



**DESARROLLO DE UN MODELO PREDICTIVO DE PROCESOS  
GAUSSIANOS PARA LA ESTIMACIÓN DE LA PRODUCTIVIDAD  
DE LA CAÑA DE AZÚCAR EN HACIENDAS DEL INGENIO  
INCAUCA**

**MAESTRÍA EN TECNOLOGÍAS DE LA INFORMACIÓN  
GEOGRÁFICA**

**UNIVERSIDAD DE MANIZALES**



**UNIVERSIDAD DE  
MANIZALES®**

 Acreditación Institucional  
de Alta Calidad

**Desarrollo de un Modelo Predictivo de Procesos Gaussianos para la Estimación de la Productividad de la Caña de Azúcar en Haciendas del Ingenio Incauca**

Jose Manuel Collazos Vargas

Director:

Ing. Aldemar Reyes Trujillo, PH. D

Codirector:

Ing. Juan Pablo Rivera Caicedo, PH. D

Universidad de Manizales

Facultad de Ciencias e Ingeniería

Maestría en Tecnologías de la Información Geográfica

Enero 31 de 2023

## **AGRADECIMIENTOS Y DEDICATORIA**

Agradezco a mis directores Aldemar Reyes y Juan Pablo Rivera por su compromiso y dedicación para guiarme durante el desarrollo de este trabajo de investigación, por aportar su conocimiento y contribuir con mi formación académica.

Agradezco a la Universidad de Manizales, a cada uno de los docentes que aportaron su conocimiento y experiencia durante mi proceso en la maestría.

Dedico este trabajo primeramente a Dios pues es mi guía y me ha brindado la bendición de seguir formándome académicamente y como persona, a mis padres, sin su apoyo incondicional, los valores inculcados, el amor y el deseo por superarme día a día, no fuera posible alcanzar esta meta.

## RESUMEN

Estimar la producción del cultivo de caña de azúcar en edades temprana es fundamental para la gestión de prácticas agronómicas, el propósito de este trabajo fue desarrollar un modelo predictivo de procesos Gaussianos para estimación de la productividad del cultivo de caña de azúcar a través del uso multitemporal (2017 – 2021) de imágenes satelitales Sentinel 1 y el cruce con variables climáticas, dado que la zona de estudio se encuentra en ambientes cálidos y húmedos con alta nubosidad, por lo que es difícil la utilización de imágenes satelitales. La información se consolidó en dos bases de datos de manera independiente, por un lado, los datos del radar de apertura sintética (SAR) y por otro las variables climáticas, de las cuales fueron extraídas las características: media, la desviación estándar, la pendiente, el área bajo la curva, sesgo y curtosis, datos con los cuales se alimentaron los modelos, que fueron evaluados por etapa fenológica del cultivo. El modelo con mejor resultado lo arrojó los datos SAR, evaluando la polarización VH, utilizando la combinación de variables (Desviación estándar, Pendiente) en la fenología 1 obteniendo un error de 11.11 TCH. El uso de datos SAR ofrece una alternativa con alto potencial para el monitoreo y estudio de la productividad en zonas con alta nubosidad, donde se limita la utilización de imágenes ópticas.

***Palabras clave:*** *Caña de azúcar, modelo predictivo, Radar de Apertura Sintética SAR, Sentinel 1, variables climáticas*

## CONTENIDO

AGRADECIMIENTOS Y DEDICATORIA .....	3
RESUMEN .....	4
LISTA DE FIGURAS.....	8
LISTA DE TABLAS .....	9
1. PLANTEAMIENTO DEL PROBLEMA DE INVESTIGACIÓN.....	10
1.1 Descripción del área problemática.....	10
1.2 Formulación del problema .....	11
1.3 Justificación .....	12
2. ANTECEDENTES .....	14
2.1 Importancia de la predicción de la productividad de los cultivos.....	14
2.2 Predicción de productividad usando imágenes satelitales. ....	15
2.3 Aplicación de Geoestadística y Maching Lerning para la predicción de productividad.....	17
2.4 Aplicación de métodos combinados para predecir la producción del cultivo de caña de azúcar.....	20
2.5 Uso de UAV para la predicción de la producción de la caña de azúcar .....	23
3. OBJETIVOS .....	25
3.1 Objetivo general.....	25
3.2 Objetivos específicos .....	25

4.	REFERENTE CONTEXTUAL .....	26
5.	REFERENTE NORMATIVO Y LEGAL .....	28
6.	REFERENTE TEÓRICO.....	30
6.1	Caña de azúcar .....	30
6.2	Agricultura de precisión.....	30
6.3	Percepción remota.....	32
6.4	Radiación electromagnética .....	35
6.5	Sistema sensor.....	36
6.6	Minería de datos.....	37
7.	HIPÓTESIS DE INVESTIGACIÓN .....	39
8.	METODOLOGIA .....	40
8.1	Enfoque metodológico .....	40
8.2	Tipo de estudio.....	40
8.3	Diseño de la investigación .....	40
8.4	Población o unidad de trabajo.....	41
8.5	Muestra y muestreo.....	41
8.6	Operacionalización de variables .....	41
8.7	Técnicas e instrumentos de recolección de información .....	42
8.8	Desarrollo metodológico.....	42
8.8.1	Imágenes satelitales .....	42

8.8.2	Preprocesado de imágenes .....	43
8.8.3	Datos climáticos.....	45
8.8.4	Datos de productividad del Cultivo. ....	46
8.8.5	Creación de la base de datos .....	46
8.8.6	Extracción de características.....	47
8.8.6.1	Etapas de análisis .....	47
8.8.6.2	Extracción de características.....	48
8.8.6.3	Algoritmo de Procesos Gaussianos (GPR) .....	50
8.8.6.4	Validación del modelo .....	51
8.8.6.5	Error cuadrático medio MSE .....	51
8.8.7	Selección de variables.....	52
9.	DISCUSIÓN Y RESULTADOS .....	53
10.	CONCLUSIONES .....	57
	BIBLIOGRAFÍA .....	58

## LISTA DE FIGURAS

Ilustración 1 Mapa de localización .....	26
Ilustración 2 Características de las haciendas donde se desarrollará la investigación.....	27
Ilustración 3 Proceso de la percepción remota. ....	34
Ilustración 4 Modelo de preprocesado Sentinel 1 .....	45
Ilustración 5 Bases de datos construidas para el desarrollo de la investigación.....	47
Ilustración 6 Características extraídas de las series temporales.....	49
Ilustración 7 Proceso de extracción de características .....	49
Ilustración 9 Modelo de TCH estimado vs TCH medido en campo.....	55

## LISTA DE TABLAS

Tabla 1 Productividad histórica a nivel de suerte .....	46
Tabla 2 Combinaciones de las bases de datos .....	52
Tabla 3 Resultados del modelo al usar las bases de datos SAR .....	53
Tabla 4 Resultados del modelo al usar las bases de datos climáticos.....	54

## 1. PLANTEAMIENTO DEL PROBLEMA DE INVESTIGACIÓN

### 1.1 Descripción del área problemática

La caña de azúcar (*Saccharum spp.*) es una materia prima que permite la obtención de azúcar, esta se utiliza como insumo en la industria de alimentos, en Colombia la agroindustria de caña de azúcar, representa el 0,6% del PIB Total Nacional y el 3,7% del PIB Agrícola nacional, en el Valle del Cauca representa el 31,4% del PIB Agrícola, generando 286 mil empleos directos e indirectos. La producción en 2020 fue 24,3 millones de toneladas de caña; 2,3 millones de toneladas de azúcar; 438 millones de litros de bioetanol a partir de caña y 1.615 GWh de energía generada a partir del bagazo de la caña, lo suficiente para para abastecer una ciudad de 1 millón de habitantes (Asocaña, 2022). Las condiciones edafoclimáticas del valle del río Cauca permiten que la caña de azúcar pueda ser cultivada a lo largo de todo el año, en esta región hay actualmente 241.205 hectáreas sembradas en caña de azúcar, de las cuales se pudo obtener una producción promedio de 127 toneladas de caña por hectárea (TCH) y rendimiento de 13,6 toneladas de azúcar por hectárea (TAH) (Asocaña, 2022).

El desarrollo del sector de la caña de azúcar implica que los proveedores y los ingenios inviertan en investigación y aplicación de nuevas tecnologías en la búsqueda de mejorar su producción y disminuir los costos en el desarrollo del cultivo, este proceso no es algo fácil, dado que el país conserva en gran proporción las prácticas de agricultura tradicional, tal como indica Corporación Colombiana de Investigación Agropecuaria (CORPOICA), (2015) “el poco desarrollo que el tema tiene en el país y del necesario cambio de paradigma que implica pasar de métodos de transferencia de tecnología lineales a procesos que reconocen que la innovación no

es un fenómeno individual sino un fenómeno social y sistémico”. La implementación de la agricultura de precisión es una de las propuestas hacia la innovación y desarrollo del sector, sin embargo, su desarrollo no ha podido ser consolidado por la falta de información asociada a los datos espectrales de gran utilidad en un programa de agricultura de precisión para la monitorización del cultivo. No obstante, las condiciones geográficas limitan su obtención y aprovechamiento por aspectos asociados a la alta nubosidad, tierras escarpadas, etc. La falta de información base y la limitada situación socioeconómica en Colombia, que hacen que la utilización de herramientas como la detección remota satelital o aerotransportada, tenga muchas limitaciones por su falta de disponibilidad real, alto costo y complejidad operativa. En Colombia las técnicas de percepción remota satelital presentan un escaso desarrollo y aplicación en sistemas biológicos vegetales que aprovechen el potencial del seguimiento espaciotemporal de los cultivos, por lo que es importante contar con información confiable que facilite la trazabilidad y el manejo óptimo de los sistemas productivos agrícolas. (Reyes-Trujillo et al., 2021), por lo que esta investigación apuesta a la implementación de técnicas de las Tecnologías de Información Geográficas – TIG, en busca de métodos novedosos de agricultura de precisión que contribuyan al mejoramiento en el rendimiento y la producción del cultivo, combinando los métodos tradicionales con las nuevas tecnologías, logrando así una a reducción en las pérdidas económicas y la afectación ambiental en el sector de la caña de azúcar.

## **1.2 Formulación del problema**

La carencia de un método fiable que permita predecir la producción del cultivo de caña de azúcar es el principal motivo para el desarrollo de esta investigación, por lo que se busca

responder al interrogante de si ¿es posible crear un modelo que permita predecir de manera multitemporal la producción del cultivo de caña de azúcar?

### **1.3 Justificación**

La predicción del rendimiento es importante para las prácticas agrícolas, la implementación de modelos de agricultura de precisión va encaminado al éxito de la producción agrícola dado que el cultivo depende en gran medida de la detección de enfermedades, malezas, estrés de los cultivos y la determinación de las condiciones de diversas variables biofísicas las cuales al ser detectadas a tiempo pueden ser corregidas para no afectar la producción final del cultivo. La Agricultura de Precisión es un campo que ha venido cobrando fuerza en los últimos años, puesto que reúne diferentes Tecnologías de la Información Geográfica -TIG dentro de las cuales se destacan la utilización de imágenes satelitales, sensores remotos, datos GNSS, los Sistemas de Información Geográfica SIG, entre otros, todas estas encaminadas a la correcta toma de decisiones entorno a la reducción en los costos de los insumos, la correcta distribución de los cultivos, la eficacia en la gestión agronómica, la precisión de las labores, la eficiencia en los equipos empleados para la producción agrícola y el cuidado de los recursos naturales, todo esto en la búsqueda de mejorar el rendimiento y la producción de los cultivos.

La implementación de nuevas estrategias y el uso de las tecnologías en desarrollo involucradas en la Agricultura de Precisión, el Internet de las Cosas (IoT), la Inteligencia Artificial (IA), Big Data, Machine learning, entre otras, potencian la precisión de los modelos predictivos y permiten gestionar adecuadamente la producción agrícola. “El desarrollo de un sistema preciso de predicción del rendimiento de los cultivos a gran escala es de suma importancia para la gestión de los recursos agrícolas y la seguridad alimentaria mundial” (Li et al, 2022). Por lo que la observación de la tierra mediante el uso métodos como los sensores

remotos, permite el monitoreo desde una diversidad de rangos espectrales, favoreciendo la gestión de los cultivos.

Estas tecnologías brindan herramientas que no solo contribuyen a la identificación de la producción agrícola, sino que traen beneficios para el ambiente en el que se desarrollan los cultivos. Antony (2021) plantea que estas tecnologías se convierten en herramientas de gestión que permiten abordar diferentes problemas como la conservación del agua y del suelo, mediante estrategias como la aplicación adecuada de abonos y plaguicidas o el riego justo a los cultivos, beneficiando la productividad. Sin embargo, el determinar la producción de los cultivos es complejo dado que influyen muchos factores como la salud de suelo, la climatología, variedad de los cultivos y la información de teledetección no refleja completamente su verdadera situación, por lo que la precisión en la predicción se verá afectada. Es por eso que el complementar la información satelital con datos como tipo de suelo, variables climáticas y variedad de cultivo mejora el rendimiento a la hora de implementar un modelo predictivo (Tian et al., 2020)

Todo el conjunto de prácticas con su notorio desarrollo beneficiarían en gran medida a los ingenios y fincas productoras de caña de azúcar, quienes serán los principales beneficiados con la identificación de un modelo que haga uso de las tecnologías de la información geográfica y el cruce de variables climáticas para predecir el rendimiento en los cultivos, lo cual permitirá la correcta toma de decisiones en la búsqueda potenciar la producción que se verá refleja en las toneladas de caña por hectárea de cada finca productora.

