

UNIVERSIDAD DE COSTA RICA
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA DE BIOSISTEMAS

ANÁLISIS CORRELATIVO DE ÍNDICES DE VEGETACIÓN CON INCIDENCIA
TEMPRANA DE LA ROYA (*hemileia vastatrix*) EN PLANTAS DE CAFÉ (*coffea arabica*) EN
LA PARCELA EXPERIMENTAL DE LA HACIENDA ALSACIA

Trabajo final de graduación tipo práctica dirigida para optar por el título de:

Licenciatura en Ingeniería de Biosistemas

Modalidad: Práctica Dirigida

Sustentante:

Andrés Eduardo Lastra Lastra

Carné: B74073

Director:

Ing. Matías Adrián Chaves Herrera, M.Sc.

Lectores:

Ing. Melissa Rojas Downing, PhD

Ing. José Francisco Aguilar Pereira, M.Sc.

Ciudad Universitaria Rodrigo Facio

San José, Costa Rica

2024

Trabajo Final De Graduación bajo modalidad de practica dirigida, sometido a revisión por la Comisión De Trabajos Finales De Graduación de la Escuela de Ingeniería de Biosistemas de la Facultad de Ingeniería de la Universidad de Costa Rica, como requisito para optar por el grado de Licenciatura en Ingeniería Agrícola y de Biosistemas



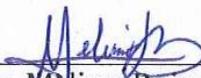
Andrés Eduardo Lastra Lastra

Postulante



Ing. Matías Adrián Chaves Herrera, M.Sc.

Director, Comité Asesor



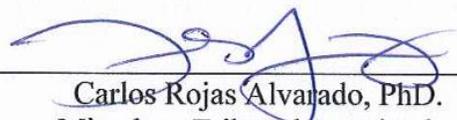
Ing. Melissa Rojas Downing, PhD.
Comité Asesor



Ing. José Francisco Aguilar Pereira, M.Sc.
Comité Asesor



Ing. Alejandra Rojas González, PhD.
Presidenta, Tribunal examinador



Carlos Rojas Alvarado, PhD.
Miembro, Tribunal examinador

DEDICATORIA

Mi familia, mis amigos, **MI PAÍS**, mi gente.

AGRADECIMIENTOS

A Dios,

A mis 33 profesores, quienes hacen esta universidad un hogar,

A mis inigualables 22 compañeros a quienes con honor llamo colegas. Junto al ingeniero y amigo Jorge Mora.

A mis cuatro padres, que dan todo por mí y yo por ellos.

A mis dos hermanos, que me han visto crecer en este camino,

A ti, quien lees este documento.

RESUMEN

La presencia del hongo de la roya (*hemileia vastatrix*) en las hojas de cafetos (*coffea Arabica*) perjudica el crecimiento y desarrollo adecuado de la planta, y de los frutos (evitando la fotosíntesis). Para luchar contra pérdidas en la cosecha, se hace una evaluación extensa de las parcelas, estudiando cafetos representativos y los patógenos presentes. Este proceso se lleva a cabo de manera visual por los trabajadores y agrónomos, lo cual es altamente demandante en tiempo y de personal calificado. Para facilitar el trabajo a los agrónomos y poder detectar remotamente el hongo, se lleva a cabo este trabajo final de graduación (TFG). El objetivo general de esta investigación es evaluar la correlación entre la incidencia temprana de la roya del café y los índices de vegetación (IV) calculados a partir de imágenes multiespectrales en la Hacienda Alsacia, Alajuela, Costa Rica. Esto se llevó a cabo por medio de la extracción de información sobre la incidencia de la roya en cafetos con la herramienta de ESRI y el apoyo del Ingeniero Jorge Mora. Se dividió la parcela en 20 bloques, donde se analizaron tres cafetos en tres tomas de datos (abril, mayo y julio del 2024). Los datos mostraron una mayor incidencia de roya en la tercera toma de datos, especialmente en las zonas del noreste (parte más alta de la parcela). Posteriormente, se generaron IV (NDVI, GNDVI, GEMI, NDVI_{re}) usando imágenes multiespectrales capturadas con un dron Mavic 3M. Las imágenes se procesaron en ArcGIS Pro para calcular los índices de vegetación, los datos se limpiaron y se generaron promedios de cada IV en cada toma de datos para cada bloque. Los resultados de IV mostraron que las plantas están en un rango de poca salud (ningún valor supera 0,6, indicando baja a media actividad vegetal y fotosíntesis). Para finalizar con la elección del mejor IV se correlacionaron los resultados anteriores con la incidencia de la Roya. Usando el software estadístico SPSS, se determinó la normalidad de los datos y se generaron tres coeficientes de correlación (CC), siendo el mejor para cada uno los siguientes: Tau-b de Kendall ($p(GEMI_{@R3}) = -0,266$) y Spearman ($r(GEMI_{@R3}) = -0,267$). Se seleccionó GEMI como el mejor IV, pero al tener una correlación débil ($0,3 > p = 0,267 > 0$) y no significativa ($\text{sig} = 0,252 > \alpha = 0,05$, no se rechaza H_0) no se puede asegurar que el IV seleccionado detecta la roya en los cafetos.

Palabras Clave: NDVI, GEMI, GNDVI, NDVI_{re}, Kendall, Café, Hongo.

SUMMARY

The presence of coffee leaf rust (*hemileia vastatrix*) on the leaves of coffee plants (*coffea arabica*) impairs the growth and proper development of the plant and the fruits (preventing photosynthesis). To combat harvest losses, extensive evaluation of the plots is carried out by studying representative coffee plants and the pathogens present. This process is conducted visually by workers and agronomists, which is highly demanding in terms of time and qualified personnel. To facilitate the work of agronomists and remotely detect the fungus, this graduation project (TFG) was undertaken. The general objective of this research is to evaluate the correlation between the early incidence of coffee rust and vegetation indices (VI) calculated from multispectral images at Hacienda Alsacia, Alajuela, Costa Rica. This was done by extracting information on the incidence of rust in coffee plants using the ESRI tool and with the support of Engineer Jorge Mora. The plot was divided into 20 blocks, where three coffee plants were analyzed in three replicates (April, May, and July 2024). The data showed a higher incidence of rust in the third replicate, especially in the northeastern areas (the highest part of the plot). Subsequently, VIs (NDVI, GNDVI, GEMI, NDVI_{re}) were generated using multispectral images captured with a Mavic 3M drone. The images were processed in ArcGIS Pro to calculate the vegetation indices, the data were cleaned, and averages of each VI were generated in each replicate for each block. The VI results showed that the plants were in a range of poor health (no value exceeded 0.6, indicating low to medium vegetation activity and photosynthesis). Finally, to select the best VI, the previous results were correlated with the incidence of rust. Using SPSS statistical software, the normality of the data was determined and three correlation coefficients (CC) were generated, with the best for each being: Kendall's Tau-b ($p(GEMI_{@R3}) = -0.266$) and Spearman ($r(GEMI_{@R3}) = -0.267$). GEMI was selected as the best VI, but due to its weak correlation ($0.3 > p = 0.267 > 0$) and non-significance ($\text{sig} = 0.252 > \alpha = 0.05$, H_0 is not rejected), it cannot be assured that the selected VI detects rust in coffee plants.

Key Words: NDVI, GEMI, GNDVI, NDVI_{re}, Kendall, Coffee, Rust.

ÍNDICE GENERAL

1. INTRODUCCIÓN	1
1.1. Justificación	2
1.2. Delimitación del problema.....	3
1.3. Objetivos.....	4
1.3.1. Objetivo general.....	4
1.3.2. Objetivos específicos	4
2. MARCO TEÓRICO	5
2.1. Marco conceptual.....	5
2.1.1. Roya	5
2.1.2. Aspectos Agroclimáticos del Café	5
2.1.2.1. Altitud	5
2.1.2.2. Temperatura	5
2.1.2.3. Precipitación	6
2.1.2.4. Suelo	6
2.1.2.5. Sombra	6
2.1.2.6. Humedad Relativa.....	6
2.1.3. Incidencia.....	6
2.1.4. Teledetección	7
2.1.5. Imágenes multiespectrales (IM).....	7
2.1.6. Ráster	7
2.1.7. Ortofotos y ortomosaicos.....	7
2.1.1. Geoproceso	8
2.1.2. Support Vector Machine (SVM).....	8
2.1.3. Índice de vegetación (IV).....	8

2.1.3.1.	NDVI.....	8
2.1.3.2.	GNDVI.....	9
2.1.3.3.	GEMI	9
2.1.3.4.	NDVIre	9
2.1.4.	Muestreo por bloques (estratificado)	9
2.1.5.	Estadística Paramétrica (EP).....	10
2.1.6.	Estadística no Paramétrica (ENP).....	10
2.1.7.	Shapiro-Wilk.....	10
2.1.8.	Coefficientes de correlación (CC).....	11
2.1.8.1.	Coefficiente de correlación de Spearman.....	11
2.1.8.2.	Coefficiente de correlación de Tau-b de Kendall.....	11
2.1.9.	Potencia estadística (PE).....	12
2.1.10.	Tamaño del efecto (TE)	12
2.2.	Antecedentes	13
2.3.	Marco de referencia empresarial.....	17
3.	METODOLOGÍA.....	17
3.1.	Sitio de estudio.....	18
	Diseño experimental	19
3.1.1.	Objetivo 1: Clasificar espacialmente la incidencia de la roya en cafetos en la parcela experimental de la Hacienda Alsacia.....	19
3.1.1.1.	Inspección de la hacienda	19
3.1.1.2.	Clasificación de datos	20
3.1.1.3.	Diseño del muestreo.....	21
3.1.1.4.	Muestreo de la toma de datos.....	23
3.1.1.5.	Toma de datos	24

3.1.2.	Objetivo 2: Recolectar imágenes multiespectrales para el cálculo de cuatro diferentes índices de vegetación que indican el estado de vigor y salud de la planta.....	26
3.1.2.1.	Muestreo de toma de datos	26
3.1.2.2.	Procesamiento de imágenes	27
3.1.3.	Objetivo 3: Seleccionar el índice de vegetación que tenga la mejor correlación con la incidencia en etapas tempranas de la roya en café por métodos estadísticos.....	30
3.1.3.1.	Coefficientes y la correlación de la Información	30
3.1.3.2.	Normalidad de los datos.....	33
4.	RESULTADOS Y DISCUSIÓN	33
4.1.	Objetivo específico 1: Clasificación de la incidencia de la roya en la parcela experimental.....	33
4.2.	Objetivo específico 2: Cálculo y Gestión de los Índices de Vegetación.....	37
4.3.	Objetivo específico 3: Selección del índice de vegetación basado en correlación estadística.....	42
5.	CONCLUSIONES.....	50
6.	RECOMENDACIONES	52
7.	BIBLIOGRAFÍA.....	54
	Apéndice A.....	61
	Apéndice B.....	81
	Apéndice C.....	83

ÍNDICE DE FIGURAS

Figura 1. Diagrama de flujo dividido por objetivos.....	18
Figura 2. Temperatura y precipitación de la zona.....	19
Figura 3. Exploración de la Hacienda Alsacia.	20
Figura 4. Hojas de cafeto de la parcela experimental de la Hacienda Alsacia.	21
Figura 5. Distribución de bloques en la parcela experimental.....	22
Figura 6. Encuesta de campo para la clasificación de la incidencia de la roya en cafeto en la plataforma Survey123 de ArcGIS.	24
Figura 7. Fechas de toma de datos con respecto a la presencia del hongo.	25
Figura 8. Gráfico de barras de la incidencia de la roya para cada bloque en cada toma de datos.	34
Figura 9. Gráfico de cajas de la incidencia de la roya en cada toma de datos.....	35
Figura 10. Gráfico de cajas de la incidencia de la roya para cada bloque.	35
Figura 11. Distribución geográfica de la incidencia de la roya: toma de datos 1, toma de datos 2 y toma de datos 3.	36
Figura 12. Gráfico de cajas para los índices de vegetación calculados (abril-julio).	38
Figura 13. Gráfico de caja para el NDVI en tres momentos en el tiempo.	39
Figura 14. Índices de vegetación calculados en la parcela experimental para las tomas de datos 1, 2 y 3.....	40
Figura 15. Gráfico de puntos de incidencia de la roya con IV calculados de la toma de datos 1.43	
Figura 16. Reclasificación de los índices de vegetación en escalas ordinales.....	45
Figura 17. Mapa de GEMI de la tercera toma de datos (Mejor correlación con la incidencia de la roya).	49

ÍNDICE DE TABLAS

Tabla 1. Orden de las bandas generadas por PIX4D para ser compuestas en ArcGIS Pro.	28
Tabla 2. Coeficiente Kappa para la clasificación realizada en las mascararas de las tres tomas de datos.	30
Tabla 3. Reclasificación de los índices de vegetación (rangos).	32
Tabla 4. Correlación de Spearman de la incidencia de la roya y los índices de vegetación calculados.	44
Tabla 5. Correlación de Tau-b de Kendall de la incidencia de la roya y los índices de vegetación calculados.	45
Tabla 6. Correlación de índices de vegetación con incidencia temprana de la roya en cafetos. ..	46
Tabla 7 Selección del índice de correlación que tenga la mejor relación con la incidencia de la roya en etapas tempranas en tres tomas de datos.	47

INDICE DE ABREVIATURAS

Coefficientes de correlación.....	CC
La Escuela de Ingeniería de Biosistemas.....	EIB
Estadística no paramétrica.....	ENP
Estadística paramétrica.....	EP
Banda verde.....	G
Índice de monitoreo ambiental global.....	GEMI
Sistemas de información geográfica	GIS
Índice de vegetación normalizado verde	GNDVI
Productividad primaria bruta	GPP
Hipótesis nula.....	H_0
Hipótesis alterna.....	H_1
Índices de correlación.....	IC
Imágenes multiespectrales.....	IM
Índices de vegetación.....	IV
Tamaño de la muestra.....	n
El índice de vegetación por diferencia normalizada.....	NDVI
Índice de vegetación de diferencia normalizada en el borde rojo.....	NDVIre
Banda cerca del infrarrojo.....	NIR
Significancia bilateral de las dos colas.....	Sig.
Statistical Package for the Social Sciences	SPSS
Potencia estadística.....	PE
Banda de borde rojo.....	RE
Banda roja.....	RED
Imagen del espectro visible (rojo, verde y azul).....	RGB
Posicionamiento cinemático en tiempo real.....	RTK
Tamaño del efecto.....	TE
Trabajo final de graduación.....	TFG
Eficiencia en el uso del agua.....	WUE
Nivel de error.....	α
Probabilidad de no rechazar una hipótesis nula cuando en realidad se debería.....	β
Tau-b de Kendall (coeficiente).....	ρ

1. INTRODUCCIÓN

El café es un producto agrícola de gran relevancia global y su cadena productiva genera empleo para millones de personas. Asimismo, tiene un impacto significativo en la economía de varios países en desarrollo ubicados en las regiones ecuatoriales y subecuatoriales de África, América y Asia. En cuanto a las especies que dominan el mercado mundial del café, son dos: *Coffe arabica L.* (café Arábica) y *C. canephora Pierre ex A. Froehner* (café Robusta), con una participación del 75% y 25%, respectivamente (Thao et al., 2022). La primera especie produce una bebida aromática y suave, muy apreciada por los consumidores; mientras que la segunda produce una bebida con un sabor más intenso y amargo (Velásquez et al., 2021).

En Costa Rica, el 100 % del café legalmente producido corresponde a la especie arábica, de las variedades *caturra* y *catuai*, las cuales presentan granos y tazas de alta calidad. Desde 1989, está prohibida por la Ley, la siembra de café de la especie robusta, debido a su baja calidad de taza. Además, en el país, se cuenta con ocho zonas productoras de café: Brunca, Turrialba, Tres Ríos, Orosi, Tarrazú, Valles Central y Occidental, y Guanacaste.

El objetivo principal de la industria cafetalera en el país es la producción de café fino, donde se destaca su mayor calidad y no la producción de grandes cantidades (ICAFE, 2022). En los últimos años, se ha experimentado un aumento de la producción de café, pasando de 1.672.510 fanegas (una fanega es 260 kg, también equivale a 20 cajuelas, cada una de 13 kg) producidas a 1.908.938 en el periodo 2022-2023; lo que significa un incremento del 14,14% (ICAFE, 2023).

Un problema común que enfrentan las fincas agrícolas es el control de patógenos de manera efectiva. Una de las enfermedades más destructiva ha sido la provocada por el hongo de la Roya (*hemileia vastatrix*), que causa el amarillamiento de las hojas y la disminución de la fotosíntesis al deshojar las plantas de café en un lapso de varios meses o años. A pesar de que se han desarrollado variedades de café resistentes a la enfermedad, estas han ido perdiendo efectividad por la mutación del hongo. Se han reportado epidemias que han generado daños desastrosos en Sri Lanka y el sur de la India (1869–1885), África occidental (1950–1970) y América Latina (1970–1989) (Katsuhama et al., 2018).

En América Latina, en épocas recientes, la enfermedad se ha estado expandiendo y se ha convertido en una gran amenaza para la industria del café (Katsuhama et al., 2018). El método tradicional para detección de la roya consiste en la inspección visual de la planta e identificando signos que muestren la presencia de la enfermedad; debido a esto, es un proceso que consume mucho tiempo.

Para mejorar la eficiencia de su detección, se han generado nuevos métodos que implementen acciones para mitigar el daño, destacando la identificación de la enfermedad en etapas tempranas para brindar un manejo adecuado (Gavito Muñoz, 2019). La detección de la roya en el sitio ha sido el único método efectivo para eliminar las plantas de café afectadas, a fin de prevenir la propagación de la infección; sin embargo, es difícil cubrir áreas extensas con precisión y velocidad mediante reconocimientos terrestres (Katsuhama et al., 2018).

1.1. Justificación

El carecer de un sistema de detección temprana y un manejo adecuado ante la roya, puede implicar una pérdida de hasta el 50 % de hojas y una reducción del 70 % en la producción de café si esta no se controla a tiempo. Por ello, conocer los síntomas, la gravedad y la distribución espacial en el campo de la roya es vital para llevar a cabo un manejo correcto de la enfermedad y reducir las pérdidas asociadas a ella (Barbosa et al., 2021).

Una de las herramientas que puede facilitar esta tarea es el uso de la agricultura de precisión, la cual permite monitorear y analizar las condiciones del cultivo y su entorno mediante sensores, GPS, drones y otros dispositivos. De una manera económica y amigable con el medio ambiente se puede obtener información específica sobre el estado de las plantas y las enfermedades que las atacan (Velásquez et al., 2021).

La teledetección es una herramienta útil para monitorear las condiciones de los cultivos, ya que detecta los cambios en la reflectancia de las plantas sanas y enfermas. Estos cambios se deben a la afectación de la clorofila y las estructuras foliares por las enfermedades y plagas. A diferencia de las imágenes RGB, las imágenes multiespectrales son más efectivas en este ámbito, ya que cubren

más rangos del espectro que lo visible por el ojo humano (Barbosa et al., 2021; Sánchez-Méndez y Arguijo-Hernández, 2018).

Esta investigación se clasifica como preexperimental, la cual tuvo como el fin el detectar la roya en etapa temprana de su desarrollo mediante sistemas de detección a distancia en la Hacienda Alsacia, a través de índices de vegetación (IV) derivados de imágenes multiespectrales. Al presentar esta investigación a la empresa, se cumplió con propiciar una herramienta alternativa de detección de la roya, permitiendo reorientar la mano de obra hacia actividades más exigentes. Según informó la empresa, cada año disminuye el número de personas dispuestas a trabajar en tareas agrícolas. El utilizar IV para calificar la salud de las plantas, le permite a la empresa poder segmentar en zonas de alto riesgo y otras más saludables, haciendo más eficiente el uso de funguicidas.

La relevancia de abordar estos aspectos en las empresas agrícolas es crucial para la tecnificación del campo. Esta investigación abre la puerta a futuros trabajos en teledetección de la roya del café, permitiendo probar diferentes IV. Además, establece las bases de una metodología que puede ser ajustada y mejorada según las necesidades de investigaciones futuras y en distintas épocas del año. Junto a esto, se contribuyó en esta investigación a los Objetivos del Desarrollo Sostenible (ODS) 2 y 12. Puede ver un desglose de esta información en el anexo (**Tabla A 11**).

1.2. Delimitación del problema

La localización de la roya en la parcela experimental fue el problema a solucionar, para su debido control posteriormente por la empresa. Para esto, se analizó la correlación entre IV y la incidencia de la roya presente en la parcela experimental de la Hacienda Alsacia. Por medio de criterio experto, herramientas geomáticas, estadísticas e imágenes multiespectrales de la parcela experimental, se calculó una correlación entre estas dos variables (IV y la incidencia de la roya).

Por otro lado, Sánchez-Méndez y Arguijo-Hernández, (2018) decidió combinar las bandas del Landsat-8 6, 5 y 4 para realizar el análisis vegetativo por medio del NDVI, a fin de filtrar los cultivos y trabajar únicamente con el café. Junto a este filtro pudo determinar que cafetos se encontraban enfermos. En el caso de la roya esta información es relevante, ya que confirma el uso

de esta tecnología en otras partes del continente, junto a el uso de este IV para determinar enfermedades en los cafetos.

Como delimitación temporal, esta investigación se realizó en un periodo de seis meses, comenzó con la toma de datos de abril a julio y finalizó con el escrito en el mes de septiembre del 2024. La enfermedad fue detectada con la asistencia del Ingeniero Agrónomo Jorge Mora, lo que requirió inspección en campo. Como no se contaba con todo el personal deseado, se escaló a un área manejable, delimitado espacialmente a una parcela en la zona norte de la Hacienda que cuenta con 2,3 ha.

Se tomaron datos en tres instantes de la plaga, por medio de un barrido de información con la ayuda de una aplicación (Survey 123) y se clasificaron según la incidencia del hongo y el criterio experto. Paralelamente se recolectaron imágenes multiespectrales por medio de un dron, para calcular cuatro IV (NDVI, GNDVI, GEMI y NDVI_{re}). Por métodos estadísticos de correlación, se determinó cuál de los IV calculados se ajustó mejor a la clasificación de la incidencia del hongo.

1.3. Objetivos

1.3.1. Objetivo general

Evaluar la correlación entre la incidencia en etapas tempranas de la roya (*hemileia vastatrix*) del café (*coffea arabica*) y los índices de vegetación calculados a partir de imágenes multiespectrales en la parcela experimental de la Hacienda Alsacia, Alajuela, Costa Rica.

1.3.2. Objetivos específicos

1. Clasificar espacialmente la incidencia de la roya en cafetos en la parcela experimental de la Hacienda Alsacia.
2. Recolectar imágenes multiespectrales para el cálculo de cuatro diferentes índices de vegetación que indican el estado de vigor y salud de la planta.
3. Seleccionar el índice de vegetación que tenga la mejor correlación con la incidencia en etapas tempranas de la roya en cafeto por métodos estadísticos.

2. MARCO TEÓRICO

2.1. Marco conceptual

2.1.1. Roya

La enfermedad fúngica conocida como roya del café (*hemileia vastatrix*), reportada por primera vez en 1869, en Sri Lanka, provoca la defoliación de los cafetos (plantas de café), además, disminuye su rendimiento y calidad. Esta patología se originó en el continente africano y se ha propagado por todas las zonas productoras de café del mundo, salvo Hawái. Su incidencia está determinada por factores ambientales, como la temperatura, la humedad y la precipitación, así como por factores genéticos, como el nivel de resistencia o susceptibilidad de las distintas variedades de café. Por su parte, las estrategias de control de la roya se fundamentan en el empleo de fungicidas, el mejoramiento genético y el manejo integrado del cultivo como indica la organización SENASICA (2019).

2.1.2. Aspectos Agroclimáticos del Café

Los aspectos Agroclimáticos son las variables climáticas ideales para que un cultivo sea exitoso. Una de las más fundamentales son las siguientes:

2.1.2.1. Altitud

La altitud define otros factores tales como la temperatura y precipitación. La altura óptima para la producción de café es entre los 500 y 1800 msnm. Si la altura es mayor, puede perjudicar al desarrollo del cafeto (CICAFE, 2020).

2.1.2.2. Temperatura

Idealmente esta variable debería oscilar entre los 17 y 23 °C. Temperaturas inferiores a 10 °C genera clorosis y paralización del crecimiento de hojas. En caso contrario, temperaturas mayores a los 24 °C genera un déficit en la producción (CICAFE, 2020).

2.1.2.3. *Precipitación*

El rango ideal es de 1500 a 2000 mm anuales. Si llueve menos de 1000 mm, se limitará el crecimiento de las plantas, afectando la futura cosecha, incluso la muerte de los cafetos. Mientras que precipitación mayor a 3000, afecta directamente al grano, junto a un aumento en los controles fitosanitarios (CICAFE, 2020).

2.1.2.4. *Suelo*

El cafeto prefiere un suelo rico en materia orgánica, con buenos drenajes y un pH entre 6 y 6,5. Esto permite un desarrollo de las raíces, y una buena absorción de los nutrientes (Garro Chacón, 2020).

2.1.2.5. *Sombra*

Aunque los cafetos pueden cultivarse a plena luz del día, la sombra proporcionada por árboles o sistemas artificiales permite mantener una temperatura más baja, reduciendo la evaporación del agua (Farfán-Valencia, 2021).

2.1.2.6. *Humedad Relativa*

La humedad relativa ideal promedio mensual se encuentra entre 60% y 80%, al alcanzar niveles de 85% se propicia el crecimiento y propagación de enfermedades fungosas (CICAFE, 2020).

2.1.3. *Incidencia*

La incidencia se define en este documento como el promedio de severidad en un área delimitada. En epidemiología, el termino severidad se utiliza para determinar qué tan grave es la presencia (en este caso) de una enfermedad en un sujeto estudiado. Para esta investigación, la severidad de la roya está basada en el estudio de SENASICA (2019).

2.1.4. Teledetección

Permite obtener información de un objeto o superficie sin entrar en contacto directo. Se aplica en la agricultura para monitorear el estado y el comportamiento de los cultivos frente a diferentes variables, como enfermedades, plagas y rendimiento, según lo exponen Thao et al., (2022).

2.1.5. Imágenes multiespectrales (IM)

Las imágenes multiespectrales son aquellas que capturan datos de imágenes dentro de rangos de longitud de onda específicos a través del espectro electromagnético en distintas bandas espectrales. Estas permiten analizar la superficie terrestre, sus características físicas y biológicas por medio de distintos índices (Sobrino & Raissouni, 2000). Las IM cuentan con varias bandas (de tres a 20), las que se utilizaron en esta investigación son las siguientes NIR ($660 \text{ nm} \pm 26 \text{ nm}$), RE ($730 \text{ nm} \pm 16 \text{ nm}$), RED ($560 \text{ nm} \pm 16 \text{ nm}$) y G ($560 \text{ nm} \pm 16 \text{ nm}$), tomadas por cámaras de 5 MP (DJI, 2023).

2.1.6. Ráster

Un ráster es una estructura de datos (puede ser visualizada como una imagen) que representa una matriz rectangular de celdas (cada píxel de la imagen), donde cada celda tiene un valor que representa información, como la intensidad de color en una imagen, la elevación en un mapa topográfico o la reflectancia de un espectro. Los datos ráster son ampliamente utilizados en sistemas de información geográfica (GIS) y en el procesamiento de imágenes. Su formato más común es “.TIFF” (López-Rivera, 2021).

2.1.7. Ortofotos y ortomosaicos

Un ortomosaico se compone de varias ortofotos como si de un rompecabezas se tratara. Cada una de ellas, representa una imagen aérea georreferenciada, la cual ha sido corregida geométricamente, esto para representar las dimensiones reales de la superficie. El uso de estos mosaicos permite medir de manera más precisas, realizar inspecciones visuales y análisis detallados de zonas de interés (Carneiro et al., 2014).

2.1.1. Geoproceso

Un geoproceso es una operación que se realiza sobre datos geográficos para analizarlos, transformarlos o gestionarlos. Estas operaciones son fundamentales en los sistemas de información geográfica (GIS) y permiten manipular datos espaciales para obtener información útil y tomar decisiones informadas (Miñambres-Vidal, 2023).

2.1.2. Support Vector Machine (SVM)

Es un método de aprendizaje supervisado utilizado para clasificación. En el contexto de los rásteres, las SVM son especialmente útiles para la clasificación de imágenes, ya que pueden manejar datos segmentados y no segmentados. Funciona identificando un hiperplano que separa las clases de datos con el mayor margen posible. El proceso se realiza dentro de la función del programa ArcGIS PRO, la función “Classification Wizard”. Se selecciona un ráster a clasificar, posteriormente se agregan ejemplos de cada clase a clasificar, ya sea, nubes, suelo, vegetación, cultivos, árboles, etc. El programa tiene varios ajustes para una clasificación más personalizada, pero en el caso de esta investigación, todos quedaron por defecto (esri, s/f).

2.1.3. Índice de vegetación (IV)

Según Revelo Luna et al., (2020), se puede obtener a partir de las imágenes multiespectrales y se asocia al vigor de la planta y su estado de salud. Se calcula mediante operaciones aritméticas de las bandas espectrales en cada píxel de una imagen, además, refleja el contenido de clorofila, la fotosíntesis y la biomasa de la vegetación. Existen diferentes tipos de índices de vegetación, como NDVI, GNDVI, GEMI, NDVI_{re} entre otros. Por lo general los valores de los IV varían entre +1,0 y -1,0.

