



# **DESARROLLO DE UN MODELO PREDICTIVO DEL BROTE DE LA SIGATOKA NEGRA PARA LAS PLANTACIONES DE PLÁTANO AL SUR DEL LAGO DE MARACAIBO.**

Autor: Juan Antonio Freitez T.  
Tutor: Magdiel Ablan Bortone  
Co-Tutor: Carlos Gómez

Trabajo presentado como requisito parcial para optar al título de  
MAGISTER SCIENTIAE

UNIVERSIDAD DE LOS ANDES  
MÉRIDA, VENEZUELA  
Mérida, 16 de Febrero 2007



## ÍNDICE

ÍNDICE DE FIGURAS .....	iii
ÍNDICE DE TABLAS.....	iv
Resumen.....	vi
Agradecimientos.....	viii
CAPITULO I.....	1
1.1 INTRODUCCIÓN.....	1
1.2 DESCRIPCION DEL PROBLEMA.....	6
1.3 ANTECEDENTES.....	7
1.4 OBJETIVOS GENERAL .....	15
1.5 OBJETIVOS ESPECÍFICOS.....	15
CAPITULO II - MARCO TEÓRICO.....	16
2 INTRODUCCIÓN.....	16
2.1 DESCRIPCIÓN DE LA ZONA DE ESTUDIO.....	16
2.2 EL PLÁTANO.....	18
2.2.1 ETAPAS DE DESARROLLO DE UNA HOJA DEL PLATANO .....	20
2.3 LA SIGATOKA NEGRA.....	22
2.3.1 EVALUACIÓN DEL ESTADO DE LA INFECCIÓN .....	26
2.3.2 FACTORES QUE INFLUYEN EN EL DESARROLLO DE LA ENFERMEDAD.....	32
2.3.3 VARIABLE CLIMATOLÓGICAS.....	34
2.4 GENERACIÓN DEL MODELO PREDICTIVO.....	38
2.5 EXPLORACIÓN Y LIMPIEZA DE DATOS .....	50
CAPITULO III - MARCO METODOLÓGICO.....	61
3 INTRODUCCIÓN.....	61
3.1 Pre-Procesamiento de los datos.....	61
3.1.1 Búsqueda de datos anómalos o extremos: .....	61
3.1.2 Búsqueda de datos ausentes: .....	62
3.1.3 Integración de las bases de datos: .....	62
3.1.4 Selección de las variables: .....	63
3.2 Procesamiento de los datos.....	66
3.2.1 Búsqueda del modelo de pronóstico temprano o preaviso: .....	67
CAPITULO IV - RESULTADOS.....	69
4 INTRODUCCIÓN.....	69
4.1 EXPLORACIÓN Y LIMPIEZA DE DATOS .....	69
4.1.1 Datos atípicos.....	69
4.1.2 Datos ausentes.....	71
4.2 INTEGRACIÓN Y SELECCIÓN DE DATOS.....	72
4.2.1 Caso de estudio 1: El lapso de exploración 15 hasta 30. 73	
4.2.2 Caso de estudio 2: El lapso de exploración 15 hasta 140 días.....	79

<a href="#"><u>4.3 MODELO PREDICTIVO.....</u></a>	<a href="#"><u>82</u></a>
<a href="#"><u>4.3.1 Caso de estudio 1: La lapso de exploración 15 hasta 30 días.....</u></a>	<a href="#"><u>83</u></a>
<a href="#"><u>4.3.2 Caso de estudio 2: La lapso de exploración 15 hasta 140. ....</u></a>	<a href="#"><u>86</u></a>
CAPITULO V - DISCUSIÓN DE RESULTADOs.....	88
capitulo vi - CONCLUSIONES y Recomendaciones.....	97
capitulo vii - REFERENCIAS.....	102
capituloviii - APÉNDICE A.....	109
APÉNDICE B.....	111

## ÍNDICE DE FIGURAS

Figura N° 1. La numeración utilizada para las hojas del plátano..	10
Figura N° 2. Sistema de calificación de la severidad. ....	12
Figura N° 3. Ventana con retardo para un lapso de tiempo de exploración dado.....	14
Figura N° 4. Costos de Producción. Composición Porcentual 4.3.2.3.....	18
Figura N° 5. Etapas de desarrollo de una hoja de banano 4.3.2.3.21	
Figura N° 6. Estadio de la Sigatoka Negra.....	22
Figura N° 7. Conteo de hojas para la obtención de la emisión foliar. ....	31
Figura N° 8. Proceso de extracción de conocimiento en bases de datos.....	40
Figura N° 9. Gráfico de caja de estado de evolución de la enfermedad.....	57
Figura N° 10. Reajuste de las mediciones de las variables.....	63
Figura N° 11. Grafico del comportamiento de la correlación y el valor p. ....	74
Figura N° 12. Temporada de lluvia.....	93
Figura N° 13. Temporada de sequía.....	93

## ÍNDICE DE TABLAS

Tabla N° 1. Síntomas de las manchas foliares de Sigatoka Negra asociados a los diferentes estadios de la enfermedad.....	23
Tabla N° 2. Coeficientes arbitrarios de severidad de la Sigatoka Negra.....	29
Tabla N° 3. Codificación de los casos de estudios para un lapso de exploración n números de días.....	65
Tabla N° 4. Tabla de calificación cualitativa.....	66
Tabla N° 5. Resultados de la prueba de Grubbs aplicada a las variables biológicas.....	69
Tabla N° 6. Resultados de la prueba de Grubbs aplicada a las variables climáticas.....	70
Tabla N° 7. Tabla del número de valores ausentes por variable climatológica.....	72
Tabla N° 8. Tabla de evaluación cualitativa de la correlación entre las variables Biológicas y climatológicas para un lapso de 30 días. ....	75
Tabla N° 9. Tabla de las variables climatológicas seleccionada para el lapso de 30 días. ....	75
Tabla N° 10. Evaluación cualitativa de la correlación para el periodo de lluvia.....	76
Tabla N° 11. Variables climatológicas seleccionadas para el periodo de lluvia. ....	76
Tabla N° 12. Evaluación cualitativa de la correlación para el periodo de sequía.....	77
Tabla N° 13. Evaluación cualitativa de la correlación para el periodo de sequía (Continuación).....	77
Tabla N° 14. Variables climatológicas seleccionadas para el periodo de sequía. ....	77
Tabla N° 15. Variables climatológicas seleccionadas para el periodo de sequía (Continuación). ....	79
Tabla N° 16. Variables climatológicas seleccionadas sin discriminación por temporada. ....	79
Tabla N° 17. Variables climatológicas seleccionadas sin discriminación por temporada (continuación). ....	80
Tabla N° 18. Variables climatológicas seleccionadas para la temporada de lluvia.....	81
Tabla N° 19. Variables climatológicas seleccionadas para la temporada de sequía.....	82
Tabla N° 20. Tabla de la correlación y porcentaje de error de los modelos generados para datos globales, lapso de 30 días. ....	83
Tabla N° 21. Tabla de la correlación y porcentaje de error de los modelos generados, caso temporada de lluvia, lapso de 30 días. .	84

Tabla N° 22. Tabla de la correlación y porcentaje de error de los modelos generados, caso de estudio temporada de sequía, lapso de 30 días. ....	86
Tabla N° 23. Tabla de la correlación y porcentaje de error de los modelos de predicción de la severidad (SE), lapso de 140 días. ...	87
Tabla N° 24. Algunas de las variables climatológicas seleccionadas para predecir la severidad de la enfermedad (SE). ....	88
Tabla N° 25. Resumen estadístico de las variables biológicas para temporada de sequía.....	91
Tabla N° 26. Resumen estadístico de las variables biológicas para temporada de lluvia.....	92
Tabla N° 27. Resumen estadístico de variables climatológica par temporada de Sequía.....	109
Tabla N° 28. Resumen estadístico de variables climatológica par temporada de lluvia.....	109
Tabla N° 29. Análisis multivariantes de relación de la severidad y las variables climáticas para un lapso de exploración de 140 días. ....	111
Tabla N° 30. Análisis multivariantes de relación de la severidad y las variables climáticas para un lapso de exploración de 140 días, periodo de lluvia. ....	111
Tabla N° 31. Análisis multivariantes de relación de la severidad y las variables climáticas para un lapso de exploración de 140 días, periodo de sequía. ....	111

## RESUMEN

El cultivo de musáceas (plátanos y bananos) es una actividad agrícola de importancia para numerosos países tropicales. En Venezuela, la Zona sur del Lago de Maracaibo es la región que más se destaca en este rubro por poseer el 55% de la superficie sembrada de Musáceas en el país. En la última década, la Sigatoka Negra causada por el hongo ***Mycosphaerella Fijiensis Morelet*** ha causado preocupación a los productores de plátano de la zona por su impacto directo en la producción del cultivo y por ende en los costos (48.6 %) necesarios para su control, siendo la aplicación aérea de agroquímicos con periodicidad mensual la técnica más utilizada para combatir la enfermedad. El desarrollo de modelos que permitan establecer un sistema de pronóstico temprano del brote o estadios de enfermedad tiene gran importancia como medio de racionalizar el uso de agroquímicos, reducir los costos de producción y mantener las plantaciones en buen estado sanitario. El objetivo de este trabajo fue elaborar un modelo que permita desarrollar un sistema bioclimático de pronóstico de la Sigatoka Negra en los cultivos de plátano en la zona Sur del Lago de Maracaibo, a fin de decidir el momento más oportuno de efectuar la aplicación de fungicida. A tal fin se elaboraron modelos que tratan de relacionar indicadores biológicos de ocurrencia de la enfermedad con datos meteorológicos recolectados en la Estación Local Chama – INIA que se encuentran en el repositorio bioclimático del Centro de Cálculo de la Universidad de los Andes (CECALCULA). Para el desarrollo del modelo se consideraron técnicas de minería de datos y herramientas que permiten la búsqueda iterativa de correlaciones entre la suma o promedio de



registros de días consecutivos (ventanas temporales) de las variables climáticas y la fecha de medición de las variables biológicas. Además del criterio de discriminación por temporada de lluvia para el estudio del comportamiento dinámico estacional de la enfermedad. Los modelos generados son de calidad variable y presentaron dinámica estacionaria. En algunos casos logran predecir la severidad de la enfermedad, con y sin discriminación por temporada climática (lluvia y sequía) con correlaciones hasta de 79,22 %, donde destacan las variables humedad relativa, precipitación, velocidad del viento y energía solar como las que mejor predicen la severidad de la enfermedad para la temporada seca ( $r^2 = 0.79$ ), mientras en el caso de la temporada lluviosa lo hacen la temperatura del aire, evapotranspiración, humedad relativa y precipitación ( $r^2 = 0.73$ ). Se recomienda validar los modelos con nuevos datos de campo a una escala piloto para hacer el diseño definitivo del sistema de pronóstico a ser utilizado por los productores de plátano.

## Agradecimientos

*Agradezco infinitamente a Dios, por darme la vida y la salud necesaria para trabajar en este proyecto de investigación a pesar de los contratiempos.*

*Gracias señor por darme el don del aprendizaje, hecho que motivó la realización de mis estudios de maestría y concretamente este proyecto de grado.*

*Agradezco de una manera muy especial con mucha sinceridad y humildad:*

*A mi familia (padres y hermanos) por el cariño incondicional y apoyo moral que me han brindado durante todos los años de mi existencia.*

*A mi tutora Magdiel y -tutor Carlos, por ser la guía precisa que necesité durante el desarrollo de la maestría. Dios me brindó nuevamente la oportunidad de adquirir nuevos sabios conocimientos a través de ustedes en esta etapa importante de mi vida.*

*A mis compañeros y amigos de postgrado, que con su paciencia y apoyo moral me auxiliaron cuando los necesitaba.*

*A la planta profesoral del Postgrado en Modelado y Simulación de Sistemas y de Computación por transmitirme de manera eficiente los conocimientos necesarios para el desarrollo de este trabajo especial de grado para optar al título de Magister Scientiae en Modelado y Simulación de Sistemas.*

*Al personal que labora en el Centro y Simulación de Modelos (CeSiMo), por el apoyo que me brindaron en los últimos años.*

*A todas aquellas personas que de alguna manera contribuyeron  
para la ejecución de esté proyecto.*



## CAPITULO I

### 1.1 INTRODUCCIÓN

El cultivo de *musáceas* (plátanos y bananos) es una actividad agrícola de importancia para numerosos países tropicales por su aporte comercial, ya que es considerado como el cuarto producto alimenticio a nivel global después del arroz, trigo y maíz en términos del producto interno bruto 4.3.2.3. En Venezuela la siembra de plátano (*Musa* AAB) se basa exclusivamente en la variedad “*Hartón*” y en el tipo de banano (*Musa* AAA) “*Cavendish*” 4.3.2.3. Al igual que otros cultivos, las musáceas se encuentran expuestas a situaciones ambientales y fitosanitarias adversas que afectan las plantaciones en cuanto a su producción final, tal es el caso de la Sigatoka Negra (*Mycosphaerella fijiensis* Morelet) 4.3.2.3.

La Sigatoka Negra es una enfermedad causada por el hongo *Mycosphaerella fijiensis* Morelet y es considerada mundialmente como una de las enfermedades más destructivas en las musáceas, ya que produce necrosis foliar en las hojas de las plantas de plátano y banano, y por consiguiente los frutos no poseen las características necesarias para su consumo o para una efectiva actividad comercial. Esta enfermedad ha logrado diseminarse a todas las áreas donde se cultiva el plátano o el banano. En 1991 en Venezuela, fue detectada por primera vez en el Municipio Catatumbo, Estado Zulia a través del Servicio Autónomo de Sanidad Agropecuaria (SASA). La diseminación, además de los factores humanos y meteorológicos, se debió al arrastre de material infectado por las corrientes del Catatumbo. Para diciembre de 1992 ya se había extendido a los estados Mérida,

#### Trujillo y Barinas 4.3.2.3.

El combate contra la Sigatoka Negra está originando un cambio del patrón tecnológico de los productores quienes, por falta de experiencia en su manejo, emplean medidas y acciones de control fitosanitario utilizadas en otros países con diferentes características edafoclimáticas y de cultivo 4.3.2.3. El control de la Sigatoka Negra es obligatorio para obtener una producción de valor comercial aceptable. Esto eleva los costos de producción y por ende los pequeños productores se ven afectados a tal magnitud que en ocasiones cambian de rubro 4.3.2.3. Los productores comerciales han logrado disminuir la evolución de la Sigatoka Negra aplicando controles basados en tratamientos químicos y prácticas culturales en paralelo, como por ejemplo, control de malezas, eliminación de las hojas secas, entre otros. Pero la aplicación de controles químicos trae como consecuencia efectos de gravedad sobre el medio ambiente y riesgo de generar resistencia del hongo al fungicida 4.3.2.3.

Es posible reducir notablemente la frecuencia de los tratamientos si éstos se realizan en el marco de un sistema de predicción de la enfermedad 4.3.2.3. La aplicación correcta de controles químicos para el control de la Sigatoka Negra, está basado en el conocimiento de la prevalencia y el progreso de la enfermedad. Según Gahul 4.3.2.3, el uso de un método de control de las enfermedades o de otro depende de diferentes factores como la cultura fitopatológica, los recursos disponibles, las formas de tenencia de las plantaciones plataneras y las formas de asociación gremial existentes 4.3.2.3.

En forma general se puede hablar de dos sistemas básicos de manejo y detección de la Sigatoka Negra, el método de Stover 4.3.2.3 y el método de preaviso biológico o de predicción 4.3.2.3. El de Stover, se basa esencialmente en determinar el índice de severidad de la enfermedad en una plantación, a partir de la hoja más joven manchada (HJM). Esto se logra realizando evaluaciones semanales de la quemazón foliar en cada una de las plantas. Mientras que el método de preaviso biológico realizado por investigadores tales como: Bureau 4.3.2.3, Chuang y Jeger 4.3.2.3, abajo, Ganry y Meyer 4.3.2.3, Jimenez 4.3.2.3, y otros; se basan en la rata de cambio de la enfermedad según la escala de Meredith y Lawrence 4.3.2.3 y usan fundamentalmente fungicidas sistémicos curativos en aceite puro: Control químico = Aceite de petróleo + fungicida 4.3.2.3. Este sistema permite determinar con anticipación, que va desde unos cinco días a tres semanas o más, cuándo es necesario aplicar un tratamiento si continúa la rata de desarrollo existente en determinado momento. A medida que se va obteniendo información adicional, si las condiciones son más favorables a la enfermedad, se determinará que los ciclos deben acortarse. Si las condiciones ambientales son estables no habrá cambios en el ciclo establecido y si es negativo el ciclo se alargará 4.3.2.3.

Dentro de los sistemas de pronósticos o de prevención que han logrado efectividad contra la enfermedad se encuentra el estudio realizado en Costa Rica en el año 1995 por el Centro Agronómico Tropical de Investigación y Enseñanza (CATIE), el cual provee alertas tempranas a los productores para el control de la

Sigatoka Negra, basándose en la correlación entre la duración acumulada de la lluvia por períodos de cuatro semanas y el nivel de infección de la hoja N°. 4. El sistema es simple, funciona con reglas de decisión objetivas y permite hacer una previsión a dos semanas, lo que favorece un combate de la enfermedad en estados de desarrollos sensibles a los fungicidas 4.3.2.3. Bureau y colaboradores 4.3.2.3, mostraron la posibilidad de realizar el control de la Sigatoka Negra en plátanos en base a tratamientos pronosticados en función del estado de evolución y la evaporación Piche. Ambos sistemas sin embargo deben ser estudiados localmente, porque la lluvia no es la única forma de humedecimiento de los tejidos foliares y en áreas donde la pluviometría no es tan elevada, los rocíos pueden jugar un papel importante en el comportamiento de la enfermedad, por lo que puede ser más útil medir la duración del humedecimiento de las hojas y no la duración de las lluvias 4.3.2.3.

Por ser la Zona sur del Lago de Maracaibo de Venezuela la región que se destaca por poseer el 55% de la superficie sembrada de Musáceas en el país, diferentes instituciones de los Estados Mérida y Zulia crearon una Red de Estaciones Bio-climáticas para ambos Estados 4.3.2.3, cuyo objetivo es proveer con apoyo tecnológico datos ordenados y clasificados para fortalecer las investigaciones en el rubro del plátano y apoyar a los productores de la zona u otras investigaciones 4.3.2.3.

La propuesta para la zona Sur del Lago de Maracaibo está basada en elaborar un modelo de pronóstico de alertas tempranas de la incidencia de la enfermedad Sigatoka Negra que sea de fácil acceso a los productores de la región. Utilizando los datos



biológicos de la enfermedad y climáticos recolectados en la Estación Local Chama del Instituto Nacional de Investigaciones Agrícolas (INIA), durante el período correspondiente al 2001 hasta mediados del 2005. La construcción de este modelo repercutirá en el uso racional de agroquímicos para el combate de la enfermedad, reduciendo los posibles impactos sobre el medio ambiente.

