



UNIVERSIDAD
DE PIURA

FACULTAD DE INGENIERÍA

**Análisis e identificación de las variables de mayor
influencia en el desarrollo y crecimiento del banano
orgánico en la Región Piura**

Tesis para optar el Título de
Ingeniero Industrial y de Sistemas

Wilson David Acuña Murrieta

Asesor:
Dra. Ing. Ana Valeria Quevedo Candela

Piura, junio de 2023

Declaración Jurada de Originalidad del Trabajo Final

Yo, Wilson David Acuña Murrieta., egresado del Programa Académico/ programa de posgrado de Ingeniería Industrial y de Sistemas de la Facultad de Ingeniería de la Universidad de Piura, identificado(a) con DNI 71475572

Declaro bajo juramento que:

1. Soy autor del trabajo final titulado:
"Análisis e identificación de las variables de mayor influencia en el desarrollo y crecimiento del banano orgánico en la Región Piura"
El mismo que presento bajo la modalidad de **Tesis**.¹ para optar el (Título profesional/Grado Académico²) de Título Profesional.
2. Que el trabajo se realizó en coautoría con los siguientes alumnos de la Universidad de Piura.
 - Haga clic o pulse aquí para escribir texto. _____, identificado con DNI N° Escribir número
 - Haga clic o pulse aquí para escribir texto. _____, identificado con DNI N° Escribir número
 - Haga clic o pulse aquí para escribir texto. _____, identificado con DNI N° Escribir número
 - Haga clic o pulse aquí para escribir texto. _____, identificado con DNI N° Escribir número
 - Haga clic o pulse aquí para escribir texto. _____, identificado con DNI N° Escribir número
3. La asesoría del trabajo estuvo a cargo de:
 - Dra. Ing. Ana Valeria Quevedo Candela _____, identificado con DNI N° 40569732
 - Haga clic o pulse aquí para escribir texto. _____, identificado con DNI N° Escribir número
4. El texto de mi trabajo final respeta y no vulnera los derechos de terceros o de ser el caso derechos de los coautores, incluidos los derechos de propiedad intelectual, datos personales, entre otros. En tal sentido, el texto de mi trabajo final no ha sido plagiado total ni parcialmente, para la cual he respetado las normas internacionales de citas y referencias de las fuentes consultadas.
5. El texto del trabajo final que presento no ha sido publicado ni presentado antes en cualquier medio electrónico o físico.
6. La investigación, los resultados, datos, conclusiones y demás información presentada que atribuyo a mi autoría son veraces.
7. Declaro que mi trabajo final cumple con todas las normas de la Universidad de Piura.

El incumplimiento de lo declarado da lugar a responsabilidad del declarante, en consecuencia; a través del presente documento asumo frente a terceros, la Universidad de Piura y/o la Administración Pública toda responsabilidad que pueda derivarse por el trabajo final presentado. Lo señalado incluye responsabilidad pecuniaria incluido el pago de multas u otros por los daños y perjuicios que se ocasionen.

Fecha: 24/07/2023



.....
Firma del autor optante³

Dedicatoria

A Dios, a mi madre
Yolanda, a quien debo
todos mis éxitos y
aciertos; a mis abuelos
Nora y Emérito, por
entregarme siempre más
de lo que debían; y a mi
asesora de tesis, Valeria,
por todo el apoyo
brindado a lo largo de
este proyecto.





Resumen

El presente trabajo de investigación busca identificar cuáles son las principales variables que inciden y condicionan el desarrollo de una planta de banano orgánico en la Región Piura. Para ello, se realizó un análisis estadístico y matemático a la conductividad eléctrica, al contenido volumétrico de agua y a la temperatura del suelo de dos microfundos en el Valle del Chira, con el fin de desarrollar modelos matemáticos que permitieran identificar dichas variables.

Todo el contenido del trabajo se ha dispuesto en cuatro capítulos: el primero de ellos recoge información sobre el banano orgánico, descripción, origen y formas de producción; y sobre la clasificación que se utiliza en la actualidad para los suelos, así como algunas de sus características.

El capítulo dos recoge toda la metodología aplicada para el desarrollo de la investigación, desde el diseño experimental desplegado para la captura de datos hasta las herramientas estadísticas utilizadas para el análisis de toda la data obtenida. Recoge, así mismo, la justificación que marca el punto de inicio del desarrollo del experimento.

El capítulo tres aborda el análisis aplicado a toda la data obtenida, tanto aquella que se obtuvo de la planta misma de banano orgánico (altura, grosor del pseudotallo y número de hojas emitidas), así como aquella obtenida de los suelos de las parcelas de los microfundos.

El capítulo cuatro, finalmente, presenta los modelos lineales desarrollados a partir del capítulo previo, así como su análisis y posterior diagnóstico. En consecuencia, en este capítulo se presentan aquellas variables que condicionan el desarrollo de una planta de banano en la Región Piura.



Tabla de contenido

Introducción	13
Capítulo 1 Marco teórico.....	15
1.1 Banano orgánico.....	15
1.1.1 Planta de banano orgánico.....	16
1.1.2 Cultivo y cosecha de banano orgánico en la región	17
1.1.3 Indicadores de producción de banano	18
1.2 Suelo	19
1.2.1 Variables del suelo que influyen en el crecimiento de las plantas.....	20
Capítulo 2 Metodología.....	23
2.1 Justificación	23
2.2 Diseño experimental	24
2.2.1 Disposición y características de parcelas piloto	25
2.2.2 Medición de variables biométricas.....	26
2.2.3 Medición de variables edáficas	28
2.3 Limpieza y análisis de datos	28
2.4 Modelamiento.....	30
Capítulo 3 Análisis de datos.....	33
3.1 Limpieza de datos.....	35
3.2 Análisis estadístico	36
3.2.1 Análisis de variables biométricas.....	37
3.2.2 Análisis de correlación	47
Capítulo 4 Resultados	59
4.1 Altura	61

4.2 Circunferencia	64
4.3 TEFD	66
Conclusiones.....	69
Recomendaciones.....	71
Referencias bibliográficas.....	73
Apéndices	77
Apéndice A Gráficas de medias para cada variable biométrica.....	79
Apéndice B Análisis de la depuración de los modelos de regresión.....	86



Lista de tablas

Tabla 1. Parcelas de estudio	26
Tabla 2. Cronograma de medición de variables biométricas por tipo de planta.....	28
Tabla 3. P-valor para las correlaciones entre las variables biométricas de las plantas hijas	49
Tabla 4. P-valor para las correlaciones entre las variables biométricas de las plantas nietas	49
Tabla 5. P-valor para la correlación de cada combinación de variables edáficas para las plantas hijas.....	50
Tabla 6. P-valor para la correlación de cada combinación de variables edáficas para las plantas nietas.....	51
Tabla 7. P-valor para la correlación de cada combinación de la conductividad eléctrica del suelo con las variables biométricas de las plantas hijas.....	53
Tabla 8. P-valor para la correlación de cada combinación de la conductividad eléctrica del suelo con las variables biométricas de las plantas nietas	53
Tabla 9. P-valor para la correlación de cada combinación de la humedad del suelo con las variables biométricas de las plantas hijas.....	55
Tabla 10. P-valor para la correlación de cada combinación de la humedad del suelo con las variables biométricas de las plantas nietas.....	55
Tabla 11. P-valor para la correlación de cada combinación de la humedad del suelo con las variables biométricas de las plantas nietas.....	57
Tabla 12. P-valor para la correlación de cada combinación de la humedad del suelo con las variables biométricas de las plantas nietas.....	57
Tabla 13. Resumen del modelo de regresión lineal mixto para predecir la altura de las plantas nietas.....	61
Tabla 14. Resumen del modelo de regresión lineal mixto para predecir la circunferencia de las plantas nietas	64



Lista de figuras

Figura 1. Planta de banano y sus partes.....	17
Figura 2. Triángulo de clases texturales del suelo.....	20
Figura 3. Partes de un diagrama de caja.....	29
Figura 4. Evolución de la altura de las plantas hijas y plantas nietas en el tiempo	39
Figura 5. Evolución de la altura de las plantas hijas y plantas nietas en el tiempo por tipo de planta y tipo de suelo	39
Figura 6. Evolución de la altura de las plantas hijas y plantas nietas en el tiempo por tipo de planta y tipo de riego.....	40
Figura 7. Evolución de la altura de las plantas hijas y plantas nietas en el tiempo por tipo de parcela	40
Figura 8. Curva de crecimiento en altura por tipo de planta, construida en base a las medias	41
Figura 9. Evolución de la circunferencia del pseudotallo de las plantas hijas y plantas nietas en el tiempo.....	42
Figura 10. Evolución de la circunferencia del pseudotallo de las plantas hijas y plantas nietas en el tiempo por tipo de planta y tipo de suelo	42
Figura 11. Evolución de la circunferencia del pseudotallo de las plantas hijas y plantas nietas en el tiempo por tipo de riego.....	43
Figura 12. Evolución de la circunferencia del pseudotallo de las plantas hijas y plantas nietas en el tiempo por parcela.....	43
Figura 13. Curva de crecimiento en circunferencia del pseudotallo por tipo de planta, construida en base a las medias.....	44
Figura 14. Evolución del número acumulado de hojas brotadas en plantas hijas y plantas nietas a lo largo del tiempo por tipo de planta.....	45
Figura 15. Evolución del número acumulado de hojas brotadas en plantas hijas y plantas nietas en el tiempo por tipo de planta y tipo de suelo	45

Figura 16. Evolución del número acumulado de hojas brotadas en plantas hijas y plantas nietas en el tiempo por tipo de planta y tipo de riego.....	46
Figura 17. Evolución del número acumulado de hojas brotadas en plantas hijas y plantas nietas en el tiempo por tipo de planta y por parcela	46
Figura 18. Panel de diagramas pareados entre variables biométricas para plantas hijas (izquierda) y plantas nietas (derecha)	48
Figura 19. Panel de diagramas pareados entre las variables edáficas para las plantas hijas	50
Figura 20. Panel de diagramas pareados entre variables edáficas para plantas nietas	51
Figura 21. Panel de diagramas pareados entre la conductividad eléctrica del suelo a distintos niveles de profundidad y las variables biométricas de las plantas hijas	52
Figura 22. Panel de diagramas pareados entre la conductividad eléctrica del suelo a distintos niveles de profundidad y las variables biométricas de las plantas nietas.....	53
Figura 23. Panel de diagramas pareados entre el contenido volumétrico de agua en el suelo a distintos niveles de profundidad y las variables biométricas de las plantas hijas	54
Figura 24. Panel de diagramas pareados entre el contenido volumétrico de agua en el suelo a distintos niveles de profundidad y las variables biométricas de las plantas nietas	55
Figura 25. Panel de diagramas pareados entre la temperatura del suelo a distintas profundidades y las variables biométricas de las plantas hijas.....	56
Figura 26. Panel de diagramas pareados entre la temperatura del suelo y las variables biométricas de las plantas nietas	57
Figura 27. Tabla de resumen estadístico para la significancia individual de variables predictivas	60
Figura 28. Prueba de factor de inflación de la varianza	61
Figura 29. Gráfico de residuos vs valores ajustados para el modelo de regresión lineal mixto para la altura de las plantas nietas.....	63
Figura 30. Gráfico cuantil-cuantil para el modelo de regresión lineal mixto para la altura de las plantas nietas	63
Figura 31. Gráfico de residuos vs valores ajustados para el modelo de regresión lineal mixto para la circunferencia del pseudotallo de las plantas nietas	65
Figura 32. Gráfico cuantil-cuantil para el modelo de regresión lineal mixto para la circunferencia del pseudotallo de las plantas nietas	66

Introducción

El banano orgánico es uno de los principales productos de exportación de la Región Piura. Este producto dinamiza parte importante de la economía piurana y ha experimentado en los últimos años un crecimiento constante y sostenido en sus niveles de producción y exportación. Sin embargo, este sector supone también un campo poco explorado en materias de investigación e innovación.

Así pues, este trabajo supone desarrollar una serie de modelos matemáticos que ayuden a identificar qué variables condicionan el crecimiento de una planta de banano orgánico en la región, de modo que estas puedan ser manipuladas en un futuro con el fin de obtener ejemplares mejor desarrollados y que muestren, en consecuencia, un nivel de productividad más alto.

Para ejecutar esta investigación se utilizaron sensores capaces de medir la conductividad eléctrica, así como el contenido volumétrico de agua y la temperatura de los suelos de una serie de parcelas pertenecientes a agroexportadores de la cooperativa APBOSMAN.

Dichos sensores fueron proporcionados por el proyecto “Optimización de del uso de agua en el sistema de riego por inundación del banano orgánico, utilizado por pequeños productores del Valle del Chira, Piura” del Concytec.



Capítulo 1

Marco teórico

Este capítulo inicial presenta una recopilación de información sobre el cultivo del banano orgánico. Se ofrece una breve revisión histórica de la producción de esta fruta, así como una descripción detallada del desarrollo biométrico de la planta. El propósito principal de esta sección es proporcionar una introducción clara y concisa de los conceptos que se abordarán en el estudio, con especial atención en la definición de las variables biométricas que se utilizarán para medir el desarrollo del banano.

Además, en este capítulo se presentan datos y análisis sobre los suelos y sus respectivas variables que tienen una influencia significativa en el crecimiento del banano. De manera similar, se explica la metodología utilizada para clasificar cada tipo de suelo que se ha estudiado en el proyecto. El objetivo de esta sección es proporcionar una comprensión precisa de las variables del suelo que se están investigando, lo que ayudará a definir y contextualizar los resultados del estudio sobre el banano orgánico.

1.1 Banano orgánico

El banano, una fruta ampliamente conocida y consumida a nivel mundial, es producido por varias especies del género *Musa*, el cual incluye tanto a las variedades cultivadas como a las silvestres. El fruto, que se caracteriza por ser alargado, curvado y carnoso, puede variar en cuanto a su tamaño, color y textura, dependiendo de la especie de la planta que lo produzca. Además, es una importante fuente de nutrientes como potasio, magnesio y vitaminas B6 y C.

En cuanto a su origen, existen diversas teorías sobre el lugar de donde proviene la domesticación de las plantas de banano. Sin embargo, un estudio realizado por Denham y otros (2003) sugiere que su origen se ubica en Papúa Nueva Guinea, donde se encontraron restos de plantas de banano con una antigüedad de entre 6 000 y 10 000 años. Estos hallazgos demostrarían que el banano fue uno de los primeros cultivos agrícolas en ser domesticados por los habitantes de estas tierras, y que tardó miles de años en llegar a otras partes del mundo (como Europa o América).

De hecho, el banano llega al continente americano a mediados del siglo XVI, en torno al año 1516 con la llegada misma de los españoles. Así pues, siguiendo la ruta que

realizaron los conquistadores en el nuevo mundo, se sabe que el banano es recibido inicialmente en zonas caribeñas, naciones que en la actualidad son de las que más banano producen alrededor del mundo.

Pese a que hoy en día existe un sinnúmero de variedades del fruto (superan las mil), pueden ser clasificados en dos grandes grupos de acuerdo con la forma de consumo que el ser humano hace de ellos:

- La primera se trata del plátano rico en azúcares, consumido generalmente crudo y a modo de postre.
- La segunda se trata del plátano rico en almidón, el cual se cuece, fríe o asa para poder ser consumido.

Dado que el presente trabajo de investigación centra su experimentación en el primero de los dos tipos previamente mencionados, la información que se abordará a partir de este punto girará en torno a dicho tipo de banano.

Para la Fundación Española de la Nutrición (2020) este fruto tiene forma alargada, la mayoría de las veces también curvada. Su peso varía entre el común de 100 y 200 g. Con una cáscara gruesa de color amarilla cuando el fruto ha terminado con su etapa de maduración que cubre una pulpa carnosa y fácil de pelar.

1.1.1 Planta de banano orgánico

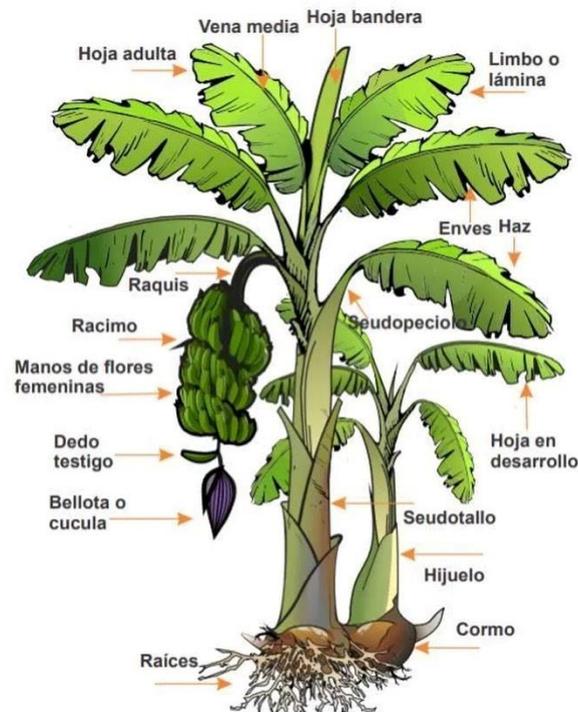
La planta que produce el banano es conocida como platanera, bananero, plátano o banano, y es una hierba perenne de gran tamaño perteneciente al género *Musa*. Como las demás plantas del mismo género, no cuenta con un verdadero tronco, sino que, en su lugar, posee un pseudotallo formado a partir de vainas foliares, el cual puede alcanzar hasta 12 metros de altura y tiene una estructura cilíndrica que puede llegar a tener un diámetro de hasta 60 cm.

Las hojas de esta planta son de las más grandes que existen en el reino vegetal, pues pueden alcanzar los 3 m de largo y hasta 90 cm de ancho. Son lisas, tiernas, oblongas, con el ápice trunco y la base redonda o ligeramente cordiforme, verdes por el haz y más claras y normalmente glaucas por el envés, con los márgenes lisos y las nervaduras pinnadas, amarillentas o verdes. El total de hojas por planta varía, normalmente, entre 5 y 15.

Las plantas de banano se caracterizan por tener una floración única y completa, en la que todas las flores en una misma inflorescencia maduran al mismo tiempo. Dicha inflorescencia nace directamente a partir del rizoma luego de que hayan brotado en ella entre 26 y 32 hojas. Esta inflorescencia se conoce en algunas zonas como "bellota". Las flores de banano se agrupan en racimos que pueden tener entre 50 y 150 frutos. Cada fruto se desarrolla a partir de una flor y contiene numerosas semillas pequeñas. Una vez

que los frutos maduran, la planta muere y es reemplazada por brotes laterales que crecen a partir de los rizomas.

Figura 1. Planta de banano y sus partes



Nota. Adaptado de infoAgro.com (2016).

1.1.2 Cultivo y cosecha de banano orgánico en la región

El estudio y experimentación de la presente investigación se llevó a cabo en parcelas de la asociación APBOSMAN. Dicha asociación sigue un proceso de cultivo y cosecha de banano orgánico con certificado internacional de calidad.

Así pues, se sigue una serie de operaciones a lo largo del proceso de siembra y cosecha del banano orgánico, las cuales son (García, 2021):

- a) Siembra: En esta primera etapa se siembra la semilla. Luego de haberse preparado el terreno para la plantación y de haberse escogido la semilla, esta se entierra de forma que quede en vertical y con las yemas adheridas a la planta orientadas hacia un solo lado, procurando de esa forma tener racimos ordenados. Aquí se asegura que se tenga acceso a la suficiente humedad y nutrientes durante la fase de instalación.
- b) Deshermane: Este proceso suele llevarse a cabo durante 4 meses, periodo de tiempo suficiente para determinar cuál de los brotes surgidos luego de la siembra será la futura madre. Se elige el brote de mayor vigor y se eliminan los restantes para evitar que interfieran con su desarrollo.
- c) Deshije: Análogicamente a la fase anterior, esta consiste en quitar hijos de la planta madre, con el fin de asegurar el sistema de producción “Madre – hija

- nieta”, dando a la madre la posibilidad de que continúe con el proceso de desarrollo de manera óptima.
- d) Deshoje: Este proceso consiste en la eliminación de las hojas secas o quebradas que podrían llegar a limitar el crecimiento de la planta. Existe una técnica específica para llevar a cabo este proceso, con el fin de no lastimar el pseudotallo, evitando generar aberturas mediante las cuales puedan penetrar enfermedades.
 - e) Deschante: Consiste en el retiro de todas las vainas del pseudotallo que se han secado y, por lo tanto, han cumplido su ciclo de vida. Tal como sucede con el deshoje, existe una técnica específica para hacer esto, con el fin de no lastimar la planta.
 - f) Eliminación de malezas: Es un proceso clave en el cultivo de banano orgánico, ya que de no ser quitadas pueden influir negativamente en el desarrollo y productividad de la planta.
 - g) Riego: La planta de banano requiere de un constante suministro de agua.
 - h) Fertilización: Actividad relacionada al uso de fertilizantes que garantice el constante suministro y correcto balance de elementos nutritivos necesarios para el crecimiento de la planta.
 - i) Control de plagas: Consiste en el manejo y eliminación de plagas mediante el uso de diversas sustancias.
 - j) Control de la pudrición del tallo y corona: Actividad que busca preservar el tallo y corona de la planta.
 - k) Cosecha: Actividad que no supone más que el recojo de la fruta brotada de la planta. El tiempo que transcurre desde el brote de la bellota hasta la disponibilidad de cosecha es de, aproximadamente, 12 semanas.

1.1.3 Indicadores de producción de banano

Obtener un registro claro sobre el volumen de producción anual de banano en el mundo es bastante difícil. Esto se debe principalmente a que gran parte de la producción de este fruto recae en pequeños agricultores que lo comercializan en mercados informales, los cuales la mayoría de las veces no cuentan con un sistema de trazabilidad. Por ejemplo, el 70 u 80% del banano producido en África provienen de plantas que han estado presentes en el continente durante más de mil años (FAO, 2020).

De acuerdo con la FAO (2020), entre los años 2000 y 2017, la producción aumentó a un ritmo de 3.2% anual, alcanzando un récord de 114 millones de toneladas producidas en el año 2017 (lo que contrasta bastante con las 67 millones producidas en el año 2000).

En ese contexto, el banano es predominantemente producido en Asia, África y Latinoamérica y el Caribe. India y China, con 29 y 11 millones de toneladas anuales en promedio respectivamente, son los países que más banano producen en el mundo.

El Perú, por su lado, es también un importante exportador de banano orgánico. Cuenta hoy en día con 15 000 hectáreas destinadas al cultivo de dicha fruta, de las cuales 12 800 se encuentran en la región Piura, principalmente en la provincia de Sullana, en la zona conocida como Valle del Chira (MINAGRI, 2019).

Entre los años 2010 y 2015, el volumen de exportaciones de banano orgánico con origen en territorio peruano experimentó un crecimiento que rondaba el 94%, a pesar de que la cantidad de terreno disponible para su cosecha representó solo el 4% del total destinado para el cultivo de banano (incluyendo el convencional) (MINAGRI, 2015).

1.2 Suelo

El suelo constituye el segundo elemento de importancia en el sistema experimental del que se obtendrán datos de mediciones. En el presente proyecto de investigación, es necesario destacar que el suelo es un componente especialmente fundamental que juega un papel determinante en el crecimiento y desarrollo de las plantas. Por tanto, es necesario establecer algunos conceptos básicos sobre el mismo.

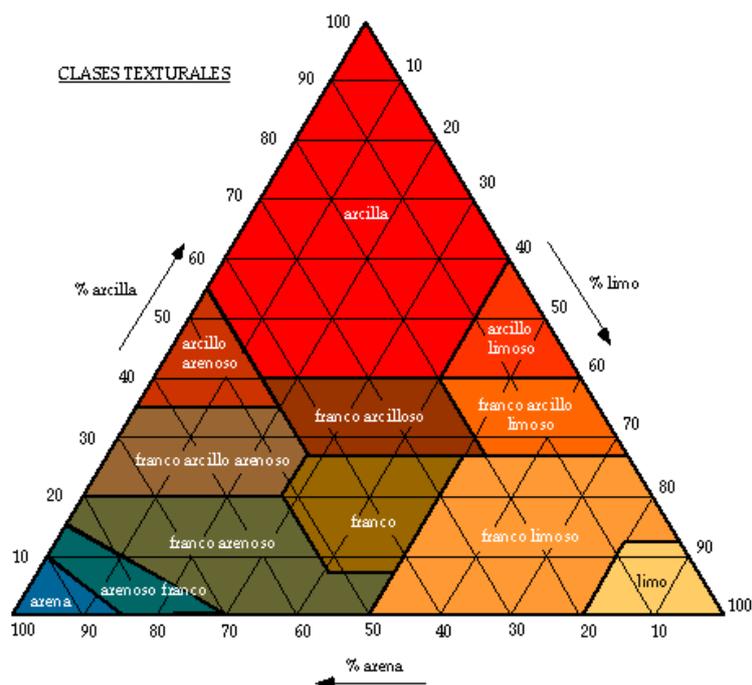
En principio, se entiende como suelo a la parte externa (superficial) de la corteza terrestre que se forma a partir de las erosiones o cambios físicos/químicos de las rocas y los residuos de las actividades de seres vivos que habitan sobre él (Villalaz, 2004).

