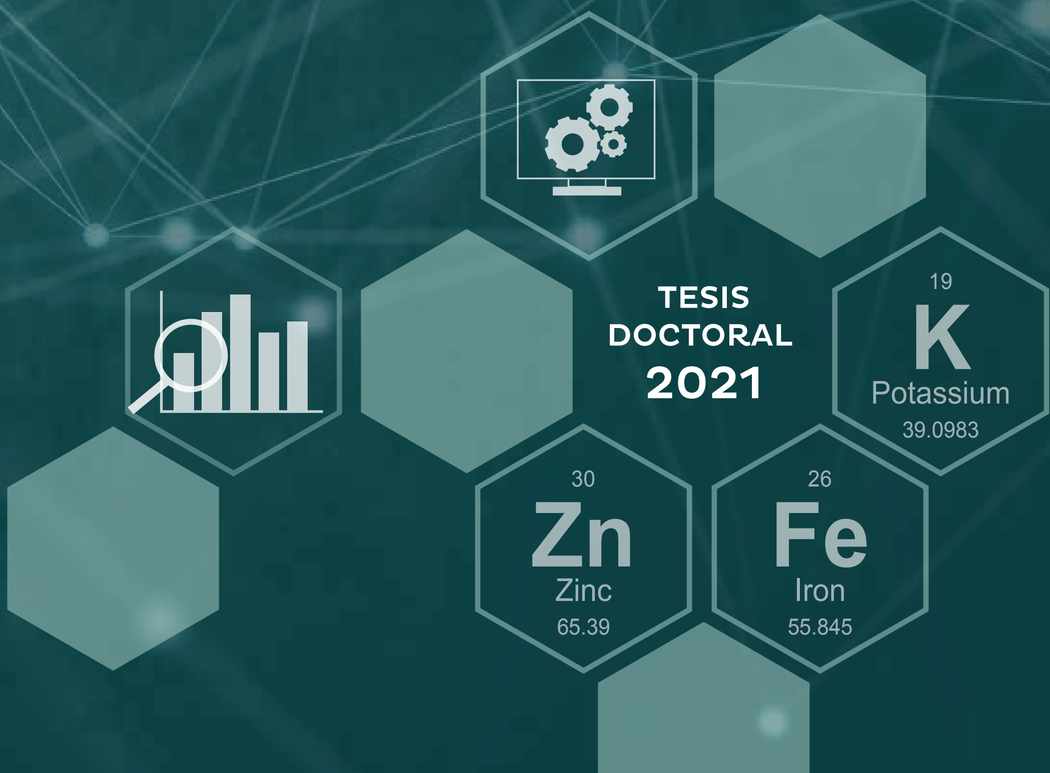
UNIVERSIDAD DE CÓRDOBA

DETERMINATION OF THE POTENTIAL INFLUENCE OF SOIL IN THE DIFFERENTIATION OF PRODUCTIVITY AND IN THE CLASSIFICATION OF SUSCEPTIBLE AREAS TO BANANA WILT IN VENEZUELA

DETERMINACIÓN DE LA INFLUENCIA POTENCIAL DEL SUELO EN LA DIFERENCIACIÓN DE LA PRODUCTIVIDAD Y EN LA CLASIFICACIÓN DE ÁREAS SUSCEPTIBLES A LA MARCHITEZ DEL BANANO EN VENEZUELA

BARLIN ORLANDO OLIVARES CAMPOS

TESIS DOCTORAL
2021



TITULO: *Determination of the potential influence of soil in the differentiation of productivity and in the classification of susceptible areas to Banana Wilt in Venezuela*

AUTOR: *Barlin Orlando Olivares Campos*

© Edita: UCOPress. 2022
Campus de Rabanales
Ctra. Nacional IV, Km. 396 A
14071 Córdoba

https://www.uco.es/ucopress/index.php/es/
ucopress@uco.es

University of Cordoba

Doctoral Program in Agricultural, Food, Forestry and Sustainable Rural Development Engineering

PhD Thesis

Determination of the potential influence of soil in the differentiation of productivity and in the classification of susceptible areas to Banana Wilt in Venezuela

Barlin Orlando Olivares Campos

Supervisors: José Alfonso Gómez Calero & Blanca B. Landa del Castillo

Córdoba, November 2021

Universidad de Córdoba

Programa de Doctorado en Ingeniería Agraria, Alimentaria, Forestal y del Desarrollo Rural Sostenible

Tesis Doctoral

Determinación de la influencia potencial del suelo en la diferenciación de la productividad y en la clasificación de áreas susceptibles a la marchitez del banano en Venezuela

Barlin Orlando Olivares Campos

Directores: José Alfonso Gómez Calero & Blanca B. Landa del Castillo

Córdoba, noviembre 2021

University of Cordoba

Doctoral Program in Agricultural, Food, Forestry and Sustainable Rural Development Engineering

PhD Thesis

Determination of the potential influence of soil in the differentiation of productivity and in the classification of susceptible areas to Banana Wilt in Venezuela

Memoria redactada para optar al grado de Doctor con Mención Internacional por la Universidad de Córdoba, por el Ingeniero Agrónomo:

Barlin Orlando Olivares Campos

Fdo.: José Alfonso Gómez Calero

Fdo.: Blanca B. Landa del Castillo

Córdoba, noviembre 2021

TÍTULO DE LA TESIS: Determination of the potential influence of soil in the differentiation of productivity and in the classification of susceptible areas to Banana Wilt in Venezuela

DOCTORANDO: Barlin Orlando Olivares Campos

INFORME RAZONADO DEL/DE LOS DIRECTOR/ES DE LA TESIS

(se hará mención a la evolución y desarrollo de la tesis, así como a trabajos y publicaciones derivados de la misma).

Don José Alfonso Gómez Calero y Doña Blanca Beatriz Landa del Castillo, ambos investigadores científicos del Instituto de Agricultura Sostenible CSIC, y como directores del alumno del programa de doctorado "Ingeniería agraria, alimentaria, forestal y de desarrollo rural sostenible" D. Barlin Orlando Olivares Campos informan que:

Durante su periodo como doctorando del programa, desde noviembre de 2016, hasta la fecha de emitir este informe ha cumplido todas las actividades formativas obligatorias del programa, así como otras actividades voluntarias, con diligencia y dedicación. Entre estas actividades destacan cinco estancias de investigación de carácter internacional en:

1. Universidad Nacional de Córdoba, Facultad de Ciencias Agropecuarias Ciudad: Córdoba, Argentina, del 11/02/2019 - 06/06/2019 con una duración de cuatro meses.
2. Universidad de Panamá. Facultad de Ciencias Agropecuarias. Ciudad: Bocas del Toro, Panamá del 23/06/2019 - 10/08/2019 con una duración dos meses.
3. Universidad de Talca. Facultad de Ciencias Agrarias. Ciudad: Talca, Chile, del 28/10/2019 - 26/01/2020 con una duración de tres meses.
4. Universidad Nacional de Costa Rica. Facultad de Ciencias de la Tierra y el Mar. Ciudad: Heredia, Costa Rica, del 17/02/2020 - 05/07/2020 con una duración de cinco meses.
5. Universidad La Gran Colombia. Facultad de Ingeniería Geográfica y Ambiental. Ciudad: Armenia, Colombia del 06/01/2021 - 07/03/2021 con una duración de dos meses.

La finalidad de estas cuatro estancias de movilidad internacional, con un total de 16 meses con tres de ellas de duración igual o superior a 3 meses, estuvo orientada a que el doctorando se formara en un ambiente internacional en áreas científicas punteras relacionadas con datos del sector agroalimentario, complementado su formación predoctoral. En su conjunto, esta formación se ha articulado alrededor del estudio de la determinación de la influencia del suelo en la productividad del banano y en la incidencia de la marchitez del banano en las condiciones de Venezuela, situándose entre los campos de la patología vegetal, la agronomía y la ciencia del suelo. Su trabajo se ha centrado en un estudio los factores que condicionan el desarrollo de la Fusariosis vascular del banano en Venezuela (Capítulo 1); la identificación de las propiedades del suelo asociadas con la incidencia de la Marchitez del banano mediante métodos supervisados (Capítulo 2); la relación entre las propiedades cuantitativas del suelo y la productividad del banano en las dos principales áreas de cultivo de Venezuela (Capítulo 3) y finalmente la correlación de los niveles de productividad del banano y las propiedades morfológicas del suelo utilizando “Regularized Optimal Scaling Regression”.

Durante todo el desarrollo de su formación predoctoral el candidato ha mostrado un elevado grado de dedicación, capacidad y habilidad para el trabajo en equipo. Todo ello, además de posibilitar la realización de los trabajos de investigación incluidos en su Tesis doctoral, le ha permitido alcanzar el grado de madurez y especialización para optar al grado de doctor.

Los resultados que se han ido generando a lo largo del desarrollo de su Tesis doctoral han sido presentados como comunicaciones orales en los siguientes congresos nacionales e internacionales:

1. Olivares, B. 2021. Soluciones novedosas para enfermedades en banano: un enfoque con Machine Learning para el desarrollo de una agricultura sostenible. In III Congreso Internacional de Ciencias Agropecuarias y Recursos Zoogenéticos. (02, 2021, Los Ríos, Ecuador). Quevedo, N. & Chávez, D. (comp.). Los Ríos, Ecuador. Instituto de Investigaciones Binario. p. 39.
2. Olivares, B. 2021. Machine Learning and the new sustainable agriculture: applications in banana production systems. In: III Congreso Internacional de Ciencias Agropecuarias, Tecnología e Innovación Industrial (08, 2021, Ciudad de Valencia, Ecuador). López, J. & Muñoz, P. (comp). Ciudad de Valencia, Ecuador. Instituto de Investigaciones Binario. p. 47.
3. Olivares, B. 2021. Novel solutions for diseases in bananas: an approach with Machine Learning for the development of a sustainable agricultura. In I International Congress of Organic Agriculture. (06, 2021, Leuven, Belgium). IASS (comp.) International Association of Students in Agriculture and Related Sciences (IAAS World). Leuven, Belgium. p. 12
4. Olivares, B. 2021. Diagnóstico e Idoneidad Ambiental de Enfermedades Tropicales del Banano en Venezuela mediante Machine Learning. In: IX Congreso Científico de Investigadores en Formación. (05, 2021, Córdoba, España). Córdoba, España: UCOPress. Editorial Universidad de Córdoba.
5. Olivares, B. 2020. Análisis de los niveles de productividad del banano basado en las propiedades morfológicas del suelo: un estudio por escalamiento óptimo. In: I Seminario Internacional de Agricultura, Agroindustria y Ambiente. (08, 2020, Riobamba, Ecuador). IASS (comp.) International Association of Students in Agriculture and Related Sciences (IAAS World). Riobamba, Ecuador.
6. Olivares, B; Araya-Alman, M; Rey, J.C. 2020. A study of the relationship between soil properties and banana productivity in Venezuela. In II Congreso Internacional Multidisciplinar de Investigadores en Formación. (12, 2020, Córdoba, España). Sánchez, F. & Serrano, R. (coord.). Memoria. Córdoba, España. Universidad de Córdoba. p. 194.
7. Olivares, B; Rueda, MA; Vega, A. 2020. Banana productivity levels analysis based on soil morphological properties: A random forest approach. In II Congreso Internacional Multidisciplinar de Investigadores en Formación. (12, 2020, Córdoba, España). Sánchez, F. & Serrano, R. (coord.). Memoria. Córdoba, España. Universidad de Córdoba. p. 194.
8. Olivares, B; Paredes-Trejo, F. 2020. Estrés hídrico y sequía en Venezuela: desde la percepción en campo hasta la satelital. In I Simposio Nacional de Recursos Hídricos. (11, 2020, Caracas, Venezuela). Silva, O. (coord.). Memoria. Caracas, Venezuela. Academia general de la ingeniería y el habitad. p. 66.

9. Olivares, B; Rey, J.C; Lobo, D; Gómez, J.A y Landa, B. 2019. Impacto del cambio climático en zonas bananeras de la Región Central de Venezuela: El futuro de los bananos en un escenario hídrico incierto. En: Chica Pérez, A. F. y Mérida García, J. (Eds). Creando Redes Doctorales: Investiga y Comunica. (pp. 367-370). Córdoba, España: UCOPress. Editorial Universidad de Córdoba.
10. Olivares, B; Rey, J.C; Lobo, D; Gómez, J.A & Landa, B. 2019. El cambio climático en zonas bananeras de la Región Central de Venezuela: El futuro de los bananos con un escenario hídrico incierto. In III Simposio venezolano de Cambio Climático, Agricultura y Seguridad Alimentaria. (10, 2019, Caracas, Venezuela). Goldwaser, M. & Morales, G. (eds.). Memoria. Caracas, Venezuela. Academias de Ciencias Físicas, Matemáticas y Naturales y de la Ingeniería y el Hábitat. p. 251.
11. Olivares, B; Cortez, A; Lobo, D; Parra, R.M; Rey, J.C & Rodríguez, M. F. 2019. Evaluación de la vulnerabilidad agrícola a la sequía meteorológica en Venezuela. In III Simposio venezolano de Cambio Climático, Agricultura y Seguridad Alimentaria. (10, 2019, Caracas, Venezuela). Goldwaser, M. & Morales, G. (eds.). Memoria. Caracas, Venezuela. Academias de Ciencias Físicas, Matemáticas y Naturales y de la Ingeniería y el Hábitat. p. 251.

y han dado lugar a las siguientes publicaciones en revistas científicas indexadas:

1. Olivares, B.O., Calero, J., Rey, J.C., Lobo, D., Landa, B.B., Gómez, J. A. 2022. Correlation of banana productivity levels and soil morphological properties using regularized optimal scaling regression. Catena, 208: 105718. https://doi.org/10.1016/j.catena.2021.105718 Q1, En 2020, Q1, 7 de 37 revistas en *Soil Science* con un factor de impacto de 5.198.
2. Olivares B, Rey JC, Lobo D, Navas-Cortés JA, Gómez JA, Landa BB. 2021. Fusarium Wilt of Bananas: A Review of Agro-Environmental Factors in the Venezuelan Production System Affecting Its Development. Agronomy, 11(5):986. https://doi.org/10.3390/agronomy11050986 En 2020, Q1, 16 de 91 revistas en *Agronomy* con un factor de impacto de 3.417.
3. Olivares, B., Araya-Alman, M., Acevedo-Opazo, C. et al. 2020. Relationship Between Soil Properties and Banana Productivity in the Two Main Cultivation Areas in Venezuela. J Soil Sci Plant Nutr. 20 (3): 2512-2524. https://doi.org/10.1007/s42729-020-00317-8 En 2020, Q1, 49 de 235 revistas en *Plant Sciences* con un factor de impacto de 3.872.

Por todo ello, se autoriza la presentación de la Tesis doctoral con mención internacional.

Córdoba, 04 de noviembre de 2021

Firma de los directores

GOMEZ CALERO JOSE ALFONSO - 30533940Z

Firmado digitalmente por GOMEZ CALERO JOSE ALFONSO - 30533940Z
Fecha: 2021.11.04 14:06:51 +01'00'

Fdo.: José Alfonso Gómez Calero

LANDA DEL CASTILLO BLANCA BEATRIZ - DNI 30542791X

Firmado digitalmente por LANDA DEL CASTILLO BLANCA BEATRIZ - DNI 30542791X
Nombre de reconocimiento (DN): c=ES, o=CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS, ou=CERTIFICADO ELECTRONICO DE EMPLEADO PUBLICO, ou=INSTITUTO DE AGRICULTURA SOSTENIBLE, ou=30542791, serialNumber=IDCES-30542791X, sn=LANDA DEL CASTILLO, givenName=BLANCA BEATRIZ, cn=LANDA DEL CASTILLO BLANCA BEATRIZ - DNI 30542791X
Fecha: 2021.11.04 13:16:40 +01'00'

Fdo.: Blanca B. Landa del Castillo

TÍTULO DE LA TESIS: Determination of the potential influence of soil in the differentiation of productivity and in the classification of susceptible areas to Banana Wilt in Venezuela

DOCTORANDO: Barlin Orlando Olivares Campos

INFORME RAZONADO DEL TUTOR

(Ratificando el informe favorable del director. Sólo cuando el director no pertenezca a la Universidad de Córdoba).

Doña María Auxiliadora Soriano Jiménez, como tutor del alumno del Programa de doctorado "Ingeniería agraria, alimentaria, forestal y de desarrollo rural sostenible" D. Barlin Orlando Olivares Campos informa que:

Durante el periodo de doctorado, desde noviembre de 2016 hasta la fecha de emitir este informe, el doctorando ha realizado todas las actividades formativas obligatorias del programa de doctorado, así como otras actividades voluntarias, con gran diligencia y dedicación. Así mismo, el doctorando ha mostrado una gran capacidad y habilidad para el trabajo en equipo. Además de la realización de los trabajos de investigación incluidos en su tesis doctoral, todo ello le ha permitido alcanzar el nivel de especialización requerido para optar al grado de doctor.

Por todo ello, se autoriza la presentación de la tesis doctoral.

SORIANO JIMENEZ MARIA AUXILIADORA - 38492225T

Firmado digitalmente por SORIANO JIMENEZ MARIA AUXILIADORA - 38492225T
Nombre de reconocimiento (DN): c=ES, serialNumber=IDCES-38492225T, givenName=MARIA AUXILIADORA, sn=SORIANO JIMENEZ, cn=SORIANO JIMENEZ MARIA AUXILIADORA - 38492225T
Fecha: 2021.11.05 07:48:17 +01'00'

Fdo.: María Auxiliadora Soriano Jiménez

Córdoba, a 5 de noviembre de 2021

Firma del responsable de la línea de investigación

FERERES CASTIEL ELIAS - 05393383K

Firmado digitalmente por FERERES CASTIEL ELIAS - 05393383K
Fecha: 2021.11.05 09:07:27 +01'00'

Fdo.: Elías Fereres Castiel

Mención de Doctorado Internacional en el título de doctor

Mediante la defensa de esta Memoria se pretende optar a la mención de **Doctorado Internacional,** debido a que el doctorando reúne los requisitos exigidos para tal mención, a saber:

1. Informes favorables de dos doctores pertenecientes a Instituciones de Enseñanza Superior de otros países:
   - Prof. Gustavo Curaqueo Fuentes. Facultad de Recursos Naturales, Departamento de Ciencias Agropecuarias y Acuícolas de la Universidad Católica de Temuco, Chile.
   - Prof. Zenaida Lozano Pérez. Facultad de Agronomía, Instituto de Edafología de la Universidad Central de Venezuela.

2. Uno de los miembros del tribunal que ha de evaluar la Tesis pertenece a un centro de Enseñanza Superior de otro país:
   - Dr. Giovanni Bubici. National Research Council, Institute for Sustainable Plant Protection, Italy.

3. La exposición y la defensa de parte de esta Tesis se realizarán en una lengua diferente a la materna: inglés.

4. Estancia de tres meses en un centro de investigación de otro país:
   - Facultad de Ciencias Agropecuarias de la Universidad Nacional de Córdoba (UNC) en Argentina bajo la supervisión de la Dra. Mónica Balzarini. Fechas: 11/02/2019 al 06/06/2019. Duración: 4 meses.
   - Centro de Investigación y Trasferencia en Riego y Agroclimatología (CITRA), Facultad de Ciencias Agrarias de la Universidad de Talca en Chile bajo la supervisión del Dr. César Antonio Acevedo Opazo. Fechas: 28/10/2019 al 26/01/2020. Duración: 3 meses.

Título de la tesis: Determination of the potential influence of soil in the differentiation of productivity and in the classification of susceptible areas to Banana Wilt in Venezuela

Doctorando: Barlin Orlando Olivares Campos

Indicios de calidad científica de la Tesis Doctoral

Olivares B, Rey JC, Lobo D, Navas-Cortés JA, Gómez JA, Landa BB. (2021). Fusarium Wilt of Bananas: A Review of Agro-Environmental Factors in the Venezuelan Production System Affecting Its Development. Agronomy, 11(5):986.
https://doi.org/10.3390/agronomy11050986. JCR Impact factor: 3.417; JIF Rank: 16/91; First Quartile (Q1) in Agronomy Category.

Olivares B, Araya-Alman M, Acevedo-Opazo C, Cañete-Salinas P. et al. (2020). Relationship Between Soil Properties and Banana Productivity in the Two Main Cultivation Areas in Venezuela. J Soil Sci Plant Nutr, 20(3): 2512-2524.
https://doi.org/10.1007/s42729-020-00317-8. JCR Impact factor: 3.872; JIF Rank: 49/235. First Quartile (Q1) in Plant Sciences Category.

Olivares B, Calero J, Rey JC, Lobo D, Landa BB, Gómez JA, (2022). Correlation of banana productivity levels and soil morphological properties using Regularized Optimal Scaling Regression. Catena, 208(1): 105718.
https://doi.org/10.1016/j.catena.2021.105718. JCR Impact factor: 5.198; JIF Rank: 7/37; First Decil (D1) in Soil Science Category.

Agradecimientos

Con la elaboración de esta Tesis Doctoral finaliza una de las etapas más intensas y bonitas de mi trayectoria, de la mano de la Universidad de Córdoba, es por ello que quiero agradecer en estas líneas a mis directores, supervisores de estancias, familia y amigos por todo el apoyo cercano, humano y académico recibido durante estos años, sin ustedes no hubiese sido posible alcanzar este logro tan importante para mí.

En primer lugar, agradecer a la Asociación Universitaria Iberoamericana de Postgrado (AUIP) por otorgarme la beca de doctorado y en especial al Dr. Arturo Chica de la Universidad de Córdoba por todo el apoyo recibido durante estos años.

En segundo lugar, a mis directores José Alfonso Gómez, Blanca B. Landa y Deyanira Lobo, de los cuales aprendí de su experiencia, su ejercicio profesional y la calidad humana, siempre acertados en la orientación, supervisión, cercanía y el apoyo que recibí. A la Dra. Mónica Balzarini, Dr. César Acevedo, Dir. Jacob Pitti, Dr. Rafael Granados y la Dra. Ximena Cifuentes, quienes me permitieron desarrollar las distintas estancias de investigación en Latinoamérica.

También quiero agradecer a mi linda familia, que desde Venezuela me apoyaron, me animaron a seguir y a vencer la nostalgia y la tristeza de estar lejos de mi tierra. A mis padres, que han sido mi inspiración y a mi hermana que siempre ha estado allí cuando la necesito. A mis tías, tíos, primas, primos y a mis abuelas Barbara y Lina por el amor incondicional.

Agradecer a quienes formaron parte de este largo pero bonito camino en España; a Patricia San Segundo, Mario y los niños Virginia y Roberto por acogerme e integrarme con todo ese cariño y simpatía; a Rosa Camacho y Luis Blanco por recibirme en su hogar múltiples veces, también a todos mis amigos y compañeros que de alguna manera estuvieron presentes en las estancias.

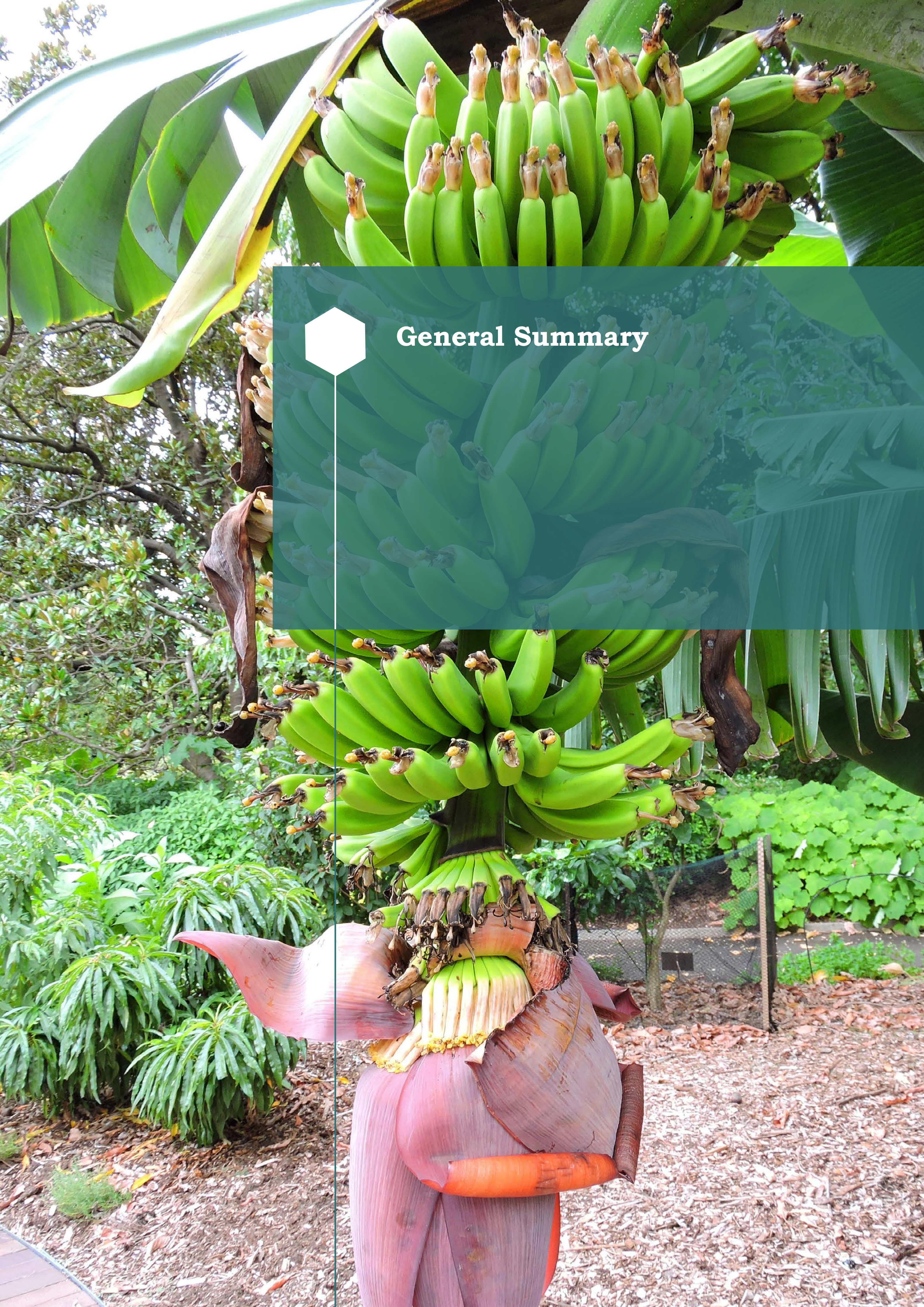
Finalmente, a Juan Francisco Gutiérrez (Juan Fran) y a su linda familia, por hacerme sentir en casa, por apoyarme emocionalmente, por no dejarme solo, y por recordarme que, a pesar de estar lejos de casa, de mi familia, de mi tierra; el amor que recibo lo compensa y me llena de felicidad.

A todos mil gracias

He aprendido que el valor no es la ausencia de miedo, sino el triunfo sobre él. El hombre valiente no es el que no siente miedo, sino el que lo domina –

Nelson Mandela

Contents



General Summary

General Summary

Banana, the edible fruit of Musaceae, is a staple food for more than 400 million people worldwide due to their nutritional and energy attributes. This makes Musaceae a crop of worldwide relevance, particularly in tropical regions, highlighting the impact of improved Musaceae cropping systems in the current efforts worldwide oriented towards a new agricultural revolution based on sustainable intensification.

To achieve this, better practices for food production based on scientific and technical research capable to consider the complexity and variability within the agri-food sector are necessary. The research presented in this PhD Thesis is oriented towards providing answers to the causes of two aspects considered of high relevance for banana production, both affecting productivity and sustainability, always addressed for the Venezuelan conditions, one of the world's largest producing countries:

1- The impact of phytosanitary risks related to *Fusarium* wilt and the influence of the soil on the incidence of Banana Wilt (BW) caused by a fungal-bacterial complex.

2- An observed trend towards loss of productivity and decline of soil quality in some commercial farms of Aragua and Trujillo states in Venezuela.

The first issue, related to banana plant health, has been covered in two consecutive studies. Firstly, in Chapter I a systematic review on the effect of agro-environmental factors on the impact of *Fusarium* Wilt of Bananas, caused by *Fusarium oxysporum* f. sp. *cubense* (*Foc*) tropical race 4 (TR4), and the implications for the Venezuelan production system of this disease is presented. This Chapter synthetically characterizes reliable information on the biotic and abiotic factors related to *Foc* TR4 occurrence, in conjunction with a risk analysis and climate suitability maps for *Foc* TR4 in Venezuela. This chapter can serve as a basic summary of the available knowledge for use by plant health technicians and professionals, as well as for other stakeholders concerning disease management.

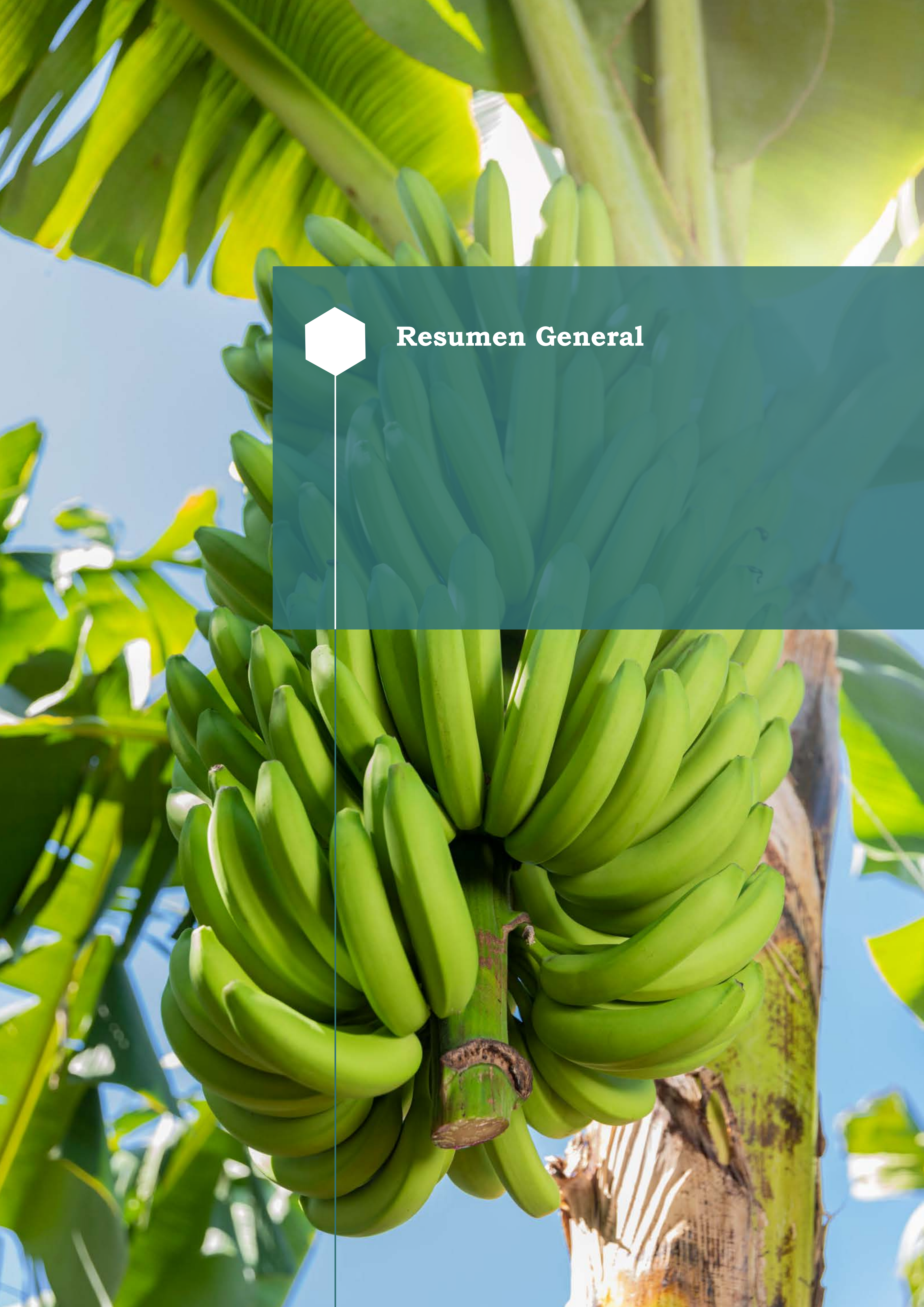
The research oriented towards the plant health issues in banana is completed with the study presented in Chapter II. This chapter analyzes the relationship between soil properties and the incidence of Banana Wilt (BW), a disease of unknow etiology, that is attributed to be caused by a fungal-bacterial complex, in a case study of a commercial banana farm in the state of Aragua in Venezuela, whose incidence has reduced the planted area by more than 35.0% in recent years. The application of the Random Forest algorithm allowed to classify with good precision the incidence of BW in lacustrine soils of Venezuela based on the physical and chemical soil properties, being an effective tool for decision-making in the field. In addition, the use of soil information in banana areas of Venezuela allowed the identification of banana lots with high and low incidence of BW using also the Random Forest algorithm. The model showed that the incidence level (low or high) of Banana Wilt could be distinguished through its relationship with Zn, Fe, K, Ca, Mn and Clay content in the soil. These results can contribute to improve our understanding of the basic mechanisms and progression of BW incidence and identify soil variables that can play a determinant role in predicting risk and evolution of BW in banana farms in tropical lacustrine soils.

The second issue, related to the relationship between banana productivity and soil properties, has been covered also in two studies. Chapter III contains the research oriented toward the development of an empirical correlation model to predict productivity based on soil characteristics. Five soil properties were found to have a clear agronomic and environmental importance: Mg, resistance to penetration, total microbial respiration, soil bulk density, and free-living omnivorous nematodes. This model could be used at the field level for the reliable identification of areas of high and low banana productivity in the studied areas of Venezuela.

Finally, Chapter IV presents a study which can broaden the usefulness of soil information derived from soil profile descriptions. It validated the hypothesis that it is possible to delimit areas of different productivity within banana farms, in the two main banana producing areas of Venezuela (Aragua and Trujillo states) using soil morphological properties (e.g., soil structure). For this, we developed a model of categorical regression prediction calibrated with soil morphological properties such as biological activity, texture, dry consistency, reaction to HCl and structure type. In the future, if further studies are conducted validating this approach in other environmental conditions, banana productivity could be improved using information which might be already available or can be acquired at a moderate cost using standard soil profile descriptions.

This PhD Thesis, has combined a systematic bibliographic review, crop and soil information from a systematic survey of different farm types in Venezuela with soil profile descriptions. Using that information, it has validated the hypothesis that by identifying the abiotic properties of the soil, the predisposition of the banana plant to the BW disease, and the potential productivity of the crop can be predicted. This approach can allow the differentiation of zones with different levels of productivity and BW risk, and as an immediate consequence, avoid areas of high risk or low productivity, or adapt agronomical practices to enhance productivity and sustainability of banana cropping systems in Venezuela.

Resumen General



Resumen General

La banana, fruta comestible de las Musáceas, es un alimento básico para más de 400 millones de personas en todo el mundo debido a sus atributos nutricionales y energéticos. Esto hace de las Musáceas cultivos de importancia global, particularmente en regiones tropicales, remarcando la importancia de la mejora de los sistemas de cultivo en Musáceas dentro de los esfuerzos actuales a nivel mundial orientados a una nueva revolución agrícola basada en la sostenibilidad productiva. Para lograrlo, son necesarias buenas prácticas para la producción de alimentos basadas en la investigación científica y técnica capaces de considerar la complejidad y variabilidad dentro del sector agroalimentario. La investigación presentada en esta Tesis Doctoral está orientada a dar respuesta a las causas de dos aspectos considerados de alta relevancia para la producción bananera, que afectan tanto la productividad como la sostenibilidad, siempre dirigidas hacia las condiciones de Venezuela, uno de los principales países productores a nivel mundial:

1- El impacto del riesgo fitosanitario relacionado con la Fusariosis Vascular y la influencia del suelo en la incidencia de la Marchitez del Banano (MB) causada por un complejo fúngico-bacteriano.

2- Una tendencia observada hacia la pérdida de productividad y la disminución de la calidad del suelo en algunas fincas comerciales de los estados de Aragua y Trujillo en Venezuela.

El primer tema, relacionado con la sanidad vegetal del banano, se ha abordado en dos estudios consecutivos. En primer lugar, en el Capítulo I se presenta una revisión sistemática sobre el efecto de los factores agroambientales en el impacto de la Fusariosis Vascular del banano, causada por *Fusarium oxysporum f. sp. cubense* (*Foc*) raza tropical 4 (TR4), y las implicaciones de esta enfermedad para el sistema de producción venezolano. Este Capítulo caracteriza sintéticamente información fiable sobre los factores bióticos y abióticos relacionados con la ocurrencia de *Foc* TR4, de forma conjunta al desarrollo de un análisis de riesgos y mapas de idoneidad climática para *Foc* TR4 en Venezuela. Este capítulo puede servir como un resumen básico del conocimiento disponible para el manejo de la enfermedad para que lo utilicen los técnicos y profesionales de la sanidad vegetal, así como para otras partes interesadas.

La investigación orientada a los aspectos fitosanitarios del banano se completa con el estudio presentado en el Capítulo II. Este capítulo analiza la relación entre las propiedades del suelo y la incidencia de la Marchitez del Banano (MB) una enfermedad de etiología desconocida, atribuida a un complejo fúngico-bacteriano, en un estudio de caso de una finca comercial bananera en el estado de Aragua en Venezuela, cuya incidencia ha reducido la superficie plantada en más de un 35,0% en los últimos años. La aplicación del algoritmo Random Forest permitió clasificar la incidencia de MB en suelos lacustres de Venezuela con base a las propiedades físicas y químicas del suelo con buena precisión, siendo una herramienta eficaz para la toma de decisiones en campo. Además, el uso de información de suelos en áreas bananeras de Venezuela permitió la identificación de lotes de banano con alta y baja incidencia de MB utilizando también el algoritmo Random Forest. El modelo mostró que el nivel de incidencia (alta o baja) de la MB se puede distinguir a través de su relación con el contenido de Zn, Fe, K, Ca, Mn y arcilla en el suelo. Estos resultados contribuyen a mejorar nuestra comprensión acerca de los mecanismos básicos y la progresión de la incidencia de MB, e identifican las variables del suelo que pueden jugar un papel determinante en la predicción del riesgo y la evolución de MB en fincas bananeras de suelos lacustres tropicales.

El segundo tema, relacionado con la productividad del banano y las propiedades del suelo, también se ha abordado en dos estudios. El Capítulo III contiene la investigación orientada al desarrollo de un modelo de correlación empírico para predecir la productividad del banano en base a las características del suelo. Se encontró que cinco propiedades del suelo tienen una clara importancia agronómica y ambiental: Mg, resistencia a la penetración, respiración microbiana total, densidad aparente del suelo y nematodos omnívoros de vida libre. Este modelo podría utilizarse a nivel de campo para la identificación confiable de áreas de alta y baja productividad bananera en las zonas estudiadas de Venezuela.

Finalmente, el Capítulo IV presenta un estudio que puede ampliar la utilidad de la información derivada de las descripciones del perfil del suelo. Se validó la hipótesis de que es posible delimitar áreas de diferente productividad dentro de las fincas bananeras, en las dos principales áreas productoras de banano de Venezuela (estados Aragua y Trujillo) utilizando propiedades morfológicas del suelo (por ejemplo, estructura del suelo). Para ello, se desarrolló un modelo de predicción de regresión categórica calibrado con propiedades morfológicas del suelo tales como actividad biológica, textura, consistencia seca, reacción al HCl y tipo de estructura. En el futuro, si se llevan a cabo más estudios que validen este enfoque en otras condiciones ambientales, la productividad del banano podría mejorarse utilizando información que podría estar ya disponible o puede adquirirse a un costo moderado utilizando descripciones estándar del perfil de suelo.

Esta Tesis Doctoral ha combinado una revisión sistemática de literatura, información de cultivos y suelos a partir de un muestreo sistemático de diferentes tipos de fincas en Venezuela con descripciones de perfiles de suelos. Con esa información, se ha validado la hipótesis de que, al identificar las propiedades abióticas del suelo, se puede predecir la predisposición de la planta de banano a la enfermedad de la MB y la productividad potencial del cultivo. Esta aproximación puede permitir la diferenciación de zonas con diferentes niveles de productividad y riesgo de la MB y, como consecuencia inmediata, evitar áreas de alto riesgo o baja productividad, incluso adaptar prácticas agronómicas para mejorar la productividad y sostenibilidad de los sistemas bananeros en Venezuela.

Objectives

Objectives

The general objective of the research carried out in this Doctoral Thesis was to determine and quantify, the influence of soil edaphic factors and agro-environmental variables in the differences of banana productivity and susceptibility to Banana Wilt disease occurring in areas of interest for banana commercial farming in Venezuela. This general objective was oriented through these four specific objectives:

a. To identify and analyze the main agro-environmental factors that can influence Fusarium Wilt disease caused by *Fusarium oxysporum f.* sp. *cubense (Foc)* tropical race 4 (TR4) and its dissemination in the Venezuelan production systems of Musaceae through a review of the current state of knowledge.

b. To evaluate the incidence of Banana Wilt and its relationship with the soil physical and chemical properties in a case study in a Venezuelan banana plantation through the performance of the supervised methods such as the Orthogonal Least Squares Discriminant Analysis (OPLS-DA) and the Random Forest algorithm.

c. To identify the main edaphic variables most correlated to banana productivity in Venezuela and explore the development of an empirical model to predict this productivity from numerical soil characteristics.

d. To study the potential use of soil morphological properties to differentiate levels of banana productivity in Venezuelan conditions using a prediction model based on soil categorical properties.

General Introduction

1. Importance of bananas in the world's food chain

Banana is the most consumed fruit in the world, and it is one of the most dynamic crops in international trade, being one of the most exported fruits. This dominance can be explained by characteristics of banana plants that give them comparative advantages over other fruits. Among the most important characteristics of this crop are that banana plants reproduce asexually sprouting shoots from an underground stem (Iskandar et al. 2018), it has a vigorous vegetative growth, shoots can produce a mature bunch in less than a year, the fruits can be harvested throughout the year (Scott et al. 2021), and shoots continue to sprout from a single plant year after year making bananas a perennial crop whose duration of plantation can last from 6 to 15 years (Varma & Bebber, 2019).

In other words, Musaceae (bananas and plantains) are perennial crops that grow quickly and can be harvested throughout the year. Around 116,781,658 t of bananas are produced in the world per year (FAO, 2020a). These figures are an approximation, since most of the world's banana production, almost 60.0 %, comes from relatively small plots and home gardens from which there are no reliable statistics (Figure 1). In many developing countries, the majority of banana production is destined for self-consumption or traded locally, thus playing an essential role in food security on those countries (Coltro & Karaski, 2019).

That is why the importance of bananas as a food crop in the tropics cannot be underestimated. In Africa, particularly in countries such as Uganda, for example, annual per capita consumption reaches about 243 kg, and in Rwanda, Gabon and Cameroon it ranges between 100 and 200 kg. In these four countries, bananas represent between 12.0-27.0% of the daily calorie intake of their populations (Sabiiti et al. 2018).

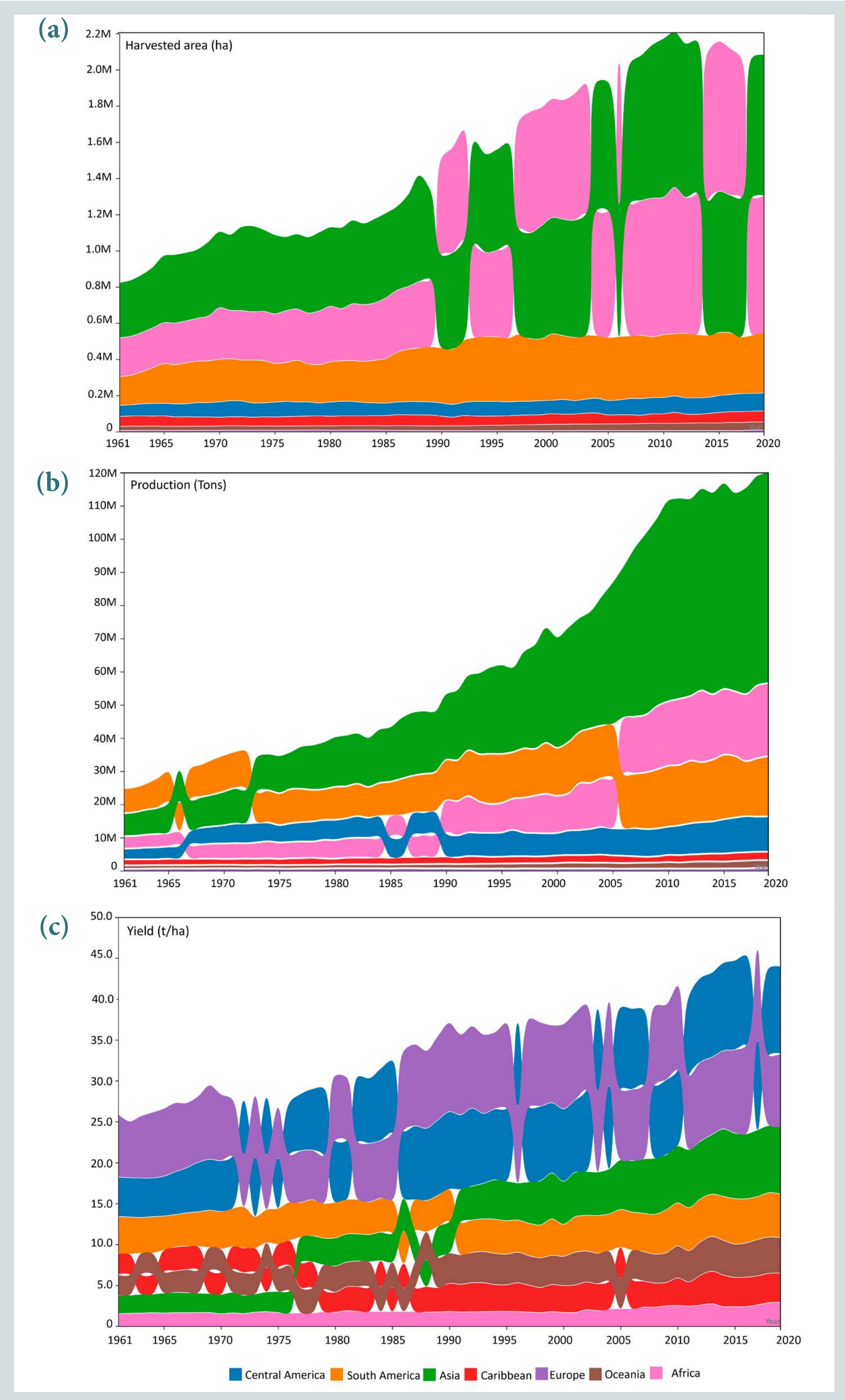


Figure 1. Bump Chart of historical evolution (1961-2019) of banana cultivation in Central America, South America, Asia, Caribbean, Europe, Oceania and Africa. (a) harvested area (ha), (b) production (t) and (c) yield (t/ha). Note: Interpreting a Bump Chart, when a line crosses another line, that is indicative of a change in rank. In other words, a crisscross in a bump chart indicates one entity (continent) has surpassed other in absolute terms even when comparison is based on relative ranks.

FAO reports (2020b) indicate that approximately 99.0% of banana world production takes place in developing countries, and practically all this world production corresponds to the Cavendish variety. Examined by region, 56.0% of banana production is originated in Asia, 26.0% in Latin America and 15.0% in Africa (Figure 1b). In the last two years, world trade grew rapidly, at 3.0% per year, and more than three-quarters of exports went to consumers in developed countries. However, Latin America is the third region in terms of banana production, behind Asia and Africa. The world's leading producing country of bananas is India with 30,460,000 tons per year, followed by China and Indonesia. In Latin America, Brazil is the leading producer with 6,812,708 t per year, followed by Ecuador, Guatemala, Colombia, Costa Rica, the Dominican Republic and Venezuela. These seven countries together account for more than half of the world production of Cavendish bananas (Figure 1b).

The world banana production has increased steadily between 1985 and 2019. Annual production increased by 49.0%, from 42.5 million tons in 1985-87 to 63.4 million tons in 1998-2000 (Figure 1b). This increase was due, firstly, to the expansion of the cultivated area (Figure 1a) and, to a lesser extent, to the increase in productivity. During this period, the average yield increased from 13.7 to 15.8 tons per hectare (15.0%) (Figure 1c).

According to the FAO (2020b), the projections suggest that world banana production will keep growing at a rate of 1.5% per year, reaching 135 million tons in 2028. Of all fruits, bananas would register the slowest growth due to the fact that its demand is saturated in most regions and so its growth is mainly driven by population growth. Banana is the main crop in terms of volume among tropical fruits and is expected to represent approximately 53.0% of total world tropical fruit production in 2028, compared to approximately 58.0% in 2009, as demand of other tropical fruits, particularly mango and avocado, has increased rapidly.

It is expected that Asia will maintain a high share of world production at approximately 54.0%, and India will continue to be the world's largest producer of bananas with an expected volume of 33 million tons a year. Production growth in India will be supported by strong domestic demand as a result of higher population growth. Production in the prominent exporting region of Latin America and the Caribbean (primarily Ecuador, Brazil, Guatemala, Colombia, Costa Rica, and Mexico) is expected to reach 34 million tons, fueled by demand for imports from key customers in the markets from developed countries.

2. Overview of banana production in Venezuela

Bananas represent a very important crop for Venezuela's economy, being one of the few alternatives to an oil producer economy. The monoculture areas in the main producing areas use banana clones (Musa AAA, Cavendish subgroup). Generally, these areas are characterized by being located at an altitude not above 600 m.a.s.l (Martínez et al. 2020a), concentrated mainly in the South of the Maracaibo Lake (Zulia state), the Central region, the

Central West and the East regions (Figure 2). However, it can be indicated that, in recent years, the introduction of the tetraploid FHIA-17 (*Musa* AAAA) has been observed in the banana cultivated areas of Aragua state (Rey et al. 2020).

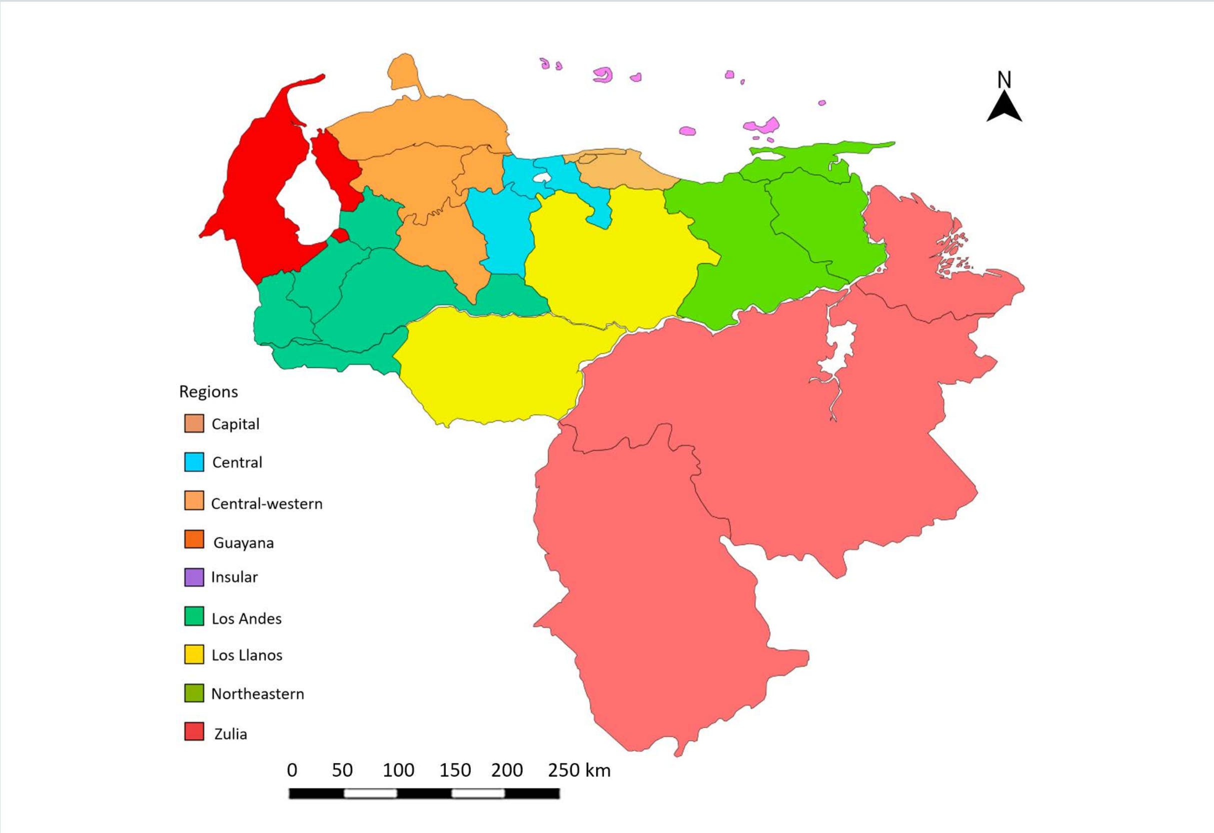


Figure 2. Map of Venezuela's states, colored according to the administrative regions.

Production data for Venezuela between 2007 and 2018 (FAO 2020b) indicate that it ranks eight among the ten important banana producing countries in America with an average of 480,103 t/year. For the year 2019, the production of bananas in the country was 650,051 t (Figure 3b) in a harvested area of 41,708 ha (Figure 3a) and with a yield of 15,58 t/ha (Figure 3c).

As can be seen in the data in Figure 3a, in 1992 the harvested area reached a maximum of 58,745 ha, from that date it suffered a sharp drop until it reached a minimum of 27,468 ha for 2008 (FAO, 2020a), basically due to bad survey of primary information and the social and political conditions of the country at that time.

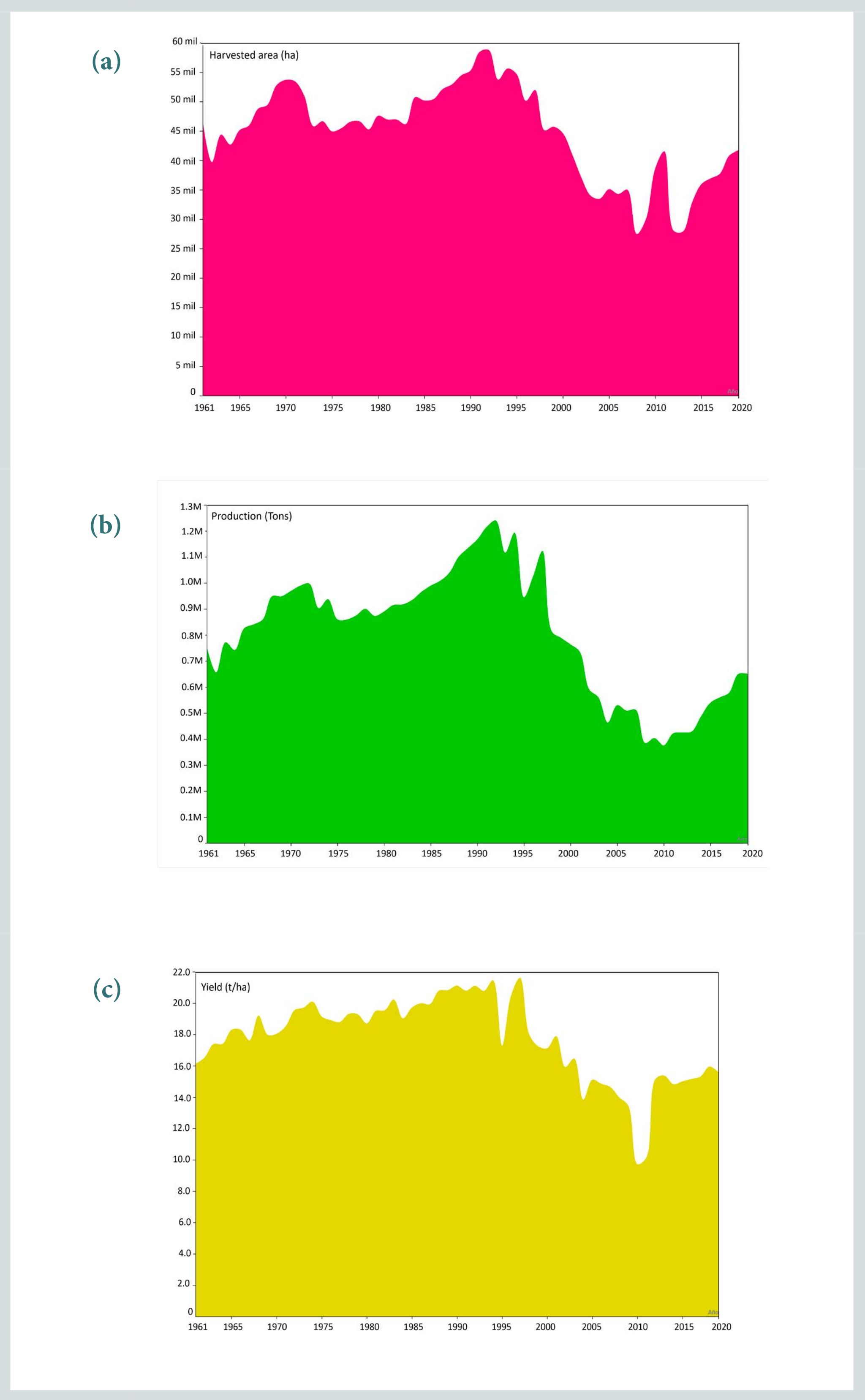


Figure 3. Historical evolution (1961-2019) of banana cultivation in Venezuela. (a) Harvested area (ha), (b) production (t) and (c) yield (t/ha) according to figures from FAOSTAT (2020).