## 1.2 DESCRIPCION DEL PROBLEMA

El plátano (*Musa* AAB cv. Hartón) constituye un producto básico en la dieta de muchos países latinoamericanos, donde es cultivado por un gran número de pequeños y medianos productores, quienes hasta hace poco habían enfrentado los problemas del cultivo con acierto. Sin embargo, la aparición de la Sigatoka Negra, *Mycosphaerella fijiensis* M. 4.3.2.3, desde la década del 90 ha traído consecuencias negativas para los productores de plátano por su impacto directo en la producción del cultivo y por ende en los costos necesarios para su control. En la actualidad los mismos, representan el 48,56 % del costo de la producción de las plantaciones de plátanos. En Venezuela, esta temible enfermedad fue detectada en el año 1991, en el Municipio Catatumbo del estado Zulia [5]. Desde entonces se diseminó a toda la zona platanera y en la actualidad se encuentra en todos los estados productores de musáceas del país, tales como Táchira, Mérida, Trujillo, y Barinas entre otros. El patógeno es capaz de ocasionar severos daños al follaje de la planta destruyendo así su capacidad de respiración y fotosíntesis, lo cual trae como consecuencia una reducción del rendimiento y calidad de la fruta. Si el ataque se produce sobre todas sus hojas jóvenes, es causal de la pérdida total de la producción. Las frutas obtenidas de plantaciones atacadas por este mal sufren maduraciones tempranas, ocasionando pérdidas en el transporte de la mercancía. Por otro lado, el tratamiento para su control con agroquímicos conduce elevados costos para los productores. Esto sin mencionar los efectos nocivos de los agroquímicos al ambiente4.3.2.3, 4.3.2.3, 4.3.2.3.

### 1.3 ANTECEDENTES

Los controles implementados para combatir la enfermedad elevan los costos de producción ya están basado primordialmente en la aplicación de fungicidas de forma aérea, de las cuales se requieren 33 a 36 aplicaciones o más por año para lograr cierta efectividad 4.3.2.3, además se debe considerar el impacto ambiental causado por el uso de los fungicidas. Entre las medidas de control efectivas que se han realizado en varios países se distinguen las tres siguientes:

- a. **Uso de los productos químicos:** Son las medidas de control más usada actualmente para atacar o reducir la enfermedad. La mayor preocupación de esta medida es el favorecer a la epidemia de Sigatoka Negra por aplicaciones inadecuadas de fungicidas o formulaciones no aptas, deficiencia en las técnicas de aplicación que resultan pobres en cobertura y los efectos sobre el medio ambiente. En tal sentido, Nava y Villarreal 4.3.2.3 implementaron una metodología que permite mantener un manejo agronómico adecuado, logrando la convivencia de las plantas de musáceas con la sigatoka negra, con nula o mínimas aplicaciones de fungicidas al follaje, al aplicar medidas que permiten controlar los niveles de fertilización nitrogenada y potásica mediante la adición de B, Mg o Zn como complementos.
- b. **Mejoramiento de las plantas:** Las técnicas de hibridación son frecuentemente usada con el fin de obtener plantas más resistentes a enfermedades y con mayor rendimiento. Sobre estas perspectivas 4.3.2.3 varios investigadores lo han venido utilizando en las musáceas para contrarrestar la sigatoka negra. En Venezuela, García y Sosa 4.3.2.3 evaluaron el

comportamiento agronómico del híbrido de plátano resistente a sigatoka negra, denominado FHIA-21, de manera de poder ofrecerle al pequeño y mediano productor una alternativa en el cultivo del plátano. Los resultados obtenidos de este trabajo indican la buena adaptación del híbrido a la zona y su resistencia a la sigatoka negra, constituyéndose en una alternativa viable de producción para el pequeño y mediano productor de plátano sin tener que recurrir a agroquímicos para el control de la enfermedad. No obstante, aunque algunos autores reportan un incremento de la resistencia a la enfermedad y un alto rendimiento de las plantas, sin embargo, posee muchos detractores que consideran que su aceptación implicará una limitación para su comercialización nacional e internacionalmente, como lo han vivido y lo siguen experimentando los agricultores. Las razones que alegan es que estos híbridos en general poseen baja adaptación al manejo post cosecha tradicional y características organolépticas diferentes a los clones comerciales y en el caso particular del FHIA-21 existe restricción para su siembra por el SASA, debido a la presencia en su genoma del virus del mosaico del banano CMV que en Venezuela no está muy diseminado 4.3.2.3.

- c. **Manejo integrado de la enfermedad:** Se trata de convivir con la enfermedad una vez ya detectada a través del mantenimiento de un programa permanente de prevención y control. Para realizar éste, se siguen algunas etapas, como por ejemplo: regulación de la población, deshoje, control de malezas, fertilización de plantaciones establecidas, deshoje

fitosanitario, manejo y distribución de desechos y control químico 4.3.2.3, 4.3.2.3.

Alternativamente también se han desarrollado trabajos (4.3.2.3, 4.3.2.3, 4.3.2.3, 4.3.2.3) que buscan hacer más eficiente la aplicación de los fungicidas basados en el pronóstico temprano de posibles brotes de esta enfermedad, permitiendo reducir costos, el impacto ambiental y daños mayores. En tal sentido, se han desarrollado diversas técnicas de pronóstico o advertencia temprana y comercialmente se han aplicado dos sistemas de pronósticos para decidir las aplicaciones de fungicidas en plantaciones de Bananos. Ambos fueron originalmente desarrollados para la Sigatoka Amarilla, *Mycosphaerella musicola*. Klein 4.3.2.3 en Honduras utilizó un método basado en el conteo de estrías amarillas en el semilimbo<sup>1</sup> izquierdo de las hojas más jóvenes de la planta, hojas 2, 3 y 4. Si el número de ellas estaba entre 55 y 100 en la hoja 4 ó 3 (Figura N° 1), se recomienda una aplicación de aceite agrícola, pues el mismo autor había determinado que el aceite era capaz de parar o retardar en forma significativa el desarrollo posterior de las estrías jóvenes 4.3.2.3. Con la llegada de la Sigatoka Negra a Honduras, el sistema se volvió ineficiente, por la mayor agresividad y potencial de inoculación<sup>2</sup> de *M. fijiensis* comparado con *M. musicola*, el primero no podía ser controlado en forma satisfactoria con fungicidas de tipo protector.

---

<sup>1</sup> La mitad izquierda o derecha del limbo o lámina de la hoja, separadas por el nervio medio

<sup>2</sup> Introducir en el organismo un virus, micro organismo, agente causal o patógeno

En la Figura N° 1, se muestra un ejemplo como la primera hoja que presenta 10 ó más lesiones discretas necrosadas y maduras es la número 4.

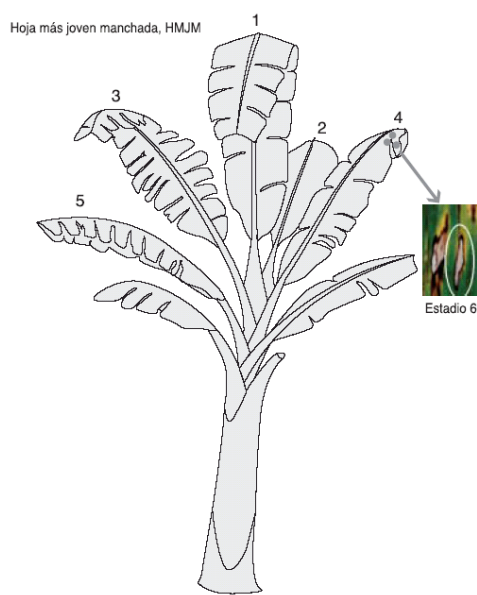


Figura N° 1. La numeración utilizada para las hojas del plátano.

En 1972, Ganry y Meyer 4.3.2.3 desarrollaron un sistema basado en mediciones semanales de evaporación Piche<sup>3</sup>, acoplado con la evaluación del desarrollo de los estadios iniciales de los síntomas en las hojas más jóvenes (Figura N° 1). La tasa de emisión foliar se registra también cada semana y a los estadios de los síntomas se les califica asignando un coeficiente de ponderación de acuerdo con la edad de la hoja donde están presentes, calificando con mayor severidad (Figura N° 2) la presencia de estadios 2 de los síntomas en una hoja 3, que en una hoja 4, por ser la primera más joven. Las sumas de los coeficientes obtenidos de acuerdo con las infecciones y la edad de la hoja, son posteriormente

<sup>3</sup> La cantidad de agua por unidad de área y por unidad de tiempo evaporada en una pequeña superficie expuesta al aire libre

corregidos para los estados de desarrollo de la hoja candela, para conjuntamente con el ritmo de emisión foliar obtener el valor de estado de evolución de la enfermedad, variable utilizada para definir las aplicaciones de fungicidas sistémicos en mezcla ó suspensión en aceite agrícola. Este sistema de evaluación de los estados tempranos de la enfermedad fue luego utilizado en la Sigatoka Negra en Camerún y fue ampliamente utilizado en Costa Rica 1988 y 1991 4.3.2.3. En las primeras temporadas de implementar esta técnica se obtuvieron resultados favorables en el control de la enfermedad. Pero posteriormente la Sigatoka Negra se volvió más resistente a tratamiento con agroquímicos y los niveles de infecciones aumentaron, por lo cual los productores optaron por volver a las aplicaciones de fungicidas con base en intervalos calendarizados, dependiendo de si el fungicida era protector o sistémico, de las condiciones de lluvia y del estado de infección de la finca, lo cual continua manejándose de esta manera hasta la actualidad. También se ha utilizado en parcelas de plátano, donde los resultados son muy satisfactorios, inclusive hasta la fecha, principalmente si el sistema de siembra de plátano es anual ó bianual, pues la uniformidad en el crecimiento de las plantas favorece mucho el control de los niveles de inóculo, y porque el plátano es menos susceptible a la Sigatoka Negra que el banano 4.3.2.3, lo cual le hace responder a los tratamientos fungicidas en mejor forma que el cultivo de banana. Jiménez 4.3.2.3 demostró en el cultivo del plátano, que las variaciones en el nivel de infección de la hoja 4 está correlacionada con la duración acumulada de las lluvias durante las primeras cuatro semanas anteriores a la fecha de evaluación de la enfermedad, usándola para pronosticar el momento de hacer las aplicaciones

de fungicidas en este cultivo. El sistema de pronóstico descrito es una derivación del sistema de Ganry y Meyer, descrito anteriormente.

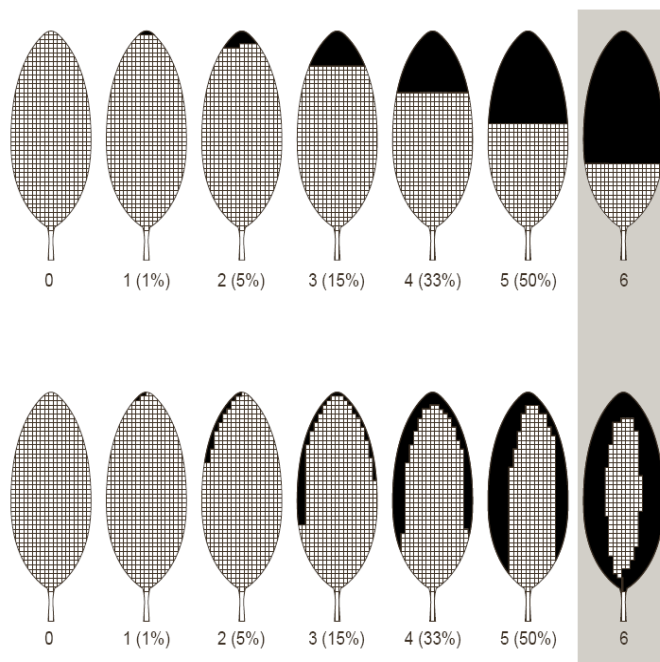


Figura N° 2. Sistema de calificación de la severidad.

En la Figura N° 2 se representa entre paréntesis el porcentaje del área de la hoja afectada por la enfermedad y su correspondiente grado de la enfermedad.

Chuang y Jeger 4.3.2.3abajodesarrollaron un modelo para pronosticar la tasa de progreso de la Sigatoka Negra en Taiwán, basado en valores previos de la incidencia de la enfermedad y en el número de días con humedad relativa del 90% y la lluvia acumulada en las semanas precedentes. El sistema de pronóstico desarrollado por ellos dio valores de precisión considerados bastante aceptables, sin embargo, su utilización a nivel comercial no se ha dado.



En general, existen básicamente dos alternativas al momento de desarrollar modelos de pronóstico de enfermedades de planta basados en variables climatológicas 4.3.2.3. La primera, son los modelos mecanicistas usados para mejorar el entendimiento de la propagación de la enfermedad. Estos modelos son generalmente funciones de las propiedades físicas del fenómeno observado y por consiguiente lo reflejan bastante bien. Sin embargo, pueden ser muy complejos involucrando muchos parámetros, lo que limita su uso para predecir la severidad de la enfermedad. La segunda alternativa son los modelos empíricos basados en la búsqueda de patrones o funciones que expliquen de manera sencilla los fenómenos, lo que se conoce como minería de datos. Su ventaja principal es que estos modelos son normalmente más simples de usar a escala de campo. Por otro lado, los mismos autores plantean el uso de la metodología desarrollada por Coakley y colaboradores 4.3.2.3abajo para encontrar estos patrones. La misma consiste en la búsqueda de funciones o correlaciones a partir de las variables climatológicas que tenga mayor relación con la severidad de la enfermedad, lo cual se logra aplicando métodos iterativos de exploración de la relación entre las variables climatológicas de una cantidad variable de días consecutivos de tiempo (ventanas) y la enfermedad. La ventana se ubica antes de la fecha de medición de la variable biológica, este lapso de tiempo entre la fecha de medición de la variable biológica y la fecha de inicio de la amplitud de la ventana se conoce como retraso, dicho retraso es variable en un lapso de exploración que antecede ala medición de la variables biológicas, por lo cual, se generan “ $n$ ” ventanas en una lapso de tiempo de exploración. En la Figura Nº 3 se muestra una ventana de

amplitud de 3 días., la cual se encuentra con un retardo temporal respecto a la fecha de la medición (E7) de 5 días.

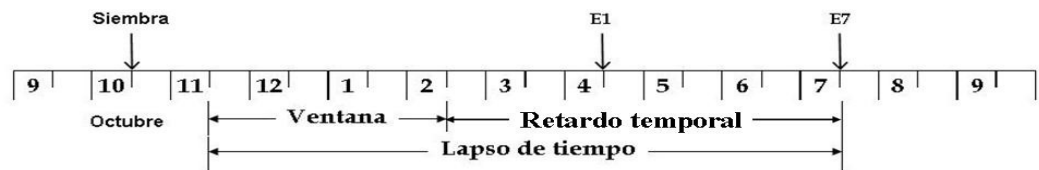


Figura N° 3. Ventana con retardo para un lapso de tiempo de exploración dado.

#### **1.4 OBJETIVOS GENERAL**

Desarrollar un modelo de pronóstico de alertas tempranas del contagio de la Sigatoka Negra en plantaciones de plátano Hartón de la zona Sur del Lago de Maracaibo, a partir de información meteorológica y biológica recopilada en la Estación Local Chama - INIA.

#### **1.5 OBJETIVOS ESPECÍFICOS**

- Determinar las variables de relevancia para el modelo, mediante el análisis y la depuración de los datos biológicos y meteorológicos obtenidos de la Estación Local Chama - INIA.
- Evaluar y seleccionar las herramientas o técnicas a implementar en la generación del modelo, mediante el estudio de las alternativas existentes en el campo del modelado.
- Elaborar el modelo de pronóstico de predicción de alertas tempranas del contagio de Sigatoka Negra.
- Validar el modelo de pronóstico en una unidad de producción experimental ubicada en las adyacencias de la Estación Local Chama - INIA.

## **CAPITULO II - MARCO TEÓRICO**

### **2 INTRODUCCIÓN**

En este capítulo se explican algunos de los conceptos necesarios para comprender las variables y los datos utilizados en el modelo de pronóstico de la Sigatoka Negra. Se describirán tanto las variables y datos biológicos como las climatológicas.

#### **2.1 DESCRIPCIÓN DE LA ZONA DE ESTUDIO**

En Venezuela, el cultivo del plátano (*Musa* AAB subgrupo plátano *cv Hartón*) tiene una significativa importancia socioeconómica, por ser uno de los productos fundamentales de la canasta familiar y fuente generadora de trabajo. En el año 1994 se reportó que este rubro generó ingresos brutos superiores a los 23.000 millones de bolívares, por concepto de exportaciones 4.3.2.3. El consumo per cápita de la población venezolana fue de 19,4 kg para el año 1993 con una producción nacional de 535.359 ton. y en exportaciones 120.000 ton. 4.3.2.3. En resumen el cultivo aporta el 50% del valor de la producción frutícola nacional y su exportación provee divisas para el país y contribuye en unos 15.000 empleos directos 4.3.2.3.

El cultivo del plátano está determinado por distintas condiciones climáticas, entre ellas, precipitaciones de 700 hasta 4000 mm/año, con ocurrencia generalmente de dos meses secos separados, y temperaturas medias anuales superiores a 21 °C, con variaciones estacionales de entre 4 y 13 °C. El volumen de precipitaciones mensuales que exige el cultivo debe estar por encima de los 50 mm, pues valores inferiores afectan las condiciones hídricas de la planta. La zona Sur de Lago de Maracaibo es ideal, al presentar cada una de ellas. Se registra una temperatura media que oscila

entre 24 y 28,8 °C; la precipitación varía entre 1300 y 2600 mm anuales, presentándose un período húmedo entre 4 y 12 meses; la evaporación media anual del área fluctúa entre 1100 y 2000 mm 4.3.2.3.

Los cultivos de plátano en la zona se ubican principalmente en la planicie y, secundariamente en el piedemonte andino. En su totalidad están por debajo de los 100 msnm, en tierras de origen sedimentario relacionadas con el levantamiento de la cordillera andina. Se localizan en varias subregiones naturales con zonas de vida desde el bosque seco tropical hasta el húmedo tropical 4.3.2.34.3.2.3.

Para el Sur del Lago de Maracaibo el cultivo del plátano, representa una actividad de importancia estratégica para el desarrollo de la región. De ella dependen directamente alrededor de 60.000 familias e indirectamente a más de 180.000 4.3.2.3.

La Sigatoka Negra causada por el hongo *Mycosphaerella fijiensis* Morelet en la última década ha causado preocupación a los productores de plátano por su impacto directo en la producción del cultivo y por ende en los costos necesarios para su control 4.3.2.3. En la actualidad los mismos, representan el 48,56 % 4.3.2.3 del costo de la producción de las plantaciones de plátanos (Figura N° 4). El costo promedio del programa de aplicación para controlarla, van desde 8.105,27 Bs/ha hasta 17.757,80 Bs/ha. Los costos por control de la enfermedad estuvieron en el orden de 97.263,24 Bs/ha/año hasta 213.093,60 Bs/ha/año. Siendo la aplicación aérea la más utilizada con frecuencia de 30 días. Las

unidades de producción con mayor superficie sembrada de plátano, tienen mayores costos, más problemas en el manejo de los recursos disponibles y requieren más insumos para el control de la enfermedad que aquellas unidades de producción con superficies menores donde su control es más racional reduciendo de esta forma sus costos y obteniendo, por lo tanto, una mayor efectividad en el uso de los productos para controlar la Sigatoka Negra 4.3.2.3, 4.3.2.3, 4.3.2.3.

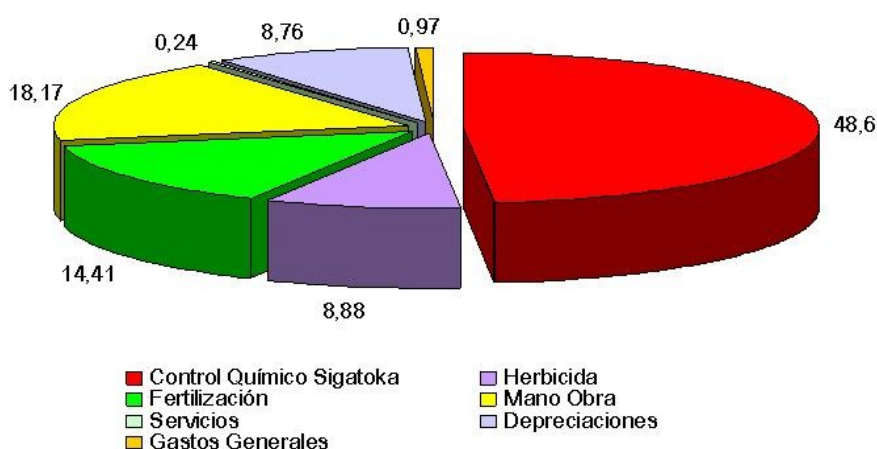


Figura N° 4. Costos de Producción. Composición Porcentual  
4.3.2.3.