Los conceptos a entender de este elemento dependen del modo de afectación de este en el proyecto. Es decir, conviene entender de qué manera afectan los suelos el crecimiento o desarrollo de una planta de banano. Para ello, es necesario explicar dos indicadores de los mismos: la clase textural del suelo y el ratio de decrecimiento.

El primer indicador de acuerdo con la FAO (2019), sirve para determinar en gran medida la capacidad de retención de agua de un suelo, así como la facilidad para trabajar el mismo. Para ello, mide el contenido relativo de partículas de tamaños distintos presentes en un mismo suelo, lo que permite indicar el porcentaje de arena, limo y arcilla de dicho suelo (agentes que determinan, además, el nombre de una gran variedad de suelos: arenosos, limosos, arcillosos, franco-arenosos, franco-limosos, franco-arcillosos, etc.).

La Figura 2 muestra una herramienta gráfica diseñada por el Departamento de Agricultura de Estados Unidos (USDA, por sus siglas en inglés) que permite clasificar un suelo de acuerdo con el porcentaje de las diferentes partículas presentes en él.

Figura 2. Triángulo de clases texturales del suelo



Nota. Adaptado de Departamento de Agricultura de Estados Unidos (1993).

El segundo indicador corresponde a uno evaluado anteriormente por García (2021) que busca básicamente cuantificar el ritmo de disminución de la cantidad de agua presente en el suelo a lo largo del tiempo. Dicha investigación hace un estudio más detallado de lo explicado en esta sección.

Ambos indicadores permitieron clasificar los suelos estudiados en las distintas parcelas en donde se desarrolló el estudio en tres tipos, todos de acuerdo con la capacidad de retención de agua que poseían:

- Suelos poco retentivos: aquellos suelos de clase textural franco arenoso con un ratio de decrecimiento mayor al presentado por suelos retentivos y muy retentivos.
- Suelos retentivos: suelos de clase franco o franco arenoso cuyo ratio de decrecimiento es mayor al de los suelos poco retentivos pero menor al de los suelos muy retentivos.
- Suelos muy retentivos: suelos de clase franco o franco limoso que poseen un ratio de decrecimiento inferior al de suelos retentivos y especialmente inferior al presentado por suelos poco retentivos.

1.2.1 Variables del suelo que influyen en el crecimiento de las plantas

A lo largo del tiempo se han ido estudiando las distintas propiedades y variables presentes en el suelo que influyen en el crecimiento de las plantas a fin de mejorar el cultivo y potenciar las cosechas en distintas partes del mundo. Para efectos prácticos del

presente trabajo de investigación, se recopiló información sobre cuáles son las principales variables del suelo que influyen en el crecimiento de las plantas (con especial hincapié en aquellas que se tomaron en cuenta para el estudio: conductividad eléctrica, humedad y temperatura del suelo).

- **Conductividad eléctrica**

Esta propiedad se define como la capacidad del suelo para conducir la corriente eléctrica y está relacionada con la cantidad de sales disueltas en presencia del suelo (Cremona & Enriquez, 2020). La concentración de sales en el suelo puede influir significativamente en la absorción de agua y nutrientes por las raíces de las plantas. La presencia de altas concentraciones de sales puede aumentar la tensión osmótica del suelo, lo que dificulta la absorción de agua por las raíces. Por otro lado, una baja concentración de sales puede generar un suelo deficiente en nutrientes, lo que también puede afectar negativamente el crecimiento de las plantas (Barbaro, Karlanian, & Mata, 2014).

- **Humedad**

El agua es esencial para la fotosíntesis, la absorción de nutrientes y el transporte de sustancias dentro de la planta. El agua es también un componente fundamental en la turgencia de las células vegetales, que es necesaria para mantener la rigidez de la planta y soportar su estructura. El contenido de agua en el suelo afecta directamente la disponibilidad de agua para las plantas y puede influir en su crecimiento y desarrollo.

Un contenido de agua insuficiente en el suelo puede generar un estrés hídrico en las plantas, lo que reduce la turgencia de las células y provoca la disminución de la fotosíntesis, el cierre de las estomas y la pérdida de biomasa. Por otro lado, un exceso de agua en el suelo puede limitar el flujo de aire en las raíces, lo que puede provocar la falta de oxígeno y la proliferación de microorganismos anaerobios, generando un suelo anegado y reduciendo la disponibilidad de nutrientes para las plantas (Escobar, y otros, 2018).

- **Temperatura**

La temperatura del suelo influye en la disponibilidad de agua y nutrientes, así como en la actividad metabólica de las raíces y microorganismos del suelo. Las raíces de las plantas necesitan una temperatura adecuada para absorber agua y nutrientes de la solución del suelo, ya que una temperatura demasiado baja o alta puede afectar la actividad enzimática y la movilidad de los iones en la solución del suelo. Además, la temperatura del suelo también puede influir en la disponibilidad de nutrientes, ya que la actividad microbiana del suelo se ve afectada por la temperatura (Fischer, Torres, & Torres, 1999).

Por otro lado, la temperatura del suelo también puede influir en la germinación de las semillas y la emergencia de plántulas. La mayoría de las plantas tienen una temperatura óptima de germinación que varía según la especie. Una temperatura demasiado baja puede retrasar la germinación o incluso inhibirla, mientras que una temperatura demasiado alta puede afectar la viabilidad de las semillas y provocar una baja tasa de germinación. Además, la temperatura del suelo también puede influir en la velocidad de crecimiento de las plántulas y en la aparición de enfermedades y plagas del suelo (Gámez Vázquez, y otros, 2015).

- **pH del suelo**

El pH del suelo es una medida de la acidez o alcalinidad del suelo y está determinado por la concentración de iones hidrógeno (H^+) en la solución del suelo. El pH del suelo puede variar desde muy ácido (pH 4.5) hasta muy alcalino (pH 9.0). La mayoría de las plantas prefieren un pH del suelo ligeramente ácido, entre 6.0 y 7.0, aunque hay algunas plantas que prefieren suelos más ácidos o alcalinos. La acidez del suelo puede afectar la disponibilidad de nutrientes esenciales para las plantas, como el fósforo, el calcio, el magnesio y el hierro. La absorción de estos nutrientes se ve limitada cuando el pH del suelo es demasiado ácido o alcalino. Además, el pH del suelo puede afectar la actividad microbiana en el suelo, lo que a su vez puede afectar la disponibilidad de nutrientes para las plantas (Cremona & Enriquez, 2020).

El pH del suelo también puede influir en la toxicidad de ciertos metales pesados en el suelo. Cuando el pH del suelo es demasiado ácido, puede haber una mayor movilidad de metales pesados como el aluminio y el manganeso, lo que aumenta su toxicidad para las plantas. Por otro lado, cuando el pH del suelo es demasiado alcalino, puede haber una mayor movilidad de metales pesados como el zinc y el cobre, lo que también puede ser tóxico para las plantas (Barbaro, Karlanian, & Mata, 2014).

Capítulo 2

Metodología

Toda investigación que se realice requiere de una justificación que la valide. En la actualidad, el campo de investigación relacionado a la agricultura del banano en el norte del Perú es limitada e incipiente, por lo que todo esfuerzo en comprender aquellos factores que influyen en el correcto desarrollo y crecimiento de dicha planta significan un avance significativo en dicho campo.

En este capítulo se presenta la justificación del trabajo experimental y de investigación desarrollado. Así pues, se explica cómo se diseñó el sistema experimental para el recojo de la información y todo el trabajo posterior de análisis y limpieza de la misma.

Se especifica, además, qué herramientas (físicas y digitales) se utilizaron en cada etapa del proceso de experimentación y exploración, desde el diseño del experimento hasta la selección del modelo estadístico.

2.1 Justificación

Así como hay variables que no se podrán controlar, existen variables que definen o condicionan el desarrollo de una planta de banano y que pueden ser manipuladas. El control de dichas variables posibilita la modificación de las características morfológicas propias de la planta de banano. Es así que, de identificar estas variables y modificar sus valores a conveniencia, es posible conseguir plantas de banano mejoradas.

Como ya había demostrado Rey, y otros (2010), existen determinadas propiedades edáficas (principalmente físicas, como la resistencia a la penetración y la infiltración) que condicionan el vigor que puedan desarrollar las plantas de banano sembradas en suelo de distintas fincas en Venezuela.

Además, otro estudio llevado a cabo por Caicedo, Balmaseda, & Proaño (2015) en una finca de Ecuador, demostró que las plantas de banano que no recibían algún tipo de riego presentaban un nivel de producción hasta 43% menor que aquellas que sí lo hacían.

Dichos estudios motivan la exploración y determinación, desde el alcance y limitaciones propias del trabajo de investigación, qué variables presentes en la región Piura condicionan y determinan el crecimiento y desarrollo de las plantas de banano, y describir dicho crecimiento mediante el análisis de su relación con dichas variables, tratando de determinar la dirección y magnitud de dicha relación.

2.2 Diseño experimental

Para desarrollar todo el proceso experimental es necesario, en principio determinar la manera de cuantificar el crecimiento o desarrollo de las plantas de banano. Esto permitiría establecer comparaciones y determinar si una planta es, objetivamente, más grande o se encuentra mejor desarrollada que otra.

Tomando como referencia investigaciones realizadas previamente, tales como la de Olivares, y otros (2020), y Martínez Varona (2013), se resolvió cuantificar el desarrollo de la planta de banano en base a los siguientes tres parámetros, a los cuales se denominará de ahora en adelante como variables biométricas:

- La altura de la planta.
- La circunferencia del pseudotallo.
- El número total de hojas emitidas por la planta.

Una vez se establecieron las variables biométricas a medir, se procedió a identificar las variables que podrían afectar o condicionar el desarrollo de la planta de banano, es decir, variables de las cuales dependería el valor que pudieran llegar a alcanzar las plantas de banano en alguno de los parámetros listados previamente.

Como parte del proyecto Concytec – Banco Mundial “Mejoramiento y Ampliación de los Servicios del Sistema Nacional de Ciencia Tecnología e Innovación Tecnológica”, basado además en investigaciones preliminares, se dispuso que el presente trabajo de investigación se modelara bajo un experimento que contemplara los siguientes factores, los cuales pueden agruparse según dos tipos de variables: uno categórico y otro cuantitativo.

Las variables categóricas seleccionadas fueron:

- El tipo de suelo en el que se sembró cada planta, con tres categorías asignadas: suelo poco retentivo, suelo retentivo, suelo muy retentivo (clasificación explicada en el Capítulo 1).
- El tipo de riego aplicado, con dos categorías asignadas: riego tradicional y riego mejorado. El riego tradicional es el riego aplicado de forma predeterminada por la mayoría de fundos de siembra y cultivo de banano orgánico en la región, esto es, un riego por inundación: se suministra toda el agua al cultivo hasta inundarlo, y no se vuelve a aplicar hasta que sea absorbida por completo. Mientras que el riego mejorado supone la

dosificación de la cantidad de agua que se suministra por cada riego, así pues, no se recurre a la inundación del cultivo y se tiene que aplicar de forma más continua que el tradicional. Se utiliza la misma cantidad de agua, pero se suministra por partes en más y distintos momentos.

Las variables cuantitativas, por otro lado, fueron:

- Tiempo.
- Contenido volumétrico de agua del suelo, a la que se hace referencia también como simplemente “humedad del suelo”.
- Conductividad eléctrica del suelo: cuyo estudio se encuentra justificado por trabajos de investigación previos, además de los presentados en el capítulo anterior, como los de Watabe, Nakano, & Dong-Hyuk (2022) o Romaneckas, y otros (2023).
- Temperatura del suelo.

Las últimas tres variables cuantitativas, dada la naturaleza de su procedencia, serán denominadas de ahora en adelante como variables edáficas.

En base a las consideraciones previamente indicadas, se solicitó el acceso a las instalaciones y sujetos de estudio propiedad de dos productores, quienes recibirán el título de Productor 1 y Productor 2 a fin de mantener su identidad como confidencial, miembros de APBOSMAM, una cooperativa encargada de vincular a productores locales de banano con consumidores de dicha fruta alrededor del mundo, cuyas instalaciones se localizan actualmente en la provincia de Sullana, departamento de Piura.

Tomando en cuenta la disponibilidad de dichas instalaciones y los recursos que estas poseían para llevar a cabo la experimentación, esta se realizó bajo las siguientes premisas, todas contempladas en un principio por el Proyecto Concytec – Banco Mundial “Mejoramiento y Ampliación de los Servicios del Sistema Nacional de Ciencia Tecnología e Innovación Tecnológica”:

- Se utilizarían 6 parcelas piloto en total.
- Habría dos parcelas por cada tipo de suelo, cubriendo así los tres tipos de suelo establecidos.
- A la mitad de ellas se les aplicaría un sistema de riego mejorado y a la otra mitad un sistema de riego tradicional.
- Los datos de medidas del suelo se tomarían de forma periódica, mientras que los de las plantas de banano se tomarían después de un tiempo mucho más prolongado y no constante.

2.2.1 Disposición y características de parcelas piloto

En este estudio preliminar, cuatro de las seis parcelas establecidas para el proceso experimental, pertenecían a las instalaciones del fundo Productor 1; mientras

que las dos restantes pertenecían al Productor 2. Dos parcelas del Productor 1 poseían un suelo poco retentivo, una un suelo retentivo y la otra un suelo muy retentivo. Por su parte, una de las parcelas del Productor 2 poseía un suelo retentivo y la otra un suelo muy retentivo. Así pues, existían dos parcelas por cada tipo de suelo, tal como se había establecido previamente en las consideraciones generales del diseño experimental.

Por otro lado, a dos de las cuatro parcelas del Productor 1 se les aplicó un riego mejorado, y las otras dos mantuvieron el sistema de riego tradicional. En cuanto a las parcelas del Productor 2, a una se le aplicó un sistema de riego mejorado y a la otra se le aplicó un sistema de riego tradicional. Con ello, se cumplía el tercer punto de las premisas generales del diseño experimental. Sin embargo, el riego mejorado no fue aplicado a las parcelas correspondientes hasta pasados cinco meses de haber iniciado con la toma de medidas a las plantas, por lo que todas las seis parcelas recibieron un riego tradicional previo a este cambio de tipo de riego.

En cada parcela se seleccionó aleatoriamente a un número de plantas de banano a las que se les hizo el seguimiento de su crecimiento a través de las variables biométricas antes mencionadas.

A continuación, la Tabla 1 sintetiza y presenta de forma clara toda la información referente a las parcelas explicadas previamente.

Tabla 1. Parcelas de estudio

N.º	Código	N.º de plantas en estudio	Instalaciones	Tipo de suelo	Tipo de riego	Estación de monitoreo
1	ME-A1	50	Productor 1	Muy retentivo	Tradicional	Estación 6
2	ME-B1	50	Productor 1	Retentivo	Tradicional – Mejorado ¹	Estación 3
3	ME-A2	25	Productor 1	Poco retentivo	Tradicional – Mejorado ¹	Estación 7
4	ME-B2	25	Productor 1	Poco retentivo	Tradicional	Estación 4
5	MG-A	50	Productor 2	Muy retentivo	Tradicional – Mejorado ¹	Estación 5
6	MG-B	50	Productor 2	Retentivo	Tradicional	Estación 2

¹ Aplicado posteriormente a la cuarta medida de variables biométricas.

2.2.2 Medición de variables biométricas

Las variables biométricas seleccionadas –altura, circunferencia y número de hojas brotadas– se midieron en tres generaciones distintas de plantas de banano: plantas madre, plantas hija y plantas nietas.

La primera toma de medidas se realizó a las plantas madre sembradas en las seis parcelas presentadas en la Tabla 1. Se procuró que todas las plantas madre que iban a ser medidas se encontraran en la misma etapa fenológica: el florecimiento o emisión de su florescencia, para el día en que la toma de medidas se llevara a cabo (16 de enero de

2020). Sin embargo, vale aclarar que los valores obtenidos de la medición aplicada a dichas plantas finalmente fueron utilizados como datos referenciales y no formaron parte de la formulación de los modelos matemáticos desarrollados.

La ventaja de elegir plantas que hubieran emitido o estuvieran cerca de emitir su florescencia (bellota, como se conoce en el argot agrícola) es que se marcaba un punto de partida similar para todos los sujetos experimentales. De esta manera, se podría asegurar que todos los sujetos de estudio tuvieran en cierta medida la misma edad (partían en igualdad de condiciones, en ese aspecto).

Adicional a la toma de medidas de las plantas madre se procedió también a la medición de las variables biométricas de aquellas plantas hijas que ya hubieran brotado para el mismo día (considerándose el tiempo para este punto como 0 meses) y, cuando llegasen a brotar las plantas nietas, se procedería a medirlas también.

La medición de las plantas nietas se inició el día 3 de junio de 2020, 5 meses después de haberse realizado la primera medida de variables biométricas a las plantas hijas. Vale aclarar que el tiempo transcurrido se establece en relación a la primera toma de las medidas (con fecha 16 de enero de 2020) y no al momento en que nacieron cada una de las plantas. Por lo tanto, no todas las plantas nietas medidas en un mismo corte temporal necesariamente van a coincidir en etapa fenológica o edad (hay algunas que nacen antes o después de la toma de medidas ejecutada en dicha fecha debido a la característica propia).

Estas mediciones se llevaron a cabo de la siguiente manera:

- La altura de la planta, independiente del tipo o generación a la que perteneciera, se midió desde el nivel de suelo hasta la copa. La herramienta utilizada fue una cinta métrica y los datos fueron registrados en metros.
- La circunferencia del pseudotallo se midió a un tercio del total de la altura de la planta. Así pues, si la altura de la planta era igual a 3 m, la medición de la circunferencia de su pseudo tallo se hacía a 1 m de altura desde el nivel del suelo. La herramienta de medición fue también una cinta métrica y los datos fueron registrados, nuevamente, en metros.
- Para medir la cantidad de hojas brotadas se contó el número de hojas nuevas florecidas desde la última vez que se tomó medidas de la planta en estudio. La primera medida de esta variable expresó la cantidad inicial de hojas de la planta. Vale aclarar, así mismo, que para admitir una hoja como brotada su ancho debía ser superior a 10 cm (ancho a partir del cual una hoja de planta de banano se considera completamente desarrollada).

La metodología establecida para la medición de las variables biométricas de las plantas de banano se ejecutó de acuerdo a lo explicado anteriormente ya que no se

tenía control total sobre los sujetos experimentales. Es decir, no se pudo controlar el momento de siembra de cada una de las plantas, por lo que se optó diseñar la toma de medidas según lo detallado en párrafos previos.

Las medidas fueron tomadas por el equipo de investigación a cargo del proyecto, y los datos obtenidos se almacenaron en archivos Excel. El cronograma de medición se llevó a cabo tal como muestra la Tabla 2. Vale aclarar que el riego mejorado no fue aplicado en el experimento hasta después de haberse realizado la cuarta toma de medidas (es decir, entre el 3 de junio y el 3 de agosto de 2020).

Tabla 2. Cronograma de medición de variables biométricas por tipo de planta

	16/01/2020	20/02/2020	15/04/2020	03/06/2020	03/08/2020	26/10/2020	21/12/2020	24/03/2021
Plantas madre	X							
Plantas hija	X	X	X	X	X			
Plantas nieta				X	X	X	X	X

2.2.3 Medición de variables edáficas

Las variables edáficas se midieron mediante estaciones de monitoreo ubicadas en cada una de las parcelas piloto. Mediante la instalación de calicatas, fue posible obtener medidas de cada variable y almacenar de forma automática los datos recolectados.

Se establecieron 4 puntos de referencia para la toma de medidas, cada uno a diferentes niveles de profundidad: a 20, 40, 60 y 80 cm de profundidad.

Los sensores instalados tomaban valores cada 2 horas. Sin embargo, y como será explicado más adelante en el capítulo correspondiente, existieron mediciones dobles que tenían que ser depuradas.

2.3 Limpieza y análisis de datos

La limpieza de datos se basó en la corrección y/o eliminación de errores de medición y en la identificación y eliminación/corrección de valores atípicos. Adicional a ello, se considera como parte de la limpieza de datos la estructuración del dataframe con la data previamente depurada que serviría como base para el análisis estadístico aplicado.

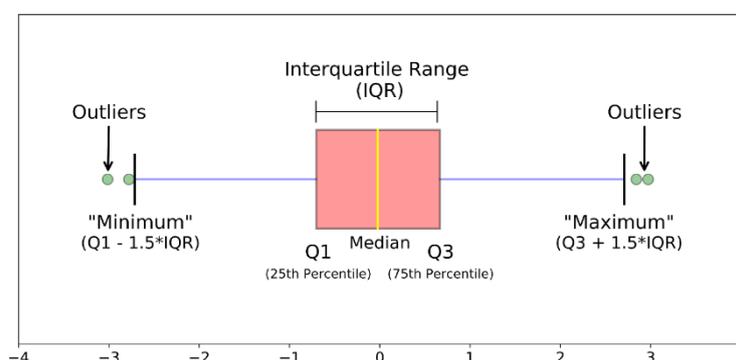
Dada la facilidad de uso y potencia, la herramienta seleccionada para la limpieza de los datos de la medición de las variables edáficas fue Microsoft Excel. El lenguaje de programación y análisis estadístico R se utilizó para la construcción del *dataframe* final y para su posterior análisis estadístico.

Las herramientas estadísticas utilizadas durante el análisis de datos fueron las siguientes:

- **Diagrama de caja**

Los diagramas de caja, o box plot en inglés, constituyen un procedimiento estandarizado para la representación gráfica de la distribución de una serie de datos numéricos en torno a 5 variables de resumen estadístico: valor mínimo, primer cuartil, mediana (segundo cuartil), tercer cuartil y valor máximo.

Figura 3. Partes de un diagrama de caja



Nota. Adaptado de Galarnyk (2018).

El uso de este método de representación gráfica permite identificar valores atípicos, evaluar la simetría, la dispersión y el nivel de sesgo del conjunto de datos que se está analizando.

- **Diagramas de dispersión**

Los diagramas de dispersión son utilizados para representar sobre un plano cartesiano los valores de dos variables de un conjunto de datos. Se emplea típicamente cuando una de estas dos variables se encuentra bajo control en un proceso de experimentación, ubicándola usualmente en el eje de las abscisas, y a la otra en el eje de las ordenadas.

El uso de este gráfico permite identificar patrones en la distribución de los puntos que denoten la existencia de algún tipo de correlación entre las variables evaluadas. Dicha correlación puede ser descrita de distintas maneras a partir del mismo análisis: positiva o negativa, fuerte o débil, lineal o no lineal.

- **Series de tiempo**

Los diagramas de serie de tiempo permiten observar tendencias o patrones de un conjunto de datos a lo largo del tiempo. Con ello, cuando la tendencia es ajustada, comportamientos anómalos permiten determinar valores atípicos en el conjunto de datos en estudio.

2.4 Modelamiento

Para desarrollar el apartado del modelamiento y determinación de las variables que condicionan el crecimiento de banano entre el conjunto que se había preseleccionado, se utilizó R nuevamente.

Se identificó que, tal como se explica más adelante en el Capítulo 4, la naturaleza de los datos trabajados implicaba el uso de un modelo lineal mixto. Se consideró así mismo que, dada la elección de cuantificar el crecimiento o desarrollo de una planta de banano de acuerdo a tres variables biométricas, cada una de ellas debía ser explicada mediante un modelo matemático en específico.

Las herramientas estadísticas utilizadas para el proceso de modelamiento fueron las siguientes:

- **El método de Estimación por Máxima Verosimilitud**

El método EMV (o MLE, por sus siglas en inglés) se emplea para determinar los valores de los parámetros de un modelo de forma que, una vez establecidos, estos maximicen la similitud entre los valores generados a partir de dicho modelo y los valores observados en un principio.

Este método es ampliamente utilizado para la generación de modelos lineales, análisis factorial (exploratorio y confirmatorio), análisis estructurales y demás y puede resumirse bajo la siguiente ecuación:

$$L(\theta|x) = \prod_{i=1}^n f(x_i, \theta)$$

- **Factor de la inflación de la varianza**

El factor de la inflación de la varianza (VIF, por sus siglas en inglés) es un indicador estadístico que cuantifica el incremento de la varianza de un coeficiente de regresión por efecto de la colinealidad.

Así pues, es un indicador de la multicolinealidad en análisis de regresión normal de mínimos cuadrados. La siguiente ecuación describe el factor de la inflación de la varianza:

$$VIF = \frac{1}{1 - R_i^2}$$

- **Criterio de Información de Akaike**

El Criterio de Información de Akaike (AIC, por sus siglas en inglés) es un indicador empleado para analizar la calidad relativa de un modelo estadístico, para un conjunto de datos. Es decir, es una herramienta utilizada para la selección del modelo.

Sin embargo, este método no significa una prueba del modelo tal como una prueba de hipótesis (donde se admite o se rechaza la significancia de un modelo), sino que facilita un medio para la comparación entre modelos para la selección de alguno de ellos. Si todos los modelos en cuestión no son significativos, el AIC no será capaz de indicarlo.

En ese sentido, este método brinda una estimación relativa de la cantidad de información que se pierde cuando se emplea un modelo u otro para reproducir los valores generados por el proceso. De acuerdo a ello, el modelo que presente el AIC más pequeño será, en principio, el mejor entre todos los modelos candidatos.