## 2. ANTECEDENTES

La literatura revisada fue de gran ayuda para conocer como los grandes productores de caña de azúcar han implementado diferentes métodos para la estimación de la producción del cultivo, estos han sido generados bajo varias metodologías como la predicción con imágenes satelitales, las predicciones con variables biofísicas, las realizadas con métodos de machine learning y los métodos combinados, tal como se evidencia a continuación:

### 2.1 Importancia de la predicción de la productividad de los cultivos

El determinar la productividad del cultivo en edades tempranas es muy útil para las enmiendas y la toma de decisiones sobre el cultivo, autores como Everingham et al. (2002), realizaron una revisión de diferentes métodos en busca de un enfoque integral para usar sistemas de pronóstico climático estacional con el objetivo de mejorar la gestión de riesgos y la capacidad de toma de decisiones en todos los sectores de la industria de la caña de azúcar, se destaca la de esta investigación el proceso de investigación participativa de 'aprendizaje activo' el cual facilitó en gran medida la integración de los sistemas de pronóstico climático en un marco de decisión.

Así mismo, Lawes & Lawn (2005), llevaron a cabo un estudio en el que se revisan las diferentes formas en que la información de la industria ha sido utilizada para facilitar la toma de decisiones y su valor en la identificación de problemas tales como una disminución crónica de CCS en los trópicos húmedos. Destacando que la información proporcionada por la industria puede brindar información valiosa sobre el sistema de producción de caña de azúcar y ayudar en la toma de decisiones dentro de la industria. los análisis de la información suministrada por la industria deben continuar para garantizar que las tendencias de productividad se controlen de

cerca y se identifiquen las amenazas y oportunidades para que se puedan tomar medidas oportunas en beneficio de la industria.

## **2.2 Predicción de productividad usando imágenes satelitales.**

La información proveniente de los diferentes sensores enriquece los estudios de productividad, Bégué et al. (2010), propusieron indicadores espectrales apropiados para la previsión de rendimiento, para hacer esto, utilizaron imágenes provenientes los satélites SPOT4 y SPOT5 adquiridas mensualmente durante un período de 2 años. Se logro mostrar que las variaciones en el Índice de Vegetación de Diferencia Normalizada (NDVI) de la caña de azúcar a escala de campo es resultado de la interacción entre el calendario de cultivo de la caña de azúcar y la fenología de la planta en un ambiente climático dado. Se destaca como principal hallazgo que la comprensión de la fisiología y las prácticas culturales de la caña de azúcar son cruciales si se quiere que las aplicaciones de teledetección tengan éxito en las predicciones de rendimiento.

Por otra parte Shendryk et al. (2021), usaron las imágenes satelitales Sentinel-1 y Sentinel-2 en combinación con datos de clima, suelo y elevación para predecir el rendimiento de la caña de azúcar, este estudio demostró el potencial de usar imágenes satelitales Sentinel-1 y Sentinel-2 junto con otros atributos geográficos para la predicción temprana del rendimiento de caña y rendimiento de azúcar, dejando como aporte que el éxito de la predicción del rendimiento de la caña de azúcar depende del análisis de series de tiempo y la consideración de atributos, como el clima, el suelo, la elevación, la variedad y la información sobre el número de retoños. Así como resaltar el beneficio que generalmente, una predicción temprana del rendimiento con una confianza "adecuada" es de mayor valor que una predicción más precisa que llega demasiado tarde.

Del mismo modo Rahman & Robson (2020), exploraron un enfoque que como el expuesto anteriormente también se basa en la teledetección, pero en esta ocasión combina imágenes Landsat y Sentinel para predecir el rendimiento de la caña de azúcar a nivel de bloque mediante el desarrollo adicional de un modelo de serie de tiempo regionalmente específico e incluyendo la fecha de siembra de cultivos individuales. El modelo de series temporales GNDVI derivado de Landsat-8 y Sentinel-2 muestra la progresión del crecimiento del cultivo de caña de azúcar en diferentes fechas de siembra o retoño. El estudio deja en evidencia que la transferencia de un modelo de series de tiempo se puede lograr desde Landsat-8 y Sentinel-2 o viceversa, así como que La comparación de GNDVI entre los dos sensores mostró una cierta cantidad de dispersión, que podría deberse a una serie de posibles causas, como el registro incorrecto de la imagen y los efectos de visualización de la geometría y la longitud de la trayectoria atmosférica, que podrían no contabilizarse por completo.

Así mismo, Muller et al. (2020), examinaron las bandas e índices del satélite multiespectral Landsat-8 que son más apropiados para modelar la fracción de radiación fotosintéticamente activa absorbida - fAPAR en entornos de producción particulares y evaluó la solidez de los modelos para su aplicación en áreas donde se cultiva caña de azúcar en condiciones agroclimáticas variables, descubriendo que los modelos basados en SWIR parecen ser más robustos con AFRI y NDMI basados en SWIR 1 funcionando consistentemente bien. También Alemán Montes (2022), evaluó los sensores satelitales MODIS, Landsat Oli 8 y Sentinel 2 para el desarrollo del sistema de monitoreo satelital en las plantaciones de caña de azúcar, la combinación de índices de vegetación (EVI y SAVI) con variables históricas de rendimiento permitieron construir un modelo de estimación de rendimiento de campo, además, se logró estimar el total de toneladas de caña de azúcar producidas. De este trabajo se concluye

que, los SIG y la Teledetección tienen un alto potencial para la agricultura costarricense a través de herramientas que permitan la toma de decisiones informadas y apoyadas en conceptos fundamentales de la agricultura de precisión como lo son, el dónde y el cuándo realizar acciones de manejo en las plantaciones agrícolas.

Finalmente Rahimi Jamnani et al. (2019), buscaron identificar la fecha adecuada para adquirir las mejores imágenes satelitales para lograr la máxima correlación entre los índices de vegetación y el rendimiento de fin de temporada a escala de finca. Así mismo se intenta el modelo óptimo y el mejor índice de vegetación para predecir el rendimiento de la caña de azúcar irrigada, anticipando la cosecha prevista tres meses antes. Los índices de vegetación muestran la máxima correlación con la planta rendimiento cuando alcanza su valor máximo que es de unos tres meses antes de la cosecha, el modelo de Red Neuronal Artificial - ANN exhibió un mejor rendimiento y el GVI proporcionó una mejor previsibilidad en comparación con GNDVI y NDVI en todos los escenarios.

### **2.3 Aplicación de Geoestadística y Maching Larning para la predicción de productividad**

Los autores de Almeida et al. (2021), desarrollaron un estudio para determinar entornos de producción de caña de azúcar utilizando un número reducido de variables de bajo costo a través de la técnica de aprendizaje automático. En el desarrollo de la investigación se hizo uso de los métodos de estadística descriptiva para comprender el comportamiento de los datos, seguida de la regresión por pasos para determinar qué variables serían útiles para el modelo, después, se aplicó una prueba de multicolinealidad y un árbol de decisión y por último se elaboró una matriz de confusión para evaluar la eficiencia del modelo. La relación entre la geoestadística y los el aprendizaje automático de variables asociadas al suelo permitieron una perspectiva prometedora

para la evaluación del rendimiento del modelo de predicción, demostrando que los atributos del suelo son los factores que más impactan la producción agrícola.

Por su parte Erazo et al. (2013), determinaron y generaron mapas de variabilidad espacial usando geoestadística para entrenar una red neuronal de perceptrón multicapa. Esta red permitió calcular el peso de cada variable utilizando técnicas de métricas relevantes, encontrando que las variables con mayor impacto en la productividad dependen de los factores formadores del suelo, dentro de las cuales se destacan el manganeso (Mn) y el boro (B). Autores como Calier et al. (2015) aplicaron el modelo de Productividad Primaria Neta (PPN) a escala regional de Kumar y Monteith, donde realizaron análisis espacio-temporales con técnicas de geomática y caracterización edafoclimáticas del entorno, también, se aplicaron monitoreos de campo, de este modelo se obtuvieron mapas de distribución espacio temporales donde se mostraron las características del suelo y biofísicas de la plantación, se calcularon las variables fPAR, PAR, EUR de la ecuación del modelo, mediante percepción remota y modelos matemáticos relacionados a través del índice de vegetación de diferencia normalizada. Dentro del trabajo se encontraron algunos vacíos dado que fue posible observar que existe un amplio desfase entre los datos calculados y los datos obtenidos después de la cosecha, lo cual podría deberse a que el manejo de la producción y en la cosecha de la misma, pues el manejo que se da al cultivo de la caña de azúcar en la zona es netamente manual y tradicional.

Así mismo Han et al. (2022), desarrollaron un enfoque de pronóstico de rendimiento de caña de azúcar basado en datos de resolución fina que combinó una amplia gama de conjuntos de datos en edades tempranas y tardías del cultivo, disponibles públicamente y en la granja, teniendo como objetivo probar la practicidad de los pronósticos de rendimiento de principios de temporada al guiar el fertilizante de N de dosis variable y comparar estas recomendaciones con

los estándares de la industria, logrando demostrar que los pronósticos de rendimiento se pueden usar como una herramienta de apoyo a la toma de decisiones para ayudar a las recomendaciones de fertilizantes nitrogenados, al incorporarlos con las pautas existentes, del estudio se recomienda tener una resolución adecuada de las variables predictoras para no ser rechazadas por el algoritmo de bosque aleatorio.

Otros métodos utilizados por Yu et al. (2022), investigaron el valor de tres variables agronómicas comunes (altura de la planta, PH; índice de área foliar, LAI; y humedad del suelo, SM) para realizar simulaciones de crecimiento de caña de azúcar y explorar cuales variables tienen el mayor contenido de información, sugiriendo una estrategia de fusión de datos LAI y PH como alternativa a la estrategia tradicional de fusión LAI y SM para lograr mejores resultados de simulación con el menor costo de mano de obra y computación. En otro caso, Cevallos et al. (2019), aplicaron la metodología del aprendizaje profundo de la convolución de redes neuronales para realizar el procesamiento de imágenes satelitales y la obtención de datos indirectos, realizando una clasificación de las características del rendimiento de los cultivos, en la región Troncal de la Costa del Ecuador. El experimento mostró resultados afirmativos de aproximadamente un 95\% de probabilidad de reconocimiento del estado del cultivo. Se considera que el uso de más capas, un mayor número de imágenes y otros filtros geográficos podrían generar predicciones mucho más precisas y no solo para cultivos de caña de azúcar.

Por su parte Everingham et al. (2007), se aplicó un procedimiento de análisis discriminante bayesiano para determinar la probabilidad de una cosecha pequeña, mediana o grande en 4 regiones principales de cultivo de caña de azúcar en Australia. El modelo bayesiano considera rendimientos potenciales simulados, índices de pronóstico climático y el tamaño de la cosecha del año anterior.

En otro estudio realizados por algunos de los autores anteriores Everingham et al. (2009), utilizaron métodos de minería de datos estadísticos para seleccionar y asignar pesos a los mejores modelos. Una técnica, llamada aproximación de lazo, produjo los mejores resultados. El enfoque de conjunto que utiliza el algoritmo avanzado por etapas demostró ser una forma eficaz y eficiente de seleccionar y ponderar modelos para las predicciones de rendimiento. Los modelos de subconjuntos obtenidos del método de conjunto tenían la ventaja adicional de que representan condiciones ambientales.

#### **2.4 Aplicación de métodos combinados para predecir la producción del cultivo de caña de azúcar**

El uso de datos climáticos y físicos del cultivo en combinación con datos espectrales es otra de las alternativas planteadas por diferentes autores para determinar la productividad del cultivo, autores como Young et al., (2019), presentaron un trabajo en el que utilizaron técnicas de aprendizaje automático aplicadas a un modelo biofísico específico para las características de la caña de azúcar, en dicho modelo se utilizaron variables tales como los datos históricos del suelo, clima, manejo, condiciones fitosanitarias, nutrición y resultados de producción. Se destaca de este estudio la combinación entre las variables biofísica, los datos espectrales de los cultivos y modelos calibrados derivados de antecedentes y pronósticos de clima, lo que permite que un único modelo pueda aprovechar información valiosa de operaciones independientes relacionadas con la caña de azúcar.

Por su parte Morel et al. (2014) realizaron un estudio donde compararon tres métodos de estimación de rendimiento basados en sensores remotos tomando como partida un conjunto de datos de campo en fincas con condiciones climáticas contrastadas y con diferentes prácticas agrícolas, el primer método fue la relación empírica con un Índice de Vegetación de Diferencia

Normalizada integrado en la temporada de crecimiento NDVI, el segundo fue un modelo de eficiencia de Kumar-Monteith y el último fue el método de acoplamiento forzado con un modelo de cultivo de caña de azúcar (MOSICAS) y una fracción de radiación fotosintéticamente activa absorbida derivada de satélites. De este estudio se destaca que el método más adecuado para la estimación del rendimiento de la caña de azúcar a escala de campo parece ser el modelo empírico NDVI porque es el modelo más simple de configurar y proporciona la estimación del rendimiento más precisa durante un ciclo completo del cultivo, así mismo se destaca la importancia de aplicar una serie de correcciones radiométricas antes del procesamiento de la imagen, para que sean comparables.