During the period 1961-1991, the banana national average production was 928,191 tons per year. In 1992, production reached its maximum level with 1,239,480 t and in 2010 its lowest point with 375,118 t (Figure 3b). According to the FAO, the volume of production registered shows a growing trend from 2010 to 2019. On the other hand, the highest yield of national banana production was located in 1997 with 21,61 t/ha until reaching a minimum in 2010 of 9,71 t/ha (Figure 3c). In general, there is a decline in the banana production and harvested area in Venezuela from 1992 to approximately 2010 and as of this date a partial recovery has been noted.

3. Main phytosanitary problems in Venezuela

During the last 25 years, banana production in Venezuela has faced severe challenges due mainly to the shortage of agro-inputs (seeds, fertilizers, agrochemicals), problems of access to foreign exchange to satisfy internal demand, and the inadequate management of agricultural policies, as well as the impact due to drought, pests and diseases (Rey et al. 2020).

A recent study developed by Martínez et al. (2020a) establishes that among the main limiting factors of banana production in Venezuela are the occurrence of meteorological droughts (Olivares et al. 2017; Cortez et al. 2018), the attack of the black banana weevil (*Cosmopolites sordidus* Germar) (Rey et al. 2006; 2009); and the severe affectation of diseases such as the sigatoka complex (*Mycosphaerella* spp.); the bacteria *Erwinia carotovora*; a fungal-bacterial complex named Banana Wilt (BW); the Moko (*Ralstonia salanacearum*) and Fusarium wilt (*Fusarium oxysporum* f. sp. *cubense*) (*Foc*) race 1 (Figure 4).

According to Martínez et al. (2020a) there is a wide differential margin between the phytosanitary limitations caused by the Black Sigatoka complex (*Mycosphaerella fijiensis* Morelet) (Hernández et al. 2006) and Yellow Sigatoka (*Mycosphaerella musicola* Leach et Mulder) and the rest of the diseases (Gómez et al. 2012). The distribution of these phytosanitary problems in regions of production of Musaceae in Venezuela is indicated in Figure 4. It is observed that in the South region of the Maracaibo Lake, Black Sigatoka occurs in a higher proportion with respect to the rest of the diseases, followed by the cases reported for *Erwinia carotovora* and *Foc* race 1, considering the rest of them of a lesser degree of importance. Black Sigatoka appears in the rest of the regions, except in the Central one, as the main phytopathological problem.

In Los Llanos region, Black Sigatoka occurs with very narrow differential ranges with respect to the number of cases reported for Moko and Yellow Sigatoka. In the Central region, unlike the rest of the regions, yellow sigatoka together with a new disease called the fungal-bacterial complex (BW), represent the main diseases in crops, even when black sigatoka is present, but its degree of attack is attenuated. due to existing environmental conditions.

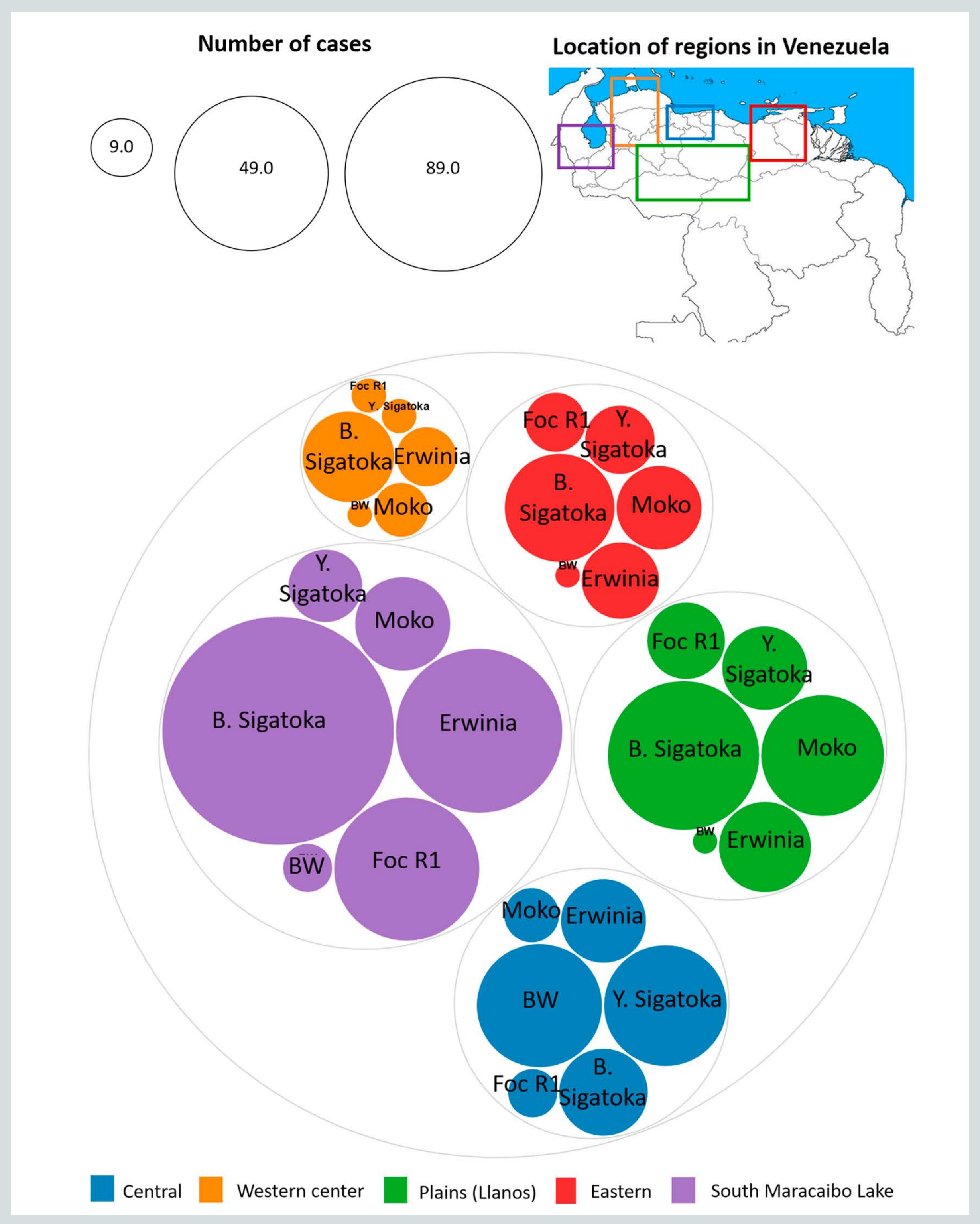


Figure 4. Distribution of the main phytosanitary problems in the Musaceae production regions in Venezuela according to Martínez et al. (2020a). B. Sigatoka = Black Sigatoka (*Mycosphaerella fijiensis* Morelet); Y. Sigatoka = Yellow Sigatoka (*Mycosphaerella musicola* Leach et Mulder); Foc: *Fusarium oxysporum* f. sp. *cubense* (Race 1). Erwinia = *Erwinia carotovora*; BW = fungal-bacterial complex called Banana Wilt; Moko (*Ralstonia salanacearum*). (Central Region, n = 40; Western Center Region, n = 15; Los Llanos Region, n = 50; Eastern Region, n = 23; South Region of Lake Maracaibo, n = 101).

In those producing areas characterized by high rainfall and high temperatures, and certain drainage limitations, there may be conditioning factors for the eventual incidence and severity of diseases, which is why it is recommended to implement integrated control with the use of propagation material. healthy (Hamed, et al. 2018), restriction of movement between zones (Martínez et al. 2020b), establishment of preventive strategies against emerging diseases to mitigate the risk of introduction and spread (Dita et al. 2018) and diversification of the production with clones resistant to new threats; disseminate this knowledge (Ahmad et al. 2020).

Biotic stress caused by emerging diseases continues to affect these production systems in various regions, causing partial damage or total destruction of the fields, and represent a considerable threat to the food security of Venezuela (Olivares et al. 2021). Unfortunately, the banana is seriously threatened worldwide by a "pandemic" caused by the *Foc* TR4 fungus, which once established in the field, can cause the complete loss of the harvest and for which there is no effective control measures until now.

The fact that the fungus is found in the soil and the pathogen can survive on it for more than 40 years, complicates its control and eradication. The only viable solution today is, through prevention and training of producers and monitoring of plantations to detect infected plants as early as possible and contain the spread of the fungus (Dita et al. 2017). In fact, a collaborative effort is underway between official institutions and the private sector so that this disease does not spread to countries such as Brazil, Ecuador, Costa Rica and Venezuela (Ghag et al. 2015). At the same time, research is continuing in the search for alternatives for effective biological or chemical treatments in order to protect plants or reduce disease severity.

The use of varieties with resistance to the disease is a key management strategy to face the problem, but supposes a great challenge: 80.0% of the world banana production comes from varieties susceptible to this disease. Furthermore, almost 50.0% of the bananas that are produced and 95.0% of what is exported originate from a Cavendish type or variety, which is susceptible to this TR4 (Ploetz, 2015). Definitely, to stop the disease caused by *Foc* TR4 an integrated approach is needed, where all the actors of the system are under alert and contributing from their possibilities.

4. Problems of banana soils in Venezuela

Most of the banana plantations in Venezuela are characterized by being aged (> 15 years), This aging has been identified since the end of the 20th century and in recent years the trend has not improved. They are generally aged plantations with low-quality materials, with a low technological level and with serious phytosanitary problems, managed by small and medium producers for subsistence purposes, and to a lesser extent for the national market (Nava, 1997; Delgado et al. 2008).

According to the studies of Rey et al. (2006); Delgado et al. (2008); Rey et al. (2009) and Delgado et al. (2010a) the application of conventional agricultural techniques and high-cost inputs have not been sufficient to reverse the trend towards reduced yields. This has also been linked to worsening of soil properties in banana plantations (Rey et al. 2020).

For example, in the last decade, the overall productivity of banana plantations in Costa Rica and Panama has shown considerable oscillations (FAO, 2020a) contrary to the stability in productivity in organically managed banana farms, which are characterized by the limited use of agricultural inputs (Serrano, 2003). Likewise, these organic production systems promote greater biological activity and therefore greater suppressiveness on pathogens associated with the root system than conventionally managed soils (Pocasangre, 2000; Menéses et al. 2003). This result in an economic advantage for the farmer, as well as an improvement of soil quality (Rosales et al. 2008).

Nowadays, the methods used to determine the productive potential of a banana soil are based on the appraisal of its physical and chemical properties (rarely considering biological properties), and on the effect of local topography and climate and agronomic management. In general, these methods are not sufficient to capture the complex interactions of the soil and its rhizosphere on banana productivity (Pattison et al. 2014). Due to this, there is a limited number of studies aimed at a broader evaluation of soil quality and its impact on banana productivity (e.g., Delgado et al. 2010a). Despite this, there is the need of further studies to enhance the understanding of soil physical, chemical and biological properties, quantitative or categorical, and its influence on the productivity of bananas, in many countries, such as Venezuela. This is a necessary step to address and solve the problem of depletion and low productivity of many banana plantations, which has been linked to the adverse impact of the conventional production system (Gauggel et al. 2003; Pattison et al. 2004; Delgado et al. 2010b).

Recent studies have demonstrated the possibility of using soil morphological properties in studies of soil quality and plant productivity, through statistical techniques and methods such as optimal scaling, categorical regression or the analysis of categorical principal components (Calero et al. 2018; Bouma, 2020), although to our knowledge this approach has never been tried in banana farms.

5. Innovative solutions using machine learning

Machine learning is used in the field of artificial intelligence to train predictive models that are used in different areas of science (Baviera, 2017). Machine Learning is a branch of Artificial Intelligence, where the computer increases its knowledge to fulfill a task. Through algorithms it has the ability to identify patterns in massive data to make predictions, allowing them to be autonomous, without the need to be programmed (Wang et al. 2020; Alimi et al. 2021).

Machine learning application uses supervised (algorithms that "learn" from data entered by a person) and unsupervised (algorithms learn from data with unlabeled elements looking for patterns or relationships between them) methods to support data analysis and calibration procedures. It consists of two phases, a training phase and a test phase, in the training phase there is a set of data (usually between 60% or 70% of the total available data) and 40% or 30% of the data remaining is used to make predictions and to validate the performance of the algorithm. (Van Engelen & Hoos, 2020). Machine Learning allows both identifying patterns among a considerable amount of data that may be of different nature and predicting behaviors through algorithms capable of learning and evolving based on their own experience (Ma et al 2017; Ye et al., 2020).

There are several examples of different techniques for machine learning that might be useful for predicting the incidence of diseases in bananas and achieve a more precise diagnosis. For instance:

a. Random Forest (RF) used by Owomugisha et al. (2014), Sangeetha et al. (2020), Ye et al. (2020), and Gómez-Selvaraj et al. (2020).

b. Support vector machines with linear or radial kernel (SVM) used by Hou et al. (2015), Aruraj et al. (2019), and Ye et al. (2020).

c. Classification trees (CART) applied by Manjunath et al. (2019)

d. Decision tree algorithm (C5.0) applied by Owomugisha et al. (2014)

e. Linear discriminant analysis (LDA) developed by Companioni et al. (2005)

f. Artificial Neural network used by Ye et al. (2020).

Many efforts have been made to calculate parameters such as vegetation indices, crop height, crop yield, leaf area index, surface soil properties, soil biomass, water stress, canopy height models, and more. A brief description of each algorithm is specified in the following paragraphs:

a. Random forest (RF): is an ensemble of many independent individual classification and regression trees (CART) and is defined as the equation 1, where, h represents the RF classifier, x is the input variable, and {θk} represents the independently identically distributed random predictor variables, which are used for generating each CART tree (Breiman, 2001; Liaw and Wiener, 2002). The final response of the RF is calculated based on the output of all decision trees.

$$\{h(x, \theta_k), k = 1,2, \dots i \dots\} \qquad (1)$$

b. Linear discriminant analysis (LDA): is a supervised classification method that is used to create machine learning models. These models based on dimensionality reduction are used in the detection of plant diseases (Xanthopoulos et al. 2013). Using Bayes' theorem, LDA estimates the probability that an observation, given a specific value of the predictors, belongs to each of the classes of the variable (Equation 2). Finally, the observation is assigned to class k for which the predicted probability is higher.

$$P(\mathrm{Y} = k|\mathrm{X} = x) \quad (2)$$

c. Classification and Regression Trees (CART): the representation used for CART is a binary tree. Predictions are made with CART by traversing the binary tree given a new input record. The tree is learned using a so call "greedy" algorithm on the training data to pick splits in the tree. Stopping criteria define how much tree learns and pruning can be used to improve a learned tree (Quinlan, 2007; Breiman et al. 1984).

d. Radial basis function kernel Support Vector Machines (RSVM): is a non-parametric supervised statistical learning classifier. The higher performance of SVM classifier makes it a favored alternative for plant disease detection. SVM considers structural risk minimization (SRM) principle to maximize the margin of class separation for better generalization performance of SVM. (Vapnik, 1995; Hearst et al., 1998). There are two parameters that need to be set when using an SVM classifier with the radial basis function kernel, i.e., the cost function (C) and the kernel width parameter (γ). The C parameter trades off the misclassification of training examples against the simplicity of the decision surface. The γ affects the smoothness of the class-dividing hyperplane (Ye et al. 2020). The equation 3 shows the Mathematical Definition of Radial Basis Kernel, where x, x´ are vector point in any fixed dimensional space.

$$K(x, x') = exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \qquad (3)$$

e. Linear kernel Support Vector Machines (LSVM): is a parametric model (Cortes and Vapnik, 1995). Given a set of samples xi (i=1,2,…M), where M is the number of samples. The set has two classes, those are positive class and negative class. We denote yi = 1 for the positive class and yi = −1 for the negative class, respectively. It is possible to find a hyperplane $f(\mathrm{x}) = 0$ that classifies the given dataset (equation 4), where w is a M-dimensional vector and b is a scalar, and they are used to define the hyperplane (Lei, 2017).

$$\mathrm{f}(x) = w^{\mathrm{T}}x + b = \sum_{j=1}^{M} w_j\, x_j + b = 0 \qquad (4)$$

f. C5.0 decision tree algorithm: The main two modes for this model are: a basic tree-based model and a rule-based model (Quinlan, 2014) C5.0 can create an initial tree model then decompose the tree structure into a set of mutually exclusive rules. These rules can then be pruned and modified into a smaller set of potentially overlapping rules (Kuhn et al. 2018). C5.0 uses the concept of entropy for measuring purity. The entropy of a sample of data indicates how mixed the class values are; the minimum value of 0 indicates that the sample is completely homogenous, while 1 indicates the maximum amount of disorder. The definition of entropy can be specified in the equation 5, for a given segment of data (S), the term c refers to the number of different class levels, and pi refers to the proportion of values falling into the class level i.

$$Entropy(S) = \sum_{I=1}^{C} -p_i \, log_2(p_i) \qquad (5)$$

Figure 5 provides an overview of the current use of these machine learning techniques for fruit crops using Scopus (2000-2020) by using VOSViewer (Van Eck & Waltman, 2010). The words machine learning, learning system and fruit are the central nodes, which establishes that they are high frequency keywords in the studies found. Four clusters are observed that represent four large thematic areas, the first cluster (red color in Figure 5) the words related to machine learning in banana, the second cluster (blue color in Figure 5) represents the learning system in agriculture, followed by the third cluster (green color in Figure 5) that is linked to Deep Learning in fruits and, finally, the cluster related to the application of classification algorithms (yellow color in Figure 5).

In short, there is scope for further research using machine learning techniques in agronomy, which within the context of this PhD Thesis could contribute to identify soil properties directly related to the incidence of diseases in banana, that could help to develop more sustainable banana production systems.

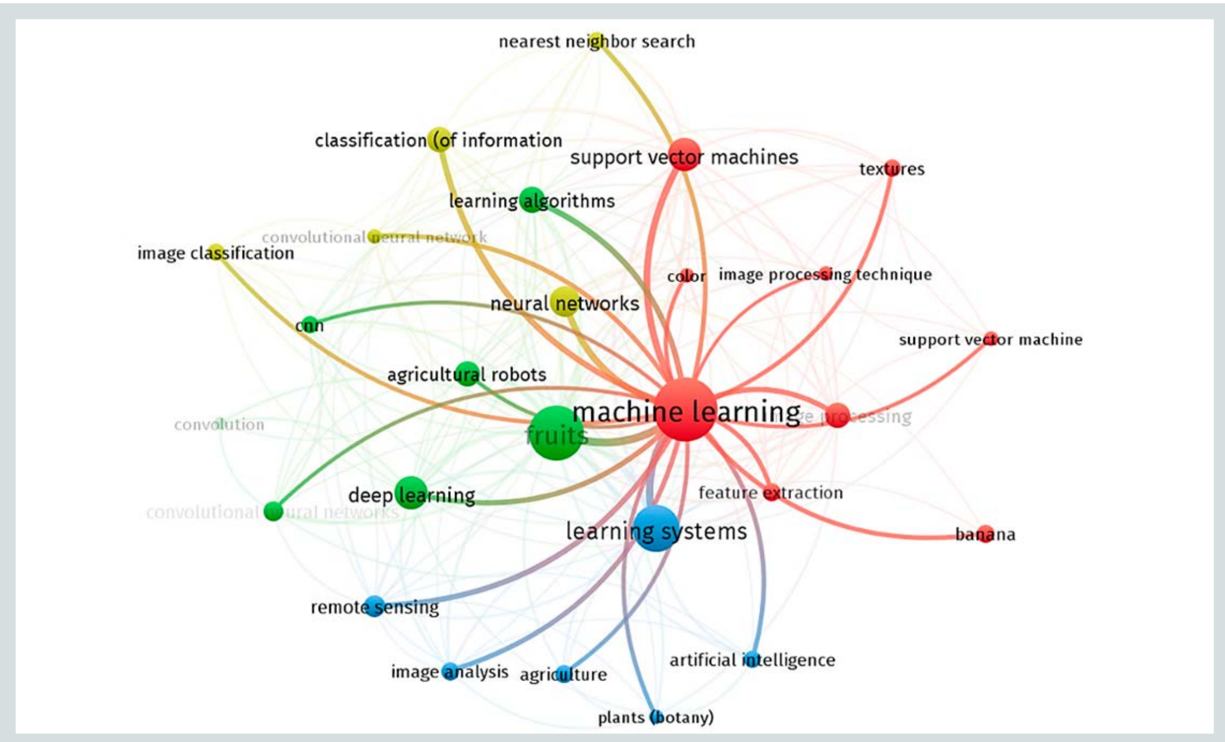
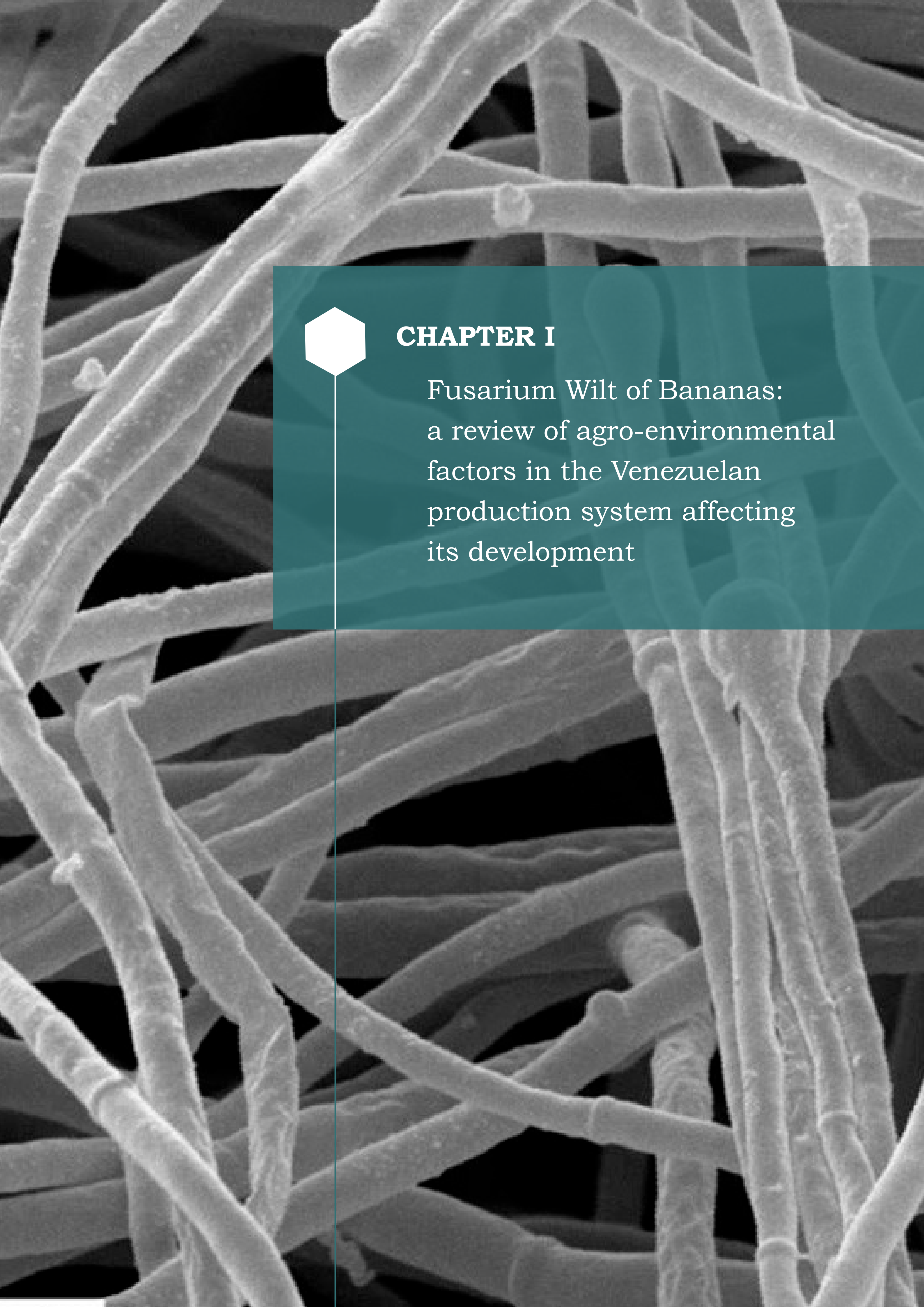


Figure 5. Co-occurrence of keywords with VOSViewer with clusters identified by different colors.



CHAPTER I

Fusarium Wilt of Bananas: a review of agro-environmental factors in the Venezuelan production system affecting its development

This is a literal reproduction of the article: Olivares B, Rey JC, Lobo D, Navas-Cortés JA, Gómez JA, Landa BB. (2021) . Fusarium Wilt of Bananas: A Review of Agro-Environmental Factors in the Venezuelan Production System Affecting Its Development. Agronomy, 11(5):986. https://doi.org/10.3390/agronomy11050986.

Fusarium Wilt of Bananas: A review of agro-environmental factors in the Venezuelan production system affecting its development

Barlin Orlando Olivares, Juan C. Rey, Deyanira Lobo, Juan A. Navas-Cortés, José A. Gómez and Blanca B. Landa

CHAPTER I. Fusarium Wilt of Bananas: A review of agro-environmental factors in the Venezuelan production system affecting its development

Abstract

Bananas and plantains (*Musa* spp.) are among the main staple of millions of people in the world. Among the main Musaceae diseases that may limit its productivity, Fusarium wilt (FW), caused by *Fusarium oxysporum* f. sp. *cubense* (*Foc*), has been threatening the banana industry for many years, with devastating effects on the economy of many tropical countries, becoming the leading cause of changes in the land use on severely affected areas. In this article, an updated, reflective and practical review of the current state of knowledge concerning the main agro-environmental factors that may affect disease progression and dissemination of this dangerous pathogen has been carried out, focusing on the Venezuelan Musaceae production systems. Environmental variables together with soil management and sustainable cultural practices are important factors affecting FW incidence and severity, excluding that the widespread dissemination of *Foc*, especially of its highly virulent tropical race 4 (TR4), is mainly caused by human activities. Additionally, risk analysis and climatic suitability maps for *Foc* TR4 in Venezuela have been developed. Although currently there are no effective management solutions available for FW control, this perspective provides an overview on the influence that environmental and agricultural variables would have on FW incidence and severity, giving some insight into management factors that can contribute to reducing its detrimental effects on banana production and how climate change may affect its development.

Keywords

banana diseases; climatic suitability; *Fusarium oxysporum* f. sp. *cubense*; pathogenic races; risk factors

1. Introduction

According to the FAO [1], the banana (*Musa* spp.) is a source of staple food for a large part of the world's population. Its annual production during the 2000–2015 period grew at a rate of 3.7%, reaching a record of 117.9 million tons in 2015, compared to 68.2 million tons in 2000. Due to the rapid growth of this crop, world banana exports, excluding plantain, reached the highest production of 20.2 million tons during 2019, with strong growth of the supply of two main producers (Ecuador and Philippines) being responsible mainly for the increased exports. World banana export volumes reached approximately 18.9 million tons during 2019. Preliminary estimates indicate a 4% growth in the largest net importer, the European Union, and a contraction of 1% in the United States.

In the Venezuelan territory, there were 82,000 productive hectares of banana 'Cavendish' (Musa AAA) and 'Hartón' (Musa AAB) in 2017, with a production of 424,649 tons destined for the local and export market, whose average yield was 13.91 tons/ha [1]. The production of Musaceae in Venezuela is concentrated in four large areas: the western (Zulia, Mérida, Táchira and Trujillo States), the southwestern (Barinas, Portuguesa and Apure States), the central (Aragua, Carabobo, Yaracuy, Vargas and Miranda States) and, to a limited extent, in the eastern (Sucre, Delta Amacuro State) zones of the country [2].

Banana production worldwide can be curtailed by several fungal diseases including aerial (e.g., Anthracnose and Fungal Scald, Botryodiplodia Finger Rot, Brown Spot and Diamond Spot, Cigar-End Rot, Cladosporium Speckle, Cordana Leaf Spot (Leaf Blotch), Pitting Disease, Sigatoka Leaf Diseases and Black Tip), soil-borne (e.g., Fusarium Wilt or Panama Disease, root rot) and postharvest (e.g., Crown Mold, Crown Rot and Pedicel Rot) diseases [3]. Among them, Fusarium wilt (FW) of bananas (FWB) caused by *Fusarium oxysporum* f. sp. *cubense* (E.F. Sm.) W.C. Snyder and H.N. Hansen (*Foc*), is the main threat and limiting factor for different banana cultivars of economic and strategic importance all over the world [4].

In recent decades, scientific interest in the FW of bananas has increased, especially in the main banana producing countries. This disease is the main phytopathological problem of banana plantations in tropical areas. However, despite the overwhelming impact that *Foc* has had over the years, and although there is extensive information concerning the biology and genetic diversity of this pathogen [4], there is still limited information available on its biogeography with concerning soil and climate, and in particular, there is no precise information on the agro-environmental factors that directly or indirectly affect the epidemiology of this disease [5]. This information would be relevant to a broader and more comprehensive understanding of the phytosanitary problem that *Foc* represents for banana plantations. Particularly, it can provide insight into its relationship with other fundamental agronomic components, which may be useful for the management of bananas and the disease. Therefore, this article aims to present an updated, reflective and practical review of the current state of knowledge of the main agro-environmental factors that affect the development and spread of FW of bananas, focusing on Venezuelan production. With the scientific knowledge collected in this report, it will be possible to design or select sustainable management strategies to prevent or help to reduce FW incidence in banana plantations.

2. The Causal Agent of FW of Banana and Its Geographical Distribution: The Risk Posed by Tropical Race 4 (TR4) of *Foc*

Historically, *Foc*, the causal agent of FW of bananas, is the main threat for different banana cultivars of economic and strategic importance worldwide. Detailed analysis and description of FW of bananas or its causal agent (*Foc*) has been thoroughly reviewed in the past [6–8], and more recently [4,9–15], although most of the analysis has devoted limited attention to the interaction between FW and agro-environmental factors.

Foc is a pathogen that inhabits the soil and produces three types of asexual spores: microconidia, macroconidia and chlamydospores [16]. The chlamydospores are highly resistant double membrane propagules that allow the pathogen to remain viable in the soil for many years in absence of a host, making it not possible to replant susceptible cultivars in the same soil once infested [17].

Three races of *Foc* (Race 1 (R1), Race 2 (R2) and Race 4 (R4)) and several Vegetative Compatibility Groups (VCG) within each race differing in virulence have been described in the populations of this pathogen [14,18,19]. *Foc* R1 is responsible for the epidemic in 'Gros Michel' and 'Manzano' clones, in addition to 'Pome' (AAB), 'Pisang Awak' (ABB), and 'Maqueño' (AAB). *Foc* R2 especially attacks bananas belonging to the subgroup 'Bluggoe' or 'Topocho' (ABB) [19]. *Foc* R4 is the most dangerous of all races because it attacks all these groups of banana plants, including the Cavendish clone [6,11].

The *Foc* originates from Southeast Asia and has coevolved in conjunction with the Musaceae in its center of origin, being reported in all the banana producing regions of the world (Figure 1), except in the south of the Pacific Islands, Somalia and riparian countries of the Mediterranean Basin [18,20]. The disease was first described by Bancroft in 1876 in Australia [9,20], and then by Ashby in 1913 in Costa Rica and Panama, where approximately 80,000 ha of the Gros Michel cultivar (AAA) were destroyed in Latin America by *Foc* R1 between 1890 and 1960 [9].

Foc R4 is the most virulent of the three races and is subdivided into Tropical (TR4) and subtropical (SR4) races. *Foc* SR4 attacks banana cultivars of the 'Cavendish' group in subtropical regions such as

Taiwan, the Canary Islands (Spain), South Africa and Australia. Studies on the recognition of TR4 as a different *Foc* pathotype have been proposed by Buddenhagen [21] and show that TR4 isolates are highly virulent even under environmental conditions non-conducive for disease development (e.g., cold climates) [19,22]. *Foc* TR4 affects banana cultivars of the 'Cavendish' group in Australia and the tropical regions of Southeast Asia including China, Indonesia, Malaysia and the Philippines (Figure 1) [8,23,24]. Indeed, very recently this *Foc* TR4 was found to be genetically distant from the other races and has been described as *Fusarium odoratissimum* [20].

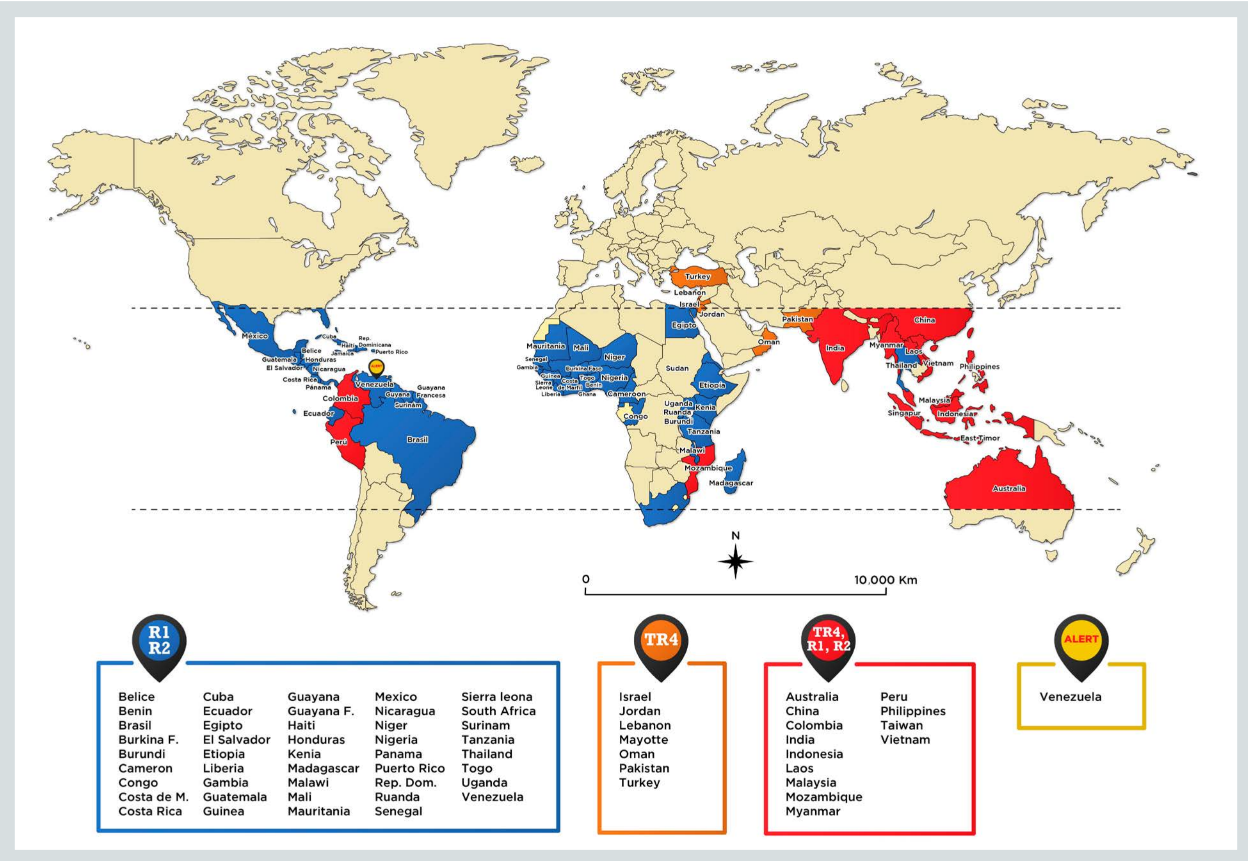


Figure 1. Geographical distribution of *Fusarium oxysporum* f. sp. *cubense* (*Foc*) races. Source: Adapted from CABI/EPPO [25], EPPO [26] and PROMUSA [27].

The fact that *Foc* R4 is destroying the cv. Cavendish in the tropics may cause unexpected harmful effects on production and exports in Southeast Asia and the 'Cavendish' market in the Western Hemisphere [4,20]. Thus, this situation threatens the production of small and medium banana farmers in Latin America and West Africa. In August 2019, the state of the national phytosanitary emergency was officially reported in Colombia due to the detection of *Foc* TR4 in 'Cavendish' banana crops in the municipalities of Dibulla and Riohacha (Guajira) [28]. In April 2021, the National Agrarian Health Service of Peru confirmed the finding of a banana orchard infected by *Foc* TR4 in the Piura Department. This situation represents a high risk for Colombia and Peru but also for other producing countries in the region.

In Venezuela, there is a latent concern of the spread of the TR4 from Colombia to bordering areas. So far, only the *Foc* R1 and R2 are present in Venezuelan producing areas [14]. Consequently, minimizing the spread of *Foc* TR4 will depend on strict compliance with established quarantine measures, such as preventing the transfer of banana shoots and rhizomes from affected to disease-free areas. In this way, the prevention strategy would protect production in the western part of Venezuela.

3. Disease Cycle of Fusarium Wilt of Bananas

Foc infects the plant through the root system reaching the xylem vessels, where it grows and multiplies, occluding them and limiting the nutrient and water absorption by the plant. The pathogen can also grow saprophytically in surrounding tissues as diseased plants die, forming chlamydospores that remain in the soil (Figure 2, phase 1). The chlamydospores are stimulated by plant root exudates, which subsequently germinate and infect the roots of nearby healthy plants (Figure 2, phase 2) [13]. After germination, and root infection, more mycelium and chlamydospores are produced, infecting secondary or tertiary banana roots (Figure 2, phase 3).

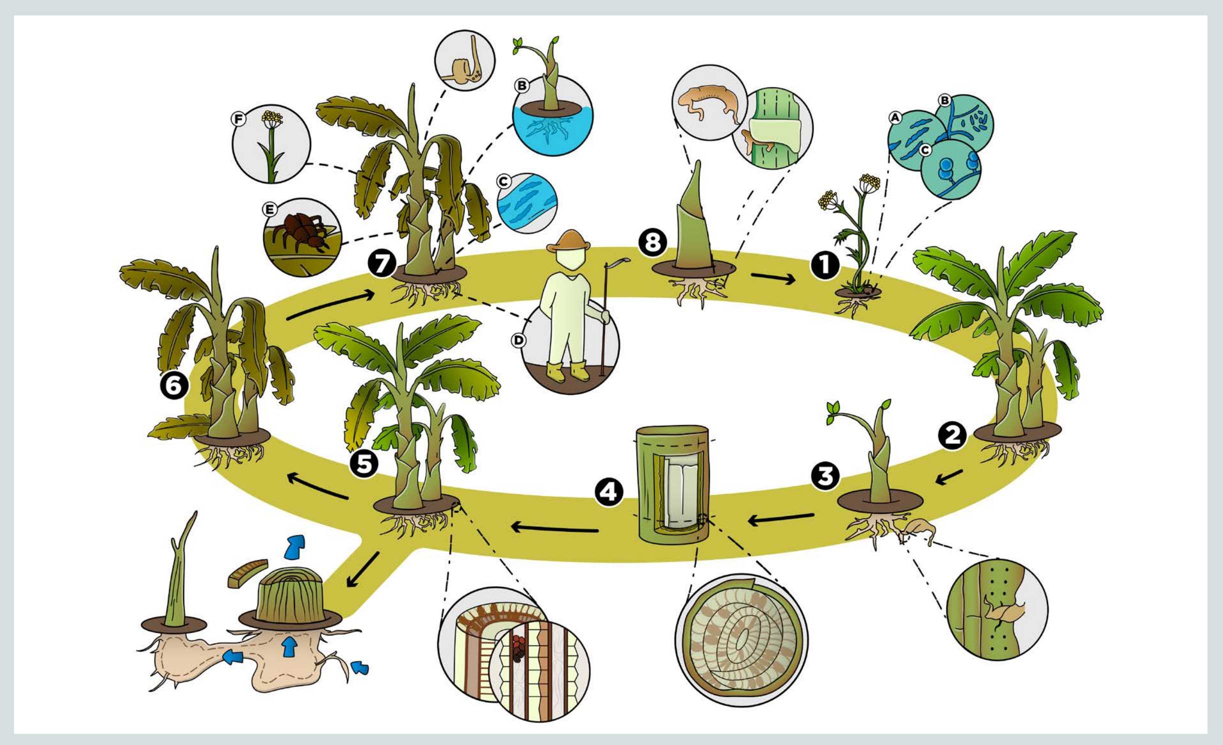


Figure 2. Schematic representation of the disease cycle of Fusarium wilt of banana, Ghag et al. [13] and Dita [14]. (1) *Foc* Spores (micro and macro conidia and chlamydospores); (2) germination of the chlamydospores; (3) colonization of the roots; (4) corm infestation; (5) development of wilt symptoms; (6) complete wilting of the mother plant; (7) pathogen dissemination: a. planting material; b. drainage water and runoff; c. irrigation water; d. workers; e. the weevil; f. weeds.

In susceptible plants, the fungus is not blocked by the host's defense mechanisms and the infection becomes systemic through the vascular system of the corm, the pseudo-stem, and the stem (Figure 2, phase 4). In resistant cultivars, the fungus is blocked by host responses and vascular occlusion of the infected xylem vessels, making the pathogen not able to continue infecting the corm [8,13]. In general, young plants have been described as most susceptible [6,7], due to the presence of younger roots that are more susceptible to *Foc* infection [15].

As *Foc* blocks the water flux in xylem vessels, the leaves turn yellow and wilt, making this effect more pronounced as leave age increases. Distinctive symptoms appear within the pseudo-stem, whose main characteristics are brown, red, or yellow ring-shaped lines. In the corm, brown stripes or specks appear (Figure 2, phase 5 and 6). Infected plants generally do not produce fruit, or the size of them is reduced [8] (Figure 3).

(a) (b)

Figure 3. *Fusarium oxysporum* f. sp. *cubense* symptoms in plants of banana clone 'Manzano' growing at 'El Diamante' Farm, El Cenizo Irrigation System, Sabana Mendoza, Trujillo State of Venezuela. (a) The first external symptoms of Fusarium Wilt are chlorosis and the death of the oldest leaves, which usually bend and collapse against the pseudo-stem. (b) Internal brown to red brick discoloration of the vascular system. [Pictures by Gustavo Martínez] (Trujillo-Venezuela, 2009).

Foc being a soil-borne pathogen, movement of infested soil can disperse soil particles infested by fungal structures (e.g., spores and mycelium). In addition, irrigation water, the weevil *Cosmopolites sordidus* and other secondary hosts such as ornamental plants and weeds can contribute to the spread of the pathogen [4,29]. The movement of plants as planting material also plays a significant role in pathogen dissemination (Figure 2, phase 7) [4]. A summary of the main *Foc* dispersal means is discussed below (see Section 4.6. Hosts and Dispersion).

4. Factors That Affects the Development of Fusarium wilt Epidemics in Banana Crops

Several abiotic and biotic factors of physical, chemical, microbial, climatic, or even sociocultural nature, have been shown to have an important role in the interactions among the banana crop, the pathogen and the soil environment. As a result of those complex and site-specific interactions, FW of banana may develop or not, leading to detrimental levels for banana production.

4.1. Environmental Conditions: Current and Future Climate Change Scenarios

In general, banana growth is not significantly limited by solar radiation or temperature, precipitation being the most important climatic factor for rain-fed plantations [30]. However, these factors can directly influence the occurrence of certain tropical diseases such as Black Sigatoka (*Mycosphaerella fijiensis* Morelet), Yellow Sigatoka (*Mycosphaerella musicola* Leach et Mulder) and False Panama disorder (of unknown etiology) [31]. However, in the case of banana-producing farms located in the extratropical zone, the temperature would represent a limiting factor for the adequate growth of the plants [32]. In the literature it is reported that banana cultivation is characteristic of tropical lowlands, in latitudes below 10°, altitudes around 100 m above sea level, minimum average temperature values of 19 °C and total rainfall exceeding 100 mm/month [33].

According to Nelson [16], the optimum temperature for in vitro growth of *F. oxysporum* isolates is between 25 to 28 °C, being restricted when the temperature exceeds 33 °C or is below 17 °C. Similarly, Pérez et al. [34] determined the effect of temperature on the growth of *Foc* isolates belonging to R1 and R2 races. They found that isolates of both races develop optimally in a wide range of temperatures between 23 to 29 °C, to rapidly decrease their growth outside this range. In a different study, Groenewald et al. [35] found differences in growth rate in vitro of *Foc* isolates at different incubation temperatures, identifying an optimum temperature of 25 °C for almost all isolates, while growth was limited at 10 and 35 °C and no growth was observed at 5 and 40 °C. Under field conditions, Brake et al. [36] showed that temperature mainly affects plant growth rather than directly affects the virulence of the pathogen. However, under temperature values around its optimal values for growth, higher severity of the disease would be expected since the growth and reproduction of the fungus would be favored by these environmental conditions. Moreover, high temperature values would also favor water stress in the plants, enhancing the severity of *Foc* symptoms. Ploetz [37] showed that temperature had an important effect on the development of FW disease. In this context, the level of susceptibility/resistance to *Foc* R1 in certain banana cultivars, such as 'Lacatan' or 'Gros Michel', would largely depend on temperature. That is, in certain banana cultivars the level of disease would increase under a temperature range optimum for disease development compared to that developed when suboptimal temperatures prevail.

Current research suggests that one of the main consequences of climate change would be the increase of temperatures and the alteration of rainfall patterns that could modify the incidence of pests, diseases and consequently impact the productivity of crops. Although the information on the occurrence records available for the different *formae speciales* of *F. oxysporum* for *Foc* in banana crops is limited, some studies have modeled the future distribution of other *formae speciales* of *F. oxysporum* estimating that at the species level, this fungus would find suitable areas for growth in North Africa, Middle Eastern and European countries for the years 2050 and 2100 [5]. In a similar analysis, Pérez-Vicente and Porras [38] have estimated for different climate change scenarios in Cuba, leading to a reduction in the geographic extent of the *Foc* pathogen, although FW would increase its severity when present. Additionally, in conditions of extreme events of heavy rains that could cause flooding, whose incidence might increase in some areas as a result of climate change, *Foc* could severely affect areas with susceptible cultivars such as the Burro (Bluggoe, ABB) types. These authors [5,38] also indicate that the reported temperature range for the pathogen exceeds the temperature range prevalent in areas where the disease occurs in Cuba, and the expected increase in temperature in the future would render these areas not suitable for FW development.

It is important to note that in rainfed cultivation, banana is very sensitive to water stress, particularly in areas without optimum soil and climate conditions, whose productivity can be reduced by up to 50% [39,40]. Waterlogging of the soil can produce symptoms of leaf yellowing and necrosis at the tips of the roots in banana plants, which can be confused with symptoms of FW. According to Lahav and Israeli [41], conditions of excessive moisture in the soil or waterlogging would predispose the plant to infection by the pathogen. The impact of oxygen deficiency on the interaction with *Foc* was evaluated by Aguilar et al. [42] based on changes in the activities of enzymes involved in phenol metabolism (phenylalanine ammonia lyase, PAL and peroxidase, PER) in banana cultivars differing in their reaction to *Foc*. Infected plants were subjected to hypoxia-induced changes in PER activity, which correlated with their resistance to FW. However, a breakdown of resistance to *Foc* of cv. Williams (a Cavendish cultivar) occurred when the soil was waterlogged. On the contrary, it has also been documented that flooding longer than 18 months destroys *Foc*'s reproductive structures [37]. On the contrary, it is well established that *Fusarium* spp. can survive in the soil for long periods when unfavorable drought conditions occur [43]. However, in arid or semi-arid conditions, with rainfall below 500 mm/year, it has been shown that soil moisture levels and temperature conditions can be unfavorable for *Foc* growth and development [40]. However, although mycelium and conidia survive for a short time in very dry soils, this stress situation represents a starting point for the development of *Foc* chlamydospores that can remain dormant for approximately 20 to 30 years [36].

4.2. Land and Soil Physical Properties

Bosman [32] studied the relationship between the slope in the field and the incidence of *Foc* in banana plantations in tropical soils of Costa Rica, finding that steep slopes have more erosion and nutrient losses than gentle slopes, where sediments often accumulate. Erosion also contributes to the spread of *Foc* inoculum that is transported with the sediments and runoff. In addition, steep and slope location influences the water availability in the soil and the amount of runoff water, and the movement of *Foc* inoculum close to the root system [44].

Soil structure, often expressed as the degree of stability of aggregates, exerts important influences on the edaphic conditions and the environment and results from the rearrangement, flocculation and cementation of soil particles. It is mediated by soil organic carbon (SOC), biota, ionic bridging, clay and carbonates [45]. Li et al. [46] analyzed two typical banana-growing soils (ultisol and inceptisol), which were either suppressive or conducive to the FW of banana from Hainan, China. They found that the suppressive soils had significantly more >2 and <0.053 mm aggregates, had a comparatively even size distribution of aggregates within the range of 0–0.25 mm, and a higher total carbon, total nitrogen and soil enzyme activity in the aggregates.

In the Canary Islands, Spain, Domínguez et al. [47] found that soils suppressive to *Foc* had high EC, higher levels of clay and soluble Na. In the same region, on volcanic soils, Domínguez et al. [48] found a clear separation between areas with and without FW, finding that the soluble K/Na ratio was always greater in affected areas, which is correlated with higher amounts of clay-sized particles and the increase of water-stable aggregate mass in these diseased areas. Moreover, the low potential buffering capacity for K observed in diseased areas suggests that massive K fertilizations might exert a negative effect on the disease development in banana plants. In Brazil, Deltour et al. [49] showed that soils with a higher level of suppressiveness to *Foc* R1 are characterized by higher clay content and higher pH, which suggests that soils with heavy texture could be less prone to the development of FW than sandy soils with lower pH.

4.3. Soil Chemical Properties

Multiple soil chemical properties including pH, organic matter content and the availability of some micronutrients in the soil are related to the suppressiveness to FW of certain soils. However, this effect largely depends on the soil type and climate, which makes it not possible to generalize the effect of those soil properties on the development of FW diseases. Nevertheless, for the specific case of FW of bananas, some research works have identified some specific chemical soil properties including pH, and organic matter, potassium, phosphorus, nitrogen and magnesium content (either by excess or deficient levels) as key factors contributing to reduce the susceptibility of banana crops to FW [50–52].

Concerning pH, *Foc* can grow in vitro in a wide pH range (optimal, 7.5–8.5) [36,46]. Under field conditions, some works have reported that, in the banana zones of Central America, the fungus seems to have different ranges of optimal pH depending on the soil type. Thus, several works have presented contradictory results. For example, Chuang [53] pointed out that the germination of *Foc* chlamydospores in the soil was negatively correlated with a pH of 8–10, which caused the pathogen to survive longer in alkaline soils (pH 8–10) than in very acid ones (pH 2–4) and Peng et al. [54] showed a higher incidence of FW in alkaline conditions. However, Bosman [32] recorded a high incidence of *Foc* in tropical banana soils with an acidic pH of 5.1, a condition in which the growth of the fungus in the soil was promoted [55].

Concerning soil nutrients, it has been observed in the banana plantations that certain fluctuations in the availability of specific nutrients may cause stress in banana plants and, consequently, increase their susceptibility to the attack of diseases, such as FW [56]. Within the macronutrients, potassium has been shown to have a direct and indirect influence on the incidence of FW. Soils with potassium deficiencies are directly correlated with a high incidence of the disease [57]. According to Domínguez et al. [48], the content of potassium in soil solution together with the presence of fine clay particles in the soil induce the suppression of FW in banana plantations in volcanic soils grown in arid and semi-arid regions, such as the Canary Islands.

In banana plantations in Costa Rica, the areas with a high incidence of *Foc* have a lower average phosphorus concentration (20 mg/L) than the areas with low incidence, resulting from this a nutrient that is very important not only for the health status of the plants but also for the incidence of FW [32,52]. In the Canary Islands, the study conducted by Borges [58] showed that zinc application to the soil significantly reduced the incidence and severity of FW in bananas in that area.

Concerning magnesium, a higher incidence of diseases in banana fields is related to high concentrations of this element in the soil. Therefore, in soil, a low level of magnesium might be recommended to increase plant resistance to FW. Despite the above, magnesium is usually applied at high concentrations as fertilizer to increase banana growth.

Interestingly, rather than the concentration of specific nutrients in the soil, some authors highlighted the importance of the Potassium/Magnesium ratio in the reaction of banana crops to *Foc*. Thus, Borges [59] reported that banana soils infected by *Foc* in the Canary Islands have the highest values of the Potassium/Magnesium ratio (0.67) compared to soils with healthy plants (0.48). The absolute values of potassium in these soils are commonly high due to the frequent applications of potassium salts as fertilizers. On the other hand, the absolute values of magnesium were low in soils that presented the disease. Other studies have indicated that Potassium/Magnesium ratio ranges between 0.55–0.81 are associated with serious losses due to FW in banana plantations in the Canary Islands [56].

4.4. Crop Management

At the banana production unit level, agroecosystem management seems to have a major role in the incidence of FW. For example, the use or establishment of a green cover in banana plantations based on fodder peanut (*Arachis pintoi*), grass carpet (*Axonopus affinis*) and indigenous native grass (*Paspalum conjugatum*) have been shown to have a positive impact on reducing the incidence of FW [60,61]. In line with this observation, it is commonly observed that a high percentage of bare soil (30.1%), which is related to a higher incidence of FW in the banana fields of Central America [32]. Experiences in this region and in Australia suggest that the presence of bare soil should be avoided and replaced by a green or brown cover [61].

Practices like biological disinfestation of the soil based on the incorporation of the cover crop or organic amendments (rice straw) of easy decomposition, flooding with irrigation and covering with a plastic film, leads to anaerobic conditions, which helps to control the attack of pathogens including *Foc* [62].

Intercropping and crop rotation are old practices that are used for disease control [39]. For the FW of banana, the rotation with Chinese chives (*Allium tuberosum*) is a cultural practice capable of reducing the incidence of *Foc* disease in China, significantly inhibiting the growth and causing the death of *Foc* spores [49,63].

Another aspect to consider is plant density. According to Bosman [32], if the plantation has a higher plant density (<3 m of plant spacing), the chances of the plants becoming infected is higher. A smaller distance between banana plants generates negative impacts on their health through competition, and increases the chances of *Foc* infection. However, the opposite effect may occur due to lower microbial activity, since the root activity is reduced and movement of water in the soil limited, resulting in a condition of less fertile soils and, therefore, more susceptibility to *Foc* attack [14].

Finally, there are studies indicating that the addition of chicken manure increases the inoculum of *Foc* in soil and the incidence of FW, despite being an agricultural practice that improves the soil fertility in banana plantations infected by *Foc* [51,64].

4.5. Soil Biota

The proportional size of abundance of microbial populations in soils appears to have a large influence of FW on bananas [54,65,66]. On this premise, microbial soil populations are essential to suppress *Foc* in this environment, while on the other hand, the physical and chemical properties of the soil affect the growth and development of the microbial population in the plant rhizosphere.

Bacterial and fungal communities were mainly determined by the organic matter content in banana soils in China [66]. These bacterial and fungal communities were significantly altered after long-term banana monoculture, indicating that the increase in fungal richness showed a significant correlation with the high incidence of FW disease. In this regard, Deng et al. [67] showed that the metabolic characteristics of the microbial communities present in a banana plantation in China, were significantly different when comparing healthy and diseased plants on the same banana plot.

