.084	70.448
14	75	71.942	74.701	73.182	79.426	76.157	77.965	73.702	75.593	77.904	71.244	70.431
15	75	75.944	73.886	76.045	70.444	76.866	74.146	76.521	74.061	75.044	69.881	69.029
16	40	64.387	56.545	65.091	61.223	59.264	56.750	61.860	56.070	42.762	67.224	71.472

Tabla B.2: Rendimiento reportado y estimados por variable para la zafra 2021/2022, empleando los datos modelados

Parcela	Yield	ndvi_max	ndvi_mean	gndvi_max	gndvi_mean	cire_max	cire_mean	ndre1_max	ndre1_mean	msi_mean
1	60	59.649	40.497	56.372	46.307	55.929	44.852	60.458	40.985	18.971
2	65	69.742	59.412	66.835	57.627	66.668	62.694	67.456	62.244	63.276
3	50	51.203	59.085	47.713	57.158	52.546	54.449	54.063	57.533	66.574
4	65	66.815	64.163	66.317	60.465	68.067	65.507	66.382	66.525	61.984
5	75	65.914	65.996	67.118	66.583	66.262	65.635	65.190	65.070	62.040
6	75	69.350	64.786	70.885	65.292	64.666	64.111	67.114	64.848	66.051
7	75	73.058	75.104	72.924	76.502	78.605	79.054	75.421	76.195	70.575
8	75	73.249	67.044	74.665	67.053	71.846	65.507	70.600	65.689	67.764
9	75	67.852	67.280	71.125	67.216	64.623	64.986	66.166	66.573	68.271
10	75	68.963	76.303	66.556	79.375	70.454	76.613	70.455	76.018	71.745
11	75	70.922	79.726	69.795	80.042	75.080	82.371	72.939	80.752	72.498
13	50	41.137	51.265	39.850	47.022	40.014	46.060	39.937	48.423	60.014
14	65	73.079	48.153	73.621	51.798	71.319	56.954	71.771	51.879	54.498
15	50	67.102	65.801	66.093	62.695	63.458	63.814	67.379	67.137	69.403
16	30	37.779	44.831	43.706	46.930	39.125	41.864	38.622	42.486	33.343

Tabla B.3: Rendimiento reportado y estimados por variable para la zafra 2022/2023, empleando los datos reales

No.	Yield	ndvi_max	ndvi_mean	gndvi_max	gndvi_mean	cire_max	cire_mean	ndrel_max	ndrel_mean	msi_mean	lluvia	etc
1	60	44.124	37.339	46.332	42.345	47.523	44.010	46.589	40.144	24.069	82.402	66.797
2	65	71.681	71.817	69.189	68.234	66.995	69.315	70.171	71.910	68.862	50.621	57.598
3	50	49.847	57.679	47.795	55.750	51.377	53.770	51.741	56.301	63.262	64.934	66.120
4	65	63.697	61.082	64.626	58.221	68.416	65.515	66.662	64.841	60.220	63.065	58.261
5	75	71.200	67.442	74.909	67.932	68.685	66.819	70.058	66.968	65.757	82.384	75.124
6	75	69.372	68.246	71.952	70.044	67.937	67.717	68.480	68.060	66.513	65.930	61.284
7	75	74.732	71.459	75.358	74.205	78.329	76.728	75.963	73.078	69.207	68.244	76.014
8	75	65.115	64.144	68.554	65.434	66.243	64.777	64.390	63.724	65.912	65.783	57.900
9	75	68.350	65.319	72.635	67.268	67.670	64.430	67.349	64.878	67.185	65.783	58.075
10	75	74.820	75.800	73.199	78.894	76.547	76.274	75.781	75.235	72.363	66.073	74.910
11	75	77.799	78.927	77.570	80.213	80.470	80.646	78.419	78.634	73.844	66.073	76.617
13	50	48.045	49.455	42.508	43.981	43.335	44.669	44.789	46.624	57.056	63.836	49.582
14	65	55.351	51.272	56.358	53.503	58.405	57.761	57.683	55.020	56.976	65.915	67.300
15	50	66.604	66.878	62.276	62.896	62.740	64.787	66.067	67.986	69.518	48.530	55.386
16	30	36.457	39.807	39.649	41.690	39.206	40.023	37.590	39.642	30.058	51.823	45.373