Del mismo modo dos Santos Luciano et al. (2021), se inclinaron por un modelo combinado el cual se centró en el desarrollo de modelos predictivos combinando información de diferente fuente, primero se usaron datos adquiridos y accesibles varios meses antes de la cosecha, la segunda fuente de información fueron datos agronómicos y meteorológicos y luego se realizó una combinación de imágenes de satélite Landsat, datos agronómicos y meteorológicos y así mismo se agregaron datos de la futura fecha de cosecha para ver como reaccionaban los modelos, se destaca que los índices de vegetación y las bandas de reflectancia de Landsat mejoraron el pronóstico del rendimiento de la caña de azúcar cuando se integraron con datos agronómicos y meteorológicos, además de esto se comprobó que el número de cosechas y el índice de vegetación NDMI fueron las variables más importantes para pronosticar el rendimiento de la caña de azúcar, así mismo el incluir la fecha de cosecha a los modelos mejoró aún más la predicción del rendimiento.

Así mismo Begue et al. (2008), realizaron un análisis de la variabilidad del cultivo de caña de azúcar a escalas de tiempo estacionales y anuales, donde vincularon la variabilidad con

factores ambientales como (clima, topografía y profundidad del suelo) y de cultivo (fecha de cosecha). Se aplicó un análisis de serie temporal del índice de vegetación de los campos de caña de azúcar (NDVI) de quince imágenes SPOT y un método de clasificación original que permitió centrarse en la variabilidad espacial de los cultivos independientemente de etapas de crecimiento del cultivo. Se logró determinar que es necesario conocer la fenología del cultivo para interpretar correctamente el patrón espacial, y que las imágenes de una sola fecha pueden ser insuficientes para el diagnóstico de las condiciones del cultivo o para las predicciones.

En otros estudios Lofton et al. (2012), determinaron si el potencial de rendimiento de la caña de azúcar podría estimarse utilizando una estimación en temporada del índice vegetativo de diferencia normalizada (NDVI). Los valores estimados de rendimiento durante la temporada (INSEY) se calcularon dividiendo el NDVI por las variables térmicas. Se determinó que el marco de tiempo óptimo para estimar el YP de la caña de azúcar es de 601 a 750 CGDD. Debido a que este marco de tiempo está fuera de las recomendaciones actuales para la fertilización con N, los productores de caña de azúcar necesitarían retrasar la fertilización con N durante la temporada por un mes para integrar el potencial de rendimiento en un esquema de manejo de N.

Por otra parte Fernandes et al. (2017), seleccionaron 60 municipios y extrajeron métricas espectrales de la serie temporal de NDVI para cada municipio desde 2003 hasta 2012. Se aplicó un envoltorio de red neuronal con eliminación secuencial hacia atrás para eliminar características irrelevantes y/o redundantes del conjunto de datos inicial, reduciendo el sobreajuste.

Posteriormente, se predijo el rendimiento de la caña de azúcar utilizando un modelo de conjunto de apilamiento con ANN. El método Stacking pudo estimar el rendimiento de la caña de azúcar para el estado de São Paulo con un RMSE más pequeño que los levantamientos de datos oficiales, anticipando el pronóstico de cosecha tres meses antes de la cosecha.

## 2.5 Uso de UAV para la predicción de la producción de la caña de azúcar

Los vehículos aéreos no tripulados (UAV) contribuyen al estudio de los diferentes cultivos, estos tienen la ventaja de poder operar en alturas inferiores a la de los satélites y obtener escenas libres de nubes lo que mejora la calidad de la información, autores como Sumesh et al. (2021), realizaron una evaluación del potencial de una cámara RGB montada UAV para la estimación del rendimiento de la caña de azúcar con un conjunto mínimo de datos de campo. Lograron mapear la altura de la planta y la densidad de tallos a nivel de una cuadrícula de 4 m x 4 m. Se utilizó un enfoque de análisis de imágenes basado en objetos (OBIA) para extraer el área de caña de azúcar integrando el modelo de altura de la planta (PHM), extraído al restar el modelo de elevación digital (DEM) del modelo de superficie de cultivo (CSM). Este estudio deja apreciar que los UAV tienen alto potencial para estimar la altura de la planta, la densidad del tallo y el rendimiento con una alta precisión lo que lo hacen un insumo importante a la industria azucarera, además de que no presentan las limitaciones de los sensores comerciales como una menor resolución espacial, nubosidad y atenuación atmosférica, lo que permite que la predicción del rendimiento solo tenga un éxito moderado, por lo que las imágenes de ultra alta resolución tomadas por UAV pueden superar las limitaciones de estos enfoques y pueden ser una solución para incorporar sensores remotos a nivel de granja.

Así mismo Yu et al. (2020), lograron asimilar los datos de altura de las plantas en el modelo de cultivo para mejorar la estimación del rendimiento y optimizar la gestión del agua agrícola. Los datos de altura fueron monitoreados por medio de UAV, a los datos obtenidos se suma el uso del modelo SWAP que considera la simulación de la altura de la planta junto con un algoritmo más suave de conjunto iterativo que se construye para simular la altura de la planta y

el procedimiento de crecimiento del cultivo. Demostrando que la variable de estado del cultivo y el rendimiento a escala de campo se puede mejorar asimilando las medidas de altura de la planta en el modelo, además de que la comparación de los datos de altura de la planta medidos en el suelo y la incorporación de las medidas derivadas de UAV.

Del mismo modo Chea et al. (2020), utilizaron un sensor MicaSense RedEdge en un UAV, recolectando imágenes que permitieron generar mapas en cinco bandas de mapas de reflectancia (azul, verde, rojo, NIR y rededge), esto permitió generar mapas con seis diferentes índices de vegetación, para que luego estos se combinaran con valores de referencia promedio para crear modelos de regresión lineal simple. Se encontró que CIrededge muestra la mejor correlación con Pol ( $R^2$  0.77) y CCS ( $R^2$  0.68), independientes de la variedad, mientras que los modelos Brix dependen de la variedad y requieren diferentes índices de vegetación.

### 3. OBJETIVOS

#### 3.1 Objetivo general

- Desarrollar un modelo de procesos Gaussianos que permita predecir la producción del cultivo de caña de azúcar a través del uso multitemporal de imágenes satelitales Sentinel 1 y el cruce de variables climáticas.

#### 3.2 Objetivos específicos

- Consolidar la información satelital, climática y de productividad del cultivo de caña de azúcar de las fincas seleccionada a través de la construcción de una base de datos multitemporal.
- Determinar cuáles son las variables climáticas que están directamente relacionadas con la producción del cultivo de caña de azúcar.
- Identificar la respuesta espectral del cultivo de caña de azúcar usando imágenes multitemporales Sentinel 1.

#### 4. REFERENTE CONTEXTUAL

La investigación se ubicó al suroccidente de Colombia en los municipios de Corinto, Miranda y Padilla, sobre seis Haciendas pertenecientes al Ingenio del Cauca S.A. en suertes con manejo agronómico convencional, con similares condiciones edafoclimáticas, variedad (CC01-1940), edad y fertilización nitrogenada, abarcando un área total de 51.7 ha. El área de estudio se encuentra en promedio a 1005 msnm, con una temperatura media de 24° grados centígrados.

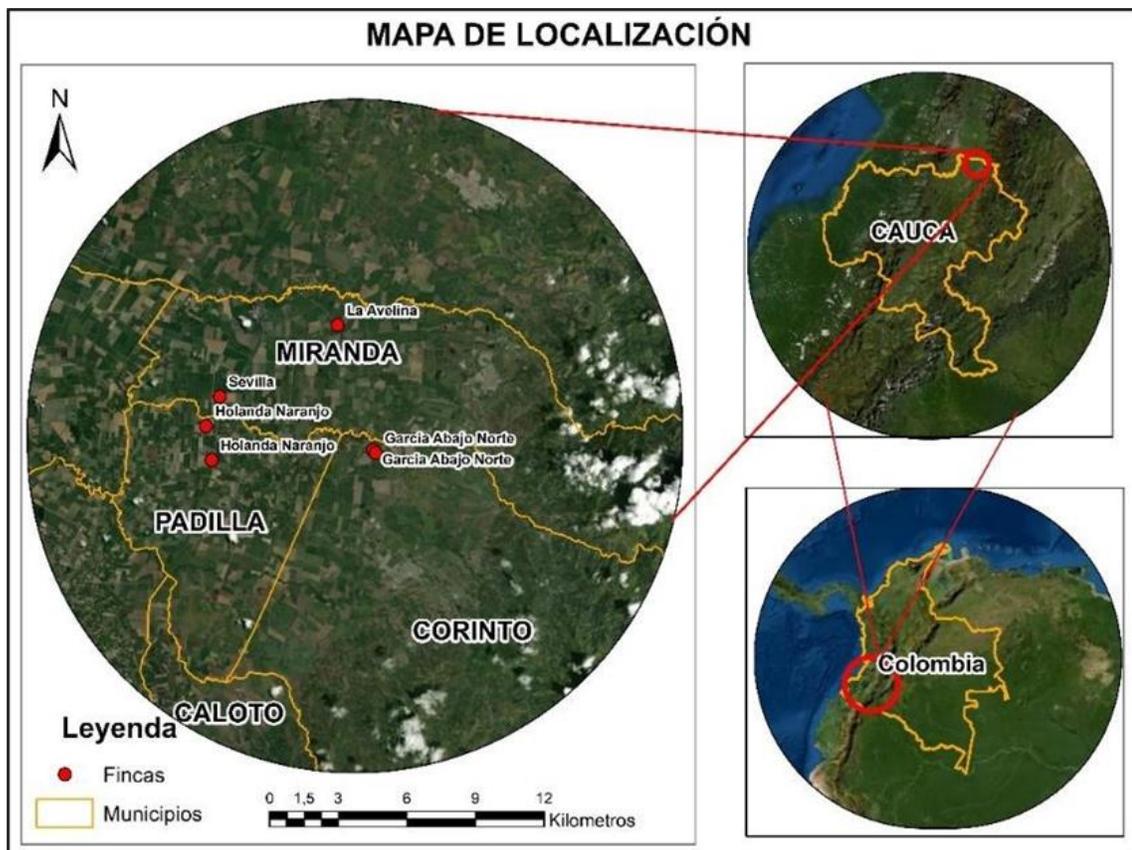


Ilustración 1 Mapa de localización

Nombre	Suerte	Área (ha)	Cortes	TCH	Textura	Fertilización Urea 46	Localización
GARCIA ABAJONORTE	009B	3.10	2	135	F-FA	350	3°13'19.4"-76°16'27.8"
GARCIA ABAJONORTE	008A	4.55	1	118	FAr-FArA-FL	350	3°13'15.0"-76°16'24.0"
LA AVELINA	002	5.85	1	86	F-FL	350	3°16'16.6"-76°17'18.0"
HOLANDA - NARANJO	015A	3.74	1	112	F-FL	350	3°13'53.4"-76°20'24.4"
HOLANDA - NARANJO	6	12.68	1	123	F-FL	375	3°13'05.7"-76°20'16.8"
SEVILLA	003A	5.4	1	132	F-FL	350	3°14'35.9"-76°20'05.6"

Ilustración 2 Características de las haciendas donde se desarrollará la investigación

## 5. REFERENTE NORMATIVO Y LEGAL

En esta parte del documento se abordan las normativas que tienen relación con la presente investigación, esto con el fin de tener referentes legales, normativos y técnicos para la utilización de imágenes satelitales. A continuación, se mencionan y listan las normas.

- **Decreto 290 DE 1957:** Mediante el Decreto 290 de noviembre de 1957, se creó el "Instituto Geográfico Agustín Codazzi", que tiene como misión producir, proveer y divulgar información y conocimiento, en materia de cartografía, agrología, catastro, geografía y tecnologías geoespaciales, y regular su gestión, en apoyo a los procesos de planificación y desarrollo integral del país y, de manera particular, a la gestión del territorio.
- **Decreto 2442 de julio de 2006:** Mediante el Decreto 2442 de julio de 2006, se creó la Comisión Colombiana del Espacio -CCE, como órgano intersectorial de consulta, coordinación, orientación y planificación, que tiene como fin orientar la ejecución de la política nacional para el desarrollo y aplicación de las tecnologías espaciales y coordinar la elaboración de planes, programas y proyectos en este campo, que contribuyan al aumento de la productividad, la eficiencia y la competitividad nacional.
- **CONPES 3585 de 2009:** "Consolidación de la Política Nacional de Información Geográfica y la Infraestructura Colombiana de Datos Espaciales –ICDE", en el cual se estableció que los datos de imágenes satelitales y de otros sensores remotos se consideran como datos fundamentales.

- **Instituto Geográfico Agustín Codazzi -IGAC, 2016:** Se crea el instructivo de funcionamiento del portal interno banco nacional de imágenes – BNI, por el grupo interno de trabajo de gestión de proyectos geográficos y cartográficos  
Para la incorporación de una imagen al aplicativo del BNI es necesario contar con la siguiente información:
  - ◆ Datos completos de la empresa que tomó la imagen
  - ◆ Grado o nivel de procesamiento en la que fue recibida la imagen
  - ◆ Fecha de la toma de la imagen
  - ◆ Siguiendo la grilla original de cada tipo de imagen, conocer la ubicación de la imagen en dicha grilla y obtener su nombre.
  - ◆ Cubrimiento Nacional, Departamental y Municipal de la imagen.
  - ◆ Resolución espacial de la imagen
  - ◆ Numero bandas multiespectrales y conocer si viene en modo pancromático
  - ◆ Resolución radiométrica de la imagen
  - ◆ Dimensión de la imagen, cantidad de pixeles en filas y columnas
  - ◆ Nivel de licenciamiento de la imagen.