2.1.3.1. NDVI

El índice de vegetación por diferencia normalizada (NDVI) describe el verdor, la densidad y la salud de la vegetación para cada píxel de una IM. Se calcula como una relación entre los valores de reflectancia en el rojo y el infrarrojo cercano; las cuales son las longitudes de onda de la luz absorbida y reflejada por las plantas verdes (Eastman et al., 2013).

Las áreas de roca, arena o tierra suelen tener valores de NDVI muy bajos (por ejemplo, 0,1 o menos). Por su parte, la vegetación escasa como arbustos y pastizales o los cultivos en senescencia pueden tener valores de NDVI moderados (aproximadamente 0,2 a 0,5). Los valores altos de NDVI (aproximadamente 0,6 a 0,9) corresponden a una vegetación densa como la que se encuentra en los bosques templados y tropicales o los cultivos en su etapa máxima de crecimiento. Además, el NDVI es uno de los índices de vegetación más utilizados para el estudio de la fenología por teledetección remota (Brown, 2018).

2.1.3.2. *GNDVI*

El método del Índice de Vegetación Normalizado Verde (GNDVI) es un índice de vegetación para estimar la actividad fotosintética y es un IV comúnmente utilizado para determinar la absorción de agua y nitrógeno en el dosel vegetal (García-Cárdenas, 2018). Para esta investigación, al GNDVI ser una modificación del NDVI, posee los mismos rangos.

2.1.3.3. *GEMI*

El método del Índice de Monitoreo Ambiental Global (GEMI) es un IV no lineal para el monitoreo ambiental global a partir de IM. Tiene una semejanza con el NDVI, pero es menos sensible a los efectos atmosféricos. Se debe tener cuidado, ya que es influenciado por el suelo sin vegetación; por esto mismo, no es recomendado usarlo en áreas de poca vegetación (Pinty & Verstraete, 1992).

2.1.3.4. *NDVI_{re}*

El método Índice de Vegetación de Diferencia Normalizada en el borde rojo (NDVI_{re}) es un índice de vegetación para estimar la salud de la vegetación usando la banda del borde rojo (RE). Es especialmente útil para estimar la salud de los cultivos en las etapas medias a tardías del crecimiento, cuando la concentración de clorofila es relativamente mayor (Turan et al., 2019). Para esta investigación, al NDVI_{re} ser una modificación del NDVI, posee los mismos rangos.

2.1.4. *Muestreo por bloques (estratificado)*

El muestreo por conglomerados o por bloques, es una técnica que se utiliza para simplificar la logística de evaluación en extensas áreas de trabajo, como manejo forestal o agrícola. Esta

metodología implica la subdivisión de la población en grupos o bloques homogéneos. Cada bloque es representado proporcionalmente en la muestra, para asegurar que toda la población esté representada. Para formar la muestra final, cada representante debe ser elegido de manera aleatoria. Con este método, se pueden mejorar la precisión de las estimaciones, reduciendo la variabilidad al garantizar que se captura la diversidad de la población (Sáenz & Tamez, 2014).

2.1.5. Estadística Paramétrica (EP)

La EP se basa en suposiciones sobre la distribución de los datos, asumiendo que esta es normal. Con esto en mente, se encarga de estimar parámetros poblacionales y probar hipótesis basadas en este tema. Utiliza datos cuantitativos y requiere que se cumplan ciertos supuestos, como la normalidad, la homogeneidad de varianzas e independencia. Ejemplo de estas son las pruebas de correlación de Pearson, Spearman (que puede ser considerada no paramétrica en ciertos contextos) y la prueba de Kendall. (Bautista Díaz et al., 2020).

2.1.6. Estadística no Paramétrica (ENP)

La ENP se considera como métodos estadísticos que se utilizan cuando los datos no cumplen con los supuestos necesarios para las pruebas paramétricas, como la normalidad. Se utilizan para el análisis de datos ordinales o de razón, sin una distribución específica. Algunas de las pruebas más comunes son Kruskal-Wallis, prueba de signos y en correlación se pueden usar de igual manera las pruebas de Kendall y Spearman (Bautista Díaz et al., 2020).

2.1.7. Shapiro-Wilk

Este es un test empleado para contrastar la normalidad de los datos cuando la muestra no supera las 50 observaciones. Es equivalente a el test de Kolmogórov-Smirnov (Trabaja con más de 50 datos). El método consiste en ordenar de menor a mayor, y calcular la mediana y la varianza de la muestra. Se rechaza la H_0 de normalidad cuando $W < \alpha = 0,05$ (Flores-Tapia & Flores-Cevallos, 2021). Los cálculos de este test de normalidad fueron realizados en el programa “Statistical Package for the Social Sciences” (SPSS).

2.1.8. Coeficientes de correlación (CC)

Son herramientas bastante usadas en diversos campos científicos, siendo esencial para el desarrollo de áreas de conocimiento empírico, incluidas las ciencias sociales, económicas, de la salud, ingenierías, física y agrícolas, incluyendo muchas más (Ortiz-Pinilla & Ortiz-Rico, 2021). Ejemplificando el uso en áreas agrícolas, Barbosa et al., (2021) demuestran el uso de distintos CC para determinar la relación entre diversos IV y el índice de área foliar en cafetos. Los CC proporcionan un medio eficaz para establecer la relación entre dos variables a través de un único valor final. Sin embargo, cada coeficiente tiene sus propias limitaciones y contextos específicos en los que su uso es apropiado o necesario.

2.1.8.1. Coeficiente de correlación de Spearman

Es un CC que se basa en los rangos de los valores de las variables. Se encarga de evaluar la asociación directa o inversa entre dos variables. En ausencia de relaciones directas o inversas, los rangos de las variables tienen distribuciones uniformes discretas e independientes (Fiallos, 2021). Este CC dependiendo del caso puede ser el sustituto de Pearson, cuando hay una ausencia en la normalidad de los datos. La diferencia entre estos dos cae en que Spearman mide la tendencia monótona de la relación, mientras que Pearson mide la linealidad (Ortiz-Pinilla & Ortiz-Rico, 2021).

2.1.8.2. Coeficiente de correlación de Tau-b de Kendall

Este CC se refiere a la discrepancia entre el orden actual de una variable y otros posibles órdenes que podrían surgir naturalmente, lo que genera un perceptible acuerdo o desacuerdo con la muestra. Se trata de una medida no paramétrica de correlación de rangos categóricos. Existen tres tipos de Tau de Kendall, pero esta investigación se centra en el tipo 'b' ('a' es el clásico, sin modificaciones no es tan refinado; 'b' es una modificación, para número de clases iguales en ambas variables y 'c' es para número de clases diferentes entre variables), ya que presenta igual número de clases de clasificación en ambas variables a correlacionar, además, es menos afectada por la presencia de empates. Un par concordante es un conjunto en el que X e Y tienen el mismo orden relativo con respecto a la clasificación asignada, mientras que un par discordante es el caso opuesto y un empate,

es donde el orden no se ve afectado ya que los pares de la siguiente posición tienen el mismo orden que el par ordenado actual (Cuevas Urióstegui et al., 1994).

2.1.9. Potencia estadística (PE)

Es la probabilidad de que una prueba o análisis pueda detectar los efectos sobre una población, siendo capaz de encontrar diferencias significativas cuando estas son reales. También se considera como una herramienta para evitar caer en errores estadísticos. Existen dos tipos de errores a evitar con la PE en pruebas de significación estadística, el error tipo I (falso positivo), que ocurre cuando $p < \alpha$, por ende, se rechaza la H_0 cuando en realidad es verdadera, asumiendo erróneamente que (en este caso) existe una correlación directa entre las dos variables.

El segundo caso, el error tipo II (Falso negativo) que se refiere a la probabilidad (denotada como β) de no rechazar una hipótesis nula cuando en realidad se debería hacer, asumiendo que no hay (en este caso) correlación entre las dos variables cuando sí. Esto se puede deber a la elección no voluntaria de los casos atípicos de la población. La potencia estadística esperada convencionalmente para un análisis es del 80% ($1 - \beta = 0,80$). Esto significa que existe un 20% de probabilidad de aceptar la hipótesis nula cuando en realidad es falsa ($\beta = 0,20$). Se considera que un valor inferior implicaría un riesgo demasiado alto de cometer un error de Tipo II. La PE se calcula con tres datos: tamaño de la muestra (n), nivel de error (α) y tamaño del efecto (TE) (Cárdenas Castro, 2014).

2.1.10. Tamaño del efecto (TE)

Esta medida estadística permite entender la magnitud de una diferencia o relación entre dos grupos (el grado con que la H_0 es falsa). El TE es una forma de interpretar las diferencias o similitudes encontradas y compararlas entre distintos estudios, sin importar las variaciones en el diseño o los tamaños de muestra. Aunque la significancia bilateral de las dos colas (sig) de los Índices de correlación (IC) nos indica si la relación se debe o no en el azar, el TE nos permite conocer que tan importante es la correlación, si es pequeña, mediana o grande (como se muestra en la **Tabla A 1**) y su valor es igual al IC calculado (Rendón-Macías et al., 2021).

2.2. Antecedentes

El artículo de Caballero Beltrán (2017) expone una investigación sobre el uso de drones y el software AgVault para obtener los índices NDVI, GNDVI y GVI que permitan controlar el riego y la enfermedad en un cultivo. El propósito fue incrementar la eficacia y la rentabilidad de la agricultura de precisión, así como disminuir el daño ambiental. El autor muestra que los índices de vegetación son herramientas valiosas para identificar problemas de estrés hídrico, nutricional, enfermedades o problemas fitosanitarios en el cultivo, por medio de los drones; los cuales son una opción “económica y sencilla”, al ponerlos en perspectiva, para la obtención de datos espaciales. Además, se tiene como sugerencia el uso de esta tecnología para mejorar la gestión agrícola y optimizar los recursos.

La dinámica de los índices de vegetación NDVI, GNDVI y GVI en un cultivo de arroz de la variedad Fedearroz 2000 se estudió mediante el geoprocesamiento de imágenes aéreas multiespectrales obtenidas por drones o UAV con una cámara Canon S100 modificada (García-Cárdenas, 2018). El objetivo fue detectar las zonas del cultivo con diferentes condiciones de estrés, salud o densidad, y evaluar el desempeño de la cámara. El cultivo, ubicado en San José de Cúcuta, norte de Santander, Colombia, tenía un área de 4,1 ha. Se realizaron 12 vuelos con el dron de la marca Iris y se generaron mosaicos para el cálculo de NDVI, GNDVI y GVI para cada fecha. A partir de lo anterior, se observó una correlación positiva y fuerte entre el NDVI y el GNDVI. Se propuso un protocolo para el análisis de cultivos agrícolas con drones como herramienta de agricultura de precisión. Este estudio permite entender la dinámica de distintos índices de vegetación en condiciones similares a las que se pueden encontrar en Costa Rica.

Otro estudio de imágenes multiespectrales del satélite Landsat-8 por parte de Sánchez-Méndez y Arguijo-Hernández (2018) fue para detectar y evaluar los cultivos de café, así como sus problemas fitosanitarios en Veracruz, México. Mediante el análisis de componentes principales y la clasificación supervisada, se logra distinguir las zonas con plantaciones de café y sus niveles de desarrollo. El interés principal fue prevenir la roya; se revisaron algunos estudios que usan IV como el NDVI y el EVI para estimar la incidencia de la roya y advertir sobre su alta prevalencia a nivel global. Los hallazgos iniciales para el NDVI con los valores entre 0,2 y 0,8 indican

vegetación sana, mientras que los valores entre 0,4 y 0,9 señalan cultivos con buena cobertura y productividad.

El estudio realizado en el desierto de Thar (India y Pakistán) utilizando técnicas geoespaciales, este se encargó de evaluar las dinámicas estacionales de la vegetación. El estudio se enfoca en el análisis de los cambios de la vegetación a lo largo de 3 años, y como estos se veían afectados por las estaciones. Alguno de los IV fueron el NDVI, NDVI_{re}, WDRVI y WDRVI_{re}. El Red-Edge NDVI (NDVI_{re}) es un índice útil para ubicar el área de vegetación en estrés (0,3 y 0,6) y lograr una gestionar la productividad agrícola de manera eficiente (Singh et al., 2021).

En el río Tarim, China, Se realizó la evaluación de varios IV para estimación de la cobertura fraccional de la vegetación fotosintética (fPV) y no fotosintética (fNPV), esto gracias a imágenes satelitales. El índice que tuvo mejor correlación fue el GEMI ($R^2 = 0,59$), en comparación con los demás índices probados (NDVI, NDI, RVI y otros), Este índice, junto con el índice de combustible muerto (DFI) para la vegetación no fotosintética, permitió analizar la variación espacial y estacional de la vegetación en la zona baja del río Tarim. La investigación generó un modelo (GEMI-DFI), el cual demostró ser capaz de identificar la variación de la vegetación en áreas áridas, proporcionando una herramienta de alto valor para la conservación de las áreas naturales (Guo et al., 2021).

El estudio de los investigadores Marin et al. (2021), en Brasil, se enfocó en estudiar la severidad de la roya del café por medio de imágenes de un dron, junto a modelos de aprendizaje automatizados basados en árboles de decisión. Este estudio evaluó un total de 63 IV, alguno de ellos fueron el GNDVI, SAVI, NDVI, RERDVI entre muchos más. Los modelos de árboles de decisión, como el "Logistic Model Tree" (LMT), demostraron su efectividad para predecir la severidad de la roya en distintas etapas (específicamente en la temprana y última etapa del hongo). Por este medio se detectaron que el NDVI, NDRE, y el GNDVI fueron los índices más sensibles a la roya con una precisión de 0,672, 0,637 y 0,578 respectivamente. Este estudio genera un gran aporte para la agricultura de precisión y la detección de enfermedades en el café.

El uso de la banda visible (RGB) para monitoreo de los cultivos se puso a prueba durante un año en el documento de Barbosa et al. (2021), el cual obtuvo buenos resultados. Se generaron los IV

MPRI y el GLI, que presentaron una correlación débil con respecto al índice de Área Foliar. A pesar de ello, si son útiles para percibir malezas y variables en el cultivo del café. Esta herramienta permite a los agricultores optimizar los recursos durante los ciclos de producción.

En Costa Rica, se utilizó tecnología de drones para capturar imágenes hiper-espectrales y térmicas con el objetivo de estimar los índices NDVI y EVI del café. El estudio, realizado por Jin et al., (2021), buscaba evaluar la productividad y la eficiencia en el uso del agua (WUE) del café tras la aplicación de “biochar” (biocarbón) en el suelo, un material que mejora la retención de agua y nutrientes. Los resultados mostraron una mejora significativa en la producción del cultivo en comparación con años anteriores sin uso de “biochar”. Además, se demostró el potencial de las imágenes hiper-espectrales y térmicas para cuantificar los efectos del “biochar” en el cultivo del café, evidenciando un aumento en los índices NDVI, EVI, la productividad primaria bruta (GPP) y la WUE. Este estudio resalta la utilidad de estas herramientas para monitorear la salud de las plantas y entender la evolución de los índices NDVI y EVI en diferentes cultivos a lo largo del tiempo.

Dentro de la región de centro América, Guatemala, se realizó un estudio que da un contexto de la enfermedad del hongo de la roya (*hemileia vastatrix*) y como ésta tuvo un aumento de 0.70% con respecto al mismo instante de un año atrás (5,17%). El informe dirigido por las entidades ANACAFÉ y CEDICAFÉ (2022) se realizó mediante un muestreo de varias parcelas en el país, el cual fue hecho por medio de una aplicación de su propiedad (Coffee Cloud). Al realizar contrastes de información, se pudo determinar que este cambio se debía a condiciones meteorológicas. Como recomendación de esta investigación, incita a reforzar la aplicación de funguicidas si se ve un cambio climático (aumento de lluvias).

En Sultepec, estado de México, México, Ramírez Dávila et al. (2023) realizó una investigación sobre la distribución espacial de la roya, en seis parcelas experimentales, por un periodo de seis meses. Por medio de geoestadística se pudo identificar y predecir el comportamiento de la enfermedad, esto gracias a la relación con su ciclo biológico y la fenología del cultivo. Esta investigación permite generar nuevas estrategias de manejo fitosanitarios, reduciendo el impacto de la enfermedad.

Chedid et al. (2024) expresa en su estudio realizado en Brasil, como se comparan distintos IV (NDVI, SAVI, ARVI, EVI y VDVI) durante un año, y como estos evalúan el estado vegetativo del cultivo del café. Por medio de la plataforma Planet, se llevó a cabo el estudio en una finca comercial. Los índices seleccionados fueron analizados utilizando estadísticas descriptivas, correlación de Pearson, propuestas de clasificación/interpretación y el índice Kappa. Como resultados de la investigación, los cuatro IV mostraron ser eficientes, con un índice de Kappa cercanos a 90% a excepción de VDVI. Este estudio concluyó que el uso de IV es una alternativa rápida y práctica para el monitoreo de los cultivos de café, dando la oportunidad de poder detectar enfermedades y plagas de forma remota.

El estudio que se llevó a cabo en la finca experimental Santa Lucía en Heredia, Costa Rica, se centró en evaluar la incidencia de la roya en relación con condiciones climatológicas, y aplicaciones de funguicida. Se encontró que el uso de funguicidas puede generar buenas resistencias al hongo, manteniendo una incidencia menor al 7%, esto a pesar de que las condiciones apoyaran al desarrollo de la enfermedad (alta precipitación y altas temperaturas). Calero Ocón (2024) concluyó en su investigación que el uso adecuado de sombras puede ayudar a reducir la incidencia de la roya. Además, esta investigación proporciona información valiosa para el manejo de la roya del café, y la importancia de tomar en cuenta las condiciones climáticas para programar las pulverizaciones.

En el otro lado del planeta, en Grecia, los autores Fan et al. (2024), decidieron generar un nuevo IV llamado “Vegetation Anomaly Spectral Texture Index” (VASTI), que combina características espectrales y texturales para mejorar la detección de áreas de vegetación quemada. Esto permite generar mejor detección de áreas vegetales quemadas. En la investigación, se compararon varios IV, entre ellos destacan GEMI, EVI, NDVI y GNDVI. Estos índices fueron fundamentales para el monitoreo de la vegetación, permitiendo evaluar la salud y densidad vegetal por medio de imágenes satelitales.

Para la detección de vegetación quemada algunos IV suelen tener limitaciones, ya que puede ocurrir una confusión en el espectro. Sin embargo, la incorporación del GEMI ayuda a reducir la confusión espectral entre diferentes tipos de cobertura terrestre, mejorando la exactitud de la detección de áreas quemadas. Como resultado de esta investigación, que incluyeron un conjunto

de datos global con 1774 pares de muestras de 10 sitios diferentes, el IV generado por Fan et al. (2024), VASTI, supera estas limitaciones al incorporar características texturales. Esto permite una mejor discriminación de las áreas afectadas por incendios, esto gracias a la integración de varios IV en su cálculo.

2.3. Marco de referencia empresarial

La empresa con la que se consolidó la investigación fue Starbucks. Esta es una multinacional que se dedica a la producción y comercialización de café y otros productos relacionados. Su origen se remonta a 1971, cuando abrió su primera tienda en Seattle, Estados Unidos. Desde entonces, ha experimentado un crecimiento exponencial, llegando a tener más de 30,000 establecimientos en más de 80 países.

Starbucks se caracteriza por ofrecer una experiencia de consumo diferenciada, basada en la calidad de sus granos de café. Además, se ha posicionado como una empresa socialmente responsable, que apoya el desarrollo sostenible de las comunidades cafetaleras, mediante programas de comercio justo, certificación orgánica y apoyo a la conservación ambiental. Así, contribuye a la industria, tanto desde el punto de vista económico como social, ambiental y, en este caso, en pioneros de nuevas tecnologías, siendo un referente para otras empresas del sector.

La Hacienda Alsacia, es una finca cafetalera ubicada en el distrito de Sabana Redonda, cantón de Poás, provincia de Alajuela, Costa Rica. Esta Hacienda pertenece a Starbucks. En 2013, la empresa adquirió la Hacienda con una extensión de 240 ha, donde se cultivan diferentes variedades de café arábica bajo un sistema agroforestal.

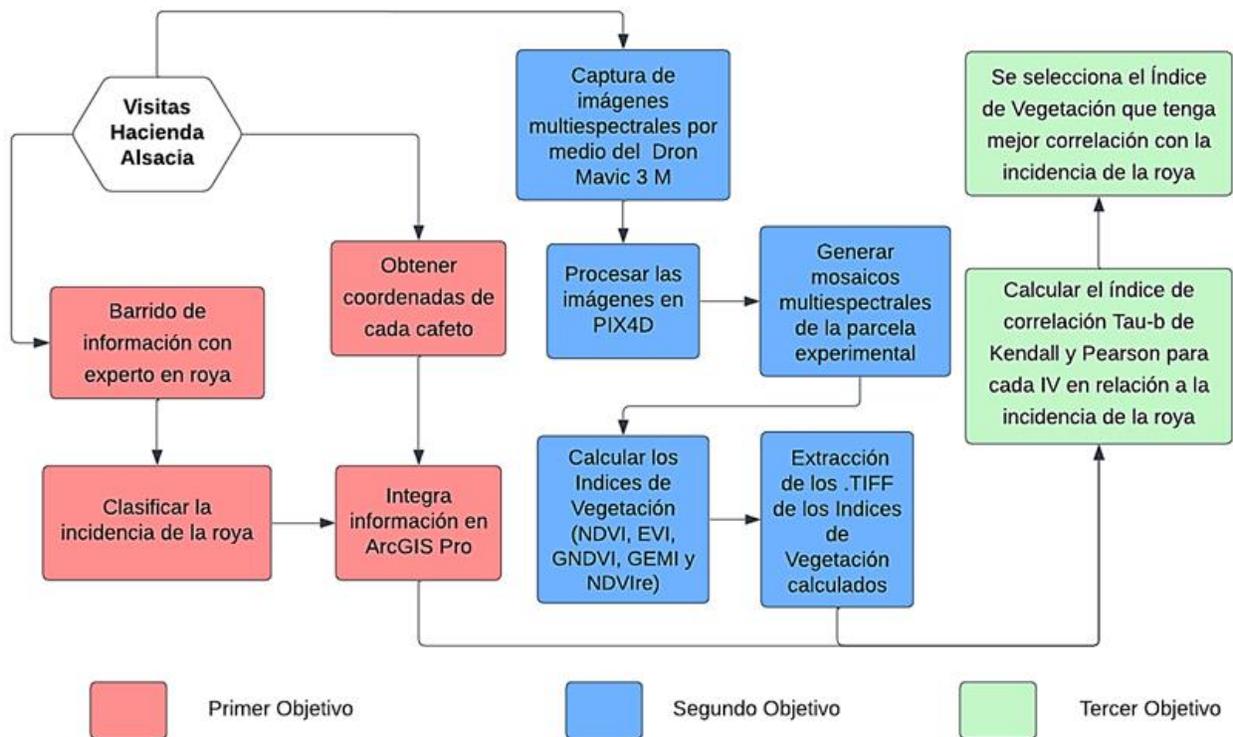
3. METODOLOGÍA

El primer objetivo de la investigación consistió en obtener la incidencia de la roya de la parcela a estudiar, para ello, por medio de un muestreo por bloques, se realizó un barrido de información geolocalizado con la ayuda de criterio experto. Esta fue clasificada en niveles según la incidencia del hongo en la planta.

Para el segundo objetivo, se utilizó un dron DJI Mavic 3M y el software PIX4D para realizar un mosaico con las imágenes multiespectrales tomadas el sitio de estudio. Estas imágenes se sometieron a los diferentes IV (NDVI, GNDVI, GEMI y NDVIre) en la herramienta ArcGIS Pro.

Para finalizar con el tercer objetivo, cada índice de vegetación se correlacionó con la incidencia de la roya clasificada en el primer objetivo, y se determinó el mejor índice para el caso de la roya en el café mediante CC (Spearman y Tau-B de Kendall). Se puede revisar el diagrama de flujo de esta metodología en la **Figura 1**.

Figura 1. Diagrama de flujo dividido por objetivos.

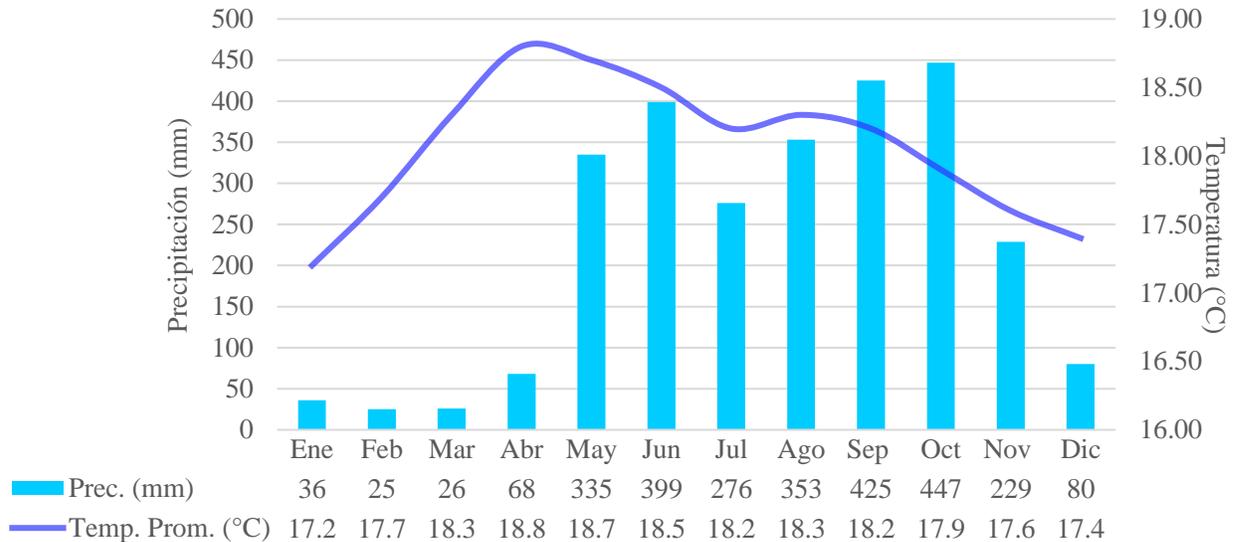


3.1. Sitio de estudio

Se estudió una parcela de 2,3 ha con cafetos arábigos en la zona norte de la Hacienda Alsacia, Sabanilla, Alajuela. Ubicada en las coordenadas 10,1053, -84,1972 a una altura de 1248 msnm. La pendiente del terreno es de 15 ° y disminuye de Noroeste al Sureste. Los cafetos predominan con edades mayores a los cinco años. El clima de la región consta de un promedio anual de 18 °C con mínimos de 14 °C y máximos de 22 °C (**Figura 2**). La precipitación de la zona es mayor en los

meses desde mayo hasta noviembre, en la cual se dan las condiciones para que la roya se genere (Climate Data, 2024).

Figura 2. Temperatura y precipitación de la zona



Nota: El gráfico fue generado con la información de Climate Data (2024), con información de 20 años, 1999-2019.

Para el cultivo del café se requieren unos aspectos agroclimáticos claves, como los mencionados en la sección 2.1.2, de los cuales se resalta que esta ubicación cumple con los requisitos de altura, temperatura, y sombra (en la mayoría de las zonas). Los aspectos de precipitación anual (2700 mm) y humedad relativa anual (86%) se encuentran un poco por encima de los valores límites deseados (2000 mm y 80% respectivamente) para un cultivo exitoso. Estos valores propician el crecimiento de distintas enfermedades fúngicas, como la roya.

Diseño experimental

3.1.1. Objetivo 1: Clasificar espacialmente la incidencia de la roya en cafetos en la parcela experimental de la Hacienda Alsacia

3.1.1.1. Inspección de la hacienda

Desde mayo del 2023 hasta noviembre del 2023, se llevaron a cabo tres visitas en la Hacienda Alsacia con el propósito de explorar sus instalaciones y familiarizarse con la ubicación de la

parcela experimental. En cada visita, se sostuvieron reuniones con Carlos Mario Rodríguez, director del departamento de Investigación y Desarrollo Global de la empresa Starbucks, así como con Sara Bogante Vargas, agrónoma de “Farmer Support Center” (FSC) en Costa Rica y responsable de la gestión agrícola en toda la Hacienda. Este proceso permitió obtener un conocimiento más profundo sobre la Hacienda y establecer conexiones clave para el desarrollo de la investigación. Se realizaron diversas actividades y reuniones con el personal para determinar la magnitud de esta investigación, alguna de ellas se puede observar en la **Figura 3**. Por ejemplo, dentro de las actividades, se percibió la roya, se trabajó con la frutilla del café y se dimensionó la Hacienda.

Figura 3. Exploración de la Hacienda Alsacia.



Nota: imágenes tomadas en las primeras visitas, se percibió la roya, se trabajó con la frutilla del café y se dimensionó la Hacienda.

La visita inicial proporcionó una perspectiva clara sobre el método más efectivo para la extracción de información. En consecuencia, se determinó que un agrónomo debería encargarse de la clasificación de la severidad de la roya en las hojas de los cafetos. Con el objetivo de llevar a cabo un sondeo eficiente, estructurado y evitando errores, se consideraron múltiples alternativas, resultando seleccionado la aplicación móvil Survey 123. Esta herramienta forma parte de la familia ESRI, desarrolladora de ArcGIS Pro, la cual permite generar encuestas georreferenciadas.

