

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

**TELEDETECCIÓN TEMPRANA DE ESTRÉS HÍDRICO EN DOS VARIEDADES
DE PLANTAS DE CAFÉ (*COFFEA ARABICA*)**

TRABAJO FINAL DE GRADUACIÓN PARA OPTAR POR EL TÍTULO DE
LICENCIATURA EN INGENIERÍA AGRÍCOLA Y DE BIOSISTEMAS

ANDRÉS BARQUERO ARAYA

CIUDAD UNIVERSITARIA RODRIGO FACIO

26 DE FEBRERO DE 2024

HOJA DE APROBACIÓN

Este Trabajo Final de Graduación fue aceptado por la Comisión de Trabajos Finales de Graduación de la Escuela de Ingeniería de Biosistemas de la Universidad de Costa Rica, como requisito para optar al grado y título de Licenciatura en Ingeniería Agrícola y de Biosistemas.

Andrés Barquero A

Andrés Barquero Araya

Estudiante



Ing. Alejandra Rojas González, Ph.D.

Directora Trabajo Final de Graduación



Ing. Melissa Rojas Downing, Ph.D.

Comité Asesor



Ing. Juan Delgado Fernández, MGA.

Comité Asesor



M.Sc. Anayansi Wong Monge

Presidente Tribunal Examinador



M. Sc. Matías Chaves Herrera

Lector Invitado

DEDICATORIA

Esta investigación está primeramente dedicada a Dios porque sin él nada sería posible. Seguidamente, a mi familia, especialmente a mi padre Gonzalo Barquero Dodero y a mi madre Ana Laura Araya Ocampo. Gracias por todo el apoyo que me han brindado a través de todo este tiempo, por su amor incondicional y por siempre creer en mí. A mi hermana, Lissa Barquero Araya por siempre mostrarme una manera diferente de ver las cosas. Sin ustedes, no sería la persona que soy hoy en día.

AGRADECIMIENTOS

La finalización de este trabajo de investigación fue posible gracias a la colaboración de muchas personas. Quiero agradecer primeramente a la Universidad de Costa Rica por mi formación como profesional y su amplia visión de lo que educación significa. A mi directora de tesis Alejandra Rojas González por su acompañamiento a través del proceso. A la profesora Anayansi Wong por su experiencia en el campo de la estadística. Un agradecimiento especial a Gilles Chaix y al CIRAD por brindar todo su conocimiento en el campo de la quimiometría. Al Ing. Carlos Mario y a Víctor Trejos por abrirme las puertas en Hacienda Alsacia para realizar la investigación. También quiero agradecer a Juan Delgado y Andrés Castillo de la empresa LIFE-RID por su generosa disposición y apoyo tanto en las mediciones realizadas como en la revisión del documento final. Por último pero no menos importante un agradecimiento muy especial a mis compañeros de carrera por acompañarme en esta etapa tan hermosa. En especial, a mi pareja Gloriana Guzmán Vindas por el apoyo incondicional, las largas horas de trabajo compartido y la motivación en momentos difíciles. Este trabajo simboliza el final de una etapa y el inicio de otra aún más emocionante y llena de retos.

RESUMEN

La teledetección tiene una aplicación amplia en diferentes campos. En el mejoramiento de cultivos en sistemas productivos agrícolas puede llegar a solventar problemáticas de otras técnicas como la aplicabilidad a diferentes escalas, la naturaleza destructiva de algunas pruebas de calidad y la alta necesidad de recursos y tiempo que se traducen en aumento de costo. El objetivo general de esta investigación fue desarrollar cuatro modelos para la detección temprana de estrés hídrico en dos variedades de plantas de café (Victoria 14 y Turrialba 4730) bajo condiciones controladas por medio de espectroscopía del espectro visible e infrarrojo cercano utilizando mediciones de transfectancia de la hoja y correlacionándolas con variables medidas de las plantas de café en un invernadero. Variables como la tasa de transpiración, la asimilación de CO₂, la temperatura de la hoja, entre otros indicadores fisiológicos y biométricos fueron evaluadas en un periodo de 38 días en diferentes condiciones de estrés hídrico. Se exploraron los efectos de dos distintos tratamientos sobre la respuesta de las plantas al estrés abiótico contrarrestándolos con el tratamiento control (A), que consistió en suministrar 1.7 litros de agua diarios a las plantas manteniendo las plantas hidratadas. Uno de los tratamientos experimentales se basó en un régimen de deshidratación sin riego (EH), mientras que el segundo añadía a este régimen los compuestos Seamaxx y Photon para intentar mitigar el estrés (EHC).

Las mediciones fueron correlacionadas con datos espectrales para la creación de 16 modelos predictivos (cuatro variables, dos estratos: alto y medio; y dos variedades de café) por medio de la regresión de mínimos cuadrados (PLSR, por sus siglas en inglés) usando además diferentes pretratamientos de firmas espectrales para encontrar la mejor combinación. Se utilizó principalmente el R² ajustado y el RMSEP para definir su rendimiento. El modelo que obtuvo los mejores resultados fue el modelo 11, para la predicción de la variable de temperatura de la hoja en la variedad Turrialba con datos del estrato medio (R² = 0.39, RMSEP = 0.36) y el pretratamiento Savitzky-Golay 2 (2da derivada) con 16 variables latentes. Los resultados de los indicadores R² y RMSEP obtenidos son similares a diferentes estudios que utilizaron la regresión de mínimos cuadrados principales (PLSR, por sus siglas en inglés). Los datos obtenidos muestran una alta variabilidad relacionada a diferentes errores metodológicos que afectaron al diseño experimental planteado. Por consiguiente, la

efectividad de los insumos agrícolas no se pudo evidenciar de manera concisa y limitó la aplicabilidad de los modelos desarrollados. A través de este análisis, se buscó comprender de manera más profunda cómo cada tratamiento incidió sobre las plantas, aportando información valiosa para el desarrollo de futuras estrategias en el manejo del estrés abiótico.

INDICE GENERAL

HOJA DE APROBACIÓN.....	Error! Bookmark not defined.
DEDICATORIA	i
AGRADECIMIENTOS	iii
RESUMEN.....	iv
1. INTRODUCCIÓN	1
2. MARCO TEÓRICO.....	4
2.1 Marco conceptual	4
2.1.1 Interacción entre la luz y las plantas	4
2.1.2 Capacidad fotosintética	5
2.1.3 Variedades de café en estudio.....	7
2.1.3.1 Variedad Victoria 14.....	8
2.1.3.2 Variedad Turrialba 4730	8
2.2 Antecedentes	9
2.2.1 Factores de plantas relacionados con el estrés hídrico	9
2.2.2 Teledetección de estrés hídrico.....	10
2.2.3 Quimiometría	11
2.2.4 Métodos de análisis para datos hiperespectrales	12
3. OBJETIVOS	14
3.1 Objetivo general	14
3.2 Objetivos específicos.....	14
4. METODOLOGIA	15
4.1 Sitio y objeto de estudio.....	16
4.1.1 Material vegetal y condiciones de crecimiento	17
4.2 Diseño Experimental	18
4.2.1 Variables Independientes.....	20
4.2.2 Variables dependientes y datos auxiliares	21
4.3 Recolección de datos.....	22
4.3.2 Altura de las muestras y cantidad de hojas.....	24
4.3.3 Datos hiperespectrales.....	24
4.4 Análisis de las variables fisiológicas.....	26
4.5 Caracterización y evaluación hiperespectral	26
4.5.1 Preprocesamiento de información espectral.....	28
4.5.1.1 Gap Remove.....	28

4.5.1.2 Pretratamientos espectrales	28
4.5.2 Creación del modelo predictivo	29
4.5.2.1 Análisis de componentes principales (PCA)	30
4.5.2.2 Separación de datos de calibración y validación	31
4.5.3 Regresión de mínimos cuadrados parciales (PLSR)	31
4.5.4 Eficiencia de los modelos.....	32
5. RESULTADOS	35
5.1 Evaluación Experimental: Comportamiento temporal de las variables fisiológicas	35
5.1.1 Potencial hídrico.....	38
5.1.1.2 Altura.....	41
5.1.1.3 Cantidad de hojas	42
5.1.1.4 Tasa de transpiración	44
5.1.1.5 Tasa de asimilación neta de CO ₂	45
5.1.1.6 Temperatura de la hoja	47
5.1.1.7 Presión de vapor de saturación a la temperatura de la hoja.....	49
5.2 Análisis de varianza (ANOVA)	51
5.3 Análisis de respuestas espectrales	54
5.3.1 Preprocesamiento de información espectral.....	54
5.3.2 Generación del modelo predictivo	56
5.3.2.1 Análisis de componentes principales (PCA)	57
5.3.2.2 Regresión de mínimos cuadrados parciales (PLSR)	63
6. CONCLUSIONES	71
7. RECOMENDACIONES	73
8. REFERENCIAS	75
APÉNDICE A:	83
RESULTADOS DE EVALUACIÓN EXPERIMENTAL	83
APÉNDICE B:	91
RESULTADOS DE ANÁLISIS DE RESPUESTAS ESPECTRALES.....	91
APÉNDICE C:	108
SECCIÓN DE CÓDIGO UTILIZADO	108

ÍNDICE DE FIGURAS

Figura 1. Interacción de la luz en las hojas de las plantas. (a) reflexión, (b) dispersión, (c) absorción y (d) transmisión.....	5
Figura 2. Diagrama de flujo sobre la metodología de la investigación.....	15
Figura 3. Estructura final para la protección de las muestras.....	17
Figura 4. Distribución espacial de las muestras, tratamientos y ubicación del sistema de riego	20
Figura 5. Línea de tiempo de la investigación 2022.....	35
Figura 6. Implementación de bolsas plásticas 26 de mayo del 2022.....	37
Figura 7. Implementación de tarimas 2 de junio del 2022.....	37
Figura 8. Potencial hídrico (bar) vs tiempo (días) para la variedad Turrialba (a) y Victoria (b).....	40
Figura 9. Muestra V33 con potencial hídrico menor de 40 bar el día 13 de junio del 2022.....	41
Figura 10. Altura promedio de las muestras (cm) vs tiempo (días) para la variedad Turrialba (a) y Victoria (a).....	42
Figura 11. Porcentaje de defoliación vs tiempo (días) para la variedad Turrialba (a) y Victoria (b).....	43
Figura 12. Tasa de transpiración ($\text{mol m}^{-2} \text{s}^{-1}$) vs tiempo (días) para la variedad Turrialba (a) y Victoria (b).....	44
Figura 13. Tasa de asimilación neta de CO_2 ($\mu\text{mol m}^{-2} \text{s}^{-1}$) vs tiempo (días) para la variedad Turrialba (a) y Victoria (b).....	47
Figura 14. Temperatura de la hoja en el balance de energía ($^{\circ}\text{C}$) vs tiempo (días) para la variedad Turrialba (a) y Victoria (b).....	48
Figura 15. Presión de vapor de saturación a la temperatura de la hoja (kPa) vs tiempo (días) para la variedad Turrialba (a) y Victoria (b).....	49
Figura 16. Datos promediados del estrato alto (a) y medio (b) luego de aplicar la función rmgap.....	55
Figura 17. Modificaciones en el comportamiento de los datos espectrales luego de aplicar diferentes pretratamientos al conjunto de datos del estrato alto. SNV (a), detrend (b), Savitzky-Golay 1era derivada (c) y Savitzky-Golay + SNV (d).....	56
Figura 18. Resultados del análisis de componentes principales (PCA) para la tasa de transpiración (E) representados con PC1, PC2 y PC3. Turrialba estrato alto (a), Victoria estrato alto (b), Turrialba estrato medio (c) y Victoria estrato medio (d).....	59

Figura 19. Coeficientes de regresión vs longitudes de onda (nm) más importantes de los modelos 1(a), 2(b), 3(c) y 4(d).....	62
Figura 20. Curvas de datos predichos (eje y) vs datos de la tasa de transpiración (eje x) para el modelo 1(a), 2(b), 3(c) y 4(d).....	70
Figura A.1. Esquema con la posición de las muestras escogidas aleatoriamente.....	85
Figura A.2. Datos de precipitación (a) y precipitación acumulada (b) en mm de la estación meteorológica (SOMBRA) de Hacienda Alsacia.....	86
Figura B.2. Modificaciones en el comportamiento de los datos espectrales luego de aplicar diferentes pretratamientos al conjunto de datos del estrato medio. SNV (a), Detrend (b), Savitzky-Golay 1era derivada (c) y Savitzky-Golay + SNV (d).....	92
Figura B.3. Resultados del análisis de componentes principales (PCA) para la tasa de asimilación neta de CO ₂ (A) representados con PC1, PC2 y PC3. Turrialba estrato alto (a), Victoria estrato alto (b), Turrialba estrato medio (c) y Victoria estrato medio (d).....	93
Figura B.4. Resultados del análisis de componentes principales (PCA) para temperatura de la hoja (Tleaf) representados con PC1, PC2, y PC3. Turrialba estrato alto (a), Victoria estrato alto (b), Turrialba estrato medio (c) y Victoria estrato medio (d).....	94
Figura B.5. Resultados del análisis de componentes principales (PCA) para la presión de vapor de saturación a la temperatura de la hoja (SVPleaf) representados con PC1, PC2 y PC3. Turrialba estrato alto (a), Victoria estrato alto (b), Turrialba estrato medio (c) y Victoria estrato medio (d).....	95
Figura B.6. Curva de datos predichos (eje y) vs datos medidos de la variable A (eje x) para el modelo 5(a), 6(b), 7(c) y 8(d).....	102
Figura B.7. Curvas de datos predichos (eje y) vs datos medidos de la variable SVPleaf (eje x) para el modelo 9(a), 10(b), 11(c) y 12(d).....	103
Figura B.8. Curvas de datos predichos (eje y) vs datos medidos de la variable Tleaf (eje x) para el modelo 9(a), 10(b), 11(c) y 12(d).....	104
Figura B.9. Coeficientes de regresión vs longitudes de onda (nm) más importantes de los modelos 5(a), 6(b), 7(c) y 8(d).....	105
Figura B.10. Coeficientes de regresión vs longitudes de onda (nm) más importantes de los modelos 9(a), 10(b), 11(c) y 12(d).....	106
Figura B.11. Coeficientes de regresión vs longitudes de onda (nm) más importantes de los modelos 13(a), 14(b), 15(c) y 16(d).....	107

ÍNDICE DE CUADROS

Cuadro 1. Resumen de variables dependientes recolectadas y sus características.....	22
Cuadro 2. Características de las firmas espectrales recolectadas.....	25
Cuadro 3. Estratos para las diferentes variedades de café.....	25
Cuadro 4. Nomenclatura definida para las combinaciones de pretratamientos utilizados.....	29
Cuadro 5. Cuadro resumen de los modelos generados.....	30
Cuadro 6. Estadísticos utilizados en la determinación del rendimiento de los modelos predictivos.....	33
Cuadro 7. Resumen de ANOVAS para las variables estudiadas.....	53
Cuadro 8. Porcentajes de varianza explicados para cada componente principal de cada modelo generado para la tasa de transpiración (E).....	57
Cuadro 9. Resultados obtenidos de los indicadores para los modelos generados.....	64
Cuadro A.1. Mediciones de potencial hídrico (bar) de todas las muestras en el periodo de investigación.....	84
Cuadro A.2. Resultados completos de análisis de varianza (ANOVA) para potencial hídrico.....	87
Cuadro A.3. Resultados completos de análisis de varianza (ANOVA) para tasa de transpiración.....	87
Cuadro A.4. Resultados completos de análisis de varianza (ANOVA) para tasa de asimilación.....	88
Cuadro A.5. Resultados completos de análisis de varianza (ANOVA) para temperatura de la hoja (Tleaf).....	88
Cuadro A.6. Resultados completos de análisis de varianza (ANOVA) para presión de vapor de saturación a la temperatura de la hoja (SVPleaf).....	89
Cuadro A.7. Resultados completos de análisis de varianza (ANOVA) para altura de las plantas.....	89
Cuadro A.8. Resultados completos de análisis de varianza (ANOVA) para cantidad de hojas.....	90
Cuadro A.9. Resumen de las relaciones porcentuales de las variables fisiológicas entre los tratamientos EH con respecto a EHC en ambas variedades para el día 31 y 38.....	90

1. INTRODUCCIÓN

La agricultura ha enfrentado diversos desafíos, tales como el rápido crecimiento de la población, la disminución de los recursos naturales, la contaminación ambiental y el cambio climático. Para abordar estas problemáticas, la agricultura de precisión surge como una alternativa prometedora que busca mejorar las prácticas productivas para reducir el uso de insumos, aumentar la producción y disminuir el impacto ambiental (Liu et al., 2005; Weiss et al., 2020).

Gracias al desarrollo tecnológico, la agricultura de precisión permite monitorear el progreso de los cultivos en tiempo real para generar información valiosa que facilita la toma de decisiones. Esta información se utiliza para mejorar la eficiencia de diferentes procesos, como la aplicación de fertilizantes, pesticidas, irrigación y control de biomasa, entre otros.

La teledetección es una herramienta valiosa para abordar las problemáticas previamente mencionadas. Su enfoque sistemático permite la obtención de información local y global de manera no destructiva. Existen una variedad de instrumentos para la toma de datos, como los satélites multiespectrales e hiperespectrales que capturan imágenes con gran cantidad de información útil tras su procesamiento. Los drones también son ampliamente utilizados y cuentan con sensores especiales para la captura y transmisión de información. Además, los sensores espectrales de campo proporcionan información de alta resolución espacial para el estudio de comportamientos sensibles de las plantas y suelos. Con ellos, es posible obtener información sobre características biofísicas, procesos bioquímicos de las plantas y su respuesta a diferentes tipos de estrés abiótico y enfermedades (Mohd Asaari et al., 2018).

El café es una planta que tiene gran relevancia tanto cultural como económica debido a sus características y efectos estimulantes generados por la cafeína. De acuerdo con el Instituto del Café de Costa Rica (ICAFFE, 2019), este es uno de los productos primarios más comercializados en todo el mundo, lo cual representa un impacto directo o indirecto en la economía de los países que lo producen. Sin embargo, se reportó una disminución del 4,26% en la producción cafetalera para el periodo 2019-2020, según la Organización Internacional

del Café (OIC) y el Departamento de Agricultura de los Estados Unidos (USDA, por sus siglas en inglés, 2019).

La pandemia del COVID-19 tuvo un impacto directo en este mercado, debido a la implementación de medidas sanitarias en los establecimientos públicos y la implementación de confinamientos preventivos. Por otro lado, el consumo de café de los países importadores se incrementó en un 4,9% durante el periodo (Ene-Dic) de los años 2018-2019 (ICAFE, 2019). Durante el periodo de 2019-2020, la producción de café en Costa Rica experimentó altibajos en términos de productividad e ingresos generados. Según cifras del Instituto del Café de Costa Rica (ICAFE, 2019), se produjeron 1.974.801,16 fanegas de café (equivalentes a entre 44,5 y 45,5 kg de café verde por cada 253 kg de café fruta), lo que representa un aumento del 14,97% en comparación con el periodo anterior.

En cuanto a los ingresos generados por la exportación del café costarricense, la Promotora de Comercio Exterior de Costa Rica (PROCOMER, 2020) reportó que en el año cafetalero 2019-2020 (Oct-Set) se generó un valor de \$322.539.696 en divisas para el país, lo que representa un incremento del 2,8% en relación con el periodo anterior. Según el ICAFE (2019), este ingreso representa el 2,8% del total de ingresos por exportaciones del país.

El desarrollo de modelos para predecir el estrés hídrico en plantas de café es crucial para asegurar su producción y desarrollo. Según Cortez & Tercero (2014), la irrigación de cafetos durante la época seca mejora su área foliar, nudos, crecimiento de bandola y calidad del grano. Sin embargo, dado el aumento de la temperatura global, se recomienda la implementación de sistemas de irrigación para garantizar las condiciones óptimas de las plantas de café.

Por otro lado, la creación de un modelo que tome en cuenta todas las variables que influyen en el estrés hídrico es esencial para identificar y cuantificar de manera eficiente y no destructiva este fenómeno en las plantas de café. Estudios han demostrado la aplicabilidad de la teledetección para detectar el estrés hídrico en las plantas de café, mediante la relación directa entre la diferencia en el contenido de agua de la planta y la reflectancia de los cultivos sometidos a diferentes tratamientos (Elvanidi et al., 2018). Por lo tanto, esta investigación representa el primer paso para relacionar datos hiperespectrales con la detección temprana de estrés hídrico en variedades de café utilizadas comercialmente en el país. Esto es de gran

importancia para el sector productivo, ya que puede ser replicado en diferentes variedades de café y zonas cafetaleras, y su relación es directa con la producción, el gasto de insumos y el impacto ambiental.

Además, se ha demostrado que el uso de riego controlado en plantas de café puede mejorar la calidad de los datos obtenidos en experimentos, reduciendo el posible ruido generado por la falta de agua en la planta (Pérez et al., 2021). Esto permite una mejor evaluación de las respuestas fisiológicas y bioquímicas de la planta al estrés hídrico y una mayor precisión en la determinación de la tolerancia al mismo. Por lo tanto, la implementación de sistemas de riego controlado puede ser un complemento importante en la creación del modelo de detección temprana de estrés hídrico en variedades de café mediante teledetección.

Este trabajo final de graduación se alinea con los Objetivos de Desarrollo Sostenible (ODS) 2 y 4, los cuales buscan promover prácticas agrícolas sostenibles para garantizar la seguridad alimentaria y la gestión sostenible de los recursos naturales. Según la Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO), se estima que alrededor de 800 millones de personas en el mundo sufren de inseguridad alimentaria y que la agricultura sostenible es esencial para abordar este problema (FAO, 2020). Además, este trabajo busca adaptar investigaciones realizadas en otras regiones a nuestro contexto local, considerando las diferencias en las condiciones climáticas, los factores que influyen en el desarrollo de los cultivos y las características de las especies vegetales contribuyendo a la generación de conocimiento sobre la agricultura de precisión en nuestro país y en la región.

La implementación de prácticas agrícolas sostenibles también tiene un impacto directo en la preservación del medio ambiente y la mitigación del cambio climático. Según la FAO, la agricultura es responsable de alrededor del 20% de las emisiones globales de gases de efecto invernadero y se estima que la implementación de prácticas agrícolas sostenibles podría reducir estas emisiones (FAO, 2020).

2. MARCO TEÓRICO

2.1 Marco conceptual

2.1.1 *Interacción entre la luz y las plantas*

La interacción de la radiación electromagnética con las plantas es diferente según la frecuencia de dicha energía, porque la actividad fotosintética se realiza en las hojas de las plantas y esta interacción es un tema de sumo interés (Jacquemoud & Ustin, 2001). Distintas regiones del espectro electromagnético están relacionadas con diferentes funciones de las plantas, por ejemplo, la región visible (400 - 700 nm) se relaciona con la interacción entre la luz y los pigmentos. La región del infrarrojo cercano (700 - 1.100 nm) con la absorción de materia seca y la región del infrarrojo de onda corta (1.100 - 2.500 nm) con los cambios en la absorción del agua (Lu et al., 2020).

Cuando la luz entra en la superficie de la hoja y los campos electromagnéticos de ambos átomos empiezan a interactuar existen cuatro tipos de interacción dependiendo de las características tanto de la fuente de luz como de la superficie de la hoja. Los cuatro tipos de interacción se pueden apreciar en la figura 1, la reflexión (Figura 1.a) ocurre en la superficie y una parte de la luz es reflejada. La dispersión (Figura 1.b) es causada por las irregularidades en la estructura interna de las hojas la cual cambia la propagación de los rayos de luz. La absorción (Figura 1.c) se refiere al proceso de transferencia de energía entre el fotón y el electrón de la planta, en donde la configuración de este se modifica, afectando de igual forma a los átomos de la cercanía en forma de energía térmica. Por último, la transmisión o transmitancia (Figura 1.d) de la luz hace alusión a la porción de luz o energía que por medio de vibraciones se transfieren por el cuerpo de las hojas y es reemitida fuera de ella (Chen et al., 2006).

En la teledetección las propiedades de las plantas pueden ser medidas con índices espectrales de vegetación (IV), estos se obtienen convirtiendo el espectro de la reflectancia en un único valor numérico. Estos índices se dividen en tres grandes grupos: (1) estructurales, (2) bioquímicas y (3) fisiológicas. Algunas de las características presentes en el grupo (1) son biomasa verde, índice del área de hoja (LAI), cobertura fraccional y senescencia de biomasa.

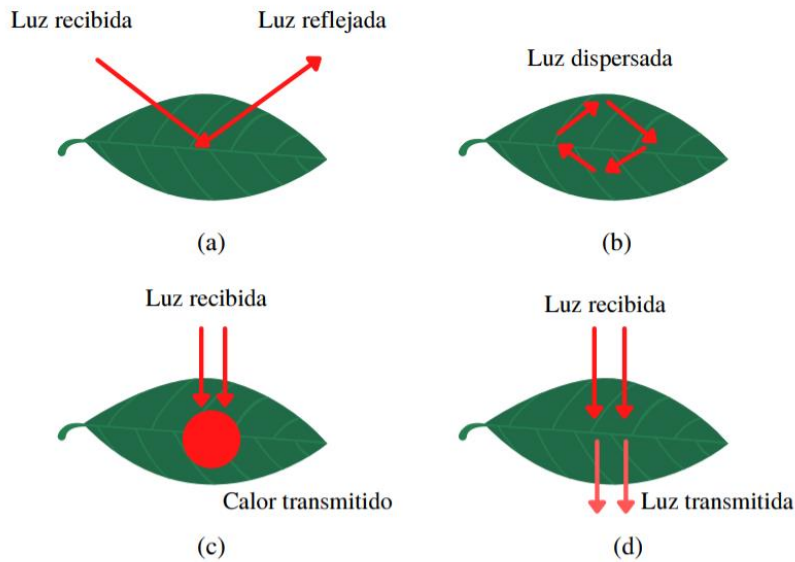


Figura 1. Interacción de la luz en las hojas de las plantas. (a) reflexión, (b) dispersión, (c) absorción y (d) transmisión.

En el grupo (2) se encuentran factores como agua, pigmentos (clorofila, carotenoides y antocianinas), compuestos ricos en nitrógeno (proteínas) y materiales estructurales de las plantas (lignina y celulosa). Por último, el grupo (3) contiene índices relacionados con mediciones de estrés biótico o abiótico provocado por cambios en las concentraciones de sustancias como clorofila, xantofilas o factores externos como fluorescencia o contenido de humedad de las hojas (Thenkabail et al., 2019b).

2.1.2 Capacidad fotosintética

La capacidad fotosintética es un término utilizado para referirse a la capacidad de una planta para realizar el proceso de fotosíntesis de manera eficiente. Está estrechamente relacionado con la capacidad de fijar CO₂ que resultará en la síntesis de carbohidratos. Estos son almacenados en estructuras de las plantas y son utilizados en momentos en los cuales obtener alimentos del sol no es posible (Tolentino et al., 2016). Se cuenta con cuatro variables medibles mencionadas en esta investigación que se relacionan con la capacidad fotosintética de la planta que son la tasa de transpiración, la tasa de asimilación neta de dióxido de carbono, la temperatura de la hoja y la presión de vapor de saturación.

La tasa de transpiración se refiere a la cantidad de agua que una planta pierde por evaporación a través de las estomas de las hojas en un período de tiempo determinado. Es un proceso importante para la absorción de nutrientes, el enfriamiento y el mantenimiento de la temperatura adecuada para el crecimiento de la planta. La tasa de transpiración puede variar según las condiciones ambientales y es un indicador de la salud de la planta y su capacidad para tolerar condiciones adversas como la sequía o el estrés hídrico (Taiz et al., 2016).

La tasa de asimilación neta de dióxido de carbono es una medida de la eficiencia fotosintética de una planta, expresada en unidades de masa de CO₂ fijado por unidad de tiempo y de masa de hoja. El estrés hídrico puede disminuir la tasa de asimilación neta de CO₂ al limitar la capacidad de la planta para absorber CO₂ del aire y llevar a cabo la fotosíntesis de manera eficiente. Esto puede ocurrir por el cierre de las estomas de la planta, una disminución en la cantidad de clorofila en las hojas, y daño en las células de la planta (Taiz et al., 2016).

La temperatura de la hoja es un factor clave en el balance de energía de una planta, que se refiere a la relación entre la energía solar que recibe y la energía que pierde por varios procesos. La temperatura de la hoja está influenciada por varios factores, como la energía solar, la transpiración, la conductividad térmica y la temperatura ambiente. Una hoja más caliente puede aumentar la pérdida de energía por radiación y conducción, mientras que una hoja más fría puede aumentar la pérdida de energía por transpiración (Taiz et al., 2016). En general, la temperatura de la hoja es importante en el balance de energía de una planta y puede ser influenciada por factores ambientales y fisiológicos.

La presión de vapor de saturación a la temperatura de la hoja es la máxima cantidad de vapor de agua que el aire puede retener antes de que se alcance la condensación. Esta presión es importante para comprender la transpiración en las plantas, ya que crea un gradiente que impulsa la salida de agua de las hojas hacia el aire (Taiz et al., 2016). Si la temperatura de la hoja es mayor que la presión de vapor de saturación, se transpira vapor de agua, pero si es menor, puede haber condensación. En resumen, la presión de vapor de saturación influye en los procesos de balance hídrico y transpiración en las plantas.

Todos los conceptos antes mencionados están muy relacionados con la fisiología de las plantas y las relaciones que estas tienen con el medio ambiente y sus recursos. Se puede

notar que estas cuatro variables influyen directamente la una de la otra y son indicadores de un funcionamiento adecuado de las plantas. Otro parámetro que también se estudiará en esta investigación y ha sido mencionado en las variables anteriores es el potencial hídrico.

El potencial hídrico del tallo es una medida que se utiliza para evaluar el estrés hídrico en las plantas. Normalmente se utilizan los kPa o los bar como unidad y las magnitudes de esta variable son negativas (Taiz et al., 2016). Cuando una planta experimenta estrés hídrico, el potencial hídrico del suelo disminuye, lo que hace más difícil para la planta absorber agua y nutrientes. La medición del potencial hídrico del tallo es útil para evaluar el estado de hidratación de la planta y tomar medidas para reducir el estrés hídrico y mejorar su supervivencia. La medición del potencial hídrico se realiza normalmente a las 6:00 am ya que a esta hora, los estomas se encuentran cerrados y la planta se encuentra hidratada al máximo; lo que significa que todas las hojas están en equilibrio hídrico con el suelo (Pino et al. 2019).

La edad de una planta de café es muy importante ya que los sistemas radiculares de las plantas jóvenes son más susceptibles a cambios físicos negativos por diferencias en el contenido de agua menores al adecuado (Chemura et al., 2017; González et al., 2017). El punto de marchitez permanente de una planta de café sin riego es alcanzado a los 21 días aproximadamente (Veintimilla, 2022). Estas dos características fueron claves para determinar la etapa fenológica del cultivo a utilizar como también el periodo de medición de la investigación.

2.1.3 Variedades de café en estudio

Existen dos principales especies de café que se cultivan y comercializan a nivel mundial, *Coffea arabica* más conocido como Arábica y *C. canephora*, comúnmente llamado Robusta. Siendo el café arábico la especie que predomina a nivel comercial y la que se encuentra más relacionada con la alta calidad de grano por sus características únicas (World Coffee Research, 2019). Existen diferentes tipos de variedades y cultivares que también son utilizados por sus diferentes cualidades genéticas.

Muchas de las variedades que se utilizan hoy en día en América provienen de un linaje genético conocido como Borbón y Típica los cuales a finales del siglo XVII fueron extraídos

de Yemen y de ahí se esparcieron por todo el mundo formando la base del cultivo de café Arábica. De igual forma otro grupo de cultivares provienen de un grupo conocido como variedades de introgresión, estas plantas poseen rasgos genéticos de otras especies, en este caso una mezcla de las especies *C. arabica* y *C. canephora*. Existen dos grupos principales de las variedades arábicas introgresadas: Catimores y Sarchimores (World Coffee Research, 2019).

2.1.3.1 Variedad Victoria 14

Este híbrido fue seleccionado como resultado del procedimiento de selección, realizado en Finca Inés, Grecia, Costa Rica, finca de CoopeVictoria. Victoria-14 es un cruce entre el Híbrido de Timor (CIFC832/2 y Villa Sarchí. Victoria-14 es una variedad seleccionada y desarrollada por Calos Mario Rodríguez.

Su producción estimada de cereza es de 10500 kg/Ha y en café oro 3000 kg/Ha. El tamaño promedio de la planta es de 1.59 m y su forma es cónica invertida con una longitud de rama de 77.6 cm. Se recomienda la siembra en una altitud de 800 – 1800 msnm y una precipitación de 1000 – 3000 mm.

La primera producción de esta variedad se da al segundo año. Su sistema radicular es promedio al igual que la maduración de su fruto. Es tolerante a la roya y posee una densidad de siembra de 5000 pl/Ha. Actualmente se encuentra en F9 estable y una de sus características más importantes es que posee un sistema radicular muy fuerte. Por lo tanto, se piensa que su comportamiento respecto al estrés hídrico va a ser más resistente.

2.1.3.2 Variedad Turrialba 4730

Esta variedad se encuentra en el catálogo del CATIE bajo el número de ascensión T.04730. Fue recolectada de Etiopia el 17 de junio de 1965 y se encuentra en el Centro Agronómico Tropical de Investigación y Enseñanza del CATIE en Turrialba, Costa Rica. No es considerada una variedad comercial por el momento y su información relacionada a rendimiento y características morfológicas es escasa.

2.2 Antecedentes

2.2.1 Factores de plantas relacionados con el estrés hídrico

La teledetección aplicada a la agricultura es una poderosa herramienta para monitorear las variaciones fisiológicas y morfológicas de los cultivos y para utilizar esa información en la toma de decisiones en sistemas productivos. Con la teledetección de corto alcance se pueden estudiar procesos muy finos en las plantas gracias a su capacidad de adquirir datos a una resolución espacial muy alta (Lu et al., 2020).

La estimación de la cantidad de clorofila del dosel (CCC por sus siglas en inglés) es una variable muy importante en la determinación de la capacidad fotosintética de diferentes cultivos. Por medio de teledetección de corto alcance se estudiaron diferentes algoritmos para el manejo de datos de datos obtenidos de seis cultivos (arroz, trigo, maíz, soja, remolacha y zacate) ubicados independientemente en Japón, Francia, Estados Unidos e Italia durante el periodo de 1990-2021. Se utilizaron el índice de diferencia espectral normalizada (NDSI por sus siglas en inglés), el índice de mapa e índice de radio espectral (RSI por sus siglas en inglés) y regresión multivariable para establecer el mejor acercamiento a la estimación de CCC. Un modelo basado en índices espectrales (SI por sus siglas en inglés) basado utilizando RSI con reflectancia entre 815 y 704 nm fue el que mostró mejores resultados midiendo la CCC considerando los datos auxiliares de contenido de agua y clorofila, área de la hoja, altura, biomasa seca y densidad de tallo (Inoue et al., 2016).