Nowadays, the use of next-generation sequencing techniques is facilitating the study and the understanding of the plant-associated microbial communities and their shifts under varying conditions. Furthermore, the plant associated microorganisms or microbiome is recognized as a key factor behind the health of the plants [68]. Although there is still a lack of knowledge concerning the relationships between the microbiome profiles of the banana plant and the rhizosphere environment, some recent works are providing insights on the effects of banana-associated microbiome and the development of FW. Thus, some studies [68–70] found a determining beneficial role of Gammaproteobacteria present in banana soils in Central America, concluding that some members of this bacterial class were associated with the lack of successful infection of banana plants by *Foc* in soils infected by this pathogen. Additionally, Köberl et al. [70] found an increase in the *Pseudomonas* and *Stenotrophomonas* populations of healthy banana plants, whereas FW diseased plants showed an increase in Enterobacteriaceae. More recent studies have found that the Acidobacteria phylum was significantly elevated, but Bacteroidetes was significantly reduced in banana soils suppressive to FW. Additionally, certain bacteria belonging to the genera *Gp4*, *Gp5*, *Chthonomonas*, *Pseudomonas* and *Tumebacillus* were specifically enriched in suppressive soils, whereas *Gp2* was reduced [71]. However, it is important to point out that the exact mechanisms responsible for FW microbial suppression are probably due to complex microbial communities more than to specific bacterial genera.

The application of biological control agents (BCAs) not only effectively controls soil-borne pathogens such as *Foc*, but also significantly promotes plant growth and increases plant biomass. For the specific case of FW of bananas, several studies have dealt with the application of BCAs at the field-testing stage, with some of them showing high effectiveness (e.g., *Pseudomonas* spp. *Trichoderma* spp., *Bacillus* spp., non-pathogenic *Fusarium* strains and arbuscular mycorrhizal fungi) [15]. The use of BCAs for controlling FW of bananas is out of the scope of this article that rather focuses on the effect of indigenous soil biota development of FW of bananas; however, for those interested in this subject, Bubici et al. [15] has recently provided a comprehensive detailed and updated revision on this topic. Overall, under field conditions, the FW of banana has been controlled by up to 79% by using *Pseudomonas* spp. strains, and up to 70% by several endophytes and *Trichoderma* spp. strains. Lower biocontrol efficacy (42–55%) has been obtained with arbuscular mycorrhizal fungi, *Bacillus* spp. and non-pathogenic *Fusarium* strains [15].

Other soil biotas, apart from microorganisms, may have a strong influence on the suppressiveness to FW of banana of certain soils. Thus, modifications in the microbiological properties of the soil caused by the activity of meso-and macro-organisms, such as nematodes, beneficial arthropods and earthworms, have been identified as relevant factors in the incidence of FW in monoculture banana systems [66]. FW of bananas can be exacerbated by the presence of certain nematode species (e.g., *Radopholus similis*), due to root lesions that weaken the plant and facilitate the penetration of the pathogen through the injuries caused by nematode feeding [66,72]. In the banana agroecosystems, nematodes are one of the main causes of production losses. *R. similis*, *Helicotylenchus* sp., *Pratylenchus coffeae*, *Meloidogyne* sp. and *Rotylenchulus reniformis* are some of the plant-parasitic nematode species frequently associated with crop losses in banana crops [73,74].

According to Duyck et al. [75] and Zhong et al. [76] banana soils with a high inoculum of *Foc* or FW incidence show a reduced diversity of total nematodes. Thus, under such conditions, the populations of bacterivores (mainly in Rhabditidae, Pangrolaimidae and Cephalobidae families), some plant parasites (mainly within Meloidogynidae, Hoplolaimidae, Pratylenchidae and Rotylenchulidae families) and omnivores or predators (mainly in Qudsianematidae family) decreased in contrast to other groups of nematodes present in non-infested-healthy soils. Table 1 shows a summary of the agro-environmental variables that influence the incidence of *Foc* and the suppressive or conducive nature of soils to Fusarium wilt of bananas.

Table 1. Summary of the agro-environmental variables that influence the incidence of *Fusarium oxysporum* f. sp. *cubense* (*Foc*) and the suppressive or conducive nature of the soils to Fusarium Wilt (FW) of bananas.

Variable	Description	Source
Climate		
Temperature	Favorable temperature ranges from 23 to 29 °C, with optimum at 25 °C; Limited growth at 10 to 35 °C and no growth ≤5 or ≥40 °C	[35] [34]
Precipitation	Water deficit: oxygen deficiency in the radical system favors *Foc* infection Water Excess: poorly drained soils with heavy textures favor FW	[42] [41]
Soil physical characteristics		
Slope	Convex curvature slope favors FW	[32]
Distribution of the size of aggregates	Conducive soils: <0.053 mm Suppressive soils: >2.0 mm	[46] [45]
Texture	Suppressive soils: higher clay content (high pH) Conducive soils: sandy texture (low pH)	[49]

Variable	Description	Source
Soil chemical characteristics		
pH	Optimal in vitro growth of *Foc*: 7.5–8.5	[36] [46]
Acidity	pH of 5.1 increases the availability of toxic aluminum and favors the growth of *Foc*	[55]
Potassium	Conducive soils: potassium deficiencies correlate with high FW incidence	[57]
Phosphorus	Conducive soils: concentration < 20 mg/L	[32] [52]
Zinc	Suppressive soils: the application of zinc in the soil significantly reduces FW incidence	[58]
Magnesium	Conducive soils: 7.9–10.6 cmol/kg	[56]
Ratio K/Mg	Conducive soils: >0.67 Suppressive soils: ≤0.48	[59]
Soil biological properties		
Bacteria	Suppressive soils: higher levels of Acidobacteria phylum, *Gp4*, *Gp5*, *Chthonomonas*, *Pseudomonas* and *Tumebacillus* genera Suppressive soils: increase in *Pseudomonas* and *Stenotrophomonas* populations Conducive soils: Increase of Enterobacteriaceae	[71] [70]
Nematodes	Conducive soils: high presence of bacterivores nematodes (Rhabditidae, Pangrolaimidae and Cephalobidae), plant parasitic (Meloidogynidae, Hoplolaimidae, Pratylenchidae and Rotylenchulidae), omnivores or predators (Qudsianematidae)	[75] [76]
Crop management		
Green cover	Green cover of forage peanut (*Arachis pintoi*), carpet of grass (*Axonopus affinis*) and indigenous native grass (*Paspalum conjugatum*) reduces the incidence of *Foc*	[60] [61]
Bare soil	Conducive soils: average bare soil of 30.13% Suppressive soils: average bare soil of 11.65%	[32] [61]
Distance between plants	Conducive soils: <3.0 m Suppressive soils: ≥4.0 m	[32]
Crop rotation	The rotation with Chinese Chives (*Allium tuberosum*) inhibits the growth of *Foc*	[63]
Chicken manure	Increases the inoculum of *Foc*	[51] [64]

4.6. Hosts and Dispersion

Plant genetic resistance is generally considered the most plausible strategy and economically feasible measure to effectively manage the FW of bananas [11,12]. However, resistant cultivars might not match the demands of the market and the available resistance may be overcome by new pathogenic strains, as was the case of Cavendish and *Foc* TR4 [22]. Once FW is established in an area, the use of resistant varieties is the most effective, if not the only means to manage the disease. In bananas, complete resistance has only been described in the Cavendish (AAA)-*Foc* R1 interaction. Other interactions, such as Prata (AAB)-*Foc* R1 and Giant Cavendish Tissue Culture Variants (GCTCV)-*Foc* TR4, show intermediate resistance, i.e., those banana genotypes develop less severe symptoms than susceptible varieties when grown under similar environmental conditions and *Foc* inoculum levels. However, under-management practices and inoculum pressure are conducive for disease development; FW and yield losses will increase gradually [14]. Currently, efforts are being made to unravel the genetic and molecular mechanisms driving resistance responses of different banana genotypes using state-of-the-art molecular approaches such as deep-RNA sequencing [77]. A description of the levels of resistance and its implications on FW of banana management is described in Dita et al. [14].

Foc is a facultative saprophytic fungus, with the ability to survive in weeds and grasses. Thus, *Foc* can invade roots of weeds present in banana plantations from hyphae growing from the tissues of senescent banana roots that remain in the soil for long periods [78]. This could explain the persistence of the fungus in soils without banana crops [7]. Thus, the pathogen can infect the roots of certain weeds without causing visible symptoms and can remain in these plants in the absence of a banana crop [79]. The study of Hennessy et al. [78] found that the monocotyledonous species *Chloris inflata* and three dicotyledonous species: *Euphorbia heterophylla*, *Tridax procumbens* and *Cyanthilium cinereum*, can sustain *Foc* inoculum in their root system (Table 2). Among the different plant species that can be hosts of *Foc*, it is of particular relevance to avoid or control the export of the ornamentals *Canna indica*, *Aglaonema pictum* and *Hedychium coronarium*, especially in areas close to FW affected fields, since these species can serve as alternative hosts for the pathogen [80]. Additionally, it is known that monocotyledons are excellent endophytic carriers of *F. oxysporum* [49] and could contribute to the increase or maintenance of *Foc* inoculum.

The systemic growth of *Foc* in xylem tissues of infected asymptomatic banana plants represents one of the primary ways in which the fungus can be introduced into a free pathogen growing area [12,13]. The influence of the anthropogenic component by the active movement of infected planting material, equipment and people between infected and not affected areas, had important repercussions in the dissemination of *Foc* R1 and TR4 [14]. According to Stover [81], the epidemic of *Foc* R1 in 'Gros Michel' was due to the absence of quarantine measures, the use of infected planting material and the establishment of new plantations using machinery and propagation material from infected fields [14]. Thus, the pathogen is dispersed by the movement of propagating material and infested agricultural tools that have contaminated soil, runoff water during rainy events, floods and irrigation [6,11]. With the rain, the spores of the pathogen, as well as the infected material, are transported to the drainage channels and through irrigation water, these spores infect new areas.

Table 2. Main host plants (type, family and species) of *Fusarium oxysporum* f. sp. *cubense.*

Type of Plant	Family	Species	Reference
Ornamental	Heliconiaceae	*Heliconia caribea* *Heliconia latispatha* *Heliconia chartacea* *Heliconia collinsiana* *Heliconia crassa* *Heliconia rostrata* *Heliconia marie* *Heliconia vellerigera*	[50] [8] [26]
	Cannaceae	*Canna indica*	[80]
	Araceae	*Aglaonema pictum*	
	Zingiberaceae	*Hedychium coronarium*	
Crop	Musaceae	*Musa sp.* *Musa schizocarpa* *Musa textiles* *Musa acuminata* *Musa balbisiana*	[50] [19] [11,12] [26]
Weeds	Commelinaceae	*Commelina diffusa*	[29] [78]
	Poaceae	*Choris inflata* *Ixophorus unisetus*	
	Asteraceae	*Tridax procumbens*	
	Euphorbiaceae	*Euphorbia heterophylla*	
Grass	Poaceae	*Paspalum fasciculatum* *Panicum purpurascens*	[29] [51,78]

5. Risk Analysis of *Fusarium oxysporum* f. sp. *cubense* Occurrence in Banana Plantations in Venezuela

Venezuela has a diverse ecology and it is divided into several agroecological zones (ZAE). This division in ZAEs is based on the geographic location, edaphic and climatic characteristics, agricultural potential and the predominant agricultural production systems [82]. We have established the delimitation of the main areas of Musaceae cultivation in Venezuela and the total area yield by using the Space Production Allocation Model (SPAM) 2005 [83], developed by the International Food Policy Research Institute (IFPRI), at a spatial resolution of 5 arc-minutes (approximately 10 km^2) (Figure 4a) [84].

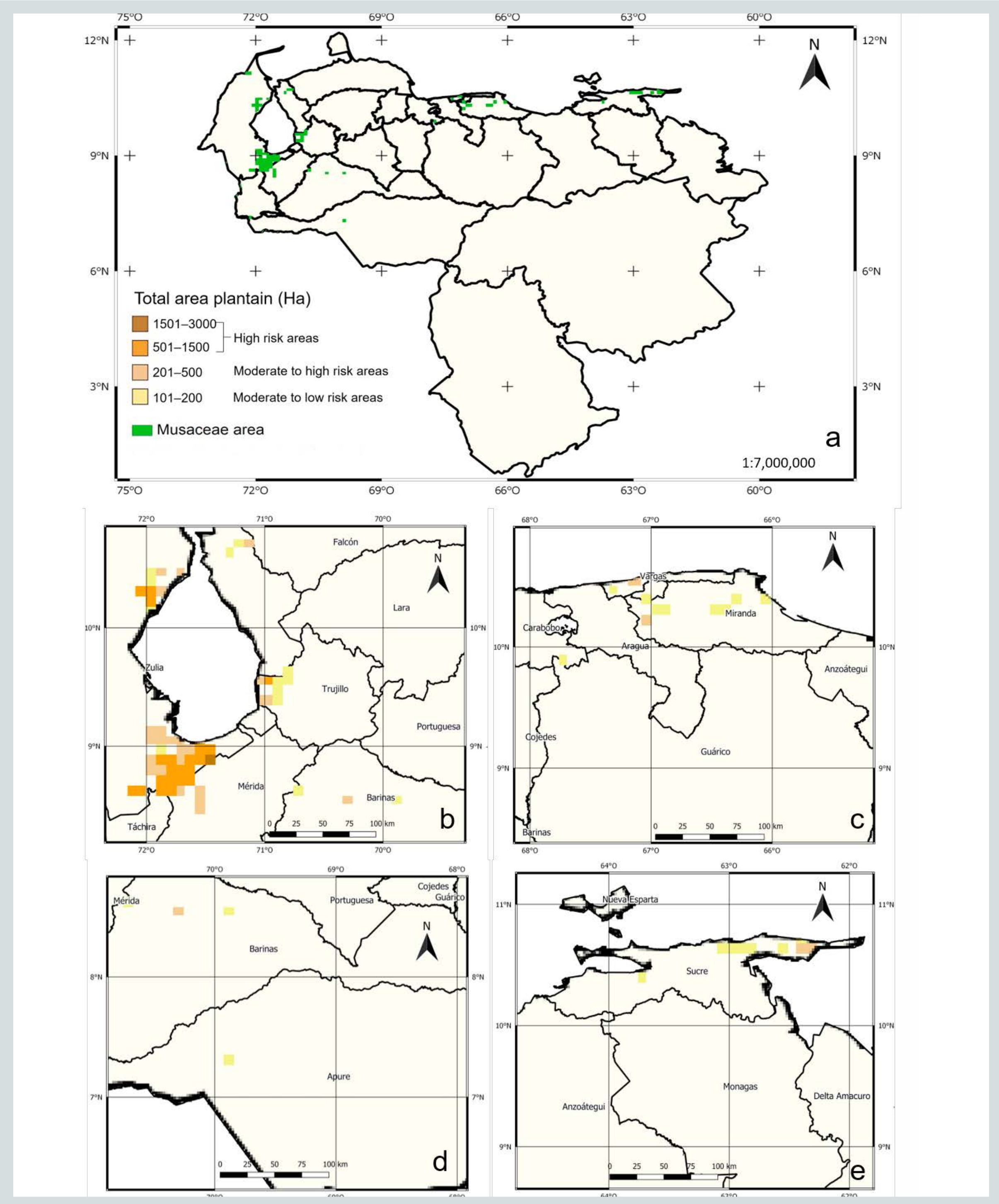


Figure 4. (a) Total area of Musaceae production in Venezuela (Source: USDA) and potential risk of *Fusarium oxysporum* f. sp. *cubense* occurrence in Venezuela in (b) the western region, (c) central region, (d) southwestern region and (e) eastern region.

The designated areas with a different potential risk of *Foc* occurrence in the main banana producing areas in Venezuela were based on the assembly of agro-environmental factors (climate, soil type and agronomic management) (Figure 4a). Additionally, these analyses were complemented by five Musaceae experts of Red Venezolana de Musáceas (Musaven), who classified production systems with different levels of risk or susceptibility (high-, moderate- and low-risk areas), according to the characteristics established by FAO [79].

The classification was based on the prevailing characteristics of the different Musaceae production systems in Venezuela with four important steps, including risk identification (sources, communities and production systems), analysis (probability and consequences), evaluation (prioritization) and assessment (response to the risk) (Table 3). Our results defined three areas with a different potential risk of *Foc* occurrence in Venezuela:

Table 3. Location and main characteristics of Musaceae producing areas in Venezuela.

Location	Climatic Characteristics *	Edaphic Characteristics	Predominant Production Systems
Eastern Zone			
Sucre State: Andrés Eloy Blanco and Andrés Mata Municipalities	Ecoclimatic Region: Subhumid premontane tropics (C1) Altitude: 0–500 m AP: 700–900 mm AAT: >24 °C	In flat areas predominate soils with good drainage and good natural fertility In areas with high slope soil quality is at risk due to erosion	Diversification of the uses of Musaceae (bananas and plantains)
Sucre State: Bermúdez Municipality	Ecoclimatic Region: semiarid (G1) Altitude: 0–500 m AP: 400–500 mm AAT: >24 °C	Areas of valleys and plains with low slope, with saline soils. Irrigation and water quality are determining factors for agricultural use	It is characterized by rainfed subsistence and semi-commercial agriculture.
Sucre State: Valdez Municipality	Ecoclimatic Region: Humid tropics low (B3) Altitude: ≤500 m AP: >1.800 mm AAT: 18–24 °C	Areas with very frequent or almost permanent flooding caused by tidal flows and river flooding	Small banana plantation areas commonly associated in indigenous conucos
Central Zone			
Miranda State: Municipalities: Carrizal and Los Salias	Ecoclimatic Region: Subhumid premontane tropics (C1) Altitude: 1.300 m AP: 1.300–1.500 mm AAT: 18–24 °C	Areas of soils with moderate drainage and moderate natural fertility	Highly productive systems and potential importance for diversification of Musaceae with vegetables and other fruits

Location	Climatic Characteristics *	Edaphic Characteristics	Predominant Production Systems
Miranda State: Municipality Acevedo	Ecoclimatic Region: Humid tropics low (B1) AP: >1.800 mm AAT: >24 °C	Areas with a predominance of flat topography, with moderate to good natural fertility soils	Intensive banana production systems
Miranda State: Brion and Buroz Municipalities Aragua State: Libertador Municipality	Ecoclimatic Region: Subhumid tropics low (A1) Altitude: <200 m AP: 700–1.800 mm AAT: >24 °C	It includes flat areas, with soils of good to moderate drainage, moderate natural fertility and risk of physical deterioration of the soil due to compaction and surface sealing	Large-scale production systems, the Buroz municipality is the most important in banana production in Miranda state
Vargas State	Ecoclimatic Region: Subhumid tropics low (A4) Altitude: <500 m AP: 700–1.200 mm AAT: >24 °C	Flat areas, with low to very low natural soil fertility and excessive tendency drainage	Medium scale production systems
Western Zone			
Trujillo State: Bolivar and La Ceiba Municipalities Zulia State: La Cañada de Urdaneta, Colón, Francisco Javier Pulgar and Miranda Municipalities	Ecoclimatic Region: Subhumid tropics low (A1) Altitude: <200 m AP: 1.200 mm AAT: >24 °C	Flat areas, with soils of good to moderate drainage, high fertility, and risk of physical deterioration of the soil due to compacted layers (plow floor), surface sealing, crusting and water erosion	Commercial plantations of Musaceae on the alluvial plains Maracaibo Lake
Trujillo State: Sucre and Candelaria Municipalities Merida State: Alberto Adriani and Sucre Municipalities Táchira State: Ayacucho Municipality	Ecoclimatic Region: Humid premontane tropics (D1) Altitude: 400–600 m AP: 1.600–1.800 mm AAT: 18–24 °C	Its main limitations are water erosion and acidity of soils, which coexist with some areas of better fertility	Commercial and semi-commercial plantations of Musaceae
Zulia State: Guajira Municipality	Ecoclimatic Region: Dry areas of the low tropics (G1) Altitude <500 m AP: 400–600 mm AAT: >24 °C	Saline soils and high risk of physical deterioration due to surface sealing. The irrigation and water quality are determining factors for agricultural use	Subsistence and semi-commercial small-scale agriculture

Location	Climatic Characteristics *	Edaphic Characteristics	Predominant Production Systems
South-western Zone			
Apure State: Paéz Municipality Barinas state: Obispos Municipality	Ecoclimatic Region: Subhumid tropics low (A5) Altitude: <500 m AP: 1000–1.100 mm AAT: >24 °C	Flat areas with good to moderate drainage soils and moderate fertility	The largest commercial systems of Musaceae in southern Maracaibo Lake and Zulia State
Barinas state: Pedraza and Barinas Municipalities	Ecoclimatic Region: Humid tropics low (B2) Altitude: <500 m AP: >1.800 mm AAT: >24 °C	Areas with a varied topography and soils with low to very low natural fertility	Systems of semi-commercial production of Musaceae from 100 to 200 ha

* Ecoclimatic Region according to INIA [81]; AP: Annual precipitation; AAT: Average annual temperature.

5.1. High-Risk Areas

Cavendish banana large-scale farms located in the municipalities Colón, Francisco Javier Pulgar and La Cañada de Urdaneta in Zulia State are at high potential risk (Figure 4b) because: (i) they include large plantations (500–1900 ha) of genetic clones highly susceptible to *Foc* TR4 like Cavendish (Pineo Gigante, Williams); (ii) they present edaphoclimatic characteristics (precipitation greater than 1200 mm and mean diurnal temperature range of 24–34 °C) suitable for the appearance and establishment of the pathogen; (iii) they implement intensive production systems, which requires the continuous movement of employees between plantations; and (iv) a single source of water for irrigation can favor the spread of the fungus within and across plantations on soil particles and irrigation water. Additionally, the physical proximity with Colombia puts these western municipalities of Venezuela at risk because the border with Venezuela is approximately 120 km from the *Foc* infected area in Colombia and at a distance of around 300 km from the banana zones at the south of the Maracaibo Lake.

5.2. Moderate to High-Risk Areas

Medium-scale Cavendish banana production systems (200–500 ha) can be considered areas of moderate to high vulnerability to *Foc* TR4. These plantations are mainly located in La Guajira in Zulia State, some areas in the Alberto Adriani and Sucre municipalities at the State of Mérida and Barinas municipality at Barinas state (Figure 4d). As indicated for large-scale production systems, the pathogen can enter and disseminate due to the movement of employees, materials, tools and irrigation water within and between farms.

5.3. Moderate to Low-Risk Areas

Production systems based on mixed banana varieties supplying local markets (100–200 ha) are considered a moderate to low risk to the *Foc* TR4. These areas are located mainly in the central and eastern parts of the country, some areas of Trujillo State, Aragua and Miranda state (Figure 4c), Sucre state (Figure 4e) and south of Maracaibo Lake (Figure 4b). This risk category also includes production systems based on intercropping.

6. Climatic Suitability for *Fusarium oxysporum* f. sp. *cubense* TR4 Occurrence in Venezuela

Maximum Entropy (MaxEnt) models are widely used to model potential distribution of organisms covering diverse aims including finding correlation of species occurrences, mapping their current geographic distributions, or predicting new times and places [85]. MaxEnt estimates a target probability distribution by finding the probability distribution with maximum entropy (i.e., that is most spread out, or closest to uniform), subject to a set of constraints that represent our incomplete information about the target distribution [85]. MaxEnt integrates species' occurrences with background data (i.e., randomly selected points) from spatial environmental gradients in the study area and generates the probability of species' presence [85]. It identifies areas that have conditions most like species' current known occurrences and ranks them from '0′(unsuitable or most dissimilar) to '1′(most suitable or most similar).

In this perspective, MaxEnt model v. 3.4.3 [85] was used to estimate the potential for the establishment of *Foc* TR4 in Venezuela. Presence-only data was obtained from the current known global distribution of *Foc* TR4 obtained from the following sources: (i) The CABI Crop Protection Compendium (CABI/EPPO, 2015) [25], (ii) the ProMusa project website (https://www.promusa.org/Tropical+race+4+-+TR4#Distribution) (07/03/2021) [27] and (iii) from reviewing the specialized literature. We only selected references that contained geographical information about the presence of the pathogen. To reduce spatial autocorrelation, presence records were submitted to spatial filtering, delimiting a minimum distance of 1 km between each locality data [86] using the spThin package in R and repeated four times. For pseudo-absences generation, a weight for presences and absences to simulate a prevalence of 0.1 was used. Climate data was obtained from Chelsa Climatology [87] that includes monthly mean temperature and precipitation patterns for the 1979–2013 period. Nineteen bioclimatic variables were derived from monthly temperature and precipitation values and are intended to approximate climate dimensions meaningful to biological species and represent annual characteristics (e.g., Temperature Annual Range), seasonality (e.g., Precipitation Seasonality) and extreme environmental factors [88].

Multicollinearity was addressed by the variance inflation factor (VIF) using a threshold of 10 [89]. The climate suitability was estimated based on bioclimatic variables representing annual trends. Nine non-correlated bioclimatic variables were selected based on VIF for model fitting. These variables included: mean diurnal range [mean monthly (maximum temperature—minimum temperature)] (bio2), isothermality (bio3), mean temperature of the wettest quarter (bio8), mean temperature of the driest quarter (bio9), mean temperature of the warmest quarter (bio10), precipitation of the coldest month (bio13), precipitation seasonality (bio15), precipitation of the warmest quarter (bio18) and precipitation of the coldest quarter (bio19). The contribution of these bioclimatic variables assessed by their variance importance based on AUC indicated that the potential geographic distribution of *Foc* TR4 in Venezuela is mostly influenced by the mean diurnal temperature range (bio2) and precipitation, but particularly during extreme cold (bio13) and warm (bio18) periods of the season.

The MaxEnt predictions were generated with the sdm package in R [90] using a five-fold cross-validation procedure. Model performance was evaluated using several widely established threshold-dependent statistics: specificity, sensitivity, and the true skill statistics (TSS), as well as threshold independent statistics: the area under the curve (AUC) of the Receiver Operating Characteristic (ROC) plot and Cohen's Kappa [91].

Results indicated that the performance of the model was considered high. The AUC was estimated in 0.93 ± 0.0236, the sensitivity in 0.94 ± 0.052, the specificity 0.87 ± 0.044, and the TSS and the Cohen's kappa were estimated in 0.81 ± 0.052 and 0.55 ± 0.085, respectively. The map with the continuous suitability scores for the *Foc* TR4 in Venezuela drawn from the MaxEnt model is presented in Figure 5a and the Figure 5b show four-level categorized suitability. Except for some areas in the central and eastern part of the Bolivar state, the northern part of the Amazonas state and the Andes Mountains in the west, for which the potential for establishment is estimated negligible, the rest of the country could have climatic conditions that would allow the establishment of *Foc* TR4, with the central part of the country being classified with low suitability between 0.1 to 0.3; interestingly, all regions in northern states are classified as moderately suitable between 0.3 to 0.6 and most of the banana producing areas in the states of Zulia, Trujillo, Miranda and Sure have climatic conditions that are estimated as highly suitable >0.6 for the establishment of *Foc* TR4 (Figure 5).

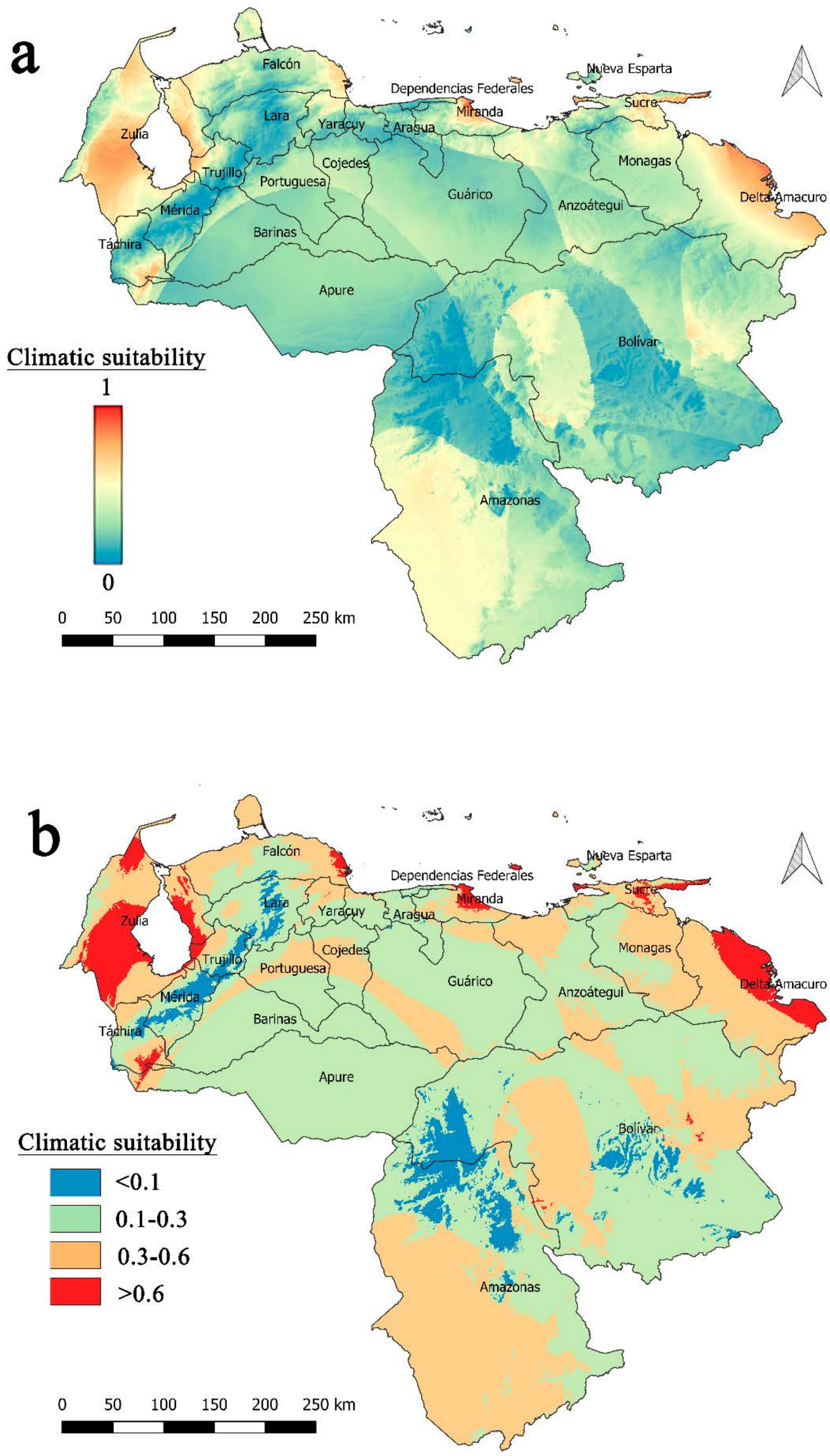


Figure 5. Estimated climatic suitability maps for *Fusarium oxysporum* f. sp. *cubense* TR4 according to a Maximum entropy model. (a) Climatic suitability index 0 to 1; (b) Climatic suitability index categorized in four levels: unsuitable (<0.1), low (0.1–0.3), moderately (0.3–0.6) and highly suitable (>0.6).

7. Integrated Approach to the Prevention of *Fusarium oxysporum* f. sp. *cubense* TR4 Occurrence in Venezuela

The recent occurrence of *Foc* TR4 in commercial farms in the Colombian Guajira region in July 2019, has increased the risk to *Foc* TR4 on the Venezuelan banana-producing areas due to proximity to the outbreak. According to Dita et al. [22,92], these border areas will be necessary to implement state policies, with the exclusion being the main priority to prevent the movement of propagation material, establish quarantine sectors and promote awareness campaigns to inform producers on the serious threat that *Foc* TR4 represents for the Venezuelan banana sector.

In Venezuela, intensive banana systems located in the western region (Zulia state) are more at risk to *Foc* TR4 as extremely strict biosafety protocols are required. Due to the absence of efficient control measures once the plant is infected in the field, the best way to protect the crop is to prevent the introduction of the pathogen from affected areas and the use of resistant banana cultivars to *Foc* TR4 such as Giant Cavendish tissue-culture variants [93].

Small-scale Cavendish banana production systems located in most of Venezuela's territory are characterized by using sprouts as a propagation system instead of tissue crops. Farmers also tend to share planting and transport tools, and neglect the sanitary practices at the time of planting, which considerably increases the chance of introducing the fungus from affected crops to pathogen-free plantations. On the contrary, production systems based on mixed banana cultivars, as well as mixed crops, subsistence and indigenous farming systems in areas of the southwest and east of Venezuela are considered less vulnerable areas to the economic damage that the occurrence of *Foc* TR4 pathogen attack, due to the diversity of banana cultivars and the possibility to replace banana by other crops.

In this context, Table 4 lists the main activities that could be implemented to prevent the entry and spread of this pathogen, and more specifically of the highly virulent *Foc* TR4 in Venezuela. In summary, results indicate that the best approach to fight against the threat that *Foc* TR4 assume includes prevention approaches, the use of resistant/tolerant cultivars, as well as the implementation of good phytosanitary practices at planting time. However, these measures are very expensive and require highly trained personnel to identify the pathogen and control it. Unfortunately, in areas still free of this highly virulent race, such as the case of Venezuela, there are few specialists working on this disease. Consequently, it is important to build technical capabilities in growers, services providers and national plant protection agents to identify *Foc* symptoms in the field, and to implement identification methodologies based mainly on the use of modern molecular diagnostic tools to unequivocally identify this highly virulent race of the pathogen.

Table 4. Activities aimed at preventing the entry and spread of *Fusarium oxysporum* f. sp. *cubense* (*Foc*) TR4 in the banana production systems of Venezuela. Adapted from FAO [79].

Type of Action	Description of the Activity	Production Systems			
		1	2	3	4
Legislation and regulation	Request in vitro plants and certificates for disease indexing				
	Strengthen border control to avoid imports of bananas and plant parts from countries in which *Foc*-TR4 is known to occur, in particular due to the short distance, from the infected plantations in Colombia				
	Strengthen surveillance for early detection of potential introductions of the pathogen				
On-border	Manage and control the movement of visitors and vehicles to farms by cleaning and disinfestation				
	Obtain clean certified planting and propagation material				
	Control the use of agricultural machinery, equipment and tools among neighboring farms				
	Use of amendments and fertilizers to overcome the constraints of the soil regarding pH, nitrogen, calcium, silicon, iron, etc				
Early detection	Capacity building in: symptoms recognition caused by *Foc* TR4 and differentiate from other banana diseases, sampling of plant parts and soil, treatment and manipulation of samples and diagnostics (isolation and new molecular diagnostic tools)				
	Restrict personnel, equipment and animal access in suspicious and infected areas				
	Establish quarantine if a new *Foc* TR4 outbreaks are confirmed, delimit control area and eradication and destruction of affected and surrounding plants				

1 = Large-scale Cavendish monoculture; 2 = small-scale Cavendish monoculture; 3 = small and medium scale mixed banana; 4 = small-scale mixed crops, subsistence and indigenous agriculture.

Very important, Moderately important, Less important.

8. Conclusions

The progress made in FW of banana research in tropical producing regions is still insufficient to adequately define the mechanisms of prevention and control of this disease [10,22]. Despite several studies finding correlations between incidence and virulence of FW with different climate variables, soil properties and management practices, there is not a comprehensive understanding of the overall impact of these properties and their interactions. Previous research carried out by scientific and academic institutions, together with the producer associations, shows that no product (biological or chemical) is capable of effectively eradicating this disease once it has been detected and established in a field, although some studies indicate a reduction of FW using biological control agents (BCAs). Consequently, the implementation of eradication or exclusion protocols in affected areas has been suggested as one of the main management strategies in order to prevent the spread of the pathogen, especially of the highly virulent tropical race 4 (TR4) of *Foc* to areas free of the pathogen or specifically of the TR4 strain. Nevertheless, the agro-environmental factors compiled in this perspective may provide an overview of the influence of environmental and agroecological variables on the susceptibility of banana plants to FW. With this information, agricultural practices in banana plantations could be improved, soil management practices promoted and new farm locations proposed in suitable areas, which will lead to more opportunities to control the coexistence between *Foc* and banana production in Venezuela as well as in other banana producing countries.

Acknowledgments

The authors wish to thank the researchers Gustavo Martínez from the National Institute of Agricultural Research of Venezuela and Cesar Acevedo from University of Talca (Chile) for providing access to information on Fusarium in Bananas and their valuable comments. Also, to José Riera for the graphic design and Miguel López Beltrán for cartographic support.

References

1. FAO. Banana Statistical Compendium year 2019. Available online: http://www.fao.org/3/cb0466en/cb0466en.pdf (accessed on 16 March 2020).

2. Olivares, B.O.; Araya-Alman, M.; Acevedo-Opazo, C.; Rey, J.C.; Cañete-Salinas, P.; Kurina, F.G.; Balzarini, M.; Lobo, D.; Navas-Cortés, J.A.; Landa, B.B.; et al. Relationship between soil properties and banana productivity in the two main cultivation areas in venezuela. *J. Soil Sci. Plant* **2020**, *20*, 2512–2524, doi:10.1007/s42729-020-00317-8.

3. Ploetz, R.C.; Nishijima, G.A.; Rohrbach, W.T.; Kg Ohr, H.D. *Compendium of Tropical Fruit Diseases*; The American Phytopathological Society: Washington, DC., USA, 1994; pp. 7–28.

4. Kema, G.H.J.; Drenth, A.; Dita, M.; Jansen, K.; Vellema, S.; Stoorvogel, J.J. Editorial: Fusarium

wilt of banana, a recurring threat to global banana production. *Front. Plant. Sci* **2021**, *11*, 1–6.

5. Shabani, F.; Kumar, L.; Esmaeili, A. Future distributions of *Fusarium oxysporum* f. spp. in European, Middle Eastern and North African agricultural regions under climate change. *Agric. Ecosyst. Environ.* **2014**, *197*, 96–105, doi:10.1016/j.agee.2014.08.005.
6. Stover, R.H. *Fusarium Wilt (Panama Disease) of Banana and Other Musa Species*, Phytopathology paper 4; Commonwealth Mycological Institute: Kew Surrey, UK, 1962; pp. 17–29.
7. Ploetz, R.C. Panama-Disease—return of the first banana menace. *Int. J Pest. Manag.* **1994**, *40*, 326–336, doi:10.1080/09670879409371908.
8. Ploetz, R.C.; Pegg, K. *Fusarium Wilt in Diseases of Banana, Abaca and Enset*; CABI Publishing: Wallingford, UK, 2000; pp. 143–159.
9. Pérez-Vicente, L. Marchitamiento por Fusarium (Mal de Panamá) en bananos: Una revisión actualizada del conocimiento presente sobre su agente causal. *Fitosanidad* **2004**, *8*, 27–38.
10. Pocasangre, L.E.; Ploetz, R.C.; Molina, A.B.; Vicente, L.P. Raising awareness of the threat of fusarium wilt tropical race 4 in latin america and the caribbean. *Acta Hort.* **2011**, *897*, 331–337, doi:10.17660/ActaHortic.2011.897.45.
11. Ploetz, R.C. Fusarium wilt of banana. *Phytopathology* **2015**, *105*, 1512–1521, doi:10.1094/phyto-04-15-0101-rvw.
12. Ploetz, R.C. Management of fusarium wilt of banana: A review with special reference to tropical race 4. *Crop. Prot.* **2015**, *73*, 7–15, doi:10.1016/j.cropro.2015.01.007.
13. Ghag, S.B.; Shekhawat, U.K.S.; Ganapathi, T.R. Fusarium wilt of banana: Biology, epidemiology and management. *Int. J. Pest. Manag.* **2015**, *61*, 250–263, doi:10.1080/09670874.2015.1043972.
14. Dita, M.; Barquero, M.; Heck, D.; Mizubuti, E.S.G.; Staver, C.P. Fusarium wilt of banana: Current knowledge on epidemiology and research needs toward sustainable disease management. *Front. Plant Sci.* **2018**, *9*, doi:10.3389/fpls.2018.01468.
15. Bubici, G.; Kaushal, M.; Prigigallo, M.I.; Cabanas, C.G.L.; Mercado-Blanco, J. Biological control agents against fusarium wilt of banana. *Front. Microbiol.* **2019**, *10*, 1, doi:10.3389/fmicb.2019.01290.
16. Nelson, P.E. History of fusarium systematics. *Phytopathology* **1991**, *81*, 1045–1048.
17. Davis, R. Fusarium wilt (Panama disease) of banana.. *Pest. Adv. Leaflet* **2005**, *42*, 1–8.
18. Bentley, S.; Pegg, K.G.; Moore, N.Y.; Davis, R.D.; Buddenhagen, I.W. Genetic variation among vegetative compatibility groups of *Fusarium oxysporum* f. sp. *cubense* analyzed by DNA fingerprinting. *Phytopathology* **1998**, *88*, 1283–1293, doi:10.1094/phyto.1998.88.12.1283.

19. Ploetz, R.C. Fusarium wilt of banana is caused by several pathogens referred to as *Fusarium oxysporum* f. sp *cubense*. *Phytopathology* **2006**, *96*, 653–656, doi:10.1094/phyto-96-0653.

20. Maryani, N.; Lombard, L.; Poerba, Y.S.; Subandiyah, S.; Crous, P.W.; Kema, G.H.J. Phylogeny and genetic diversity of the banana Fusarium wilt pathogen *Fusarium oxysporum* f. sp. *cubense* in the Indonesian centre of origin. *Stud. Mycol.* **2009**, *92*, 155–194, doi:10.1016/j.simyco.2018.06.003.

21. Buddenhagen, I. Understanding Strain Diversity in *Fusarium oxysporum* f. sp *cubense* and History of Introduction of 'Tropical Race 4' to Better Manage Banana Production. *Acta Hortic*. **2009**, 828, 193–204, doi: 10.17660/ActaHortic.2009.828.19

22. Dita, M.; Perez-Vicente, L.; Guzman, M.; Urias, C.; Staver, C. Fusarium tropical race 4 and it threat to food security and the banana industry in Latin America and the Caribbean. *Phytopathology* **2017**, *107*, 8–8.

23. Molina, A.B.; Fabregar, E.; Sinohin, V.G.; Yi, G.; Viljoen, A. Recent Occurrence of *Fusarium oxysporum* f. sp *cubense* Tropical Race 4 in Asia. *Acta Hortic*. **2009**, 828, 109–115, doi: 10.17660/ActaHortic.2009.828.10

24. Garcia-Bastidas, F.; Ordonez, N.; Konkol, J.; Al-Qasim, M.; Naser, Z.; Abdelwali, M.; Salem, N.; Waalwijk, C.; Ploetz, R.C.; Kema, G.H.J. First report of *Fusarium oxysporum* f. sp *cubense* tropical race 4 associated with panama disease of banana outside southeast asia. *Plant. Dis.* **2014**, *98*, 694–694, doi:10.1094/pdis-09-13-0954-pdn.

25. CABI/EPPO. *Fusarium oxysporum* f. sp. *cubense*. *Distribution Maps of Plant Diseases, Map 31*, 7th ed.; CABI: Wallingford, UK, 2015.

26. EPPO. PQR database. Paris, France: European and Mediterranean Plant Protection Organization, 2014. Available online: https://gd.eppo.int/taxon/FUSACB/distribution (accessed on 16 March 2021).

27. PROMUSA. Available online: http://www.promusa.org/Tropical+race+4+-+TR4#Distribution (accessed on 16 March 2021).

28. FONTAGRO. Available online: https://www.fontagro.org/new/proyectos/fusarium-R4 (accessed on 16 March 2021).

29. Waite, B.H.; Dunlap, V.C. Preliminary host range studies with *Fusarium oxysporum* f. *cubense. Plant. Dis. Rep.* **1953**, *37*, 79–80.

30. Olivares, B.O. Tropical rainfall conditions in rainfed agriculture in Carabobo, Venezuela. *La Granca* **2018**, *27*, 86–102, doi:10.17163/lgr.n27.2018.07.

31. Martínez, G.E.; Rey-Brina, J.C.; Rodríguez, D.; Jiménez, C.; Rodríguez, Y.; Rumbos, R.; Pargas-Pichardo, R.; Martínez, E. Análisis de la situación fitopatológica actual de las musáceas comestibles en Venezuela. *Agr. Trop.* **2020**, *70*, 1–20, doi:10.5281/4323273.

32. Bosman, M. *Role of the Environment on the Incidence of Panama Disease in Bananas*; Wageningen University: Wageningen, The Netherlands, 2016; pp. 12–68.

33. Robinson, J.C.; International, C.A.B. *Bananas and Plantains*; CAB International: Wallingford, Conn, USA, 1996; p. 312.

34. Pérez, L.; Batlle, A.; Fonseca, J. *Fusarium oxysporum* f. sp. *cubense* en Cuba: Biología de las poblaciones, reacción de los clones híbridos de la FHIA y biocontrol. In Proceedings of the Taller Internacional manejo convencional y alternativo de la Sigatoka negra, nematodos y otras plagas asociadas al cultivo de Musáceas, MUSALAC/INIBAP, Guayaquil, Ecuador, 11–13 August, 2003.

35. Groenewald, S.; van den Berg, N.; Marasas, W.F.O.; Viljoen, A. Biological, physiological and pathogenic variation in a genetically homogenous population of *Fusarium oxysporum* f.sp. *cubense*. *Australas. Plant Pathol.* **2006**, *35*, 401–409, doi:10.1071/AP06041.

36. Brake, V.M.; Pegg, K.G.; Irwin, J.A.G.; Chaseling, J. The influence of temperature, inoculum level and trace of *Fusarium-oxysporum* f sp *cubense* on the disease reaction of banana cv cavendish. *Aust. J. Agric. Res.* **1995**, *46*, 673–685, doi:10.1071/ar9950673.

37. Ploetz, R.C. Population biology of *Fusarium oxysporum* f sp *cubense*. In *Fusarium Wilt of Banana*; Ploetz, R.C., Ed.; APS Press: St. Paul, Minnesota, USA, 1990; Volume 1, pp. 63–76.

38. Pérez-Vicente, L.; Porras, Á. Impacto potencial del cambio climático sobre las plagas de bananos y plátanos en Cuba. *Fitosanidad* **2015**, *19*, 201–211.

39. Olivares, B.; Cortez, A.; Lobo, D.; Parra, R.; Rey, J.; Rodriguez, M. Evaluation of agricultural vulnerability to drought weather in different locations of Venezuela. *Rev. Fac. Agron.* **2017**, *34*, 103–129.

40. Nansamba, M.; Sibiya, J.; Tumuhimbise, R.; Karamura, D.; Kubiriba, J.; Karamura, E. Breeding banana (*Musa* spp.) for drought tolerance: A review. *Plant. Breed.* **2020**, *139*, 685–696.

41. Lahav, E.; Israeli, Y. Mineral Deficiencies of Banana. In *Diseases of banana, Abacá and enset*; Jones, D., Ed.; CABI Publishing: Wallingford, Oxon, UK, 2000; pp. 339–350.

42. Aguilar, E.A.; Turner, D.W.; Sivasithamparam, K. Fusarium oxysporum f. sp cubense inoculation and hypoxia alter peroxidase and phenylalanine ammonia lyase activities in nodal roots of banana cultivars (*Musa* sp.) differing in their susceptibility to Fusarium wilt. *Aust. J. Bot.* **2000**, *48*, 589–596, doi:10.1071/bt99009.

43. Summerell, B.A.; Leslie, J.F.; Liew, E.C.Y.; Laurence, M.H.; Bullock, S.; Petrovic, T.; Burgess, L.W. Fusarium species associated with plants in Australia. *Fungal Divers.* **2011**, *46*, 1–27, doi:10.1007/s13225-010-0075-8.

44. Rishbeth, J. Fusarium wilt of bananas in jamaica: II. Some aspects of host-parasite relationships. *Ann. Bot.* **1957**, *21*, 215–245. doi:10.1093/oxfordjournals.aob.a083561.

45. Bronick, C.J.; Lal, R. Soil structure and management: A review. *Geoderma* **2005**, *124*, 3–22, doi:10.1016/j.geoderma.2004.03.005.

46. Li, Z.; Deng, Z.; Chen, S.; Yang, H.; Zheng, Y.; Dai, L.; Zhang, F.; Wang, S.; Hu, S. Contrasting physical and biochemical properties of orchard soils suppressive and conducive to Fusarium wilt of banana. *Soil Use Manag.* **2018**, *34*, 154–162, doi:10.1111/sum.12390.

47. Domínguez, J.; Negrin, M.A.; Rodriguez, C.M. Aggregate water-stability, particle-size and soil solution properties in conducive and suppressive soils to Fusarium wilt of banana from Canary Islands (Spain). *Soil Biol. Biochem.* **2001**, *33*, 449–455, doi:10.1016/s0038-0717(00)00184-x.

48. Domínguez-Hernandez, J.; Negrin, M.A.; Rodriguez, C.M. Soil potassium indices and clay-sized particles affecting banana-wilt expression caused by soil fungus in banana plantation development on transported volcanic soils. *Comm. Soil Sci. Plant. Anal.* **2008**, *39*, 397–412, doi:10.1080/00103620701826522.

49. Deltour, P.; Franca, S.C.; Pereira, O.L.; Cardoso, I.; De Neve, S.; Debode, J.; Hofte, M. Disease suppressiveness to Fusarium wilt of banana in an agroforestry system: Influence of soil characteristics and plant community. *Agric. Ecosyst. Environ.* **2017**, *239*, 173–181, doi:10.1016/j.agee.2017.01.018.

50. Pegg, K.G.; Moore, N.Y.; Bentley, S. Fusarium wilt of banana in Australia: A review. *Aust. J. Agric. Res.* **1996**, *47*, 637–650, doi:10.1071/ar9960637.

51. Pittaway, P.A.; Nasir, N.; Pegg, K.G. Soil receptivity and host-pathogen dynamics in soils naturally infested with *Fusarium oxysporum* f. sp. *cubense* the cause of Panama disease in bananas. *Aust. J. Agric. Res.* **1999**, *50*, 623–628, doi:10.1071/a98152.

52. Segura, R.A.; Serrano, E.; Pocasangre, L.; Acuna, O.; Bertsch, F.; Stoorvogel, J.J.; Sandoval, J.A. Chemical and microbiological interactions between soils and roots in commercial banana plantations (*Musa* AAA, cv. Cavendish). *Sci. Hortic.* **2015**, *197*, 66–71, doi:10.1016/j.scienta.2015.10.028.

53. Chuang, T.Y. Soil suppressive of banana Fusarium wilt in Taiwan. *Plant. Prot. Bull.* **1991**, *33*, 133–141.

54. Peng, H.X.; Sivasithamparam, K.; Turner, D.W. Chlamydospore germination and Fusarium wilt of banana plantlets in suppressive and conducive soils are affected by physical and chemical factors. *Soil Biol. Biochem.* **1999**, *31*, 1363–1374, doi:10.1016/S0038-0717(99)00045-0.

55. Segura Mena, R.; Stoorvogel, J.J.; Garcia-Bastidas, F.; Salacinas-Niez, M.; Sandoval, J.A.; Kema, G. Soil management as an effective strategy for crop disease management: The case of panama disease in banana. In Proceedings of the Book of Abstract of the Wageningen Soil Conference: Soil Science in a Changing World, Wageningen, The Netherlands, 2015; p. 48.

56. Gutiérrez Jerez, F.; Borges-Pérez, A.; Trujillo Jacinto del Castillo, I. Estudio sobre el mal de panamá en las islas canarias. I. Características físicas y químicas de los suelos y su relación con la aparición de la enfermedad. *Fruits* **1983**, *38*, pp. 677–682.

57. Huber, D.; Römheld, V.; Weinmann, M. Relationship between nutrition, plant diseases and pests. In *Mineral. Nutrition of Higher Plants*, 3rd ed.; Marschner, P., Ed.; Academy Press: Stuttgart, Germany, 2012; pp. 283–298. doi:10.1016/B978-0-12-384905-2.00010-8.

58. Borges, A.F.; Bravo, J.J.; Perez, J.F.; Lopez, I. Enhanced resistance of banana plants (Dwarf Cavendish) to *Fusarium oxysporum* f.sp. *cubense* by controlled Zn nutrition under field conditions. *Banan. Newsl.* **1991**, *14*, 24–26.

59. Borges-Pérez, A.; Angulo Rodríguez, D.; Gutiérrez Jerez, F.; Trujillo Jacinto del Castillo, I. Estudio sobre el Mal de Panamá en las Islas Canarias. II. Influencia de los desequilibrios nutritivos P-Zn y K-Mg del suelo, en la alteración de los mecanismos de resistencia de la platanera (Cavendish enana) al Mal de Panamá. *Fruits* **1983**, *38*, 755–758.

60. Rasiah, V.; Armour, J.D.; Moody, P.W.; Pattison, A.B.; Lindsay, S.; Florentine, S. Characterizing and improving the deteriorating trends in soil physical quality under banana. *Aust. J. Soil Res.* **2009**, *47*, 574–584, doi:10.1071/sr08256.

61. Pattison, A.B.; Wright, C.L.; Kukulies, T.L.; Molina, A.B. Ground cover management alters development of Fusarium wilt symptoms in Ducasse bananas. *Australas. Plant. Pathol.* **2014**, *43*, 465–476, doi:10.1007/s13313-014-0296-5.

62. Huang, X.; Wen, T.; Zhang, J.; Meng, L.; Zhu, T.; Cai, Z. Toxic organic acids produced in biological soil disinfestation mainly caused the suppression of *Fusarium oxysporum* f. sp *cubense*. *Biocontrol* **2015**, *60*, 113–124, doi:10.1007/s10526-014-9623-6.

63. Zhang, H.; Mallik, A.; Zeng, R.S. Control of panama disease of banana by rotating and intercropping with chinese chive (*Allium tuberosum* Rottler): Role of plant volatiles. *J. Chem. Ecol.* **2013**, *39*, 243–252, doi:10.1007/s10886-013-0243-x.

64. Nasira, N.; Pittaway, P.A.; Pegg, K.G. Effect of organic amendments and solarization on Fusarium wilt in susceptible banana plantlets, transplanted into naturally infested soil. *Aust. J. Agric. Res.* **2003**, *54*, 251–257, doi:10.1071/ar02099.

65. Shen, Z.; Penton, C.R.; Lv, N.; Xue, C.; Yuan, X.; Ruan, Y.; Li, R.; Shen, Q. Banana fusarium wilt disease incidence is influenced by shifts of soil microbial communities under different monoculture spans. *Microb. Ecol.* **2018**, *75*, 739–750, doi:10.1007/s00248-017-1052-5.

66. Shen, Z.Z.; Ruan, Y.Z.; Xue, C.; Zhong, S.T.; Li, R.; Shen, Q.R. Soils naturally suppressive to banana fusarium wilt disease harbor unique bacterial communities. *Plant. Soil* **2015**, *393*, 21–33, doi:10.1007/s11104-015-2474-9.

67. Deng, X.; Li, Q.F.; Hou, X.W.; Wu, C.Y. Soil microbial functional diversity from different infection grades of banana fusarium wilt (*Fusarium oxysporum* f. sp *cubense*). *Appl. Mech. Mater.* **2013**, *295*, 2274–2280, doi:10.4028/www.scientific.net/AMM.295-298.2274.

68. Pirttila, A.M. Commentary: Agroforestry leads to shifts within the gammaproteobacterial microbiome of banana plants cultivated in Central America. *Front. Microbiol.* **2016**, *7*, doi:10.3389/fmicb2016.006.56.

69. Koberl, M.; Dita, M.; Martinuz, A.; Staver, C.; Berg, G. Agroforestry leads to shifts within the gammaproteobacterial microbiome of banana plants cultivated in Central America. *Front. Microbiol.* **2015**, *6*, doi:10.3389/fmicb.2015.00091.

70. Koberl, M.; Dita, M.; Martinuz, A.; Staver, C.; Berg, G. Members of Gammaproteobacteria as indicator species of healthy banana plants on Fusarium wilt-infested fields in Central America. *Sci. Rep.* **2017**, *7*, 9, doi:10.1038/srep45318.

71. Shen, Z.; Wang, B.; Lv, N.; Sun, Y.; Jiang, X.; Li, R.; Ruan, Y.; Shen, Q. Effect of the combination of bio-organic fertiliser with *Bacillus amyloliquefaciens* NJN-6 on the control of banana Fusarium wilt disease, crop production and banana rhizosphere culturable microflora. *Biocontrol. Sci. Technol.* **2015**, *25*, 716–731, doi:10.1080/09583157.2015.1010482.

72. Dinesh, B.M.; Ravichandra, N.G.; Reddy, B.M.R.; Somasekhara, Y.M. Interactions between *Radopholus similis* and *Fusarium oxysporum f. sp. cubense* causing wilt Complex on Banana. *Int. J. Adv. Res.* **2014**, *2*, 976–985.

73. Ritzinger, C.H.S.P.; Fancelli, M. Integrated management of nematodes in the banana tree culture. *Rev. Bras. Frutic.* **2006**, *28*, 331–338, doi:10.1590/s0100-29452006000200041.

74. Almeida, N.O.; Teixeira, R.A.; Carneiro, F.A.; de Oliveira, C.M.; Ribeiro, V.A.; Lobo Junior, M.; da Rocha, M.R. Occurrence and correlations of nematodes, *Fusarium oxysporum* and edaphic factors on banana plantations. *J. Phytopathol.* **2018**, *166*, 265–272, doi:10.1111/jph.12683.

