



**INSTITUTO TECNOLÓGICO DE CD. GUZMÁN**

TEMA:  
**IDENTIFICACIÓN DE ANOMALÍAS DEL DESARROLLO  
VEGETATIVO DE CAÑA DE AZÚCAR POR MEDIO DE DATOS  
SENTINEL 2**

QUE PARA OBTENER EL TÍTULO DE:  
**MAESTRO EN CIENCIAS DE LA COMPUTACIÓN**

PRESENTA:  
**HUMBERTO CRUZ SANABRIA**

DIRECTORES:  
**DRA. MARÍA GUADALUPE SÁNCHEZ CERVANTES  
DR. JUAN PABLO RIVERA CAICEDO**

CD. GUZMÁN JALISCO, MÉXICO, AGOSTO DE 2021



**EDUCACIÓN**  
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO  
NACIONAL DE MÉXICO

Instituto Tecnológico de Ciudad Guzmán

Ciudad Guzmán, **11/agosto/2021**

DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

Asunto: Autorización de impresión de Tesis

**ING. HUMBERTO CRUZ SANABRIA**  
**CANDIDATO AL GRADO DE MAESTRO EN CIENCIAS DE LA COMPUTACIÓN**  
**PRESENTE**

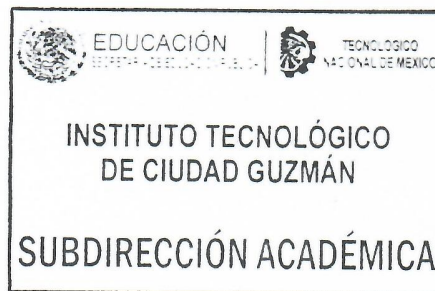
De acuerdo con los Lineamientos para la Operación de los Estudios de Posgrado en el Tecnológico Nacional de México y las disposiciones en este Instituto, habiendo cumplido con todas las indicaciones que la Comisión Revisora realizó con respecto a su Trabajo de Tesis titulado "Identificación de anomalías del desarrollo vegetativo de caña de azúcar por medio de datos Sentinel-2", la División de Estudios de Posgrado e Investigación de este Instituto, concede la Autorización para que proceda a la impresión del mismo.

Sin otro particular, quedo de Usted.

**ATENTAMENTE**

Excelencia en Educación Tecnológica  
"Innova, Transforma y Crea para ser Grande"

**M.C. CARLOS RUVALCABA MÁRQUEZ**  
**SUBDIRECTOR ACADÉMICO**



ccp. Archivo  
CRM/MC65/megg



Av. Tecnológico No. 100 C.P. 49100 A.P. 150  
Cd. Guzmán, Jal. Tel. Conmutador (341) 575200  
tecnm.mx | itcg.edu.mx





## Resumen

En México la caña de azúcar es un cultivo de gran importancia comercial y económica. Uno de los principales problemas al cual se enfrentan los productores de caña de azúcar es como mejorar el rendimiento del cultivo. Con el fin de contribuir a la solución del problema, en este estudio se presentan un método constituido por dos procesos para monitorear el cultivo de caña de azúcar basado en el uso de datos del sensor *MultiSpectral Instrument* a bordo del satélite Sentinel-2. El primero trata sobre identificar las etapas fenológicas de cultivos de caña de azúcar utilizando diversos índices de vegetación y el índice de área foliar como características que describen la fenología del cultivo. Para el desarrollo de este método, se evaluaron los métodos de clasificación: *k-Nearest Neighbors*, *Random Forest*, *Support Vector Machine* y *Naïve Bayes*; los resultados fueron validados utilizando la técnica de validación cruzada con  $k = 10$  iteraciones. Los resultados indican que es posible identificar las etapas fenológicas del cultivo de caña de azúcar con un  $F - measure = 0.923$  utilizando el clasificador *Random Forest*, siendo éste el que mejor se ajusta al proceso de identificar etapas de crecimiento de cultivos de caña de azúcar. El segundo proceso trata de identificar anomalías mediante el uso de perfil de referencia basado en el Índice de Vegetación Perpendicular (PVI, por sus siglas en inglés) y el índice de área foliar (LAI, por sus siglas en inglés). Para probar ambos procesos, se han definido dos áreas de estudio. Por un lado, seleccionamos una zona con control detallado del cultivo, sobre la que hemos definido el perfil de referencia. Por otro lado, seleccionamos un área de prueba para detectar áreas anómalas en el cultivo. Las pruebas mostraron que desde el día 0 al 60, la caña de azúcar en el área de prueba estuvo mayormente dentro del rango con el perfil de salud, y en el día 120, alcanza el crecimiento máximo más rápido que el perfil de salud; a partir de este día el área de prueba comenzó a mostrar algunas deficiencias. Por tanto, se concluye que es posible determinar las áreas con alguna deficiencia en el cultivo de caña de azúcar a través del método propuesto.

**Palabras clave:** Fenología, Anomalías, Sentinel-2.

## Abstract

In Mexico, sugarcane is a crop of great commercial and economic importance. One of the main problems sugarcane producers face is how to improve crop yield. In order to contribute to solving the problem, this study presents a method consisting of two processes to monitor the cultivation of sugarcane based on the use of data from the sensor *textit* MultiSpectral Instrument on board the Sentinel satellite. 2. The first one deals with identifying the phenological stages of sugarcane crops using various vegetation indices and the leaf area index as characteristics that describe the phenology of the crop. For the development of this method, the following classification methods were evaluated: *textit* k-Nearest Neighbors, *textit* Random Forest, *textit* Support Vector Machine and *textit* Naïve Bayes; the results were validated using the cross-validation technique with  $k = 10$  iterations. The results indicate that it is possible to identify the phenological stages of sugarcane cultivation with an  $F - measure = 0.923$  using the *textit* Random Forest classifier, this being the one that best fits the process of identifying growth stages of sugar cane crops. The second process tries to identify anomalies by using a reference profile based on the Perpendicular Vegetation Index (PVI) and the Leaf Area Index (LAI). To test both processes, two study areas have been defined. On the one hand, we select an area with detailed crop control, on which we have defined the reference profile. On the other hand, we select a test area to detect abnormal areas in the crop. The tests showed that from day 0 to day 60, the sugarcane in the test area was mostly within the range with the health profile, and on day 120, it reached the maximum growth faster than the health profile; From this day on the test area began to show some deficiencies. Therefore, it is concluded that it is possible to determine the areas with some deficiency in the cultivation of sugarcane through the proposed method.

**Keywords:** Phenology, Anomalies, Sentinel-2.

# Agradecimientos

*Gracias a mis padres y hermano por su apoyo.*

*Gracias a mi directora de tesis Dra. María Guadalupe Sánchez Cervantes, a mi Codirector Dr. Juan Pablo Rivera Caicedo y al Dr. Himer Avila George por sus enseñanzas y guía.*

*Gracias a todos mis maestros.*

*Gracias al doctor Santiago Belda del IPL(Laboratorio de Procesamiento de Imágenes) por su colaboración en mi estancia virtual en la Universidad de Valencia.*

# Índice general

Resumen

Abstract

Agradecimientos

Lista de figuras

Lista de tablas

<b>1. Introducción</b>	<b>1</b>
1.1. Planteamiento del problema . . . . .	4
1.2. Justificación . . . . .	5
1.3. Objetivos . . . . .	6
1.3.1. General . . . . .	6
1.3.2. Objetivos específicos . . . . .	6
1.4. Estructura del trabajo . . . . .	7
<b>2. Marco teórico</b>	<b>8</b>
2.1. Caña de azúcar . . . . .	8
2.1.1. Etapas fenológicas del cultivo . . . . .	8
2.2. Teledetección . . . . .	10
2.3. Espectro electromagnético . . . . .	10
2.3.1. Firmas Espectrales . . . . .	11
2.3.2. Índices de Vegetación . . . . .	12
2.3.2.1. Índice de Vegetación de Diferencia Normalizada (NDVI) . . . . .	12
2.3.2.2. Índice de Vegetación ajustado al Suelo (SAVI) . . .	13
2.3.2.3. Índice de Área Foliar (LAI) . . . . .	13
2.3.2.4. Índice de Vegetación de diferencia normalizada Ver- de (GNDVI) . . . . .	14
2.3.2.5. Índice de Vegetación Perpendicular (PVI) . . . . .	14

2.4.	Imágenes Satelitales . . . . .	15
2.4.1.	Resolución espacial . . . . .	15
2.4.2.	Resolución espectral . . . . .	15
2.4.3.	Resolución temporal . . . . .	16
2.4.4.	Series temporales . . . . .	17
2.4.5.	Sentinel-2 . . . . .	17
2.5.	Aprendizaje Automático . . . . .	18
2.5.1.	Regresión . . . . .	18
2.5.1.1.	Smoothing Splines. . . . .	18
2.5.1.2.	Cubic smoothing spline. . . . .	19
2.5.2.	Clasificación . . . . .	19
2.5.3.	Métodos de remuestreo . . . . .	20
2.5.4.	Métricas . . . . .	21
<b>3.</b>	<b>Antecedentes</b>	<b>23</b>
3.1.	Fenología . . . . .	23
3.1.1.	Fenología de bosques . . . . .	23
3.1.2.	Fenología en cultivos . . . . .	25
3.1.3.	Fenología en pastizales . . . . .	26
3.2.	Anomalías . . . . .	27
<b>4.</b>	<b>Metodología y materiales</b>	<b>30</b>
4.1.	Identificación de las etapas fenológicas del cultivo de caña de azúcar	33
4.1.1.	Descarga . . . . .	33
4.1.2.	Pre-Procesamiento . . . . .	33
4.1.3.	Cálculo de índices . . . . .	35
4.1.4.	Extracción de la zona de interés . . . . .	36
4.1.5.	Cálculo de series . . . . .	37
4.1.6.	Clasificación . . . . .	37
4.2.	Identificación de anomalías . . . . .	38
4.2.1.	Identificación de zonas anómalas . . . . .	39
<b>5.</b>	<b>Resultados</b>	<b>41</b>
5.1.	Caso de estudio . . . . .	41
5.2.	Descarga de imágenes . . . . .	42
5.3.	Pre-Procesamiento . . . . .	43
5.3.1.	Corrección atmosférica . . . . .	43
5.3.2.	Remuestreo, Recorte y Calculo de índices . . . . .	44
5.3.3.	Extracción de la zona de interés . . . . .	47
5.4.	Calculo de series . . . . .	48
5.5.	Identificación de las etapas fenológicas . . . . .	49
5.5.1.	Seleccionado las mejores características . . . . .	50
5.5.2.	Evaluación con las mejores características . . . . .	50
5.6.	Identificación de anomalías . . . . .	51
5.6.1.	Regresión . . . . .	53



<b>6. Conclusiones y trabajos futuros</b>	<b>56</b>
6.1. Conclusiones . . . . .	56
<b>A. Estancia</b>	<b>64</b>
A.1. Estancia virtual . . . . .	64
A.1.1. Lectura de imágenes . . . . .	65
A.1.2. Extracción de la zona de interés . . . . .	65
A.1.3. Interpolación . . . . .	66
A.1.3.1. Series de Fourier . . . . .	66
A.1.3.2. Whittaker . . . . .	66
A.1.4. Serie temporal . . . . .	66
A.1.4.1. Parámetros fenológicos . . . . .	67
A.1.5. Reconstrucción de imágenes . . . . .	68
<b>B. Productos de la tesis</b>	<b>69</b>
<b>Bibliografía</b>	<b>74</b>

# Lista de figuras

1.1. Principales etapas fenológicas de la caña de azúcar. . . . .	3
2.1. Comparación de longitud de onda del espectro electromagnético. . .	11
2.2. Firma espectral de la vegetación (green vegetation), suelo (soil) y agua (water). . . . .	12
2.3. Características de la resolución espacial y espectral de Sentinel-2. .	16
2.4. Sentinel-2 . . . . .	17
4.1. Metodología para identificar las etapas fenológicas en caña de azúcar	31
4.2. Metodología para identificar anomalías en caña de azúcar . . . . .	32
4.3. Fase de preprocesamiento . . . . .	33
4.4. Fase de cálculo de índices . . . . .	35
4.5. Interfaz gráfica de GeoJSON . . . . .	36
4.6. Fase de identificación de zonas anómalas . . . . .	39
5.1. Identificación de tiles . . . . .	43
5.2. Proceso de corrección atmosférica de sen2cor . . . . .	44
5.3. Corrección atmosférica con sen2cor . . . . .	44
5.4. Construcción del grafo . . . . .	45
5.5. Grafo utilizado en la fase de pre-procesamiento y calculo de índices	46
5.6. Estructura XML del grafo . . . . .	46
5.7. Interfaz gráfica de GeoJson . . . . .	47
5.8. Extracción de los píxeles de la zona de interés . . . . .	48
5.9. Series temporales de los índices de vegetación NDVI, GNDVI, SAVI, PVI y LAI en el área $Z_1$ . . . . .	49
5.10. Ejemplo de comparación de dos características, MinLAI y MaxPVI	49
5.11. Rango para el índice LAI. . . . .	52
5.12. Rango para el índice PVI . . . . .	52
5.13. Perfil de referencia para el índice LAI. . . . .	53
5.14. Perfil de referencia para el índice PVI. . . . .	53
5.15. Detección de anomalías; a) 02/02/2019, b) 04/03/2019, c) 06/02/2019, d) 09/20/2019, e) 11/04/2019, f) 02/20/2020 . . . . .	55
A.1. Metodología del script . . . . .	65
A.2. Interpolación: (a) Serie de fourier de grado 4 efectuada sobre la media, (b) Método de Whittaker efectuada sobre la media . . . . .	67
A.3. Parámetros fenológicos . . . . .	67

A.4. Interpolación de imagen para el día 100: (a) Serie de Fourier de grado 4, (b) Método de Whittaker . . . . .	68
--	----

# Lista de tablas

1.1. Municipios del estado de Jalisco en los que se cultiva la caña de azúcar. . . . .	2
2.1. Resolución de las bandas del Sentinel-2 . . . . .	18
5.1. Distribución de las etapas fenológicas del cultivo en la zona de estudio . . . . .	42
5.2. Resultados de la clasificación de las 15 características de las series temporales. . . . .	50
5.3. Resultados de la clasificación considerando las mejores características . . . . .	51

## ***Dedicatoria***

*A mi y a mi familia.*

# Capítulo 1

## Introducción

La caña de azúcar, llamada también *Saccharum officinarum* L., es una gramínea originaria de Nueva Guinea; se cultivó por primera vez el Sureste Asiático y la India occidental alrededor del año 327 d.de C., se introdujo en Egipto alrededor del año 647 d.de C. y alrededor de un siglo más tarde a España (755 d.de C.). Desde ese momento, el cultivo de la caña de azúcar se expandió a la mayoría de las regiones tropicales y sub-tropicales, a México llegó alrededor del año 1522 en la época de la conquista, de esta manera la primera plantación se realizó en el estado de Veracruz, estableciéndose posteriormente los primeros ingenios azucareros en las partes cálidas del país como consecuencia de la colonización (CONADESUCA, 2015).

México es uno de los principales productores de caña de azúcar a nivel internacional, ocupando el quinto lugar en su producción. Esta planta se cultiva en más de 130 países a nivel mundial. (López-Vázquez et al., 2016). Los mayores sembradíos se concentran en seis regiones: Noroeste (estado de Sinaloa), Pacífico (estados de Nayarit, Jalisco, Michoacán y Colima), Centro (estados de Puebla y Morelos), Noreste (estados de Tamaulipas y San Luis Potosí), Golfo (estados de Veracruz, Oaxaca y Tabasco) y Sur (estados de Chiapas, Campeche y Quintana Roo) (CONADESUCA, 2013).

Según el Comité Nacional para el Desarrollo Sustentable de la Caña de Azúcar (CONADESUCA, 2007), en el estado de Jalisco, 39 municipios cultivan caña de azúcar, siendo el mayor productor el municipio de Tala. En la tabla 1.1 se listan

los municipios que cultivan caña de azúcar en Jalisco.

TABLA 1.1: Municipios del estado de Jalisco en los que se cultiva la caña de azúcar.

Acatlán de Juárez	Ahualulco del Mercado	Etzatlán	San Juanito E.
Amatitlán	Ameca	Gómez Farías	San Marcos
Atoyac	Autlán de Navarro	Hostotipaquillo	San Martín H.
Casimiro Castillo	Cocula	La Huerta	Sayula
Cuaautitlán de García	El arenal	Magdalena	Tala
El Grullo	El Limón	Pihuamo	Tamazula
Tecalitlán	Tequila	Tlajomulco	Tonaya
Teocuitatlán de Corona	Teuchitlán	Tonila	Tuxcacuesco
Tuxpan	Villa Corona	Villa Purificación	Zacoalco
Zapopan	Zapotiltic	Zapotitlán de V.	

Jalisco es de los principales estados que producen la caña de azúcar, en 2018 en conjunto con Veracruz aportaron 28.3 millones de toneladas; esto es, 50.6 % del total nacional. La caña de azúcar se utiliza principalmente en la elaboración de azúcar, mediante la obtención del jugo de sus tallos. Además, algunos subproductos de la industria azucarera como por ejemplo, el bagazo, la melaza o la miel, son utilizados como insumos fundamentales para otras industrias, por ejemplo, la utilización de la melaza como elemento principal en la producción del alcohol, el bagazo en la industria del papel y en la producción de energía eléctrica.

La industria azucarera tiene gran impacto económico en México, ya que a través de ella se generan empleos directos e indirectos. La ley de desarrollo rural sustentable reconoce a la caña de azúcar como un cultivo fundamental en México y ordena atender los requerimientos de investigación de cada sector integrante de la agroindustria de la caña de azúcar para garantizar una agroindustria rentable, competitiva y sustentable. En 2017, CONADESUCA creó y publicó la Agenda Nacional de Investigación, Innovación y Transferencia de Tecnología de Caña de Azúcar, en la que se detallan restricciones tecnológicas de la cadena de valor de la agroindustria de la caña de azúcar, donde se destacan: (1) Bajos rendimientos de campo, (2) Degradación del suelo, (3) Excesivo uso de agroquímicos, (4) Fertilización química sin considerar necesidades específicas del cultivo, ciclo y clima. Estas restricciones afectan directamente a los productores que no disponen de las herramientas tecnológicas necesarias para garantizar que sus cultivos rindan conforme a las demandas de la industria (CONADESUCA, 2019). Por lo tanto, es de vital importancia el comprender como las plantas responden y perciben a su entorno y los cambios morfológicos y anatómicos que se presentan durante su desarrollo.

Para poder identificar estos cambios es de vital importancia el reconocer las diferentes etapas fenológicas de la planta, pues esto permite reconocer etapas críticas las cuales pueden ayudar al desarrollo de medidas para mejorar el rendimiento del cultivo de caña de azúcar en campo (Endres et al., 2018).

Después del primer corte (socas), se identifican tres etapas fenológicas en caña de azúcar: (1) brotación y amacollamiento, (2) rápido crecimiento y (3) maduración (Aguilar Rivera, 2015); ver la Figura 1.1 modificada desde (McCollam, 2018). Un problema reiterado en el cultivo de caña de azúcar, es que existen productores los cuales no tienen un registro exacto de la fecha de corte, por lo cual, comúnmente se ignora la etapa fenológica en la que se encuentra el cultivo de caña, lo que dificulta el crear medidas agrícolas que ayuden a mejorar el rendimiento en campo.

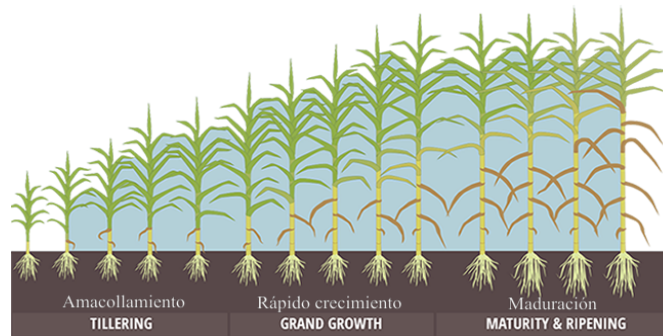


FIGURA 1.1: Principales etapas fenológicas de la caña de azúcar.  
Modificada desde (McCollam, 2018)

En los últimos años, se ha popularizado el uso de vehículos aéreos no tripulados para el desarrollo de aplicaciones agrícolas mediante el registro de la energía electromagnética que interactúa con los cultivos con los cuales es posible identificarlos, delinearlos y distinguirlos, esto se conoce como teledetección. Comúnmente, las propiedades de los cultivos en términos de sus niveles asociados de energía electromagnética son recolectados por sensores montados en plataformas aéreas lo que brinda la oportunidad de capturar grandes áreas geográficas con una sola observación o escena. Específicamente, las plataformas espaciales como los satélites tienen la ventaja de que las áreas geográficas de interés se revisan en un ciclo regular, lo que facilita la adquisición de datos que revelan condiciones cambiantes en estas áreas a lo largo del tiempo. A partir de la década de los 70, se han desarrollado nuevos sensores con mejoras en la resolución espectral, espacial y radiométrica, lo que ha permitido enriquecer el conocimiento de la tierra. El Instrumento Multiespectral (MSI, por sus siglas en inglés) es uno de los sensores más actuales y fue



desarrollado por la Agencia Espacial Europea (ESA, por sus siglas en inglés), este sensor es capaz de capturar imágenes en 13 bandas distintas del espectro electromagnético, con resoluciones de las capturas de 10, 20 y 60 metros. El MSI viaja a bordo de los satélites Sentinel-2 <sup>1</sup>(A y B), ambos satélites son idénticos y viajan en la misma órbita, con una separación de 180 grados entre si con la finalidad de lograr una cobertura y una descarga de datos óptimos. Cada cinco días los satélites cubren todas las superficies terrestres, grandes islas y aguas costeras (Delegido et al., 2018). Gracias a sus prestaciones estos satélites han tenido un gran impacto en la comunidad, contribuyendo significativamente al estudio y desarrollo de aplicaciones agrícolas, siendo preferido a otras plataformas de datos ópticos.