## 2.2 EL PLÁTANO

El nombre de plátano, banano, cambur o guineo agrupa a un gran número de plantas herbáceas del género *Musa*, tanto híbridos obtenidos horticulturalmente a partir de las especies silvestres del género *Musa acuminata* y *Musa balbisiana* como cultivares genéticamente puros de estas especies. Clasificado originalmente por Linneo como *Musa paradisiaca* en 1753, la especie tipo del género *Musa*, estudios posteriores han llevado a la conclusión de que la compleja taxonomía del género incluye numerosos híbridos, de variada composición genética, y se ha desarrollado un sistema

estrictamente *sui generis* de clasificación para dar cuenta de esta variación. Sin embargo, de acuerdo a las reglas del Código Internacional de Nomenclatura Botánica, el nombre *linneano* cuenta con prioridad, y sigue siendo usado para designar genéricamente a estas variedades 4.3.2.3.

En la nomenclatura vernácula a veces se traza una diferencia entre las bananas, consumidas crudas como fruta de postre, y los plátanos, que por su superior contenido en fécula deben asarse o freírse antes de su ingesta. La diferencia no se corresponde exactamente con ningún criterio genético; aunque las variedades con mayor presencia genética de *M. balbisiana* suelen estar comprendidas en este segundo grupo, no se puede determinar si una planta producirá plátanos o bananas simplemente por su constitución genética. La confusión aumenta por el hecho de que en otras regiones los términos se consideran perfectamente sinónimos 4.3.2.3, 4.3.2.3, 4.3.2.3.

En todo caso, este grupo de vegetales conforman la fruta tropical más consumida del mundo. Se trata de una falsa baya, de forma falcada o alongada, que crece en racimos de hasta cien unidades y 50 kg de peso; de color amarillo cuando está maduro, es dulce y carnoso, rico en carbohidratos, potasio, vitamina A y vitamina C. Es mucho más rico en calorías que la mayoría de las frutas por su gran contenido en fécula; de los 125 g que pesa en promedio, el 25% es materia seca, que aporta unas 120 calorías 4.3.2.3.

Se cultivan en más de 130 países, desde el sudeste asiático de donde son nativas, hasta Oceanía y Sudamérica; el principal

productor mundial es la India, de donde proceden casi un cuarto de los frutos comercializados en el mundo, aunque buena parte de los mismos son para consumo doméstico. El principal exportador es Ecuador, que genera casi un tercio de las exportaciones globales. El volumen de producción de bananas y plátanos sólo es superado por el trigo (*Triticum spp.*), el arroz (*Oryza sativa*) y el maíz (*Zea mays*). La producción continua de fruto a lo largo de todo el año los hace especialmente valiosos como alimento en la época entre cosechas en los países tropicales; son los plátanos de freír los que cumplen este papel principalmente 4.3.2.3.

#### **2.2.1 ETAPAS DE DESARROLLO DE UNA HOJA DEL PLATANO**

Las hojas se cuentan entre las más grandes del reino vegetal; son de color verde o amarillo verdoso claro, con los márgenes lisos y las nervaduras pinnadas. Las hojas tienden a romperse espontáneamente a lo largo de las nervaduras, dándoles un aspecto desaliñado. El lapso que toma la hoja para abrirse es variable. Bajo condiciones climáticas favorables, toma aproximadamente siete días, pero puede extenderse hasta 15 o 20 días bajo condiciones adversas (sequía, mala nutrición, etc.) 4.3.2.3 ; las hojas viven no más de dos meses, y en los trópicos se renuevan a razón de una por semana en la temporada de crecimiento.

Para comprender mejor este proceso, es importante recordar que la formación de la hoja ocurre dentro del pseudotallo antes de la parición. La hoja nueva se encuentra muy enroscada, es de color blancuzco y muy frágil 4.3.2.3.



La salida de la hoja da como resultado un crecimiento extraordinariamente rápido de la vaina de la hoja (4 m en 10 días para *Gros Michel*<sup>4</sup>). La hoja joven se desliza por el canal peciolar de la hoja anterior y es así como el desarrollo de una hoja corresponde a dos fenómenos sucesivos: el de crecimiento y el de apertura.

Para facilitar la descripción del proceso de apertura, éste se ha dividido en cinco estadios sucesivos, los cuales se definen arbitrariamente puesto que en realidad es un proceso continuo (Figura N° 5). Se puede decir que los primeros dos estadios corresponden a la etapa de crecimiento, el tercero representa el final del crecimiento y el inicio del proceso de apertura y los estadios cuarto y quinto son en sí la fase misma de la apertura. Los estadios mencionados anteriormente se definen a continuación 4.3.2.3:



Figura N° 5. Etapas de desarrollo de una hoja de banano 4.3.2.3.

- **Estadio A:** La hoja 'candela', de aproximadamente 10 cm de longitud, todavía se encuentra unida a la hoja anterior.
- **Estadio B:** La hoja 'candela' es más grande, pero aún no ha alcanzado su tamaño completo.

---

<sup>4</sup> *Gros Michel* es una variedad del plátano comercial

- **Estadio C:** La hoja ‘candela’ está completamente libre. Alcanza su tamaño total y el diámetro de su ápice ha aumentado considerablemente después de soltarse del espiral.
- **Estadio D:** El lado izquierdo ya está abierto y su apertura ocurre en el extremo del ápice.
- **Estadio E:** La parte de arriba de la hoja se abre y la base tiene la forma de una corneta abierta.

### 2.3 LA SIGATOKA NEGRA

La Sigatoka Negra, *Mycosphaerella fijiensis* M, es una enfermedad que ataca inicialmente en el envés de las hojas formando lesiones necróticas o manchas con halos amarillos y centro gris claro, en grandes áreas del tejido foliar (Figura N° 6). Afecta a la planta en las hojas más jóvenes y por lo tanto, ocasiona daños más importantes en los tejidos fotosintéticos 4.3.2.3. La mancha sigue avanzando en su desarrollo y evolución; se hace más grande y ancha de forma elíptica y se rodea de un borde café oscuro visible cuando la hoja está mojada. Luego de este estado la mancha se seca en el centro, se torna gris y se deprime; la lesión se rodea de un borde angosto negro bien definido. Al unirse todas las lesiones la hoja se torna negra y muere en 3 ó 4 semanas después de asomar los primeros síntomas 4.3.2.3.

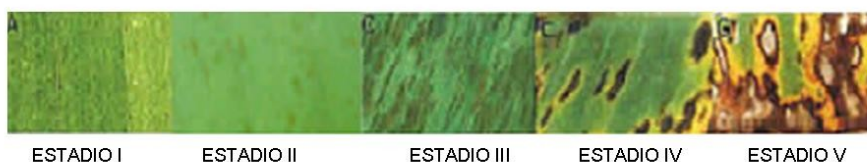


Figura N° 6. Estadio de la Sigatoka Negra.

Tabla N° 1. Síntomas de las manchas foliares de Sigatoka Negra asociados a los diferentes estadios de la enfermedad

ESTADIO	SÍNTOMAS
I	Puntos pequeños, angulares, de color marrón rojizo de 0.5 mm. de diámetro, que aparecen solamente en la cara inferior de la hoja y son visibles mediante luz transmitida.
II	Rayas de 1-2 mm de ancho por 10 mm de longitud de color marrón rojizo y tienen sus bordes muy bien definidos. Son más visibles en la cara inferior de la hoja que en la superior.
III	La raya se alarga un poco más y cambia a un color marrón muy oscuro o casi negro y ya son claramente visibles en la cara superior de la hoja
IV	La raya se alarga aún más, pero comienza a tomar una forma elíptica y está rodeada de una zona de color marrón claro que se observa húmeda, como impregnada de agua, especialmente temprano en la mañana.
V	El centro de la mancha se ha hundido, la zona húmeda es ahora más nítida y tiene un halo amarillento.
VI	El centro de la mancha se seca, toma un color gris claro rodeado de un anillo negro bien definido rodeado de un halo amarillo. Este último estado es muy similar al equivalente de la Sigatoka Amarilla, pero las manchas son aisladas.

La Sigatoka Negra es una típica enfermedad policíclica, en la que tanto conidios como ascosporas juegan su papel en la dispersión de la enfermedad. Los conidios aparecen en conidióforos sencillos y se forman en lesiones jóvenes como estrías (estadios 2 y 3) y en el primer estadio de mancha (estadio 4, según la escala de Fouré 4.3.2.3). Se producen más abundante durante períodos de alta humedad, sobre todo si una película de agua libre está presente en la hoja. Los conidios no son desprendidos por el viento, se dispersan por medio del salpique de la lluvia y el escurrimiento del agua por la superficie de las hojas. Por lo anterior, los conidios son

asociados principalmente con infecciones a corta distancia, entre hojas de una misma planta, de la planta madre a los hijos y entre plantas cercanas. Debido a que comparativamente con *M. musicola*, *M. fijiensis* producen relativamente pocos conidios, las ascosporas son consideradas más importantes en la dispersión de la enfermedad. No obstante, los conidios pueden jugar un rol muy importante en la prevalencia de la enfermedad durante periodos de baja precipitación 4.3.2.3, en los que se forma abundante rocío en horas de la madrugada y las primeras horas de la mañana. Hasta ahora no ha sido posible establecer diferencias entre los síntomas de infecciones causadas por conidios u ascosporas. Definitivamente las ascosporas son la principal forma de dispersión a distancias mayores, por efecto del viento y son las responsables de la introducción paulatina de la enfermedad en nuevas áreas, sobre todo, durante períodos de alta precipitación. Las ascosporas se producen en cuerpos fructíferos denominados pseudotecios, en lesiones maduras, generalmente en las hojas más viejas. Los pseudotecios se forman en ambas superficies de la hoja, la información disponible sobre la abundancia relativa en cada superficie es controversial. Para la liberación de las ascosporas se requiere de suficiente agua de lluvia o rocío como para humedecer bien el pseudotecio. En estudios desarrollados en Hawaii y Costa Rica se encontró que la liberación de ascosporas es más abundante durante las horas de la noche, decreciendo significativamente durante el día. Sin embargo, durante días lluviosos se pueden producir picos cortos de liberación de ascosporas poco después de que ha iniciado la lluvia. No obstante, la conocida capacidad de dispersión de las ascosporas por el viento, se estima que la dispersión a largas distancias es limitada a unos pocos cientos de

kilómetros, debido a su susceptibilidad a la radiación ultravioleta 4.3.2.3.

Las ascosporas se depositan eficientemente en la superficie abaxial de la hoja en desarrollo (hoja candela o cigarro), evidencia de que una superficie cilíndrica es muy apropiada para la deposición de este tipo de propágulos que son diseminados por el viento. Tanto conidios como ascosporas deben infectar la hoja vía estomas y estos son más abundantes en la superficie abaxial de las hojas. Se puede asumir, en forma general, que por cada estoma presente en haz de la hoja existen tres en el envés. Tanto conidios como ascosporas requieren de una alta humedad relativa para germinar, aunque el rango óptimo para los conidios es más amplio (92-100%) que para las ascosporas (98-100%) 4.3.2.3.

La temperatura tiene un efecto cuadrático sobre la germinación de los conidios y las ascosporas, diversos estudios han determinado que el rango óptimo oscila entre 26,5 y 28 °C. Existe poca información acerca de la viabilidad de los conidios y ascosporas 4.3.2.3.

Estudios recientes desarrollados en Brasil demostraron que los conidios pueden permanecer viables hasta más de 60 días sobre la superficie de la hoja y hasta 18 días en la epidermis de los frutos en condiciones de sombra 4.3.2.3.

#### **El sistema de preaviso biológico.**

El carácter permanente de las plantaciones de banano y plátano, el continuo crecimiento de las plantas o sus retoños y las condiciones

climáticas cambiantes, características de las zonas de producción de bananos y plátanos, permiten el establecimiento de sistemas de preaviso. Estos sistemas se basan en el análisis de descriptores biológicos y climáticos para la aplicación oportuna de los fungicidas, en períodos en los cuales la severidad de la enfermedad comienza a incrementar y las condiciones climáticas conducen a un favorable desarrollo del patógeno. El sistema de preaviso biológico para Sigatoka negra, es el producto de adaptaciones hechas de los sistemas desarrollados por los franceses en las Antillas para la Sigatoka amarilla. El preaviso es basado en observaciones semanales de los síntomas en las hojas jóvenes de plantas en crecimiento activo. Coeficientes arbitrarios se asignan a las tres hojas más jóvenes, según la incidencia y severidad de la enfermedad y con ellos se calculan dos variables: la Suma Bruta (SB) y el Estado de Evolución (EE). La SB se refiere al estado presente de la infección y es un valor arbitrario que aumenta con el avance de los síntomas y la juventud de las hojas. El EE es calculado usando la SB y el ritmo de emisión foliar de las plantas. El sistema de preaviso se implementó y adaptó a finales de los 80's e inicios de los 90's en diferentes países de América Latina y con él se logró reducir significativamente el número de aplicaciones de fungicidas. Posteriormente, por varias razones, incluyendo la resistencia a fungicidas, el sistema perdió vigencia 4.3.2.3.

### **2.3.1 EVALUACIÓN DEL ESTADO DE LA INFECCIÓN**

La evaluación continua de la evolución de la enfermedad es necesaria para lograr mantener una idea clara y precisa del estado sanitario de las plantaciones de musáceas y poder prevenir daños severos al cultivo y su producción, por lo cual deben hacerse

evaluaciones periódicas, semanales o quincenales, sobre la severidad o incidencia de la Sigatoka Negra en cada finca 4.3.2.3. Por otro lado, para la obtención de la información biológica de la manchas foliares causadas por *Mycosphaerella fijiensis* M. es necesario conocer las etapas tanto del proceso de apertura de la hoja (Figura Nº 5), como también del desarrollo de los síntomas de la enfermedad. Para esto último, se usa la metodología de preaviso biológico, la cual se basa en la observación del desarrollo de la enfermedad, medida como: el ritmo de emisión foliar, la suma bruta y el estado de evolución., Esta información es obtenida para cada planta de un total de 10 ó 15 plantas seleccionadas, y a cada una se le hace seguimiento detallado cada semana. Las investigaciones llevadas a cabo en la Estación Local Chama - INIA usan las metodología de preaviso biológico y Stover 4.3.2.3 modificado por Gaulh 4.3.2.3 para evaluar la incidencia y severidad de Sigatoka Negra en la zona Sur del Lago de Maracaibo. La incidencia es el número de hojas afectadas y la severidad es el área foliar manchada por el ataque de Sigatoka Negra en el plátano. Para llevar a cabo tal evaluación en necesario registrar los parámetros o variables que permiten dar una perspectiva clara de la enfermedad, para lo cual normalmente son registradas las siguientes variables 4.3.2.3:

- Fecha de la observación o registro de la evaluación de la infección.
- Total de hojas de las plantas estudiadas (TH).

## **VARIABLES CLIMATOLÓGICAS**

- **SEVERIDAD DE LA ENFERMEDAD (SE)**

La severidad de la enfermedad es la extensión del área de la hoja infectada por el patógeno.

Este se puede expresar en porcentajes o en grados de la enfermedad. Los grados de la enfermedad en la hoja se deben calificar utilizando el sistema de Stover 4.3.2.3 modificado por Gauhl 4.3.2.3 (Figura N° 2). El porcentaje del área de la hoja afectado por el patógeno se expresa en grados de infección, esta evaluación es para cada hoja en cada planta 4.3.2.3.

- **ESTADO EVALUACIÓN DE LA ENFERMEDAD (EE).**

Es la expresión del desarrollo de la enfermedad. Se obtiene a partir de la evaluación semanal del crecimiento vegetativo de la planta (emisión foliar) y la presencia de los síntomas de la enfermedad 4.3.2.3. Se calcula de la siguiente manera

$$EE = SEV \times REF_x \quad \text{Ec. 1}$$

Donde:

SEV: Suma de evolución

REF<sub>x</sub>: Ritmo de emisión foliar ponderado

**Suma de evolución para toda la planta (SEV)**

$$SEV = SB - CE \quad \text{Ec. 2}$$

Donde:

SB: Suma Bruta

CE: Coeficiente de evolución

Coeficiente de evolución (CE)

$$CE = 2 \times \sum_{i=1}^n (EC_i \times HE_i) \quad \text{Ec. 3}$$

Donde:



EC: Estadio de la hoja candela por planta

HE: Número de hojas enfermas por planta.

Siendo la sumatoria para  $n$  número de plantas

### Suma bruta de daño en las hojas (SB)

$$SB = \sum_{i=1}^6 (S_i \times Ca) \quad \text{Ec. 4}$$

Donde:

S: Sumatoria del número de síntomas de cada grado por hoja.

Ca: coeficiente arbitrario de severidad (ver Tabla N° 2).

En la Tabla N° 2, se presentan los coeficiente arbitrarios de severidad sugeridos por Bureau et al, 1992 4.3.2.3, en la cual aparece la cuantificación del estadio más avanzado de la enfermedad sobre cada hoja observada. Las hojas observadas son las hojas 2, 3, y 4. La pareja de números de los estados de las hojas que contienen cercosporas se determina por un coeficiente base que caracteriza la velocidad de evolución de la enfermedad en función del tiempo 4.3.2.3. Los parámetros para determinar la evolución de la enfermedad son los coeficientes arbitrarios sugeridos en la Tabla N° 2, el número total de la emisión foliar y el estado de la hoja candela en cada plantación.

Tabla N° 2. Coeficientes arbitrarios de severidad de la Sigatoka Negra

Calificación del ataque	Hojas		
	2	3	4
-1	60	40	20
+1	80	60	40
-2	10	80	60

+2	12 0	10 0	80
-3		12 ^	10 12
+3			0

Fuente: Bureau et al (1992).

La observación de los diferentes estados de la enfermedad comienza por el envés del semilimbo izquierdo del ápice foliar y a partir de esta zona se observa el resto de la hoja. Se anota únicamente el grado del síntoma más desarrollado y se coloca el signo (+) en caso de existir una densidad de síntomas superior a 50 y si el número de síntomas es inferior a 50 se coloca el signo (-) 4.3.2.3.

- **HOJA MÁS JOVEN MANCHADA (HMJM)**

Esta corresponde a la primera hoja totalmente abierta que presenta 10 ó más lesiones discretas necrosadas y maduras o un área grande necrosada con 10 centros secos de color claro, contando las hojas de arriba hacia abajo (Figura N° 1).

- **RITMO O ÍNDICE DE EMISIÓN FOLIAR (REF)**

El REF es un índice que corresponde al número de hojas producido por semana, el cual deberá ser menor que uno. Este cálculo deberá efectuarse regularmente (por lo menos una vez al mes) para cada planta de prueba y de referencia a partir de los tres meses después de la siembra y hasta la emergencia del racimo (parición). El ritmo de emisión foliar (REF) se determina mediante la identificación de cada hoja colocando en la base de la vena central un número progresivo comenzando con el N° 1 (Figura N° 7), que corresponde a la emisión foliar en ese

momento. También se identifica el estado de desarrollo de la hoja candela (bandera), según la descripción de Brun, citado por Marín y Romero 4.3.2.3, Figura N° 5, el cual se anota como la fracción decimal del total de hojas emitidas, por ejemplo: 12,4.