El AIC se define bajo la siguiente ecuación:

$$AIC = 2k - 2\ln(L)$$

Donde:

k: número de parámetros libres del modelo.

L: máximo valor de la función de verosimilitud para el modelo estimado.

- **Coefficiente de determinación (R cuadrado)**

El coeficiente de determinación indica la proporción de la varianza total de la variable respuesta que es explicada por el modelo propuesto. En ese sentido, este coeficiente refleja la bondad del ajuste de un modelo a la variable que pretende explicar.

El coeficiente de determinación, también conocido como R cuadrado, puede calcularse bajo la siguiente ecuación:

$$R^2 = \frac{\sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2}{\sum_{t=1}^T (Y_t - \bar{Y})^2}$$

Dada la naturaleza de los datos de las variables predictivas, los modelos propuestos fueron modelos lineales mixtos. Ello supuso la existencia de dos tipos de R cuadrado, uno condicional y uno marginal (McCulloch, Searle, & Neuhaus, 2008).

El R cuadrado marginal cuantifica la proporción de la varianza total que es explicada solo por los factores fijos del modelo. Dicho coeficiente puede expresarse de la siguiente forma:

$$R_{GLMM(m)}^2 = \frac{\sigma_f^2}{\sigma_f^2 + \sum_{l=1}^u \sigma_l^2 + \sigma_e^2 + \sigma_d^2}$$

Por su parte, el R cuadrado condicional describe la proporción de la varianza total que es explicada tanto por los factores fijos del modelo como por los aleatorios. En consecuencia, este puede expresarse del siguiente modo:

$$R_{GLMM(c)}^2 = \frac{\sigma_f^2 + \sum_{l=1}^u \sigma_l^2}{\sigma_f^2 + \sum_{l=1}^u \sigma_l^2 + \sigma_e^2 + \sigma_d^2}$$

- **Gráficas**

Adicional a todas las herramientas descritas previamente, se utilizaron dos gráficas como apoyo para la validación del modelo seleccionado: la primera una gráfica de dispersión entre los valores ajustados y los residuos estándar del modelo elegido y la segunda una gráfica de ajuste de normalidad de los residuos estándar del modelo.

La primera gráfica permite evaluar el comportamiento de la varianza del modelo para identificar si se cumple el principio de homocedasticidad: no se muestran patrones anormales y es constante a lo largo del tiempo. La segunda, por su parte, permite analizar qué tanto se ajustan los errores a una distribución normal.



Capítulo 3

Análisis de datos

En el presente capítulo se describe el proceso de limpieza y análisis de toda la data obtenida durante el proceso de experimentación y recolección de datos del trabajo de investigación. Para ello, resulta necesario explicar primero qué tipos de datos son los que se están analizando y el proceso que se realizó para llegar hasta dicho análisis.

El uso del análisis exploratorio de datos (EDA, por sus siglas en inglés) para el reconocimiento de relaciones entre variables, la existencia de subgrupos de datos o la identificación de valores atípicos, significa una poderosa herramienta estadística para la comprensión de la naturaleza y el comportamiento de los datos en los cuales se basa la construcción de cualquier modelo matemático.

Entendiendo aquello como un paso fundamental para el correcto establecimiento del modelo matemático a construir, se aplicó el uso de dicha herramienta a los datos obtenidos en el proceso de experimentación. El proceso de aplicación y los resultados obtenidos se recogen en el presente capítulo.

La exploración y el análisis de datos se ha realizado mediante el uso de diversas herramientas de representación gráfica y resumen estadístico que forman parte del análisis exploratorio de datos: series de tiempo, gráficos caja bigote (boxplots en inglés), diagramas de dispersión y análisis de correlación.

Tomando en cuenta la naturaleza de las variables cuyos datos han sido explorados mediante el uso de estos recursos estadísticos, se puede establecer dos grandes etapas de aplicación del EDA. La primera referencia al análisis realizado a los datos obtenidos en la medición de las variables biométricas de las plantas de banano y la segunda al análisis realizado a los datos obtenidos en la medición de las variables edáficas

Resulta necesario recordar que los valores registrados para cada variable edáfica –conductividad eléctrica, contenido volumétrico de agua y temperatura del suelo– fueron medidos cada dos horas por medio de estaciones de control ubicadas en cadauna de las parcelas de prueba a distintos niveles de profundidad (20, 40, 60 y 80 cm por debajo del nivel de suelo).

Así mismo, se recuerda que los datos registrados en la medición de cada variable biométrica de las plantas de banano –altura de la planta, circunferencia del pseudotallo y número acumulado de hojas brotadas– se midieron en un punto específico en el tiempo, y el intervalo de tiempo transcurrido entre un punto y otro ha sido representado en cantidad de meses.

- **Datos recopilados**

Los datos recopilados durante todo el proceso de experimentación llevado a cabo en el presente trabajo de investigación se pueden agrupar en dos grandes categorías de acuerdo con la naturaleza de su origen.

El primero de ellos, obtenido directamente de la medición de parámetros específicos de la planta de banano, se llamará “variables biométricas del banano” y corresponden básicamente a la medición de determinadas variables de la planta que cuantifican su crecimiento. Estas, que han sido mencionadas previamente, son:

- Altura de la planta: Cuantifica el desarrollo del pseudotallo de la planta en base a la distancia desde el suelo hasta la punta del mismo. Su unidad de medida es en metros.
- Circunferencia del pseudotallo: Cuantifica el desarrollo de la planta en base al grosor que alcance su pseudotallo. Este se mide a un tercio de la altura de la planta y su unidad de medida es en metros también.
- Número de hojas brotadas: Cuantifica el desarrollo de la planta en base al número total de hojas que han brotado en ella. No posee unidad de medida.

Todos los datos de estas medidas fueron tomados por el equipo colaborador encargado del proyecto, quienes estructuraron una matriz de tomas de medidas en un libro Excel, donde registraban los datos especificando la fecha de la toma de medida y el valor.

El segundo grupo de datos corresponden a los extraídos del suelo de las parcelas donde se plantaron los sujetos de estudio (las plantas de banano). Este grupo se conocerá como “variables edáficas”:

- Contenido volumétrico de agua: Que se puede homologar simplemente a la humedad del suelo. Indica la cantidad de agua presente en el suelo para un momento específico. Su unidad de medida es m^3/m^3 .
- Conductividad eléctrica: Indica la capacidad del suelo de conducir corriente eléctrica en un momento específico. Su unidad de medida es $\mu S/cm$.
- Temperatura: Corresponde al registro tomado de la temperatura del suelo en un momento específico. Su unidad es de $^{\circ}C$.

Todos los datos obtenidos de estas tres variables se registraron en archivos correspondientes al formato CSV (Comma Separated Values), los cuales fueron tratados posteriormente en Excel.

3.1 Limpieza de datos

El primer gran reto del análisis de datos obtenidos de las variables biométricas y edáficas fue la gestión y estructuración de la data de modo que esta pudiera ser trabajada de forma óptima por un lenguaje de programación orientado al análisis estadístico (R, en este caso).

Para ello, fue necesario consolidar toda la data obtenida en un solo dataframe. Las dos grandes fuentes de información fueron los archivos que contenían los registros de las variables biométricas, por un lado, y de las variables edáficas por otro.

Para poder estructurar las primeras solo fue necesario transformar los puntos en el tiempo (fechas de la toma de medidas) en una variable cuantitativa. Para ello, solo hizo falta determinar el tiempo transcurrido entre toma y toma y transformar el mismo en cantidad de meses.

Sin embargo, para las segundas el proceso fue un tanto más extenso. Cabe recordar que los datos obtenidos de la medición de las variables edáficas se tomaban cada dos horas de forma automática por medio de las estaciones de monitoreo ubicadas en cada parcela de prueba, por lo tanto, la base de datos de estas variables era mucho más amplia en cuanto a volumen de datos que la correspondiente a las variables biométricas.

Por otro lado, era típico encontrar datos duplicados y tomas erróneas de las medidas (valores iguales a 0 o inexistentes). Por lo tanto, fue imprescindible ejecutar una limpieza previa de toda esta data, para depurar todos estos errores existentes.

Una vez ejecutada esta limpieza, se procedió a consolidar todos los datos recogidos con el fin de colocar uno solo que resumiera el valor en el dataframe elaborado. Para ello, basado en un trabajo de investigación previo, se procedió a colocar el promedio ponderado de los 28 días previos a una toma de medidas de las variables biométricas (García, 2021).

Con los datos de las variables biométricas y edáficas a incluir en el dataframe estructurado, se procedió a concluir con su construcción. Esta base contenía distintos campos (columnas), la mayoría de ellos relacionados a las variables medidas a lo largo del experimento, los cuales se listan y detallan a continuación:

- idPlanta: Corresponde a un registro único de identificación para cada planta estudiada.
- Planta: Indica el número de planta asignado de acuerdo a la parcela piloto.

- Tiempo: Indica el número de meses transcurrido desde el inicio de proyecto hasta la toma de medidas.
- Tipo: Variable categórica de valores “Hija” que designa a las plantas hijas y “Nieta” que designa a las plantas nietas.
- Parcela: Variable categórica que designa la parcela a la que pertenece una planta.
- TipoSuelo: Variable categórica que indica el tipo de suelo que posee la parcela en la que se sembró la planta de banano.
- TipoRiego: Variable categórica que indica el tipo de riego que se aplicó a la parcela en la que se sembró la planta de banano.
- Altura: Variable cuantitativa. Contiene el valor de la altura de la planta de banano en un tiempo específico.
- Circunferencia: Variable cuantitativa. Contiene el valor del grosor del pseudotallo de la planta de banano en un tiempo específico.
- Hojas: Variable cuantitativa. Contiene el valor del total de hojas brotadas en la planta de banano desde el inicio del proyecto hasta un punto específico en el tiempo.
- Estación: Variable categórica. Contiene el número de estación con la que se monitoreó y se gestionó la toma de datos de las variables edáficas.
- Cond20, Cond40, Cond60, Cond80: Variables cuantitativas. Registran el promedio de los datos de conductividad eléctrica del suelo obtenidos durante 28 días previos a la toma de medidas en las plantas de banano para cada uno de los niveles de profundidad establecidos en el diseño de experimentos.
- Hum20, Hum40, Hum60, Hum80: Variables cuantitativas. Registran el promedio de los datos de humedad del suelo obtenidos durante 28 días previos a la toma de medidas en las plantas de banano, para cada nivel de profundidad.
- Temp20, Temp40, Temp60, Temp80: Variables cuantitativas. Registran el promedio de los datos de temperatura del suelo obtenidos durante 28 días previos a la toma de medidas en las plantas de banano correspondientes a cada nivel de profundidad.

3.2 Análisis estadístico

Este apartado se ha dividido en dos secciones enfocadas al análisis de ambos grupos de variables: las variables biométricas y las variables edáficas. La primera de ellas se ha denominado como “Análisis de variables biométricas”, y la segunda como “Análisis de correlación”, ya que, al establecer las variables edáficas como variables predictivas del modelo que se busca construir, importa saber el nivel de correlación existente entre ellas.

3.2.1 Análisis de variables biométricas

Para analizar los datos registrados en la medición de las variables biométricas se hizo uso de diversos gráficos caja bigote. Mediante el resumen gráfico de dichos datos bajo el empleo de esta herramienta se buscó determinar el grupo de individuos que registraron los valores más altos para cada una de las variables biométricas en estudio, atendiendo especialmente a las características de clasificación de cada grupo de plantas.

Se analizó la evolución en el tiempo de cada variable biométrica para cada grupo. Se usaron distintas agrupaciones para la visualización:

- El tipo de planta: el primer criterio de agrupación se utilizó para separar los datos en plantas hijas o plantas nietas. Esta regla de agrupación, para efectos prácticos será denominada como *agrupación 1*.
- El tipo de suelo: los datos separados según plantas hijas o nietas, fueron separados nuevamente de acuerdo al tipo de suelo en el que crecieron: poco retentivo, retentivo o muy retentivo. Esto corresponde a la *agrupación 2*.
- El tipo de riego: los datos por tipo de planta se agruparon de acuerdo al tipo de riego que se aplicó a la parcela en la que fueron cultivadas: tradicional o mejorado. De acuerdo con la nominación, corresponde a la *agrupación 3*.
- La parcela: los datos agrupados por tipo de planta, se segregaron de acuerdo a la parcela en la que crecieron: ME-A1, ME-A2, ME-B1, ME-B2, MG-A o MG-B. Esta es la *agrupación 4*.

A lo largo de esta primera sección se recogen gráficos boxplot que muestran el comportamiento y evolución de la altura, la circunferencia del pseudotallo y el número acumulado de hojas brotadas de las plantas de banano a lo largo del tiempo (Figura 4 a Figura 17). Este desarrollo, además, se representa de acuerdo con el tipo de planta (hijas o nietas), y con los datos agrupados según los criterios previamente descritos.

En líneas generales y de acuerdo con la lógica de lo esperado, las tres variables biométricas evolucionan positivamente en el tiempo. Como muestra, tanto en la Figura 4 como en la Figura 9 se observa que los valores de altura y circunferencia del pseudotallo para las plantas hijas registrados a 7.25 meses de haber iniciado el proyecto poseen todos sus cuartiles por encima de cualquier cuartil de los grupos de valores de altura y circunferencia del pseudotallo registrados a 0 o 1.25 meses de realizada la primera medida de altura para el mismo tipo de plantas.

Además, tanto en altura como en grosor, parece ser que el rango de crecimiento es mayor para las plantas nietas que para las plantas hijas. Por ejemplo, en la Figura 4 se aprecia que en tanto los valores ubicados entre el primer y el tercer cuartil del conjunto de datos de altura de las plantas hijas parecen variar entre 1 a 1.5 m y los 2.75

a 3.5 m, los valores de altura de las plantas nietas ubicados entre los mismos cuartiles parecen hacerlo entre los 0.25 a 0.75 m y los 2.5 a 3.5 m.

Sin embargo, se debe tener en cuenta que, tanto para la deducción anterior como para el resto que se recogen en este capítulo, el tiempo transcurrido entre la primera y la última toma de medidas entre plantas hijas y plantas nietas es distinto: el segmento de tiempo para las primeras es de 7.25 meses, mientras que para las segundas es de 10.5 meses, lo que implicaría en primera instancia que las plantas nietas “gozaron de más tiempo para crecer”.

Para finalizar, es prácticamente una regla general para la altura de las plantas y la circunferencia de sus pseudotallos que aquellas que se cultivaron en un suelo de tipo retentivo y con un riego de tipo mejorado presenten valores máximos superiores a los máximos alcanzados por aquellas plantas que se cultivaron en alguno de los otros dos tipos de suelo (muy y poco retentivos) y a las que se les aplicó el riego tradicional a lo largo de todo su proceso de cultivación. Esta afirmación se refleja en las gráficas correspondientes a los box-plot en los que se agrupan los datos por parcela que se muestran posteriormente.

Sin embargo, sucede también para ambas variables biométricas que estas mismas plantas presentan valores muy distintos entre sí, lo que evidencia una clara dispersión entre los datos de altura de la planta y grosor de su pseudotallo (esta dispersión se refleja en los boxplot que agrupan los datos de ambas variables según el tipo de suelo en el que fueron cultivadas, es decir, en la Figura 10 y la Figura 15).

Así mismo, es una constante también en ambas variables que las parcelas MG-A y MG-B presenten siempre valores inferiores en comparación con el resto de parcelas monitoreadas.

En los tres apartados presentados a continuación se recogen deducciones específicas que se han podido obtener del análisis simple de los boxplots elaborados a partir de la data recogida en el proceso de medición de cada una de las variables biométricas contempladas en el experimento (altura, circunferencia y número acumulado de hojas brotadas).

Dichas deducciones, a diferencia de las generales planteadas en los párrafos previos que se presentan como una constante entre las dos primeras variables biométricas, son específicas para cada variable, y no nacen a partir de la observación de un comportamiento similar o generalizado en dos o más de ellas.

3.2.1.1 Altura.

Figura 4. Evolución de la altura de las plantas hijas y plantas nietas en el tiempo

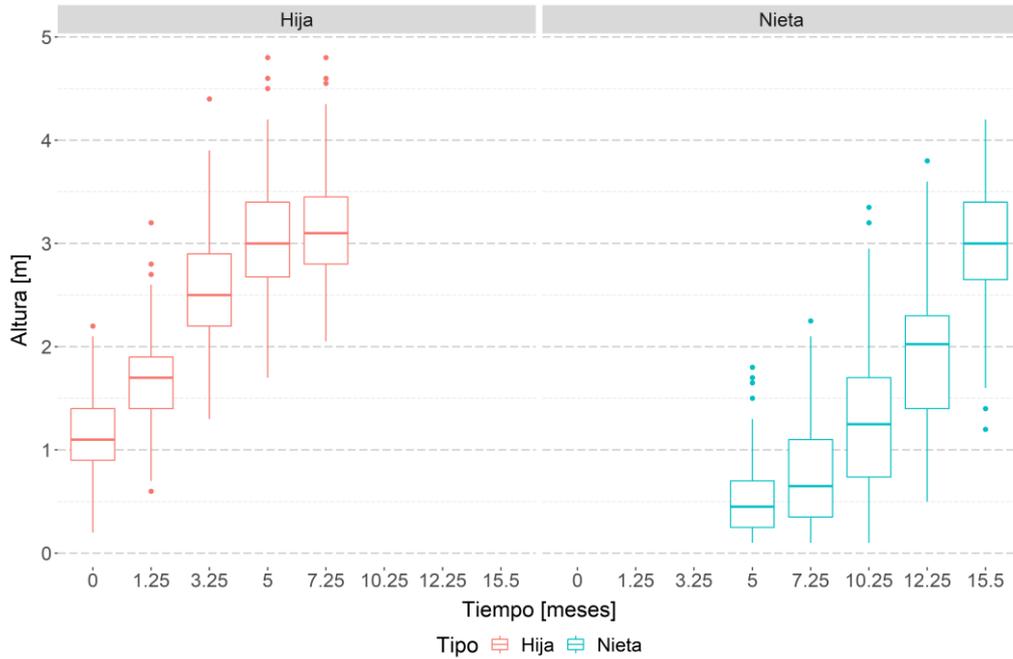


Figura 5. Evolución de la altura de las plantas hijas y plantas nietas en el tiempo por tipo de planta y tipo de suelo

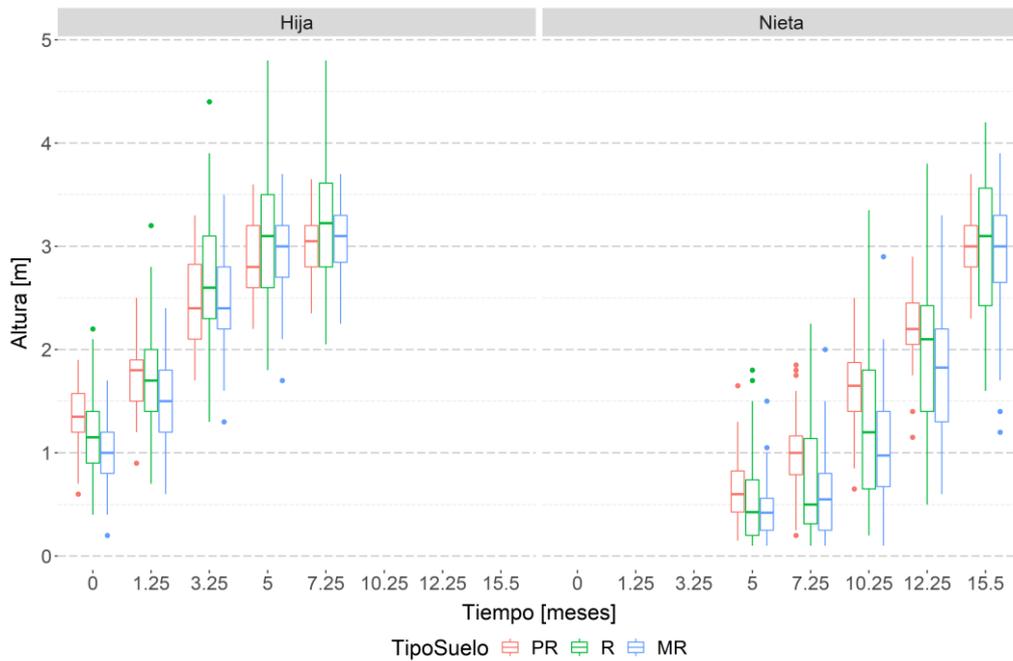


Figura 6. Evolución de la altura de las plantas hijas y plantas nietas en el tiempo por tipo de planta y tipo de riego

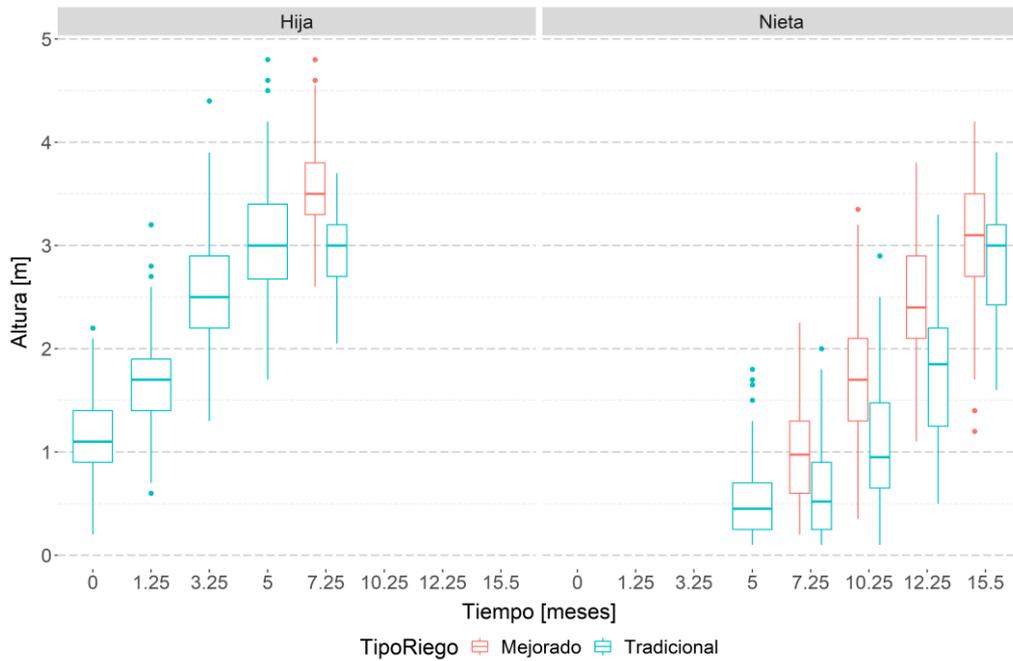
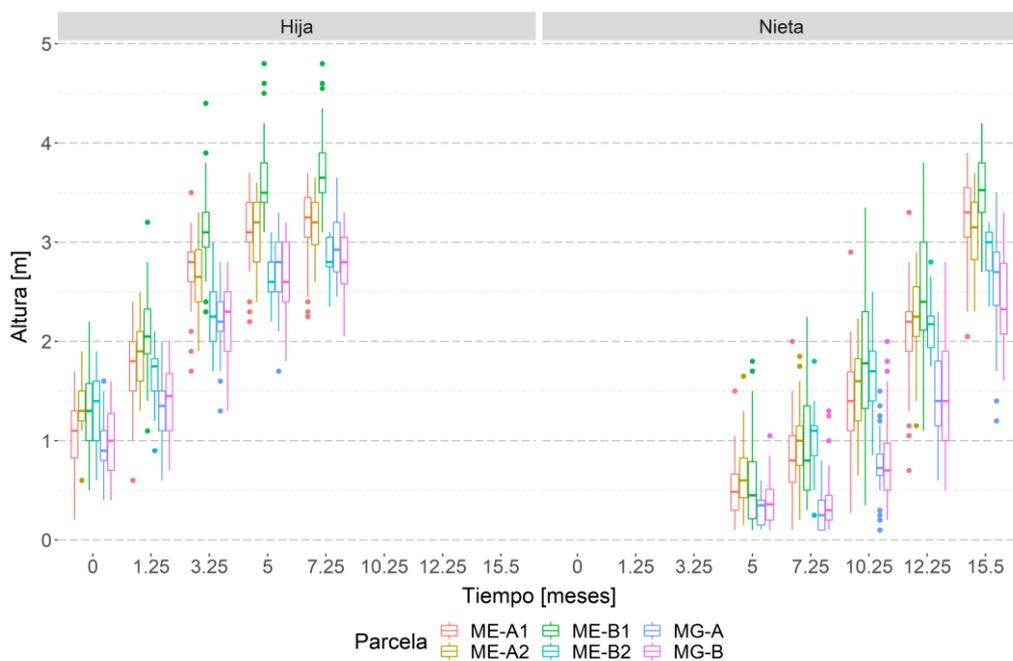


Figura 7. Evolución de la altura de las plantas hijas y plantas nietas en el tiempo por tipo de parcela



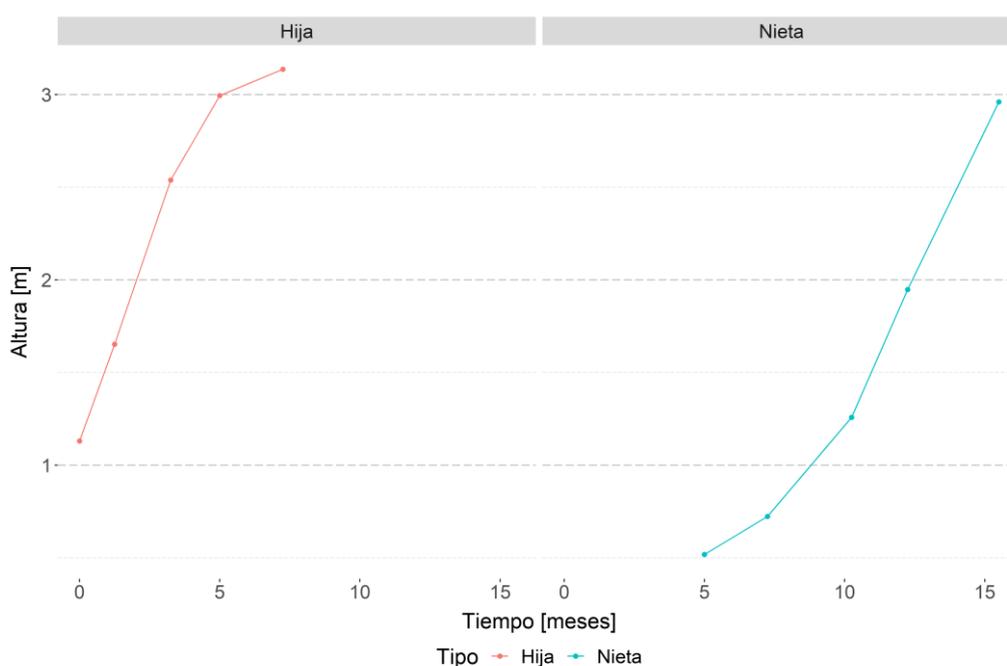
En cuanto a la primera de las variables biométricas medidas, en la gráfica de la Figura 4 se puede inferir que las plantas nietas han alcanzado la misma altura que las plantas hijas para el final del proceso de medición.