En este trabajo de tesis, se desarrolló un sistema de software que utiliza imágenes proporcionadas por el sensor MSI del satélite Sentinel-2, con las que se caracterizan las diferentes etapas fenológicas de la caña de azúcar mediante el uso de índices de vegetación con la finalidad de identificarlas. Adicionalmente se presenta un método para identificar anomalías en el cultivo de la caña de azúcar a través de un perfil de salud construido con el Índice de Vegetación Perpendicular (PVI, por sus siglas en inglés) y el Índice de Área Foliar (LAI, por sus siglas en inglés). Para la detección de anomalías se utilizaron dos áreas de estudio con condiciones similares de sistemas de riego y plantación de caña de azúcar. Por un lado, seleccionamos una zona con control detallado del cultivo, de la que tenemos definido el perfil sano. Por otro lado, seleccionamos un área de prueba para detectar posibles áreas anómalas en el cultivo. El área de prueba fue monitoreada durante la temporada 2019-2020.

## 1.1. Planteamiento del problema

La ley de desarrollo sustentable de la caña de azúcar publicada en el diario oficial de la federación (DOF 22-08-2005), ordena atender los requerimientos de investigación de los sectores integrantes de la industria de la caña de azúcar para garantizar que esta sea rentable, competitiva y sustentable. Su marco operativo se establece en el Programa Nacional de la Caña de Azúcar (PRONAC, 2014-2018). Con fundamento en lo anterior, CONADESUCA, en 2017, desarrolló y publicó

---

<sup>1</sup><https://sentinel.esa.int/web/sentinel/missions/sentinel-2>

la agenda nacional de investigación, innovación y transferencia de tecnología de caña de azúcar (CONADESUCA, 2017), en su trabajo se detallan alrededor de 18 restricciones tecnológicas de la cadena de valor de la agroindustria de la caña de azúcar en campo, donde se destacan:

- Bajos rendimientos de campo.
- Degradación del suelo.
- Excesivo uso de agroquímicos.
- Fertilización química sin considerar necesidades específicas del cultivo, ciclo y clima.

Estas restricciones afectan al eslabón mas débil de la cadena, los productores, los cuales no cuentan con productos tecnológicos que contribuyan a disminuir la brecha tecnológica (Legazpi, Rangel, y Núñez, 2017). Específicamente en el estado de Jalisco, esta problemática se ve reflejada en el rendimiento en campo, pues desde 2015 no se alcanza el máximo histórico que fue de 98.79 t/ha, además de que este máximo histórico está por debajo del peor rendimiento del estado de Puebla que es de los estados con mejor rendimiento en campo del país (CONADESUCA, 2019). Es por esto la importancia de identificar anomalías del desarrollo vegetativo de la caña de azúcar.

Las diferentes técnicas de percepción remota satelital pueden brindar un apoyo a los productores para identificar los efectos de los limitantes tanto de variedades, clima, suelo y plagas en grandes extensiones de cultivos y a bajo costo (Gerritsen, Bernal, y Rodríguez, 2019).

## 1.2. Justificación

En México la caña de azúcar es uno de los principales cultivos. En el 2018, Jalisco fue uno de los principales estados productores, que en conjunto con Veracruz aportaron 28.3 millones de toneladas; esto es, 50.6 % del total nacional. La caña de azúcar se utiliza principalmente en la obtención de azúcar, mediante la extracción

del jugo de sus tallos, y de los principales subproductos de la industria azucarera que son la melaza y el bagazo, representan la materia prima para diversas industrias, por ejemplo, la melaza es utilizada para la producción de alcohol y en la industria del papel el bagazo es utilizado como materia prima (Romero, 2012).

Dada a la fuente importante de materia que proporciona la caña de azúcar a diversas industrias, especialmente la agroindustria azucarera que es de suma importancia en la economía de México, pues representa una importante fuente de empleos directos e indirectos (CONADESUCA, 2015). CONADESUCA en 2017 creó y publicó la agenda nacional de investigación, innovación y transferencia de tecnología de caña de azúcar, en donde se reúne la mayoría de los requerimientos tecnológicos y económicos del cultivo, que se consideran deben ser atendidos con el objetivo de alcanzar el mayor potencial de desarrollo de la caña de azúcar, pues en los últimos años el rendimiento en campo del cultivo de caña apenas creció en un 2.4

### **1.3. Objetivos**

A continuación, se describe el objetivo general y los objetivos específicos.

#### **1.3.1. General**

Diseñar, implementar y validar un sistema de software que estime anomalías en el desarrollo fenológico del cultivos de caña de azúcar como método de alerta temprana sobre los deficientes en suelo o en cultivos que afecten la productividad de los productores del estado de Jalisco.

#### **1.3.2. Objetivos específicos**

- Identificar las diferentes etapas fenológicas del cultivo de caña de azúcar utilizando índices espectrales.

- Caracterizar las diferentes etapas fenológicas para la caña de azúcar utilizando índices espectrales para determinar el perfil fenológico de referencia para el cálculo de las anomalías en el desarrollo vegetativo del cultivo.
- Validar la metodología propuesta para el monitoreo de anomalías en caña de azúcar en el estado de Jalisco para el periodo 2019-2020 como estudio de caso.

## 1.4. Estructura del trabajo

El capítulo 2 de Marco teórico tiene como objetivo el introducir los conceptos básicos que enmarcan el presente trabajo de tesis, con la finalidad de presentar los conocimientos indispensables para comprender el desarrollo de este.

En el capítulo 3 de Metodología y materiales se expone la metodología diseñada e implementada para el monitoreo de la caña de azúcar. Se presentan 2 procesos que son identificación de las etapas fenológicas e identificación de anomalías, así como los materiales y herramientas utilizadas para el procesamiento de las imágenes en el software.

El capítulo 4 de Resultados muestra los resultados logrados para cada uno de los procesos incluidos en el software, para ello se eligieron 2 zonas en las cuales se efectuaron los ambos procesos.

Por último en el capítulo 5 de Conclusiones se muestra las conclusiones obtenidas del funcionamiento y resultados del software desarrollado, además de presentar futuras mejoras y añadidos para robustecer el estudio.

# Capítulo 2

## Marco teórico

### 2.1. Caña de azúcar

La caña de azúcar es una gramínea la cual pertenece al genero *Saccharum*, esta crece con forma de tallos o macollos los cuales se propagan por partes vegetativas (en forma asexual) (Van Dillewijn, 1951), es una planta con producción alta de hojas y tallos, en su etapa de madurez tiene la mitad de su biomasa en forma de fibra y azúcares. Según CONADESUCA (2015) la caña de azúcar se puede clasificar de la siguiente manera.

Reino: Plantae; Division: Magnoliophyta; Clase: Liliopsida; Subclase: Commelinidae; Orden: Poales; Familia Poaceae; Subfamilia: Panicoideae; Tribu: Andropogoneae; Genero: Saccharum; Especie: *S. officinarum* L.

#### 2.1.1. Etapas fenológicas del cultivo

El ciclo plantilla de la caña de azúcar se caracteriza por un desarrollo vegetativo de duración variable, ya que este es dependiente de factores como la variedad y la influencia del clima. La duración del cultivo hasta la cosecha dura desde 14 a 17 meses. La planta de caña de azúcar en este ciclo pasa por cuatro etapas:

germinación y/o emergencia, amacollamiento o ahijamiento, rápido crecimiento y maduración.

Mientras que el desarrollo después del primer corte (socas) tiene un periodo de duración de 11 a 13 meses, en la cual se pueden distinguir tres etapas: brotación y amacollamiento, rápido crecimiento y maduración, ver figura 1.1. A continuación, se describe cada una de estas etapas fenológicas según Romero (2012):

**Germinación y emergencia.** La germinación es la etapa en la que se da lugar al crecimiento de los órganos primordios latentes en la yema a un estado activo de crecimiento y desarrollo. La duración de esta etapa suele variar, comienza entre los 7 a 10 días después de la siembra. El crecimiento se prolonga hasta los 35 días.

#### **Amacollamiento o ahijamiento.**

Esta etapa inicia cerca de los 35 a los 40 días después de la siembra, se caracteriza por el brote de diversos tallo a partir de las articulaciones nodales que se encuentran en la base de los tallos primarios.

**Rápido crecimiento.** En esta etapa se presenta una gran acumulación de masa seca, además la planta alcanza su máxima área foliar, es decir, mayor cantidad de hojas. La formación y elongación de la caña se presenta con mayor rapidez en esta etapa. En esta fase el cultivo expresa la máxima respuesta a los factores ambientales y de manejo.

**Maduración.** En la etapa de maduración, la planta de caña de azúcar da paso al proceso de síntesis y acumulación de sacarosa en los tallos. Esta etapa tiene una duración aproximada de dos a tres meses.

**Cosecha.** La cosecha se realiza entre los 11 y 16 meses desde el inicio de la plantación. En este punto los tallos dejan de desarrollarse, las hojas se marchitan y caen, la corteza de la capa se vuelve quebradiza. Generalmente se quema la plantación para eliminar la maleza que dificulta la recolección y corte de la caña, así como posibles plagas (ratas de campo, víboras, tuzas, etc.) que pudiesen causar daño a los cortadores.

## 2.2. Teledetección

La teledetección se define como la adquisición de información sobre un objeto a distancia mediante el uso de sensores (instrumentos capaces de captar la radiación electromagnética) localizados en plataformas móviles como un satélite, esto sin que exista contacto material entre el objeto o sistema observado y el observador. Según Sobrino (2001), la teledetección tiene como objetivo el reconocer las características de la superficie terrestre y de los fenómenos que tienen lugar en ella. Esto mediante los datos captados por el sensor. Existen tres elementos principales en la teledetección, los cuales son:

- Un instrumento o sensor para observar y obtener información del objeto.
- Una plataforma para sostener el instrumento.
- Un objeto a observar.

Otro elemento importante en la teledetección es la transformación de los datos captados por el sensor en magnitudes físicas en información útil.

## 2.3. Espectro electromagnético

Según Ordóñez (2012) a la corriente saliente de energía de un origen con estructura de ondas electromagnéticas se le define como radiación electromagnética. Esta radiación puede provenir de fuentes naturales o artificiales. El espectro electromagnético es el conjunto de todas las frecuencias (número de ciclos de la onda por unidad de tiempo) posibles a las que se produce radiación electromagnética, De esta manera, el límite teórico inferior del espectro electromagnético es cero (no existen frecuencias negativas) y el teórico superior es infinito. En la figura 2.2 se puede observar el rango de las longitudes de onda. Cada objeto en el planeta tiene una manera de interactuar con el espectro de manera diferente (firma espectral).

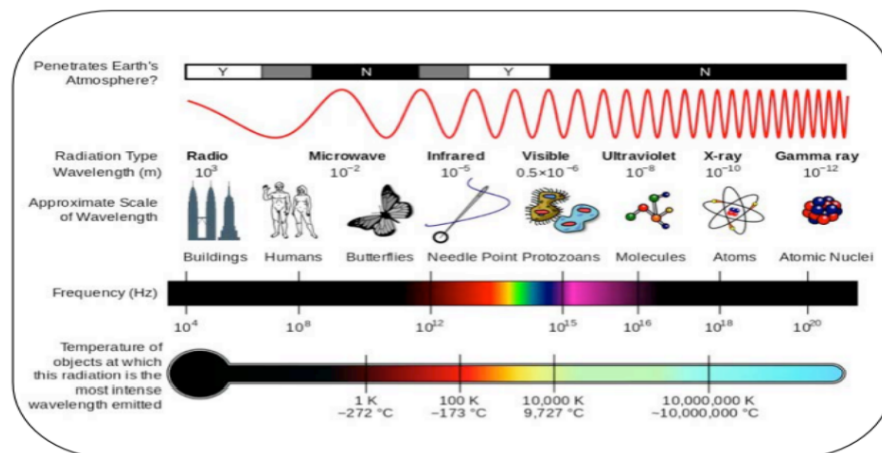


FIGURA 2.1: Comparación de longitud de onda del espectro electromagnético.  
Fuente:(CESAR, 2018)

### 2.3.1. Firmas Espectrales

Los distintos rangos del espectro electromagnético

La energía que llega a la superficie terrestre (por absorción, reflexión, o transmisión), define la firma espectral particular de cada objeto o superficie en los distintos rangos que se encuentran en el espectro electromagnético (Hernández y Montaner, 2009). La firma espectral es muy utilizada en procesos de análisis de imágenes.

En la figura 2.3 se pueden observar las firmas espectrales de tres elementos, soil, green vegetation y water. Las firmas de estos elementos tienen una variación distinta en el espectro visible (1, 2 y 3), pero la firma de la vegetación tiene mayor interacción en el infrarrojo cercano (4), por lo que el uso de esta banda es de gran relevancia en el estudio de los cultivos (Rodríguez Ramirez, 2019).



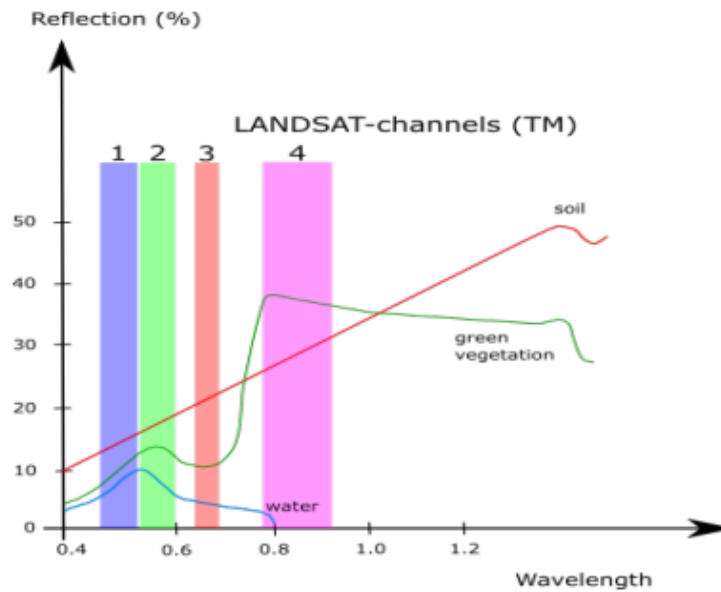


FIGURA 2.2: Firma espectral de la vegetación (green vegetation), suelo (soil) y agua (water).  
Fuente:(ESA, 2018)

### 2.3.2. Índices de Vegetación

Un índice de vegetación se define como la combinación algebraica de algunos de los rangos del espectro electromagnético (bandas en sensores multiespectrales), estos permiten realizar un fenómeno o la respuesta biológica del cultivo (Sanseechan et al., 2019).

A continuación se muestran algunos de los índices mas utilizados.

#### 2.3.2.1. Índice de Vegetación de Diferencia Normalizada (NDVI)

El NDVI representado en la Ec. 2.1, fue presentado por Rouse (1974). Este índice representa la relación de las diferencias en las reflectividades para la banda del infrarrojo cercano (NIR) y la banda roja (RED).

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \quad (2.1)$$

NDVI disminuye efectos topográficos y produce una escala lineal de medición

que va de  $-1$  a  $1$ . Esta normalización, reduce el impacto de la degradación de calibración del sensor y la influencia de los efectos atmosféricos. El valor cero indica aproximadamente áreas en donde comienza la falta de vegetación, los valores negativos indican áreas faltas de vegetación. Los valores muy bajos de NDVI, del orden de  $0.1$ , comúnmente indican áreas arenosas, rocosas o nevadas. Valores de  $0.2$  a  $0.3$  suelen indicar zonas con pasturas naturales o arbustos. Después de estos valores de NDVI, se encuentran los niveles que indican zonas con praderas o cultivos, según el valor obtenido Yao et al. (2018). NDVI tiene la desventaja de ser sensible a la reflectancia del suelo sobre el que se sitúa la planta, esto limita su potencial de discriminación, lo cual representa un problema bastante grave para cuando la cubierta vegetal analizada es menor al  $50$ .

### 2.3.2.2. Índice de Vegetación ajustado al Suelo (SAVI)

El índice SAVI presentado en la Ec. 2.2, es usado para corregir la influencia del brillo del suelo en áreas donde la cobertura vegetativa es baja en el índice NDVI. Con la finalidad de reducir el impacto del brillo del suelo presente en el índice NDVI, el índice SAVI obtiene los efectos del suelo en el NDVI por medio de la constante  $L = 0.428$  que toma valores de  $0$  a  $1$  inversos a la densidad de la vegetación (Huete, 1988).

$$SAVI = \frac{(NIR - RED)}{(NIR + RED + L)}(1 + L). \quad (2.2)$$

El índice SAVI está ajustado a estudios de análisis de vegetación en etapas de crecimiento inicial o vegetación dispersa (Ren, Zhou, y Zhang, 2018; Ma et al., 2018).

### 2.3.2.3. Índice de Área Foliar (LAI)

LAI hace referencia a la cantidad de masa foliar de una cubierta vegetal, por lo que se trata de un parámetro biofísico que por tanto se refiere a la capacidad de fotosíntesis relacionada con el conjunto de pigmentos de la cubierta vegetal (Navarro-Cerrillo, 2013). LAI se calcula de la siguiente manera:

$$LAI = -\frac{\log \frac{0.69 - SAVI}{0.59}}{0.91} \quad (2.3)$$

#### 2.3.2.4. Índice de Vegetación de diferencia normalizada Verde (GNDVI)

El GNDVI hace referencia a la variación de clorofila en la cubierta vegetal de un cultivo, este es una variación del NDVI utilizando la banda verde (GREEN) en lugar de la banda roja. La formula del indice puede observarse en la Ec. 2.4. Este índice es utilizado normalmente para identificar el gasto de agua y nitrógeno de la cubierta vegetal del cultivo (Gitelson et al., 1996).

$$GNDVI = \frac{(NIR - GREEN)}{(NIR + GREEN)}. \quad (2.4)$$

Los valores obtenidos por dicho índice van entre -1 y 1. Valores entre -1 y 0 están ligados a la presencia de agua o suelo desnudo (Cárdenas et al., 2018).

#### 2.3.2.5. Índice de Vegetación Perpendicular (PVI)

Este índice fue desarrollado por Perry Jr y Lautenschlager (1984), con el objetivo de reducir los efectos causados por el brillo del suelo en cultivos con baja densidad de vegetación. PVI parte por utilizar la distancia perpendicular desde cada coordenada de píxel presente en la imagen, hasta la línea del suelo; la línea del suelo desnudo se obtiene a partir de la reflectividad en el infrarrojo cercano y el rojo de una serie de píxeles de suelo desnudo localizados en la imagen, el cálculo se hace mediante un análisis de regresión en el cual la banda roja se toma como variable independiente y la banda infrarroja como variable dependiente, ver Ec. 2.5.

$$PVI = \frac{(aNIR - RED + b)}{\sqrt{(a^2 + 1)}}. \quad (2.5)$$

Donde  $a$  es la pendiente de la línea del suelo y  $b$  es el gradiente de la línea del suelo.

De esta manera las zonas sin vegetación serán representadas sobre la línea del suelo. Si existe cierta cantidad de vegetación, la reflectividad de esa zona en la banda roja será más baja que la del suelo desnudo, y en el infrarrojo será más alta, por lo que el punto estará representado alejado de la línea del suelo. Sin embargo, dado a que el índice PVI depende de la exactitud con la que se defina la línea del suelo, este puede exhibir una dispersión elevada. Además, este índice sólo debe ser utilizado en imágenes con corrección atmosférica, ya que es sensible a los cambios atmosféricos.

Sin embargo, debido a que PVI depende de la precisión con que se determine la línea del suelo, puede presentar una elevada dispersión. Además, este índice sólo debería ser utilizado en imágenes que se han corregido atmosféricamente, ya que es sensible a los cambios atmosféricos.

## 2.4. Imágenes Satelitales

### 2.4.1. Resolución espacial

La resolución espacial define el tamaño mínimo de los elementos distinguibles en una imagen. A menor es el tamaño del píxel, mejor resolución tendrá la imagen y será posible discriminar más objetos (Sobrino, 2001). Sentinel-2 captura imágenes en tres resoluciones distintas, 10, 20 y 60 metros por píxel. La resolución espacial es un elemento muy importante a la hora de caracterizar la potencialidad de un sensor, pues define el nivel de detalle que ofrece la imagen.