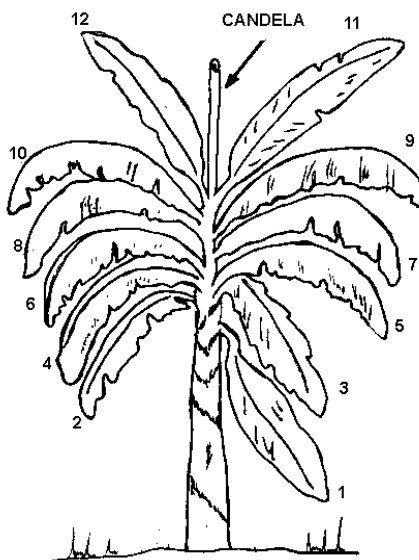


Figura N° 7. Conteo de hojas para la obtención de la emisión foliar.

Fuente: Gómez, C. et al. 2001

### **Estimación del Ritmo de Emisión foliar ponderado actual ( $REF_x$ )**

$$REF_x = REF_A - REF_P \quad \text{Ec. 5}$$

Donde:

$REF_A$ : Ritmo de emisión foliar actual

$REF_P$ : Ritmo de emisión foliar ponderado pasado

### **Ritmo de emisión foliar actual ( $REF_A$ )**

$$REF_A = \frac{REF}{N} \quad \text{Ec. 6}$$

Donde:

$REF$ : Ritmo de emisión foliar

$N$ : Número de días entre evaluaciones

### **Ritmo de emisión foliar (REF)**

$$REF = EF_A - EF_P \quad \text{Ec. 7.}$$

Donde:

$EF_A$ : Emisión foliar actual

$EF_P$ : Emisión foliar pasada

### **2.3.2 FACTORES QUE INFLUYEN EN EL DESARROLLO DE LA ENFERMEDAD.**

La epidemiología de la Sigatoka Negra depende de factores bióticos y abióticos. Entre los factores abióticos son importantes los patrones de temperatura y humedad (básicamente el número de horas que la superficie de la hoja permanezca humedecida) y la disponibilidad de evolución de la enfermedad 4.3.2.3. Aunado a estos, numerosos investigadores se han dado a la tarea determinar la influencia de algunas variables climatológicas en la germinación y propagación de la Sigatoka Negra, entre estos se puede mencionar a:

- Fouré 4.3.2.3 planteó la existencia de una estrecha relación entre algunos factores climáticos como la humedad relativa, temperatura, precipitación y el patógeno, los cuales condicionan la incidencia y severidad de la enfermedad.
- Mourichon y Zapater 4.3.2.3, plantearon que la enfermedad presenta una dinámica estacional determinada por las variaciones de temperatura y precipitación a lo largo del año. La estructura reproductiva se desarrolla mediante inoculación cruzada que se facilita cuando hay agua libre sobre las hojas.
- Gauhl y colegas 4.3.2.3, expresó que la liberación de ascosporas ante la presencia de lluvias es alta, atribuido a la existencia de

una capa de agua en la superficie de la hoja donde existe una mayor cantidad de manchas en el envés. Las hojas secas adheridas a las plantas representan una excelente fuente del inoculo.

- Jácome y Schuh 4.3.2.3, determinaron que, la temperatura y la humedad relativa favorecen el desarrollo de la epidemia, ya que las temperaturas entre 20-35 °C contribuyen a la germinación de conidios y ascosporas de los hongos, ocurriendo máxima germinación si existe un rango de temperatura entre 25-28 °C y una alta humedad relativa, especialmente cuando hay presencia de la película húmeda sobre la hoja, como se expresó anteriormente.
- Pérez 4.3.2.3, estimó que las ascósporas de *Mycosphaerella fijiensis* germinan en un rango amplio entre 10-38 °C, considerándose óptimo 27 °C, observándose que la velocidad relativa del crecimiento de los tubos germinativos de esta se deprimen fuertemente a temperaturas menores de 20 °C.

Con respecto al viento, se ha observado que la concentración de las conidiósporas en las plantaciones es alta en las capas inferiores del aire, en comparación con el follaje, mientras que las ascosporas en el aire es la misma en ambas alturas, lo cual indica su importancia en el ciclo de la enfermedad 4.3.2.3.

Las esporas de la Sigatoka Negra son dispersadas por el viento y depositadas en las hojas más jóvenes de la planta 4.3.2.3. Las esporas depositadas germinan, si las condiciones de humedad favorecen su desarrollo, emitiendo un tubo germinativo que penetra por las estomas de las hojas, para luego ramificarse y

colonizar varias células vecinas, produciendo el síntoma característico de pizca y posteriormente la mancha necrosis.

La lluvia posee un papel muy importante en la liberación del inóculo. La precipitación provee condiciones de humedad que favorecen el desarrollo de las infecciones, permitiendo establecer una época relativa de baja y otra de alta incidencia 4.3.2.3.

La humedad relativa es importante en proveer las condiciones hídricas de las esporas y el desarrollo de las infecciones, y el viento es el factor que permite la dispersión de las esporas del patógeno, una vez que estas han sido liberadas 4.3.2.3.

### **2.3.3 VARIABLE CLIMATOLÓGICAS.**

Las variables registradas por el sistema automatizado de la Estación Local Chama y transmitidas al Centro de Cálculo de la Universidad de los Andes (CECALCULA), donde son colocados a disposición del público en servidor web [http://www.cecalc.ula.ve/redbc/colecciones/coleccion\\_estacion\\_chama/datos\\_estacion\\_chama.html](http://www.cecalc.ula.ve/redbc/colecciones/coleccion_estacion_chama/datos_estacion_chama.html), algunas de las variables a utilizar en el desarrollo del modelo son las siguientes 4.3.2.3:

- Fecha. Está expresada en DD/MM/YY que indica el día, mes año.
- Hora. Se expresa de la siguiente manera: HH/MIN m que indica la hora, minutos y si ésta corresponde a la mañana (m) o a la tarde (p)
- Temperatura del aire (°C): es la temperatura que se puede medir con un termómetro o sensores y que se toma del medio ambiente.

- Temperatura máxima del aire (Tmax) (°C): Es la temperatura más alta registrada durante el intervalo de tiempo.
  - Temperatura mínima del aire (Tmin) (°C). Es la temperatura más baja registrada durante el intervalo de tiempo.
  - Temperatura promedio (Tp) (°C). Es la temperatura promediada entre la máxima y la mínima temperatura registrada durante el intervalo de tiempo.
- Evapotranspiración (mm)(ET): Es una medida de la cantidad de vapor de agua retornado por el aire en un área dada. Combina la cantidad de vapor de agua retornado a través de la evaporación (desde la superficie de la vegetación y estomas de las hojas) con la cantidad de vapor de agua retornado a través de la transpiración para arribar a un total para el área. Esta variable no es medida sino calculada. Los cálculos de la ET actual son realizados por el GroweatherLink ET/Datos Logger (se calcula entre una hora y otra). Para calcular y desplegar la ET la estación utiliza el GroweatherLink Software. La Groweather usa la temperatura del aire, humedad relativa, funcionamiento del viento y radiación solar para estimar la ET, basado en un modelo de Penman-Monteith 4.3.2.3.

$$ET_0 = \frac{1}{\lambda} \cdot \left( \frac{\Delta(R_n - G) + \rho_a \cdot C_p \cdot \left( \frac{e_s - e_a}{r_a} \right)}{\Delta + \gamma \cdot \left( 1 + \frac{r_s}{r_a} \right)} \right) \quad \text{Ec. 8}$$

- Radiación solar (R.S): Se conoce por radiación solar al conjunto de radiaciones electromagnéticas emitidas por el Sol que alcanzan la superficie de la Tierra. La unidad práctica que describe la radiación solar que llega a la Tierra es la irradiancia, o unidad de potencia por metro cuadrado (W/m²) 4.3.2.3..

- Energía solar incidente (E.S): Es la cantidad de energía solar incidente recibida durante un periodo de tiempo. La unidad de medida es el Langley (Ly) el cual equivale a 41,84 kilojoules por metro cuadrado, equivale también a 11.622 Watt-hora por metro cuadrado. La estación meteorológica mide la energía recibida en la banda espectral entre 400 y 1100 nm 4.3.2.3..
- Presión barométrica (P): Es la presión o el peso que ejerce la atmósfera en un punto determinado, el instrumento de medición es el barómetro. La medición puede expresarse en varias unidades de medidas: hectopascales, milibares, pulgadas o milímetros de mercurio (Hg). También se conoce como presión atmosférica 4.3.2.3..
- Velocidad del viento (Vv): El viento es el movimiento del aire. Los vientos globales se generan como consecuencia del desplazamiento del aire desde zonas de alta presión a zonas de baja presión, determinando los vientos dominantes de un área o región. Aún así hay que tener en cuenta numerosos factores locales que influyen o determinan los caracteres de intensidad y periodicidad de los movimientos del aire. La medición de la velocidad y dirección del viento se efectúa con instrumentos registradores llamados anemómetros, que disponen de dos sensores, uno para medir la velocidad y otro para medir la dirección del viento. Las mediciones se registra en anemógrafos. Se mide en Km/s 4.3.2.3.
- Dirección del viento: La dirección es designada por el rumbo del cual viene el viento y no por aquel al cual se dirige. Así, un viento noreste es un viento que viene del noreste 4.3.2.3.
- Enfriamiento del viento: Es la temperatura que nota un cuerpo humano (o animal) debido a la combinación de la temperatura y



la velocidad del viento. Excepto a temperaturas altas, el viento sirve para aumentar la sensación de frío, ya que el viento favorece la evaporación de la piel y para ello se necesita calor que roba al cuerpo. Se mide en °C 4.3.2.3.

- Precipitación (P.A)) (mm). Precipitación equivale al espesor de la lámina de agua que se formaría, a causa de la precipitación sobre una superficie plana e impermeable 4.3.2.3.
- Humedad relativa (Hr) (%): Es la humedad que contiene una masa de aire, en relación con la máxima humedad absoluta que podría admitir, sin producirse condensación, conservando las mismas condiciones de temperatura y presión atmosférica. La humedad relativa se puede medir mediante un instrumento denominado higrómetro o un psicrómetro. Un psicrómetro está formado por un termómetro de bulbo húmedo y un termómetro seco. Se mide la temperatura húmeda y puede obtenerse la humedad relativa mediante el ábaco higrométrico, comparándola con la temperatura real del termómetro seco 4.3.2.3.
- Punto de rocío (Pr) (°C). Es la temperatura a la que empieza a condensar el vapor de agua contenido en el aire, produciendo rocío, neblina o, en caso de que la temperatura sea lo suficientemente baja, escarcha. Para el cálculo se puede utilizar esta fórmula 4.3.2.3:

$$Pr = \left( \sqrt[8]{\frac{H}{100}} \right) (0,9 \cdot T + 11,2) + (0,9 \cdot T - 11,2) \quad \text{Ec. 9}$$

Pr = Punto de rocío.

T = Temperatura en grados Celsius

H = Humedad relativa.

- Índice del calor. Es un índice que combina temperatura del aire y humedad relativa para determinar una temperatura evidente. También conocido como temperatura aparente. El índice de calor es una medida de cuánto calor se siente cuando la humedad relativa se combina con los efectos de la temperatura del aire. Se mide cuando la temperatura del aire está por encima de los 20°C 4.3.2.3.

## **2.4 GENERACIÓN DEL MODELO PREDICTIVO**

La búsqueda de un modelo predictivo para el brote de la enfermedad se fundamenta en la extracción de patrones de comportamiento de la misma respecto a las variables climatológicas; por consiguiente, se implementa el uso de herramientas estadísticas y matemáticas que permitan obtener dichos patrones. Una de las más utilizadas en la actualidad para la búsqueda de patrones en bases de datos es la minería de datos. El término, aunque relativamente moderno, integra numerosas técnicas de análisis de datos y extracción de información antes utilizadas. A pesar de estar fundamentado sobre varias disciplinas tiene su distinción en su orientación al fin, aun más que al medio. Por ello se facilita su enriquecimiento de todas estas. El fin al que se orienta es la capacidad de extracción de patrones, descubrimiento de tendencias, predicción de comportamientos y en general al aprovechamiento de la información computarizada con la que cuentan instituciones y organizaciones, y que generalmente se encuentra en grandes volúmenes y cantidades 4.3.2.3.

La minería de datos se encuentra enmarcada dentro del proceso de descubrimiento o extracción de conocimiento en bases de datos

(*Knowledge Discovery in Databases*, KDD), entendiendo por tal a la extracción no trivial de información potencialmente útil, válida, novedosa y comprensible a partir de un gran volumen de datos 4.3.2.3. Las investigaciones en estos temas incluyen análisis estadístico de datos, técnicas de representación del conocimiento, razonamiento basado en casos, razonamiento aproximado, adquisición de conocimiento, redes neuronales y visualización de datos. Tareas comunes en KDD son la inducción de reglas, los problemas de clasificación y clustering, el reconocimiento de patrones, el modelado predictivo, la detección de dependencias, entre otros.

Los datos recogen un conjunto de hechos (una base de datos) y los patrones son expresiones que describen un subconjunto de los datos (un modelo aplicable a ese subconjunto). KDD involucra un proceso (Figura N° 8) iterativo e interactivo de búsqueda de modelos, patrones o parámetros. Los patrones descubiertos han de ser válidos, novedosos para el sistema (para el usuario siempre que sea posible) y potencialmente útiles 4.3.2.3.

Se han de definir medidas cuantitativas para los patrones obtenidos (precisión, utilidad, beneficio obtenido, etc.). Se debe establecer alguna medida de interés que considere la validez, utilidad y simplicidad de los patrones obtenidos mediante alguna de las técnicas de minería de datos. El objetivo final de todo esto es incorporar el conocimiento obtenido en algún sistema real, tomar decisiones a partir de los resultados alcanzados o, simplemente, registrar la información conseguida y suministrársela a quien esté interesado 4.3.2.3.

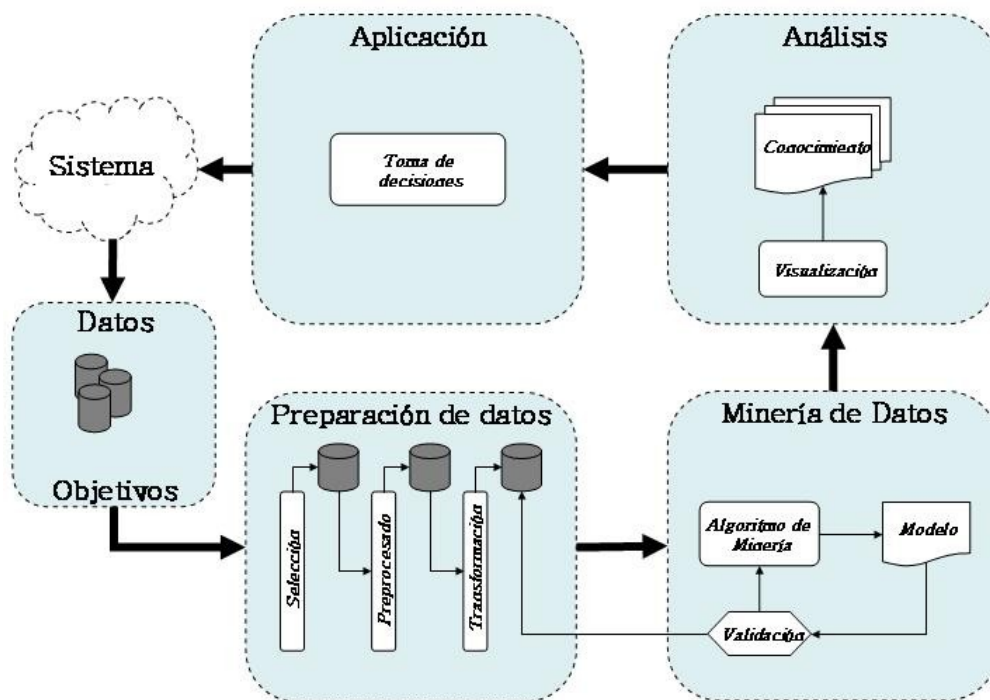


Figura N° 8. Proceso de extracción de conocimiento en bases de datos.

Los algoritmos de minería de datos (una etapa dentro del proceso completo de KDD) intentan obtener patrones o modelos a partir de los datos recopilados. Decidir si los modelos obtenidos son útiles o no suele requerir una valoración subjetiva por parte del usuario. Los algoritmos de minería de datos suelen tener tres componentes:

- El MODELO, que contiene parámetros que han de fijarse a partir de los datos de entrada.
- El CRITERIO DE PREFERENCIA, que sirve para comparar modelos alternativos
- El ALGORITMO DE BÚSQUEDA, como cualquier otro programa de Inteligencia Artificial (IA).

El criterio de preferencia suele ser algún tipo de heurística y los algoritmos de búsqueda empleados suelen ser los mismos que en otros programas de IA. Las principales diferencias entre los algoritmos de minería de datos se hallan en el modelo de representación escogido y la función del mismo (el objetivo perseguido). Por ejemplo, un modelo de clasificación basado en árboles de decisión suele utilizar un algoritmo *greedy* (una búsqueda sin vuelta atrás) y una heurística que favorezca la construcción de árboles de decisión con pocos nodos 4.3.2.3.

Las herramientas de minería de datos empleados en el proceso de KDD se pueden clasificar en dos grandes grupos: técnicas de verificación (en las cuales el sistema se limita a comprobar hipótesis suministradas por el usuario) y métodos de descubrimiento (en los que se han de encontrar patrones potencialmente interesantes de forma automática, incluyendo en este grupo todas las técnicas de predicción). El resultado obtenido con la aplicación de algoritmos de minería de datos (pertenecientes al segundo grupo, el de técnicas de descubrimiento) puede ser de carácter descriptivo o predictivo. Las predicciones sirven para prever el comportamiento futuro de algún tipo de entidad mientras que una descripción puede ayudar a su comprensión. De hecho, los modelos predictivos pueden ser descriptivos (hasta donde sean comprensibles por personas) y pueden emplearse para realizar predicciones 4.3.2.3.

El objetivo en este trabajo es desarrollar un modelo que permita predecir el comportamiento de la Sigatoka Negra, para la cual se usaran las variables biológicas que brindan una información

descriptiva sobre la misma. Entre las variables que han venido destacándose (4.3.2.3, 4.3.2.3, 4.3.2.3, 4.3.2.3) cómo variable a predecir se encuentran la severidad de la enfermedad, hoja más joven mancha, suma bruta en la hojas 2, 3, 4 y el estado de evolución de la enfermedad, siendo así, es imperante saber que existe una diversas alternativas de acuerdo a como se quiere predecir ese valor, es decir, que de acuerdo con el objetivo o tarea perseguida con respecto a cómo se quiere reportar la información, la siguientes alternativas :

- Clasificación o discriminación 4.3.2.3: los ejemplos se presentan como un conjunto de pares de elementos de dos conjuntos,  $\mathcal{D}=\{\langle e, s \rangle : e \in E, s \in S\}$ , donde  $S$  es el conjunto de valores de salida, y  $E$  es el conjunto de ejemplos. Los ejemplos  $e$ , al ir acompañados de un valor de  $S$ , denominan comúnmente ejemplos etiquetados  $\langle e, s \rangle$  y, en consecuencia,  $\mathcal{D}$  se denomina conjunto de datos etiquetado. El objetivo es aprender una función  $\lambda : E \rightarrow S$ , denominada clasificador, que represente la correspondencia existente en los ejemplos, es decir, para cada valor de  $E$  tiene un único valor para  $S$ . Además,  $S$  es nominal, es decir, puede tomar un conjunto de valores  $c_1, c_2, \dots, c_m$ , denominados clases (cuando el número de clases es dos, tenemos lo que se llama clasificación binaria). La función aprendida será capaz de determinar la clase para cada nuevo ejemplo sin etiquetar, es decir dará un valor de  $S$  para cada valor de  $e$ .
- Clasificación suave 4.3.2.3: la presentación del problema es la misma que la de la clasificación, pares de elementos de dos conjuntos,  $\mathcal{D}=\{\langle e, s \rangle : e \in E, s \in S\}$ . Además de la función  $\lambda : E \rightarrow S$ , se aprende otra función  $\theta : E \rightarrow \mathbb{R}$  que significa el grado de

certeza de la predicción hecha por la función  $\lambda$ . Lógicamente, siempre es preferible tener un clasificador suave que acompañe a las predicciones de una medida de certeza o de fiabilidad de dichas predicciones (aunque sea una mera estimación). Este tipo de extensión permite realizar otras aplicaciones, como son los rankings de predicciones o la selección de los  $n$  mejores ejemplos.