Por otra parte, se puede observar que el crecimiento de las plantas hijas parece haberse desacelerado para el final del proceso de toma de medidas. Así pues, la

diferencia entre los grupos de valores capturados a los 5 y 7.25 meses no es igual de considerable que entre los grupos de dos puntos de tiempo consecutivos anteriores.

Este comportamiento no es observable en las plantas nietas, por lo que se podría inferir, en principio, que dichas plantas no desaceleraron su crecimiento al final de la toma de medidas (este comportamiento se puede observar mejor en la Figura 8, construida a partir de las medias de los valores para cada punto en el tiempo). Sin embargo, se debe observar que el tiempo transcurrido entre los últimos puntos de tiempo de las plantas hijas (2.25 meses) es menor que al tiempo de las plantas nietas (3.25 meses), por lo que no se podría afirmar que efectivamente las plantas nietas no desaceleraron en ningún momento su crecimiento.

Figura 8. Curva de crecimiento en altura por tipo de planta, construida en base a las medias



Por otra parte, se observa en la Figura 6 que, a pesar de contar con un periodo de tiempo de medición más corto (por ende, un tiempo de crecimiento más reducido), las plantas hijas que recibieron un riego mejorado en su último periodo de medición presentaron valores de altura por encima de los presentados por las plantas nietas, sea que estas hayan recibido un riego tradicional o uno mejorado.

Finalmente, cabe mencionar que un comportamiento llamativo en esta variable es el que presenta el grupo de plantas cultivadas en la parcela MG-A. Como se puede apreciar en la Figura 7, esta parcela es la que poseía la caja de valores de menor altura entre todas, sin embargo, al finalizar la medición de la altura de las plantas hijas, esta logró tener plantas con valores que superaban a otras dos parcelas monitoreadas (ME-B2 y MG-B).

En resumen, se puede inferir que el tipo de riego y, en especial, el tipo de suelo influyen en la altura que puede alcanzar una planta de banano. Esto es observable en cada una de las gráficas mostradas anteriormente.

3.2.1.2 Circunferencia.

Figura 9. Evolución de la circunferencia del pseudotallo de las plantas hijas y plantas nietas en el tiempo

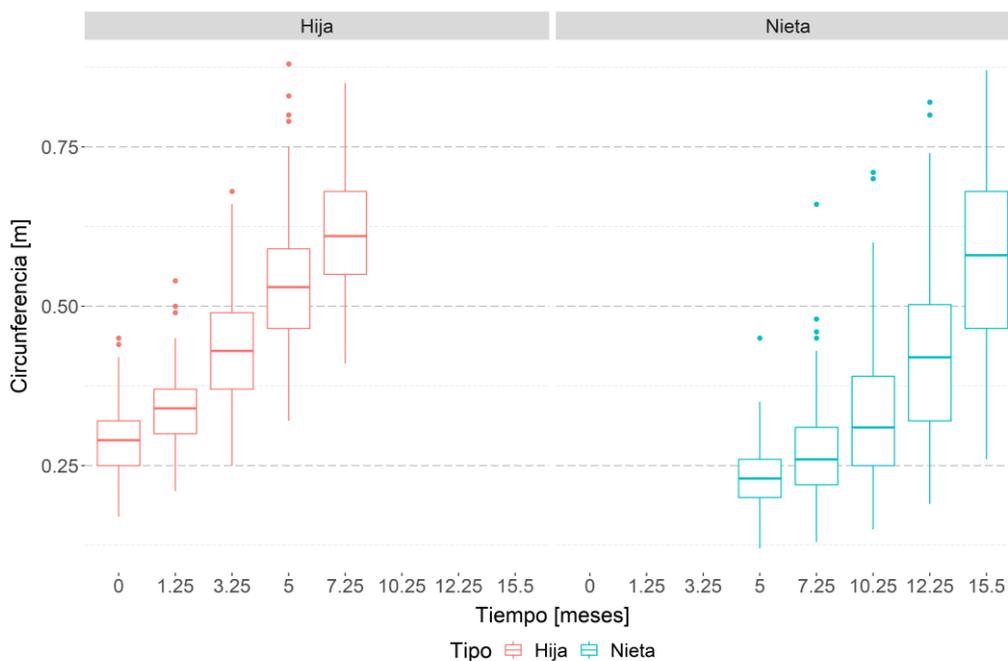


Figura 10. Evolución de la circunferencia del pseudotallo de las plantas hijas y plantas nietas en el tiempo por tipo de planta y tipo de suelo

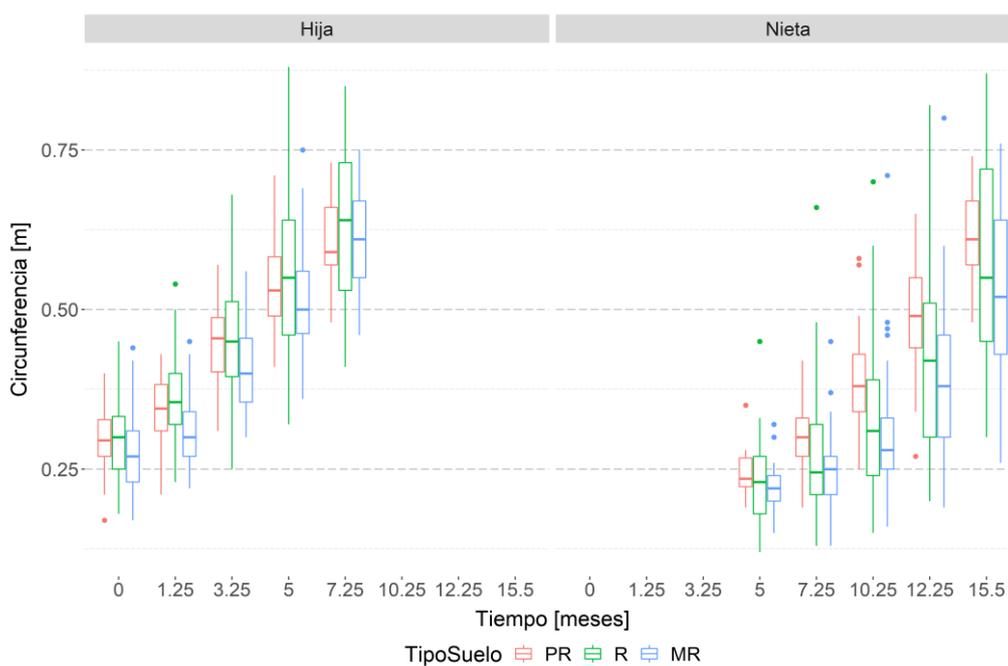


Figura 11. Evolución de la circunferencia del pseudotallo de las plantas hijas y plantas nietas en el tiempo por tipo de riego

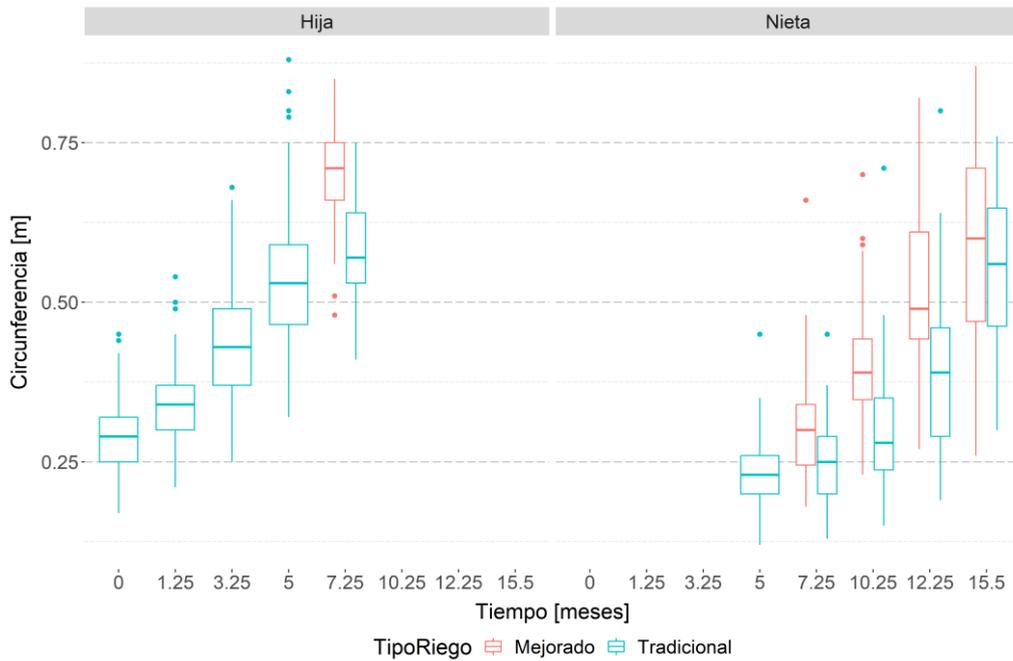
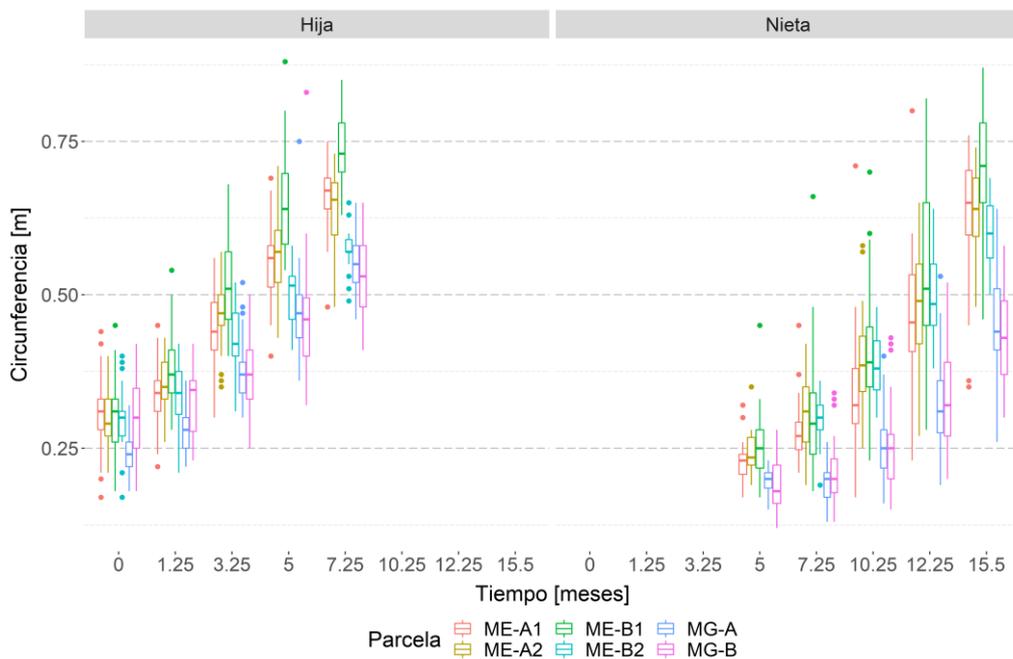


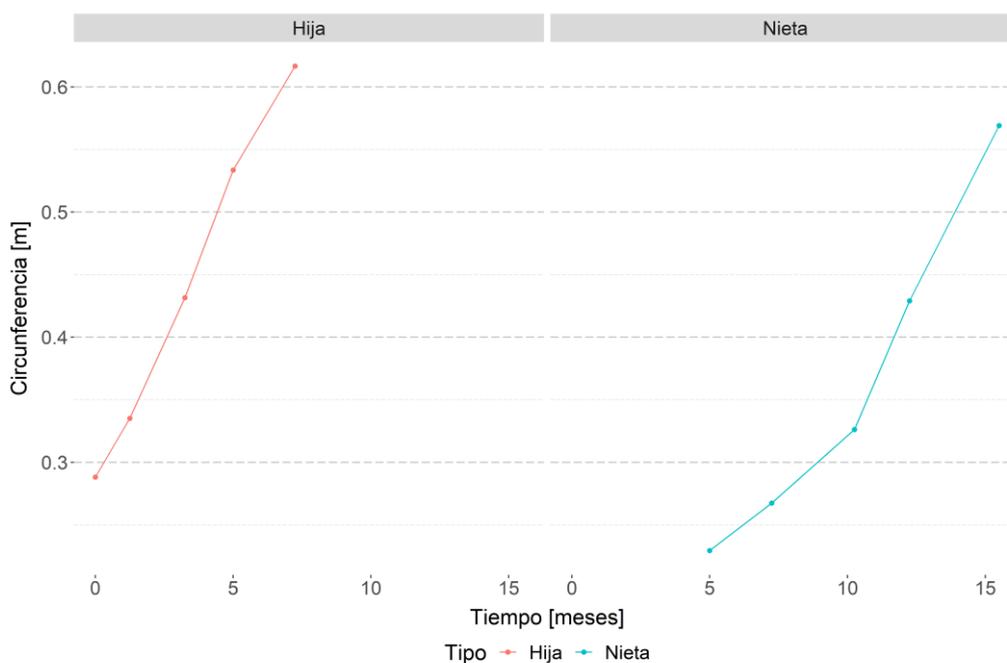
Figura 12. Evolución de la circunferencia del pseudotallo de las plantas hijas y plantas nietas en el tiempo por parcela



En lo que respecta a la circunferencia del pseudotallo, en la Figura 9 se observa que las curvas formadas por los conjuntos de datos de plantas hijas y plantas nietas son parecidas entre sí y describen, además, un comportamiento exponencial (mucho más pronunciado para las segundas que, para las primeras, cabe recalcar).

Este comportamiento se puede observar de mejor manera en la Figura 13, ya que la pendiente formada por la curva es evidentemente más empinada que la mostrada por la variable de altura en el apartado anterior.

Figura 13. Curva de crecimiento en circunferencia del pseudotallo por tipo de planta, construida en base a las medias



Así mismo se evidencia que la dispersión de los datos en la última toma de medidas de circunferencia es mayor para las plantas nietas que para las plantas hijas, ocasionando que, en consecuencia, la caja de valores que se dibuja entre los cuartiles 1 y 3 de estas sea más alargada que para las otras.

De hecho, esta diferencia en el largo de la caja se torna más evidente en los últimos 3 grupos de valores medidos para las plantas nietas, situación que no se presentaba para la variable biométrica de altura.

Por otro lado, en la Figura 10, a diferencia de lo que se puede apreciar con respecto a las plantas hijas, para las plantas nietas los cuartiles de valores de circunferencia de los pseudotallos de aquellas plantas que crecieron en suelos muy retentivos nunca superan los valores del tercer cuartil de la circunferencia de aquellas plantas crecidas en suelos poco retentivos.

Se observa también en la Figura 11 que luego de 12 meses de iniciado el proyecto de experimentación, la dispersión en los datos obtenidos de las plantas nietas para ambos tipos de riego es evidentemente mayor que para las medidas que se hicieron en los puntos en el tiempo anteriores. Situación distinta a la observada en el gráfico equivalente en el apartado de la altura, donde la dispersión de datos no es evidente entre grupo y grupo de valores.

Además, en la misma gráfica es de rescatar que en la última toma de medidas los valores de circunferencia para plantas con riego tradicional se acercaron bastante a los que se obtuvieron de las plantas con riego mejorado. Sin embargo, la caja de valores de estas últimas alcanzó valores superiores de todos modos.

Cuando se compara la evolución de la circunferencia del pseudotallo de acuerdo con el tratamiento de riego entre ambos tipos de plantas, se observa que en el último punto en el tiempo en el que existen medidas de circunferencia para las plantas hijas, dicha variable biométrica alcanza valores superiores a los que tienen las plantas nietas, aunque esta superioridad no es tan evidente como la observada en el gráfico equivalente en la altura.

Finalmente, cabe mencionar que, de acuerdo con la Figura 12, aquellas plantas sembradas en la parcela ME-B1 alcanzaron valores de circunferencia en su pseudotallo más altos que el de las plantas de banano cultivadas en las otras parcelas, tal como sucediera también con la altura. Además, las cajas correspondientes a las dos parcelas de MG-A y MG-B fueron las que alcanzaron los mínimos valores, tanto para plantas hijas como para plantas nietas.

En resumen, tal como sucediera con la altura, la circunferencia de una planta de banano es una variable biométrica que también se ve influenciada por el tipo de riego que se le aplica y el tipo de suelo en el que se siembra.

3.2.1.3 Número acumulado de hojas brotadas.

Figura 14. Evolución del número acumulado de hojas brotadas en plantas hijas y plantas nietas a lo largo del tiempo por tipo de planta

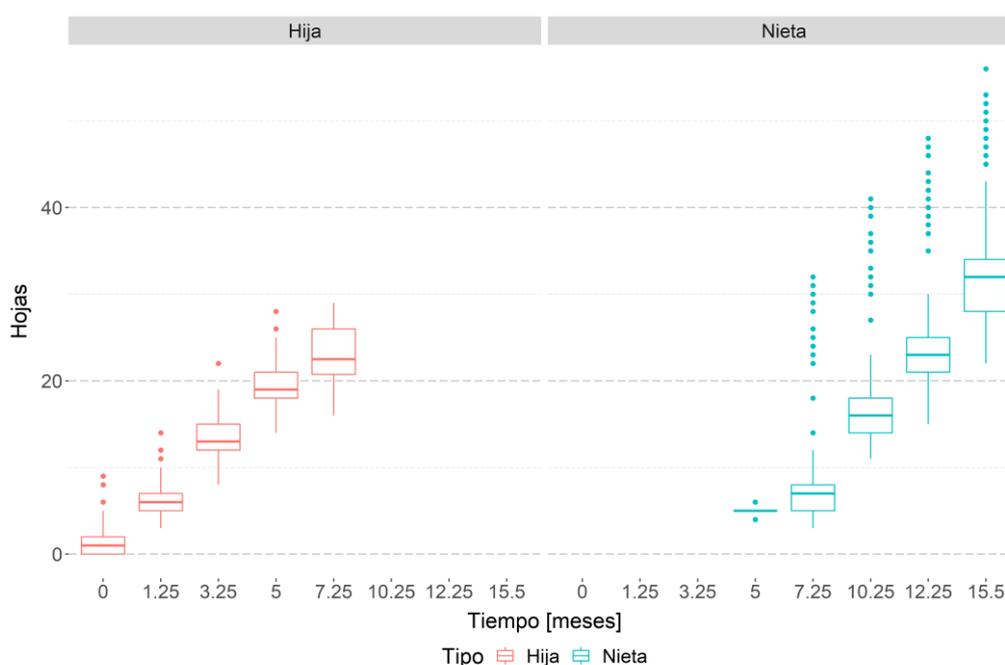


Figura 15. Evolución del número acumulado de hojas brotadas en plantas hijas y plantas nietas en el tiempo por tipo de planta y tipo de suelo

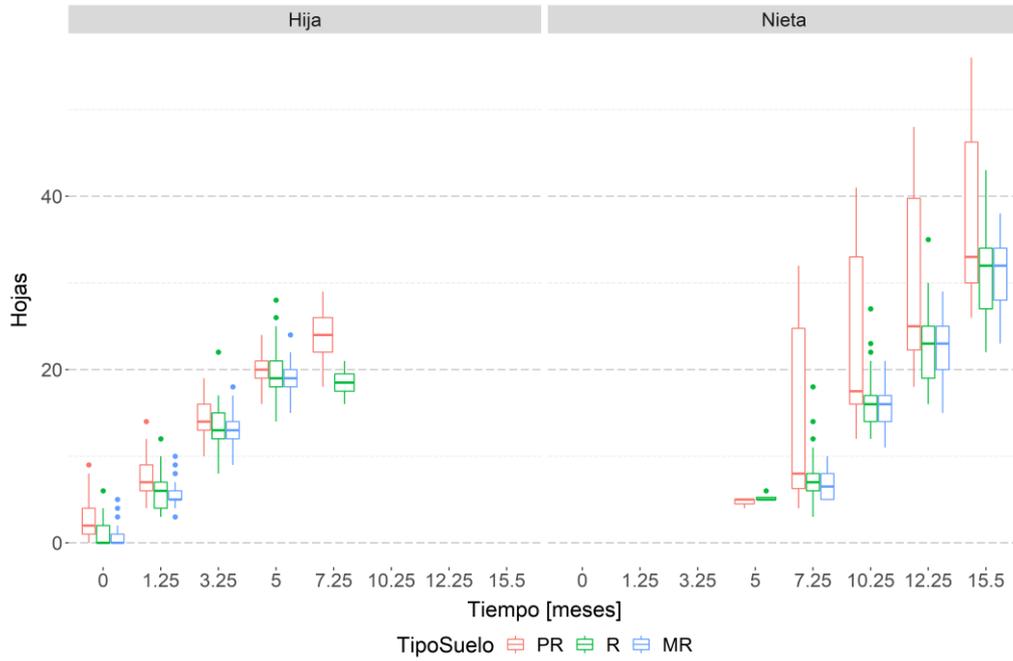


Figura 16. Evolución del número acumulado de hojas brotadas en plantas hijas y plantas nietas en el tiempo por tipo de planta y tipo de riego

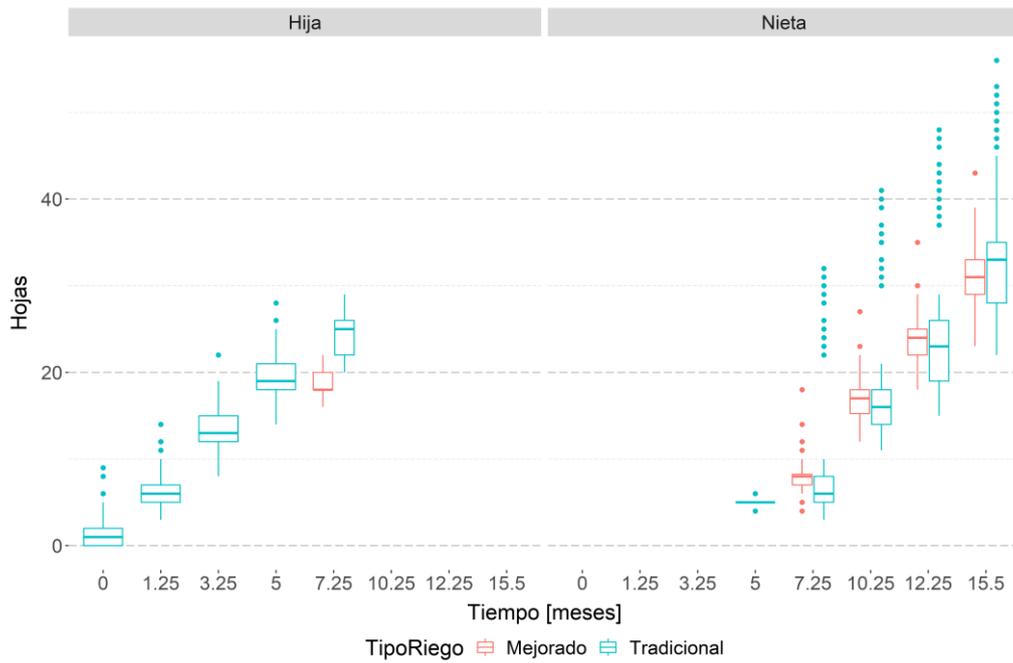
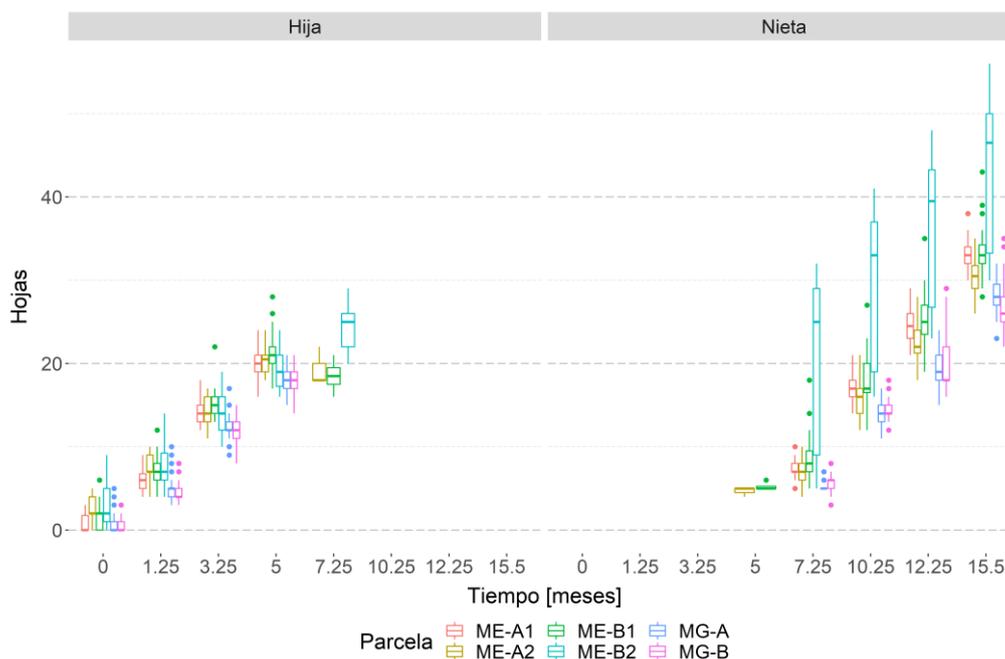


Figura 17. Evolución del número acumulado de hojas brotadas en plantas hijas y plantas nietas en el tiempo por tipo de planta y por parcela



Aunque se puede elaborar un análisis completo a cada una de las gráficas presentadas en este apartado, tal cual como se realizó con las otras dos variables biométricas anteriores, este pierde toda validez cuando existe un evidente problema con la data recolectada. Así pues, a diferencia de lo observado con la altura y la circunferencia del pseudotallo, dos de las gráficas correspondientes al análisis de la data del número acumulado de hojas brotadas muestran una cantidad excesiva de valores atípicos (Figura 14 y Figura 16) y otras dos un grupo de cajas excesivamente alargadas, es decir, valores excesivamente dispersos (Figura 15 y Figura 17).