75. Duyck, P.-F.; Pavoine, S.; Tixier, P.; Chabrier, C.; Queneherve, P. Host range as an axis of niche partitioning in the plant-feeding nematode community of banana agroecosystems. *Soil Biol. Biochem.* **2009**, *41*, 1139–1145, doi:10.1016/j.soilbio.2009.02.020.

76. Zhong, S.; He, Y.D.; Zeng, H.C.; Mo, Y.W.; Jin, Z.Q. Effects of banana wilt disease on soil nematode community structure and diversity. *Afr. J. Biotechnol.* **2011**, *10*, 12759–12767, doi:10.5897/ajb11.1558.

77. Li, W.M.; Dita, M.; Rouard, M.; Wu, W.; Roux, N.; Xie, J.H.; Ge, X.J. Deep RNA-seq analysis reveals key responding aspects of wild banana relative resistance to Fusarium oxysporum f. sp. cubense tropical race 4. *Funct. Integr. Genom.* **2020**, *20*, 551–562. doi:10.1007/s10142-020-00734-z.

78. Hennessy, C.; Walduck, G.; Daly, A.; Padovan, A. Weed hosts of *Fusarium oxysporum* f. sp *cubense* tropical race 4 in northern Australia. *Australas. Plant. Pathol.* **2005**, *34*, 115–117, doi:10.1071/ap04091.

79. Pérez-Vicente, L.; Dita M.A.; Martínez de la Parte, E. Technical Manual: Prevention and diagnostic of *Fusarium* Wilt (Panama disease) of banana caused by *Fusarium oxysporum* f. sp. *cubense* Tropical Race 4 (TR4). Prepared for the FAO/CARDI Regional Workshop on the Diagnosis of Fusarium Wilt (Panama disease) caused by *Fusarium oxysporum* f. sp. *cubense* Tropical Race 4: Mitigating the Threat and Preventing it Spread in the Caribbean. University of West Indies, Trinidad and Tobago, 2014.

80. Waman, A.A.; Bohra, P.; Sathyanarayana, B.N.; Chandrashekar, S.C.; Rani, R.T. Are bananas (*Musa* spp.) really safe from their aesthetic relatives? Screening potential alternative hosts of *Fusarium oxysporum* f. sp *cubense*. *J. Hortic. Sci. Biotech.* **2013**, *88*, 559–562, doi:10.1080/14620316.2013.11513006.

81. Stover, R.H. Fusarium wilt (panama disease) of bananas—Current status and control strategies. *Acta Hortic.* **1990**, *275*, 707–712, doi:10.17660/ActaHortic.1990.275.88.

82. INIA (Instituto Nacional de Investigaciones Agrícolas, VE). *Aproximación agroecológica para el nuevo modelo de producción agrícola en Venezuela*; Editorial del Instituto Nacional de Investigaciones Agrícolas: Maracay, Venezuela, 2016; pp. 12–90.

83. Wood-Sichra, U.; Joglekar, A.B.; You, L. *Spatial production allocation model (SPAM) 2005*: Technical Documentation. Harvest Choice Working Paper., Harvest Choice: St. Paul, Minnesota, USA, 2016.

84. You, L.; Wood, S.; Wood-Sichra, U.; Wu, W. Generating global crop distribution maps: From census to grid. *Agric. Syst.* **2014**, *127*, 53–60, doi:10.1016/j.agsy.2014.01.002.

85. Phillips, S.J.; Anderson, P.R.; Schapire, R.E. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **2006**, *190*, 231–259, doi:10.1016/j.ecolmodel.2005.03.026.

86. Boria, R.A.; Olson, L.E.; Goodman, S.M.; Anderson, R.P. Spatial filtering to reduce sampling bias can improve performance of ecological niche models. *Ecol. Model.* **2014**, *275*, 73–77, doi:10.1016/j.ecolmodel.2013.12.012.

87. Karger, D.N.; Conrad, O.; Böhner, J.; Kawohl, T.; Kreft, H.; Soria-Auza, R.W.; Zimmermann, N.E.; Linder, H.P.; Kessler, M. Climatologies at high resolution for the earth's land surface areas. *Sci. Data* **2017**, *4*, 170122, doi:10.1038/sdata.2017.122.

88. Hijmans, R.J.; Cameron, S.E.; Parra, J.L.; Jones, P.G.; Jarvis, A. Very high-resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **2005**, *25*, 1965–1978, doi:10.1002/joc.1276.

89. Chatterjee, S.; Hadi, A.S. *Regression Analysis by Example*; John Wiley and Sons: New Jersey, USA, 2006; pp. 80–127.

90. Naimi, B.; Araujo, M.B. Sdm: A reproducible and extensible R platform for species distribution modelling. *Ecography* **2016**, *39*, 368–375, doi:10.1111/ecog.01881.

91. Allouche, O.; Tsoar, A.; Kadmon, R. Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* **2006**, *43*, 1223–1232, doi:10.1111/j.1365-2664.2006.01214.x.

92. Dita, M.; Echegoyén Ramos, P.E.; Pérez Vicente, L.F. *Plan. de Contingencia Ante un Brote de la Raza 4 Tropical de Fusarium oxysporum f. sp. Cubense, en un País de la Región Del OIRSA*; OIRSA Editorial: San Salvador, El Salvador, 2013; p. 155.

93. Huang, B.; Xu, L.; Molina, A.B., Jr. *Preliminary evaluation of IMTP-III varieties and local cultivars against Fusarium wilt disease in South China*; International Plant Genetic Resources Institute (IPGRI): Rome, Italy, 2005; pp. 187–192.

CHAPTER II

Identification of soil properties associated with the incidence of Banana Wilt using supervised methods

Identification of soil properties associated with the incidence of Banana Wilt using supervised methods

Barlin Orlando Olivares, Andrés Vega, M. Angélica Rueda Calderón, Juan Carlos Rey, Deyanira Lobo, José A. Gómez, Juan Antonio Navas-Cortés and Blanca B. Landa

CHAPTER II. Identification of soil properties associated with the incidence of Banana Wilt using supervised methods

Abstract

For decades, a growing incidence of Banana Wilt (BW) has been detected in banana producing areas of the central zone of Venezuela. This disease is thought to be caused by a fungal-bacterial complex coupled with the influence of specific soil properties. However, until now, there is no consensus on the soil characteristics associated to a high incidence of BW. The objective of this study was to identify soil properties potentially associated with BW incidence using supervised methods. For that, soil samples associated to banana plant lots in Venezuela, showing low (n= 25) and high (n = 53) incidence of BW were collected during two consecutive years (2016 and 2017). On those soils, sixteen soil variables including percentage of sand, silt and clay, pH, electrical conductivity, organic matter, available contents of K, Na, Mg, Ca, Mn, Fe, Zn, Cu, S and P were determined. The Wilcoxon test was used to determine the occurrence of significant differences on soil variables between the two groups. In addition, Orthogonal Least Squares Discriminant Analysis (OPLS-DA) was applied to find soil variables capable of distinguishing banana lots showing high or low BW incidence. Finally, the Random Forest (RF) algorithm was used as a machine learning approach for classifying lots with high and low incidence of BW. The analysis of the Receptor Operating Characteristics (ROC) by RF revealed that the combination of Zn, Fe, Ca, K, Mn and Clay was able to accurately differentiate 85.4% of the banana lots with a sensitivity of 92.5% and a specificity of 88.0%. So far, this is the first study that identifies these six soil variables as possible new indicators associated to BW incidence in soils of lacustrine origin in Venezuela.

Keywords

Calcium, Clay, Iron, Machine Learning, Random Forest, Zinc.

1. Introduction

Bananas (*Musa* spp.) represent an important crop for Venezuela's economy, predominantly based on oil. During the last 20 years, banana production has undergone slight reductions, reaching 650,051 tons in 2019, with a cultivated area of around 41,708 ha, partially due to the shortage of agricultural inputs (fertilizers and agrochemicals), problems of access to foreign currency to meet domestic demand, and the inadequate management of agricultural policies, and the impact of drought, pests and diseases (FAO, 2020).

Banana Wilt (BW), also called "False Panama Disease" was first described in South Africa by Deacon et al. (1985). Although some *Fusarium* species have been associated with BW plants, pathogenicity tests using those strains were not successful and the etiology of BW could not be established.

Both biotic and abiotic factors (Rey et al. 2016), including some physical and chemical soil characteristics and potentially pathogenic soil microorganisms (Domínguez et al. 2008) have been referred as potential causes of BW. However, BW is a disease on unknown etiology up to date, and is mainly considered a physiological and metabolic plant disorder, which symptoms can be easily confounded with those of Fusarium wilt caused by *Fusarium oxysporum* f.sp. *cubense* (Foc), considered one of the most destructive disease of bananas worldwide (Dita et al. 2018).

In the Aragua state of Venezuela, one of the main producing areas in the country, the yields of Cavendish bananas have been decreasing since 2006 associated to the BW disease (Martínez et al. 2020; Rey et al. 2020) increasing the concerns of farmers. However, since the causal agent of this disease has not been properly identified jet, its prevention and control are difficult.

The scientific literature and the evidence in the field of Martínez et al. (2020) and Rey et al. (2020) in Venezuela suggest that there is possibly a relationship between the properties of the soil that generates a stress situation in the plant caused by specific abiotic factors, which would enhance a deleterious effect of certain microorganisms such as fungi and bacteria (fungal-bacterial complex) inducing the expression of symptoms in the plant.

Thus, according to Rey et al (2020), BW is associated with a fungal-bacterial complex and with some agroecological conditions characterized by silty soils presenting drainage problems and with nutritional imbalances, typical of lacustrine soils that are accentuated by unappropriated fertilization regimens in the last years. Additionally, the appearance and increase of the disease is associated with an average annual precipitation decrease and maximum temperatures increase (Olivares et al. 2021).

Despite technological advances, it is difficult to find studies that relates soil properties to disease incidence through the use of supervised methods such as Random Forest (RF), Orthogonal Least Squares of Discriminant Analysis (OPLS-DA) and other algorithms. RF ia a supervised learning classifier that can be used in complex situations (Hou et al. 2015; Ye et al. 2020) and has been proved to be a highly accurate classifier, but it has rarely been applied in the identification of soil properties associated with the incidence of diseases such as BW (Gomez-Selvaraj et al. 2020; Yuan et al. 2020).

In order to anticipate the potential occurrence of BW disease, it will be very valuable if certain soil characteristics can be associated to a mayor risk of occurrence of BW incidence. This study presents a study aimed to validate the hypothesis that it is possible to identify specific soil properties associated with high incidence of BW using supervised methods such as RF and OPLS-DA, whose results can be of straightforward agronomic and environmental interpretation.

2. Materials and methods

2.1 Study area

The study was carried out in a banana plantation located in the Aragua state, with 205 ha planted with Cavendish cv. Pineo Gigante (67,58º W, 10,14º N; Figure 1). These plants had at time of sampling: i) a leaf number from 16 to 18; ii) height values ranging from 3.5 to 4.5 m; and iii) a growth period from 9 to 10 months. This region is characterized by a Tropical Savanna climate (Aw). The annual mean rainfall is 980 mm (Olivares et al. 2020) and shows a marked seasonal pattern with a wet season from May to October. The mean annual temperature is 26.2 °C, whereas the mean annual relative humidity is 70.0% (Olivares, 2018). The terrain relief is mostly flat (slope ranging 0-2%). The predominant types of soil are Mollisol and Entisol, which are mostly of lacustrine origin, with medium textures, high nutrient availability, moderate to good drainage, soil pH varying from neutral to alkaline, good fertility and high soil organic matter content (Delgado et al. 2010).

2.2 Soil sampling

A systematic soil sampling was carried out in 39 banana lots sampled during January 2016 and 2017 (total banana lots sampled, n=78). Sampling was conducted following the guidelines of Lozano et al. (2004), with an approximate distance of 150 m between sampling sites. Composite soil samples were obtained in each of the banana lots, in the first horizon at a depth of 0 to 20.0±5.0 cm. The samples were subjected to soil analysis for fertility characterization purposes, in total 16 soil variables were determined including: percentage of sand, silt and clay (Gee & Or, 2002), soil reaction (pH), electrical conductivity (EC, dS/m) in suspension 1: 2 (soil: water) (Soil Survey Staff, 2017), organic matter (OM, %) (Heanes, 1984); available contents of potassium (K, mg/kg); sodium (Na, mg/kg); magnesium (Mg, mg/kg); calcium (Ca, mg/kg); manganese (Mn, mg/kg); iron (Fe, mg/kg); zinc (Zn, mg/kg), copper (Cu, mg/kg); sulfur (S, mg/kg) and phosphorus (P, mg/kg) (Mehlich, 1984).

Figure 1. Geographical location of the study area with banana lots (marked with yellow color boundaries).

2.3 Banana Wilt incidence

Before the beginning of the study, plants with typical symptoms of BW disease were located and identified in all the lots of the farm, from which tissue samples were taken from the pseudostem and roots, for the identification of pathogenic microorganisms. The isolation method in PDA culture medium and humid chamber was used, in the laboratory of the Faculty of Agronomy of the Central University of Venezuela.

For the identification of BW incidence in the field, in each banana lot each banana plant was individually inspected on a monthly basis for the presence of symptoms compatible with BW. Banana plants showing BW symptoms were eliminated in each lot and each evaluation period. Therefore, in the next monthly inspection, only the number of plants with new BW symptoms to that date were counted. The cumulative incidence of BW was determined in each of the 78 banana lots sampled during 2016 and 2017 using the guidelines by Bosman (2016). The main aim of the continuous monitoring of BW incidence was to determine the new cases of BW that occur in the total population of plants planted in each banana lot in a given plot and sampling time (equation 1). Hence, the cumulative incidence rate is calculated as the sum of the monthly incidence of BW values in percentage for each banana lot in a particular year.

$$\text{Cumulative incidence rate (\%)} = \left(\frac{n^{o}\ of\ diseased\ plants\ (BW)}{Total\ plants\ planted}\right) \text{x } 100 \qquad (1)$$

In the scientific literature there are no information describing threshold values to establish categories for BW incidence for the study area, nor in other banana areas of Venezuela. Therefore, we established a threshold value of 1.80% for BW incidence that allowed classifying the lots as having low (<1.8% of affected plants) or high (>1.8% of affected plants) incidence of BW in these lacustrine soils. with this. This high incidence value would represent a decrease of up to 13,300 kg/ha/year in those bananas lots showing an incidence of BW of 1.80% and was selected based on the information provided by J. C Rey (personal communication, September 28, 2019) and several years of experience observing yield losses associated to BW.

2.4 Data analysis

Before data analysis, we checked the data integrity. The normalization of the soil variables was carried out using the statistical package in R software version 4.0.2 (R Core Team, 2020) based on the geometric mean, and a generalized logarithmic transformation using "glog" function in R was performed to make the variables comparable among them due to differences on units to measure them (Chong et al. 2019; Yang et al 2021). Figure 2 shows the general scheme of the data analysis procedures followed in this work.

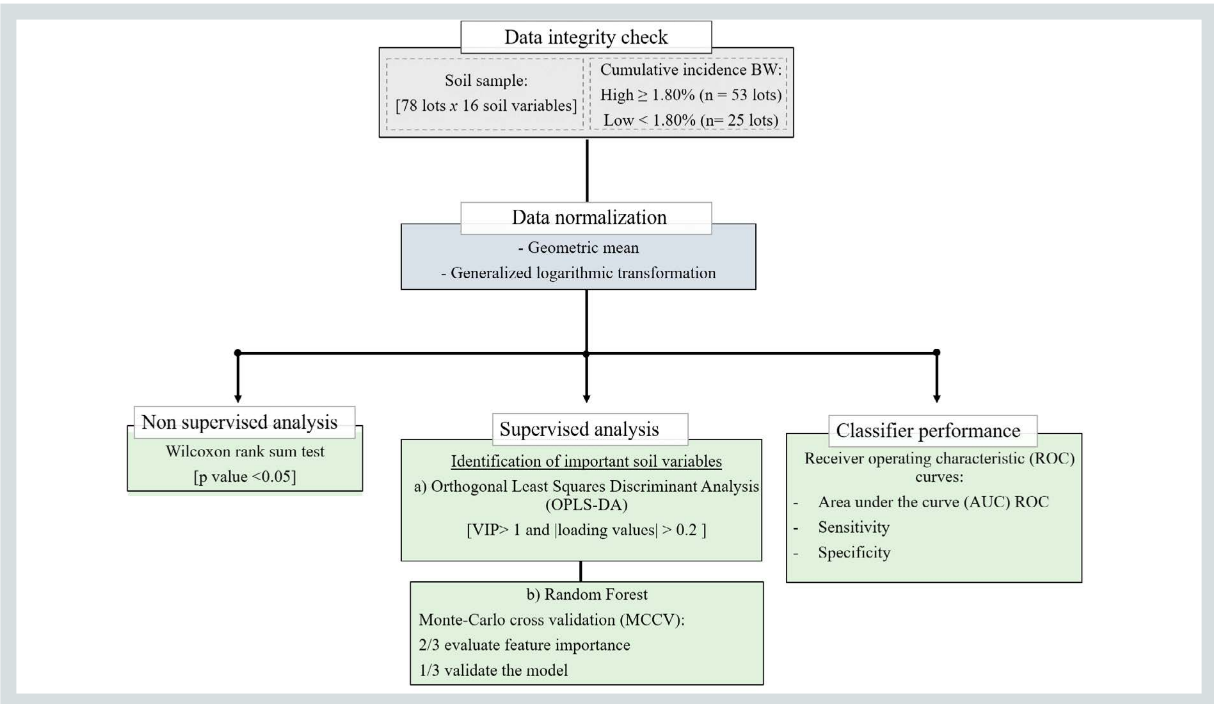


Figure 2. General scheme of the data analysis procedures (sample size, n = 78; variable size, n= 16).

2.4.1 Identification of important soil variables

For identification of relevant soil variables characterizing the incidence of BW, a Wilcoxon rank sum test was performed to find the most important features of soil variables at a threshold *p* value <0.05 (Chong et al. 2019) showing differences between the group of bananas lots with low and high incidence. Next, an Orthogonal Least Squares Discriminant Analysis (OPLS-DA) was used to reduce the number of soil variables in high-dimensional data to produce a robust and easy-to-interpret model, and to identify the main soil characteristics that drive the separation of plant lots based on BW incidence (Low or high). This multivariate statistical analysis was carried out using "ropls" R packages (Bylesjö et al. 2006).

The variable importance in projection (VIP) > 1, and corresponding |loading values| > 0.2 in the model were used to identify the variables responsible for distinguishing both BW categories (Szymańska et al. 2012). Furthermore, a permutation test with 100 permutations was employed to validate the performance of OPLS-DA model. For quality criterion we chose in the OPLS-DA model, the R^2Y (goodness of fit parameter) and Q^2 (predictive ability parameter) > 0.5 (Triba et al. 2015).

2.4.2 Classifier performance and accuracy assessment

The random forest (RF) algorithm was used as a machine learning approach for classifying lots with high and low incidence of BW (Breiman et al. 2003). RF models allow prediction of unknown samples (i.e., test dataset) after training on a known dataset (i.e. training dataset). Receiver operating characteristic (ROC) curves were generated by Monte-Carlo cross validation (MCCV) (Xu and Liang, 2001) that is a cross validation approach which creates multiple random splits of the dataset into training and validation data. In each MCCV, 2/3 of the samples were used to evaluate feature importance, and the remaining 1/3 were used to validate the model created in the first step (Xia and Wishart, 2016; Paraskevaidi et al. 2020).

To determine the predictive performance of the model, the graphs of the ROC curve were used, from which the sensitivity defined as the relationship between the number of P correctly classified and the total P observed, against "1 - specificity" (specificity is the relationship between the number of N correctly classified and the total N observed). A model will have high predictive performance if at low values of "1 − specificity" a high sensitivity is obtained, that is, a good capacity to correctly classify P with a low number of false positives. This yields a curve closer to the upper left corner (Garosi et al., 2019). The Area under the ROC curve (AUC) quantifies this relationship, so that a model is considered acceptable if AUC ≥ 0.7, excellent if AUC ≥ 0.8 and outstanding if AUC ≥ 0.9.

3. Results

3.1. Incidence of BW in experimental lots

The analysis of the identification of pathogenic microorganisms revealed the presence of bacteria (*Pectobacterium* and *Erwinia* genera) and fungi (*F. moniliforme*, *F. oxysporum*, and *F. solani*), these microorganisms have also been found by Sabadell (2003) in tissues with wilt symptoms from the Canary Islands (Spain) and recently by Rey et al. (2020) in lacustrine banana soils of Venezuela, but no vascular *Fusarium* isolates were recovered from the internal symptoms which indicated that symptoms observed in the field plot were associated to BW and not to Fusarium wilt.

Symptoms of BW disease are shown in Figure 3. Generally, yellowing begins on the lower or older leaves. The margin of each leaf turns pale green to yellow, necrotic stripes appear surrounded by a yellow margin, and the leaf eventually dies (Figure 3a). The lower leaves die and hang from the pseudostem like a skirt (Figure 3b). According to Beer et al. (2001) the base of the leaf remains green and healthy, while its distal part dies. Often 1 to 4 upper leaves remain green, but are smaller in size and their development stops. New leaf growth can occur but the bunches in this case are generally small with short and thin bananas, which generates economic losses due to the rejection of the fruit in the market.

(a) (b)

Figure 3. Symptoms of Banana Wilt disease in study area. (a) The yellow margins on the leaves and the necrotic stripes surrounded by the yellow margins on the lower or older leaves. (b) Set of dead leaves hanging from the pseudostem of a plant affected with Banana Wilt disease.

All the lots evaluated (n=78) in the study area have BW disease. The percentage of lots with a low incidence (<1.80%) of BW reached 32.05% (n = 25), while the lots with high incidence (≥1.80%) represented 67.95% (n = 53) (Figure 4). The highest incidence values were found in lot 36 with 8.47%, lot 32 (5.97%) and lot 34 (5.13%) for the year 2017, while during 2016 the maximum incidence values were registered in the lots 38 and 45 with 5.57% and 5.03%, respectively. On the other hand, lots 12, 13 and lot 17 presented low incidence values that did not exceed 1.0% in both years of evaluation (Figure 4).

3.2 Description of soil properties in experimental lots

Figure 5 shows the result of the heat map of the soil data classified into the high and low incidence groups. The heat map provides an intuitive visualization of the data used, each colored cell in the map corresponds to a concentration value in the data table, with the soil properties in the rows and the 78 banana lots in the columns.

In general, the soils with a high incidence of BW presented loam to silty loam textures, with a predominance of particles with an equivalent diameter between 2 and 50 µm. In these soils, the banana lots classified as high incidence of BW showed high values of Na, Fe and Mn slightly higher pH values (Figure 5).

On the other hand, the characteristics of the parental material of these soils produce very high levels of Ca. The limitations for the development of roots in these soils with a high incidence of BW could be associated with chemical conditions, such as the presence of high $CaCO_3$ content, the limiting ratios being Ca/Mg and Ca/K (*data not shown*). Sodium levels were high in most of the lots with high incidence of BW, which could generate toxicity problems for plants and low structural stability in soils. Likewise, low levels of Cu were observed in batches with a low incidence of BW. The metabolic nature of these elements means that their deficiency can greatly affect the development of the crop. It is important to highlight that in some batches with a high incidence of BW, high levels of P were present on the surface, possibly due to overfertilization.

In very loamy soils, with low permeability and limited drainage, and with nutrient imbalance, BW disease was more frequent. The differences in the incidence of the disease may be due to the fact that banana batches 32 to 38 have been planted for more than 10 years, while most of the batches 9 to 11 and a sector of batches 15 to 18 had were about 5-years oldat the time this study was carried out. So, it is to be expected that in these last batches the disease can advance rapidly.

Likewise, it can be inferred that, in soils with high incidence of BW, the clay content is slightly higher, whereas the K and Zn contents is slightly lower. However, a high incidence of BW occurs in those plant lots where the Ca content is higher, while the soils are more saline in depth.

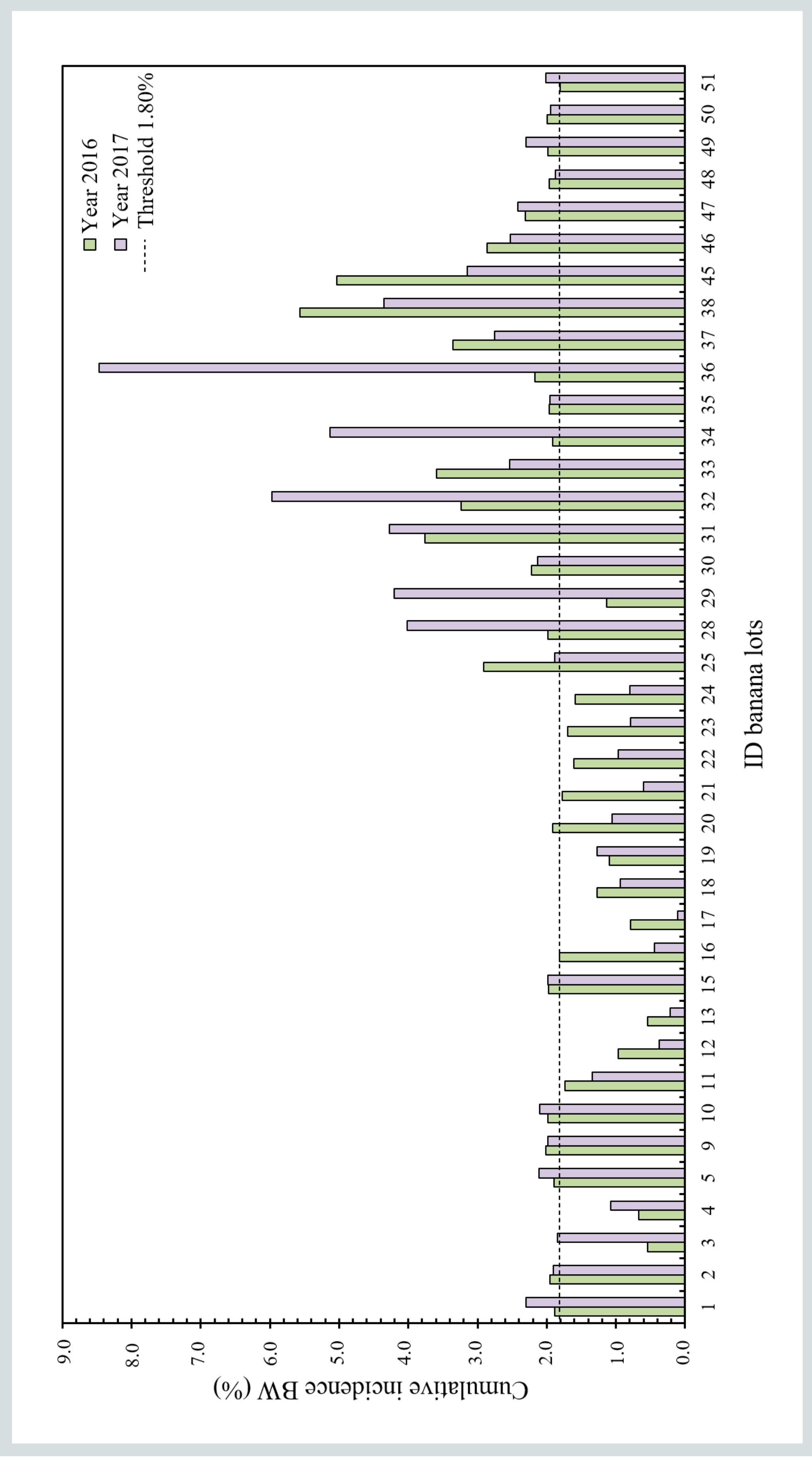


Figure 4. Cumulative incidence (%) of Banana Wilt in the study area during 2016 and 2017 (n=78).

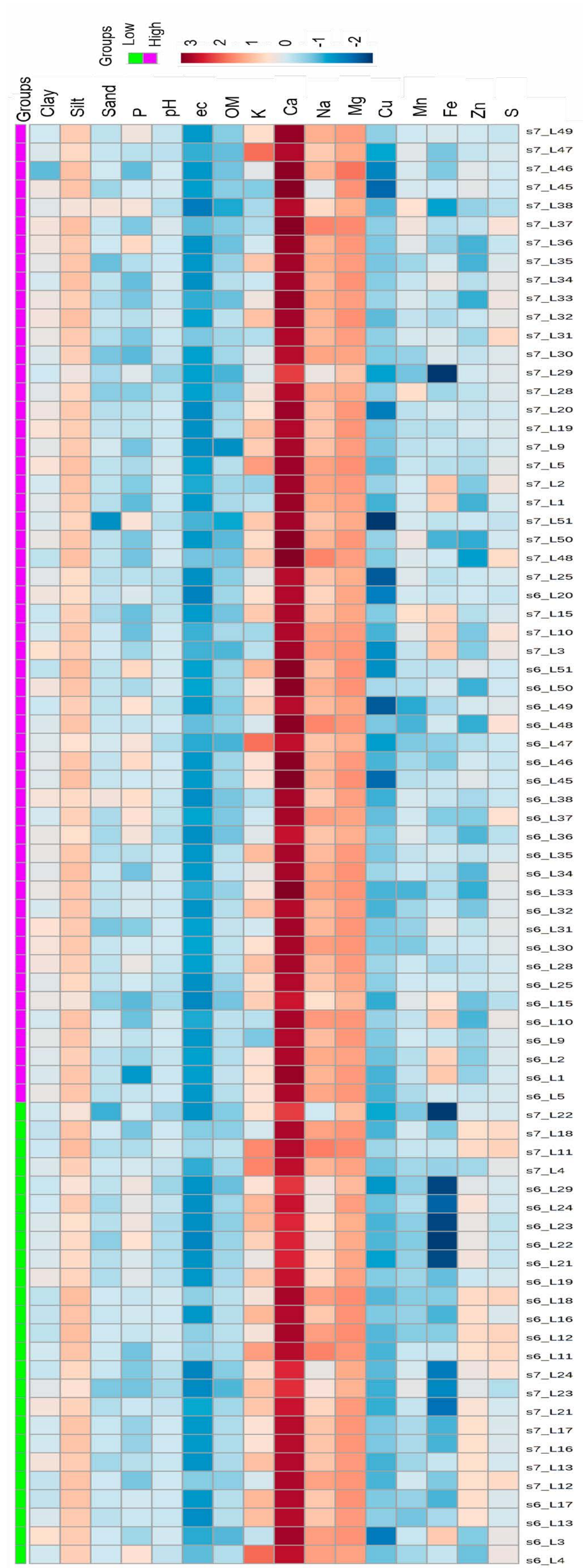


Figure 5. Heatmap generated from soil data of the banana lots with low (green) or high (purple) incidence of BW evaluated in year 2016 (s6) and year 2017 (s7), which represents increasing concentration values of the soil variables (blue to red color) for the study periods.

3.3 Wilcoxon Rank Test

For direct comparison of the levels of the soil variables, the Wilcoxon analysis was used to identify critical variables between the groups with low and high incidence of BW. The analysis revealed a total of six significant soil variables (adjusted p value <0.05) (Table 1): Zn, Ca, Fe, Clay, Mn and K. In our study, a small fraction of false positives could be accepted to substantially increase the total number of discoveries, therefore the false discovery rate (FDR) obtained is usually appropriate and useful. The FDR is the rate at which the so-called significant features are actually null. An FDR of 5% means that, among all the characteristics called significant, 5% of these are actually null. For example, if a study publishes statistically significant results for an FDR of 5.0%, the reader is confident that at most 5.0% of the results deemed significant are actually false positives. The significant and most important soil variables that were responsible for the observed differentiation between the two BW incidence groups are shown in Figure 6.

Table 1. Important variables selected by Wilcoxon Rank Test with threshold 0.05.

Variable	V	*p*-value	-log10(p)	FDR
Zn	1199	9.52E-05	80.215	1.52E-03
Ca	222	2.46E-02	56.086	1.38E-01
Fe	223	2.60E-02	55.858	1.38E-01
Clay	357	0.001	29.618	0.004
Mn	386	0.003	25.052	0.009
K	921	0.005	22.414	0.015

Note: V: The V-statistic. These values are based on the pairwise difference between the cases in two groups. FDR: The false discovery rate.

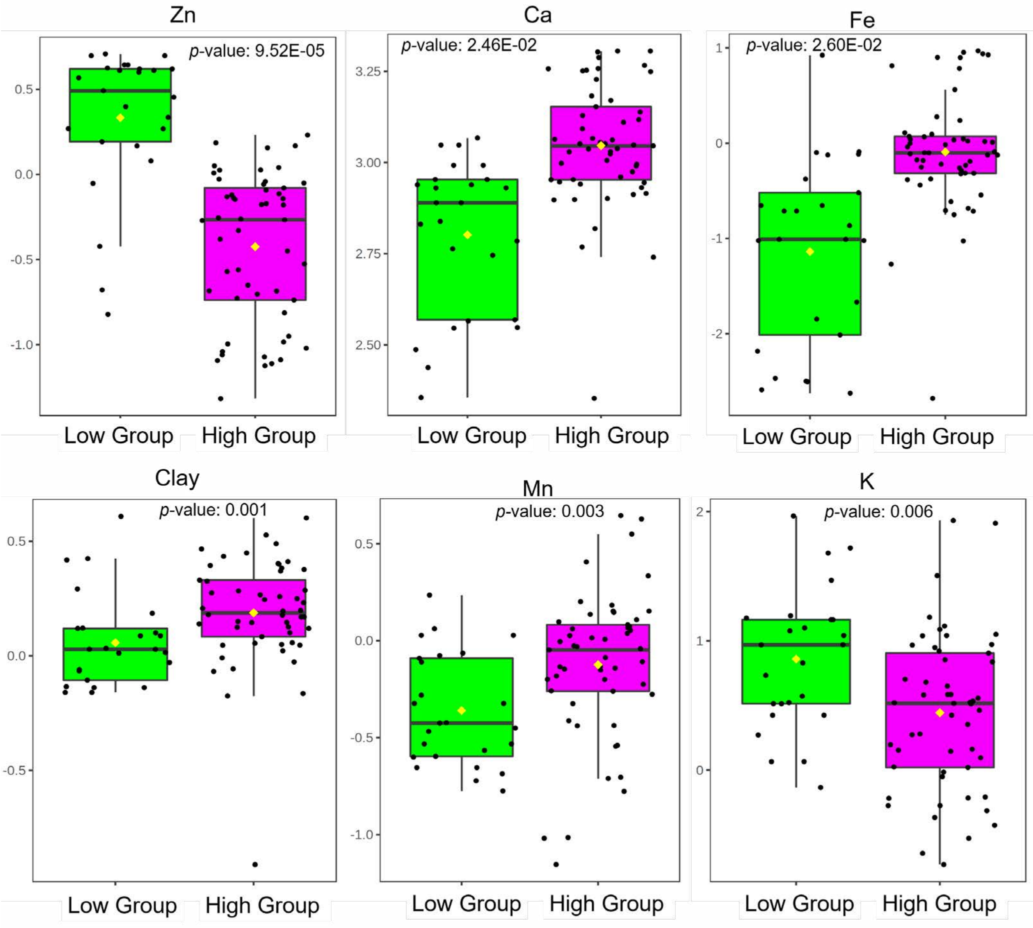


Figure 6. Box plots of levels of significant soil variables based on Wilcoxon test.

3.4 Identification of important soil variables

The results of the descriptive analysis (Table 2) indicated important differences between the characteristics of the soils of the banana lots sampled. The variable importance in the projection (VIP) values were obtained from the OPLS-DA model. The VIP was taken for selection, and those variables with VIP> 1 were considered as possible candidate variables for group discrimination (Table 2). Accordingly, the analysis revealed prominent values in three variables: K, Fe and Zn. On the other hand, as shown in Figure 7a, the OPLS-DA allowed us to analyze the information collected in the predictive component independently of the orthogonal components. That is, it allowed to separate the variability responsible for the discrimination from the noise generated by the uncorrelated variability. For this reason, the OPLS-DA was the method chosen for the selection of variables considered relevant in the discrimination of groups. Also, based on the loading values > 0.2, OPLS-DA identified six critical variables: clay, Mn, K, Ca, Fe and Zn (Figure 7b). Besides, this OPLS-DA model showed a proper fitting of the data ($R^2Y = 0.61$, p-value < 0.01), and exhibit good predictive power ($Q^2 = 0.50$, p-value < 0.01) (Figure 7c).

Table 2. Input variables used in model's construction (mean ± standard deviation, coefficient of variation, maximum and minimum) and the variable importance in the projection (VIP) values obtained from the OPLS-DA model.

Variable	Mean ± SD	Median	CV (%)	Min	Max	VIP
Clay (%)	16.10±7.86	15.00	48.78	1.00	40.00	0.36
Silt (%)	76.82±9.75	78.30	12.70	39.93	90.84	0.08
Sand (%)	7.08±4.98	6.27	70.33	0.38	38.07	0.01
pH	7.86±0.22	7.85	2.74	7.41	8.55	0.15
EC (dS/m)	0.65±0.53	0.45	82.51	0.21	2.58	0.28
OM (%)	3.39±1.52	3.50	44.86	0.25	6.33	0.32
P (mg/kg)	12.98±15.48	5.67	119.32	0.35	54.97	0.24
K (mg/kg)	110.57±229.26	35.60	207.34	1.48	1336.00	1.16
Ca (mg/kg)	9,704.51±2,968.46	8,892.00	30.59	4,936.00	16,648.00	0.68
Na (mg/kg)	152.80±97.07	132.40	62.33	10.72	472.00	0.55
Mg (mg/kg)	300.47±55.34	296.00	18.65	216.00	640.00	0.06
Cu (mg/kg)	1.41±0.87	1.60	61.20	0.03	3.20	0.15
Mn (mg/kg)	9.47±10.81	5.60	114.12	0.80	58.40	0.66
Fe (mg/kg)	13.39±22.59	5.20	168.72	0.04	78.40	2.91
Zn (mg/kg)	13.21±13.28	7.60	100.57	0.36	36.80	2.11
S (mg/kg)	17.09±11.86	11.84	69.43	6.47	48.80	0.34

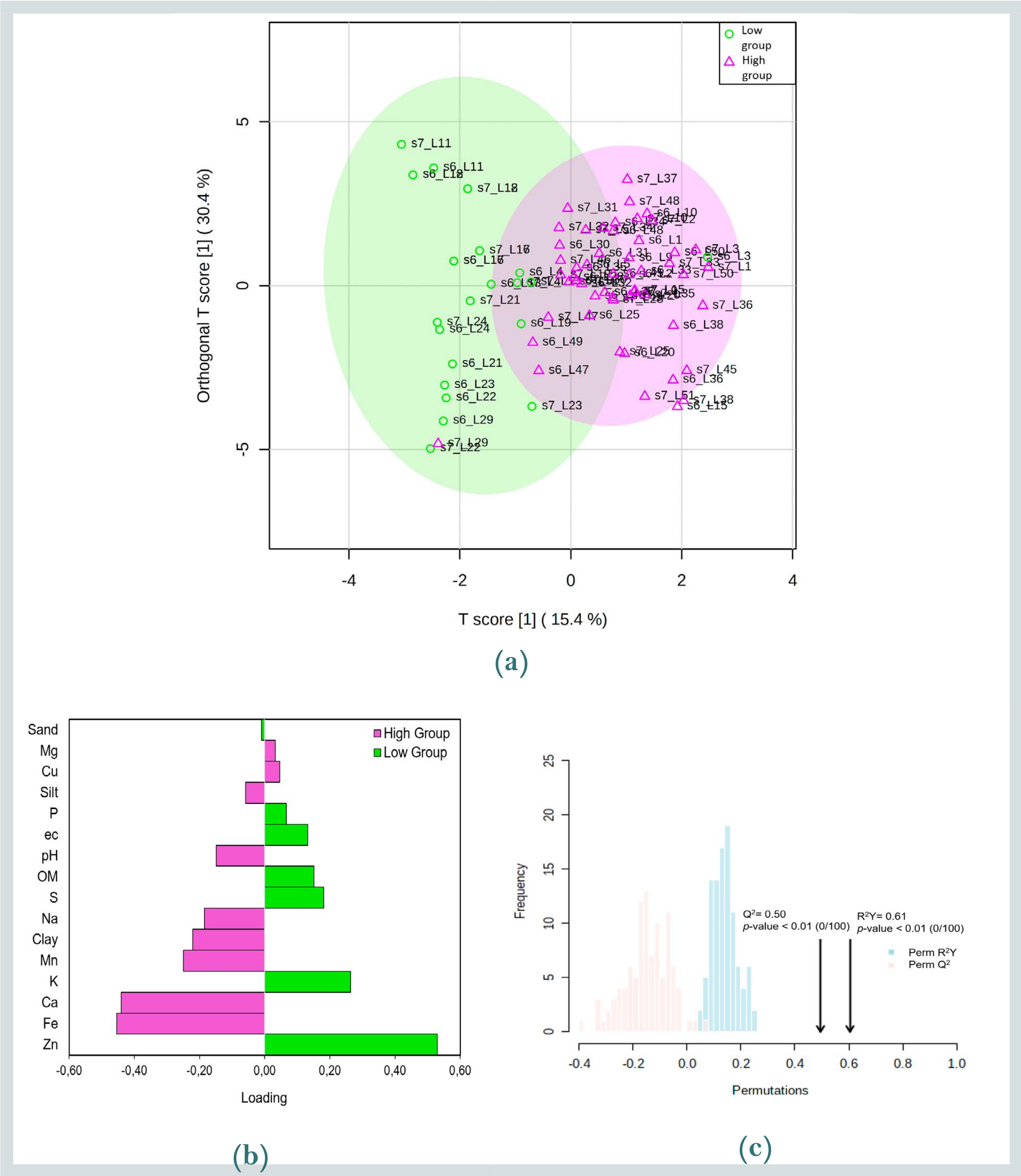


Figure 7. (a) OPLS-DA score plot of all soil variables, based separation of the groups (low incidence of BW group, n=25; high incidence of BW group, n=53). (b) Loading plot weights of each variable selected from OPLS-DA; The color indicates the class in which the variable has the maximum level of expression. (c) internal validation of the corresponding OPLS-DA model by permutation analysis (n = 100); fraction of the variance of descriptor class response (Y) (R^2Y) = 0.61 (blue bars), p-value < 0.01; fraction of the variance predicted (cross-validated) (Q^2) = 0.50 (Red bars), p-value < 0.01.

3.5 Classifier performance and accuracy assessment

Table 3 shows the measures of importance of the soil variables selected by the RF model. The results establish the frequency that an independent variable is selected greater-than/equal-to a defined importance threshold (0.5). The Mean Decrease Accuracy (MDA) allows to visualize the relative impact on the performance of the RF classifier by subtracting each specific soil variable. Figure 6 shows the classification results after RF analysis; the receiver operating characteristic (ROC) curve of the best-performing model indicated an area under the curve (AUC) of 0.93 (95% confidence interval CI: 0.85% to 0.99%) (Figure 8a). The scores plot (Figure 8b) shows the predicted class probabilities for all samples included in the analysis, indicating correct classification of 49 banana lots out of 53 with high incidence of BW and 22 banana lots out of 25 with low incidence. The misclassified lots of the low BW incidence group were: S6L3, S6L4 (year 2016) and S7L4 (year 2017), while in the high BW incidence group the lots that were not correctly classified were the S7L29 and S7L46 (year 2017) and S6L49 and S6L47 (year 2016).

Our results showed the great power of the RF classifier to correctly differentiate lots of bananas with high or low BW incidence. Furthermore, our proposed system reached 92.5% of sensitivity and 88.0% specificity in the test dataset, which implies that most banana lots with low BW incidence were correctly classified with a false negative (FN) rate of 4/53, and most of the banana lots with high BW incidence were also correctly classified with a false positive (FP) rate of 3/25 (Figure 8c).

4. Discussion

Banana Wilt is a disease of unknow etiology that has not been properly studied yet. Indeed, only in few countries the incidence of BW has been assessed, as is the case of Costa Rica were a BW incidence of 7.3% has been reported (Lichtemberg et al. 2010); in Colombia, where an incidence of 0.31% has been reported in some banana producing areas with a prevalence of 4.30% (Merchán 2002), and in Indonesia, were the average incidence of BW in 15 provinces was as high as 24% (Hermanto et al 2011).

In the case of the banana areas located in the Aragua state of Venezuela, Martínez et al. (2016), Ramírez et al. (2016) and Rey et al. (2020) reported incidences of BW ranging from 0.32% to 11.41% in different plant lots. These values coincide with that obtained in our study where the vast majority of the foci showing incidence of BW are centralized between lots 31 to 46 of the farm sampled and for both years evaluated. This could suggest that the spread of the disease may be linked to specific to soil physical-chemical characteristics combined in some degree with poor agronomic management (unappropriated fertilization) that generates a significant nutritional imbalance in the soil.

Table 3. Frequencies of variables being selected (%), Mean Decrease Accuracy and descriptive statistics of the model soil variables with Random Forest (Accuracy: 85.40%).

			Low incidence group (n=25)		High incidence group (n=53)	
Variables	**Frequencies of being selected (%)**	**Mean Decrease Accuracy**	**Mean±SD**	**Range**	**Mean±SD**	**Range**
Zn (mg/kg)	1.00	0.18	28.90±11.28	(1.60 - 36.80)	5.81±5.32	(0.36-30.40)
Fe (mg/kg)	1.00	0.05	4.43±12.64	(0.04 - 64.00)	17.62±25.0	(0.06-78.40)
Ca (mg/kg)	0.98	0.04	6,848.00±1,070.69	(4,936.0 – 8,920.0)	11051.92±2597.26	(6,472.0-16,448.0)
Clay (%)	0.89	0.01	14.52±8.52	(5.00 – 31.00)	16.85±7.49	(1.00- 40.00)
K (mg/kg)	0.65	0.01	156.04±211.79	(5.60 – 984.00)	89.12±235.94	(1.50-1,336.0)
Mn (mg/kg)	0.50	0.01	6.18 ± 6.51	(1.60 – 33.60)	11.02±12.05	(0.80-58.40)

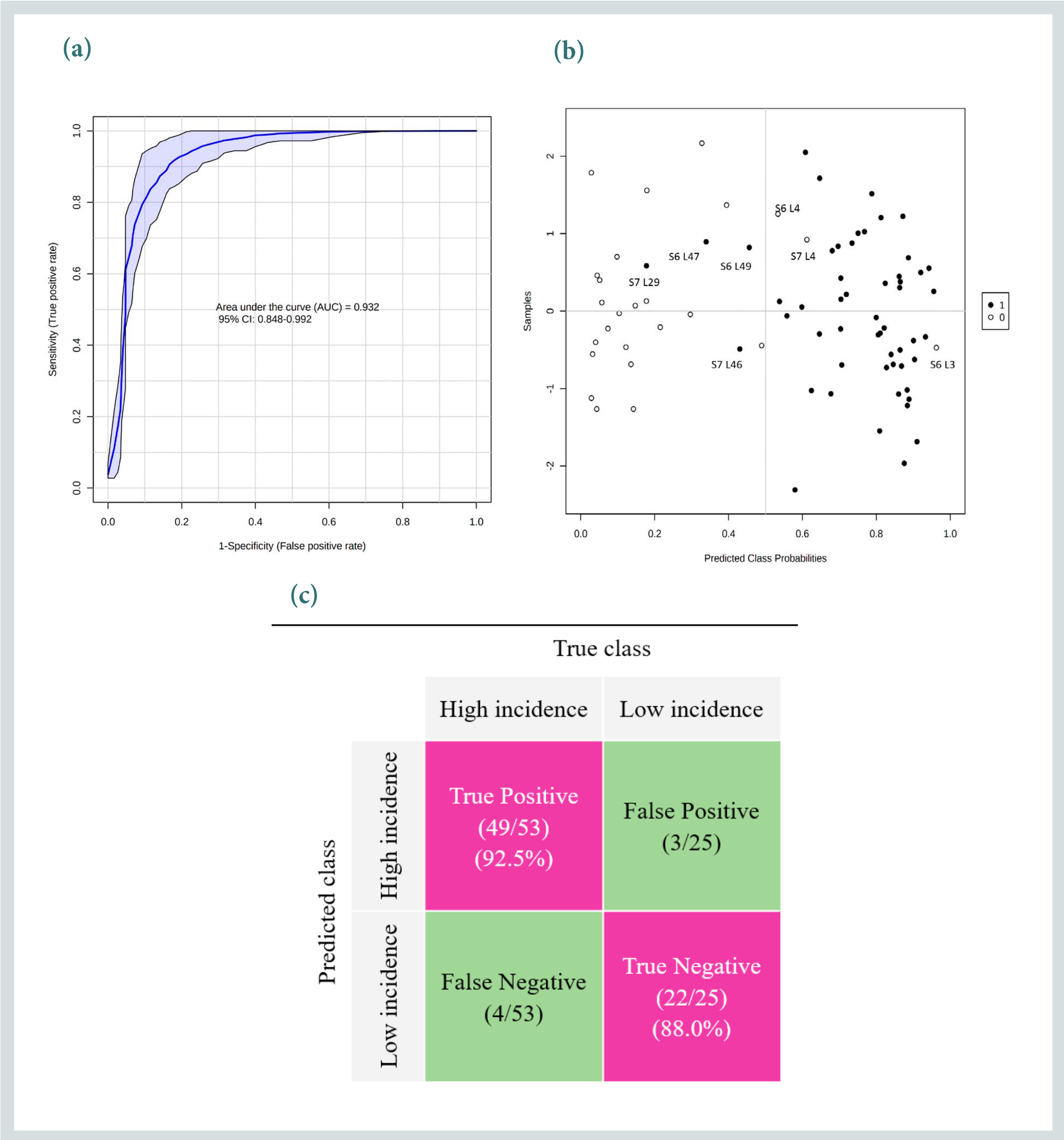


Figure 8. Classification of bananas lots according to the incidence of banana wilt (BW). (a) Receiver operating characteristic (ROC) curve after obtained by Random Forest as classification method. The values generated for the area under the curve (AUC) (0.93) along with the 95% confidence intervals (CI) (0.85-0.99) are given within the graph and accuracy: 85.4%. (b) Predicted class probabilities for each banana lots, allowing display of misclassified bananas lots (lots of high BW incidence are shown as black dots; lots of low BW are shown as white dots). Since a balanced subsampling approach is used for model training, the classification limit is always in the center (x = 0.5, the dotted line). (c) Confusion matrix showing the number of true positives (49/53), true negatives (22/25), false positives (3/25) and false negatives (4/53). Sensitivity and specificity are given in the regions highlighted in purple, being 92.5% and 88.0% respectively.

The identification of symptoms associated to BW represented the first step to understand and identify the causes of the disease in the field and distinguish the areas affected by the disease, to later perform a classification based on certain previously established statistical, economic and agronomic management parameters. In our study, we established two levels (low and high) for describing the incidence of BW based on previous experience in banana field plots in Venezuela presenting similar type of soils and agronomical practices (J. C Rey, personal communication). This threshold incidence value of 1.8% was selected as that inducing severe yield loss.

The studies by Beer et al. (2001), Sabadell (2003), Martínez et al. (2020) and Rey et al. (2020) indicated that soil factors, specifically its physical and chemical properties, are closely associated with the occurrence of BW in bananas. In the present study using a RF model we identified soil differences in six soil variables (i.e., Zn, Fe, Ca, K, Mn and Clay) between zones with different levels of BW incidence. K contents are highest (5.0 - 984 mg/kg) in the group of lots with low incidence of BW. However, Ca contents are excessively high in both groups, being more notable the concentrations in the lots of high incidences of BW (6,472- 16,448 mg/kg), due to the lacustrine origin of the soils, which can generate K and Mg deficiencies in the plants (López and Espinosa 1995).

In relation to microelements, Fe (0.1-78 mg/kg) and Mn (0.8-58 mg/kg) are present at high levels in the group of lots with high incidence of BW, while Zn is at low levels (0.3-30 mg/kg) (Table 3), these high Fe and Mn contents could be associated with the higher clay content that can generate drainage problems. Under these conditions of excess humidity, the solubility of Fe^{2+} and Mn^{2+} increases (Gutiérrez-Boem, 2016).

Regarding Zn, in the Canary Islands, Borges Pérez (1991) demonstrated that the application of Zn in the soil notably reduces the incidence and severity of BW because this type of soils shows Zn deficiency. Therefore, in our study conducted in soils of Aragua, Venezuela the low levels of this element in plant lots with high incidence of BW may be favoring the appearance of BW symptoms.

According to Domínguez (2008) the banana soils in the Canary Islands that presented severe BW problems showed a tendency to the formation of stable aggregates of clays, that with excess of irrigation favored anaerobiosis in the soil and high concentrations of Fe, which causes compaction when the soil got dry.

These relationships of clay content (1-40 %) with water and the detrimental effect of compaction in banana soils results in a decrease in productivity, plant height and a reduction in the number of children plants in the banana production unit. Additionally, according to the results of Dorel (1993) and Sabadell (2003) the most significant effect would be related to the reduction of the absorption of N, P, K, Ca and the massive absorption of Mn.

The results of our analysis established that the heavy texture in the lots with high incidence of BW favors the appearance of symptoms, agreeing with other studies that have found that this disease develops in the presence of soils with high humidity and heavy texture (Deltour et al. 2017) and poor drainage (Lahav and Israel, 2019), favoring infection by deleterious microorganisms in the lateral rootlets.

The study by Rey et al. (2020) establishes that the variables that showed the highest significant correlation with the incidence of BW were the sand and silt content, organic carbon, exchangeable Mg content and the Ca/Mg ratio. The authors found that only for the silt content and Ca/Mg in banana soils of Aragua, a positive correlation was observed with BW incidence, indicating that, in very silty soils, with low permeability and limited drainage, and with nutrient imbalance, it was more frequent to find a high incidence of BW. Likewise, they found that in addition to the mentioned variables, the C/N ratio, the K content, the nutritional relationships between the exchangeable cations (Ca, Mg and K) and the Zn content were the variables that had the greatest importance in the differentiation of field areas with different levels of BW incidence, coinciding with the results of this study.

Our results also showed that the incidence of the disease is not uniform throughout the farm; the most affected areas have very silty soils with drainage problems, certain nutrient deficiencies and nutritional imbalances, related to the natural condition of lacustrine soils and surely the lack of appropriate fertilization cycles in recent years (Rey et al., 2020).

In recent times, modern approaches such as machine learning and deep learning algorithm have been employed to identify characteristics of banana agroecosystems that could be affecting productivity and the appearance of diseases in the field. Several investigations have been carried out in the field of machine learning for the detection and diagnosis of banana diseases, using RF (Owomugisha et al. 2014, Ma et al. 2017; Sangeetha et al. 2020; Ye et al. 2020; Gomez-Selvaraj et al. 2020), artificial neural networks (Ye et al. 2020), support vector machine (SVM) (Vipinadas and Thamizharasi, 2016; Hou et al. 2015; Aruraj et al. 2019; Ye et al. 2020;), and decision trees (Owomugisha et al. 2014), among others. For that, this study was aimed to use RF model analysis's strategy to determine soil variables that can favor development of BW disease with the final aim to help avoiding the use of those soils or to promote the application of appropriate corrective fertilization treatments.

On those studies reported above the machine learning analysis approaches were used to detect Fusarium wilt and Black Sigatoka diseases using aerial images, but none of them used in situ soil data to predict occurrence of a banana disease as is the case of our study. This evidences the existence of an information gap regarding the application of these novel algorithm-based techniques using data from sampled soils. Our study is pioneer in showing results from the application of supervised methods such as OPLS-DA and the RF algorithm to identify soil variables associated to BW incidence. According to our results, it is reported for the first time that soil variables such as Zn, Fe, Ca, K, Mn and Clay content could be promising new soil indicators to classify lots of bananas prone to show higher incidence to BW disease in lacustrine soils in Venezuela.

The RF classifier achieved a significant advantage over the classifiers used in previous works (Sangeetha et al. 2020; Ye et al. 2020; Gomez-Selvaraj et al. 2020). The characteristics of the RF classifier and the way in which the most important soil variables are selected through the OPLS-DA determine the performance of the RF classifier. However, the precision of classifying banana lots with different levels of BW incidence can be affected by many different factors, such as the quality and representativeness of the information obtained, the performance of the characteristic extraction algorithm, and the subsets used with training and testing purposes, as established by studies of Gomez-Selvaraj et al. (2020) and Ye et al. (2020). The results of our study showed that RF performed well in differentiating banana lots with high or low BW incidence. More interestingly, our model provides an easy, fast and unexpensive method to accurately identify risk of incidence of BW in bananas.