3.1.1.2. Clasificación de datos

Los cafetos se clasificaron según la severidad del hongo presente en las plantas, para posteriormente calcular la incidencia de la roya en la parcela experimental. Cada cafeto fue clasificado según la roya presente en tres de sus hojas, con el fin de tener una clasificación final

promedio para cada planta seleccionada. Esta fue llevada a cabo según la metodología ya aplicada en países de Norteamérica como México (SENASICA, 2019) e indicada en la **Figura A 1**. A continuación, en la **Figura 4**, se muestra un ejemplo de hojas de cafeto clasificadas en la parcela experimental. Como se puede observar de izquierda a derecha las hojas se clasificaron con una severidad del hongo presente en la hoja de niveles 3, 4 y 6, respectivamente.

Figura 4. Hojas de cafeto de la parcela experimental de la Hacienda Alsacia.



Nota: Ejemplo de la hoja del cafeto para ser clasificada según la severidad de la roya

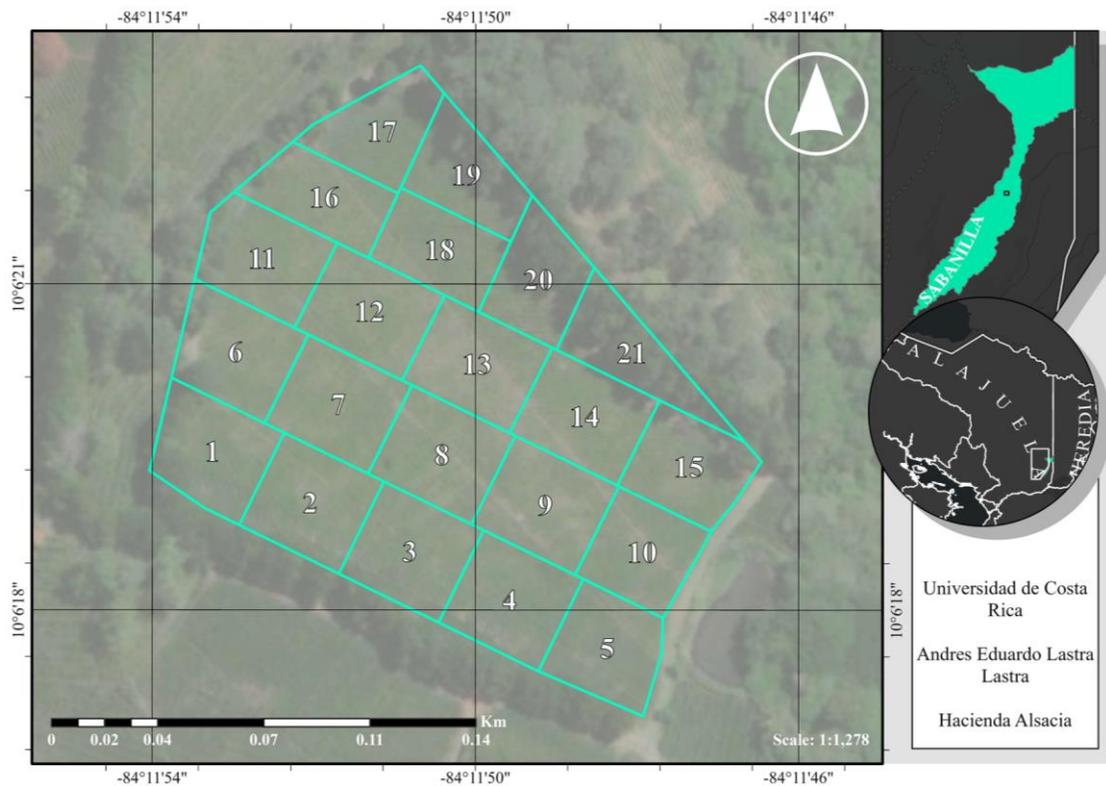
Durante la primera visita, el Ing. Jorge Mora Monge impartió una capacitación práctica sobre la detección del hongo de la roya y explicó en qué estrato del cafeto suele encontrarse, específicamente en la zona media. El ingeniero acompañó el proceso de clasificación de los cafetos según la severidad de la roya (**Figura A 1**). El Ing. Mora es egresado de la Escuela de Agronomía de la Universidad de Costa Rica y cuenta con más de 15 años de experiencia en el campo del café.

3.1.1.3. Diseño del muestreo

Se usó el muestreo por bloques (conglomerados) descrito en el apartado 2.1.4, este consiste en seleccionar elementos consecutivos en una población como muestra representativa. Este método es eficiente cuando la población es muy grande y dispersa, ya que permite estimar la precisión de los resultados de la muestra con base en la variabilidad entre los bloques, con una inversión menor de recursos (Sáenz & Tamez, 2014). Junto a la evaluación realizada, se tomaron las coordenadas GPS del punto evaluado, el cual fue de suma importancia para el muestreo por bloques.

El barrio de información fue realizado desarrollando el tipo de muestreo seleccionado. Para ello, se debió hacer una segmentación de la parcela en los bloques a estudiar. Por medio de ArcGIS Pro, se generó un polígono, delimitando la zona de trabajo, de la siguiente manera: primero, se genera un nuevo “shapefile”, con una geometría de tipo polígono. Seguidamente, en la pestaña de editar, se utilizó la herramienta de polígono, dentro de la sección de crear, esta permitió generar la delimitación de la parcela experimental. Ya con la zona a delimitar, se debió de generar la división en bloques de igual tamaño, para ello, volviendo a la pestaña Editar, se utilizó la herramienta dividir. Al aparecer la nueva ventana emergente, se dividió el polígono, generando 21 bloques, en este caso en seis bloques con áreas de 1000 m² y 15 bloques con áreas de 1200 m² (ver en **Figura 5**).

Figura 5. Distribución de bloques en la parcela experimental.



Es importante destacar que los bloques de mayor tamaño se localizaron en la zona este de la parcela, caracterizada por una mayor densidad de árboles que rodean la parcela y por ser una zona más angosta. Estas condiciones dificultan la recolección de datos. Sin embargo, el hecho de contar con una superficie mayor resulta beneficioso para la actividad. Se debe señalar que el bloque 21 se

descartó después de la primera captura, debido a que en las imágenes tomadas con el dron (**Figura C 2**) de este bloque estaba completamente oculto por el dosel arbóreo, impidiendo así la extracción de información para su comparación.

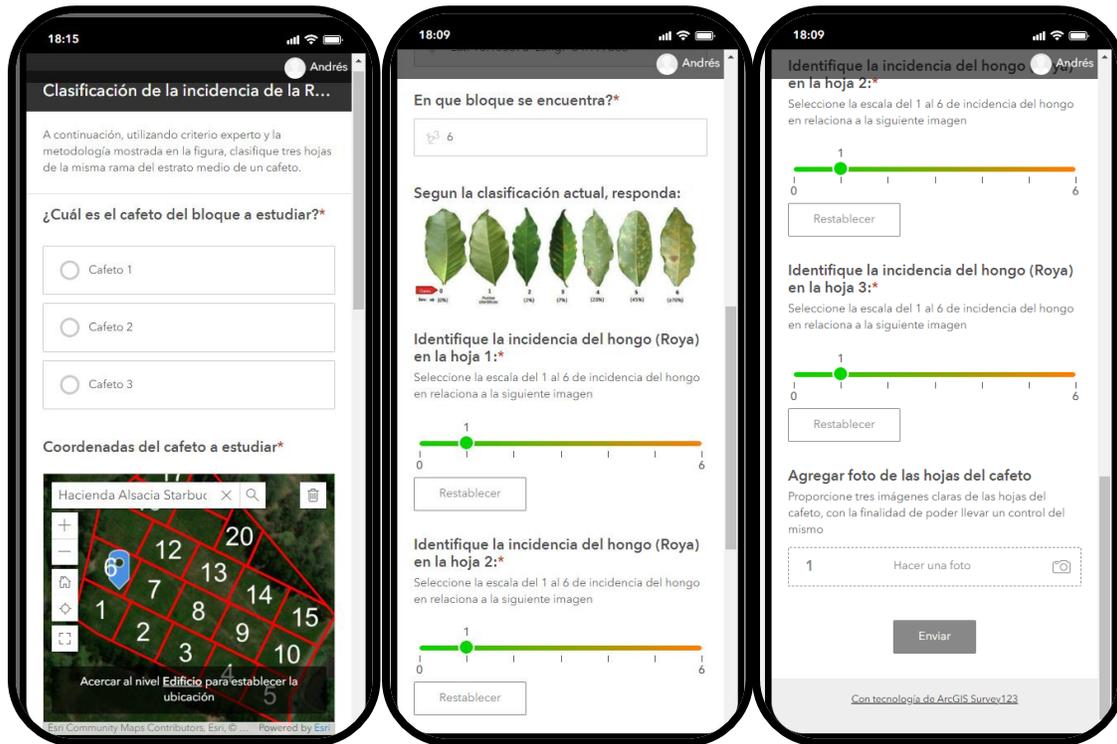
3.1.1.4. Muestreo de la toma de datos

La severidad del hongo hace referencia a el área de la hoja cubierta por el mismo, mientras la incidencia de la roya expresa que tan presente se encuentra la enfermedad en un área definida. Por ende, de ahora en adelante, este es el término empelado en toda la investigación cuando se quiere hacer referencia de la presencia del hongo en los bloques estudiados.

El muestreo constó de tres cafetos (unidad experimental) por cada uno de los bloques de la **Figura 5**. En cada uno de ello se realizó la clasificación a tres hojas del estrato medio de la planta que pertenezcan a la misma bandola (metodología aplicada actualmente en la Hacienda, **Figura A 2**), para darle un valor promedio al cafeto. Teniendo un total de 60 cafetos por toma de datos (tres cafetos por 20 bloques, ya que el bloque 21 es descartado por presencia de árboles). Se debe aclarar que actualmente la metodología aplicada en la hacienda (metodología OIRSA) es similar en la identificación, (2 bandolas por cafeto, medio superior y medio inferior). En este caso varía espacialmente, ya que es ir parcela por parcela observando cafetos al azar, y según el criterio experto del agrónomo de la empresa, se decide pulverizar o no ese día, repitiendo esto para todas las parcelas, proceso que consume mucho tiempo (Avelino et al., 2019).

Para la toma de datos se desarrolló, en la plataforma ya mencionada Survey123 de ArcGIS, una encuesta, la cual permitió extraer las coordenadas del cafeto a evaluar, junto a la clasificación de tres hojas del estrato medio según la metodología explicada. Además de esto, le da al usuario, desde su celular, la opción de subir fotos para tener un respaldo de la clasificación, todo esto almacenado en la web y fácil de exportar a hojas de cálculo (**Figura B 1**). Con el objetivo de recopilar la información más relevante y necesaria para esta investigación. Las preguntas diseñadas para la encuesta pueden ser revisadas en el Apéndice B. El polígono mostrado en la **Figura 5**, se cargó a la herramienta de Survey 123 para utilizarlo dentro de los mapas como se aprecia en la **Figura 6**.

Figura 6. Encuesta de campo para la clasificación de la incidencia de la roya en café en la plataforma Survey123 de ArcGIS.



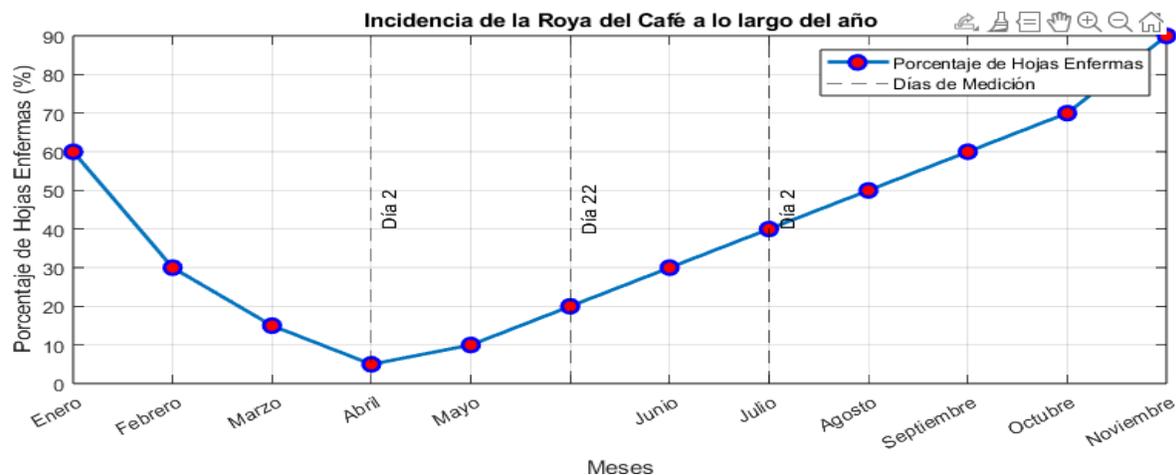
3.1.1.5. Toma de datos

Se realizaron tres toma de datos de la misma parcela experimental en diferentes épocas del año, con el objetivo de obtener un muestreo representativo de la afectación del hongo en sus etapas más tempranas del desarrollo (ANACAFÉ y CEDICAFÉ, 2022). Con la herramienta lista, se realizaron tres visitas a la hacienda Alsacia, en las fechas siguientes: 2 de abril del 2024 (época seca), 22 de mayo del 2024 (comienzo de época lluviosa) y el 2 de julio del 2024 (época lluviosa) como se muestra en la **Figura 7**. Las fechas fueron separadas mes y medio de cada una para tener una diferencia más significativa en el tiempo entre cada muestra. Para cada toma de datos, se visitó la parcela experimental con el apoyo del director de esta investigación, Ing. Matías Chaves, M.Sc. y un agrónomo experto en la roya del café, el Ing. Jorge Mora Monge.

En el contexto de esta investigación la selección de estas fechas permitió contrastar los cambios en una parcela experimental con una presencia baja de roya (segunda toma de datos) y la misma área con una presencia mayor (tercera toma de datos). Aunque el estudio no fue llevado a cabo

durante los picos de infestación (lo cual requeriría un estudio prolongado de nueve a 12 meses) se buscó cumplir con las necesidades actuales y solicitud de la empresa Starbucks.

Figura 7. Fechas de toma de datos con respecto a la presencia del hongo.



Nota: Esta figura fue elaborada en base a los datos de ANACAFÉ (2022).

En el contexto de esta investigación la selección de estas fechas permitió contrastar los cambios en una parcela experimental con una presencia baja de roya (segunda toma de datos) y la misma área con una presencia mayor (tercera toma de datos). Aunque el estudio no fue llevado a cabo durante los picos de infestación (lo cual requeriría un estudio prolongado de nueve a 12 meses) se buscó cumplir con las necesidades actuales y solicitud de la empresa Starbucks.

La empresa buscaba la capacidad de identificar plagas en sus primeras etapas de aparición (**Figura 7**). Esta necesidad surge por razones económicas y de eficiencia laboral, con el objetivo de implementar un control riguroso de inmediato. La intención es no esperar hasta que la plaga alcance su punto máximo a finales de año para tomar acciones, ya que esta situación ha generado problemas económicos en el pasado. Se estableció la resolución temporal de 1,5 meses, debido a que es necesario percibir un cambio en la parcela y limitaciones económicas.

Finalmente, la información de cada toma de datos es extraída de la plataforma y trabajada con tablas dinámicas en Excel, con el fin de conseguir los promedios y desviación estándar de cada parcela, con la cual se pudo calcular los CC en el tercer objetivo.

3.1.2. Objetivo 2: Recolectar imágenes multiespectrales para el cálculo de cuatro diferentes índices de vegetación que indican el estado de vigor y salud de la planta

3.1.2.1. Muestreo de toma de datos

Para el cálculo de los IV fue fundamental contar con información de los espectros visibles y no visibles (infrarrojo). En este caso, utilizando la cámara del dron Mavic 3M y su radio control (RC Pro), fue posible capturar la información necesaria. El RC Pro permite delimitar el área de estudio y configurar, según las necesidades del proyecto, diversos parámetros como: altura del dron a 100 m sobre el suelo (1248 msnm), solapado de imágenes (70%), uso de cámaras (RGB + MS), ángulos del recorrido y velocidad de desplazamiento (por defecto) (DJI, 2023). Estas configuraciones permitieron trabajar con una distancia de muestreo del suelo (GSD) de 4 cm/píxel, lo cual proporcionó una resolución espacial adecuada para esta investigación.

En las visitas realizadas a la hacienda se realizaron varios vuelos de prueba, con la finalidad de delimitar el área a trabajar. Como primer contacto en el sitio de estudio, se decidió hacer un mapeo del 100% de la hacienda (aproximadamente 200 ha). Este proceso tenía una duración de 54 min, el cual se extendía a dos horas ya que a los 25 min las baterías debían ser remplazada. Como resultado de este proceso se extrajo la información multiespectral de toda la hacienda. Este proceso presentaba diversos inconvenientes, entre los cuales se incluye la necesidad de procesar más de 50 gigabytes de fotografías por toma de datos, lo que requería una considerable capacidad de cómputo y almacenamiento.

Al tener la parcela experimental seleccionada por Don Carlos Mario Rodríguez (CEO de la empresa), se delimitó la zona de trabajo, que abarcaba 2,30 ha. Utilizando la aplicación DJI4 en el controlador, se creó una plantilla de la parcela y se configuró el dron para iniciar los sobrevuelos. Esta tarea tuvo una duración de 15 minutos ininterrumpida, generando un total de cuatro gigabytes de datos, un tamaño manejable para transferir entre dispositivos y realizar copias de seguridad en servicios en línea. Se realizó el levantamiento por parte de los operarios Ing. Matías Adrián Chaves Herrera, M.Sc. (director de esta investigación) y Andrés Eduardo Lastra Lastra, Ing. (autor) en tres momentos en el tiempo. En cada momento se asignaron la misma configuración para tener una homogeneidad en los datos. Este proceso se debe calibrar cada vez que se va a realizar, pero para

el momento que esta investigación fue llevada a cabo, la EIB no contaba con este equipo. El calibrador consta de varias placas con un código QR, el cual el programa puede detectar para hacer las debidas correcciones.

3.1.2.2. *Procesamiento de imágenes*

Las imágenes capturadas por el dron cuentan con distintos metadatos como geolocalización, fecha, hora y etiquetas especiales en el nombre de cada archivo, con el fin de que el programa pueda identificar en dónde, cuándo y con cuál cámara fue capturada cada imagen, generando correcciones y clasificaciones. El programa PIX4D cuenta con la capacidad de poder clasificar cada imagen automáticamente en los espectros a trabajar: GREEN, NIR, RED, RED EDGE Y RGB. Además de identificar el sistema de coordenadas utilizado, en el caso de esta investigación se utilizó datum WGS84. El software tiene varias aplicaciones en el campo de la agricultura, como nube de puntos, renderizados en 3D, cámara de calor y procesos multispectrales, generando ortofotos.

El software provee una alta variedad de geo procesos, el seleccionado para esta investigación fue el “Arg Multispectral”, esto debido a que es el proceso que permite generar ortomosaicos para cada banda trabajada. La versión gratuita del software permite realizar un procesamiento básico de los datos, incluyendo el reajuste de tomas y la generación de mallas de datos. Sin embargo, para la generación de ortomosaicos, es necesario utilizar la versión de pago, la cual fue proporcionada por la distribuidora para este proyecto.

Al tener todas las imágenes cargadas, se debió ajustar varios parámetros dentro de la configuración del proceso. Para esto, se seleccionó la opción de DSM, Ortomosaicos, y las siguientes opciones:

- Ráster DSM: casillas “GEO TIFF” y “Merge Tiles”.
- Ortomosaicos: casillas “GEO TIFF”, “Merge Tiles”, “GEO TIFF WITHOUT TRANSPARENCY” (generara una sola banda por espectro por ráster).

Al tener ya la configuración ajustada, se procedió a la compilación de las imágenes. El programa realizó tres procesos (validación de datos, nube de puntos y generación de los ortomosaicos) para completar la tarea. Este proceso utiliza de manera demandante la CPU como la GPU del ordenador y tiene una duración aproximada de una hora (varía según las cualidades del equipo). Suele tomar de uno a dos intentos que el proceso se cumpla satisfactoriamente. Ya con el proceso culminado

en Pix4D, da como resultado cinco archivos tipo ráster, uno por cada espectro tomada (uno de ellos contara con tres bandas: verde, rojo y azul). Este último es de sumo cuidado para los procesos a realizar más adelante, ya que puede llevar a confusión.

Estas cinco imágenes (**Figura A 14**) o rásteres fueron importadas al programa ArcGIS Pro, y se trataron cada una como bandas. Para el cálculo de GEMI, como se puede apreciar en la **Tabla A 2**, se requiere la banda NIR y RED. El obstáculo está en que este índice está diseñado para imágenes satelitales específicamente, y no es sensible a efectos atmosféricos, eliminándolos por medio de un ajuste llamado eta (Pinty & Verstraete, 1992). Al tener una imagen de Dron, sin presencia de efectos atmosféricos (cúmulos, tormentas, polvo y la atmósfera), los resultados llegan a valores de 10^6 , saliendo del rango de estudio de -1 a 1.

Esto conllevó hacer una modificación de estas bandas, exclusivamente para el cálculo de este índice. El proceso se realizó por medio de la herramienta “Ráster Calculator”, la cual permite hacer operaciones con cada uno de los pixeles presentes en el ráster, en este caso, ambas capas fueron divididas por 10^4 . Se obtuvo como resultado dos rásteres modificadas.

Para continuar con el proceso, se realizó la composición de bandas por medio del geoprocso “composite bands”, dando como resultado un solo ráster que contenga las siete bandas (cinco rásteres en uno solo). El orden en este paso es de suma importancia ya que los cálculos son con el número de banda y esta debe coincidir con la necesaria para el IV. Se debe tener en cuenta que el ráster de RGB presenta tres bandas dentro del mismo ráster. En paralelo, las bandas modificadas también fueron compuestas en una sola, y su uso fue exclusivo para el cálculo del IV GEMI, pero todos los procesos fueron iguales. En la **Tabla 1** se puede observar el orden que tuvieron las bandas en los nuevos rásteres. El diagrama del proceso seguido se encuentra en la **Figura A 13**.

Tabla 1. Orden de las bandas generadas por PIX4D para ser compuestas en ArcGIS Pro.

Numero de banda	Bandas sin ajustar	Bandas ajustadas
1	Green	NIR (Ajustada)
2	RED (RGB)	RED (Ajustada)
3	GREEN (RGB)	-
4	BLUE (RGB)	-
5	NIR	-

Numero de banda	Bandas sin ajustar	Bandas ajustadas
6	RED	-
7	RED EDGE	-

Existen diversas maneras de hacer la extracción los IV para cada bloque. En este caso se realizó por la herramienta “Zonal Statistics as Table”, el cual permite generar estadísticas de los ráster dentro de un polígono delimitado. Si se calculan los IV directamente sobre todo el terreno, los valores del suelo y arboles pueden perjudicar el dato, disminuyendo el valor presente a los cafetos. Para solucionar esto, se generó una máscara previa a todo el proceso de cálculo de los IV, la cual solo contempló los cafetos, esto reduce el error y la carga computacional del proceso.

La máscara se generó utilizando la herramienta “Classification Wizard” y el ráster de “Bandas sin ajustar”. En su configuración supervisada y basada en objetos, se dejaron los ajustes por defecto. Con la capa de “Bandas sin ajustar” generada previamente, se clasificaron varias zonas de la imagen para entrenar el proceso, enfocándose principalmente en “Plantaciones”, que corresponde a los cafetos de la parcela experimental. Se seleccionó el método de clasificación “Support Vector Machine” con sus parámetros por defecto, obteniendo un archivo ráster con tres clasificaciones del terreno.

Posteriormente, se extrajo únicamente la clasificación de “Plantaciones” mediante el geoprocso “Extract by Attributes”, estableciendo el polígono de los bloques como límite en el apartado de “Environments”. Finalmente, se obtuvo la máscara de la zona de interés, correspondiente a los cafetos de la parcela experimental (**Figura A 3**). Esta mascara fue evaluada por medio de una matriz de confusión, para determinar su coeficiente de Kappa, y determinar si la clasificación fue realizada de la manera correcta. En la **Tabla 2**, se puede apreciar cómo se obtuvo una buena correlación Kappa para las tres mascarar (una para cada toma de datos). En la **Tabla A 12**, **Tabla A 13** y **Tabla A 14** se encuentra desplegada la matriz de confusión para cada una de las máscaras realizadas.

Con esta máscara, se filtraron los ráster de Bandas y Bandas ajustadas utilizando el geoprocso “Extract by Mask”. Al contar los rásteres filtrados con los objetos de interés, la herramienta “Band Arithmetic Properties” permitió realizar el cálculo de los IV. Ubicada en la pestaña “Analysis”, en

el apartado “Raster”, dentro de “Raster Functions”. Al seleccionar esta opción, se desglosa una nueva ventana donde se buscó la herramienta deseada en la categoría “Math”.

Tabla 2. Coeficiente Kappa para la clasificación realizada en las mascararas de las tres tomas de datos.

Toma de datos	Kappa	Error estándar	I.C 95%	Fuerza de la concordancia
Toma 1	0,76	0,06	(0,65,0,88)	Buena
Toma 2	0,92	0,04	(0,85,0,99)	Muy buena
Toma 3	0,73	0,06	(0,61,0,85)	Buena

Nota: Los cálculos fueron generados en la herramienta “Compute Confusion matrix” en el programa ArcGIS Pro.

Dentro de ella, la herramienta solicita el índice que se desea calcular (NDVI, GEMI, GNDVI y NDVIre), el ráster de Bandas (o Bandas Ajustadas que será usado solo para GEMI) y los números de las bandas (**Tabla 1**) solicitadas según el IV seleccionado (para una mejor visualización del proceso, revisar la **Figura A 4**). Esta herramienta utiliza los cálculos presentes en la **Tabla A 2**. Con los IV calculados por cada pixel para los cafetos de la parcela experimental, se generó el promedio, desviación estándar, máximos y mínimos de cada índice para cada bloque con la herramienta ya mencionada “Zonal Statistics as Table”. Lo cual da un valor de [-1,1] para cada bloque, en cada toma de datos. Cada IV puede representar la salud y vigor de la planta en diferentes rangos. Sin embargo, todos comparten una característica común: cuanto más cerca esté el valor de +1, mayor será la densidad, vigor y salud de las plantas, mientras que valores por debajo de cero y cercanos a -1 indican superficies no vegetadas.

3.1.3. *Objetivo 3: Seleccionar el índice de vegetación que tenga la mejor correlación con la incidencia en etapas tempranas de la roya en café por métodos estadísticos*

3.1.3.1. *Coeficientes y la correlación de la Información*

Se empleo los CC de Spearman y Tau-b de Kendall para determinar la relación entre la salud y vigor de las plantas (con base en los IV) con la incidencia del hongo de roya. La aplicación de Spearman se utiliza en este caso ya que los datos del objetivo 1 son escalares (rangos) ya que clasifican la incidencia del hongo en 7 clases. Este es un procedimiento de pruebas de ENP., la normalidad no afecta su uso (Bautista Díaz et al., 2020).

$$p = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad [1]$$

Donde: p = Coeficiente de correlación de Spearman., d_i = diferencia entre los dos rangos observados y n = numero de observaciones. Este proceso fue realizado en el software SPSS de la empresa IBM, de la siguiente forma: se cargaron los pares ordenados (IV, Incidencia de la roya) como una hoja de cálculo, posteriormente, en la pestaña “Analyze”, se eligió la opción “correlación”. Dentro de esta pestaña se seleccionó a la opción “Bivariate”.

En la ventana emergente se seleccionaron ambas variables mencionadas anteriormente y se seleccionó la casilla que indica “Spearman”. El programa generó una tabla resumen indicando el coeficiente calculado. Además de esto, nos dio un valor de “sig.” (nivel de significancia bilateral) el cual indicará si se debe rechazar la hipótesis nula.

$H_0 =$ no hay asociación entre las variables

$H_1 =$ hay asociación entre las variables

si $p \leq \alpha$, se rechaza la H_0 , si $p > \alpha$, no se rechaza la H_0 . $\alpha = 0,05$

El CC Tau-B de Kendall (τ_b) indica la fuerza y dirección de la correlación entre las dos variables (Ecuación 2).

$$\tau_b = \frac{C - D}{\sqrt{(C + D + T_x)(C + D + T_y)}} \quad [2]$$

Donde: C: Pares concordantes; D: Pares discordantes; T: Empates. Para el uso de este CC fue necesario realizar una reclasificación de los valores del ráster de cada IV como la ejemplificada en la **Tabla 3**, con el fin de tener la misma cantidad de clases en ambos tipos de datos. Esta reclasificación se realizó en el programa de SPSS, por medio de la herramienta de “Transformar y Recodificar en la Misma Variable”. Dentro de ella se colocaron los rangos y se le asignó un nuevo valor (del 1 al 7). En la pestaña de vista de variables, existe el espacio de “valores”, permitiendo asignarle nombres a cada valor, terminado con la reclasificación de las variables y convirtiéndolas en ordinales, originalmente intervalos o de razón.

Tabla 3. Reclasificación de los índices de vegetación (rangos).