La reducción en la cantidad de biomasa o el índice de área de la hoja (LAI por sus siglas en inglés) de las plantas está relacionada con diferentes tipos de estrés que afectan a los pigmentos responsables de realizar los procesos fotosintéticos. El contenido de clorofila de las plantas está directamente relacionado con el contenido de nitrógeno por lo que en el estudio de Clevers & Kooistra (2012) se utilizaron diferentes modelos (PROSAIL y un modelo de transferencia radiativa) e índices para calcular la CCC en cultivos de papa y pasto. El estudio se llevó a cabo en el sur de Los Países Bajos y en la región de Achterhoek respectivamente por medio de teledetección de corto alcance. Se utilizó el radiospectrómetro de campo ASD FieldSpec® 3, con un rango espectral de 350-2500 nm y una resolución de ancho de banda de 3 nm a VNIR y 10 nm para SWIR. El índice de clorofila de borde rojo ($CI_{red\ edge}$) fue relacionado linealmente con el CCC con un valor de $R^2 = 0,94$ y

en el caso de cantidad de nitrógeno del dosel fue relacionado linealmente con el $CI_{red\ edge}$ con un valor de $R^2 = 0,88$ para papa y $0,77$ para pasto. De igual forma se encontró una relación exponencial con un $R^2 = 0,85$ para plantas fertilizadas. Se utilizaron datos establecidos para la simulación del modelo PROSAIL, entre estas variables se encuentran concentración de clorofila, LAI, espesor de agua equivalente (EWT), materia seca y parámetro de estructura de la hoja, parámetro de punto caliente, brillo del suelo, radiación directa/difusa, ángulo cenital solar y de vista (Clevers & Kooistra, 2012).

2.2.2 Teledetección de estrés hídrico

En el caso específico de teledetección de estrés hídrico existen diferentes metodologías para la creación de modelos predictivos. Se puede trabajar con diferentes longitudes de ondas para monitorear cambios y correlacionar estos con estrés abiótico.

Chemura et al. (2017) trabajó con el espectro visible y el infrarrojo cercano para predecir el contenido de agua de la planta (PWC por sus siglas en inglés) de café utilizando un algoritmo de bosques aleatorios (random forest). De todas las bandas solo tres fueron identificadas como significantes (485 nm, 670 nm y 885 nm). Las longitudes de onda fueron entrenadas con $n = 36$ casos y probadas en datos independientes ($n = 24$) antes de ser integradas en el algoritmo. Los resultados muestran que las bandas sensibles fueron las que tuvieron mejores rendimientos en la detección de estrés hídrico ($r = 0.87$, raíz del error cuadrático medio (RMSE) = 4.91% y porcentaje de desviación (pBias) = 0.9%), al compararlo con la diferencia de la reflectancia ($r = 0.79$, RMSE = 6.19 y pBias = 2.5%) y las bandas seleccionadas de la validación cruzada ($r = 0.75$, RMSE = 6.52 y pBias = 1.6%). Estos resultados muestran que la región visible y del infrarrojo cercano puede predecir el PWC de las plantas de café y que las bandas escogidas por análisis de sensibilidad producen los resultados de menores errores.

De manera similar por medio de la teledetección de corto alcance se estudiaron los cambios espectrales en la reflectancia de tomates en hidroponía expuestos a diferentes tratamientos de irrigación. La reflectancia aumentó al igual que el déficit de agua. Por medio de índices que fueron correlacionados con el contenido volumétrico de agua del sustrato y la cantidad de clorofila en la hoja. Se obtuvieron resultados que muestran que cuando los valores del índice modificado simple de radio rojo ($mrSRI$ por sus siglas en inglés) y del

índice modificado de vegetación rojo normalizado (*mr*NDVI por sus siglas en inglés) aumentaron por más de 2,5% y 23% respectivamente, el contenido volumétrico de agua del sustrato disminuyó por más de 3%. Además, cuando el valor del índice de reflectancia de absorción de clorofila transformado (TCARI por sus siglas en inglés) se incrementó alrededor de 16%, el contenido de clorofila de la hoja disminuyó alrededor de un 3% (Elvanidi et al., 2018).

2.2.3 Quimiometría

La quimiometría hace referencia a la utilización de métodos estadísticos y/o matemáticos para resolver problemas químicos (Brown et al., 2020). Es una disciplina con diversas aplicaciones que abarcan distintas áreas de la química. Desde la física química, donde se aplica en la cinética y el equilibrio de reacciones, hasta la química orgánica, donde se emplea para optimizar procesos reactivos. En el ámbito de la química teórica, la quimiometría encuentra su principal uso en campos como la cromatografía y la espectroscopía. Estas herramientas analíticas permiten resolver problemas relacionados con otras disciplinas, como el monitoreo ambiental, la biología y las ciencias forenses (Brereton, 2018).

La aplicación de la quimiometría en la cromatografía y la espectroscopía es especialmente relevante para el análisis de muestras complejas. Al utilizar técnicas quimiométricas en estas áreas, es posible obtener información detallada sobre la composición de sustancias presentes en una muestra, lo que resulta invaluable para la identificación y cuantificación de compuestos.

En el campo del monitoreo ambiental, la quimiometría es una herramienta esencial para evaluar la calidad del aire, el agua y el suelo. Permite identificar y cuantificar contaminantes, así como analizar tendencias y patrones en los datos ambientales. Por ejemplo, en el estudio realizado por Lourenco et al. (2020) se comprobó la presencia de hidrocarburos ligeros en una zona de Brasil en donde las empresas petroleras no habían reportado ningún accidente.

Además, la aplicación de la quimiometría en biología y ciencias forenses ha demostrado ser de gran utilidad. En biología, se emplea para el análisis de datos genéticos,

proteómicos y metabolómicos, lo que permite una comprensión más profunda de los sistemas biológicos. En el estudio realizado por Jalalvand et al. (2019) se estudiaron las interacciones de la cabergolina (CBG) con la albúmina de suero humano (HSA) con un modelo multivariado alternado de mínimos cuadrados (MCR-ALS por sus siglas en inglés).

En el ámbito forense, la quimiometría se aplica en el análisis de evidencias, como huellas dactilares, muestras de ADN y análisis de drogas, lo que contribuye a la resolución de casos judiciales de manera más eficiente y precisa. En el artículo publicado por Malegori et al. (2020) se combina la quimiometría con la recolección de imágenes hiperespectrales para la detección de sangre deshidratada, orina y semen en diferentes materiales. Las muestras simuladas mostraron resultados robustos que luego fueron aplicados en casos reales para obtener resultados prometedores.

En el estudio de Zhang (2022) se realizó un análisis multivariado con datos hiperespectrales utilizando la regresión de mínimos cuadrados (PLSR, por sus siglas en inglés) para predecir el contenido de agua relativo (RWC), grosor de agua equivalente (EWT), “fuel moisture content” (FMC) y conductancia estomática (STC) de 10 especies diferentes de árboles. Entre las características de las maderas se encontraban las no porosas (NP), anillo porosas (RP) y las difuso porosas (DP). Los indicadores R^2 y RMSE para la variable RWC fueron de 0 (2.3), 0.88 (4.19) y 0.98 (4.99) para los tipos de madera NP, RP y DP respectivamente. En el caso de la variable EWT indicadores fueron 0.06 (0.22), 0.99 (0.03) y 0.99 (0.05). Seguidamente para la variable FMC 0.07 (8.98), 0.96 (14.71) y 0.99 (4.76). Por último, para STC fueron 0.02 (8.13), 0.2 (45.88) y 0.07 (18.74).

La quimiometría es una poderosa herramienta analítica con aplicaciones diversas en la química y otros. Su capacidad para resolver problemas y proporcionar información detallada la convierte en una aliada imprescindible en diversos campos. Su uso continuo y desarrollo prometen abrir nuevas posibilidades en la comprensión y aplicación de la ciencia en distintos ámbitos.

2.2.4 Métodos de análisis para datos hiperespectrales

La cantidad de información contenida en los datos hiperespectrales es muy amplia y para cuestiones de análisis estadísticos redundante en algunos casos. Por lo tanto, es necesario

seguir una serie de procedimientos para sintetizar la información, este flujo se divide en tres partes: (1) Pre-procesamiento, el cual consiste en transferir los datos de luminosidad en el sensor a reflectancia y aplicar las correcciones necesarias para disminuir la influencia de factores externos; (2) Extracción de información espectral pura relacionada con longitudes de onda únicas para la interpretación y diferentes acercamientos; (3) Clasificación espectral para la creación de mapas de distribución (Kurz & Buckley, 2016).

El preprocesamiento de los datos es importante ya que esta información no deseada puede afectar a la interpretación y habilidad predictiva del modelo. Entre los pretratamientos más conocidos se encuentran los métodos derivativos y multiplicativos de primer y segundo orden y la transformación variable normal estándar (SNV por sus siglas en inglés). El efecto de las primeras derivadas de ambos métodos es eliminar líneas base adicionales. Las segundas derivadas remueven el efecto multiplicativo de la línea de base (Brown et al., 2009).

La técnica de “detrend” (o eliminación de tendencia) se utiliza para eliminar una tendencia lineal o polinómica de los datos espectrales. En el estudio realizado por Hu & Qi (2018) se comprobó que la utilización de este pretratamiento mejoró el rendimiento de un modelo predictivo (PLSR) para la predicción del contenido orgánico del suelo. Las bandas espectrales de 450-619, 760-909 y 1968-2011 nm, fueron las más importantes en la determinación de este parámetro.

La “Standard Normal Variate” (SNV) es una técnica utilizada para corregir la desviación de la intensidad espectral debida a diferencias en la concentración y en la variación de la muestra. Básicamente, la SNV normaliza cada espectro individual restando el promedio del espectro y dividiendo por la desviación estándar. Esto ayuda a eliminar los efectos de intensidad y mejora la comparabilidad entre los espectros (Hobley et al., 2018). El filtro Savitzky-Golay es una técnica utilizada para suavizar los datos espectrales y reducir el ruido mientras se mantiene la forma general de los picos espectrales. Se basa en el ajuste de polinomios locales a los datos para realizar la suavización. El filtro Savitzky-Golay tiene diferentes órdenes y tamaños de ventana que se pueden ajustar según las necesidades del análisis (John, 2021). Los órdenes utilizados en esta investigación fueron polinomios de grado 2, ventanas de 15 y derivadas de primer y segundo grado.

El análisis de componentes principales (PCA, por sus siglas en inglés) es un método utilizado para la exploración de los datos, detección de datos extremos y clústeres, reducción dimensional de los datos y clasificación. En quimiometría es la primera herramienta utilizada para explorar los datos de la variable x (longitudes de onda) y la variable y (datos de laboratorio). Para los datos correlacionados el PCA proyecta una estructura de datos ocultos (latentes) en dos subespacios llamados “score” y “loadings”. La primera proyección es denominada primer componente y el mismo procedimiento es repetido en segundo, tercero y así sucesivamente hasta alcanzar la suma cuadrada de la matriz residual. Cada componente representa el máximo porcentaje de datos representados en cada distribución (Brown, 2009).

3. OBJETIVOS

3.1 Objetivo general

Desarrollar cuatro modelos para la detección temprana de estrés hídrico en dos variedades de plantas de café (Victoria 14 y Turrialba 4730) bajo condiciones controladas por medio de espectroscopía del espectro visible e infrarrojo cercano y datos auxiliares.

3.2 Objetivos específicos

1. Recolectar y analizar datos de estrés hídrico y factores externos que influyen en las plantas de café de dos variedades expuestas a diferentes niveles de estrés hídrico en condiciones controladas, como cantidad de hojas, altura, capacidad fotosintética y potencial hídrico para calibrar los modelos.

2. Recolectar y analizar datos hiperespectrales en condiciones controladas de las plantas de café de dos variedades expuestas a diferentes niveles de estrés hídrico por medio de un espectrorradiómetro como base para los modelos a desarrollar.

3. Generar modelos estadísticos con la correlación de la transreflectancia y datos de capacidad fotosintética en plantas de café de dos variedades expuestas a diferentes niveles de estrés hídrico.

4. METODOLOGIA

El resumen de la metodología planteada para este estudio se resume en la Figura 2, la cual será detallada en las siguientes secciones.

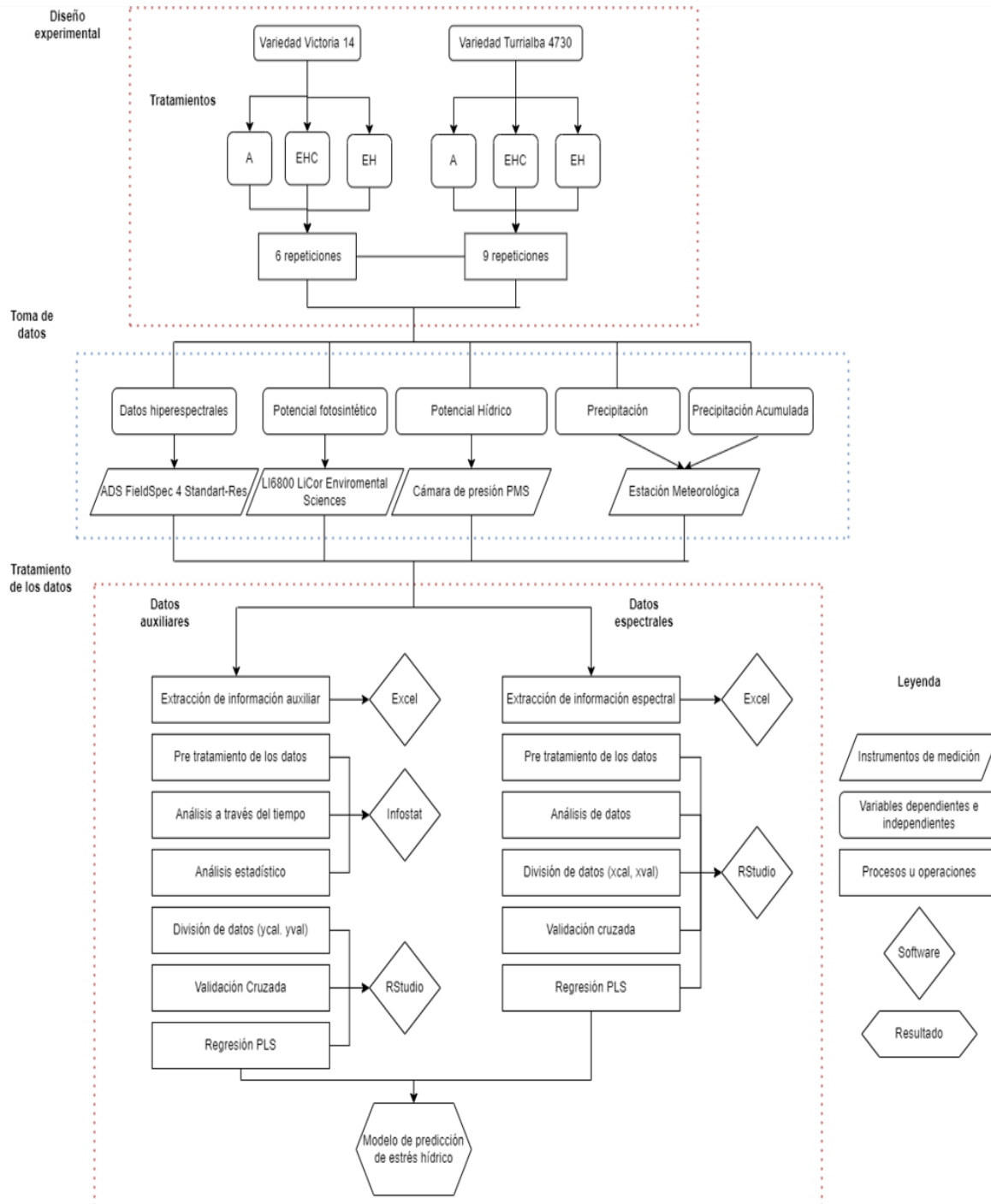


Figura 2. Diagrama de flujo sobre la metodología de la investigación.

4.1 Sitio y objeto de estudio

El sitio de estudio se ubicó en Hacienda Alsacia, la cual es una finca cafetalera de la empresa Starbucks, ubicada en las laderas del Volcán Poás, Sabanilla, Alajuela, Costa Rica. La finca se extiende a lo largo de 240 hectáreas, de las cuales 170 ha están dedicadas al cultivo de diversas variedades de café y 14 ha son utilizadas para la investigación y el mejoramiento genético de esta planta. El departamento de investigación de la empresa proporcionó 144 plantas de café en etapa de crecimiento, 72 de la variedad Turrialba y 72 de Victoria, las cuales se ubicaron en un área de 300 m² dentro de un invernadero tropical asimétrico.

La ubicación geográfica del invernadero es 10°05'49.4"N 84°12'00.0"W y las dimensiones son 10 m de frente y 30 m de fondo. La estructura no se encontraba totalmente aislada de las condiciones externas y presentaba aberturas en ciertas partes causadas por el ingreso de personas y animales. El techo del invernadero fue reemplazado por sarán verde por lo que en temporada lluviosa no se puede controlar el ingreso de las fuertes lluvias características de la temporada. La temporada lluviosa en la zona del Volcán Poás inicia a finales de mayo y se extiende hasta noviembre y la presencia de vientos fuertes también es característica de esta época. Estas dos condiciones fueron limitantes e influyeron directamente con los objetivos de la investigación y sus consecuencias se ven reflejadas en los resultados.

El invernadero contaba con un sistema de riego por aspersión dividido en dos sectores, y el encargado del invernadero era responsable de tomar decisiones relacionadas con el riego del material genético. Sin embargo, no se contaba con información detallada sobre la lámina bruta que recibían las plantas antes del experimento, por lo que era difícil saber si se estaba aplicando el agua necesaria para el óptimo crecimiento de la planta. El sistema de riego estaba controlado por una electroválvula que activa el riego dos veces al día. Ambas variedades que se estudiaron habían sido previamente sometidas a diferentes tiempos de riego durante su desarrollo, normalmente entre 3 y 4 horas diarias dependiendo de las condiciones climáticas.

La estructura inicialmente fabricada para la protección de las muestras tuvo diferentes modificaciones debido a las condiciones climáticas de la zona; en especial el viento característico de la temporada.

Se decidió hacer un diseño más adecuado, conformado de varillas de construcción en las bases y tubos PVC de 6 m como se aprecia en la Figura 3. Luego de realizar pruebas para ver la efectividad de la estructura se decidió empezar el régimen de estrés hídrico de la investigación el día 6 de mayo del 2022. Para esta fecha las plantas tenían alrededor de un año y medio de edad en las condiciones de crecimiento anteriormente mencionadas. La ubicación de la estructura en el espacio presentó un desnivel topográfico leve, el cual con el exceso de precipitación facilitó el flujo de agua por la parte inferior de las macetas.



Figura 3. Estructura final para la protección de las muestras.

4.1.1 Material vegetal y condiciones de crecimiento

Para este estudio se utilizaron plantas de café de dos variedades diferentes: Victoria 14 y Turrialba 4730. La variedad Victoria es propia de Starbucks, está en F9 y estable; mientras que la variedad Turrialba es una línea pura de una colección del CATIE y aún no se considera una variedad comercial. Cabe resaltar que, según el Ing. Agrónomo Carlos Mario Rodríguez, director del departamento de mejoramiento genético de Hacienda Alsacia, la variedad Victoria posee un sistema radicular robusto al ser un híbrido en comparación con la

variedad Turrialba. Esta fue una de las razones principales a la hora de escoger el material genético adecuado para el estudio. Además, es importante generar información para Hacienda Alsacia sobre su variedad exclusiva del mejoramiento genético de la empresa, con dos sistemas radiculares diferentes.

La edad de las plantas era de alrededor de un año al momento de la primera visita a Hacienda Alsacia para discutir los objetivos de la investigación. Se encontraban en bolsas negras de plástico con un sustrato compuesto de una mezcla de suelo, turba y arena en proporciones de 1:1:1 ubicadas en el suelo del invernadero. La utilización de plantas de café en esta etapa fenológica radica en la sensibilidad a cambios en su metabolismo causado por su sistema radicular en desarrollo. El monitoreo desde una temprana etapa fenológica asegura el desarrollo óptimo y el buen rendimiento en su etapa productiva.

En su crecimiento las plantas fueron fertilizadas con Oscomote como fertilizante de liberación lenta durante seis meses, seguido de dos aplicaciones al año de abono granulado con fórmula de 18-5-15 y 6-0-2 (N-P-K) y dos aplicaciones al año de fertilizante para hidroponía de macronutrientes y micronutrientes. Las plantas fueron trasplantadas el 15 de marzo del 2022 a maceteros de plástico con medidas de 22.5 cm de boca, 22.5 cm de altura y 17.0 cm de diámetro inferior.

Las muestras no se encontraban en condiciones óptimas de crecimiento ya que muchas de ellas presentaban en su mayor parte plagas y enfermedades que afectaban la sanidad de las plantas. Se pudo apreciar en su mayoría presencia de cochinilla y mancha de hierro. Las 144 plantas donadas por Hacienda Alsacia formaban parte del material genético reservado para estudios de investigación y es el número de plantas que se podía utilizar para la realización de esta investigación.

4.2 Diseño Experimental

La unidad experimental para este estudio es una planta de café. Para la ubicación de estas, se utilizó el diseño de bloques al azar balanceado, que es uno de los más comunes en la experimentación agrícola y aumenta la precisión del resultado al reducir los errores mediante el agrupamiento en bloques (Hernández et al., 2019). En total, se utilizaron 144 plantas, manejando la mitad por variedad (72 unidades).

Se aplicaron tres tratamientos: el control o tratamiento A, tratamiento con estrés hídrico con producto (EHC) y el tratamiento con estrés hídrico y sin aplicación de producto (EH). Cada tratamiento está conformado de 24 plantas. En la Figura 4 se muestra un diagrama con la distribución de las muestras respecto al espacio físico. Cada fila está compuesta por dos grupos de cuatro plantas para cada tratamiento distribuidos de forma aleatoria.

Estos tratamientos fueron diseñados para someter a las unidades experimentales a un régimen de estrés hídrico progresivo y estudiar la respuesta de la capacidad fotosintética y las diferencias en las firmas espectrales y otras variables de respuesta como los datos auxiliares de altura y cantidad de hojas.

En el tratamiento A las plantas fueron sometidas a 30 minutos de riego por goteo cada 12 horas desde el 6 de mayo hasta el 13 de junio del 2022.

El tratamiento EHC estuvo en un régimen sin riego con la aplicación de un bioestimulante a base de extractos de algas llamado Seamaxx (Registro SFE #7553) y un protector del estrés abiótico que contiene óxido de zinc, óxido de silicio e ingredientes orgánicos llamado Photon (Registro SFE #10144). Los insumos agrícolas fueron aplicados 7 días antes y 10 días después de iniciado el periodo de déficit hídrico (PDH) de manera similar a la metodología seguida por Fernández et al. (2020). Ambos productos fueron aplicados de manera foliar mezclados a una dosis de 3 ml/L de Seamaxx y 30 ml/L de Photon. La mitigación del estrés hídrico sobre el potencial hídrico y la reducción en el porcentaje de defoliación de las plantas de café bajo condiciones de estrés hídrico generada por la aplicación de Seamaxx ha sido estudiada con detenimiento en el artículo de Fernández (2020). El protector solar Photon refleja y dispersa la radiación ultravioleta y estimula procesos fisiológicos de las plantas en diferentes etapas fenológicas. Entre sus ventajas se encuentra el mejoramiento del proceso fotosintético y la consolidación de mecanismos de resistencia a daños abióticos.

Por último, las plantas de café del tratamiento EH no recibieron irrigación por goteo ni tampoco aplicaciones de productos adicionales en todo el periodo de 38 días.

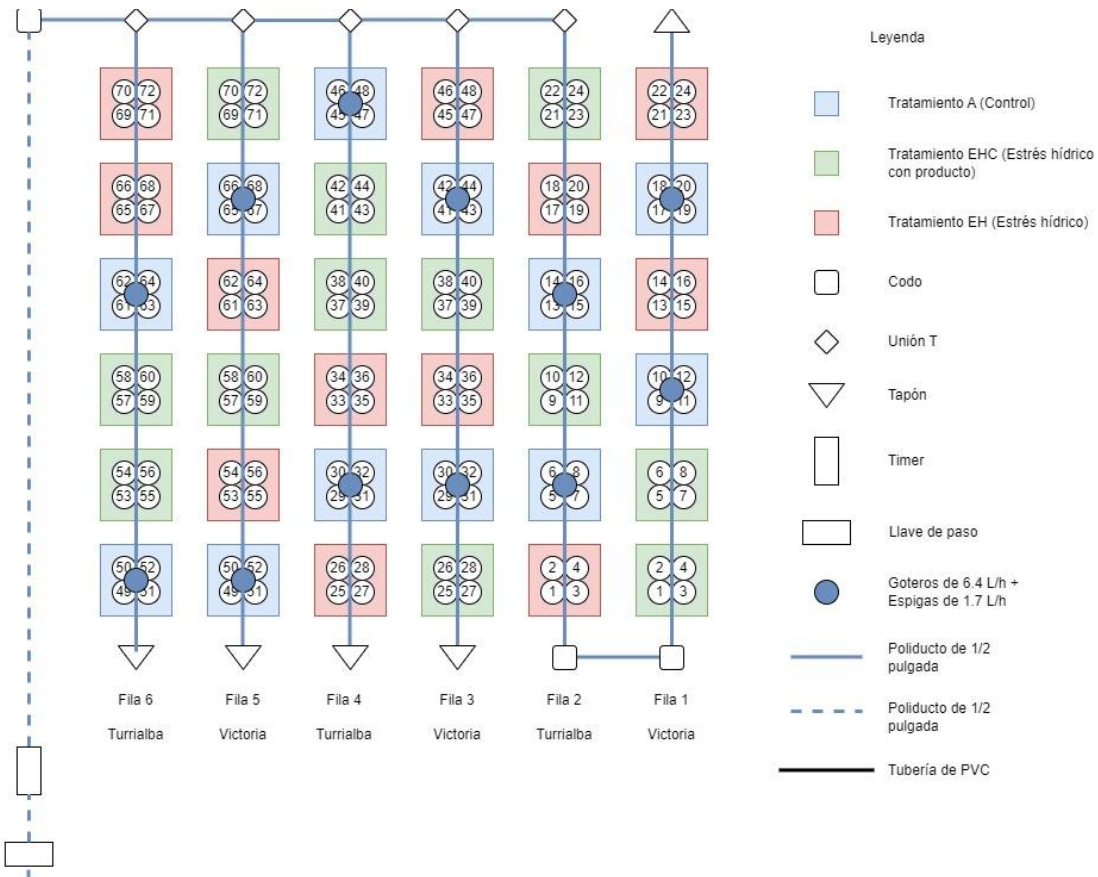


Figura 4. Distribución espacial de las muestras, tratamientos y ubicación del sistema de riego.

4.2.1 Variables Independientes

Las variables independientes son aquellas que se presumen tener un efecto causal sobre las variables dependientes. El control o manipulación en un experimento es fundamental para comprobar esas relaciones (Galarza, 2021).

Para esta investigación se considera como variables independientes: el tiempo, la variedad del café (Victoria y Turrialba), los tratamientos (A, EHC y EH) y las firmas espectrales obtenidas de cada hoja.

Es importante destacar que, en esta etapa experimental, la variable del tiempo jugó un papel crucial. Se prestó especial atención al comportamiento de las variables de respuesta a lo largo del tiempo y a las interacciones que puedan surgir con otras variables independientes y auxiliares. Se estableció un periodo de tiempo de 38 días en total entre los

meses de mayo y junio. Las visitas fueron realizadas los días 6, 13 y 26 de mayo y 6 y 13 de junio.

4.2.2 Variables dependientes y datos auxiliares

Las variables dependientes en una investigación científica son los resultados o fenómenos que se estudian y se espera que cambien en respuesta a las variables independientes. Su importancia radica en que nos permiten comprender, explicar o predecir fenómenos, así como obtener información sobre relaciones causales y efectos de intervenciones o tratamientos (Galarza, 2021).

Para esta investigación en el desarrollo de los modelos de PLSR se considera como variables dependientes las asociadas a la capacidad fotosintética: tasa de transpiración, tasa de asimilación neta de CO₂, temperatura de la hoja y presión de vapor de saturación; y como datos auxiliares el potencial hídrico del tallo, la altura de las plantas y cantidad de hojas.

En el caso del análisis ANOVA se consideraron como variables dependientes el potencial hídrico del tallo, altura de las muestras y cantidad de hojas para encontrar diferencias significativas que ayuden a describir el comportamiento a través del tiempo entre variedades y los tratamientos.

En el Cuadro 1 se muestran las variables dependientes recolectadas en esta sección experimental con sus días de medición respectivos y diferentes características.

Cuadro 1. Resumen de variables dependientes recolectadas y sus características.

<i>Dato Recolectado</i>	<i>Día de visita</i>	<i>Medición por unidad experimental</i>	<i>Repeticiones</i>	<i>Total de mediciones por día</i>
<i>Tasa de transpiración</i>	0, 7, 17 y 38	mol m ⁻² s ⁻¹	6	54
<i>Tasa de asimilación</i>	0, 7, 17 y 38	μmol m ⁻² s ⁻¹	6	54
<i>Temperatura de la hoja</i>	0, 7, 17 y 38	°C	6	54
<i>Presión de vapor de saturación</i>	0, 7, 17 y 38	kPa	6	54
<i>Potencial hídrico</i>	0, 7, 17, 31 y 38	bar	3	18
<i>Altura</i>	0, 7, 17, 31 y 38	cm	-	144
<i>Cantidad de hojas</i>	0, 7, 17, 31 y 38	unidad	-	144

4.3 Recolección de datos

La recolección de datos se realizó en visitas al invernadero los días 6, 13 y 26 de mayo y 6 y 13 de junio. Las mediciones iniciaron a las 6:00 am y se extendieron hasta las 10:00 am aproximadamente. A continuación, se detalla el procedimiento para la toma de información.

4.3.1 Datos de capacidad fotosintética y potencial hídrico

Las variables dependientes relacionadas a la capacidad fotosintética: tasa de transpiración (E), tasa de asimilación neta de CO₂ (A), temperatura de la hoja (Tleaf), presión de vapor de saturación (SVPLleaf) se registraron utilizando el medidor portátil de fotosíntesis LI-6800 LiCor Environmental Sciences (Bluestem OS v2.0).

El medidor de fotosíntesis tiene una tasa de flujo a granel de 680-1700 $\mu\text{mol s}^{-1}$, una uniformidad de fuente de luz de $<\pm 10\%$ de variación por encima del 90% de apertura y un rango de control de temperatura en la cámara de 10°C por encima o por debajo de la temperatura ambiente.

Para la medición de las variables se planteó la utilización de hojas maduras del segundo al tercer nudo del meristemo apical. Las hojas enfermas o muy jóvenes no fueron consideradas en la investigación.

Durante el día de la recolección del dato, antes de tomar la primera medición, se calentó el equipo por al menos 30 minutos para estabilizar las condiciones. Se seleccionó una hoja por planta para llevar a cabo las mediciones, y después de un minuto de estabilización de las condiciones externas de la hoja, se registraron los valores y se continuó con la siguiente planta.

La toma de datos de capacidad fotosintética se vió afectada por las condiciones climáticas de la zona y la rehidratación no deseada de las muestras, lo cual modificó los días de medición y los datos considerados. El día 31 (03 de junio) no se realizaron las mediciones de capacidad fotosintética y los datos recolectados el día 7 (13 de mayo) no fueron considerados en el análisis de las variables de capacidad fotosintética. Estas decisiones fueron tomadas conforme la investigación avanzó, al analizar los datos de los días anteriores.

El potencial hídrico del tallo se midió con una cámara de presión Scholander PMS. La cámara tiene un rango de medición de 0-40 bar, un manómetro clase 0.5 y una cámara cilíndrica de acero inoxidable. La preparación de las muestras para la medición del potencial hídrico del tallo es clave para obtener un resultado preciso. Un día antes de la medición se forraron las hojas con papel aluminio para evitar el contacto con la luz y se embolsaron con bolsas de cierre hermético. Se aseguraron con un clip para evitar el ingreso de agua y mantener la estabilidad en la presión del tallo.

La lectura del potencial hídrico del tallo consistió en aplicar presión constante del gas dentro de la cámara hasta la aparición de la primera gota de savia visible en la superficie del corte del pecíolo, momento en el cual se registró la presión del tallo en el manómetro de la cámara (Scholander et al., 1965).

Los equipos utilizados para medir la capacidad fotosintética (LI-6800 LiCor Environmental Sciences) y el potencial hídrico (Cámara de presión Scholander) fueron proporcionados por LIFE-RID, que es el departamento de Investigación, Innovación y Desarrollo de la empresa AMVAC LatAm.

4.3.2 Altura de las muestras y cantidad de hojas

La medición de la altura de las plantas se realizó con una cinta métrica desde el comienzo del tallo hasta el último nudo, y se registraron los datos separados por planta presencia de dos plantas por maceta. De igual forma, para la cantidad de hojas se tomaron en cuenta ambas plantas por separado. Para facilitar el procesamiento y visualización de datos se utilizaron únicamente los datos de la planta dominante o más vigorosa.

4.3.3 Datos hiperespectrales

Se realizaron mediciones hiperespectrales con el radioespectrómetro de campo ASD FieldSpec® 4 Standard-Res proporcionado por la Escuela de Ingeniería de Biosistemas de la Universidad de Costa Rica. Este instrumento tiene un rango espectral de 350-2500 nm y una resolución de ancho de banda de 3 nm en VNIR y 10 nm en SWIR. El rango espectral de los 700 a los 1300 nm corresponde al infrarrojo cercano (NIR, por sus siglas en inglés) y es crucial para capturar los procesos fisiológicos de captación y retención de agua en las plantas (Wijewardana, 2019). Se utilizó el accesorio "Plant contact probe" que tiene su propia fuente de luz, lo que evita variaciones en la radiación externa y humedad que puedan afectar las mediciones.

Para la toma de datos hiperespectrales, primeramente, se calentó el equipo por 30 minutos antes de iniciar el muestreo y se realizó una optimización por calentamiento del equipo y una corrección con el blanco de referencia (Spectralon) cada 30 min para efectuar las calibraciones del instrumento, eliminar ruidos por calentamiento y obtener la transfectancia (T) de la hoja. Se decidió utilizar la transfectancia por la naturaleza de las hojas jóvenes que dejan pasar parte de la luz cuando son iluminadas (con algún grado de transparencia). De igual forma, esta investigación busca innovar métodos de análisis en la espectroscopía y plantea establecer una línea base para la continuación de estudios en donde se utilice transfectancia.