## **6. REFERENTE TEÓRICO**

### **6.1 Caña de azúcar**

Según Ramírez (2008), la caña de azúcar es una gramínea tropical perenne con tallos gruesos y fibrosos que pueden crecer entre 3 y 5 metros de altura, estos contienen una gran cantidad de sacarosa que se procesa para la obtención de azúcar, la caña de azúcar es uno de los cultivos agroindustriales más importantes en las regiones tropicales. Es una materia prima de la cual es posible obtener además del azúcar, otros productos derivados de su procesamiento para los que el azúcar sirve como insumo, tales como la industria de alimentos, chocolates, confites, bebidas, alcoholes, papel, alimentos para animales y la generación de energía. Las ventajas ofrecidas por el valle geográfico del río Cauca, tanto en suelo como en clima hace que este cultivo pueda ser sembrado y cosechado a lo largo de todo el año, ofreciendo beneficios a los ingenios, productores y cultivadores obteniendo así una gran productividad comparada con otros países, lo que permite la generación de ingresos y empleo a la población local, por todos los beneficios que se logran obtener de este cultivo la agroindustria debe plantearse estrategias que potencialicen y optimicen la producción.

### **6.2 Agricultura de precisión**

La presente investigación se encuentra enmarcada dentro del contexto agrícola, el cual representa a nivel mundial un importante renglón en la actividad económica, así lo indica la Organización de las Naciones Unidas para la Alimentación y la Agricultura, quien expresa que “Actualmente el 30% del PIB en el mundo se sustenta en actividades basadas en la agricultura” (Yearbook, FAO Statistical, 2013). Por lo tanto, se debe considerar la importancia de la misma,

teniendo en cuenta que desde el estudio mencionado se tiene un crecimiento poblacional y se requieren nuevas técnicas o estrategias tecnológicas que permitan mejorar la producción y optimizar recursos, buscando la sostenibilidad productiva. Como respuesta a los nuevos desafíos generados por la dinámica mundial entorno al desarrollo agrícola, han surgido instrumentos que buscan dar respuesta a las necesidades de sostenibilidad, optimización y toma de decisiones acertadas para mejorar la producción y el rendimiento de los cultivos, una de los principales sistemas emergentes se denomina agricultura de precisión, la cual remonta su origen a:

La agricultura de precisión surgió en Estados Unidos como un ciclo de realimentación anual donde se obtenían resultados relacionados con aplicación de insumos en función de las necesidades específicas del terreno (...) La AP se define como un conjunto de procedimientos y procesos que buscan optimizar espacial y temporalmente el ciclo de vida de diferentes cultivos a través de tecnologías, elementos y estudios de manera amigable con el medio ambiente (Orozco & Llano, 2016)

Por su parte, Vecchio et al., (2020) definen la agricultura de precisión como “un sistema de gestión agrícola integrado y sostenible que utiliza tecnologías modernas para aumentar la rentabilidad de la granja mediante la reducción del impacto ambiental”, el uso de nuevas tecnologías permite al agricultor y al productor contar con datos de calidad sobre el cultivo que promuevan mejores prácticas y permitan identificar elementos negativos como enfermedades o malezas que presentan afectación en el cultivo. Así mismo, Murillo Sandoval & Carbonell González, (2012) conciben la agricultura de precisión como el conjunto de practicas empleadas para establecer la variabilidad de un mismo lote, usadas para la correcta toma de decisiones sobre

el cultivo, estas prácticas incluyen la recolección, interpretación y manejo de una gran cantidad de datos, los cuales se usan para ajustar y mejorar la producción de los cultivos.

Los elementos propuestos por los autores evidencian el potencial de la agricultura de precisión, que con su tecnificación aplicada a múltiples cultivos permite la obtención gran cantidad de datos, los cuales al ser procesados correctamente brindan información fiable para la correcta toma de decisiones enfocada al rendimiento, producción, rentabilidad y prácticas amigables con el medio ambiente. Este tipo de desarrollo no es ajeno a Colombia, dado la rápida y creciente aplicación de estas técnicas a nivel sudamericano, lo que hace que el país incurriere en las mismas buscando mejorar el desarrollo agrícola y lograr estrategias que beneficien la conservación de los recursos naturales del país. Es por esto que el sector cañero, quien representa una de las principales actividades económicas en Colombia, está incursionando en prácticas esenciales para el desarrollo agrícola del país, aprovechando las múltiples ventajas que puede ofrecer dicho cultivo, por lo que se busca el desarrollo de tecnologías modernas combinadas con los datos propios del terreno, logrando una mayor eficiencia en los cultivos, al punto de lograr el desarrollando de las haciendas en unidades productivas empresariales. (Ministerio de Agricultura y Desarrollo Rural, Bogotá, 2005) .

### **6.3 Percepción remota**

Como resultado a la búsqueda de tecnologías que mejoren el rendimiento, la optimización y producción de caña de azúcar podemos encontrar la aplicación de tecnologías como lo es el uso de percepción remota, esta técnica consiste en la obtención de información espacial, temporal y espectral a través de la medición de la energía que es reflejada o emitida por diferentes cuerpos u objetos sin estar en contacto con ellos, (Murillo Sandoval & Carbonell

González, 2012), la energía que se emite viaja por el espacio en forma de ondas electromagnéticas y se diferencian por su longitud, esta es captada por medio de los sensores, cámaras y espectro radiómetros, transformando esa energía en información. En consecuencia, Villegas, (2008), la define como “La ciencia de adquirir información sobre objetos lejanos sin hacer contacto físico con ellos, restringida a medidas de radiación electromagnética, incluyendo luz visible e infrarroja, al igual que ondas de radio y calor”.

Dentro de los sistemas de percepción remota se destacan dos fuentes de energía, discriminada en sistemas activos y los sistemas pasivos, en el primero se emite un flujo energético el cual rebota en el cuerpo u objeto y permite medir la característica de la señal reflejada, un ejemplo de este sistema es la percepción por radar, la cual emite ondas de radio mediante un transmisor y luego son recogidas por un receptor una vez son reflejadas por el objeto, la ventaja de este sistema es que puede ser usada inclusive en días nublados (Carbonell González, 2011; Von Martini et al., 2022). Por su parte el sistema de percepción pasiva recibe las señales externas a los sensores, estas son emitidas naturalmente y reflejadas por los objetos percibidos, la energía generada por la radiación natural provee información muy valiosa sobre los objetos observados, esta última es la más común en la evaluación de producción agrícola. (Von Martini et al., 2022).

Existen diferentes enfoques para la aplicación de la percepción remota en la agricultura de precisión y una de ellas como lo indica Murillo Sandoval & Carbonell González (2012), es la correlación entre las imágenes y variables específicas como propiedades del suelo o deficiencia de nitrógeno (N). Esta correlación permite emplear métodos como los índices de vegetación o métodos no convencionales, los cuales proporcionan imágenes que se convierten en mapas de

fertilización para la aplicación de dosis, por lo que se puede concluir que los datos obtenidos mediante la percepción remota pueden ser usados como entrada en modelos que permiten determinar el crecimiento y la predicción de producción del cultivo en grandes áreas. La información que se puede obtener por medio de las tecnologías de la información derivadas de las imágenes capturadas por los diferentes satélites y los índices espectrales que se derivan del postproceso de las imágenes son un insumo enriquecedor y de vital importancia en la estimación de la producción del cultivo de caña de azúcar, dado que nos brindan una gran cantidad de datos que al ser combinados con variables biofísicas del cultivo permitirán generar un modelo predictivo en la búsqueda de la identificación temprana de la productividad en una plantación de caña de azúcar.

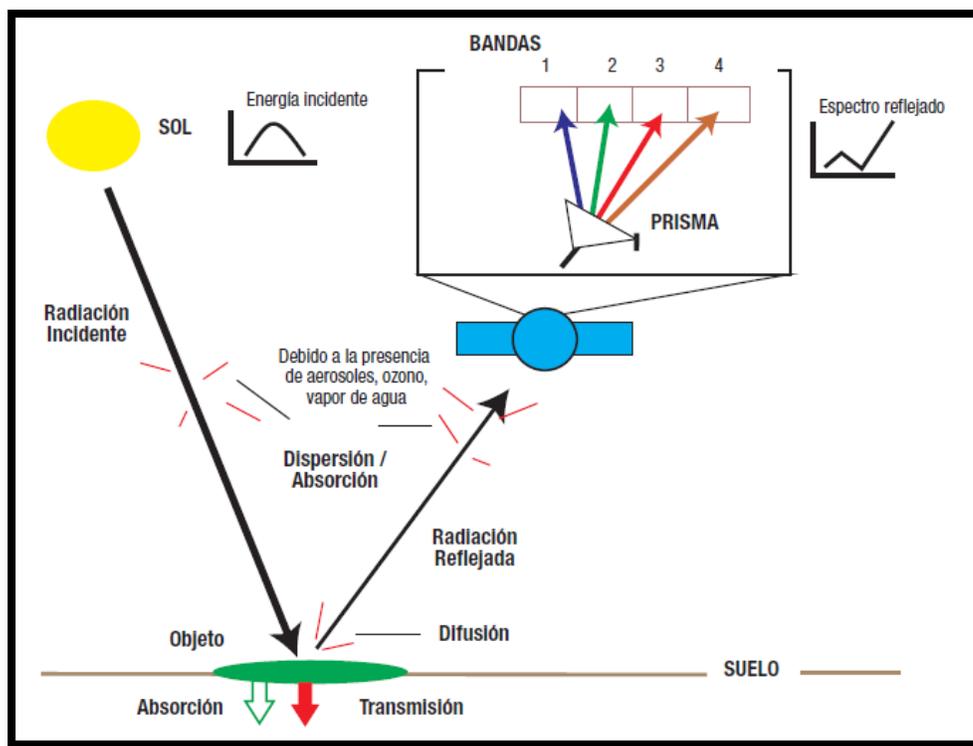


Ilustración 3 Proceso de la percepción remota.

Tomado de (Murillo Sandoval & Carbonell González, 2012)

#### 6.4 Radiación electromagnética

Para que sea posible la obtención de información remota, es fundamental la interacción entre los objetos y el sensor, a lo que se denomina radiación electromagnética, definida por Murillo Sandoval & Carbonell González (2012) como “un modo de propagación de la energía a través del vacío en forma de interacción que avanza entre campos eléctricos y magnéticos (...) se refiere a la energía transportada por ondas electromagnéticas y está compuesta por partículas energizadas llamadas cuantos”. Acorde a la definición anterior, Villegas (2008) plantea que “la energía electromagnética se transmite de un lugar a otro siguiendo un modelo armónico y continuo, a la velocidad de la luz, y conteniendo dos campos de fuerzas ortogonales entre sí: el eléctrico y el magnético. A su vez, Murillo Sandoval & Carbonell González (2012) sostienen que “el grado de energía y frecuencia de los cuantos determina la longitud de onda y el color de la radiación. Las ondas electromagnéticas se diferencian por su respectiva longitud, que es la distancia entre dos crestas consecutivas”.

Suelen establecerse una serie de bandas en donde la radiación electromagnética manifiesta un comportamiento similar, la organización de estas bandas o rangos entre valores específicos de longitudes de onda o frecuencias, se denomina espectro electromagnético” (Villegas, 2008). Las longitudes de onda están reunidas en el espectro electromagnético, y este se subdivide en grupos familiares: rayos X, rayos ultravioletas (UV), visibles, infrarrojos (IR), microondas y ondas de radio” (Murillo Sandoval & Carbonell González, 2012). Murillo y Carbonell (2012) sostienen que las franjas del visible e infrarrojo tienen mayor aplicación agrícola puesto que, la luz del sol es la fuente de energía que activa el proceso de fotosíntesis en los cuerpos vegetales, por lo que son las más utilizadas en la percepción remota. La radiación

que incide en los objetos de la superficie terrestre se convierte en la energía reflejada, absorbida y transmitida, esta presenta un comportamiento particular acorde al objeto y a los materiales que conforman el mismo, así mismo, plantean que “la huella espectral que captan los diversos sensores utilizados en teledetección o percepción remota refleja el comportamiento de los diferentes objetos analizados en función de su mayor o menor capacidad de absorción, transmisión o reflexión de la energía que reciben”.

### **6.5 Sistema sensor**

La aplicación y el uso de percepción remota viene ligado estrechamente con la capacidad para discriminar y registrar la información detallada, a lo cual se conoce como sistema sensor, Murillo Sandoval & Carbonell González, (2012) expresan que “Se denomina sistema sensor porque la resolución se refiere al conjunto del equipo y no a cada una de sus partes”. Por su parte Villegas, (2008) los clasifican “Tomando como criterio la fuente de energía, pueden dividirse en pasivos y activos. Los primeros detectan la radiación solar reflejada o emitida por los objetos de la superficie (...) y los activos son aquellos que producen su propia radiación”. Los sistemas sensores aéreos o satelitales pueden ser empleados en plataformas tales como: satélites, aeronaves, globos, cometas, dirigibles y aeromodelos, para el caso puntual de esta investigación, nos centraremos en la aplicación de los satélites, definidos por Murillo Sandoval & Carbonell González, (2012) los definen como vehículos colocados en órbita alrededor de la tierra, los cuales tienen como ventaja la capacidad de recolectar información regular de la superficie terrestre, entre sus características se encuentra que pasan por determinada área a la misma hora local, su órbita es circular y tienen una altitud entre 700 km y los 900 km.