5. Conclusion

In the present study by using a random forest analysis approach we identified that the risk of low or high incidence of BW in a banana farm in Venezuela could be associated to differences in six key soil variables including Zn, Fe, K, Ca, Mn and clay content. The findings may contribute to increase our understanding of the basic mechanisms and progression of BW incidence and indicated that these soil variables are potentially determining factors of a risk of high BW incidence in tropical lacustrine soils of Venezuela.

Although the RF analysis performed well in this particular study, and its performance in other banana areas in Venezuela has not been proven yet, we consider that this machine learning algorithm using soil properties as indicators has the potential to be further explored as a simple and effective tool in in banana areas with risk of develop BW.

Acknowledgements

The authors recognize the financial support for international mobility of the Ibero-American Secretary General with Fundación Carolina and Action KA107 of Erasmus+ Program from Agri-food Campus of International Excellence (ceiA3) (2020). Also, by project “Technological innovations for the management and improvement of the quality and health of banana soils in Latin America and the Caribbean” financed by FONTAGRO and coordinated by Bioversity International (before INIBAP) and project SHui (European Commission Grant Agreement number: 773903).

References

Aruraj, A., Alex, A., Subathra, M. S. P., Sairamya, N. J., George, S. T., & Ewards, S. E. V. (2019). Detection and Classification of Diseases of Banana Plant Using Local Binary Pattern and Support

Vector Machine. 2019. 2nd International Conference on Signal Processing and Communication (ICSPC). pp. 231-235. Doi: 10.1109/ICSPC46172.2019.8976582

Beer, Z, Hernández, J; & Sabadell, S. (2001). Enfermedad del falso mal de Panamá en banano. Enfermedades de musa: Hoja divulgativa N° 9 (en línea). INIBAP, Montpellier, Francia. 4 p. Available online at https://bit.ly/386KkUS. Last accessed 07 February 2021.

Borges-Pérez, A.F.; Bravo, J.J.; Perez, J.F.; & Lopez, I. (1991). Enhanced resistance of banana plants (*Dwarf Cavendish*) to *Fusarium oxysporum* f.sp. *cubense* by controlled Zn nutrition under field conditions. Banan. Newsl., *14*, 24–26.

Bosman, M. (2016). Role of the environment on the incidence of Panama disease in bananas. MSc thesis - Soil Geography and Landscape, Wageningen University. Netherlands., pp. 12-68.

Breiman, L., Last, M., & Rice, J. (2003) Random forests: Finding quasars. In: Statistical Challenges in Astronomy. Springer, New York, NY. https://doi.org/10.1007/0-387-21529-8_16

Bylesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J. K., Holmes, E., & Trygg, J. (2006). OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. Journal of Chemometrics, 20(8-10), 341-351. https://doi.org/10.1002/cem.1006

Chong, J., Wishart, D.S., & Xia, J. (2019). Using MetaboAnalyst 4.0 for comprehensive and integrative metabolomics data analysis. Current Protocols in Bioinformatics, 68(1): e86, https://doi.org/10.1002/cpbi.86.

Deacon, J.W., Herbert, J.A., Dames, J. (1985). Fals Panama disorder of bananas. In: Citrus and Subtropical Fruit Research Institute Information Bulletin. pp. 15-18.

Delgado, E., Trejos, J., Villalobos, M., Martínez, G., Lobo, D., Carlos Rey, J., Rodríguez, G., Rosales, F. E., & Pocasangre, L. E. (2010). Determination of a soil quality and health index for banana plantations in Venezuela. Interciencia, *35*(12), 927-933.

Deltour, P., Franca, S. C., Pereira, O. L., Cardoso, I., De Neve, S., Debode, J., & Hofte, M. (2017). Disease suppressiveness to Fusarium wilt of banana in an agroforestry system: Influence of soil characteristics and plant community. Agriculture Ecosystems & Environment, 239, 173-181. https://doi.org/10.1016/j.agee.2017.01.018

Dita, M., Barquero, M., Heck, D., Mizubuti, E. S. G., & Staver, C. P. (2018). Fusarium wilt of banana: Current knowledge on epidemiology and research needs toward sustainable disease management. Frontiers in Plant Science, *9*, 1468. https://doi.org/10.3389/fpls.2018.01468

Domínguez, J., Negrin, M. A., & Rodriguez, C. M. (2008). Soil potassium indices and clay-sized particles affecting banana-wilt expression caused by soil fungus in banana plantation development on transported volcanic soils. Communications in Soil Science and Plant Analysis, *39*(3-4), 397-412. https://doi.org/10.1080/00103620701826522

Dorel, M., 1993. Banana development in an andosol in Guadeloupe: effect of soil compaction. Fruits, 48, 83-88. Available online at https://n9.cl/vahwkz. Last accessed 07 February 2021.

FAO. 2020. FAOSTAT Database: Area, production and export of bananas and plantains (online). Consulted 5 mar. 2020. Available at https:// bit.ly/34z8QN8

Garosi, Y., Sheklabadi, M., Conoscenti, C., Pourghasemi, H.R., Van Oost, K., 2019. Assessing the performance of GIS- based machine learning models with different accuracy measures for determining susceptibility to gully erosion. Science of The Total Environment, 664, 1117–1132. https://doi.org/10.1016/j.scitotenv.2019.02.093

Gee, G. W., & Or, D. (2002). 2.4 Particle-Size Analysis. In *Methods of Soil Analysis: Part 4 Physical Methods* (pp. 255-293). Soil Science Society of America. https://doi.org/10.2136/sssabookser5.4.c12

Gomez-Selvaraj, M., Vergara, A., Montenegro, F., Alonso Ruiz, H., Safari, N., Raymaekers, D., Ocimati, W., Ntamwira, J., Tits, L., Omondi, A. B., & Blomme, G. (2020). Detection of banana plants and their major diseases through aerial images and machine learning methods: A case study in DR Congo and Republic of Benin. ISPRS Journal of Photogrammetry and Remote Sensing, 169, 110-124. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2020.08.025

Gutiérrez-Boem, F.H. 2016. Micronutrientes. In Carina R. Álvarez C.R and Rimski-Korsakov, H. (Eds) Manejo de la fertilidad del suelo en planteos orgánicos, pp. 106 – 113. Editorial Facultad de Agronomía, Universidad de Buenos Aires, Argentina.

Heanes, D. L. (1984). Determination of total organic-C in soils by an improved chromic acid digestion and spectrophotometric procedure. Communications in Soil Science and Plant Analysis, 15(10), 1191-1213. https://doi.org/10.1080/00103628409367551

Hermanto, C., Sutanto, A., Jumjunidang, Edison, H. S., Daniells, J. W., O'Neill, W. T., Sinohin, V. G. O., Molina, A. B., & Taylor, P. (2011). Incidence and distribution of fusarium wilt disease of banana in Indonesia. Acta Horticulturae, 897, 313-322. https://doi.org/10.17660/ActaHortic.2011.897.43

Hou, J. C., Hu, Y. H., Hou, L. X., Guo, K. Q., & Satake, T. (2015). Classification of ripening stages of bananas based on support vector machine. International Journal of Agricultural and Biological Engineering, *8*(6), 99-103. https://doi.org/10.3965/j.ijabe.20150806.1275

Lahav, E. & Israeli, Y. (2019), *Mineral deficiencies of banana*. In Diseases of banana, Abacá and Enset. (Ed. D. Jones). CAB International Publishing: Wallingford, Oxon, UK, pp. 475-486.

Lichtemberg, P.S., Pocasangre, L.E., Sikora, R.A., & Staver, C. (2010). Current status of Fusarium Wilt (*Fusarium oxysporum* f. sp. *cubense*) in bananas Gros Michel (AAA) at smallholder level of Costa Rica. XIX Reunión Internacional ACORBAT 2010. Medellín. Colombia, pp 32-44.

López, A., & Espinoza, J. (1995). Manual de nutrición y fertilización de banano. International Plant Nutrition Institute. Quito, Ecuador. 86 p.

Lozano P, Z., Bravo, C., Ovalles, F., Hernández, R. M., Moreno, B., Piñango, L., & Villanueva, J. G. (2004). Selección de un diseño de muestreo en parcelas experimentales a partir del estudio de la variabilidad espacial de los suelos. Bioagro, *16*(1), 61-72

Ma, L., Fu, T., Blaschke, T., Li, M., Tiede, D., Zhou, Z., Ma, X., & Chen, D. (2017). Evaluation of feature selection methods for object-based land cover mapping of unmanned aerial vehicle imagery using Random Forest and Support Vector Machine classifiers. ISPRS International Journal of Geo-Information, 6, 51. https://doi.org/10.3390/ijgi6020051

Martínez, G. E., Rey-Brina, J. C., Rodríguez, D., Jiménez, C., Rodríguez, Y., Rumbos, R., Pargas-Pichardo, R., & Martínez, E. (2020). Análisis de la situación fitopatológica actual de las musáceas comestibles en Venezuela. Agronomía Tropical, 70, 1-20. https://doi.org/10.5281/zenodo.4323273

Martínez, G., Rey J.C., Castro L., Micale E., López O., Pargas R. & Manzanilla E. (2016). Cavendish banana wilt in the Central Region of Venezuela associated with a fungus-bacteria complex. In Resúmenes de Ponencias, pp 34, ACORBAT Meeting, 2016. Miami USA.

Mehlich, A. (1984). Mehlich 3 soil test extractant: a modification of Mehlich 2 extractant. Communications in soil science and plant analysis, *v. 15*(no. 12), pp. 1409-1416-1984 v.1415 no.1412. https://doi.org/10.1080/00103628409367568

Merchán, V.M. (2002). Manejo Integrado de plagas del plátano y banano. In: Acorbat. Memorias XV Reunión. Asociación de Bananeros de Colombia AUGURA. Medellín Colombia. 354 -361p.

Olivares, B. O. (2018). Tropical rainfall conditions in rainfed agriculture in Carabobo, Venezuela. La Granja-Revista de Ciencias de la Vida, *27*(1), 86-102. https://doi.org/10.17163/lgr.n27.2018.07

Olivares, B., Paredes, F., Rey, J., Lobo, D., Galvis-Causil, S. (2021). The relationship between the normalized difference vegetation index, rainfall, and potential evapotranspiration in a banana plantation of Venezuela. STJSSA, 18, 58-64. http://dx.doi.org/10.20961/stjssa.v18i1.50379

Olivares, B.O., Araya-Alman, M., Acevedo-Opazo, C., Rey, J.C., Cañete-Salinas, P., Kurina, F.G., Balzarini, M., Lobo, D., Navas- Cortés, J.A., Landa, B.B & Gómez J.A. (2020). Relationship between soil properties and banana productivity in the two main cultivation areas in Venezuela. Journal of Soil Science and Plant Nutrition, 20, 2512–2524. https://doi.org/10.1007/s42729-020-00317-8

Owomugisha, G., Quinn, J., Mwebaze, E., & Lwasa, J. (2014). Automated Vision-Based Diagnosis of Banana Bacterial Wilt Disease and Black Sigatoka Disease. In Booysen, MJ. (ed). 2014. Proceedings of the 1rst International Conference on the use of Mobile Information and Communication Technology (ICT) in Africa UMICTA 2014.

Paraskevaidi, M., Cameron, S. J., Whelan, E., Bowden, S., Tzafetas, M., Mitra, A., ... & Kyrgiou, M. (2020). Laser-assisted rapid evaporative ionisation mass spectrometry (LA-REIMS) as a metabolomics platform in cervical cancer screening. EBioMedicine, 60, 103017. https://doi.org/10.1016/j.ebiom.2020.103017

R Core Team (2020) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria

Ramírez, H., Rey, J.C., & Martinez, G. (2016). Relationship between the incidence of banana wilt and properties of a lacustrine soil. ACORBAT Meeting, 2016. Miami USA. Memories.

Rey, J. C.; Martínez, G., Pizzo N., Micale, E., & Fernández, N. (2016). Susceptible areas to false Panama disease in Cavendish bananas, in the Central Region of Venezuela. In Resúmenes de Ponencias, pp 21, ACORBAT Meeting, 2016. Miami USA

Rey, J.C; Martínez-Solórzano, G., Ramírez, H., & Pargas-Pichardo, R. (2020). Relación de las condiciones agroecológicas de un lote de planicie lacustrina con la marchitez del banano Cavendish en Aragua, Venezuela. Agronomía tropical, 70, 1–12. http://doi.org/10.5281/zenodo.4346252

Sabadell, S. (2003). Etiología y epidemiologia del falso mal de Panamá en Canarias (en línea). Tesis Dr. Tenerife, España, Universidad Autónoma de Barcelona. 286 p. Consultado 12 sep. 2020. https://bit.ly/3nBp3cx

Sangeetha, T., Lavanya G., Jeyabharathi, D., Rajesh Kumar T., & Mythili K. (2020). Detection of pest and disease in banana leaf using convolution Random Forest. Test Engineering and Management, 83 (2), 3727 – 3735

Soil Survey Staff. (2017). Soil Survey Manual Hadbook 18. United States Department of Agriculture. Washington D.C.

Szymańska, E., Saccenti, E., Smilde, A.K., & Westerhuis, J.A. (2012). Double-check: Validation of diagnostic statistics for PLS-DA models in metabolomics studies. Metabolomics, 8, 3–16, https://doi.org/10.1007/s11306-011-0330-3

Triba, M.N., Le Moyec, L., Amathieu, R., Goossens, C., Bouchemal, N., Nahon, P., Rutledge, D.N., & Savarin, P. (2015). PLS/OPLS models in metabolomics: The impact of permutation of dataset rows on the K-fold cross-validation quality parameters. Molecular Biosystem, 11, 13–19, https://doi.org/10.1039/C4MB00414K

Vipinadas M.J. & Thamizharasi A. (2016). Banana leaf disease identification technique. International Journal of Advanced Engineering Research and Science, 3(6), 120-124.

Xia, J., & Wishart, D. S. (2016). Using MetaboAnalyst 3.0 for comprehensive metabolomics data analysis. Current Protocols in Bioinformatics, 55(1), 14.10.1-14.10.91. https://doi.org/10.1002/cpbi.11

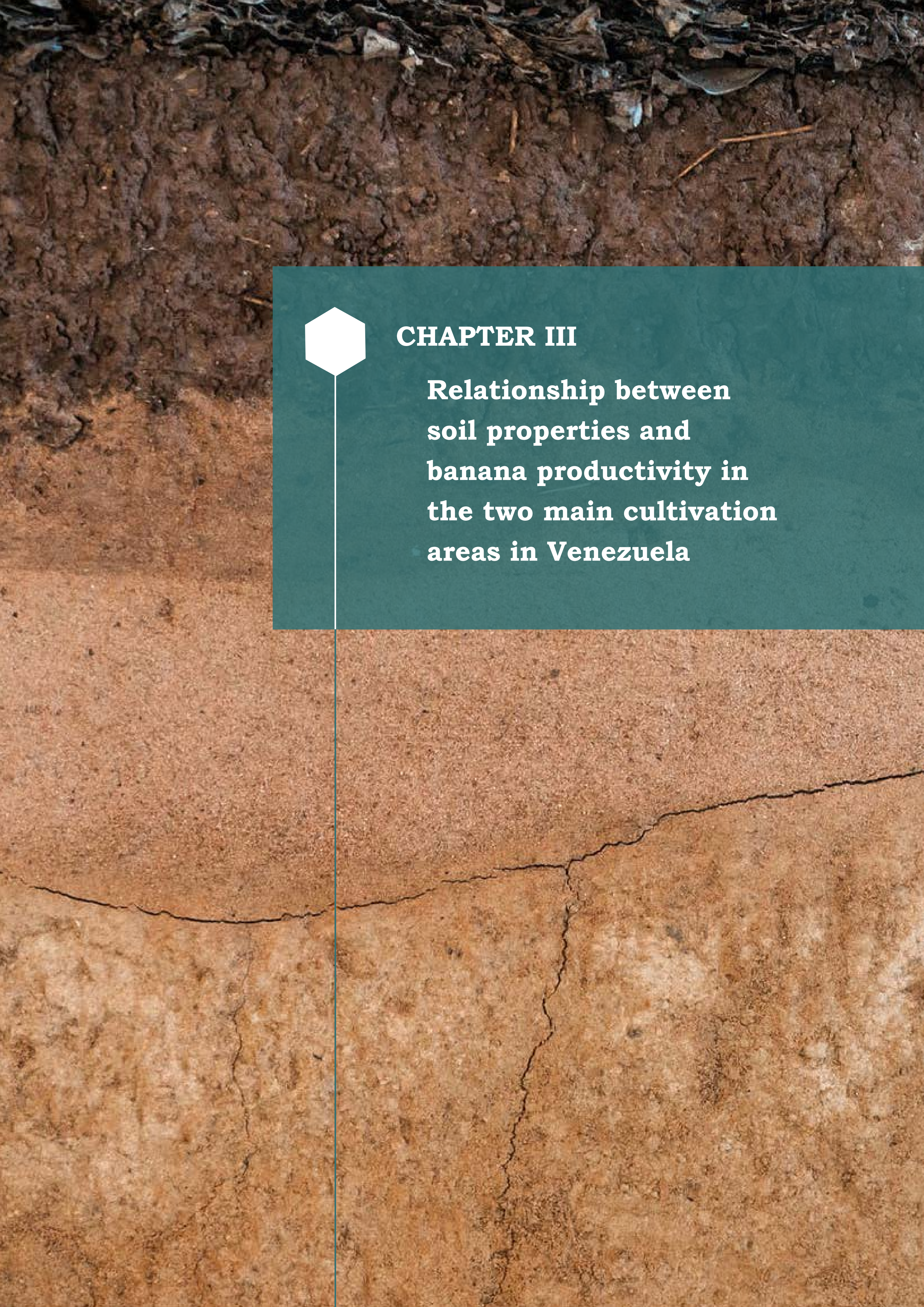
Xu Q.S., Liang Y.Z. (2001). Monte Carlo cross validation. Chemometrics and Intelligent Laboratory Systems, 56 (1):1–11. https://doi.org/10.1016/S0169-7439(00)00122-2

Yang, B.; Zhang, C.; Cheng, S.; Li, G.; Griebel, J.; Neuhaus, J. (2021). Novel Metabolic Signatures of Prostate Cancer Revealed by 1H-NMR Metabolomics of Urine. Diagnostics, 11, 149. https://doi.org/10.3390/diagnostics11020149

Ye, H. C., Huang, W. J., Huang, S. Y., Cui, B., Dong, Y. Y., Guo, A. T., Ren, Y., & Jin, Y. (2020).

Identification of banana fusarium wilt using supervised classification algorithms with UAV-based multi-spectral imagery. *International Journal of Agricultural and Biological Engineering*, *13*(3), 136-142. https://doi.org/10.25165/j.ijabe.20201303.5524

Yuan, J., Wen, T., Zhang, H., Zhao, M., Penton, C. R., Thomashow, L. S., & Shen, Q. (2020). Predicting disease occurrence with high accuracy based on soil macroecological patterns of Fusarium wilt. *The ISME Journal*, *14*(12), 2936-2950. https://doi.org/10.1038/s41396-020-0720-5



CHAPTER III

Relationship between soil properties and banana productivity in the two main cultivation areas in Venezuela

This is a literal reproduction of the article: Olivares B, Araya-Alman M, Acevedo-Opazo C, Cañete-Salinas P. et al. (2020). Relationship Between Soil Properties and Banana Productivity in the Two Main Cultivation Areas in Venezuela. J Soil Sci Plant Nutr, 20 (3): 2512-2524. https://doi.org/10.1007/s42729-020-00317-8.

Relationship between soil properties and banana productivity in the two main cultivation areas in Venezuela

Barlin Orlando Olivares, Miguel Araya-Alman, Cesar Acevedo-Opazo, Juan Carlos Rey, Paulo Cañete-Salinas, Franca Giannini Kurina, Mónica Balzarini, Deyanira Lobo, Juan A. Navas-Cortés, Blanca B. Landa and José Alfonso Gómez

CHAPTER III. Relationship between soil properties and banana productivity in the two main cultivation areas in Venezuela

Abstract

Purpose: To identify the main edaphic variables most correlated to banana productivity in Venezuela and explore the development of an empirical correlation model to predict this productivity based on soil characteristics. **Methods:** Six agricultural fields located in two of the main banana production areas of Venezuela were selected. The experimental sites were in large farms (≥50 ha) with four productivity levels in "Gran Nain" bananas, with an area of 4 ha for each of four productive levels: High - High, High - Low, Low - High and Low - Low. Sixty sampling points were used to characterize the soils under study. Additionally, a Productivity Index (PI) based on three different biometric data on plant productivity was proposed. **Results:** Through hierarchical statistical analysis, the first 16 soil variables that best explained the PI were selected. Thus, five multiple linear regression models were estimated, using the Stepwise regression method. Subsequently, a performance analysis was used to compare the prediction quality range and the error associated with the number of soil variables selected for the proposed models. The selected model included the following soil variables: Mg, penetration resistance, total microbial respiration, bulk density and omnivorous free-living nematodes. These variables explain the PI with an R^2 of 0.55, the Mean Absolute Error (MAE) of 0.8 and the Root of the Mean Squared Error (RMSE) of 1.0. **Conclusion:** The five selected variables are proposed to characterize the soil productivity index in banana and could be used in a site-specific soil management program for the banana areas of Venezuela.

Keywords

Musaceae, free-living nematodes, penetration resistance, bulk density, soil quality, total microbial respiration.

1. Introduction

Banana (*Musa* spp.) is an essential food source for a large part of the world's population. Its annual world production grew at a rate of 3.7%from 2000 to 2015, increasing from 68.2 million tons in 2000 to 117.9 million tons in 2015. Banana productivity levels vary depending on the producing country and banana variety. According to FAO (2019) in the Venezuelan territory there were 30,544 productive hectares of bananas (*Musa* AAA) in 2017, with a production of 424,649 tons destined for both the local market and for export, with an average yield of 13.91tons ha^{-1}. The 'Inter-American Network of Academies of Sciences' (IANAS, 2017) recognizes that banana production in Venezuela is stagnant due to the combination of a shortage of agricultural inputs (fertilizers and agrochemicals), difficult access to foreign exchange to meet internal demand and external supplies of machinery or equipment, and also due to the absence of sustainable agricultural practices over time.

Conventional cultivation of bananas has been noted as driver of soil degradation, resulting in loss of natural fertility due to inappropriate practices, demanding intensive use of agrochemicals aimed to maximize yield and enlarge cultivated areas (Viloria et al. 2003). In banana plantations in Venezuela, the most common practices are intensive tillage, monoculture, mechanization and chemical fertilization characterized by high doses of fertilizers (Rey et al. 2006; Delgado et al. 2010a). Previous studies in banana and other tree crops have shown a direct relationship between the loss of soil quality and the reduction in crop productivity due to inappropriate conventional production system (e.g. Segura et al. 2015; Calero et al. 2018). Within this field, the interactions among soil properties, yield and yield quality is still an active field of research with remaining open questions. For instance, Vega-Ávila et al. (2018) identified a set of variables (amylase, cellulase, and xylanase activities, bacterial abundance, microbial biomass and water content, pH and electric conductivity) to be used in future studies relating soil quality and grape quality. Sometimes these studies tried to identify the relationship between soil properties modified by specific agronomic practices, with yield and plant physiological responses, as in the case of Martinez et al. (2018) which explored the effect of organic amendments in grapevines on soil quality and grapevine root development.

Despite these studies, more complete information on optimum microbiological, physical and chemical conditions for banana roots and their relationship to banana plant development and productivity is lacking, as well as a deeper understanding of the relationship between soil properties and banana productivity in the main growing areas of Venezuela. So far, the methods used to measure the productive potential of banana soils in Venezuela have been based on general soil physical and chemical properties, geomorphology, topography, and climate (Delgado et al. 2010b; Rey et al. 2009), while rarely considering soil quality and its interactions with the rhizosphere (Rodriguez et al. 2006). This will contribute to facilitate the delimitation of zones with different soil quality on which to apply differentiated management practices within the same banana plantation, a subject which is of great interest to farmers.

This manuscript presents a study aimed to validate the hypothesis that it is possible to delineate areas of different productivity within banana fields in the two of the main banana-producing areas in Venezuela (Aragua and Trujillo States) using an index derived from easily measured soil properties with a straightforward agronomic and environmental interpretation.

2. Material and methods

2.1. Description of the study areas and banana fields

Six banana fields located in the states of Aragua and Trujillo in Venezuela were selected (Table 1). The banana plantations of the Aragua state (Fields A1, A2, A3 and A4) are in the Valencia Lake Basin, characterized by Tropical Savanna (Aw) type of climate with an average annual rainfall depth of 980 mm (Olivares et al. 2018). Rains in this area are seasonal, concentrated between May and October. The average mean annual temperature value is 26.2 °C with an average annual relative humidity of 70.0% (Olivares 2018). The terrain relief is flat (slope ranging 0-2%). The predominant soils types are Mollisol and Entisol, with moderate to good drainage, soil pH neutral to alkaline, with good fertility and medium to high organic matter (Delgado et al. 2010b).

The second study area is located at Trujillo state (Fields T1 and T2), in the Southeast Region of Maracaibo Lake, and is also characterized by a Tropical dry-winter (Aw) type of climate. The average annual precipitation is 950 mm, with two marked wet periods (dry and rainy season) (Olivares et al. 2017). Its annual average mean temperature is 27.5 °C and the annual average relative humidity 78.0%. This area of the Trujillo state is an alluvial plain with slopes below 1% and mainly Entisol soils. The drainage is moderate to poor with neutral to alkaline pH soils, with moderate fertility and average organic matter content, around 2.75% (Rodríguez et al. 2006; Rey et al. 2009).

For field sampling, areas of high and low productivity were identified on each study area based on farmers' information and preliminary productivity assessment combined with data on the potential production and vigor obtained at each banana site from available records. The final location and number of the sampling sites were determined according to the size of the farm.

To estimate banana productivity, sampling was performed according to the guidelines proposed by Rosales et al. (2008). In large fields (≥50 ha, fields A1, T1 and T2), four yield levels of "Gran Nain" bananas, determined previously based on the productivity history of all Venezuelan farms, were identified. These levels were: High - High (HH) (≥40 t/ha/year), High - Low (HL) (35-40 t/ha/year), Low - Low (LL) (≤30 t/ha/year) and Low - High (LH) (30-35 t/ha/year). In these large field areas, a plot for each of these four yield levels were identified and established according to average fruit production. Each plot having an area of 4 ha with four repetitions sampled in each plot. In the remaining small fields (<25 ha, fields A2, A3 and A4) only two plots, with a surface area of 1 ha each were identified: High- High (≥ 30 t/ha/year) and Low–Low productivity (< 30 t/ha/year), with two repetitions sampled per plot. In this way the indicators were measured and evaluated for their discriminatory quality in two productive environments at the macro level (differences among fields) and at the micro level (differences within the fields).

Table 1. Geomorphological description and taxonomic classification of banana soils sampled in the study in each state by productivity level. *Classification nomenclature of banana soil productivity according to average fruit production: HH= High - High yield (≥40 t/ha/year); HL= High - Low yield (35-40 t/ha/year); LL= Low - Low yield (≤30 t/ha/year); LH= Low - High yield (30-35 t/ha/year). † Soil Survey Staff (2010).

Fields	Productivity level*	Geographic coordinates	Geomorphological environment				Soil taxonomy†
			Natural region	State	Relief	slope (%)	
Aragua State							
A1	HH	10°08'52"N 67°35'23"W	Depression of Valencia Lake	Aragua	Lacustrine plain	0.5-1.0	Mollic Ustifluvents
	HL	10°08'55"N 67°35'13"W				0.5-1.0	Fluventic Haplustolls
	LL	10°08'44"N 67°35'09"W				0.5-1.0	Fluventic Haplustolls
	LH	10°08'52"N 67°35'03"W				1.0-2.0	Aquic Ustifluvents
A2	HH	10°12'36"N 67°24'44"W	Depression of Valencia Lake	Aragua	Terrace	0.5-1.0	Fluventic Haplustepts
	LL	10°12'55"N 67°23'43"W				2.0	Fluventic Haplustolls
A3	HH	10°11'30"N 67°31'06"W	Depression of Valencia Lake	Aragua	Alluvial Plain	0.5	Fluventic Haplustolls
	LL	10°11'38"N 67°31'06"W				0.5	Fluventic Haplustolls
A4	HH	10°11'24"N 67°31'26"W	Depression of Valencia Lake	Aragua	Alluvial Plain	0.5	Fluventic Haplustolls
	LL	10°11'19"N 67°31'27"W				0.5-1.0	Fluventic Haplustolls
Trujillo State							
T1	HH	09°29'14"N 70°57'05"W	Recent alluvial depression of Maracaibo Lake	Trujillo	Alluvial Plain	0.5-1.0	Oxyaquic Ustifluvents
	HL	09°29'55"N 70°56'50"W				0.5-1.0	Typic Ustifluvents
	LL	09°30'20"N 70°57'12"W				0.5	Typic Ustifluvents
	LH	09°30'24"N 70°57'32"W				1.0	Fluvaquentic Haplustolls
T2	HH	09°29'09"N 70°56'25"W	Recent alluvial depression of Maracaibo Lake	Trujillo	Alluvial Plain	0.5-1.0	Oxyaquic Ustifluvents
	HL	09°28'32"N 70°55'53"W				0.5-1.0	Fluaquentic Haplustolls
	LL	09°29'08"N 70°56'14"W				0.5-1.0	Typic Ustifluvents
	LH	09°29'02"N 70°56'09"W				0.5-1.0	Oxyaquic Ustifluvents

2.2. Soil sampling

In each sampled site, soil pits (60 cm wide, 60 cm long and 60 cm deep) were made to characterize the soils under study. From them a total of 1,080 disturbed and undisturbed soil samples were taken from the six evaluated sites in the first three soil horizons according to the Soil Quality and Health Diagnostic Guide of Rosales et al. (2008). In each plot of the larger fields (A1, T1 and T2) four soil pits were made at each plot and productivity class, accounting for a total of 48 soil pits (3 fields x 4 plots x 4 replications). In the case of the smaller fields (A2, A3 and A4) two soil pits were performed for each plot and productivity class, accounting for a total of 12 soil pits (3 fields x 2 plots x 2 replications).

The undisturbed samples were collected with an Uhland type sampler, obtaining three replications with an average depth of 29.6±17.6 cm, and nine samples for each soil pit. In this way, a total of 540 undisturbed samples (432 samples for the three larger fields and 108 samples for the three smaller fields) were obtained for the determination of each physical property of the soil. These physical properties were as bulk density, total porosity, saturated hydraulic conductivity, modulus of rupture according to Pla (1983), resistance to penetration using 150 kPa as a critical value for banana roots (Vaquero 2005) and moisture index (W) expressed as the gravimetric soil water content (Cassel and Nielsen 1986).

The same design was used for obtaining disturbed samples for measurement of the chemical and biological analysis of soils, totaling 540 disturbed samples (432 samples for larger fields and 108 samples for smaller fields) were taken and placed in plastic bags; sampling approximately 2 kg of soil. From these disturbed samples several chemical properties were determined according to the protocols indicated in Table 2. Biological parameters were selected according to Rosales et al. (2008) using protocols indicated in Table 2.

Table 2. Methods for the determination of the chemical, biological and root properties of the sampled banana plots

Indicator	Method	Reference
Chemical properties		
pH of soil	McLean method, soil-water ratio 1: 2.5	USDA (1995)
Organic Matter (%)	From the oxidizable organic carbon by the Walkley and Black method.	USDA (1995)
Electrical conductivity of the soil ($dS.m^{-1}$)	Conductivity meter	USDA (1995)
Nitrate ($cmol.kg^{-1}$)	Extractor solution KCl, K2SO4	Brinford et al. (1992)
Calcium carbonate ($cmol.kg^{-1}$)	Horton and Newsom method	Horton and Newsom (1953)
Aluminum, Exchangeable acidity, Calcium, Magnesium, Potassium, Sodium, phosphorus, Iron, Copper, Zinc, Manganese ($cmol.kg^{-1}$)	Extracted with Mehlich 3	Mehlich (1984)
Biological properties		
Total count populations of bacteria, fungi and actinomycetes (Colony-forming units. g^{-1} soil)	Direct counting techniques	Weaver et al. (1994)
Microbial respiration (mg $100.g^{-1}.10\ days^{-1}$)	CO2 release	Alef (1995)
Microbial biomass (mg)	Fumigation-Extraction method with chloroform	Joergensen (1995)
Carbon associated with microbial biomass (mg)	(mg C in 100 g dry soil at 105 °C; flow of C in flow-C / factor 0.45; fumigated and not fumigated)	Joergensen (1995)
Radical weight (functional, non-functional and total roots) (g)	Total extraction plant roots	Bongers (1990)
Plant-parasitic nematodes: *Radopholus similis and Helicotylenchus multicinctus* (Logarithm (N+1))	Nematode extraction from plant roots	Bongers (1990)
Total free-living nematodes (Number of nematodes in 100 g root)	Modified Baermann's Method	Bongers (1990)
Number of genera of free-living nematodes		

2.3. Biometric productivity data

In each of the evaluated banana plots (60 plots in total), four representative plots of 1,000 m^2 were selected, to estimate biometric data on plant productivity: circumference of the pseudo-stem of the mother plant at one meter in height (CircM, cm), number of hands per bunch (Nhand, n), and the height of the succession plant (AltH, cm) according to the methodology proposed by Rosales et al. (2008). A total of 20 plants were selected per plot, with their bunch close to harvesting (age of bunch fluctuated between 10-13 weeks after its appearance). To analyze the biometric productivity, the results obtained for the selected agronomic variables in each of the study plots were averaged.

2.4 Statistical analyses

2.4.1 Productivity index

A Productivity Index (PI) was developed based on the biometric production data evaluated in the plots (i.e., CircM, Nhand and AltH). The developed PI was generated from a principal component analysis (PCA) in which the variables best represented in the first axis (Principal Component 1) were retained. PCA is a technique for the construction of multivariate indices used in different areas of knowledge, e.g., Primpas et al. (2010) and Araya-Alman et al. (2019). The objective of the PCA analysis is to reduce the dimensionality of the data set. In addition, the linear combination of the three biometric variables evaluated was used as a synthesis variable for the analysis using XLSTAT software (Addinsoft 2010) with the coefficients of this linear combination being the loadings of the variables in PC1.

2.4.2 Relative importance of predictor variables

Random forest is a technique that allows selecting the most important predictors; however, the most usual procedure in forward stepwise estimation is the analysis of the correlation matrix and multicollinearity testing (for example, analysis of tolerance or variance inflation factor, VIF). This random forest technique allows reducing multicollinearity problems between variables. As a matter of fact, a high degree of multicollinearity can have substantive effects on the regression coefficients estimation and their statistical significance.

A random forest is a collection of predictor trees h ($\mathbf{x}$; θk), $k= 1 \ldots K$ where $\mathbf{x}$ represents the observed input vector (covariate) of length p with associated random vector $\mathbf{x}$ and the θ_k are independent and identically distributed (*iid*) random vectors. This type of model optimizes the variables selected in the analysis according to the percentage of contribution of the variable response (Breiman et al. 2003).

The relative importance of the soil variables used in the study through random forest was obtained using the equations developed by Friedman (2001) and implemented in the *gbm* library in the R software version 3.6.0 (R Core Team 2015). The measures are based on the number of times a variable is selected for splitting, weighted by the squared improvement to the model as a result of each split, and averaged over all trees (Friedman and Meulman 2003). This relative importance (or contribution) of each variable is scaled so that the sum adds up to 100, with higher numbers indicating a greater influence on the response; in this case, a threshold value 3.0% was used (Elith et al. 2008).

2.4.3 Multiple Linear Regression approach (MLR)

The soil variables selected through Random Forest were subjected to an automatic selection of predictive variables, in order to determine which independent variables are best related to banana productivity. For this reason, a multiple linear regression model was performed, using the forward stepwise regression method, as proposed by Miller (2002). The model used for multiple linear regression is presented in equation (1) using *lm* function in Rpackagelme4.

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_k X_i + \varepsilon \qquad (1)$$

where Y_i is the banana productivity index (PI), β_0 is the intercept, β_1 to β_k are the model regression parameters, X_1 to X_i are the soil variables and ε represents the random error.

2.4.4 Checking the MLR solution by bootstrapping.

Bootstrapping was applied to the regression models, to check the stability of the method, calculating the confidence intervals of the regression coefficient using the package bootstrap version 2019.6 in R. The model was considered stable when 95% of the values of the sampled coefficients are within the range (lower - upper) of the estimated confidence intervals. This would demonstrate that the method is stable and generalizable, saving the tedious verification of all the basic assumptions of the regression (normality, homogeneity of variances, etc.) and having the certainty that overfitting has been avoided without having to resort to the complex cross-validation (Efron and Tibshirani 1993).

2.4.5 Performance analysis of the soil variables selected

A performance analysis was used to compare the quality range of prediction and the associated error in the number of soil variables selected. In the first case, a specific number of variables were used to estimate the PI that was related to different Multiple Linear Regression models. In the second case, different error rates were calculated to obtain the optimal PI model. Then, the performance analysis considered three statistical measurements to assess the accuracy of the models finally proposed: i) the Root of the Mean Squared Error (RMSE), ii) the Mean Absolute Error (MAE) and iii) the coefficient of determination (R^2) (Acevedo-Opazo et al. 2013).

3. Results

3.1. Productivity index

Figure 1 shows the results of the PCA for the first two principal components (PC 1 and PC 2) that account for most of the variance in biometric data (96.83%) on plant productivity. PC1 and PC2 explain 74.86% and 21.97% of the total data variability, respectively. In our study, only PC1 was selected to propose the PI, since this PC better explains the behavior of the variables: Circumference pseudo-stem of the mother plant at one meter in height and number of hands per bunch, as well as clusters of the different productivity classes (Fig. 1a). PC2 would be mainly explained by the height of the succession plant (Fig. 1a). Therefore, the PI will depend on the magnitude (weight and sign) of the vectors under study. It is also apparent in Figure 1 how the different classes tend to overlap along PC2, especially for the larger spots, although there is a clear separation between the smaller and larger spots along this axis.

In general, the high productivity sites are in the right side of Fig, 1b with higher scores on PC1, whose IP ranges from 0.06 to 2.70 (Fig. 1b). Sites located on the left panel, with lower PC1 values, would be the plots with the lowest productivity whose PI ranges from -4.21 to 0.11. The T2, A1 and A4 sites are those with the highest levels of banana productivity; these results contrast with the productivity observed in the A2 and A3 sites, which have the lowest productivity indices in the study (Figure 1b).

3.2. Relative importance of predictor variables

In total, 16 soil variables that accounted for the highest proportion of the variance of PI were selected in the Random Forest analysis (Table 3) including: Penetration resistance (PR), Copper (Cu), Magnesium (Mg) and soil pH from both soil horizons, while the variables omnivorous free-living nematodes (NVLomc), calcium carbonate ($CaCO_3$), silt, total microbial respiration (TMRc), moisture index (W), soil bulk density (BD), percentage of functional roots (% RF) were selected from data on soil horizon 1 only, and Potassium (K) from soil horizon 2.

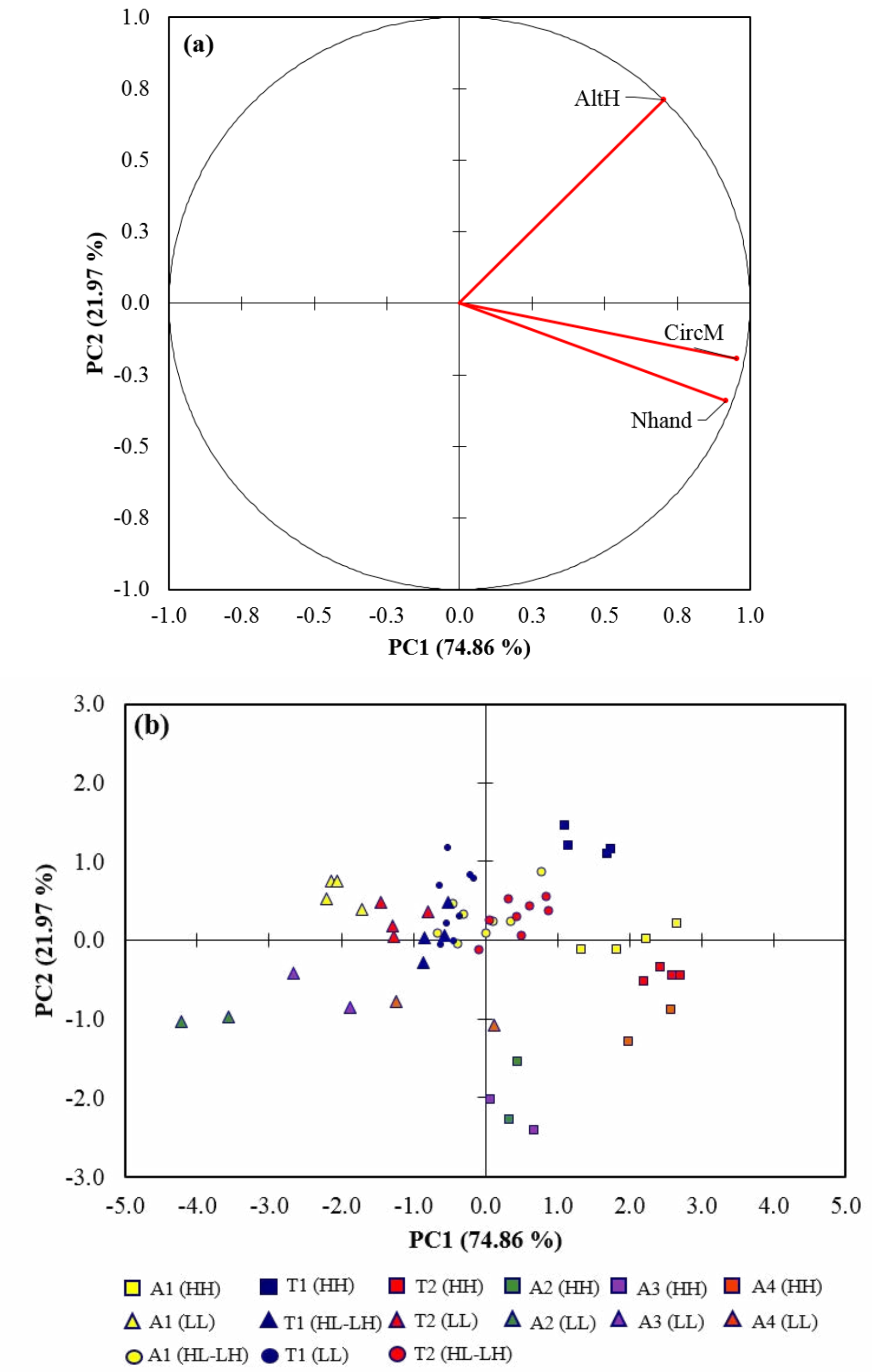


Figure 1 (a) Projection of the circumference of the pseudo-stem of the mother plant (CircM), number of hands per bunch (Nhand) and the height of the succession plant (AltH) in principal components, indicating the relative weight of these variables into PC1 and PC2 through principal component analysis. Values in brackets in the axis legend indicates the proportion of variance explained by PC1 and PC2 through principal component analysis. (b) Bivariate representation of the results of the six banana sites (A1, A2, A3, A4, T1 and T2, with banana productivity according to the average fruit production of each plot evaluated (HH= High - High yield (≥40 t/ha/year); HL= High - Low yield (35-40 t/ha/year); LL= Low - Low yield (≤30 t/ha/year); LH= Low - High yield (30-35 t/ha/year)) within PC1 and PC2.

Table 3. Summary of the relative contributions (%), mean and range of the main predictor variables using Random Forest Analysis with a tree complexity of 5 and a learning rate of 0.005. Omnivorous free-living nematodes (NVLomc); soil pH, Magnesium (Mg), Penetration resistance (PR), Copper (Cu), total microbial respiration (TMRc), soil bulk density (BD), Potassium (K), moisture index (W), percentage of functional roots (% RF). *Data from soil horizon 1; **Data from soil horizon 2

Variable	Relative contribution (%)	Mean	Range
PR (kPa)*	12.31	0.22	(0.04-0.53)
Cu ($cmol.kg^{-1}$)*	10.63	4.57	(0.00-12.60)
NVLomc (Log10)*	8.42	18.92	(0.00-140.00)
Mg ($cmol.kg^{-1}$)**	7.70	3.22	(0.44-8.63)
pH**	6.27	8.14	(6.37-8.98)
BD ($Mg.m^{-3}$)*	6.13	1.30	(0.67-1.65)
K ($cmol.kg^{-1}$)**	5.71	0.73	(0.08-2.48)
pH*	5.22	8.03	(6.33-8.77)
$CaCO_3$($cmol.kg^{-1}$)*	4.86	8.56	(0.38-24.94)
Silt (%)*	4.56	53.85	(24.63-70.20)
TMRc (mg $100.g^{-1}.10\ days^{-1}$)*	4.53	3.48	(0.12-8.50)
Cu ($cmol.kg^{-1}$)**	4.38	3.72	(0.00-12.10)
PR (kPa)**	3.77	0.24	(0.00-0.62)
Mg ($cmol.kg^{-1}$)*	3.73	3.79	(0.32-8.55)
W*	3.58	0.30	(0.09-0.69)
%RF*	3.57	88.63	(70.84-99.26)

3.3 Multiple Linear Regression

A functional relationship was established between the PI and the statistically significant variables ($P <$ 0.05). A total of nine variables among those identified in the Random Forest Analysis were included in the model that showed the highest goodness of fit ($R^2 = 0.69$), indicating that there is a good degree of association between the PI and the predictor variables in the model (Equation 2). The variables in equation 2 are described in Table 4. For this model the coefficients are robust, being within the bootstrap confidence interval, so the stability of the model is assured, and the results can be generalized.

$$\mathrm{PI} = 10.03 - 0.02(\mathrm{NVLomc}^{*}) - 0.91(\mathrm{pH}^{*}) + 0.38(\mathrm{Mg}^{*}) - 0.25(\mathrm{Mg}^{**}) + 0.17(\mathrm{Cu}^{**}) - 7.63(\mathrm{PR}^{*}) - 0.26(\mathrm{RMTc}^{*}) - 3.46(\mathrm{BD}^{*}) + 0.05\,(\%\mathrm{RF}^{*}) \quad (2)$$

Table 4 Coefficients (estimate, standard error, *T* statistic, *P*-value) bias and 95% percentile confidence intervals for bootstrap of the model parameters of the population mean based on the 2,000 bootstrap samples. Omnivorous free-living nematodes (NVLomc); soil pH, Magnesium (Mg), Penetration resistance (PR), Copper (Cu), total microbial respiration (TMRc), soil bulk density (BD), percentage of functional roots (% RF). *Data from soil horizon 1; **Data from soil horizon 2

Parameter	Estimate	Standard error	*T* statistic	*P*-value	Bootstrap for coefficients		
					Bias	95% confidence intervals for bootstrap	
						lower	upper
Constant	10.03	2.94	3.41	0.001	-0.254	4.100	15.256
NVLomc*	-0.02	0.01	-3.66	0.001	-0.001	-0.030	-0.008
pH*	-0.91	0.32	-2.88	0.006	-0.005	-1.658	-0.305
Mg*	0.38	0.12	3.21	0.002	0.009	0.171	0.654
Mg**	-0.25	0.12	-2.15	0.037	-0.006	-0.551	-0.036
Cu**	-0.17	0.07	-2.47	0.017	0.008	-0.294	0.020
PR*	-7.63	1.69	-4.52	0.000	-0.070	-10.745	-4.805
TMRc*	-0.26	0.09	-2.84	0.007	0.016	-0.474	-0.005
BD*	-3.46	0.76	-4.56	0.000	0.091	-5.075	-1.543
%RF*	0.05	0.02	2.41	0.019	0.001	0.013	0.100

3.4 Performance analysis of the selected soil variables

Performance analysis was carried out for five models developed from the model described in Table 4 reducing the number of variables. For the first model, nine variables were used according to equation 2. From the more complex model (model 1 in Table 5), at each iteration the variable whose level of significance was lower was gradually discarded from the model. Table 5 shows the differences among regression models in terms of R^2, MAE and RMSE. The results of this analysis showed that as the number of variables in the model decreases, R^2 decreases to the point where the model performed poorly (model 5, which only explains about 25% of the total variability). In general, PI models 1 to 4 present an acceptable performance. The best performed model explained 70% of the variation in the measured PI. Models 1-3 had the lowest values of MAE (generally < 0.8) and RMSE (generally ≤ 1.0 PI) (Table 5). The fourth model showed slightly higher MAE and RMSE values than the previous models, while the fifth model had the highest MAE and RMSE values (1.05 and 1.30, respectively).

Table 5 Performance analysis represented by the coefficient of determination, the mean absolute error and the root of the mean squared error applied to the five models evaluated. Magnesium (Mg), Copper (Cu), percentage of functional roots (% RF), soil pH, soil bulk density (BD), penetration resistance (PR), total microbial respiration (TMRc), omnivorous free-living nematodes (NVLomc). The Root of the Mean Squared Error (RMSE), the Mean Absolute Error (MAE) and the coefficient of determination (R^2). * Data from soil horizon 1; ** Data from soil horizon 2.

Model	Explanatory variables (Number)	R^2	MAE	RMSE	Variables excluded
1	Mg**, PR*, TMRc*, BD*, Mg*, Cu**, %RF*, pH*, NVLomc* (9)	0.69	0.63	0.82	-
2	PR*, TMRc*, BD*, Mg*, pH*, NVLomc* (6)	0.59	0.79	0.96	Mg**, Cu**, %RF*
3	PR*, TMRc*, BD*, Mg*, NVLomc* (5)	0.55	0.80	1.00	Mg**, Cu**, %RF*, pH*
4	PR*, TMRc*, BD*, NVLomc* (4)	0.50	0.85	1.05	Mg**, Cu**, %RF*, pH*, Mg*
5	PR*, TMRc*, NVLomc* (3)	0.23	1.05	1.30	Mg**, Cu**, %RF*, pH*, Mg*, BD*

To visualize the predictive ability of the proposed models, the values of the observed and predicted PI of all plots were plotted (Fig. 2). In Figure 2a, model 1 (in Table 5) presented the highest R^2 value. It cannot be stated that this is the best model, since although it has an R^2 greater than the other models, there may be an over-adjustment or simply not being parsimonious. For models 1 and 3 (Fig. 2 a, b), the data was clustered near the 1:1 line, a similar dispersion was observed between the data observed and predicted. It is also apparent how model 5 presented a poor predictive capability with the lowest R^2 value (Fig. 2c) and far from the 1:1 line.

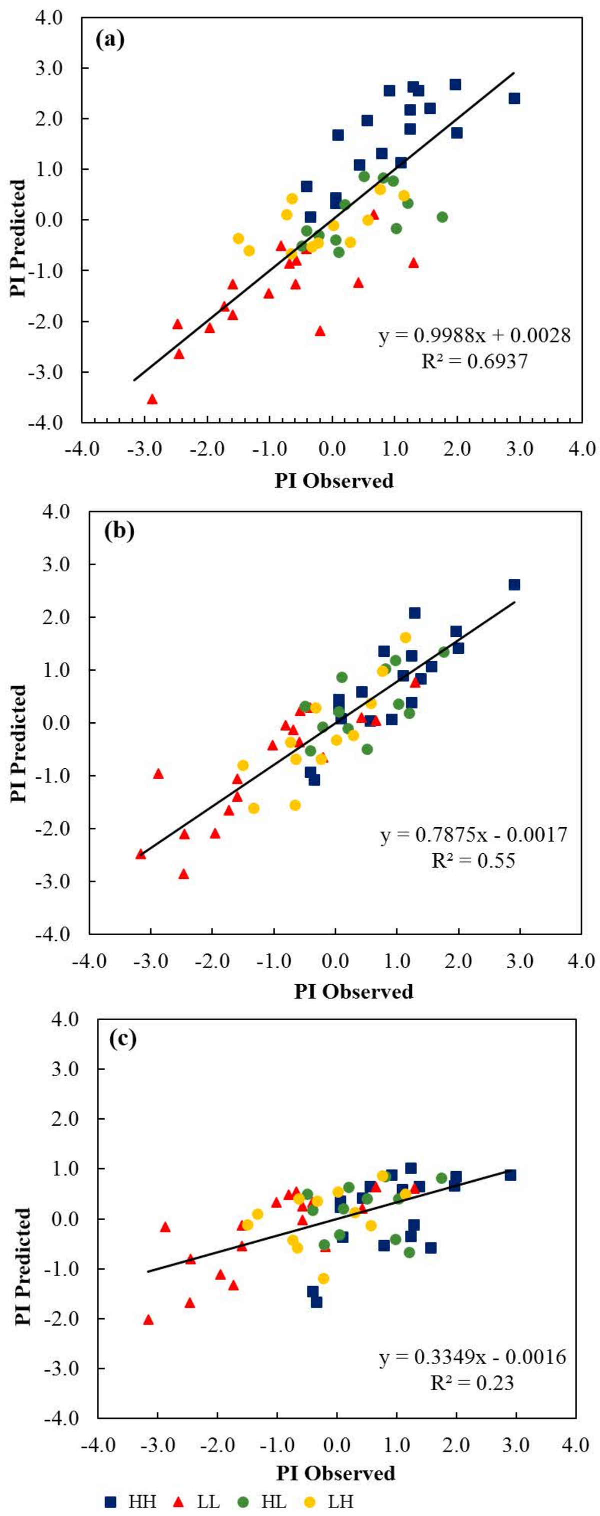


Figure 2. Results of productivity index (PI) observed in the six banana sites with banana productivity according to the average productivity index of each plot evaluated (HH= High - High yield (≥40 t/ha/year); HL= High - Low yield (35-40 t/ha/year); LL= Low - Low yield (≤30 t/ha/year); LH= Low - High yield (30-35 t/ha/year)) versus the predicted PI of the best selected models in the study: (**a**) model 1, (**b**) model 3 and (**c**) model 5. The black line indicates the correlation between that observed and predicted.

Results suggested that the PI may have a strong correlation with BD given the large reduction in R^2 when this variable was not included in model 5. It is important to note that the error (MAE: 1.05 and RMSE: 1.30) associated with the model 5 (without the BD variable) was the highest, while the models that considered the BD had a more acceptable, and similar, performance (Table 4). According to all previous results, the selected model would be model 3 (equation 3), which is composed of five variables that explain the PI with an R^2 of 0.55, a MAE of 0.8 and a RMSE of 1.0.

$$PI= 9.03+ 0.19(Mg^*)-8.78(PR^*)-0.40(TMRc^*)-4.70(BD^*)+0.02(NVLomc^*) \quad (3)$$

Where Mg= Magnesium, PR= Penetration resistance, TMRc= total microbial respiration, BD=soil bulk density, NVLomc= omnivorous free-living nematodes. *All soil variables were obtained from soil horizon 1.

The purpose of this analysis was to look for parsimony using fewer variables (losing some precision) but that would best represent the PI. Although the R^2 of 0.55 is not completely satisfactory, it is an effective indicator at the field level, where you can usually see coefficients of variation close to 40% (regardless of which variable is being measured). Using five variables instead of nine (model 1) will allow the banana producers to simplify the number of soil analysis to perform, with the consequent saving of money and time, and the making more feasible for field implementation.

3.5 The MLR solution by bootstrapping

Table 6 shows the bootstrapping performed on model 3, which indicates that the coefficients have a small bias (that is, very little difference between the mean value of the coefficients in the 2000 resamples and the original coefficients). The bootstrap coefficients are still significant, and 95% of the values of the resampled coefficients are within the range (lower - upper) of the estimated confidence intervals. This shows that the model is stable and might be generalizable, so the results can be generalized for similar situations. These results support the decision to select model 3 instead of model 1 composed of 9 variables, precisely due to the parsimony criterion.

Table 6. Bootstrap coefficients (B), bias, standard error, significance values and confidence interval (95%) for regression coefficients of model 3 for the population mean based on the 2,000 bootstrap samples. Omnivorous free-living nematodes (NVLomc). Magnesium (Mg). Penetration resistance (PR). Total microbial respiration (TMRc). Soil bulk density (BD). Note: All soil variables belong to horizon 1.

			Bootstrap				
						95% confidence intervals	
Model	Parameter	B (Estimate)	Bias	Standard error	Significance (bilateral)	lower	upper
3	(Constant)	9.03	-0.097	1.210	0.000	6.893	11.168
	NVLomc	-0.02	-0.001	0.008	0.012	-0.038	-0.008
	Mg	0.19	-0.008	0.082	0.031	0.038	0.316
	PR	-8.78	-0.070	1.752	0.000	-12.429	-5.689
	TRMc	-0.40	0.018	0.102	0.002	-0.625	-0.152
	BD	-4.70	0.068	0.732	0.000	-6.399	-2.851

4. Discussion

Nowadays there is a great interest from banana farmers in Venezuela, among other producing countries, to optimize input use in plantations. For this, characterizing the productive efficiency of banana plots will facilitate the delimitation of zones for differentiated management (Castañeda et al. 2010). This study identified the soil properties most associated with banana productivity in two producing areas if Venezuela to explore the potential development of empirical simple models that could help in predicting the banana productivity by estimating a PI as a function of soil characteristics. In these conditions, the three biometric variables associated with the PI used (circumference of the pseudo-stem of the mother plant, number of hands per bunch, and the height of the succession plant) confirmed the prior designation of each sampled area in different categories according to the production of each of those spots, indicating its validity as proxies of yield. This is in line with previous studies which also have established good accuracy in the estimation of banana productivity using biometric characteristics of the crop (Sumner 1990; Rodríguez et al. 2004, 2005; Segura et al. 2015), but not in the regions covered by our study.