En el estudio de Piccolo (2022) se utilizó tanto la reflectancia como transflectancia con técnicas de “*machine learning*” para determinar el contenido total de antocianinas (ACC, por sus siglas en inglés) en hojas de *Prunus cerasifera*. Los resultados muestran medidas de RMSEP de 0.3 y menores con respecto a la fracción de los datos de validación con la totalidad de datos.

En el Cuadro 2 se muestra la firma espectral como variable independiente recolectada en esta sección. La altura de la planta puede influir en las mediciones de variables debido a su impacto en la disponibilidad de nutrientes y agua. A medida que la planta crece, sus raíces exploran diferentes capas del suelo, lo que puede afectar la absorción de nutrientes y agua (Peña et al., 2018). Por lo tanto, se tomaron tres mediciones por estrato en las unidades experimentales seleccionadas al azar para el día medición. Se asume que la medición relacionada a la capacidad fotosintética es la misma para las tres hojas de los diferentes estratos. Los estratos definidos para las variedades de café se presentan en el Cuadro 3.

Cuadro 2. Características de las firmas espectrales recolectadas.

<i>Dato Recolectado</i>	<i>Día de visita</i>	<i>Medición por unidad experimental</i>	<i>Repeticiones</i>	<i>Total de mediciones por día</i>
<i>Datos espectrales</i>	0, 7, 17, 31 y 38	nm	6	1296

Cuadro 3. Estratos para las diferentes variedades de café

<i>Variedad</i>	<i>Estrato alto (cm)</i>	<i>Estrato medio (cm)</i>	<i>Estrato bajo (cm)</i>
<i>Turrialba</i>	30 – 40	15 – 30	0 – 15
<i>Victoria</i>	24 – 40	12 – 24	0 – 12

Al igual que en la primera etapa experimental solo se consideraron las hojas desarrolladas para el muestreo, independientemente de su coloración. Las hojas nuevas o secas no se incluyeron en la investigación.

4.4 Análisis de las variables fisiológicas

Se analizó el comportamiento de todas las variables dependientes a través del tiempo y se escogió el rango de tiempo con el cual se realizaron los análisis basados en el comportamiento y datos meteorológicos de la zona durante la investigación. Esto con la finalidad de analizar los cambios de las variables en un régimen de estrés hídrico bajo los diferentes tratamientos para las variedades de café Victoria y Turrialba. En esta sección el periodo de déficit hídrico real (PDH real) únicamente toma en cuenta los días en donde las plantas de los tratamientos EHC y EH estuvieron bajo un régimen de estrés hídrico. Esto hace referencia a las mediciones de los días 23 de mayo y 13 de junio del 2022. Por contrario, el periodo de déficit hídrico teórico (PDH teórico) fue el que se tomó en cuenta para el análisis de esta sección. Este comprende los días 0, 17 y 38 que corresponden al 6 y 23 de mayo y al 13 de junio del 2022 (los dos últimos corresponderían al día 0 y 21 del PDH real).

El análisis de varianza o ANOVA se utilizó para determinar si hay diferencias significativas entre los grupos, mientras que la prueba de Tukey se utilizó para identificar qué grupos difieren significativamente entre sí. En otras palabras, el ANOVA proporciona una estimación de la significancia global del efecto, y la prueba de Tukey proporciona información más detallada sobre las diferencias específicas entre grupos (Ross, 2018). Ambos análisis fueron realizados con un nivel de significancia de 0.05 para las variables citadas en el Cuadro 1.

4.5 Caracterización y evaluación hiperespectral

Los datos hiperespectrales representan las características y propiedades de las unidades experimentales que se utilizan como entrada en un modelo predictivo. Como se mencionó en la sección 4.2.2, la variable independiente para el análisis de modelos espectrales es el conjunto de firmas obtenidas de las plantas de café. Así mismos, las variables dependientes son las variables que se quieren predecir, en este caso: la tasa de transpiración, tasa de asimilación neta de CO₂, temperatura de la hoja en el balance de energía y la presión de saturación a la temperatura de la hoja. El potencial hídrico del tallo, la altura de la planta y la defoliación son datos auxiliares que ayudan a explicar los fenómenos generados. Es importante mencionar que en esta sección la variable independiente no es el tiempo por lo

tanto se utilizaron tanto los datos hiperespectrales como los de la capacidad fotosintética de los días 0, 17 y 38.

Para la creación del modelo predictivo con datos hiperespectrales se debe de seguir una metodología sistemática que abarca varias etapas. En primer lugar, se realiza la recopilación de datos espectrales como se indicada en la sección 4.3, la cual involucra el uso de tecnologías como la espectroscopía o la teledetección. Posteriormente se realiza la preparación de los datos, que incluye el preprocesamiento, la normalización y la selección de características relevantes. Luego, se procede a la construcción del modelo, utilizando técnicas estadísticas adecuadas, como la regresión de mínimos cuadrados parciales (PLSR, por sus siglas en inglés). Durante esta etapa, es esencial dividir los datos en conjuntos de entrenamiento (70% de las muestras) y prueba (30%) para evaluar la capacidad de generalización del modelo (Thenkabail, 2019b).

Una vez construido el modelo, se lleva a cabo su validación utilizando medidas de rendimiento, como la raíz del error medio cuadrático de la predicción (RMSEP, por sus siglas en inglés) o el coeficiente de determinación ajustado (R^2). Finalmente, se realiza una etapa de optimización del modelo, que implica ajustar los parámetros (variables latentes) y validar nuevamente el rendimiento. Esta metodología proporciona un enfoque estructurado y riguroso para la creación de modelos predictivos basados en datos espectrales, permitiendo la generación de modelos precisos y confiables para la predicción de variables de interés (Thenkabail, 2019a).

Parte de la metodología utilizada en este estudio fue respaldada con la capacitación práctica sobre espectroscopia de infrarrojos para el monitoreo de carbono del suelo. Esta capacitación fue realizada por el Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD, por sus siglas en francés) en las fechas del 31 de octubre al 4 de noviembre del 2022 en la estación experimental Alfredo Volio Mata de la Universidad de Costa Rica (UCR). Esta actividad fue parte de un programa de asistencia técnica a las instituciones públicas de Costa Rica en la implementación del Plan Nacional de Descarbonización. La capacitación fue dirigida por Julien Demenois (CIRAD, CATIE) y Gilies Chaix (CIRAD) en donde se abordaron los temas más relevantes de la creación de modelos predictivos.

4.5.1 Preprocesamiento de información espectral

Las firmas espectrales de transreflectancia (T) obtenidas se procesaron con el software View Spec Pro para obtener el log (1/T) antes de la transformación de datos tipo .asd a tipo .txt para ser manejados con el lenguaje de programación R con el ambiente de entorno de R Studio.

Se utilizó el entorno de RStudio para la exploración de los datos espectrales; una sección del código utilizado se encuentra en el apéndice en la sección C. Se utilizaron los paquetes: readxl, readr (Hadley, sf), remotes, rnirs y rchemo. Los primeros dos son paquetes con funciones para trabajar con archivos de Excel y datos con formatos tabulares o rectangulares. El paquete “remotes” permite importar paquetes de repositorios locales, un ejemplo muy popular es la plataforma GitHub. Los últimos dos paquetes son exclusivos para la reducción, regresión y discriminación de datos espectrales (Lesnoff, 2021).

4.5.1.1 Gap Remove

Primeramente, los datos de las firmas espectrales se promediaron para cada medición del estrato alto y medio. La eliminación de brechas o “*gap remove*” en datos espectrales es una operación necesaria para que el cambio de sensor no afecte la calidad del análisis. En este caso, se utilizó la función en las longitudes de onda de 1000 nm y los 1800 nm, ya que, en estas longitudes de onda existe un cambio de sensor del equipo y se presenta una baja señal (Brown et al., 2020). Seguidamente, se reduce la longitud de onda con la cual se registró en cada medición.

4.5.1.2 Pretratamientos espectrales

El uso de diferentes técnicas de pretratamientos de datos hiperespectrales antes de generar el modelo aborda diferentes problemas en el preprocesamiento de datos espectrales y tiene un efecto distinto en los datos y pueden afectar el desempeño de este. El tipo de pretratamiento y el orden de aplicación son factores que pueden inducir a cambios en los resultados.

En esta investigación se probaron diferentes combinaciones de pretratamientos de datos espectrales antes mencionados: “detrend”, standar normal variate (SNV) y Savitzky-

Golay con primera y segunda derivada para determinar, cuál combinación da el mejor rendimiento del modelo de la variable, dando como resultado un total de 10 combinaciones para cada modelo de cada variable y variedad.

A continuación, en el Cuadro 4 se presentan los pretratamientos aplicados y su nomenclatura respectiva. En algunos casos se presentan combinaciones de algoritmos y diferentes órdenes de estos. Se compararon los resultados de los modelos con los diferentes pretratamientos para encontrar el que mejor se ajusta a cada variedad y variable a predecir.

Cuadro 4. Nomenclatura definida para las combinaciones de pretratamientos utilizados.

<i>Código</i>	<i>Pretratamiento</i>
<i>no</i>	Sin pretratamiento
<i>snv</i>	Standard Normal Variate
<i>detrend</i>	Detrend
<i>savgol</i>	Savitzky-Golay
<i>savgolsnv</i>	Savitzky-Golay + Standard Normal Variate
<i>savgol1</i>	Savitzky-Golay 1
<i>savgol1snv</i>	Savitzky-Golay 1 + Standard Normal Variate
<i>snvsavgol1</i>	Standard Normal Variate + Savitzky-Golay 1
<i>savgol2</i>	Savitzky-Golay 2
<i>savgol2snv</i>	Savitzky-Golay 2 + Standard Normal Variate
<i>snvsavgol2</i>	Standard Normal Variate + Savitzky-Golay 2

NOTA. Los números del pretratamiento Savitzky-Golay hacen referencia al número de derivada utilizada.

4.5.2 Creación del modelo predictivo

Tomando como base los resultados de las pruebas del ANOVA de los parámetros de capacidad fotosintética, se realizan modelos predictivos para cada variedad y parámetro. También es importante mencionar que para la conformación de los modelos predictivos se utilizaron tanto las firmas espectrales como los datos de capacidad fotosintética de los días 1, 7 y 38 y para cada uno de estos se probaron las combinaciones de pretratamientos definidos

en el Cuadro 4. Como referencia para el lector se designará una nomenclatura para cada modelo generado. A continuación se presenta un cuadro resumen de los modelos:

Cuadro 5. Cuadro resumen de los modelos generados.

<i>Modelo</i>	<i>Variable</i>	<i>Descripción</i>
1	E	Turrialba + datos estrato alto
2	E	Victoria + datos estrato alto
3	E	Turrialba + datos estrato medio
4	E	Victoria + datos del estrato medio
5	A	Turrialba + datos estrato alto
6	A	Victoria + datos estrato alto
7	A	Turrialba + datos estrato medio
8	A	Victoria + datos del estrato medio
9	Tleaf	Turrialba + datos estrato alto
10	Tleaf	Victoria + datos estrato alto
11	Tleaf	Turrialba + datos estrato medio
12	Tleaf	Victoria + datos del estrato medio
13	SVPleaf	Turrialba + datos estrato alto
14	SVPleaf	Victoria + datos estrato alto
15	SVPleaf	Turrialba + datos estrato medio
16	SVPleaf	Victoria + datos del estrato medio

4.5.2.1 Análisis de componentes principales (PCA)

Previo al PCA se utilizó la función `savgol` del paquete `rnirs` (Lesnoff, 2021) con un número de filtro $n = 25$, orden polinomial $p = 3$ y orden de derivación $m = 2$. La razón de esto es para que los resultados del PCA sean más precisos. Se utilizó la función `pcasvd` del paquete `rchemo` (Lesnoff, 2021) de RStudio. Esta función es un algoritmo centrado y con peso de análisis de componentes principales (PCA, por sus siglas en inglés).

Se calcularon 20 componentes principales (PC) para los datos hiperespectrales, con esta cantidad se logra capturar una proporción sustancial de la variabilidad presente en los

datos originales. El utilizar un número excesivo de componentes principales puede conducir a un sobreajuste (*overfitting*) de los datos de entrenamiento, lo que puede llevar a un rendimiento deficiente. Se utilizó adicionalmente la función descomposición de valor singular (SVD, por sus siglas en inglés) para proporcionar una forma estable de calcular los componentes principales. Además, en términos de eficiencia computacional permite realizar el PCA para grandes conjuntos de manera rápida y eficiente en comparación con otros métodos. Por último, este algoritmo da como resultado una solución única, lo que significa que no hay ambigüedad en los resultados (Thenkabail, 2019b).

El PCA se aplica como un análisis exploratorio a los datos para la verificación de datos atípicos y la comprobación de la partición de los datos de calibración y validación de los modelos, de tal forma que se busca que las muestras se encuentren uniformemente contenidas en ambos sets de datos. Por otro lado, los puntajes representan la posición de cada observación en relación con los componentes principales y se pueden utilizar para clasificar las observaciones en diferentes estados o grupos (Beattie, 2021).

4.5.2.2 Separación de datos de calibración y validación

Para entrenar el modelo de regresión fue necesario, primeramente, dividir tanto la matriz de datos espectrales (x) y las variables de respuesta (y) (ejemplo, la tasa de transpiración) en un set de datos de calibración y validación. Para los datos de calibración se tomó el 80% y el 20% restante corresponde a los datos tomados para el set de validación. Estos fueron llamados para mayor facilidad como y_{cal} , y_{val} , x_{cal} y x_{val} .

La validación cruzada es una técnica utilizada en el análisis de datos y el aprendizaje automático para evaluar el rendimiento de un modelo estadístico o predictivo en un conjunto de datos independiente. Se utilizó la función `segmkf` del paquete `rchemo` (Lesnoff, 2021). Esta función construye los segmentos de observación para la validación cruzada con Monte Carlo (K-Fold y Cross Validation (CV)). Los parámetros de entrada son el número de grupos o folds ($k=2$), número de repeticiones ($nrep=20$) y el set de datos de x_{cal} .

4.5.3 Regresión de mínimos cuadrados parciales (PLSR)

La regresión de mínimos cuadrados parciales (PLS, por sus siglas en inglés) es un método estadístico y de modelado utilizado en el análisis de datos multivariados. PLS se

utiliza principalmente en la regresión y en la modelización de relaciones complejas entre múltiples variables independientes y una variable dependiente (Hanseler, 2018).

En este análisis se utilizó la función `cvfit` del paquete `rnirs` (Lesnoff, 2021). Esta función implementa la validación cruzada en un modelo predictivo con diferentes métodos, en este caso, PLSR. Como parámetros de entrada se tienen las matrices `xcal` y `yca` y una lista de segmentos que usualmente es el output de la función `segmkf` de CV. Seguidamente, el código busca el valor de RMSEP menor con respecto al número de variable latente utilizada para mostrar el mejor resultado de R^2 ajustado.

4.5.4 Eficiencia de los modelos

Los análisis de regresión conforman una parte muy importante en el aprendizaje supervisado y consisten en la predicción de una variable independiente en particular de un set de datos de otras variables de predicción. Muchos estudios utilizan el error cuadrado promedio (MSE, por sus siglas en inglés) y la raíz del error medio cuadrado (RMSE, por sus siglas en inglés). En el artículo de Chicco (2021) se explora la capacidad predictiva del estadístico R^2 y de otros parámetros. Los resultados obtenidos muestran que el R^2 y RMSEP tienen una capacidad descriptiva alta para modelos predictivos que otros índices.

Los índices y sus respectivas fórmulas utilizados para evaluar la calidad y el rendimiento de los modelos de regresión en términos de precisión, error y ajuste a los datos observados se mencionan en el Cuadro 6. El número de componentes principales (NCOMP), número de muestras (NBPRED), error medio cuadrado de predicción (MSEP), raíz del error medio cuadrado de predicción (RMSE), error estándar de la predicción (SEP), pendiente de la línea de regresión (b) y el coeficiente de determinación (R^2) representan métricas y estadísticas claves utilizadas para evaluar la calidad y el rendimiento de los modelos de regresión en términos de precisión, error y ajuste a los datos observados.

Cuadro 6. Estadísticos utilizados en la determinación del rendimiento de los modelos predictivos.

<i>Estadístico</i>	<i>Unidades</i>	<i>Ecuación</i>
<i>Coficiente de determinación (R²)</i>	Adimensional	$r^2 = \frac{(\sum_{i=1}^n \bar{y}_i y_i - \sum_{i=1}^n \bar{y}_i \sum_{i=1}^n \frac{y_i}{n})^2}{(\sum_{i=1}^n \bar{y}_i^2 - \frac{(\sum_{i=1}^n \bar{y}_i)^2}{n}) (\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n})}$ (1)
<i>Error estándar de predicción (SEP)</i>	Igual al valor de referencia	$SEP = \sqrt{\frac{\sum_{i=1}^n (\bar{y}_i - y_i - bias)^2}{n - 1}}$ (2)
<i>Raíz media cuadrada del error de la predicción (RMSEP)</i>	Igual al valor de referencia	$RMSEP = \sqrt{\frac{\sum_{i=1}^n (\bar{y}_i - y_i)^2}{n}}$ (3)
<i>Bias (d)</i>	Igual al valor de referencia	$d = \frac{\sum_{i=1}^n (\bar{y}_i - y_i)^2}{n}$ (4)
<i>Radio del rendimiento de la desviación (RPD)</i>	Adimensional	$RPD = \frac{Sd_y}{SEP}$ (5)

En donde:

$$RMSEP^2 = SEP^2 + Bias^2 \quad (6)$$

$\bar{y}_i = i^{th}$ valor predicho de la muestra de validación.

$y_i = i^{th}$ valor de referencia de la muestra de validación.

$Sd_y =$ desviación estándar de los valores de referencia del conjunto de validación.

$n =$ número de muestras en el conjunto de validación.

En esta investigación se define el coeficiente de determinación ajustado (R^2) como el parámetro principal para determinar la capacidad predictiva de los modelos. Se utilizó por su capacidad para penalizar el agregar variables independientes que no ayudan a predecir la variable dependiente. Entre otras de sus ventajas se encuentra que está ajustado para el

número respectivo de grados de libertad y considera el número total de la muestra y el número de variables independientes (Butvinik, sf). El valor negativo de R^2 está asociado al cálculo de este. La diferencia entre el cálculo de R^2 y R^2 ajustado se presenta a continuación:

$$R^2 = 1 - \frac{RSS}{TSS} \quad (0.1) \quad (7)$$

$$ajustado - R^2 = \left\{ 1 - \left[\frac{(1 - R^2)(n - 1)}{n - k - 1} \right] \right\} \quad (0.1) \text{ o negativo} \quad (8)$$

Siendo:

n = número de observaciones.

RSS = Suma de cuadrados residual.

TSS = Suma de cuadrados total.

ESS = Suma de cuadrados explicada.

5. RESULTADOS

5.1 Evaluación Experimental: Comportamiento temporal de las variables fisiológicas

A lo largo de un periodo de 38 días se realizaron cinco visitas, iniciando el 06 de mayo y finalizando el 13 de junio del 2022. Se recolectaron 7 variables dependientes en dos variedades de café bajo condiciones protegidas: altura, cantidad de hojas, potencial hídrico, tasa de transpiración, tasa de asimilación neta de CO₂, temperatura de la hoja y presión de vapor de saturación a la temperatura de la hoja. Se definieron tres tratamientos: control (A), estrés hídrico con producto (EHC) y estrés hídrico sin producto (EH) y se midieron datos hiperespectrales de las hojas. Es importante mencionar que el periodo de investigación establecido (38 días) se afectó por diferentes factores, mencionados a continuación. En la Figura 5 se muestra una línea de tiempo de la investigación en donde se detallan fechas y acontecimientos importantes, los cuales se detallaran a continuación.



Figura 5. Línea de tiempo de la investigación 2022

Primeramente, se presentó una rehidratación de las muestras del día 7 al 17 por filtraciones de precipitación, lo cual se puede visualizar en los datos del Cuadro A.1. Esto se corroboró con datos de precipitación incremental y acumulada de la estación meteorológica más cercana. En la Figura A.2 se muestra un total de 57.5 mm de precipitación acumulada desde el día 14 de mayo hasta el 23 de mayo del 2022. Estas fechas coinciden con el día 7

(13 de mayo) y el 17 (23 de mayo) de recolección de datos. Se realizó una inspección exhaustiva en la estructura donde se encontraban ubicadas las unidades experimentales para verificar cualquier afectación en el experimento. Por lo tanto, se decidió definir dos periodos de deficit hídrico o deshidratación: uno teórico (PDF teórico) y uno real (PDF real). Estos se identificaron en la Figura 5 con una coloración celeste y roja respectivamente.

El diseño de la estructura de protección en contra de la lluvia no permitió una deshidratación homogénea de todas las unidades experimentales. El terreno tenía una leve inclinación en dirección Noreste-Suroeste. El suelo estaba cubierto por sarán negro, el cual mantenía la humedad luego de fuertes lluvias. Por último, la posición de las unidades experimentales influía en el desarrollo de la deshidratación. Las plantas de café ubicadas en los extremos de la estructura (fila 1 y 6) presentaron una lenta deshidratación a diferencia de las muestras ubicadas en el centro. Por consiguiente, se tomaron varias medidas para aislar las plantas de posibles filtraciones y acelerar la deshidratación.

En el día 20 se instalaron bolsas plásticas para evitar el contacto directo con el suelo (Figura 6). Rápidamente, se notó que las bolsas propiciaban el encharcamiento de agua y fueron reemplazadas por tarimas el día 30 (Figura 7). Los resultados positivos del nuevo sistema se pueden justificar con los datos del potencial hídrico del tallo el día 31 y 38 mostrados en el Cuadro A.1. A manera de resumen en el Cuadro A.9 se presentan las relaciones porcentuales de ambas variedades entre el tratamiento EH con respecto a EHC para el día 31 y 38. Este cuadro hace referencia numérica al comportamiento entre los tratamientos de interés para los días finales de la investigación.



Figura 6. Implementación de bolsas plásticas 26 de mayo del 2022.



Figura 7. Implementación de tarimas 2 de junio del 2022.

5.1.1 Potencial hídrico

El potencial hídrico del tallo es una medida importante para evaluar la disponibilidad de agua en una planta y su capacidad para mantener la turgencia celular (Taiz et al., 2015). Se registró el potencial hídrico del tallo en dos variedades de café a lo largo de un periodo de 38 días, con cinco visitas iniciando el 06 de mayo y finalizando el 13 de junio del 2022. El periodo de déficit hídrico (PDH) fue experimentado únicamente por los tratamientos EHC y EH principalmente en las últimas dos visitas de campo.

En la Figura 8 se muestra el comportamiento del potencial hídrico a través del tiempo para ambas variedades de café. Los valores de potencial hídrico de las muestras se pueden consultar en el Cuadro A.1. De igual forma, un esquema con la posición de las muestras escogidas aleatoriamente y sus días de medición se presentan en la Figura A.1.

Las plantas de café del tratamiento A fueron sometidas a una irrigación por goteo de 1.7 L por día todos los días. Este tratamiento sirvió como referencia para comparar los efectos de los otros dos tratamientos. El potencial hídrico del tallo se mantuvo relativamente constante a lo largo del periodo de investigación, ya que las plantas recibieron una cantidad adecuada de agua para satisfacer sus necesidades. La variedad Turrialba (Figura 8.a) experimentó una variación del potencial hídrico de -3.17 a -1.00 bar, mientras que la variedad Victoria (Figura 8.b) entre -3.33 a -1.17 bar. Estos resultados concuerdan con el rango de un potencial hídrico óptimo (-3 a -1 bar) según Taiz et al. (2015).

Las plantas de café del tratamiento EHC se les aplicó un bioestimulante a base de extractos de algas (Seamaxx®) y un protector del estrés abiótico que contiene óxido de zinc, óxido de silicio e ingredientes orgánicos (Photon®) con el objetivo de reducir el daño causado por el estrés hídrico. El potencial hídrico del tallo en este tratamiento mostró una disminución gradual a lo largo del tiempo variando de -1.67 a -26.67 bar para la variedad Turrialba y de -1.75 a -15.75 bar para Victoria.

En el tratamiento EH, las plantas de café no recibieron irrigación, aunque fueron capaces de obtener agua de diferentes formas como se explicó en la sección 5.1. Estas plantas experimentaron un déficit hídrico significativo en los últimos días principalmente y, por lo tanto, muestran una disminución considerable en su potencial hídrico del tallo a medida que

pasa el tiempo. La variación es de -1.33 a -13.50 bar para la variedad Turrialba y de -1.50 a -30.13 bar para Victoria.

En general, se puede observar que los tratamientos de EHC y EH, muestran magnitudes más bajas de potencial hídrico en comparación con el control. Esto indica que las plantas sometidas a escasez de agua experimentaron un mayor estrés hídrico y, por lo tanto, una mayor probabilidad de sufrir daños relacionados con la falta de agua, lo cual se visualiza en la firma espectral.

En el caso de la variedad Turrialba (Figura 8.a) se evidencia una deshidratación lenta al día 7. Los valores registrados para los tratamientos A, EHC y EH este día son: -3.17, -3.50 y -2.67 bar respectivamente. El día 17 las plantas se rehidrataron y se reportaron valores de -1.33, -1.67 y -1.33 bar para los tratamientos A, EHC y EH, respectivamente. De este día en adelante el potencial hídrico de los tratamientos EHC y EH disminuyen rápidamente, mientras que el tratamiento A mantiene un rango de potencial hídrico constante. Para el día 38 el tratamiento EHC registra los valores más negativos (-26.67 bar), mientras que el tratamiento EH reporta un valor de -13.50 bar. La diferencia de potencial hídrico de estos tratamientos es de -13.17 bar.

La variedad Victoria (Figura 8.b) experimentó un comportamiento en su deshidratación similar y para el día 17 se reportan magnitudes de -1.17, -2.50 y -2.83 bar en los tratamientos A, EHC y EH, respectivamente. Para el día 38 el potencial hídrico del tratamiento EH fue de -30.13 bar, mientras que el tratamiento EHC obtuvo un valor de -15.75 bar. Este comportamiento es el esperado por la aplicación de productos para la reducción del estrés hídrico.

Para el día 31 el potencial hídrico de la variedad Turrialba experimentó un aumento relativo de 74% entre el tratamiento EH con respecto al ECH. Por otro lado, la variedad Victoria registró una disminución del 4%. Para el día 38 los valores cambiaron drásticamente y se obtuvo una disminución de 97.53% en la variedad Turrialba y un aumento de 47.72% en la variedad Victoria. En la investigación de Fernández et al. (2020) se obtuvieron disminuciones de potencial hídrico relativas de 11.02%, 18.46% y 17.85% para las variedades Costa Rica 95, Obatá y Catuaí Rojo.

En comparación, se observa que las disminuciones porcentuales en la presente investigación son más pronunciadas y presentan una mayor variabilidad en comparación con la investigación de Fernández et al. (2020). Esto puede indicar diferencias en los factores que afectan el potencial hídrico entre los dos estudios, como las condiciones ambientales y de infraestructura, la heterogénea humedad, las prácticas de manejo o las variedades de plantas utilizadas.

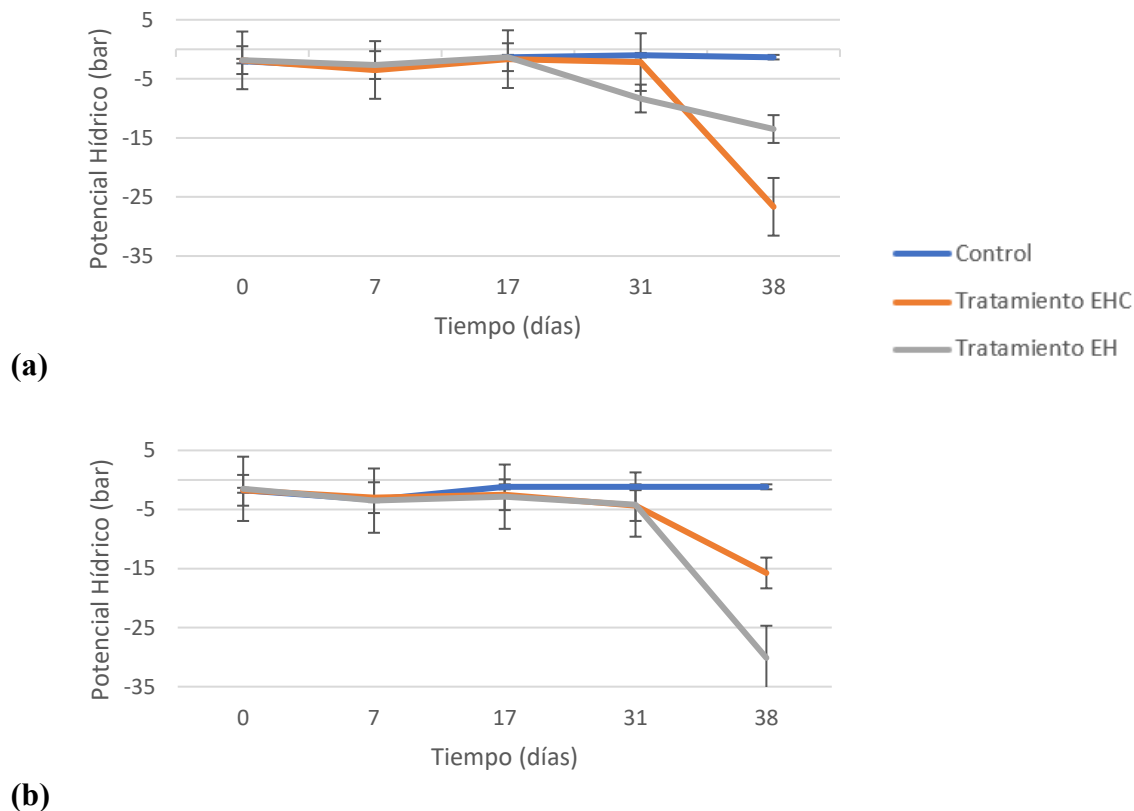


Figura 8. Potencial hídrico (bar) vs tiempo (días) para la variedad Turrialba (a) y Victoria (b).



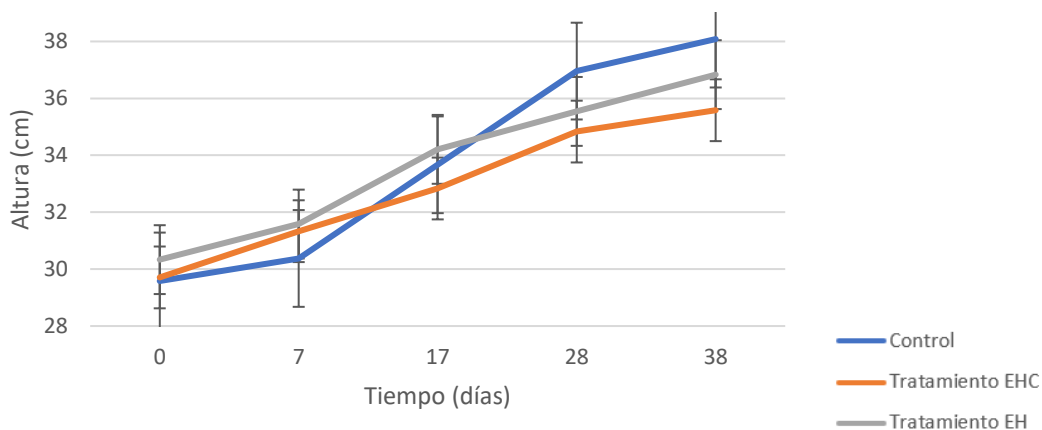
Figura 9. Muestra V33 con potencial hídrico menor de -40 bar el día 13 de junio del 2022.

5.1.1.2 Altura

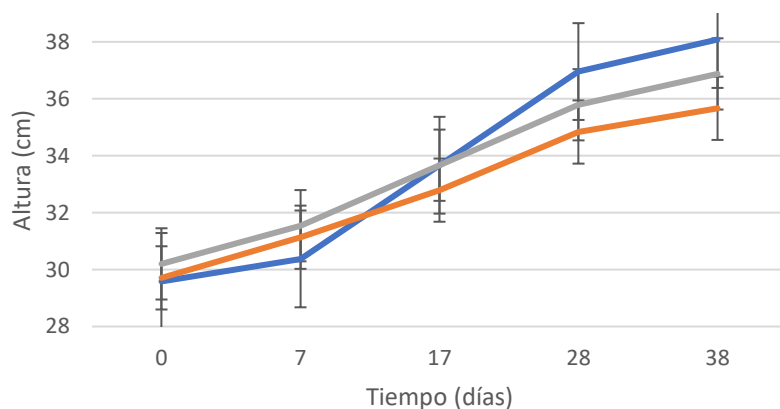
La altura de las plantas de café puede ser un indicador importante de su salud y desarrollo. Un incremento en la altura generalmente indica un crecimiento activo y vigoroso, mientras que un incremento de altura reducido puede ser indicativo de problemas de salud o desarrollo (Taiz et al., 2015). La medición de la altura de las plantas es una herramienta esencial para comparar diferentes variedades de plantas y evaluar su respuesta a diferentes condiciones ambientales. Proporciona información cuantitativa sobre el crecimiento y el desarrollo, así como indicios sobre posibles afectaciones o adaptaciones. Se registró la altura de todas las unidades experimentales (144 plantas) con una cinta métrica durante el periodo de 38 días iniciando el 06 de mayo y finalizando el 13 de junio del 2022. En la Figura 10 se presentan los datos de la altura promedio para ambas variedades de café a través del tiempo.

En la variedad Turrialba se obtuvieron cambios de altura del día 0 al 38 de 8.50, 5.88 y 6.50 cm para los tratamientos A, EHC y EH respectivamente. Además, es importante resaltar que el tratamiento EHC fue el que tuvo menor crecimiento. De igual forma, en la variedad Victoria se registraron valores de 8.50, 5.96 y 6.68 cm para los tratamientos A, EHC y EH. Las plantas de la variedad Victoria tuvieron un crecimiento más rápido aunque poco notable ya que las diferencias son muy pequeñas. Esto se relaciona a la morfología y genética de las diferentes variedades y al tiempo entre mediciones. La variedad Turrialba posee una

morfología más alargada con un sistema radicular débil, mientras que Victoria posee un tamaño menor y un sistema radicular más robusto y resistente.