Por ello se puede observar que la mayoría de plantas nietas han alcanzado un acumulado de entre 20 y 30 hojas brotadas en total al final de la toma de datos, pero un solo grupo de ellas (el correspondiente a la parcela ME-B2) han alcanzado un acumulado de hasta 50 hojas brotadas. La diferencia es abismal y no tiene lógica aparente.

3.2.2 Análisis de correlación

Los gráficos mostrados en este apartado presentan el nivel de correlación entre las variables de respuesta y las variables predictivas del sistema de datos establecido. Este análisis permite conocer, a groso modo, si existe una relación lineal entre dos variables numéricas. Esto nos puede ayudar a identificar relaciones entre las variables y usar este conocimiento cuando se estime el modelo de regresión.

De acuerdo con las gráficas y tablas mostradas más adelante, el p-valor para la inmensa mayoría de correlaciones estudiadas es cercano a cero, lo que supondría que existe una relación estadísticamente significativa entre dichas variables (a excepción de

determinados casos, como la altura y la conductividad eléctrica del suelo a 40 cm de profundidad, casos que son muy puntuales).

Sin embargo, a pesar de que el p-valor para la mayoría de correlaciones analizadas es cercano a cero, el valor de la correlación en muchos de dichos casos es cercano a cero también, lo que implica que la fuerza de relación es débil. Es decir, aunque es poco probable que la correlación observada se deba al azar, la correlación en sí misma es baja y no indica una relación lineal fuerte entre las dos variables estudiadas.

En muchos otros casos, además, no es posible determinar el tipo de correlación que desarrollan dos variables (una biométrica con una edáfica) debido a que esta presenta distinto sentido y magnitud a diferentes niveles de profundidad, como se puede observar con el análisis de correlación entre la circunferencia del pseudotallo y la conductividad eléctrica del suelo en cada uno de sus rangos de profundidad.

En resumen, el análisis desarrollado en los siguientes apartados de la correlación de las distintas variables del proyecto nos permite suponer que no existe una relación lineal importante entre dos variables que no sean biométricas (altura de la planta, circunferencia de su pseudotallo o número de hojas brotadas), o que no pertenezcan al mismo tipo de variable edáfica (conductividad eléctrica, humedad y temperatura del suelo en distintos niveles de profundidad). Haciendo salvedad en lo último de la conductividad eléctrica y la humedad del suelo, ya que son dos variables que sí presentan una alta y significativa correlación al mismo nivel de profundidad.

3.2.2.1 Nivel de correlación entre variables biométricas.

La Figura 18 se construyó en base a los datos de las medidas de las variables biométricas de las plantas de banano.

Figura 18. Panel de diagramas pareados entre variables biométricas para plantas hijas (izquierda) y plantas nietas (derecha)

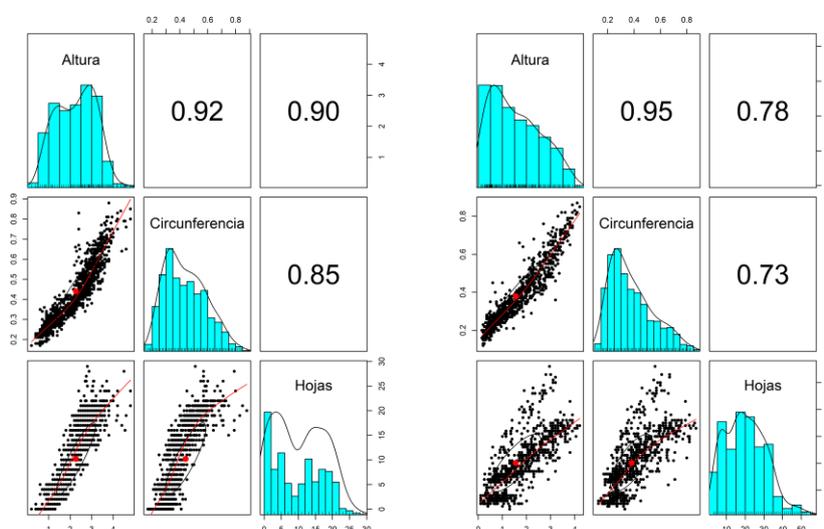


Tabla 3. P-valor para las correlaciones entre las variables biométricas de las plantas hijas

	Altura	Circunferencia	Hojas
Altura		0.000	0.000
Circunferencia	0.000		0.000
Hojas	0.000	0.000	

Tabla 4. P-valor para las correlaciones entre las variables biométricas de las plantas nietas

	Altura	Circunferencia	Hojas
Altura		0.000	0.000
Circunferencia	0.000		0.000
Hojas	0.000	0.000	

Se puede observar en la ilustración anterior que existe un importante nivel de correlación entre las variables biométricas entre sí, tanto para las plantas hijas como para las plantas nietas. Es decir, se puede explicar el comportamiento de una en base a la otra con una correlación bastante importante. Dicho nivel, vale aclarar, es mucho más elevado en plantas hijas que en plantas nietas (a excepción de la relación entre la altura y la circunferencia del pseudotallo, la cual es más importante en plantas nietas que en plantas hijas).

Sin embargo, no se puede dejar de lado que la combinación de las variables biométricas entre sí mismas, por simple sentido lógico, producen valores de correlación muy altos, es decir: mientras más alta es una planta, más grueso se puede esperar que sea su pseudotallo, o que, mientras más tiempo transcurra, se espera que broten más hojas de dicha planta.

Además, de acuerdo con la Tabla 3 y la Tabla 4, todas las correlaciones entre dichas variables son significativas, lo que implica que ese nivel de correlación no se debe al azar puro.

3.2.2.2 Nivel de correlación entre variables edáficas.

La Figura 19 y la Figura 20 se construyeron en base a los datos de las variables edáficas obtenidos a lo largo del proceso de medición durante el proyecto de investigación.

Figura 19. Panel de diagramas pareados entre las variables edáficas para las plantas hijas

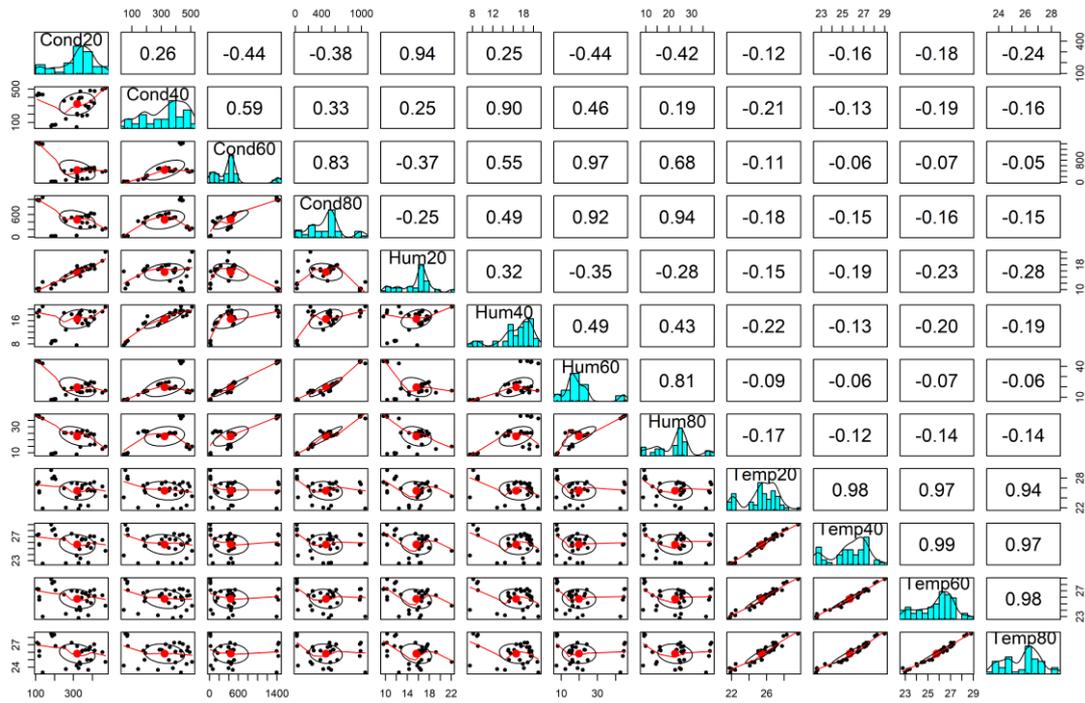


Tabla 5. P-valor para la correlación de cada combinación de variables edáficas para las plantas hijas

	Cond 20	Cond 40	Cond 60	Cond 80	Hum 20	Hum 40	Hum 60	Hum 80	Temp 20	Temp 40	Temp 60	Temp 80
Cond20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Cond40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Cond60	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.003
Cond80	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Hum20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Hum40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Hum60	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.001
Hum80	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Temp20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Temp40	0.000	0.000	0.001	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000
Temp60	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Temp80	0.000	0.000	0.003	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000

Figura 20. Panel de diagramas pareados entre variables edáficas para plantas nietas

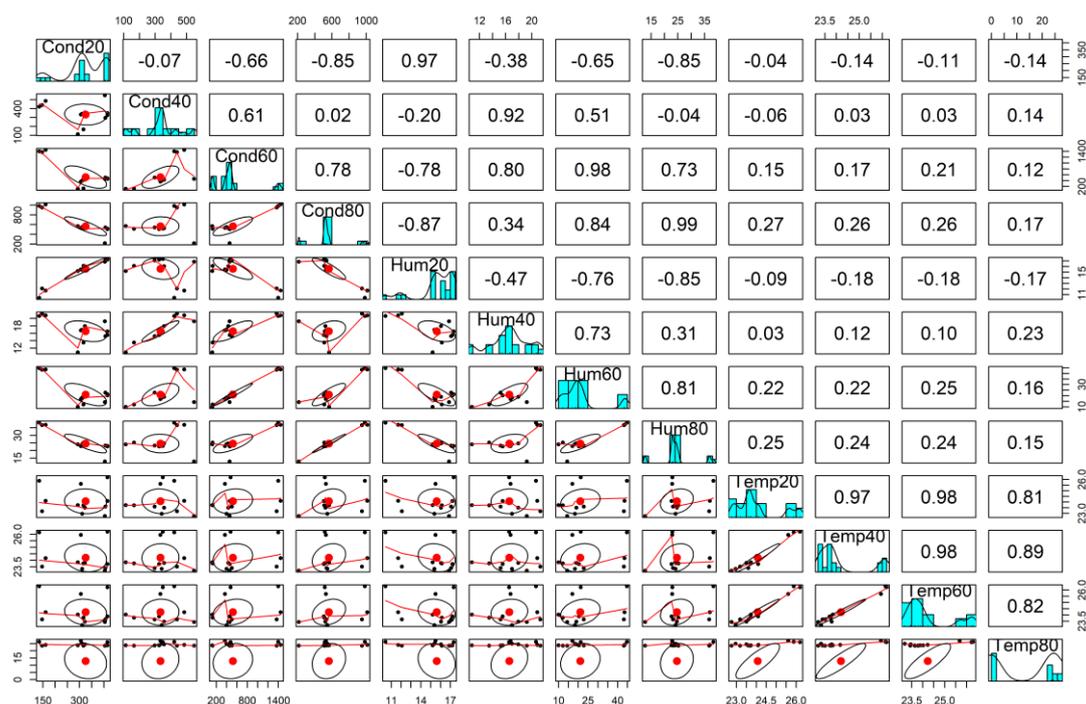


Tabla 6. P-valor para la correlación de cada combinación de variables edáficas para las plantas nietas

	Cond 20	Cond 40	Cond 60	Cond 80	Hum 20	Hum 40	Hum 60	Hum 80	Temp 20	Temp 40	Temp 60	Temp 80
Cond20		0.014	0.000	0.000	0.000	0.000	0.000	0.000	0.106	0.000	0.000	0.000
Cond40	0.014		0.000	0.325	0.000	0.000	0.000	0.140	0.036	0.322	0.238	0.000
Cond60	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Cond80	0.000	0.325	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Hum20	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.001	0.000	0.000	0.000
Hum40	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.274	0.000	0.000	0.000
Hum60	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000
Hum80	0.000	0.140	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000
Temp20	0.106	0.036	0.000	0.000	0.001	0.274	0.000	0.000		0.000	0.000	0.000
Temp40	0.000	0.322	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000
Temp60	0.000	0.238	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000
Temp80	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	

Se puede observar que existe un nivel de correlación importante entre las variables de conductividad eléctrica medidas a distintas profundidades tanto para las plantas hijas como para las plantas nietas, excepto por el nivel de correlación presente entre la conductividad eléctrica a 20 y 40 cm de profundidad (la cual es prácticamente despreciable para las plantas nietas).

También se puede observar en ambos tipos de planta que los valores del índice de correlación entre las variables de humedad a distintos niveles de profundidad no

siguen un patrón constante: entre dos únicos niveles de profundidad puede tratarse de una correlación negativa, mientras que entre otros dos de una positiva bastante fuerte.

Además, llama la atención que ninguna de las tres variables edáficas (independiente del nivel de profundidad en el que fueron medidas) tiene un nivel de correlación importante con alguna otra, a excepción de la humedad y la conductividad eléctrica: a los mismos niveles de profundidad el índice de correlación supera el valor de 0.9 en plantas hijas y nietas. Básicamente, se podría afirmar que la humedad del suelo determina su nivel de conductividad eléctrica.

Por otra parte, la temperatura a distintos niveles de profundidad tiene un nivel de correlación muy alto para ambos tipos de planta: el índice de correlación entre todas ellas es prácticamente de 1 en el caso de las plantas hijas, y no es inferior a 0.8 en el caso de las plantas nietas. Lo que implicaría que la temperatura en cualquier nivel de profundidad entre 0 y 80 cm define la temperatura de los otros niveles.

En los apartados siguientes que conforman el resto del capítulo se muestra el mismo tipo de gráfico presentado para analizar la correlación entre las variables de medición del suelo, pero centrado esta vez en analizar el nivel de correlación de cada una de estas variables edáficas (conductividad eléctrica del suelo, humedad del suelo y temperatura del suelo) con las variables biométricas de la planta de banano (altura de la planta, circunferencia del pseudotallo, número acumulado de hojas brotadas).

3.2.2.3 Conductividad eléctrica del suelo.

Figura 21. Panel de diagramas pareados entre la conductividad eléctrica del suelo a distintos niveles de profundidad y las variables biométricas de las plantas hijas

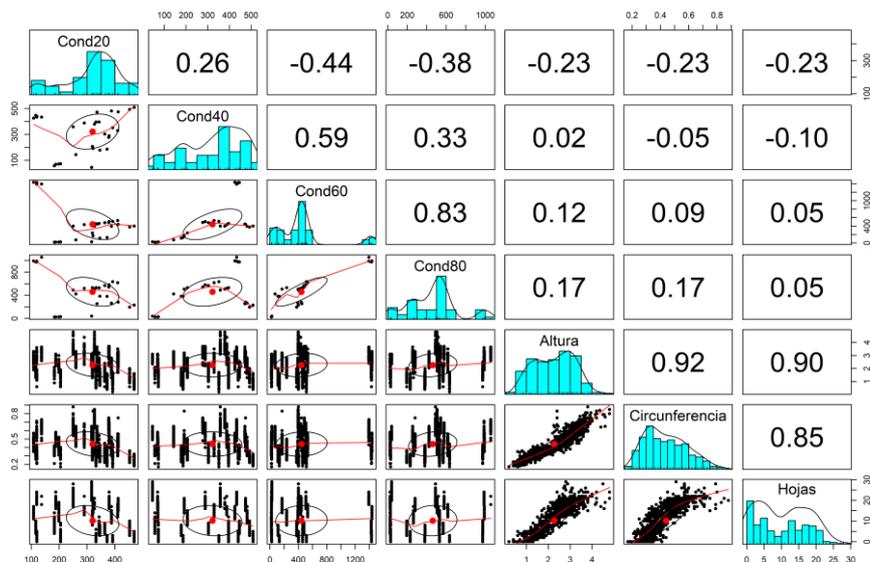


Tabla 7. P-valor para la correlación de cada combinación de la conductividad eléctrica del suelo con las variables biométricas de las plantas hijas

	Cond20	Cond40	Cond60	Cond80	Altura	Circunf.	Hojas
Cond20		0.000	0.000	0.000	0.000	0.000	0.000
Cond40	0.000		0.000	0.000	0.210	0.007	0.000
Cond60	0.000	0.000		0.000	0.000	0.000	0.022
Cond80	0.000	0.000	0.000		0.000	0.000	0.008
Altura	0.000	0.210	0.000	0.000		0.000	0.000
Circunf.	0.000	0.007	0.000	0.000	0.000		0.000
Hojas	0.000	0.000	0.022	0.008	0.000	0.000	

Figura 22. Panel de diagramas pareados entre la conductividad eléctrica del suelo a distintos niveles de profundidad y las variables biométricas de las plantas nietas

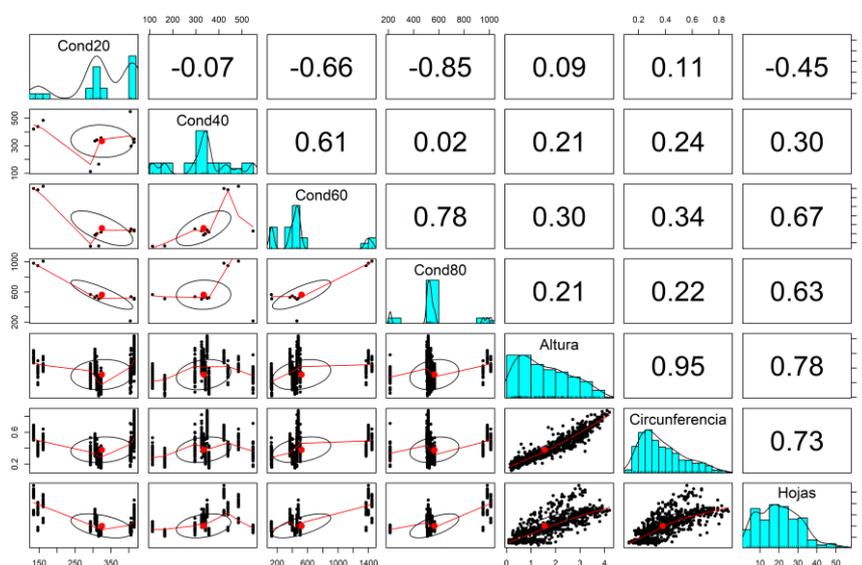


Tabla 8. P-valor para la correlación de cada combinación de la conductividad eléctrica del suelo con las variables biométricas de las plantas nietas

	Cond20	Cond40	Cond60	Cond80	Altura	Circunf.	Hojas
Cond20		0.014	0.000	0.000	0.002	0.000	0.000
Cond40	0.014		0.000	0.325	0.000	0.000	0.000
Cond60	0.000	0.000		0.000	0.000	0.000	0.000
Cond80	0.000	0.325	0.000		0.000	0.000	0.000
Altura	0.002	0.000	0.000	0.000		0.000	0.000
Circunf.	0.000	0.000	0.000	0.000	0.000		0.000
Hojas	0.000	0.000	0.000	0.000	0.000	0.000	

Como se puede apreciar tanto en la Figura 21 como en la Figura 22, no existe un nivel de correlación importante (ni un tipo único de correlación) entre la conductividad del suelo y la altura o el grosor del pseudotallo de las plantas de banano hijas o nietas.

Por otra parte, los valores para la correlación entre esta variable y alguna de las variables biométricas para las plantas hijas alcanza un máximo de apenas 0.23 (ya sea esta positiva o negativa), siendo que la mayoría de los valores son muy cercanos a 0 (una correlación inexistente). Aunque este nivel de correlación en plantas nietas sea más elevado, no es igual de importante que el observado en casos anteriores (como el existente entre las temperaturas a distinta profundidad).

En lo que respecta al número acumulado de hojas brotadas, el nivel de correlación entre esta variable y la conductividad eléctrica del suelo sí es bastante más alto a comparación de la altura y la circunferencia del pseudotallo. Sin embargo, dado que se identificó un problema serio con la validez y lógica real de la data capturada de esta variable, cualquier deducción que se pueda hacer sobre el nivel de correlación desarrollada con respecto a cualquier variable edáfica tiene poca validez. Es por ello que en los apartados que recogen el análisis de correlación de las otras dos variables edáficas se prescinde de emitir observaciones o deducciones con relación a esta variable.

Finalmente, centrando un poco más el análisis en las dos primeras variables biométricas de la planta y basándose en lo observado anteriormente, se puede inferir que la conductividad del suelo no es una variable que pueda explicar por sí sola qué tan alta puede llegar a ser una planta de banano o qué tan grueso será su pseudotallo.