Reclasificación numérica	Valor del promedio de IV calculado por parcela			
	NDVI	GNDVI	GEMI	NDVIRE
1	[0,220;0,239)	[0,010;0,039)	[0,358;0,377)	[-0,115;-0,095)
2	[0,239;0,257)	[0,039;0,067)	[0,377;0,397)	[-0,095;-0,075)
3	[0,257;0,276)	[0,067;0,096)	[0,397;0,416)	[-0,075;-0,054)
4	[0,276;0,294)	[0,096;0,124)	[0,416;0,436)	[-0,054;-0,034)
5	[0,294;0,313)	[0,124;0,153)	[0,436;0,455)	[-0,034;-0,014)
6	[0,313;0,331)	[0,153;0,182)	[0,455;0,474)	[-0,014;0,006)
7	(0,331;0,350]	(0,182;0,210]	(0,474;0,494]	(0,006;0,027]

Nota: los rangos de los valores de ráster dependen de los valores máximos y mínimos de cada IV calculado, y a partir de estos, se generan 7 subdivisiones.

En este caso, se obtuvieron las variables: 1) la clasificación del cafeto según la incidencia de la roya del objetivo 1 (según la metodología de la **Figura A 1**; 2) Salud de la planta (Objetivo 2), valor dado por el IV reclasificada como se expresa en la **Tabla 3**, con siete clases cada variable, con un orden escalar, una jerarquía (de sano a enfermo) con la misma cantidad de clases para las dos variables, como lo requerido en el marco teórico de Tau-b de Kendall. El proceso en el software fue idéntico al usado para calcular Spearman, con la única diferencia de que se seleccionó la casilla que indica Tau-b de Kendall en el último paso.

Ambos coeficientes indican que mientras más cerca de +1 o -1 (inversa), la correlación es más fuerte y directa, a diferencia, si es igual o muy cerca de 0, se indica que no hay correlación entre las dos variables o es muy débil (Ortiz-Pinilla & Ortiz-Rico, 2021). Los cálculos de estos coeficientes se llevaron a cabo con los cuatro IV seleccionados. Como resultado de este objetivo, se generó una tabla que permitió comparar y determinar cuál IV se correlaciona mejor para las etapas tempranas del hongo de la roya en la parcela experimental.

La información generada en cada uno de los objetivos fue ordenada en hojas de cálculo y clasificada según la fecha de extracción, permitiéndole a la empresa ver la evolución del hongo en el transcurso de tres meses. Además, los rásteres permiten visualizar la salud y vigor de la planta de toda la parcela experimental y fueron compartidos con la empresa en una presentación con los ejecutivos.

3.1.3.2. Normalidad de los datos

Dado que hay 20 datos por variable, se aplicó la prueba de Shapiro-Wilk ($n \leq 50$). La herramienta permite identificar diversos comportamientos de los datos, incluye todas las variables que permitan calcular su normalidad, en este caso la prueba de normalidad mencionada; este proceso se encuentra en el apartado de análisis dentro de SPSS. La prueba de Shapiro-Wilk presenta las siguientes Hipótesis:

$$H_0 = \text{La muestra sigue una distribución normal.}$$
$$H_1 = \text{La muestra no sigue una distribución normal.}$$

Se tuvo en cuenta la regla de decisión. Donde p-valor es el valor de probabilidad y α es el nivel de significancia (0,05). Si $p \leq \alpha$ se rechaza la H_0 . Si $p > \alpha$ se acepta la hipótesis nula (Padilla, 24 de diciembre de 2018). Al tomar en cuenta que se está trabajando con ENP, la normalidad no es un criterio de selección, como si es para la EP, pero de igual manera en un proceso que siempre se debe hacer en cualquier investigación.

4. RESULTADOS Y DISCUSIÓN

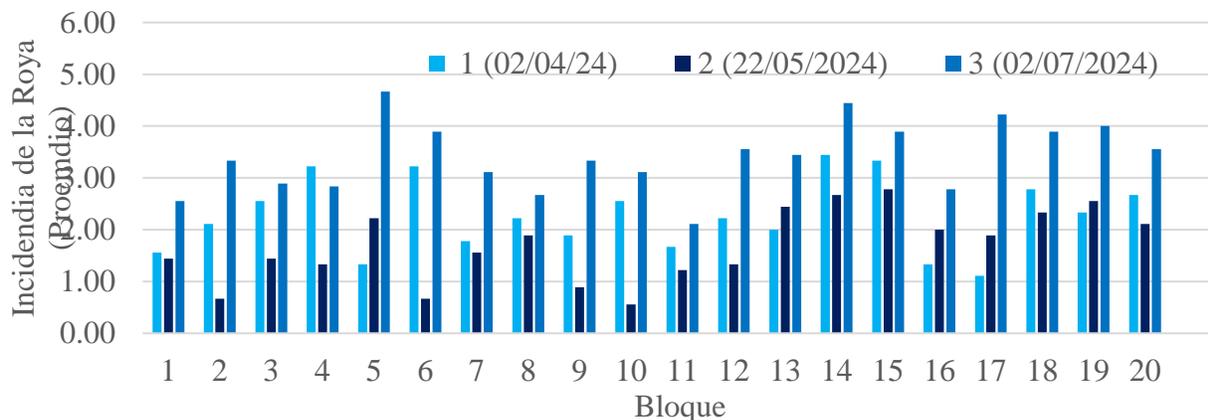
4.1. Objetivo específico 1: Clasificación de la incidencia de la roya en la parcela experimental

Por medio del Ing. Jorge Mora y una plataforma de encuestas (Survey123), se realizó el levantamiento de la incidencia de la roya en la parcela experimental de la Hacienda Alsacia. Este proceso se realizó en tres tomas de datos, y se evaluaron tres hojas por cada cafeto, tres cafetos por cada bloque, 20 bloques por toma de datos (ya que el bloque número 21 cuenta con gran área de este cubierto por árboles, lo cual impide que el dron cumpla su propósito).

La incidencia de la roya fue percibida por medio de la metodología ya explicada (**Figura A 1**), la cual indica que cero es una planta sin roya y 6 una hoja con más del 70% de su superficie cubierta por el hongo. Este proceso de evaluación se realizó en cada hoja de manera individual y por la parte trasera de la hoja, que es donde se encuentra el hongo. La **Figura 8** ilustra gráficamente los datos presentados en la **Tabla A 3**, proporcionando una representación visual más clara del comportamiento de la roya en cada bloque.

La **Figura 8** indica que la incidencia de roya es mayor en el mes de julio (Toma de datos 3), seguida por el mes de abril (Toma de datos 1), y finalmente, el mes de mayo (Toma de datos 2). Expresaron los funcionarios y expertos de la Hacienda que un comportamiento en zigzag del hongo es lo más común, no es normal un crecimiento lineal, esto debido a los patrones de reproducción del hongo (esporas) que se ven afectados con los cambios de lluvias.

Figura 8. Gráfico de barras de la incidencia de la roya para cada bloque en cada toma de datos.



Nota: La desviación estándar de cada bloque en cada toma de datos se encuentra en la **Tabla A 4**.

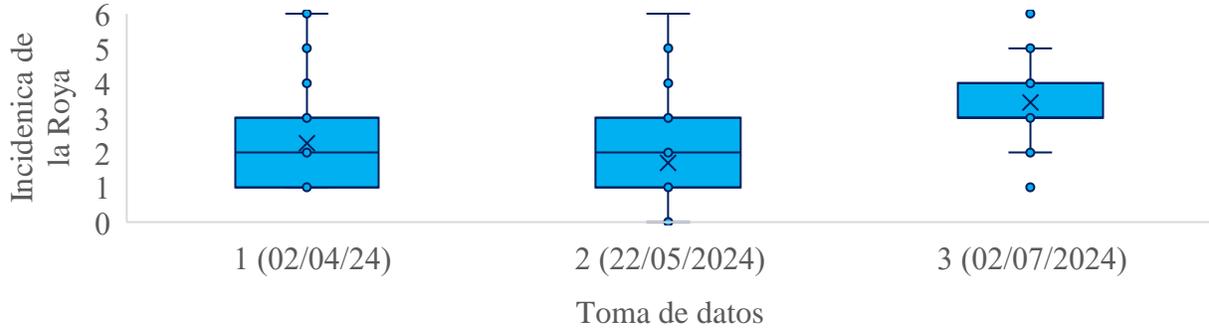
Este comportamiento coincide con los estudios de Calero Ocón (2024), en el cual se identificó que la roya tiene este mismo comportamiento y se ve gravemente afectada por la sombra, las lluvias y humedad relativa de la región. A este comportamiento se le puede agregar las tormentas y temporales que sufrió el país en los meses que se realizó esta investigación (INM, 2024).

Esta tendencia de zigzag en la parcela experimental es más visible en la **Figura 9**. Para sintetizar la información, sin dejar ninguno por fuera, se presentan dos gráficos de caja, con el fin de dar una ejemplificación de los datos conseguidos por toma de datos (**Figura 9**) y por bloque (**Figura 10**).

A diferencia de la **Figura 8**, donde se muestran los promedios del bloque (tres incidencias por café, tres cafés por bloque), en la **Figura 9**, se muestran todos los datos, con el fin de poder apreciar si hubo presencia de incidencias de roya de alto nivel. Para ello en la se puede apreciar que desde la primera toma de datos ha habido un nivel seis (máximo nivel) de incidencia de la roya, gracias a los límites superiores del gráfico de caja. La presencia de altos niveles de incidencia de la roya suele aumentar con el paso del año, consecuencia de factores favorables para la roya, como las lluvias y la humedad relativa (ANACAFÉ & CEDICAFÉ, 2022). Esta información puede

ser relacionada con la toma de datos tres, la cual presenta el mayor promedio de nivel de incidencia de la roya (3,44) de las tres tomas de datos.

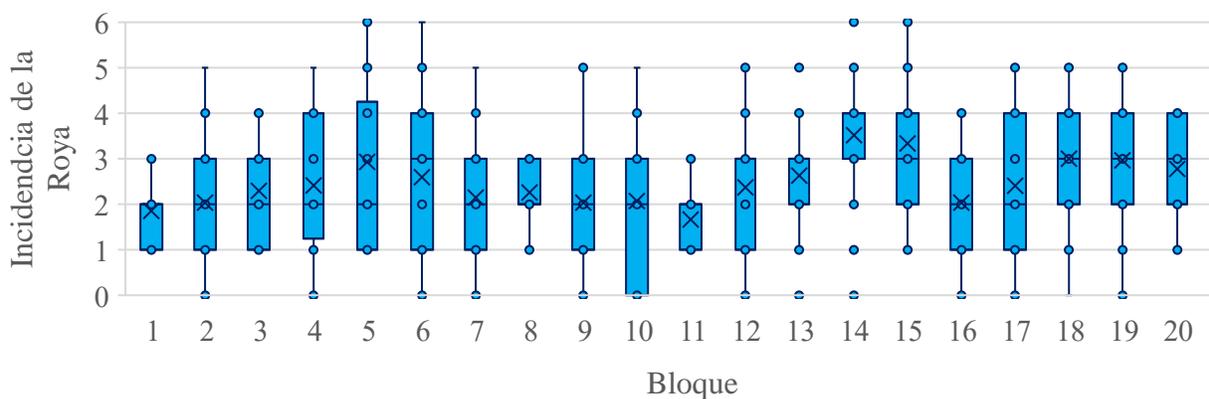
Figura 9. Gráfico de cajas de la incidencia de la roya en cada toma de datos de datos.



La toma de datos dos es la que presenta valores más bajos de las tres, con un promedio de 1,7. Por otro lado, en esta misma toma de datos se observa como también se presentan valores de nivel seis (máximo nivel). Se puede entender que, aunque sea la toma de datos con una incidencia de roya más leve, no se descarta la presencia fuerte del patógeno en algunos bloques.

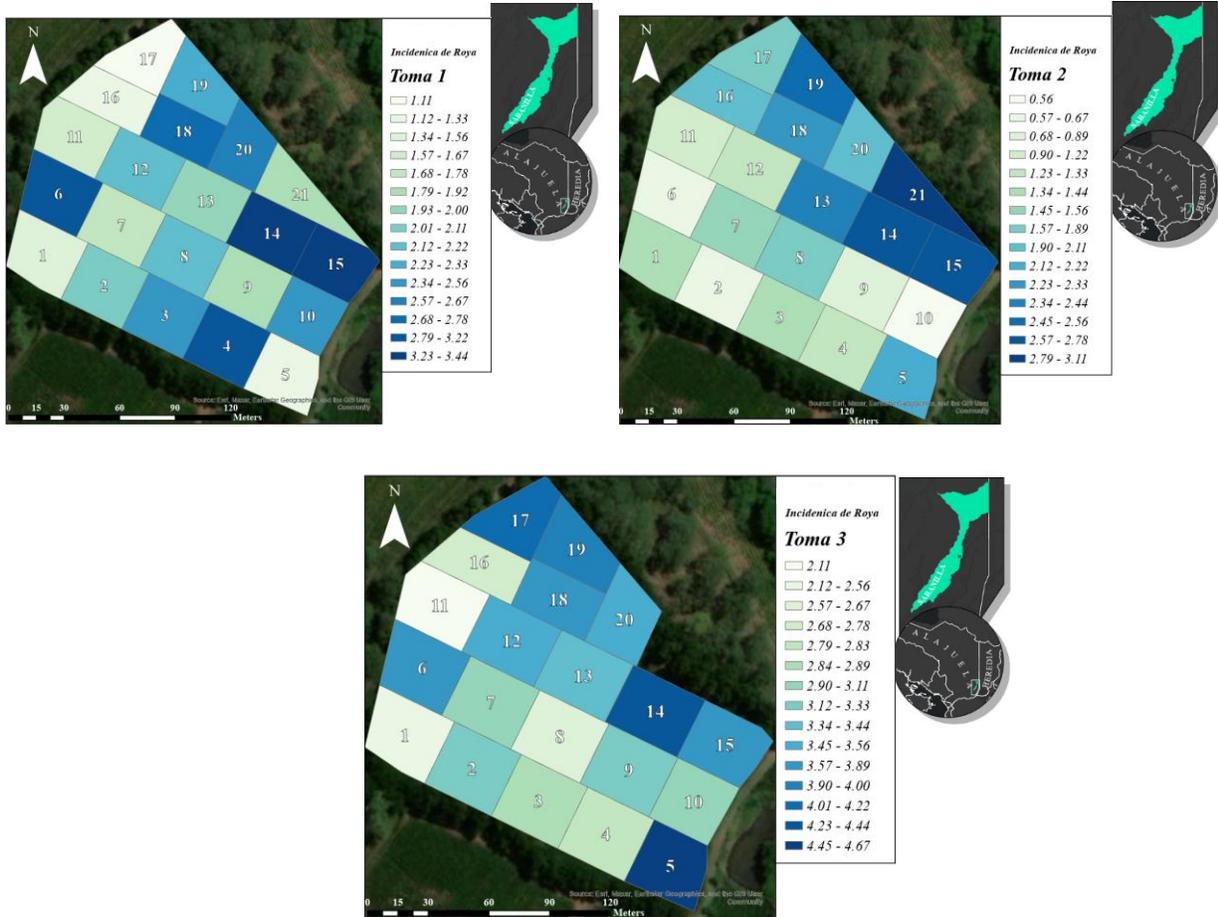
La **Figura 10** demuestra la varianza existente dentro de cada bloque a lo largo de las tres tomas de datos. Se puede apreciar como los bloques del exterior (principalmente bloques de la zona noreste, 16, 17, 18, 19 y 20) son más vulnerables al hongo. Esto se puede dar por diversos factores: 1) alta humedad relativa, 2) temperaturas menores de 20 a 22 °C, debido a la sombra generada por los árboles (Ramírez Dávila et al., 2023) y 3) terreno más alto, todos estos Aspectos Agroclimáticos del Café están definidos en el marco teórico de esta investigación.

Figura 10. Gráfico de cajas de la incidencia de la roya para cada bloque.



Si comparamos las imágenes de la **Figura 11**, se puede observar que los bloques 14, 15, 18, 19, y 20, son los que presentan una incidencia de la roya mayor en todo momento, esto es perceptible de igual manera en la **Figura 10**, ya que las cajas de estos bloques se encuentran desplazadas hacia arriba, indicando mayor incidencia.

Figura 11. Distribución geográfica de la incidencia de la roya: toma de datos 1, toma de datos 2 y toma de datos 3.



Nota: Estas figuras fueron hechas por medio de la plataforma ArcGIS PRO.

Los bloques mencionados anteriormente, se encuentran ubicados en la zona más poblada de árboles de la parcela experimental, junto con la colina más alta de la misma. Por el contrario, los bloques 7, 9, 8, 12 y 13 son los bloques que se encuentran en el centro de la parcela experimental, al mismo tiempo son los bloques que presentan una incidencia leve de la roya. Estos bloques, sin importar la hora, se mantienen constantemente recibiendo luz a lo largo del día, esto gracias a que

las sombras de los árboles no penetran en estos bloques. En la investigación de Calero Ocón (2024) se expresa como la sombra es uno de los principales factores que favorecen al desarrollo de la enfermedad.

El comportamiento de ciertos bloques no cambio en función del paso del tiempo, ejemplo de estos son los bloques 1, 8, 11, 13, y 14 que a pesar de las circunstancias mantuvieron un rango de datos similares. En la **Figura 10** se aprecia como la caja que representa Q1 y Q3 de los bloques anteriormente mencionados son las más pequeñas, indicando menor variabilidad en los datos, lo que promueve que la incidencia de estos bloques no fue tan cambiante. En contraste, los demás bloques si tuvieron un cambio más pronunciado entre las tomas de datos. Se puede destacar que los bloques 4, 5, 6, 10 y 17 los cuales tienen una separación de los Q1 y Q3 de aproximadamente 3 niveles de incidencia.

En resumen, en la tercera toma de datos existe una mayor incidencia de roya, debido al aumento de las lluvias y el comportamiento de zigzag del hongo (**Figura 9**). El aumento de la incidencia de la roya es más significativo en las zonas altas (noreste de la parcela experimental) debido a la menor cantidad de horas de luz diarias causadas por la sombra de los árboles. Este factor contribuye al desarrollo de la enfermedad, pero es necesario para disminuir el estrés de las plantas, por ende se debe generar una buena gestión de las sombras (Farfán-Valencia, 2021). La parcela experimental posee zonas de incidencia leve de la roya, las cuales se encuentran en la zona central de la misma, donde reciben más luz a lo largo del día.

4.2. Objetivo específico 2: Cálculo y Gestión de los Índices de Vegetación

Por medio de un dron con cámara multiespectral, se capturaron imágenes de la parcela experimental, las cuales fueron sometidas a procesos geomáticos en la plataforma PIX4D. Al finalizar se obtuvieron cuatro ortomosaicos (**Figura A 14**) por cada toma de datos. Se realizó el cálculo de los IV en la plataforma ArcGIS Pro, teniendo como resultado rásteres con valores por pixel entre -1 (superficies no vegetadas) y 1 (densidad, actividad fotosintética).

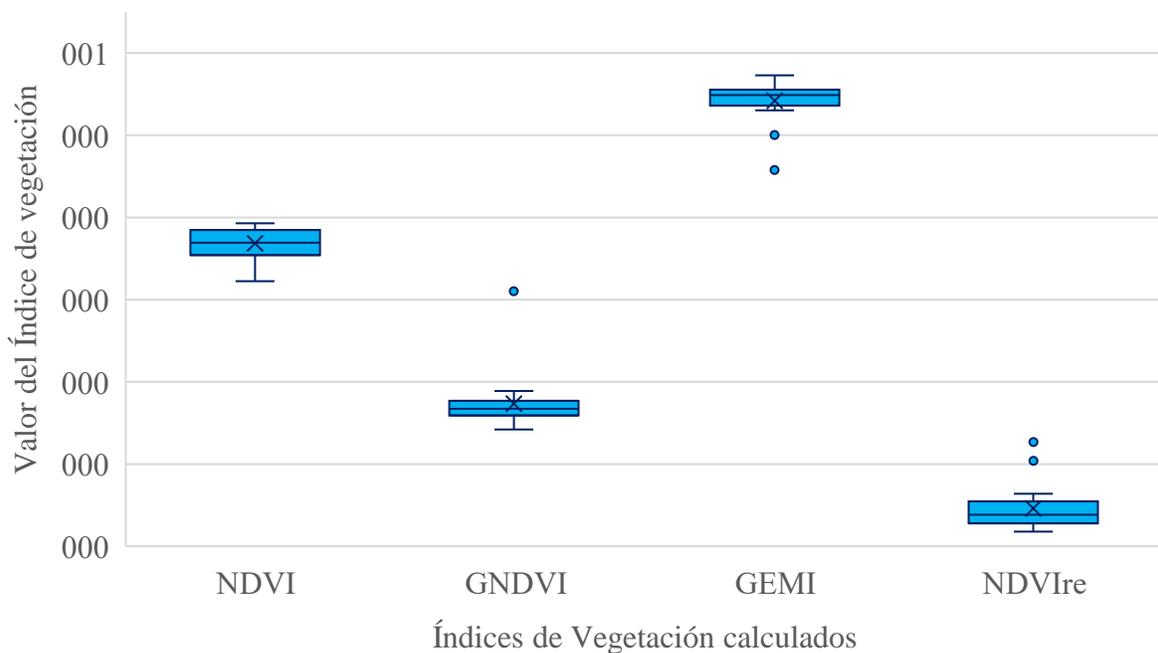
Cada IV puede representar la salud y vigor de la planta en diferentes rangos. Esta información se extrajo a hojas de cálculo, se generaron promedios desviación estándar, máx, min, rangos y

mediana para cada IV, en cada bloque, en cada toma de datos. Esta información puede ser visualizada en la **Tabla A 6**. Para ejemplificar de una manera visual la distribución de los datos, se presenta un gráfico de cajas en la **Figura 12**, el cual es la representación visual de los promedios por bloques calculados en la parcela experimental en las 3 tomas de datos. La **Figura A 11**, **Figura A 12** y **Figura A 13**, presentan individualmente cada toma de datos. Cabe recalcar que, este IV es el promedio calculado de todos pixeles dentro de cada bloque, que se identificaron como cafetos, para visualizar la información puede revisar la **Tabla A 5**.

Los IV calculados se encuentran dentro de un rango de -0,12 (NDVIre) y 0,49 (GEMI). A excepción de NDVIre, los IV se encuentran por encima de cero, siendo esto lo esperado, ya que se realizó una máscara para evitar que el suelo perjudicara el promedio de cada bloque, generalmente con valores inferiores a cero en la mayoría de los IV.

Por otro lado, GEMI tendió a valores más altos (promedio de 0.44). Esto puede deberse a que cada IV toma distintas capas para realizar su cálculo y es más sensible a ciertos factores, ejemplo, GEMI es sensible las condiciones climáticas, ya que se basa en los espectros más alejados (NIR y RED) (Pinty & Verstraete, 1992).

Figura 12. Gráfico de cajas para los índices de vegetación calculados (abril-julio).

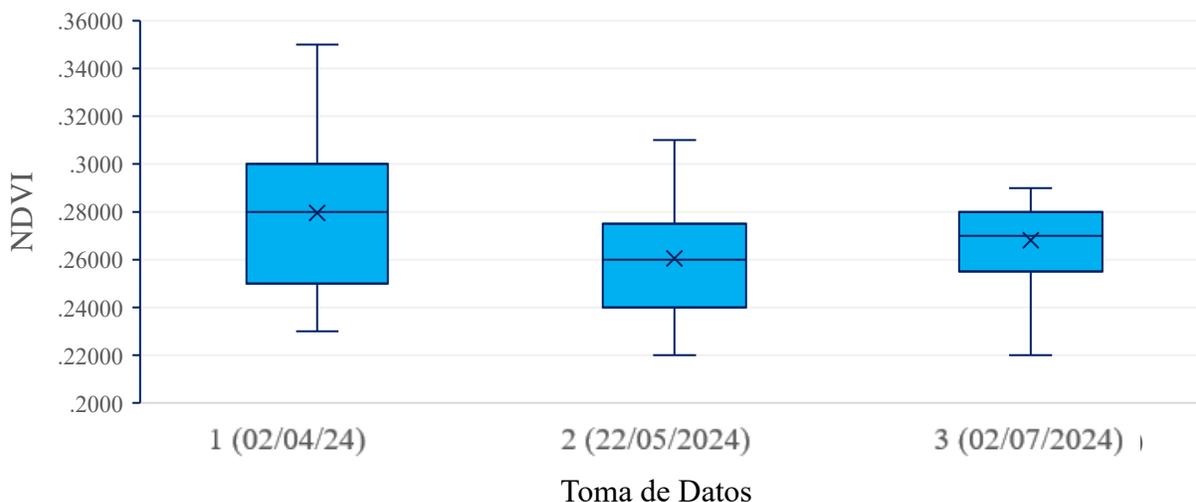


Al comparar los IV GNDVI y NDVI se observa que los valores se encuentran en la zona inferior del gráfico. Estos dos IV no son muy distintos del NDVI, siendo GNDVI y NDVI una modificación de NDVI (usando GREEN y RE en vez de RED, respectivamente). Estos valores tan bajos nos indican que en la mayoría de los casos el nivel de reflectancia de RE \geq NIR durante las tomas de datos. Teóricamente indica plantas con poca actividad fotosintética o zonas áridas, esto debido al estrés en las plantas, lo cual varía según cada temporada y según la planta. NDVI es uno de los índices más afectados por la densidad vegetal, ya que su principal uso es para determinar el estrés de grandes zonas de plantaciones (Singh et al., 2021).

Al hacer una comparación con el estudio de Chedid et al.(2024), el cual fue realizado en Romaria (MG), Brasil, se indicó que los cafetos con una actividad fotosintética pobre se encuentran con valores de NDVI inferiores a 0,6. Con valores mayores a este, las plantas se encuentran en actividad vegetal (fotosíntesis) y se consideran saludables.

De igual manera, si se comparan los datos de Chedid et al. (2024) con los de esta investigación, los meses de lluvia en la zona donde el informe fue hecho, comienzan en septiembre, el cual sería comparable a la primera toma de datos de esta investigación (abril). Como se muestra en la **Figura 13**, se tiene un promedio de NDVI mayor en la primera toma de datos (0,28), el cual disminuye en la segunda (0,26) y aumenta de nuevo en la tercera (0,27), la información puede ser consultada en la **Tabla A 5**.

Figura 13. Gráfico de caja para el NDVI en tres momentos en el tiempo.

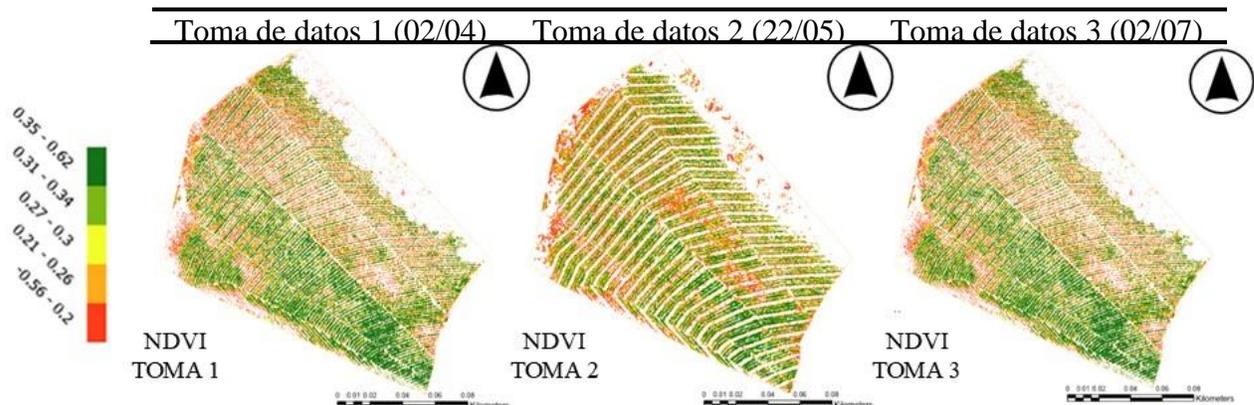


En la investigación de Chedid et al. (2024) se presenta una tendencia similar, ya que el primer mes de lluvia comienza con valores de NDVI de 0,68, baja a 0,61 y aumenta de nuevo a 0,76 a los 3 meses de la primera toma. Se puede destacar la tendencia del GNDVI (**Figura A 5**), que es similar a las anteriormente descritas. En la primera toma de datos el valor es mayor, luego disminuye y en la tercera vuelve a subir. Si contemplamos los demás índices, GEMI (**Figura A 6**) tiene una tendencia descendiente (siendo esta la esperada si la incidencia de la roya aumenta con el paso del tiempo), mientras NDVIre (**Figura A 7**) tiene una tendencia ascendiente.

En la **Figura 14**, se pueden apreciar de manera visual los IV calculados por cada toma de datos. Es apreciable como en la primera toma de datos de todos los índices se ve una franja verde en el lado oeste de la parcela, la cual indica valores altos, tal y como muestra la leyenda de cada representación (mejor estado de vigor de la planta).