(a)



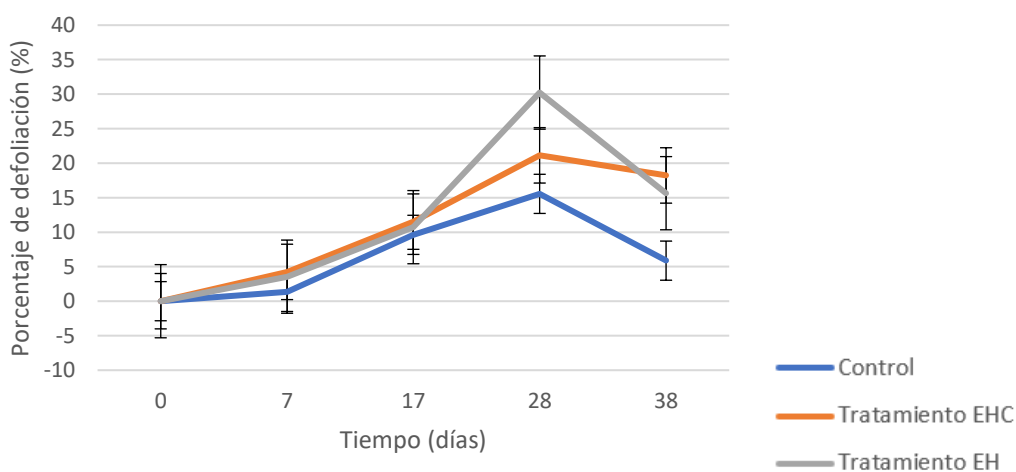
(b)

Figura 10. Altura promedio de las muestras (cm) vs tiempo (días) para la variedad Turrialba (a) y Victoria (a).

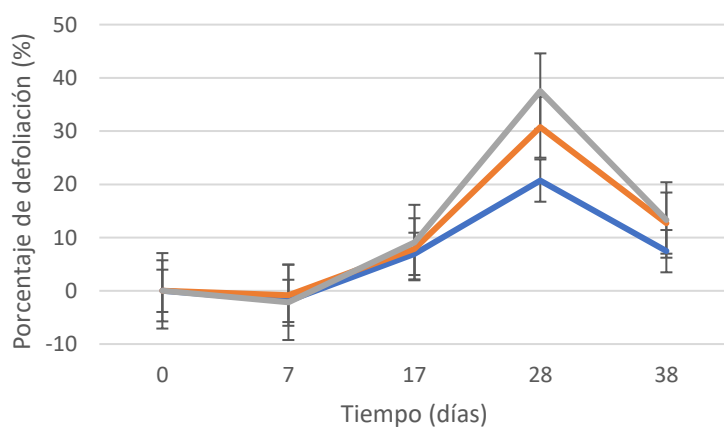
5.1.1.3 Cantidad de hojas

La cantidad de hojas de una planta puede proporcionar una indicación de su salud y el impacto de factores estresantes, enfermedades o plagas en su desarrollo. Un alto porcentaje de defoliación puede ser una señal de estrés ambiental, presencia de enfermedades o plagas, o una nutrición inadecuada (Taiz et al., 2015). Evaluar y monitorear el porcentaje de defoliación de las plantas es esencial para tomar medidas adecuadas y promover su salud y productividad. Se registró la cantidad de hojas de todas las unidades experimentales (144

plantas) durante el periodo de 38 días iniciando el 06 de mayo y finalizando el 13 de junio del 2022. En la Figura 11 se muestra el porcentaje de defoliación respecto al día 0 (06/05/24). Se encontró generalmente que la variedad Victoria presentó valores mayores los largo de los 38 días y el día 31 (03/06/24) se registró el valor más alto para el tratamiento EH y variedad Victoria con un 38% de porcentaje de defoliación. Para el día 38 el porcentaje de defoliación del tratamiento EHC para la variedad Turrialba es de 18%, mientras que para la variedad Victoria es de 13%. Por otro lado, el tratamiento EH obtuvo 16% y 13% para las variedades mencionadas anteriormente.



(a)



(b)

Figura 11. Porcentaje de defoliación vs tiempo (días) para la variedad Turrialba (a) y Victoria (b).

5.1.1.4 Tasa de transpiración

La tasa de transpiración se refiere a la cantidad de agua que una planta pierde por evaporación a través de las estomas de las hojas en un período de tiempo determinado. Es un proceso importante para la absorción de nutrientes, el enfriamiento y el mantenimiento de la temperatura adecuada para el crecimiento de la planta. La tasa de transpiración puede variar según las condiciones ambientales y es un indicador de la salud de la planta y su capacidad para tolerar condiciones adversas como la sequía o el estrés hídrico (Taiz et al., 2016).

Recordando el efecto de rehidratación que sufrieron las plantas por la filtración de la lluvia en el sistema, se decidió no realizar las mediciones de capacidad fotosintética el día 31 (03 de junio), ya que no se cuantificó hasta qué punto se deshidrataron las muestras en ese lapso. El nuevo periodo de medición está conformado por el día 0 (6 de mayo), 17 (23 de mayo) y 38 (13 de junio). En la Figura 12 se muestra la tasa de transpiración a través del tiempo para la variedad Turrialba (a) y Victoria (b).

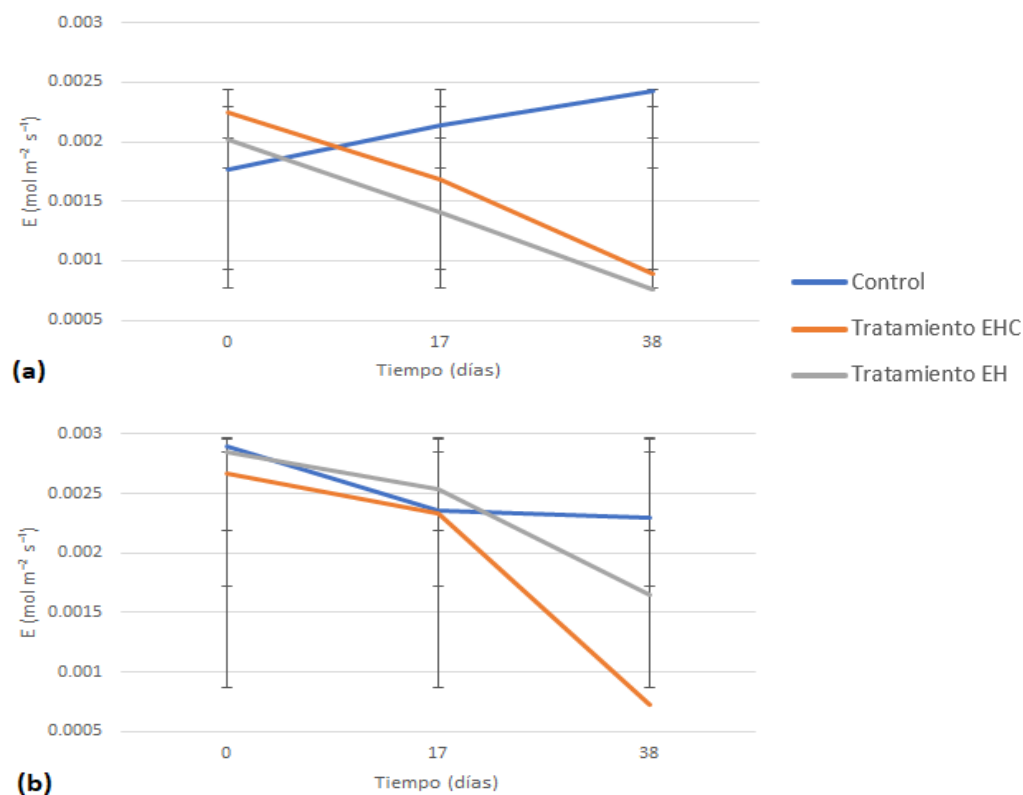


Figura 12. Tasa de transpiración ($\text{mol m}^{-2} \text{s}^{-1}$) vs tiempo (días) para la variedad Turrialba (a) y Victoria (b).

El comportamiento de las plantas de café del tratamiento A difiere en ambas variedades. Las plantas de la variedad Turrialba (Figura 12.a) mostraron un aumento en la tasa de transpiración del 27.11% desde el día 0 hasta el día 38, mientras que la variedad Victoria (figura 12.b) sufrió una disminución del 20.51% en el mismo periodo de tiempo. Los valores mínimos y máximos para la variedad Turrialba son 0.00177 y 0.00243 mol m⁻² s⁻¹ medidos el día 0 y 38. Para la variedad Victoria se registraron valores de 0.00289 mol m⁻² s⁻¹ en el día 0 y de 0.00230 en el día 38.

Para la variedad Turrialba los tratamientos EHC y EH presentan disminuciones porcentuales de 60.48% y 62.21% en el periodo de 38 días. En este caso, el comportamiento de ambos tratamientos al igual que en el potencial hídrico es el esperado obteniéndose un menor porcentaje de disminución en el tratamiento EHC. La aplicación de los productos Seamaxx® y Photon® reduce los efectos negativos del estrés abiótico en un 13.96%.

El comportamiento de ambos tratamientos es contrario para las muestras de la variedad Victoria. Se registraron disminuciones porcentuales de 72.92% y 42.12% en el periodo de 38 días para los tratamientos EHC y EH respectivamente. Para el día 38 se tiene una diferencia del 56.09% en la tasa de transpiración del tratamiento EH respecto a EHC. En este caso, el tratamiento EHC obtiene el valor mínimo de toda la investigación el día 38 con 0.00072 mol m⁻² s⁻¹. Este comportamiento es contrario al esperado a partir de los resultados obtenidos para el potencial hídrico (sección 5.1.1). Estas variables se encuentran estrechamente relacionadas y aun así se muestra una respuesta contraria para la tasa de transpiración en ambas variedades. La alta variabilidad generada por el tiempo de medición es una de las posibles causantes, aunque no se descarta el error humano y las limitaciones de infraestructura como parte de la problemática.

5.1.1.5 Tasa de asimilación neta de CO₂

La tasa de asimilación neta de CO₂ es la velocidad a la cual las plantas capturan y utilizan el CO₂ durante la fotosíntesis para producir carbohidratos. Es un proceso fundamental para el crecimiento, desarrollo y supervivencia de las plantas, ya que proporciona energía y construye biomasa (Taiz et al., 2015).

Los resultados obtenidos de la tasa de asimilación neta de CO₂ revelaron un comportamiento similar entre las diferentes variedades de café y los tratamientos aplicados con respecto a las variables potencial hídrico y tasa de transpiración. Esto indica que, independientemente de la variedad de café utilizada y los diferentes tratamientos implementados, las plantas presentaron patrones similares en cuanto a su capacidad para capturar dióxido de carbono y llevar a cabo la fotosíntesis.

El rango de tasa de asimilación registrado para el tratamiento A fue de 9.92 y 7.72 $\mu\text{mol m}^{-2} \text{s}^{-1}$ para la variedad Turrialba y 9.68 y 8.30 $\mu\text{mol m}^{-2} \text{s}^{-1}$ para la variedad Victoria. Seguidamente, el tratamiento EHC presentó valores de 0.00225 y 0.00089 $\mu\text{mol m}^{-2} \text{s}^{-1}$ para Turrialba y 0.00267 y 0.00072 $\mu\text{mol m}^{-2} \text{s}^{-1}$ para la variedad Victoria. Por último, el tratamiento EH obtuvo datos de 0.00202 y 0.00076 $\mu\text{mol m}^{-2} \text{s}^{-1}$ para la variedad Turrialba y 0.00285 y 0.0165 $\mu\text{mol m}^{-2} \text{s}^{-1}$ para Victoria.

El tratamiento A presentó un aumento sostenido hasta el 22.24% en el día 38 de medición para la variedad Turrialba (Figura 13.a) y para la variedad Victoria (Figura 13.b) se registró una disminución de 14.27% en los primeros 17 días, y posteriormente se registró un aumento del 12.91% para el día 38.

Para la variedad Turrialba los tratamientos EHC y EH registraron disminuciones de 32.04% y 36.73% respectivamente. Por otro lado, en la variedad Victoria las disminuciones fueron de 38.81% y 16.40%, respectivamente en el periodo de 38 días.

El efecto beneficioso de los insumos agrícolas aplicados (EHC) con respecto al tratamiento EH en la tasa de asimilación neta de CO₂ es de un 13.27% para la variedad Turrialba. Por el contrario, en la variedad Victoria se cuantificó una disminución del 25.58% en dicha variable con respecto a los tratamientos anteriormente mencionados. El estudio de Fernández. et al. (2020) midió la capacidad fotosintética de diferentes variedades de café robusta (*Coffea canephora*) que fueron irrigadas cada tres días. Se registraron valores de tasa de asimilación de 9.5 y 15 $\mu\text{mol m}^{-2} \text{s}^{-1}$. Estos valores son similares a los obtenidos en la investigación para el tratamiento A (control). Al igual que en la sección anterior, teóricamente existe una relación estrecha y lógica entre la tasa de asimilación neta de CO₂ y el potencial hídrico que es directamente proporcional, sin embargo, nuevamente se puede apreciar como

los efectos de los insumos en los tratamientos son contrarios a los obtenidos para el potencial hídrico (sección 5.1.1).

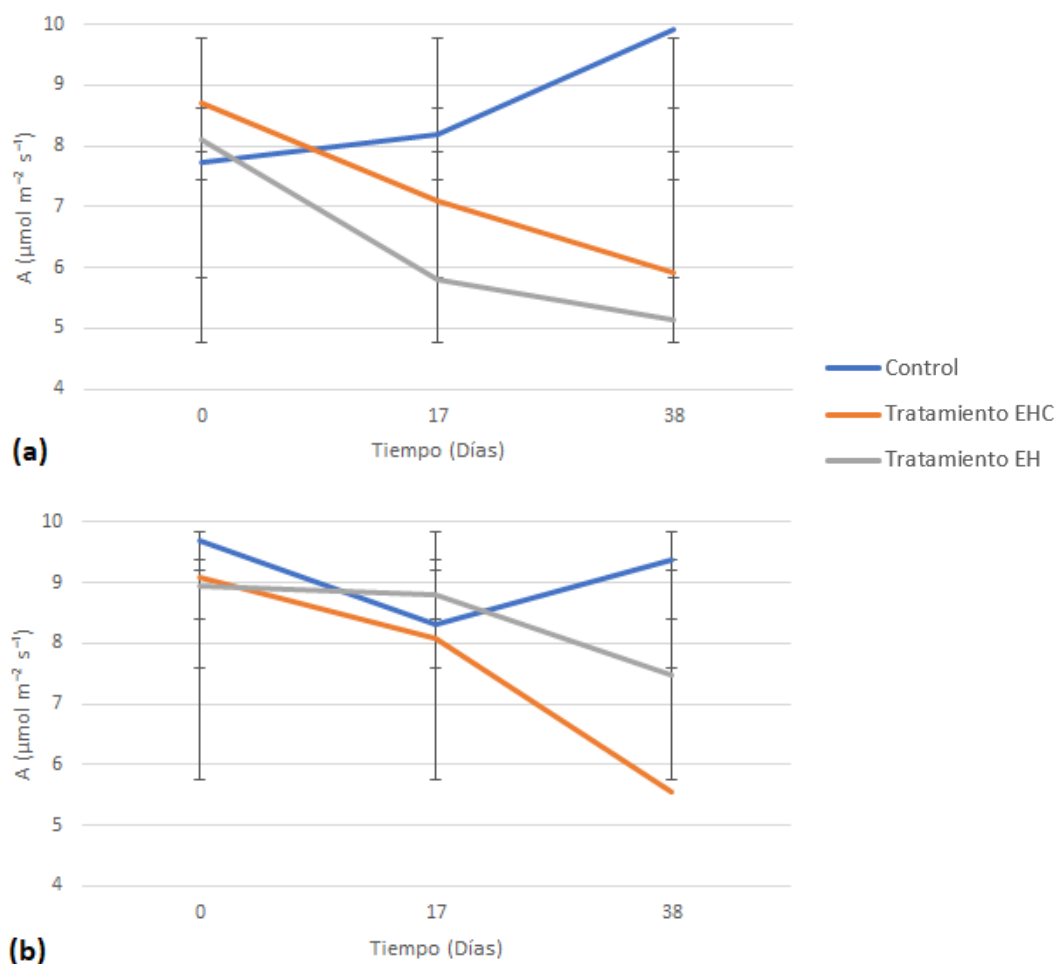


Figura 13. Tasa de asimilación neta de CO₂ ($\mu\text{mol m}^{-2} \text{s}^{-1}$) vs tiempo (días) para la variedad Turrialba (a) y Victoria (b).

5.1.1.6 Temperatura de la hoja

La temperatura de la hoja en el balance de energía de las plantas es el resultado de la interacción entre la ganancia de energía solar, la pérdida de energía por transpiración y radiación térmica. Su regulación es esencial para el funcionamiento eficiente de la planta y la optimización de la fotosíntesis y la transpiración (Alemán et al., 2019).

En la variedad Turrialba se registró una disminución de la temperatura de la hoja de 8.73%, 6.29% y 6.28% para los tratamientos A, EHC y EH respectivamente (Figura 14.a).

Para la variedad Victoria (Figura 14.b) los valores registrados fueron de 8.25%, 6.92% y 7.73%. En general, todos los tratamientos presentan un comportamiento similar independiente de la variedad. En el tratamiento A se nota una pequeña diferencia de temperatura al día 38 con relación a los otros tratamientos. La diferencia respecto al tratamiento EHC es de 40.45% para la variedad Turrialba y 40.73% para Victoria. Los tratamientos EHC y EC tienen valores muy similares a travez del tiempo. En el día 17 existen diferencias del 18.32% y 8.32% entre tratamientos para la variedad Turrialba y Victoria respectivamente. En general la hoja está casi medio grado más fresca en el tratamiento control.

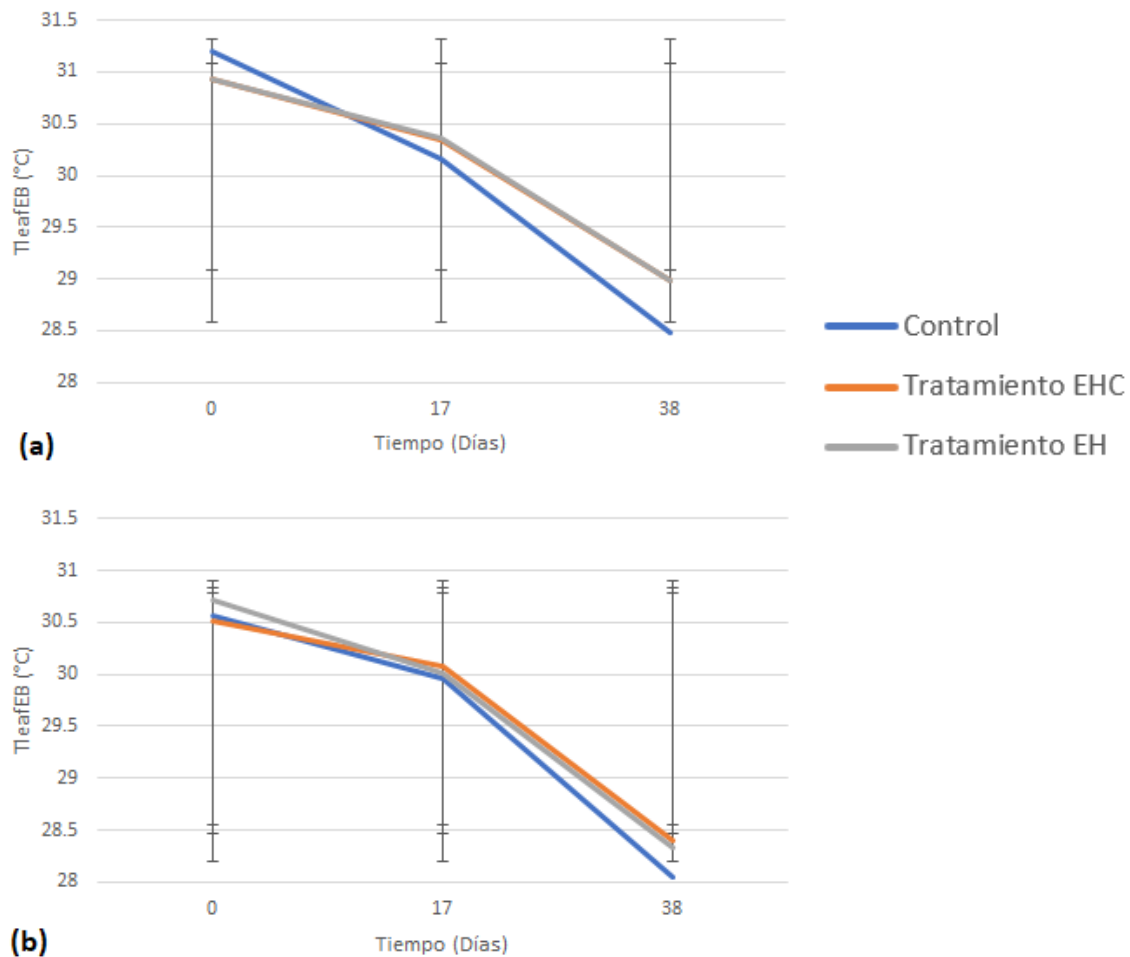


Figura 14. Temperatura de la hoja en el balance de energía (°C) vs tiempo (días) para la variedad Turrialba (a) y Victoria (b).

5.1.1.7 Presión de vapor de saturación a la temperatura de la hoja

La presión de vapor de saturación a la temperatura de la hoja indica la cantidad máxima de vapor de agua que el aire puede retener antes de que se produzca la condensación. Esta relación es relevante para comprender el equilibrio de vapor de agua y la transpiración en las plantas (Taiz et al., 2015).

El comportamiento de esta variable es similar al de la temperatura de la hoja. Se registran disminuciones del 13.86%, 10.05% y 9.93% para los tratamientos A, EHC y EH en la variedad Turrialba (Figura 15.a) al día 38. Por otro lado, las disminuciones para la variedad Victoria (Figura 15.b) son de 12.76%, 10.82% y 12.29%. El día 38 se registró una diferencia del 0.12% entre los tratamientos EH y EHC en la variedad Turrialba. Esta diferencia es de 0.61% para los tratamientos EHC y EH en la variedad Victoria.

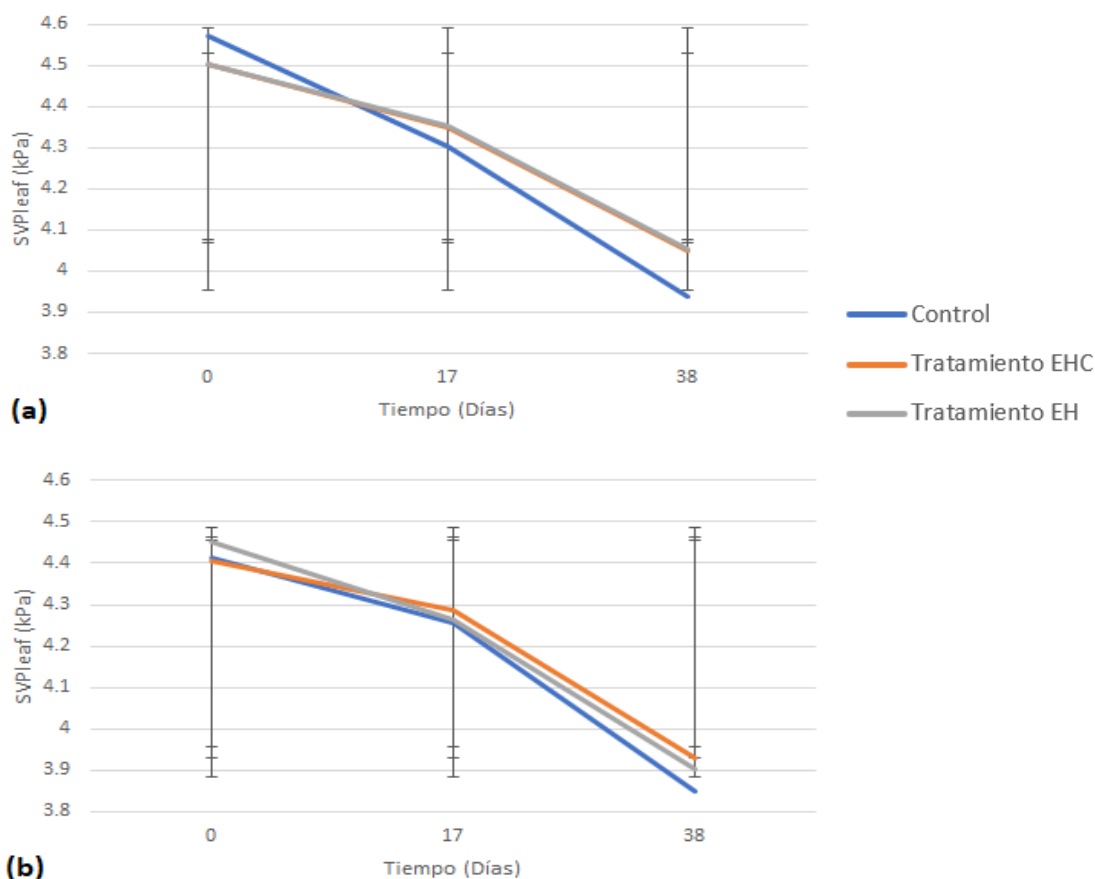


Figura 15. Presión de vapor de saturación a la temperatura de la hoja (kPa) vs tiempo (días) para la variedad Turrialba (a) y Victoria (b).

Las condiciones climáticas de la zona características de la época lluviosa afectaron directamente los resultados de la investigación. Las fechas establecidas para las visitas de medición coincidieron con el final de la época seca y el inicio de la época lluviosa. Por lo tanto, hubo una presencia muy marcada de fuertes lluvias y vientos pronunciados que influyeron en el diseño experimental. La faltante de techo del invernadero fue clave en la rehidratación de las muestras y otras variables como la topografía y la humedad del suelo afectaron en la deshidratación homogénea. Por otro lado, las condiciones iniciales del material genético afectaron al crecimiento óptimo de las plantas y, por ende, las variables fisiológicas reflejaron estas relaciones. Por último, la alta variabilidad detectada en las mediciones realizadas con el medidor de fotosíntesis portátil Li-6800 generó dificultades para interpretar los resultados y en algunos casos los comportamientos predichos no fueron los obtenidos. La variabilidad fue causada en mayor parte por la diferencia de tiempo entre las mediciones realizadas entre muestras los días de visita. Las mediciones iniciaron a las 6:00 am y terminaron a las 11:00 am en la mayoría de los casos. En un periodo de 5 horas las condiciones climáticas varían evidentemente y las variables que se estudiaron son muy sensibles a factores como temperatura y humedad relativa.

Como resultado, el comportamiento de las variables: potencial hídrico, tasa de transpiración, tasa de asimilación neta de CO₂, temperatura de la hoja, presión de vapor de saturación y cantidad de hojas presentan un cambio marcado a partir del día 17 del inicio teórico del estrés hídrico. En el caso de la altura de las plantas no se evidencia un aumento o disminución respecto al tiempo, únicamente se evidencia la diferencia entre los tratamientos planteados. Siendo el tratamiento EHC el que presentó un menor crecimiento en ambas variedades.

Las variables que determinan la capacidad fotosintética, es decir la tasa de transpiración, tasa de asimilación neta de CO₂, temperatura de la hoja y la presión de vapor de saturación presentan un comportamiento contrario al del potencial hídrico. En la variedad Turrialba los tratamientos EHC y EC muestran una diferencia significativa respecto a los de la variedad Victoria. En el caso de Turrialba la aplicación de los productos Seamaxx ® y Photon ® para reducir el daño causado por estrés abiótico del tratamiento EHC disminuyeron las medidas de potencial hídrico en un 49.38% para el día 38, respecto al tratamiento EH.

Por contrario, en la variedad Victoria el potencial hídrico aumentó en un 47.72% en el tratamiento EHC, respecto al EH.

De manera similar, los resultados de la tasa de transpiración para la variedad Turrialba mostraron en el día 38, una disminución de 16.23% del tratamiento EH respecto a EHC. Por otro lado, la variedad Victoria tuvo un aumento del 56.09% entre los tratamientos anteriormente mencionados. La tasa de asimilación neta de CO₂ tuvo una diferencia de 15.30% para la variedad Turrialba y un aumento del 25.58% para la variedad Victoria. En el caso de la temperatura de la hoja y la presión de vapor de saturación las diferencias obtenidas son muy bajas. Para la variedad Turrialba se obtuvo un aumento del 0.015% y una disminución del 0.22% en la temperatura la de hoja. Para la variedad Turrialba se obtuvo un aumento de 0.12% y una disminución de 0.61% en la presión de vapor de saturación.

5.2 Análisis de varianza (ANOVA)

Se realizó un análisis de varianza con un nivel de confianza de 95% y un nivel de $p < 0.01$, con el objetivo de encontrar diferencias significativas en las medias de los diferentes conjuntos de datos. La importancia de este paso radica en la toma de decisiones para el posterior tratamiento tanto de los datos auxiliares como de los datos espectrales para la conformación del modelo. Para las pruebas de ANOVA que se muestran a continuación se utilizó el mismo conjunto de datos que en la sección anterior (5.1).

En el Cuadro 7 se muestra un resumen de los resultados obtenidos para las variables dependientes: potencial hídrico (PH), tasa de transpiración (E), tasa de asimilación neta de CO₂ (A), temperatura de la hoja (Tleaf), presión de vapor de saturación a la temperatura de la hoja (SVPlleaf), altura (h) y cantidad de hojas (n). Los grupos más importantes son: fecha (X₁), tratamiento (X₂) y variedad (X₃). Las interacciones entre grupos no fueron tomadas en cuenta en este análisis ya que todas estas variables son muy dependientes. En la sección respectiva del apéndice se muestra el cuadro completo del análisis ANOVA para cada una de las variables mencionadas anteriormente.

El análisis de varianza (ANOVA) reveló diferencias estadísticamente significativas en todas las variables y el grupo fecha (X₁). No se observaron diferencias significativas en la variable cantidad de hojas (n) y el grupo tratamiento (X₂) y variedad (X₃). La altura tampoco

presentó diferencias significativas respecto al grupo tratamiento (X_2). De igual forma, la temperatura de la hoja (T_{leaf}) y la presión de vapor de saturación a la temperatura de la hoja (SV_{Pleaf}) no presentaron diferencias significativas con los grupos tratamiento (X_2) y variedad (X_3).

Para la variable potencial hídrico (PH), se encontró diferencias estadísticamente significativas entre los grupos fecha (X_1) ($F = 36.59$, $p < 0.001$), tratamiento (X_2) ($F = 11.34$, $p = 0.001$). Por el contrario, el grupo X_3 ($F = 0.37$, $p = 0.549$) no presentó diferencias significativas. Estos resultados sugieren que los grupos difieren en términos de la variable PH, lo que indica la influencia de algún factor subyacente en esta diferencia.

En el caso del tiempo, es natural pensar que conforme avanza el déficit hídrico en los tratamientos EHC y EH las plantas de café se van a deshidratar. Por lo tanto, la relación entre los grupos: tiempo y tratamiento respecto al potencial hídrico es bastante evidente. Finalmente, los resultados de la prueba de Tukey muestran que el control difiere de manera significativa para el PH en comparación con el tratamiento EHC y EH. Este resultado concuerda con los datos recolectados en los 38 días, ya que el control mantuvo valores de potencial hídrico de -3.33 y -1.00 bar en ambas variedades de café.

Los resultados del ANOVA indican diferencias significativas en la tasa de transpiración (E) y la tasa de asimilación (A) entre los grupos X_1 ($F = 15.49$, $p < 0.001$ y $F = 5.77$, $p = 0.0039$), X_2 ($F = 5.69$, $p = 0.0042$ y $F = 7.47$, $p = 0.0008$) y X_3 ($F = 15.17$, $p = 0.0001$ y $F = 7.14$, $p = 0.0084$), respectivamente. Estos hallazgos respaldan la hipótesis de que estas variables son influenciadas por factores como el tiempo de medición, los diferentes tratamientos aplicados y las distintas variedades de plantas de café.

Los resultados de las pruebas de Tukey muestran diferencias significativas entre los grupos. Específicamente, para la variable E se encontró que el día 38 difiere significativamente de los días 0 y 17, lo cual indica que hay una disparidad clara entre estos dos grupos en relación con la variable en estudio. Por otro lado, para la variable A se observó que el día 17 presenta una similitud estadísticamente significativa con el día 0 y 38. Por el contrario, los días 0 y 38 no presentan una similitud en sus medias.

Se encontraron diferencias significativas en las interacciones de ambas variables con el grupo X₂. La variable E del control presenta disminuciones de 51.60% y 47.69% respecto a los tratamientos EHC y EH en la variedad Turrialba. Para la variedad Victoria las diferencias son de 69.56% y 50.53%. De manera similar, en la variable A se presentaron aumentos de 99.94% al comparar el tratamiento C con EHC y EH en la variedad Turrialba. Para la variedad Victoria los aumentos fueron de 99.86% y 99.91% respecto a los tratamientos mencionados anteriormente.

En conclusión, los resultados del ANOVA han revelado una estrecha relación entre los valores de la tasa de transpiración y los de la tasa de asimilación neta de CO₂ en las plantas investigadas. La tasa de transpiración, la tasa de asimilación y el potencial hídrico están relacionados con el crecimiento de las plantas. La transpiración ayuda a absorber agua y nutrientes, mientras que la asimilación de dióxido de carbono es necesaria para producir alimentos. El potencial hídrico de la planta es importante porque afecta la transpiración y la absorción de nutrientes (Taiz et al., 2015). Además, ya que existen diferencias significativas entre los tratamientos y las variedades, se va a tomar en cuenta a la hora de la división y pretratamiento de los datos espectrales.

Cuadro 7. Resumen de ANOVA para las variables estudiadas.

<i>Descripción</i>	<i>Suma de Cuadrados</i>	<i>Grados de Libertad</i>	<i>Cuadrado Medio</i>	<i>Valor de F</i>	<i>Valor de P</i>
<i>PH</i>	Potencial hídrico				
<i>Fecha, X₁</i>	2050.08	3	683.36	36.59	<0.001*
<i>Tratamiento, X₂</i>	423.50	2	211.75	11.34	0.001*
<i>Variedad, X₃</i>	6.82	1	6.82	0.37	0.549
<i>E</i>	Tasa de transpiración				
<i>Fecha, X₁</i>	2.5E-05	2	1.3E-05	15.49	<0.0001*
<i>Tratamiento, X₂</i>	9.2E-06	2	4.6E-06	5.69	0.0042*
<i>Variedad, X₃</i>	1.2E-05	1	1.2E-05	15.17	0.0001*
<i>A</i>	Tasa de asimilación neta de CO₂				
<i>Fecha, X₁</i>	61.06	2	30.53	5.77	0.0039*
<i>Tratamiento, X₂</i>	79.01	2	39.50	7.47	0.0008*
<i>Variedad, X₃</i>	37.77	1	37.77	7.14	0.0084*
<i>Tleaf</i>	Temperatura de la hoja				

<i>Fecha, X₁</i>	147.16	2	73.58	21.42	<0.0001*
<i>Tratamiento, X₂</i>	0.78	2	0.39	0.11	0.8927
<i>Variedad, X₃</i>	6.99	1	6.99	2.03	0.1559
SVPleaf	Presión de vapor de saturación a la temperatura de la hoja				
<i>Fecha, X₁</i>	7.57	2	3.78	20.93	<0.0001*
<i>Tratamiento, X₂</i>	0.04	2	0.02	0.10	0.9037
<i>Variedad, X₃</i>	0.37	1	0.37	2.04	0.1552
h	Altura				
<i>Fecha, X₁</i>	4101.15	4	1025.29	51.40	<0.001*
<i>Tratamiento, X₂</i>	74.06	2	37.03	1.86	0.1570
<i>Variedad, X₃</i>	26199.67	1	26199.67	1313.46	<0.001*
n	Cantidad de hojas				
<i>Fecha, X₁</i>	127567.38	4	31891.84	238.40	<0.001*
<i>Tratamiento, X₂</i>	905.41	2	452.70	3.38	0.0345
<i>Variedad, X₃</i>	170.87	1	170.87	1.28	0.2588
* diferencias significativas encontradas.					