Los sensores aplicados al cultivo de caña de azúcar varían según sea la necesidad del proyecto que se quiera aplicar, puesto que así mismo se requerirá una resolución espacial, temporal y espectral para cada caso de estudio, Murillo Sandoval & Carbonell González (2012) consideran que “En el caso de la caña de azúcar, y en general en la agricultura, el uso de la PR satelital se ha hecho con los programas más antiguos como Landsat, NOAA y SPOT 4 y 5”. La información que se puede obtener por medio de las tecnologías de la información derivadas de las imágenes capturadas por los diferentes satélites y los índices espectrales que se derivan del postproceso de las imágenes son un insumo enriquecedor y de vital importancia en la estimación de la producción del cultivo de caña de azúcar, dado que nos brindan una gran cantidad de datos que al ser combinados con variables climáticas del cultivo permitirán generar un modelo predictivo en la búsqueda de la identificación temprana de la productividad en una plantación de caña de azúcar.

## **6.6 Minería de datos**

La necesidad por obtener información útil de un conjunto de datos ha llevado a hablar de la minería de datos, la cual se refiere a todo el proceso de extracción de conocimiento a partir de una base de datos y marca un cambio de paradigma en el que lo importante es el conocimiento útil que seamos capaces de descubrir a partir de los datos (Riquelme Santos et al., 2006), así mismo autores como Pérez López & Santin González (2007), la definen como un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencia al examinar grandes cantidades de datos. Esta técnica que conforma las diferentes metodologías que permiten extraer la información procesable, con el objetivo de buscar relación entre las variables para lograr predecir tendencias y patrones de comportamiento, pudiendo aplicarse sobre cualquier evento



desconocido, ya sea en el pasado, presente o futuro, son hoy por hoy muy utilizadas en la agricultura de precisión y específicamente en caña de azúcar.

## 7. HIPÓTESIS DE INVESTIGACIÓN

H<sub>1</sub>: Un modelo predictivo de procesos Gaussianos, desarrollado mediante el uso multitemporal de imágenes satelitales Sentinel 1 y datos climáticos, permitirá la estimación de la productividad del cultivo de caña de azúcar en edades tempranas, en la búsqueda de información confiable y útil para la toma de decisiones sobre el manejo del cultivo.

H<sub>0</sub>: Un modelo predictivo de procesos Gaussianos, desarrollado mediante el uso multitemporal de imágenes satelitales Sentinel 1 y datos climáticos, no lograra la estimación de la productividad del cultivo de caña de azúcar en edades tempranas, en la búsqueda información confiable y útil para la toma de decisiones sobre el manejo del cultivo.

## 8. METODOLOGIA

### 8.1 Enfoque metodológico

El presente trabajo se enmarca dentro del enfoque cualitativo, puesto que se busca responder a una hipótesis a través de la valoración numérica y el análisis estadístico inferencial. Se desarrolló un modelo de estimación utilizando el método de procesos gaussianos, como variable dependiente: el rendimiento de caña medido en toneladas por hectárea y como variables independientes: la retrodispersión en las polaridades VV y VH del sensor Sentinel 1 y variables climáticas de temperatura, brillo solar y precipitación correspondientes a la estación con área de influencia para cada hacienda. Como método de validación del modelo de estimación se utilizó la metodología de validación cruzada Leave-one-out.

### 8.2 Tipo de estudio

Este trabajo corresponde a un estudio correlacional, dado que se busca la correlación de la información satelital y los datos climáticos con la productividad del cultivo de caña de azúcar, mediante un modelo de procesos Gaussianos.

### 8.3 Diseño de la investigación

Conforme a la definición de (Hernández Sampieri et al., 2010), el diseño de este trabajo es fundamentalmente no experimental – transversal, puesto que con datos ya existentes se harán los procedimientos para relacionarlos y plantear una estrategia metodológica.

#### **8.4 Población o unidad de trabajo**

La población corresponde a las haciendas: La Avelina, Holanda Naranjo, García Abajo Norte y Sevilla, pertenecientes al ingenio del Cauca, localizadas en los municipios de Corinto, Miranda y Padilla, al norte del departamento del Cauca.

#### **8.5 Muestra y muestreo**

Los datos considerados para el desarrollo del modelo corresponden a la respuesta espectral del Radar de Apertura Sintética (SAR) de banda C de la misión Sentinel 1 y los valores diarios de precipitación, radiación solar y temperatura media, para el área de cada hacienda. Para el desarrollo y validación del modelo de estimación de la variable de producción: Toneladas de caña por hectárea se utilizaron todos los píxeles en cada parcela y se validaron usando el método de validación cruzada dejando uno por fuera (LOOCV por sus siglas en inglés de Leave-one-out cross-validation).

#### **8.6 Operacionalización de variables**

Datos satelitales de retrodispersión SAR y las variables climáticas fueron las variables objeto de estudio, obtenidas mediante técnicas diferentes, los datos satelitales se encuentran en formato raster con una frecuencia de 12 días, mientras que los parámetros climáticos son datos diarios que se han tratado como datos raster según el área de influencia de los polígonos de Thiessen a la misma escala espacial que los datos satelitales. Se construyeron series temporales desde la fecha de siembra hasta la fecha de corte para cada set de datos y se extrajeron de cada serie la media, la tendencia (pendiente), el área bajo la curva y la desviación estándar. El

segundo grupo de variables son los valores diarios de las estaciones climáticas, donde se analizó la serie temporal desde la fecha de siembra y corte y se extrajeron los valores estadísticos diarios entregados por estaciones climáticas en campo, extrayendo la media, la tendencia (pendiente), el área bajo la curva, la desviación estándar y la sequía. Por último se contó con los valores de productividad de cada finca siendo la variable a estimar dentro del modelo predictivo.

### **8.7 Técnicas e instrumentos de recolección de información**

La recolección de la información se realizó de dos formas, los datos satelitales fueron descargados de la plataforma de Sentinel 1, mientras que la información climática y los datos de productividad fueron entregados por el Ingenio del Cauca. La información obtenida para el desarrollo de esta investigación se realizó de manera multitemporal.

## **8.8 Desarrollo metodológico**

### **8.8.1 Imágenes satelitales**

Los datos utilizados fueron de la misión Sentinel 1, la cual comprende una constelación de dos satélites en órbita polar y cuenta con un sensor de Radar de Apertura Sintética (SAR) de banda C, este radar opera en longitudes de onda que no se ven obstaculizadas por la nubosidad o la falta de iluminación y puede adquirir datos sobre un sitio durante el día o la noche en todas las condiciones climáticas (Agency T. E., 2022). Las imágenes Sentinel 1 son Ground Range Detected (GRD) de alta resolución y poseen una resolución temporal de hasta 12 días, estas fueron descargadas en el periodo del 15 de julio del 2017 al 30 de septiembre del 2021 obteniendo un total de 339 mosaicos.

### *8.8.2 Preprocesado de imágenes*

Una vez se obtuvieron las imágenes Sentinel 1, se generó un modelo de preprocesado en la aplicación SNAP, la cual es uno de los recursos gratuitos que se pueden emplear para el procesamiento y análisis de imágenes satelitales junto a las herramientas Toolboxes propias de la flota Sentinel, la ilustración 4 muestra un esquema del desarrollado por Filipponi (2019) el cual este modelo permite el procesamiento de grandes lotes de datos y consta de siete pasos de procesamiento, estos son descritos a continuación.

- Aplicación de Archivos de Órbita, este proceso permite tener una actualización automática sobre los datos de vectores de orbita, los cuales se encuentran disponibles semanas después de la toma de la imagen, lo cual ayuda a corregir los datos de posición y velocidad del satélite.
- Eliminación del Ruido Térmico, permite la reducción de ruido térmico, el cual está asociado a la cantidad de electrones que se generan por el efecto térmico cuando el sensor o sus elementos se exponen a altas temperaturas, logrando la normalización de la señal de retrodispersión dentro de toda la escena Sentinel-1
- Eliminación de Ruido de Borde, elimina el ruido generado entre las franjas de las imágenes sobre los bordes de las escenas, proceso que sucede con el ajuste de la curvatura terrestre y la dirección del del azimut.

- Calibración, permite convertir los valores de los píxeles digitales en datos SAR calibrando radiométricamente la retrodispersión, permite la conversión simple de valores de intensidad de imagen en valores sigma cero.
- Filtrado de Manchas, Mejora la calidad de las escenas limpiando el moteado granular generado por la interferencia en de las ondas reflejada por dispersores elementales.
- Corrección del terreno de rango Doppler, permite la corrección de las distorsiones geométricas causadas por la topografía, ayuda a corregir la ubicación de cada pixel sobre imagen, respecto a la posición de su zona territorial, ortorrectificando las escenas
- Conversión a dB, en el último paso el coeficiente de retrodispersión se convierte en decibelios (dB) utilizando una transformación logarítmica.

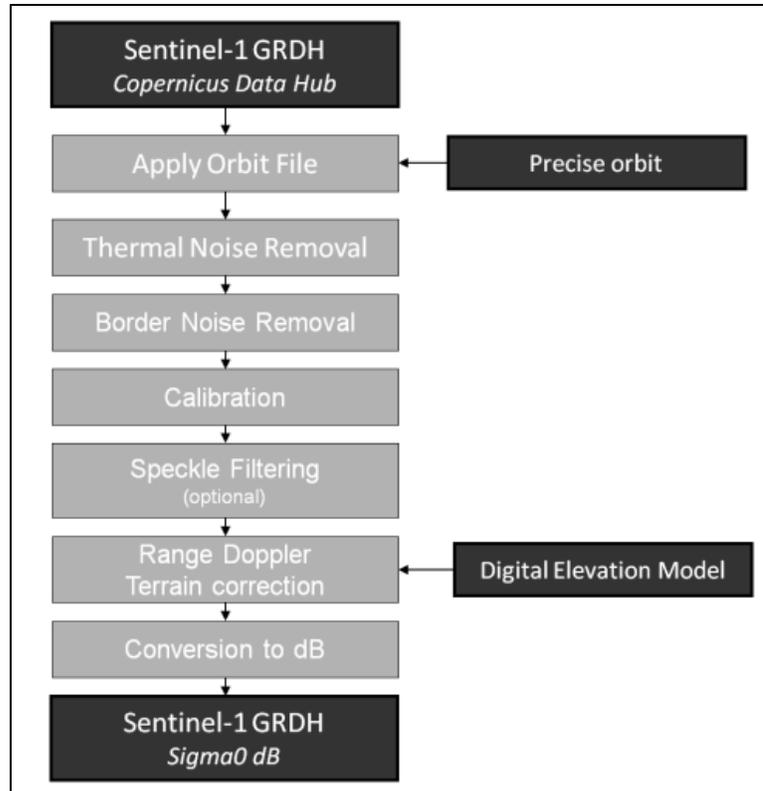


Ilustración 4 Modelo de preprocesado Sentinel 1

Fuente: (Filipponi, 2019)

### 8.8.3 Datos climáticos

Los datos climáticos se obtuvieron del Metaportal de Cenicaña, para las estaciones de Miranda y Ortigal, las cuales entregan información por área de influencia, esta área ha sido generada por el método de polígonos de Thiessen, los datos comprenden valores diarios de precipitación, radiación solar y temperatura media. Las seis haciendas se ubican en la zona agroecológica 11H3, la cual se caracteriza por suelos de texturas franca fina y franca fina sobre arcillosa de orden Mollisol, distribuidos en el cuerpo y pie de abanicos aluviales, de régimen

údicó, con precipitaciones entre 200-600 mm/año, temperatura media de 23.7 °C y altura promedio 1005 m.s.n.m. (Carbonell González, 2011).

#### **8.8.4 Datos de productividad del Cultivo.**

Los datos de productividad del cultivo de caña son medidos en toneladas de caña por hectárea – TCH, esta información fue suministrada por el ingenio del Cauca, estos datos fueron utilizados como la variable a estimar dentro del modelo, la unidad espacial de los mismo corresponde a la suerte de cada hacienda y se adquirieron entre los años 2017 a 2021 para los periodos de siembra y cosecha o corte y cosecha.

HACIENDA	SUERTE	TCH			
		2017/2018	2018/2019	2019/2020	2020/2021
LA AVELINA	002	101.65	86.22	103.68	119.75
SEVILLA	003A	149.25	131.98	122.30	144.09
HOLANDA - NARANJO	006	138.98	123.26	118.80	126.79
GARCIA ABAJO NORTE	009B	98.62	134.56	103.34	96.08
HOLANDA - NARANJO	015A	138.35	112.02	114.09	107.79
GARCIA ABAJO NORTE	008A	110.05	123.08	107.47	94.93

Tabla 1 Productividad histórica a nivel de suerte

#### **8.8.5 Creación de la base de datos**

La base de datos se conformó por ocho variables (figura 2), las variables climáticas: radiación solar, temperatura y precipitación medidas a escala diaria y en los datos satelitales se analizaron las polaridades VV y VH y los índices espectrales VV/VH, VH/VV, por último, se

calculó la diferencia normalizada  $\frac{VH-VV}{VH+VV}$ .