Tabla B.4: Rendimiento reportado y estimados por variable para la zafra 2022/2023, empleando los datos modelados

Parcela	Yield	ndvi_max	ndvi_mean	gndvi_max	gndvi_mean	cire_max	cire_mean	ndre1_max	ndre1_mean	msi_mean
2	50	52.485	56.000	49.576	55.013	50.874	55.144	50.962	55.583	55.339
3	65	53.811	56.299	53.160	55.767	54.994	55.193	55.179	57.320	59.364
4	65	52.698	56.209	52.232	55.808	52.537	55.401	52.568	57.426	59.736
5	65	58.528	52.566	62.247	53.532	61.300	53.808	61.459	53.903	55.189
7	65	56.473	52.299	57.973	51.847	56.596	53.170	57.387	50.492	49.604
8	65	59.068	54.302	63.079	53.811	62.522	53.201	62.323	52.886	53.162
9	65	52.580	57.274	53.675	57.455	53.629	55.211	53.156	59.033	64.922
10	65	55.816	50.340	56.421	51.826	54.663	53.158	55.569	50.032	48.180
11	40	56.968	51.419	58.183	51.419	58.626	52.271	58.647	48.611	44.993
13	40	50.698	55.706	47.718	56.215	49.539	55.598	48.614	56.916	56.055
15	45	54.570	53.526	53.396	53.711	51.898	54.398	52.205	53.035	49.163

Tabla B.5: Rendimiento reportado y estimados por variable para la zafra 2023/2024, empleando los datos reales

No.	Yield	ndvi_max	ndvi_mean	gndvi_max	gndvi_mean	cire_max	cire_mean	ndre1_max	ndre1_mean	msi_mean	lluvia	etc
2	50	55.004	55.051	54.102	54.412	52.450	53.346	53.866	54.744	53.396	52.612	54.599
3	65	55.419	56.537	54.203	54.438	53.556	53.444	54.019	55.024	56.276	51.791	53.246
4	65	56.338	57.200	53.985	54.450	52.598	53.117	53.639	55.171	56.623	51.556	51.347
5	65	52.619	52.270	54.987	54.258	57.378	55.907	55.310	53.648	51.504	56.349	58.691
7	65	52.914	51.974	54.719	54.276	55.649	55.359	54.897	53.737	51.512	51.761	51.272
8	65	51.103	51.306	55.431	54.224	59.044	56.760	55.859	53.348	51.746	56.732	59.641
9	65	56.314	57.899	53.966	54.509	54.030	54.048	53.998	54.981	63.772	66.772	65.513
10	65	52.776	51.581	54.722	54.282	55.152	55.259	54.861	53.650	51.548	52.414	53.847
11	40	52.247	51.372	54.883	54.265	57.171	56.242	55.337	53.405	52.247	52.414	53.685
13	40	55.823	56.280	53.954	54.481	51.916	52.776	53.538	55.135	54.452	51.742	51.442
15	45	53.741	52.955	54.405	54.349	53.807	54.204	54.415	54.179	51.493	51.794	48.450

Tabla B.6: Rendimiento reportado y estimados por variable para la zafra 2023/2024, empleando los datos modelados