3.2.2.4 Contenido volumétrico de agua en el suelo.

Figura 23. Panel de diagramas pareados entre el contenido volumétrico de agua en el suelo a distintos niveles de profundidad y las variables biométricas de las plantas hijas

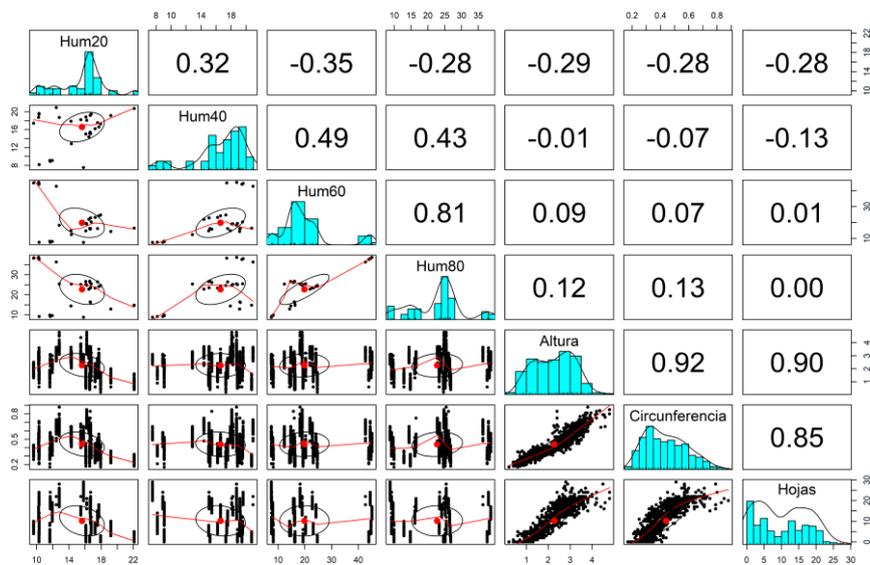


Tabla 9. P-valor para la correlación de cada combinación de la humedad del suelo con las variables biométricas de las plantas hijas

	Hum20	Hum40	Hum60	Hum80	Altura	Circunf.	Hojas
Hum20	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Hum40	0.000	0.000	0.000	0.000	0.483	0.000	0.000
Hum60	0.000	0.000	0.000	0.000	0.000	0.000	0.558
Hum80	0.000	0.000	0.000	0.000	0.000	0.000	0.932
Altura	0.000	0.483	0.000	0.000	0.000	0.000	0.000
Circunf.	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Hojas	0.000	0.000	0.558	0.932	0.000	0.000	0.000

Figura 24. Panel de diagramas pareados entre el contenido volumétrico de agua en el suelo a distintos niveles de profundidad y las variables biométricas de las plantas nietas

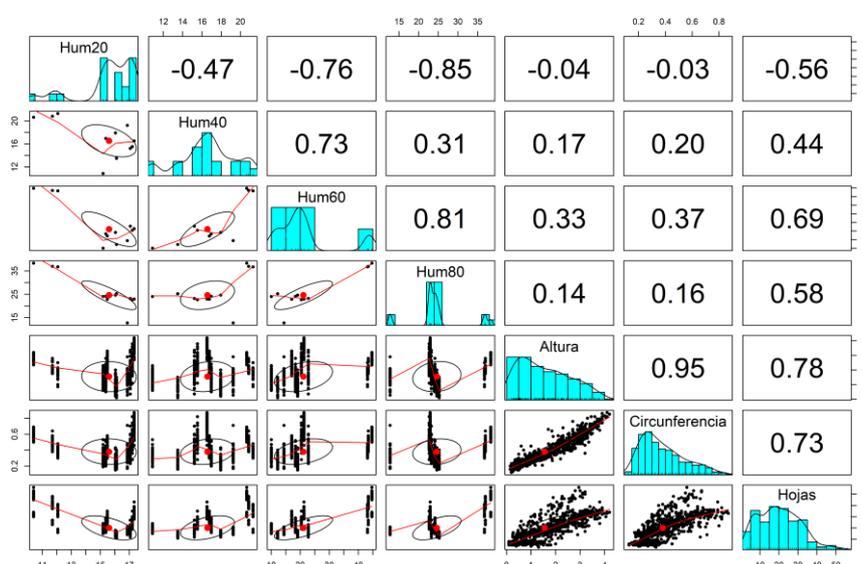


Tabla 10. P-valor para la correlación de cada combinación de la humedad del suelo con las variables biométricas de las plantas nietas

	Hum20	Hum40	Hum60	Hum80	Altura	Circunf.	Hojas
Hum20	0.000	0.000	0.000	0.000	0.199	0.298	0.000
Hum40	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Hum60	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Hum80	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Altura	0.199	0.000	0.000	0.000	0.000	0.000	0.000
Circunf.	0.298	0.000	0.000	0.000	0.000	0.000	0.000
Hojas	0.000	0.000	0.000	0.000	0.000	0.000	0.000

En lo que respecta al contenido volumétrico de agua en el suelo, se puede observar que la correlación desarrollada entre esta variable edáfica y las variables biométricas de altura y circunferencia del pseudotallo de ambos tipos de planta es

prácticamente una calca de la desarrollada por la variable edáfica anterior (la conductividad eléctrica del suelo).

Es así como, tal como se observa en la Figura 23 y la Figura 24, no existe un nivel de correlación significativo entre la altura de las plantas hijas y nietas y el contenido volumétrico de agua en el suelo medido en alguno de los cuatro niveles de profundidad establecidos.

Además, tampoco se puede establecer un único tipo de correlación entre estas variables, ya que a ciertos niveles de profundidad desarrollan una correlación directa mientras que en otros esta es inversa.

En resumen, se puede llegar a la misma conclusión que con la variable edáfica anterior: el contenido volumétrico de agua en el suelo no es una variable que pueda explicar en sí misma qué altura puede alcanzar una planta de banano, o cuan grueso será su pseudotallo.

Lo mencionado previamente es algo a remarcar teniendo en cuenta que ya se había deducido en el análisis correspondiente a la correlación entre variables edáficas que una podía ser explicada en base a la otra.

3.2.2.5 Temperatura del suelo

Figura 25. Panel de diagramas pareados entre la temperatura del suelo a distintas profundidades y las variables biométricas de las plantas hijas

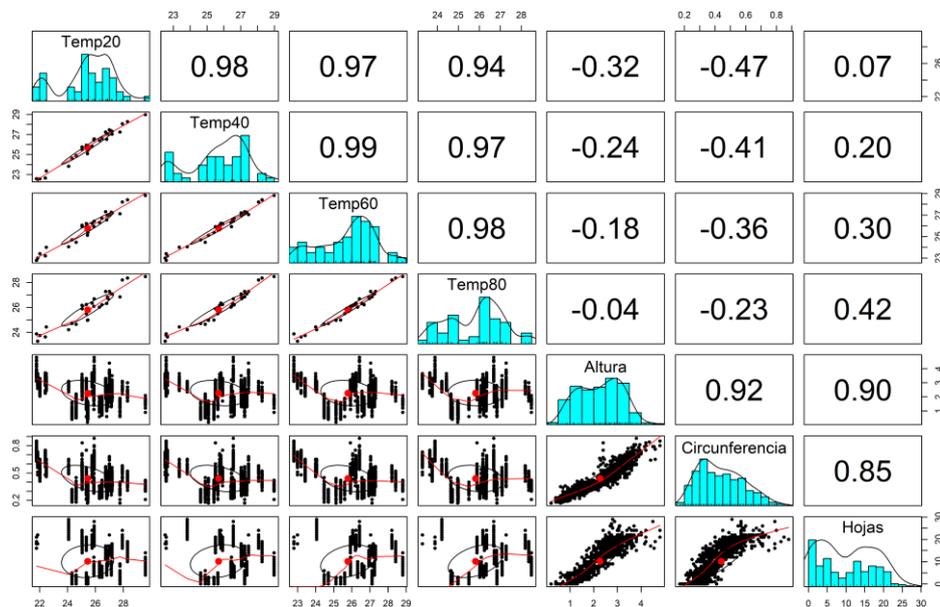


Tabla 11. P-valor para la correlación de cada combinación de la temperatura del suelo con las variables biométricas de las plantas hijas

	Temp20	Temp40	Temp60	Temp80	Altura	Circunf.	Hojas
Temp20		0.000	0.000	0.000	0.000	0.000	0.001
Temp40	0.000		0.000	0.000	0.000	0.000	0.000
Temp60	0.000	0.000		0.000	0.000	0.000	0.000
Temp80	0.000	0.000	0.000		0.019	0.000	0.000
Altura	0.000	0.000	0.000	0.019		0.000	0.000
Circunf.	0.000	0.000	0.000	0.000	0.000		0.000
Hojas	0.001	0.000	0.000	0.000	0.000	0.000	

Figura 26. Panel de diagramas pareados entre la temperatura del suelo y las variables biométricas de las plantas nietas

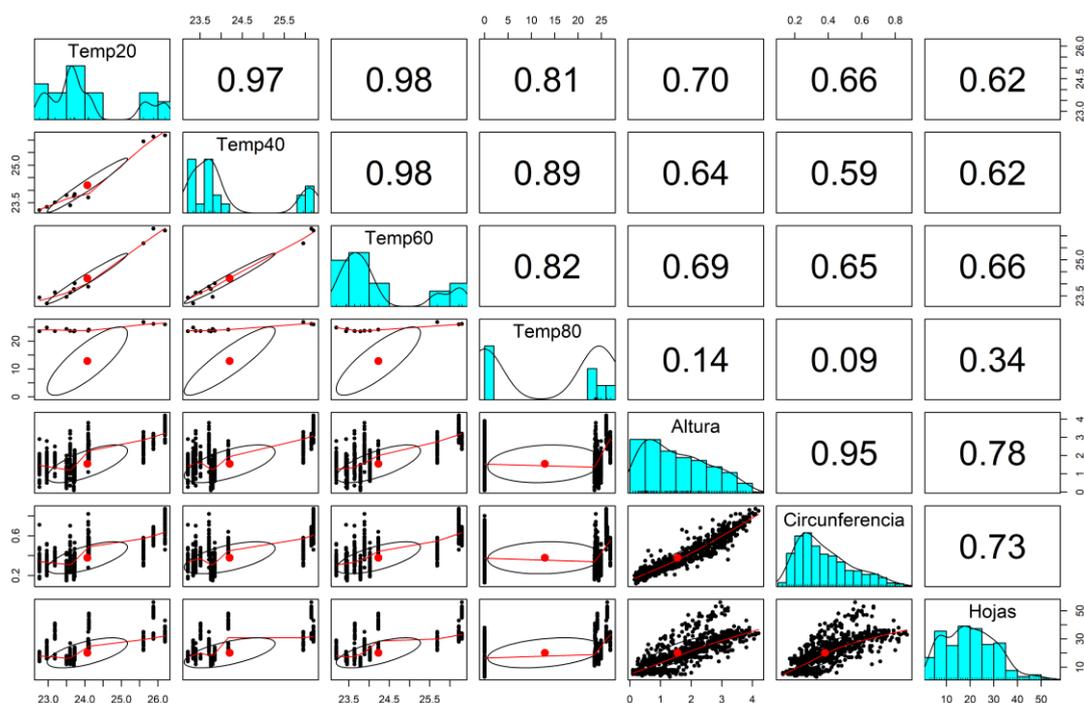


Tabla 12. P-valor para la correlación de cada combinación de la temperatura del suelo con las variables biométricas de las plantas nietas

	Temp20	Temp40	Temp60	Temp80	Altura	Circunf.	Hojas
Temp20		0.000	0.000	0.000	0.000	0.000	0.000
Temp40	0.000		0.000	0.000	0.000	0.000	0.000
Temp60	0.000	0.000		0.000	0.000	0.000	0.000
Temp80	0.000	0.000	0.000		0.000	0.000	0.000
Altura	0.000	0.000	0.000	0.000		0.000	0.000
Circunf.	0.000	0.000	0.000	0.000	0.000		0.000
Hojas	0.000	0.000	0.000	0.000	0.000	0.000	

En lo que respecta a la última variable edáfica pendiente de analizar, la Figura 25 y la Figura 26 muestran que la variable biométrica de altura comparte niveles de

correlación más importantes con la temperatura que los que había notado previamente dicha variable con las otras dos variables edáficas, tanto para plantas hijas como para plantas nietas, siendo más altos los presentados en relación con la altura de las segundas que con la altura de las primeras.

Sin embargo, se debe notar que, si bien tienen niveles de correlación más altos, su tipo varía según la clase de planta que se trate. Así pues, mientras que los valores de todos los índices denotan una correlación inversa entre la temperatura a cualquier nivel de profundidad y la altura de las plantas hijas, para las plantas nietas denotan una correlación de tipo directa.

Lo mismo sucede con la circunferencia del pseudotallo de ambos tipos de planta: el nivel de correlación con la temperatura del suelo es más alto que el alcanzado con la conductividad eléctrica o el contenido volumétrico de agua en el suelo (a excepción del indicado a 80 cm de profundidad, donde no es significativo), y en cuanto al tipo, para las plantas hijas se trata nuevamente de una correlación inversa y para las plantas nietas de una directa.

Por otra parte, cabe destacar que existe una tendencia clara para la altura y el grosor del pseudotallo de las plantas hijas: el nivel de correlación con esta variable edáfica empieza a ser menos significativa a mayores niveles de profundidad. Esta tendencia no está presente como tal en el caso de las plantas nietas, salvo por el índice de nivel de correlación de ambas variables biométricas con la temperatura del suelo a 80 cm de profundidad, donde el valor de este es menor a 0.15.

Para finalizar, se puede deducir que, a pesar de mostrar mayores niveles de correlación que las otras dos variables edáficas, la temperatura del suelo no es una variable que por sí sola pueda explicar enteramente (o de manera suficiente) el desarrollo de una planta de banano en función de su altura y grosor de su pseudotallo.

Capítulo 4

Resultados

Durante el proceso de modelación de los datos extraídos de la medición de las variables biométricas de la planta, se observó que existe correlación entre las distintas variables. Además, se debe tomar en cuenta que las observaciones tomadas de las variables biométricas en el tiempo para una misma planta no serán independientes entre sí, pues existirá correlación entre ellas.

Por lo tanto, para tratar de diseñar un modelo que reproduzca los valores observados resulta necesario tener en cuenta la varianza generada desde dos fuentes distintas: la creada por la diferencia entre los puntos como tal, y la creada por los grupos de puntos correspondientes a cada planta. Así pues, se habla de una varianza generada entre grupos y, además, dentro de los grupos.

Tomando en consideración todo lo mencionado anteriormente, se optó por generar un MLM, "Mixed Linear Model", Modelo Lineal Mixto, que toma en cuenta no solo la varianza generada por la diferencia dentro de los grupos, sino también la generada entre puntos.

Se generaron tres modelos predictivos para el desarrollo y crecimiento de la planta de banano, uno para cada variable biométrica estudiada. Para el desarrollo y selección de cada uno de dichos modelos, se partió de la generación de modelos prototipo bajo la iteración en el cambio de variables predictivas. La ecuación general que describe cada modelo es la siguiente:

$$y_{ijkt} = \beta_0 + \alpha_i + \beta_{1j}TipoRiego_j + \beta_{2k}TipoSuelo_k + \beta_3Tiempo_t + \beta_{4j}(TipoRiego_j \times Tiempo_t) + \beta_{5k}(TipoSuelo_k \times Tiempo_t) + \beta_6(TipoRiego_j \times TipoSuelo_k) + \gamma_1x_{1jkt} + \dots + \gamma_px_{pjkt} + \epsilon_{ijkt},$$
$$i = 1,2, \dots, n; \quad j = 1,2; \quad k = 1,2,3; \quad t = 1, \dots, 5,$$

Donde:

- y_{ijkt} es la variable biométrica respuesta para la planta i para el tipo de irrigación j y tipo de suelo k para el tiempo t .
- β_0 es el promedio de los interceptos de todas las plantas
- α_i es el intercepto de la planta i . Se asume que $\alpha_i \sim N(0, \sigma_1^2)$.

- β' s representa los efectos fijos del tipo de riego, suelo, tiempo y sus interacciones.
- z_{ij} representa el tiempo; el tiempo j y planta i
- $\gamma_1 \dots \gamma_p$ es el coeficiente de cada variable predictora: Conductividad eléctrica (Cond20, Cond40, Cond60, Cond80), contenido volumétrico de agua en el suelo (Hum20, Hum40, Hum60, Hum80), temperatura del suelo (Temp20, Temp40, Temp60, Temp80).
- $x_{1jkt} \dots x_{pjkt}$ representa los valores de cada variable predictora para cada parcela en el tiempo t .
- ϵ_{ijkt} es el error. Se asume $\epsilon_{ijkt} \sim N(0, \sigma^2)$ y que son independientes de α_i .

Como detalle del proceso de generación y selección de los modelos matemáticos, vale indicar que se partió en principio de generar un modelo matemático por cada nivel de profundidad en el que se habían tomado medidas de las variables edáficas de cada parcela, adicional a uno que no incluía ninguna de estas. Se tenía como base, entonces, un total de 5 modelos iniciales para cada una de las variables biométricas.

Posterior a ello y, teniendo en cuenta el análisis previo desarrollado en el Capítulo 3, en el que se observó que algunas variables incluidas en los modelos propuestos tenían un nivel de correlación significativo con otras del mismo modelo, se aplicó una depuración de variables a cada uno de ellos haciendo uso del factor de inflación de la varianza. Finalizada esta depuración, se realizó la prueba de significancia individual de variables predictivas (que se obtiene de ejecutar un resumen estadístico a un modelo de regresión en R) y una prueba adicional del factor de inflación de la varianza que certificara la inexistencia de colinealidad entre variables, tal como se muestra en la Figura 27 y en la Figura 28.

Figura 27. Tabla de resumen estadístico para la significancia individual de variables predictivas

	Value	Std.Error	DF	t-value	p-value
(Intercept)	0.16971999	0.03258898	893	5.20790	0.0000
Tiempo	0.04907141	0.00094489	893	51.93339	0.0000
factor(TipoSuelo)PR	0.04170919	0.01283185	247	3.25044	0.0013
factor(TipoSuelo)R	0.04399562	0.01029432	247	4.27378	0.0000
factor(TipoRiego)Tradicional	-0.06237634	0.00852335	893	-7.31829	0.0000
Cond40	0.00004250	0.00003189	893	1.33249	0.1830
Temp40	0.00519677	0.00118440	893	4.38766	0.0000
Tiempo: factor(TipoSuelo)PR	-0.00620902	0.00159253	893	-3.89883	0.0001
Tiempo: factor(TipoSuelo)R	-0.00279055	0.00132786	893	-2.10153	0.0359

Figura 28. Prueba de factor de inflación de la varianza

	GVIF	Df	GVIF^(1/(2*Df))
Tiempo	114.58505	1	10.704441
factor (TipoSuelo)	1367.91810	2	6.081560
factor (TipoRiego)	3.81671	1	1.953640
Hum20	181.74773	1	13.481385
Cond20	71.75233	1	8.470675
Temp20	30.53768	1	5.526091
Tiempo: factor (TipoSuelo)	6995.03467	2	9.145290

Finalmente, de entre los modelos propuestos que cumplieran satisfactoriamente con las pruebas anteriormente mencionadas, se eligió aquel que tuviera el R² condicional y marginal más alto, y el menor valor de AIC (criterio de información de Akaike).

Vale aclarar que, dado que el riego mejorado no se aplicó a las plantas hijas hasta la quinta toma de medidas (es decir, hasta pasados 7.25 meses de iniciada la medición, véase Figura 6), estas no fueron sujetos de prueba en su totalidad del experimento diseñado, ya que, tal como se explicó en el Capítulo 2, las plantas a analizar (fuesen hijas o nietas) debían experimentar una variación tanto en el tipo de suelo, como en el tipo de riego, y estas no experimentaron dicha variación en ninguna de las parcelas hasta el final de su ciclo de vida. Por lo tanto, los modelos propuestos y analizados fueron construidos en base a la data obtenida específicamente de plantas nietas.

4.1 Altura

Para la primera variable biométrica establecida, el modelo de regresión seleccionado es el siguiente:

$$\hat{y}_{ijkt} = \beta_0 + \alpha_i + \beta_{1j}TipoRiego_j + \beta_{2k}TipoSuelo_k + \beta_3Tiempo_t + \beta_{5k}(TipoSuelo_k \times Tiempo_t),$$

Dicho modelo posee un R² marginal igual a 0.715 y un R² condicional igual a 0.92. A continuación, la Tabla 13 resume el valor de los coeficientes para cada una de las variables del modelo seleccionado.

Tabla 13. Resumen del modelo de regresión lineal mixto para predecir la altura de las plantas nietas

Variable	Valor	Error estándar	p-Valor
Intercepto [β_0]	-1.203	0.066	0.000
Tiempo [β_3]	0.263	0.004	0.000
Tipo de suelo [β_{21}]: PR	0.709	0.111	0.000
Tipo de riego [β_{11}]: Tradicional	-0.143	0.036	0.000
Tiempo * Tipo de suelo [β_{51}]: PR	-0.039	0.007	0.000

El resumen arriba mostrado toma como base comparativa el supuesto de una planta creciendo en un tipo de suelo muy retentivo (MR) y aplicándole un riego mejorado. Entonces, los coeficientes de las variables de Tipo de suelo y Tipo de riego

colocadas en dicha tabla muestran su comportamiento en comparación a este supuesto. Por lo tanto, de acuerdo con el modelo seleccionado, se puede esperar que una planta que crece en un suelo poco retentivo llegue a ser 71 cm más alta que una planta que crece en un suelo muy retentivo.

En lo que respecta al riego, se espera que una planta que fue cultivada aplicándosele un riego mejorado alcance una altura 14 cm mayor a la que puede alcanzar una cultivada con un riego tradicional.

Además, de acuerdo con el coeficiente de la interacción del tiempo y el tipo de suelo, se puede esperar que, en promedio, para cada aumento de un mes en el tiempo, la altura de una planta de banano que crece en suelos poco retentivos disminuirá en 4 centímetros en comparación con las plantas de banano que crecen en suelos retentivos.

En términos generales, la planta de banano que crece en suelos poco retentivos tendrá una tasa de crecimiento más baja en términos de altura a lo largo del tiempo en comparación con las plantas que crecen en suelos retentivos.

En resumen, el modelo de regresión establece que aquellas plantas que se cultiven en un suelo de tipo poco retentivo y con un riego mejorado, alcanzarán una altura mayor que cualquier planta cultivada en alguno de los otros tipos de suelo y con un riego tradicional (además, que la diferencia en altura entre ambos tipos de suelo es no significativa).

a. Diagnóstico del modelo de regresión

La diagnóstico del modelo previamente descrito se llevó a cabo haciendo uso de dos herramientas gráficas estadísticas. Así pues, la Figura 29 recoge el gráfico de residuos (eje de ordenadas) vs valores ajustados (eje de abscisas), y la Figura 30 recoge la prueba gráfica cuantil-cuantil de la distribución de los residuos respecto a la distribución normal teórica.

Figura 29. Gráfico de residuos vs valores ajustados para el modelo de regresión lineal mixto para la altura de las plantas nietas

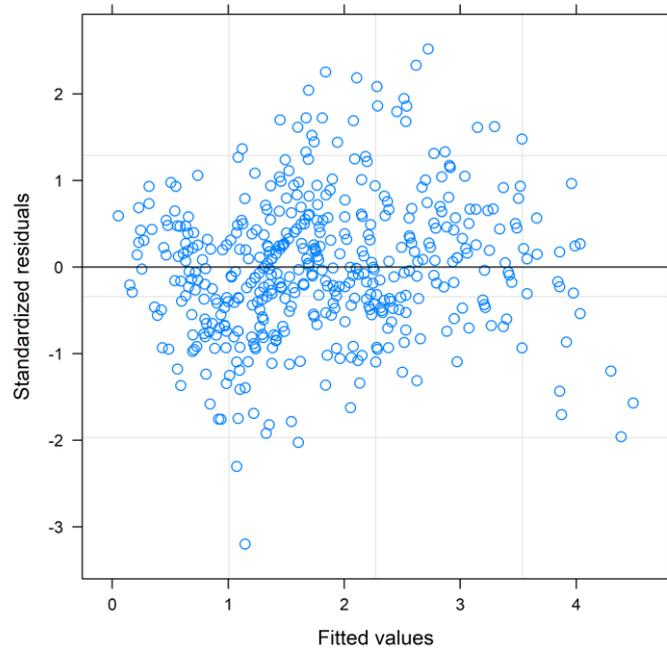
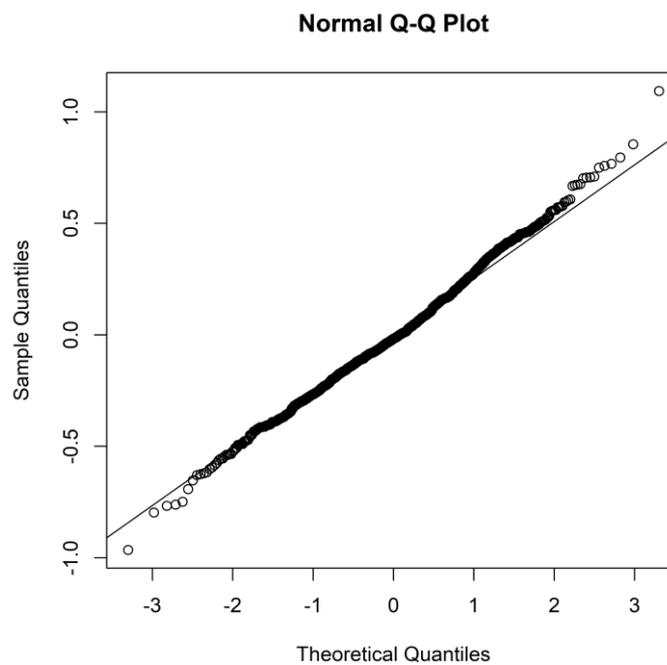


Figura 30. Gráfico cuantil-cuantil para el modelo de regresión lineal mixto para la altura de las plantas nietas



En lo que respecta al primero de ellos, no se observa un patrón evidente en la distribución de los puntos, ni alguna tendencia presente en ellos. Además, la varianza se presenta constante a lo largo del diagrama, lo que certifica la homocedasticidad que supone un modelo de regresión válido.

Por otra parte, se observa que, en el gráfico de prueba de normalidad de los residuos, estos se ajustan bien a la recta de normal, a excepción de algunos puntos en ambas colas de la línea.

Además, la aplicación del test Durbin-Watson presenta un valor de 1.9162, lo que supone un valor muy cercano al ideal (2) y una prueba satisfactoria más para la validez del modelo desarrollado.