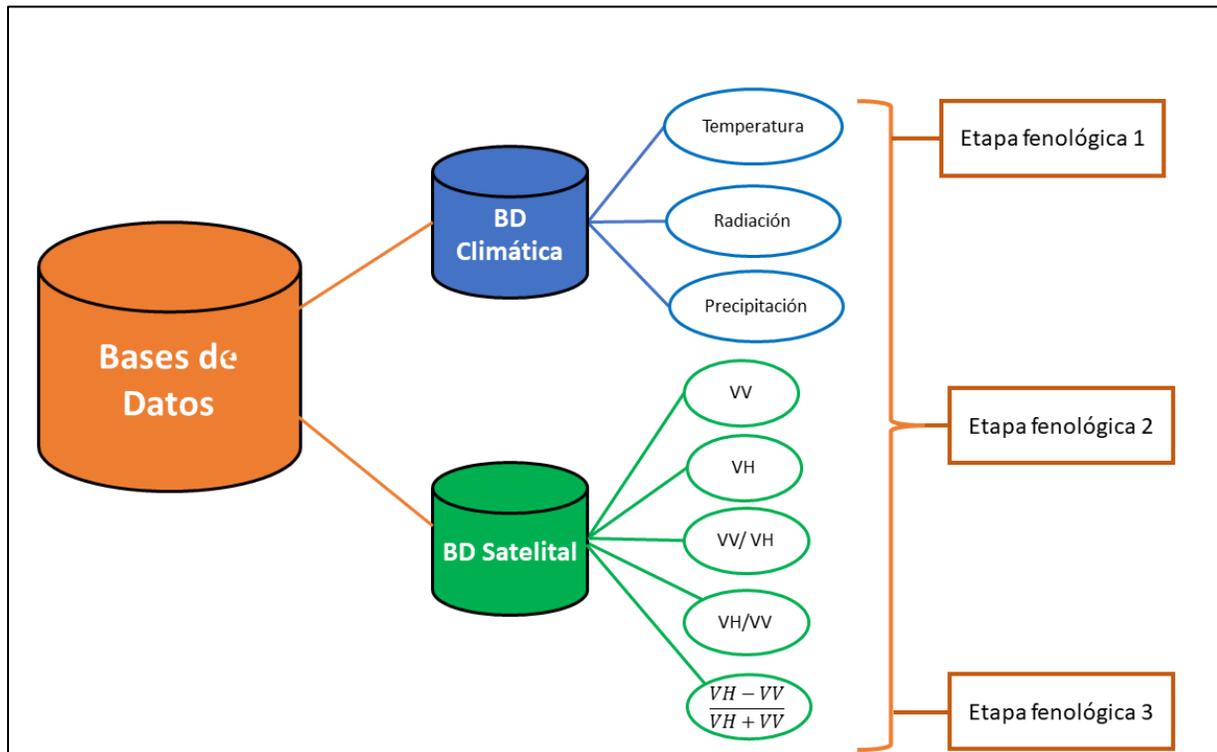


Ilustración 5 Bases de datos construidas para el desarrollo de la investigación

## 8.8.6 Extracción de características

### 8.8.6.1 Etapas de análisis

Una vez se construyeron las base de datos se dividió la información acorde a las etapas fenológicas del cultivo durante los periodos de siembra y cosecha entre los años 2017 a 2021, el proceso vegetativo de la caña de azúcar oscila entre los 12 -14 y meses, contemplando cuatro etapas: germinación (0 – 35 días), macollamiento (35 días – 120 días), rápido crecimiento (120 días – 180 días) y maduración (180 días – cosecha), (CONADESUCA, 2015), sin embargo para este estudio se contemplaron solo tres etapas, la primera de ellas agrupa las etapas de germinación y macollamiento, recogiendo los datos desde el día 0 al día 120, la segunda etapa

se consideró desde el día 121 al día 180 y la última etapa de análisis se estableció entre el día 181 al día de cosecha.

#### **8.8.6.2 Extracción de características**

Con la clasificación de la información por fenología para cada hacienda y por serie temporal, se procedió a evaluar y seleccionar las mejores características para alimentar el modelo de predicción, para lo cual se emplearon métricas estadísticas como la media, la desviación estándar, la pendiente y el área bajo la curva por el método del trapecio. Estas características fueron extraídas para las series temporales de las bases de datos climática y satelital como se muestra en la ilustración 7. Otra de las características extraídas fue la “sequía” definida como la cantidad de días sin lluvia durante cada serie temporal, aquí fue asignado el valor acumulativo de días sin lluvia para el periodo de análisis. Para la serie temporal de sequía se calcularon las estadísticas sesgo y curtosis, las cuales fueron incorporadas al modelo predictivo, tal como se observa en la ilustración 6.

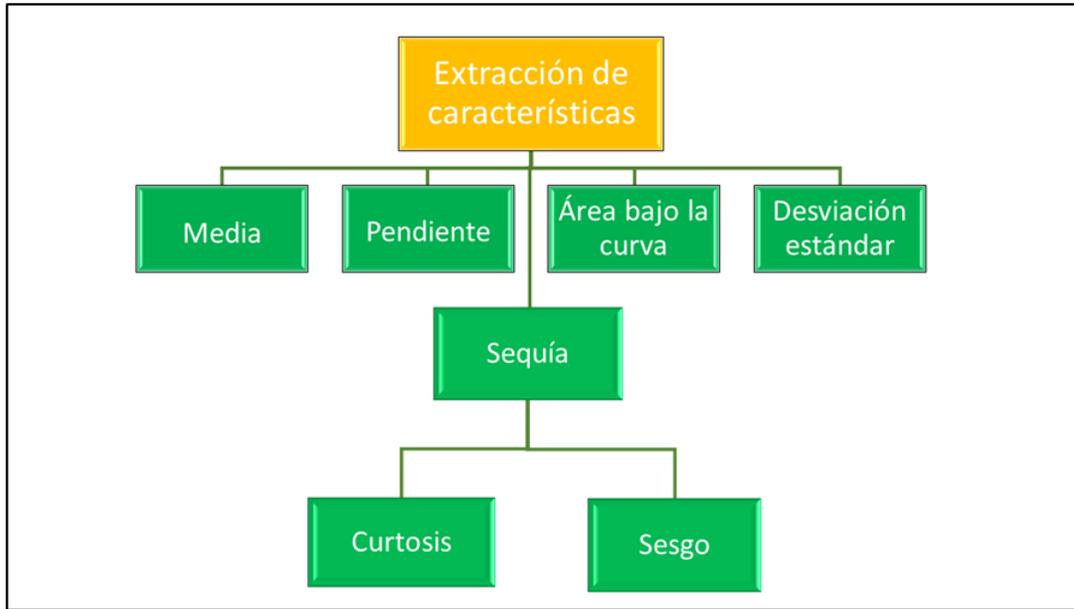


Ilustración 6 Características extraídas de las series temporales

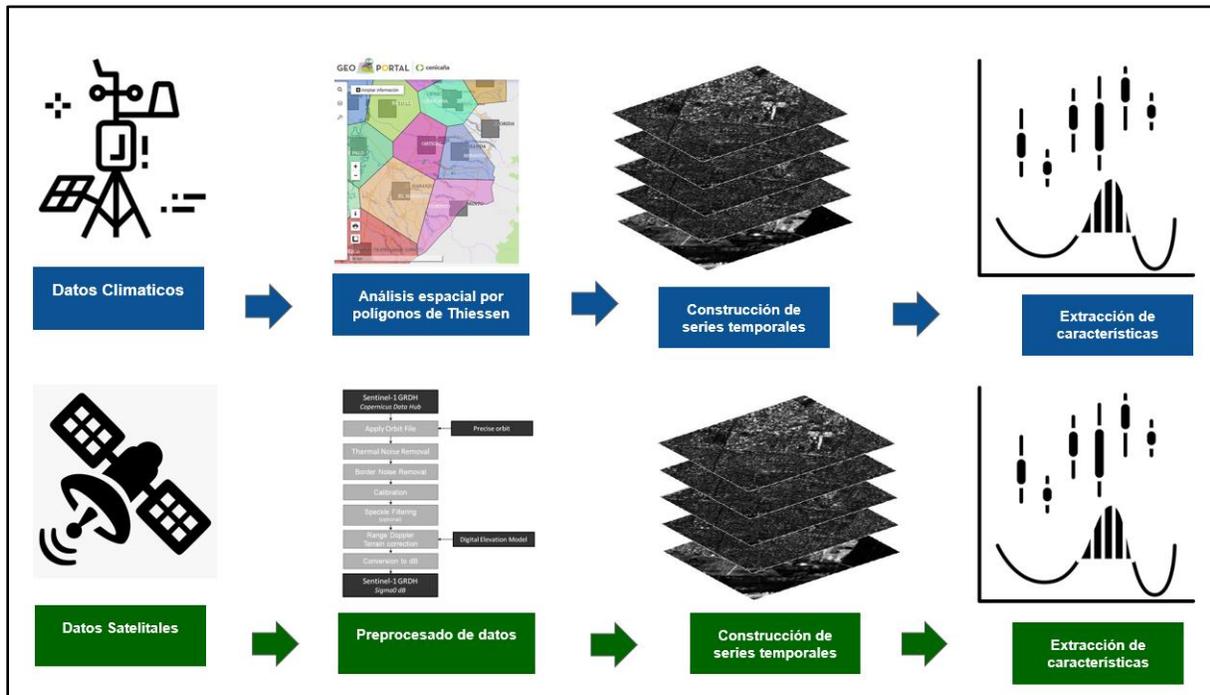


Ilustración 7 Proceso de extracción de características

### 8.8.6.3 Algoritmo de Procesos Gaussianos (GPR)

Es uno de los algoritmos de aprendizaje automático más potente y robusto de la actualidad, dado que, a diferencia de otros modelos, GPR se basa en pocos parámetros para hacer predicciones. Sander (2021), argumenta que debido a que GPR es (casi) no paramétrico, se puede aplicar de manera efectiva para resolver una amplia variedad de problemas de aprendizaje supervisado, incluso cuando hay pocos datos disponibles. GPR requiere un conjunto de datos de entrenamiento relativamente pequeño y puede adoptar núcleos muy flexibles, identifica las bandas y observaciones relevantes al establecer relaciones con una variable, y finalmente junto con estimaciones de píxeles, GPR proporciona intervalos de confianza complementarios. (Verrelst et al., 2013).

GPR proviene de su naturaleza Bayesiana, por lo que no solo proporciona predicciones ante un espectro dado, sino una distribución predictiva que permite calcular los primeros momentos como la media (predicción media) y la desviación estándar (intervalo de confianza de las predicciones). (Campos-Taberner, 2013).

El modelo GPR establece una relación entre la entrada (B-espectros de bandas)

$\mathbf{X} \in \mathbb{R}^B$  y la variable de salida (TCH)  $\mathbf{y} \in \mathbb{R}$  de la forma:

$$\hat{\mathbf{y}} = \mathbf{f}(\mathbf{X}) = \sum_{i=1}^N \alpha_i \mathbf{K}(\mathbf{X}_i, \mathbf{X}) + \alpha_0$$

Donde  $\{\mathbf{X}_i\}_{i=1}^N$  son los espectros en la base de datos usados en el entrenamiento,  $\alpha_i \in \mathbb{R}$  los pesos de cada espectro en la base de datos de entrenamiento,  $\alpha_0$  es la ordenada en el origen de la regresión, y  $\mathbf{K}$  es la función que evalúa la similitud entre los espectros test y los de entrenamiento.

La clave a la hora de realizar con éxito el GPR, es la selección del Kernel adecuado. En este trabajo se propone el uso de un Kernel RBF (Radial Basis Function) del tipo:

$$K(X_i, X_j) = v \exp \left( - \sum_{b=1}^B \frac{(X_i^b - X_j^b)^2}{2\sigma_b^2} \right) + \sigma_n^2 \delta_{XX}$$

Donde  $v$  es el factor de escala,  $B$  es el número de bandas y el superíndice  $b$  en  $X_i$  indica la banda  $b$  del espectro  $X_i$ ,  $\sigma_b$  es un factor que controla la propagación de cada banda, y  $\sigma_n$  es la desviación estándar.

#### 8.8.6.4 Validación del modelo

El método utilizado para la validación del modelo fue la validación cruzada dejando uno por fuera o por sus siglas en inglés LOOCV - Leave-One-Out cross-validation, implica dividir aleatoriamente el conjunto de datos entre datos de entrenamiento y datos de prueba, dejando solo una interacción para prueba y el resto serán los datos de entrenamiento, entrenando tantos modelos existentes como número de muestras existan (Sammut & Webb, 2011).

#### 8.8.6.5 Error cuadrático medio MSE

La métrica que se utilizó para la evaluación del desempeño del modelo fue el MSE, la cual es utilizada en la evaluación del desempeño en modelos de regresión dado que penaliza los grandes errores, es decir, busca un modelo que no se desvíe demasiado del comportamiento de la variable objetivo (Cabrera, 2022)

$$MSE = \frac{1}{n} \left( \sum_{i=1}^n Y_i - \widehat{Y}_t \right)$$

### 8.8.7 Selección de variables

Una vez fueron evaluadas todas las posibles combinaciones de las características extraídas de las series de temperatura, precipitación, radiación solar y los datos SAR como se muestran en la tabla 1, se hallan los mejores resultados de manera individual y luego por combinación entre variables, seleccionando la combinación que obtuvo el menor error medio cuadrático.