### 2.4.2. Resolución espectral

Resolución espectral se refiere al número y ancho de bandas espectrales que puede captar un sensor. Cuanto más estrechas sean estas bandas mayores será la resolución espectral (Sobrino, 2001), en la figura 2.3 se ejemplifica los canales

espectrales del satélite Sentinel-2, en donde el eje  $y$  representa en metros, las tres resoluciones espaciales del sensor y el eje  $x$  representa en nanómetros, la longitud de onda central del espectro electromagnético. La radiación electromagnética (EMR) se define como toda energía que se mueve a la velocidad de la luz en un patrón de onda armónica (es decir, todas las ondas están espaciadas igual y repetidamente en el tiempo). La luz visible (VIS, por sus siglas en inglés), el infrarrojo cercano visible (VNIR, por sus siglas en inglés), los infrarrojos de onda corta (SWIR, por sus siglas en inglés), son algunas de las categorías del EMR. Cuando un sensor es capaz de registrar radiación electromagnética en diferentes longitudes de onda se dice que este es un sensor multispectral. Estos dispositivos cuentan con características especiales que se ven reflejados en una buena resolución de las capturas.

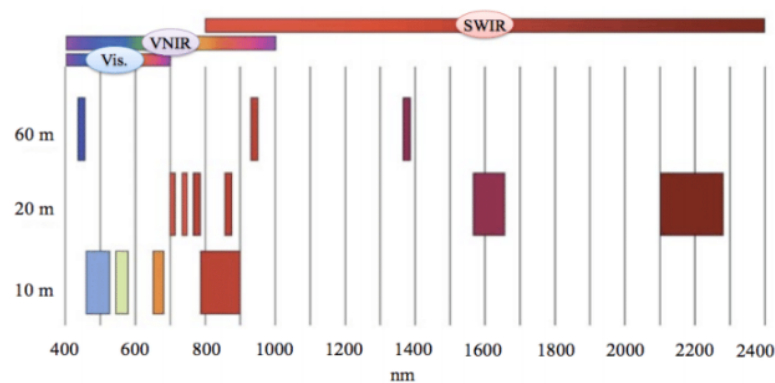


FIGURA 2.3: Características de la resolución espacial y espectral de Sentinel-2.

Fuente: (Rivera Caicedo, 2014)

### 2.4.3. Resolución temporal

Al tiempo que le toma a una plataforma el volver a visitar una ubicación geográfica específica se conoce como su resolución temporal. Generalmente, la resolución temporal, o tiempo de revisión, se expresa en términos de días (Khorram et al., 2012). Por ejemplo, Sentinel-2 entrega datos tomados sobre todas las superficies terrestres y zonas costeras cada cinco días en condiciones libres de nubes, y generalmente cada 15-30 días considerando la presencia de nubes (Rivera Caicedo, 2014).

#### 2.4.4. Series temporales

Una serie temporal es definida como la reunión de observaciones de una variable obtenidas secuencialmente en el tiempo (Weigend, 2018). Estas observaciones generalmente se obtienen en momentos de tiempo equiespaciados. Un ejemplo es el comportamiento de algún índice de vegetación o parámetro biofísico como el LAI durante el crecimiento de un cultivo, con la finalidad de monitorizarlo.

#### 2.4.5. Sentinel-2

La misión Sentinel-2 consiste de una constelación de dos satélites en órbita polar colocados en la misma órbita, en fases de  $180^\circ$  entre sí, en la figura 2.4 se muestra el diseño de ambos satélites. El objetivo de ambos satélites es monitorear la variabilidad en las condiciones de la superficie terrestre, con su gran ancho de franja y alto tiempo de revisión (10 días en el ecuador con un satélite y 5 días con dos satélites en condiciones sin nubes, lo que da como resultado 2-3 días a mediados de latitudes).

Una de las ventajas principales de este satélite, es el acceso gratuito a la información censada, el usuario puede descargar las imágenes capturadas desde la plataforma de la ESA. Otra de sus ventajas es que ofrece una amplia resolución espacial en las capturas, que van desde 10, 20 y hasta 60 m, en 13 bandas del espectro electromagnético (ver tabla 2.1), lo que permite captar mejor los detalles de la cubierta del planeta.



FIGURA 2.4: Sentinel-2

Fuente: ESA

TABLA 2.1: Resolución de las bandas del Sentinel-2

Banda	Intervalos de longitud de onda	Resolución de píxel
1	443 nm	60 m
2	540 nm	10 m
3	560 nm	10 m
4	665 nm	10 m
5	705 nm	20 m
6	740 nm	20 m
7	783 nm	20 m
8	842 nm	10 m
8a	965 nm	20 m
9	940 nm	60 m
10	1375 nm	60 m
11	1610 nm	20 m
12	2190 nm	20 m

Aprovechando el tipo y calidad de datos brindados por el programa Copernicus, empresas y universidades desarrollan herramientas para utilizar esta información y poder ofrecer servicios y generar conocimiento sobre la cubierta vegetal del planeta.

## 2.5. Aprendizaje Automático

### 2.5.1. Regresión

El análisis de regresión crea modelos matemáticos que examinan la relación de una variable dependiente con una o más variables independientes. Estos modelos se pueden utilizar para predecir respuestas a valores no observados y/o futuros de las variables independientes (James et al., 2013).

#### 2.5.1.1. Smoothing Splines.

Las *smoothing splines* son estimaciones de funciones,  $\hat{f}(x)$  obtenidas de un conjunto de observaciones ruidosas  $y_i$  del objetivo  $f(x_i)$ , con el objetivo de equilibrar una medida de bondad de ajuste (De Boor et al., 1978). Estas proporcionan una

forma suavizar datos ruidosos en una serie. El ejemplo más conocido es el *cubic smoothing spline*, pero existen muchas otras.

### 2.5.1.2. Cubic smoothing spline.

Esta regresión se basa en ajustar una función polinomial de grado 3 entre cada dato de tal manera que haya un equilibrio entre este ajuste y una suavidad especificada (De Boor et al., 1978). *cubic smoothing spline* es la clase de funciones que minimiza la Ec. 2.6.

$$\frac{1}{n} \sum_{i=1}^n \{Y_i - f(x_i)\}^2 + \lambda \int f''(x)^2 dx \quad (2.6)$$

## 2.5.2. Clasificación

La clasificación trata los problemas de identificar elementos en una de  $K$  clases. Las variables se pueden caracterizar como cuantitativas o cualitativas (también conocidas como categóricas). Las variables cuantitativas toman valores numéricos. Los ejemplos incluyen la edad, la altura o los ingresos de una persona, el valor de una casa, la categoría y el precio de una acción. En contraste, las variables cualitativas toman valores en una de  $K$  clases o categorías diferentes. A continuación se presentan algunos de los clasificadores más utilizados en el área de la teledetección.

**k-Nearest Neighbors.** Dado un entero positivo  $k$  y una observación de prueba  $x_0$ , el clasificador KNN identifica primero los puntos  $k$  en los datos de entrenamiento que están más cerca de  $x_0$ , representados por  $N_0$ . Luego estima la probabilidad condicional para la clase  $j$  como la fracción de puntos en  $N_0$  cuyos valores de respuesta son iguales a  $j$  (James et al., 2013).

**Random Forest.** Random Forest se trata de un algoritmo de clasificación el cual está constituido por muchos árboles de decisión. Este método de clasificación utiliza *bagging* y la aleatoriedad de las características para construir cada árbol de decisión para crear un bosque de árboles no correlacionado y cuya predicción por



comité es más precisa que la de cualquier árbol de decisión de manera individual (James et al., 2013).

**Support Vector Machine.** Este método de clasificación se basa en encontrar un hiperplano en un espacio  $N$ -dimensional ( $N$  - el número de características) el cual pueda clasificar claramente los puntos de datos en el vector de características. Para separar las clases de puntos de datos, hay muchos hiperplanos posibles que podrían elegirse, el objetivo es encontrar un hiperplano el cual contenga el margen máximo, es decir, la distancia máxima entre los puntos de datos de cada clase (James et al., 2013).

**Naïve Bayes.** Es un modelo probabilístico de aprendizaje automático que se basa en el teorema de Bayes. Usando el teorema de Bayes, podemos encontrar la probabilidad de que ocurra  $A$ , dado que  $B$  ha ocurrido. Aquí,  $B$  es la evidencia y  $A$  es la hipótesis. La suposición hecha aquí es que las características son independientes. Es decir, la presencia de una característica particular no afecta a la otra (Rish y others, 2001).

### 2.5.3. Métodos de remuestreo

Los métodos de remuestreo consisten en extraer muestras desde un conjunto de entrenamiento y reajustar un modelo elegido sobre cada muestra para conseguir información complementaria sobre el modelo ajustado. Por ejemplo, para estimar la variabilidad de un ajuste de regresión lineal, se puede extraer repetidamente muestras diferentes de los datos de entrenamiento, ajustar una regresión lineal a cada nueva muestra y luego examinar hasta qué punto difieren los ajustes resultantes. Tal enfoque permite obtener información que no estaría disponible al ajustar el modelo solo una vez usando la muestra de entrenamiento original. A continuación se describen uno de los métodos de remuestreo más utilizados, *Cross-validation* Por ejemplo, para estimar la variabilidad de un ajuste de regresión lineal, se puede extraer repetidamente muestras diferentes de los datos de entrenamiento, ajustar una regresión lineal a cada nueva muestra y luego examinar hasta qué punto difieren los ajustes resultantes. Tal enfoque permite obtener información que no estaría disponible al ajustar el modelo solo una vez usando la

muestra de entrenamiento original. A continuación se describen uno de los métodos de remuestreo más utilizados, *Cross-validation*.

**Cross-validation.** Este método parte por dividir de manera aleatoria el conjunto de observaciones en  $k$  grupos, o pliegues, de aproximadamente la misma dimensión. El primer pliegue es tratado como un conjunto de validación y el método se ajustara a los  $k-1$  pliegues restantes. A partir de las observaciones en el pliegue extendido se calcula el error cuadrático medio,  $MSE_1$ . Este proceso se reitera  $k$  veces; por cada vez, un grupo diferente de observaciones se trata como un conjunto de validación. Este proceso da como resultado  $k$  estimaciones del error de prueba,  $MSE_1, MSE_2, \dots, MSE_k$ . La estimación de *k-fold cross-validation* se calcula promediando estos valores (James et al., 2013).

#### 2.5.4. Métricas

*Accuracy* representa la relación entre el número de predicciones correctas y el número total de muestras utilizadas (Hackeling, 2017), ver Ec. 2.7.

$$accuracy = \frac{tp + tn}{ts} \quad (2.7)$$

donde:  $tp$  son los positivos verdaderos, es decir los casos en los que el clasificador predijo una clase y ésta instancia realmente pertenecía a esa clase;  $tn$  son los negativos verdaderos, es decir los casos en los que el clasificador predijo que la instancia no pertenecía a una clase y esta realmente no pertenecía a esa clase; finalmente,  $ts$  es el total de instancias del conjunto.

Por otra parte, *precision* cuya fórmula se muestra en la Ec. 2.8, denota la proporción de casos positivos predichos que son correctamente positivos reales (Powers, 2007).

$$precision = \frac{tp}{tp + fp} \quad (2.8)$$

donde  $fp$  son los positivos falsos, es decir, los casos en los que el clasificador predijo una instancia a una clase que no era la correcta.

El *recall* es la proporción de casos positivos reales que se predicen correctamente (Powers, 2007), ver Ec. 2.9.

$$recall = \frac{tp}{tp + fn} \quad (2.9)$$

donde *fn* son los falsos negativos, es decir, las instancias de una clase clasificadas incorrectamente en otras clases.

Finalmente, el *F – measure* es una métrica que combina la *precision* y *recall* en la media armónica de estos (Hackeling, 2017), ver Ec. 2.10.

$$F - measure = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (2.10)$$

# Capítulo 3

## Antecedentes

### 3.1. Fenología

El uso de plataformas espaciales, como los satélites Sentinel-2 con su alta resolución temporal y resolución espacial, han contribuido significativamente al estudio y desarrollo de aplicaciones agrícolas, mediante el registro de la energía electromagnética que interactúa con la vegetación. Una área que ha cobrado reciente interés científico es el de la teledetección de la fenología vegetal como indicador del cambio climático y para cartografiar la cobertura del suelo. El inicio de la temporada (SOS, por sus siglas en inglés), el pico mas alto (POS, por sus siglas en inglés) y el final de la temporada (EOS, por sus siglas en inglés), son algunos de los parámetros generales que describen las tendencias de la fenología. A continuación se presentan algunos trabajos de investigación relacionada con la fenología en bosques, cultivos y pastizales utilizando datos de Sentinel-2.

#### 3.1.1. Fenología de bosques

El uso de datos Sentinel-2 cada vez se esta volviendo mas popular para los estudiar la fenología de los bosques gracias a su resolución espacial y temporal que han ayudado a superar algunas de las limitaciones de la teledetección óptica (Lange et al., 2017).

Un ejemplo de esto se encuentra en (Kollert et al., 2021), donde se identificaron tres grupos de especies arbóreas rodiales de hoja ancha, pino, pino enano y abeto en la región de los Alpes de Europa del Este. El método propuesto utiliza el sensor MSI (bandas infrarrojo, rojo y verde). Esta adquisición de Sentinel-2 fue seleccionada debido a su baja cantidad de nubosidad. Posteriormente se utilizó un mapa generado en el software de información geográfica SAGA GIS, el cual contiene las zonas en las cuales se encuentran los bosques de los tres grupos de especies arbóreas. El procedimiento resultó en un total de alrededor de 47,000 píxeles. Después los datos obtenidos se partitionaron de la siguiente manera: primero se crearon medias compuestas, es decir se tomaron todos los píxeles del año sensado y se dividieron en trozos los cuales abarcan 3 meses y a partir de ahí se obtuvo la media, el proceso se repitió solo que los trozos fueron por cada mes. En segundo lugar se generaron series de tiempo en base al índice de vegetación NDVI por cada píxel de la zona y se obtuvieron las métricas fenológicas de inicio y final de temporada para el 10 %, 20 % y 50 % de la amplitud anual del NDVI, del pico de temporada y la amplitud anual. Con estas características se entrenó el clasificador *Random Forest*, con el cual se generaron varios modelos configurándolos de distinta manera con las características para medir el impacto de estas, estos modelos se evaluaron mediante *Cross Validation*. Los resultados obtenidos para el mejor modelo arrojaron una precisión del 87.0 % y el incluir las métricas fenológicas solo mejoró de un 1 % a un 2 %. Finalmente se concluye que a pesar de que el método propuesto no mejora mucho el mapeo respecto a otras técnicas, se recomienda considerar las métricas fenológicas y medias compuestas cuando la zona a sensar sea nubosa.

Adicionalmente, en (Kowalski et al., 2020) se propone el detectar la fenología en bosques latifoliados templados, con la finalidad de medir y cuantificar las respuestas fenológicas de estos a la variabilidad climática. El método propuesto utiliza series temporales combinando datos Sentinel-2 y Landsat para modelar y caracterizar la fenología primaveral de estos bosques en Alemania y estimar el inicio de la temporada (SOS) para el año 2018. se probaron dos enfoques de modelado estadístico (modelos logísticos y aditivos generalizados) y los índices de vegetación NDVI y el índice de vegetación mejorado EVI. Los resultados mostraron que las estimaciones del SOS de los modelos logísticos tuvieron un mayor impacto al elegir el índice que el método de ajuste elegido ( $rEVI=86\%$ ;  $rNDVI=65\%$ ) respecto a las observaciones terrestres. Se concluye la importancia de medir cambios fenológicos

causados por el cambio climático, pues alteran la estructura, composición y función de los ecosistemas en todo el mundo, además de enfatizar en el potencial de uso de datos Sentinel-2 y Landsat en la estimación del SOS anualmente.

### 3.1.2. Fenología en cultivos

El análisis de la fenología de cultivos es diferente al de los bosques en varios aspectos. Aunque como se pudo ver en los artículos incluidos para el estudio de la fenología en bosques, en ambos se requiere y es importante diferenciar las señales de fondo (por ejemplo, del suelo desnudo) de la vegetación, mediante el uso de los índices de vegetación. La diferencia se da debido a que el análisis de las curvas de las series temporales para los cultivos pueden ser particularmente desafiantes debido a las prácticas de manejo. Además, las curvas de las series temporales de algún índice de vegetación observadas en áreas con alta intensidad de cultivo, o más de una temporada de cultivo al año, contrastan fuertemente con las curvas de series de tiempo de los bosques. Por lo tanto, el procesamiento de datos de series de tiempo debe ajustarse en consecuencia y permitir disminuciones en los valores del índice de vegetación que indiquen eventos tales como la cosecha (Atkinson et al., 2012). Si bien estos problemas presentan un desafío para cualquier enfoque de la fenología de cultivos por teledetección, las resoluciones espaciales, temporales y espectrales de Sentinel-2 ofrecen mejoras potenciales que han sido aprovechadas en diferentes estudios.

Tal es el caso de (Wang et al., 2019b), en donde se utilizaron datos MSI en conjunto con la plataforma *Google Earth Engine* para identificar zonas sembradas con caña de azúcar en áreas complejas con frecuentes condiciones climáticas lluviosas, el método propuesto parte por obtener imágenes por cada una de las 3 etapas fenológicas de la caña de azúcar (macollamiento, rápido crecimiento y maduración), después se calculo el índice de vegetación NDVI por cada etapa. Finalmente, se genero un conjunto de datos NDVI de tres bandas junto con 4,000 muestras de entrenamiento y 2,000 muestras de validación aleatorias para el mapeo de la caña de azúcar. Para medir la robustez del método se utilizaron cinco clasificadores basados en aprendizaje automático, que incluyen dos clasificadores support vector machine (Polynomial-SVM y RBF-SVM), un clasificador random forest (RF), un clasificador de red neuronal artificial (ANN) y un clasificador de árbol de decisión

(CART-DT), la zona de estudio se ubica en el condado de Longzhou de la provincia de Guangxi de China. Los resultados mostraron que, a excepción del clasificador ANN, los clasificadores Polynomial-SVM, RBF-SVM, RF y CARTDT mostraban mapas resultantes de caña de azúcar con una precisión del 91 %. Finalmente, se enfatizó en la importancia de considerar la fonología en el mapeo rápido de la caña de azúcar y sugiere el potencial de los datos Sentinel-2 y los enfoques de aprendizaje automático en la gestión del uso de la tierra y la toma de decisiones de alta precisión.

Otro estudio es el de Tian et al. (2020), en el que se estudia el impacto en los cambios del microclima inducidos por grandes ciudades en la fonología del trigo de invierno. Dicho estudio se realizó en diferentes ciudades de China con distintas extensiones de tierra. El método calcula el índice de vegetación NDVI obtenido de los datos del satélite Sentinel-2 y se extrajo el inicio de la temporada (SOS) para cinco periodos de crecimiento completos del año 2015-2020. El estudio tuvo tres resultados principales: (1) El SOS muestra una tendencia creciente significativa a lo largo de los cinco periodos monitorizados tanto en Shijiazhuang como en Baoding, lo que sugiere que los aumentos de temperatura inducidos por las zonas urbanas avanzan la fenología primaveral del trigo de invierno. (2) El tamaño máximo de influencia de los efectos de la temperatura inducidos por la ciudad en el SOS se correlaciona positivamente con el tamaño de la ciudad, es decir, 27 km para Shijiazhuang, 14 km para Baoding y 7 km para Linqing y (3) La tasa de cambio del SOS respecto al tamaño de la ciudad es significativamente mayor en la ciudad más grande (Shijiazhuang: 0,26 días/km) que en las ciudades de mediana y pequeña escala (Baoding: 0,21 día / km y Linqing : 0,11 día/km), lo que sugiere que las ciudades más grandes propagan el calor a un ritmo más rápido que el de las ciudades más pequeñas. Finalmente, se concluye que la siembra y el manejo del trigo de invierno alrededor de las ciudades deben considerar la influencia del tamaño de la ciudad para optimizar los rendimientos.

### 3.1.3. Fenología en pastizales

Los pastizales son una fuente de recursos valiosos al ecosistema en donde se producen y representan más de un tercio de todas las tierras agrícolas del planeta (Hill, 2013; Watson, Restrepo-Coupe, y Huete, 2019). Caracterizar la fenología de

los pastizales es particularmente difícil, ya que son sensibles tanto a las prácticas de manejo (es decir, frecuentes eventos de pastoreo y corta) como el clima (Wang et al., 2019a).

En (Gómez-Giráldez et al., 2020) se plantea un método que relaciona la fonología de los pastizales anuales con la dinámica del agua presente en el suelo en una granja denominada “Santa Clotilde” ubicada en España del sur, para ello se utilizó una combinación de fotografía terrestre, imágenes de satélite (Sentinel-2) y mediciones hidrológicas terrestres. A partir de estos se obtuvieron parámetros fenológicos clave del ciclo del pastizal a partir de datos de cámaras terrestres utilizando el índice de coordenadas cromáticas verdes (GCCc, por sus siglas en inglés). Estos parámetros se compararon con los proporcionados por series temporales de los índices de vegetación (NDVI, GNDVI, EVI, SAVI) obtenidos del satélite Sentinel-2 y series temporales de variables abióticas (temperatura del aire, radiación solar entrante y saliente, déficit de presión de vapor, precipitación, humedad volumétrica del suelo), que definieron la hidrología del sistema. El ajuste se llevó a cabo mediante una matriz de correlación de Pearson y el análisis de componentes principales (PCA, por sus siglas en inglés). Los resultados mostraron que los parámetros fenológicos estimados por el NDVI y la humedad del suelo presentaron el mejor ajuste con las observaciones terrestres comparadas a los proporcionados por otros índices de vegetación y variables abióticas. Se concluye que es posible monitorear el estado hídrico del suelo a partir de los parámetros fenológicos obtenidos con NDVI de imágenes Sentinel-2.