We have demonstrated that it is possible to develop different models to characterize the PI in banana plantations based on soil characteristics and provided guidance for selecting the best choice from a stakeholder's perspective. In this approach Model 1, although showing better correlation with PI, was ruled out as a practical alternative for estimating banana productivity because it included too many soil variables - nine - and consequently would be more difficult to be implemented at the field level. Therefore, the best models to be implemented at the field level in terms of information needed and ability of PI prediction were models 2 and 3 (with six and five soil variables included in each model, respectively). Model 3 was finally selected due to similar performance to model 2, and because it was based on only five soil variables: magnesium, penetration resistance, total microbial respiration, soil bulk density and omnivorous free-living nematodes from soil horizon 1. Delgado et al. (2010b) also estimated an index of quality and health for Venezuelan soils, in Aragua and Trujillo, associated with pH, magnesium, copper, % sand, penetration resistance, free-living nematodes, total microbial respiration and organic matter. Although he did not link it to a productivity index. These soil variables are largely consistent with the results reported in our study. It is worth noting that variables that were significant in our model to estimate the PI are not usually considered by Venezuelan producers, especially for banana production, and so they provide additional insight form improving agronomical practices. For instance, magnesium is an element almost not considered in fertilization programs, low attention is played to physical conditions (soil bulk density and penetration resistance) and negligible consideration is given to biological aspects, such as free-living nematodes which have shown to be an adequate variable to estimate productivity. There is no need to emphasize that the kind of models linking soil quality and yield are site-specific, as indicated previously by Delgado et al. (2010b), Villarreal-Nuñez (2013) and Rey et al (2009). Nevertheless, these models aimed to identify variables related to banana productivity use variables similar to the ones identified in our study: pH, magnesium, copper, sand content, penetration resistance, total microbial respiration, organic matter, radical weight, total bacteriophages free-living nematodes and abundance of *Trichoderma*.

As noted above, our analysis can provide an overall insight into key variables that might be hampering banana productivity. So, Mg levels were moderate to low in all the plots sampled in our study, but always lower in the high-productivity plots (Table A.1, Appendix A), indicating that there may be problems of nutritional balance and absorption of these nutrients by part of the plants. Our results have shown a clear correlation of Mg with the PI in the study areas, even though the literature indicates that Mg requirements for crop growth are smaller than Ca requirements Fageria (2009) suggests that in these areas there is a response to differences in available Mg, as noted in other banana growing regions because it is a neglected element in fertilization. An example of previous findings in this direction is that of Irizarry et al. (1990) in Ultisol soils in Puerto Rico that indicated a significant increase in yield of bananas to Mg fertilization. In our study, the physical variables most related to crop productivity levels were BD and PR. According to Blanco-Sepúlveda (2009) there is a direct relationship between the PR and the BD. In the light of the results obtained in this work, the soil BD can be used as a relevant soil property to estimate the potential banana productivity in different producing areas in Venezuela. The high BD values observed in some plots are associated with the lacustrine origin of these which can affect the normal development of banana roots. Bananas show a high sensitivity to root damage and decreased yields due to soil compaction (Villarreal-Núñez et al. 2013) also confirmed by studies of Rey et al. (2006) in Venezuela and of Pattison et al. (2005) in Queensland (Australia) noting that most of the problems related to soil quality and low banana plantation productivity were related to some of the physical properties of the soil restricting root development. Thus, the weight of the roots and shoots of the banana decreases significantly as the BD of the soil increases. Similarly, banana at high soil compaction presents a significant decrease in the total number of leaves, the average leaf area, the total leaf area, the diameter of the pseudo stem, and the height of the succession plant (Villarreal-Núñez et al. 2013). Drastic changes in soil aggregation, in the presence of expandable clays, or soil compaction associated with less porosity and poor drainage, could damage roots and affects banana production (Araya and Blanco, 2001, Castañeda et al. 2010). Textural extremes (clay and sand) are also factors that limit the growth of the banana root (Aguirre et al. 2012). Indeed, for heavy textures, where there is high bulk density and resistance to root penetration, water availability and nutrient absorption by crops is highly restricted.

The inclusion of biological soil parameters such as the abundance of free-living nematodes, and microbial abundance and activity in soil quality studies, as we have done in this study, can provide significant knowledge of the link between the ecological role of soil fauna (Vega-Ávila et al. 2018) and effect of these organisms on banana productivity. The large number of culturable microorganisms and nematodes found in the soil associated with the banana rhizosphere in soils of the two localities sampled in our study could be attributed to the changes made by the plant roots in the surrounding rhizosphere microenvironment, resulting in higher populations in the rhizosphere of the most producing trees, something also found by Segura et al. (2015) in banana plantations in Costa Rica. Soil nematodes are a group of invertebrates of high ecological importance, which have attributes that make them valuable tools as biological indicators (Salas and Achinelly 2020). In the sampled soils in our study, free-living fungivore nematodes were predominant, followed by omnivores and lower proportions of bacteriophage and predatory microorganisms (*data not shown*). The low proportion of this last trophic group could reflect the presence of soil disturbance due to intensification and poor management of soils in the experimental spots evaluated. Our results agree with those of Castilla-Díaz et al. (2017) which found the genus c.f. Dorylaiminae (Omnivorous) as the most abundant nematode in Colombia. Millán et al. (2016) showed that greater diversity and richness of nematode populations, mainly phytophagous and omnivorous is observed in medium texture soils, whereas soils of heavier texture presented more abundance of the genus c.f. Dorylaiminae in areas in which banana is cultivated. Other studies have found that omnivore and predator nematodes are less abundant in disturbed soils or intensive banana plantations (Bautista et al. 2015; Ferris et al. 1996).

The metabolic activity of soil microorganisms is responsible for important processes such as mineralization and humification of soil organic matter (Martínez et al. 2018; Trujillo-Narcía et al. 2019). In our research, high productivity plots (HH) showed low microbial respiration in all the sites evaluated, and in fact soil organic carbon was not among the soil properties selected to be part of the predictive model of banana productivity based on soil properties. These results are consistent with the experiences of Gauggel et al. (2005) and González-Pedraza et al. (2014) that indicated the lower amount of microbial carbon and total organic carbon observed in Venezuelan soils with high productivity plants. This indicates that specific strategies to increase soil organic carbon content in banana fields for the provision of other ecosystem services, such as for instance carbon sequestration, need to be identified, since farmers might find little incentive to increase soil organic carbon if their link to yield is not straightforward in the short or medium term. The lower concentration of microorganisms observed in the roots of high productivity plots (HH) could be a result of biochemical relationships between different nutrients (Al, Cu, Zn, Hg and Ag) and the other microorganisms; these elements can be part of important enzymatic processes in the rhizosphere and root tissues. In typical Latin American areas of Musaceae production, such as Venezuela, Colombia, and Costa Rica, several authors (Delgado et al. 2008; Hernández et al. 2007; Delgado et al. 2010c), have reported similar results to those presented in this research. The differences found in the values of total soil microbial respiration between the plots of high and low productivity could be due to differences in the composition of the microbial communities in the rhizosphere area in both soils, which deserves further study.

The results in our research demonstrate the complex interactions between the chemical, physical, and microbiological variables of the soil and how they affect banana productivity. This information, and particularly the methodological approach to develop the model linking selected soil properties with banana productivity, is innovative and new in the banana cultivation of Venezuela. It shows that the microbial respiration and the concentration of free-living nematodes of the banana soil are suppressed in areas with high productivity (i.e., HH plots). Practices to increase and maintain soil quality and stimulate microbiological activity in those soils could have a positive effect on agricultural banana production, not only for low productivity fields but also for sustainable use of high productivity lots.

5. Conclusions

Our study modeled banana production capability based on plant biometric data developing a Productivity Index (PI) that was then linked to soil properties by developing an empirical model which predicted the PI based on five soil properties commonly used in soil quality indicators. These five soil properties have clear agronomic and environmental significance: magnesium content, penetration resistance, total microbial respiration, soil bulk density, and omnivorous free-living nematodes. This model might be used at the field level for reliable identification of areas of high and low banana productivity in the studied areas of Venezuela, that might be extended to other banana production sites in the different States of Venezuela (e.g., Yaracuy, Monagas, Sucre, Zulia, Barinas) with additional field validation. Finally, the identification in this study of the main soil properties that are most related to banana productivity, may contribute to provide guidance to farmers on the selection of sustainable soil management practices, considering the natural variability of soils, to improve the long-term sustainability of banana production in Venezuela.

Appendix A

Table A.1 Means values ± standard deviations for the variables selected in model 3: Magnesium content (Mg), Penetration resistance (PR), total microbial respiration (TMR), soil bulk density (BD) and Omnivorous free-living nematodes (NVLomc) in the six banana sites included in the study. All soil properties correspond to horizon 1. *Classification nomenclature of banana soil productivity according to average fruit production: HH= High - High yield (≥40 t/ha/year); HL= High - Low yield (35-40 t/ha/year); LL= Low - Low yield (≤30 t/ha/year); LH= Low - High yield (30-35 t/ha/year).

Fields	Productivity level*	Mg ($cmol.kg^{-1}$)	PR (kPa)	TMRc (mg $100.g^{-1}$)	BD ($Mg.m^{-3}$)	NVLomc (Log10)
Aragua State						
A1	HH	7.22± 0.90	0.35±0.05	5.78±1.67	0.79±0.06	2.75±1.25
	HL	5.44±1.37	0.32±0.03	7.61±0.75	0.77±0.09	9.25±8.53
	LL	7.23±0.90	0.41±0.08	5.34±0.72	1.26±0.03	28.00±13.85
	LH	6.82±0.61	0.29±0.07	6.28±1.51	1.14±0.05	25.75±16.85
A2	HH	1.56±0.11	0.11±0.02	0.29±0.26	1.45±0.02	130.00±14.14
	LL	1.76±0.01	0.30±0.01	2.67±0.02	1.40±0.10	35.00±7.07
A3	HH	0.97±0.38	0.13±0.01	1.61±0.57	1.54±0.05	5.00±7.08
	LL	1.00±0.32	0.17±0.00	2.53±0.67	1.36±0.11	10.00±14.14
A4	HH	0.96±0.09	0.07±0.04	2.83±0.01	1.35±0.04	40.00±14.13
	LL	1.83±0.03	0.23±0.01	4.30±0.07	1.29±0.01	70.00±70.70
Trujillo State						
T1	HH	5.15±0.28	0.23±0.12	1.34±0.62	1.42±0.07	2.00±2.31
	HL	4.92±0.30	0.23±0.03	1.44±1.30	1.50±0.06	4.00±3.27
	LL	4.50±0.71	0.16±0.01	2.26±1.28	1.53±0.08	12.00±5.66
	LH	5.85±1.00	0.24±0.08	3.11±0.74	1.54±0.07	10.00±10.07
T2	HH	1.49±0.76	0.09±0.04	1.85±0.81	1.26±0.10	12.00±0.01
	HL	0.66±0.37	0.13±0.05	2.38±1.34	1.38±0.12	9.00±1.10
	LL	1.82±0.32	0.22±0.02	3.56±1.50	1.33±0.03	12.00±9.70
	LH	1.69±0.52	0.12±0.02	4.12±2.19	1.29±0.14	12.00±9.79

Funding: The authors recognize the financial support for international mobility of the Ibero-American scholarship program (2018-2019) of Banco Santander. Also, by project "Technological innovations for the management and improvement of the quality and health of banana soils in Latin America and the Caribbean" financed by FONTAGRO and coordinated by Bioversity International (before INIBAP) and project SHui (European Commission Grant Agreement number: 773903).

References

Acevedo-Opazo C, Valdés-Gómez H, Taylor JA, Avalo A, Verdugo-Vásquez N, Araya M, Jara-Rojas F, Tisseyre B (2013) Assessment of an empirical spatial prediction model of vine water status for irrigation management in a grapevine field. Agric Water Manag 124: 58-68. https://doi.org/10.1016/j.agwat.2013.03.018

Addinsoft TM (2010) XLSTAT software. version 5.03. Addinsoft. Barcelona. Spain. http://www.xlstat.es

Aguirre ES, Piraneque N, Menjivar Flores CJ (2012) Relationship between soil and edaphoclimatic properties and the Black Sigatoka incidence (*Mycosphaerella fijiensis* Morelet) in the Banana Region of Magdalena - Colombia. Rev Investig Agrar Ambient 3:1-13.https://doi.org/10.22490/21456453.970

Alef K (1995) Field methods. In: Alef K, Nannipieri P (ed) Methods in Applied Soil Microbiology and Biochemistry. Academic Press, London, pp 464-470.

Araya M, Blanco F (2001) Changes in the stratification and spatial distribution of the banana (Musa AAA cv. Grand Naine) root system of poor regular and good developed plants. J Plant Nutr 24:405-405. https://doi.org/10.1081/PLN-100107306

Araya-Alman M, Leroux C, Acevedo-Opazo C, Guillaume S, Valdés-Gómez H, Verdugo-Vásquez N, Pañitrur-De la Fuente C, Tisseyre B (2019) A new localized sampling method to improve grape yield estimation of the current season using yield historical data. Precis Agric 20:445-459. https://doi.org/10.1007/s11119-019-09644-y

Bautista LG, Bolaños MM, Massae Asakawa N, Villegas EB (2015) Plantain *Musa* AAB simmonds phytonematodes response to soil integrated management strategies and nutrition. Rev Luna Azul 40:69-84. http://dx.doi.org/10.17151/luaz.2015.40.6.

Blanco-Sepúlveda R (2009) Relationship between bulk density and mechanical resistance as an indicator of soil compaction. Agrociencia 43: 231-239.

Bongers T (1990) The maturity index: an ecological measure of environmental disturbance based on nematode species composition. Oecologia 83:14-19. https://doi.org/10.1007/BF00324627

Breiman L, Last M, Rice J (2003) Random forests: Finding quasars. In: Statistical Challenges in Astronomy. Springer, New York, NY. https://doi.org/10.1007/0-387-21529-8_16

Brinford G, Blackmer A, Cerrato M (1992) Relationship between corn yields and soil nitrate late spring. Agron J 84:59-63. https://doi:10.2134/agronj1992.00021962008400010012x

Calero J, Aranda V, Montejo-Raez A, Martin-Garcia JM (2018) A new soil quality index based on morpho-pedological indicators as a site-specific web service applied to olive groves in the Province of Jaen (South Spain). Comput Electron Agric 146:66-76. https://doi.org/10.1016/j.compag.2018.01.016

Cassel DK, Nielsen DR (1986) Field Capacity and Available Water Capacity. Methods of Soil Analysis: Part 1—Physical and Mineralogical Methods. American Society of Agronomy-Soil Science Society of America, Madison. https://doi.org/10.2136/sssabookser5.1.2ed.c36

Castañeda DA, Jaramillo DF, Cotes Torres JM (2010) Components of the spatial variability in site-specific management in banana. Pesqui Agropecu Bras 45:836-845. https://doi.org/10.1590/S0100-204X2010000800009

Castilla-Díaz E, Millán-Romero E, Mercado-Ordoñez J, Millán-Páramo C (2017) Relation of soil parameters on the diversity and spatial distribution of free-living nematodes. Tecnol Marcha 30: 24-34. https://doi.org/10.18845/tm.v30i3.3270

Delgado E, Rosales F, Trejos J, Villalobos M, Pocasangre L (2010a) An index of soil quality and health for banana plantations in four countries of Latin America and the Caribbean. Bioagro 22:53-60

Delgado E, Trejos J, Villalobos M, Martínez G, Lobo D, Carlos Rey J, Rodríguez G, Rosales FE, Pocasangre LE (2010b) Determination of a soil quality and health index for banana plantations in Venezuela. Interciencia 35:927-933

Delgado EJ, Carlos Rey J, Martínez G, Lobo Luján D, Sapuky M, Chacin M, Nuñez E, Pocasangre L, Rosales F (2010c) Propiedades del suelo que determinan el vigor del banano (Musa AAA) de producción local en Venezuela. XIX Reunión Internacional ACORBAT Medellín, Colombia, nov 8-12 (In Spanish)

Delgado E, Gómez N, González O, Marín C (2008) On farm evaluation of effect of high plant density on plantain (*Musa* AAB cv. Subgroup Horn plantain), Obispo County, Barinas state, Venezuela. Rev Fac Agron (LUZ) 25:603-616

Efron B, Tibshirani RJ (1993) An Introduction to the Bootstrap. Chapman &Hall, New York

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. J Anim Ecol 77:802-813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Fageria NK (2009) The use of Nutrients in Crop Plants. CRC Press, Florida

FAO (2019) Banana Market Review and Banana Statistics 2018. Rome. http://www.fao.org/economic/est/est-commodities/bananas/en/ Accessed jan122020

Ferris H, Venette RC, Lau SS (1996) Dynamics of nematode communities in tomatoes grown in conventional and organic farming systems. and their impact on soil fertility. Appl Soil Ecol 3:161-175. https://doi.org/10.1016/0929-1393(95)00071-2

Friedman JH (2001) Greedy function approximation: A gradient boosting machine. Ann. Stat 29:1189-1232. https://doi.org/10.1214/aos/1013203451

Friedman JH, Meulman JJ (2003) Multiple additive regression trees with application in epidemiology. Stat Med 22:1365-1381. https://doi.org/10.1002/sim.1501

Gauggel CA, Sierra F, Arévalo G (2005) The problem of banana root deterioration and its impact on production. In: D.W. and F.E. Rosales (eds) Banana Root System: towards a better understanding for its productive management: Proceedings of an international symposium. International Network for the Improvement of Banana and Plantain, Montpellier, France, pp 13-22

González-Pedraza AF, Atencio J, Cubillan K, Almendrales R, Ramirez L, Barrios O (2014) Microbial activity in soils cultivated with plantain (*Musa* AAB plantain subgroup cv. Horn) with different vigor of plants. Rev Fac Agron (LUZ) 31:526-538

Hernández Y, Marín M, García J (2007) Response to the plantain (*Musa* AAB cv. Horn) yield as a function of the mineral nutrients and its phenological cycle. Part I. Growth and Production. Rev Fac Agron (LUZ) 24:607-626.

Horton JH, Newsom DW (1953) A rapid gas evolution method for calcium carbonate equivalent in liming materials. Soil Sci Soc Am J 17: 414–415.

IANAS (Interamerican Network of Academies of Sciences. México) (2017) Retos y oportunidades de la seguridad alimentaria y nutricional en las Américas: El punto de vista de las Academias de Ciencias. IANAS, México (In Spanish).

Irizarry H, Rivera E, Rodriguez JA (1990) Response of bananas (*Musa-acuminata*. AAA) to magnesium fertilization in an Ultisol. J Agr U Puerto Rico 74:419-426

Joergensen RG (1995) Microbial biomass. In: Alef K. Nannipieri P (eds) Methods in Applied Soil Microbiology and Biochemistry. Academic Press. London. pp 375-417

Martínez M, Ortega R, Janssens M, Fincheira P (2018) Use of organic amendments in table grape: effect on plant root system and soil quality indicators. J Soil Sci Plant Nutr 18: 100-112. https://dx.doi.org/10.4067/S0718-95162018005000501

Mehlich A (1984) Mehlich 3 soil test extractant: a modification of Mehlich 2 extractant. Comm Soil Sci Plant Anal 15:1409-1416

Millán E, Castilla D, Millán C (2016) Comunidades de nematodos de vida libre del suelo y su correspondencia con la calidad. Ingenieria Reg 16: 25-34 (In Spanish) https://doi.org/10.25054/22161325.1282

Miller A (2002) Subset Selection in Regression. New York: Chapman and Hall/CRC. https://doi.org/10.1201/9781420035933

Olivares BO, Cortez A, Lobo D, Parra R, Rey J, Rodriguez M (2017) Evaluation of agricultural vulnerability to drought weather in different locations of Venezuela. Rev Fac Agron (LUZ) 34:103-129

Olivares BO (2018) Tropical rainfall conditions in rainfed agriculture in Carabobo, Venezuela. Lgr Lif Sci J 27:86-102. https://doi.org/10.17163/lgr.n27.2018.07

Olivares BO, Hernández R, Coelho R, Molina JC, Pereira Y (2018) Analysis of climate types: Main strategies for sustainable decisions in agricultural areas of Carabobo, Venezuela. Sci. agropecu 9:359-369.http://dx.doi.org/10.17268/sci.agropecu.2018.03.07

Pattison T, Smith L, Moody P, Armour J, Badcock K, Cobon J, Rasiah V, Lindsay S, Gulino L (2005) Banana root and soil health project - Australia. In: D.W. and F.E. Rosales (eds) Banana root system: towards a better understanding for its productive management. Proceedings of an International Symposium International Network for the Improvement of Banana and Planta. Montpellier, France, pp 149-165

Pla I (1983) Metodología para la caracterización física con fines de diagnóstico de problemas de manejo y conservación de suelos en condiciones tropicales. Rev Fac Agron 32: 1-91 (In Spanish)

Primpas I, Tsirtsis G, Karydis M, Kokkoris GD (2010) Principal component analysis: Development of a multivariate index for assessing eutrophication according to the European water framework directive. Ecol. Indic 10:178-183. https://doi.org/10.1016/j.ecolind.2009.04.007

R Core Team (2015) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria

Rey JC, Chacín M, Sapuky M, Núñez M, Martínez G, Rodríguez G, Espinoza J, Arturo M, Pocasangre L, Delgado E, Rosales F (2006) Land suitability for banana in Venezuelan soils and its relation to productivity. XVII Reunión Internacional ACORBAT: Banano un negocio sustentable. Joinville. Santa Catarina, Brasil. Nov 15-20. p. 362.

Rey JC, Martínez G, Lobo D, Trejos J, Pocasangre L, Rosales F (2009) Aspectos sobre calidad y salud de suelos bananeros en Venezuela. Producción Agropecuaria 2: 52-55 (In Spanish)

Rodríguez G, Núñez MC, Lobo D, Martínez G, Rey JC, Espinoza J, Muñoz N, González D, Rosales F, Pocasangre L, Delgado E (2006) Banana root health in lots with different productivity levels in a soil at the oriental coast of Maracaibo Lake, Venezuela. XVII Reunión Internacional ACORBAT: Banano un negocio sustentable. Joinville, Santa Catarina, Brasil. Nov 15-20. p. 355

Rodríguez V, Malavolta E, Sánchez A, Lavoranti O (2004) Reference nutritional balance of soils and leaves in the Horn plantain crop. Bioagro 16:39-46

Rodríguez V, Silva A, Rodríguez O (2005) Nutritional balance and number of leaves as variables for yield prediction of the Horn plantain. Pesqui Agropecu Bras 40:175-177. https://doi.org/10.1590/S0100-204X2005000200011

Rosales FE, Pocasangre LE, Trejos J, Serrano E, Peña W (2008) Guía de Diagnóstico de la Calidad y Salud de Suelos. Bioversity International. Montpellier, France (In Spanish)

Salas A, Achinelly MF (2020) Community structure of soil nematodes associated with the rhizosphere of *Solanum Lycopersicum* in a major production area in Argentina: A Case Study Among Agroecosystem Types. J Soil Sci Plant Nutr 20: 43–54. https://doi.org/10.1007/s42729-019-00099-8

Segura RA, Serrano E, Pocasangre L, Acuna O, Bertsch F, Stoorvogel JJ, Sandoval JA (2015) Chemical and microbiological interactions between soils and roots in commercial banana plantations (*Musa* AAA. cv. Cavendish). Sci Hortic 197: 66-71. https://doi.org/10.1016/j.scienta.2015.10.028

Soil Survey Staff (2010) Keys to soil taxonomy. 11th ed. NRCS. Washington DC

Sumner ME (1990) Advances in the use and application of plant analysis. Comm Soil Sci Plant Anal 21:1409-1430. https://doi.org/10.1080/00103629009368313

Trujillo-Narcía A, Rivera-Cruz MC, Magaña-Aquino M (2019) The burning of sugarcane plantation in the tropics modifies the microbial and enzymatic processes in soil and rhizosphere. J Soil Sci Plant Nutr 19: 906–919.https://doi.org/10.1007/s42729-019-00089-w

USDA (1995) Soil survey laboratory methods manual. USDA. USA

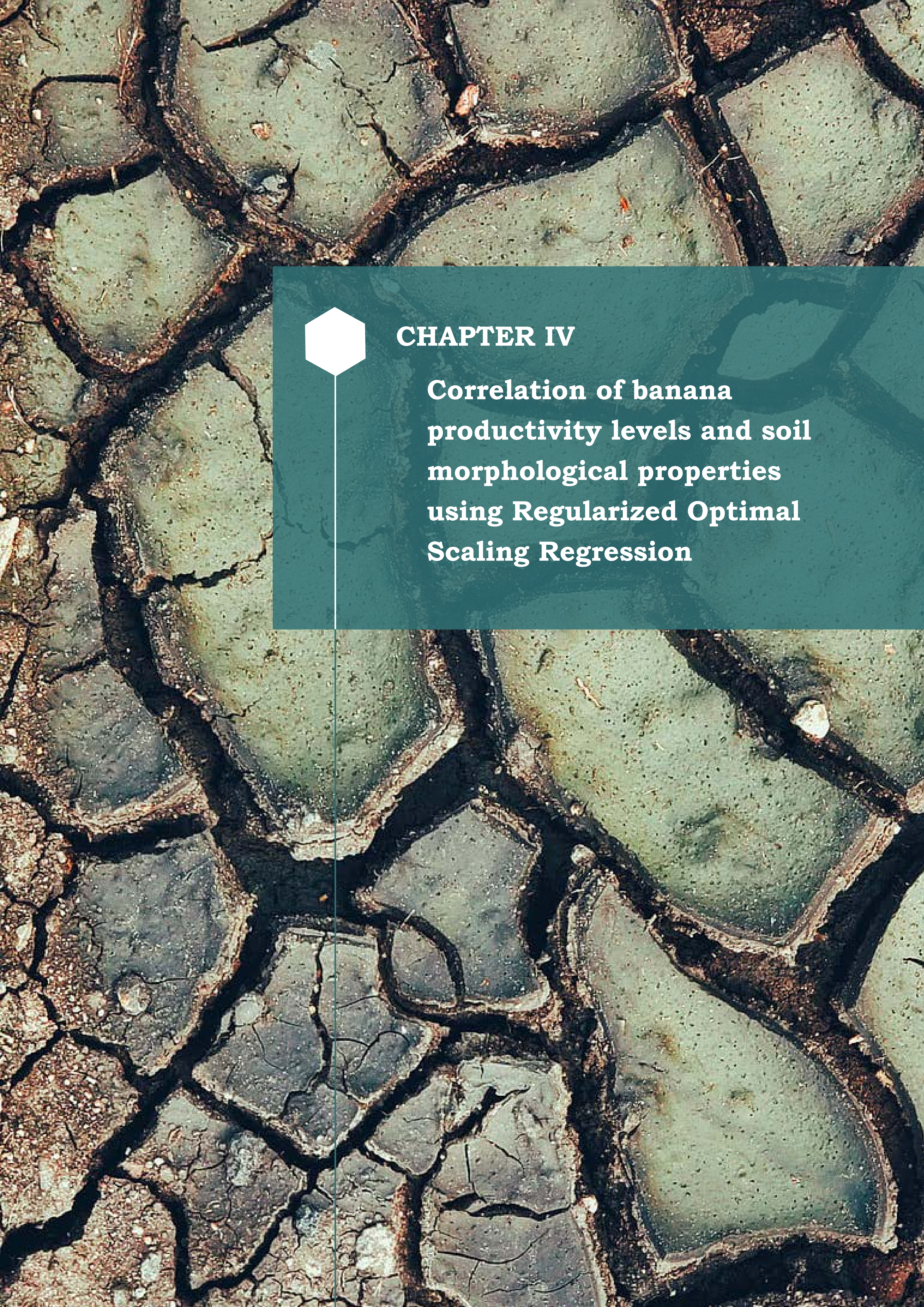
Vaquero MR (2005) Soil physical properties and banana root growth. In Turner DW, Rosales FE (ed). Banana Root System: towards a better understanding for its productive management: Proceedings of an international symposium. International Network for the Improvement of Banana and Plantain, Montpellier, France, pp 125 – 131

Vega-Ávila A, Medina E, Paroldi H, Toro M, Baigori M, Vázquez F (2018). Bioindicators of soil quality of open shrubland and vineyards. J Soil Sci Plant Nutr 18: 1065-1079. https://dx.doi.org/10.4067/S0718-95162018005003002

Villarreal-Núñez J, Pla-Sentis I, Agudo-Martínez L, Villaláz-Perez J, Rosales F, Pocasangre L (2013) Soil quality indexes in areas cultivated with banana in Panama. Agron Mesoam 24:301-315

Viloria JA, Pineda C, López D, Gabaldón O (2003) Land suitability of the Taguaiguay irrigation system (Aragua-Venezuela) for sugarcane and banana. Agron Trop 53:189-208

Weaver F, Angle J, Bottomley P (1994) Methods of Soil Analysis. Part 2. Microbiological and Biochemical Properties. Book Series Number 5. Soil Sci Soc Am, USA



CHAPTER IV

Correlation of banana productivity levels and soil morphological properties using Regularized Optimal Scaling Regression

This is a literal reproduction of the article: Olivares B, Calero J, Rey JC, Lobo D, Landa BB, Gómez JA, (2022). Correlation of banana productivity levels and soil morphological properties using Regularized Optimal Scaling Regression. Catena, 208(1): 105718. https://doi.org/10.1016/j.catena.2021.105718.

Correlation of banana productivity levels and soil morphological properties using Regularized Optimal Scaling Regression

Barlin Orlando Olivares, Julio Calero, Juan Carlos Rey, Deyanira Lobo, Blanca B. Landa, José Alfonso Gómez

CHAPTER IV. Correlation of banana productivity levels and soil morphological properties using Regularized Optimal Scaling Regression

Abstract

Soil morphological properties described in the field, such as texture, consistence or structure, provide a valuable tool for the evaluation of soil productivity potential. In this study, we developed a regression model between the soil morphological variables of banana plantations and a crop Productivity Index (PI) previously developed for the same areas in Venezuela. For this, we implemented categorical regression, an optimal scaling procedure in which the morphological variables are transformed into a numerical scale, and can thus be entered in a multiple regression analysis. The model was developed from data from six plantations growing “Gran Nain” bananas, each with two productivity levels (high and low), in two 4-ha experimental plots, one for each productivity level. Sixty-three A horizons in thirty-six soils were described using 15 field morphological variables on a nominal scale for structure type, texture and hue, and an ordinal scale for the rest (structure grade, structure size, wet and dry consistence, stickiness, plasticity, moist value, chroma, root abundance, root size, biological activity and reaction to HCl). The optimum model selected included biological activity, texture, dry consistence, reaction to HCl and structure type variables. These variables explained the PI with an R^2 of 0.599, an expected prediction error (EPE) of 0.645 and a standard error (SE) of 0.135 using bootstrapping, and EPE of 0.662 with a SE of 0.236 using 10-fold cross validation. Our study showed how soil quality is clearly related to productivity on commercial banana plantations, and developed a way to correlate soil quality indicators to yield by using indicators based on easily measured soil morphological parameters. The methodology used in this study might be further expanded to other banana-producing areas to help identify the soils most suitable for its cultivation, thereby enhancing its environmental sustainability and profitability.

Keywords

biological activity, sustainability, qualitative soil indicators, dry consistence, soil structure, texture.

1. Introduction

The banana is one of the most important crops in the world after rice, wheat and corn, both in terms of production yields and area cultivated. This fruit constitutes the basis of the diet in tropical countries like Costa Rica, Colombia, Ecuador, Panama and Venezuela. It is also an important source of income for producers (FAO, 2020). Identification of the most suitable areas for banana cultivation is essential to increasing its productivity in tropical regions. Therefore, soil properties must be characterized to understand their relationship to crop productivity (Villarreal-Núñez et al., 2013; Delgado et al., 2010b).

Soil morphological field properties have been recognized as a valuable tool for studying a broad range of soil characteristics, including those related to soil development in agricultural areas, for the ease and speed with which they can be described (Soil Survey Staff 2017; Calero et al., 2008; Pulido-Moncada et al., 2017). Soil morphological properties are relatively easily and economically characterized in soil pits, almost all are included in soil databases, and they can be easily determined by technicians (Delgado et al., 2010a). Unlike soil chemistry and most of its biological properties, field morphological data are generally nonnumerical (nominal or ordinal) and measured on evaluation scales with origins difficult to establish (Vaughan & Ormerod, 2005; Meulman et al., 2002). Similarly, numerical relationships of different categories are still poorly developed. Such data cannot be subjected to rigorous treatment by statistical methods, such as factor analysis or multiple regression, like numerical quantitative data (Andrews et al., 2004; Linting et al., 2007; Calero et al., 2005). Therefore, there is little background for calculating soil quality indices from categorical indicators established by decision rules based on existing knowledge (Pulido-Moncada et al., 2014, Calero et al., 2018).

Soils in banana plantations in Venezuela were first characterized at the beginning of the 20th century. Other regional and farm characterization studies were carried out under Venezuelan research and development programs from 1970 to 1998 (Martínez et al., 2008). Some of the most recent soil characterization studies in banana zones are those by Hernández et al. (2007), Rey et al. (2009), Delgado et al. (2010a, 2010b), González-Pedraza et al. (2014), Olivares et al. (2020), González García et al. (2021a, 2021b, 2021c) and González-Pedraza & Escalante (2021). From the 1970s onwards, the worldwide trend toward intensification of banana production systems was also observed in Venezuela, and the relationship between reduction in productivity and loss of soil quality became apparent. As a result, growers are looking for innovative sustainable management methods that can simultaneously maintain or increase productivity. Improved use of available, or newly acquired, soil information, is essential for achieving such sustainable banana production in tropical countries.

The link between soil quality and current or potential productivity is critical in the search increased productivity and sustainability in agriculture. As mentioned, most approaches are based on quantitative variables, while the potential of morphological variables has still only been moderately explored. Several methods have been studied, in which the high potential of categorical regression (CATREG), already in use in such fields as education, marketing, and agricultural economics, among others (van der Kooij, 2007; Meulman et al., 2019; Sevinç et al., 2019), has been highlighted. However, to our knowledge, it has not been extensively used in agronomic sciences and never in banana fields. This is a novel element in our study, in which soil morphological properties were explored as promising new soil indicators for assessing banana productivity in Venezuelan soils. Our study is a pioneer in the application of CATREG, an optimal scaling procedure, to the transformation of morphological variables into a numerical scale, which is completely novel in banana soils. Nevertheless, beyond its application in the case of a specific crop (banana) and geography (Venezuela), we have developed a scientific rationale that is easily transferred to other areas, not only in agriculture, but soil science in general.

According to many authors (i.e., Mc Ewan and Fitzpatrick, 1996; Lal et al., 1998; and just recently Vasu et. al., 2021), it is important for soil assessment to be based on such characteristics as morphological properties, that are measured easily and inexpensively. Beyond its taxonomic utility, morphological properties can improve soil assessment. A wealth of such data is available from agricultural services (universities, research centers, etc.), just waiting to be usefully employed, and our study is a fine example of how it can be used.

This study aimed to validate the hypothesis that soil morphological properties can differentiate banana productivity levels in large areas of Venezuela using a categorical regression prediction model. In this case, qualitatively estimated soil morphological properties could be used for improving the assessment of banana productivity. Categorical regression analysis can provide an operating model for bananas in an area where there is little background of soil quality indices with categorical soil properties.

2. Materials and methods

2.1. Description of the study areas and banana plantations

Six banana plantations in the states of Aragua and Trujillo in Venezuela were selected (Table 1). The plantations in the State of Aragua (PL, SM, PZ and CH) are in the Lake Valencia Basin. They are characterized by a tropical savanna (Aw) climate with a mean annual rainfall depth of 980 mm. Rainfall in this area is seasonal for five to six rainy months, concentrated between May and October (Olivares et al., 2021). The mean annual temperature is 26.2°C and the mean annual relative humidity is 70.0% (Olivares, 2018). The terrain relief is flat (slope 0-2%). Plantation PL is located on the fourth level of the lacustrine terrace produced by drying of Lake Valencia, while Plantations SM, PZ and CH are on alluvial soils, all with medium to silty textures. Soils in these farms are Mollisols and Inceptisols, generally with moderate to good drainage, soil pH neutral to alkaline, fertile, with medium to high organic matter content (Delgado et al., 2010a).

Table 1. Geographic location and planted area of bananas (ha) of the sampled plantations in Venezuela.

Plantation's code	Geographical coordinates	Height (masl)	Sites	Soil Taxonomy†	State	planted area (ha)
PL	10º 12' N; 67º 30' W	435	H	Mollic Ustifluvents Fluventic Haplustolls Cumulic Haplustolls	Aragua	135
			L	Oxyaquic Ustifluvents Fluventic Haplustolls		
SM	10º 12' N; 67º 23' W	502	H	Fluventic Haplustepts Fluventic Haplustolls	Aragua	11
			L	Fluventic Haplustolls		
PZ	10º 11' N; 67º 31' W	514	H	Fluventic Haplustolls	Aragua	20
			L	Fluventic Haplustepts		
CH	10º 11' N; 67º 31' W	498	H	Fluventic Haplustolls	Aragua	9
			L			
BA	09º 29' N; 70º 57' W	16	H	Oxyaquic Ustifluvents Aeric Fluvaquents	Trujillo	300
			L	Typic Ustifluvents Oxyaquic Ustifluvents		
KA	09º 28' N; 70º 55' W	17	H	Oxyaquic Ustifluvents Fluventic Haplustoll	Trujillo	270
			L	Typic Ustifluvents Oxyaquic ustifluvents		

† Soil Survey Staff (2014). Sites: H: High and L: Low productivity

The second study area is located in the State of Trujillo (Plantations BA and KA), in the region southeast of Lake Maracaibo, also characterized by a tropical savanna (Aw) climate. The mean annual precipitation is 1094 mm, with two rainfall peaks, one in April-May (monthly precipitation approximately 120 mm) and the other in October (monthly precipitation approximately 145 mm). The driest months are January and February when the monthly precipitation is less than 50 mm (Olivares et al., 2017). The area has a mean annual temperature of 27.5°C and a mean relative humidity of 78.0%. This area of the State of Trujillo is an alluvial plain with slopes of less than 1.0% with mainly Entisol soils (Rodríguez et al., 2006). These soils have moderate to poor drainage with neutral to alkaline pH. They are moderately fertile and average organic matter content is around 2.75% (Rey et al., 2009). In both areas, soil management was concentrated on fertilization, and no reclamation action was taken to improve drainage or increase organic matter.

2.2. Soil sampling

Banana productivity was estimated in sampling areas delimited by productivity level following the guidelines proposed by Rosales et al. (2008). On all the plantations, two plots or productivity levels were identified a priori as High (H) and Low (L) for estimating the Productivity Index (PI). The "Gran Nain" variety was the only variety grown, and each productivity plot had an area of 4 ha (on the large >50 ha plantations, PL, BA and KA) with four replicated plots in each field. On large plantations (≥ 50 ha, PL, BA and KA), the average yield of high productivity plots was 69.8 ± 5.0 t ha^{-1} yr^{-1} and in low productivity plots, it was 59.7 ± 5.3 t ha^{-1} yr^{-1}. On the remaining small plantations (< 25 ha, SM, PZ and CH) these two levels of productivity were identified in 1-ha fields, with two replicated plots in each field, for a total of 36 plots. The average yield on small plantations was 11.5 ± 0.7 t ha^{-1} yr^{-1} for high productivity plots and 1.6 t ha^{-1} yr^{-1} for those with low productivity.

2.3. Productivity index (PI)

Our study used the banana productivity index (PI) previously developed by Olivares et al. (2020) to estimate productivity in each of the evaluation plots. The PI is based on the morphometric characteristics of banana plants, such as circumference of the mother plant pseudo-stem at 1 m height (cm), number of hands per bunch (n), and height of the succession plant (cm), following the methodology proposed by Rosales et al. (2008). The PI was generated from a principal component analysis (PCA) in which the variables best represented on the first axis (Principal Component 1) were retained. In addition, the linear combination of the three biometric variables evaluated was used as a synthesis variable, where the coefficients of this linear combination were the variable loadings in PC1. Then, in the categorical regression the ordinal scaling level was considered the response variable.

2.4. Morphological characterization of soil properties

A soil profile was evaluated in each of the thirty-six replicated plots indicated in Section 2.2. The A horizon morphological properties were described and characterized in the field following the FAO (2006) and the Soil Survey Staff (2017) methodologies. Color was determined according to the Munsell soil color charts (Munsell Color Company, 1999). Texture was determined in the laboratory by sieving and sedimentation, using a Robinson pipette, (Soil Survey Staff, 2017). Fifteen morphological variables in these soil profile descriptions were studied: texture class, structure size, structure grade, structure type, moist hue, moist value, moist chroma, moist consistence, dry consistence, stickiness, plasticity, biological activity, root abundance, root size and reaction to HCl.

2.4.1. Sample analysis

Our analysis was applied to the A horizons (N = 63) in each soil. The average thickness of these horizons is 31±12 cm, which roughly coincides with the area of maximum concentration of the banana tree (30 cm). We hypothesized that focusing our analysis on the A horizon would make more agronomic sense, since this is the area that most influences plant development and yield. The functional roots of the banana in our profile descriptions were concentrated in the A horizons which ranged, from about the top 30 to 40 cm deep, a root concentration range similar to the description of banana in the tropics by Gauggel et al. (2005).

2.5. Statistical modelling of field data

2.5.1. Regression with optimal scaling

First, a regression model was fitted to be able to predict the banana productivity index (PI) (*outcome*) based on the soil morphological properties (*predictors*) on the plantations sampled. However, the categorical or nonmetric nature of the predictors prevents proper application of classical regression analysis. Various techniques have been applied in an attempt to include categorical variables in multiple regression, such as the creation of dummy variables. These variables, however, introduce high multicollinearity and hinder interpretation of the model, reducing the significance of the regression coefficients (Wissmann et al., 2011) and the predictive power of the model (Hair et al., 1999; Xu et al., 2010).

The problem of multicollinearity between predictors can be dealt with by selecting a good theoretical model, which enters only those predictors with a high degree of unique variance with the dependent variable (Hair et al., 1999). However, it is hard to select categorical predictors using the usual stepwise procedure, where each variable is selected or discarded before its transformation to a numeric scale. For rigorous treatment of categorical variables, such as field morphological variables, in multivariate methods, optimal scaling methods need to be applied (Calero et al., 2018).

Categorical regression (CATREG) (van der Kooij et al., 2006; Meulman et al., 2019) is an optimal scaling technique that can transform the k_j categories ($s = 1, \ldots, j$) of the j^{th} nonmetric predictors, and the k_r categories of the response variable r, by means of nonlinear numerical functions, while minimizing the error (that is, increasing the coefficient of determination R^2 of the model). CATREG takes the classical linear regression model with nonlinear optimal scaling (Equation 1):

$$\varphi(\mathrm{y}) = \sum_{j=1}^{p} \beta_j \varphi(\mathrm{x}) + e \qquad (1)$$

where β_j are the regression coefficients of the j^{th} predictor, φ (x) and φ (y) the transformation functions of predictors (x) and response variable (y), respectively, and e is the model error. In this study, the only restriction applied to φ(x) and φ(y) was monotonicity, which makes it possible to distinguish between field soil morphological properties measured on an ordinal scale (having an intrinsic categorical order *i.e.*, stickiness: from not sticky to very sticky) from nominal variables not having this restriction. Hue, structure type and texture class were the second type. Rewriting Equation 1 in terms of indicator matrices and category quantifications yields Equation 2:

$$\boldsymbol{G_r y_r} = \sum_{j=1}^{p} \beta_j \boldsymbol{G_j y_j} + \mathrm{e} \qquad (2)$$

where $\boldsymbol{\beta}$ ($\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p$), in order p (the number of predictor variables), is the vector containing regression coefficients, $\boldsymbol{y}_j$ and $\boldsymbol{y}_r$, in order k_j and k_r, the optimal scaling's or numerical transformations of the categories for the predictors and response variable, respectively, and $\boldsymbol{G}_j$ and $\boldsymbol{G}_r$, in order $n \times k_j$ and $n \times k_r$ (where n is the number of cases), indicator matrices, such that 1 is when the i^{th} object is in the k_j category of variable j and 0 otherwise. CATREG estimates the regression coefficients by minimizing the least squares loss function (van der Kooij et al., 2006) in Equation 3:

$$\sigma(\boldsymbol{y_r}; \boldsymbol{\beta}; \boldsymbol{y_j}) = \left\| \boldsymbol{G_r y_r} - \textstyle\sum_{j=1}^{p} \beta_j \boldsymbol{G_j y_j} \right\|^2 \qquad (3)$$

The multiple correlation coefficient R^2 can be found from the ratio between the regression sum of squares and the total sum of squares (Gifi, 1990; van der Kooij, 2007) in Equation 4:

$$R^2 = N^{-1/2} (\boldsymbol{G_r y_r})' \, \mathbf{v} \, (\mathbf{v}\,\mathbf{v}')^{-1/2} \qquad (4)$$

where N is the number of observations and **v** is the accumulated contribution of predictor variables so that:

$$\mathbf{v} = \textstyle\sum_{j=1}^{p} \beta_j \boldsymbol{G_j y_j} \qquad (5)$$

The statistics (t, F values) and fit and error measures, as well as the correlation matrices R, partial correlation and predictor tolerance, which will be used to assess the goodness of the model, are found from β_j and R^2. Tolerance is defined as one less the determination coefficient R^2 of the prediction of any predictor by the others, considering them independent variables, so it should be high to avoid multicollinearity.

2.5.2. Accuracy of categorical regression (CATREG)

The error term in the regression model ($1\text{-}R^2$) is usually not a good estimator of the prediction error, since it is found from the same data used to fit the model. To obtain a better estimate, resampling methods, such as cross validation or bootstrapping, may be used.

In CATREG, the goal is to minimize the dataset prediction error. The expected prediction error (EPE) must be known to be able to assess the reliability of predictions of future observations (van der Kooij, 2007). Therefore, ideally, a training dataset would be used to estimate a model and a test dataset. However, in this case, there is no test dataset, and therefore two resampling methods were used: K-Fold Cross-Validation (CV) and the 0.632 Bootstrap, using the mean square error (MSE) for assessing the EPE.

In K-Fold Cross-Validation, data are divided into randomized k groups of approximately the same size. K-1 groups are used to train the model and another is used for validation. This process is repeated k times using a different group as validation in each iteration. The process generates k estimates of the error which is averaged as the final estimate (Meulman et al., 2019). In bootstrapping, N cases of the same size as the sample are taken at random from the full dataset in each resampling to obtain B bootstrapped samples. Then, a model is fitted for each bootstrapped sample, estimating its prediction error from the original (not resampled) dataset. Contrary to CV, random sampling is performed with replacement, that is, an observation can be repeated in the bootstrapped sample, while another is excluded (*out-of-bag* set). The simple bootstrap estimate of the expected prediction error is found by averaging the error estimates of the B bootstrapped samples. Since the bootstrapped (training set) and the original sample (validation set) have many observations in common, the bias seems to be greater than with CV and, in general, it performs worse. Efron (1983) proposed a modification of the bootstrap prediction error called the 0.632 method, which includes the bootstrapped model error estimates in the simple bootstrap in their own *out-of-bag* observations (not the original dataset). This corrects the bias and improves model performance (van der Kooij, 2007). Here, we used 10 folds for CV and 50 bootstrap samples in 0.632 bootstrapping. In addition to the reliability of the prediction estimated by the MSE, a more flexible approach can be used to evaluate the CATREG model accuracy (Hartmann et al., 2009). For this, the dependent variable (PI, which usually ranged from -4 to +4) was distributed into two classes or groups: N, negative with PI <0 and P, positive with PI> 0, using zero as the cut-off point. Thus, model accuracy could be addressed by such classification accuracy measures as Efficiency and the Receiver Operating Characteristic (ROC) curve (Garosi et al., 2019). Efficiency can therefore be defined as:

$$Efficiency = \frac{TP+TN}{TP+TN+FP+FN} \qquad (6)$$

where TP (true positive) and TN (true negative) represent the number of correctly classified P and N plots, respectively, FP (false positive) the number of N plots that have been classified as P and FN (false negative) those P considered N.

The ROC curve plots the sensitivity (the ratio of the number of correctly classified P to total observed P) versus "1 – specificity" (specificity is the ratio between the number of correctly classified N and the total observed N). A model's predictive performance is high if high sensitivity is obtained at low values of "1 – specificity", that is, good capacity for correctly classifying P with a low number of false positives. This yields a curve closer to the upper left-hand corner (Garosi et al., 2019). The Area under the

ROC curve (AUC) quantifies this relationship, so that a model is considered acceptable if AUC ≥ 0.7, excellent if AUC ≥ 0.8 and outstanding if AUC ≥ 0.9.

2.5.3. Regularization of CATREG

Regularization is used for selecting the model and avoiding overfitting in predictive techniques, since estimation of the regression coefficients by least squares may present collinearity (James et al., 2013). It is especially useful in categorical regression (Meulman et al., 2019). Of the three main regularization techniques, Ridge, Lasso and Elastic Net, Lasso is one of the most widely employed for optimal scaling regression. As first developed by Tibshirani (1996), Lasso regularization can deal with complex models with many predictors and high multicollinearity, when ordinary least squares show instability and overfitting. Both effects (instability and overfitting) are common in optimal scaling regression and can make it difficult to select a satisfactory theoretical model (Hartmann et al., 2009).

Lasso applies a λ penalty to the CATREG loss function that reduces the estimated regression coefficients, shrinking them to zero as the penalty increases (van der Kooij, 2007) in Equation 7.

$$\sigma^{lasso}(\boldsymbol{\beta}) = \left\| \boldsymbol{G}_y \boldsymbol{y}_r - \sum\nolimits_{j=1}^{p} \beta_j \boldsymbol{G}_j \boldsymbol{y}_j \right\|^2 + \lambda \sum_{j=1}^{p} |\beta_j| \qquad (7)$$

Predictors with the most stable estimate of the coefficient shrink to zero more slowly, so Lasso regularization can be used to advantage for exploratory analysis instead of stepwise procedures to obtain a set of predictors with low multicollinearity. The advantage of Lasso over alternatives like Ridge regularization is that it shrinks the coefficients to zero, which the others do not. By cancelling coefficients, model interpretation (coefficient selection) becomes straightforward. Lastly, the Elastic Net regularization combines the ridge penalty rule and Lasso, but involves a complicated calculation, and its interpretation is not as immediate as Lasso. For details of the mathematics of these three regularization techniques in optimal scaling regression, see Meulman et al. (2019).

Since each λ penalty involves a regression model, the model that the regularized regression coefficients will be based on must be selected. We followed the procedure specified in Meulman et al. (2019) for this, selecting the most parsimonious model with the lowest prediction error. As an increase in the penalty usually yields lower prediction errors, the best model would be the one in which most of the coefficients are cancelled out, that is, when the sum of the standardized regression coefficients is very close to zero. However, a local minimum might be selected, such that it keeps a suitable number of coefficients, enabling an adequate theoretical model to be established. The one standard error (*1-SE*) rule can be applied for this (James et al., 2013). We chose the regularized model with the lowest number of predictors (the more parsimonious model) within one standard error of the model with the lowest EPE (the optimal model). All statistical analyses, including the regularized regression, were performed with IBM SPSS 24 (IBM Corp., 2016).

3. Results

3.1. Morphological soil properties

Tables 2 and 3 show the values of morphological soil properties: texture, colour, structure, consistence, stickiness, plasticity, biological activity, roots, and reaction to HCl of representative profiles on the high and low productivity banana sites, respectively. The characteristics of high (Mollic Ustifluvents) and low (Fluventic Haplustolls) productivity sites on Plantation PL are similar, except for texture: High productivity soils are silty loam (Table 2), while low productivity soils are generally silty clay loam (Table 3), with the same very dark greyish brown colour (2.5 Y 3/2), and a blocky subangular structure, with moderate-to-medium particle size. At both sites vegetation growth is limited by the high calcium carbonate content from its parent material, which originated in Lake Valencia. At Plantation SM, high productivity soils (Fluventic Haplustepts) usually have high biological activity, and a moderate reaction to HCl, with weak structural stability and lithological discontinuity (Tables 2 and 3). However, low productivity soils (Fluventic Haplustolls) were limited by their susceptibility to sealing and compaction due to the high proportion of silt and fine sand (100 and 200 µm) and very fine sand (50 and 100 µm) (Table 3).

The high productivity soils (Fluventic Haplustolls) at Plantation PZ have geomorphological characteristics favouring high biological activity (Table 2). The low productivity (Fluventic Haplustepts) sites show an abrupt textural change coupled with heavier soil texture, which is a significant limitation for banana productivity (Table 3). At Plantation CH, friable consistence along with weakly adhesive and plastic characteristics favour high biological activity and the common presence of fine roots at both productivity sites (Table 2). In the low productivity soils, characteristics are related to the sharp textural change in depth and poor structure. Soils at both sites are classified as Fluventic Haplustolls (Table 3).

At Plantation BA in the State of Trujillo, the high productivity sites have dark greyish brown (2.5 Y 4/2) silty clay soils (Oxyaquic Ustifluvents). The consistence is firm when moist, adhesive and plastic with high biological activity, no reaction to HCl and abundant roots (Table 2). On the other hand, low productivity soils (Typic Ustifluvents) on this plantation also have certain limitations associated with poor drainage and the presence of a water table. They are generally unstructured with abrupt textural changes, and few very fine roots (Table 3). At Plantation KA, the high and low productivity soils are silty clay loam, predominantly Entisols (Oxyaquic Ustifluvents) (Table 2), while in lower productivity soils the limitations are associated with poor soil structure (Table 3).

Table 2. Values of morphological soil properties in the A horizons of a representative profile of each high productivity level of plantations: texture, color, structure, consistence when moist and dry, stickiness and plasticity.

Plantations	Horizon†	Depth (cm)	Texture	Color	Structure	Consistence							
						When Dry	When moist	Stickiness	Plasticity	Biological activity	Root abundance	Root size	Reaction to HCl
PL	Ap	0-20	sil	2,5 Y 3/2	sbk/m/gm	dsh	mfr	wss	wps	ah	raf	rsm	rhs
SM	Ap	0-17	sil	2,5 Y 4/2	sbk/f/gw	dsh	mvfr	wss	wps	ah	ra	rsm	rhm
PZ	Ap	0-22	ls	2,5 Y 3/2	sbk/f/gw	dsh	mvfr	wso	wpo	ah	ram	rsvf	rhno
PZ	A1	22-44	ls	2,5 Y 4/2	sbk/f/gw	ds	mvfr	wss	wpo	am	raf	rsvf	rhno
CH	Ap	0-18	sil	2,5 Y 3/2	abk/m/gw	dsh	mfr	wss	wps	ah	ram	rsf	rhno
CH	A1	18-38	sil	2,5 Y 4/2	abk/m/gm	dsh	mfr	wss	wps	am	raf	rsm	rhw
BA	Ap	0-18	sic	2,5 Y 4/2	abk/m/gm	dsh	mfi	ws	wp	ah	ram	rsm	rhno
BA	AC	18-40	sil	2,5 Y 4/4	m/ns/ng	ds	mfr	wss	wps	am	raf	rsm	rhno
KA	Ap	0-24	sicl	2,5 Y 3/2	sbk/m/gm	dsh	mfi	ws	wp	ah	ram	rsm	rhw
KA	A/C	24-42	sicl	2,5 Y 4/2	m/ns/ng	ds	mfr	ws	wp	am	raf	rsf	rhw

† Soil Survey Staff (2014). Abbreviations: **Texture:** ls= sandy loam; cl= clay loam; s= sand; L= loam; lvfs= very fine sandy loam; sic= silty clay; sicl= silty clay loam; sil= silty loam; sc= sandy clay. **Structure: Size:** ns=Structureless**;** vf= very fine; f=fine; m=medium; c=coarse. **Grade:** ng= structureless; gw= weak; gm= moderate; gs= strong. **Type:** m= massive**;** gr= granular; sbk= subangular blocky; abk= angular blocky; pr= prismatic. **Consistence: when dry:** dl=loose; ds=soft; dsh=slightly hard; dh=hard. **When moist:** mvfr=very friable; mfr=friable; mfi=firm. **Stickiness (consistence when wet):** wso=non-sticky; wss=slightly sticky; ws=stick. **Plasticity (consistency when wet):** wpo=non-plastic; wps=slightly plastic; wp=plastic. **Root abundance**: raf= few roots; ram: many roots. **Root size**: rsvf: very fine; rsf: fine; rsm: medium. **Biological activity**: ah= high activity; am=medium activity. **Reaction to HCl**: rhno= no reaction; rhw= weak reaction; rhm= moderate reaction; rhs= strong reaction.