EXTRACCIÓN DE CARACTERÍSTICAS								
VARIABLES	Precipitación	Temperatura	Radiación	VV	VH	VV/VH	VH/VV	$\frac{VH - VV}{VH + VV}$
CARACTERÍSTICAS	Área bajo la curva							
	Curtosis	Desviación estándar						
	Desviación estándar	Media						
	Media	Pendiente						
	Pendiente							
	Sesgo							

Tabla 2 Combinaciones de las bases de datos

## 9. DISCUSIÓN Y RESULTADOS

Una vez se obtuvieron los mejores resultados de los modelos para las bases de datos evaluadas por cada fenología, se encontró que el valor con mayor potencial de predicción corresponde a los datos SAR, evaluando la polarización VH , utilizando la combinación de variables (Desviación estándar / Pendiente) en la fenología 1 obteniendo un RMSE de 11.11 (ver tabla 4). En cuanto a los datos climáticos se halló que el mejor resultado lo entrega la base de datos de temperatura en la etapa fenológica 4, utilizando tres variables para la evaluación (media, desviación estándar y Pendiente) alcanzando un RMSE de 11.82 (ver tabla 5).

<b>RESULTADOS DEL MODELO DE PREDICCIÓN PARA LOS DATOS SAR</b>					
<b>Fenología</b>	<b>Base de Datos</b>	<b>BD – Polaridad</b>	<b>VARIABLES UTILIZADAS</b>	<b>Numero de variables</b>	<b>RMSE</b>
fen1	SAR	SAR_VH	Desviación Estándar, Pendiente	2	11.11
fen2	SAR	SAR_VV	Área bajo la curva, Pendiente	2	12.36
fen3	SAR	SAR_VH	Área bajo la curva	1	14.64
fen4	SAR	SAR_VV	Media, Pendiente	2	13.66

Tabla 3 Resultados del modelo al usar las bases de datos SAR

RESULTADOS DEL MODELO DE PREDICCIÓN PARA LAS VARIABLES CLIMÁTICAS				
Fenología	Base de Datos	Variables utilizadas	Numero de variables	RMSE
fen1	Radiación	media	1	14.99
fen1	Precipitación	Sesgo, promedio, desviación estándar	3	16.80
fen1	Temperatura	Pendiente	1	17.20
fen2	Temperatura	Media, desviación estándar, Pendiente	3	14.82
fen2	Radiación	Media, desviación estándar, Pendiente	3	15.96
fen2	Precipitación	Curtosis, sesgo, promedio, desviación estándar	4	16.50
fen3	Radiación	Área bajo la curva, media, desviación estándar, pendiente	4	16.33
fen3	Precipitación	Curtosis, sesgo, promedio, desviación estándar	4	16.88
fen3	Temperatura	Pendiente	1	17.20
fen4	Temperatura	Media, desviación estándar, Pendiente	3	11.82
fen4	Precipitación	Promedio, desviación estándar	2	14.35
fen4	Radiación	Media, desviación estándar, Pendiente	3	16.81

Tabla 4 Resultados del modelo al usar las bases de datos climáticos

Los resultados muestran un modelo con potencial para la estimación de la producción de caña de azúcar en edades tempranas y como el uso de la percepción remota en torno a series temporales posibilita una estimación de producción mucho más confiable, si bien se sabe que la capacidad de la caña de azúcar para producir sacarosa está directamente relacionada con la variedad, manejo del cultivo y factores climáticos como la precipitación, luminosidad y temperatura (Larrahondo & Villegas, 1995), la investigación logro demostrar que el uso de imágenes SAR de manera multitemporal contribuye para mejorar los modelos predictivos, en este caso el modelo con polaridad VH evaluando la combinación de variables (Desviación estándar y Pendiente) en la fenología 1 del cultivo, hallo un error de 11.11 TCH, entre el valor medido en campo y el valor de TCH estimado.

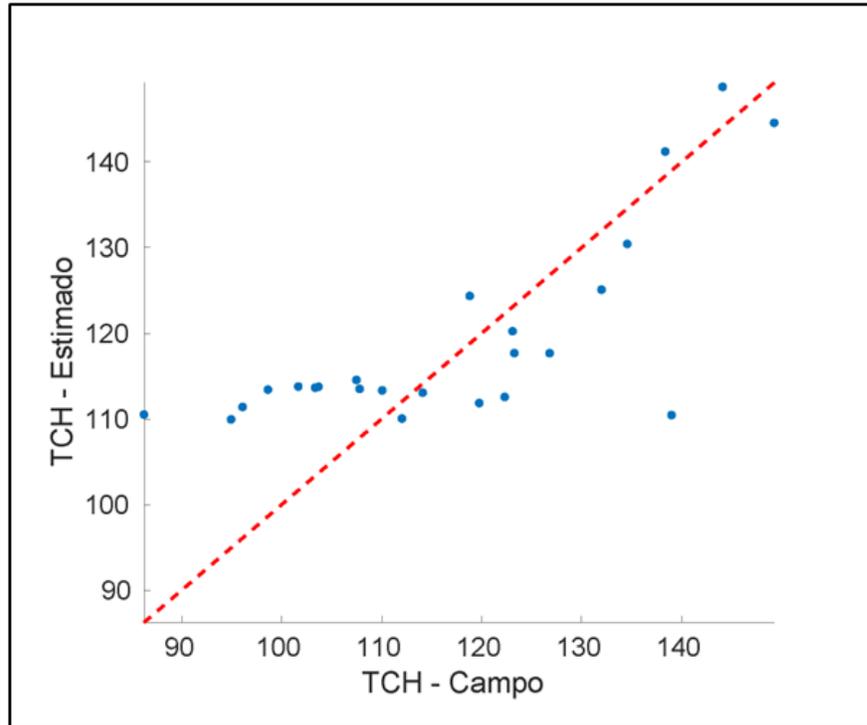


Ilustración 8 Modelo de TCH estimado vs TCH medido en campo

Los modelos desarrollados en esta investigación fueron generados de manera independiente, tanto para los datos SAR como para las variables climáticas, destacándose la polarización VH como el mejor modelo, logro que va en concordancia con algunos autores quienes hallaron un mejor resultado usando los canales de polarización cruzada (VH y HV) para la detección de la cosecha de caña de azúcar (Baghdadi et al., 2009), así como en la caracterización de los estados fenológicos de la caña de azúcar, donde se encontró que la polarización cruzada es mejor que las copolarizaciones (Baghdadi et al., 2010). Estos hallazgos tal vez estén relacionados con lo encontrado por Paloscia, (1998) y Picoli et al., (2013) quienes argumentan que los datos polarizados están más estrechamente relacionados con el índice de área foliar o la biomasa.

Dada la importancia del cultivo de la caña de azúcar para la región, es muy significativo predecir la producción en edades tempranas, lo que permitiría administrar, controlar y regular los diferentes recursos para lograr mantener la producción promedio, anticipar estrategias de ventas, regular mercados internos y externos y gestionar operaciones de campo a tiempo, puesto que el cultivo depende en gran medida de la detección de enfermedades, arvenses, estrés y la determinación de diversas variables biofísicas, las cuales al ser detectadas a tiempo pueden ser corregidas para no afectar la producción final del cultivo.

## 10. CONCLUSIONES

Se desarrolló un modelo de estimación del rendimiento del cultivo en caña de azúcar usando el algoritmo de procesos gaussianos. Los parámetros de entrada para poder estimar toneladas de caña por hectárea son las características de tendencia (pendiente) y desviación estándar extraídas a partir de series temporales de los datos de la banda de polaridad VH del sensor Sentinel 1 en la etapa fenológica que comprende los primeros 120 días después de siembra.

La caña de azúcar es un cultivo con gran biomasa área, por lo anterior, la polarización VH genera un modelo con buen desempeño y de potencial aplicación para cubrir las grandes áreas sembradas de caña de azúcar en el valle del río Cauca y tener un diagnóstico temprano del comportamiento de la producción que permita gestionar oportunamente las labores agrícolas.

Este estudio proporciona uno de los primeros acercamientos a la investigación de los datos SAR para la construcción de modelos predictivos en caña de azúcar, si bien es conocido la dificultad de la adquisición de escenas satelitales libres de nubes en las zonas tropicales, el uso de los datos SAR ofrece una alternativa con alto potencial para el monitoreo y estudio de la productividad, en busca de información que permita tomar decisiones tempranas sobre el manejo del cultivo.

Para próximas investigaciones se recomienda realizar un estudio experimental donde sea posible establecer datos de productividad a nivel de parcela, dado que para este estudio solo se obtuvieron datos a nivel de suertes por hacienda, por lo que dividió la productividad total entre cada pixel, se considera que entre mayor sea la densidad de datos para adicionar al modelo se logre un resultado mucho mejor.

## BIBLIOGRAFÍA

Agency, E. S. (06 de 08 de 2022). *European Space Agency*.

<https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1/overview>

Agency, E. S. (5 de 8 de 2022). *European Space Agency*.

<https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1/overview>

Agency, T. E. (29 de 01 de 2022). *The European Space Agency*. The European Space Agency:

[https://www.esa.int/Applications/Observing\\_the\\_Earth/Copernicus/Sentinel-1/Introducing\\_Sentinel-1](https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1/Introducing_Sentinel-1)

Alemán Montes, B. (28 de 01 de 2022). *Sistema de monitoreo espacio-temporal del cultivo de caña de azúcar (Saccharum officinarum), a partir de información satelital, en Coopevictoria RL Grecia, Costa Rica*. Repositorio Kérwá:

<http://repositorio.ucr.ac.cr/handle/10669/82808>

Antony, B. (2021). Prediction of the production of crops with respect to rainfall. *Environmental Research*, 202(111624). <https://doi.org/https://doi.org/10.1016/j.envres.2021.111624>

Asocaña. (29 de 11 de 2022). *asocana.org*. [asocana.org](http://asocana.org):

<https://www.asocana.org/publico/info.aspx?Cid=215#:~:text=La%20agroindustria%20de%20la%20ca%C3%B1a,mil%20empleos%20directos%20e%20indirectos.>

Asocaña. (29 de 01 de 2022). *Asocaña*. Asocaña:

<https://www.asocana.org/documentos/1782021-3772D9B2-00FF00,000A000,878787,C3C3C3,FF00FF,2D2D2D,A3C4B5.pdf>

Baghdadi, N., Boyer, N., Todoroff, P., & El Hajj, M. B. (2009). Potential of SAR sensors

TerraSAR-X, ASAR/ENVISAT and PALSAR/ALOS for monitoring sugarcane crops on

- Reunion Island. *Remote sensing of environment*, 113(8), 1724-1738.  
<https://doi.org/https://doi.org/10.1016/j.rse.2009.04.005>
- Baghdadi, N., Cresson, R., Todoroff, P., & Moinet, S. (2010). Multitemporal observations of sugarcane by TerraSAR-X images. *Sensors*, 10(10), 8899-8919.  
<https://doi.org/10.3390/s101008899>
- Bégué, A., Lebourgeois, V., Bappel, E., Todoroff, P., Pellegrino, A., Baillarin, F., & Siegmund, B. (2010). Spatio-temporal variability of sugarcane fields and recommendations for yield forecast using NDVI. *International Journal of Remote Sensing*, 31(20).  
<https://doi.org/https://doi.org/10.1080/01431160903349057>
- Begue, A., Todoroff, P., & Pater, J. (2008). Multi-time scale analysis of sugarcane within-field variability: improved crop diagnosis using satellite time series. *Precision Agriculture*, 9(3), 161-171. <https://doi.org/10.1007/s11119-008-9063-3>
- Bocca, F. F., Rodrigues, L. H., & Arraes, N. A. (2015). When do I want to know and why? Different demands on sugarcane yield predictions. *Agricultural Systems*.
- Cabrera, D. S. (2022). *Técnicas de aprendizaje supervisado en la predicción de producción de un cultivo caña de azúcar en el Valle del Cauca, Colombia. Caso de aplicación*. Bogotá.
- Calier, F. R., Mallungo, L. A., Vargas, W. L., & Báez, S. A. (2015). Aplicación de una metodología de análisis de datos obtenidos por percepción remota orientados a la estimación de la productividad de caña para panela al cuantificar el NDVI (índice de vegetación de diferencia normalizada). *Ciencia y tecnología agropecuaria*, 16(1), 25 - 40.
- Campos-Taberner, M. (2013). *Evaluación de procesos Gaussianos en la estimación de parámetros biofísicos*.