## 3.2. Anomalías

Adicionalmente al estudio de la fenología, el presente trabajo de tesis trata la identificación de anomalías en la caña de azúcar. Según Abdel-Rahman y Ahmed (2008), las técnicas de percepción remota han sido utilizadas de manera exitosa en estudios relacionados con cultivos de caña de azúcar, un ejemplo es el monitoreo de deficiencias nutricionales y estrés hídrico. La mayoría de estos estudios se han realizado en grandes extensiones de cultivos utilizando sensores con baja resolución, i.e., 30m por píxel.



A partir del lanzamiento de los satélites Sentinel-2, al igual como ocurre con la fenología, estos han sido aprovechados por la comunidad científica y se han registrado diversos avances en el monitoreo de la salud de los cultivos. Un ejemplo de esto es el estudio de Rodríguez Ramírez (2019), en donde el estudio se enfocó en obtener mapas LAI de Ameca Jalisco, utilizando imágenes satelitales proporcionadas por Sentinel-2 mediante una cadena de procesamiento, la contribución está en que estos mapas pueden ser utilizados en conjunto con el contenido de clorofila en hojas (CHL, por sus siglas en inglés) para conocer la etapa fenológica y estado de salud en que se encuentra el cultivo.

Susantoro et al. (2018) determinaron los mejores índices de vegetación que pueden mapear el estado de la plantación de caña de azúcar Jatitujuh que está rodeada por campos de petróleo y gas, en Regencia Indramayu y Majalengka en la provincia de Java Occidental Indonesia, el objetivo fue el detectar condiciones anómalas en la caña de azúcar ocasionadas por el petróleo y gas determinando cual índice de vegetación representa mejor esta condición, en total se analizaron 23 índices los cuales fueron evaluados usando desviación estándar y regresión lineal, se utilizaron imágenes proporcionadas por Landsat-8, los resultados arrojaron que los índices que mejor mapean las condiciones del cultivo de caña bajo las condiciones que generan los campos de petróleo y gas en Jatitujuh son cinco : NDVI y la versión mejorada (ENDVI, por sus siglas en inglés), GDVI, LAI y el índice de pigmento insensible a la estructura (SIPI, por sus siglas en inglés).

Los estudios antes mencionados muestran un vistazo al gran impacto que tuvo el lanzamiento de Sentinel-2, pues la mayor resolución espacial, espectral y temporal de esta plataforma es preferible a muchas de las otras plataformas espaciales de datos ópticos disponibles gratuitamente, ya que ha mostrado capacidades mejoradas para el mapeo de la vegetación y la estimación de la fenología. Sin embargo un área que aún no se ha explorado lo suficiente es la identificación de las etapas fenológicas de la caña de azúcar (amacollamiento, rápido crecimiento y maduración) en México, y esto representa una gran oportunidad de contribuir al desarrollo de mejores técnicas de manejo en campo, pues al conocer la etapa fenológica en la que se encuentra la caña de azúcar se puede comprender como la caña percibe y

responde a su entorno, además de comprender los cambios morfológicos y anatómicos que se presentan durante su desarrollo vegetativo, lo que permite el identificar de etapas críticas que pueden ser de utilidad en la elaboración de medidas para mejorar el rendimiento del cultivo (Endres et al., 2018) en campo. En este trabajo se propone un método para identificar la etapa fenológica de la caña de azúcar además de identificar anomalías en su crecimiento.

# Capítulo 4

## Metodología y materiales

En este capítulo se presentan dos metodologías diseñadas para el monitoreo de la caña de azúcar. La primera, aborda un método que identifica las etapas fenológicas del cultivo de caña de azúcar en el ciclo de corte resoca 1, esto a través el uso de series temporales construidas a partir de diversos índices de vegetación y el LAI. La segunda, presenta un método para identificar anomalías en el cultivo de caña de azúcar a través de un perfil de salud construido con el índice PVI y el LAI de un cultivo de caña en una zona con control detallado del manejo del cultivo. En las figuras 4.1 y 4.2, se muestran las fases que conforman cada una de las metodologías.

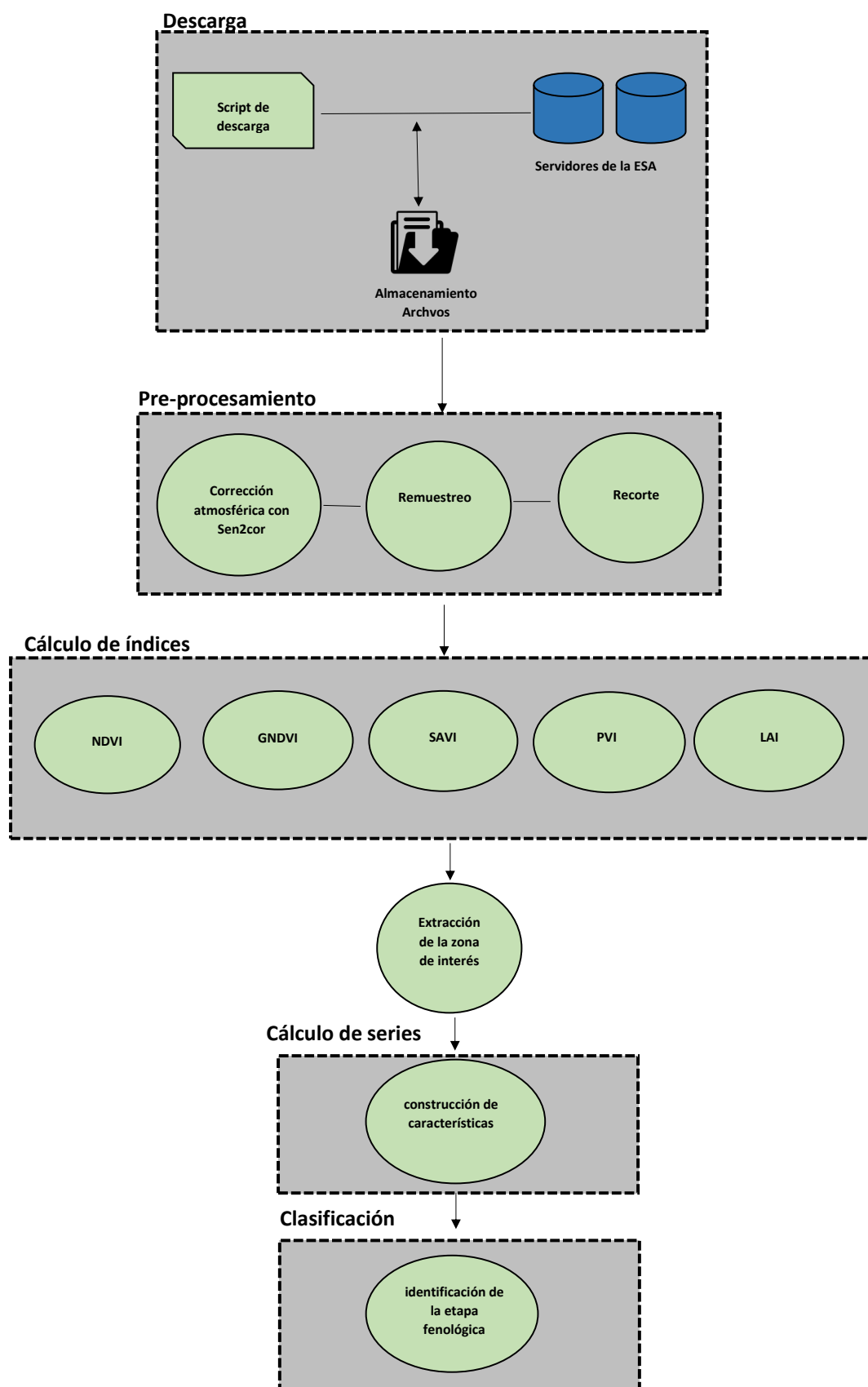


FIGURA 4.1: Metodología para identificar las etapas fenológicas en caña de azúcar

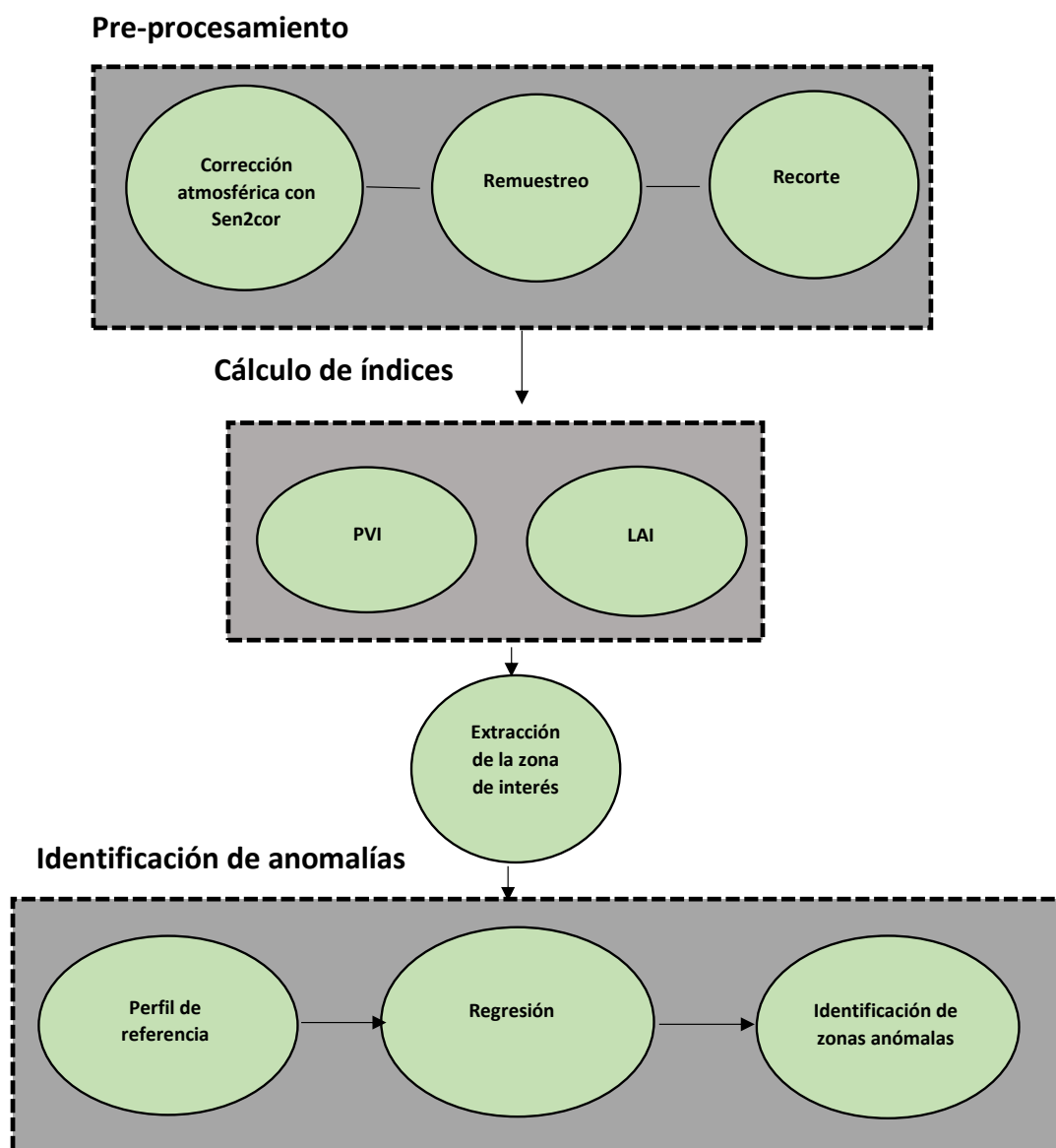


FIGURA 4.2: Metodología para identificar anomalías en caña de azúcar

Todas las fases de las metodologías están codificadas en *Python* ya que el lenguaje cuenta con una gran cantidad de bibliotecas que facilitan las tareas de identificación, manipulación y visualización, con librerías como *numpy*, *sklearn*, *GDAL*, *matplotlib*, entre otras.

A continuación se describe cada una de las fases de las metodologías.

## 4.1. Identificación de las etapas fenológicas del cultivo de caña de azúcar

La identificación de las etapas fenológicas del cultivo de caña de azúcar esta conformada por cinco fases, en la primera fase se desarrolla la descarga de las imágenes satélites, en la segunda se realiza un pre-procesado a las imágenes, en la tercera se calculan índices de vegetación y parámetros biofísicos, la cuarta consiste en construir series temporales y finalmente en la etapa cinco se identifica la etapa fenológica del cultivo mediante el uso de clasificadores.

### 4.1.1. Descarga

Para la descarga de los datos del sensor MSI abordo del satélite Sentinel-2, se elaboro un script en Python, este se conecta al repositorio de *Sentinels Scientific Data Hub* de la ESA, este script como parámetros de entrada recibe el periodo de tiempo para la descarga de las imágenes, la zona sobre la cual se descargarán las imágenes y el nivel de nubosidad deseado.

### 4.1.2. Pre-Procesamiento

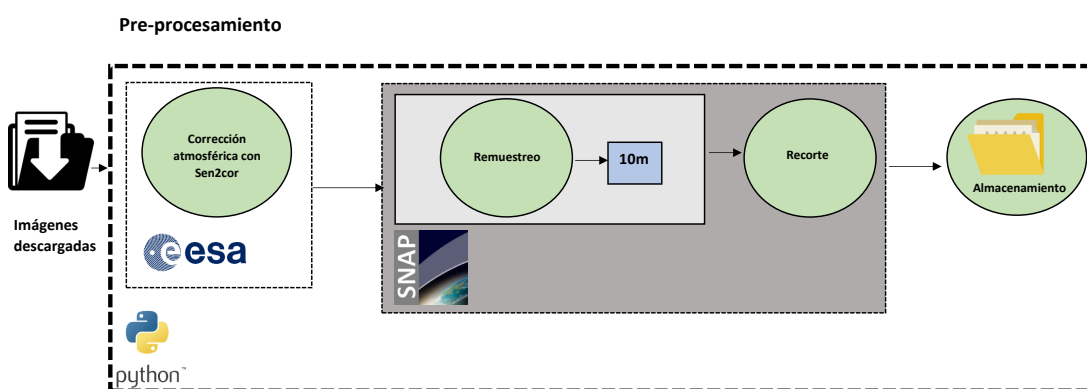


FIGURA 4.3: Fase de preprocesamiento

La etapa de pre-procesamiento esta compuesto por tres funciones (1) corrección atmosférica, (2) remuestreo y (3) recorte; cada función se explica a continuación.

La corrección atmosférica es un proceso fundamental en la extracción de datos en imágenes satelitales, esto es porque la energía reflejada por un cultivo agrícola en su trayecto hacia el sensor, recibe contribuciones de la interacción de la energía con la atmósfera, estas provienen del vapor de agua y los gases en la atmósfera; estas contribuciones de energía se consideran como ruido, el cual tiene repercusiones negativas en la extracción de datos. Por lo tanto, es importante eliminar la mayor cantidad del ruido para obtener la energía reflejada por los cultivos. La corrección atmosférica se realizó con la herramienta Sen2Cor<sup>1</sup>, la cual fue desarrollada por la ESA.

Posteriormente, la función de remuestreo homogeneiza la resolución espacial de las imágenes (detalles visibles en la imagen). Las imágenes descargadas se encuentran en tres distintas resoluciones espaciales, ver Tabla 2.1; lo que significa que un píxel en la imagen descargada representa 10, 20 o 60 metros reales del terreno sensado. Cuanto menor sea el área terrestre representada por cada píxel, mayores serán los detalles captados Khorram et al. (2016).

Con la finalidad de obtener un mayor detalle de los cultivos de caña de azúcar, se selecciono la resolución de 10 m, por lo tanto fue necesario homogeneizar las imágenes de 20 m y 60 m a 10 m. Dicho tratamiento digital se hizo con la herramienta SNAP<sup>2</sup>.

En seguida, las imágenes se recortaron utilizando la herramienta *subset* del software SNAP; la zona de estudio se delimito utilizando un polígono generado mediante la aplicación en línea GeoJSON<sup>3</sup>.

---

<sup>1</sup><https://step.esa.int/>

<sup>2</sup><http://step.esa.int/>

<sup>3</sup><http://geojson.io/>

### 4.1.3. Cálculo de índices

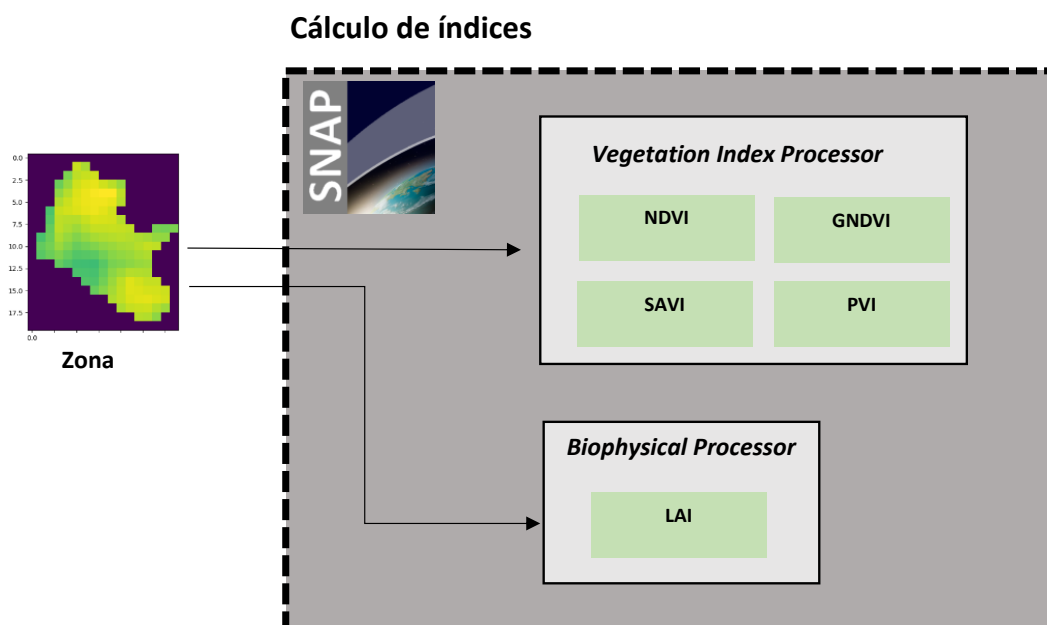


FIGURA 4.4: Fase de cálculo de índices

Una vez extraído los píxeles que comprenden la zona de interés, el método propone el uso de índices que permitan obtener características para clasificar los cultivos de caña de azúcar según su etapa fenológica. Los índices permiten incrementar las diferencias de reflectancia entre usos del suelo; los índices utilizados para la identificación de las etapas fenológicas fueron los siguientes:

- NDVI
- GNDVI
- SAVI
- PVI
- LAI

El cálculo de los índices considerados se llevo a cabo mediante el software SNAP, específicamente para los índices de vegetación NDVI, GNDVI, SAVI y



PVI, se realizó usando la función de *Vegetation Index Processor*; respecto al índice LAI, éste se obtuvo usando la función *Biophysical Processor*. La estimación del LAI se realiza mediante una red neuronal de tres capas con la siguiente topología: La capa de entrada que consiste de 11 neuronas (8 bandas y 3 ángulos de la geometría de observación e iluminación), la capa oculta con 5 neuronas y función de transferencia tipo tangente sigmoidal, ver Ec. 4.1, finalmente como capa de salida una neurona con una función de activación lineal.

$$f(x) = \frac{1}{1 - e^{-x}} \quad (4.1)$$

#### 4.1.4. Extracción de la zona de interés

Una vez que se han calculado los índices, es necesario extraer los píxeles que comprenden solo la zona a analizar, para ello se hace uso de un polígono en formato shapefile que contendrá las coordenadas de la zona de interés, este polígono se construye a través de la herramienta en línea GeoJSON, en la Figura 4.5 se ejemplifica la interfaz gráfica de esta herramienta.