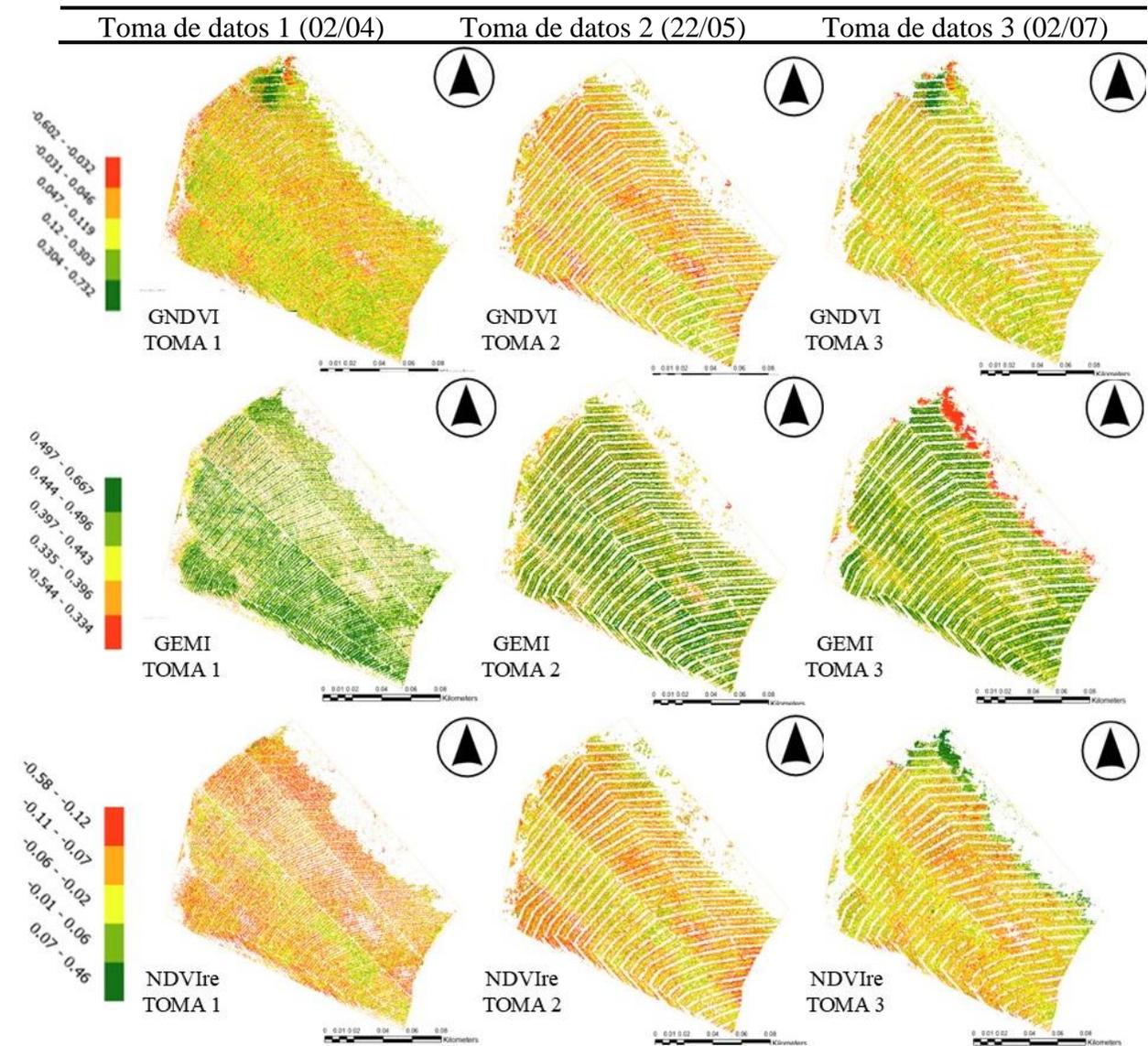
Si volvemos a la **Figura 11** (resultado del objetivo 1), los bloques que se encuentran con valores más bajos de incidencia de la roya se ubican en esta misma zona. Por el otro lado, se aprecian valores más amarillentos y rojizos en todos los IV en la zona noreste de la parcela, en la cual, como se mencionó en el primer objetivo, son las zonas más afectadas por la incidencia de la roya. Se aprecia que los distintos IV, aunque se calcularon con bandas distintas y dieron valores distintos, el comportamiento o representación de estos en el campo es muy similar.

Figura 14. Índices de vegetación calculados en la parcela experimental para las tomas de datos 1, 2 y 3.



*Nota: los rangos para cada IV se realizaron con el propósito de que en cada uno existiera la misma cantidad de datos, esto explica por qué los rangos no son iguales en ningún momento. Los valores usados para esta visualización se ven expresados en la **Tabla A 6**.*

Figura 14. Continuación.



En la segunda toma de datos se puede apreciar un cambio de la densidad de índices de color rojo, como se muestra en la leyenda de cada imagen, estos indican valores bajos del IV calculado. La presencia de puntos rojos aumenta en la zona central, pero desplazada al este. En el caso de GNDVI se aprecia una franja roja del lado noreste de la parcela, la cual, si nos dirigimos a la **Figura 11**, se aprecia como esa misma zona presenta los bloques con mayor incidencia de la roya. Mientras que el lado suroeste de la parcela aun presenta un IV mayor (tonos verdes), tal como se ve en la **Figura 14**.

El GEMI, sigue el patrón anteriormente mencionado. Ahora, en la tercera toma de datos, los valores más bajos de GEMI (en este caso valores entre -0.53 y 0.39) se encuentran en una nube de puntos en el centro de la parcela (los valores marginales del lado noreste de la parcela no son tomados en cuenta para realizar un análisis, ya que es sombra de los árboles). Este patrón se repite en todos los IV estudiados.

En IV que tuvo la mejor similitud en la tercera toma de datos con relación a la **Figura 11** es el NDVI_{re}, el cual presenta mayor densidad de puntos rojos en la zona donde se encuentran los bloques con mayor incidencia de la roya. El contraste y cambio es perceptible entre las tres tomas de datos, al igual que los datos calculados en el primer objetivo, se ve una “mejora” en la segunda toma, y una recaída en la tercera, esto debido a el comportamiento de zigzag del hongo, y los temporales que sufrió el territorio nacional.

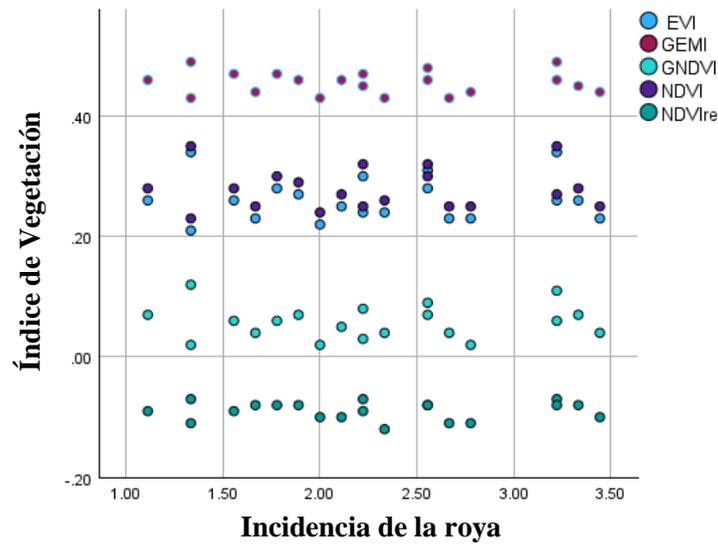
Los IV permiten visualizar el dinamismo de las parcelas, siendo esta información importante para poder llevar el control temporal del estado de salud de las plantas, y poderlo contrastar con tratamientos o diversas aplicaciones de fungicidas. En resumen, la mayoría de los promedios de los IV muestran una tendencia similar a la vista en la literatura, comenzando alto, bajando y remontando en la tercera toma de datos. Chedid et al. (2024) indicó que las plantas se encuentran en un rango no óptimo para la actividad fotosintética, lo que impide generar frutos de una calidad alta. Esto debido a que ningún valor de NDVI supera 0,6. Valores mayores a 0,6 indican que las plantas se encuentran en actividad vegetal (saludables) (Chedid et al., 2024). La presencia del hongo de la roya perjudica la actividad fotosintética del café, ya que, al cubrir las hojas con su presencia, estas impiden la entrada directa de los rayos de sol a los fotorreceptoras (SENASICA, 2019).

4.3. Objetivo específico 3: Selección del índice de vegetación basado en correlación estadística

Como se desea seleccionar el IV más adecuado para la detección de roya en etapas tempranas se utilizó SPSS, con el fin de identificar alguna tendencia clara de los datos antes de aplicar cualquier prueba de correlación. Se debe tener en cuenta que la escala de los datos dificulta este proceso.

Al examinar la **Figura 15**, se puede ver cómo se relacionan las dos variables (incidencia de la roya y IV) en la toma de datos 1. En el eje X se encuentra la incidencia de la roya (objetivo 1) y en el eje Y los cuatro IV calculados (objetivo 2). Con la escala de trabajo mostrada, no se ve una tendencia positiva o negativa predominante, lo que nos indica que la correlación puede ser baja (si es que existe). Lo mismo sucede en las demás tomas de datos (Figura A 10 y la Figura A 11). Para validar la información, se utilizaron los CC.

Figura 15. Gráfico de puntos de incidencia de la roya con IV calculados de la toma de datos 1.



*Nota: puede apreciar las demás tomas de datos en la **Figura A 8** y la **Figura A 9**.*

En el programa SPSS, se realizó un análisis de correlación, integrando todas las variables del objetivo 1 y 2 (incidencia de la roya e IV, respectivamente), seleccionando la prueba de Spearman. La **Tabla 4** muestra la correlación generada, además, en la **Tabla A 10** indica que todas las variables son normales, a excepción de GNDVI3, GEMI3, NDVIre3. Se debe realizar la prueba de Spearman para todas las variables, ya que se está aplicando ENP (Ortiz-Pinilla & Ortiz-Rico, 2021), la cual indica que esta se puede llevar a cabo sin importar la normalidad de los datos. Para el CC de Spearman se tomaron en cuenta las siguientes Hipótesis:

H_0 = No existe relación significativa entre la incidencia de la roya y el índice de vegetación calculado.

H_a = Si existe relación significativa entre la incidencia de la roya y el índice de vegetación calculado.

Tabla 4. Correlación de Spearman de la incidencia de la roya y los índices de vegetación calculados.

Toma de datos	NDVI	GNDVI	GEMI	NDVire
1 (abril)	0,022	0,016	-0,138	-0,008
2 (mayo)	0,125	0,029	-0,239	0,163
3 (julio)	-0,191	0,142	-0,267	0,245

Nota: Ver tabla de significancia en la **Figura A 6**.

Las correlaciones obtenidas son bajas, siendo la mayor de -0,267, una correlación débil dentro del espectro (Fiallos, 2021). Esta correlación corresponde a la toma de datos de julio, perteneciendo al IV GEMI. En la **Figura A 6**, se puede observar que todos los valores de significancia (p) de todos los IV son mayores a 0,05, lo cual nos lleva a no rechazar la H_0 . Por lo tanto, no existe relación significativa entre la incidencia de la roya y los índices de vegetación. Si bien si existe una correlación baja, esta se puede dar por situaciones del azar. Para comprobar la veracidad de estos datos, se calculó el TE y PE de cada uno.

En la prueba de correlación de Tau-b de Kendall, se utilizaron las mismas hipótesis nulas (H_0) y alternativas (H_1) que en el caso de la correlación de Spearman. Las pruebas tuvieron el mismo objetivo en esta investigación: determinar la correlación entre las variables (incidencia de la roya e IV, respectivamente). Sin embargo, hay diferencias importantes. En el caso de Tau-b de Kendall, se empleó una ENP y se deben cumplir un requisito. Los datos no pueden ser de tipo rango o razón; deben ser de tipo ordinal. Una variable ordinal es una variable cualitativa, la cual expresa una cualidad u orden, en este caso de muy sana a muy enferma. En la variable de incidencia de la roya, el cero es el valor que representa la menor cantidad de roya, quiere decir la más sana, en contraparte, en los IV el valor más grande o más cerca de uno, indica mayor salud. Con esto en mente, la correlación debería ser negativa, ya que la salud está en el extremo opuesto de la contraparte.

Para lograr esto, se realizó una reclasificación de los datos. Al separar los datos en siete rangos de igual tamaño (**Figura A 8**) se cumple con el requisito de tipo ordinal. Es relevante mencionar que estos rangos serán diferentes para cada IV, tomando como referencia los valores máximos y mínimos de cada tipo de índice en las tres mediciones. Para el caso de la incidencia de roya (variables del objetivo 1), los siete rangos ya están establecidos según **Figura A 1**.

En la **Figura 16** se muestra la distribución de los valores en la nueva clasificación según cada IV calculado. En este caso, la prueba de normalidad para las variables reclasificadas no es necesaria para el criterio de decisión de pruebas estadísticas no paramétricas (Cuevas Urióstegui et al., 1994). De igual manera se puede encontrar en la **Tabla A 9**. Esta indica que solo una variable, NDVI2, es normal. El criterio de ENP indica que esta se puede llevar a cabo sin importar la normalidad de los datos.

Ya con las clasificaciones de IV hechas, se cuenta con una muestra cuadrada (7x7) ya que las variables de incidencia cuentan con siete clases y las de IV cuentan con siete clases también. Por ende, se puede calcular el Coeficiente de correlación de Tau-b de Kendall. Se realizó un análisis de correlación, integrando todas las variables, y seleccionando la prueba de Tau-b de Kendall, en el programa SPSS. El resultado de esta prueba puede ser observada en la **Tabla 5**.

Figura 16. Reclasificación de los índices de vegetación en escalas ordinales.

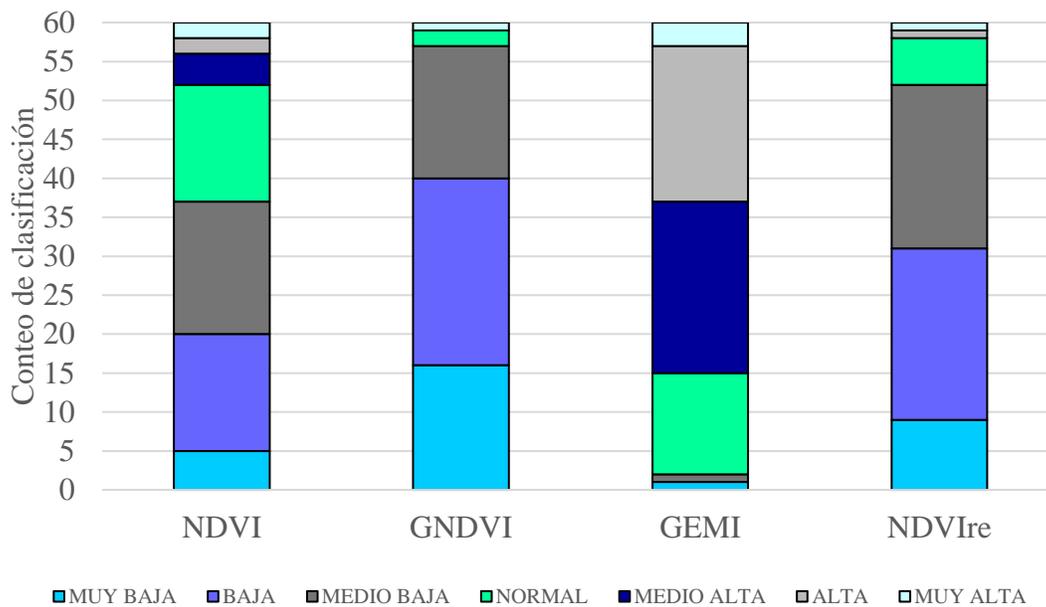


Tabla 5. Correlación de Tau-b de Kendall de la incidencia de la roya y los índices de vegetación calculados.

Toma de datos	NDVI	GNDVI	GEMI	NDVIre
1 (abril)	0,055	0,029	0,030	-0,042
2 (mayo)	0,115	-0,016	-0,169	0,159
3 (julio)	-0,135	0,133	-0,266	0,024

Nota: Ver tabla de significancia en la **Figura A 6**.

Al igual que en la **Tabla 4**, tanto para la toma de datos dos y tres, GEMI presenta el mayor CC (-0,266), pero como se puede observar en la **Tabla A 7**, no existe ningún valor significativo, lo que indica que este resultado no es 100% fiable y cuenta con más del 5% de valores al azar generando la correlación. Para darle más veracidad a los datos, se generó un análisis de TE y PE en la herramienta GPower.

Dentro de la aplicación, se seleccionó en la sección de “test”, modelo normal “bi-variable” de correlación, además se indicó el tipo de análisis “Post Hoc”, ya que los cálculos ya fueron realizados en otro programa. Por medio de los datos del CC de la **Tabla 4** y la **Tabla 5** junto con los valores de significancia de la **Tabla A 7**, se calcularon tanto el TE ($1-\beta$) y la PE (Sig) para todas las pruebas, esta información está presente en la **Tabla 6**, en la cual se identifica los CC calculados con respecto a la incidencia de roya y los IV individualmente. Ha de tenerse en cuenta que el coeficiente de determinación es el CC elevado al cuadrado.

Tabla 6. Correlación de índices de vegetación con incidencia temprana de la roya en cafetos.

Toma de datos	Incidencia de Roya	Índices de vegetación			
		NDVI	GNDVI	GEMI	NDVIre
1 (abril)	Tau-B de Kendall	0,055	0,029	0,030	-0,042
	Sig	0,774	0,883	0,883	0,843
	1- β	0,780	0,884	0,884	0,846
	Spearman's rho	0,022	0,016	-0,138	-0,008
	Sig	0,927	0,947	0,561	0,975
	1- β	0,927	0,947	0,627	0,975
2 (mayo)	Tau-B de Kendall	0,115	-0,016	-0,169	0,159
	Sig	0,559	0,938	0,410	0,443
	1- β	0,606	0,938	0,524	0,542
	Spearman's rho	0,125	0,029	-0,239	0,163
	Sig	0,599	0,902	0,311	0,492
	1- β	0,650	0,903	0,533	0,591
3 (julio)	Tau-B de Kendall	-0,135	0,133	-0,266	0,024
	Sig	0,510	0,537	0,194	0,906
	1- β	0,578	0,601	0,453	0,907
	Spearman's rho	-0,191	0,142	-0,267	0,245
	Sig	0,421	0,551	0,255	0,298
	1- β	0,562	0,621	0,524	0,530

*Nota: Valores negativos indican una correlación de la misma índole. Valores en **negrita** representan el CC mayor de cada caso.*

Para evaluar la relación entre la incidencia de la roya en etapas tempranas (julio) y el IV GEMI calculado mediante bandas multiespectrales, se realizó una correlación de Spearman. Los resultados obtenidos son $r(20) = -0,267, sig = 0,255$, con un tamaño de efecto pequeño ($<0,30$) y una potencia estadística de $1 - \beta = 0,524$.

En etapas tempranas de la incidencia de la roya, el IV GEMI dio una correlación (Spearman) leve. La significancia al ser mayor que α nos indica que no hay una correlación significativa, lo cual quiere decir que la relación entre las variables se debe en 20% al azar, aceptando la H_0 . Esta es un TE pequeño, indicando distancia entre las dos variables. Por otro lado, la PE nos indica que no hay un error II (falso negativo) ya que existe un 47,6% de probabilidad de rechazar la H_0 cuando en realidad no se deba, siendo el mínimo justificado 20%. Por ende, se mantiene la H_0 .

En el caso de la incidencia de la roya y el IV GNDVI en el mes de mayo, existe una correlación (Tau-b de Kendall) de -0,016, con una probabilidad de que los datos se relacionen al azar de 93%, esto nos lleva a aceptar la H_0 , decretando que no existe una correlación entre las dos variables. Para corroborar la información, el TE nos indica que ($>,50$) es muy grande, demostrado la distancia entre ambas curvas, junto a la PE de un 6,2% de probabilidad de rechazar la H_0 cuando en realidad no se deba, siendo el mínimo justificado 20%. En este caso se rechaza la H_0 . Esto da a entender que la poca relación que existe entre las variables no se debe al azar, sino a la distancia tan grande que existe entre las dos curvas de las variables, lo que incita a aumentar la muestra.

Luego de hacer el análisis de cada uno de los IV, se seleccionó el índice de vegetación que tenga la mejor correlación con la incidencia en etapas tempranas de la roya en café. Esto se realizó por medio de una comparativa de los CC obtenidos, seleccionando el índice que en cada toma de datos tuvo un valor absoluto mayor, siendo el más acertado el que más se repita para las tres tomas de datos, este proceso se observa en la **Tabla 7**.

Como resultado de la tabla anterior, se tiene que el mejor IV para la detección de la roya en etapas tempranas es el GEMI. Como se puede apreciar en la **Tabla 6**, en ninguno de las tres tomas de datos GEMI presenta una $sig. < \alpha$. Por ende, aunque sea el que tiene la mejor relación con la incidencia temprana de la roya no se puede descartar la H_0 .

Tabla 7. Selección del índice de correlación que tenga la mejor relación con la incidencia de la roya en etapas tempranas en tres tomas de datos.

Toma de datos	Coefficiente de Correlación	Mejor Correlación por coeficiente	Valor de la correlación	Mejor correlación de la toma
1 (abril)	Tau-B de Kendall	NDVI	0,055	GEMI
	Spearman's rho	GEMI	-0,138	
2 (mayo)	Tau-B de Kendall	GEMI	-0,169	GEMI
	Spearman's rho	GEMI	-0,239	
3 (julio)	Tau-B de Kendall	GEMI	-0,266	GEMI
	Spearman's rho	GEMI	-0,267	

Para ello, la única forma de hacerlo sería con el error tipo II de la PE, pero se tendría que aumentar el número de muestras. El programa de GPower permite hacer pruebas a priori, con el fin de determinar que tamaño de muestra es necesaria para que el PE ($1-\beta$) sea $>0,80$, con el fin de descartar la H_0 y poder utilizar esta correlación en toda la población objetivo. El cálculo indica que se necesitarían como mínimo 60 (bloques) muestras por variable, en esta investigación se realizaron 20 subparcelas.

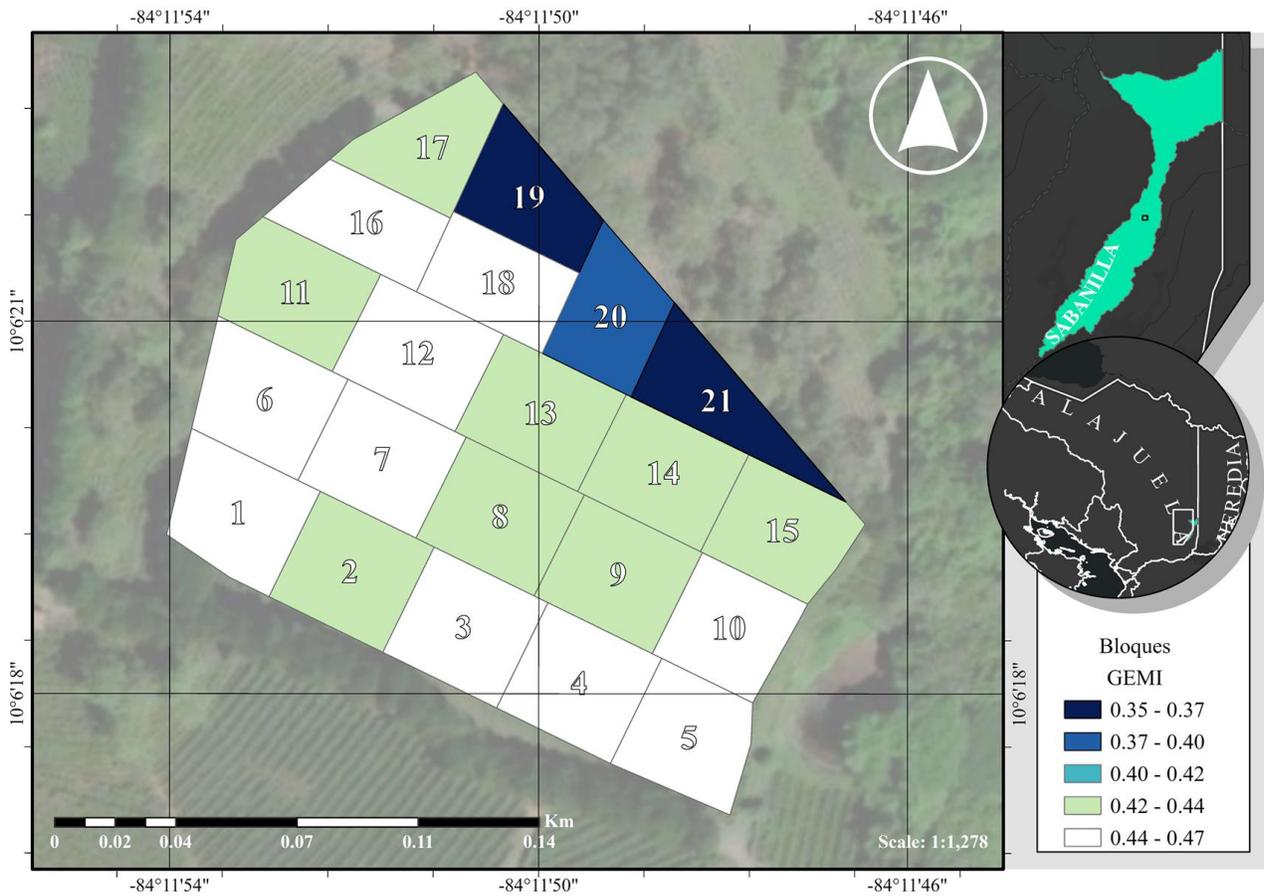
Por lo tanto, se selecciona GEMI como el mejor IV para la incidencia en etapas temprana de la roya (**Tabla 7**), pero se tiene en cuenta que es una correlación de poca fuerza (-0,266), y en todos los casos $sig. > \alpha$, por ende, no se rechaza la H_0 . Es decir, no existe relación significativa entre la incidencia de la roya y el índice de vegetación calculado, por lo que se concluye que ningún IV se puede utilizar para identificar la incidencia de la roya en etapas tempranas de la enfermedad. Como resultado final, la **Figura 17** muestra el mejor IV (GEMI) calculado, en el mejor momento (tercera toma de datos).

Con el fin de ejemplificar y poder comparar con el mapa de incidencia de la roya en la **Figura 11**, se ha de tener en cuenta que valores altos de incidencia de la roya representan enfermedad (Azul), mientras que valores bajos de IV (Azul) representa alta actividad fotosintética (vigor de la planta), por ende, la relación entre ellas es inversa.

Se tiene que los bloques de la **Figura 17** que se encuentran en la zona noreste tienen valores más bajos (menor actividad fotosintética). En contra parte, los bloques de la **Figura 11**, con colores

azul oscuro, representan una incidencia del hongo alta, estos se encuentran de igual manera en la zona noreste de la parcela. Se ve claramente una relación (inversa) en los bloques 1, 2, 9, 13, 15, 16, 19 y 20. Pero existe discordancia en los bloques 5, 6, 11, 12, 17 y 18. Por último, se aprecia cercanía al valor en los bloques 3, 4, 7, 8, 10 y 14. Siendo más los bloques que mantienen una relación inversa, a los que no la tienen.

Figura 17. Mapa de GEMI de la tercera toma de datos (Mejor correlación con la incidencia de la roya).



5. CONCLUSIONES

Algunas zonas de la parcela experimental (bloques) de la Hacienda Alsacia, se ven más afectadas que otras por las condiciones climáticas, ya sea por motivos de altura, sombra, o mal drenaje (aumento de la Humedad Relativa). Estas condiciones perjudican a los cafetos, haciéndolos más vulnerables a las enfermedades, en este caso, a la roya (*hemileia vastatrix*). Este problema se acentúa espacialmente en el noreste de la parcela experimental.

La incidencia de la roya (*hemileia vastatrix*) posee un patrón de crecimiento habitual, a partir de las primeras etapas del hongo, se presenta una incidencia promedio en toda la parcela de 2,27, 1,70 y 3,44 en las tomas de datos uno, dos y tres, respectivamente, lo que marca una tendencia en zigzag, la cual es la esperada para este cultivo.

Los valores promedios del GEMI fueron los más altos (0,46, 0,45, 0,44), en contra parte, NDVI fue el IV con los valores promedios más cercanos a cero (-0,09, -0,08, -0,05), esto para las tomas de datos uno, dos y tres, respectivamente. Cada índice tiene sus fortalezas, debilidades y rangos de operación.

El Índice de Vegetación Mejorado Global (GEMI) obtuvo el valor de correlación más alto para relacionar la incidencia de la roya en las etapas tempranas del cultivo, esto se determinó por medio de los CC de Tau-b de Kendall (-2,266) y Spearman (-0,267). Sin embargo, es importante destacar que, aunque GEMI muestra una tendencia a detectar la presencia de la enfermedad, la correlación no es estadísticamente significativa. Esto sugiere que, GEMI no puede ser utilizado de manera confiable para predicciones precisas en las primeras etapas de la infección.

A medida que el tiempo avanza en la parcela experimental, la correlación entre los índices de vegetación y la incidencia de la enfermedad se vuelve más fuerte. Este hallazgo permite indicar que los IV no son tan útiles para la detección temprana, más si para un monitoreo global y la evaluación de la salud en general de la planta.

Los IV examinados en este estudio no han demostrado ser métodos precisos para la detección temprana de la roya, debido a la falta de una correlación significativa (como expresa la **Tabla 6**, con $0,158 \geq sig. > 0,05$) en las etapas iniciales de la infección. Esta información indica que estos

índices no son sensibles a los cambios sutiles en la vegetación causados por la roya en sus primeras fases.