5.3 Análisis de respuestas espectrales

Las firmas espectrales de las plantas de café fueron obtenidas con el espectroradiómetro ASD FieldSpec 4 y el aditamento de la sonda de contacto “Plant contact probe” para cada hoja. Una repetición constó de tres mediciones por estrato (tres hojas) y la firma fue tomada en los estratos alto, medio e inferior. En total se generaron 216 firmas espectrales por estrato para la variedad Turrialba y 216 firmas para Victoria. En este análisis se trabaja solamente con los estratos alto y medio, los cuales representan mejor la planta.

5.3.1 Preprocesamiento de información espectral

Los pretratamientos en el análisis de datos espectrales desempeñan un papel fundamental para mejorar la calidad y la interpretabilidad de los resultados. Los datos espectrales sufren transformaciones en su comportamientos a la hora de aplicar diferentes pretratamientos.

El primero que se aplica es la remoción de saltos (“Gap remove”) en la firma espectral por cambios en los sensores de detección internos del equipo ASD Fieldspec (1000 nm y 1800 nm). Este es el primer paso antes de iniciar con los consecuentes análisis y generación

de modelos. Esta remoción se realizó con la aplicación de la función del paquete rchemo denominada “rmgap” en R Studio.

Posteriormente se realizó un promedio de 3 firmas espectrales para los estratos alto y medio. El estrato bajo se excluyó ya que las hojas presentaban mucha variabilidad en sus dimensiones y las mediciones de las variables fisiológicas fueron tomadas esencialmente en el estrato alto de las plantas de café.

Los datos de trasflectancia (T) se transformaron a $\log(1/T)$. En la Figura 16 se muestran los datos de las firmas espectrales promediadas para el estrato alto y medio donde en el eje y, se representan los datos de $\log(1/T)$ y en el eje x, las longitud de onda del sensor.

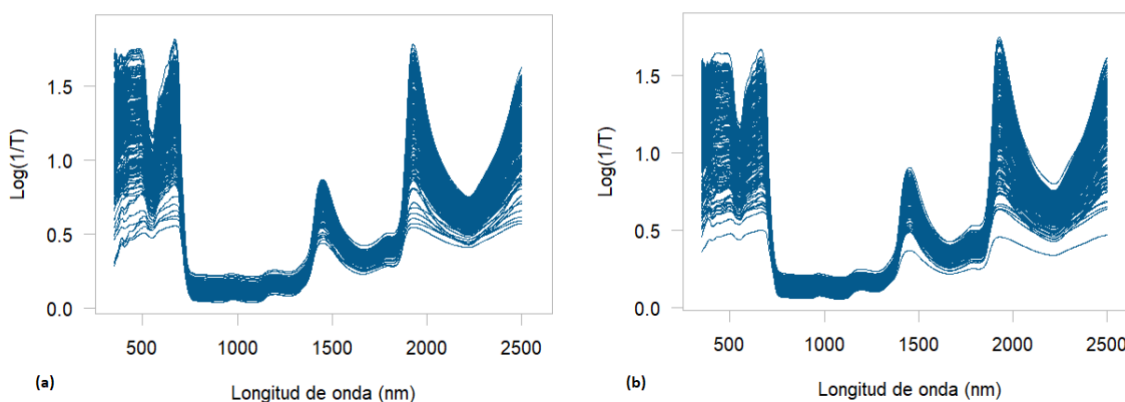


Figura 16. Datos promediados del estrato alto (a) y medio (b) luego de aplicar la función rmgap.

Se puede notar en la región de los 350 a los 700 nm aproximadamente existe una variación significativa de la absorbancia en las mediciones en ambos estratos, esto debido a la detección de cambios en la actividad fotosintética manifestado por los pigmentos de la clorofila a y b (Peña, 2019).

Con el fin de evaluar modelos con diferentes pretratamientos se procedió a aplicar métodos aditivos como lo son el SNV y el detrend (Figura 17 a y b) a las firmas espectrales de $\log(1/T)$. El SNV busca normalizar los datos mientras que detrend busca eliminar tendencias lineales y polinómicas.

El metodo multiplicativo aplicado fue Savitzky-Golay donde se utilizó la primera derivada (Figura 17.c). Además se probó una combinación de Savitzky-Golay + SNV

(Figura 17.d). El método de Savitzky-Golay es un método de suavizado de señales mediante un ajuste de polinomios en pequeñas secciones de la señal donde también se aplica la primera o segunda derivada. El comportamiento de las firmas espectrales utilizando los datos del estrato medio presentaron comportamientos similares a los del estrato alto de la Figura 16 con sus respectivos pretratamientos. Las figuras relacionadas a los datos del estrato medio se presentan en la Figura B.2 del Apéndice en la sección respectiva.

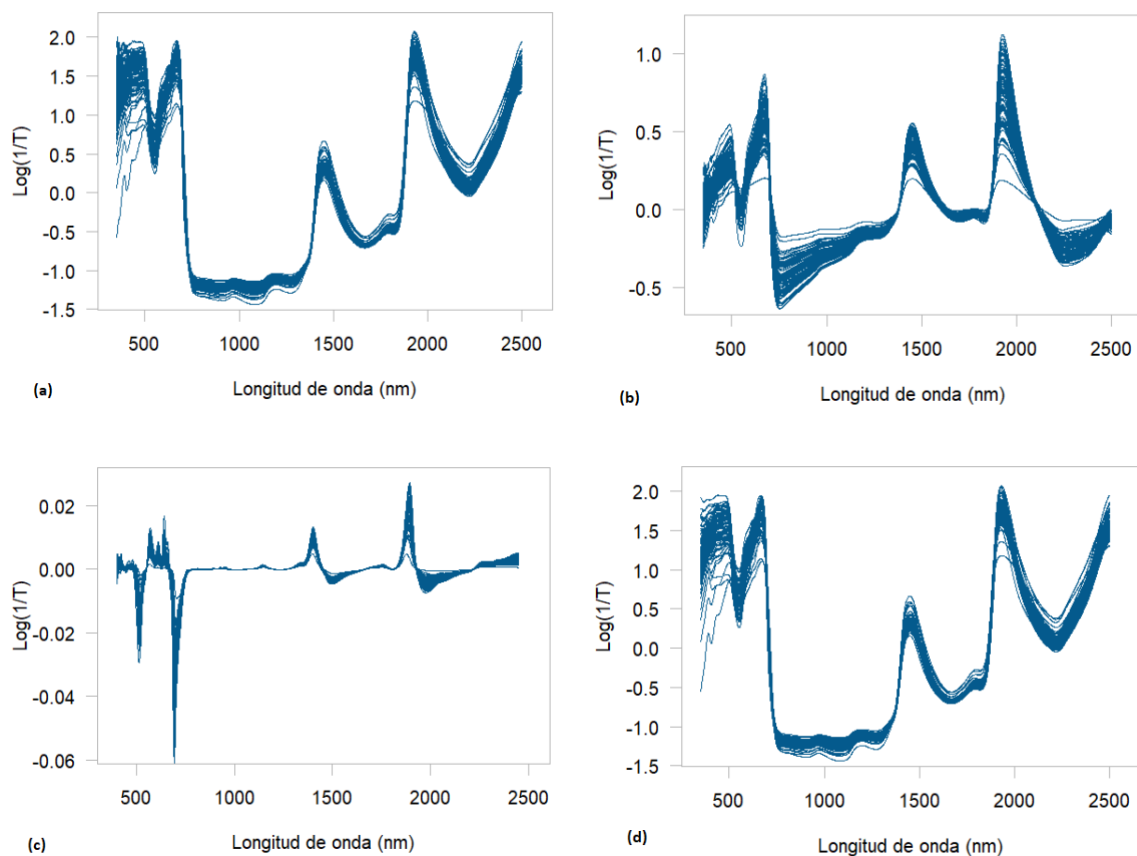


Figura 17. Modificaciones en el comportamiento de los datos espectrales luego de aplicar diferentes pretratamientos al conjunto de datos del estrato alto. SNV (a), Detrend (b), Savitzky-Golay 1era derivada (c) y Savitzky-Golay + SNV (d).

5.3.2 Generación del modelo predictivo

Es importante recordar que se creó un modelo para cada variable por variedad de planta con datos del estrato alto y medio. Es decir, la variable tasa de transpiración (E) tiene cuatro modelos: dos para los diferentes estratos de la variedad Turrialba y dos para Victoria. La metodología fue la misma para las cuatro variables relacionadas a la capacidad

fotosintética de las plantas, aunque únicamente se presentarán los gráficos para la tasa de transpiración (E), esto a causa de la similitud entre los resultados de los PCA y las PLSR para cada variable.

Los resultados de los modelos para las variables tasa de asimilación neta de CO₂ (A), temperatura de la hoja (Tleaf) y presión de vapor de saturación a la temperatura de la hoja (SVPlleaf) se presentaran en las Figuras B.2 hasta la B.10 y los Cuadros B.1 hasta B.11. Adicional, se puede apreciar la tendencia de los estadísticos RMSEP y R² en la sección de 5.3.2.2 que se presenta mas adelante.

5.3.2.1 Análisis de componentes principales (PCA)

Se realizó un análisis de componentes principales (PCA) que relaciona los datos de las firmas espectrales y parámetros fisiológicos de la planta para explorar la distribución espacial de los conjuntos de validación y calibración. Con esta herramienta se pueden identificar “clústeres” o agrupamientos de datos y se puede comprobar que los datos de validación estén contenidos en los datos de calibración.

Los componentes principales representan las combinaciones lineales de las variables originales que capturan la mayor parte de la variabilidad en los datos. En el Cuadro 8 se puede apreciar que se necesitan tres componentes principales para representar una varianza mayor al 85% de los datos. En la Figura 18, se logran apreciar de manera tridimensional tanto los datos de calibración como los de validación.

Cuadro 8. Porcentajes de varianza explicados para cada componente principal de cada modelo generado para la tasa de transpiración (E).

MODELO	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
1	68.90	15.07	7.47	1.98	1.95	1.16	0.78	0.58	0.53	0.43
2	67.74	14.26	10.39	2.06	1.58	1.03	0.67	0.46	0.39	0.34
3	52.44	24.95	11.48	5.07	1.38	1.13	0.82	0.60	0.41	0.33
4	53.03	23.99	11.53	5.33	1.47	1.05	0.86	0.62	0.41	0.35

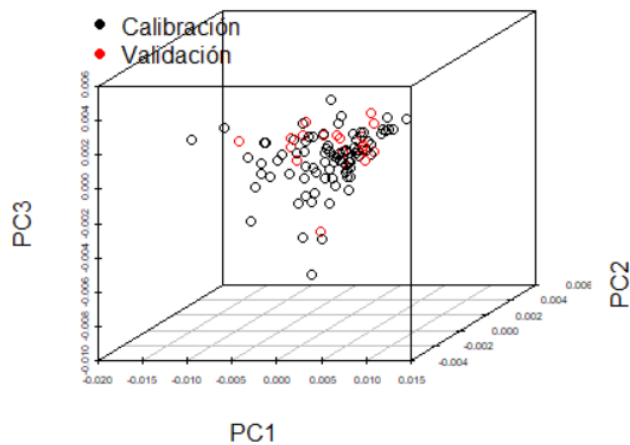
NOTA. La abreviatura PC se refiere a componente principal y Modelo 1, 2, 3 y 4 representan los modelos definidos anteriormente para la variable (E).

Los resultados de los PCA y su representación tridimensional de las variables A, SVPleaf y Tleaf se encuentran en las Figuras B.2, B.3 y B.4 y en los Cuadros B.1, B.2 y B.3, respectivamente. Al igual que para la variable E, fue necesario tomar tres componentes principales para representar el 85% de los datos de las variables A, Tleaf y SVPleaf. De forma similar, en el estudio de Zhou (2021), un porcentaje del 97.55% de los datos espectrales sin tratamientos fue representado con tres componentes principales, lo cual reveló una baja e inesperada diversidad de propiedades espectrales.

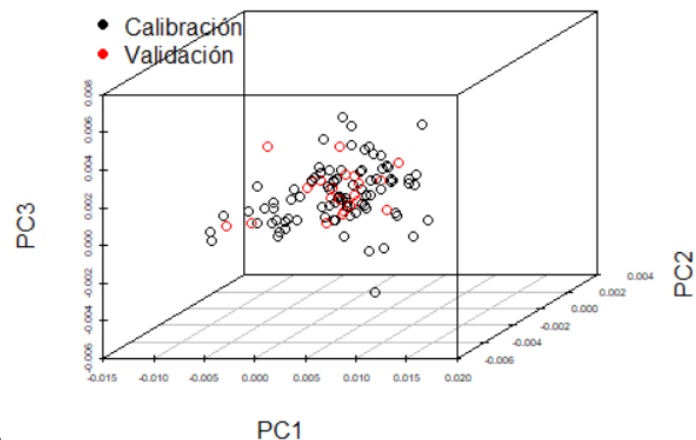
Los resultados obtenidos del PCA para la variable E se muestran en la Figura 18. En la Figura 18.a y 18.b se puede apreciar un agrupamiento de los datos en el espacio definido por los componentes principales. Esta acumulación indica que los datos espectrales y los parámetros fisiológicos están relacionados y se pueden usar para identificar y clasificar diferentes estados en las plantas de café. En el caso de la Figura 18.c y 18.d se logra diferenciar una separación clara de los datos en el espacio. Es posible que la interacción de las variables es más compleja o que otros factores están influyendo en la variabilidad observada.

Por último, se analizaron las longitudes de onda más importantes para cada modelo. En la Figura 19 se muestran las regiones espectrales más importantes de los cuatro modelos generados. En los cuatro modelos se encuentran picos importantes en la región que abarca desde los 350 – 500 nm. De igual forma, en la región cercana a los 750 nm los cuatro modelos presentan un comportamiento similar. Los picos que abarcan la región de los 1500 – 2500 nm también están presentes en todos los modelos.

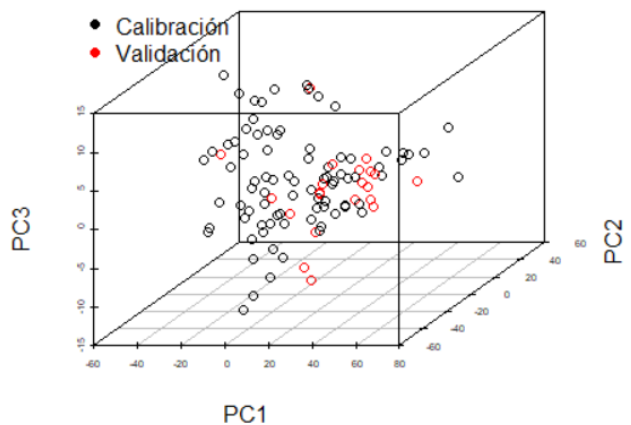
La región de onda corta del infrarrojo (SWIR, por sus siglas en inglés) abarca desde los 1000 nm a los 2500 nm. En esta región se encuentran las interacciones de los enlaces O-H de absorción y por lo tanto las interacciones relacionadas con agua.



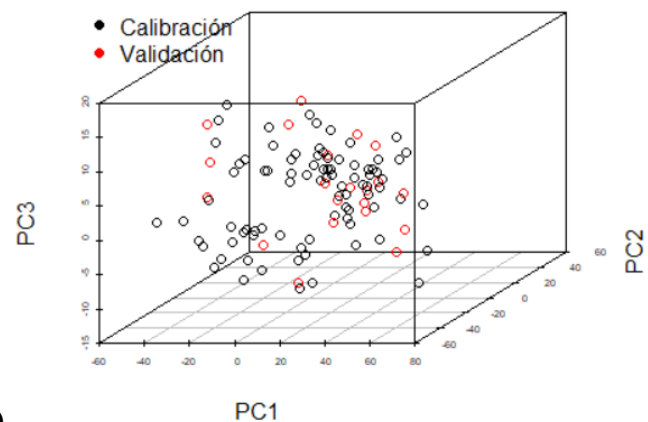
(a)



(b)



(c)



(d)

Figura 18. Resultados del análisis de componentes principales (PCA) para la tasa de transpiración (E) representados con PC1, PC2 y PC3. Turrialba estrato alto (a), Victoria estrato alto (b), Turrialba estrato medio (c) y Victoria estrato medio (d).

En el estudio de Villesseche (2022) se encontró que las bandas de 1450 nm y 1950 muestran alta correlación con la absorción de agua. También, la región de los 750–1000 nm en donde no hay absorción de agua, mantiene los mismos valores a lo largo de la senescencia de las hojas. Por último, la longitud de onda de los 680 nm es donde se pueden notar más claramente los efectos en el cambio de la hoja a causa de la degradación de clorofila que aumenta la absorción de la luz hasta que los pigmentos mueren. Este cambio en la capacidad de absorción de luz puede influir en el proceso de fotosíntesis y, por ende, en la regulación de la apertura estomática, lo que podría afectar la tasa de transpiración de la planta.

En la investigación de Zhou (2021) se encontraron 5 regiones con correlaciones altas similares a las encontradas en este estudio. La primera de los 400 – 480 nm, la segunda de los 500 – 660 nm, la tercera de los 720 – 1400 nm, la cuarta de los 1450 – 1800 nm y la quinta de los 1900 – 2500 nm. Estas características espectrales presentan un reto para la creación de modelos predictivos robustos ya que los valores de R^2 y RMSEP son menor, por lo que la capacidad predictiva de los mismos es menor.

La investigación de Zhou (2021) presenta similitudes con la mayoría de los modelos generados en esta investigación por las longitudes de onda en donde se encontraban esas correlaciones. De igual forma, en el estudio llevado por Villesseche (2022) se muestran similitudes en las cantidades de correlaciones encontradas para los modelos de las cuatro variables estudiadas en esta investigación. A continuación, se discuten los resultados obtenidos observando las gráficas del coeficiente de regresión vs longitudes de onda (nm) para los 12 modelos restantes.

En los cuatro modelos para la variable E presentados en la Figura 19 se pueden apreciar diferencias en la cantidad de regiones con altas correlaciones. En el caso del modelo 1 (Figura 18.a) se muestran 4 regiones: una correlación en la región de los 250 - 400 nm, una correlación en la región de los 600 – 700 nm y dos correlaciones en la región SWIR. Una en el rango de los 1400 – 1500 nm y otra entre los 1800 – 1900 nm.

El modelo 2 (Figura 19.b) al igual que modelo 3 presentan la mayor cantidad de correlaciones para la variable E. Tres ubicadas en la región VIS (400 – 700 nm) y tres en la región SWIR (1300 – 2500 nm) para un total de 6.

La Figura 19.c muestra las regiones con altas correlaciones para el modelo 3. Similar a los resultados del modelo 2 se presentan 6 correlaciones: tres en la región VIS y tres en la región SWIR. La única diferencia sería que para el modelo 2 las correlaciones son más fuertes y esto se puede apreciar en el comportamiento de los picos en las regiones mencionadas. Por último, la Figura 19.d muestra los resultados del modelo 4 en donde se identifican 5 regiones: tres en la región VIS e inferior y dos para la región SWIR. Los resultados de los modelos relacionados a las variables A, SVPleaf y Tleaf se presentan en la sección las Figuras B.8, B.9 y B.10.

Los resultados de los modelos para la variable A muestran las siguientes correlaciones. El modelo 5 (Figura B.8.a) presenta cuatro correlaciones: dos en la región VIS y dos en la región SWIR. El modelo 6 (Figura B.8.b) presenta cuatro: dos en la región VIS y dos muy bajas en la región SWIR. El modelo 7 (Figura B.8.c) contiene la mayor cantidad con un total de 6: tres en la región VIS y tres en la región SWIR. Por último, el modelo 8 (Figura B.8.d) presenta cinco: tres en la región VIS y dos en la región SWIR. La Figura B.9 muestra los resultados para la variable SVPleaf. En este caso, se presentan cinco correlaciones para los modelos 9(a), 10(b), 11(c) y 12(d): tres en la región VIS y dos en la región SWIR. De igual forma los resultados para los modelos generados de la variable Tleaf se presentan en la Figura B.10. Al igual que la variable anterior se presentan 5 correlaciones: tres en la región VIS y dos en SWIR. La única diferencia sería para el modelo 15 (Figura B.10.c) en donde las correlaciones son muchas y bajas: alrededor de tres en la región VIS y 4 en SWIR.

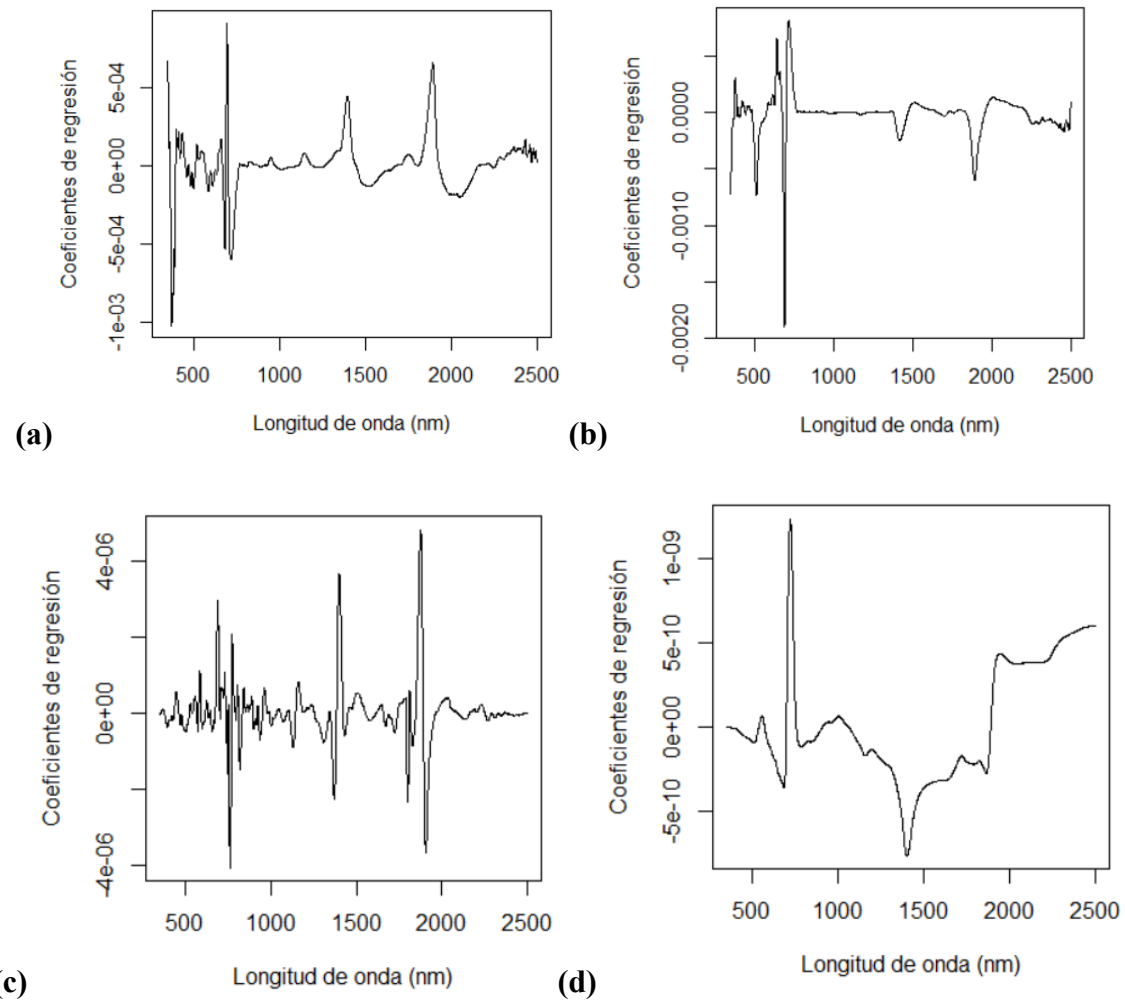


Figura 19. Coeficientes de regresión vs longitudes de onda (nm) más importantes de los modelos 1(a), 2(b), 3(c) y 4(d).

5.3.2.2 Regresión de mínimos cuadrados parciales (PLSR)

A continuación, se presentan y discuten los resultados obtenidos a través del análisis de Regresión por Mínimos Cuadrados Parciales (PLSR, por sus siglas en inglés) aplicado al conjunto de datos de estudio. La PLSR es una técnica poderosa que permite explorar y modelar relaciones complejas entre múltiples variables predictoras y una variable respuesta.

En el Cuadro 9 se muestran los resultados de los indicadores relacionados a los cuatro modelos generados para cada una de las variables de capacidad fotosintética. En este cuadro se presentan únicamente 5 pretratamientos: sin pretratamiento, SNV, Detrend, Savitzky-Golay y Savitzky-Golay + SNV. Esto a causa de la similitud de los resultados al utilizar diferentes combinaciones de pretratamientos. Por lo tanto, los resultados para todos los pretratamientos mencionados en el Cuadro 5 se muestran en los Cuadros B.4 hasta B.11.

Se utilizó el R^2 ajustado para determinar cuál es el mejor pretratamiento o la combinación que posee el mejor rendimiento. Se muestra la cantidad de variables latentes utilizadas para obtener el mejor resultado para cada modelo. En algunos casos los valores de RMSEP y R^2 ajustado presentan las mismas magnitudes, esto es causado por la alta variabilidad de los datos. Como consecuencia, la utilización de diferentes pretratamientos no mejora el rendimiento de los modelos, esto se puede ver claramente en los modelos que utilizan 0 variables latentes.

En el caso de los modelos con número de variables latentes mayor a 2 se puede apreciar como los rendimientos aumentan y los diferentes pretratamientos muestran diferentes magnitudes.

La modelación optimiza la cantidad de variables latentes que genera el mejor resultado, por lo tanto, si se empieza a reducir la cantidad de variables latentes el rendimiento se verá afectado. De igual forma, si se aumenta el número de variables latentes la diferencia entre los datos de calibración y validación aumentará y la capacidad de predicción del modelo se reducirá.

Cuadro 9. Resultados obtenidos de los indicadores para los modelos generados.

<i>Variable</i>	<i>Modelo</i>	<i>Pretratamiento</i>	<i>Variables Latentes</i>	<i>RMSEP</i>	<i>R²*</i>
<i>Tasa de transpiración(E)</i>	1.Turrialba	Sin	1	0.0075	-0.03072
	Estrato alto				
		SNV	0	0.00076	-0.04666
		Detrend	1	0.00075	-0.03559
		Savitzky-Golay	1	0.00075	-0.03070
		Savitzky-Golay SNV	0	0.00076	-0.04666
	2. Victoria	Sin	0	0.00105	-0.03215
	Estrato alto				
		SNV	0	0.00105	-0.03215
		Detrend	0	0.00105	-0.03215
		Savitzky-Golay	0	0.00105	-0.03215
		Savitzky-Golay SNV	0	0.00105	-0.03215
	3. Turrialba	Sin	0	0.00075	-0.03543
	Estrato medio				
		SNV	6	0.00073	0.02371
	Detrend	3	0.00074	0.00431	
	Savitzky-Golay	0	0.00075	-0.03543	
	SNV Savitzky-Golay	4	0.00073	0.02928	
	2				
4.Victoria	Sin	0	0.00105	-0.03773	
Estrato medio					
	SNV	0	0.00105	-0.03773	
	Detrend	0	0.00105	-0.03773	
	Savitzky-Golay	0	0.00105	-0.03773	
	Savitzky-Golay SNV	0	0.00105	-0.03773	
<i>Tasa de asimilación neta de CO₂ (A)</i>	1.Turrialba	Sin	1	2.24694	-0.02574
	Estrato alto				
	SNV	0	2.26568	-0.04293	

<i>Variable</i>	<i>Modelo</i>	<i>Pretratamiento</i>	<i>Variables Latentes</i>	<i>RMSEP</i>	<i>R²*</i>
<i>Temperatura de la hoja (Tleaf)</i>		Detrend	1	2.26336	-0.04093
		Savitzky-Golay	1	2.24688	-0.02572
		Savitzky-Golay SNV	0	2.26568	-0.04293
	2. Victoria	Sin	0	2.53691	-0.05359
	Estrato alto				
		SNV	0	2.53691	-0.05359
		Detrend	0	2.53691	-0.05359
		Savitzky-Golay	0	2.53691	-0.05359
		Savitzky-Golay SNV	0	2.53691	-0.05359
	3. Turrialba	Sin	0	2.23963	-0.01908
	Estrato medio				
		SNV	7	2.23848	-0.01804
		Detrend	7	2.23639	-0.01613
		Savitzky-Golay	0	2.23963	-0.01908
		SNV Savitzky-Golay	4	2.18112	0.03347
		2			
	4. Victoria	Sin	0	2.5062	-0.02824
	Estrato medio				
		SNV	0	2.5062	-0.02824
		Detrend	0	2.5062	-0.02824
		Savitzky-Golay	0	2.5062	-0.02824
		Savitzky-Golay SNV	0	2.5062	-0.02824
	1. Turrialba	Sin	2	0.44990	0.01741
	Estrato alto				
		SNV	1	0.45841	-0.02012
		Detrend	0	0.46484	-0.04894
		Savitzky-Golay	2	0.44977	0.01795
		Savitzky-Golay SNV	1	0.45825	-0.01944
2. Victoria	Sin	0	0.50247	-0.02060	
Estrato alto					
	SNV	3	0.49599	0.00556	

<i>Variable</i>	<i>Modelo</i>	<i>Pretratamiento</i>	<i>Variables Latentes</i>	<i>RMSEP</i>	<i>R²*</i>
<i>Presión de vapor de saturación (SV_{Pleaf})</i>		Detrend	3	0.49497	0.00965
		Savitzky-Golay	0	0.50247	-0.02060
		Savitzky-Golay SNV	3	0.49581	0.00628
	3. Turrialba	Sin	7	0.42382	0.12803
	Estrato medio				
		SNV	2	0.43881	0.06524
		Detrend	6	0.41589	0.16035
		Savitzky-Golay 2	16	0.355505	0.38802
		Savitzky-Golay SNV	2	0.43871	0.06565
	4. Victoria	Sin	0	0.49987	-0.01006
	Estrato medio				
		SNV	2	0.49212	0.02082
		Detrend	0	0.49987	-0.01006
		Savitzky-Golay	0	0.49987	-0.01006
		Savitzky-Golay SNV	2	0.49207	0.02100
	1. Turrialba	Sin	2	1.91393	0.01970
	Estrato alto				
		SNV	1	1.94106	-0.00829
		Detrend	0	1.96079	-0.02889
		Savitzky-Golay	2	1.91362	0.02001
		Savitzky-Golay SNV	1	1.94040	-0.00760
	2. Victoria	Sin	0	2.17286	-0.02354
	Estrato alto				
		SNV	2	2.15988	-0.01134
	Detrend	3	2.15092	-0.00297	
	Savitzky-Golay	0	2.17286	-0.02354	
	Savitzky-Golay SNV	2	2.15977	-0.01124	
3. Turrialba	Sin	8	1.79536	0.13739	
Estrato medio					
	SNV	4	1.84739	0.08667	

<i>Variable</i>	<i>Modelo</i>	<i>Pretratamiento</i>	<i>Variables Latentes</i>	<i>RMSEP</i>	<i>R²*</i>
		Detrend	7	1.76389	0.16737
		Savitzky-Golay 2	17	1.53460	0.36977
		Savitzky-Golay SNV	4	1.84816	0.08590
4.Victoria		Sin	7	2.14624	0.00139
Estrato medio		SNV	2	2.12431	0.02169
		Detrend	2	2.18280	-0.03292
		Savitzky-Golay	7	2.14583	0.00177
		Savitzky-Golay SNV	2	2.12396	0.02202

NOTA. La abreviatura R^{2*} se refiere al indicador R^2 ajustado. Los valores con los mejores resultados para cada modelo se señalan con negrita.

En la Figura 20 se presentan las gráficas de las curvas de los valores predichos versus los valores de laboratorio para todos los modelos de la tasa de transpiración. Los modelos predictivos para ambas variedades utilizando los datos espectrales del estrato alto (modelo 1 y 2) obtuvieron un $R^2 = -0.031$ y -0.032 , respectivamente.

Para el modelo 1 se utilizó el pretratamiento Savitzky-Golay con 1 variable latente. Los otros pretratamientos generaron resultados similares y el número de variables latentes se mantuvo entre 1 y 0 para todas las combinaciones. Para el modelo 2 todos los pretratamientos obtuvieron la misma magnitud y el número de variables latentes utilizadas se mantuvo en 0. Los modelos generados con los datos del estrato medio para ambas variedades (modelo 3 y 4) presentaron valores de $R^2 = 0.029$ y -0.038 . Para el modelo 3 el mejor resultado se obtuvo con la combinación SNV + Savitzky-Golay 2 (2da derivada) con 4 variables latentes.

En el caso del modelo 4 al igual que en el modelo 2 todos los pretratamientos obtuvieron el mismo resultado con un número de variables latentes igual a 0. Estos valores resultan en modelos que no son lo suficientemente robustos para predecir las relaciones entre las propiedades fotosintéticas de las plantas de café con las firmas espectrales. De igual forma, las gráficas de los datos predichos versus datos medidos de las variables A, Tleaf y SVPleaf se presentan en las Figuras B.5, B.6 y B.7. Los valores del R^2 ajustado para la

variable A son de $R^2 = -0.026, -0.054, 0.033$ y -0.028 para los modelos 5, 6, 7 y 8, respectivamente. El mejor resultado de R^2 ajustado fue el del modelo 7 utilizando la combinación de pretratamientos SNV + Savitzky-Golay2 con 4 variables latentes.

Para la variable Tleaf los valores de R^2 ajustado son de $R^2 = 0.018, 0.001, 0.39$ y 0.021 . El pretratamiento Savitzky-Golay 2 (2da derivada) con 16 variables latentes generó un R^2 ajustado de 0.39 que representa al mejor modelo de la variable Tleaf. Por último, para la variable SVTleaf los resultados de R^2 ajustado son de $R^2 = 0.020, -0.003, -0.37$ y 0.022 . El mejor resultado de R^2 ajustado de 0.37 se obtuvo con el pretratamiento Savitzky-Golay 2 (2da derivada) y 17 variables latentes.