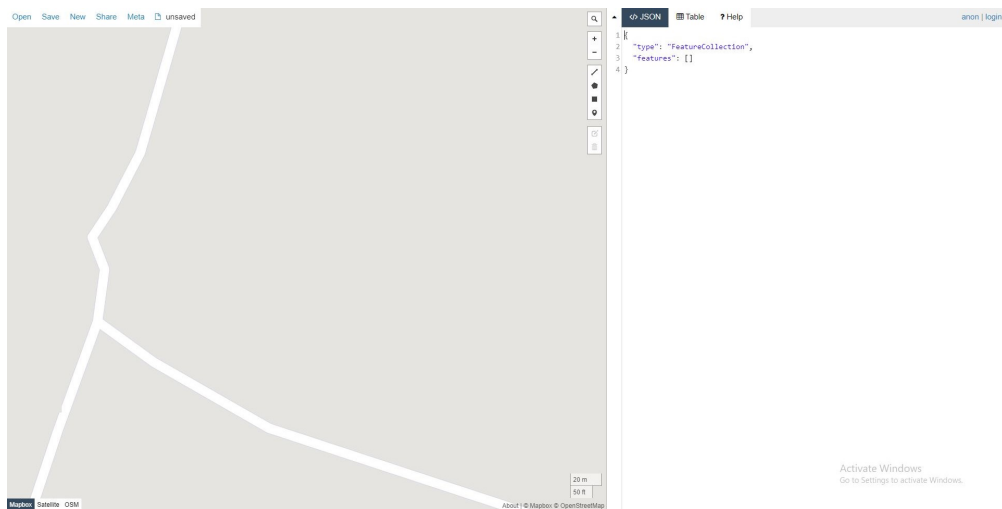


FIGURA 4.5: Interfaz gráfica de GeoJSON

### 4.1.5. Cálculo de series

Una vez obtenido solo los píxeles que comprenden la zona de interés, por cada índice (NDVI, GNDVI, SAVI, PVI y LAI) se crean tres series utilizando python, correspondientes a los valores máximo, medio y mínimo de cada índice obtenido.

En total, se obtienen 15 series, i.e., 15 características. La información que proporcionaron las series se utilizó para construir un conjunto de datos el cual contiene la etiqueta correspondiente a la etapa fenológica en cada fecha. Este conjunto de datos se utilizó para la etapa de clasificación, las series temporales se graficaron mediante el paquete *Matplotlib* de *Python*.

### 4.1.6. Clasificación

Los clasificadores considerados para esta etapa fueron los siguientes: *Naiïve Bayes* (NB), *Random Forest* (RF), *k-Nearest Neighbors* (KNN) y *Support Vector Machine* (SVM). La selección de estos clasificadores principalmente es porque son de las técnicas de clasificación más frecuentes y que han sido exitosas en estudios agrícolas. Por ejemplo, en Barbosa et al. (2015), se utilizaron los clasificadores RF y NB para definir la legitimidad de muestras de caña de azúcar orgánica; en Döker y Boşgelmez (2015) se propone el empleo de series temporales y el clasificador RF para monitorear cultivos de caña de azúcar; Guedes y Pereira Guedes y Pereira (2019) plantean el uso del clasificador KNN para identificar la caña de azúcar cruda en presencia de pequeñas cantidades de impurezas sólidas; recientemente, Wang et al. Wang et al. (2019b), proponen modelos para crear mapas de cultivos de caña de azúcar usando la integración de imágenes multitemporales del satélite Sentinel-2 y métodos de aprendizaje automático basados en SVM y RF.

La técnica de validación usada fue *k-fold cross-validation* con  $k=10$ .

Para evaluar los clasificadores las métricas que se consideraron fueron las siguientes: *accuracy*, *precision*, *recall*, y *F-measure*.

## 4.2. Identificación de anomalías

La metodología para la identificación de anomalías de la caña de azúcar incluye tres partes principales:

- Pre-Procesamiento.
- Cálculo de índices.
- Identificación de zonas anómalas.

Primero, las imágenes descargadas del sensor MSI se corrigen atmosféricamente, se homogeneizan resoluciones y se recortan para extraer el área de interés, y luego se calculan los índices LAI y PVI. Estas dos primeras fases se tomaron de la metodología para identificar las etapas fenológicas de la caña de azúcar. Finalmente en la fase 3 se define un umbral de un cultivo considerado sano para identificar zonas en la que otro cultivo se considere anómalo. Se utiliza un modelo de regresión con la finalidad de obtener datos para cada día del ciclo de cultivo sano, y posteriormente se utiliza el perfil para detectar zonas anómalas en otro cultivo.

A continuación se describe la fase 3 de la metodología.

### 4.2.1. Identificación de zonas anómalas

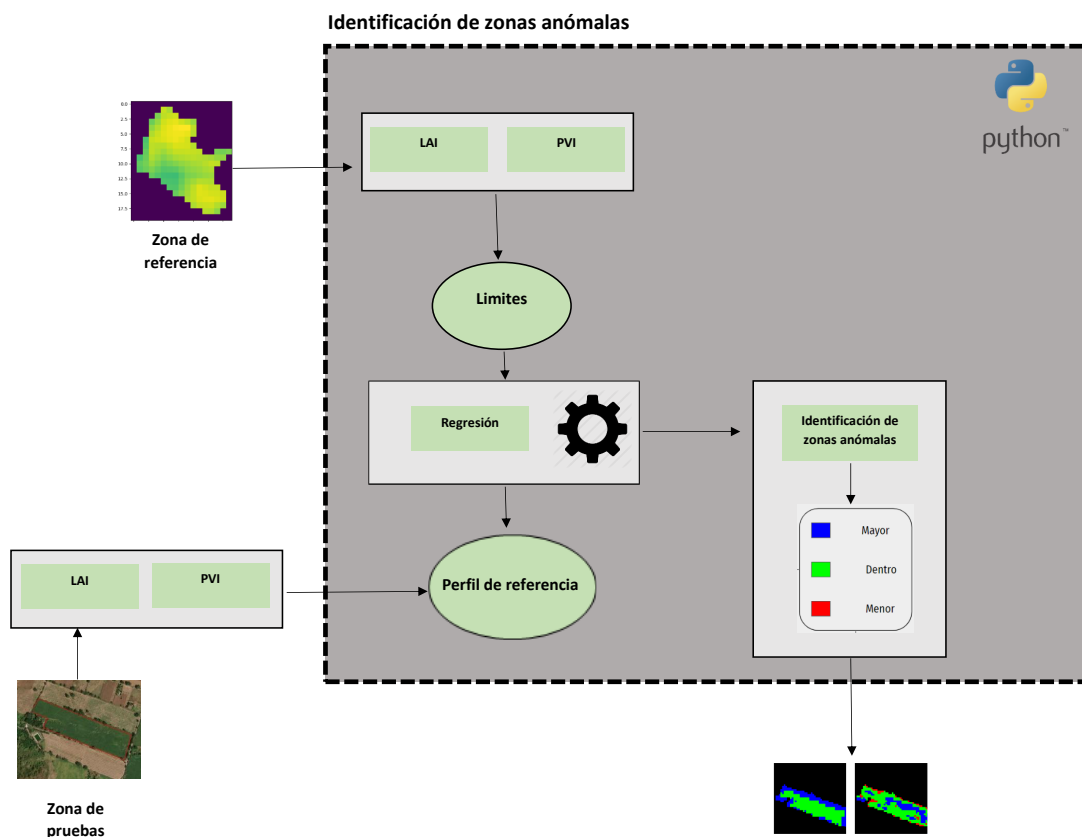


FIGURA 4.6: Fase de identificación de zonas anómalas

Esta fase parte por tomar los índices LAI y PVI calculados para una zona de referencia  $Z_1$ . Los valores de ambos índices en todo el ciclo de vida de la caña en  $Z_1$  se tomarán como referencia para determinar límite inferior y superior en donde el cultivo de caña de azúcar se considera normal, los límites se construyeron a partir de la media y la desviación estándar de la intensidad de los píxeles en  $Z_1$ , el límite inferior se calculó restando la desviación estándar de la media, y para el límite superior, la desviación estándar se sumó a la media.

Debido a que obtener datos para cada día del ciclo de vida del cultivo es imposible, debido a la resolución temporal de Sentinel-2 y a que en temporada de lluvias la presencia de nubes genera ruido, se utilizó un método de regresión para obtener datos de los días en los que no había imágenes del cultivo. El método de regresión utilizado fue *cubic smoothing spline*.

Para verificar el método propuesto, hemos monitorizado otra zona de caña de azúcar  $Z_2$ . Se calculan los índices PVI y LAI, y cada píxel de  $Z_2$  se compara con los límites correspondientes a ese día. Gráficamente, utilizamos un mapa de colores para definir la localidad dentro del área de interés en la que los índices están dentro o fuera del umbral. Se utiliza verde para aquellos píxeles que están dentro de los límites inferior y superior, se utiliza el color azul para aquellos píxeles en donde uno o ambos índices son mayores que el límite superior y finalmente se utiliza el color rojo para aquellos píxeles en donde uno o ambos índices son menores que el límite inferior.

# Capítulo 5

## Resultados

En este capítulo se muestran los resultados de cada una de las etapas de las metodologías (identificación de las etapas fenológicas de la caña de azúcar e identificación de anomalías), además se presenta el apartado del caso de estudio para evaluar las metodologías.

### 5.1. Caso de estudio

En este trabajo se utilizaron dos zonas de estudio diferentes: (1) el área de referencia ( $Z_1$ ) ( $19^\circ 40'01.4''$  N,  $103^\circ 24'42.6''$  W) y; (2) el área de prueba ( $Z_2$ ) ( $19^\circ 40'09.2''$  N,  $103^\circ 24'50.7''$  W).

El área de referencia ( $Z_1$ ) se encuentra sembrada con la variedad de caña de azúcar AT MEX 96-40, cuenta con un sistema de riego por goteo y los datos registrados corresponden al ciclo de corte resoca 1. El área de prueba  $Z_1$  fue en donde se efectuó el proceso de identificar las etapas fenológicas en la caña de azúcar, la tabla 5.1 muestra la distribución de las etapas fenológicas en la área de prueba, además esta zona se utilizó en el proceso de identificar anomalías como zona de referencia. Hemos considerado  $Z_1$  un área de referencia porque el productor de caña de azúcar ha tenido un control detallado de su cultivo, produciendo en la temporada 2019-2020 153.755 toneladas por hectárea, que fue superior al rendimiento del campo nacional en la temporada 2018-2019 (CONADESUCA, 2019).

TABLA 5.1: Distribución de las etapas fenológicas del cultivo en la zona de estudio

Etapa fenológica	Periodo
Brotación y Amacollamiento	Febrero 2019
	Marzo 2019
	Abril 2019
	Mayo 2019
Rápido Crecimiento	Junio 2019
	Julio 2019
	Agosto 2019
	Septiembre 2019
Maduración	Octubre 2019
	Noviembre 2019
	Diciembre 2019
	Enero 2020
	Febrero 2020

A partir de esto, definimos un proceso en el que un cultivo se considera sano o no. Se ha considerado una zona diferente de seguimiento ( $Z_2$ ) para realizar las pruebas, esta área esta también esta ubicada en Zapotiltic, Jalisco, México. Además, contar con el mismo sistema de riego por goteo, variedad de caña AT MEX-96-40, y el mismo número de cortes (resoca 1). La superficie total cultivada en  $Z_1$  es de 2,9 ha. y en  $Z_2$  es 2,7 ha.

## 5.2. Descarga de imágenes

Debido a que la descarga de imágenes requiere conocer el área sobre la cual se descargaran las imágenes es necesario identificar dentro de la malla de barrido de sentinel-2 el cuadrante en donde se encuentra la zona de interés, esto se puede efectuar por medio de una plataforma en linea en donde se puede elegir el satélite del cual se quiere conocer la malla de barrido, la dirección web para ingresar a la plataforma es la siguiente: <https://search.remotepixel.ca>. En la Figura 5.1 se ejemplifica la interfaz de la plataforma. Adicionalmente sera necesario proporcionar nuestro usuario y contraseña. Para registrarse sera necesario ingresar al siguiente link: <https://scihub.copernicus.eu/dhus//home>

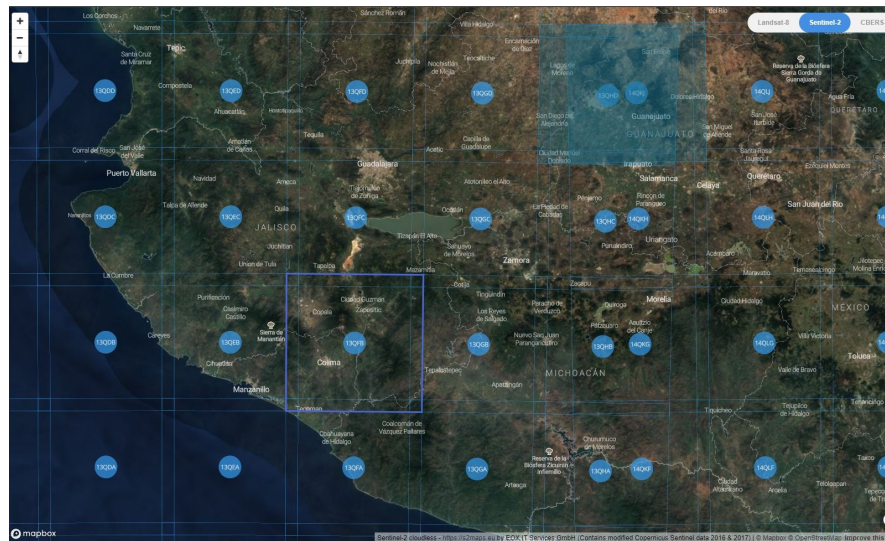


FIGURA 5.1: Identificación de tiles

Una vez identificado el tile sobre el cual se encuentra la zona de interés y registrar un usuario, basta con ejecutar el script y proporcionar los parámetros; usuario, contraseña, tile o tiles a descargar así como la cantidad de nubosidad máxima con la cual se quieren descargar las imágenes. A continuación el script de descarga de las imágenes hará la conexión a los servidores de la ESA por medio de la biblioteca `sentinelsat` y comenzará la descarga de las imágenes. El proceso puede ser tardado ya que la cantidad de imágenes a almacenar generalmente será alta además de que el peso de cada imagen suele ser mayor a 1GB. Una vez descargadas las imágenes, estas estarán en archivos comprimidos, cada archivo comprimido contiene las imágenes (bandas) además de su metadatos.

### 5.3. Pre-Procesamiento

A continuación se presentan los resultados de la fase de pre-procesamiento.

#### 5.3.1. Corrección atmosférica

Una vez que se cuenta con todas las imágenes, se requiere que todos los archivos se descompriman, una vez realizado este proceso, será necesario abrir una ventana



de consola de comandos y ubicarse en la carpeta del proyecto y mediante el comando `for /D %s in (S2*) do L2A_Process.bat %s resolution = 10` se efectuara el proceso de corrección atmosférica mediante la herramienta `sen2cor`, la Figura 5.2 ejemplifica este proceso. Una vez terminado este proceso las imágenes estarán listas para ser procesadas en las siguientes fases.

```

Command Prompt - L2A_Process.bat S2B_MSIL1C_20210427T171849_N0300_R012_T13QFB_20210427T205937.SAFE --resolutions=10
C:\Users\Humberto\Desktop\Sen2Cor-02.08.00-win64>L2A_Process.bat S2B_MSIL1C_20210427T171849_N0300_R012_T13QFB_20210427T205937.SAFE --resolution=10
Sentinel-2 Level 2A Processor (Sen2Cor). Version: 2.8.0, created: 2019.02.20, supporting Level-1C product version 14.2 - 14.5 started ...
Product version: 14.5
Operation mode: TOOLBOX
Processing baseline: 99.99
Progress[%]: 0.00 : Generating datastrip metadata
1C datastrip found, L2A datastrip successfully generated
Selected resolution: 10 m
Progress[%]: 0.08 : PID-13668, L2A_Processfile: 20 m resolution must be processed first, elapsed time[s]: 1.021, total: 0:00:04.224000
Progress[%]: 0.11 : PID-13668, L2A_Processfile: processing with resolution 20 m, elapsed time[s]: 0.453, total: 0:00:04.677000
Progress[%]: 0.11 : PID-13668, L2A_Processfile: start of pre processing, elapsed time[s]: 0.004, total: 0:00:04.681000
Progress[%]: 0.12 : PID-13668, L2A_Tables: start import, elapsed time[s]: 0.147, total: 0:00:04.828000
Progress[%]: 0.17 : PID-13668, L2A_Tables: band B01 imported, elapsed time[s]: 0.679, total: 0:00:05.507000
Progress[%]: 0.44 : PID-13668, L2A_Tables: band B02 imported, elapsed time[s]: 3.608, total: 0:00:09.107000
Progress[%]: 0.81 : PID-13668, L2A_Tables: band B03 imported, elapsed time[s]: 4.231, total: 0:00:14.038000
Progress[%]: 1.46 : PID-13668, L2A_Tables: band B04 imported, elapsed time[s]: 8.735, total: 0:00:22.773000
Progress[%]: 1.61 : PID-13668, L2A_Tables: band B05 imported, elapsed time[s]: 2.030, total: 0:00:24.803000
Progress[%]: 2.89 : PID-13668, L2A_Tables: band B06 imported, elapsed time[s]: 6.348, total: 0:00:31.143000
Progress[%]: 2.21 : PID-13668, L2A_Tables: band B07 imported, elapsed time[s]: 1.597, total: 0:00:32.740000
Progress[%]: 2.48 : PID-13668, L2A_Tables: band B08 imported, elapsed time[s]: 3.692, total: 0:00:36.432000
Progress[%]: 2.57 : PID-13668, L2A_Tables: band B09 imported, elapsed time[s]: 1.136, total: 0:00:37.568000
Progress[%]: 2.78 : PID-13668, L2A_Tables: band B10 imported, elapsed time[s]: 2.799, total: 0:00:40.367000
Progress[%]: 3.01 : PID-13668, L2A_Tables: band B11 imported, elapsed time[s]: 3.149, total: 0:00:42.516000
Progress[%]: 3.11 : PID-13668, L2A_Tables: band B12 imported, elapsed time[s]: 1.352, total: 0:00:44.069000
Progress[%]: 3.11 : PID-13668, L2A_Processfile: start of Scene Classification, elapsed time[s]: 0.010, total: 0:00:44.079000
Progress[%]: 3.12 : PID-13668, L2A_Tables: band B02 must be resampled, elapsed time[s]: 0.123, total: 0:00:45.001000
Progress[%]: 3.63 : PID-13668, Pre process , elapsed time[s]: 0.836, total: 0:00:51.037000
Progress[%]: 3.69 : PID-13668, L2A_Tables: band B01 must be resampled, elapsed time[s]: 0.840, total: 0:00:52.677000
Progress[%]: 4.25 : PID-13668, L2A_Tables: band B03 must be resampled, elapsed time[s]: 7.400, total: 0:01:00.077000
Progress[%]: 4.42 : PID-13668, L2A_Tables: band B04 must be resampled, elapsed time[s]: 2.263, total: 0:01:02.340000
Progress[%]: 4.63 : PID-13668, L2A_Tables: band B09 must be resampled, elapsed time[s]: 2.914, total: 0:01:05.254000

```

FIGURA 5.2: Proceso de corrección atmosférica de `sen2cor`

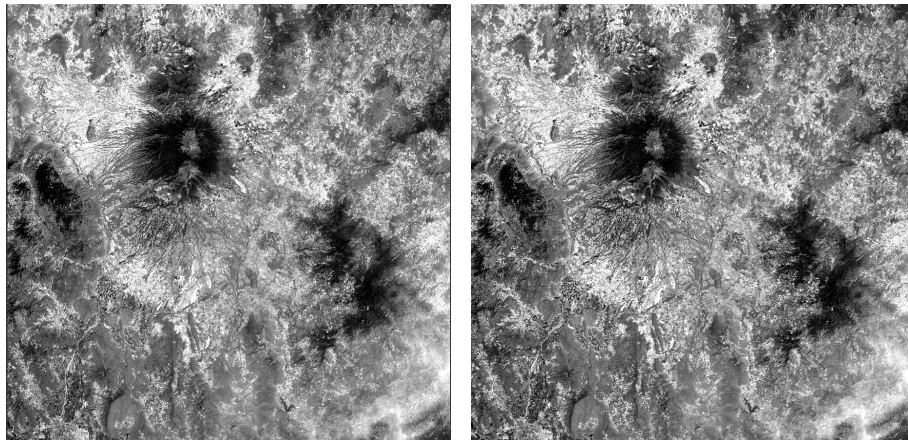


FIGURA 5.3: Corrección atmosférica con `sen2cor`

### 5.3.2. Remuestreo, Recorte y Calculo de índices

Estos procesos se realizaron mediante el software SNAP utilizando su archivo (compilado) de instrucciones escritas en java que pueden ser llamadas desde Python o la consola de comandos del sistema.

SNAP las permite ejecutar la mayor parte de sus funciones a través de un procesamiento por lotes a por medio de la creación de grafos, estos grafos nos permiten

enlazar funciones incluidas en SNAP las cuales se guardarán en un archivo xml y después podrán ejecutarse sin la necesidad de inicializar la interfaz gráfica de SNAP.

Para la creación del grafo, dentro del software SNAP hay una herramienta llamada Graph Builder, esta ofrece una pequeña interfaz donde se puede construir la estructura para el pre-procesamiento, como puede observarse en la Figura 5.4.