El análisis de correlación utilizando el coeficiente de Spearman reveló los índices de correlación más altos en comparación con otros métodos estadísticos ($Spearman_{@R3}(p) = -0,267$). Esta información permite guiar el estudio a como deberían ser tratados los datos, ya que permitiría, a futuros estudios, la selección de un solo CC ahorrando trabajo y tiempo, ya que las reclasificaciones no tendrían que hacerse. El inconveniente está en que los datos deberían ser todos no normales o tener más de 50 datos (bloques en este caso).

6. RECOMENDACIONES

Aunque el programa GPowerNT indica que para mejorar la robustez y la validez estadística de los resultados se recomienda aumentar el número de muestras a 60, esto permitirá una mayor variabilidad y representatividad de los datos. Además, reducirá el error experimental y aumentará la precisión de las conclusiones. Esto se podría hacer por medio de aumentar el número de cafetos por bloques y disminuyendo el número de bloques a estudiar, lo cual, al concentrar el trabajo en una zona más limitada, reduce el tiempo de traslados, y disminuyendo la dispersión de los datos.

Es recomendable replicar el experimento en distintas etapas del desarrollo de la roya. Como indican los CC en la **Tabla 6**, hay una tendencia al aumento de la correlación mientras más avanza la enfermedad. Además, comparar los resultados de las diferentes etapas proporcionará una visión más completa de la dinámica de la enfermedad y su impacto en los cultivos.

Explorar y probar otros índices de vegetación puede ofrecer nuevas perspectivas y mejorar la detección de la roya. Índices como el Índice de Vegetación Ajustado al Suelo (SAVI), Índice de vegetación atmosféricamente resistente (ARVI) y el Índice de vegetación de diferencia visible (VDVI) podrían ser evaluados para determinar su eficacia en la detección de la roya en diferentes etapas.

Se recomienda el uso de calibradores en las imágenes multiespectrales, para garantizar la precisión y consistencia de los datos obtenidos. Los calibradores permiten corregir variaciones en la iluminación y otros factores ambientales, mejorando la precisión de las imágenes y la fiabilidad de los índices de vegetación calculados. Esta práctica es especialmente importante en estudios a largo plazo y en diferentes condiciones climáticas, asegurando que los resultados sean comparables y reproducibles en el tiempo.

Llevar a cabo el experimento en parcelas no experimentales, es decir, en condiciones de campo reales, es fundamental para validar la aplicabilidad de los resultados en situaciones prácticas. Esto ayudará a evaluar la efectividad de los índices de vegetación en un entorno más variado y menos controlado, proporcionando una mejor comprensión de su utilidad en la agricultura comercial.

Además, permitirá identificar posibles limitaciones y desafíos que puedan surgir en la implementación práctica de las técnicas de monitoreo.

Se recomienda explorar otros métodos de detección, como el uso de sensores hiperespectrales, huellas espectrales o técnicas de teledetección más avanzadas, que puedan ofrecer una mayor precisión y sensibilidad en la identificación de la enfermedad en sus etapas iniciales.

Al presentar esta investigación a la empresa, se realizaron varias observaciones, entre ellas la aplicación de una evaluación previa de falsos positivos con grupos de control, lo que permitiría ajustar mejor los índices de vegetación. Además, se recomendó generar una huella espectral de la roya para facilitar su detección mediante imágenes multispectrales.

7. BIBLIOGRAFÍA

- ANACAFÉ, & CEDICAFÉ. (2022). Reporte de Roya y Recomendaciones. *Centro de investigaciones en Café de Anecafé*, 4.
- Arcilla, J. (2007). Crecimiento y desarrollo de la planta de café. *Sistemas de producción de café en Colombia*, 2, 22–60. <http://www.cenicafe.org/es/documents/LibroSistemasProduccionCapitulo2.pdf>
- Avelino, J., Treminio, E., Casanoves, F., Vílchez, S., Cárdenas, J., & Lizardo, C. (2019). *Programa Centroamericano de Gestión Integral de la Roya del Café; Guía para la vigilancia de la roya del café (Hemileia vastatrix)*. *March* 2022, 29. <https://doi.org/10.13140/RG.2.2.25907.30244>
- Barbosa, B. D. S., Araújo e Silva Ferraz, G., Mendes dos Santos, L., Santana, L. S., Bedin Marin, D., Rossi, G., & Conti, L. (2021). Application of RGB Images Obtained by UAV in Coffee Farming. *Remote Sensing*, 13(12), 2397. <https://doi.org/10.3390/rs13122397>
- Bautista Díaz, Victoria Rodríguez, Vargas Estrella, & Hernández Chamosa. (2020). Pruebas estadísticas paramétricas y no paramétricas. *Educación y Salud Boletín Científico Instituto de Ciencias de la Salud Universidad Autónoma del Estado de Hidalgo*, 9(17), 78–81.
- Brown, J. (2018, noviembre 27). *NDVI, the Foundation for Remote Sensing Phenology*. USGS.
- Caballero Beltrán, D. A. (2017). Aplicación de imágenes NDVI para el control de riego y enfermedad en cultivos Agrícolas mediante el uso de aeronaves no tripuladas (UAV) y el software AgVault. *APORTE SANTIAGUINO*, 10(2), 305. <https://doi.org/10.32911/as.2017.v10.n2.172>
- Calero Ocón, I. (2024). Respuesta de la roya (*Hemileia vastatrix*) frente a condiciones climáticas y aplicaciones de fungicidas en el cultivo de café (*Coffea arabica*) bajo sombra. *Perspectivas Rurales Nueva Época*, 22(43), 1–20. <https://doi.org/10.15359/prne.22-44.8>
- Cárdenas Castro, J. M. (2014). Potencia estadística y cálculo del tamaño del efecto en G*Power:

- complementos a las pruebas de significación estadística y su aplicación en psicología. *Salud & Sociedad*, 5(2), 210–244. <https://doi.org/10.22199/S07187475.2014.0002.00006>
- Carneiro, D., Silva, D. A., Willem, G., Toonstra, A., Lacet, H., & Souza, S. (2014). Qualidade de Ortomosaicos de Imagens de VANT Processados com os Softwares APS , Pix4D e Photoscan. *V Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação, April*, 747–754.
- Chedid, V., Cortez, J. W., & Arcoverde, S. N. S. (2024). MONITORING THE VEGETATIVE STATE OF COFFEE USING VEGETATION INDICES. *Engenharia Agrícola*, 44. <https://doi.org/10.1590/1809-4430-eng.agric.v44e20220212/2024>
- CICAFE. (2020). Guía técnica para el cultivo del café. *Instituto de Café de Costa Rica*, 2(1), 94. <https://www.icafe.cr/wp-content/uploads/cicafe/documentos/GUIA-TECNICA.pdf>
- Climate Data. (2024). *Clima Sabanilla: Temperatura, Climograma y Tabla climática para Sabanilla*. <https://es.climate-data.org/americadelnorte/costa-rica/alajuela/sabanilla-495106/>
- Cuevas Urióstegui, M. L., Reynaga Obregón, J., & Garduño Espinosa, J. (1994). Tau b de Kendall para datos agrupados . Señalamiento de discrepancia. *Médica del Instituto Mexicano del Seguro Social*, 32, 299–302. <http://137.184.230.111:8080/bitstream/62000/3350/1/22.pdf>
- DJI. (2023). *Mavic 3M - Manual de usuario 1.0* (1a ed.). https://dl.djicdn.com/downloads/DJI_Mavic_3_Enterprise/20230303/DJI_Mavic_3M_User_Manual_ES.pdf
- Eastman, J., Sangermano, F., Machado, E., Rogan, J., & Anyamba, A. (2013). Global Trends in Seasonality of Normalized Difference Vegetation Index (NDVI), 1982–2011. *Remote Sensing*, 5(10), 4799–4818. <https://doi.org/10.3390/rs5104799>
- esri. (s/f). *Train Support Vector Machine Classifier (Image Analyst)*. <https://pro.arcgis.com/en/pro-app/latest/tool-reference/image-analyst/train-support-vector-machine-classifier.htm>

- Fan, J., Yao, Y., Tang, Q., Zhang, X., Xu, J., Yu, R., Liu, L., Xie, Z., Ning, J., & Zhang, L. (2024). A Hybrid Index for Monitoring Burned Vegetation by Combining Image Texture Features with Vegetation Indices. *Remote Sensing*, 16(9), 1539. <https://doi.org/10.3390/rs16091539>
- Farfán-Valencia, F. (2021). Caficultura bajo sombrío. En *Guía más agronomía, más productividad, más calidad* (pp. 117–131). Cenicafé. https://doi.org/10.38141/10791/0014_8
- Fiallos, G. (2021). La Correlación de Pearson y el proceso de regresión por el Método de Mínimos Cuadrados. *Ciencia Latina Revista Científica Multidisciplinar*, 5(3), 2491–2509. https://doi.org/10.37811/cl_rcm.v5i3.466
- Flores-Tapia, C. E., & Flores-Cevallos, K. L. (2021). PRUEBAS PARA COMPROBAR LA NORMALIDAD DE DATOS EN SHAPIRO-WILK Y KOLMOGÓROV-SMIRNOV TESTS TO VERIFY THE NORMALITY OF DATA IN PRODUCTION PROCESSES. *Rev ciencias sociales y humanísticas*, 23(2), 83–106. https://www.researchgate.net/profile/Carlos-Flores-Tapia/publication/371337094_Pruebas_para_comprobar_la_normalidad_de_los_datos_en_procesos_productivos_Anderson-Darling_Ryan-Joiner_Shapiro-Wilk_y_Kolmogorov-Smirnov/links/647f577b2cad460a1bf9fda4/Pruebas-
- García-Cárdenas, D. A. (2018). *Dinámica de los índices de vegetación NDVI, GNDVI Y GVI en un cultivo de arroz a partir de imágenes aéreas multiespectrales tomadas por drones*. [Universidad de Pamplona]. <http://repositoriodspace.unipamplona.edu.co/jspui/handle/20.500.12744/4603>
- Garro Chacón, L. E. (2020). *Diseño de un sistema agroforestal en fincas ecológicas de café para la conectividad de dos áreas protegidas en providencia de copey, costa rica*. 60. https://repositoriotec.tec.ac.cr/bitstream/handle/2238/12399/TFG_Luis_Garro_Chacón.pdf?sequence=1&isAllowed=y
- Gavito Muñoz, J. cesar. (2019). *Sistema Inteligente para Detección Temprana de la Roya en Cafetales* [Centro de Ingeniería y Desarrollo Industrial]. <http://cidesi.repositorioinstitucional.mx/jspui/handle/1024/422>

- Guo, Z., Kurban, A., Ablekim, A., Wu, S., de Voorde, T. Van, Azadi, H., De Maeyer, P., & Dufatanye Umwall, E. (2021). Estimation of photosynthetic and non-photosynthetic vegetation coverage in the lower reaches of tarim river based on sentinel-2a data. *Remote Sensing*, 13(8). <https://doi.org/10.3390/rs13081458>
- ICAFE. (2022). *El Mejor Café del Mundo*. <https://www.icafe.cr/nuestro-cafe/el-mejor-cafe-del-mundo/>
- ICAFE. (2023). *Producción de café de la cosecha 22-23 supera en un 14.14% a la producción de la cosecha anterior*. <https://www.icafe.cr/cosecha22-23/>
- INM. (2024). *Ingreso de la onda tropical n.º15*. Aviso Meteorológico. https://www.imn.ac.cr/web/imn/avisos-meteorologicos/-/asset_publisher/eaLoNV1oHSZ2/content/aviso-meteorologico-imn-020724-0930am?redirect=https%253A%252F%252Fwww.imn.ac.cr%252Fweb%252Fimn%252Favisos-meteorologicos%253Fp_p_id%253D101_INSTANCE_eaLoNV1oHSZ2%2526p_p_lifecycle%25
- Jin, H., Köppl, C. J., Fischer, B. M. C., Rojas-Conejo, J., Johnson, M. S., Morillas, L., Lyon, S. W., Durán-Quesada, A. M., Suárez-Serrano, A., Manzoni, S., & Garcia, M. (2021). Drone-based hyperspectral and thermal imagery for quantifying upland rice productivity and water use efficiency after biochar application. *Remote Sensing*, 13(10), 1866. <https://doi.org/10.3390/rs13101866>
- Katsuhama, N., Imai, M., Naruse, N., & Takahashi, Y. (2018). Discrimination of areas infected with coffee leaf rust using a vegetation index. *Remote Sensing Letters*, 9(12), 1186–1194. <https://doi.org/10.1080/2150704X.2018.1519269>
- López-Rivera, L. A. (2021). *Un Modelo De Crecimiento Urbano Vertical Con Factores Caracteristicos Basado En Inteligencia Artificial*. [UAEMEX]. <http://hdl.handle.net/20.500.11799/112219>

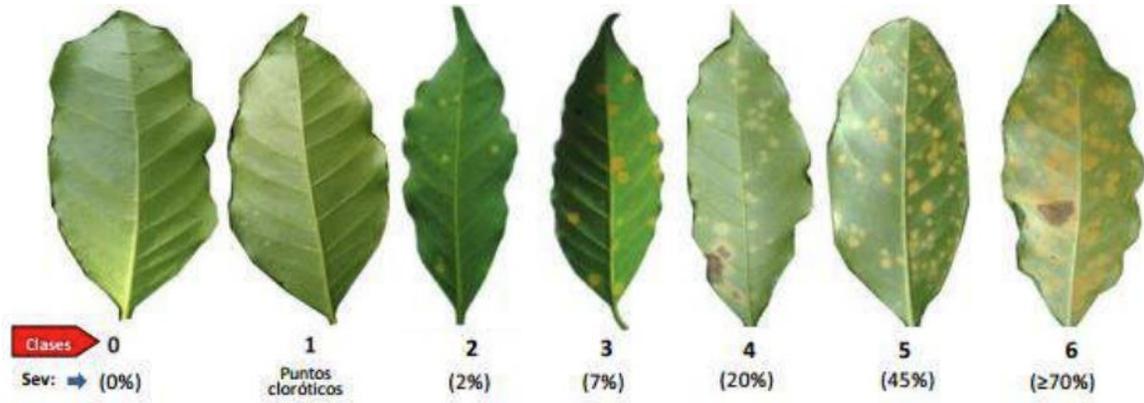
- Marin, D. B., Ferraz, G. A. e. S., Santana, L. S., Barbosa, B. D. S., Barata, R. A. P., Osco, L. P., Ramos, A. P. M., & Guimarães, P. H. S. (2021). Detecting coffee leaf rust with UAV-based vegetation indices and decision tree machine learning models. *Computers and Electronics in Agriculture*, *190*(September). <https://doi.org/10.1016/j.compag.2021.106476>
- Miñambres-Vidal, M. (2023). *GIS-based methodology and software tools for oil pipeline routing planning. A case study of an hypothetical connection between Valencia and Alicante*. [UNIVERSITAT POLITÈCNICA DE VALÈNCIA]. <https://riunet.upv.es/handle/10251/197551>
- Ortiz-Pinilla, J., & Ortiz-Rico, A. F. (2021). ¿Pearson y Spearman, coeficientes intercambiables? *Comunicaciones en Estadística*, *14*(1), 53–63. <https://doi.org/10.15332/23393076.6769>
- Padilla, J. A. (2018). *Correlación de Pearson, Spearman, tamaño del efecto, potencia estadística [Archivo de video]*. <https://www.youtube.com/watch?v=VnNbVVoN4qo&t=702s>
- Pinty, B., & Verstraete, M. M. (1992). GEMI: a non-linear index to monitor global vegetation from satellites. *Vegetatio*, *101*(1), 15–20. <https://doi.org/10.1007/BF00031911>
- Ramírez Dávila, J. F., Pérez-Constantino, A., Gutiérrez-Rodríguez, F., & Pérez-López, D. de J. (2023). Comportamiento espacial de roya del cafeto en Amatepec, Estado de México. *Acta Universitaria*, *33*(February 2021), 1–14. <https://doi.org/10.15174/au.2023.3870>
- Rendón-Macías, M. E., Zarco-Villavicencio, I. S., & Villasís-Keever, M. Á. (2021). Métodos estadísticos para el análisis del tamaño del efecto. *Revista Alergia México*, *68*(2), 128–136. <https://doi.org/10.29262/ram.v658i2.949>
- Revelo Luna, D. A., Mejía Manzano, J., Montoya-Bonilla, B., & Hoyos García, J. (2020). Análisis de los índices de vegetación NDVI, GNDVI y NDRE para la caracterización del cultivo de café (*Coffea arabica*). *Ingeniería y Desarrollo*, *38*(2), 298–312.
- Sáenz, K., & Tamez, G. (2014). Muestreo estratificado. En *Métodos y Técnicas Cualitativas y Cuantitativas Aplicables a la Investigación en Ciencias Sociales*. <http://www.tirant.com/mex/libro/metodos-y-tecnicas-cualitativas-y-cuantitativas-aplicables->

- Sánchez-Méndez, A. G., & Arguijo-Hernández, S. P. (2018). Análisis de imágenes multiespectrales para la detección de cultivos y detección de plagas y enfermedades en la producción de café. *Research in Computing Science*, 147(7), 309–317. <https://doi.org/10.13053/rscs-147-7-24>
- SENASICA. (2019). Roya del cafeto *Hemileia vastatrix* Verkeley & Brome Ficha Técnica No. 40. *Dirección General de Sanidad Vegetal-Programa de Vigilancia Epidemiológica Fitosanitaria*. Ciudad, 40, 1–23. <http://www.cesaveson.com/files/docs/campanas/vigilancia/fichas2016/ROYAcafeto.pdf>
- Singh, S. K., Meraj, G., Mondal, N., Bera, A. K., Verma, M. K., Tomar, J. S., & Kanga, S. (2021). Assessing seasonal vegetation dynamics over parts of Thar Desert using geospatial techniques. *ANGRAU Research Journal*, 49(3), 105–109.
- Sobrino, J. A., & Raissouni, N. (2000). *Teledetección* (J. A. Sobrino (ed.)). AECID.
- Thao, N. T. T., Khoi, D. N., Denis, A., Viet, L. Van, Wellens, J., & Tychon, B. (2022). Early Prediction of Coffee Yield in the Central Highlands of Vietnam Using a Statistical Approach and Satellite Remote Sensing Vegetation Biophysical Variables. *Remote Sensing*, 14(13), 2975. <https://doi.org/10.3390/rs14132975>
- Turan, J., Višacki, V., Oparnica, S., Sedlar, A., Pavol, F., & Burg, P. (2019). Variability of normalized vegetative indices in sugar beet. *Savremena poljoprivredna tehnika*, 45(3), 119–126. <https://doi.org/10.5937/SavPoljTeh1903119T>
- Velásquez, D., Sánchez, A., Sarmiento, S., Velásquez, C., Toro, M., Montoya, E., Trefftz, H., Maiza, M., & Sierra, B. (2021). A Cyber-Physical Data Collection System Integrating Remote Sensing and Wireless Sensor Networks for Coffee Leaf Rust Diagnosis. *Sensors (Basel, Switzerland)*, 21(16), 5474. <https://doi.org/10.3390/s21165474>

APÉNDICE

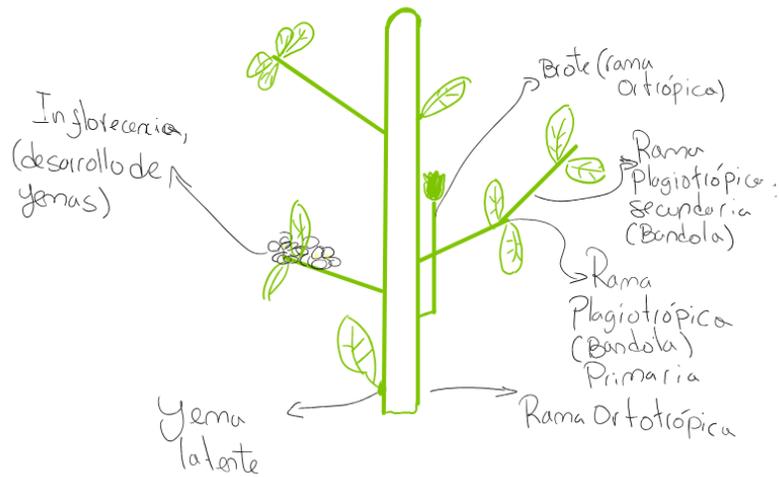
Apéndice A

Figura A 1. Clasificación con relación a la severidad de la roya presente en la hoja.



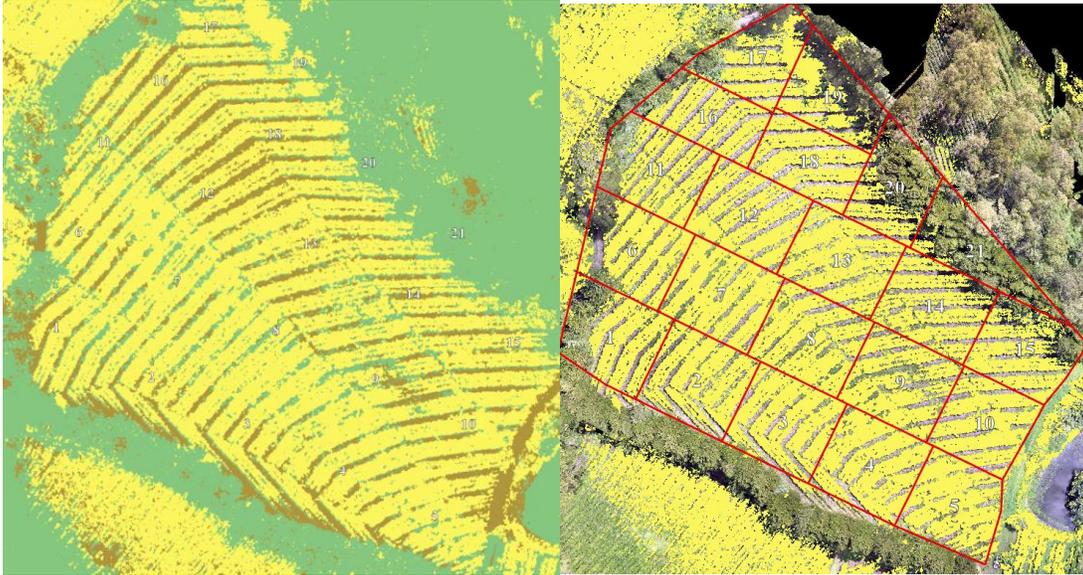
Fuente: Servicio Nacional de Sanidad, Inocuidad y Calidad Agroalimentaria (2016).

Figura A 2. Anatomía de un cafeto.



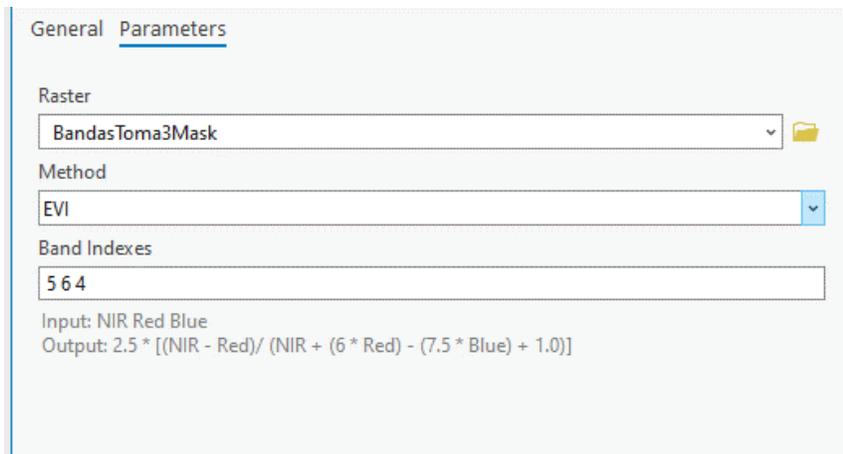
Nota: Elaboración propia, Fuente de la Información: (Arcilla, 2007)

Figura A 3. Máscara de Cafetos.



Nota: la zona amarilla representa los sembradíos, en este caso, los cafetos. Captura de pantalla del programa ArcGIS Pro.

Figura A 4. Cálculo de índices de vegetación con la herramienta “Band Arithmetic Properties”.



Nota: Captura de pantalla del programa ArcGIS Pro.

Figura A 5. Gráfico de caja para el IV GNDVI en tres momentos en el tiempo

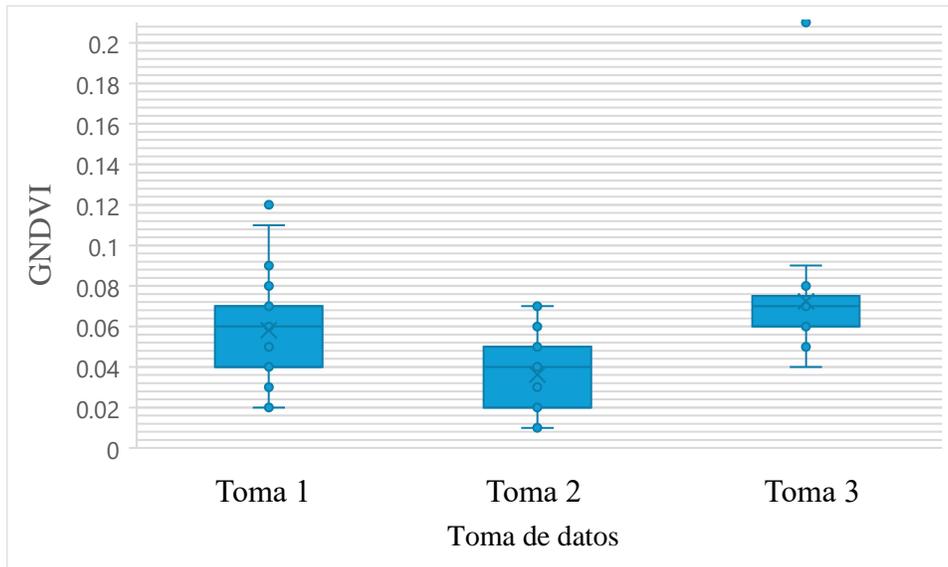


Figura A 6. Gráfico de caja para el IV GEMI en tres momentos en el tiempo

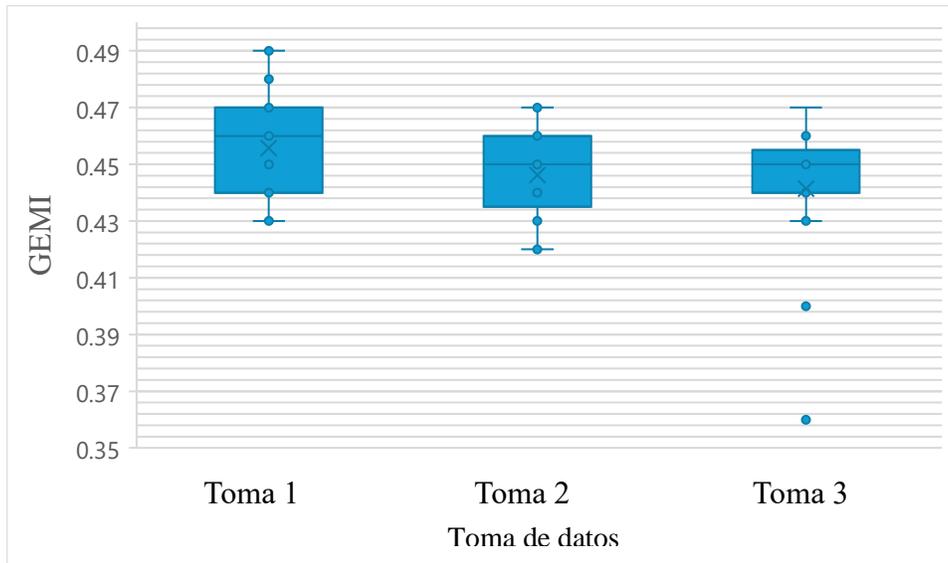


Figura A 7. Gráfico de caja para el IV NDVIre en tres momentos en el tiempo (tomas de datos).