- Estimación de probabilidad de clasificación 4.3.2.3: se trata, en realidad, de una generalización de la clasificación suave. La presentación del problema es la misma que la de la clasificación normal y suave, pares de elementos de dos conjuntos,  $\mathcal{D} = \{\langle \mathbf{e}, \mathbf{s} \rangle : \mathbf{e} \in E, \mathbf{s} \in S\}$ . La función a aprender, sin embargo, es distinta de la clasificación y de la función suave. Se trata de aprender exclusivamente  $m$  funciones  $\theta : E \rightarrow \mathbb{R}$ , donde  $m$  es el número de clases. Es decir, cada función a aprender retorna para cada ejemplo  $m$  un valor real  $\mathbf{p}$ . Cada uno de estos valores  $\mathbf{p}$  se denomina probabilidad de la clase  $i$  y significa el grado de certeza de que un ejemplo sea de la clase  $i$ . Idealmente, si además se cumple que  $\forall \mathbf{p} : 0 \leq \mathbf{p}$ , y además,  $\sum \mathbf{p} = 1$ , estas  $\mathbf{p}$  representan la probabilidad de que un ejemplo sea de la clase 1. El conjunto de funciones aprendidas se denomina estimador de probabilidad. Nótese que si se posee un estimador de probabilidades, para predecir la clase sólo se tiene que hacer  $\arg\max \mathbf{p}$ . A veces, especialmente para dos clases, sólo se proporciona una de las probabilidades (ya que la otra se obtiene como  $1 - \mathbf{p}$ ). En este caso un estimador de probabilidades y un clasificador suave vienen a ser lo mismo, ya

que con la probabilidad de una de las clases, podemos saber la clase predicha y su certeza.

- Categorización 4.3.2.3: no se trata de aprender una función, sino una correspondencia. Es decir, cada ejemplo de  $\delta = \{\langle e, s \rangle : e \in E, s \in S\}$ , así como la correspondencia a aprender  $\lambda : E \rightarrow S$ , pueden asignar varias categorías a un mismo  $e$ , a diferencia de la clasificación, que sólo asigna una y sólo una. Dicho de otra manera, un ejemplo podría tener varias categorías asociadas. La categorización se puede presentar también en forma de categorización suave (cada categoría asignada va acompañada de su certeza) o en forma de un estimador de probabilidades (se estima una probabilidad para todas las categorías), en este caso la suma de probabilidades puede ser mayor que 1. En este caso, nótese que, aunque en  $\delta$  podemos tener varias etiquetas para el mismo ejemplo, la función aprendida por un estimador de probabilidades de clasificación y un estimador de probabilidades de categorización viene a ser prácticamente lo mismo. En realidad usarlo como clasificador consiste en seleccionar la clase de mayor probabilidad y realizar un categorizador consiste simplemente en seleccionar las  $k$  mejores categorías o las categorías que superen un cierto valor.
- Preferencias o priorización 4.3.2.3: el aprendizaje de preferencias consiste en determinar a partir de dos o más ejemplos, un orden de preferencia. La definición formal es un poco más compleja. Cada ejemplo es en realidad una secuencia:  $\langle e_1, e_2, \dots, e_k \rangle, e_i \in E, k \geq 2$ , donde el orden de la secuencia representa la predicción. Un conjunto de datos para este problema es, por tanto, un conjunto de secuencias



$\delta = \{\langle \mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_k \rangle : \mathbf{e} \in E\}$ . Otra manera alternativa de presentar los datos es mediante un orden parcial, que se puede considerar un caso particular de la anterior, donde las secuencias sólo tienen dos elementos ( $k=2$ ). Ejemplos: dados una serie de candidatos para un trabajo, dar un orden priorizado para cubrir el puesto (el modelo de preferencia se habrá estimado a partir de selecciones anteriores (priorizaciones) o comparaciones de grupos de candidatos anteriores, etc.). Una simplificación del problema sería obtener sólo preferencias entre dos elementos. El modelo aprendido sólo sería capaz de decir cuál prefiere entre dos objetos (decir si  $\mathbf{e}_i$  es mejor que  $\mathbf{e}_j$ ), pero no ordenar más de dos objetos. En otros casos, lo que se aprende es un valor real de la magnitud del ejemplo, y el resultado se parece mucho a la regresión (aunque el valor de salida es ficticio). Quizá lo más característico de esta tarea es la presentación de los datos, ya que, por ejemplo, con un clasificador suave o un estimador de probabilidades también se pueden hacer priorizaciones, aunque aquí lo que se prioriza es la clase no el ejemplo completo.

- Regresión 4.3.2.3: es quizá la tarea más sencilla de definir. El conjunto de evidencias son correspondencias entre dos conjuntos  $\delta : E \rightarrow S$ , donde  $S$  es el conjunto de valores de salida. Al igual que con la clasificación, los ejemplos, al ir acompañados de un valor de  $S$ , se denominan comúnmente ejemplos etiquetados y  $\delta$  es un conjunto de datos etiquetado. El objetivo es aprender una función  $\lambda : E \rightarrow S$  que represente la correspondencia existente en los ejemplos, es decir, para cada valor de  $E$  se posee un único valor para  $S$ . La diferencia respecto a la clasificación es que  $S$  es numérico, es decir puede

ser un valor entero o real. La regresión se conoce con otros nombres: interpolación (generalmente cuando el valor predicho está en medio de otros) o estimación (cuando se trata de algo futuro). Es importante destacar de nuevo las conexiones entre distintas tareas; por ejemplo, un modelo de regresión se puede convertir en un clasificador binario suave si se establece un umbral  $u$  para la salida y de la función  $\lambda$ , es decir, si  $y < u$  se tiene una clase y si  $y \geq u$  se está en presencia de la otra clase. Este es el caso del método denominado “regresión logística” que en realidad se utiliza para clasificación.

No obstante cada una de las tareas mencionada anteriormente requiere de técnicas, método o algoritmos para resolverlas. Esto lógicamente permite que una tarea pueda usar un abanico de técnicas y una técnica pueda dar solución a más de una tarea. Entre algunas tareas se pueden mencionar:

- Técnicas algebraicas y estadísticas 4.3.2.3: se basan, generalmente, en expresar modelos y patrones mediante formulas algebraicas, funciones lineales, funciones no lineales, distribuciones o valores agregados estadísticos tales como media, varianza correlaciones, etc. Frecuentemente, estas técnicas, cuando obtienen un patrón, lo hacen a partir de un modelo ya predeterminado del cual, se estima los parámetros, de ahí surge el nombre de técnicas paramétricas. Algunas de los algoritmos más conocidos dentro de las técnicas son la regresión lineal, la regresión logarítmica y la regresión logística. La discriminantes lineales y no lineales, basado en funciones predefinidas, es decir, discriminante paramétricos, entran a esta categoría.

- Técnicas Bayesianas 4.3.2.3: está basada en estimar la probabilidad de pertenecía [a una clase o grupo], mediante la estimación de probabilidades condicionales inversas o a priori, utilizando para ello el teorema de Bayes. Algunos algoritmos muy populares son el clasificados bayesiano *naive*, los métodos basados en máxima verisimilitud y el algoritmo EM. Las redes bayesianas generalizan las topologías de las interacciones probabilísticas entre las variables y permiten representar gráficamente dichas interacciones.
- Técnicas basadas en conteos de frecuencias y tablas de contingencia 4.3.2.3: estas técnicas se basan en contar la frecuencia en la que dos o más sucesos se presenten conjuntamente. Cuando el conjunto de sucesos posibles es muy grande, existen algoritmos que van comenzando por pares de sucesos e incrementando los conjuntos solo en aquellos casos que las frecuencias conjuntas superen un cierto umbral. Ejemplos de estos algoritmos son el algoritmo A priori o similares.
- Árboles basados en árboles de dedición y sistemas de reglas 4.3.2.3: son técnicas que, además de su representación en forma de reglas, se basan en dos tipos de algoritmos: los algoritmos denominados *divide y vencerás*, como el ID3/C4.5 o el CART, y los algoritmos denominados *separa y vencerás*, como el CN2.
- Técnicas relacionales, declarativas y estructurales 4.3.2.3: la característica principal de este conjunto de técnicas es que representan los modelos mediante lenguajes declarativos, como los lenguajes lógicos, funcionales o lógicos-funcionales. Las técnicas de programación lógicas inductiva ILP son más

representativas y las que han dado nombre a un conjunto de técnicas denominadas minería de datos relacional.

- Técnicas basadas en redes neuronales artificiales 4.3.2.3: se trata de técnicas que aprenden un modelo mediante entrenamiento de los pesos que conecten un conjunto de nodos o neuronas, La topología de la red y los pesos de las conexiones determinan el patrón aprendido. Existen innumerables variantes de organización: perceptrón simple, redes multicapa, redes de base radial, redes de kohonen, etc., con no menos algoritmos diferentes para cada organización; el más conocido es el retro propagación (*backpropagation*).
- Técnicas basadas en núcleo y máquinas de soporte vectorial 4.3.2.3: se trata de técnicas que intentan maximizar el margen entre los grupos o clases formadas. Para ellos se basan en unas transformaciones que puedan aumentar la dimensionalidad. Estas transformaciones se llaman núcleos (*kernels*). Existen muchísimos variantes, dependiendo del núcleo utilizado y de la manera de trabajar con el margen.
- Técnicas estocásticas y difusas 4.3.2.3: Son técnicas en las que o bien los componentes aleatorios son fundamentales, como el *simulated annealing*, los métodos evolutivos y genéticos o bien al utilizar funciones de pertenencia difusas (*fuzzy*). Bajo este paraguas se incluyen la mayoría de las técnicas que, junto a las redes neuronales, forman lo que se denomina computación flexible.
- Técnicas basadas en caso, en densidad o distancia 4.3.2.3: son métodos que se basan en distancia al resto de elementos, ya sea directamente, como el vecino más próximo (los casos más similares), de unas maneras más sofisticadas, mediante,

mediante la estimación, mediante la estimación de funciones de densidad. Además de los vecinos más próximos, algunos algoritmos muy conocidos son los jerárquicos, como Two-step o COBWEB, y los no jerárquicos, como K medias.

Además de todo lo anterior existen multitud de híbridos que dificultan más realizar una taxonomía razonable.

### **PREPARACIÓN DE DATOS.**

El proceso de recopilación y almacenado de los datos tanto climatológicos y biológicos difieren sustancialmente. Por un lado, las variables climatológicas son registradas por un sistema totalmente automático, que realiza mediciones y registro cada media hora durante las 24 horas al día, los 365 días del año. Este registro se ha llevado a cabo desde septiembre del 2001. Sin embargo, no se ha mantenido la consistencia estructural de los registros de las variables, aunado a estos algunas variables se dejan de registrar debido a desperfectos en el sistema. Esto ha conllevado a dejar de hacer las mediciones por algunos períodos de tiempo y a omitir los registros de las mediciones de algunas variables. Esto último modifica el registro de las variables en su estructura, es decir, que existen diferencias sustanciales en los datos de las variables climatológicas tanto como presencia de datos ausentes como de variables.

Por otro lado, el registro y almacenamiento de las variables biológicas es de forma manual y las mediciones son realizadas de forma periódica, cada 7 o 8 días, es decir, una medición por semana. Se cuenta con registros desde enero del 2002 hasta junio

del 2006, lo cual arroja 3 años de mediciones de las variables biológicas.

El objetivo de este trabajo es desarrollar un modelo predictivo para el brote de la Sigatoka Negra, basado en las posibles relaciones entre las variables climatológicas y las variables biológicas, lo cual implica una integración entre estos dos conjuntos de variables o mediciones. Por lo descrito en el párrafo anterior, existen varias incompatibilidades entre estos conjuntos de datos. La primera es de forma, por ejemplo, ambos conjuntos de datos difieren en la periodicidad de sus mediciones, no permitiendo una integración directa, lo cual implica que se debe tratar uno de los dos conjuntos de datos sin perder información. Por otro lado, cada subconjunto posee sus propios errores de medición, en términos de variables no consideradas, datos ausentes, y atípicos. Por lo tanto, en la preparación de los datos hay que implantar pretratamiento por separado a cada conjunto de datos. Este pretratamiento consiste en una limpieza y transformación de los datos por separado. Luego implementar la integración, exploración y selección de las variables para el desarrollo posterior del modelo predictivo.

## **2.5 EXPLORACIÓN Y LIMPIEZA DE DATOS**

### **Análisis exploratorio de los datos**

El análisis exploratorio de datos es un conjunto de técnicas estadísticas cuya finalidad es conseguir un entendimiento básico de los datos y de las relaciones existentes entre las variables analizadas. El análisis exploratorio de datos proporciona métodos sistemáticos sencillos para organizar y prepararlos, detectar fallos en el diseño y recogida de los mismos, tratamiento y evaluación de

datos ausentes (*missing*), identificación de casos atípicos (*outliers*) y comprobación de los supuestos subyacentes en la mayor parte de las técnicas multivariantes (normalidad, linealidad, homocedasticidad). Para realizar un análisis exploratorio de datos conviene seguir las siguientes etapas 4.3.2.3:

- Preparar los datos para hacerlos accesibles a cualquier técnica estadística.
- Realizar un examen gráfico de la naturaleza de las variables individuales a analizar y un análisis descriptivo numérico que permita cuantificar algunos aspectos gráficos de los datos.
- Realizar un examen gráfico de las relaciones entre las variables analizadas y un análisis descriptivo numérico que cuantifique el grado de interrelación existente entre ellas.
- Evaluar, si fuera necesario, algunos supuestos básicos subyacentes a muchas técnicas estadísticas como, por ejemplo, la normalidad, linealidad y homocedasticidad.
- Identificar los posibles casos atípicos (*outliers*) y evaluar el impacto potencial que puedan ejercer en análisis estadísticos posteriores.
- Evaluar, si fuera necesario, el impacto potencial que pueden tener los datos ausentes (*missing*) sobre la representatividad de los datos analizados.

### **Evaluación y Tratamiento de datos atípicos**

Los casos o datos atípicos son observaciones con características diferentes de las demás. Los mismos no pueden ser caracterizados categóricamente como benéficos o problemáticos sino que deben ser contemplados en el contexto del análisis y debe evaluarse el tipo de información que pueden proporcionar. Su principal

problema radica en que son elementos que pueden no ser representativos de la población pudiendo distorsionar seriamente el comportamiento de los contrastes estadísticos. Por otra parte, aunque diferentes a la mayor parte de la muestra, pueden ser indicativos de las características de un segmento válido de la población y, por consiguiente, una señal de la falta de representatividad de la muestra 4.3.2.3.

### **Tipos de datos atípicos**

Los casos atípicos pueden clasificarse en 4 categorías.

- La primera categoría contiene aquellos casos atípicos que surgen de un error de procedimiento, tales como la entrada de datos o un error de codificación. Se deberían eliminar en el filtrado de los datos, y si no se puede, deberían eliminarse del análisis o recodificarse como datos ausentes.
- La segunda clase es la observación realizada como consecuencia de un acontecimiento extraordinario. En este caso, el outlier no representa ningún segmento válido de la población y puede ser eliminado del análisis.
- La tercera clase contiene las observaciones cuyos valores caen dentro del rango de las variables observadas pero que son únicas en la combinación de los valores de las mismas, debiendo ser retenidas en el análisis pero, estudiando qué influencia ejercen en los procesos de estimación de los modelos considerados.
- La cuarta y última clase comprende las observaciones extraordinarias para las cuales el investigador no tiene explicación. En estos casos lo mejor que se puede hacer es replicar el análisis con y sin dichas observaciones con el fin de



analizar su influencia sobre los resultados. Si las mismas son influyentes el analista debería reportarlo en sus conclusiones y debería averiguar el por qué de ellas 4.3.2.3..

### **Identificación de datos atípicos**

Los casos atípicos pueden identificarse desde una perspectiva univariante o multivariante. La primera examina la distribución de observaciones para cada variable, seleccionando como casos atípicos aquellos cuyos valores caigan fuera de los rangos de la distribución. La cuestión principal consiste en el establecimiento de un umbral para la designación de caso atípico. Esto se puede hacer gráficamente mediante histogramas o diagramas de caja o bien numéricamente, mediante el cálculo de puntuaciones tipificadas. Para muestras pequeñas (de 80 o incluso menos observaciones), las pautas sugeridas identifican como tales aquellos casos con valores estándar de 2,5 o superiores. Cuando los tamaños muestrales son mayores, las pautas sugieren que el valor umbral sea 3 4.3.2.3.

Para la detección de los datos atípicos:

- a. Gráfico de caja y bigotes.
- b. Diagrama de control. Representación gráfica con una línea central que representa el valor medio de la variable y dos líneas horizontales llamadas límite superior e inferior de control (LSC, LIC). Si los valores se encuentran entre los límites de control, se considera que no hay atípicos. Estos límites se suelen fijar a 3 desviaciones típicas. (En la distribución normal el intervalo a 1, 2, 3 y 4 sigmas incluye el 68.26%, 95.46%, 99.73%, 99.99% de los valores) 4.3.2.3..
- c. Prueba de Dixon, Grubs, Rosmer o Chi Cuadrado 4.3.2.3. La

hipótesis nula es no existen atípicos, p-valores menores que 0.05 nos indica que hay atípicos al 95% de confianza.

## **PRUEBAS PARA LA DETERMINACIÓN DE DATOS ATÍPICOS**

### **Prueba de Rosner**

La prueba de Rosner (4.3.2.3 4.3.2.3 4.3.2.3 4.3.2.3) para detectar k valores atípicos puede ser utilizada cuando el número de los datos de referencias son 25 o más. Esta prueba identifica los valores atípicos que están por encima o por debajo de la media. Los datos se ordenan ascendente y se determina la media y la desviación estándar. El procedimiento exige el quitar el conjunto de observación,  $x$ , que es el más lejano de la media. Entonces una prueba estadística,  $R$ , se calcula:

$$R_{i+1} = \frac{|x_i - \bar{x}|}{S_i}$$

La estadística de  $R$  entonces se compara con el valor crítico. La hipótesis nula, indicando que los datos se ajustan a una distribución normal, entonces se prueba. Si  $R$  es menos que el valor crítico, la hipótesis nula no puede ser rechazada, y por lo tanto no hay valores atípicos. Si  $R$  es mayor que el valor crítico, se rechaza la hipótesis nula y la presencia de los k valores atípicos se acepta. Esta prueba se puede también utilizar con datos distribuidos registro-normal, cuando los logaritmos de los datos se utilizan para el cómputo.

### **Prueba de Grubbs**

Prueba de Grubbs (4.3.2.3 4.3.2.3 4.3.2.3 4.3.2.3) es recomendada por la agencia Norteamérica de Protección del Medio Ambiente (EPA) como prueba estadística para los valores atípicos. EPA sugiere tomar los logaritmos de los datos ambientales, que son a

menudo registro-normal distribuido. Los datos se alinean en orden ascendente y se calcula la media y la desviación estándar. El punto de referencias más bajo o más alto que pueden ser usados para probar los valores atípicos.