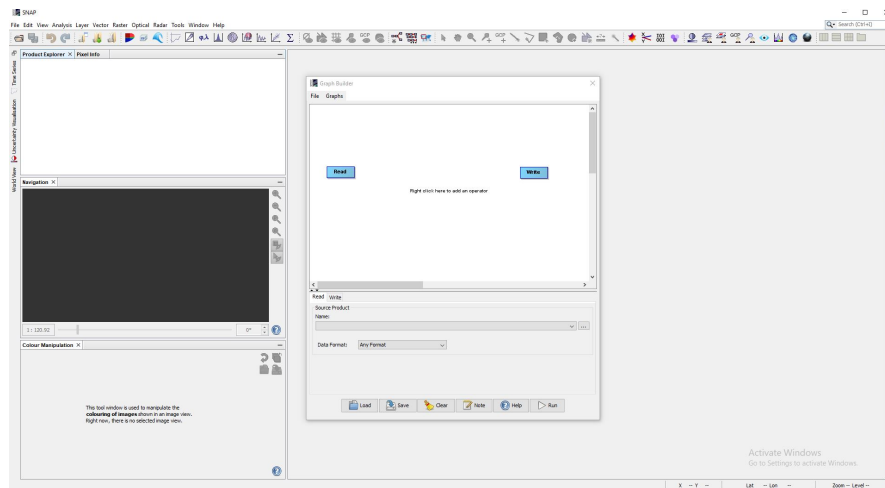


FIGURA 5.4: Construcción del grafo

Dentro de esta interfaz se eligen los procesos que se requieren realizar, no es necesario indicar archivos de entrada o salida ya que son campos que el script en Python modificara.

Una vez terminado el grafo se guarda y se obtiene un archivo XML que tendrá una estructura general como se muestra en la Figura 5.6.

El grafo de la Figura 5.5 es el utilizado por la fase de pre-procesamiento y calculo de índices.

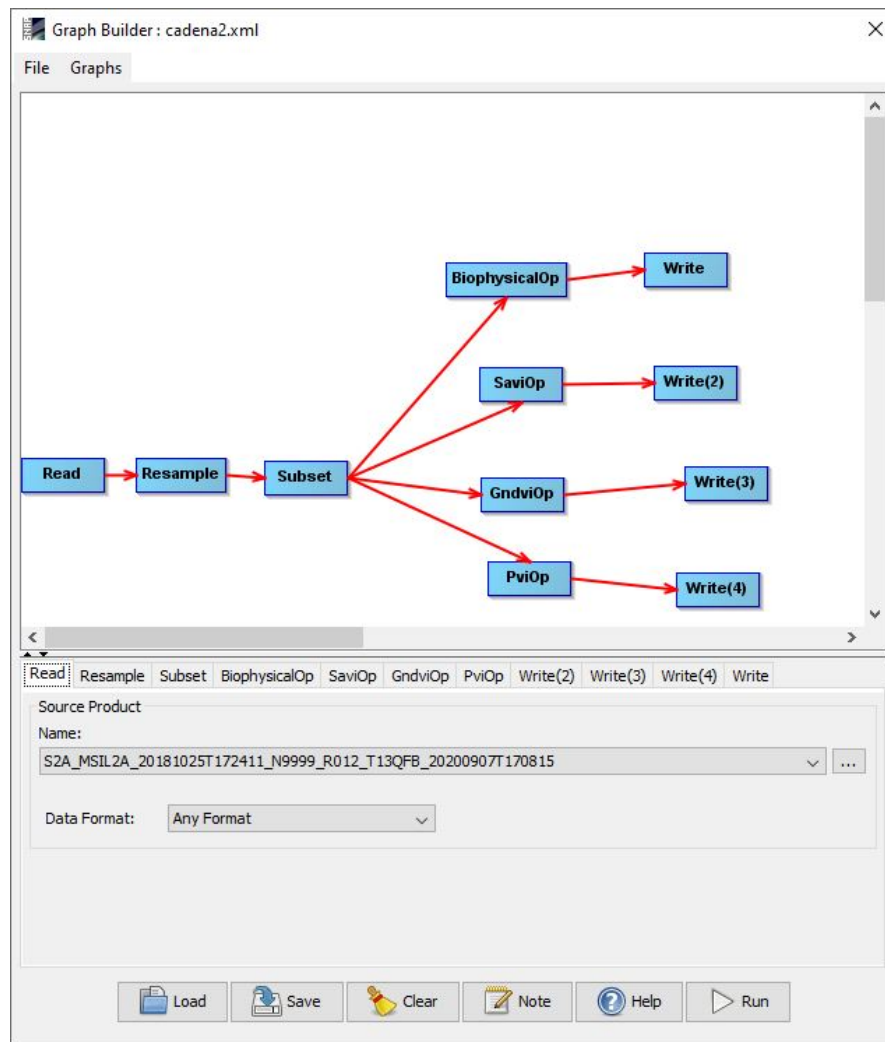


FIGURA 5.5: Grafo utilizado en la fase de pre-procesamiento y calculo de índices

```

<graph id="Graph">
  <version>1.0</version>
  <node id="Read">
    <\node>
  <node id="Resample">
    <operator>Resample</operator>
  <node id="Subset">
    <operator>Subset</operator>
  <node id="ConvertDataType">
    <operator>DataType</operator>
  <node id="Write">
    <operator>Write</operator>
  </node>
</graph>

```

FIGURA 5.6: Estructura XML del grafo

### 5.3.3. Extracción de la zona de interés

Posteriormente a la adquisición de los índices, se obtienen los valores del área de interés. Para ello, se utiliza un polígono generado en la plataforma en línea GeoJson <sup>1</sup>. Esta plataforma nos provee de una interfaz gráfica, como se puede observar en la Figura 5.7, en esta podremos dibujar una polilínea sobre el área de interés y generar un polígono. Una vez creado el polígono basta con seleccionar la opción de guardarlo como shapefile, este archivo contendrá las coordenadas sobre la cual se encuentra el área de interés y mediante Python se realizara la extracción de los píxeles para su posterior análisis, la Figura 5.8 ejemplifica el resultado de esta fase.

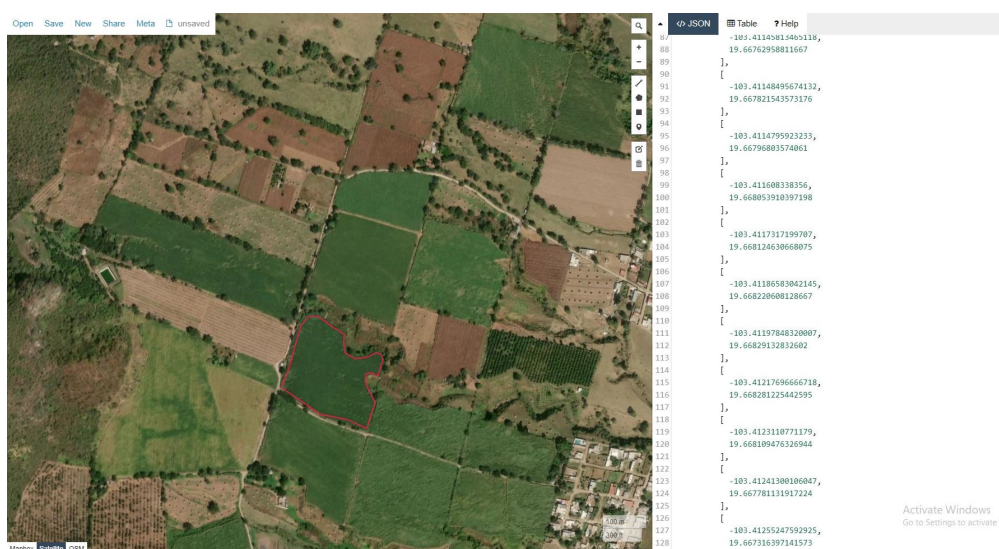


FIGURA 5.7: Interfaz gráfica de GeoJSON

<sup>1</sup><https://geojson.io/map=2/20.0/0.0>.

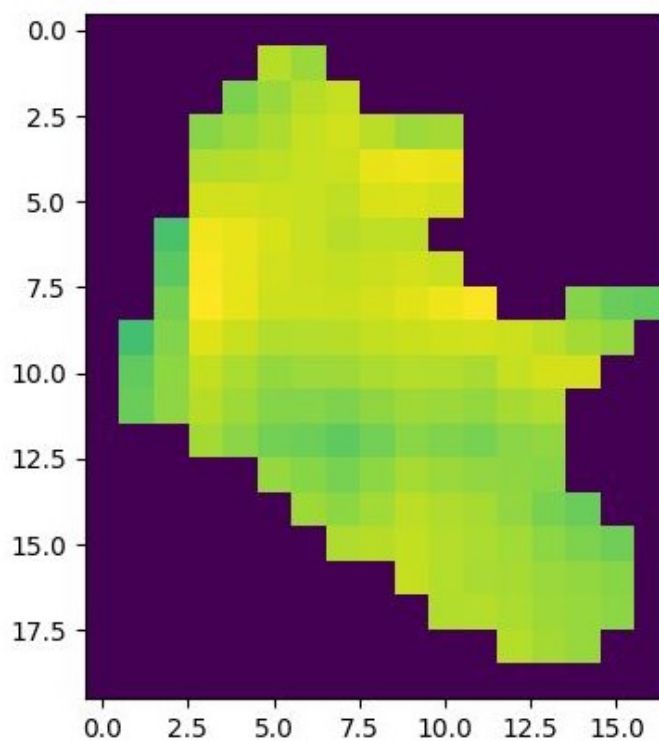


FIGURA 5.8: Extracción de los píxeles de la zona de interés

## 5.4. Cálculo de series

Una vez se obtiene la zona de interés, en base de los valores de la media, máximo y mínimo de cada índice considerado NDVI, GNDVI, SAVI, PVI y LAI, se construyen cada una de las series. En la Figura 5.9 se muestran las series temporales para el valor de la media de cada índice de vegetación, las líneas verticales punteadas muestran el rango de cada una de las etapas fenológicas, de izquierda a derecha, en primer lugar se encuentra la etapa de brote o amacollamiento, seguida por la etapa de rápido crecimiento y finalmente la etapa de maduración.

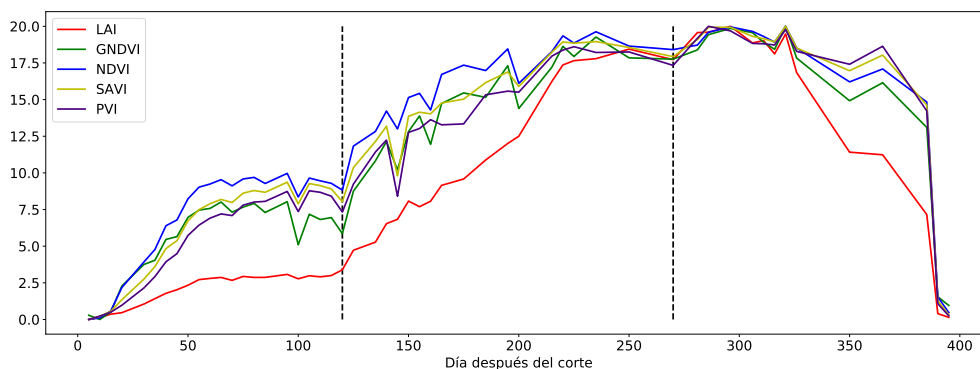


FIGURA 5.9: Series temporales de los índices de vegetación NDVI, GNDVI, SAVI, PVI y LAI en el área  $Z_1$ .

Las series se utilizaron como características para identificar la etapa fenológica del cultivo de caña de azúcar, en la Figura 5.10 se muestra un ejemplo de comparación entre el mínimo en LAI y el máximo en PVI.

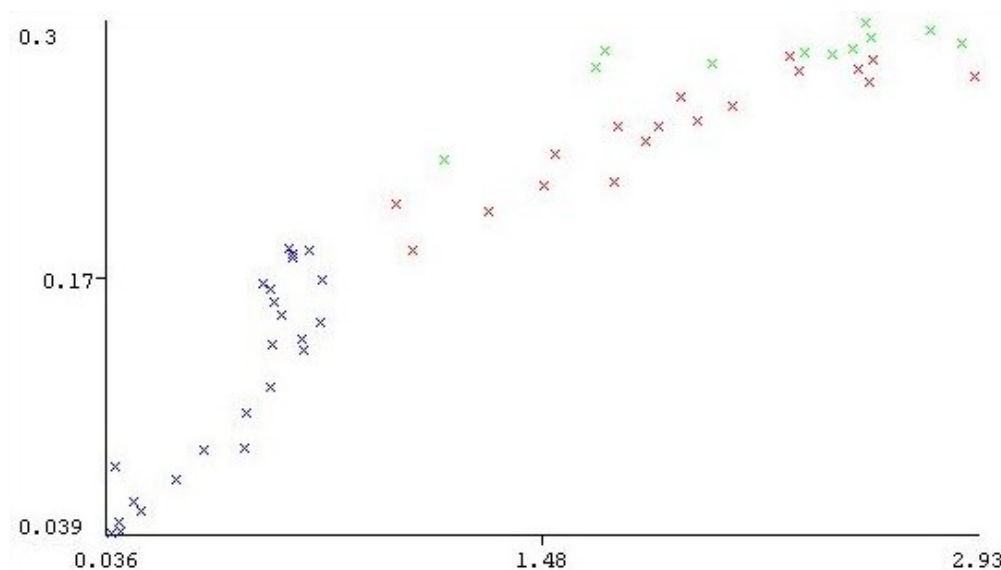


FIGURA 5.10: Ejemplo de comparación de dos características, MinLAI y MaxPVI

## 5.5. Identificación de las etapas fenológicas

La Tabla 5.2, muestra los resultados obtenidos por los clasificadores considerados (KNN, RF, SVM y NB) considerando las 15 series (características) compuestas



por los valores máximo, medio y mínimo calculados para cada uno de los 5 índices analizados en el presente trabajo.

TABLA 5.2: Resultados de la clasificación de las 15 características de las series temporales.

Métrica	KNN-5	RF	SVM	NB
Accuracy	84.906	92.453	79.245	81.132
Precision	0.845	0.926	0.798	0.834
Recall	0.849	0.925	0.792	0.811
F-Measure	0.843	0.923	0.787	0.808

Como puede observarse en la Tabla 5.2, el clasificador RandomForest (RF) fue el que proporciono mejores resultados en la clasificación, con un F-Measure de 0.923%. Pese a que el resultado es bueno, algunas de las características seleccionadas pueden estar proporcionando poca información e información redundante respecto a la etapa fenológica del cultivo, por lo que se procede a realizar una selección de las mejores características.

### 5.5.1. Seleccionado las mejores características

El procedimiento de seleccionar características se realizó mediante el uso de el método *CfsSubsetEval* el cual se incluye en el software Weka (Kotthoff, Thornton, y Hutter, 2017). Este método mide la importancia de las características en base de la capacidad predictiva de estas y su grado de redundancia.

*CfsSubsetEval* selecciona un subconjunto el cual tenga menos intercorrelación pero que a su vez este altamente correlacionado con la clase objetivo (Kotthoff, Thornton, y Hutter, 2017). Siguiendo estos criterios, los resultados mostraron que las mejores características son: LAIMin, GNDVIMean, LAIMin, SAVIMax, PVI-Mean, PVIMAx, LAIMax.

### 5.5.2. Evaluación con las mejores características

En la Tabla 5.3 se muestran los resultados por cada clasificador utilizando las seis mejores características obtenidas en el proceso anterior.

TABLA 5.3: Resultados de la clasificación considerando las mejores características

<b>Métrica</b>	<b>KNN-5</b>	<b>RF</b>	<b>SVM</b>	<b>NB</b>
Accuracy	84.906	92.453	83.019	83.019
Precision	0.845	0.926	0.826	0.847
Recall	0.849	0.925	0.830	0.830
F-Measure	0.843	0.923	0.820	0.828

Como puede observarse en la Tabla 5.3, el efecto de eliminar información redundante y poco descriptiva respecto a la etapa fenológica del cultivo es positiva, ya que solo considerando 6 de las 15 características, se mantuvo el rendimiento de KNN-5 y de RF. Adicionalmente, los clasificadores NB y SVM produjeron mejores resultados con la selección de éstas características.

## 5.6. Identificación de anomalías

En esta sección se muestran los resultados obtenidos al aplicar el proceso de detección de anomalías utilizando el área de prueba ( $Z_2$ ). Como primer paso se calcularon los índices PVI y LAI en la zona de referencia ( $Z_1$ ), el valor de estos índices en cada imagen capturada definirá los valores esperados de la zona de prueba ( $Z_2$ ). Una vez calculado los valores se construye una serie que define un límite inferior y un superior en ambos índices como se puede ver en las Figuras 5.11 y 5.12.



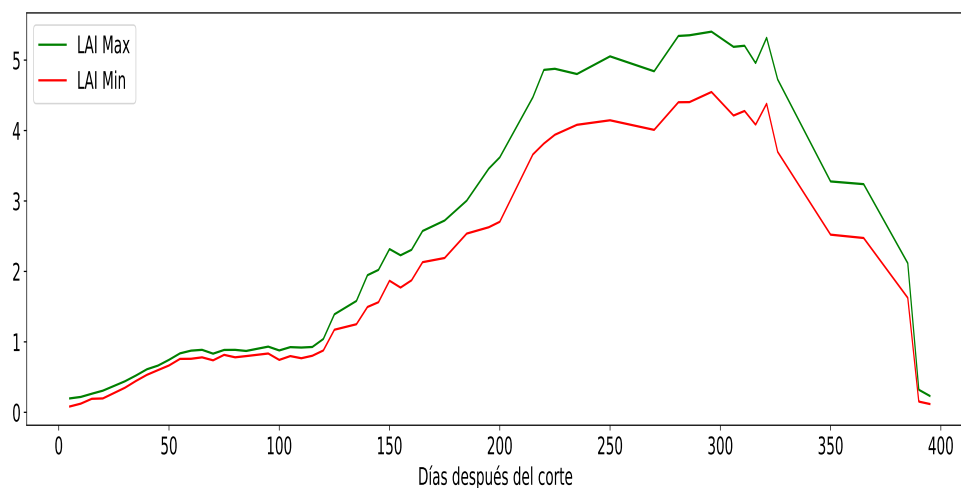


FIGURA 5.11: Rango para el índice LAI.

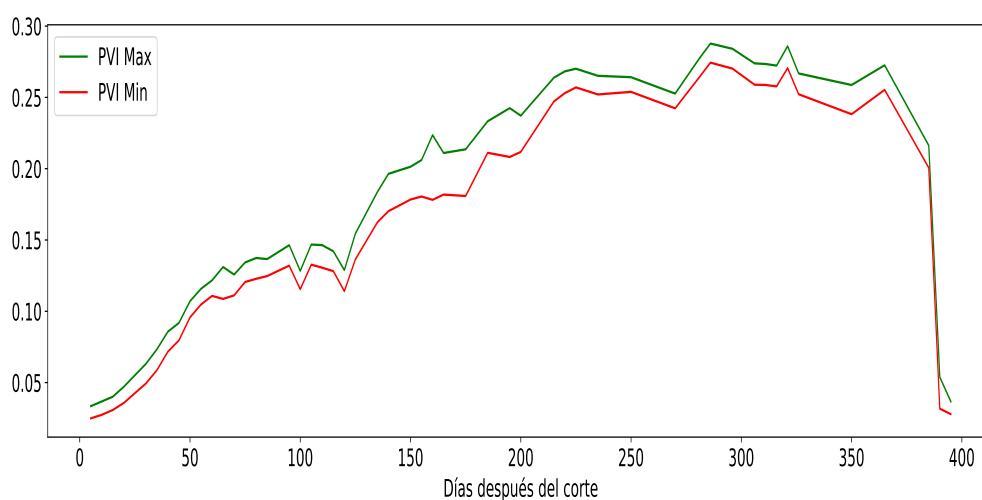


FIGURA 5.12: Rango para el índice PVI

Dado que no se cuenta con datos de cada día del ciclo del cultivo en la zona ( $Z_1$ ), es necesario utilizar un método de regresión que permita interpolar para aquellos días en los que no se cuenta con imágenes, dicho proceso se realizó con el método *cubic smoothing spline*.

### 5.6.1. Regresión

Las Figuras 5.13 y 5.14 muestran el umbral obtenido a partir de  $Z_1$  que define los valores esperados para cada índice considerados saludables en todos los días del ciclo de vida de la caña de azúcar utilizando el método de regresión *cubic smoothing spline*.

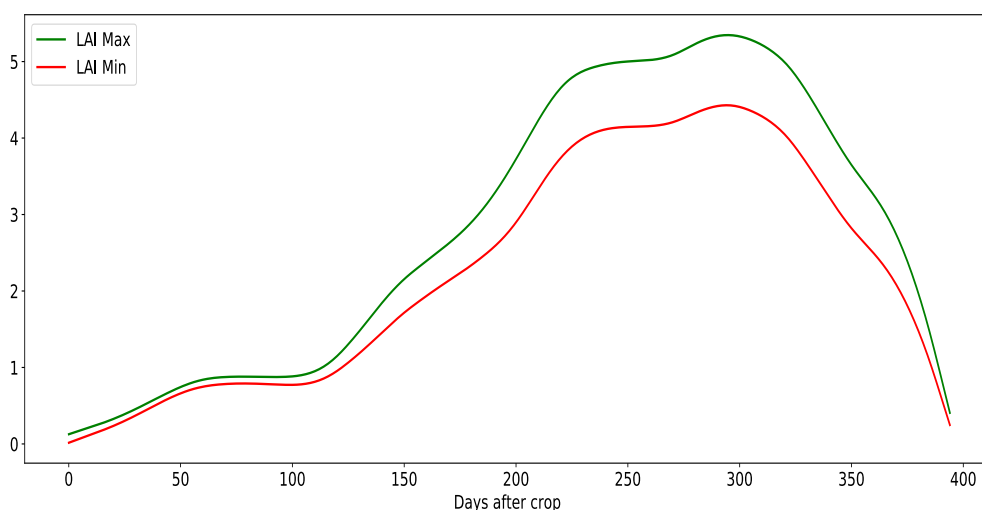


FIGURA 5.13: Perfil de referencia para el índice LAI.