En general, se puede notar que la cantidad de variables latentes influye directamente al modelo predictivo. En el libro de Hastie (2017) se habla de la importancia de seleccionar la cantidad correcta ya que si se utilizan demasiadas se puede provocar un sobreajuste del modelo mientras que si se utilizan pocas se genera un subajuste. La combinación de la cantidad de variables latentes y los pretratamientos definidos modificaron los rendimientos de los modelos generados.

En la investigación llevada a cabo por el CIRAD y aún no publicada a la fecha de esta tesis denominada: Le projet CReSi PAPRICA (Phénotypage pour l'Atténuation des limitations Photosynthétiques chez le RIz dans un contexte d'élévation du CO_2 Atmosphérique) se obtuvieron resultados similares (Chaix, sf). En esta investigación se realizaron mediciones con los mismos equipos ASD FieldSpec Pro 4, Li-Cor LI-6800 y un medidor SPAD. Se midió en ambiente controlado la cantidad de diferentes azúcares y una variedad de variables asociadas al proceso fotosintético de plantas con diferentes genotipos. Algunas de las variables analizadas son la tasa máxima de carboxilación, tasa de transporte de electrones, conductancia del mesófilo, fotorrespiración. Se utilizaron tres diferentes algoritmos de aprendizaje supervisado Transformer, PLSR y Decon_Sep. Los indicadores R^2 obtenidos después de la PLSR para las variables anteriormente mencionadas se encontraron en el rango de $R^2 = 0 - 0.28$. A pesar de los diferentes pretratamientos aplicados a los sets de datos, los resultados fueron de igual forma bajos. Entre las conclusiones de este estudio se menciona la compleja relación entre las variables a predecir. Se expone la posibilidad de problemas con la variabilidad espacial y temporal del medidor de fotosíntesis portátil Li-

6800 (Chaix, sf). En el caso de los problemas relacionados a la variabilidad de las mediciones del Li-6800 fue que el gradiente de la hoja es bastante alto en comparación con la diferencia del genotipo.

Las mediciones del SPAD y NIR son más bajas y cuatro veces menor en el caso de NIRS. La repetibilidad del SPAD y NIRS es de cuatro a cinco veces mejor que el Li-6800. Se cree que las mediciones n del Li-6800 tienen un efecto n+1 en las variables a causa de la fisiología de la hoja (Chaix, sf). De igual forma, cabe resaltar que la variabilidad temporal reflejada en los datos de las variables fisiológicas está estrechamente relacionada con la hora asociada a cada medida realizada el día de visita. Es decir, Las mediciones iniciaron a las 6 de la mañana y normalmente terminaron alrededor de las 12 md. En este periodo las variables estudiadas que dependen proporcionalmente de la temperatura y la humedad relativa variaron a través del tiempo. Por último, al agregar el efecto de la deshidratación no homogénea de las muestras se obtienen datos con alta variabilidad que van a afectar la capacidad predictiva.

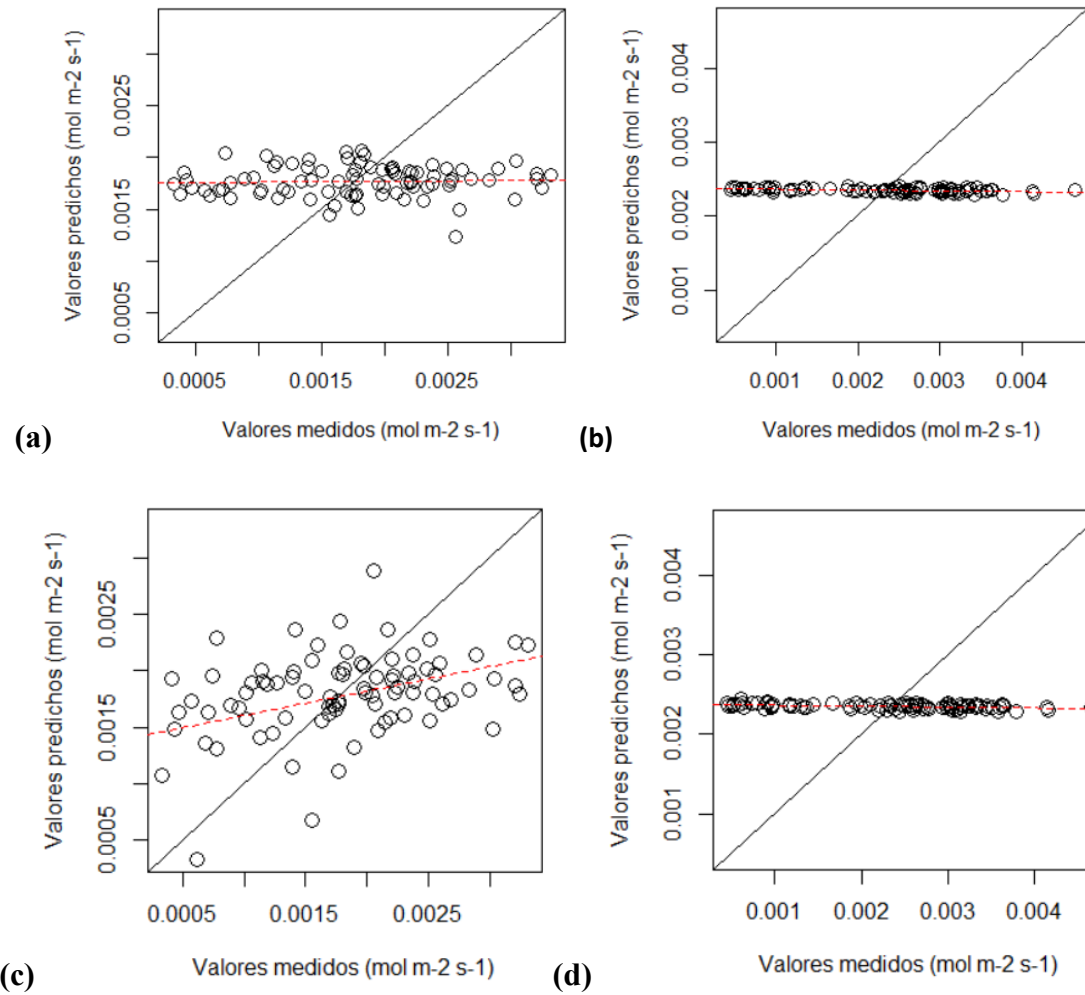


Figura 20. Curvas de datos predichos (eje y) vs datos de la tasa de transpiración (eje x) para el modelo 1(a), modelo 2(b), modelo 3(c) y modelo 4(d).

6. CONCLUSIONES

- Se desarrollaron cuatro modelos para cada una de las variables que definen la capacidad fotosintética: tasa de transpiración (E), tasa de asimilación neta de CO₂ (A), temperatura de la hoja (Tleaf) y presión de vapor de saturación (SVPl_{leaf}). Las condiciones controladas establecidas en el diseño experimental no se lograron cumplir ya que las condiciones climáticas de la zona en la época en la que se realizó la investigación influyeron directamente en el estado hídrico de las muestras. Por lo tanto, la variabilidad de los datos tuvo un impacto negativo en el rendimiento de los modelos predictivos.
- La relación entre el estrés hídrico de las muestras se define por la correspondencia de las variables que componen la capacidad fotosintética y como estas son directamente influenciadas por el estado hídrico de las plantas de café.
- En la investigación se logró comprobar una rehidratación incidente de las muestras que se atribuye a la interacción de las condiciones climáticas de la zona con las muestras. De igual forma, las muestras se deshidrataron de una manera no homogénea y rápida en los últimos días de medición. Esto generó un escenario en donde los tratamientos EHC y EH mostraron comportamientos atípicos.
- Los análisis de varianza ANOVA reafirmaron el comportamiento de las variables a través del tiempo y sus relaciones entre los grupos: tiempo, tratamiento y variedad. En todas las variables se encontraron diferencias significativas respecto al primer grupo (tiempo), en donde el análisis de Tukey mostró que la fecha que se diferenciaba más era el día 38 (último día).
- El rendimiento de los modelos se ve afectado por diferentes factores. Los pretratamientos espectrales utilizados y la cantidad de componentes principales (PC) o variables latentes (LV's) afectan directamente a los valores de RMSEP y R². En algunos casos la alta variabilidad de los datos generó la misma magnitud en los valores de los indicadores independientemente del pretratamiento utilizado, igualmente el número de variables latentes óptimo se mantuvo en cero.

- Los resultados obtenidos no son comparables con los encontrados en la literatura para variables fisiológicas similares. El resultado de mejor rendimiento de los modelos fue de un R^2 ajustado = -0.031 y RMSEP = 0.36. Fue generado utilizando los datos del estrato medio para predecir la temperatura de la hoja (Tleaf) en la variedad Turrialba. Para lograr este resultado se utilizó el pretratamiento Savitzky-Golay 2 (2da derivada) con 16 variables latentes.
- La alta variabilidad de los datos se atribuye principalmente a errores metodológicos en el periodo de medición y también se cree que está relacionada a la poca repetibilidad del medidor de fotosíntesis Li-Cor Li-6800 en comparación con el espectroradiómetro. Las condiciones de laboratorio de la mayoría de los artículos en donde se logran obtener correlaciones altas son muy diferentes a las condiciones de la investigación, las cuales se asemejan más a las condiciones reales de un invernadero. La complejidad de las interacciones fotosintéticas de las plantas y como estas se relacionan con otras variables es clara en diferentes referencias bibliográficas.
- Las condiciones explicadas al inicio de esta sección mostraron un escenario en donde los productos Photon (Registro SFE #10144) y Seamaxx (Registro SFE #7553) generaron un aumento del potencial hídrico para la variedad Victoria 14. Por el contrario, la variedad Turrialba obtuvo mejoría en las variables restantes. Estos resultados presentaron diferencias a los comportamientos esperados y en algunos casos variables fisiológicas que se encuentran estrechamente relacionadas registraron datos que muestran no tener relación entre ellas.
- La alta variabilidad de los datos hace difícil interpretar con certeza los datos y generar resultados que no se encuentren sesgados. Por lo tanto, no se logró comprobar la efectividad de ambos insumos agrícolas de manera concisa.

En conclusión, la investigación no alcanzó su objetivo principal en su totalidad debido a que los modelos predictivos exhibieron una capacidad de predicción limitada y no se logró comprobar las interacciones reales entre los tratamientos propuestos y las variables dependientes. A pesar de estos resultados, el estudio permitió identificar las limitaciones enfrentadas, estableciendo una línea base y aportando valiosas lecciones para futuras investigaciones. Este reconocimiento de las restricciones del estudio sienta las bases para

abordajes más efectivos en el futuro, enfatizando la importancia de la mejora continua en la investigación científica.

7. RECOMENDACIONES

Teniendo en cuenta los objetivos de la investigación es recomendable hacer un análisis previo de las condiciones del espacio físico. En caso de que el espacio sea reducido, el diseño de una estructura resiliente a las condiciones de la zona es lo óptimo. Lo más recomendable es un invernadero en donde las condiciones sean las más controladas posibles y las condiciones externas estén lo suficientemente asiladas para que no afecten a las muestras. También es importante tomar en cuenta la época en la que se va a realizar el estudio para disminuir los efectos climáticos aún más.

En cuanto al material genético es recomendable que todas las plantas sean de la misma edad y se encuentren en condiciones óptimas (libres de plagas y enfermedades y con adecuados niveles nutricionales). Esto contribuiría a que los resultados de las variables cuantificadas sean únicamente atribuidos al estrés abiótico y no al biótico. Una planta de alrededor de un año es óptima en casos similares a este estudio ya que la planta es más sensible a cambios relacionado con la absorción de agua y nutrientes. Por lo tanto, a través del tiempo la diferencia en las variables se puede cuantificar de manera más eficiente.

La cantidad de muestras debe ser previamente preparada y establecida según el diseño experimental a desarrollar. En este caso, para un análisis de tres tratamientos con dos variedades de plantas de café se recomienda aumentar la muestra y los días.

De igual forma, se recomienda un equipo de trabajo con previa capacitación para realizar las mediciones de manera correcta reduciendo el error humano causado por diferentes razones. También se recomienda que las interacciones de las variables a predecir sean simples y no dependan de mecanismos complejos que puedan sesgar los resultados y por lo tanto disminuir la capacidad predictiva de los modelos.

En el caso de la capacidad fotosintética se recomienda utilizar diferentes algoritmos de aprendizaje asistido para el procesamiento de los datos obtenidos y la implementación de índices espectrales como alternativa para evaluar la efectividad de ambos. Aunque es más

importante reducir los errores metodológicos al máximo para disminuir la variabilidad de los datos y obtener resultados más precisos.

La previa calibración y capacitación en el uso de los instrumentos definidos para el estudio es clave para realizar mediciones precisas. En el caso del espectrorradiómetro se recomienda realizar pruebas anteriores con los diferentes fondos del plant probe y crear una línea base antes de iniciar las mediciones para reducir el ruido en diferentes longitudes de onda generadas por condiciones externas. Además, la utilización de instrumentos adicionales como tensiómetros o la implementación de pruebas de laboratorio para determinar el estado inicial y final de las muestras. Estas estrategias pueden apoyar la toma de decisiones críticas que pueden llevar a la investigación por el camino correcto.

8. REFERENCIAS

- Alemán, S. G., Solano, E. & López, P. J. (2019). La arquitectura hidráulica de las plantas vasculares terrestres, una revisión. *Madera y bosques*, 25(3).
- Beattie, J. R., & Esmonde-White, F. W. (2021). Exploration of principal component analysis: deriving principal component analysis visually using spectra. *Applied Spectroscopy*, 75(4), 361-375.
- Brereton, R. G. (2018). *Chemometrics*. Wiley.
- Brown, S. D., Tauler, R., & Walczak, B. (Eds.). (2020). *Comprehensive chemometrics: chemical and biochemical data analysis*. Elsevier.
- Castillo, A.A. (2020). Efecto del punto de liberación del estrés hídrico sobre la fenología, la fisiología, el rendimiento y la calidad de naranja dulce (*Citrus sinensis* Osb. Var. Valencia) injertada sobre dos patrones contrastantes en el trópico húmedo. *Revista Científica LIFE-RID-2020 I Edición*.
- Chemura, A., Mutanga, O., & Dube, T. (2017). Remote sensing leaf water stress in coffee (*Coffea arabica*) using secondary effects of water absorption and random forests. *Physics and Chemistry of the Earth*, 100, 317–324.
<https://doi.org/10.1016/j.pce.2017.02.011>
- Chen, Z. P., Morris, J., & Martin, E. (2006). Extracting chemical information from spectral data with multiplicative light scattering effects by optical path-length estimation and correction. *Analytical Chemistry*, 78(22), 7674–7681.
<https://doi.org/10.1021/ac0610255>
- Chicco D, Warrens MJ, Jurman G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science* 7:e623 <https://doi.org/10.7717/peerj-cs.623>
- Clevers, J. G. P. W., & Kooistra, L. (2012). Using hyperspectral remote sensing data for retrieving canopy chlorophyll and nitrogen content. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2), 574–583.
<https://doi.org/10.1109/JSTARS.2011.2176468>

- Cortez Mojica G, & Tercero Mercado J. (2014). *Efecto de tres láminas de riego por goteo en época seca sobre el estímulo de la floración y producción de café (Coffea arabica L. Var. CATRENIC), finca San Dionisio, San Marcos, Carazo 2013.*
- Dai, H. (2022). Analysis of Variance (ANOVA). En: Y. Ding, X. Li, J. Li, Y. Li, & H. Dai (Eds.), *Handbook of Data Science* (pp. 1-12). Springer. https://doi.org/10.1007/978-981-16-7023-3_93-1
- Damatta, F. M. (2004). Exploring drought tolerance in coffee: a physiological approach with some insights for plant breeding. In *Braz. J. Plant Physiol* (Vol. 16, Issue 1).
- Elvanidi, A., Katsoulas, N., Ferentinos, K. P., Bartzanas, T., & Kittas, C. (2018). Hyperspectral machine vision as a tool for water stress severity assessment in soilless tomato crop. *Biosystems Engineering*, 165, 25–35.
<https://doi.org/https://doi.org/10.1016/j.biosystemseng.2017.11.002>
- Fernández, L., Zavala, D., Castillo, A. y Delgado, J. (2020). Efecto mitigante de Seamaxx (*Ascophyllum nodosum* L.) al estrés hídrico en tres variedades de café (*Coffea arabica* L.) Costa Rica 95, Obatá y Catuaí Rojo. *Revista Científica LIFE-RID*, 80-89.
- Fernández, W. A. T., Cortez, P. J. M., Guerrero, W. D. L., Chila, V. H. R., & Ortega, M. J. B. (2020). Capacidad fotosintética de 15 clones de café robusta (*Coffea canephora*). *Revista ESPAMCIENCIA ISSN 1390-8103*, 11(2), 57-67.
- Fitzgerald, G., Rodriguez, D., & O’Leary, G. (2010). Measuring and predicting canopy nitrogen nutrition in wheat using a spectral index-The canopy chlorophyll content index (CCCI). *Field Crops Research*, 116(3), 318–324.
<https://doi.org/10.1016/j.fcr.2010.01.010>
- Galarza, A. R. (2021). Diseños de investigación experimental. *CienciAmérica: Revista de divulgación científica de la Universidad Tecnológica Indoamérica*, 10(1), 1-7.
- González, R. F., Cisneros, Z. E., & Montilla, E. (2017). Respuesta al déficit hídrico del cafeto (*Coffea arabica* L.) en diferentes fases de desarrollo. *Revista Ciencias Técnicas Agropecuarias*, 26(3), 4-11.
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2017). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: springer.

- Henseler, J. (2018). Partial least squares path modeling: Quo vadis?. *Quality & Quantity*, 52(1), 1-8.
- Hernandez, R. D., Tapia, E. M., & Hernandez, G, S. (2019). *Estadística inferencial 2: aplicaciones para ingeniería*. Grupo Editorial Patria.
<https://elibro.net/es/lc/sibdi/titulos/121281>
- Hobley, E., Steffens, M., Bauke, S. L., & Kögel-Knabner, I. (2018). Hotspots of soil organic carbon storage revealed by laboratory hyperspectral imaging. *Scientific Reports*, 8(1), 13900.
- Hu, T., & Qi, K. (2018, July). Using vis-nir spectroscopy to estimate soil organic content. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 8263-8266). IEEE.
- Inoue, Y., Guérif, M., Baret, F., Skidmore, A., Gitelson, A., Schlerf, M., Darvishzadeh, R., & Olioso, A. (2016). Simple and robust methods for remote sensing of canopy chlorophyll content: a comparative analysis of hyperspectral data for different types of vegetation. *Plant, Cell & Environment*, 39(12), 2609–2623.
<https://doi.org/https://doi.org/10.1111/pce.12815>
- Instituto del Café de Costa Rica (ICAFFE). (2020). Informe sobre la actividad cafetalera de Costa Rica. *Congreso Nacional Cafetalero 10 Dic 2020*. San José (Costa Rica).
- Jacquemoud, S., & Ustin, S. L. (2001). Leaf optical properties: A state of the art. In *8th International Symposium of Physical Measurements & Signatures in Remote Sensing* (pp. 223-332). CNES Aussois France.
- Jalalvand, A. R., Ghobadi, S., Akbari, V., Goicoechea, H. C., Faramarzi, E., & Mahmoudi, M. (2019). Mathematical modeling of interactions of cabergoline with human serum albumin for biosensing of human serum albumin. *Sensing and Bio-Sensing Research*, 25, 100297.
- John, A., Sadasivan, J., & Seelamantula, C. S. (2021). Adaptive Savitzky-Golay filtering in non-Gaussian noise. *IEEE Transactions on Signal Processing*, 69, 5021-5036.
- Katsoulas, N., Savvas, D., Tsirogiannis, I., Merkouris, O., & Kittas, C. (2009). Response of an eggplant crop grown under Mediterranean summer conditions to greenhouse fog

- cooling. *Scientia Horticulturae*, 123(1), 90–98.
<https://doi.org/10.1016/j.scienta.2009.08.004>
- Kurz, T. H., & Buckley, S. J. (2016). A review of hyperspectral imaging in close range applications. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 41, 865–870.
<https://doi.org/10.5194/isprsarchives-XLI-B5-865-2016>
- Lesnoff, M. (2021). R package rnirs: Dimension reduction, Regression and Discrimination for Chemometrics. <https://github.com/mlesnoff/rnirs>. CIRAD, UMR SELMET, Montpellier, France
- Lesnoff, M. 2021. R package rchemo: Dimension reduction, Regression and Discrimination for Chemometrics. <https://github.com/mlesnoff/rchemo>. CIRAD, UMR SELMET, Montpellier, France
- León Rojas, F. R. (2022). Estrés hídrico y alta temperatura en Café (*Coffea arabica* L.) en condiciones de vivero. Tesis para optar por el título de Ing. Agrónomo. Lima, Perú.
- Liu, J., Miller, J. R., Haboudane, D., Pattey, E., Nolin, M. C., Liu, J., Pattey, E., Miller, J., & Haboudane, D. (2005). Variability of seasonal CASI image data products and potential application for management zone delineation for precision agriculture. In *Can. J. Remote Sensing* (Vol. 31, Issue 5).
- Lourenco, R. A., Combi, T., da Rosa Alexandre, M., Sasaki, S. T., Zanardi-Lamardo, E., & Yogui, G. T. (2020). Mysterious oil spill along Brazil's northeast and southeast seaboard (2019–2020): Trying to find answers and filling data gaps. *Marine Pollution Bulletin*, 156, 111219.
- Lu, B., Dao, P. D., Liu, J., He, Y., & Shang, J. (2020). Recent advances of hyperspectral imaging technology and applications in agriculture. In *Remote Sensing* (Vol. 12, Issue 16). MDPI AG. <https://doi.org/10.3390/RS12162659>
- Malegori, C., Alladio, E., Oliveri, P., Manis, C., Vincenti, M., Garofano, P., ... & Berti, A. (2020). Identification of invisible biological traces in forensic evidences by hyperspectral NIR imaging combined with chemometrics. *Talanta*, 215, 120911.
- Mohd Asaari, M. S., Mishra, P., Mertens, S., Dhondt, S., Inzé, D., Wuyts, N., & Scheunders, P. (2018). Close-range hyperspectral image analysis for the early detection of stress responses in individual plants in a high-throughput phenotyping

- platform. *ISPRS Journal of Photogrammetry and Remote Sensing*, 138, 121–138.
<https://doi.org/10.1016/j.isprsjprs.2018.02.003>
- Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO). (2022). Versión resumida de El estado de la seguridad alimentaria y la nutrición en el mundo 2022. Adaptación de las políticas alimentarias y agrícolas para hacer las dietas saludables más asequibles. Roma, FAO. <https://doi.org/10.4060/cc0640es>
- Peña, L., Rentería, V., Velásquez, C., Ojeda, M.L., & Barrera, E.. (2019). Absorbancia y reflectancia de hojas de Ficus contaminadas con nanopartículas de plata. *Revista mexicana de física*, 65(1), 95-105. Epub 09 de noviembre de 2019.
- Peña, R. K., Donoso, S., Pacheco, C., Riquelme, A., Gangas, R., Guajardo, A., & Durán, S. (2018). Respuestas morfo-fisiológicas de plantas de *Lithraea caustica* (Anacardiaceae) sometidas a restricción hídrica controlada. *Bosque (Valdivia)*, 39(1), 27-36.
- Pérez, L. U., Ávila, G. W., Luna. J. G., Rivera. L. A., & Barrera. G. A. (2021). Impact of Water Stress on Plant Growth and Development: Physiological, Biochemical and Molecular Approaches. *Environ Exp Biol*, 19(1), 1-12. DOI: 10.22364/eeb.19.01.01.
- Piccolo, S. Matteoli, M. Landi, L. Guidi, R. Massai and D. Remorini, "Measurements of Anthocyanin Content of Prunus Leaves Using Proximal Sensing Spectroscopy and Statistical Machine Learning," in *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-10, 2022, Art no. 2508110, doi: 10.1109/TIM.2022.3167796.
- Pino, E., Montalván, I., Vera, A., & Ramos, L. (2019). La conductancia estomática y su relación con la temperatura foliar y humedad del suelo en el cultivo del olivo (*Olea europaea* L.), en periodo de maduración de frutos, en zonas áridas. *La Yarada, Tacna, Perú. Idesia (Arica)*, 37(4), 55-64.
- PROCOMER. (2019). Precios internacionales del café continúan aumentando. *Global Coffee Report*. https://www.procomer.com/alertas_comerciales/exportador-alerta/precios-internacionales-del-cafe-continuan-aumentando/
- Ross, S. M. (2018). *Introducción a la estadística*. Reverté.

- Sankaran, S., Mishra, A., Ehsani, R., & Davis, C. (2010). A review of advanced techniques for detecting plant diseases. In *Computers and Electronics in Agriculture* (Vol. 72, Issue 1, pp. 1–13). <https://doi.org/10.1016/j.compag.2010.02.007>
- Santesteban, L. G., Miranda, C., Marín, D., Sesma, B., Intrigliolo, D. S., Mirás-Ávalos, J. M., ... & Royo, J. B. (sf). Medir el potencial hídrico: cuándo, ¿dónde y por qué?
- Scholander, P. F., Bradstreet, E. D., Hemmingsen, E. A., & Hammel, H. T. (1965). Sap Pressure in Vascular Plants: Negative hydrostatic pressure can be measured in plants. *Science*, 148(3668), 339-346.
- T. Hu, K. Qi & Y. Hu. (2018). Using Vis-Nir Spectroscopy to Estimate Soil Organic Content," IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain. pp. 8263-8266, doi: 10.1109/IGARSS.2018.8519021.
- Taiz, L., Zeiger, E., Møller, I. M., & Murphy, A. (2015). Plant physiology and development (No. Ed. 6). Sinauer Associates Incorporated.
- Thenkabail, P. S., Lyon, J. G., y Huete, A. (2019a). *HYPERSPECTRAL REMOTE SENSING OF VEGETATION SECOND EDITION Fundamentals, sensor systems, spectral libraries and data mining for vegetation*.
- Thenkabail, P. S., Lyon, J. G., y Huete, A. (2019b). *HYPERSPECTRAL REMOTE SENSING OF VEGETATION SECOND EDITION Hyperespectral Indices and Image Classifications for Agriculture and Vegetation*.
- Thenkabail, P. S., Lyon, J. G., y Huete, A. (2019c). *HYPERSPECTRAL REMOTE SENSING OF VEGETATION SECOND EDITION Biophysical and Biochemical Characterization and Plant Species Studies*.
- Tolentino, G. L., Arguello, B. M., & Ricardo, H. L. S. (2016). Medición de intercambio gaseoso, área foliar e índice de clorofila en plantas elicidadas con nanopartículas.
- Vega-Vilca, J. C., & Guzmán, J. (2011). Regresión PLS y PCA como solución al problema de multicolinealidad en regresión múltiple. *Revista de Matemática Teoría y Aplicaciones*, 18(1), 09-20.
- Veintimilla, V., & Isabel, R. (2022). *Comportamiento agronómico de materiales de café robusta, Coffea canephora, tolerantes al estrés hídrico bajo condiciones de marchitez permanente* (Bachelor's thesis, La Libertad: Universidad Estatal Península de Santa Elena, 2022).

- Villesseche, H., Ecartot, M., Ballini, E. *et al.* (2022). Unsupervised analysis of NIRS spectra to assess complex plant traits: leaf senescence as a use case. *Plant Methods* **18**, 100. <https://doi.org/10.1186/s13007-022-00927-6>
- Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, 236. <https://doi.org/10.1016/j.rse.2019.111402>
- Wijewardana, C., Alsajri, F. A., Irby, J. T., Krutz, L. J., Golden, B., Henry, W. B., ... & Reddy, K. R. (2019). Physiological assessment of water deficit in soybean using midday leaf water potential and spectral features. *Journal of Plant Interactions*, 14(1), 533-543.
- World Coffee Research. (2019). *LAS VARIETADES DEL CAFÉ ARÁBICA Un catálogo global de variedades que abarca: Costa Rica, El Salvador, Guatemala, Honduras, Jamaica, Kenia, Malawi, Nicaragua, Panamá, Perú, República Dominicana, Rwanda, Uganda, Zambia, Zimbabue.*
- Zhang, Y., Wu, J., & Wang, A. (2022). Comparison of various approaches for estimating leaf water content and stomatal conductance in different plant species using hyperspectral data. *Ecological Indicators*, 142, 109278.
- Zhou J-J, Zhang Y-H, Han Z-M, Liu X-Y, Jian Y-F, Hu C-G, Dian Y-Y. (2021). Evaluating the Performance of Hyperspectral Leaf Reflectance to Detect Water Stress and Estimation of Photosynthetic Capacities. *Remote Sensing* 13, 2160.

APÉNDICE

APÉNDICE A:
RESULTADOS DE EVALUACIÓN EXPERIMENTAL

Cuadro A.1 Mediciones de potencial hídrico (bar) de todas las muestras en el periodo de investigación.

06/05/23		13/05/23		23/05/23		06/06/23		13/06/23	
Muestra	PH	Muestra	PH	Muestra	PH	Muestra	PH	Muestra	PH
T14	-1.5	T30	-4.5	T8	-1.0	T5	-1.0	T6	-1.0
T63	-2.5	T51	-2.5	T31	-1.5	T45	-1.0	T29	-1.5
V41	-2.0	T14	-2.5	T51	-1.5	T49	-1.0	T50	-1.5
V11	-2.5	V10	-3.5	V19	-1.0	V18	-1.5	V12	-1.0
V49	-1.5	V68	-3.0	V30	-1.5	V32	-1.0	V29	-1.5
V51	-1.0	V44	-3.5	V66	-1.0	V52	-1.0	V65	-1.0
T11	-2.0	T9	-4.5	T21	-1.0	T10	-3.0	T12	-25.5
T38	-1.5	T57	-4.0	T44	-1.0	T41	-1.5	T43	-28.0
T41	-2.5	T37	-2.0	T58	-3.0	T54	-2.0	T60	-26.5
T53	-1.5	V2	-3.5	V1	-2.0	V5	-4.0	V4	-6.5
V6	-1.5	V71	-3.0	V39	-4.0	V37	-7.5	V59	-25.0
V27	-2.0	V26	-2.5	V60	-1.5	V57	-1.5	T2	-4.5
T4	-3.0	T1	-3.0	T20	-2.0	T17	-4.0	T36	-32.0
T69	-1.5	T27	-2.5	T33	-1.0	T34	-19.0	T69	-4.0
T26	-1	T66	-2.5	T70	-1.0	T67	-2.0	V23	-21.5
V23	-1.5	V15	-5.0	V24	-3.5	V21	-7.0	V33	-40.0
V36	-1.5	V54	-3.0	V35	-2.5	V47	-2.5	V40	-32.0
V55	-1.5	V48	-2.5	V61	-2.5	V64	-3.0	V53	-27.0

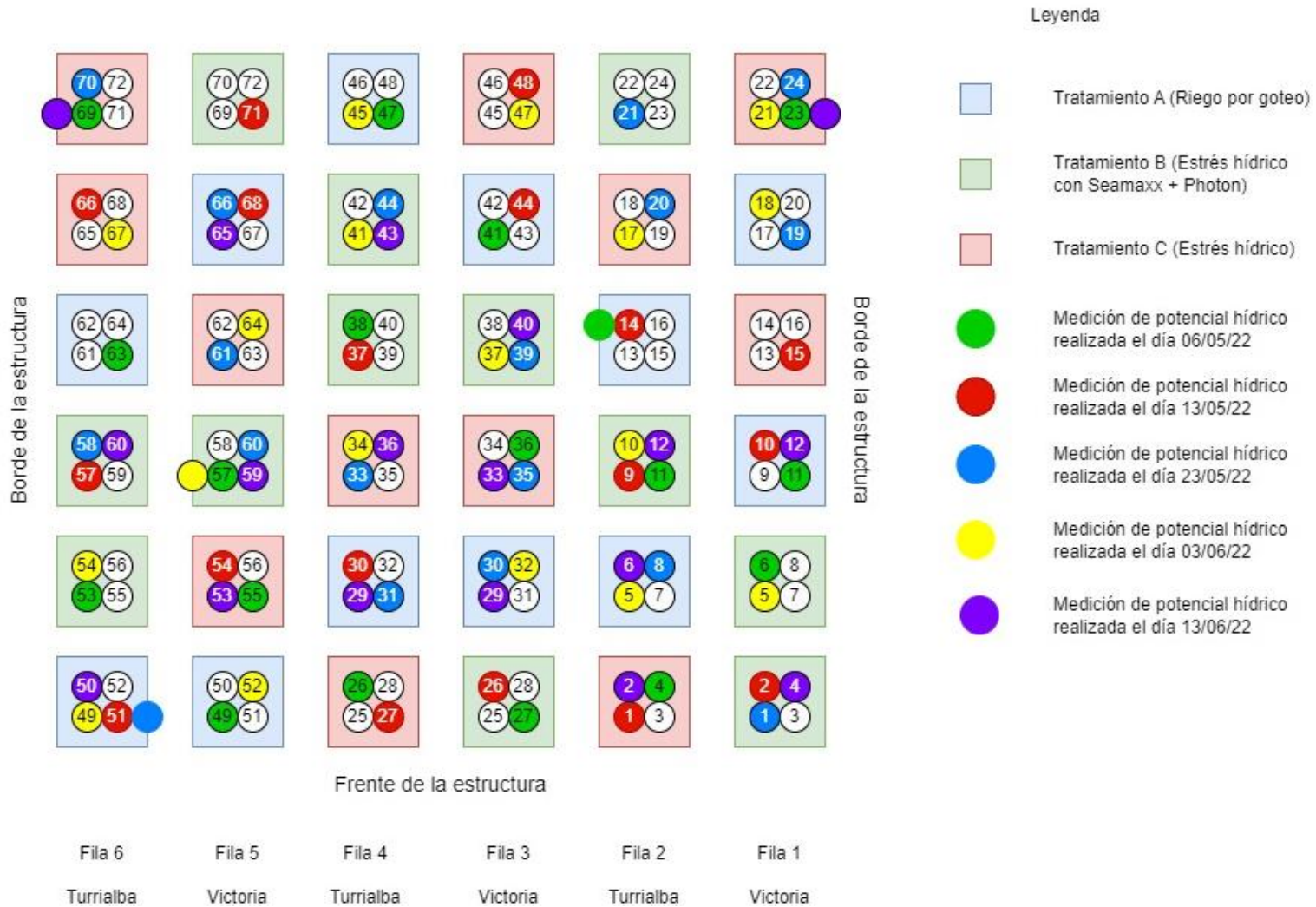


Figura A.1 Esquema con la posición de las muestras escogidas aleatoriamente.