- La estadística del  $\tau$  para el valor más bajo es:  $\tau = \frac{\bar{x} - x_1}{S}$
- La estadística del  $\tau$  para el valor más alto es:  $\tau = \frac{x_n - \bar{x}}{S}$

La estadística del  $\tau$  se compara con un valor crítico del  $\tau$  para el tamaño de muestra y el  $\alpha$  seleccionado. Si la estadística del  $\tau$  es mayor que el  $\tau$  crítico, se rechaza la hipótesis nula y la conclusión es que el dato bajo consideración es un valor atípico.

### Prueba de Dixon.

La prueba de Dixon (4.3.2.3 4.3.2.3 4.3.2.3 4.3.2.3) se utiliza generalmente para detectar una pequeña cantidad de valores atípicos. Esta prueba puede ser utilizada cuando el tamaño de la muestra está entre 3 y 25 observaciones. Los datos se alinean en la orden ascendente, entonces basada en el tamaño de muestra, la estadística del  $\tau$  para el valor más alto o el valor más bajo se estima, de la manera siguiente:

Observaciones	Valor sospechoso más alto	Valor sospechoso más bajo
3 a 7	$\tau = \frac{x_n - x_{n-1}}{x_n - x_1}$	$\tau = \frac{x_2 - x_1}{x_n - x_1}$
8 a 10	$\tau = \frac{x_n - x_{n-1}}{x_n - x_2}$	$\tau = \frac{x_2 - x_1}{x_{n-1} - x_1}$
11 a 13	$\tau = \frac{x_n - x_{n-2}}{x_n - x_2}$	$\tau = \frac{x_3 - x_1}{x_{n-1} - x_1}$

14 a 20 -30	$\tau = \frac{x_n - x_{n-2}}{x_n - x_3}$	$\tau = \frac{x_3 - x_1}{x_{n-2} - x_1}$
-------------	--	--

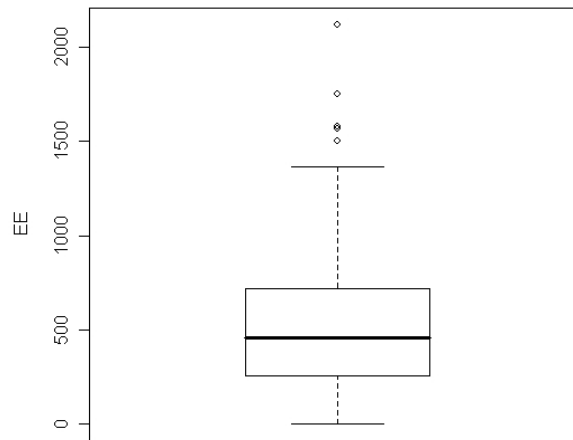
La estadística del  $\tau$  se compara a un valor crítico en un valor elegido del  $\alpha$ . Si la estadística del  $\tau$  es menos que el valor crítico, la hipótesis nula no se rechaza, y la conclusión es que no hay valores atípicos presentes. Si la estadística del  $\tau$  es mayor que el valor crítico, se rechaza la hipótesis nula, y la conclusión es el valor más extremo es un valor atípico. Para comprobar si hay otros valores atípicos, la prueba de Dixon se puede repetir, sin embargo, Eficacia de esta prueba disminuye mientras que el número de repeticiones aumenta.

### **Regla del gráfico de caja y bigotes**

La regla del gráfico (4.3.2.3 4.3.2.3 4.3.2.3 4.3.2.3) de caja y bigotes es una prueba visual a examinar para saber si hay valores atípicos (Figura N° 9). La gama interquartile es incluida en una caja y los intervalos de la confianza del 5% y del 95% se indican con las barras del error fuera de la caja. Los valores que están fuera del intervalo de la confianza son posibles valores atípicos

Límite del intervalo de la confianza del 95%:  $\frac{x - Q_3}{Q_3 - Q_1} > k$

Límite del intervalo de la confianza del 5%:  $\frac{Q_1 - x}{Q_3 - Q_1} > k$



Estado de evolución de la enfermedad

Figura N° 9. Gráfico de caja de estado de evolución de la enfermedad.

### Tratamiento de los datos atípicos

Si se han detectado los valores atípicos e idealmente sus causas, los datos pueden ser tratados. Las posibles acciones sobre los valores atípicos pueden ser:

- 1) Ignorar (dejar pasar): Algunos algoritmos son robustos a valores atípicos.
- 2) Filtrar (eliminar o reemplazar) columna: solución extrema, la posible existencia de otra columna dependiente con datos de mayor calidad. Preferible a eliminar una columna es reemplazarla por una columna discreta diciendo si el valor era normal o atípico. Sustituyendo el caso atípico por el caso no atípicos.
- 3) Filtrar la fila: Puedes sesgar los datos, porque muchas veces las causas de un dato atípico puede estar relacionando con casos especiales.
- 4) Reemplazar el valor: por el valor “nulo” si el método o algoritmos estadísticos lo puede tratar muy bien o reemplazarlo por el

máximo o el mínimo valor, dependiendo de donde se encuentre el valor atípico (sea mayor a la media o menor a la media), o por media. A veces se puede predecir el valor a partir de otros datos, utilizando cualquier técnica ML.

- 5) Discretizar: Transformar un valor continuo en un valor discreto (por ejemplo: Muy alto, alto, medio, bajo, muy bajo).

## **TRATAMIENTO Y EVALUACIÓN DE DATOS AUSENTES**

Los datos ausentes son algo habitual en el análisis de datos; de hecho, rara es la investigación donde no aparecen. La ocupación primaria del investigador debe ser determinar las razones que subyacen en el dato ausente buscando entender el proceso principal de esta ausencia para seleccionar el curso de acción más apropiado. Para ello se debe determinar cuál es el proceso de datos ausentes, entendido como cualquier evento sistemático externo a la variable observada quedando como consecuencia la ausencia de datos. En particular, el investigador debe analizar si existe algún patrón no aleatorio en dicho proceso que pueda sesgar los resultados obtenidos debido a la pérdida de representatividad de la muestra analizada 4.3.2.3.

### **Tipos de valores ausentes**

Se distinguen las dos situaciones siguientes:

- a) Datos ausentes prescindibles: son resultado de procesos que se encuentran bajo el control del investigador y pueden ser identificados explícitamente. En estos casos no se necesitan soluciones específicas para la ausencia de datos pues es inherente a la técnica usada.
- b) Datos ausentes no prescindibles: son resultado de procesos que

no se encuentran bajo el control del investigador y/o no pueden ser identificados explícitamente. En estos casos se debe analizar si existen o no patrones sistemáticos en el proceso que puedan sesgar los resultados obtenidos.

Si los datos ausentes son no prescindibles conviene, por lo tanto, analizar el grado de aleatoriedad presente en los mismos. Según este grado el proceso de datos ausentes se puede clasificar del siguiente modo:

a) Datos ausentes completamente aleatorios (MCAR): es el mayor grado de aleatoriedad y se da cuando los datos ausentes son una muestra aleatoria simple de la muestra sin un proceso subyacente que tiende a sesgar los datos observados. En este caso se podría solucionar el problema sin tener en cuenta el impacto de otras variables.

b) Datos ausentes aleatorios (MAR): aquí el patrón de los datos ausentes en una variable  $Y$  no es aleatorio sino que depende de otras variables de la muestra  $X$ . Ahora bien, para cada valor de  $X$ , los valores observados de  $Y$  sí representan una muestra aleatoria de  $Y$ . Si los datos ausentes son MAR cualquier solución al problema deberá tener en cuenta los valores de  $X$  dado que afectan al proceso generador de datos ausentes.

c) Datos ausentes no aleatorios: en este caso existen patrones sistemáticos en el proceso de datos ausentes y habría que evaluar la magnitud del problema calibrando, en particular, el tamaño de los sesgos introducidos por dichos patrones. Si éstos son grandes habría que atacar el problema directamente intentando averiguar cuáles son dichos valores.

## **Tratamiento de los datos ausentes**

Una vez establecidos los datos ausentes, idealmente, sus causas, se pueden tratar los datos de la manera siguiente:

- a. Ignorar (dejar pasar): Algunos algoritmos son robustos a valores ausentes.
- b. Filtrar (eliminar o reemplazar) columna: solución extrema, pero a veces la proporción de valores ausentes es tan alta que la columna no tienen arreglo. Por otra parte, la posible existencia de otra columna dependiente con datos de mayor calidad.
- c. Filtrar la fila: Puedes sesgar los datos, porque muchas veces la causa de un dato faltante están relacionados con casos o tipos muy especiales.
- d. Reemplazar el valor: el reemplazo de los datos ausentes puede usarse en el caso de que no sean muchos datos por valor y se haría por valores que preserve la media o la varianza, en el caso de valores numérico, o en el moda, en el caso nominal. Otra más sofisticada de estimar el valor de reemplazo es “predecirlo” a partir del resto de valores o datos utilizando técnicas de predicción o algoritmos estadísticos para esto.
- e. Segmentar: se segmentan las duplas por valores que tienen disponibles. Se obtienen modelos diferentes para cada segmento y luego se combinan.
- f. Modificar las políticas de calidad de datos y esperar hasta que los datos ausentes estén disponibles.



## **CAPITULO III - MARCO METODOLÓGICO**

### **3 INTRODUCCIÓN**

A continuación se describe de manera general el procedimiento al cual serán sometidos los datos para obtener como resultado un modelo que prediga el brote de la Sigatoka Negra. Primeramente se describe la preparación de los datos que parte desde la determinación de los valores faltantes y atípicos, con el objetivo de corregir los datos antes de ser procesados por cualquier herramienta informática. Se continua con la selección de las variables más importantes de acuerdo con el criterio de escoger aquella variable que individualmente mejor prediga a una de las variables descriptora de la Sigatoka Negra tales como severidad de la enfermedad, hoja más joven manchada, suma bruta en la hojas 2, 3, 4 y el estado de evolución de la enfermedad. Por último se describen las principales técnicas de generación del modelo usadas en este trabajo.

#### **3.1 Pre-Procesamiento de los datos**

##### **3.1.1 Búsqueda de datos anómalos o extremos:**

Para detectar valores extremos o anómalos en los datos se utilizan las herramientas estadísticas tales como caja de bigotes, el test de Grubbs, test de Nixón entre otros. Esta búsqueda se facilita mediante la utilización de una herramienta informática como el software estadístico R (versión 2.2.1), dicha herramienta cuenta con una librería de funciones que permite llevar a cabo la exploración y análisis de datos anómalos o extremos, tales como las funciones: *mvoutlier* (Función que permite la detección Multivariante de los valores anómalos basado en los métodos robustos) y *outliers* (Función que contempla un conjunto de

pruebas para la identificación de los valores anómalos). Además cuenta con la capacidad de programar las rutinas requeridas por el usuario, lo cual hace que este software se convierta en una pieza clave en la búsqueda de un modelo de preaviso para el brote de la Sigatoka Negra.

### **3.1.2 Búsqueda de datos ausentes:**

La determinación de los datos ausentes se lleva a cabo mediante una exploración visual de los datos estudiados. Como se mencionó en la capítulo del marco teórico, existen varias alternativas para tratar los valores ausentes, pueden ser omitidos en las estimaciones, o pueden ser estimados por métodos estadísticos. Este último tratamiento requiere que el volumen de datos ausentes sea considerablemente pequeño comparado con el volumen total.

### **3.1.3 Integración de las bases de datos:**

Debido a la heterogeneidad en la secuencia de las fechas de medición de las variables biológicas (cada 7 u 8 días) y climatológicas (mediciones cada 30 o 60 minutos) no es directa la integración de las dos bases de datos existentes. Por lo cual se procesa cada conjunto de datos por separado para llevar ambas bases de datos a una misma escala temporal de mediciones. En el caso de las variables climatológicas, se requiere que su base de medición sea diaria. Para obtener los valores diarios se determinan los promedios de las variables, exceptuando la variable precipitación en la que se considera la acumulación diaria.

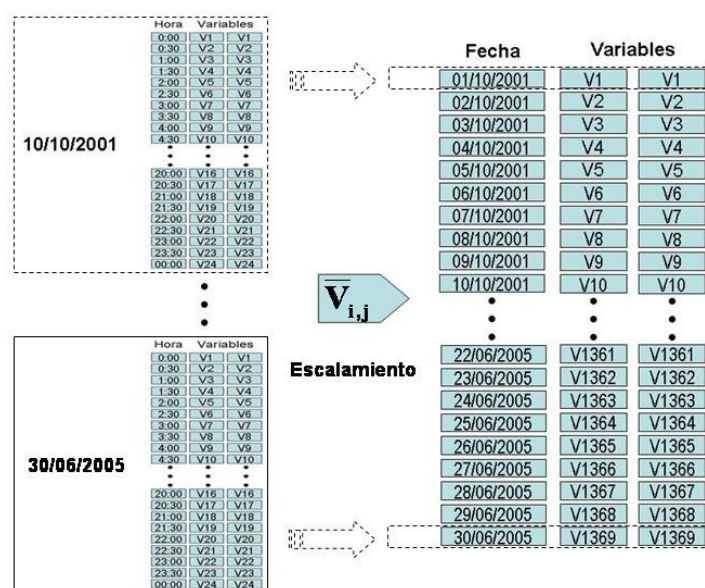


Figura N° 10. Reajuste de las mediciones de las variables.

### 3.1.4 Selección de las variables:

La selección de la variables es primordial para el desarrollo del modelo predictivo o de pronóstico temprano del brote la Sigatoka Negra. Para llevar a cabo la selección de las variables se implementará el método Win Panel descrito con más detalle en el marco 4.3.2.3. Se presupone la existencia de una correlación entre las condiciones meteorológicas y el brote de la Sigatoka Negra. Esta relación viene dada por los efectos que deberían ejercer las variables climatológicas durante el proceso de inoculación y propagación de la enfermedad en el cultivo. No obstante, como toda enfermedad de origen fungoso, los efectos no son detectables o visibles inmediatamente después que ha ocurrido la infección sino luego de un tiempo posterior a ella. Esto hace necesario obtener la relación que existe en los parámetros climatológicos registrados mucho tiempo antes de la medición de las variables biológicas. La estrategia que se plantea es la búsqueda de promedios o acumulados de las variables climatológicas para períodos de tiempo (conocidos como ventanas) a lo largo del lapso

de tiempo que antecede a la medición de la variable biológica (Figura N° 3). Así es necesario realizar esta búsqueda para diferentes tamaños de la ventana y lapsos de tiempo previos a la ventana y la medición de la variable biológica. La ventana y lapso de tiempo óptimo son los que arrojan mayor valor correlación. Esto permite estudiar los efectos de las variables por día y de varios días. De acuerdo a las recomendaciones técnicas suministradas por el Ing. Carlos Gómez, la búsqueda de la ventana óptima se realiza en un lapso de 14 a 140 días previos a la medición de la severidad enfermedad, pero para el caso de las variables biológicas que dependen del ciclo de vida de la hoja, como estado de evolución de la enfermedad, hoja más joven manchada y la suma bruta en las hojas 2, 3, el máximo periodo de exploración es de 4 semanas. La amplitud de las ventanas consideradas está entre 3 y 14 días. La combinación del lapso de exploración, el retraso y la amplitud de la ventana generan el conjunto de casos o escenarios que se estudiarán en este trabajo.

En la Tabla N° 3 se presenta de forma resumida la estructura de los casos utilizados para el análisis de la correlación entre las variables biológicas, las variables dependientes o a predecir en este trabajo, y las variables climatológicas, las variables predictoras evaluadas. Para una amplitud de ventana dada se pueden considerar hasta ***n*** retrasos, siendo el máximo valor de ***n*** la diferencia entre el período de exploración de los datos y la amplitud de la ventana. Por ejemplo, C-3-2 indica el caso de estudio para una ventana de amplitud 3 días, es decir, los estadísticos climáticos corresponden a un resumen de 3 días, y un retraso de 2 días de antelación a la fecha de medición de la

variable biológica considerada. Esto se aplica en la exploración de cada una de las variables climatológicas y su relación con cada una de las variables biológicas, repitiéndose un total de 70 (la combinación de 10 variables climáticas con 7 variables biológicas) veces para cada uno de los elementos de la Tabla N° 3.

Tabla N° 3. Codificación de los casos de estudios para un lapso de exploración **n** números de días.

Amplitud de ventana (días)	Retraso de la ventana (días)					
	1	2	3	4	** *	n
2	C - 2 - 1	C - 2 - 2	C - 2 - 3	C - 2 - 4	** *	C - 2 - n
3	C - 3 - 1	C - 3 - 2	C - 3 - 3	C - 3 - 4	** *	C - 3 - n
4	C - 4 - 1	C - 4 - 2	C - 4 - 3	C - 4 - 4	** *	C - 4 - n
5	C - 5 - 1	C - 5 - 2	C - 5 - 3	C - 5 - 4	** *	C - 5 - n
6	C - 6 - 1	C - 6 - 2	C - 6 - 3	C - 6 - 4	** *	C - 6 - n
7	C - 7 - 1	C - 7 - 2	C - 7 - 3	C - 7 - 4	** *	C - 7 - n
8	C - 8 - 1	C - 8 - 2	C - 8 - 3	C - 8 - 4	** *	C - 8 - n
9	C - 9 - 1	C - 9 - 2	C - 9 - 3	C - 9 - 4	** *	C - 9 - n
10	C - $\frac{1}{0}$ - 1	C - $\frac{1}{0}$ - 2	C - $\frac{1}{0}$ - 3	C - $\frac{1}{0}$ - 4	** *	C - $\frac{1}{0}$ - n
11	C - $\frac{1}{1}$ - 1	C - $\frac{1}{1}$ - 2	C - $\frac{1}{1}$ - 3	C - $\frac{1}{1}$ - 4	** *	C - $\frac{1}{1}$ - n
12	C - $\frac{1}{2}$ - 1	C - $\frac{1}{2}$ - 2	C - $\frac{1}{2}$ - 3	C - $\frac{1}{2}$ - 4	** *	C - $\frac{1}{2}$ - n
13	C - $\frac{1}{3}$ - 1	C - $\frac{1}{3}$ - 2	C - $\frac{1}{3}$ - 3	C - $\frac{1}{3}$ - 4	** *	C - $\frac{1}{3}$ - n
14	C - $\frac{1}{4}$ - 1	C - $\frac{1}{4}$ - 2	C - $\frac{1}{4}$ - 3	C - $\frac{1}{4}$ - 4	** *	C - $\frac{1}{4}$ - n

Una vez implementados todos los casos de estudio entre las variables en cuestión, se seleccionan aquellos que posean mayor valor de correlación, obteniéndose la amplitud y retraso de la ventana óptima. La implantación de estos procedimientos se llevo a cabo en R.

Para visualizar mejor la información obtenida como resultado del análisis anterior. se desarrolló una valoración cualitativa de los resultados obtenidos. Esta implementación consistió en valorar el comportamiento a lo largo de la del lapso establecido para la exploración de la correlación. La escala valoración establecida fue la siguiente:

Tabla Nº 4. Tabla de calificación cualitativa.

Clasificación cualitativa de la correlación	Valoración
Excelente (0,91 - 1,00)	5
Muy buena (0,80 - 0,90)	4
Buena( 0,60 - 0,80)	3
Regular ( 0,41 - 0,60)	2
Mala ( 0,21 - 0,40)	1
Muy mala (0,00 - 0,20)	0

Adicionalmente todo el proceso se repitió para los datos considerados globalmente y por temporada climatológica. Se consideró como temporada de lluvia los registros de los meses de Mayo a Octubre y como temporada de sequía los datos recolectados entre los meses Noviembre y Abril.