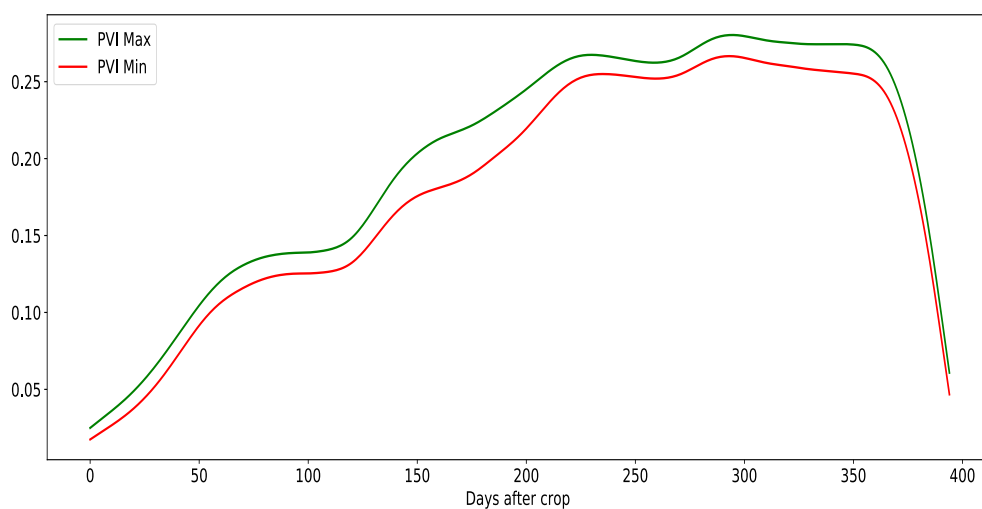


FIGURA 5.14: Perfil de referencia para el índice PVI.

Para identificar zonas anómalas en  $Z_2$ , seleccionamos seis días para las pruebas; estos son 0, 60, 120, 230, 276 y 390 que corresponden al 2 de febrero de 2019, 3 de abril de 2019, 2 de junio de 2019, 20 de septiembre de 2019, 4 de noviembre de 2019 y 20 de febrero de 2020, del período de cultivo, respectivamente.

La Figura 5.15 muestra el mapa de colores para monitorear las seis imágenes de prueba. El color rojo se utiliza cuando uno o los dos índices están por debajo del valor límite del umbral, el color verde cuando están dentro del umbral y el color azul cuando uno o los dos índices están por encima del umbral. Podemos observar en la Figura 5.15 (a) que la mayor parte del cultivo está dentro del rango o por encima del umbral; por lo que se considera que es un cultivo sano en esta fecha. La Figura 5.15 (b) muestra que hay píxeles por encima del umbral y dentro del umbral. Además, podemos ver que hay píxeles rojos en los bordes del área, lo que indica que uno o ambos índices están por debajo del umbral. En este punto, aun podemos considerarlo un cultivo sano.

El monitoreo de la Figura 5.15 (c) muestra que la cosecha de caña de azúcar es saludable ya que el LAI y PVI están por encima del umbral. Además, podemos observar que  $Z_2$  alcanza el crecimiento máximo más rápido que  $Z_1$ . La Figura 5.15 (d) muestra un cultivo poco saludable ya que tiene muchos píxeles en rojo.

La mayoría de los bordes en noviembre (Figura 5.15 (e)) están en rojo, lo que indica alguna deficiencia en el cultivo. Los otros píxeles están dentro del rango. También podemos observar que de septiembre a noviembre, la cosecha se ha recuperado, ya que, en este último mes, la zona central se encuentra dentro del umbral.

Finalmente, en febrero de 2020, se muestra el cultivo de la zona  $Z_2$  previo a la cosecha. Podemos ver que uno o ambos índices están por debajo del umbral, por lo que definimos este cultivo como insalubre. En 5.15 d, 5.15 e, 5.15 f, es donde podemos ver más deficiencias en el cultivo.

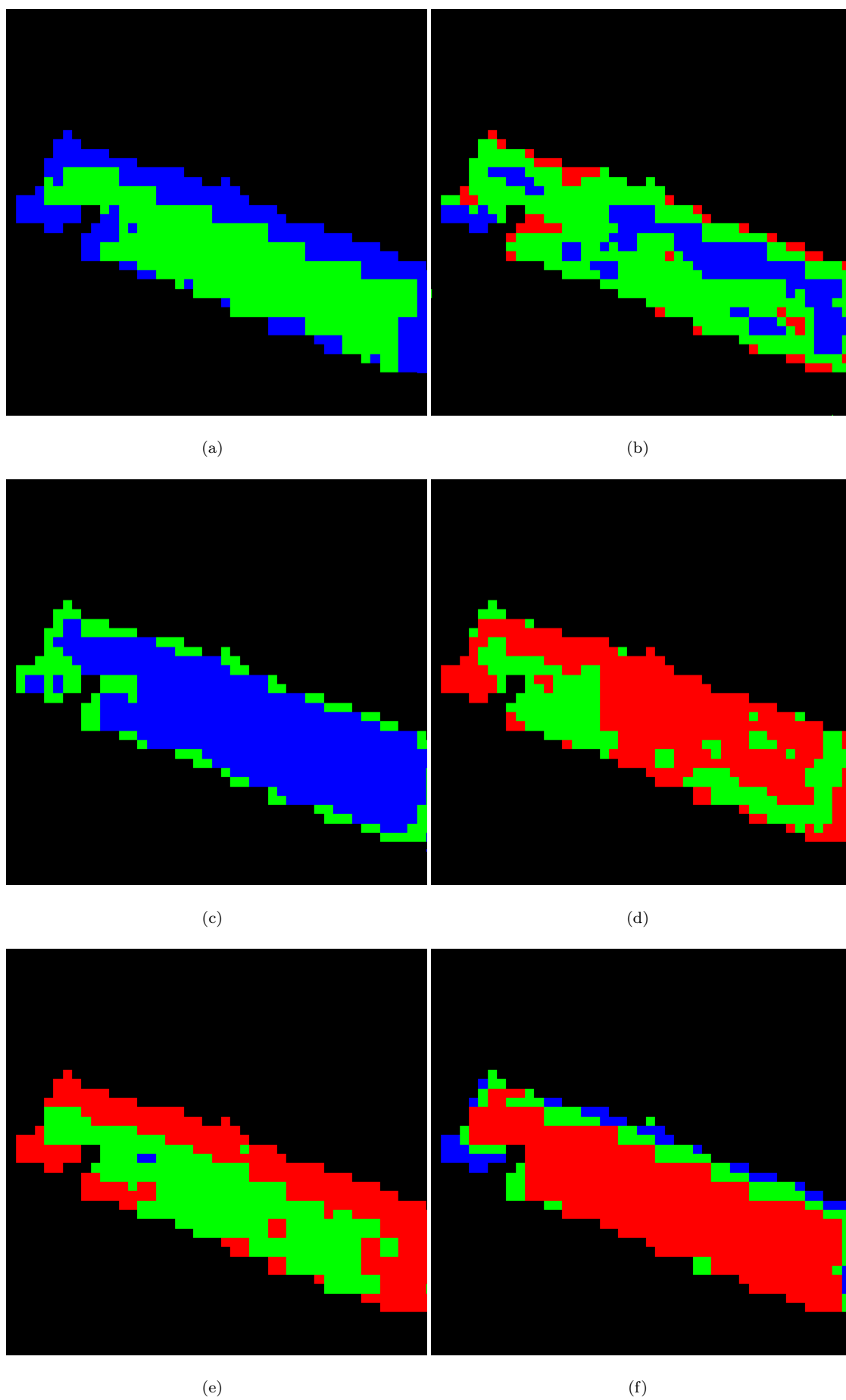


FIGURA 5.15: Detección de anomalías; a) 02/02/2019, b) 04/03/2019, c) 06/02/2019, d) 09/20/2019, e) 11/04/2019, f) 02/20/2020

# Capítulo 6

## Conclusiones y trabajos futuros

A continuación se muestran las conclusiones obtenidas en el desarrollo y resultados del proyecto, así como los trabajos a desarrollar en un futuro.

### 6.1. Conclusiones

En éste trabajo de investigación, se presentó un método para identificar las etapas fenológicas de cultivos de caña de azúcar usando datos del sensor *MultiSpectral Instrument* a bordo del satélite Sentinel-2. Para desarrollar el método propuesto, se evaluaron los clasificadores: *Random Forest*, *Support Vector Machine*, *k-Nearest Neighbors* y *Naïve Bayes*, utilizando como características 15 series temporales de cinco índices de vegetación. Adicionalmente se realizó un proceso para seleccionar las mejores características, basado en su relevancia descriptiva y su grado de redundancia, los resultados para este proceso indicaron que es posible lograr mantener el rendimiento de los clasificadores con solo 6 de las 15 características, lo cual influyó positivamente en la cantidad de datos a procesar. Los resultados mostraron que el método de clasificación *Random Forest* logrando un  $F - measure = 0.923$ , es el que mejor se adapta al proceso de identificar las etapas de crecimiento de cultivos de caña de azúcar. La caracterización del estado fenológico es una fase clave para realizar el monitoreo de cultivos, pues permite identificar presencia de plagas, deficiencias de nutrientes a nivel de suelo y cultivo, esto debido a los cambios en la respuesta espectral que se realzan con los índices de vegetación.

Adicionalmente se desarrollo un método para identificar zonas anómalas en cultivos de caña usando imágenes obtenidas del sensor Sentinel-2 MSI. Definimos un perfil de salud de un cultivo que sabíamos previamente que tenía buenas prácticas de campo. Probamos este método en un área diferente para verificar la eficiencia. Indicamos las zonas en las que el cultivo se encuentra por debajo del umbral definido como anómalo mediante un mapa de colores. Las pruebas mostraron que desde el día 0 al 60, la caña de azúcar en el área de prueba estuvo mayormente dentro del rango con el perfil de salud, y en el día 120, alcanza el crecimiento máximo más rápido que el perfil de salud; a partir de este día el área de prueba comenzó a mostrar algunas deficiencias. Por tanto, concluimos que es posible determinar las áreas con alguna deficiencia en el cultivo de caña de azúcar a través del método propuesto.

Como trabajos futuros se implementará la reunión de desarrollos tecnológicos previos como la adecuación de los requerimientos básicos para caña de azúcar, de la aplicación PulAm (Perez-Mena et al., 2018) y la cadena de procesado de imágenes Sentinel-2 (Rodriguez-Ramirez et al., 2018). Este desarrollo permitirá crear mapas de salud vegetal para caña de azúcar. Además de contar con la metodología para escalar el análisis de otros cultivos en el estado de Jalisco.

# Bibliografía

- Abdel-Rahman, E. M., y 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(13):3753–3767.
- Aguilar Rivera, N. 2015. Ficha técnica del cultivo de la caña de azúcar (*saccharum officinarum* l.). Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación.
- Atkinson, P. M.; Jeganathan, C.; Dash, J.; y Atzberger, C. 2012. Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *Remote sensing of environment* 123:400–417.
- Barbosa, R. M.; Batista, B. L.; Barião, C. V.; Varrique, R. M.; Coelho, V. A.; Campiglia, A. D.; y Barbosa Jr, F. 2015. A simple and practical control of the authenticity of organic sugarcane samples based on the use of machine-learning algorithms and trace elements determination by inductively coupled plasma mass spectrometry. *Food chemistry* 184:154–159.
- Cárdenas, D. A. G.; Valencia, J. A. R.; Velásquez, D. F. A.; y Gonzalez, J. R. P. 2018. Dynamics of the indices ndvi and gndvi in a rice growing in its reproduction phase from multi-spectral aerial images taken by drones. In *International Conference of ICT for Adapting Agriculture to Climate Change*, 106–119. Springer.
- CESAR. 2018. El espectro electromagnético. <http://cesar.esa.int>.
- CONADESUCA. 2007. Listado de municipios asociados a la producción de caña de azúcar en México.
- CONADESUCA. 2015. Ficha técnica del cultivo de la caña de azúcar (*saccharum officinarum* l.).

- CONADESUCA. 2017. Technical report, CONADESUCA.
- CONADESUCA. 2019. 6to. informe estadístico del sector agroindustrial de la caña de azúcar en México, zafas 2009-2010, 2018-2019. *CONADESUCA* 0:1–125.
- De Boor, C.; De Boor, C.; Mathématicien, E.-U.; De Boor, C.; y De Boor, C. 1978. *A practical guide to splines*, volume 27. Springer-Verlag New York.
- Delegido, J.; Pezzola, A.; Casella, A.; Winschel, C.; Urrego, E.; Jimenez, J.; Soria, G.; Sobrino, J.; y Moreno, J. 2018. Estimación del grado de severidad de incendios en el sur de la provincia de Buenos Aires, Argentina, usando Sentinel-2 y su comparación con Landsat-8. *Revista de Teledetección* (51):47–60.
- Döker, S., y Boşgelmez, İ. İ. 2015. Rapid extraction and reverse phase-liquid chromatographic separation of mercury (ii) and methylmercury in fish samples with inductively coupled plasma mass spectrometric detection applying oxygen addition into plasma. *Food chemistry* 184:147–153.
- Endres, L.; Moura dos Santos, C.; Verissimo de Souza, G.; Menossi, M.; Marcelino dos Santos, J. C.; et al. 2018. Morphological changes recorded in different phenophases of sugarcane plants subjected to water stress in tropical field conditions. *Australian Journal of Crop Science* 12(7):1041.
- ESA. 2018. Snap. <https://step.esa.int/main/toolboxes/snap/>.
- Gerritsen, P. R.; Bernal, V. S.; y Rodríguez, G. T. 2019. “la cara oculta de la seguridad alimentaria” estudio de caso del sur del estado de Jalisco, occidente de México/“the hidden side of food security case study from the south of the state of Jalisco, west of Mexico. *Brazilian Journal of Development* 5(7):7569–7585.
- Gitelson, A. A.; Kaufman, Y. J.; Merzlyak, M. N.; et al. 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote sensing of Environment* 58(3):289–298.
- Gómez-Giráldez, P. J.; Pérez-Palazón, M. J.; Polo, M. J.; y González-Dugo, M. P. 2020. Monitoring grass phenology and hydrological dynamics of an oak–grass savanna ecosystem using Sentinel-2 and terrestrial photography. *Remote Sensing* 12(4):600.
- Guedes, W. N., y Pereira, F. M. V. 2019. Raw sugarcane classification in the presence of small solid impurity amounts using a simple and effective digital imaging system. *Computers and Electronics in Agriculture* 156:307–311.



- Hackeling, G. 2017. *Mastering Machine Learning with scikit-learn*. Packt Publishing Ltd.
- Hernández, J., y Montaner, D. 2009. Patrones de respuesta espectral. *Tecnologías Geoespaciales: Experiencias Aplicadas al Estudio y Gestión del Territorio*. Carlos Mena (Ed.) ISBN 978–956.
- Hill, M. J. 2013. Vegetation index suites as indicators of vegetation state in grassland and savanna: An analysis with simulated sentinel 2 data for a north american transect. *Remote Sensing of Environment* 137:94–111.
- Huete, A. 1988. A soil-adjusted vegetation index (SAVI). *Remote sensing of environment* 25:295–309.
- James, G.; Witten, D.; Hastie, T.; y Tibshirani, R. 2013. *An introduction to statistical learning*, volume 112. Springer.
- Khorram, S.; Koch, F. H.; van der Wiele, C. F.; y Nelson, S. A. 2012. *Remote sensing*. Springer Science & Business Media.
- Khorram, S.; Van der Wiele, C. F.; Koch, F. H.; Nelson, S. A.; y Potts, M. D. 2016. *Principles of applied remote sensing*. Springer.
- Kollert, A.; Bremer, M.; Löw, M.; y Rutzinger, M. 2021. Exploring the potential of land surface phenology and seasonal cloud free composites of one year of sentinel-2 imagery for tree species mapping in a mountainous region. *International Journal of Applied Earth Observation and Geoinformation* 94:102208.
- Kotthoff, L.; Thornton, C.; y Hutter, F. 2017. User guide for auto-WEKA version 2.6. *Dept. Comput. Sci., Univ. British Columbia, BETA lab, Vancouver, BC, Canada, Tech. Rep 2*.
- Kowalski, K.; Senf, C.; Hostert, P.; y Pflugmacher, D. 2020. Characterizing spring phenology of temperate broadleaf forests using landsat and sentinel-2 time series. *International Journal of Applied Earth Observation and Geoinformation* 92:102172.
- Lange, M.; Dechant, B.; Rebmann, C.; Vohland, M.; Cuntz, M.; y Doktor, D. 2017. Validating modis and sentinel-2 ndvi products at a temperate deciduous forest site using two independent ground-based sensors. *Sensors* 17(8):1855.

- Legazpi, J. d. J. S.; Rangel, B. L. F.; y Núñez, J. V. 2017. Opinión y perspectiva de los productores de caña de azúcar del valle grullo-aatlán, jalisco, en relación a los apoyos de las organizaciones cañeras locales. *Revista Iberoamericana de las Ciencias Biológicas y Agropecuarias: CIBA* 6(11):1.
- López-Vázquez, J.; Valdez-Balero, A.; Silva-Rojas, H.; Flores-Revilla, C.; y Rangel-Ortega, C. 2016. Evaluación a la escaldadura (*xanthomonas albilineans* (ashby) downson) de la hoja de variedades de caña de azúcar (*saccharum* spp.). *Agroproductividad* 9(3).
- Ma, H.; Jing, Y.; Huang, W.; Shi, Y.; Dong, Y.; Zhang, J.; y Liu, L. 2018. Integrating early growth information to monitor winter wheat powdery mildew using multi-temporal Landsat-8 imagery. *Sensors* 18(10):3290.
- McCollam, G. 2018. Correlating nitrogen application rates in sugarcane with low-cost normalized difference vegetation index (NDVI). Technical Report FS14-282, Ellendale Farms LLC.
- Navarro-Cerrillo, R. M. 2013. Evaluación de variables biofísicas mediante teledetección en pinares mediterráneos. *Ambienta* 105:78–88.
- Ordóñez, J. L. 2012. Espectro electromagnético y espectro radioeléctrico. *Manual formativo de ACTA* (62):17–31.
- Perez-Mena, A.; Fernández-Zepeda, J. A.; Rivera-Caicedo, J. P.; y Avila-George, H. 2018. Pulam: An app for monitoring crops. In *International Conference on Software Process Improvement*, 196–205. Springer.
- Perry Jr, C. R., y Lautenschlager, L. F. 1984. Functional equivalence of spectral vegetation indices. *Remote sensing of environment* 14(1-3):169–182.
- Powers, D. M. 2007. Evaluation: From Precision, Recall and F-Factor to ROC, informedness, markedness & correlation. techreport SIE-07-001, Flinders University, Adelaide, Australia.
- Ren, H.; Zhou, G.; y Zhang, F. 2018. Using negative soil adjustment factor in soil-adjusted vegetation index (SAVI) for aboveground living biomass estimation in arid grasslands. *Remote Sensing of Environment* 209:439–445.
- Rish, I., et al. 2001. An empirical study of the naive bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, volume 3, 41–46.

- Rivera Caicedo, J. P. 2014. Optimized and automated estimation of vegetation properties: Opportunities for sentinel-2.
- Rodriguez-Ramirez, R.; Sánchez, M. G.; Rivera-Caicedo, J. P.; Fajardo-Delgado, D.; y Avila-George, H. 2018. Automating an image processing chain of the Sentinel-2 satellite. In *International Conference on Software Process Improvement*, 216–224. Springer.
- Rodriguez Ramirez, R. 2019. Cadena de procesamiento para obtener mapas lai de la caña de azúcar en el estado de jalisco utilizando imágenes sentinel-2.
- Romero, E. R., J. S. P. A. D. M. F. N. J. A. J. d. U. S. D. C. M. J. L. G. P. A. 2012. <http://www.sugarcane crops.com/introduction/>.
- Rouse, J. W. 1974. Monitoring the vernal advancement of retrogradation of natural vegetation, NASA/GSFG, Type III. *Final Report 371*.
- Sanseechan, P.; Saengprachathanarug, K.; Posom, J.; Wongpichet, S.; Chea, C.; y Wongphati, M. 2019. Use of vegetation indices in monitoring sugarcane white leaf disease symptoms in sugarcane field using multispectral uav aerial imagery. In *IOP Conference Series: Earth and Environmental Science*, volume 301, 012025. IOP Publishing.
- Sobrino, J. A. 2001. *Teledetección*. Universitat de València.
- Susantoro, T. M.; Wikantika, K.; Saepuloh, A.; y Harsolumakso, A. H. 2018. Selection of vegetation indices for mapping the sugarcane condition around the oil and gas field of north west java basin, indonesia. In *IOP Conference Series: Earth and Environmental Science*, volume 149, 012001. IOP Publishing.
- Tian, J.; Zhu, X.; Shen, Z.; Wu, J.; Xu, S.; Liang, Z.; y Wang, J. 2020. Investigating the urban-induced microclimate effects on winter wheat spring phenology using sentinel-2 time series. *Agricultural and Forest Meteorology* 294:108153.
- Van Dillewijn, C. 1951. Botánica de la caña de azúcar.
- Wang, J.; Xiao, X.; Bajgain, R.; Starks, P.; Steiner, J.; Doughty, R. B.; y Chang, Q. 2019a. Estimating leaf area index and aboveground biomass of grazing pastures using sentinel-1, sentinel-2 and landsat images. *ISPRS Journal of Photogrammetry and Remote Sensing* 154:189–201.