BIBLIOGRAFÍA

- Abdel-Rahman, E. M., & Ahmed, F. B. (2008). The application of remote sensing techniques to sugarcane (*Saccharum* spp. hybrid) production: a review of the literature. *International Journal of Remote Sensing*, 29, 3753-3767. <https://doi.org/10.1080/01431160701874603>
- Aguilar, C. G. (2014). Aplicación de índices de vegetación derivados de imágenes satelitales Landsat 7 ETM+ y ASTER para la caracterización de la cobertura vegetal en la zona centro de la provincia de Loja, Ecuador. <https://doi.org/10.35537/10915/34487>
- Aguilar, N., Galindo, G., Fortanelli, J., & Contreras, C. (2010). Índice normalizado de vegetación en caña de azúcar en la Huasteca Potosina. *Avances en Investigación Agropecuaria*. <https://www.redalyc.org/articulo.oa?id=83715125003>
- Aguilar Rivera, N. (2011). *Competitividad de la agroindustria azucarera de la huasteca México* [Tesis doctoral, Universidad Autónoma de San Luis Potosí].
- Aguilar-Rivera, N., Rodríguez L., D. A., Enríquez R., V., Castillo M., A., & Herrera S., A. (2012). The Mexican Sugarcane Industry: Overview, Constraints, Current Status and Long-Term Trends. *Sugar Tech*, 14, 207-222. <https://doi.org/10.1007/s12355-012-0151-3>
- Aguilar-Rivera, N., Algara-Siller, M., Olvera-Vargas, L. A., & Michel-Cuello, C. (2018). Land management in Mexican sugarcane crop fields. *Land Use Policy*, 78, 763-780. <https://doi.org/10.1016/J.LANDUSEPOL.2018.07.034>
- Akbarian, S., Jamnani, M. R., Xu, C., Wang, W., & Lim, S. (2023). Plot level sugarcane yield estimation by machine learning on multispectral images: A case study of Bundaberg, Australia. *Information Processing in Agriculture*. <https://doi.org/10.1016/J.INPA.2023.06.004>
- Akbarian, S., Xu, C., Wang, W., Ginns, S., & Lim, S. (2022). Sugarcane yields prediction at the row level using a novel cross-validation approach to multi-year multispectral images. *Computers and Electronics in Agriculture*, 198, 107024. <https://doi.org/10.1016/J.COMPAG.2022.107024>
- Akbarian, S., Xu, C., Wang, W., Ginns, S., & Lim, S. (2023). An investigation on the best-fit models for sugarcane biomass estimation by linear mixed-effect modelling on unmanned aerial vehicle-based multispectral images: A case study of Australia. *Information Processing in Agriculture*, 10, 361-376. <https://doi.org/10.1016/J.INPA.2022.03.005>
- Allen, R., Pereira, L., Raes, D., & Smith, M. (2006). Evapotranspiración del cultivo. Guías para la determinación de los requerimientos de agua de los cultivos. *Estudio FAO Riego Y Drenaje No 56*. <https://www.fao.org/4/x0490s/x0490s00.htm>

- Araghi, A., Martinez, C. J., & Olesen, J. E. (2022). Evaluation of multiple gridded solar radiation data for crop modeling. *European Journal of Agronomy*, *133*, 126419. <https://doi.org/10.1016/J.EJA.2021.126419>
- Baez-Gonzalez, A. D., Kiniry, J. R., Meki, M. N., Williams, J., Alvarez-Cilva, M., Ramos-Gonzalez, J. L., Magallanes-Estala, A., & Zapata-Buenfil, G. (2017). Crop parameters for modeling sugarcane under rainfed conditions in Mexico. *Sustainability (Switzerland)*, *9*. <https://doi.org/10.3390/SU9081337>
- Barbieri, V., & Villa Nova, N. A. (1993). *Condicionamento climatico da produtividade potencial da cana-de-acucar (saccharum spp.): um modelo matematico-fisiologico de estimativa* [Tesis de maestría, Universidade de São Paulo].
- Barnes, E. M., Clarke, T., Richards, S. E., Colaizzi, P. D., Haberland, J., Kostrzewski, M., Waller, P. M., Choi, C. Y., Riley, E., Thompson, T. L., Lascano, R. J., Li, H., Moran, M. S., Robert, P. C., Rust, R. H., & Larson, W. E. (2000). Coincident detection of crop water stress, nitrogen status and canopy density using ground-based multispectral data. <https://api.semanticscholar.org/CorpusID:128773162>
- Bégué, A., Lebourgeois, V., Bappel, E., Todoroff, P., Pellegrino, A., Baillarin, F., & Siegmund, B. (2010). Spatio-temporal variability of sugarcane fields and recommendations for yield forecast using NDVI. *International Journal of Remote Sensing*, *31*, 5391-5407. <https://doi.org/10.1080/01431160903349057>
- Boogaard, H., Schubert, J., De Wit, A., Lazebnik, J., Hutjes, R., Van der Grijn, G., et al. (2020). Agrometeorological indicators from 1979 to present derived from reanalysis. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*, *10*.
- Brown, D., de Sousa, K., & van Etten, J. (2023). ag5Tools: An R package for downloading and extracting agrometeorological data from the AgERA5 database. *SoftwareX*, *21*, 101267. <https://doi.org/10.1016/J.SOFTX.2022.101267>
- Calera, A., & González-Piqueras, J. (2007). Parámetros biofísicos de la cubierta vegetal: relaciones operativas para obtención de mapas de estos parámetros desde las imágenes de satélite. *Ciudad Real, España: Grupo de Teledetección, Universidad Castilla la Mancha. Documento de PLEIADeS*.
- Canata, T. F., Wei, M. C. F., Maldaner, L. F., & Molin, J. P. (2021). Sugarcane yield mapping using high-resolution imagery data and machine learning technique. *Remote Sensing*, *13*, 1-14. <https://doi.org/10.3390/rs13020232>
- Chuvieco S., E. (2008). *Teledeteccion ambiental*. Ariel. <https://books.google.com.mx/books?id=aKsNXCVCtcQC>
- Comisión Veracruzana de Comercialización Agropecuaria. (2010). *Monografía de la caña de azúcar*. Consultado el 29 de mayo de 2024, desde <https://www.siiba.conadesuca.gob.mx/siica/Consulta/verDoc.aspx?num=119>
- CONABIO. (1998). *Precipitación media anual*.
- CONABIO. (2001). *Edafología*.
- CONADESUCA. (2015). *Ficha técnica del cultivo de la caña de azúcar*. Consultado el 19 de mayo de 2024, desde https://www.gob.mx/cms/uploads/attachment/file/141823/Ficha_Tcnica_Ca_a_de_Az_car.pdf
- CONADESUCA. (2024a). *Balance Mundial de Azúcar (Oct - Sep)*. Consultado el 29 de mayo de 2024, desde https://www.gob.mx/cms/uploads/attachment/file/987308/Balance_Mundial_2025t1_20250401.pdf
- CONADESUCA. (2024b). *Reporte final de producción de caña y azúcar*. Consultado el 29 de mayo de 2024, desde https://www.gob.mx/cms/uploads/attachment/file/944009/Reporte_cifras_finales_2023_2024_a.pdf