### 3.2 Procesamiento de los datos

### **3.2.1 Búsqueda del modelo de pronóstico temprano o preaviso:**

Una vez seleccionadas las variables que individualmente mejor predicen a las variables climatológicas, se implementa la construcción de una nueva tabla de datos que contiene los datos de estas variables. Esta tabla se procesa de manera de almacenar los datos en el formato ARFF, requerido por la herramienta informática seleccionada. En nuestro caso se utiliza WEKA (versión 3.5) para la búsqueda de patrones o conocimiento que subyace en los datos.

La segunda etapa es la exploración y búsqueda del modelo de pronóstico temprano aplicando los algoritmos disponibles WEKA. De la diversidad de esquemas de aprendizaje disponibles en la herramienta, solo trabajamos con aquellos esquemas que permiten generar un modelo predictivo como los relativos a algoritmos de clasificación y regresión.

Entre los esquemas con que cuentan WEKA para la clasificación por técnicas de árboles, fueron usadas:

- Tabla de decisión
- Modelo de aprendizaje M5.(*M5P*)
- Árboles de decisión de un nivel (*DecisionStump*).
- Reglas conjuntivas.

Entre los esquemas usados para las técnicas de regresión:

- Mínimos cuadrados. (*leastMedSq*)
- Regresión lineal.

Debido a no contar con los datos suficientes, todo el conjunto de datos se usó en el entrenamiento y se usó validación cruzada de 10 particiones. Las métricas consideradas para evaluar la bondad de

ajuste de los modelos son el error relativo, el error absoluto y la correlación.



## CAPITULO IV - RESULTADOS

### 4 INTRODUCCIÓN

A continuación se describen y se discuten de manera general los principales resultados obtenidos durante el desarrollo del trabajo. Primeramente se exponen y discuten los resultados obtenidos en la exploración y limpieza de los datos. Seguidamente se describen los resultados del proceso de selección de las variables a considerar en la base de datos integrada. Estos resultados se presentan para los diferentes períodos de tiempo considerados: la estación lluviosa, la estación seca y sin discriminar entre las estaciones. Por último se muestra el modelo ajustado en cada uno de los casos, se indica cual fue el esquema de aprendizaje y se reportan métricas que permitan estimar la bondad del modelo. En la última sección se discuten los resultados obtenidos y se presentan las conclusiones del trabajo.

### 4.1 EXPLORACIÓN Y LIMPIEZA DE DATOS

#### 4.1.1 Datos atípicos

Basados en la prueba de Grubbs, para valores atípicos y el análisis de los gráficos de cajas y bigotes se obtuvieron los siguientes resultados para los datos biológicos (Tabla N° 5) y climáticos (Tabla N° 6).

Tabla N° 5. Resultados de la prueba de Grubbs aplicada a las variables biológicas.

Variables Biológicas	G	U	p-valor	Valores Atípicos
Hoja más joven manchada	2.7458	0.9574	4.99E-01	14.60
Severidad de la	4.848	0.867	4.75E-	150.80

enfermedad	3	2	05	
Estado de evolución	4.206 4	0.900 0	1.45E- 03	2120.00
Suma Bruta	4.327 2	0.894 2	7.96E- 04	2181.76
Suma Bruta de la Hoja 2	10.37 07	0.392 4	0.00E+ 00	100.00
Suma Bruta de la Hoja 3	5.025 0	0.857 3	1.67E- 05	880.00
Suma Bruta de la Hoja 4	3.827 6	0.917 2	8.43E- 03	1240.00

Tabla N° 6. Resultados de la prueba de Grubbs aplicada a las variables climáticas.

Variables climatológicas	G	U	p-valor	valor atípico
Temperatura promedio del aire (°C)	6.189 7	0.967 2	2.56E- 07	0.00
Temperatura máxima del aire (°C)	3.202 5	0.991 2	7.78E- 01	43.10
Temperatura mínima del aire (°C)	6.206 2	0.967 0	2.29E- 07	0.00
Temperatura de la evapotranspiración (°C)	6.379 1	0.959 6	5.89E- 08	0.52
Radiación solar (W/m²)	8.885 1	0.932 6	0.00E+ 00	875.67
Energía solar incidente (Ly).	6.259 0	0.966 6	1.63E- 07	58.16
Precipitación acumulada (mm)	10.15 34	0.912 0	0.00E+ 00	104.40
Humedad relativa (%)	2.884 2	0.991 7	1.00E+ 00	62.64
Temperatura del punto de rocío (°C)	5.973 1	0.964 5	8.48E- 07	0.00

Pero luego de consultar al ingeniero Carlos Gómez del responsable de la Estación Local Chama del Instituto Nacional de

Investigaciones Agrícolas, se determinó que la mayoría de estos datos deberían ser considerados, ya que son valores posibles de dichas variables. La inspección visual de los registros de datos, conjuntamente con el ingeniero, determinó algunos pocos valores fuera del rango que fueron considerados errores de transcripción y corregidos.

#### **4.1.2 Datos ausentes.**

La Tabla N° 7 resume la cantidad de valores ausentes por variable climatológica, así como el porcentaje que esta representa en la muestra total de los datos existentes (1451 valores). En esta tabla algunas de las variables se encuentran solapadas, es decir, la variable temperatura representa a la temperatura promedio, máxima y mínima mientras que la variable solar representan la energía y la radiación solar, las cuales poseen igual cantidad y porcentaje de valores ausentes.

Tabla N° 7. Tabla del número de valores ausentes por variable climatológica.

Variable climatológica	Nº de datos ausentes	Porcentaje de ausencia
Temperatura	283	19,50%
Evapotranspiración	443	30,50%
Solar	279	19,20%
Velocidad del viento	283	19,50%
Precipitación acumulada	278	19,20%
Humedad	445	30,70%
Punto de Rocío	445	30,70%

En promedio el porcentaje de valores ausentes es de  $23.8 \pm 5.8 \%$ , esta cantidad de datos ausentes se debe mayormente a problemas con la estación meteorológica que presentó fallas. Debido a la cantidad de datos ausentes y a su disposición en los datos como cúmulos se hace prácticamente imposible estimar estos valores ausentes con cualquier método estadístico 4.3.2.3. Por lo cual, al momento de ser procesados estos registros no fueron tomados en consideración.

## 4.2 INTEGRACIÓN Y SELECCIÓN DE DATOS

A continuación se describe los resultados del proceso de selección de las variables climatológicas y biológicas que muestran una mayor asociación y luego sirven de base para la generación del modelo predictivo. Se presentan los casos para las variables biológicas que tiene sentido evaluar en un lapso de 30 días y las que pueden ser evaluadas en un lapso de hasta 140 días. En cada uno de estos casos, se muestran los resultados para todos los datos y para la temporada lluviosa y de sequía.

En este trabajo se seleccionaron solo dos lapsos de exploración de los datos uno que comprende de los 15 a 30 días previos a la evolución biológica de la enfermedad y otro de 15 a 140 días, dando como resultados los siguientes caso de estudios:

#### **4.2.1 Caso de estudio 1: El lapso de exploración 15 hasta 30.**

##### **4.2.1.1 Caso de exploración de los datos globales.**

La implementación de la visualización del comportamiento de la correlación entre las variables estudiadas, mediante el uso de una evaluación cualitativa, permitió vislumbrar cuales variables biológicas y climatológicas poseen mejor correlación haciendo posible una selección previa. La valoración consistió en analizar visualmente el comportamiento de la correlación para una amplitud fija y retraso variable para la ventana, calificándose la relación entre las variables estudiadas con un valor acorde con la magnitud relativa de la correlación según la Tabla N° 4. Este procedimiento se aplicaba para cada variable biológica con respecto a todas las variables climáticas. Por ejemplo, la Figura N° 11 muestra la correlación y la significancia estadística asociada entre la severidad y la humedad relativa para una amplitud de ventana de 2 días y un lapso de exploración de 30 días (el retraso máximo es 28 días). Se puede apreciar la existencia de valores de correlación que superan el 0.4, es decir, que según la Tabla N° 4 se clasificaría como *Regular* (2) la relación existente entre ambas variables, y que si existe al menos valor de correlación para un retraso que posee un valor de correlación entre 0.4 y 0.6, recibe esta clasificación. Así mismo se calificaron las demás relaciones.

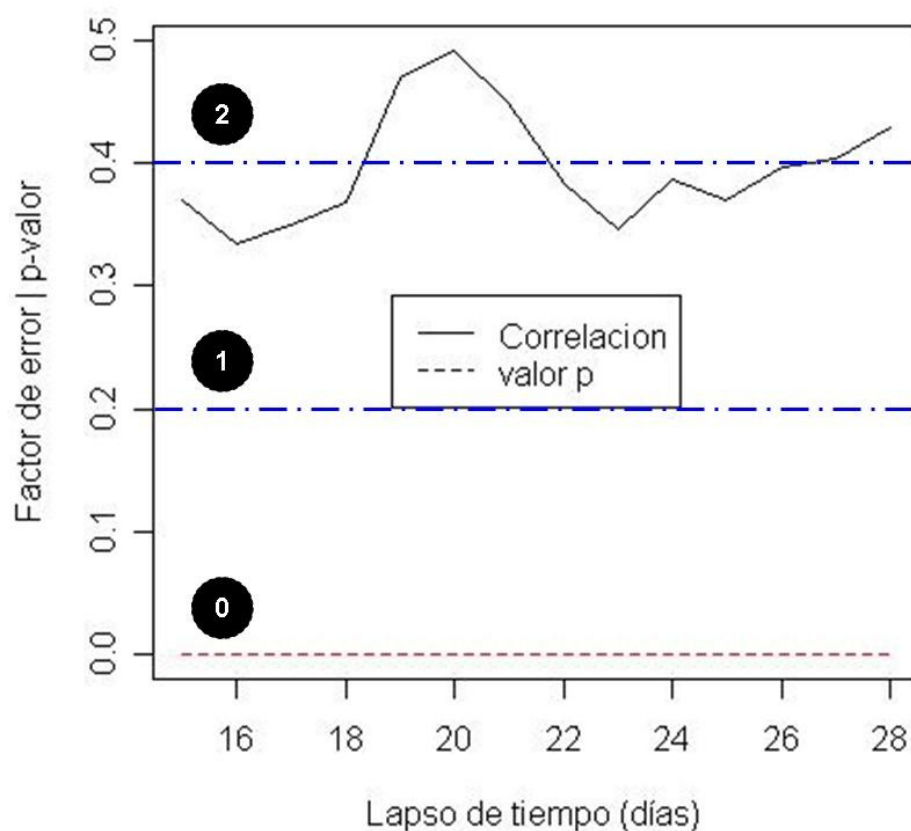


Figura N° 11. Grafico del comportamiento de la correlación y el valor p.

En el caso de la selección de las variables para un lapso que va de 14 a 30 días previos a la fecha de medición de las variables biológicas, donde se contemplaron los datos sin discriminación por periodo, los resultados de la evaluación anterior se reportan en la Tabla N° 8. En esta tabla se puede observar que las variables más preponderantes entre las biológicas son: la severidad (SE) de la enfermedad y la hoja más joven manchada (HMJM), mientras que en el caso de las climatológicas se encuentra la humedad relativa y energía solar. La Tabla N° 9 muestra en detalle las mejores asociaciones detectadas en la Tabla N° 8. Nótese que a pesar de no ser muy alta, las correlaciones son estadísticamente significantes

Tabla N° 8. Tabla de evaluación cualitativa de la correlación entre las variables Biológicas y climatológicas para un lapso de 30 días.

Variables Climatológicas	Variables Biológicas					
	SE	H MJM	SBH 3	EE	SBH 4	SB
Humedad Relativa	2		1	1	1	1
Energía Solar		2	1	1		
Precipitación Acumulada	1		1	1	1	1
Evapotranspiración		1	1	1		1
Radiación Solar		1	1	1	1	
Punto de rocío	1				1	1
Velocidad del viento	1		1			
Temperatura promedio			1			
Temperatura más alta			1			
Temperatura más baja			1			

Tabla N° 9. Tabla de las variables climatológicas seleccionada para el lapso de 30 días.

Variables		Ventanas (días)		Correlación	p-valor
Biológicas	Climatológicas	Amplitud	Retraso		
HMJM	Energía Solar	13	15	0.54401596	4.29E-14
SE	Humedad Relativa	2	20	0.4914427	2.18E-09

#### 4.2.1.2 Caso de exploración de los datos en la estación lluviosa.

Aplicando el mismo procedimiento anterior en el conjunto de datos correspondiente a la estación lluviosa se obtiene la evaluación cualitativa de la correlación de la Tabla N° 10. Se observa que las variables que se destacan como las mejores variables biológica a

predecir son: suma bruta en la hoja 4 (SBH 4), el estado de evolución de la enfermedad (EE) y repite la hoja más joven manchada (HMJM), mientras que en el caso de las variables predictivas se destacan la energía solar y la precipitación acumulada.

Tabla N° 10. Evaluación cualitativa de la correlación para el periodo de lluvia.

Variables Climatológicas	Variables Biológicas						
	SBH 3	SB	EE	HMJM	SBH 4	SE	SBH 2
Precipitación Acumulada	2	2	2	1	1	1	
Energía Solar	2	1	1	2	1		
Evapotranspiración	1	1	1	1	1	1	1
Radiación Solar	1	1	1	1	1	1	1
Humedad Relativa		1			1	1	

La Tabla N° 11, muestra las características relevantes asociadas a las variables que se destacan en la Tabla N° 10.

Tabla N° 11. Variables climatológicas seleccionadas para el periodo de lluvia.

Variables		Ventanas (días)		Correlación	p-valor
Biológicas	Climatológicas	Amplitud	Retraso		
HMJM	Energía Solar	10	15	0,4763	6,9819E-06
SBH 3	Precipitación Acumulada	3	27	0,4053	1,9186E-04
EE	Precipitación Acumulada	3	27	0,4045	1,9837E-04
SB	Precipitación Acumulada	3	27	0,4010	2,2758E-04



SBH 3	Energía Solar	4	17	-0,4258	1,2588E-04
-------	---------------	---	----	---------	------------

#### 4.2.1.3 Caso de exploración de los datos en la estación seca.

El mismo procedimiento anterior, produce la evaluación cualitativa resumida en la Tabla N° 12 y ampliada en Tabla N° 13.

Tabla N° 12. Evaluación cualitativa de la correlación para el periodo de sequía.

Variables Climatológicas	Variables Biológicas						
	S E	HMJM	EE	SBH 4	SBH 3	S B	SBH 2
Humedad Relativa	3	1	2	2	2	2	
Energía Solar	1	3					
Temperatura más baja	2	1	2	2	2	2	
Punto de rocío	2	1	2	2	2	2	

Tabla N° 13. Evaluación cualitativa de la correlación para el periodo de sequía (Continuación).

Variables Climatológicas	Variables Biológicas						
	S E	HMJM	EE	SBH 4	SBH 3	S B	SBH 2
Temperatura promedio	2		2	2	2	1	
Velocidad del viento	2		2	2	2	1	
Temperatura más alta	2		2	2	2	1	
Evapotranspiración	1	1	1	1		1	
Precipitación Acumulada	2		1	1	1	1	1
Radiación Solar		1				1	

Tabla N° 14. Variables climatológicas seleccionadas para el periodo de sequía.

Variables		Ventanas (días)		Correlación	p-valor
Biológicas	Climatológicas	Amplitud	Retraso		
EE	Humedad Relativa	5	16	0,4536	7,11E-05
	Punto de rocío	2	16	0,525	5,09E-06
	Temperatura promedio	2	19	0,4502	1,17E-04
	Velocidad del viento	2	19	0,45	1,18E-04
HMJM	Humedad Relativa	2	24	0,4052	1,80E-05
	Energía Solar	7	23	0,6043	2,33E-09
SB	Humedad Relativa	5	16	0,4852	1,80E-05
	Punto de rocío	2	16	0,5602	8,24E-07
	Temperatura más baja	2	19	0,4009	7,04E-04
SBH 3	Humedad Relativa	6	15	0,468	2,98E-05
	Humedad Relativa	5	23	0,471	2,61E-05
	Punto de rocío	2	16	0,5199	6,52E-06
	Temperatura promedio	2	19	0,4922	2,02E-05
	Velocidad del viento	4	16	0,4385	4,22E-05
	Velocidad del viento	2	19	0,4919	2,04E-05
SBH 4	Humedad Relativa	2	16	0,4716	5,62E-05
	Punto de rocío	2	16	0,5209	6,20E-06
	Temperatura promedio	2	19	0,4272	2,80E-04
	Velocidad del viento	2	19	0,4271	2,81E-04

Tabla N° 15. Variables climatológicas seleccionadas para el periodo de sequía (Continuación).

Variables		Ventanas (días)		Correlación	p-valor
Biológicas	Climatológicas	Amplitud	Retraso		
Se	Humedad Relativa	6	17	0,6594	2,2E-10
	Humedad Relativa	14	29	0,6184	7,67E-10
	Precipitación Acumulada	14	29	0,4182	8,34E-05
	Punto de rocío	2	21	0,5457	1,24E-06
	Punto de rocío	14	29	0,4858	4,29E-06
	Temperatura promedio	7	16	0,4620	1,42E-05
	Velocidad del viento	7	16	0,4581	1,71E-05

#### 4.2.2 Caso de estudio 2: El lapso de exploración 15 hasta 140 días.

##### 4.2.2.1 Caso de exploración de los datos globales.

En este caso sólo se considera como variable biológica la severidad (SE), que es la única que tiene sentido evaluar para este lapso de tiempo pues no depende del ciclo de vida de la planta. La Tabla N° 16 muestra la amplitud de la ventana y el retraso obtenidos para las mejores correlaciones encontradas.

Tabla N° 16. Variables climatológicas seleccionadas sin discriminación por temporada.

Variables		Ventanas (días)		Correlación	p-valor
Biológicas	Climatológicas	Amplitud	Retraso		
SE	Temperatura promedio	3	94	0,4607	2,08E-09
	Temperatura promedio	2	102	0,4821	1,24E-09
	Punto de rocío	5	91	0,4917	9,43E-11

Tabla N° 17. Variables climatológicas seleccionadas sin discriminación por temporada (continuación).

Variables		Ventanas (días)		Correlación	p-valor
Biológicas	Climatológicas	Amplitud	Retraso		
SE	Humedad Relativa	9	50	0,5240	1,36E-12
	Humedad Relativa	4	70	0,5249	3,88E-12
	Precipitación Acumulada	12	41	0,4609	2,52E-10
	Velocidad del viento	3	94	0,4598	2,24E-09
	Velocidad del viento	2	102	0,4819	1,26E-09

Como en el caso anterior, también se implementa la selección de las variables climáticas, aplicando la discriminación por temporada climatológica.

#### 4.2.2.2 Caso de exploración de los datos en la estación lluviosa.

Tabla N° 18. Variables climatológicas seleccionadas para la temporada de lluvia.

Biológicas	Variables	Ventanas (días)		Correlación	p-valor
	Climatológicas	Amplitud	Retraso		
SE	Temperatura promedio	2	122	0,4037	2,05E-04
	Punto de rocío	2	72	0,4130	2,55E-04
	Evapotranspiración	9	93	-0,5099	1,16E-06
	Humedad Relativa	3	71	0,5379	6,47E-07
	Precipitación Acumulada	11	43	0,4278	3,98E-05
	Precipitación Acumulada	13	57	0,4573	1,08E-05
	Velocidad del viento	2	122	0,4038	2,04E-04

#### 4.2.2.3 Caso de exploración de los datos en la estación seca.

Tabla N° 19. Variables climatológicas seleccionadas para la temporada de sequía.