4.2 Circunferencia

En lo que corresponde a la circunferencia del pseudotallo, el modelo seleccionado fue el siguiente:

$$\hat{y}_{ijkt} = \beta_0 + \alpha_i + \beta_{2k} \text{TipoSuelo}_k + \beta_3 \text{Tiempo}_t + \beta_{5k} (\text{TipoSuelo}_k \times \text{Tiempo}_t)$$

El modelo desarrollado para la circunferencia tiene un R^2 marginal de 0.607 y un R^2 condicional de 0.899. Se presentan los coeficientes de las variables de dicho modelo en la Tabla 14:

Tabla 14. Resumen del modelo de regresión lineal mixto para predecir la circunferencia de las plantas nietas

Variable	Valor	Error estándar	p-Valor
Intercepto [β_0]	-0.064	0.013	0.000
Tiempo [β_3]	0.038	0.001	0.000
Tipo de suelo [β_{21}]: PR	0.089	0.017	0.000
Tipo de suelo [β_{22}]: R	0.012	0.019	0.538
Tiempo * Tipo de suelo [β_{52}]: R	0.002	0.001	0.042

Tal como sucede con el modelo para la altura de las plantas, el seleccionado para la circunferencia supone que el grosor esperado del pseudotallo de las plantas nietas tendrá un valor negativo siempre que no se vea afectada por el resto de las variables incluidas en dicho modelo, lo cual una vez más supone una imposibilidad en el mundo real.

Nuevamente se espera que una planta cultivada en un suelo poco retentivo desarrolle un pseudotallo 9 cm más grueso que una cultivada en un suelo muy retentivo. Sin embargo, a diferencia de lo que se observa con el modelo desarrollado para la variable de altura, no existe una diferencia significativa entre la circunferencia del pseudotallo de una planta cultivada con riego tradicional y una que fue cultivada con un riego mejorado, aunque en principio se esperase un pseudotallo más grueso para aquellas a las que se le aplicó el segundo tipo de riego (ver Apéndices).

Finalmente, se observa que el coeficiente de la interacción entre el tiempo y un tipo de suelo retentivo es positivo e igual a 0.002, lo que indica que la circunferencia del pseudotallo de una planta de banano que crece en suelos retentivos aumentará en promedio en 2 milímetros por cada aumento de un mes en el tiempo. Esto significa que la tasa de crecimiento en términos de la circunferencia del pseudotallo en las plantas de banano que crecen en suelos retentivos es mayor en comparación con las plantas que crecen en suelos poco retentivos a lo largo del tiempo.

Vale recalcar que, si bien el modelo contiene una variable no significativa (tipo de suelo retentivo), esta no ha sido depurada del mismo debido a la significancia que supone su interacción con el tiempo.

Diagnos de modelo de regresión

Se usaron las mismas herramientas para la diagnos de este segundo modelo de regresión seleccionado. La Figura 31 y la Figura 32 recogen los gráficos de dichas herramientas aplicadas a este modelo de regresión:

Figura 31. Gráfico de residuos vs valores ajustados para el modelo de regresión lineal mixto para la circunferencia del pseudotallo de las plantas nietas

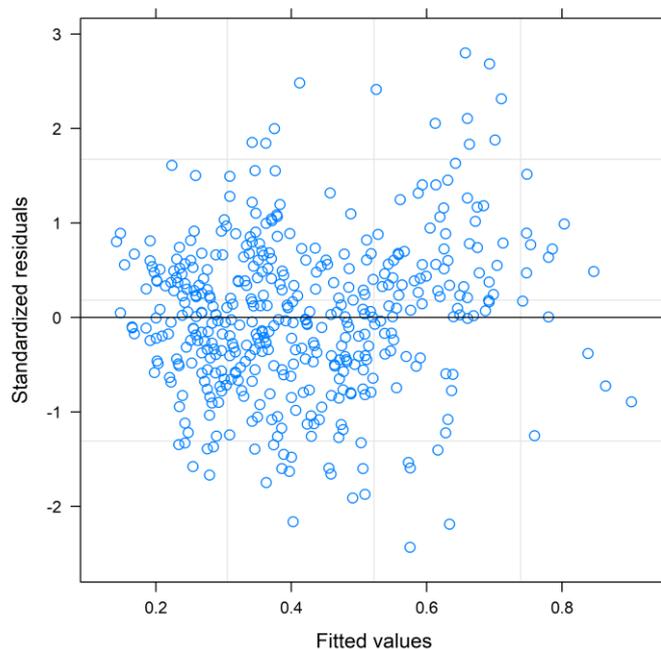
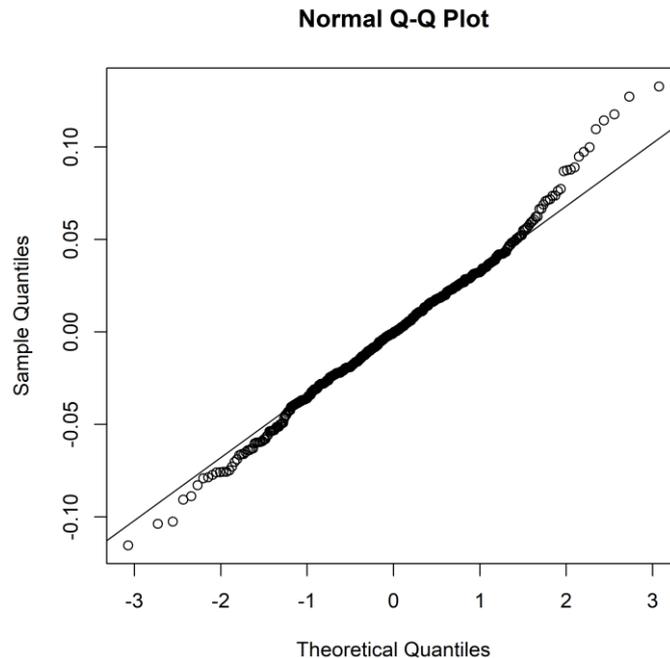


Figura 32. Gráfico cuantil-cuantil para el modelo de regresión lineal mixto para la circunferencia del pseudotallo de las plantas nietas



Tal como sucedía con el modelo de regresión seleccionado para la altura, el primer gráfico presentado en la diagnosis de modelo de este apartado no muestra que exista algún patrón o tendencia en la distribución de los puntos, en cambio, sí grafica una varianza relativamente constante a lo largo del diagrama.

Y, casi como una calca a lo sucedido con los residuos del modelo propuesto para la altura, los residuos de este modelo se ajustan bien a la recta de la normal teórica en el segundo gráfico, a excepción de algunos puntos en las colas que no se ajustan del todo bien a la recta de la normal.

Por otra parte, el test de Durbin-Watson aplicado a este modelo de regresión presenta un valor de 1.6545, el cual, si bien no muestra un valor tan cercano al ideal como el del modelo desarrollado para la altura de la planta, sigue siendo un valor cercano a este.

4.3 TEFD

El número acumulado de hojas brotadas, denominado en este apartado como la variable de Tasa de Emisión Foliar Diaria (TEFD), es una variable que se decidió no tomar en cuenta para generar un modelo de regresión lineal mixto debido a la poca fiabilidad de la data con la que se cuenta.

Así pues, tal como declaraba Carbonell (1983), es necesario contar con data confiable para poder desarrollar un modelo de regresión también confiable. Característica que no se puede asegurar con la data obtenida y analizada.

De hecho, tal como se puede observar en la Figura 16 del Capítulo 3, existe una gran cantidad de valores para el número de hojas brotadas que se recogen como atípicos en el box-plot allí representado. Estos mismos atípicos causan, además que las cajas diagramadas para las plantas nietas sembradas en suelos poco retentivos en la Figura 17 del mismo capítulo sean evidentemente más largas que para los otros tipos de suelo, lo que indica una mayor variabilidad.





Conclusiones

Existen determinadas variables que condicionan las dimensiones que puede alcanzar una planta de banano a lo largo de su crecimiento, las cuales, además, configuran un modelo matemático capaz de predecir dicho crecimiento, aunque esto no fuese parte esencial de los objetivos que justificaron el desarrollo de la investigación.

Las variables en cuestión son, de acuerdo con los modelos seleccionados para la altura y el grosor del pseudotallo: el tiempo, el tipo de suelo y el tipo de riego aplicado.

Estos resultados coinciden con lo expuesto en la justificación del experimento: sí es importante el tipo de suelo en el que se siembra una planta de banano, y, más allá de la cantidad de agua que se le suministre a una planta, es más importante el método en el que se realice.

Estos resultados coinciden, además, con diversos medios de divulgación científica y escritos relacionados al mundo de la agricultura. Así pues, tal como recoge Soto (1992) las plantas de banano necesitan ser sembradas en suelo de textura ligeramente arenosa, con buena capacidad de drenaje (que sean poco retentivos).

Por otro lado, todas las variables edáficas no están incluidas en el modelo final. Esto debido a la redundancia que supone el mantener las variables de humedad y conductividad eléctrica al mismo tiempo, por un lado, y porque existe un grado de colinealidad alto entre las variables que ocasiona que ninguna de ellas (incluida la temperatura del suelo) pueda figurar en alguno de los modelos seleccionados.

Lo anteriormente comentado se validó con la prueba de inflación de la varianza, donde todos los modelos válidos suponían la ausencia completa de todas las variables edáficas. Ello supuso también la validación del uso de las herramientas gráficas en el análisis de los datos obtenidos, puesto que significó la confirmación de lo que se podía

observar en los paneles de dispersión entre variables (especialmente para la conductividad eléctrica y para la humedad).

Esta validez única de los modelos base (aquellos que excluían las variables edáficas) en la selección aplicada a los prototipos de modelo construidos con los datos correspondientes a las plantas nietas puede ser explicada por la ausencia de gran parte de datos de las variables edáficas en la segunda parte del experimento.

Es probable, por ello, que el efecto de las variables edáficas fuera absorbido por el efecto de variables categóricas relacionadas a estas. Por ejemplo, la cantidad de agua que tenga el suelo depende básicamente del nivel de retención que este posea, entonces, el efecto de la humedad del suelo es absorbido por la variable categórica “tipo de suelo”.

Ello ocasiona que esta variable resulte redundante, pues su efecto viene dado por otras variables en el modelo. Esto explicaría, además, los valores tan altos en la prueba de inflación de la varianza.

Es de destacar también que, pese a que en el Capítulo 3 se observa que los valores máximos tanto en altura como en grosor del pseudotallo de las plantas nietas fueron alcanzados por ejemplares cultivados en suelos retentivos, ambos modelos seleccionados proponen como suelo ideal el poco retentivo.

Esto podría explicarse en la dispersión propia de los valores medidos en cada tipo de suelo, ya que los valores correspondientes a las plantas cultivadas en suelos retentivos son mucho más dispersos que los obtenidos de plantas sembradas en parcelas con suelos poco retentivos.

Finalmente, es de destacar que los modelos de regresión generados para las plantas nietas son capaces de explicar hasta un 80% del total de variación de los valores de crecimiento y desarrollo de las plantas de banano.

Eso implica que, con incluir el tiempo, el tipo de suelo y el tipo de riego aplicado se puede consolidar un modelo efectivo para predecir el desarrollo y crecimiento de una planta de banano.

Recomendaciones

Observando las limitaciones que supuso la ausencia de datos para la construcción de modelos matemáticos para las plantas nietas, la primera recomendación parte del cuidado de la toma de datos. Es imprescindible contar con la mayor cantidad de data explotable posible.

También se puntualiza que no es importante solo contar con una cantidad significativa de datos para la elaboración de un modelo de regresión, sino que, además, esta debe ser confiable. Así pues, se enfatiza que además de contar con cantidad de data, esta debe ser de calidad.

Por otro lado, es importante que, de contar con los medios para hacerlo, se explore la posibilidad de integrar al experimento variables adicionales a las trabajadas en este proyecto, tales como: presencia de minerales específicos en el suelo, respiración microbiana, resistencia a la penetración, número de manojos de banano por planta, entre otras. Esto debido a que, si bien han sido abordadas en investigaciones previas relacionadas, sería ideal medir y estudiar estas variables en el suelo piurano.



Referencias bibliográficas

- Barbaro, L., Karlanian, M., & Mata, D. (2014). *Importancia del pH y la Conductividad Eléctrica (CE) en los sustratos para plantas*. Buenos Aires: Instituto Nacional de Tecnología Agropecuaria.
- Caicedo, O., Balmaseda, C., & Proaño, J. (2015). Irrigation schedules of banana (*Musa paradisiaca*) in San José 2 ranch, Los Ríos province, Ecuador. *Ciencias Técnicas Agropecuarias*, 18-22.
- Carbonell, E. (1983). *Regresión lineal: un enfoque conceptual y práctico*. Madrid: Instituto Nacional de Investigaciones Agrarias.
- Cremona, M. V., & Enriquez, A. S. (2020). Algunas propiedades del suelo que condicionan su comportamiento: El pH y la conductividad eléctrica. *Presencia XXXI*, 5-8.
- Denham, T., Haberle, S. G., Lentfer, C., Fullagar, R., Fuller, J., Porch, N., . . . Golson, J. (2003). *Multi-disciplinary Evidence for the Origins of Agriculture from 6950-6440 Cal BP at Kuk Swamp in the Highlands of New Guinea*. New York: Science.
- Escobar, H., Solís de Ovando, L., Contreras, D., Baginsky, C., Arenas, J., & Silva, H. (2018). Efecto de la disponibilidad de agua de riego en el intercambio gaseoso, rendimiento de semillas, biomasa y eficiencia del uso del agua en dos fenotipos de chíá establecidos en el valle de Azap, Arica, Chile. *Interciencia*, 55-61.
- FAO. (2019). *Textura del suelo*. Obtenido de FAO Web Site: http://www.fao.org/tempref/FI/CDrom/FAO_Training/FAO_Training/General/x6706s/x6706s06.htm
- FAO. (20 de Abril de 2020). *Banana facts and figures*. Obtenido de FAO: <http://www.fao.org/economic/est/est-commodities/bananas/bananafacts/en/#.X2Kva2hKjIU>
- Fischer, G., Torres, F., & Torres, J. (1999). Efecto de la temperatura del suelo sobre la planta. *Revista Comalfi*, 78-92.

- Fundación Española de la Nutrición. (2020). *Plátano*. Obtenido de Fundación Española de la Nutrición: <https://www.fen.org.es/MercadoAlimentosFEN/pdfs/platano.pdf>
- Galarnyk, M. (12 de Setiembre de 2018). *Understanding Boxplots*. Obtenido de Towards Data Science: <https://towardsdatascience.com>
- Gámez Vázquez, A., De la O Olán, M., Avila Perches, M., Virgen Vargas, J., Ruiz Torres, N., Gámez Vázquez, F., & Ascencio Álvarez, A. (2015). Calidad fisiológica de semilla y desarrollo de plántulas de maíz a temperaturas bajas. *Revista mexicana de ciencias agrícolas*, 1769-1779.
- García, L. (2021). *Propuesta de un sistema de riego por inundación optimizado basado en el monitoreo de las características del suelo*. Piura: PIRHUA.
- Martínez Varona, R. (2013). Controlled deficit irrigation effect in the productivity of the banana tree. *Ciencias Técnicas Agropecuarias*, 51-55.
- McCulloch, C. E., Searle, S. R., & Neuhaus, J. M. (2008). *Generalized, Linear, and Mixed Models*. Wiley-Interscience.
- MINAGRI. (30 de Marzo de 2015). Minagri: Exportación de banano orgánico peruano creció 94% en últimos 5 años. Guayaquil, Ecuador.
- MINAGRI. (4 de Noviembre de 2019). Exportaciones de banano orgánico superan los US\$ 117 millones hasta setiembre de este año. Perú.
- Olivares, B., Araya-Alman, M., Acevedo-Opazo, C., Rey, J. C., Cañete-Salinas, P., Giannini, F., . . . Landa, B. G. (2020). Relationship Between Soil Properties and Banana Productivity in the Two Main Cultivation Areas in Venezuela. *Journal of Soil Science and Plant Nutrition*, 2512-2524.
- Rey, J. C., Martínez, G., Rodríguez, G., Lobo, D., Sapuky, M., Núñez, E., . . . Rosales, F. (2010). Soil properties that determine the banana vigor (Musa AAA) of local production in Venezuela. *XIX Reunión Internacional ACORBAT 2010* (págs. 519-524). Medellín: ACORBAT.
- Reyes, J., Martínez, D., Rueda, R., & Rodríguez, T. (2014). Efecto del estrés hídrico en plantas de frijol. *Revista Iberoamericana de Ciencias*, 191-203.
- Romanekas, K., Buragiene, S., Kazlauskas, M., Steponavicius, D., Naujokiene, V., Bruciene, I., & Srauskis, E. (2023). Effects of Soil Electrical Conductivity and Physical Properties on Seeding Depth Maintenance and Winter Wheat Germination, Development and Productivity. *Agronomy*, 1-13.
- Soto, M. (1992). *Bananos, cultivo y comercialización*. San José: Litografía LIL.
- Villalaz, C. (2004). *Mecánica de suelos y cimentaciones*. México DF: Limusa.

- Watabe, T., Nakano, Y., & Dong-Hyuk, A. (2022). The Effect of Electrical Conductivity on Fruit Growth Pattern in Hydroponically Grown Tomatoes. *Crops*, 323-335.
- Zribi, W., Faci, J., & Aragüés, R. (2011). Efectos del acolchado sobre la humedad, temperatura, estructura y salinidad de suelos agrícolas. *Información Técnica Económica Agraria*, 148-162.





Apéndices





Apéndice A

Gráficas de medias para cada variable biométrica

A continuación, se muestran las curvas de medias para la altura de las plantas, la circunferencia de su pseudotallo y el número de hojas brotadas en cada una de ellas, calculadas según los modos de agrupación descritos en el Capítulo 2.

Se debe tomar en cuenta que en la Figura 2 y en la Figura 6 la curva de las medias de la altura de aquellas plantas de banano que crecieron en suelos poco retentivos se ubica por encima de las cultivadas en los otros dos tipos de suelo.

Así mismo, en la Figura 3 y en la Figura 7 se observa que aquellas plantas nietas a las que se les aplicó un riego mejorado ubican su curva de medias por encima de aquellas plantas nietas a las que se les aplicó un riego tradicional.

Figura 1. Curvas de medias para la altura de plantas hijas y nietas

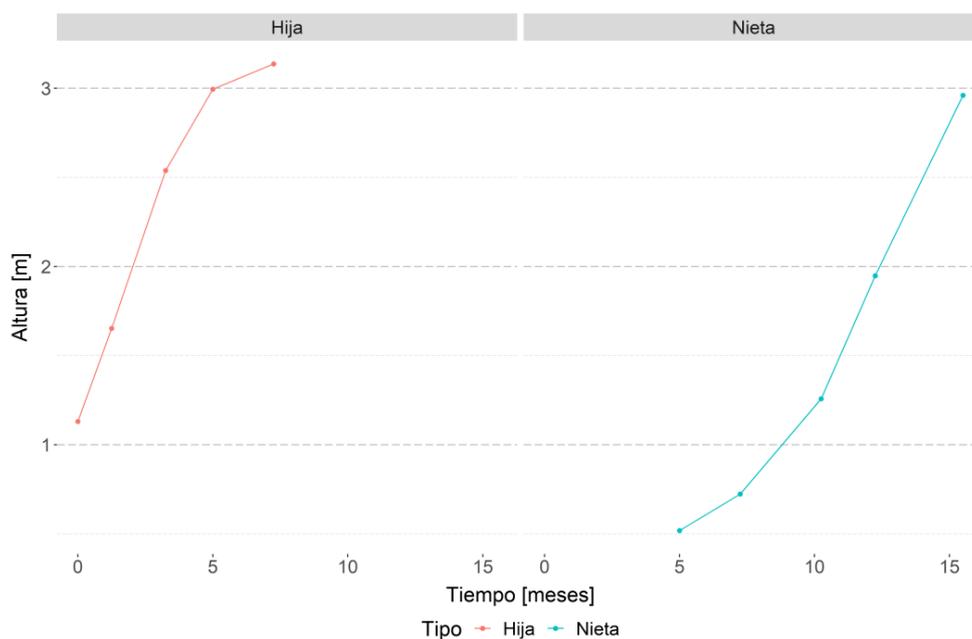


Figura 2. Curvas de medias para la altura de plantas hijas y nietas según el tipo de suelo

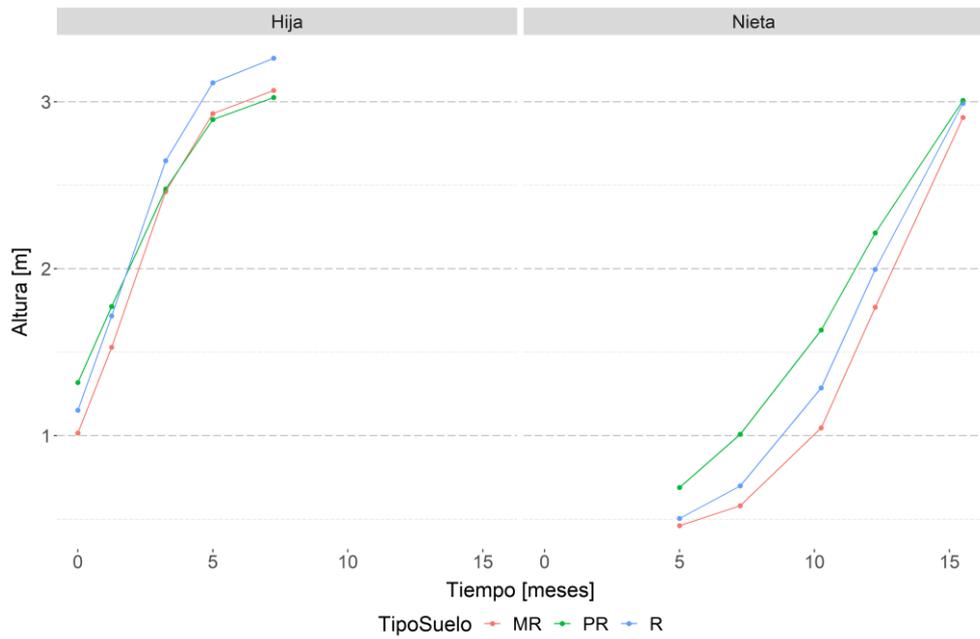


Figura 3. Curvas de medias para la altura de plantas hijas y nietas según el tipo de riego

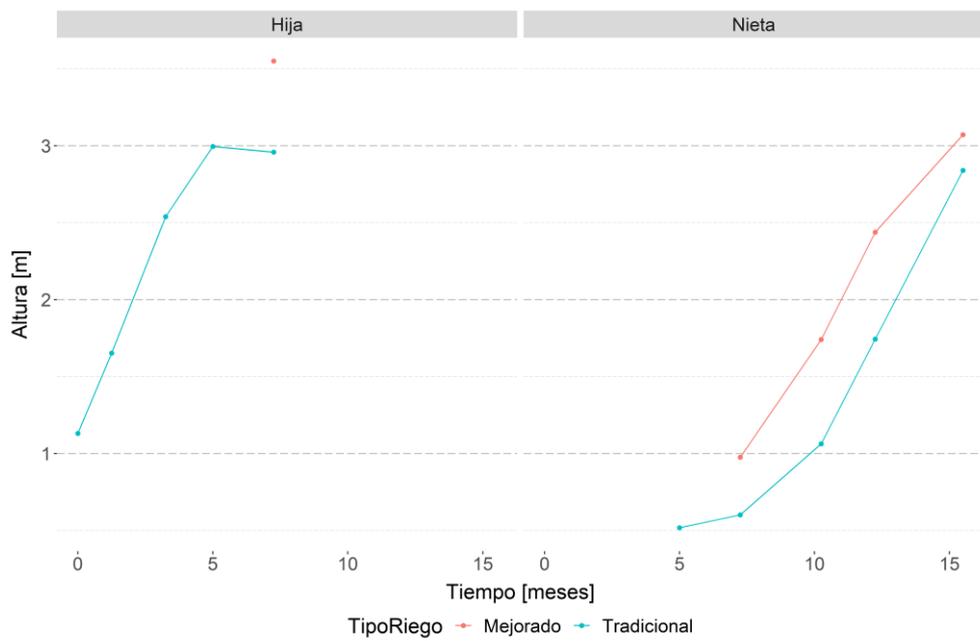


Figura 4. Curvas de medias para la altura de plantas hijas y nietas según la parcela