- Cruz, G. A. Z., Vélez, E., Chávez, L. T., & Magdaleno, H. F. (2017). Aplicación de tecnologías de percepción remota para la estimación del rendimiento en caña de azúcar. *Revista Mexicana de Ciencias Agrícolas*, 8, 1575-1586. <https://doi.org/10.29312/remexca.v8i7.513>
- Cuervo-Robayo, A. P., Téllez-Valdés, O., Gómez-Albores, M. A., Venegas-Barrera, C. S., Manjarrez, J., & Martínez-Meyer, E. (2015). *Temperatura media anual en México (1910-2009)*. http://www.conabio.gob.mx/informacion/metadatos/gis/tman13gw.xml?_httpcache=yes&_xsl=/db/metadatos/xsl/fgdc.html.xsl&_indent=no
- Dengia, A., Dechassa, N., Wogi, L., & Amsalu, B. (2023). A simplified approach to satellite-based monitoring system of sugarcane plantation to manage yield decline at Wonji-Shoa Sugar Estate, central Ethiopia. *Heliyon*, 9, e18982. <https://doi.org/10.1016/J.HELIYON.2023.E18982>
- Dimov, D., Uhl, J. H., Löw, F., & Seboka, G. N. (2022). Sugarcane yield estimation through remote sensing time series and phenology metrics. *Smart Agricultural Technology*, 2. <https://doi.org/10.1016/j.atech.2022.100046>
- Dingre, S. K., Gorantiwar, S. D., & Kadam, S. A. (2021). Correlating the field water balance derived crop coefficient (Kc) and canopy reflectance-based NDVI for irrigated sugarcane. *Precision Agriculture*, 22, 1134-1153. <https://doi.org/10.1007/s11119-020-09774-8>
- Earth Resources Observation and Science (EROS) Center. (2015). *Comparison of Landsat 7 and 8 bands with Sentinel-2*. Consultado el 29 de mayo de 2024, desde <https://www.usgs.gov/media/images/comparison-landsat-7-and-8-bands-sentinel-2>
- European Spatial Agency. (s.f.-a). *MSI Products Overview* [s.f.]. Consultado el 11 de septiembre de 2024, desde <https://sentiwiki.copernicus.eu/web/s2-products#S2Products-ProcessingLevelsS2-Products-Processing-Levelstrue>
- European Spatial Agency. (s.f.-b). *Overview of Sentinel-2 Mission* [s.f.]. Consultado el 11 de septiembre de 2024, desde <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2/satellite-description>
- Fagan, M. E., & DeFries, R. S. (2024). Remote Sensing and Image Processing. *Reference Module in Life Sciences*, 432-445. <https://doi.org/10.1016/B978-0-12-822562-2.00060-8>
- Fauconnier, R., & Bassereau, D. (1975). *La caña de azúcar*. Blume. <https://books.google.com.mx/books?id=mMzDAAAACAAJ>
- Gao, B. C. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58, 257-266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3)
- Gillies, S., et al. (2016). Rasterio: geospatial raster I/O for Python programmers. *Mapbox*. <https://github.com/mapbox/rasterio>
- Gitelson, A., Kaufman, Y., & Merzlyak, M. (1996). Use of a green channel in remote sensing of global vegetation from EOS- MODIS. *Remote Sensing of Environment*, 58, 289-298. [https://doi.org/10.1016/S0034-4257\(96\)00072-7](https://doi.org/10.1016/S0034-4257(96)00072-7)
- Gitelson, A., & Merzlyak, M. N. (1994). Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. *Journal of Photochemistry and Photobiology B: Biology*, 22, 247-252. [https://doi.org/10.1016/1011-1344\(93\)06963-4](https://doi.org/10.1016/1011-1344(93)06963-4)
- Gitelson, A. A., Viña, A., Ciganda, V., Rundquist, D. C., & Arkebauer, T. J. (2005). Remote estimation of canopy chlorophyll content in crops. *Geophysical Research Letters*, 32. <https://doi.org/10.1029/2005GL022688>
- Gomarasca, M. A. (2009). *Basics of geomatics*. Springer Netherlands. <https://doi.org/10.1007/978-1-4020-9014-1>