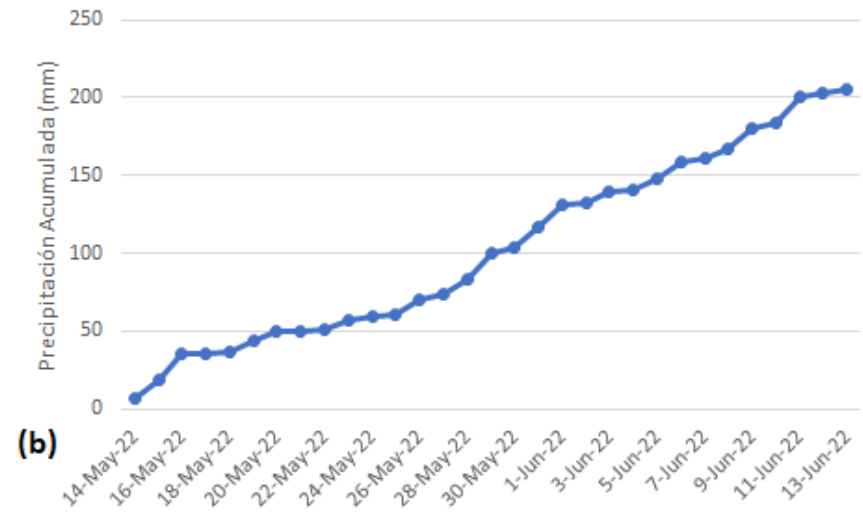
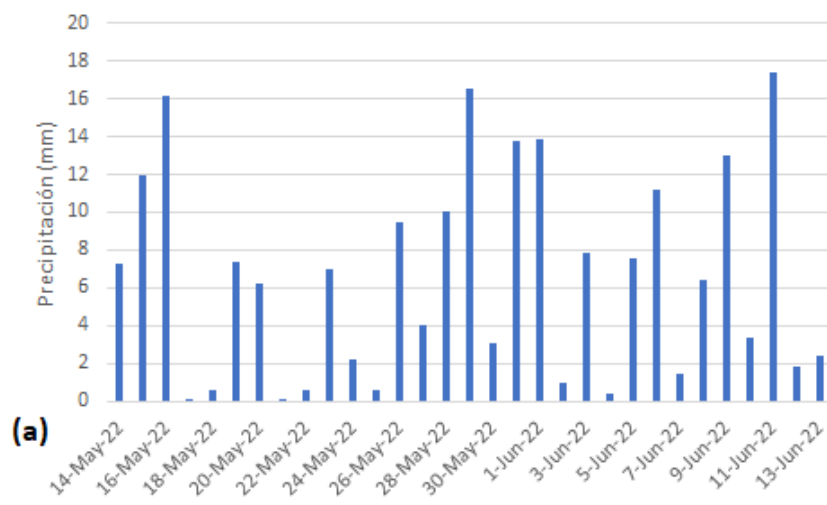


Figura A.2 Datos de precipitación (a) y precipitación acumulada (b) en mm de la estación meteorológica (SOMBRA) de Hacienda Alsacia

Cuadro A.2 Resultados completos de analisis de varianza (ANOVA) para potencial hidrico.

<i>Descripción</i>	<i>Suma de Cuadrados</i>	<i>Grados de Libertad</i>	<i>Cuadrado Medio</i>	<i>Valor de F</i>	<i>Valor de P</i>
<i>Fecha, X₁</i>	2050.08	3	683.36	36.59	<0.001
<i>Tratamiento, X₂</i>	423.50	2	211.75	11.34	0.001
<i>Variedad, X₃</i>	6.82	1	6.82	0.37	0.5485
<i>X₁ X₂</i>	1194.55	6	199.09	10.66	<0.001
<i>X₁ X₃</i>	10.52	3	3.51	0.19	0.9042
<i>X₂ X₃</i>	163.17	2	81.58	4.37	0.0181
<i>X₁ X₂ X₃</i>	404.42	6	67.40	3.61	0.0049

Cuadro A.3 Resultados completos de analisis de varianza (ANOVA) para tasa de transpiración.

<i>Descripción</i>	<i>Suma de Cuadrados</i>	<i>Grados de Libertad</i>	<i>Cuadrado Medio</i>	<i>Valor de F</i>	<i>Valor de P</i>
<i>E</i>	<i>Tasa de transpiración</i>				
<i>Fecha, X₁</i>	2.5E-05	2	1.3E-05	15.49	<0.0001
<i>Tratamiento, X₂</i>	9.2E-06	2	4.6E-06	5.69	0.0042
<i>Variedad, X₃</i>	1.2E-05	1	1.2E-05	15.17	0.0001
<i>X₁X₂</i>	1.5E-05	4	3.8E-06	4.68	0.0014
<i>X₁X₃</i>	2.6E-06	2	1.3E-06	1.63	0.1998
<i>X₂X₃</i>	3.2E-06	2	1.6E-06	2.00	0.1393
<i>X₁X₂X₃</i>	3.0E-06	4	7.4E-07	0.91	0.4576

Cuadro A.4 Resultados completos de analisis de varianza (ANOVA) para tasa de asimilación.

<i>Descripción</i>	<i>Suma de Cuadrados</i>	<i>Grados de Libertad</i>	<i>Cuadrado Medio</i>	<i>Valor de F</i>	<i>Valor de P</i>
A	Tasa de asimilación neta de CO₂				
<i>Fecha, X₁</i>	61.06	2	30.53	5.77	0.0039
<i>Tratamiento, X₂</i>	79.01	2	39.50	7.47	0.0008
<i>Variedad, X₃</i>	37.77	1	37.77	7.14	0.0084
<i>X₁X₂</i>	91.94	4	22.98	4.34	0.0024
<i>X₁X₃</i>	5.51	2	2.75	0.52	0.5953
<i>X₂X₃</i>	24.51	2	12.26	2.32	0.1023
<i>X₁X₂X₃</i>	25.50	4	6.38	1.21	0.3113

Cuadro A.5 Resultados completos de analisis de varianza (ANOVA) para temperatura de la hoja en el balance de energía (Tleaf).

<i>Descripción</i>	<i>Suma de Cuadrados</i>	<i>Grados de Libertad</i>	<i>Cuadrado Medio</i>	<i>Valor de F</i>	<i>Valor de P</i>
Tleaf	Temperatura de la hoja				
<i>Fecha, X₁</i>	147.16	2	73.58	21.42	<0.0001
<i>Tratamiento, X₂</i>	0.78	2	0.39	0.11	0.8927
<i>Variedad, X₃</i>	6.99	1	6.99	2.03	0.1559
<i>X₁X₂</i>	1.76	4	0.44	0.13	0.9720
<i>X₁X₃</i>	0.58	2	0.29	0.08	0.9195
<i>X₂X₃</i>	2.5E-03	2	1.3E-03	3.7E-04	0.9996
<i>X₁X₂X₃</i>	0.58	4	0.15	0.04	0.9966

Cuadro A.6 Resultados completos de analisis de varianza (ANOVA) para presión de vapor de saturación a la temperatura de la hoja (SVPleaf).

<i>Descripción</i>	<i>Suma de Cuadrados</i>	<i>Grados de Libertad</i>	<i>Cuadrado Medio</i>	<i>Valor de F</i>	<i>Valor de P</i>
SVPleaf	Presión de vapor de saturación a la temperatura de la hoja				
<i>Fecha, X₁</i>	7.57	2	3.78	20.93	<0.0001
<i>Tratamiento, X₂</i>	0.04	2	0.02	0.10	0.9037
<i>Variedad, X₃</i>	0.37	1	0.37	2.04	0.1552
<i>X₁X₂</i>	0.09	4	0.02	0.12	0.9743
<i>X₁X₃</i>	0.02	2	0.01	0.06	0.9442
<i>X₂X₃</i>	2.3E-04	2	1.2E-04	6.4E-04	0.9994
<i>X₁X₂X₃</i>	0.04	4	0.01	0.05	0.9943

Cuadro A.7 Resultados completos de analisis de varianza (ANOVA) para altura de las plantas.

<i>Descripción</i>	<i>Suma de Cuadrados</i>	<i>Grados de Libertad</i>	<i>Cuadrado Medio</i>	<i>Valor de F</i>	<i>Valor de P</i>
h	Altura				
<i>Fecha, X₁</i>	4101.15	4	1025.29	51.40	<0.0001
<i>Tratamiento, X₂</i>	74.06	2	37.03	1.86	0.1570
<i>Variedad, X₃</i>	26199.67	1	26199.67	1313.46	<0.0001
<i>X₁X₂</i>	235.60	8	29.45	1.48	0.1621
<i>X₁X₃</i>	73.41	4	18.35	0.92	0.4516
<i>X₂X₃</i>	11.85	2	5.92	0.30	0.7432
<i>X₁X₂X₃</i>	21.85	8	2.73	0.14	0.9975

Cuadro A.8 Resultados completos de analisis de varianza (ANOVA) para cantidad de hojas.

<i>Descripción</i>	<i>Suma de Cuadrados</i>	<i>Grados de Libertad</i>	<i>Cuadrado Medio</i>	<i>Valor de F</i>	<i>Valor de P</i>
<i>n</i>	Cantidad de hojas				
<i>Fecha, X₁</i>	127567.38	4	31891.84	238.40	<0.0001
<i>Tratamiento, X₂</i>	905.41	2	452.70	3.38	0.0345
<i>Variedad, X₃</i>	170.87	1	170.87	1.28	0.2588
<i>X₁X₂</i>	402.22	8	50.28	0.38	0.9335
<i>X₁X₃</i>	887.10	4	221.78	1.66	0.1580
<i>X₂X₃</i>	1059.13	2	529.56	3.96	0.0195
<i>X₁X₂X₃</i>	21.59	8	2.70	0.02	>0.9999

Cuadro A.9 Resumen de las relaciones porcentuales de las variables fisiológicas entre los tratamientos EH con respecto a EHC en ambas variedades para el día 31 y 31.

<i>Variable</i>	<i>Variedad</i>	<i>Al día 31</i>	<i>Al día 38</i>
<i>PH</i>	Turrialba	74%	-97.53%
<i>PH</i>	Victoria	-4%	47.72%
<i>E</i>	Turrialba	N/A	-16.23%
<i>E</i>	Victoria	N/A	56.09%
<i>A</i>	Turrialba	N/A	-15.30%
<i>A</i>	Victoria	N/A	25.58%
<i>Tleaf</i>	Turrialba	N/A	0.01%
<i>Tleaf</i>	Victoria	N/A	-0.22%
<i>SVPleaf</i>	Turrialba	N/A	0.12%
<i>SVPleaf</i>	Victoria	N/A	0.61%

NOTA. N/A hace referencia a que en ese día no se realizaron mediciones para esas variables.

APÉNDICE B:

RESULTADOS DE ANÁLISIS DE RESPUESTAS ESPECTRALES

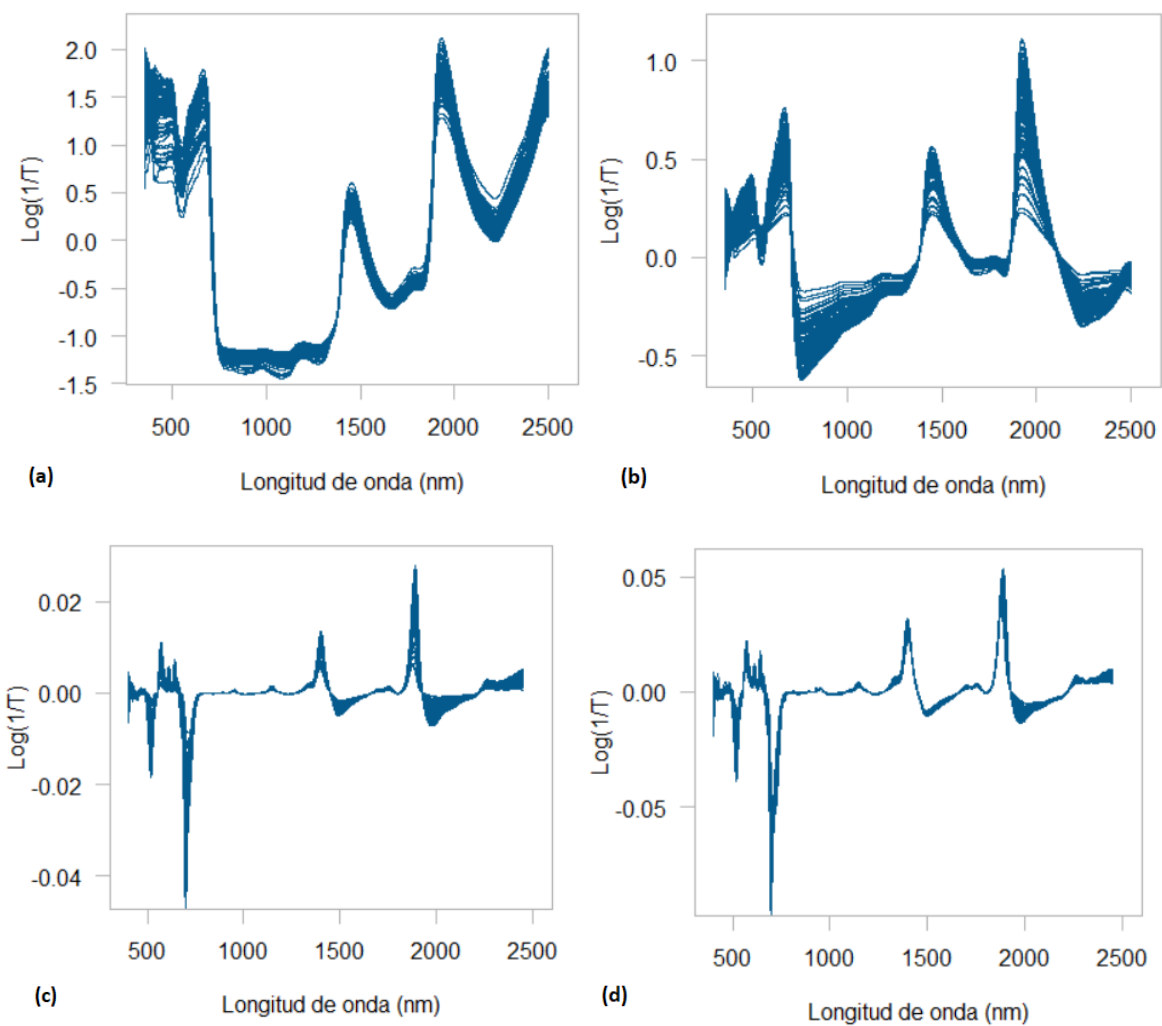


Figura B.2 Modificaciones en el comportamiento de los datos espectrales luego de aplicar diferentes pretratamientos al conjunto de datos del estrato medio. SNV (a), Detrend (b), Savitzky-Golay 1era derivada (c) y Savitzky-Golay + SNV (d).

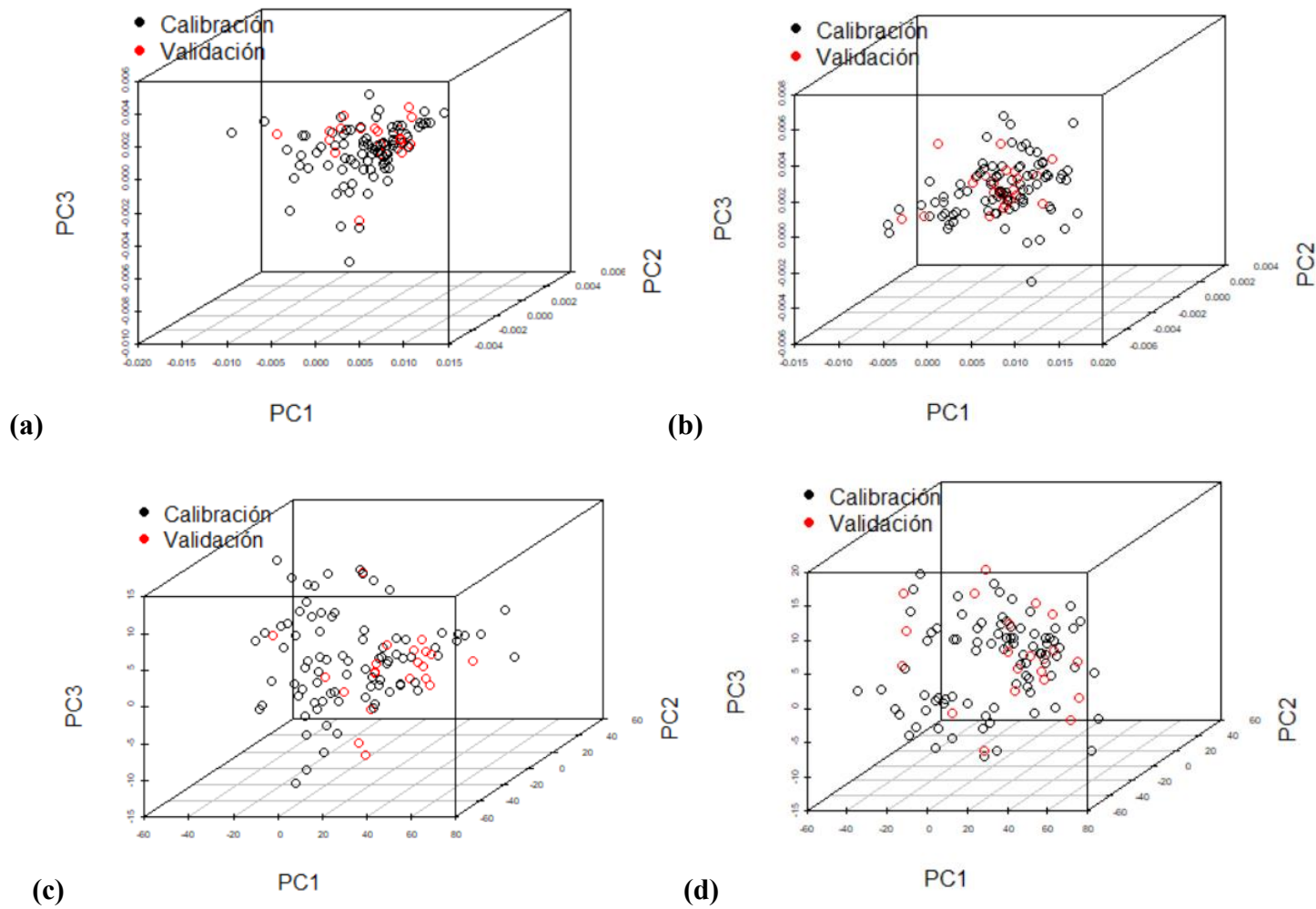


Figura B.2 Resultados del análisis de componentes principales (PCA) para la tasa de asimilación neta de CO₂ (A) representados con PC1, PC2 y PC3. Turrialba estrato alto (a), Victoria estrato alto (b), Turrialba estrato medio (c) y Victoria estrato medio (d).

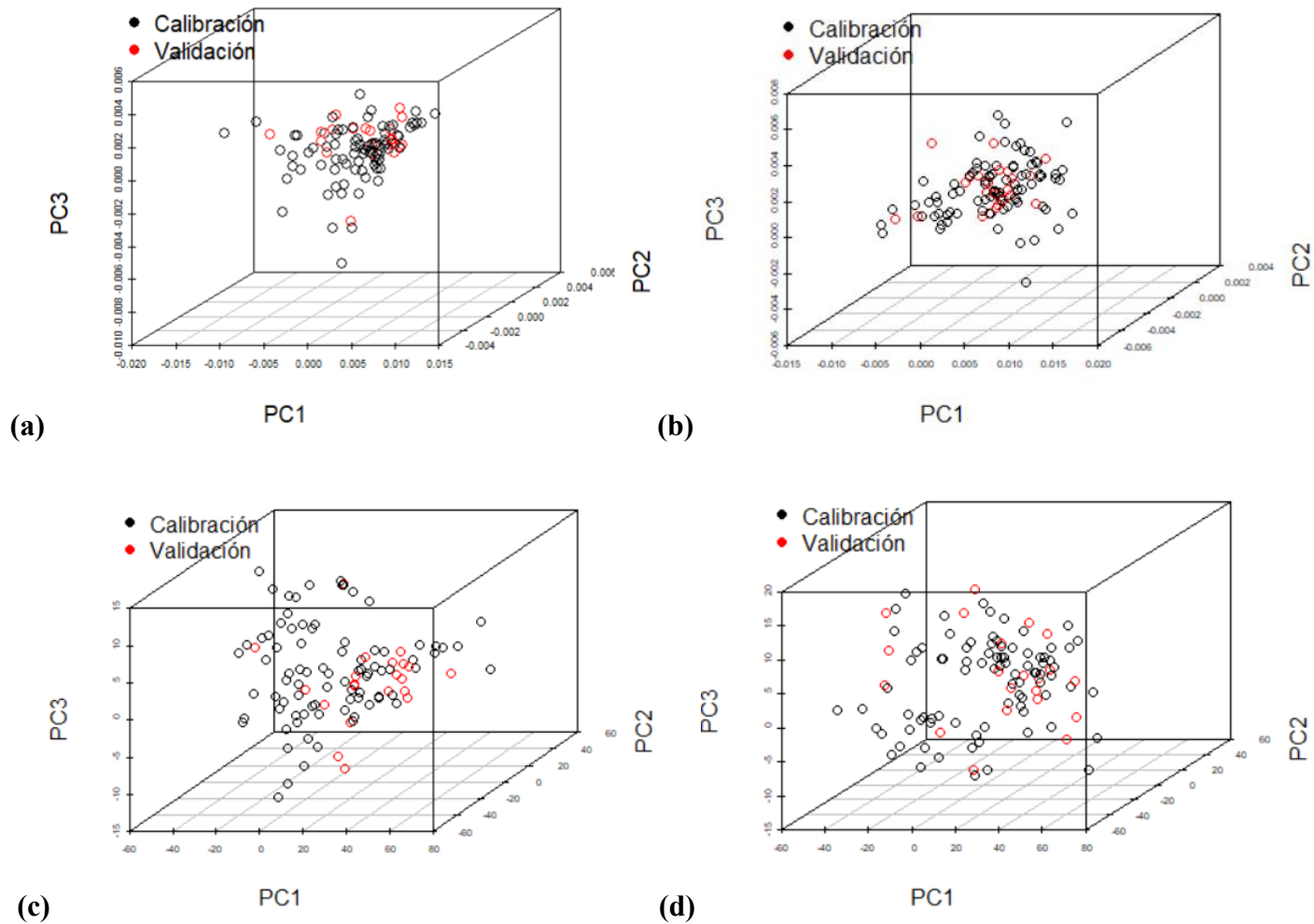


Figura B.3 Resultados del análisis de componentes principales (PCA) para la presión de vapor de saturación a la temperatura de la hoja (SVPleaf) representados con PC1, PC2 y PC3. Turrialba estrato alto (a), Victoria estrato alto (b), Turrialba estrato medio (c) y Victoria estrato medio (d).

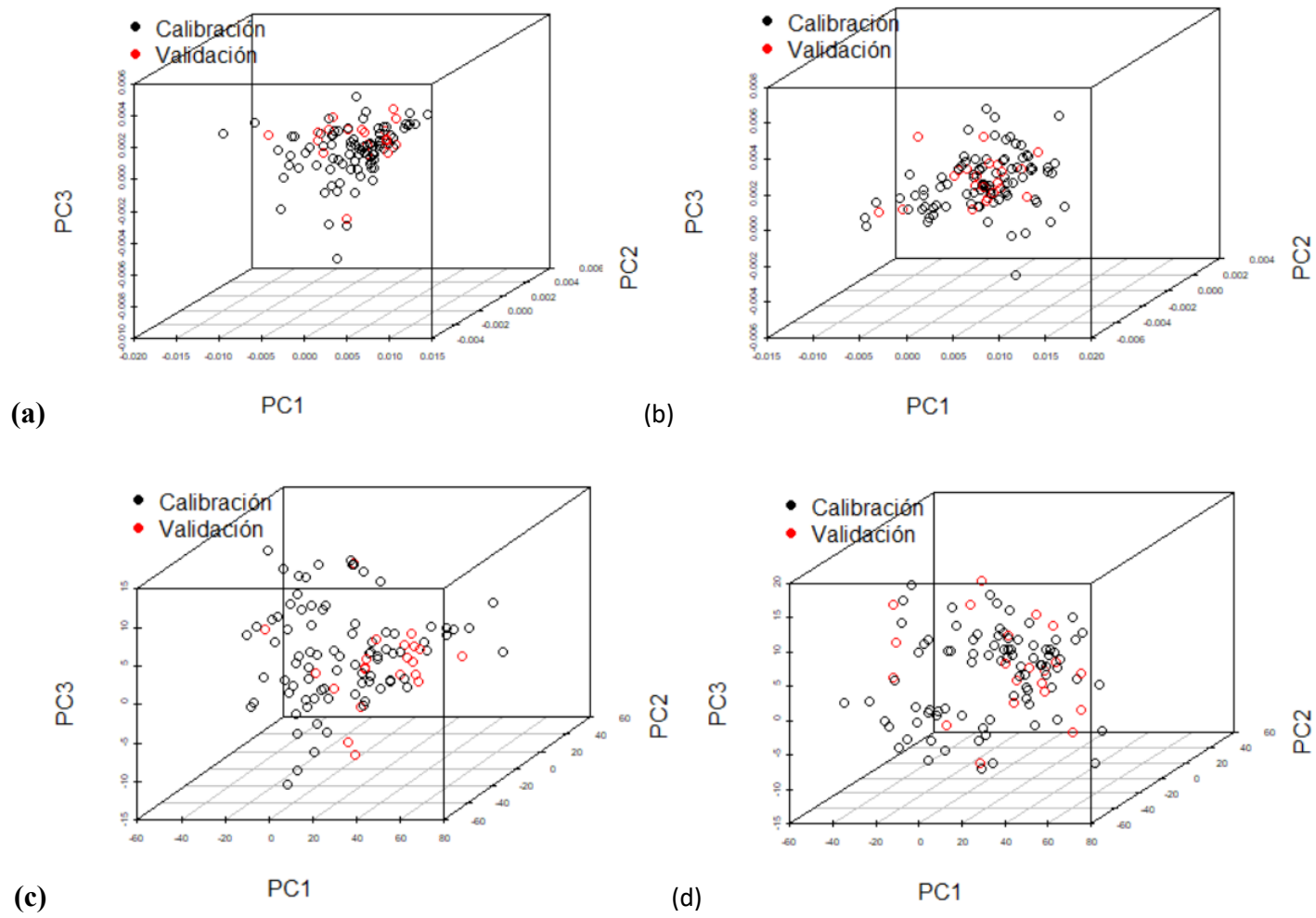


Figura B.4 Resultados del análisis de componentes principales (PCA) para temperatura de la hoja (Tleaf) representados con PC1, PC2 y PC3. Turrialba estrato alto (a), Victoria estrato alto (b), Turrialba estrato medio (c) y Victoria estrato medio (d).

Cuadro B.1 Porcentajes de varianza explicados para cada componente principal de cada modelo generado para la tasa de asimilación (A).

MODELO	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
5	67.30	15.64	8.84	2.25	1.67	1.02	0.71	0.54	0.51	0.34
6	66.36	14.67	11.15	1.96	1.74	1.04	0.67	0.53	0.44	0.32
7	53.36	22.95	12.06	5.32	1.41	1.10	0.93	0.61	0.41	0.41
8	53.26	22.30	12.55	5.71	1.42	1.09	0.88	0.56	0.43	0.37

Cuadro B.2 Porcentajes de varianza explicados para cada componente principal de cada modelo generado para la temperatura de la hoja (Tleaf).

MODELO	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
9	68.10	14.77	7.92	2.38	1.90	1.19	0.85	0.63	0.56	0.41
10	67.75	13.52	11.01	1.94	1.61	1.02	0.68	0.48	0.46	0.35
11	55.42	21.77	11.28	5.26	1.66	1.10	0.79	0.61	0.43	0.38
12	55.84	22.47	11.34	4.33	1.41	1.10	0.81	0.57	0.43	0.34

Cuadro B.3 Porcentajes de varianza explicados para cada componente principal de cada modelo generado para la presión de vapor de saturación a la temperatura de la hoja (SVPlleaf).

MODELO	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
13	69.60	14.72	7.66	2.03	1.71	1.14	0.65	0.59	0.42	0.36
14	69.58	13.34	9.58	1.86	1.53	1.01	0.69	0.53	0.45	0.35
15	53.11	22.80	13.10	4.83	1.53	1.16	0.74	0.57	0.40	0.40
16	58.95	18.42	12.09	4.58	1.45	1.13	0.80	0.58	0.38	0.34

Cuadro B.4 Resultados completos obtenidos de los indicadores para el modelo 1 y 2.

<i>Pretratamiento</i>	<i>MODELO 1</i>			<i>MODELO 2</i>		
	LV	RMSEP	R ^{2*}	LV	RMSEP	R ^{2*}
<i>no</i>	1	0.00075	-0.03072	0	0.00105	-0.03215
<i>snv</i>	0	0.00076	-0.04666	0	0.00105	-0.03215
<i>detrend</i>	1	0.00075	-0.03559	0	0.00105	-0.03215
<i>savgol</i>	1	0.00075	-0.03070	0	0.00105	-0.03215
<i>savgolsnv</i>	0	0.00076	-0.04666	0	0.00105	-0.03215
<i>savgol1</i>	0	0.00076	-0.04666	0	0.00105	-0.03215
<i>savgol1snv</i>	0	0.00076	-0.04666	0	0.00105	-0.03215
<i>snvsavgol1</i>	0	0.00076	-0.04666	0	0.00105	-0.03215
<i>savgol2</i>	0	0.00076	-0.04666	0	0.00105	-0.03215
<i>savgol2snv</i>	0	0.00076	-0.04666	0	0.00105	-0.03215
<i>snvsavgol2</i>	0	0.00076	-0.04666	0	0.00105	-0.03215

Cuadro B.5 Resultados completos obtenidos de los indicadores para el modelo 3 y 4.

<i>Pretratamiento</i>	<i>MODELO 3</i>			<i>MODELO 4</i>		
	LV	RMSEP	R ^{2*}	LV	RMSEP	R ^{2*}
<i>no</i>	0	0.00075	-0.03543	0	0.00105	-0.03773
<i>snv</i>	6	0.00073	0.02371	0	0.00105	-0.03773
<i>detrend</i>	3	0.00074	0.00431	0	0.00105	-0.03773
<i>savgol</i>	0	0.00075	-0.03543	0	0.00105	-0.03773
<i>savgolsnv</i>	6	0.00073	0.02375	0	0.00105	-0.03773
<i>savgol1</i>	4	0.00074	0.01338	0	0.00105	-0.03773
<i>savgol1snv</i>	0	0.00075	-0.03543	0	0.00105	-0.03773
<i>snvsavgol1</i>	3	0.00074	-0.00510	0	0.00105	-0.03773
<i>savgol2</i>	4	0.00074	0.01651	0	0.00105	-0.03773
<i>savgol2snv</i>	4	0.00074	0.00016	0	0.00105	-0.03773
<i>snvsavgol2</i>	4	0.00073	0.02928	0	0.00105	-0.03773

Cuadro B.6 Resultados completos obtenidos de los indicadores para el modelo 5 y 6.

<i>Pretratamiento</i>	<i>MODELO 5</i>			<i>MODELO 6</i>		
	LV	RMSEP	R ^{2*}	LV	RMSEP	R ^{2*}
<i>no</i>	1	2.24694	-0.02574	0	2.53691	-0.05359
<i>snv</i>	0	2.26568	-0.04293	0	2.53691	-0.05359
<i>detrend</i>	1	2.26352	-0.04093	0	2.53691	-0.05359
<i>savgol</i>	1	2.24691	-0.02572	0	2.53691	-0.05359
<i>savgolsnv</i>	0	2.26568	-0.04293	0	2.53691	-0.05359
<i>savgol1</i>	0	2.26568	-0.04293	0	2.53691	-0.05359
<i>savgol1snv</i>	0	2.26568	-0.04293	0	2.53691	-0.05359
<i>snvsavgol1</i>	0	2.26568	-0.04293	0	2.53691	-0.05359
<i>savgol2</i>	0	2.26568	-0.04293	0	2.53691	-0.05359
<i>savgol2snv</i>	0	2.26568	-0.04293	0	2.53691	-0.05359
<i>snvsavgol2</i>	0	2.26568	-0.04293	0	2.53691	-0.05359

Cuadro B.7 Resultados completos obtenidos de los indicadores para el modelo 7 y 8.

<i>Pretratamiento</i>	<i>MODELO 7</i>			<i>MODELO 8</i>		
	LV	RMSEP	R ^{2*}	LV	RMSEP	R ^{2*}
<i>no</i>	0	2.23963	-0.01908	0	2.5062	-0.02824
<i>snv</i>	7	2.23848	-0.01804	0	2.5062	-0.02824
<i>detrend</i>	7	2.23639	-0.01613	0	2.5062	-0.02824
<i>savgol</i>	0	2.23963	-0.01908	0	2.5062	-0.02824
<i>savgolsnv</i>	7	2.23962	-0.01907	0	2.5062	-0.02824
<i>savgol1</i>	0	2.23963	-0.01908	0	2.5062	-0.02824
<i>savgol1snv</i>	0	2.23963	-0.01908	0	2.5062	-0.02824
<i>snvsavgol1</i>	4	2.23718	-0.01685	0	2.5062	-0.02824
<i>savgol2</i>	4	2.19315	0.02278	0	2.5062	-0.02824
<i>savgol2snv</i>	4	2.21197	0.00594	0	2.5062	-0.02824

<i>snvsavgol2</i>	4	2.18112	0.03347
-------------------	----------	----------------	----------------

Cuadro B.8 Resultados completos obtenidos de los indicadores para el modelo 9 y 10.

<i>Pretratamiento</i>	<i>MODELO 9</i>			<i>MODELO 10</i>		
	LV	RMSEP	R ² *	LV	RMSEP	R ² *
<i>no</i>	2	0.44990	0.01741	0	0.50247	-0.02060
<i>snv</i>	1	0.45841	-0.02012	3	0.49599	0.00556
<i>detrend</i>	0	0.46484	-0.04894	3	0.49497	0.00965
<i>savgol</i>	2	0.44977	0.01795	0	0.50247	-0.02060
<i>savgolsnv</i>	1	0.45825	-0.01944	3	0.49581	0.00628
<i>savgoll</i>	0	0.46484	-0.04894	0	0.50247	-0.02060
<i>savgollsnv</i>	0	0.46484	-0.04894	3	0.50007	-0.01088
<i>snvsavgoll</i>	0	0.46484	-0.04894	0	0.50247	-0.02060
<i>savgol2</i>	0	0.46484	-0.04894	0	0.50247	-0.02060
<i>savgol2snv</i>	0	0.46484	-0.04894	0	0.50247	-0.02060
<i>snvsavgol2</i>	0	0.46484	-0.04894	0	0.50247	-0.02060

Cuadro B.9 Resultados completos obtenidos de los indicadores para el modelo 11 y 12.