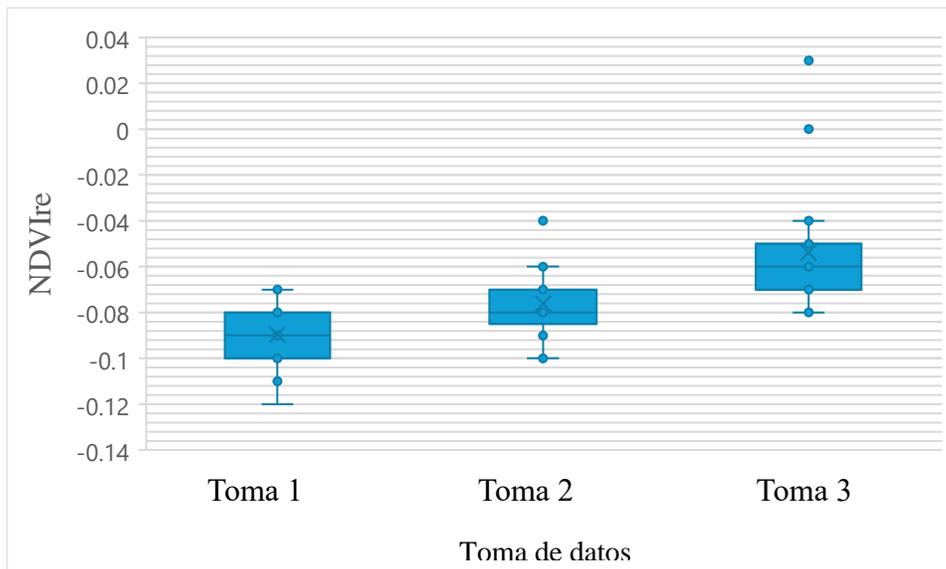
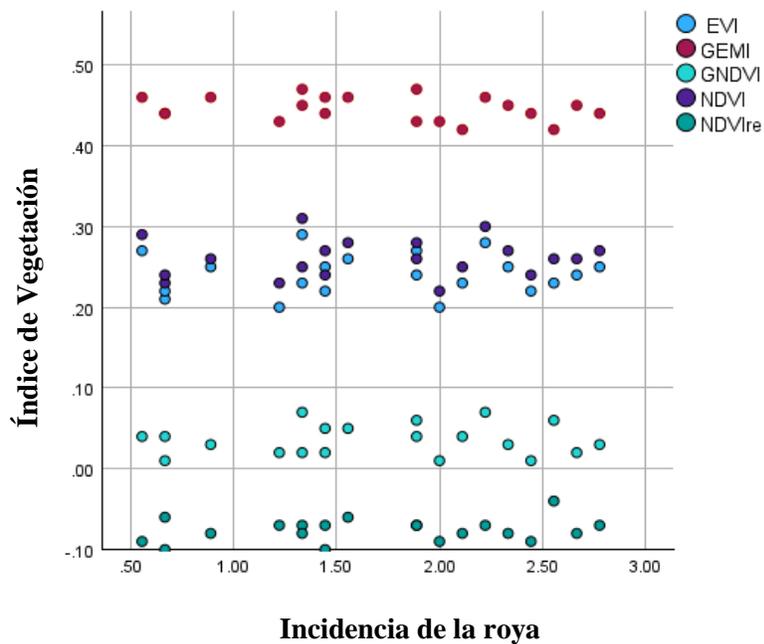
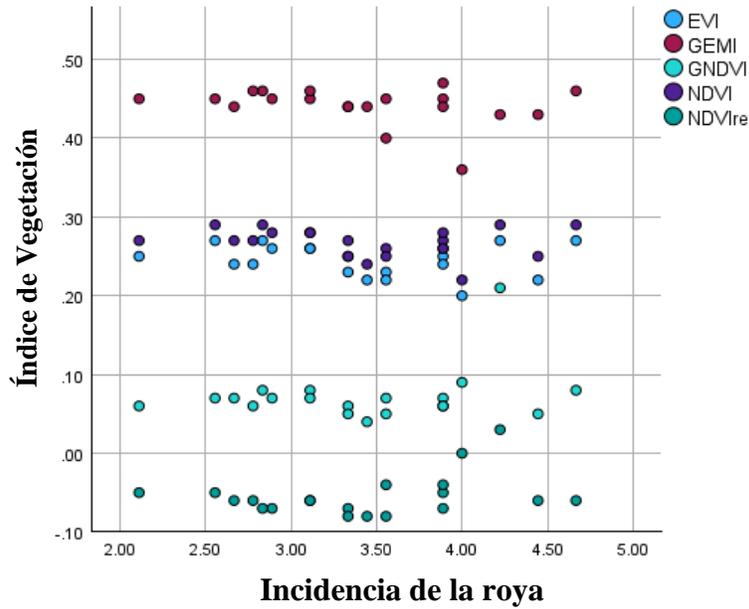


Figura A 8. Gráfico de puntos de Incidencia de la roya con IV calculados de la toma de datos 2.



Nota: Captura de pantalla del programa SPSS.

Figura A 9. Gráfico de puntos de Incidencia de la roya con IV calculados de la toma de datos 3.



Nota: Captura de pantalla del programa SPSS.

Figura A 10. Gráfico de caja para Los IV calculados (Toma de datos 1).

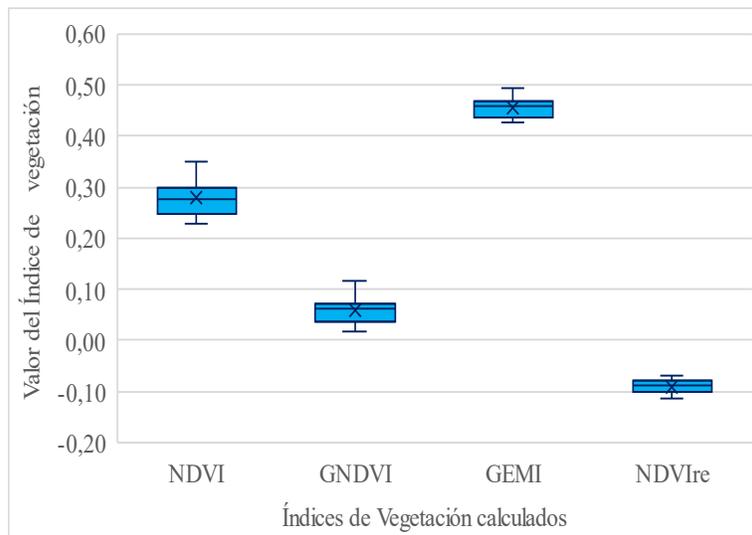


Figura A 11. Gráfico de caja para Los IV calculados (Toma de datos 2).

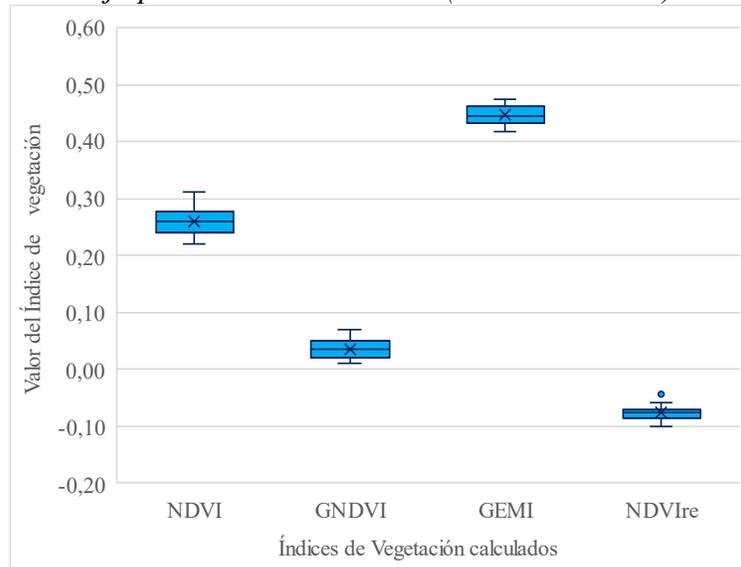


Figura A 12. Gráfico de caja para Los IV calculados (Toma de datos 3).

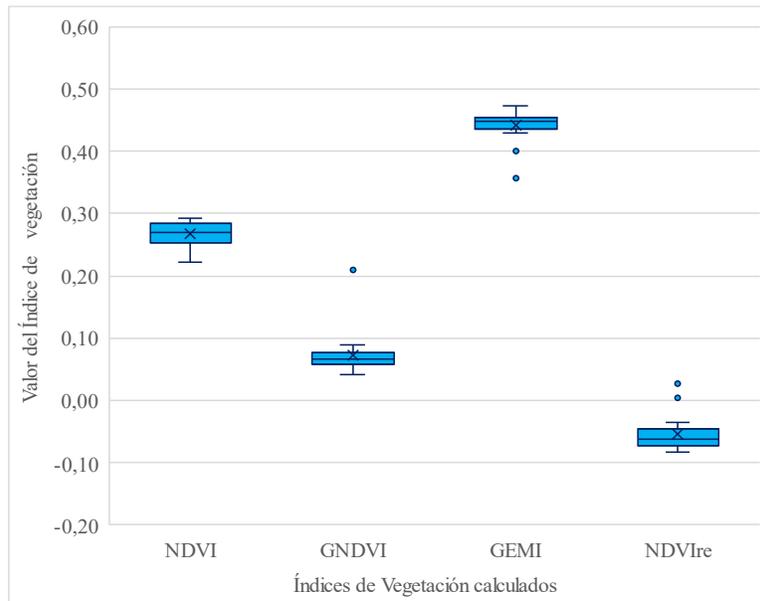


Figura A 13. Proceso de bandas para el cálculo de los IV.

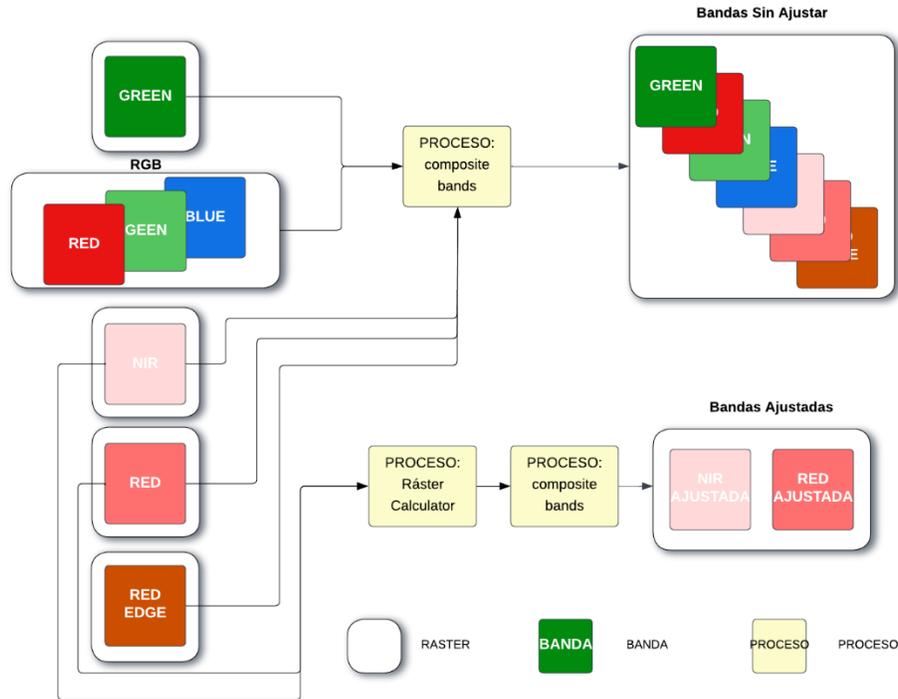


Figura A 14. Ortomosaicos de la toma de datos 1.



Nota: de izquierda a derecha tenemos GREEN, RGB, NIR, RED, RED EDGE. Para ver cada imagen individualmente, revisar el [Figura C 1](#) [Figura C 2](#) [Figura C 3](#) [Figura C 4](#) y [Figura C 5](#) respectivamente.

Tabla A 1. Valores referenciales para el tamaño del efecto de las diferentes pruebas de significación estadística.

Prueba	Símbolo	Pequeño	Mediano	Grande
Pruebas T:	d	0,20	0,50	0,80
ANOVA	f	0,10	0,25	0,40
ANOVA Factorial:	η_p^2 / f	0,01	0,06	0,14
Chi Cuadrado:	w / φ	0,10	0,30	0,50
Regresión	f^2	0,02	0,15	0,35

Nota: Para los CC se utilizan los parámetros de Chi Cuadrado (Cárdenas Castro, 2014).

Tabla A 2. Índices de vegetación a evaluar.

Nombre	Siglas	Fórmula
Normalized Difference Vegetation Index	NDVI	$NDVI = \frac{(NIR - RED)}{(NIR + RED)}$
Green Normalized Difference Vegetation Index	GNDVI	$GNDVI = \frac{(NIR - Green)}{(NIR + Green)}$
Global Environmental Monitoring Index	GEMI	$GEMI = eta * (1 - 0,25 * eta) - \frac{(Red - 0,125)}{(1 - Red)}$
		$eta = \frac{2 * (NIR^2 - Red^2) + 1,5 * NIR + 0,5 * Red}{NIR + Red + 0,5}$
The Red-Edge NDVI	NDVIre	$NDVIre = \frac{NIR - RE}{NIR + RE}$

Fuente: ArcGIS Pro, (s. f.). Nota: **NIR:** Banda cerca del infrarrojo (generalmente banda 5), **RED:** Banda roja (generalmente banda 4), **BLUE:** Banda azul (generalmente banda 2), **G:** Ganancia (valor estándar de 2,5), **L:** Factor de corrección del suelo (1,0), **C1:** Correctores de la dispersión atmosférica (6,0), **C2:** Correctores de la dispersión atmosférica (7,5), **GREEN:** Banda verde.

Tabla A 3. Promedio de incidencia de la roya por bloque.

Bloque	Toma de datos 1 (02/04)	Toma de datos 2 (22/05)	Toma de datos 3 (02/07)
1	1,56	1,44	2,56
2	2,11	0,67	3,33
3	2,56	1,44	2,89
4	3,22	1,33	2,83
5	1,33	2,22	4,67
6	3,22	0,67	3,89
7	1,78	1,56	3,11
8	2,22	1,89	2,67
9	1,89	0,89	3,33
10	2,56	0,56	3,11
11	1,67	1,22	2,11
12	2,22	1,33	3,56
13	2,00	2,44	3,44
14	3,44	2,67	4,44
15	3,33	2,78	3,89
16	1,33	2,00	2,78
17	1,11	1,89	4,22
18	2,78	2,33	3,89
19	2,33	2,56	4,00
20	2,67	2,11	3,56
Promedio	2,27	1,70	3,44
Des.	1,11	1,36	1,40

Tabla A 4. Desviación estándar de la incidencia de la Roya en los bloques.

Bloque	Toma de datos 1	Toma de datos 2	Toma de datos 3	Promedio
1	0,53	0,73	0,53	0,77
2	0,78	0,71	0,87	1,34
3	1,13	0,53	0,78	1,03
4	1,30	1,22	0,98	1,44
5	0,71	0,67	1,50	1,82
6	1,20	0,87	1,27	1,78
7	0,67	1,33	1,17	1,26
8	0,83	0,78	0,50	0,76
9	0,60	1,17	1,32	1,45
10	0,73	1,13	1,05	1,47
11	0,71	0,44	0,60	0,68
12	0,97	1,22	0,88	1,36
13	0,87	1,74	0,73	1,31
14	1,13	1,41	0,88	1,34
15	1,22	2,05	1,05	1,52

Bloque	Toma de datos 1	Toma de datos 2	Toma de datos 3	Promedio
16	0,71	1,41	0,83	1,16
17	0,33	1,54	0,83	1,67
18	1,09	1,66	0,78	1,36
19	1,32	1,74	0,87	1,51
20	0,87	0,93	0,53	0,97
Promedio	1,11	1,36	1,12	1,40

Tabla A 5. Promedio de cada IV para cada bloque, en tres tomas de datos.

Toma de datos	Bloque	NDVI	GNDVI	GEMI	NDVIre
1	1	0,28	0,06	0,47	-0,09
	2	0,27	0,05	0,46	-0,10
	3	0,32	0,09	0,48	-0,08
	4	0,35	0,11	0,49	-0,07
	5	0,35	0,12	0,49	-0,07
	6	0,27	0,06	0,46	-0,08
	7	0,30	0,06	0,47	-0,08
	8	0,32	0,08	0,47	-0,07
	9	0,29	0,07	0,46	-0,08
	10	0,30	0,07	0,46	-0,08
	11	0,25	0,04	0,44	-0,08
	12	0,25	0,03	0,45	-0,09
	13	0,24	0,02	0,43	-0,10
	14	0,25	0,04	0,44	-0,10
	15	0,28	0,07	0,45	-0,08
	16	0,23	0,02	0,43	-0,11
	17	0,28	0,07	0,46	-0,09
	18	0,25	0,02	0,44	-0,11
	19	0,26	0,04	0,43	-0,12
	20	0,25	0,04	0,43	-0,11
	Promedio	0,28	0,06	0,46	-0,09
2	1	0,24	0,02	0,44	-0,10
	2	0,23	0,01	0,44	-0,10
	3	0,27	0,05	0,46	-0,07
	4	0,31	0,07	0,47	-0,07
	5	0,30	0,07	0,46	-0,07
	6	0,24	0,04	0,44	-0,06
	7	0,28	0,05	0,46	-0,06
	8	0,28	0,06	0,47	-0,07

Toma de datos	Bloque	NDVI	GNDVI	GEMI	NDVIre
	9	0,26	0,03	0,46	-0,08
	10	0,29	0,04	0,46	-0,09
	11	0,23	0,02	0,43	-0,07
	12	0,25	0,02	0,45	-0,08
	13	0,24	0,01	0,44	-0,09
	14	0,26	0,02	0,45	-0,08
	15	0,27	0,03	0,44	-0,07
	16	0,22	0,01	0,43	-0,09
	17	0,26	0,04	0,43	-0,07
	18	0,27	0,03	0,45	-0,08
	19	0,26	0,06	0,42	-0,04
	20	0,25	0,04	0,42	-0,08
	Promedio	0,26	0,04	0,45	-0,08
	1	0,29	0,07	0,45	-0,05
	2	0,27	0,06	0,44	-0,07
	3	0,28	0,07	0,45	-0,07
	4	0,29	0,08	0,46	-0,07
	5	0,29	0,08	0,46	-0,06
	6	0,27	0,07	0,45	-0,05
	7	0,28	0,08	0,45	-0,06
	8	0,27	0,07	0,44	-0,06
	9	0,25	0,05	0,44	-0,08
	10	0,28	0,07	0,46	-0,06
3	11	0,27	0,06	0,45	-0,05
	12	0,26	0,05	0,45	-0,08
	13	0,24	0,04	0,44	-0,08
	14	0,25	0,05	0,43	-0,06
	15	0,26	0,06	0,44	-0,04
	16	0,27	0,06	0,46	-0,06
	17	0,29	0,21	0,43	0,03
	18	0,28	0,06	0,47	-0,07
	19	0,22	0,09	0,36	0,00
	20	0,25	0,07	0,40	-0,04
	Promedio	0,27	0,07	0,44	-0,05

Tabla A 6, Información de los índices de vegetación de las tres tomas de datos (abril, mayo, junio)

Toma	Índice	Bloque	MIN	MAX	Rango	Promedio	STD	Mediana	PCT90
1	NDVI	1	-0,486	0,566	1,052	0,278	0,112	0,292	0,410
		2	-0,498	0,581	1,079	0,273	0,113	0,288	0,401
		3	-0,452	0,557	1,009	0,325	0,088	0,336	0,425
		4	-0,325	0,569	0,894	0,350	0,088	0,361	0,453
		5	-0,368	0,573	0,941	0,349	0,094	0,364	0,456
		6	-0,478	0,574	1,052	0,274	0,109	0,281	0,409
		7	-0,367	0,581	0,948	0,299	0,101	0,309	0,420
		8	-0,348	0,555	0,903	0,316	0,096	0,330	0,427
		9	-0,325	0,549	0,874	0,288	0,114	0,304	0,421
		10	-0,375	0,553	0,928	0,298	0,099	0,308	0,416
		11	-0,494	0,552	1,046	0,246	0,130	0,260	0,397
		12	-0,488	0,578	1,066	0,255	0,115	0,266	0,392
		13	-0,392	0,532	0,924	0,241	0,111	0,253	0,371
		14	-0,426	0,526	0,952	0,249	0,119	0,265	0,382
		15	-0,378	0,527	0,905	0,279	0,104	0,292	0,397
		16	-0,461	0,585	1,046	0,228	0,138	0,239	0,396
		17	-0,457	0,582	1,039	0,280	0,110	0,293	0,408
		18	-0,496	0,619	1,116	0,248	0,118	0,264	0,382
		19	-0,611	0,516	1,127	0,258	0,108	0,274	0,381
		20	-0,430	0,508	0,938	0,247	0,113	0,263	0,373
		Total	-0,611	0,619	0,992	0,279	0,109	0,292	0,406
1	GNDVI	1	-0,446	0,475	0,921	0,064	0,092	0,066	0,182
		2	-0,368	0,388	0,755	0,048	0,087	0,050	0,159
		3	-0,330	0,359	0,688	0,088	0,072	0,090	0,179
		4	-0,337	0,364	0,701	0,114	0,073	0,117	0,206
		5	-0,283	0,399	0,682	0,115	0,078	0,122	0,208
		6	-0,346	0,389	0,735	0,062	0,080	0,062	0,166
		7	-0,352	0,388	0,740	0,064	0,082	0,065	0,170
		8	-0,287	0,365	0,652	0,080	0,078	0,082	0,180
		9	-0,356	0,386	0,742	0,069	0,093	0,076	0,186
		10	-0,280	0,381	0,662	0,069	0,079	0,072	0,169
		11	-0,374	0,451	0,825	0,042	0,088	0,044	0,155
		12	-0,471	0,385	0,855	0,026	0,082	0,026	0,133
		13	-0,310	0,339	0,649	0,016	0,077	0,014	0,116
		14	-0,337	0,362	0,699	0,036	0,078	0,035	0,137
		15	-0,275	0,369	0,644	0,072	0,075	0,071	0,169
		16	-0,435	0,397	0,831	0,017	0,096	0,013	0,144
		17	-0,420	0,433	0,853	0,068	0,088	0,069	0,181
		18	-0,484	0,385	0,870	0,020	0,078	0,021	0,118
		19	-0,368	0,330	0,698	0,045	0,078	0,049	0,142
		20	-0,382	0,376	0,758	0,038	0,074	0,040	0,130
		Total	-0,484	0,475	0,748	0,058	0,081	0,059	0,161
1	GEMI	1	-1,437	0,669	2,106	0,465	0,070	0,470	0,550
		2	-0,473	0,661	1,134	0,458	0,068	0,463	0,537
		3	-0,042	0,647	0,689	0,484	0,055	0,486	0,552

Toma	Índice	Bloque	MIN	MAX	Rango	Promedio	STD	Mediana	PCT90
		4	-0,258	0,660	0,918	0,494	0,054	0,496	0,562
		5	0,079	0,649	0,571	0,485	0,056	0,486	0,557
		6	-1,470	0,654	2,124	0,460	0,065	0,459	0,544
		7	-0,021	0,650	0,671	0,469	0,059	0,470	0,542
		8	0,079	0,643	0,564	0,474	0,054	0,476	0,541
		9	-0,008	0,648	0,656	0,460	0,063	0,461	0,537
		10	-0,052	0,641	0,693	0,463	0,054	0,463	0,532
		11	-0,333	0,654	0,988	0,443	0,076	0,448	0,532
		12	-1,472	0,653	2,126	0,447	0,069	0,449	0,528
		13	-0,622	0,627	1,248	0,434	0,059	0,436	0,505
		14	-0,163	0,631	0,794	0,437	0,063	0,441	0,510
		15	-0,063	0,613	0,676	0,452	0,055	0,454	0,517
		16	-0,640	0,662	1,302	0,434	0,084	0,439	0,533
		17	-0,525	0,661	1,186	0,461	0,067	0,466	0,539
		18	-1,372	0,632	2,003	0,437	0,066	0,442	0,512
		19	-0,270	0,614	0,884	0,428	0,072	0,441	0,504
		20	-0,101	0,606	0,707	0,428	0,070	0,436	0,507
		Total	-1,472	0,669	1,102	0,456	0,064	0,459	0,532
		1	-0,345	0,138	0,482	-0,095	0,045	-0,098	-0,036
		2	-0,273	0,138	0,412	-0,097	0,044	-0,100	-0,041
		3	-0,265	0,118	0,383	-0,083	0,042	-0,086	-0,028
		4	-0,260	0,177	0,437	-0,070	0,046	-0,073	-0,010
		5	-0,277	0,129	0,405	-0,070	0,044	-0,070	-0,014
		6	-0,329	0,172	0,501	-0,082	0,046	-0,084	-0,023
		7	-0,303	0,148	0,451	-0,081	0,047	-0,084	-0,020
		8	-0,278	0,153	0,431	-0,075	0,045	-0,077	-0,016
		9	-0,287	0,156	0,443	-0,078	0,053	-0,079	-0,009
		10	-0,270	0,135	0,405	-0,085	0,044	-0,087	-0,027
1	NDVIre	11	-0,373	0,229	0,602	-0,082	0,050	-0,083	-0,019
		12	-0,402	0,245	0,647	-0,089	0,049	-0,091	-0,026
		13	-0,338	0,115	0,453	-0,103	0,043	-0,108	-0,047
		14	-0,288	0,109	0,397	-0,100	0,041	-0,104	-0,046
		15	-0,326	0,148	0,474	-0,077	0,041	-0,081	-0,022
		16	-0,411	0,260	0,671	-0,106	0,054	-0,109	-0,040
		17	-0,518	0,211	0,729	-0,092	0,053	-0,096	-0,023
		18	-0,410	0,225	0,635	-0,111	0,053	-0,112	-0,045
		19	-0,392	0,190	0,581	-0,115	0,060	-0,122	-0,028
		20	-0,410	0,125	0,535	-0,111	0,050	-0,116	-0,044
		Total	-0,518	0,260	0,504	-0,090	0,047	-0,093	-0,028
		1	-0,359	0,503	0,863	0,241	0,103	0,256	0,361
		2	-0,441	0,486	0,927	0,231	0,114	0,250	0,357
		3	-0,359	0,505	0,864	0,273	0,100	0,293	0,380
2	NDVI	4	-0,314	0,506	0,820	0,305	0,096	0,324	0,408
		5	-0,270	0,509	0,779	0,301	0,098	0,320	0,405
		6	-0,401	0,520	0,922	0,245	0,111	0,258	0,376
		7	-0,405	0,529	0,935	0,277	0,103	0,292	0,394

Toma	Índice	Bloque	MIN	MAX	Rango	Promedio	STD	Mediana	PCT90
		8	-0,309	0,512	0,821	0,285	0,095	0,302	0,388
		9	-0,358	0,528	0,886	0,264	0,114	0,287	0,390
		10	-0,337	0,501	0,838	0,285	0,103	0,308	0,394
		11	-0,432	0,506	0,938	0,226	0,113	0,242	0,355
		12	-0,419	0,516	0,935	0,251	0,110	0,273	0,371
		13	-0,408	0,495	0,903	0,244	0,099	0,258	0,357
		14	-0,292	0,517	0,809	0,263	0,093	0,282	0,362
		15	-0,313	0,465	0,778	0,270	0,091	0,290	0,364
		16	-0,386	0,489	0,876	0,223	0,115	0,242	0,354
		17	-0,327	0,504	0,831	0,261	0,106	0,286	0,371
		18	-0,356	0,508	0,864	0,267	0,109	0,295	0,376
		19	-0,460	0,490	0,949	0,257	0,101	0,275	0,368
		20	-0,293	0,484	0,777	0,248	0,099	0,267	0,359
		Total	-0,460	0,529	0,866	0,261	0,104	0,280	0,374
		1	-0,359	0,293	0,651	0,024	0,075	0,027	0,117
		2	-0,366	0,320	0,686	0,010	0,080	0,015	0,109
		3	-0,368	0,291	0,659	0,053	0,069	0,058	0,137
		4	-0,324	0,298	0,622	0,071	0,069	0,077	0,154
		5	-0,323	0,289	0,611	0,066	0,075	0,074	0,155
		6	-0,347	0,325	0,672	0,036	0,076	0,038	0,133
		7	-0,304	0,336	0,640	0,053	0,075	0,057	0,147
		8	-0,285	0,337	0,622	0,055	0,071	0,059	0,143
		9	-0,349	0,291	0,640	0,029	0,084	0,035	0,134
		10	-0,322	0,298	0,620	0,038	0,077	0,044	0,133
2	GNDVI	11	-0,363	0,296	0,658	0,016	0,076	0,019	0,112
		12	-0,308	0,271	0,578	0,015	0,075	0,019	0,108
		13	-0,342	0,292	0,633	0,006	0,069	0,008	0,092
		14	-0,332	0,294	0,626	0,017	0,070	0,021	0,103
		15	-0,353	0,285	0,638	0,026	0,068	0,032	0,108
		16	-0,374	0,325	0,699	0,008	0,081	0,011	0,109
		17	-0,429	0,340	0,768	0,035	0,081	0,038	0,137
		18	-0,305	0,289	0,594	0,028	0,072	0,033	0,117
		19	-0,308	0,352	0,660	0,063	0,071	0,065	0,152
		20	-0,364	0,304	0,668	0,040	0,070	0,043	0,126
		Total	-0,429	0,352	0,647	0,034	0,074	0,039	0,126
		1	-0,001	0,615	0,616	0,436	0,065	0,438	0,518
		2	-0,004	0,604	0,607	0,435	0,070	0,444	0,517
		3	-0,114	0,615	0,729	0,457	0,065	0,468	0,533
		4	0,104	0,624	0,520	0,473	0,062	0,483	0,545
		5	0,162	0,619	0,457	0,464	0,061	0,473	0,536
2	GEMI	6	-0,462	0,633	1,095	0,442	0,069	0,446	0,527
		7	-0,079	0,639	0,718	0,462	0,067	0,471	0,543
		8	0,141	0,622	0,482	0,468	0,064	0,477	0,544
		9	0,035	0,637	0,603	0,458	0,072	0,467	0,544
		10	-0,192	0,614	0,806	0,465	0,065	0,476	0,540
		11	-0,084	0,618	0,702	0,432	0,070	0,437	0,520