Table 3. Values of morphological soil properties in the A horizons of a representative profile of each low productivity level of plantations: texture, color, structure, consistence when moist and dry, stickiness and plasticity.

Site	Horizon†	Depth (cm)	Texture	Color	Structure	Consistence							
						When Dry	When moist	Stickiness	plasticity	Biological activity	Root abundance	Root size	Reaction to HCl
PL	Ap	0-28	sicl	2,5 Y 3/2	sbk/m/gm	dh	mfi	wss	wps	ah	raf	rsf	rhs
SM	Ap	0 - 18	sicl	2,5 Y 3/2	abk/m/gm	dsh	mfi	ws	wp	ah	raf	rsm	rhm
SM	A1	18-44	sicl	2,5 Y 4/2	sbk/m/gm	dsh	mfi	ws	wp	am	raf	rsm	rhm
PZ	Ap	0-18	sicl	2,5 Y 3/2	abk/m/gm	dsh	mfi	wss	wps	am	raf	rsf	rhno
CH	Ap	0-20	sil	2,5 Y 3/2	sbk/m/gw	dsh	mfr	wss	wps	am	ram	rsf	rhno
CH	A1	20-42	sil	2,5 Y 4/2	abk/m/gw	dsh	mfr	wss	wp	al	raf	rsm	rhno
BA	Ap	0-14	sicl	2,5 Y 4/2	abk/m/gm	dsh	mfi	ws	wp	am	raf	rsvf	rhno
KA	Ap	0-22	sicl	2,5 Y 4/2	sbk/m/gs	dsh	mvfi	ws	wp	ah	ram	rsm	rhno
KA	A/C	22-42	sicl	2,5 Y 4/4	sbk/f/gm	dsh	mfi	ws	wp	al	raf	rsf	rhw

† **S**oil Survey Staff (2014). Abbreviations: **Texture:** Ls= sandy loam; cl= clay loam; s= sand; L= loam; lvfs= very fine sandy loam; sic= silty clay; sicl= silty clay loam; sil= silty loam; sc= sandy clay. **Structure: Size:** ns=Structureless**;** vf= very fine; f=fine; m=medium; c=coarse. **Grade:** ng= structureless; gw= weak; gm= moderate; gs= strong. **Type:** m= massive**;** gr= granular; sbk= subangular blocky; abk= angular blocky; pr= prismatic. **Consistence: when dry:** dsh=slightly hard; dh=hard. **When moist:** mvfr=very friable; mfr=friable; mfi=firm; mvfi= very firm. **Stickiness (consistence when wet):** wss=slightly sticky; ws=sticky. **Plasticity (consistency when wet):** wps=slightly plastic; wp=plastic. **Root abundance**: nra= no root; raf= few; ram: many. **Root size**: nra= no root; rsvf: very fine; rsf: fine; rsm: medium. **Biological activity**: ah= high activity; am=medium activity; al= low activity. **Reaction to HCl**: rhno= no reaction; rhw= weak reaction; rhm= moderate reaction; rhs= strong reaction.

3.2. Model selection

From the original set of 15 predictors, we obtained a significant model with a high coefficient of determination, $R^2 = 0.746$ (p <0.007). However, as most of the regression coefficients in the that first model was not significant (Table 4), it had to be refined. The coefficients of texture, structure type, dry consistence, biological activity and reaction to HCl were significant ($p < 0.05$). These variables should be considered in a stepwise procedure since they respond to soil processes related to plant physiology and eventual yield. Therefore, to fit a better model with significant coefficients while keeping R^2 as high as possible, we checked the MSE as a measure of the expected prediction error (EPE) given by alternative models in which predictors were progressively removed (Figure 1). Thus, the optimum model included five predictors (Figure 1). Table 5 shows the expected prediction errors in the nonregularized and regularized models. Considering the average EPEs of both bootstrap and cross-validation estimations, the most accurate models seem to be Elastic Net (average EPE of 0.676) and Lasso (average EPE of 0.677), because they had the lowest EPE. We used the Lasso model to select the optimal set of predictors since it is simpler and easier to interpret, and in addition, there are no appreciable differences in accuracy from the Elastic Net model.

Table 4. Regression coefficients, correlation with the outcome and tolerance of the model from the initial set of fifteen predictors.

	Coefficients					Correlation with the outcome		Tolerance	
Predictors	Beta	B*	gl	*F*	*p-value*	*r*	partial *r*	after	before
Texture	0.397	0.224	5	3.127	**0.023**	0.118	0.556	0.723	0.704
Moist hue	0.193	0.195	2	0.980	0.388	-0.017	0.275	0.557	0.617
Moist value	0.175	0.203	1	0.742	0.396	0.049	0.272	0.663	0.641
Moist chroma	0.245	0.277	2	0.780	0.468	-0.092	0.298	0.414	0.411
Structure type	0.409	0.216	3	3.598	**0.026**	0.275	0.575	0.748	0.263
Structure size	-0.005	0.397	1	0.000	0.989	-0.251	-0.007	0.397	0.148
Structure grade	0.281	0.547	1	0.264	0.612	-0.298	0.299	0.316	0.151
Dry consistence	-0.612	0.234	2	6.829	**0.004**	-0.297	-0.645	0.483	0.628
Moist consistence	-0.515	0.349	3	2.178	0.113	-0.253	-0.508	0.334	0.312
Stickiness	-0.609	0.462	2	1.735	0.195	-0.193	-0.469	0.193	0.189
Plasticity	0.430	0.383	2	1.256	0.300	-0.186	0.334	0.172	0.189
Biological activity	0.792	0.311	2	6.480	**0.005**	0.273	0.688	0.364	0.221
Root abundance	-0.112	0.452	1	0.061	0.807	0.172	-0.137	0.392	0.192
Root size	0.052	0.409	3	0.016	0.997	0.119	0.072	0.493	0.370
Reaction to HCl	-0.532	0.240	2	4.898	**0.015**	-0.252	-0.645	0.638	0.398

* Bootstrap SE estimates of Beta coefficients (1,000 boostrap resamplings)

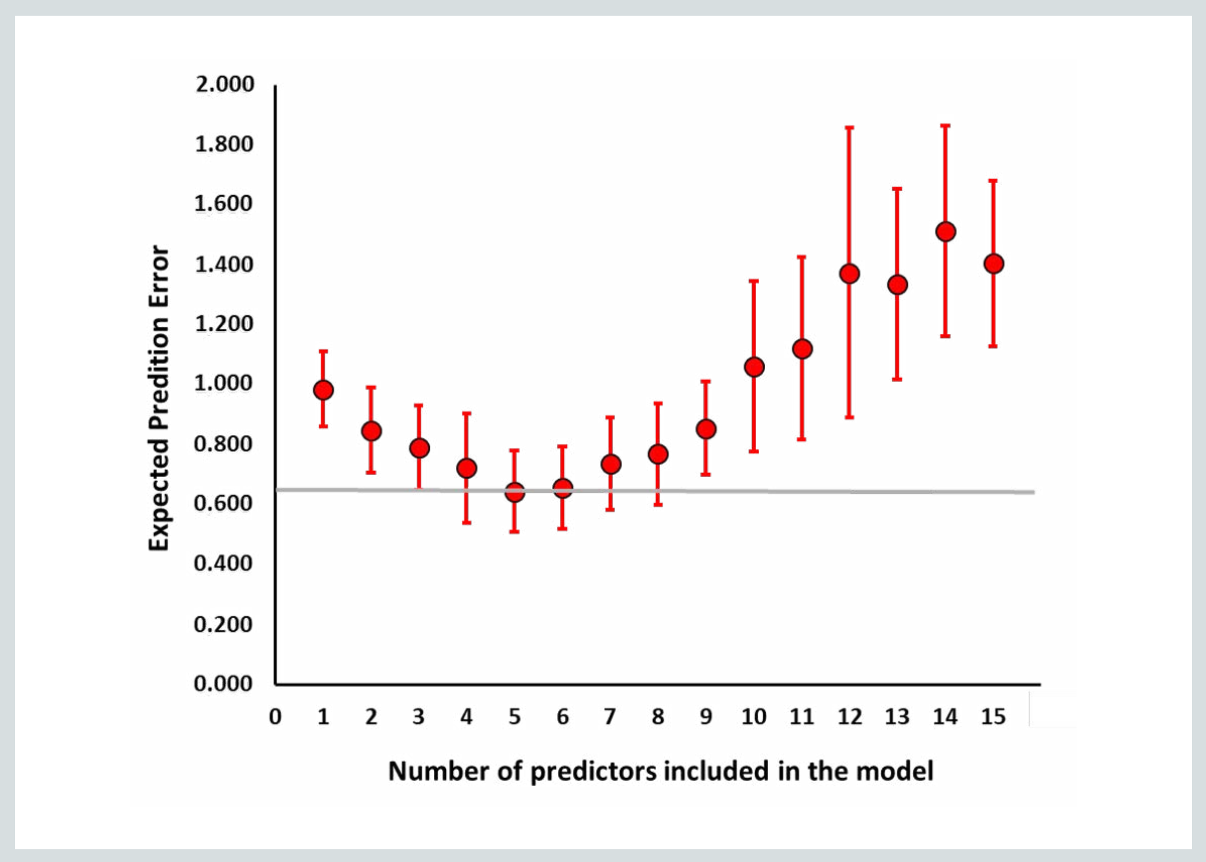


Figure 1. Expected Prediction Error (EPE) of 15 different no-regularized models, employing the 0.630 Bootstrapping procedure (standard errors are showed by the vertical bars. The grey line shows the lowest EPE).

Table 5. Expected Prediction Error (EPE) and Standard Error (SE) of regularized CATREG models for the initial set of fifteen predictors.

	0.632 Bootstrapping		10-fold Cross Validation	
Regularization	**EPE**	**SE**	**EPE**	**SE**
No regularized	1.405	0.277	1.398	0.334
Ridge	0.792	0.125	0.836	0.116
Lasso	0.861	0.104	0.493	0.087
Elastic Net	0.859	0.104	0.493	0.087

The Lasso paths (Figure 2) drawn represent the 42 regularized models performed by increasing λ (Equation 7) by 0.02 per step, from a standardized sum of coefficients of 1 (unshrunk coefficients, right side of the graph) to 0 (left side of the graph). The variables with the regression coefficients that are shrunken earliest, lowering their penalized coefficients to zero, should be taken as those with the lowest predictive power, regardless of their starting value (right of the plot), so they can be removed from the model. On the contrary, coefficients that keep non-zero values at 1-SE of the regularized optimal model may be left in the theoretical model. Our optimum model included eight predictors with a λ of 0.260, an EPE of 0.789 and a SE of 0.122, while the most parsimonious, with an EPE of 0.861 (λ of 0.400) was fitted using seven predictors. In this case, there was no important difference in simplicity (parsimony) between the two regularized models. Only the standardized coefficient of root abundance was reduced to zero in the most parsimonious model with respect to the optimum. The eight predictors for the optimum model were: 1) biological activity, 2) dry consistency, 3) texture, 4) structure grade, 5) structure type, 6) HCl reaction, 7) stickiness, and 8) root abundance. From this starting set of predictors (Model 4 in Table 6), alternative nonregularized models of seven (Model 3), six (Model 2) and five (Model 1) predictors were tested in a backward stepwise regression to find the statistically significant model with the highest R^2 and minimum prediction error (Table 6). Model 1 was finally selected, because it met all these conditions: statistical significance of R^2 ($p < 0.0001$) and of all its coefficients ($p < 0.050$), while yielding the minimum bootstrapped EPE (0.645, see also Figure 1) and a good enough R^2 (0.599) close to the model with the highest determination coefficient (Model 4, $R^2 = 0.645$).

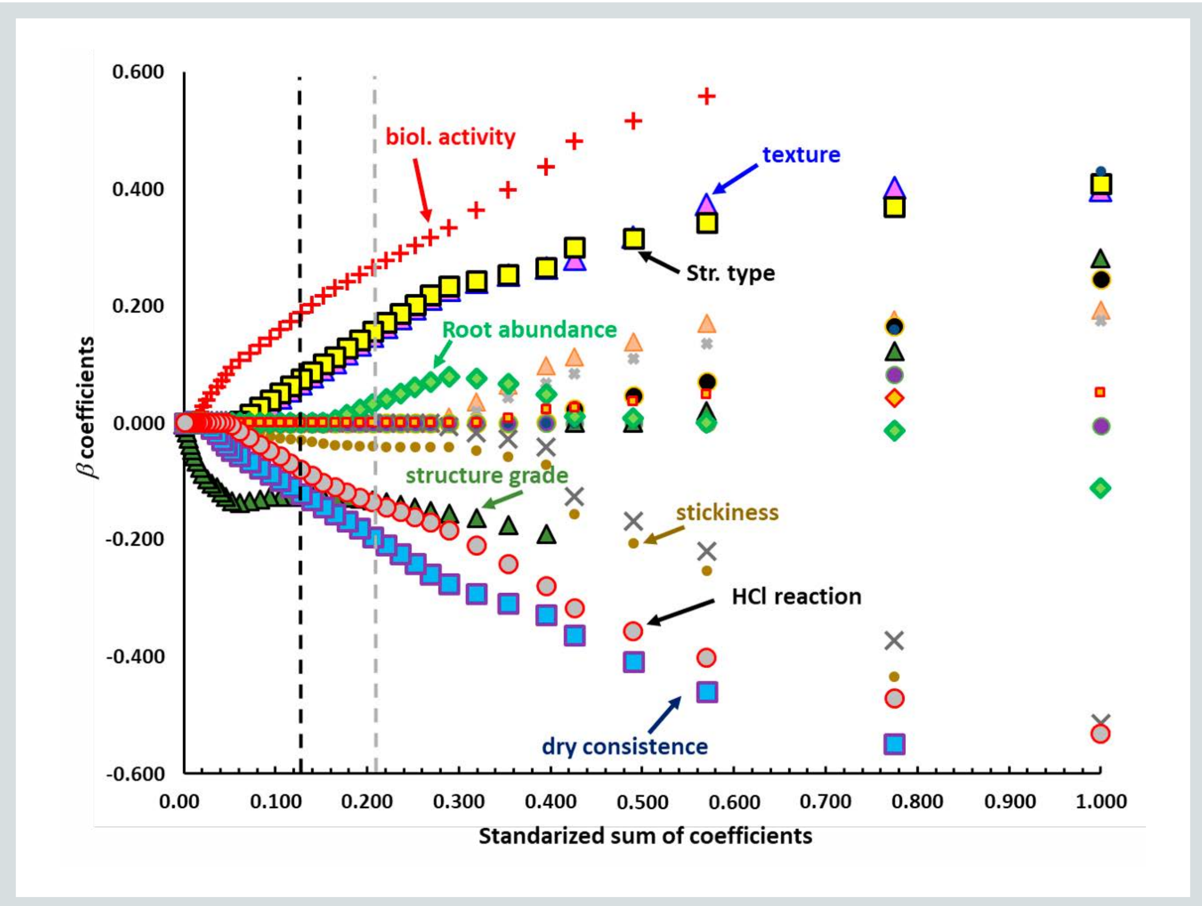


Figure 2. Categorical regression (CATREG)-Lasso coefficients path estimated with the 0.632 bootstrap (dashed bar shows the most parsimonious –black vertical line – and the optimal –grey vertical line– models within one standard error). The coefficients in the right side of the plot (standardized sum of coefficients of 1.000) are the same of that in table 4.

Table 6. Selection of the optimal CATREG model.

Model	number of predictors	R^2	*p*-value	EPE (SE) [1] Bootstrapped	EPE (SE) [1] 10-fold Cross validation	Predictors (*p*-value for regression coefficients)[2]
#1	5	0.599	<0.0001	0.645 (0.135)	0.662 (0.236)	Bio. Act. (<0.0001), Texture (<0.0001), Dry cons. (0.001), Str. Type (0.000), HCl (0.017)
#2	6	0.633	<0.0001	0.656 (0.145)	0.651 (0.134)	Bio. Act. (<0.0001), Texture (0.001), Dry cons. (0.001), Str. Type (0.001), HCl (0.013), Str. Grade (0.065)
#3	7	0.636	<0.0001	0.738 (0.115)	0.685 (0.139)	Bio. Act. (<0.0001), Texture (0.008), Dry cons. (0.001), Str. Type (0.003), HCl (0.013), Str. Grade (0.449), Stickiness (0.838)
#4	8	0.645	<0.0001	0.770 (0.169)	0.778 (0.153)	Bio. Act. (0.044), Texture (0.006), Dry cons. (0.001), Str. Type (0.002), HCl (0.182), Stickiness (0.818), Str. Grade (0.478), Root abundance (0.657)

[1] EPE: expected prediction error; SE: standard error (in bracket).

[2] Bio. Act.: biological activity; Str. Grade: structure grade; Str. Type: structure type; Dry cons.: dry consistence; HCl: reaction to HCl.

Table 7. Regression coefficients, correlation with the outcome and Tolerance of the optimal CATREG model.

5 predictors (Model #1)	Beta	B*	*p*-value	Correlation with outcome		Tolerance	
				r	partial *r*	after	before
Texture	0.371	0.157	<0.0001	0.221	0.495	0.945	0.963
Structure type	0.463	0.154	<0.0001	0.309	0.571	0.904	0.856
Dry consistence	-0.440	0.151	0.001	-0.232	-0.551	0.902	0.793
Biological activity	0.581	0.108	<0.0001	0.292	0.645	0.849	0.907
Reaction to HCl	-0.336	0.160	0.017	-0.302	-0.463	0.968	0.976

* Bootstrap SE estimate of Beta coefficients (1,000 samples)

The statistical significance of the regression coefficients was the main issue with these alternative models, because they all yielded quite similar prediction errors (within one SE) and significant overall fit ($p < 0.0001$). In Model 4, with eight predictors, elimination of root abundance was supported by the *1-SE* rule, as discussed above, while stickiness and structure grade presented coefficients far from statistical significance ($p = 0.195$ and 0.612, respectively) in the starting model (Table 4), as well as lower tolerances of all predictors (except plasticity). Therefore, Model 1 with five predictors seemed to be the best model for predicting the PI, and it was chosen as the definitive CATREG optimal scaling model (Table 7).

According to the sign of the coefficients, the ordinal variable, biological activity, is positively correlated with the dependent variable, PI, which implies that higher biological activity is correlated with higher productivity, while harder (drier consistence) soils with carbonates (reaction to HCl) are associated with lower productivity. The transformation functions (optimal scaling) indicate the linearity of this relationship, as well as the direction of the variation of the nominal variables, texture and structure (Figure 3).

3.3. Optimal scaling of the field soil morphological variables

Figure 3 shows optimal scaling of the selected categorical regression model. Of the transformation functions, dry consistence and biological activity were practically linear, which implies a proportional increase (biological activity) or decrease (dry consistency) in productivity (PI) with these variables. Two soil productivity groups were formed according to the acid reaction to their carbonate content: 1) no reaction (rhno) and weak reaction (rhwo), and 2) moderate (rhm) or strong (rhs) reaction. As this predictor has a negative correlation (beta coefficient) with PI, moderate or strong reactions must be associated with a decrease in productivity.

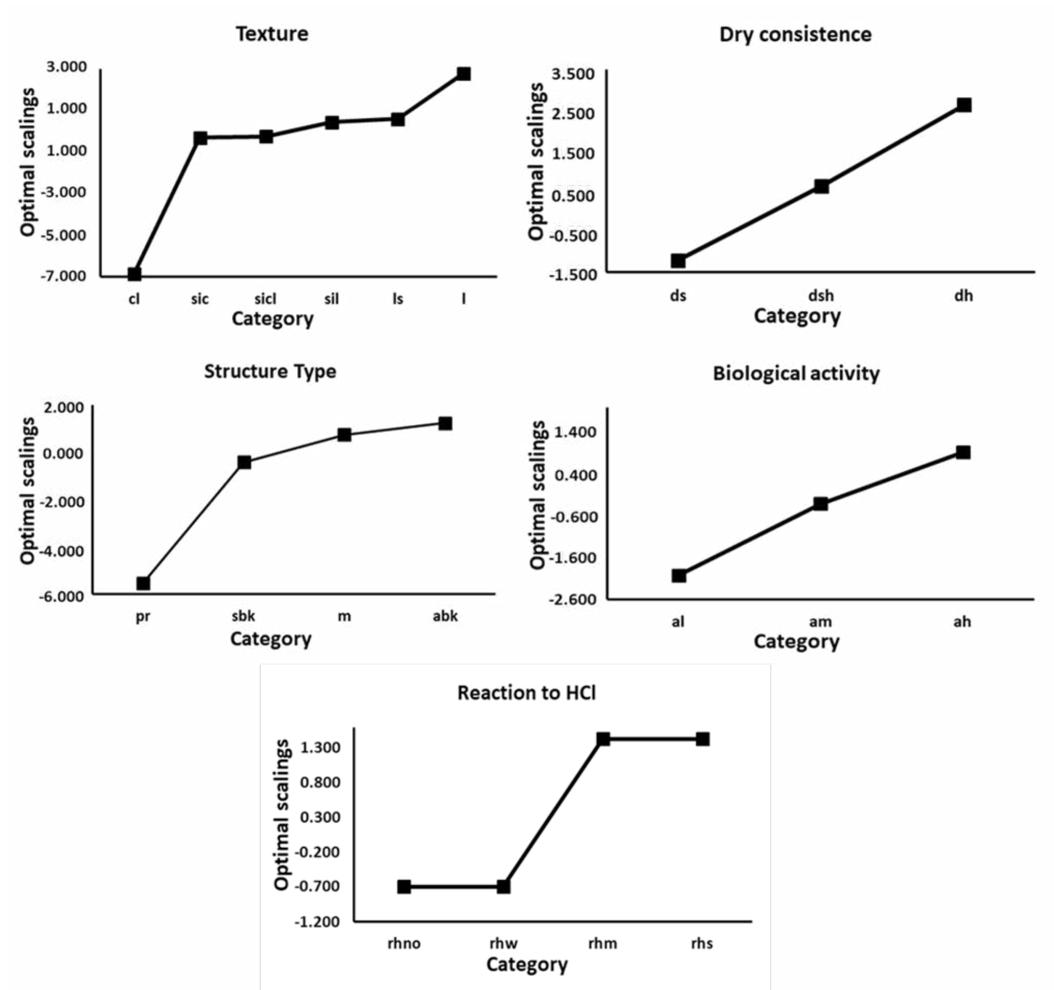


Figure 3. Optimal scaling's of texture, dry consistence, structure type, biological activity, and reaction to HCl. (Abbreviations: Texture: ls= sandy loam; cl= clay loam; l= loam; sic= silty clay; sicl= silty clay loam; sil= silty loam. Dry consistence: ds=soft; dsh=slightly hard; dh=hard. Structure type: m= massive; pr= prismatic; sbk= subangular blocky; abk= angular blocky. Biological activity: al= low activity; am=medium activity; ah= high activity. Reaction to HCl: rhno= no reaction; rhw= weak reaction; rhm= moderate reaction; rhs= strong reaction).

For structure type, categories were ordered (entered in the model as nominal, and therefore, with unrestricted order) as prismatic (pr) <subangular blocks (sbk) <massive (m) <angular blocks (abk). As the beta coefficient was positive, a positive correlation was defined between more developed (angular) structures and productivity, while prismatic structure, "extreme" development of the angular structure, was as unfavorable as massive structure, which could be explained by its characteristic loss of porosity (macro porosity). Texture categories, also entered as nominal, were ordered on a rough gradient from finer clay (cl) to loamy textures (l), and as the beta coefficient was positive, this was the most closely correlated with productivity (PI). (Figure 3).

3.4. Prediction accuracy

Figure 4 shows the prediction plot of observations versus predictions. As mentioned above, the CATREG model with the lowest bootstrap estimate of the prediction error was selected. However, it yielded a relatively high MSE (EPE) of 0.645 (RMSE of 0.803), accounting for 11.7% of the outcome ranking, which varies from -4.178 to 2.680 (X-axis in Figure 4).

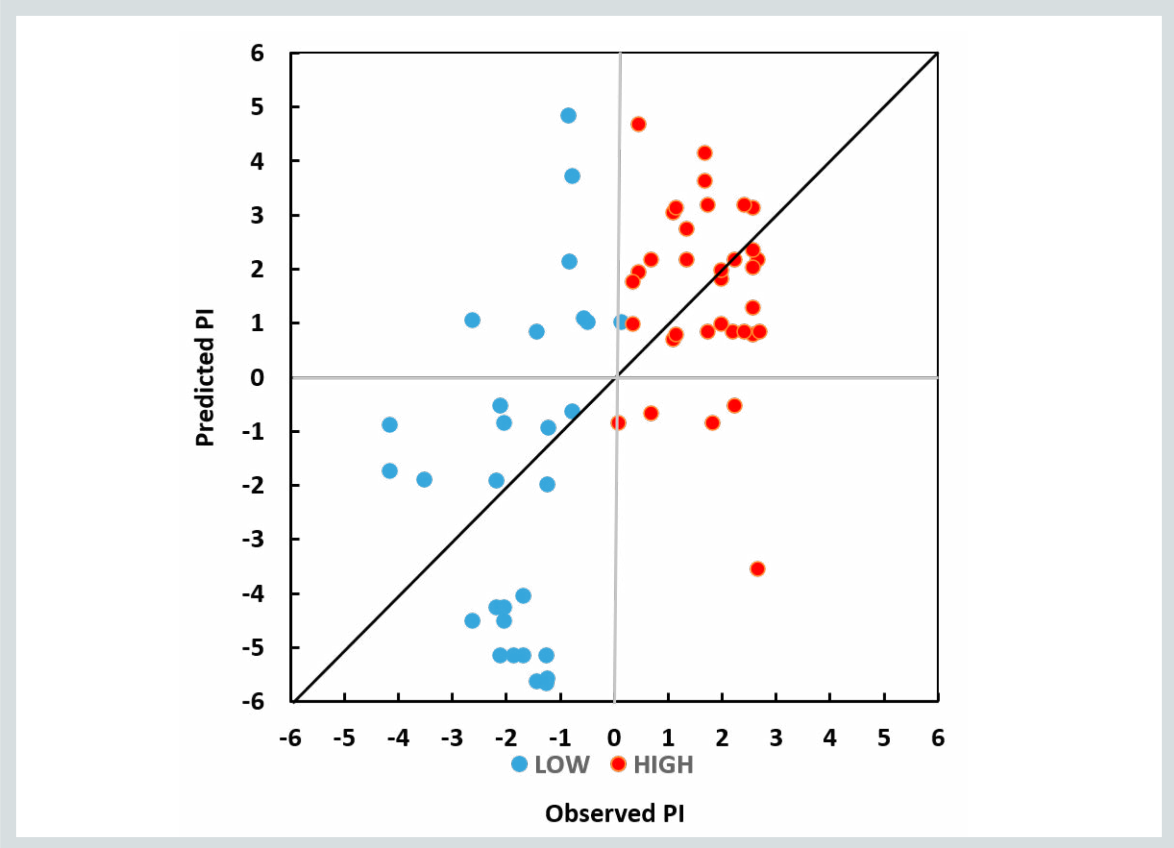


Figure 4. Observed vs. Predicted values of the productivity index (PI) for the five predictors optimal categorical regression model. The colour of each point shows whether the plot belongs to a High (red) or Low (blue) productivity plantation. Grey lines: cut-off value for dependent variable classification in P (PI >0) and N (PI < 0) groups.

For a more general assessment of model performance, the observed and predicted dependent variables were classified into two groups or classes, with a PI of zero as the cut-off point: one level, N, with PI <0 (low PI) and another with PI > 0 (high PI). It should be mentioned that most of the horizons of soils in high productivity sites (Table 1) also had positive predicted PI values. Table 8 shows the confusion matrix of the resulting classification, and various measures of their discrimination accuracy.

Table 8. Confusion matrix and discrimination accuracy of classified values of the outcome variable: Sensitivity, Specificity, Efficiency and Area under the curve ROC (AUC).

		Predicted PI level		Total observed
		P	N	
Observed PI level	P	30	5	35
	N	7	21	28
Total predicted		37	26	63

P: positive PI (> 0), N: negative PI (< 0)
Sensitivity = 0.86, Specificity = 0.75, Efficiency = 0.81, AUC = 0.834

Under these conditions, the model showed high sensitivity of 86%, and misclassified only five cases recognized as positive or high PI (bottom right-hand quadrant in Figure 4). Specificity, although slightly lower (75%), was also satisfactory, as was classification efficiency (81%) and the area under the ROC curve (AUC), which was 0.834, considered excellent. Therefore, although the prediction based on continuous PI values is not as reliable as might be desired, prediction of the level of productivity of the plot from the morphological indicators would be quite suitable. Finally, as deduced from Figure 4, the statistical relationship between predicted PI and *a priori* plot type, High or Low productivity, was also significant: The mean predicted PI for High productivity was 1.579 (SD = 3.018), and mean predicted PI for Low productivity was -1.872 (SD = 1.642), t = -5.530 (p = 0.000).

4. Discussion

The objective of this study was to validate the hypothesis that a quantitative relationship between banana productivity and key soil morphological properties is feasible. This was demonstrated using categorical regression analysis with transformation (Table 7 and Figure 3). Categorical regression with optimal scaling was implemented to find nonlinear transformations (Lasso) to select a sparse model with stable predictors and bootstrap 0.632 to evaluate prediction accuracy. With this approach we identified a subset of five variables that best predicts banana productivity levels in two areas of Venezuela. These variables included texture, soil structure type, dry consistence, biological activity indicators and HCl reaction. The model developed enables biophysical interpretation, clearly related to banana productivity. The categorical regression developed was able to correctly discriminate between areas of high and low productivity on the same plantation, and captured the trend in variation in productivity among plantations as their soil morphological variables changed (Figure 4). The accuracy of this model is in line with the prediction accuracy found by van der Kooij (2007).

An interesting feature of the regression model developed in our study is that, combined with the optimal scaling developed, it can be easily interpreted by banana production technicians in relation to some key soil properties. In addition, the key soil variables used can usually be found in soil profile descriptions, and with little explanation can also be interpreted by growers or other stakeholders. Texture, structure (grade and type) and biological activity are closely related to productivity. Soil texture influences the availability of water and nutrients, as well as aeration, drainage and accessibility in the use of agricultural implements. The model calibrated identified this, with a higher score in productivity for the loamy textures. This is in line with observations at Plantations PL, BA and KA, where the high productivity plots had loamy textures (L/L-SL) and low productivity plots had a moderate to fine texture in the A horizon. In the largest plantations, the high productivity plots were in the alluvial plains (BA, KA, SM, PZ and CH) which had a higher clay content than the lower productivity plots on the same plantations. Overall, this agrees with Vaquero (2005) who concluded that the areas with low banana productivity had soil horizons with coarse textures (very sandy soils) and very fine textures (clayey, with content of clay greater than 60%) due to the direct influence on the water retention capacity and permeability. In this regard, medium-to-fine textured soils (loam to silty loam) with good structure and porosity, developed a deeper, more extensive root system (Vaquero, 2005; Rey et al., 2009; Delgado et al., 2010a). Higher biological activity, an indicator of healthy soil, was also positively correlated to higher productivity as indicated by the regression coefficient and the scaling ranking. This is not only an important technical result, but also for spreading good agricultural practices among banana stakeholders, since it shows a clear link between soils with good biological activity and higher productivity.

The fourth variable which contributed to our predictive model of banana productivity was the type of soil structure (Table 7 and Figure 3). This is not surprising, since it reflects the process of soil formation and anthropogenesis (Hernández et al., 2010). Voorhees et al. (1971) stated that soil structure is a fundamental function in pedogenesis and for plant nutrition, because of its enormous significance in improving fertility and regulating the microbial activity of soils. In our analysis, soils with massive or highly developed prismatic structure had a lower score in productivity. This agrees with the results of Gauggel et al. (2005), who found rapid deterioration of the banana root system in coarse and very coarse blocks and prismatic structures, as is the case of the low productivity soils on the plantations located in the State of Trujillo (BA and KA). It also agrees with the results of Gauggel et al. (2005) and Villarreal-Nuñez et al. (2013), who noted deterioration of the root system where soil with massive clay structures at shallow depths forms barriers. Dry consistency was the most significant variable associated with the PI (Table 7), where consistence was strong in lower productivity soil (Table 7 and Figure 3). In alluvial soils, firm and very firm consistence with weak or no structure can cause compaction, which according to Dorel (1993), lowers banana productivity. The results of Vaquero (2005), who found that banana root density was higher in soils with friable to very friable consistence, low bulk density and low resistance to penetration, also coincide with ours. In our study, root density was drastically reduced in areas with high penetration resistance associated with a firm or very firm soil consistence and bulk density over 1.2 g cm^{-3} (Data not shown).

High carbonate content, indicated by the reaction to HCl, the fifth parameter in our model (Table 7 and Figure 3), can also decrease banana productivity (Cigales & Pérez, 2011). This is associated with the limited response to fertilization of very calcareous soils, which can even prevent bunch development and reduce the size of the pseudo stem and plant height, and sometimes facilitate the appearance of diseases (Olivares et al., 2020). Phosphorus, iron, zinc, and nitrogen deficiencies can be explained by excessive presence of carbonates. When carbonate accumulates at a certain depth in the soil profile, the apical bud can die, even after normal initial development (Vanilarasu and Balakrishnamurthy, 2014; El-Khawaga, 2013).

Overall, our study has shown how the use of categorical regression analysis with optimal scaling can deliver an operating model able to incorporate the effect of qualitative soil information into banana productivity. When properly scaled to other soil types and farms, it has the potential of being a useful tool for farmers, technicians or investors for identifying the best areas for banana plantations. It can also contribute to independent management in different areas within the same plantation based on relatively easily acquired soil information. One of the major advantages of this model is that it is based on relatively simple field evaluations at a moderate cost, and soil information from field surveys carried out previously for other purposes can be used.

5. Conclusions

The five morphological properties of the soil (soil texture, soil structure type, dry consistence, biological activity and HCl reaction) in our empirical categorical regression model have a clear agronomic relationship with banana productivity. The proposed model could be used in the field for reliable identification of areas of high and low potential banana productivity in other banana growing areas such as, the states of Barinas, Sucre and Zulia in Venezuela after local assessment. Identification of the main soil morphological properties associated with banana productivity by applying categorical regression can contribute to the long-term sustainability of banana soils in Venezuela, and other tropical areas.

Our results suggest the potential for further studies of quantitative transformation of soil morphological properties and application of categorical regression, as carried out in this study, with different levels of banana productivity This methodology can be easily applied to other crops, requiring little or no expert knowledge.

This study can serve as an example of a relatively straight forward way to quantitatively assess the effect of soil properties on banana productivity using information generated from soil surveys which are relatively inexpensive and often already available for other reasons. Calibrating similar correlations between banana productivity indicators, or even actual yield records, with soil morphological properties in specific areas, using the methodology proposed in this manuscript could be done in a few months at relatively moderate cost, which would be compensated by the savings in planting the banana plots in the most suitable areas.

Acknowledgements

The authors recognize the financial support for international mobility of the Ibero-American Secretary General with Fundación Carolina and Action KA107 of Erasmus+ Program from Agrifood Campus of International Excellence (ceiA3) (2020). Also, by project "Technological innovations for the management and improvement of the quality and health of banana soils in Latin America and the Caribbean" financed by FONTAGRO and coordinated by Bioversity International (before INIBAP) and project SHui (European Commission Grant Agreement number: 773903).

References

Andrews, S.S., Karlen, D.L., Cambardella, C.A., 2004. The soil management assessment framework: A quantitative soil quality evaluation method. Soil. Sci. Soc. Am. J. 68, 1945-1962. https://doi.org/10.2136/sssaj2004.1945

Calero, J., Serrano J.M., Aranda V., Sánchez D., Vila M.A., Delgado G. 2005. Analysis and characterization of olive tree cultivation system in Granada province (South of Spain) with optimal scaling and multivariate techniques. Agrochimica, 49, 118-131. Available online at https://n9.cl/k3rdk. Last accessed 07 July 2020.

Calero, J., Aranda, V., Montejo-Raez, A., Martin-Garcia, J.M., 2018. A new soil quality index based on morpho-pedological indicators as a site-specific web service applied to olive groves in the Province of Jaen (South Spain). Comput Electron Agric, 146, 66-76. https://doi.org/10.1016/j.compag.2018.01.016

Calero, J., Delgado, R., Delgado, G., Manuel Martin-Garcia, J., 2008. Transformation of categorical field soil morphological properties into numerical properties for the study of chronosequences. Geoderma, 145, 278-287. https://doi.org/10.1016/j.geoderma.2008.03.022

Cigales, M., Pérez, O., 2011. Soil variability and water requirement of banana cultivation in a locality in the Pacific of Mexico. Av. Investig. Agropecu, *15*, 21-31. Available online at https://n9.cl/hxjtg. Last accessed 07 July 2020.

Delgado, E. Trejos, J, Villalobos, M, Martinez, G., Lobo, D., Rey, J., Rodríguez, G., Rosales, F. and Pocasangre, L. 2010b. Determination of a soil quality and health index for banana plantations in Venezuela. Interciencia, 35, 927-933. Available online at https://n9.cl/5ly9u. Last accessed 07 February 2021.

Delgado, E., Rosales, F., Trejos, J., Villalobos, M., Pocasangre, L., 2010a. Soil quality and health index for banana plantations in four countries of Latin America and the Caribbean. Bioagro, 22, 53-60. Available online at https://n9.cl/qm3e. Last accessed 07 February 2021.

Dorel, M., 1993. Banana development in an andosol in Guadeloupe: effect of soil compaction. Fruits, 48, 83-88. Available online at https://n9.cl/vahwkz. Last accessed 07 February 2021.

Efron, B., 1983. Estimating the error rate of a prediction rule: Improvement on cross-validation. J. Amer. Statist. Assoc, 78, 316–331. https://doi.org/10.2307/2288636

El-Khawaga, A. S., 2013. Response of Grand Naine banana plants grown under different soil moisture levels to antitranspirants application. Asian J. Crop Sci, 5, 238-250. https://doi.org/10.3923/ajcs.2013.238.250

FAO, 2006. Guidelines for soil profile description. Fourth edition. FAO, Roma.

FAO, 2020. Banana Market Review and Banana Statistics 2018-20. Rome. Available online at http://www.fao.org/economic/est/est-commodities/bananas/en/. Last accessed 07 February 2021.

Garosi, Y., Sheklabadi, M., Conoscenti, C., Pourghasemi, H.R., Van Oost, K., 2019. Assessing the performance of GIS- based machine learning models with different accuracy measures for determining susceptibility to gully erosion. Sci. Total Environ, 664, 1117–1132. https://doi.org/10.1016/j.scitotenv.2019.02.093

Gauggel, C.A., Sierra, F., Arévalo, G., 2005. The problem of banana root deterioration and its impact on production: Latin America’s experience. In Turner, D.W. and Rosales, F.E. (Eds) Banana root system: towards a better understanding for its productive management. Proceedings of an International Symposium held in San José, Costa Rica on 3-5 November 2003. International Network for the Improvement of Banana and Plantain (INIBAP), Montpellier, pp. 13-22. Available online at https://n9.cl/b3wi8. Last accessed 07 February 2021.

Gifi, A., 1990. Nonlinear Multivariate Analysis, Wiley, Chichester.

González García, H., González Pedraza, A. F., Atencio, J., & Soto, A. 2021a. Evaluation of quality of banana soils through microbial activity in the south the lake of Maracaibo, Zulia state, Venezuela. Rev. Fac. Agron. (LUZ), 38(2), 216-240. Available online at https://n9.cl/wdnr5. Last accessed 19 February 2021.

González García, H. G., Pedraza, A. F. G., Yzquierdo, G. R., Pacheco, R. L., Vásquez, M. B. 2021b. Vigor of plantain plants (*Musa* AAB cv. Harton) and its relationship with physical, chemical and biological characteristics of the soil. Agron. Costarricense, 45(2), 115-134. https://doi.org/10.15517/rac.v45i2.47772

González García, H. G., González, A. F., Pineda, M., Escalante, H., Yzquierdo, G. A. R., Bracho, A. S. 2021c. Edaphic microbiota in plantain lots of contrasting vigor and relationships with soil properties. Bioagro, 33(2), 143-148. https://doi.org/10.51372/bioagro332.8

González-Pedraza, A.F. Atencio, J., Cubillán, K., Almendrales, R., Ramírez, L., Barrios, O. 2014. Microbial activity in soils cultivated with plantain (Musa AAB plantain subgroup cv. Harton) with different vigor of plants. Rev. Fac. Agron. (LUZ), 31, 526-538. https://doi.org/10.51372/bioagro332.8

González-Pedraza, A. F., Carlos Escalante, J. 2021. Nitrogen mineralization in soils cultivated with plantain (*Musa* AAB Subgroup plátano cv. Hartón), Zulia state, Venezuela. Rev. Fac. Agron. (LUZ), 38(3), 525-547. Available online at https://n9.cl/so6u. Last accessed 19 July 2021.

Hair, J. F., Anderson, R. E., Tatham, R. L., BlLack, W. C., 1999. Análisis Multivariante. 5° ed. Prentice. Hall, Madrid. España

Hartmann, A., Zeeck, A., van der Kooij A, J., 2009. Severity of Bulimia Nervosa. Measurement and classification into health or pathology. Psychopathology, 42, 22-31. https://doi.org/10.1159/000173700

Hernández, A. Bojorquez, J., Morell, F., Cabrera, A., Ascanio, M., Garcia, J., Madueño, A., Najera, O. 2010. Fundamentos de la estructura de suelos tropicales. Universidad Autónoma de Nayarit. Mexico.

Hernández, Y., Marín, M., García, J., 2007. Response to the plantain (*Musa* AAB cv. Horn) yield as a function of the mineral nutrients and its phenologycal cycle. Part I. Growth and Production. Rev. Fac. Agron. (LUZ), 24, 607-626. Available online at https://n9.cl/uwuxm. Last accessed 19 February 2021.

IBM Corp. Released 2016. IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp.

James, G., D. Witten, T. Hastie, and R. Tibshirani, 2013. An Introduction to Statistical Learning. Springer, vol. 112. https://doi.org/10.1007/978-1-4614-7138-7

Lal, R., 1998. Soil quality and agricultural sustainability. In: Lal, R. (Ed.), Soil quality and agricultural sustainability. Sleeping Bear Press, Inc, Chelsea, MI, pp. 3-12.

Linting, M., Meulman, J.J., Groenen, P.J.F., van der Kooij, A.J., 2007. Nonlinear principal components analysis: Introduction and application. Psychological Methods, 12, 336-358. https://doi.org/10.1037/1082-989X.12.3.336

MacEwan, R. J., Fitzpatrick, R. W., 1996. The Pedological Context for Assessment of Soil Quality. In: MacEwan, R. J., Carter, M. R. (Eds.), Soil quality is in the hands of the land manager. Proceedings of an international symposium. Advances in soil quality for land management: Science, Practice and Policy, 17-19 April 1996, University of Ballarat, Ballarat, Victoria, Australia, pp. 10–16.

Martínez, G., Delgado, E., Rodríguez, D., Hernández, J., Del Valle, R., 2008. Brief Analysis of Musaceae Production in Venezuela. Prod. Agrop, 1, 24-29. Available online at https://n9.cl/d48mt. Last accessed 19 February 2021.

Meulman, J.J., van der Kooij, A.J., Babinec, A., 2002. New features of categorical principal components analysis for complicated data sets, including data mining. in: Gaul, W., Ritter, G. (Eds.), Classification, Automation, and New Media, pp. 207-217. https://doi.org/10.1007/978-3-642-55991-4_22

Meulman, J.J., van der Kooij, A.J., Duisters, K.L.W., 2019. ROS Regression: Integrating Regularization with Optimal Scaling Regression. Statistical Science, 34, 361-390. https://doi.org/10.1214/19-STS697

Munsell Color Company, 1999. Munsell Soil Color Charts. Munsell Color, Macbeth Division of Kollmorgen, Maryland, USA.

Olivares, B., 2018. Tropical conditions of seasonal rain in the dry-land agriculture of Carabobo, Venezuela. Lgr, 27,86-102. http://doi.org/10.17163/lgr.n27.2018.07

Olivares, B., Cortez, A., Parra, R., Lobo, D., Rodríguez, M.F., Rey, J.C., 2017. Evaluation of agricultural vulnerability to drought weather in different locations of Venezuela. Rev. Fac. Agron. (LUZ), 34,103-129. Available online at https://n9.cl/hc5xs. Last accessed 19 February 2021.

Olivares, B., Paredes, F., Rey, J., Lobo, D., Galvis-Causil, S., 2021. The relationship between the normalized difference vegetation index, rainfall, and potential evapotranspiration in a banana plantation of Venezuela. STJSSA, 18, 58-64. http://dx.doi.org/10.20961/stjssa.v18i1.50379

Olivares, B.O., Araya-Alman, M., Acevedo-Opazo. C., Rey, J.C., Cañete, P., Giannini, F., Balzarini, M., Lobo, D., Navas-Cortés, J.A., Landa, B.B., Gómez, J.A., 2020. Relationship between soil properties and banana productivity in the two main cultivation areas in Venezuela. J. Soil Sci. Plant Nutr, 20, 2512-2524. https://doi.org/10.1007/s42729-020-00317-8

Pulido-Moncada, M., Helwig Penning, L., Timm, L.C., Gabriels, D., Cornelis, W.M., 2014. Visual examinations and soil physical and hydraulic properties for assessing soil structural quality of soils with contrasting textures and land uses. Soil Tillage Res, 140, 20-28. https://doi.org/10.1016/j.still.2014.02.009

Pulido-Moncada, M., Penning, L.H., Timm, L.C., Gabriels, D., Cornelis, W.M., 2017. Visual examination of changes in soil structural quality due to land use. Soil Tillage Res, 173, 83-91. https://doi.org/10.1016/j.still.2016.08.011

Rey, J.C. Martínez G., Rodríguez G., Lobo D., Delgado E. Trejos J., Pocasangre L. Rosales F., 2009. Aspects on soil quality and welfare in Venezuela bananas. Prod. Agrop., 2, 52-55. Available online at https://n9.cl/5qmzy. Last accessed 19 February 2021.

Rodríguez, G., Núñez, MC., Lobo, D., Martínez. G., Rey, JC., Espinoza, J., Muñoz, N., González D., Rosales, F., Pocasangre, L., Delgado, E., 2006. Banana root health in lots with different productivity levels in a soil at the oriental coast of Maracaibo Lake, Venezuela. XVII Reunión Internacional ACORBAT: Banano un negocio sustentable. Joinville, Santa Catarina, Brasil. Nov 15–20. p. 355

Rosales, F.E., Pocasangre, L.E., Trejos, J., Serrano, E., Peña, W., 2008. *Guía de Diagnóstico de la Calidad y Salud de Suelos bananeros*. Montpellier, France: Bioversity International. Available online at https://n9.cl/z52a. Last accessed 19 February 2021.

Sevinç, G., Aydogdu, M.H., Cancelik, M., Sevinc, M.R., 2019. Farmers' Attitudes toward Public Support Policy for Sustainable Agriculture in GAP-Sanliurfa, Turkey. Sustainability, 11, 6617. https://doi.org/10.3390/su11236617

Soil Survey Staff., 2017. Soil Survey Manual Hadbook 18. United States Department of Agriculture. Washington D.C.

Soil Survey Staff., 2014. Keys to Soil Taxonomy. Keys to Soil Taxonomy, 12th ed. USDA-Natural Resources Conservation Service, Washington, DC.

Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso. J. Roy. Statist. Soc. Ser. B, 58, 267–288. Available online at https://n9.cl/pqv2k. Last accessed 19 February 2021.

Van der Kooij, A. J., 2007. Prediction accuracy and stability of regression with optimal scaling transformations. Unpublished doctoral dissertation, Leiden University. Netherlands. Available online at https://n9.cl/ew8pz. Last accessed 19 February 2021.

van der Kooij, A.J., Meulman, J.J., Heiser, W.J., 2006. Local minima in categorical multiple regression. Computational Statistics & Data Analysis, 50, 446-462. https://doi.org/10.1016/j.csda.2004.08.009

Vanilarasu, K., Balakrishnamurthy, G., 2014. Influences of organic manures and amendments in soil physiochemical properties and their impact on growth, yield and nutrient uptake of banana. The Bioscan, 9, 525-529. Available online at https://n9.cl/awlm. Last accessed 19 February 2021.

Vaquero, MR., 2005. Soil physical properties and banana root growth. In: Turner, D.W and Rosales, F.E (Eds) Banana root system: towards a better understanding for its productive management: proceedings of an International Symposium held in San José, Costa Rica on 3-5 November 2003. International Network for the Improvement of Banana and Plantain, Montpellier, pp 125–131. Available online at https://n9.cl/szt4m. Last accessed 19 February 2021.

Vasu, D., Tiwari, G., Sahoo, S., Dash, B., Jangir, A., Sharma, R. P., Chandran, P., 2021. A minimum data set of soil morphological properties for quantifying soil quality in coastal agroecosystems. Catena, 198, 105042, 1-12. https://doi.org/10.1016/j.catena.2020.105042

Vaughan, I.P., Ormerod, S.J., 2005. Increasing the value of principal components analysis for simplifying ecological data: a case study with rivers and river birds. J. Appl. Ecol, 42, 487–497. Available online at https://n9.cl/2wnp0. Last accessed 19 February 2021.

Villarreal-Núñez, J. et al., 2013. Soil quality indexes in areas cultivated with banana in Panama. Agron. Mesoam, 24, 301-315. Available online at https://n9.cl/6dsn. Last accessed 19 February 2021.

Voorhees, WB, Amemiya, M., Allmaras, RR y Larson, WE., 1971. Some Effects of Aggregate Structure Heterogeneity on Root Growth. Soil Sci. Soc. Am. J, 35(4), 35, 638-643. https://doi.org/10.2136/sssaj1971.03615995003500040043x

Wissmann, M., Shalabh, D., Toutenburg, H., 2011. Role of categorical variables in multicollinearity in linear regression model. J. Appl. Stat, 19, 99 – 115. Available online at https://n9.cl/r9d7u. Last accessed 19 February 2021.

Xu, Jie, Capretz, L.F., Ho, D., 2010. Building an OSS Quality Estimation Model with CATREG. International J Comput Sci Eng, 2, 1952 – 1958. Available online at https://n9.cl/t3wlt. Last accessed 19 February 2021.

Summary discussion of main results

Summary discussion of main results

In this this PhD Thesis, the relationships between soil, productivity, climatic suitability and *Fusarium* wilt in banana has been studied giving rise to several outcomes and results of scientific relevance and also importantly, of practical implications for banana producers in Venezuela.

Based on current knowledge of the epidemiology of *Foc*, the management practices oriented to increase the quality and suppressiveness of the soil is described in Chapter I. These approaches can contribute to suppress the inoculum of *Foc*, reduce the incidence and severity of the disease and increase productivity. This will contribute to address the major challenge of identifying the most effective practices for decreasing the inoculum of a pathogen with long-term survival capacity that affects a perennial crop with continuous flowering and harvest cycles (Dita et al. 2018; Kema et al. 2021; García-Bastidas et al. 2020)

Secondly, an analysis of the climatic suitability of *Foc* RT4 for Venezuela producing areas was performed in the study included in Chapter I; an issue not properly assessed before, despite the amount of information available on this pathogen. Our results showed how the Maximum Entropy model estimated that all the geographic areas described for banana production in Venezuela are climatically suitable for the occurrence of *Foc* RT4. Thus, in the current climate conditions, the areas of great commercial importance of the municipalities La Cañada Urdaneta and La Guajira in the Zulia state show the highest suitability scores (> 0.90), followed by the municipalities Buroz (0.80) and Brión (0.70) from the Miranda state; as well as the banana areas of the Trujillo state, specifically in the Bolívar and La Ceiba municipalities (0.80). These results are very useful for the design of new strategies for the efficient use and greater effectiveness of establishment of legislative and regulatory prevention activities on borders to avoid introduction of *Foc* RT4 and or its early detection.

The intensification of research on *Foc* RT4 in Latin America and the Caribbean is extremely relevant, especially in aspects related to its epidemiology and control, particularly in Venezuela (Chapter I). However, the unexpected recent detection of *Foc* RT4 in Colombia and Peru, with very different producing situations in each of these countries, means that in all producing countries resources must be dedicated to investigating the risks posed by *Foc* and on best management methods to address them (García-Bastidas et al. 2020). This chapter identified the most vulnerable areas in Venezuela (Zulia, Miranda, Trujillo and Sucre state) for *Foc* RT4 occurrence according to a maximum entropy model, and the main consequences for banana production due to a potential expansion in the geographical distribution of this pathogen in the future. This should help to develop the appropriate measures for control of the pathogen mainly based on early detection and eradication of *Foc* RT4, since no cure is available for this highly virulent race once is introduced in an area.

Chapter II, was focused in an analysis of the key soil properties that play an important role in the incidence of BW. So far, crop disease detection models primarily focus on leaf symptoms through image recognition technology. This means that diseases can be detected only after they have appeared. Our case study carried out using a machine learning technique for soils of lacustrine origin in Venezuela provided a complementary approach based on the relationship between BW and soil properties. This could be used to identify areas with different potential risk oh high incidence of BW, optimizing the use of other techniques for BW detection and orienting best management practices to prevent its appearance. Our results showed that BW incidence is related to Zn, Fe, Ca, K, Mn and Clay content and provided a significant success (around 90%) in distinguishing between areas of high and low BW incidence. The soil properties identified are related to the drainage capacity of the soil (clay) or reflect indirectly the influence of difference soil drainage conditions (e.g. Zn and Fe solubility).

Several studies pointed out that the determination of the soil properties in the banana zone studied in this Thesis reveal similar information according to the reports by Rey et al. (2020) and Martinez et al. (2020). Machine learning analysis is considered an alternative to the more complex applications that deal with the prediction of banana diseases, as demonstrated by the performance of Random Forest in the investigations by Gómez-Selvaraj et al. (2020), Sangeetha et al. (2020) and Ye et al. (2020).

Our results open the field for further research in which we could quantitatively predict the risk of BW in banana fields based on available, or relatively easy to gather, information which in turn could allow farm managers to implement preventive measures to minimize BW risk and target other techniques (e.g. plant sampling, withdrawal of infested material) to the areas where there is maximum risk.

The way of how soil properties determine banana productivity was quantified in Chapter III. The penetration resistance, Mg, bulk density and microbiological (biotic) properties such as microbial respiration, and omnivorous free-living nematodes were linked to biometric and productive responses in commercial Cavendish banana plantations. This relationship with production can also be related to the response of plants to diseases. However, the variability of the systems must be considered in an eventual implementation of practices to manage soil properties for both production and disease management.

It was observed that in large banana farms (> 50 ha) soil management is a critical factor that cannot be neglected since the productivity indicators evaluated in this study such as the number of hands per bunch and the circumference of the mother plant showed to have a close relationship with the physical properties: resistance to penetration and bulk density followed by the microbiological properties reported in recent studies of González-García et al. (2021a; 2021b) and Acevedo et al. (2021) which establish that the deterioration of the soil could endanger the production of the crop in these soils.

The regression model obtained allowed the development of an easy-to-implement methodology for the rest of the banana areas of Venezuela and Latin America, that i when further validated, it would represent a tool for sustainable land management. When adapted to local considerations of each ecosystem, will allow monitoring and evaluating the quality of the soil and implement actions that prevent the advance of degradation with a comprehensive and sustainable approach oriented also towards yield optimization.

Likewise, we consider that this study represents an important contribution to the knowledge of the banana soils of Venezuela, contributing to other recent studies focused on the quality of banana soil (González-García et al. 2021a; 2021b; Rondón et al. 2021). We are aware that the study can be improved through its systematic use in new locations, so it is our intention and we hope that other research groups from the international scientific community will join in this task to produce improved versions of it.

In addition to corroborating the importance of individual soil properties and their interactions, Chapter IV showed that it is possible to also consider soil morphological properties in the differentiation of banana productivity. There is little history of soil quality indices using categorical soil properties, and our study so practical example to consider field morphological variables transformed by the optimal scale to capture the differences in banana productivity among sites.