- Gómez, R. A. (2009). *Conceptos de Geomática y estudios de caso en México*. Instituto de Geografía.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2017.06.031>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., . . . Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*, 1999-2049. <https://doi.org/10.1002/qj.3803>
- Huete, A. (2012). Vegetation Indices, Remote Sensing and Forest Monitoring. *Geography Compass*, *6*, 513-532. <https://doi.org/10.1111/J.1749-8198.2012.00507.X>
- INEGI. (2023). *Anuario estadístico y geográfico por entidad federativa 2022*. <https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=889463910534>
- Jensen, J. R. (2007). *Remote sensing of the environment: an earth resource perspective (2nd ed)*. Pearson Prentice Hall.
- Jutz, S., & Milagro-Pérez, M. P. (2018). Copernicus Program. *Comprehensive Remote Sensing*, *1-9*, 150-191. <https://doi.org/10.1016/B978-0-12-409548-9.10317-3>
- Karaman, Ç. H., & Akyürek, Z. (2023). Evaluation of near-surface air temperature reanalysis datasets and downscaling with machine learning based Random Forest method for complex terrain of Turkey. *Advances in Space Research*, *71*, 5256-5281. <https://doi.org/10.1016/J.ASR.2023.02.006>
- Keating, B., Carberry, P., Hammer, G., Probert, M., Robertson, M., Holzworth, D., Huth, N., Hargreaves, J., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J., Silburn, D., Wang, E., Brown, S., Bristow, K., Asseng, S., & Smith, C. (2003). An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy*, *18*, 267-288. [https://doi.org/10.1016/S1161-0301\(02\)00108-9](https://doi.org/10.1016/S1161-0301(02)00108-9)
- Leite, G. H. P., Crusciol, C. A. C., de Almeida Silva, M., Filho, W. G. V., & Suriano, A. (2009). Qualidade tecnológica, produtividade e margem de contribuição agrícola da cana-de-açúcar em função da aplicação de reguladores vegetais no início da safra. *Ciência Rural*, *39*, 726-732. <https://doi.org/10.1590/S0103-84782009000300015>
- Liang, S., Wang, J., & Jiang, B. (2020, enero). A systematic view of remote sensing. Elsevier. <https://doi.org/10.1016/B978-0-12-815826-5.00001-5>
- Lin, H., Chen, J., Pei, Z., Zhang, S., & Hu, X. (2009). Monitoring Sugarcane Growth Using ENVISAT ASAR Data. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(8), 2572-2580. <https://doi.org/10.1109/TGRS.2009.2015769>
- McKinney, W. (2010). Data Structures for Statistical Computing in Python. En S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 56-61). <https://doi.org/10.25080/Majora-92bf1922-00a>
- Morel, J., Todoroff, P., Bégué, A., Bury, A., Martiné, J.-F., & Petit, M. (2014). Toward a Satellite-Based System of Sugarcane Yield Estimation and Forecasting in Smallholder Farming Conditions: A Case Study on Reunion Island. *Remote Sensing*, *6*, 6620-6635. <https://doi.org/10.3390/rs6076620>
- Mulianga, B., Bégué, A., Simoes, M., & Todoroff, P. (2013). Forecasting Regional Sugarcane Yield Based on Time Integral and Spatial Aggregation of MODIS NDVI. *Remote Sensing*, *5*, 2184-2199. <https://doi.org/10.3390/rs5052184>
- Mutanga, S., van Schoor, C., Olorunju, P. L., Gonah, T., & Ramoelo, A. (2013). Determining the Best Optimum Time for Predicting Sugarcane Yield Using Hyper-Temporal Satellite