Toma	Índice	Bloque	MIN	MAX	Rango	Promedio	STD	Mediana	PCT90
		12	-0,013	0,637	0,649	0,454	0,072	0,463	0,540
		13	0,041	0,611	0,570	0,445	0,065	0,449	0,526
		14	0,149	0,605	0,456	0,445	0,060	0,453	0,517
		15	0,114	0,578	0,464	0,443	0,058	0,450	0,513
		16	-0,249	0,616	0,865	0,431	0,073	0,437	0,522
		17	0,131	0,593	0,462	0,432	0,064	0,442	0,507
		18	0,019	0,606	0,587	0,449	0,069	0,461	0,528
		19	-0,302	0,579	0,881	0,417	0,058	0,423	0,486
		20	0,026	0,569	0,543	0,418	0,058	0,422	0,492
		Total	-0,462	0,639	0,641	0,446	0,065	0,454	0,525
		1	-0,374	0,289	0,663	-0,100	0,053	-0,105	-0,033
		2	-0,284	0,150	0,435	-0,097	0,047	-0,099	-0,038
		3	-0,280	0,159	0,438	-0,073	0,044	-0,075	-0,018
		4	-0,243	0,174	0,417	-0,073	0,040	-0,075	-0,022
		5	-0,254	0,146	0,401	-0,074	0,041	-0,074	-0,024
		6	-0,283	0,191	0,474	-0,062	0,044	-0,065	-0,005
		7	-0,285	0,185	0,470	-0,060	0,045	-0,063	-0,001
		8	-0,261	0,167	0,427	-0,065	0,044	-0,067	-0,009
		9	-0,286	0,164	0,450	-0,077	0,048	-0,077	-0,016
		10	-0,256	0,129	0,385	-0,093	0,041	-0,095	-0,041
2	NDVIre	11	-0,301	0,226	0,527	-0,073	0,049	-0,076	-0,011
		12	-0,317	0,171	0,488	-0,082	0,050	-0,085	-0,018
		13	-0,262	0,106	0,369	-0,090	0,039	-0,093	-0,040
		14	-0,231	0,146	0,377	-0,077	0,040	-0,081	-0,025
		15	-0,245	0,160	0,405	-0,073	0,044	-0,077	-0,016
		16	-0,364	0,192	0,556	-0,088	0,053	-0,092	-0,019
		17	-0,379	0,244	0,624	-0,070	0,059	-0,075	0,006
		18	-0,327	0,214	0,541	-0,081	0,046	-0,085	-0,021
		19	-0,303	0,231	0,535	-0,043	0,055	-0,047	0,029
		20	-0,464	0,462	0,926	-0,077	0,049	-0,080	-0,015
		Total	-0,464	0,462	0,495	-0,077	0,046	-0,079	-0,017
		1	-0,437	0,550	0,988	0,293	0,089	0,307	0,393
		2	-0,315	0,508	0,823	0,272	0,082	0,283	0,366
		3	-0,262	0,477	0,739	0,285	0,079	0,299	0,374
		4	-0,277	0,482	0,759	0,287	0,083	0,303	0,378
		5	-0,262	0,490	0,752	0,293	0,079	0,308	0,380
		6	-0,490	0,487	0,977	0,268	0,099	0,289	0,370
		7	-0,412	0,479	0,891	0,280	0,080	0,293	0,368
3	NDVI	8	-0,216	0,487	0,702	0,267	0,075	0,276	0,356
		9	-0,282	0,475	0,757	0,253	0,083	0,263	0,353
		10	-0,205	0,485	0,690	0,285	0,077	0,298	0,372
		11	-0,314	0,487	0,801	0,271	0,085	0,285	0,365
		12	-0,394	0,492	0,886	0,257	0,080	0,268	0,350
		13	-0,213	0,507	0,720	0,241	0,080	0,248	0,340
		14	-0,247	0,456	0,703	0,248	0,074	0,256	0,336
		15	-0,223	0,477	0,700	0,263	0,073	0,270	0,349

Toma	Índice	Bloque	MIN	MAX	Rango	Promedio	STD	Mediana	PCT90
		16	-0,421	0,537	0,958	0,267	0,093	0,279	0,373
		17	-0,382	0,597	0,979	0,287	0,111	0,301	0,413
		18	-0,466	0,507	0,972	0,282	0,090	0,297	0,377
		19	-0,359	0,528	0,887	0,222	0,100	0,224	0,355
		20	-0,466	1,000	1,466	0,247	0,089	0,250	0,351
		Total	-0,490	1,000	0,858	0,268	0,085	0,280	0,366
		1	-0,321	0,363	0,684	0,071	0,069	0,074	0,156
		2	-0,306	0,335	0,641	0,061	0,065	0,066	0,141
		3	-0,240	0,322	0,562	0,075	0,060	0,079	0,148
		4	-0,228	0,314	0,542	0,078	0,064	0,084	0,153
		5	-0,281	0,312	0,593	0,083	0,062	0,088	0,157
		6	-0,349	0,468	0,817	0,075	0,078	0,080	0,168
		7	-0,238	0,303	0,541	0,079	0,059	0,083	0,151
		8	-0,270	0,338	0,609	0,066	0,061	0,070	0,141
		9	-0,315	0,299	0,614	0,054	0,065	0,057	0,136
3	GNDVI	10	-0,258	0,316	0,574	0,068	0,062	0,069	0,146
		11	-0,325	0,402	0,728	0,063	0,070	0,067	0,147
		12	-0,308	0,377	0,685	0,054	0,066	0,056	0,134
		13	-0,319	0,332	0,651	0,042	0,065	0,042	0,125
		14	-0,265	0,316	0,581	0,046	0,062	0,048	0,122
		15	-0,251	0,308	0,559	0,063	0,063	0,065	0,141
		16	-0,528	0,556	1,085	0,058	0,092	0,058	0,163
		17	-0,604	0,732	1,336	0,210	0,235	0,205	0,530
		18	-0,305	0,401	0,706	0,062	0,071	0,063	0,149
		19	-0,356	0,511	0,867	0,089	0,074	0,094	0,171
		20	-0,336	1,000	1,336	0,069	0,073	0,073	0,149
		Total	-0,604	1,000	0,736	0,073	0,076	0,076	0,167
		1	-0,064	0,648	0,711	0,450	0,063	0,455	0,528
		2	0,097	0,605	0,508	0,444	0,053	0,448	0,511
		3	0,120	0,607	0,487	0,454	0,054	0,461	0,520
		4	0,076	0,608	0,532	0,456	0,055	0,464	0,522
		5	0,165	0,605	0,441	0,462	0,053	0,467	0,528
		6	-0,195	0,590	0,784	0,450	0,063	0,461	0,520
		7	-0,545	0,591	1,135	0,451	0,052	0,458	0,512
		8	0,169	0,579	0,409	0,438	0,047	0,443	0,497
		9	0,106	0,579	0,473	0,435	0,050	0,440	0,498
3	GEMI	10	0,162	0,602	0,440	0,463	0,051	0,470	0,525
		11	0,091	0,582	0,491	0,448	0,053	0,456	0,511
		12	0,019	0,648	0,630	0,452	0,055	0,456	0,519
		13	0,171	0,612	0,442	0,444	0,054	0,444	0,514
		14	0,170	0,589	0,418	0,432	0,060	0,441	0,502
		15	0,195	0,599	0,404	0,440	0,066	0,452	0,516
		16	0,054	0,622	0,568	0,457	0,061	0,464	0,531
		17	-0,019	0,652	0,670	0,430	0,103	0,460	0,551
		18	-0,328	0,637	0,964	0,473	0,062	0,483	0,541
		19	0,019	0,605	0,586	0,358	0,092	0,322	0,506

Toma	Índice	Bloque	MIN	MAX	Rango	Promedio	STD	Mediana	PCT90
		20	-0,107	0,599	0,706	0,401	0,083	0,403	0,510
		Total	-0,545	0,652	0,590	0,442	0,062	0,447	0,518
3	NDVIre	1	-0,307	0,300	0,607	-0,045	0,050	-0,049	0,018
		2	-0,269	0,161	0,430	-0,068	0,041	-0,070	-0,016
		3	-0,227	0,147	0,374	-0,067	0,038	-0,069	-0,019
		4	-0,241	0,126	0,368	-0,074	0,039	-0,077	-0,024
		5	-0,241	0,154	0,395	-0,061	0,039	-0,063	-0,012
		6	-0,203	0,174	0,377	-0,046	0,039	-0,049	0,003
		7	-0,238	0,177	0,415	-0,058	0,038	-0,060	-0,010
		8	-0,251	0,138	0,389	-0,063	0,038	-0,064	-0,016
		9	-0,240	0,123	0,362	-0,077	0,039	-0,080	-0,027
		10	-0,226	0,154	0,380	-0,064	0,040	-0,068	-0,011
		11	-0,276	0,199	0,475	-0,052	0,040	-0,053	-0,001
		12	-0,302	0,210	0,512	-0,078	0,046	-0,081	-0,020
		13	-0,327	0,196	0,523	-0,082	0,042	-0,086	-0,027
		14	-0,305	0,189	0,494	-0,058	0,048	-0,066	0,014
		15	-0,226	0,193	0,419	-0,036	0,057	-0,048	0,050
		16	-0,546	0,227	0,773	-0,062	0,048	-0,065	-0,002
		17	-0,586	0,437	1,022	0,027	0,093	0,013	0,146
		18	-0,243	0,164	0,407	-0,074	0,047	-0,079	-0,012
		19	-0,214	0,197	0,411	0,004	0,065	0,019	0,079
		20	-0,351	1,000	1,351	-0,040	0,069	-0,043	0,042
		Total	-0,586	1,000	0,524	-0,054	0,048	-0,057	0,008

Tabla A 7. Significancia (Sig.) de las correlaciones realizadas.

Toma	Coefficiente	NDVI	GNDVI	GEMI	NDVIre
1 (abril)	Tau-B de Kendall	0,774	0,883	0,883	0,843
	Spearman's rho	0,927	0,947	0,561	0,975
2 (mayo)	Tau-B de Kendall	0,559	0,938	0,410	0,443
	Spearman's rho	0,599	0,902	0,311	0,492
3 (julio)	Tau-B de Kendall	0,510	0,537	0,194	0,906
	Spearman's rho	0,421	0,551	0,255	0,298

Tabla A 8. Rangos para clasificación ordinal de los IV.

Intervalo	NDVI	GNDVI	GEMI	NDVIRE
MIN	0,220	0,010	0,358	-0,115
1	0,239	0,039	0,377	-0,095
2	0,257	0,067	0,397	-0,075
3	0,276	0,096	0,416	-0,054
4	0,294	0,124	0,436	-0,034
5	0,313	0,153	0,455	-0,014
6	0,331	0,182	0,474	0,006
MAX	0,350	0,210	0,494	0,027

Tabla A 9. Prueba de normalidad a variables reclasificados.

Variable	Estadística	df	Sig.	Variable	Estadística	df	Sig.
NDVI1	0,915	20	0,078	NDVI3	0,836	20	0,003
GNDVI1	0,871	20	0,012	GNDVI3	0,561	20	0,000
GEMI1	0,879	20	0,017	GEMI3	0,777	20	0,000
NDVIre1	0,780	20	0,000	NDVIre3	0,788	20	0,001
NDVI2	0,922	20	0,108	ABRIL	0,800	20	0,001
GNDVI2	0,760	20	0,000	MAYO	0,809	20	0,001
GEMI2	0,803	20	0,001	JULIO	0,798	20	0,001
NDVIre2	0,855	20	0,006				

Nota: Si $p \leq \alpha$ se rechaza la H_0 . Si $p > \alpha$ no se rechaza la hipótesis nula. $\alpha = 0,05$.

Tabla A 10. Prueba de normalidad Shapiro-Wilk en SPSS de los datos de incidencia de la roya e índices de vegetación.

Variable	Estadística	df	Sig. (p)	Variable	Satisfice	df	Sig. (p)
INCIDENCIA1	0,961	20	0,570	GNDVI2	0,931	20	0,162
INCIDENCIA2	0,953	20	0,409	GEMI2	0,935	20	0,190
INCIDENCIA3	0,985	20	0,983	NDVIre2	0,931	20	0,163
NDVI1	0,921	20	0,079	NDVI3	0,911	20	0,067
GNDVI1	0,941	20	0,103	GNDVI3	0,572	20	<0,001
GEMI1	0,926	20	0,250	GEMI3	0,755	20	<0,001
NDVIre1	0,906	20	0,129	NDVIre3	0,766	20	<0,001
NDVI2	0,974	20	0,833				

Nota: el número al final de cada variable corresponde a la toma de datos en la que se calculó el set de datos.

Tabla A 11. Objetivos del desarrollo sostenible presentes en la investigación.

ODS	Nombre	Descripción Del ODS	Meta de la investigación
ODS 2	Hambre cero	Este objetivo se enfoca en promover la sostenibilidad de los sistemas alimentarios.	Esta investigación contribuye a prácticas agrícolas más sostenibles y seguridad alimentaria.
ODS 12	Producción y consumo responsable	Este objetivo se enfoca en promover métodos más sostenibles para el cultivo del café.	Esta investigación contribuye reduciendo el impacto de enfermedades como la roya y promoviendo prácticas agrícolas responsables.

Tabla A 12. Cálculo del Coeficiente de Kappa para la toma de datos 1.

Observador	Asignados			Total	Precisión Asignados	Kappa
	Suelo	Árbol	Plantación			
Suelo	14	0	1	15	0,93	
Árbol	0	25	7	32	0,78	
Plantación	4	2	47	53	0,89	
Total	18	27	55	100		
Precisión Observados	0,78	0,93	0,85		0,86	
Kappa						0,76

Tabla A 13. Cálculo del Coeficiente de Kappa para la toma de datos 2.

Observador	Asignados			Total	Precisión Asignados	Kappa
	Suelo	Árbol	Plantación			
Suelo	10	0	0	10	1,00	
Árbol	0	38	0	38	1,00	
Plantación	2	3	53	58	0,91	
Total	12	41	53	106		
Precisión Observados	0,83	0,93	1,00		0,95	
Kappa						0,92

Tabla A 14. Cálculo del Coeficiente de Kappa para la toma de datos 2.

Observador	Asignados			Total	Precisión asignados	Kappa
	Arboles	Suelo	Plantación			
Arboles	31	6	7	44	0,70	
Suelo	0	8	2	10	0,80	
Plantación	1	0	44	45	0,98	
Total	32	14	53	99		
Precisión Observados	0,97	0,57	0,83		0,84	
Kappa						0,73

Apéndice B

Preguntas de la encuesta hecha en Survey 123.

- ¿Cuál es el cafeto del bloque a estudiar?: se da la opción de elegir tres opciones, las cuales permite al agrónomo tener un orden de cómo se van tomando los datos.
- Coordenadas del cafeto a estudiar: La aplicación permite usar las coordenadas del dispositivo móvil para incluir esta información en la encuesta, con una precisión de 2m.
- ¿En qué bloque se encuentra?: Gracias a la interacción con ArcGIS PRO, se adjuntó el polígono de bloques (**Figura 5**) que delimita la parcela experimental y la subdivide. Además, mediante las coordenadas, se pudo realizar este paso automáticamente, ya que el programa permite extraer el ID del bloque en el que el usuario se encuentra en ese momento.
- Identifique la incidencia del hongo (Roya) en la hoja 1: se utiliza la metodología ya explicada con una clasificación del 0 al seis con relación a el porcentaje de incidencia de la roya presente en la hoja estudiada.
- Identifique la incidencia del hongo (Roya) en la hoja 2: se repite el paso anterior con otra hoja en la misma bandola (rama plagiotrópica primaria, **Figura A 2**. Anatomía de un cafeto.) del cafeto.
- Identifique la incidencia del hongo (Roya) en la hoja 3: se repite el paso anterior con otra hoja en la misma bandola del cafeto.
- Incluir una imagen por cada una de las hojas estudiadas: este apartado permite tener un control extra de las hojas analizadas, ya sea para hacer correcciones o para un futuro trabajos de investigación.

Nota: Toda esta información es almacenada en la nube y accedida por medio de las credenciales universitarias.

Figura B 1. Captura de pantalla del portal survey123.

Roya Alsacia, TFG Lastra

EditDate	En que bloque se en...	Identifique la inciden...	Iden
2/7/2024, 10:35 a.m.	1	3	2
2/7/2024, 10:35 a.m.	1	3	3
22/5/2024, 12:11 p.m.	4	0	3
22/5/2024, 12:10 p.m.	4	0	2
22/5/2024, 12:10 p.m.	4	2	1
22/5/2024, 12:09 p.m.	2	1	0
22/5/2024, 12:09 p.m.	2	1	0
22/5/2024, 12:09 p.m.	2	1	0
22/5/2024, 12:09 p.m.	7	4	1
22/5/2024, 12:09 p.m.	7	0	0
22/5/2024, 12:07 p.m.	6	0	0
22/5/2024, 12:06 p.m.	6	1	1
22/5/2024, 12:06 p.m.	6	2	0
22/5/2024, 12:05 p.m.	5	1	2
22/5/2024, 12:05 p.m.	5	2	2
22/5/2024, 12:05 p.m.	5	3	2
22/5/2024, 12:05 p.m.	3	1	1
22/5/2024, 12:05 p.m.	2	2	1

1 de 184 seleccionados

Roya Alsacia, TFG Lastra

Enviado por: andres.lastra@ucr.ac.cr
 Hora de envío: 22 may, 2024 12:09:55

¿Cuál es el cafeto del bloque a estudiar?
Cafeto 3

Coordenadas del cafeto a estudiar
Lat.: 10.105087 Long: -84.197733

En que bloque se encuentra?
2

Identifique la incidencia del hongo (Roya) en la hoja 1:
1

Identifique la incidencia del hongo (Roya) en la hoja 2:
0

Identifique la incidencia del hongo (Roya) en la hoja 3:
2

Agregar foto de las hojas del cafeto

Nota: el panel de la derecha es la encuesta respondida por el usuario. Y a la izquierda se encuentran todas las respuestas en formato de hoja de cálculo.

Apéndice C

Figura C 1. Ráster de toma de datos 1, espectro Green.

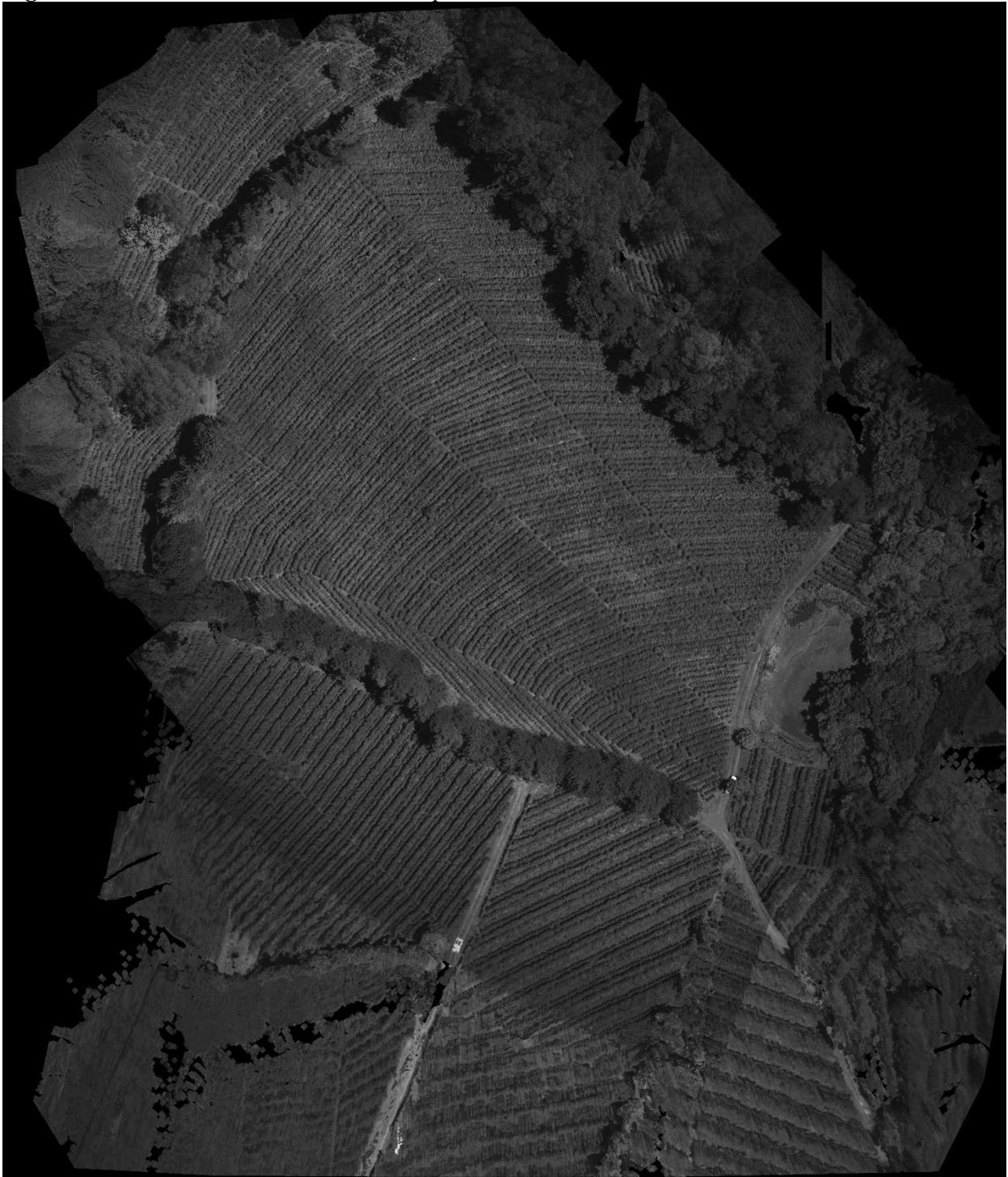


Figura C 2. Ráster de toma de datos 1, espectro Luz visible (RGB)



Figura C 3. Ráster de toma de datos 1, espectro Nir.

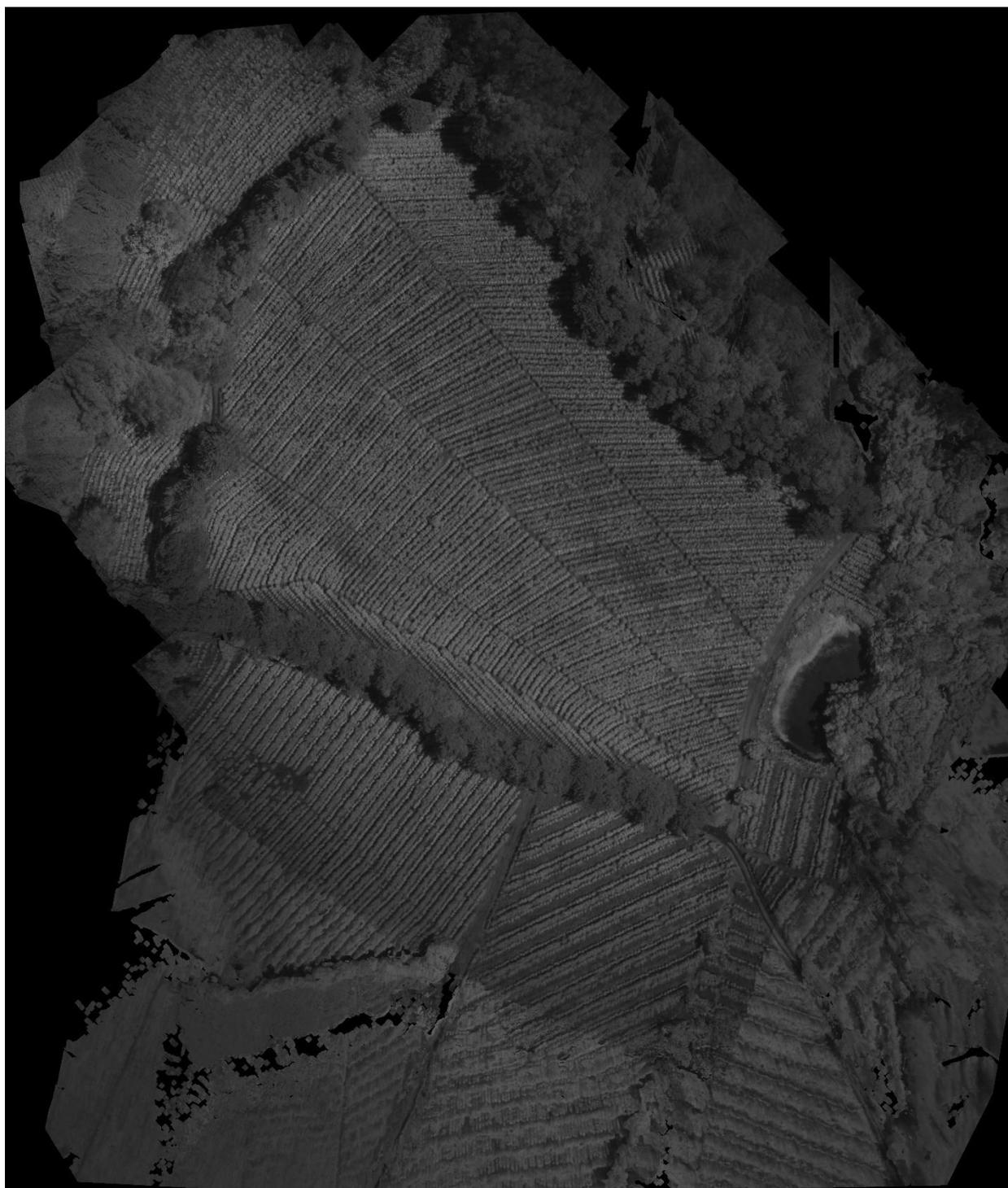


Figura C 4. Ráster de toma de datos 1, espectro Red.

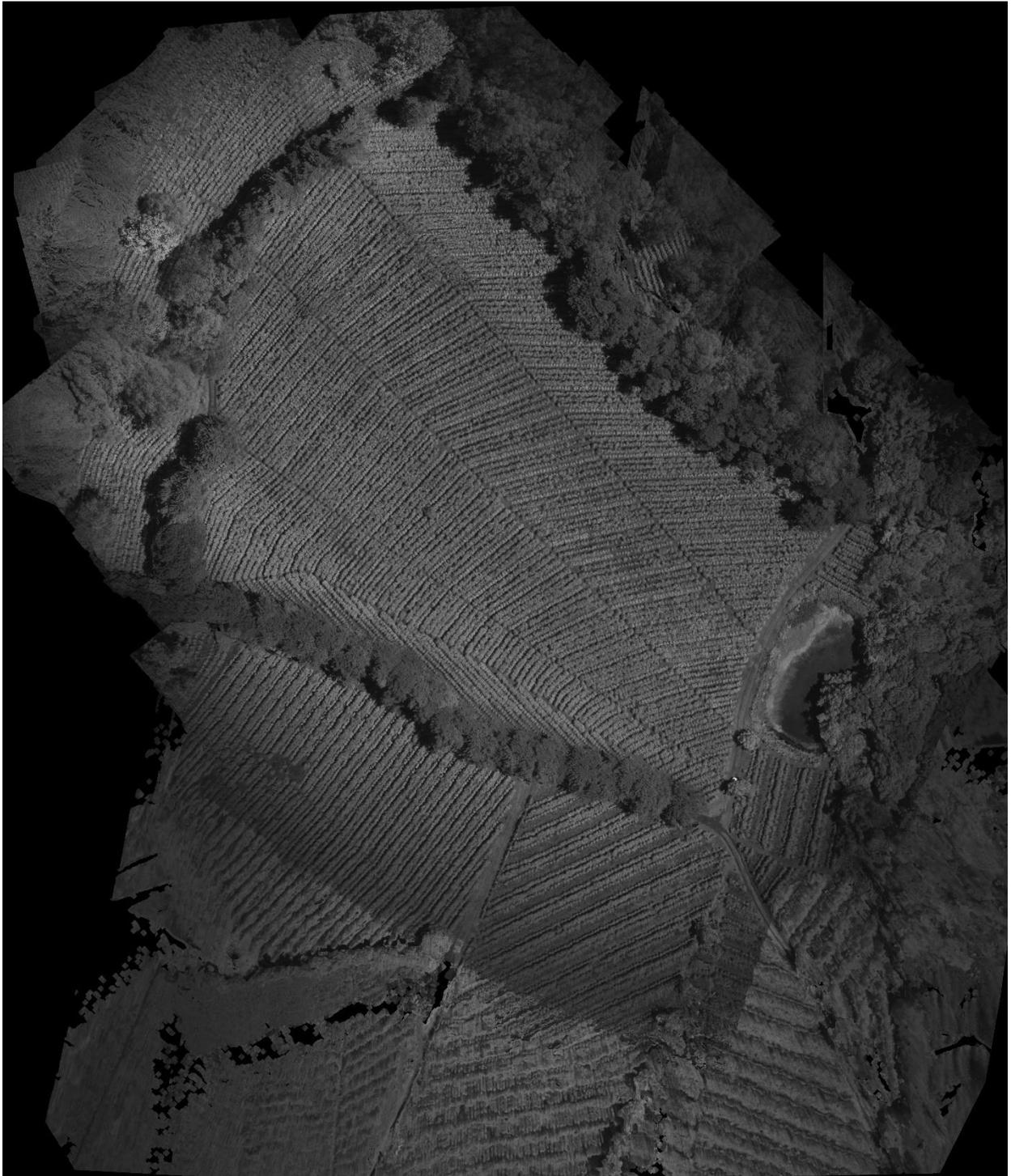


Figura C 5. Ráster de toma de datos 1, espectro Red Edge.