- 
- Wang, M.; Liu, Z.; Baig, M. H. A.; Wang, Y.; Li, Y.; y Chen, Y. 2019b. Mapping sugarcane in complex landscapes by integrating multi-temporal Sentinel-2 images and machine learning algorithms. *Land Use Policy* 88:104190.
- Watson, C. J.; Restrepo-Coupe, N.; y Huete, A. R. 2019. Multi-scale phenology of temperate grasslands: improving monitoring and management with near-surface phenocams. *Frontiers in Environmental Science* 7:14.
- Weigend, A. S. 2018. *Time series prediction: forecasting the future and understanding the past*. Routledge.
- Yao, J.; Chen, Y.; Zhao, Y.; Mao, W.; Xu, X.; Liu, Y.; y Yang, Q. 2018. Response of vegetation NDVI to climatic extremes in the arid region of Central Asia: a case study in Xinjiang, China. *Theoretical and applied climatology* 131(3-4):1503–1515.

# Apéndice A

## Estancia

### A.1. Estancia virtual

Durante la estancia en la Universidad de Valencia se trabajo en la implementación en Python de diversas técnicas para el análisis series temporales utilizando imágenes satelitales de Sentinel-2.

Los metodos implementados fueron:

- Series de Fourier.
- Whittaker.

El objetivo fue desarrollar un script en python el cual sea capaz de obtener series temporales a partir de mapas de algún índice. La estructura del script se basa en la Figura A.1.

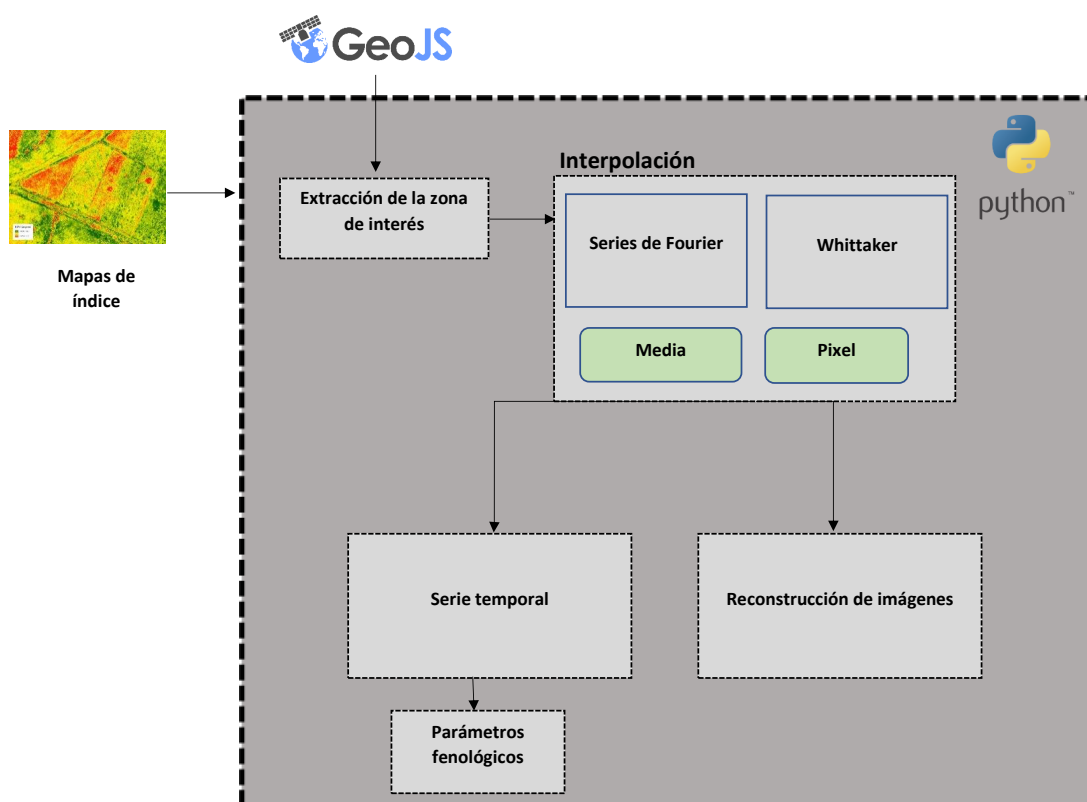


FIGURA A.1: Metodología del script

Cada una de las etapas se describe a continuación:

### A.1.1. Lectura de imágenes

El script parte por leer los mapas del índice a utilizar, las imágenes deben estar en formato *tif* y el nombre de cada imagen debe estar en el siguiente formato, índice\_AñoMesDía.

### A.1.2. Extracción de la zona de interés

Ya leída las imágenes el script utiliza un polígono en formato *shapefile* para extraer los píxeles que solo comprenden la zona de interés. El polígono se puede generar mediante la herramienta en línea *GeoJSON*.

### **A.1.3. Interpolación**

Una vez extraída la zona de interés de las imágenes el script utiliza dos métodos de interpolación, Series de Fourier y Whittaker. El usuario deberá elegir alguno de estos métodos y a continuación se le pedirán algunos datos.

#### **A.1.3.1. Series de Fourier**

Para efectuar la regresión de la serie temporal con este método, el script le pedirá al usuario el grado de la serie de Fourier. Una vez proporcionado el grado, el script presentará 2 opciones para efectuar la regresión, estas serán, media y píxel. La opción de media tomará la intensidad de los píxeles de la zona de interés y por cada imagen calculará la media con la que efectuará el proceso de regresión.

#### **A.1.3.2. Whittaker**

Para efectuar este método de regresión el usuario solamente deberá proporcionar si desea que la serie sea en base a la media o por píxel.

### **A.1.4. Serie temporal**

Una vez efectuada la regresión para alguno de los métodos, el script presentará un gráfico en donde se podrá visualizar la serie temporal con los datos originales y además de la serie temporal generada con el método seleccionado, ver Figura A.2.

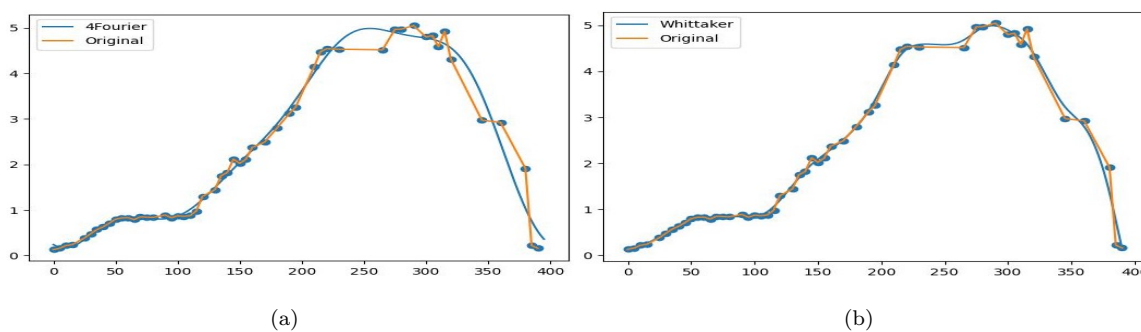


FIGURA A.2: Interpolación: (a) Serie de fourier de grado 4 efectuada sobre la media, (b) Método de Whittaker efectuada sobre la media

### A.1.4.1. Parámetros fenológicos

Adicionalmente el script presenta esta opción la cual permite obtener los parámetros fenológicos de inicio de temporada, pico mas alto y final de la temporada, esto mediante una gráfica como se puede observar en la Figura A.3.

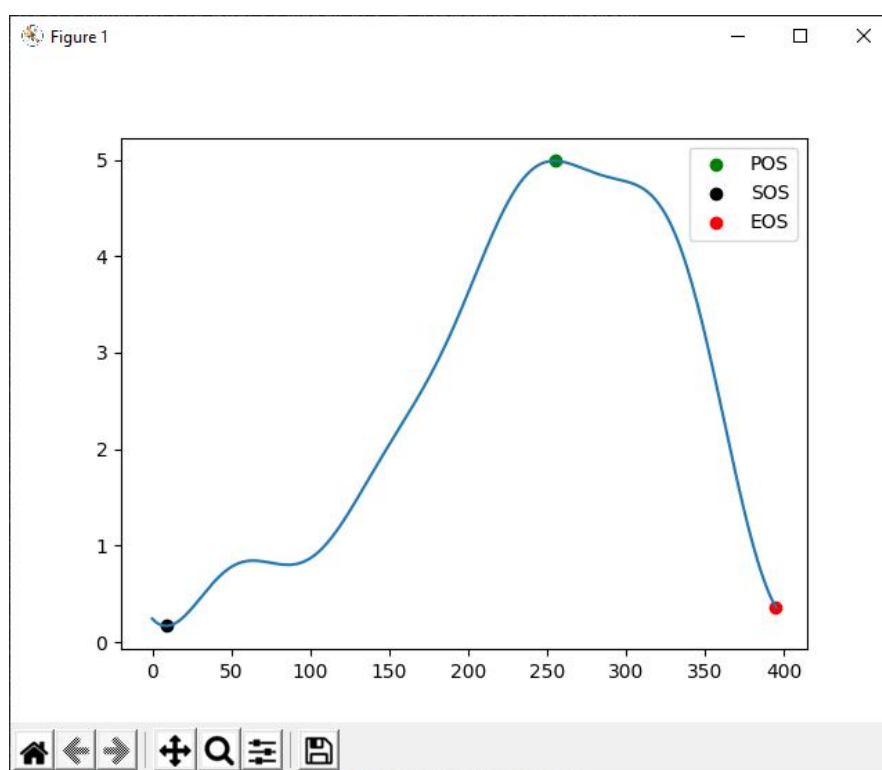
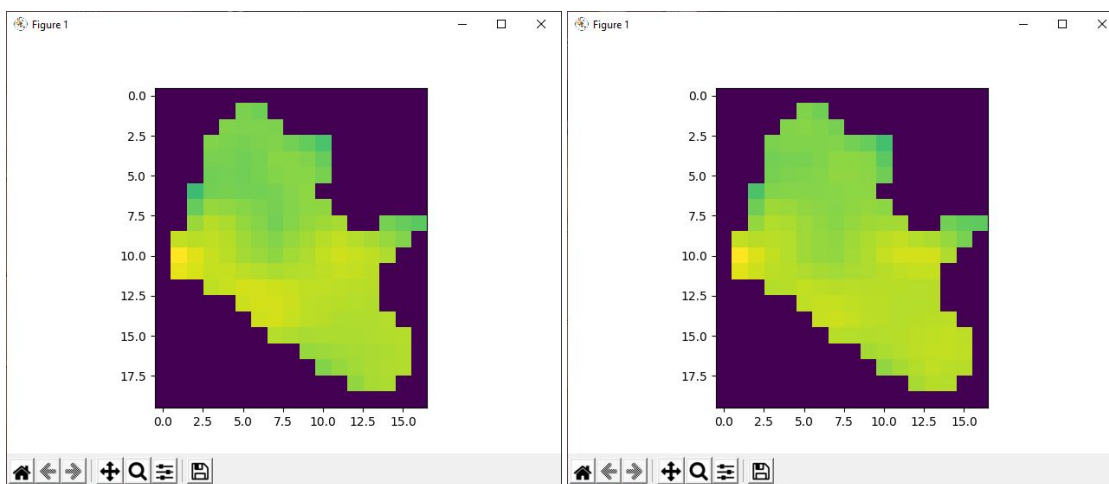


FIGURA A.3: Parámetros fenológicos



### A.1.5. Reconstrucción de imágenes

Otra opción que se da al usuario es la de generar imágenes a partir de alguno de los dos métodos de interpolación (Whittaker y Serie de Fourier). La manera en que esto funciona es mediante la interpolación de todos los píxeles que están presentes en la zona de interés y proporcionando el día del cual se requiere reconstruir la imagen, en la Figura A.4 se puede observar el resultado de efectuar este proceso.



(a)

(b)

FIGURA A.4: Interpolación de imagen para el día 100: (a) Serie de Fourier de grado 4, (b) Método de Whittaker

# Apéndice B

## Productos de la tesis



El Centro de Investigación en Matemáticas A.C.  
& la Facultad de Informática Mazatlán

In recognition and appreciation to:

**Humberto Cruz-Sanabria, María Guadalupe Sánchez, Juan Pablo Rivera-Caicedo, Himer Avila-George**

In the international Conference CIMPS 2020

with the article entitled as:

**Identificación de las etapas fenológicas del cultivo de caña de azúcar usando imágenes Sentinel-2.**

CIMPS was held at the Faculty of Informatics of Mazatlán, Sinaloa, México, October 21st-23rd,

**M. C. Rogelio Estrada Lizarraga**  
Director of the Faculty of Informatics  
Mazatlán, Sinaloa, México

**Dr. Jezreel Mejía Miranda**  
CIMPS Chair  
CIMAT A.C. Unidad Zacatecas, México



**CIMPS 2020**



# MAPPING SUGARCANE USING VEGETATION INDICES AND TIME SERIES OF SENTINEL-2 IMAGES

Humberto Cruz-Sanabria<sup>1</sup>, María Guadalupe Sánchez<sup>1</sup>, Juan Pablo Rivera-Caicedo<sup>2</sup>, Himer Avila-George<sup>3</sup>

<sup>1</sup>Departamento de Sistemas y Computación, TecNM - Instituto Tecnológico de Ciudad Guzmán, Ciudad Guzmán, Jalisco, México.

<sup>2</sup>CONACYT - Secretaría de Investigación y Posgrado, Universidad Autónoma de Nayarit, Tepic, Nayarit, México.

<sup>3</sup>Departamento de Ciencias Computacionales e Ingenierías, Universidad de Guadalajara, Ameca, Jalisco, México.

## ABSTRACT

Sugarcane is one of the most important crops in Mexico; but it is facing increasing issues in its productivity, making necessary implementing monitoring techniques to improve crop management. To contribute to solve the problem, a method to identify anomalies in sugarcane cultivations was proposed through image data obtained from MultiSpectral Instrument sensor (MSI) onboard Sentinel-2 satellites. Three main steps were defined in the methodology: (1) image pre-processing, (2) Index calculation, and (3) anomaly detection. Two study areas were defined. On one hand, an area with a thorough control of the crop was selected, of which a health profile was defined by using PVI and LAI indices. On the other hand, a test area was selected to detect possible anomalous zones in the crop. Results showed that test area of sugarcane crop showed anomalies from day 200. To conclude, determining areas presenting deficiencies in sugarcane crops was possible through the proposed method.

**Index Terms**— Sugarcane, Sentinel-2, spectral indices

## 1. INTRODUCTION

Sugarcane (*Saccharum officinarum*, L.) is a significant crop in Mexico. In 2017-2018 harvest, sugarcane agribusiness represented 0.34% of the national gross domestic product, 2.14% from the manufacturing industry, and 9.80% from the food industry [1]. Sugarcane cultivation also generates many direct and indirect jobs, mainly in economically vulnerable populations.

One of the main problems faced by sugarcane producers is the improvement of crop yield, since growth rate rose to barely 2.4% in 2017, mainly due to the increase in surface dedicated to this crop [2] indicating a need to implement technological improvements that guarantee a profitable, competitive, and sustainable agribusiness.

Given the importance of sugarcane in Mexico, extracting information from data obtained by sensors transported by satellites is a highly relevant research area that can provide low-cost support to producers to identify limitations in the development of sugarcane crops that can affect crop yield; some examples can be found on [3], [4], and [5].

This study aimed to present a method to identify anomalies in sugarcane cultivation through a health profile constructed with Perpendicular Vegetation Index (PVI) and Leaf Area Index (LAI) using Sentinel-2 MSI sensor data. Two study areas with similar conditions of irrigation systems and sugarcane plantation were used for anomaly detection. On one hand, an area with a thorough control of the crop was selected, of which a health profile was defined. On the other hand, a test area was selected to detect possible anomalous areas in the crop. The test area was monitored during the 2019-2020 season.

## 2. MATERIALS AND METHODS

### 2.1. Study area

Two different study areas were selected for this research: (1) the reference area ( $Z_1$ ) (19° 40'01.4" N, 103° 24'42.6" W) and (2) the test area ( $Z_2$ ) (19° 40'09.2" N, 103° 24'50.7" W).  $Z_1$  was considered as a reference area since the sugarcane producer had a thorough control of his crop, producing 153.755 tons of sugarcane per hectare during the 2019-2020 season, being higher than the national field performance during the 2018-2019 season [2].  $Z_2$  was selected as a different zone for monitoring results while carrying out the tests. Both areas were located in Zapotiltic, Jalisco, Mexico and had the same drip irrigation system, cane variety AT MEX-96-40, and the same number of cuts. The total cultivated area in  $Z_1$  is 2.9 Ha and in  $Z_2$  is 2.7 Ha. A process was defined in order to classify a crop as healthy or unhealthy.

## IGARSS 2021 Review Results [Paper #4092]

 Traducir mensaje a: Español | No traducir nunca de: Inglés

12

IGARSS 2021 <papers@igarss2021.com>

Mar 16/03/2021 13:05

Para: HUMBERTO CRUZ SANABRIA; Maria Guadalupe Sánchez Cervantes; Juan Pablo Rivera <jprivera@uan.edu.mx>; Himer Avila <himer.avila@academicos.udg.mx>

PLEASE only reply to this email if absolutely necessary. Read this email carefully as it probably contains all of the information you need!

Dear Humberto Cruz, María Guadalupe Sanchez, Juan Pablo Rivera, Himer Avila,

The IGARSS 2021 Technical Program Committee has completed the review process, and we are pleased to inform you that your manuscript

#4092, titled 'Mapping sugarcane in Jalisco, Mexico using vegetation indices and time series of Sentinel-2 images'

has been ACCEPTED for inclusion as a multi-media presentation in the IGARSS 2021 program. [Congratulations!](#)

What should you do now?

1) Improve your paper

If specific comments about your manuscript were received from the reviewers or the technical chair managing your submission, they are provided at the end of this email. Please consider such comments and recommendations when preparing the final 4-page paper. In addition, we encourage authors to make data and source code available to advance the field and ensure reproducibility of the results. It is the responsibility of all authors of accepted papers to prepare the best possible communications to make a successful IGARSS 2021 meeting!

All attendees of IGARSS 2021 will receive electronic versions of the 4-page full papers accepted for the technical program. You will have until 1 June 2021 to revise your 4-page manuscript, formatted according to the paper kit ([https://igarss2021.com/Papers/paper\\_kit.php](https://igarss2021.com/Papers/paper_kit.php)). Only papers that are presented at IGARSS 2021 will be included in the published Proceedings.

IMPORTANT: During the paper submission process, there was a section to indicate if you wished NOT to publish your paper in the official proceedings. When you submit a revision to your accepted paper, please double-check that you have indicated your desired response. If you indicate a response of "DO NOT PUBLISH manuscript in proceedings," there will be no way to change that decision after 1 June 2021.

2) Check your schedule

A preliminary schedule of technical sessions will be posted in the near future on the IGARSS 2021 website (<https://igarss2021.com/>). A final programme will be published on June 4th, 2021.

3) Register to the conference

IGARSS 2021 registration is offered at attractive reduced rates according to the following tariffs:

Activate Windows  
Go to Settings to activate Windows.

**Asunto: Carta de Terminación de estancia virtual**

**A QUIEN CORRESPONDA**

Por este medio informo que el alumno **Humberto Cruz Sanabria**, estudiante de la Maestría en Ciencias de la Computación en el Instituto Tecnológico de Ciudad Guzmán Jalisco con número de matrícula **M19291006**, ha realizado una estancia virtual de investigación en el IPL (Laboratorio de Procesado de Imágenes) bajo mi supervisión en el periodo comprendido del 11 de febrero al 11 de marzo del 2021.

Durante esta estancia Humberto realizó la implementación en Python de diversas técnicas para el análisis series temporales utilizando imágenes satelitales.

Sin más por el momento agradezco al Ing. Humberto Cruz Sanabria el trabajo realizado a cabo y por haber decidido realizar una estancia en el IPL.

Atentamente,



Valencia, España a 17 de marzo de 2021

-----

Dr. Santiago Belda

Post doc

University of Valencia

Image Processing Laboratory (IPL)

C/ Catedrático José Beltrán, 2, 46980 Paterna (Valencia)

<http://ipl.uv.es/leo/>