- Carbonell González, J. (2011). *Zonificación agroecológica para el cultivo de la caña de azúcar en el valle del río Cauca (cuarta aproximación): principios metodológicos y aplicaciones*. Cenicaña.
- Cevallos, J. C., Villagomez, J. A., & Andryshchenko, I. S. (2019). Convolutional neural network in the recognition of spatial images of sugarcane crops in the Troncal region of the coast of Ecuador. *Procedia Computer Science*, 150, 757-763.  
<https://doi.org/https://doi.org/10.1016/j.procs.2019.02.001>
- Chea, C., Saengprachatanarug, K., Posom, J., Wongphati, M., & Taira, E. (2020). Sugar yield parameters and fiber prediction in sugarcane fields using a multispectral camera mounted on a small unmanned aerial system (UAS). *Sugar Tech*, 22, 605–621.
- CONADESUCA (Comité Nacional para el Desarrollo sustentable de la Caña de Azúcar). (2015). *Ficha técnica del cultivo de la caña de azúcar (Saccharum officinarum L.)*.
- Corporación Colombiana de Investigación Agropecuaria (CORPOICA). (2015). *Misión para la transformación del campo*. Departamento Nacional de Planeación - DNP.
- de Almeida, G. M., Pereira, G. T., de Souza Bahia, A. S., Fernandes, K., & Júnior, J. M. (2021). Machine learning in the prediction of sugarcane production environments. *Computers and Electronics in Agriculture*, 190(106452).  
<https://doi.org/https://doi.org/10.1016/j.compag.2021.106452>
- Díaz García-Cervigón, J. J. (2015). *Estudio de Índices de vegetación a partir de imágenes aéreas tomadas desde UAS/RPAS y aplicaciones de estos a la agricultura de precisión*. UNIVERSIDAD COMPLUTENSE DE MADRID.
- dos Santos Luciano, A. C., Picoli, M. C., Duft, D. G., Rocha, J. V., Leal, M. R., & Le Maire, G. (2021). Empirical model for forecasting sugarcane yield on a local scale in Brazil using

- Landsat imagery and random forest algorithm. *Computers and Electronics in Agriculture*, 184(106063). <https://doi.org/https://doi.org/10.1016/j.compag.2021.106063>
- Erazo, E., Carbonell, J., & Isaacs, C. (2013). Factors inducing spatial variability in the productivity of sugarcane crops. *International Society of Sugar Cane Technologists: Proceedings of the XXVIIIth Congress*, (págs. 503-514). São Paulo: Hogarth, D. M. <https://www.cabdirect.org/cabdirect/abstract/20193249757>
- Espinal, C., Martínez, H., Pinzón, N., & Barrios, C. (2005). *Ministerio de Agricultura y Desarrollo Rural: Observatorio Agrociencias Colombia*.
- European Space Agency. (21 de 01 de 2023). *European Space Agency*. <https://www.esa.int/eseach?q=snap>
- Everingham, Y., Inman-Bamber, N., Thorburn, P., & McNeill, T. (2007). A Bayesian modelling approach for long lead sugarcane yield forecasts for the Australian sugar industry. *Australian Journal of Agricultural Research*, 58(2), 87-94. <https://doi.org/https://doi.org/10.1071/AR05443>
- Everingham, Y., Muchow, R., Stone, R. C., Inman Bamber, N., Singels, A., & Bezuidenhout, C. (2002). Enhanced risk management and decision-making capability across the sugarcane industry value chain based on seasonal climate forecasts. *Agricultural Systems*, 74(3), 459-477. [https://doi.org/https://doi.org/10.1016/S0308-521X\(02\)00050-1](https://doi.org/https://doi.org/10.1016/S0308-521X(02)00050-1)
- Everingham, Y., Muchow, R., Stone, R. C., Inman-Bamber, N., Singels, A., & Bezuidenhout, C. (2002). Enhanced risk management and decision-making capability across the sugarcane industry value chain based on seasonal climate forecasts. *Agricultural Systems*, 74(3), 459-477. [https://doi.org/https://doi.org/10.1016/S0308-521X\(02\)00050-1](https://doi.org/https://doi.org/10.1016/S0308-521X(02)00050-1)

- Everingham, Y., Smyth, C., & Inman-Bamber, N. (2009). Ensemble data mining approaches to forecast regional sugarcane crop production. *Agricultural and forest meteorology*, 149(3-4), 689-696. <https://doi.org/10.1016/j.agrformet.2008.10.018>
- Fernandes, J. L., Ebecken, N. F., & Esquerdo, J. C. (2017). Sugarcane yield prediction in Brazil using NDVI time series and neural networks ensemble. *International Journal of Remote Sensing*, 38(16). <https://doi.org/10.1080/01431161.2017.1325531>
- Filipponi, F. (2019). Sentinel-1 GRD Preprocessing Workflow. *Multidisciplinary digital publishing institute proceedings*.
- Gilabert, M. A., González-Piqueras, J., & García-Haro, J. (1997). Acerca de los índices de vegetación. *Revista de teledetección*, 1 - 10.
- Han, S. Y., Bishop, T. F., & Filippi, P. (2022). Data-driven, early-season forecasts of block sugarcane yield for precision agriculture. *Field Crops Research*, 276(108360). <https://doi.org/10.1016/j.fcr.2021.108360>
- Hernández Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, M. (2010). *Metodología de la Investigación. Quinta Edición*. The McGraw-Hill.
- Larrahondo, J. E., & Villegas, F. (1995). Control y características de maduración. En CENICAÑA, *El Cultivo de la Caña de Azúcar en la Zona Azucarera de Colombia* (págs. 297 - 313).
- Lawes, R., & Lawn, R. (2005). Applications of industry information in sugarcane production systems. *Field Crops Research*, 92(2-3), 353-363. <https://doi.org/10.1016/j.fcr.2005.01.033>
- Li, Z., Ding, L., & Xu, D. (2022). Exploring the potential role of environmental and multi-source satellite data in crop yield prediction across Northeast China. *Science of The Total*

*Environment*, 815(152880).

<https://doi.org/https://doi.org/10.1016/j.scitotenv.2021.152880>

Lofton, J., Tubana, B. S., Kanke, Y., Teboh, J., Viator, H., & Dalen, M. (2012). Estimating sugarcane yield potential using an in-season determination of normalized difference vegetative index. *Sensors*, 12(6), 7529-7547;. <https://doi.org/10.3390/s120607529>

Martínez Covalada, H. J., Espinal, C. F., & Ortiz Hermida, L. (2005). *La cadena del azúcar en Colombia: una mirada global de su estructura y dinámica 1991-2005*. Bogota: Ministerio de Agricultura y Desarrollo Rural Observatorio Agrocadenas.

Ministerio de Agricultura y Desarrollo Rural, Bogotá. (2005). *La cadena del azúcar en Colombia :una mirada global de su estructura y dinámica 1991 - 2005*.

Morel, J., Todoroff, P., Bégué, A., Bury, A., Martiné, J. F., & Petit, M. (2014). Toward a satellite-based system of sugarcane yield estimation and forecasting in smallholder farming conditions: A case study on Reunion Island. *Remote Sensing*, 6(7), 6620-6635. [https://doi.org/ https://doi.org/10.3390/rs6076620](https://doi.org/https://doi.org/10.3390/rs6076620)

Muller, S. J., Sithole, P., Singels, A., & Van Niekerk, A. (2020). Assessing the fidelity of Landsat-based fAPAR models in two diverse sugarcane growing regions. *Computers and Electronics in Agriculture*, 170(105248).

<https://doi.org/https://doi.org/10.1016/j.compag.2020.105248>

Muñoz-Aguayo, P. (2013). *Apuntes de Teledetección: Índices de Vegetación*. Centro de Información de Recursos Naturales (Chile).

<http://bibliotecadigital.ciren.cl/bitstream/handle/20.500.13082/26389/Tema%20Indices%20de%20vegetaci%c3%b3n%2c%20Pedro%20Mu%c3%b1oz%20A.pdf?sequence=1&isAllowed=y>

- Murillo Sandoval, P. J., & Carbonell González, J. A. (2012). *Principios y aplicaciones de la percepción remota en el cultivo de la caña de azúcar en Colombia*. Cenicaña. Servicio de cooperación técnica y transferencia de tecnología.
- [https://doi.org/https://www.cenicana.org/pdf\\_privado/documentos\\_no\\_seriados/libro\\_percepcion\\_remota/principios-y-aplicaciones\\_percepcion-remota.pdf](https://doi.org/https://www.cenicana.org/pdf_privado/documentos_no_seriados/libro_percepcion_remota/principios-y-aplicaciones_percepcion-remota.pdf)
- Murillo Sandoval, P., & Carbonell González, J. (2012). *Principios y aplicaciones de la percepción remota en el cultivo de la caña de azúcar en Colombia (No. U40-01)*. cenicaña (Colombia) Departamento Administrativo de Ciencia, Tecnología e Innovación (Colombia).
- Orozco, Ó. A., & Llano, G. (2016). Sistemas de información enfocados en tecnologías de agricultura de precisión y aplicables a la caña de azúcar, una revisión. *Revista Ingenierías Universidad de Medellín*, 15(28), 103-124. <https://doi.org/10.22395/rium.v15n28a6>
- Paloscia, S. (1998). An Empirical Approach to Estimating Leaf Area Index from Multifrequency SAR Data. *International Journal of Remote Sensing*, 19(2), 359-364.
- <https://doi.org/https://doi.org/10.1080/014311698216323>
- Pérez López, C., & Santin González, D. (2007). *Minería de datos. Técnicas y herramientas: técnicas y herramientas*. Ediciones Paraninfo, SA.
- Picoli, M. C., Lamparelli, R. A., Sano, E. E., Mello, J. R., & Rocha, J. V. (2013). Effect of sugarcane-planting row directions on ALOS/PALSAR satellite images. *GIScience & Remote Sensing*, 50, 349-357.
- <https://doi.org/https://doi.org/10.1080/15481603.2013.808457>
- Rahimi Jamnani, M., Liaghat, A., & Sadeghi Loyeh, N. (2019). Sugarcane yield prediction at farm scale using remote sensing and artificial neural network. En *11th World Congress*

*on Water Resources and Environment: Managing Water Resources for a Sustainable Future-EWRA 2019. Proceedings.*

Rahman, M. M., & Robson, A. (2020). Integrating Landsat-8 and Sentinel-2 time series data for yield prediction of sugarcane crops at the block level. *Remote Sensing*, 12(8).

<https://doi.org/10.3390/rs12081313>

Ramírez, M. Á. (2008). *Cultivos para la producción sostenible de biocombustibles: una alternativa para la generación de empleos e ingresos Modulo 5: Caña de azúcar*. Servicio Holandes de Cooperación para el Desarrollo, Tegucigalpa (Honduras).

Reyes-Trujillo, A., Daza-Torres, M. C., Galindez-Jamioy, C. A., Rosero-García, E. E., Muñoz-Arboleda, F., & Solarte-Rodríguez, E. (2021). Estimating canopy nitrogen concentration of sugarcane crop using in situ spectroscopy. *Heliyon*, 7(3).

<https://doi.org/10.1016/j.heliyon.2021.e06566>

Riquelme Santos, J. C., Ruiz, R., & Gilbert, K. (2006). Minería de datos: Conceptos y tendencias. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial*, 10(29), 11-18.

Sander, R. S. (14 de Marzo de 2021). *Towards Data Science*. Towards Data Science:

<https://towardsdatascience.com/gaussian-process-regression-from-first-principles-833f4aa5f842>

Shendryk, Y., Davy, R., & Thorburn, P. (2021). Integrating satellite imagery and environmental data to predict field-level cane and sugar yields in Australia using machine learning. *Field Crops Research*, 260(107984).

<https://doi.org/10.1016/j.fcr.2020.107984>

- Sumesh, K., Ninsawat, S., & Som-ard, J. (2021). Integration of RGB-based vegetation index, crop surface model and object-based image analysis approach for sugarcane yield estimation using unmanned aerial vehicle. *Computers and Electronics in Agriculture*, 180(105903). <https://doi.org/https://doi.org/10.1016/j.compag.2020.105903>
- Tian, L., Wang, C., Li, H., & Sun, H. (2020). Yield prediction model of rice and wheat crops based on ecological distance algorithm. *Environmental Technology & Innovation*, 20(101132). <https://doi.org/https://doi.org/10.1016/j.eti.2020.101132>
- Vecchio, Y., De Rosa, M., Adinolfi, F., Bartoli, L., & Masi, M. (2020). Adoption of precision farming tools: A context-related analysis. *Land Use Policy*, 94(104481). <https://doi.org/https://doi.org/10.1016/j.landusepol.2020.104481>
- Verrelst, J., Alonso, L., Rivera Caicedo, J., Moreno, J., & Camps-Valls, G. (2013). Gaussian Process Retrieval of Chlorophyll Content From Imaging Spectroscopy Data. *IEEE*, 867-874.
- Villegas, H. (2008). *Introducción a la percepción remota y sus aplicaciones geológicas*. Servicio Geológico Colombiano.
- Von Martini, A., Bragachini, M., Bianchini, A., Martellotto, E., & Méndez, A. (30 de 11 de 2022). *Instituto Nacional de Tecnología Agropecuaria - INTA*. Instituto Nacional de Tecnología Agropecuaria - INTA: <https://inta.gob.ar/sites/default/files/script-tmp-inta-percepcion-remota.pdf>
- Yearbook, FAO Statistical. (2013). World food and agriculture. *World Food and Agriculture. Food and Agriculture Organization of the United Nations, Rome*.
- Young, K., Israel, G. J., & Marco, J. (2019). ESTIMADOS DE PRODUCCIÓN DE CAÑA DE AZÚCAR (*Saccharum* spp.) UTILIZANDO UN MÉTODO BIOFÍSICO. *XLI*

*CONVENCIÓN Y EXPOATAM 2019*. <https://www.atamexico.com.mx/wp-content/uploads/2019/10/11.CAMPO-XLI.pdf>

Yu, D., Zha, Y., Shi, L., Jin, X., Hu, S., Yang, Q., . . . Zeng, W. (2020). Improvement of sugarcane yield estimation by assimilating UAV-derived plant height observations. *European Journal of Agronomy*, *121*(126159).  
<https://doi.org/https://doi.org/10.1016/j.eja.2020.126159>

Yu, D., Zha, Y., Shi, L., Ye, H., & Zhang, Y. (2022). Improving sugarcane growth simulations by integrating multi-source observations into a crop model. *European Journal of Agronomy*, *132*(126410). <https://doi.org/https://doi.org/10.1016/j.eja.2021.126410>

Yu, D., Zha, Y., Shi, L., Ye, H., & Zhang, Y. (2022). Improving sugarcane growth simulations by integrating multi-source observations into a crop model. *European Journal of Agronomy*.