- Imagery. *Advances in Remote Sensing*, 02, 269-275. <https://doi.org/10.4236/ars.2013.23029>
- National Aeronautics and Space Administration, Earth Observatory. (s.f.). *What is Remote Sensing* [s.f.]. Consultado el 11 de septiembre de 2024, desde <https://www.earthdata.nasa.gov/learn/backgrounders/remote-sensing>
- National Aeronautics and Space Administration, Science Mission Directorate. (2010). *Introduction to the Electromagnetic Spectrum*. Consultado el 11 de septiembre de 2024, desde http://science.nasa.gov/ems/01_intro
- Nisbet, R., Elder, J., & Miner, G. (2009). Numerical Prediction. *Handbook of Statistical Analysis and Data Mining Applications*, 259-284. <https://doi.org/10.1016/B978-0-12-374765-5.00012-7>
- Ochoa Neira, M., Reyes García, M., & Manríquez Núñez, J. (2010). Producción sostenible de caña de azúcar (FIRA). *Boletín informativo*.
- Palacios-Vélez, E., Palacios-Sánchez, J. E., & Palacios-Sánchez, L. A. (2011). Agricultura de riego asistida con satélites. *Tecnología y Ciencias del Agua*. <https://www.redalyc.org/articulo.oa?id=353531973005>
- Perich, G., Turkoglu, M. O., Graf, L. V., Wegner, J. D., Aasen, H., Walter, A., & Liebisch, F. (2023). Pixel-based yield mapping and prediction from Sentinel-2 using spectral indices and neural networks. *Field Crops Research*, 292, 108824. <https://doi.org/10.1016/J.FCR.2023.108824>
- Perry, M. T. (2015a). *Raster Stats: User Manual*. <https://pythonhosted.org/rasterstats/manual.html>
- Perry, M. T. (2015b). *Rasterstats manual*. Consultado el 11 de septiembre de 2024, desde <https://pythonhosted.org/rasterstats/manual.html>
- Picoli, M. C. A., Lamparelli, R. A. C., Sano, E. E., & Rocha, J. V. (2014). The use of ALOS/PALSAR data for estimating sugarcane productivity. *Engenharia Agrícola*, 34, 1245-1255. <https://doi.org/10.1590/S0100-69162014000600019>
- Planet Labs PBC. (2018). Planet Application Program Interface: In Space for Life on Earth. *Planet*. <https://api.planet.com>
- Pour, A. B., Ranjbar, H., Sekandari, M., El-Wahed, M. A., Hossain, M. S., Hashim, M., Yousefi, M., Zoheir, B., Wambo, J. D. T., & Muslim, A. M. (2023). Remote sensing for mineral exploration. *Geospatial Analysis Applied to Mineral Exploration: Remote Sensing, GIS, Geochemical, and Geophysical Applications to Mineral Resources*, 17-149. <https://doi.org/10.1016/B978-0-323-95608-6.00002-0>
- Rahman, M. M., & Robson, A. (2020). Integrating Landsat-8 and Sentinel-2 Time Series Data for Yield Prediction of Sugarcane Crops at the Block Level. *Remote Sensing*, 12, 1313. <https://doi.org/10.3390/rs12081313>
- Rahman, M. M., & Robson, A. J. (2016). A Novel Approach for Sugarcane Yield Prediction Using Landsat Time Series Imagery: A Case Study on Bundaberg Region. *Advances in Remote Sensing*, 05, 93-102. <https://doi.org/10.4236/ars.2016.52008>
- Reed, B. C., Brown, J. F., VanderZee, D., Loveland, T. R., Merchant, J. W., & Ohlen, D. O. (1994). Measuring phenological variability from satellite imagery. *Journal of Vegetation Science*, 5, 703-714. <https://doi.org/10.2307/3235884>
- Rock, B., Vogelmann, J., Williams, D., Vogelmann, A., & Hoshizaki, T. (1986). Remote detection of forest damage. *BioScience*, 36, 439-445. <https://doi.org/10.2307/1310339>
- Roffe, S. J., & van der Walt, A. J. (2023). Representation and evaluation of southern Africa's seasonal mean and extreme temperatures in the ERA5-based reanalysis products. *Atmospheric Research*, 284, 106591. <https://doi.org/10.1016/J.ATMOSRES.2022.106591>