<i>Pretratamiento</i>	<i>MODELO 11</i>			<i>MODELO 12</i>		
	LV	RMSEP	R ² *	LV	RMSEP	R ² *
<i>no</i>	7	0.42382	0.12803	0	0.49987	-0.01006
<i>snv</i>	2	0.43881	0.06524	2	0.49212	0.02082
<i>detrend</i>	6	0.41589	0.16035	0	0.49987	-0.01006
<i>savgol</i>	7	0.42325	0.13037	0	0.49987	-0.01006
<i>savgolsnv</i>	2	0.43871	0.06565	2	0.49207	0.02100
<i>savgoll</i>	13	0.40768	0.19315	0	0.49987	-0.01006
<i>savgollsnv</i>	13	0.43168	0.09538	3	0.49343	0.01578
<i>snvsavgoll</i>	13	0.43893	0.06473	3	0.49279	0.01836
<i>savgol2</i>	16	0.35505	0.38802	0	0.49987	-0.01006

<i>savgol2snv</i>	14	0.36707	0.34591	3	0.49688	0.00197
<i>snvsavgol2</i>	15	0.37450	0.31915	0	0.49987	-0.01006

Cuadro B.10 Resultados completos obtenidos de los indicadores para el modelo 13 y 14.

<i>Pretratamiento</i>	<i>MODELO 13</i>			<i>MODELO 14</i>		
	LV	RMSEP	R ^{2*}	LV	RMSEP	R ^{2*}
<i>no</i>	2	1.91393	0.01970	0	2.17286	-0.02354
<i>snv</i>	1	1.94106	-0.00829	2	2.15988	-0.01134
<i>detrend</i>	0	1.96079	-0.02889	3	2.15092	-0.00297
<i>savgol</i>	2	1.91362	0.02001	0	2.17286	-0.02354
<i>savgol2snv</i>	1	1.94040	-0.00760	2	2.15977	-0.01124
<i>savgoll</i>	0	1.96079	-0.02889	0	2.17286	-0.02354
<i>savgollsnv</i>	0	1.96079	-0.02889	0	2.17286	-0.02354
<i>snvsavgoll</i>	0	1.96079	-0.02889	0	2.17286	-0.02354
<i>savgol2</i>	0	1.96079	-0.02889	0	2.17286	-0.02354
<i>savgol2snv</i>	0	1.96079	-0.02889	0	2.17286	-0.02354
<i>snvsavgol2</i>	0	1.96079	-0.02889	0	2.17286	-0.02354

Cuadro B.11 Resultados completos obtenidos de los indicadores para el modelo 15 y 16.

<i>Pretratamiento</i>	<i>MODELO 15</i>			<i>MODELO 16</i>		
	LV	RMSEP	R ^{2*}	LV	RMSEP	R ^{2*}
<i>no</i>	8	1.79536	0.13739	7	2.14624	0.00139
<i>snv</i>	4	1.84739	0.08667	2	2.12431	0.02169
<i>detrend</i>	7	1.76389	0.16737	2	2.18280	-0.03292
<i>savgol</i>	8	1.79409	0.13834	7	2.14583	0.00177
<i>savgol2snv</i>	4	1.84816	0.08590	2	2.12396	0.02202
<i>savgoll</i>	15	1.69032	0.23276	6	2.16131	-0.01268
<i>savgollsnv</i>	13	1.83015	0.09914	3	2.12746	0.01879
<i>snvsavgoll</i>	14	1.89231	0.03988	3	2.14916	-0.00132

<i>savgol2</i>	17	1.53460	0.36977	3	2.18183	-0.03200
<i>savgol2snv</i>	15	1.59370	0.32029	3	2.16132	-0.01269
<i>snvsavgol2</i>	16	1.60346	0.31195	2	2.17402	-0.02463

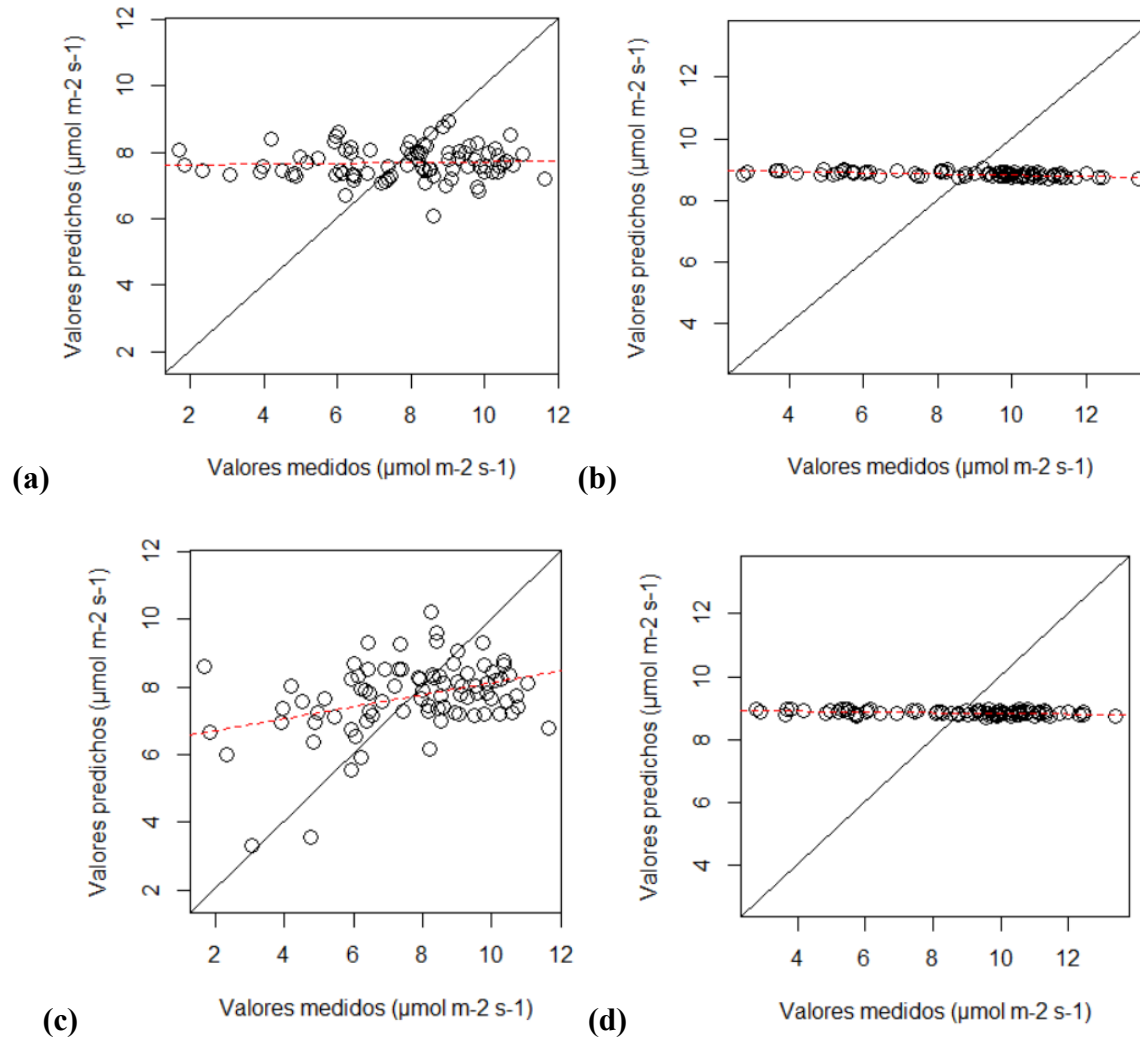


Figura B.5. Curvas de datos predichos (eje y) vs datos medidos de la variable A (eje x) para el modelo 5(a), modelo 6(b), modelo 7(c) y modelo 8(d).

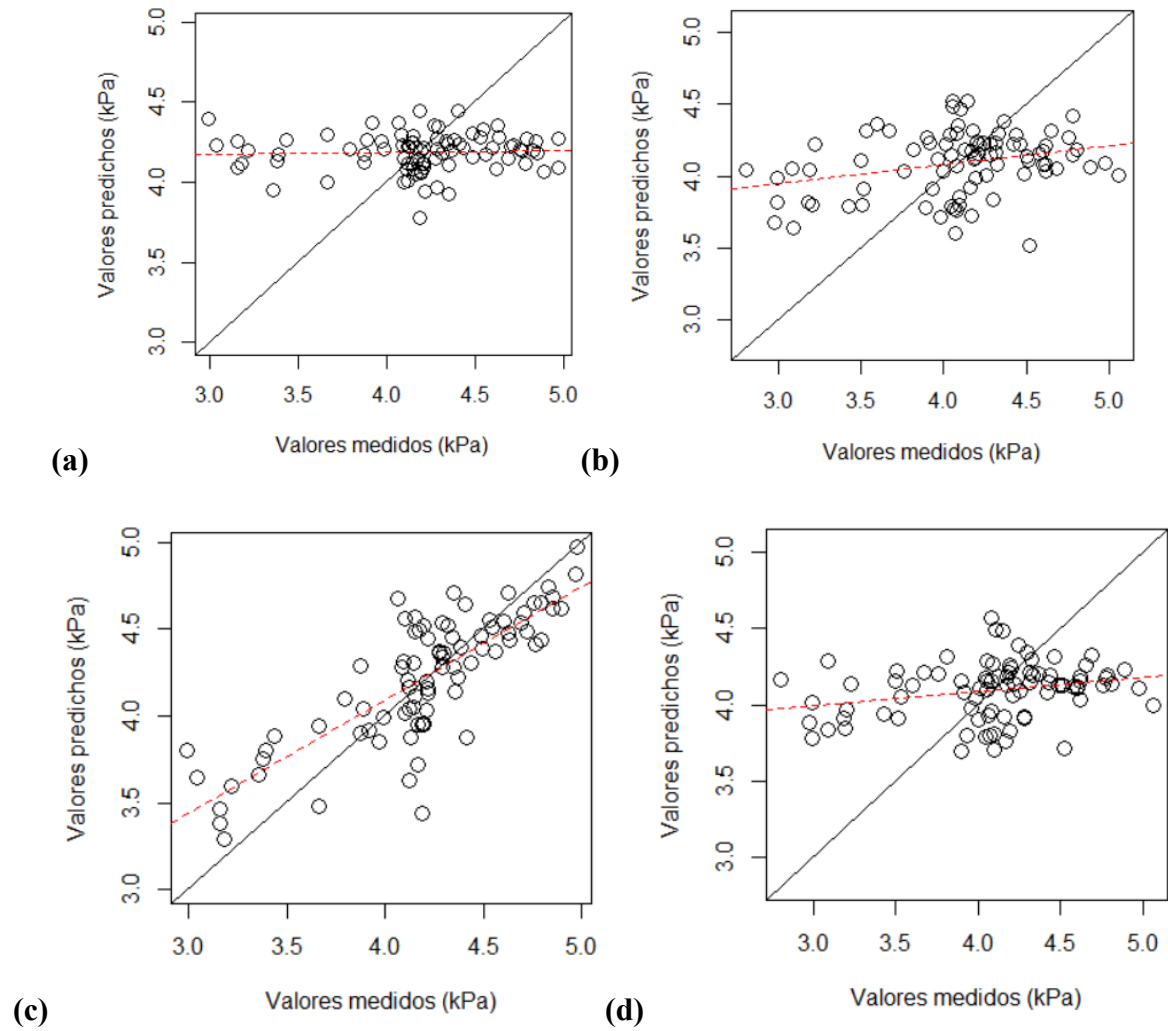


Figura B.6. Curvas de datos predichos (eje y) vs datos medidos de la variable SVleaf (eje x) para el modelo 9(a), modelo 10(b), modelo 11(c) y modelo 12(d).

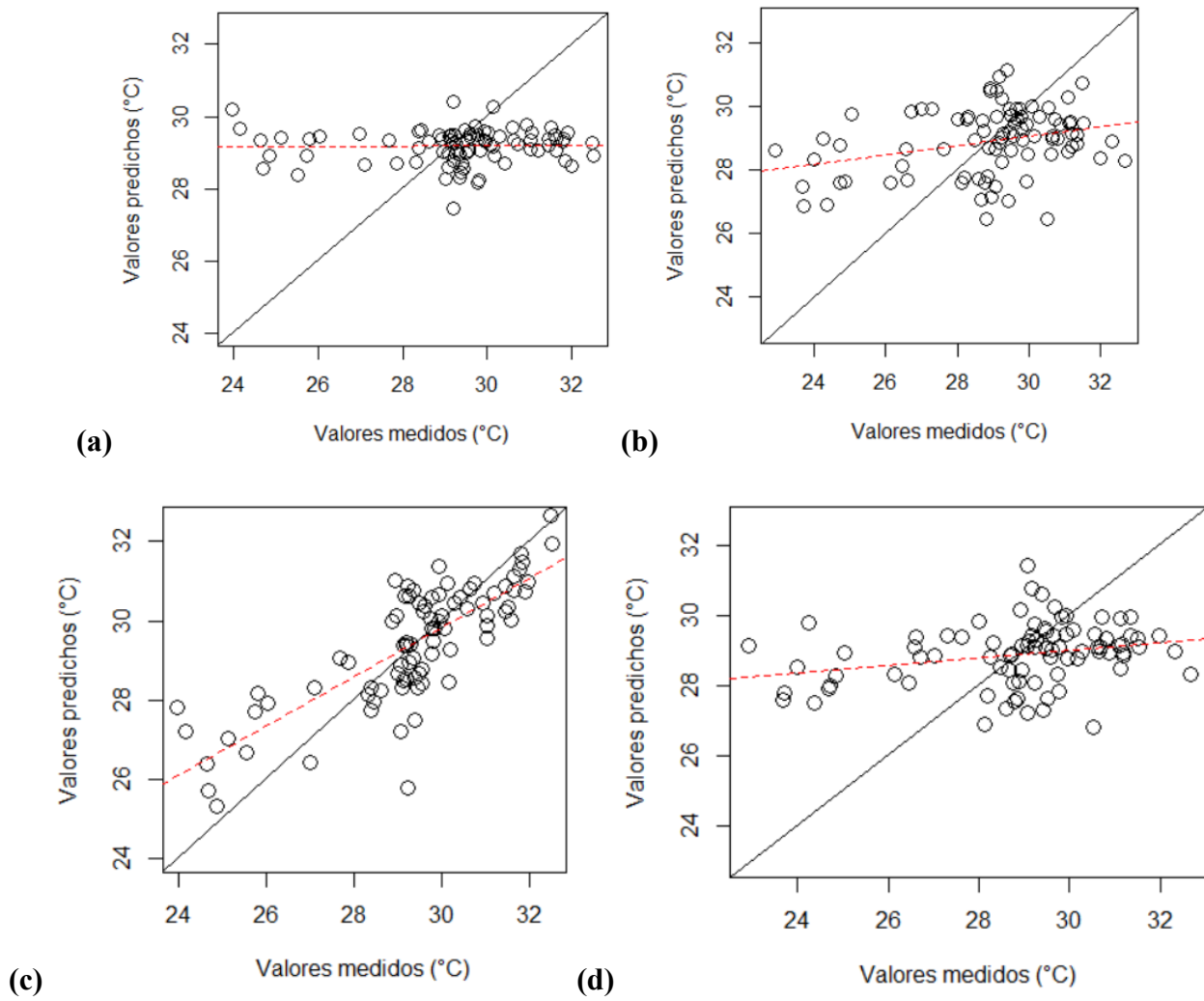


Figura B.7. Curvas de datos predichos (eje y) vs datos medidos de la variable Tleaf (eje x) para el modelo 9(a), modelo 10(b), modelo 11(c) y modelo 12(d).

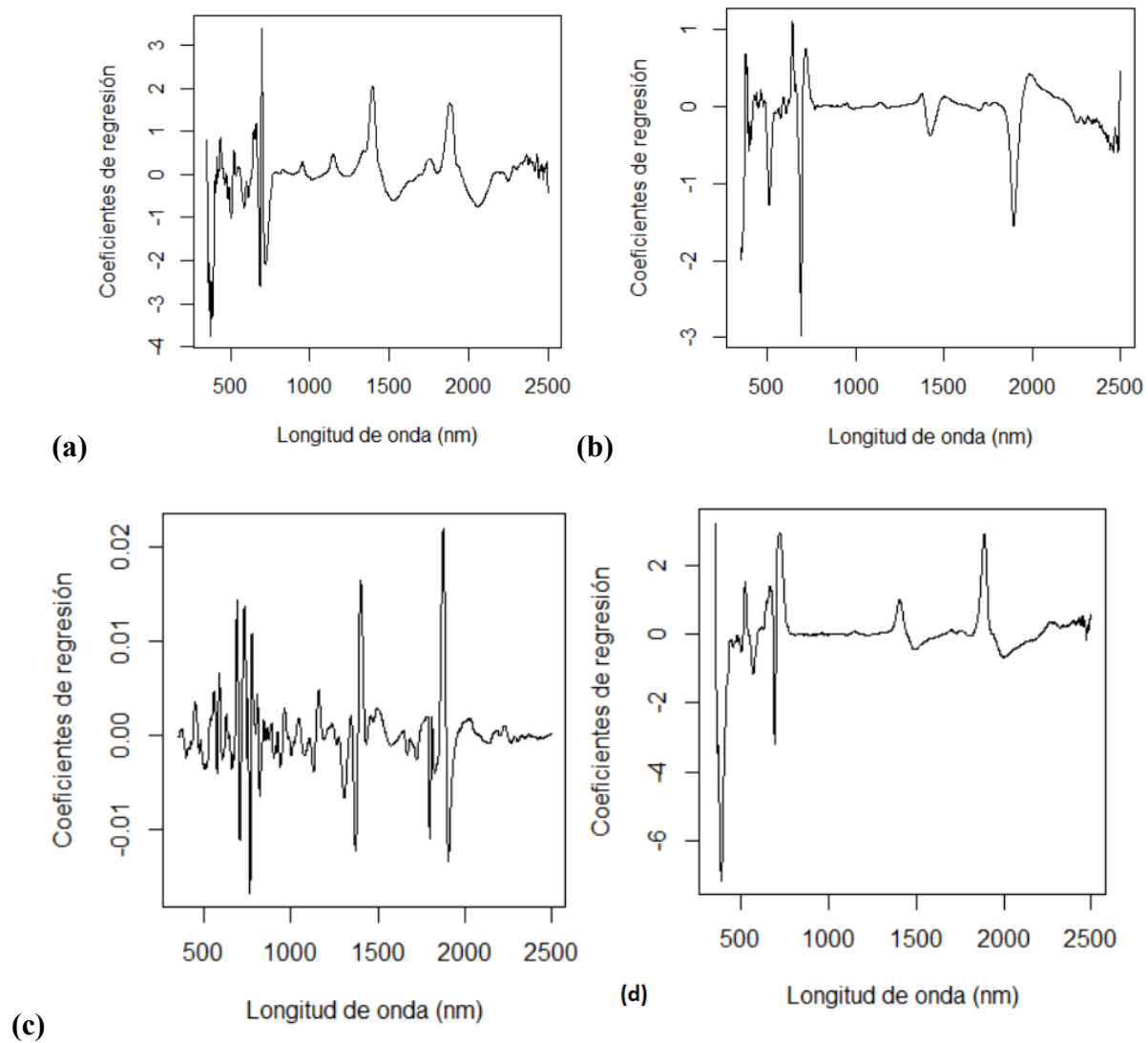
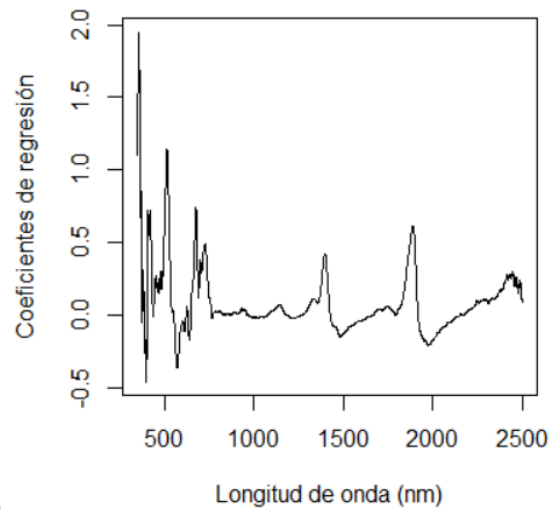
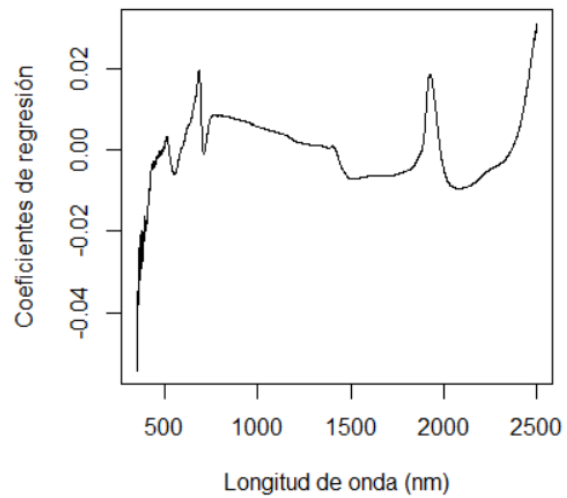


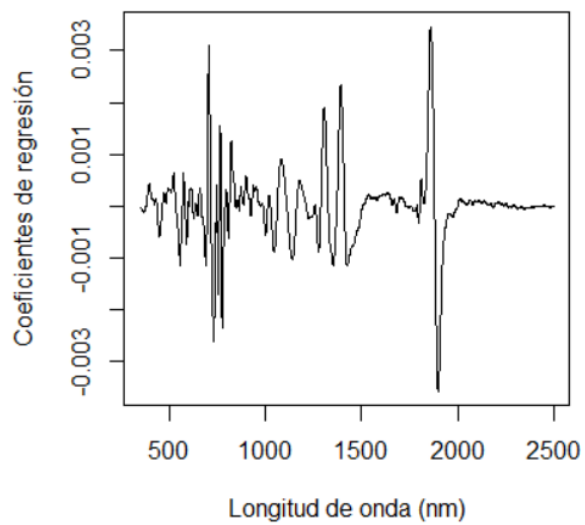
Figura B.8. Coeficientes de regresión vs longitudes de onda (nm) más importantes de los modelos 5(a), 6(b), 7(c) y 8(d).



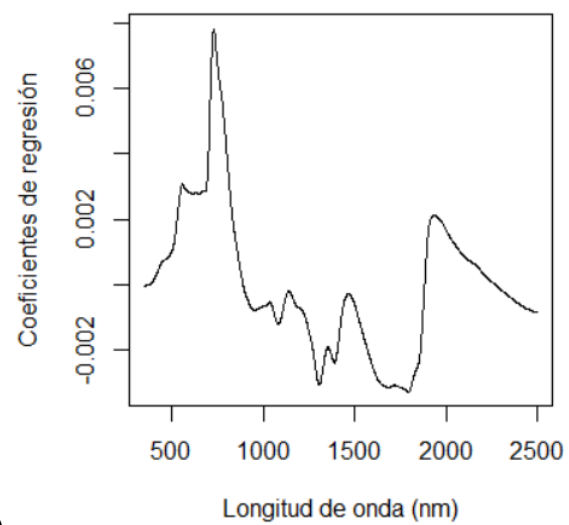
(a)



(b)



(c)



(d)

Figura B.9. Coeficientes de regresión vs longitudes de onda (nm) más importantes de los modelos 9(a), 10(b), 11(c) y 12(d).

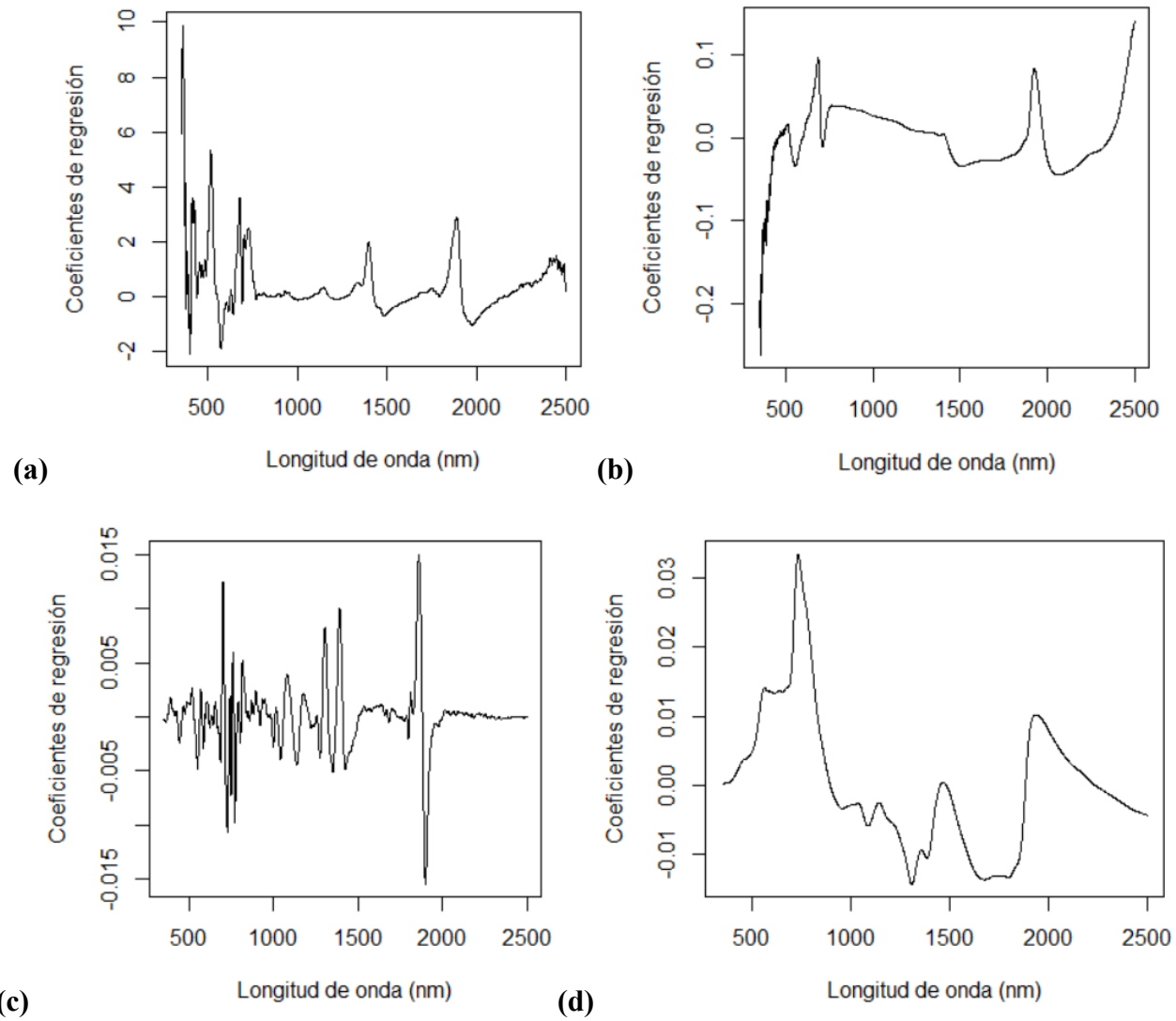


Figura B.10. Coeficientes de regresión vs longitudes de onda (nm) más importantes de los modelos 13(a), 14(b), 15(c) y 16(d).

APÉNDICE C:
SECCIÓN DE CÓDIGO UTILIZADO

Preparación de entorno

```
install.packages(remotes)

install.packages("scatterplot3d")

library(remotes)

remotes::install_github("mlesnoff/rchemo", dependencies = TRUE, build_vignettes =
TRUE)

remotes::install_github("mlesnoff/rnirs", dependencies = TRUE, build_vignettes = TRUE)

library(scatterplot3d)

library(readxl)

library(rnirs)

library(rchemo)

library(readr)

setwd("C:/Users/Admin/Desktop/Gilles/For_andres") ### change for you

data1 <- as.data.frame(read_xlsx(path="C:/Users/Admin/Desktop/xdata.xlsx"))

data2 <- as.data.frame(read_xlsx(path="C:/Users/Admin/Desktop/ydata_old.xlsx"))
```

Preparación de los datos espectrales

```
spectre <- data1[,2:2152]

col_namevisnir <- colnames(spectre)

plotsp(spectre, xlab="Longitud de onda (nm)", ylab= "Absorbancia")

indexcol <- which(colnames(spectre) == "1000" | colnames(spectre) == "1800" )

spectre <- rmgap(spectre, indexcol)

plotsp(spectre, xlab="Longitud de onda (nm)", ylab= "log(1/T)")
```



```
Xmean_sample <- aggregate(x=spectre, by=list(data2$rep), FUN = "mean")
```

```
Xmean <- Xmean_sample[,2:2152]
```

```
plotsp(Xmean, xlab="Longitud de onda (nm)", ylab= "Log(1/T)")
```

Preparación de los datos de laboratorio

```
ymean_sample1 <- data2[1,]
```

```
ymean_sample2 <- data2[4,]
```

```
ymean_sample3 <- data2[7,]
```

```
ymean_sample4 <- data2[13:114,]
```

```
ymean_sample4 <- ymean_sample4[seq(from=1,to=nrow(ymean_sample4),by=3),]
```

```
ymean_sample5 <- data2[115,]
```

```
ymean_sample6 <- data2[121:648,]
```

```
ymean_sample6 <- ymean_sample6[seq(from=1,to=nrow(ymean_sample6),by=3),]
```

```
ymean <-
```

```
  rbind(ymean_sample1,ymean_sample2,ymean_sample3,ymean_sample4,ymean_sample5,y  
  mean_sample6)
```

Selección de los diferentes tratamientos y variables dependientes

```
spectre_T <- subset(Xmean,ymean$TV=="T")
```

```
y_1 <- ymean[ymean$TV=="T",7]
```

```
hist(y_1,10,xlab="Tasa de transpiración (mol m-2 s-1)",main="")
```

```
y_2 <- ymean[ymean$TV=="T",8]
```

```
hist(y_2,10,xlab="Tasa de asimilación de CO2 ( $\mu\text{mol m}^{-2} \text{s}^{-1}$ )",main="")
```

```
y_3 <- ymean[ymean$TV=="T",9]
```

```
hist(y_3,10,xlab="Presión de vapor de saturación a la temperatura de la hoja  
(kPa)",main="")
```

```
y_4 <- ymean[ymean$TV=="T",10]
```

```
hist(y_4,10,xlab="Temperatura de la hoja en el balance de energía ( $^{\circ}\text{C}$ )",main="")
```

```
y <- y_1
```

```
yord <- y[order(y,decreasing=F)]# Classer y en ordre croissant
```

```
Xmean_ord <- Xmean[order(y,decreasing=F),]
```

```
sel <- seq(from=2,to=nrow(Xmean_ord),by=5)# a partir du 2ème prendre 1 sur 5 c'st dire  
20%
```

```
max(sel) # v?rifier que l'on a PAS pris le dernier
```

División del set de datos

```
xcal <- Xmean_ord[-sel,]
```

```
xval <- Xmean_ord[sel,]
```

```
yca <- yord[-sel]
```

```
yval <- yord[sel]
```

```
hist(yval, main = "", xlab = 'yval', ylab = 'Frecuencia')
```

```
hist(yval, main = "", xlab = 'yval', ylab = 'Frecuencia')
```

```
zxcal <- rnirs::savgol(xcal,n=25,p=3,m=2)
```

```
zxval <- rnirs::savgol(xval,n=25,p=3,m=2)
```

```
fmpca <- pcasvd(zxcal,nlv=10)
```

```
loading <- fmpca$P
```

```
scores <- fmpca$T
```

```
eigenvalues <- fmpca$eig
```

```
porcentajes_varianza_explicada <- (eigenvalues / sum(eigenvalues)) * 100
```

```
primeros10 <- porcentajes_varianza_explicada[1:10]
```

```
print(primeros10)
```

```
Tcal <- fmpca$T
```

```
Tval <- transform(fmpca, zxval)
```

```
T <- rbind( Tcal,Tval)
```

```
col <- c(rep(rgb(0,0,0), nrow(xcal)), rep(rgb(1,0,0), nrow(xval)))
scatterplot3d(T[,1], T[,2], T[,3], xlab = "PC1", ylab = "PC2", zlab = "PC3", color = col,
              cex.axis = 0.45)
legend("topleft",c("Calibración","Validación"),pch=c(16,16),col=c(rgb(0,0,0),rgb(1,0,0)),bt
y="n")
```

Validación Cruzada

```
K <- 2 # number of group
nrep <- 20 # number of repetitions
segmkfold <- segmkf(n = nrow(xcal), K = K, nrep = nrep)
ncomp <- 30
Xr <- xcal
yr <- ycal
hist(yr)
```

```
wavemin <- 400
wavemax <- 2450
```

Aplicación de pretratamientos a los modelos

```
zXr<-Xr
fmcv_no <- cvfit(zXr,yr,fun=plsr,ncomp=ncomp,segm=segmkfold,print=FALSE)
z_no <- rnirs::mse(fmcv_no, ~ ncomp,digits=5)
z_no[z_no$rmsep == min(z_no$rmsep), ]
```

```

zXr <- snv(Xr)

plotsp(zXr, xlab="Longitud de onda (nm)" , ylab= "Log(1/T)")

fmcv_snv <- cvfit(zXr,yr,fun=plsr,ncomp=ncomp,segm=segmkfold,print=FALSE)

z_snv <- rnirs::mse(fmcv_snv, ~ ncomp,digits=5)

z_snv[z_snv$rmsep == min(z_snv$rmsep), ]

zXr <- rnirs::detrend(Xr,method ="poly",degree=2)

plotsp(zXr, xlab="Longitud de onda (nm)" , ylab= "Log(1/T)")

fmcv_detrend <- cvfit(zXr,yr,fun=plsr,ncomp=ncomp,segm=segmkfold,print=FALSE)

z_detrend <- rnirs::mse(fmcv_detrend, ~ ncomp,digits=5)

z_detrend[z_detrend$rmsep == min(z_detrend$rmsep), ]

zXr <- rnirs::savgol(Xr,n=15,p=1,m=0)

plotsp(zXr, xlab="Longitud de onda (nm)" , ylab= "Log(1/T)")

fmcv_savgol <- cvfit(zXr,yr,fun=plsr,ncomp=ncomp,segm=segmkfold,print=FALSE)

z_savgol <- rnirs::mse(fmcv_savgol, ~ ncomp,digits=5)

z_savgol[z_savgol$rmsep == min(z_savgol$rmsep), ]

```