The results of the study in Chapter 4 emphasizes the importance of integrating morphological data in numerical evaluation of soil quality. Our study is perhaps the first to present a categorical regression model of soil morphological properties to evaluate banana productivity. Therefore, the studied morphological indicators could be used effectively to assess soil quality and productivity in banana cropping systems a as it has been made previously only for soil quality purposes, e.g. Pulido-Moncada et al. (2014) and Vasu et al. (2021).

Our results also showed that the categorical regression model with morphological properties is a potential option to quickly quantify the effect of land use type on soil quality, as it explains a system variability similar to that of other studies based on physical, chemical and biological properties (Delgado et al. 2010a; 2010b), the later ones usually not feasible to be measured in a large number routinely for soil monitoring. So, the pedo-morphological properties selected in this study could serve as surrogates for other properties. Banana farmers in Venezuela can easily implement morphological soil quality assessment for rapid assessment of soil quality and link it to productivity. The approach we propose in this study reduces the time required for field evaluation of soil quality in banana growing regions and, in any region, where soil data is limited.

Using these morphological properties can promote the participation of banana growers in the appraisal of soil quality and its link with banana productivity, and so spur the interest of banana farmers in the introduction of best management practices. The results of Although the participation of farmers in the visual evaluation of soil quality is sometimes limited restricted by the lack of knowledge of pedo-morphology, current research shows the potential for addressing this limitation using digital soil morphometry with proximal sensors Vasu et al. (2021). Looking to the future, the quantitative transformation of the morphological properties of the field soil by categorical regression will be useful in to compare the potential productivity of crops under different conditions.

Finally, most of the techniques developed in this PhD Thesis show the way to develop practical strategies for agricultural extension agents, providing a way to evaluate risk regarding banana diseases by agricultural services based on soil and climatic conditions to assist producers. This could also be used by producers and technical personnel of large and medium-sized farms in banana production systems in Latin America and the Caribbean.

Key Conclusions

Key Conclusions

In summary, the present PhD Thesis has identified the relationships between soil properties, productivity, climatic suitability and the incidence of diseases in bananas in Venezuela. Compared to previous studies, the most interesting and novel findings were:

a. The approach of a systematic study that included a literature review, field observation, and systematic soil sampling on farms was effective in analyzing the role of soil in banana productivity and Fusarium wilt incidence.

b. A review of the current state of knowledge concerning *Foc* TR4 worldwide and an analysis of the environmental suitability in Venezuela was provided, which made it possible to understand the real risk posed by *Foc* TR4 at the national and global level.

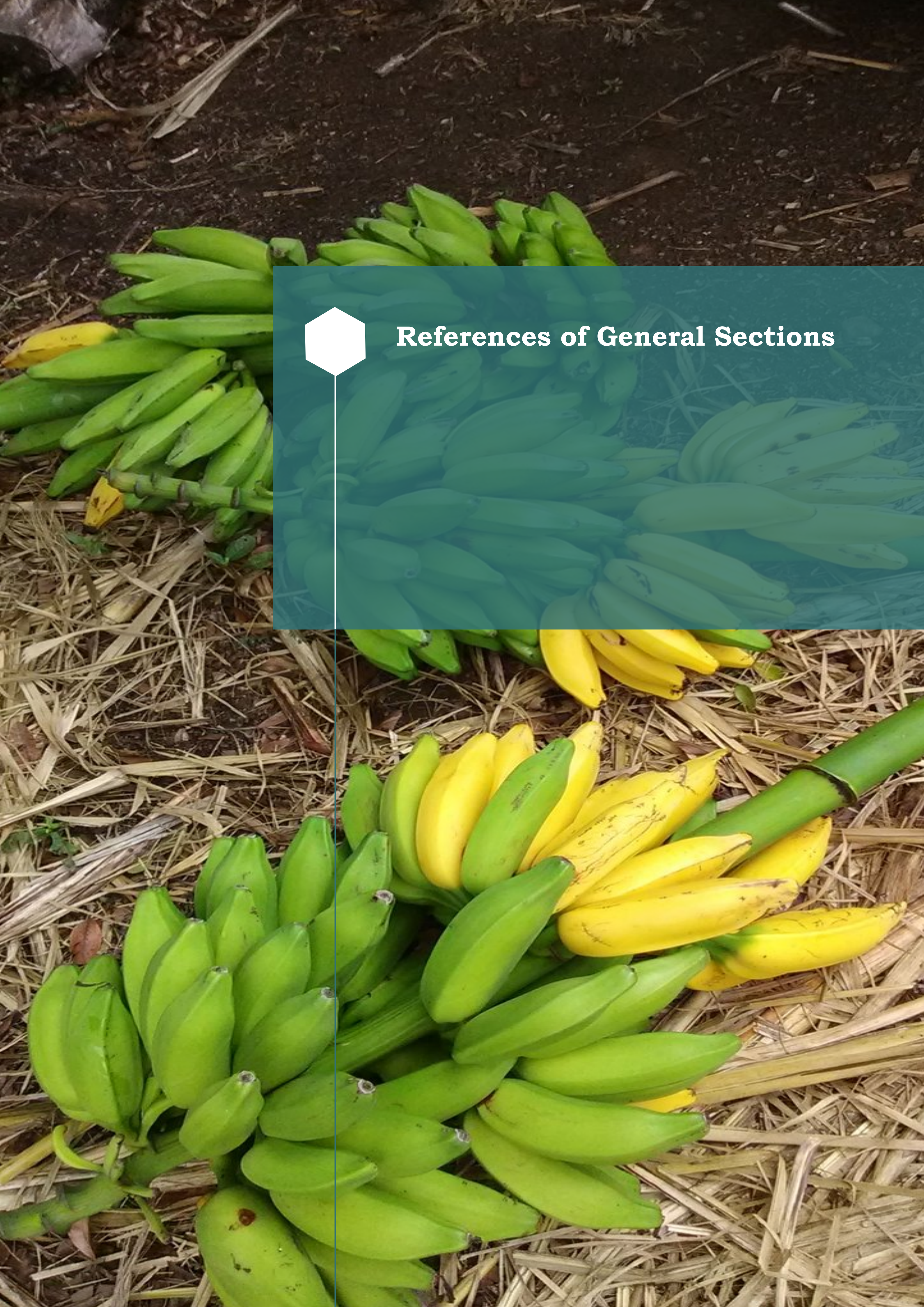
c. The presence-only data of the *Foc* pathogen on a world scale helped to model the future distributions of this species in Venezuela and to inform about the approximate risk of appearance and transmission of this lethal disease in bananas, indicating that the banana areas with greater commercial importance in Venezuela present adequate climatic conditions for survival and possibly the establishment of *Foc* TR4 in case of its introduction.

d. The potential of using soil characteristics (Zn, Fe, K, Ca, Mn and clay content) as predictors to avoid establishment of banana plantations in soils with high risk of incidence of Banana Wilt disease has been demonstrated using the Random Forest algorithm.

e. The influence of specific soil properties such as Mg, penetration resistance, total microbial respiration, bulk density and omnivorous free-living nematodes on banana productivity was demonstrated, with the interactions among them appearing to be more significant than the effect of a single property on the productivity.

f. The soil type (lacustrine and alluvial) can naturally affect the banana productivity, so that each type of soil should be considered individually when identifying the soil properties that can explain differences in banana productivity.

g. The categorical regression analysis generated an operational model in bananas in which the soil morphological properties, such as biological activity, texture, dry consistency, reaction to HCl and structure type differentiated different banana productivity levels. The results of this analysis showed for the first time that soil morphological variables could be promising new soil indicators for assessment banana productivity in Venezuelan soils.



References of General Sections

Acevedo, I., Sánchez, A. & Mendoza, B. (2021). Evaluación del nivel de degradación del suelo en dos sistemas productivos en la depresión de Quíbor. II. Calidad del suelo. Bioagro 33(2): 127-134. https://doi.org/10.51372/bioagro332.6

Ahmad, F., Martawi, N. M., Poerba, Y. S., de Jong, H., Schouten, H., & Kema, G. H. (2020). Genetic mapping of Fusarium wilt resistance in a wild banana *Musa acuminata* ssp. *malaccensis* accession. Theoretical and Applied Genetics, 133(12), 3409-3418. https://doi.org/10.1007/s00122-020-03677-y

Alimi, O. A., Ouahada, K., Abu-Mahfouz, A. M., Rimer, S., & Alimi, K. O. A. (2021). A Review of Research Works on Supervised Learning Algorithms for SCADA Intrusion Detection and Classification. Sustainability, 13(17), 9597. https://doi.org/10.3390/su13179597

Aruraj, A., Alex, A., Subathra, M. S. P., Sairamya, N. J., George, S. T., & Ewards, S. V. (2019). Detection and Classification of Diseases of Banana Plant Using Local Binary Pattern and Support Vector Machine. In 2019 2nd International Conference on Signal Processing and Communication (ICSPC) (pp. 231-235). IEEE. https://doi.org/10.1109/ ICSPC46172.2019.8976582

Baviera, T. (2017). Técnicas para el Análisis de Sentimiento en Twitter: Aprendizaje Automático Supervisado y SentiStrength. Revista Dígitos, 1(3), 33-50. http://dx.doi.org/10.7203/rd.v1i3.74

Bouma, J. (2020). Contributing pedological expertise towards achieving the United Nations Sustainable Development Goals. Geoderma 375, 114508. https://doi.org/10.1016/j.geoderma.2020.114508

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/a:1010933404324

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and regression trees. CRC press.

Calero, J., Aranda, V., Montejo-Raez, A., Martin-Garcia, J.M. (2018.) A new soil quality index based on morpho-pedological indicators as a site-specific web service applied to olive groves in the Province of Jaen (South Spain). Comput Electron Agric, 146, 66-76. https://doi.org/10.1016/j.compag.2018.01.016

Coltro, L. & Karaski, T. U. (2019). Environmental indicators of banana production in Brazil: *Cavendish* and *Prata* varieties. Journal of cleaner production, 207, 363-378. https://doi.org/10.1016/j.jclepro.2018.09.258

Companioni, B., Mora, N., Díaz, L., Pérez, A., Arzola, M., Espinosa, P., M. Hernández, M., Ventura, J., Pérez, M.C., Santos, R. & Lorenzo, J. C. (2005). Identification of discriminant factors after treatment of resistant and susceptible banana leaves with *Fusarium oxysporum* f. sp. *cubense* culture filtrates. Plant breeding, 124(1), 79-85. https://doi.org/10.1111/j.1439-0523.2004.00997.x

Cortes, C. & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273-297. doi:10.1007/BF00994018

Cortez, A., Olivares, B., Parra, R., Lobo, D., Rodríguez, M.F. & Rey, J.C. (2018). Descripción de los eventos de sequía meteorológica en localidades de la cordillera central, Venezuela. Ciencia, Ingenierías y Aplicaciones. I (1):22-44. DOI: http://dx.doi.org/10.22206/cyap.2018.vlil.pp23-45

Delgado, E., Gómez, N., González, O. & Marín, C. (2008). Evaluación a nivel de finca del efecto de la alta densidad de siembra en plátano (*Musa* AAB cv. Subgrupo plátano Hartón), municipio Obispo, Barinas, Venezuela. Rev. Fac. Agron. (LUZ), 25 (4): 27-38. https://n9.cl/poml2

Delgado, E., Rosales, F., Trejos, J., Villalobos, M. & Pocasangre, L. (2010b). Índice de calidad y salud de suelos para plantaciones bananeras de cuatro países de América Latina y el Caribe. Bioagro, 22 (1): 53-60. https://n9.cl/qm3e

Delgado, E., Trejos, J., Villalobos, M., Martínez, G Lobo, D., Rey, J., Rodríguez, G., Rosales, F. & Pocasangre, L. (2010a). Determinación de un índice de calidad y salud de suelos para plantaciones bananeras en Venezuela. Interciencia, *35* (12): 927-933. https://n9.cl/nv7vv

Demey, J.R., Adams, M. & Freites, H. (1994). Uso del método de análisis de componentes principales para la caracterización de fincas agropecuarias. Agronomía Trop, 44(3): 475–497.

Dita, M., Barquero, M., Heck, D., Mizubuti, ESG. & Staver, C.P. (2018). Fusarium wilt of banana: Current knowledge on epidemiology and research needs toward sustainable disease management. Front. Plant Sci, 9 (1468), 1-21. https://doi.org/10.3389/fpls.2018.01468

Dita, M., Perez-Vicente, L., Guzmán, M., Urias, C., & Staver, C. (2017). Fusarium tropical race 4 and it threat to food security and the banana industry in Latin America and the Caribbean. Phytopathology, *107*(7), 8-8.

Fajardo, J., Aguilar, L., Flores, M., Parra, D., & Acurio, C. (2017). Aspectos socio-económicos influyentes en el sector agrícola y su relación con la creación de empresas populares y solidarias de Quevedo. Ciencia Sociales y Económicas, 1(1), 107-116. https://doi.org/10.18779/csye.v1i1.254

FAO. (2020a). FAOSTAT Base datos: Superficie, producción y exportación de banano y plátano (en línea). Consultado 5 mar. 2020. Disponible en http://www.fao.org/faostat/es/#data/QCL

FAO. (2020b). Perspectivas a mediano plazo: perspectivas para la producción y el comercio mundial de bananos y frutas tropicales 2019-2028. Roma. https://n9.cl/xfd23

García-Bastidas, F. A., Quintero-Vargas, J. C., Ayala-Vasquez, M., Schermer, T., Seidl, M. F., Santos-Paiva, M., et al. (2020). First report of Fusarium wilt Tropical race 4 in Cavendish bananas caused by *Fusarium odoratissimum* in Colombia. Plant Dis. 104, 994–994. https://doi.org/10.1094/PDIS-09-19-1922-PDN

Gauggel, C.A., Sierra, F. & Arévalo, G. (2003). The problems of banana root deterioration and their impact on production: production experience in Latin America. *In:* Turner D.W. and F.E. Rosales (eds). Banana Root System: towards a better understanding for its productive management: Proceedings of an International Symposium/ International Network for the Improvement of Banana and Plantain, Montpellier, France. https://n9.cl/b3wi8

Ghag, S. B., Shekhawat, U. K. S. & Ganapathi, T. R. (2015). Fusarium wilt of banana: biology, epidemiology and management. International Journal of Pest Management, 61(3), 250-263. https://doi.org/10.1080/09670874.2015.1043972

Gómez-Selvaraj, M., Vergara, A., Montenegro, F., Alonso Ruiz, H., Safari, N., Raymaekers, D., Ocimati, W., Ntamwira, J., Tits, L., Omondi, A. B. & Blomme, G. (2020). Detection of banana plants and their major diseases through aerial images and machine learning methods: A case study in DR Congo and Republic of Benin. ISPRS Journal of Photogrammetry and Remote Sensing, 169, 110-124. doi: https://doi.org/10.1016/j.isprsjprs.2020.08.025

Gómez, C., Surga, J., Rumbos, R., Rosales, H., Vera, J., Pino, N., Castellano G. & Marin, C. (2012). Efecto del postdeshoje para la reducción de *Mycosphaerella fijiensis*, en plátano cv. Hartón (*Musa* AAB), Venezuela. *Revista de la Facultad de Agronomía LUZ, 29*(2), 173-182. https://n9.cl/r5jqs

González-García, H., González Pedraza, A.F., Atencio J. y Soto A. (2021a). Evaluación de calidad de suelos plataneros a través de la actividad microbiana en el sur del lago de Maracaibo, estado de Zulia, Venezuela. Revista de la Facultad de Agronomía (LUZ) 38. 2, 216-240. https://doi.org/10.47280/RevFacAgron(LUZ).v38.n2.01

González-García, H., González Pedraza, A.F., Pineda M., Escalante-García H., Gustavo Rodríguez-Yzquierdo G., y Soto-Bracho A. (2021b). Microbiota edáfica en lotes de plátano con vigor contrastante y su relación con propiedades del suelo. Bioagro 33, 2. 143-148. https://doi.org/10.51372/bioagro332.8

Hamed, A. A., Tawfik, E., Abdel-Razik, A. B., Rashed, M. A., Abdel-Aziz, S. H., & El-Shafie, M. A. (2018). Production of Transgenic Cavendish Banana (*Musa Acuminata* L.) Resistant to Fungal Infection using Chitosan Nanoparticles. Benha Journal of Applied Sciences, 3(1), 131-137. https://doi.org/10.21608/BJAS.2018.191091

Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J. & Scholkopf, B. (1998). Support vector machines. IEEE Intelligent Systems and their Applications, 13(4), 18-28. https://doi.org/10.1109/5254.708428

Hermy, M. & Paul. J.L. (1991). Multivariate ratio analysis, a graphical method for ecological ordination. Ecology, 72:735-738. https://doi.org/10.2307/2937213

Hernández, Y., Portillo, F., Portillo, M., Navarro, C., Rodríguez, M., & Velazco, J. (2006). Densidad estomática en materiales de plátano (*Musa* AAB, AAAB y ABB) susceptibles y resistentes a Sigatoka Negra (*Mycosphaerella fijiensis*, Morelet). Revista de la Facultad de Agronomía LUZ, 23(3), 294-300. https://n9.cl/gl9tq

Hou, J. C., Hu, Y. H., Hou, L. X., Guo, K. Q., & Satake, T. (2015). Classification of ripening stages of bananas based on support vector machine. International Journal of Agricultural and Biological Engineering, 8(6), 99-103. https://doi.org/10.3965/j.ijabe.20150806.1275

Iskandar, J., Kusmoro, J., Mubarokah, M., & Partasasmita, R. (2018). Ethnobotany of banana plants (*Musa x paradisiaca*) of Palintang Hamlet, Cipanjalu Village, Bandung, West Java, Indonesia. Biodiversitas Journal of Biological Diversity, 19(6), 2059-2072. https://doi.org/10.13057/biodiv/d190611

Kema, G.H.J.; Drenth, A.; Dita, M.; Jansen, K.; Vellema, S.; Stoorvogel, J.J. (2021). Editorial: Fusarium wilt of banana, a recurring threat to global banana production. Front. Plant Sci, 11, 628888. https://doi.org/10.3389/fpls.2020.628888

Kuhn M, Weston S, Culp M, Coulter N, Quinlan R. (2018). Paquete «C50». https://cran.r-project.org/web/packages/C50/C50.pdf

Lei, Y. (2017). Individual intelligent method-based fault diagnosis. In Y. Lei (Ed.), Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery (pp. 67-174). Butterworth-Heinemann. doi:10.1016/B978-0-12-811534-3.00003-2

Liaw, A., & Wiener, M. (2002). Classification and Regression by random Forest. R news, 2, 18-22. https://n9.cl/b57nr

Ma L., Fu T., Blaschke T., Li M., Tiede D., Zhou Z., Ma X. & Chen D. (2017). Evaluation of feature selection methods for object-based land cover mapping of unmanned aerial vehicle imagery using random forest and support vector machine classifiers. ISPRSIsprs Int. J. Geo-Inf, 6(2), 51. https://doi.org/10.3390/ijgi6020051

Manjunath, C. B., Gnanappazham, L., Vijayakumar, R. M., Kavino, M., Jagadeeswaran, R., & Ramsubramoniam, S. (2019). Machine learning classifiers on sentinel-2 satellite image for the classification of banana (Musa Sp.) plantations of Theni district, Tamil Nadu, India. International Journal of Chemical Studies, 7(5), 1419-1425. https://n9.cl/hdr2b

Martínez, G. E., Rey-Brina, J. C., Rodríguez, D., Jiménez, C., Rodríguez, Y., Rumbos, R., Pargas-Pichardo, R., & Martínez, E. (2020a). Análisis de la situación fitopatológica actual de las Musáceas comestibles en Venezuela. Agronomía Tropical, 70, 1-20. https://doi.org/10.5281/zenodo.4323273

Martínez, G., Rey-Brina, J. C., Pargas-Pichardo, R. E., & Domínguez, M. (2020b). Actualización sobre el manejo de la marchitez por Fusarium Raza 4 Tropical en musáceas. Agronomía

Tropical, 70, 1-18. https://doi.org/10.5281/zenodo.4291108

Menéses, A., Pocasangre, L.E., Somarriba, E., Riveros, A.S., and Rosales,F.E. (2003). Diversidad de hongos endofíticos y abundancia de nematodos en plantaciones de banano y plátano de la parte baja de los territorios indígenas de Talamanca. Agroforestería de las Américas. 10(37):59-62. https://n9.cl/gqy7y

Moreno-Carriles, R. M. (2018). Big data; But what is it? Angiología, 70(5), 191-194. https://doi.org/10.1016/j.angio.2018.05.001

Nava, C. (1997). El plátano, su cultivo en Venezuela. Ediciones Astro Data S.A. Maracaibo, Venezuela. 134 p.

Olivares, B., Cortez, A., Parra, R., Lobo, D., Rodríguez, M.F y Rey, J.C. (2017). Evaluation of agricultural vulnerability to drought weather in different locations of Venezuela. Revista de la Facultad de Agronomía (LUZ) 34 (1): 103-129. https://n9.cl/d827w

Olivares, B., Paredes, F., Rey, J., Lobo, D., Galvis-Causil, S. (2021). The relationship between the normalized difference vegetation index, rainfall, and potential evapotranspiration in a banana plantation of Venezuela. *SAINS TANAH - Journal of Soil Science and Agroclimatology, 18*(1), 58-64. http://dx.doi.org/10.20961/stjssa.v18i1.50379

Owomugisha, G., Quinn, J. A., Mwebaze, E., & Lwasa, J. (2014). Automated vision-based diagnosis of banana bacterial wilt disease and black sigatoka disease. In: International Conference on the Use of Mobile ICT in Africa (pp. 1-5). https://n9.cl/ftc3n

Pattison, A. B., Wright, C. L., Kukulies, T. L., & Molina, A. B. (2014). Ground cover management alters development of Fusarium wilt symptoms in Ducasse bananas. Australasian Plant Pathology, 43(4), 465-476. https://doi.org/10.1007/s13313-014-0296-5

Pattison, T., Badcock, K., Lindsay, S. Armour, A., Velupillai, R., Moody, P., Smith, L., Gullino, L. & Cobon, J. (2004). Banana root and soil health project – field workbook. Department of Primary Industries and Fisheries, Queensland, Australia. 15 p. https://n9.cl/nrtvc

Ploetz, R. C. (2015). Management of Fusarium wilt of banana: A review with special reference to tropical race 4. Crop Protection, 73, 7-15. http://dx.doi.org/10.1016/j.cropro.2015.01.007

Pocasangre, L.E., Sikora, R.A., Vilich, V. and Schuster, R-P. (2000). Survey of banana endophytic fungi from Central America and screening for biological control of *Radopholus similis*. Acta Horticulturae, 531: 283-289. https://doi.org/10.17660/ActaHortic.2000.531.47

Pulido-Moncada, M., Gabriels, D., Lobo, D., Rey, B.J.C. & Cornelis, W.M., (2014). Visual field assessment of soil structural quality in tropical soils. Soil Tillage Res. 139, 8–18. https://doi.org/10.1016/j.still.2014.01.002

Quinlan, J. R. (2014). C4. 5: programs for machine learning. Amsterdam, Netherlands: Elsevier.

Quinlan, J.R. (2007). Decision Trees as Probabilistic Classifiers. In: Proceedings of the Fourth International Workshop on Machine Learning. Massachusetts, United States: Morgan Kaufmann Publishers, Inc. https://doi.org/10.1016/b978-0-934613-41-5.50007-6

Rey J.C., Chacín M., Sapuky M., Núñez M., Martínez G., Rodríguez G., Espinoza J., Arturo M., Pocasangre L., Delgado E. & Rosales F. (2006). Aptitud de las tierras para banano en suelos de Venezuela y su relación con la productividad. XVII Reunión Internacional ACORBAT: Banano un negocio sustentable. Joinville. Santa Catarina, Brasil. Nov 15-20. p. 362.

Rey J.C., Martínez G., Lobo D., Trejos J., Pocasangre L. & Rosales F. (2009). Aspectos sobre calidad y salud de suelos bananeros en Venezuela. Producción Agropecuaria, 2: 52-55. https://n9.cl/zthj2

Rey, J.C., Martínez-Solórzano, G., Ramírez, H., & Pargas-Pichardo, R. (2020). Relación de las condiciones agroecológicas de un lote de planicie lacustrina con la marchitez del banano Cavendish en Aragua, Venezuela. Agronomía Tropical, 70, 1–12. http://doi.org/10.5281/zenodo.4346252

Rondón, T., Hernandez, R.M. & Guzman, M. (2021) Soil organic carbon, physical fractions of the macro-organic matter, and soil stability relationship in lacustrine soils under banana crop. PLOS ONE 16(7): e0254121. https://doi.org/10.1371/journal.pone.0254121

Rosales, F.E., Pocasangre, L.E., Trejos, J., Serrano, E. & Peña, W. (2008). Guía de Diagnóstico de la Calidad y Salud de Suelos. Bioversity International. Montpellier, Francia. 79 pp.

Sabiiti, G., Ininda, J. M., Ogallo, L. A., Ouma, J., Artan, G., Basalirwa, C. & Nanteza, J. (2018). Adapting agriculture to climate change: suitability of banana crop production to future climate change over Uganda. In Limits to Climate Change Adaptation (pp. 175-190). Springer, Cham. https://doi.org/10.1007/978-3-319-64599-5_10

Sangeetha, T., Lavanya G., Jeyabharathi, D., Rajesh Kumar T., & Mythili K. (2020). Detection of pest and disease in banana leaf using convolution random forest. Test Engineering and Management, 83(2), 3727 – 3735. https://n9.cl/yeffq

Scott, G. J. (2021). A review of root, tuber and banana crops in developing countries: Past, present and future. International Journal of Food Science & Technology, 56(3), 1093-1114. https://doi.org/10.1111/ijfs.14778

Serrano, E. (2003). Relación entre los contenidos de raíz funcional y la productividad de banano en Costa Rica. In: Turner D.W. and F.E. Rosales (eds). 2005. Banana Root System: towards a better understanding for its productive management: Proceedings of an international symposium. International Network for the Improvement of Banana and Plantain, Montpellier, France. p. 28.

Van Eck, N.J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523-538. https://doi.org/10.1007/s11192-009-0146-3

Van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. Machine Learning, 109(2), 373-440. https://doi.org/10.1007/s10994-019-05855-6

Vapnik. V. N. (1995). The nature of statistical learning theory. Berlin, Heidelberg: Springer-Verlag.

Varma, V., & Bebber, D. P. (2019). Climate change impacts on banana yields around the world. Nature Climate Change, 9(10), 752-757. https://doi.org/10.1038/s41558-019-0559-9

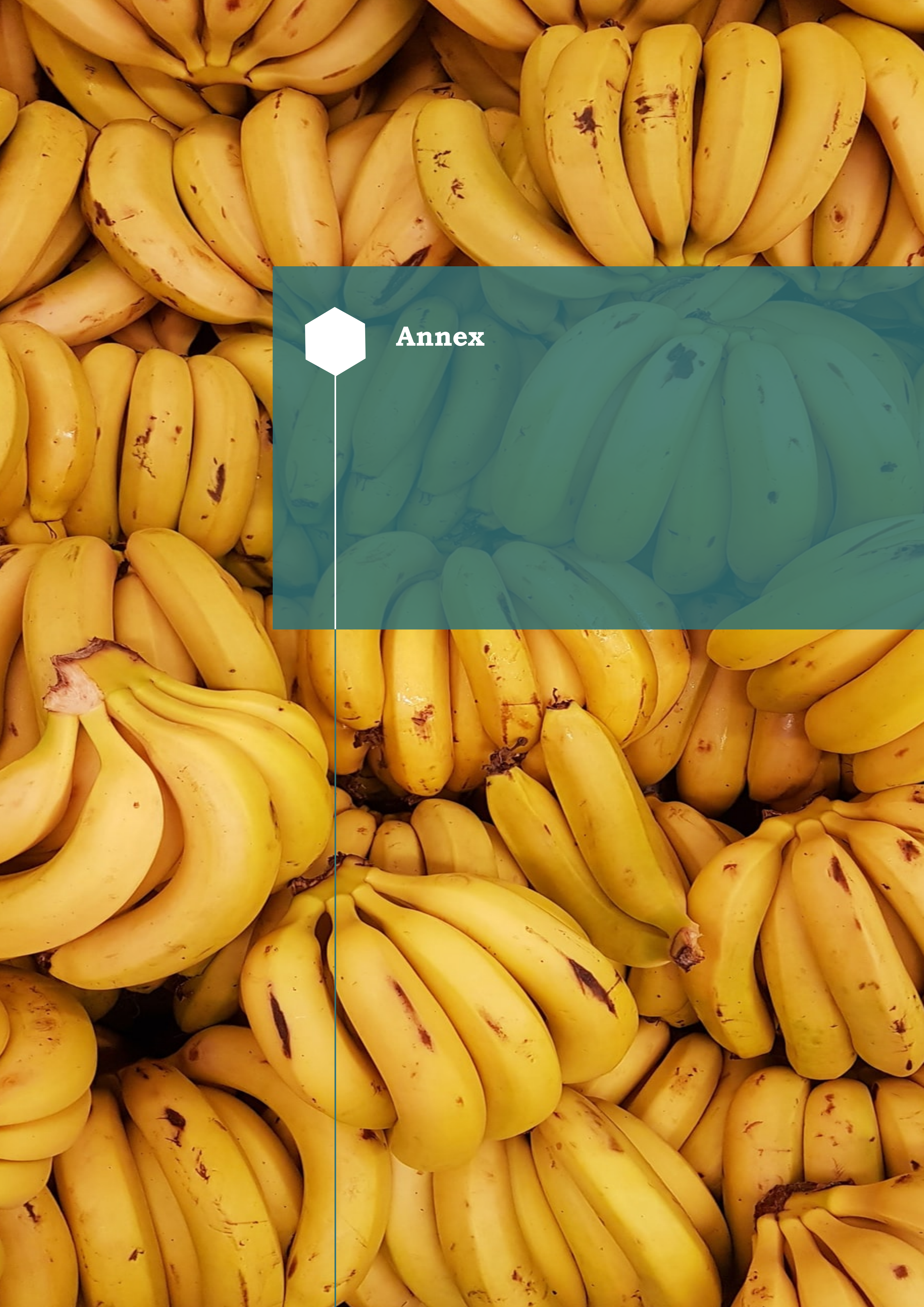
Vasu, D., Tiwari, G., Sahoo, S., Dash, B., Jangir, A., Sharma, R. P., Chandran, P., (2021). A minimum data set of soil morphological properties for quantifying soil quality in coastal agroecosystems. Catena, 198, 105042, 1-12. https://doi.org/10.1016/j.catena.2020.105042

Vite Cevallos, H., Carvajal Romero, H., & Barrezueta Unda, S. (2020). Application of machine learning algorithms to classify the fertility of a banana soil. Revista Conrado, 16(72), 15-19. https://n9.cl/qsc4l

Wang, X., Lin, X., & Dang, X. (2020). Supervised learning in spiking neural networks: A review of algorithms and evaluations. Neural Networks, 125, 258-280. https://doi.org/10.1016/j.neunet.2020.02.011

Xanthopoulos, P., Pardalos, P. M., & Trafalis, T. B. (2013). Linear discriminant analysis. In: Robust data mining (pp. 27-33). Springer, New York, NY. doi:10.1007/978-1-4419-9878-1_4

Ye H. C., Huang W. J., Huang S. Y., Cui B., Dong Y. Y., Guo A. T., Ren Y., & Jin Y. (2020). Identification of banana fusarium wilt using supervised classification algorithms with UAV-based multi-spectral imagery. Int J Agric & Biol Eng, 13(3), 136–142. https://doi.org/10.25165/j.ijabe.20201303.5524



Annex

Annex. Informe de seguimiento del doctorando 2017-2021

A. Publicaciones científicas (Artículos indexados o no, comunicaciones a congresos, libros, capítulos de libros derivados directamente o no de la tesis)

En revistas indexadas en el SCI

1. **Olivares, B.O.**, Calero, J., Rey, J.C., Lobo, D., Landa, B.B., Gómez, J. A. 2022. Correlation of banana productivity levels and soil morphological properties using regularized optimal scaling regression. Catena, 208: 105718. https://doi.org/10.1016/j.catena.2021.105718

2. **Olivares B**, Rey JC, Lobo D, Navas-Cortés JA, Gómez JA, Landa BB. 2021. Fusarium Wilt of Bananas: A Review of Agro-Environmental Factors in the Venezuelan Production System Affecting Its Development. Agronomy, 11(5):986. https://doi.org/10.3390/agronomy11050986

3. **Olivares, B.**, Araya-Alman, M., Acevedo-Opazo, C. et al. 2020. Relationship Between Soil Properties and Banana Productivity in the Two Main Cultivation Areas in Venezuela. J Soil Sci Plant Nutr. 20 (3): 2512-2524. https://doi.org/10.1007/s42729-020-00317-8

4. Casana, S., **Olivares, B**. 2020. Evolution and trend of surface temperature and windspeed (1994 - 2014) at the Parque Nacional Doñana, Spain. Rev. Fac. Agron. (LUZ), 37(1):1-25. https://n9.cl/c815e

5. Cortez, A., **Olivares, B.**, Parra, M., Lobo, D., Rey, JC., Rodriguez, MF. 2019. Systematization of the calculation of the Standardized Precipitation Index as a methodology to generate meteorological drought information. Rev. Fac. Agron. (LUZ), 36(2):209-223. https://n9.cl/prf63

6. **Olivares, B.**, López-Beltrán, M., Lobo-Luján, D. 2019. Cambios de usos de suelo y vegetación en la comunidad agraria Kashaama, Anzoátegui, Venezuela: 2001-2013. Revista Geográfica De América Central. 2(63):269-291. https://doi.org/10.15359/rgac.63-2.10

7. **Olivares, B.**, Hernández, R; Coelho, R., Molina, JC., Pereira, Y. 2018. Analysis of climate types:

Main strategies for sustainable decisions in agricultural areas of Carabobo, Venezuela. Scientia Agropecuaria. 9(3): 359 – 369. http://dx.doi.org/10.17268/sci.agropecu.2018.03.07

8. **Olivares, B**., Hernández, R; Coelho, R., Molina, JC., Pereira, Y. 2018. Spatial analysis of the water index: an advance in the adoption of sustainable decisions in agricultural territories of Carabobo, Venezuela. Revista Geográfica de América Central. 60 (1): 277-299. https://doi.org/10.15359/rgac.60-1.10

9. **Olivares, B.** Parra, R y Cortez, A. 2017. Caracterización de los patrones de precipitación en el estado Anzoátegui, Venezuela. Ería 3 (3): 353-365. https://doi.org/10.17811/er.3.2017.353-365

10. **Olivares, B.**, Lobo, D., Cortez, A., Rodríguez, M.F. y Rey, J.C. 2017. Socio-economic characteristics and methods of agricultural production of indigenous community Kashaama, Anzoategui, Venezuela. Rev. Fac. Agron. (LUZ) 34 (2): 187-215. https://n9.cl/p2gc5

11. **Olivares, B.** 2018. Condiciones tropicales de la lluvia estacional en la agricultura de secano de Carabobo, Venezuela. La Granja: Revista de Ciencias de la Vida. 27(1):86-102. https://doi.org/10.17163/lgr.n27.2018.07

12. **Olivares, B.**, Cortez, A., Parra, R., Lobo, D., Rodríguez, M.F y Rey, J.C 2017. Evaluation of agricultural vulnerability to drought weather in different locations of Venezuela. Rev. Fac. Agron. (LUZ) 34 (1): 103-129. https://n9.cl/d827w

En revistas indexadas en Scopus

13. **Olivares, B**., Paredes, F., Rey, J., Lobo, D., Galvis-Causil, S. 2021. The relationship between the normalized difference vegetation index, rainfall, and potential evapotranspiration in a banana plantation of Venezuela. *SAINS TANAH - Journal of Soil Science and Agroclimatology, 18*(1), 58-64. http://dx.doi.org/10.20961/stjssa.v18i1.50379

14. Pittí Rodríguez, J., **Olivares, B**., Montenegro, E., Miller, L., & Ñango, Y. 2021. The role of agriculture in the Changuinola District: a case of applied economics in Panama. Tropical and Subtropical Agroecosystems, 25(1), 1-11. https://n9.cl/quyl2

15. Montenegro-Gracia, E.J., Pitti-Rodríguez, J.E. & **Olivares, B.** 2021. Adaptation to climate change in indigenous food systems of the Teribe in Panama: a training based on CRISTAL 2.0. Luna Azul, 51 (2): 182-197. https://doi.org/10.17151/luaz.2020.51.10

16. **Olivares, B**., Hernández, R. 2020. Application of multivariate techniques in the agricultural land's aptitude in Carabobo, Venezuela. Tropical and Subtropical Agroecosystems, 23(2):1-12. https://n9.cl/zeedh

17. **Olivares, B**.; Hernandez, R.; Arias, A; Molina, JC., Pereira, Y. 2020. Eco-territorial adaptability of tomato crops for sustainable agricultural production in Carabobo, Venezuela. Idesia, 38(2):95-102. http://dx.doi.org/10.4067/S0718-34292020000200095

18. **Olivares, B**., Pitti, J., Montenegro, E. 2020. Socioeconomic characterization of Bocas del Toro in Panama: an application of multivariate techniques. Revista Brasileira de Gestao e Desenvolvimento Regional, 16(3):59-71. https://n9.cl/cugz

19. Bertorelli, M., **Olivares, B**. 2020. Population fluctuation of *Spodoptera frugiperda* (J.E. Smith) (Lepidoptera: Noctuidae) in sorghum cultivation in Southern Anzoátegui, Venezuela**.** Journal of Agriculture University of Puerto Rico, 104(1):1-16. https://n9.cl/9vpkz

20. **Olivares, B**., Hernández, R. 2019. Análisis regional de zonas homogéneas de precipitación en Carabobo, Venezuela. Revista Lasallista de Investigación, 16(2):90-105. https://doi.org/10.22507/rli.v16n2a9

21. **Olivares, B**., Hernández, R. 2019. Ecoterritorial sectorization for the sustainable agricultural production of potato (*Solanum tuberosum* L.) in Carabobo, Venezuela. Agricultural Science and Technology. 20(2): 339-354. https://doi.org/10.21930/rcta.vol20_num2_art:1462

22. **Olivares, B**., López, M. 2019. Normalized Difference Vegetation Index (NDVI) applied to the agricultural indigenous territory of Kashaama, Venezuela. UNED Research Journal. 11(2): 112-121. https://doi.org/10.22458/urj.v11i2.2299

23. **Olivares, B., Zingaretti, M.L.** 2019. Aplicación de métodos multivariados para la caracterización de periodos de sequía meteorológica en Venezuela. Revista Luna Azul. 48, 172:192. https://doi.org/10.17151/luaz.2019.48.10

24. **Olivares, B.,** Hernández, R; Arias, A; Molina, JC., Pereira, Y. 2018. Identificación de zonas agroclimáticas potenciales para producción de cebolla (*Allium cepa* L.) en Carabobo, Venezuela. Journal of the Selva Andina Biosphere. 6 (2): 42-54. https://n9.cl/ugm2e

25. **Olivares, B.** Hernández, R; Arias, A; Molina, JC., Pereira, Y. 2018. Zonificación agroclimática del cultivo de maíz para la sostenibilidad de la producción agrícola en Carabobo, Venezuela. Revista Universitaria de Geografía. 27 (2): 139-159. https://n9.cl/i0upn

26. Cortez, A., **Olivares, B.**, Parra, R., Lobo, D., Rodríguez, M.F. y Rey, J.C. 2018. Descripción de los eventos de sequía meteorológica en localidades de la cordillera central, Venezuela. Ciencia, Ingenierías y Aplicaciones. I (1):22-44. https://10.22206/CYAP.2018.V1I1.PP23-45

27. Parra, R., **Olivares, B.**, Cortez, A., Lobo, D., Rodríguez, M.F. y Rey, J.C. 2018. Características de la sequía meteorológica (1980-2014) en dos localidades agrícolas de los andes venezolanos**.** Revista de Investigación. 42(95):38-55. https://n9.cl/mm6d

28. **Olivares, B**. y Zingaretti, ML. 2018. Análisis de la sequía meteorológica en cuatro localidades agrícolas de Venezuela mediante la combinación de métodos multivariados. UNED Research Journal. 10 (1):181-192. http://dx.doi.org/10.22458/urj.v10i1.2026

29. **Olivares, B.**, Zingaretti, M.L., Demey Zambrano, J.A. y Demey, J.R. 2017. Aplicación del método STATIS-ACT al régimen de lluvias en la Región Oriental Venezolana. UNED Research Journal 9(1): 97-106. https://n9.cl/3mls7

Capítulos de libros

1. Paredes-Trejo, F y **Olivares, B**. 2018. El desafio de la sequía en Venezuela. En: Núñez Cobo, J. y K. Verbist (Eds.). Atlas de Sequía de América Latina y el Caribe. (pp.127-136). Paris: UNESCO/ CAZALAC.

Libros

1. **Olivares, B**. 2017. La sequía meteorológica en territorios agrícolas de Venezuela: un análisis temporal del fenómeno meteorológico y su impacto en la agricultura venezolana. Saarbrücken, Alemania, Editorial Académica Española. 145 p.

2. Hernández, R; Pereira, Y; Molina, JC; Coelho, R; **Olivares, B** y Rodríguez, K. 2017. Calendario de siembra para las zonas agrícolas del estado Carabobo en la República Bolivariana de Venezuela. Sevilla, España, Editorial Universidad Internacional de Andalucía. 247 p.

Comunicaciones en congresos

1. **Olivares, B**; 2021. Soluciones novedosas para enfermedades en banano: un enfoque con Machine Learning para el desarrollo de una agricultura sostenible. In **III Congreso Internacional de Ciencias Agropecuarias y Recursos Zoogenéticos**. (02, 2021, Los Ríos, Ecuador). Quevedo, N. & Chávez, D. (comp.). Los Ríos, Ecuador. Instituto de Investigaciones Binario. p. 39.

2. **Olivares, B**. 2021. Machine Learning and the new sustainable agriculture: applications in banana production systems. In: **III Congreso Internacional de Ciencias Agropecuarias, Tecnología e Innovación Industrial** (08, 2021, Ciudad de Valencia, Ecuador). López, J. & Muñoz, P. (comp). Ciudad de Valencia, Ecuador. Instituto de Investigaciones Binario. p. 47.

3. **Olivares, B**. 2021. Novel solutions for diseases in bananas: an approach with Machine Learning for the development of a sustainable agricultura. In **I International Congress of Organic Agriculture. (**06, 2021, Leuven, Belgium). IASS (comp.) International Association of Students in Agriculture and Related Sciences (IAAS World). Leuven, Belgium. p. 12

4. **Olivares, B**. 2021. Diagnóstico e Idoneidad Ambiental de Enfermedades Tropicales del Banano en Venezuela mediante Machine Learning. In: **IX Congreso Científico de Investigadores en Formación.** (05, 2021, Córdoba, España). Córdoba, España: UCOPress. Editorial Universidad de Córdoba.

5. **Olivares, B. 2020.** Análisis de los niveles de productividad del banano basado en las propiedades morfológicas del suelo: un estudio por escalamiento óptimo. In: **I Seminario Internacional**

de Agricultura, Agroindustria y Ambiente. (08, 2020, Riobamba, Ecuador). IASS (comp.) International Association of Students in Agriculture and Related Sciences (IAAS World). Riobamba, Ecuador.

6. **Olivares, B**; Araya-Alman, M; Rey J.C. 2020. A study of the relationship between soil properties and banana productivity in Venezuela. In **II Congreso Internacional Multidisciplinar de Investigadores en Formación**. (12, 2020, Córdoba, España). Sánchez, F. & Serrano, R. (coord.). Memoria. Córdoba, España. Universidad de Córdoba. p. 194.

7. **Olivares, B**; Rueda, MA; Vega, A. 2020. Banana productivity levels analysis based on soil morphological properties: A random forest approach. In **II Congreso Internacional Multidisciplinar de Investigadores en Formación**. (12, 2020, Córdoba, España). Sánchez, F. & Serrano, R. (coord.). Memoria. Córdoba, España. Universidad de Córdoba. p. 194.

8. **Olivares, B**; Paredes-Trejo, F. 2020 Estrés hídrico y sequía en Venezuela: desde la percepción en campo hasta la satelital. In **I Simposio Nacional de Recursos Hídricos**. (11, 2020, Caracas, Venezuela). Silva, O. (coord.). Memoria. Caracas, Venezuela. Academia general de la ingeniería y el habitad. p. 66.

9. **Olivares, B**; Rey, J.C; Lobo, D; Gómez, J.A y Landa, B. 2019. Impacto del cambio climático en zonas bananeras de la Región Central de Venezuela: El futuro de los bananos en un escenario hídrico incierto. En: Chica Pérez, A. F. y Mérida García, J. (Eds). **Creando Redes Doctorales: Investiga y Comunica.** (pp. 367-370). Córdoba, España: UCOPress. Editorial Universidad de Córdoba.

10. **Olivares, B**; Rey J.C; Lobo· D; Gómez, J.A & Landa, B. 2019. El cambio climático en zonas bananeras de la Región Central de Venezuela: El futuro de los bananos con un escenario hídrico incierto. In **III Simposio venezolano de Cambio Climático, Agricultura y Seguridad Alimentaria**. (10, 2019, Caracas, Venezuela). Goldwaser, M. & Morales, G. (eds.). Memoria. Caracas, Venezuela. Academias de Ciencias Físicas, Matemáticas y Naturales y de la Ingeniería y el Hábitat. p. 251.

11. **Olivares, B**; Cortez, A; Lobo, D; Parra, R.M; Rey, J.C & Rodríguez, M. F. 2019. Evaluación de la vulnerabilidad agrícola a la sequía meteorológica en Venezuela. In **III Simposio venezolano de Cambio Climático, Agricultura y Seguridad Alimentaria**. (10, 2019, Caracas, Venezuela). Goldwaser, M. & Morales, G. (eds.). Memoria. Caracas, Venezuela. Academias de Ciencias Físicas, Matemáticas y Naturales y de la Ingeniería y el Hábitat. p. 251.

B. Estancias de investigación

1. **Universidad La Gran Colombia**. Facultad de Ingeniería Geográfica y Ambiental. Ciudad: Armenia, Colombia. Fecha de inicio-fin: 06/01/2021 - 07/03/2021. Duración: 2 meses

2. **Universidad Nacional de Costa Rica**. Facultad de Ciencias de la Tierra y el Mar. Ciudad: Heredia, Costa Rica. Fecha de inicio-fin: 17/02/2020 - 05/07/2020. Duración: 5 meses.

3. **Universidad de Talca**. Facultad de Ciencias Agrarias. Ciudad entidad realización: Talca, Chile. Fecha de inicio-fin: 28/10/2019 - 26/01/2020. Duración: 3 meses

4. **Universidad de Panamá**. Facultad de Ciencias Agropecuarias. Ciudad: Bocas del Toro, Panamá. Fecha de inicio-fin: 23/06/2019 - 10/08/2019. Duración: 2 meses.

5. **Universidad Nacional de Córdoba**. Facultad de Ciencias Agropecuarias. Ciudad: Córdoba, Argentina. Fecha de inicio-fin: 11/02/2019 - 06/06/2019. Duración: 4 meses.

6. **Instituto de Agricultura Sostenible del Centro Superior de Investigaciones Científicas (IAS-CSIC)**-España. Ciudad: Córdoba, Andalucía, España. Fecha de inicio-fin: 27/03/2017 - 27/06/2017. Duración: 3 meses

C. Actividades de formación

1. **Curso Superior de Especialización en Riego. Manejo, Diseño y Evaluación de Sistemas de Riego.** Lugar: IFAPA. Fecha de finalización: 29/10/2021. Duración:120 horas

2. **Curso técnicas de análisis de datos en humanidades y ciencias sociales.** Lugar: Baeza, España. Fecha: 23 al 26/08/2021. Duración: 25 horas.

3. **Curso COVID-19 and the 2030 agenda in Iberoamerica: development and cooperation beyond the pandemic.** Lugar: San Lorenzo del Escorial. Fecha:12 al 16 de julio de 2021. Duración: 30 horas.

4. **Curso: Herramientas para la investigación.** Lugar: Universidad de Córdoba. Fecha de finalización: 07/05/2021. Duración:150 horas

5. **Curso de Verano "Modelos agrarios sostenibles. Alternativas a la actual crisis agroalimentaria".** Lugar: Baeza, España. Fecha: 17 al 20/08/2020. Duración: 32 horas.

6. **Curso de Verano " Ciencia de Datos y Machine Learning con Aplicaciones".** Lugar: Baeza, España. Fecha: 24 al 28/08/2020. Duración: 40 horas.

7. **Taller "Representación gráfica de resultados científicos en Python".** Lugar: Córdoba, España. Fecha: 2 al 4/12/2020. Duración: 10 horas.

8. **Taller "Diseño y validación de instrumentos de recolección de datos". Lugar:** Córdoba, España. Fecha: 30/11/2020 al 4/12/2020. Duración: 10 horas.

9. **Taller "Agroclimatología práctica: información y medios asequibles para la planificación hídrica de cultivos".** Lugar: Maracay, Venezuela. Fecha: 04 al 06/03/2021. Duración: 12 horas.

10. **Jornada técnica sobre *Phytophthora spp*. en las dehesas.** Lugar: Córdoba, España. Fecha: 01/10/2019. Duración: 8 horas.

11. **Curso de Verano: Métodos y técnicas instrumentales con Stata**. Lugar: Baeza, España. Fecha: 19 al 22/08/2019. Duración: 32 horas.

12. **Curso de datos multivariados.** Lugar: Córdoba, Argentina. Fecha: 25 al 29/03/2019. Duración: 40 horas.

13. **4th Forum of Russian and Ibero-American Rectors and Presidents.** Lugar: Sevilla, España. Fecha: 17-18/10/2019. Duración: 16 horas.

14. **Seminario-taller:** Herramientas para la identificación de riesgos climáticos en el sistema Agroalimentario: una mirada para su aplicación. Lugar: Bocas del Toro, Panamá. Fecha: 22 al 26/07/2019. Duración: 40 horas

15. **Jornada INIA- Biovegen: Technology attraction: INIA, tecnología al servicio de la agricultura.** Lugar: Madrid, España. Fecha: 23/10/2019. Duración: 5 horas

16. **Taller para la mejora de la empleabilidad.** Fecha: 05 de abril de 2018

17. **Seminario de actualidad:** la investigación con especies forestales (Quercus y Pinus): selvicultura, fisiología, biología molecular y teledetección. Fecha: 06 de junio de 2018

18. **Seminario de actualidad:** Evolución histórica de las técnicas de explotación extractiva y metalurgia en minas de Almadén. Fecha: 15 de junio de 2018

19. **Seminario de actualidad:** VIII Encuentro de estudiantes de doctorado. Fecha: 28 de junio de 2018.

20. **Seminario de actualidad: New studies on recycled concrete, CO_2 capture and future research lines.** Fecha: 09 de octubre de 2018

21. **Seminario de actualidad:** Jornada sobre interpretación y valoración del patrimonio inmaterial y monumental de montilla, dedicada a D. José M.ª Sánchez Molero. Fecha: 09 de noviembre de 2018

22. **Jornada formativa doctoral:** el Doctorado en la Universidad de Córdoba: Marco Normativo, Procesos y Procedimientos. Fecha: 12 de diciembre de 2018

23. **Taller Climatones Jóvenes Ecoemprendedores 2018.** Fecha: 04-05 de mayo de 2018

24. **VIII Jornadas de divulgación de la investigación en biología molecular, celular, genética y biotecnológica.** Fecha: 13-15 de junio de 2018

25. **Curso de verano:** el sector agroalimentario del 2030: agroindustria 4.0 y economía verde. Fecha: 20-23 de agosto de 2018

26. **Jornada Científico-Técnica CeiA3 y VI Jornada AEL: Leguminosas en la agricultura y la alimentación.** Fecha: 22-23 de octubre de 2018

27. **III Jornada de la Cátedra AgroBank "El futuro de la agricultura de regadío en un escenario hídrico incierto".** Fecha: 14 de noviembre de 2018

28. **II Seminario/Taller para escritores independientes.** Fecha: 04 de diciembre de 2018

29. Curso: **técnicas avanzadas para la propagación masiva de semillas de Musáceas**. Fecha: 29 noviembre 2016. Duración: 8 horas.

30. Jornada: **Avances en los métodos de análisis de suelos y plantas (4ª edición)**. Fecha: 12 al 14 de junio de 2017. Duración: 24 horas.

31. Seminario de actualidad UCO: **Impacto del cambio climático en la agricultura.** Fecha: 19 de mayo de 2017. Duración: 3 horas.

32. Seminario de actualidad UCO: **La Biotecnología como herramienta para superar los desafíos de la agricultura del futuro.** Fecha: 2 de junio de 2017.

33. Seminario de actualidad UCO: **La agricultura y el mercado eléctrico Evolución 1894-2017.** Fecha: 11 de mayo de 2017.

34. Seminario de actualidad UCO: **Metodología de trabajo para reconstrucciones virtuales y uso de hardware libre para la documentación de la Mezquita-Catedral de Córdoba.** Fecha: 12 de mayo de 2017.

35. Seminario de actualidad UCO: **La investigación sobre el Patrimonio Industrial: antecedentes y estado de la cuestión.** Fecha: 31 de mayo de 2017.

D. Becas y distinciones

1. **Becario** (Modalidad estudiante). Beca de matrícula y alojamiento para el curso de verano denominado: técnicas de análisis de datos en humanidades y ciencias sociales. Sede: Universidad Internacional de Andalucía, Antonio Machado. Convocatoria 2021.

2. **Becario** (Modalidad estudiante). Beca completa para el curso de verano denominado: COVID-19 and the 2030 agenda in Iberoamerica: development and cooperation beyond the pandemic. Sede: Universidad Complutense de Madrid. Convocatoria 2021.

3. **Becario** (Modalidad alumnado de doctorado). Beca para la realización de estancias de investigación en la Universidad La Gran Colombia. Convocatoria del Campus de Excelencia Internacional de Medio Ambiente, la Biodiversidad y el Cambio Global (CEICambio). Convocatoria 2020.

4. **Becario** (Modalidad estudiante). Beca de matrícula y alojamiento para el curso de verano denominado: Modelos agrarios sostenibles. Alternativas a la actual crisis agroalimentaria. Sede: Universidad Internacional de Andalucía, Antonio Machado. Convocatoria 2020.

5. **Becario** (Modalidad estudiante). Beca de inscripción del Congreso Nacional de Medio Ambiente (CONAMA-2021) y Encuentro Iberoamericano sobre Desarrollo Sostenible. Convocatoria 2020.

6. **Becario** (Modalidad alumnado de doctorado). Beca para la realización de estancias de investigación en la Universidad de Panamá. Convocatoria del Campus de Excelencia Internacional de Medio Ambiente, la Biodiversidad y el Cambio Global (CEICambio). Convocatoria 2019.

7. **Becario** (Modalidad alumnado de doctorado). Beca para realización de movilidad en la Universidad de Talca, Chile. Convocatoria Erasmus+ para la movilidad de estudiantes entre el Campus De Excelencia Internacional Agroalimentario (ceiA3) y países asociados (Acción KA107).

8. **Becario** (Modalidad alumno de doctorado). Beca de investigación de la SEGIB y la Fundación Carolina para realizar estancia en la Universidad Nacional de Costa Rica. Convocatoria 2019.

9. **Becario** (Modalidad estudiante). Beca de inscripción y matricula en curso de verano denominado: Métodos y técnicas instrumentales con Stata. Sede: Universidad Internacional de Andalucía, Antonio Machado. Convocatoria 2019.

10. **Premio anual "Evaluaciones Destacadas 2018"**. Reconocimiento a la labor y colaboración con la difusión de la investigación científica y tecnológica en Iberoamérica durante el año 2018. Centro de Información Tecnológica, La Serena, Chile.

11. **Becario** (Modalidad alumnado de doctorado). Becas Iberoamérica. Santander Investigación. Movilidad Internacional. Convocatoria 2018-2019.

12. **Becario** (Modalidad investigador iberoamericano). Beca completa de participación en el 14º Congreso Nacional de Medio Ambiente (CONAMA) y 14º Encuentro Iberoamericano sobre Desarrollo Sostenible (EIMA). Convocatoria 2018.

13. **Becario** (Modalidad estudiante). Beca de inscripción en curso de verano denominado el sector agroalimentario del 2030: agrioindustria 4.0 y economía verde de la Universidad Internacional de Andalucía, sede Antonio Machado. Convocatoria 2018.

14. **Reconocimiento honorifico** de Cónsul de Córdoba. Catedra Córdoba, Ciudad Mundo. Vicerrectorado de estudiantes y programas de movilidad. Universidad de Córdoba. Periodo II-2018.

15. **Becario** (Modalidad estudiante) para estudios de doctorado en la UCO por la Asociación Universitaria Iberoamericana de Postgrado (España). Convocatoria 2015-2016.