- SADER. (2018). *La caña de azúcar y su importancia para la industria azucarera*. <https://www.gob.mx/agricultura/nayarit/articulos/la-cana-de-azucar-y-su-importancia-para-la-industria-azucarera?idiom=es>
- Santillán-Fernández, A., Santoyo-Cortés, V. H., García-Chávez, L. R., Covarrubias-Gutiérrez, I., & Merino, A. (2016). Influence of drought and irrigation on sugarcane yields in different agroecoregions in Mexico. *Agricultural Systems*, 143, 126-135. <https://doi.org/10.1016/J.AGSY.2015.12.013>
- Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. *9th Python in Science Conference*.
- Shankar, D. R. (2017). *Remote Sensing of Soils*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-53740-4>
- Somard, J., Atzberger, C., Izquierdo-Verdiguier, E., Vuolo, F., & Immitzer, M. (2021). Remote Sensing Applications in Sugarcane Cultivation: A Review. *Remote Sensing*, 13, 4040. <https://doi.org/10.3390/rs13204040>
- Tucker, C. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8, 127-150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)
- USGS. (s.f.). *What is radiometric resolution?* [s.f.]. Consultado el 11 de septiembre de 2024, desde <https://www.usgs.gov/faqs/what-radiometric-resolution>
- Ustin, S., Roberts, D., Gamon, J., Asner, G., & Green, R. (2004). Using Imaging Spectroscopy to Study Ecosystem Processes and Properties. 54, 523-534. [https://doi.org/10.1641/0006-3568\(2004\)054\[0523:UISTSE\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2004)054[0523:UISTSE]2.0.CO;2)
- Villegas, J. V. S., & Suárez, G. A. P. (2022). *Cómo usar Google Earth Engine y no Fallar en el Intento*. Universidad Nacional Autónoma de México, Centro de Investigaciones en Geografía Ambiental. <https://doi.org/10.22201/ciga.9786073066969e.2022>
- Wang, M., Liu, Z., Baig, M. H. A., Wang, Y., Li, Y., & Chen, Y. (2019). Mapping sugarcane in complex landscapes by integrating multi-temporal Sentinel-2 images and machine learning algorithms. *Land Use Policy*, 88, 104190. <https://doi.org/10.1016/J.LANDUSEPOL.2019.104190>
- Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, 236, 111402. <https://doi.org/10.1016/J.RSE.2019.111402>
- Wright, D., Rowland, D., & Mulvaney, M. (2017). Water Use and Irrigation Management of Agronomic Crops. *EDIS*, 2017. <https://doi.org/10.32473/edis-aa131-2002>
- Zhang, J., & Li, J. (2023). Spacecraft. *Spatial Cognitive Engine Technology*, 129-162. <https://doi.org/10.1016/B978-0-323-95107-4.00004-4>