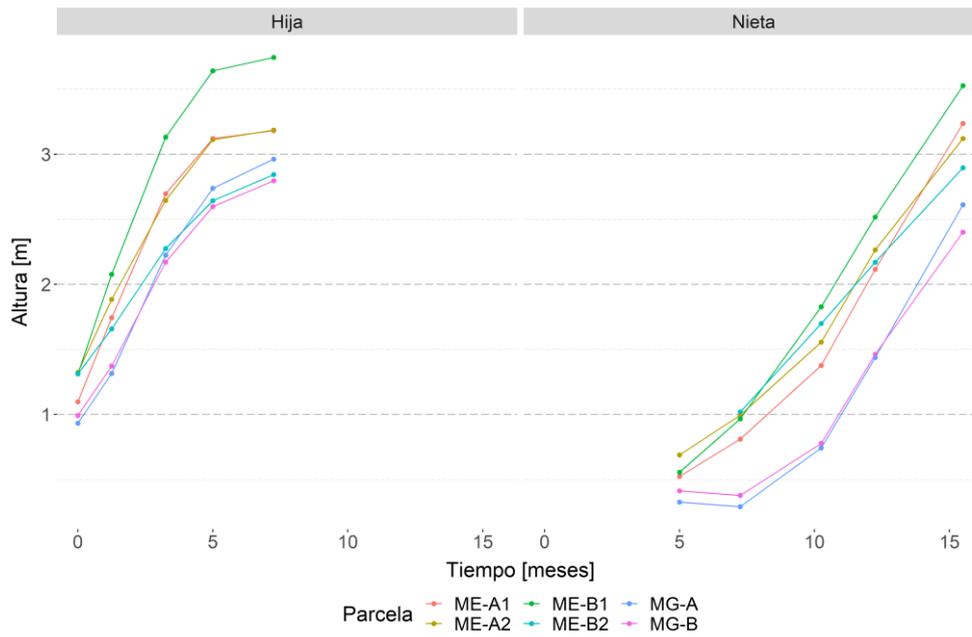


Figura 5. Curvas de medias para la circunferencia de plantas hijas y nietas

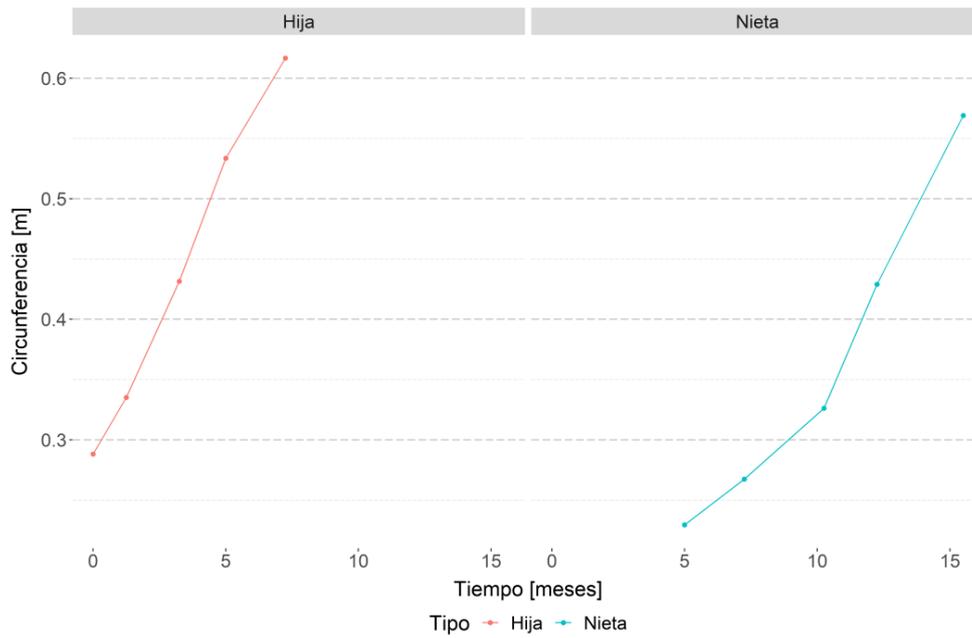


Figura 6. Curvas de medias para la circunferencia de plantas hijas y nietas según el tipo de suelo

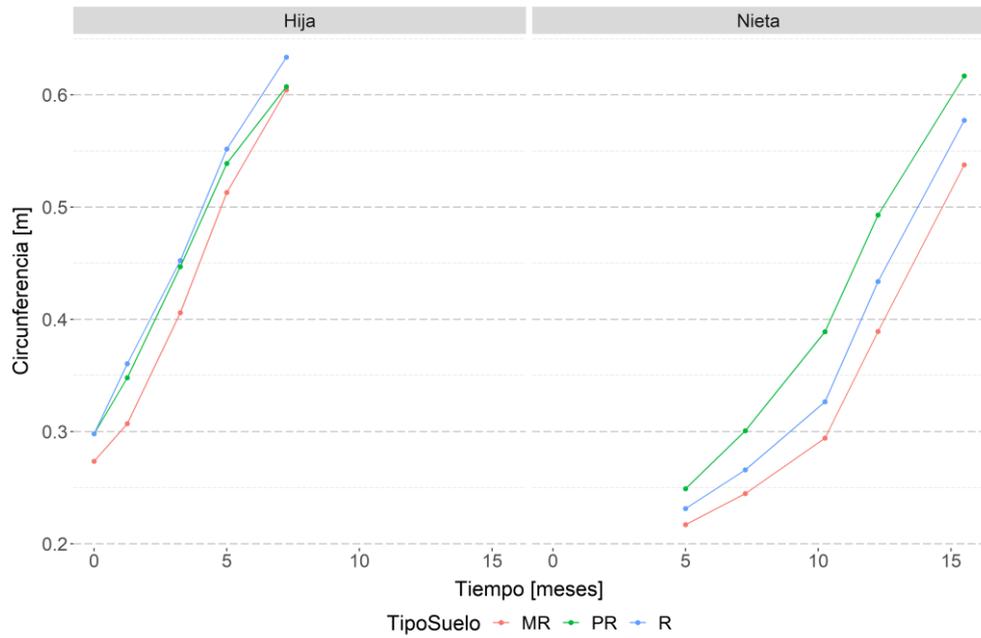


Figura 7. Curvas de medias para la circunferencia de plantas hijas y nietas según el tipo de riego

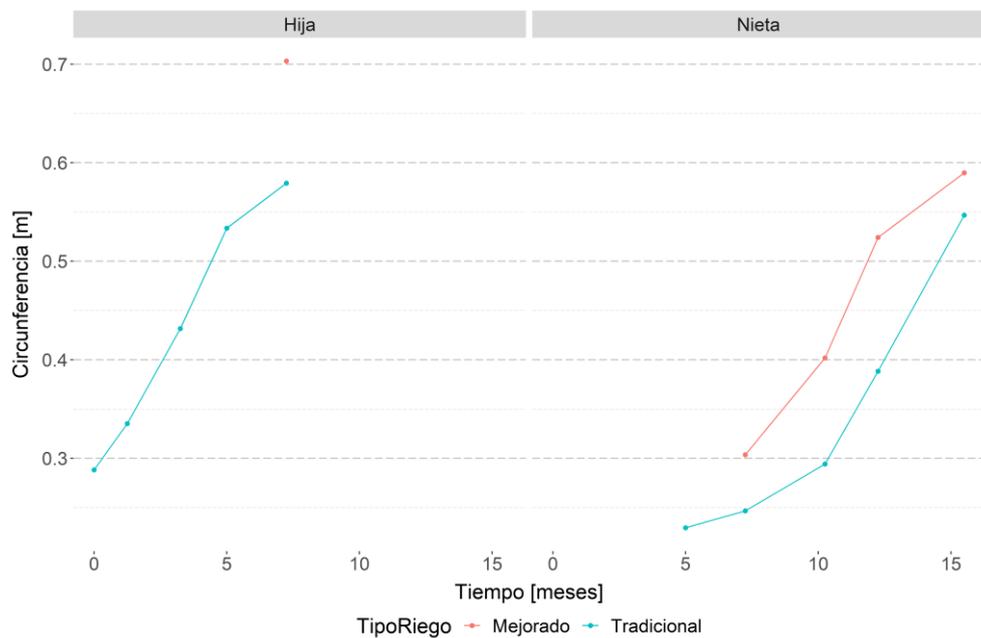


Figura 8. Curvas de medias para la circunferencia de plantas hijas y nietas según la parcela

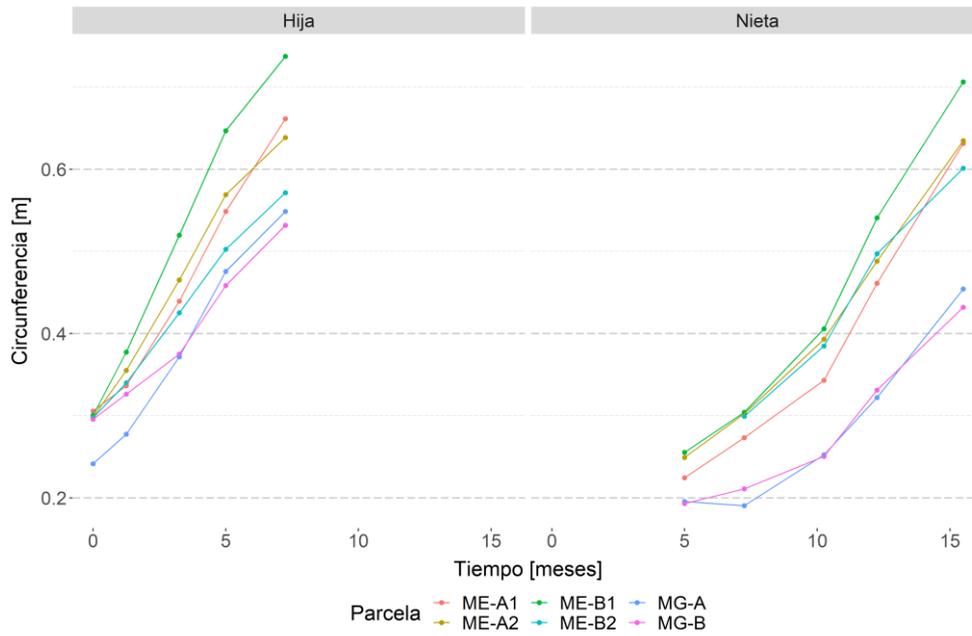


Figura 9. Curvas de medias para el número acumulado de hojas brotadas de plantas hijas y nietas

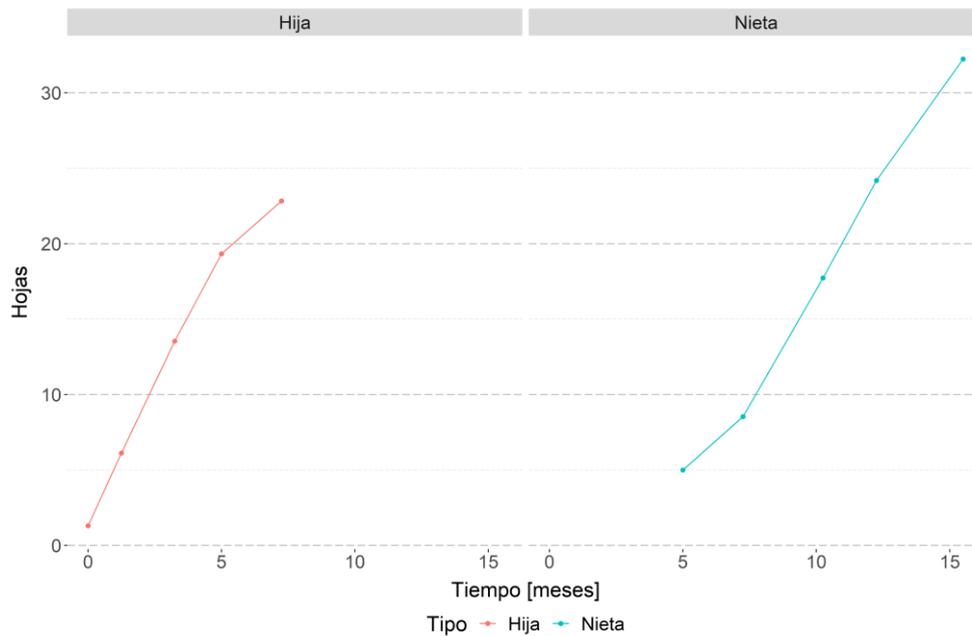


Figura 10. Curvas de medias para el número acumulado de hojas brotadas de plantas hijas y nietas según el tipo de suelo

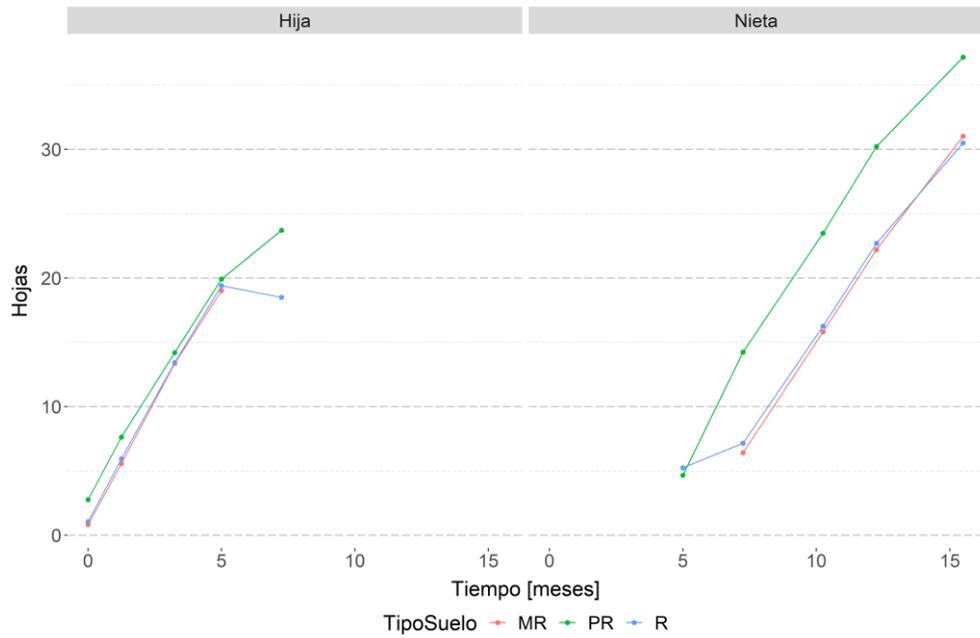


Figura 11. Curvas de medias para el número acumulado de hojas brotadas de plantas hijas y nietas según el tipo de riego

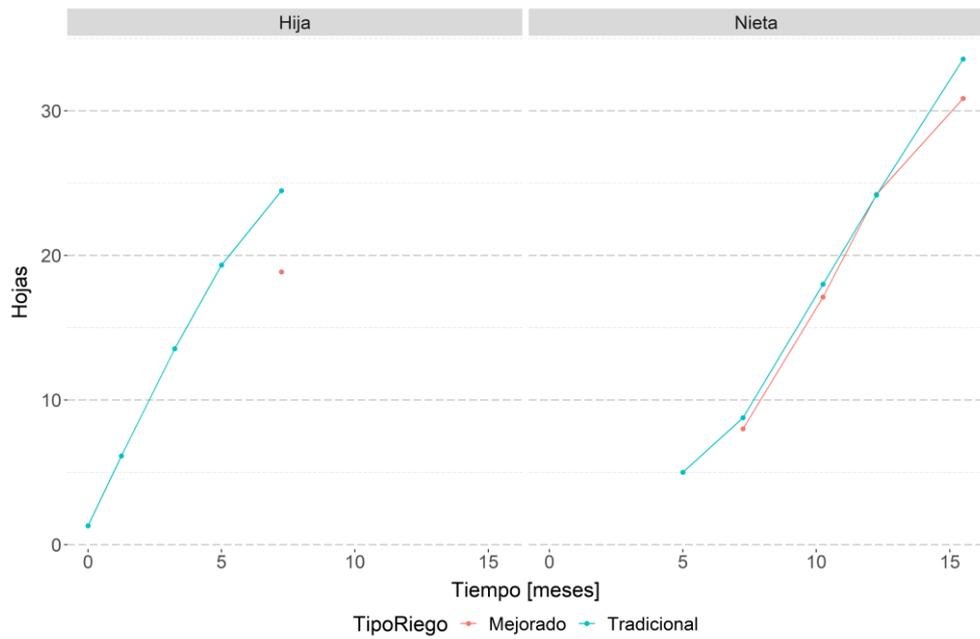
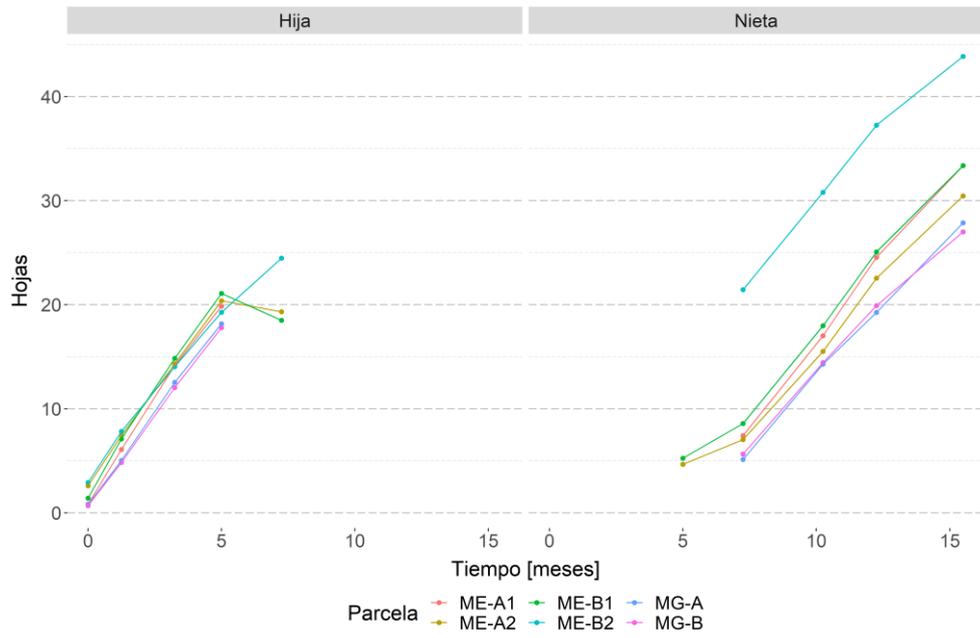


Figura 12. Curvas de medias para el número acumulado de hojas brotadas de plantas hijas y nietas según la parcela





Apéndice B

Análisis de la depuración de los modelos de regresión

Una vez seleccionados los modelos de regresión para la altura de las plantas de banano y para la circunferencia de sus pseudotallos, se procedió con la depuración sucesiva de aquellas variables no significativas hasta llegar a modelos con variables puramente significativas.

En este apartado se evalúan las variaciones entre los modelos “en bruto” y los modelos depurados, las variables que fueron retiradas de los modelos iniciales y cómo afectó dicho retiro al resto de variables y al modelo en sí.

Altura

El modelo seleccionado inicialmente para la altura tenía la siguiente estructura:

$$\hat{y}_{ijkt} = \beta_0 + \alpha_i + \beta_{1j}TipoRiego_j + \beta_{2k}TipoSuelo_k + \beta_3Tiempo_t + \beta_{5k}(TipoSuelo_k \times Tiempo_t),$$

Dicho modelo posee un R² marginal igual a 0.718 y un R² condicional igual a 0.921. A continuación, la Tabla 1 resume el valor de los coeficientes para cada una de las variables del modelo seleccionado.

Tabla 1. Resumen del modelo de regresión lineal mixto inicial para predecir la altura de las plantas nietas

Variable	Valor	Error estándar	p-Valor
Intercepto [β_0]	-1.262	0.087	0.000
Tiempo [β_3]	0.261	0.005	0.000
Tipo de suelo [β_{21}]: PR	0.759	0.123	0.000
Tipo de suelo [β_{22}]: R	0.096	0.100	0.338
Tipo de riego [β_{11}]: Tradicional	-0.134	0.036	0.000
Tiempo * Tipo de suelo [β_{51}]: PR	-0.037	0.008	0.000
Tiempo * Tipo de suelo [β_{52}]: R	0.004	0.006	0.496

Así pues, la variable Tipo de suelo [β_{22}]: R y la interacción Tiempo * Tipo de suelo [β_{52}]: R fueron depuradas del modelo final. Las variables que se mantuvieron experimentaron las variaciones que se resumen en la Tabla 2:

Tabla 2. Análisis de variación de los coeficientes y errores estándares de las variables para la depuración del modelo de regresión lineal mixto para la altura

Variable	Coeficiente			Error estándar		
	Valor 1	Valor 2	Variación	Valor 1	Valor 2	Variación
Intercepto [β_0]	-1.262	-1.203	-5%	0.087	0.066	-24%
Tiempo [β_3]	0.261	0.263	1%	0.005	0.004	-20%
Tipo de suelo [β_{21}]: PR	0.759	0.709	-7%	0.123	0.111	-10%
Tipo de riego [β_{11}]: Tradicional	-0.134	-0.143	7%	0.036	0.036	0%
Tiempo * Tipo de suelo [β_{51}]: PR	-0.037	-0.039	5%	0.008	0.007	-13%

Como se puede observar en la tabla anterior, los coeficientes de las variables que se mantuvieron en el modelo de regresión experimentaron una variación menor a los 10 puntos porcentuales, lo que supone una variación muy pequeña e implica la robustez del modelo desarrollado desde un principio.

Así mismo, el error estándar de cada variable, a excepción de Tipo de riego [β_{11}]: Tradicional, disminuyó en más de diez puntos porcentuales, lo que implica una ganancia en la precisión de los coeficientes estimados para cada una de ellas.

Finalmente, también vale aclarar que los R^2 del modelo, tanto marginal como condicional, no sufren una disminución significativa, lo que supone que la capacidad para explicar la varianza de la altura de la planta de banano del modelo inicial no se vio comprometida luego de depurar aquellas variables no significativas, tal como se muestra en la Tabla 3:

Tabla 3. Análisis de variación del R^2 marginal y condicional para la depuración del modelo de regresión lineal mixto para la altura

Parámetro	Valor 1	Valor 2	Variación
R^2 marginal	0.718	0.715	-0.4%
R^2 condicional	0.921	0.92	-0.1%

Circunferencia

El modelo propuesto inicialmente para la circunferencia tenía la siguiente estructura:

$$\hat{y}_{ijkt} = \beta_0 + \alpha_i + \beta_{1j}TipoRiego_j + \beta_{2k}TipoSuelo_k + \beta_3Tiempo_t + \beta_{5k}(TipoSuelo_k \times Tiempo_t)$$

Dicho modelo tiene, además, un R^2 marginal de 0.609 y un R^2 condicional de 0.899. Se presentan los coeficientes de las variables de dicho modelo en la Tabla 4:

Tabla 4. Resumen del modelo de regresión lineal mixto inicial para predecir la circunferencia de las plantas nietas

Variable	Valor	Error estándar	p-Valor
Intercepto [β_0]	-0.061	0.017	0.000
Tiempo [β_3]	0.038	0.001	0.000
Tipo de suelo [β_{21}]: PR	0.101	0.024	0.000
Tipo de suelo [β_{22}]: R	0.014	0.020	0.469
Tipo de riego [β_{11}]: Tradicional	-0.005	0.007	0.433
Tiempo * Tipo de suelo [β_{51}]: PR	-0.001	0.002	0.403
Tiempo * Tipo de suelo [β_{52}]: R	0.002	0.001	0.135

Así pues, las variables Tipo de riego [β_{11}]: Tradicional y la interacción Tiempo * Tipo de suelo [β_{51}]: PR fueron depuradas del modelo final. Las variables que se mantuvieron experimentaron las variaciones que se resumen en la Tabla 5:

Tabla 5. Análisis de variación de los coeficientes y errores estándares de las variables para la depuración del modelo de regresión lineal mixto para la circunferencia

Variable	Coeficiente			Error estándar		
	Valor 1	Valor 2	Variación	Valor 1	Valor 2	Variación
Intercepto [β_0]	-0.061	-0.064	5%	0.017	0.013	-24%
Tiempo [β_3]	0.038	0.038	0%	0.001	0.001	0%
Tipo de suelo [β_{21}]: PR	0.101	0.089	-12%	0.024	0.017	-29%
Tipo de suelo [β_{22}]: R	0.014	0.012	-14%	0.02	0.019	-5%
Tiempo * Tipo de suelo [β_{52}]: R	0.002	0.002	0%	0.001	0.001	0%

De acuerdo con la tabla anterior, los coeficientes de las variables que se mantuvieron en el modelo de regresión experimentaron una variación menor a los 15 puntos porcentuales, lo que si bien supone una variación superior a la experimentada por los coeficientes de las variables del modelo anterior (el de la altura), sigue suponiendo una variación no muy significativa e implica que este modelo también es robusto.

Por otra parte, el error estándar de dos de las cinco variables que se mantuvieron en el modelo disminuyó en más de veinte puntos porcentuales, lo que implica una ganancia en la precisión de los coeficientes estimados para cada una de ellas.

Finalmente, también vale aclarar que los R^2 del modelo, tanto marginal como condicional, no sufren una disminución significativa tal como sucedía con el modelo anterior, lo que supone que la capacidad para explicar la varianza de la circunferencia del pseudotallo de la planta de banano del modelo inicial no se vio comprometida luego de depurar aquellas variables no significativas, tal como se muestra en la Tabla 6:

Tabla 6. Análisis de variación del R^2 marginal y condicional para la depuración del modelo de regresión lineal mixto para la circunferencia

Parámetro	Valor 1	Valor 2	Variación
R^2 marginal	0.609	0.607	-0.3%
R^2 condicional	0.899	0.899	0.0%