Biológicas	Variables	Ventanas (días)		Correlación	p-valor
	Climatológicas	Amplitud	Retraso		
Se	Temperatura promedio	2	95	0,6145	7,35E-09
	Punto de rocío	2	99	0,6387	1,20E-09
	Evapotranspiración	3	78	0,4211	1,69E-04
	Evapotranspiración	8	85	0,4660	1,33E-05
	Evapotranspiración	2	99	0,4683	2,94E-05
	Humedad Relativa	6	17	0,6594	2,25E-10
	Humedad Relativa	14	29	0,6184	7,66E-10
	Precipitación Acumulada	13	41	0,5310	2,03E-07
	Precipitación Acumulada	3	57	0,4783	7,21E-06
	Energía Solar	14	138	0,5421	1,01E-07
	Radiación Solar	9	78	0,4614	1,131E-05
	Radiación Solar	14	134	0,4470	1,791E-05
	Velocidad del viento	2	95	0,6143	7,45E-09

#### 4.3 MODELO PREDICTIVO.

Obtenidas las variables en el paso anterior y luego de construir una nueva base de datos con dichas variables se procedió a generar el modelo de predicción para cada variable biológica, en los diferentes períodos considerados usando la herramienta informática WEKA.

Para cada uno de los casos anteriores (combinaciones de estación climática y lapso a evaluar), se aplicaron las diversas técnicas estadísticas predictivas, obteniéndose diferentes modelos para la predicción de las variables biológicas en función de las climáticas. Luego se seleccionó el mejor modelo utilizando como criterio, la maximización de la correlación entre los valores medidos, predicciones y el criterio de simplicidad del modelo para facilitar su implementación en el campo.

#### 4.3.1 Caso de estudio 1: La lapso de exploración 15 hasta 30 días.

##### 4.3.1.1 Modelo predictivo global.

Los modelos obtenidos para este caso de estudio, se presentan a continuación; La correlación y la técnica utilizada para obtener el modelo que se muestra en la Tabla N° 17:

- Hoja más joven manchada:

$$HMJM = 0.173 E.S_{(13;15)} + 6.921 \quad Ec.$$

10

- Severidad de la enfermedad:

$$SE = 0.3572 H_{(2;20)} - 8.801 \quad Ec.$$

11

Tabla N° 20. Tabla de la correlación y porcentaje de error de los modelos generados para datos globales, lapso de 30 días.

Variable Biológica	Correlación	Error Relativo	Algoritmos
HMJM	0.5243	1.2537	Mínimos cuadrados

SE	0.4041	4.0602	Mínimos cuadrados
----	--------	--------	----------------------

#### 4.3.1.2 Modelo predictivo para la temporada de lluvia

Los modelos obtenidos para este caso de estudio, se presentan a continuación así como la correlación y la técnica utilizada para generar el modelo en la Tabla N° 18:

- Hoja más joven manchada:

$$E.S_{(10;15)} \begin{cases} <= 10.467 \text{ HMJM} - 0.0588 E.S_{(10;15)} + 8.1043 \\ > 10.467 \text{ HMJM} - 0.0306 E.S_{(10;15)} + 10.278 \end{cases} \quad \text{Ec.}$$

12

- Suma Bruta en la Hoja 3

$$SBH3 = -7.8507 E.S_{(4;17)} + 7.4193 P.A_{(3;27)} + 188.999 \quad \text{Ec.}$$

13

- Estado de evolución de la enfermedad:

$$EE = 14.9826 P.A_{(3;27)} + 266.498 \quad \text{Ec.}$$

14

- Modelo de la suma bruta del total de hojas (SB)

$$SB = 17.3893 P.A_{(3;27)} + 372.681 \quad \text{Ec.}$$

15

Tabla N° 21. Tabla de la correlación y porcentaje de error de los modelos generados, caso temporada de lluvia, lapso de 30 días.

Variable	Correlaci	Media del Error	Algoritmos
----------	-----------	-----------------	------------



Biológica	ón	Absoluto	
HMJM	0.5101	1.3039	Arbol M5
SBH 3	0.4245	95.2038	Regresión Lineal
EE	0.2741	293.1627	Mínimos cuadrados
SB	0.2459	284.8454	Mínimos cuadrados

#### 4.3.1.3 Modelo para la temporada de sequía, lapso de 30 días.

Los modelos obtenidos para este caso de estudio en la temporada de sequía, se presentan a continuación así como la correlación y la técnica utilizada para hacerlo en la Tabla N° 19:

- Estado de evolución de la enfermedad:

$$EE = 39.49Pr_{(2;16)} - 389.9 \quad Ec.$$

16

- Hoja más joven manchada:

$$HMJM = 0.2308ES_{(7;23)} - 0.0883Hr_{(2;24)} + 13724 \quad Ec.$$

17

- Severidad de la enfermedad:

$$SE = 0.4487Hr_{(6;17)} - 16.1251[68/77.66] \quad Ec.$$

18

- Suma Bruta en la Hoja 3

$$SBH3 = 20.75Pr_{(2;16)} - 360.7 \quad Ec.$$

19

- Suma Bruta en la Hoja 4

$$SBH4 = 26.26Pr_{(2;16)} - 217.1 \quad Ec.$$

20

- Modelo de la suma bruta del total de hojas (SB)

$$SB = 44978Pr_{(2,16)} - 489949 [62/84778\%] \quad \text{Ec.}$$

21

Tabla N° 22. Tabla de la correlación y porcentaje de error de los modelos generados, caso de estudio temporada de sequía, lapso de 30 días.

Variable Biológica	Correlación	Media del Error Absoluto	Algoritmos
EE	0.3753	230.1964	Regresión Lineal
HMJM	0.705	1.1125	Regresión Lineal
SE	0.5956	1.608	Reglas M5
SBH 3	0.3506	102.1124	Regresión Lineal
SBH 4	0.4265	133.9107	Regresión Lineal
SB	0.4570	230.1964	Regresión Lineal

#### 4.3.2 Caso de estudio 2: La lapso de exploración 15 hasta 140.

Los modelos obtenidos para este caso de estudio, se presentan a continuación así como la correlación y la técnica utilizada para hacerlos en la Tabla N° 20.

##### 4.3.2.1 Modelo sin discriminación o general:

- Severidad de la enfermedad:

$$SE = 0.3906Pr_{(9,50)} + 0.3625P.A_{(12,41)} - 13.230 \quad \text{Ec.}$$

22

#### 4.3.2.2 Modelo de preaviso para la temporada de lluvia:

- Severidad de la enfermedad:

$$SE = 0.5825T_{(2;12)} - 75.573ET_{(9;9)} + 0.1867Hr_{(3;7)} + 0.2671P.A_{(11;4)} + 0.535P.A_{(13;5)} - 0.2757$$

Ec.

23

#### 4.3.2.3 Modelo de preaviso para la temporada de sequía:

- Severidad de la enfermedad:

$$SE = 0.4095Hr_{(14;2)} + 0.7443P.A_{(13;4)} + 0.0654R.S_{(9;7)} + 0.4655V_{(2;9)} - 41.1931$$

Ec.

24

Tabla N° 23. Tabla de la correlación y porcentaje de error de los modelos de predicción de la severidad (SE), lapso de 140 días.

Caso de estudio	Correlación	Media del Error Absoluto	Algoritmos
Global	0.5662	3.6635	Regresión Lineal
Lluvia	0.7254	3.2696	Regresión Lineal
Sequía	0.7944	2.5787	Regresión Lineal

## CAPITULO V - DISCUSIÓN DE RESULTADOS

Análisis preliminares de la correlación entre las variables biológicas y climatológicas El método evaluación preliminar de las relaciones individuales entre las variables biológicas y climáticas, permitió determinar cuáles de la variables climáticas predicen mejor a la variables biológicas, y también se pudo explorar la suma de eventos no relacionados entre sí que afectan a la enfermedad, esto es por ejemplo reflejado en la Tabla N° 24 donde existen más de una mismas variable climática, con diferente amplitud y retraso, que influyen sobre la misma variable biológica.

Tabla N° 24. Algunas de las variables climatológicas seleccionadas para predecir la severidad de la enfermedad (SE).

Variables		Ventana (días)		Correlación	p-valor
Biológicas	V. Climatológicas	Amplitud	Retraso		
SE	Humedad Relativa	6	17	0,66	2,25E-10
		14	29	0,62	7,67E-10
		9	50	0,52	1,36E-12
		3	71	0,54	6,47E-07
	Punto de rocío	2	21	0,55	1,24E-06
		14	29	0,49	4,29E-06
		2	72	0,41	2,55E-04
		5	91	0,49	9,43E-11
		2	99	0,64	1,20E-09
	Temperatura promedio	7	16	0,46	1,42E-05
		2	95	0,61	7,35E-09
		2	102	0,48	1,24E-09
		2	122	0,40	2,05E-04
	Velocidad del viento	7	16	0,46	1,71E-05
		2	95	0,48	1,26E-09
		2	102	0,40	2,04E-04
		2	122	0,61	7,45E-09

En la evaluación de la relación entre las variables biológicas y climáticas sin discriminación por temporada climática para un lapso de 30 días (Tabla N° 9) se obtuvieron que las variables más significativas para el desarrollo del modelo fueron la variable energía solar (promedio durante 13 días consecutivos) que está correlacionada (15 días de antelación) con la hoja más joven manchada y la humedad relativa (promedio durante 2 días consecutivos) que está asociada (20 días de antelación) la severidad de la enfermedad, pero a pesar que son las mejores correlaciones (0.54 y 0.49 respectivamente) del conjunto de total analizados para este caso, su valor de correlación es considerado bajo, sin embargo poseen valores de significancia relevantes (menores a  $1e-5$ ), lo cual permite ser considerados aceptables (ya que el sistema de producción del cultivo del plátano requiere tratar la enfermedad antes que cause daños severo), por lo cual, a pesar de ser considerada baja estadísticamente puede brindar información muy valiosa a nivel de campo.

En el caso de la temporada de sequía (Tabla N° 11), se observa la existencia de una correlación más pronunciada que el caso de estudio para la temporada lluvia (Tabla N° 13), destacando como las variables climáticas con mayor relación o efecto sobre el patógeno causante de la enfermedad Sigatoka Negra (*M. fijiensis*): la humedad relativa, temperatura, precipitación, el punto de rocío, lo cual se ajusta a lo reportado por Fouré [14], Gauhl [18], Jácome y Schuh 4.3.2.3, entre otros. Pero surgen otras nuevas variables como la radiación solar, variable que no había sido destacada por los autores anteriormente mencionados, jugando estas un papel muy importante en el proceso de infección del patógeno ([12], [22])

por lo cual es considerada como una buena variable para el desarrollo del modelo. Las variables biológicas que han servido como buenos indicadores de la enfermedad son: severidad de la enfermedad (SE), estado de evolución de la enfermedad (EE), la suma bruta en la hoja 3, 4 y total (SBH 3, SBH 4 y SB) y con menor número de variables climatológicas involucradas repite para ambos periodo de estudios la hoja más joven machada (HMJM). También se puede observar en el análisis que las variables que intervienen en los diferentes modelos varían por periodo o estación climática (sequía y de lluvia), es decir, que los parámetros climáticos implementados para predecir una variable biológicas son diferentes según la temporada, comportamiento similar detectó Hernández [22] en su investigación en la Zona de Yaracuy. Esto se debe a que la enfermedad presenta una dinámica estacional determinada por las variaciones de temperatura y precipitación a lo largo del año, según trabajo realizado por Mourichon y Zapater [33]. Este comportamiento puede ser observado comparando las tablas Nº 25 y 26, donde se muestran los estadísticos de las variables biológicas que permiten describir la enfermedad para las temporadas de sequía y lluvia respectivamente. Observándose que para la temporada de sequía la variables biológicas presentan mayor rango de variabilidad que la temporada de lluvia, lo cual, coincide con el mismo comportamiento de las variables climatológicas, partiendo de este hecho, podemos pensar que los efectos son más fáciles de detectar en la temporada de sequía que en la temporada lluvia, lo cual coinciden con los resultados de la correlaciones obtenidas por temporada.

Tabla Nº 25. Resumen estadístico de las variables biológicas para temporada de sequía

	HMJM	SE	EE	SB	SBH_2	SBH_3	SBH_4
Mínimo	6.4	11.3	0	0	0	0	0
Máximo	13	35.3	2120	2181.8	80	880	1240
Mediana	9.2	23.55	500.9	544.2	0	100	380
Promedio	9.393	23.11	568.5	586	0.8696	146.3	407.8
1stQu,	8.2	19.63	317.3	323.9	0	0	260
3rdQu,	10.4	27.05	725	760	0	200	520
NA's	0	0	0	0	0	0	0

Tabla N° 26. Resumen estadístico de las variables biológicas para temporada de lluvia

	HMJM	SE	EE	SB	SBH_2	SBH_3	SBH_4
Mínimo	6.2	9.6	0	0	0	0	0
Máximo	9.2	17.15	180	214.2	0	0	170
Mediana	10.4	21.5	360	478.7	0	0	320
Promedio	10.22	21.83	480.4	524.4	1.149	97.24	353.3
1stQu,	11.2	26.25	690	740.2	0	170	500
3rdQu,	14.6	39.6	1748.4	1760	100	560	1220
NA's	0	00	0	0	0	0	0

Este comportamiento puede verse mejor en las Figura N° 12 y Figura N° 13 donde se muestran el comportamiento de la correlación entre la temperatura y la severidad para una misma amplitud de venta y el igual lapso de tiempo para la exploración, mostrando que existe una mayor relación entre la temperatura en el periodo de sequía que en el de lluvia.

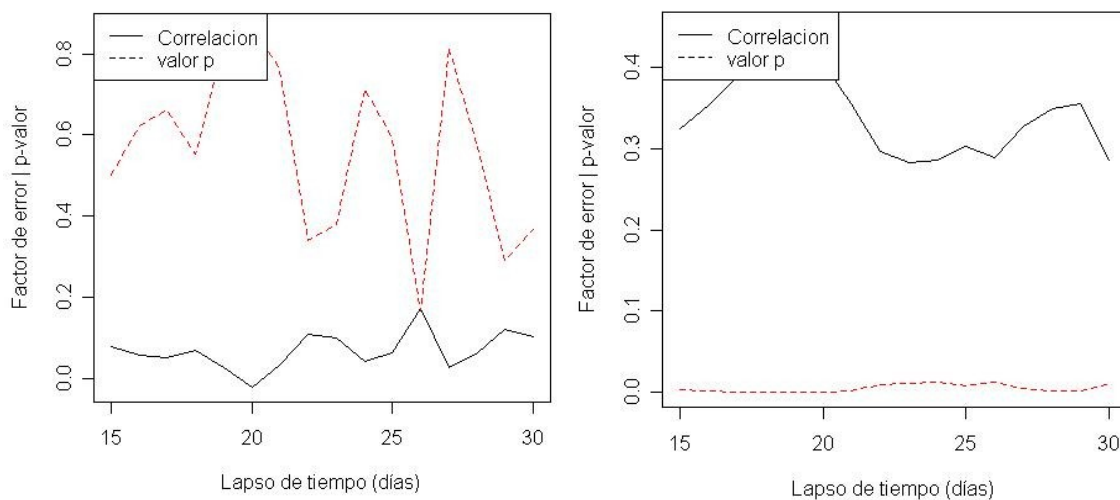




Figura N° 12. Temporada de  
lluvia

Comportamiento de la correlación entre la temperatura y la  
severidad

Figura N° 13. Temporada de  
sequía

Los métodos utilizados para la obtención de los modelos permitieron establecer ecuaciones de regresión que relacionaron, a diferentes niveles, el comportamiento de la enfermedad en función de las condiciones climáticas (Ecuación 10 a la 24). El método más adecuado para determinar el modelo que explique el comportamiento de la enfermedad, resultó ser aquel donde se obtuvo mejor correlación y que generó un modelo con una facilidad de implantación en el campo, en la mayoría de los casos, la técnica fue regresión lineal.

La ecuación que explicó mejor la severidad de la enfermedad Sigatoka Negra (SE) en la zona en estudio, fue aquella que consideró el promedio de las variables en la temporada seca. Destacando como las variables predictivas: la humedad relativa (Hr) durante 14 días consecutivos y con un retraso de 29 días, precipitación acumulada (P.A.) promedio por día durante 13 días y con un retraso 41 días, radiación solar (R.S.) durante 9 días y con un retraso de 78 días y la velocidad del viento (Vv) durante 2 días con un retraso de 95 días, con un coeficiente de determinación bajo, que podría considerarse aceptable ( $R^2 = 0,7922$ ) con alta significancia estadística para un  $\alpha \leq 0,05$  (Tabla N° 31) en cada uno de los coeficientes de la ecuación. Esto indica, que el 79,22% del comportamiento de la severidad de enfermedad para el periodo de sequía es explicado por la ecuación N° 25. Aunque se puede

afirmar para este caso que el 20,78 % restante, podría ser atribuido a otros factores no estudiados en este trabajo.

$$SE = 0.4095Hr_{(14;29)} + 0.7443P.A_{(13;41)} + 0.0654R.S_{(9;78)} + 0.4655V_{(2;93)} - 41.1931 \quad \text{Ec. 25}$$

La ecuación N° 25 es el modelo con mayor y mejor correlación de los modelos obtenidos. Sin embargo, existe otras dos ecuaciones que también pueden explicar la severidad de la enfermedad Sigatoka Negra (SE) en la zona en estudio, estos modelos son los obtenidos tanto temporada lluvia y como para él no discriminado o general. Destacando como las variables predictivas para el modelo (ecuación N° 26) de temporada de lluvia: la temperatura promedio (Tp) durante 2 días consecutivos y con un retraso de 122, días, Evapotranspiración (ET) durante 9 días consecutivos y con un retraso de 93 días, precipitación acumulada, la humedad relativa (Hr) durante 3 días consecutivos y con un retraso de 71 días, precipitación acumulada (P.A.) promedio por día durante 11 y 13 días y con un retraso 43 y 57 días respectivamente, con un coeficiente de 0,7254 y con alta significancia estadística para un  $\alpha \leq 0,05$  (Tabla N° 30) en cada uno de los coeficientes de la ecuación.

$$SE = 0.5825Tp_{(2;122)} - 75.573ET_{(9;93)} + 0.1867Hr_{(3;71)} + 0.2671P.A_{(11;43)} + 0.535P.A_{(13;57)} - 0.2757 \quad \text{Ec. 26}$$

Mientras que para el modelo (ecuación N° 27) general las variables que predicen el comportamiento de la severidad son: (Hr) durante 9 días consecutivos y con un retraso de 50 días,

precipitación acumulada (P.A.) promedio por día durante 12 días y con un retraso 41 días, teniendo una correlación 0,5662 y con alta significancia estadística para un  $\alpha \leq 0,05$  (Tabla N° 29) en cada uno de los coeficientes de la ecuación.

$$SE = 0.3906Hr_{(9;50)} + 0.3625P.A_{(12;41)} - 13.230 \quad \text{Ec. 27}$$

Otro modelo (ecuación N° 28) que obtuvo una correlación considerable fue el modelo que explicaba la hoja más joven manchada (HMJM), que al igual que los demás modelos consideró el promedio de los valores de las variables: humedad relativa durante 2 días consecutivos y con un retraso de 24 días y la radiación solar durante 7 días y con un retraso de 23 días, con un coeficiente de determinación bajo, que podría considerarse aceptable ( $R^2 = 0,705$ ) con alta significancia estadística para un  $\alpha \leq 0,05$  en cada uno de los coeficientes de la ecuación.

$$HMJM = 0.2308ES_{(7;23)} - 0.0883Hr_{(2;24)} + 13724 \quad \text{Ec. 28}$$

Ambos casos, son los modelos con mayor valor de correlación para la variable a predecir, y a su vez estos pertenecen al caso de estudio para la temporada de sequía, ratificando que la enfermedad presenta una dinámica estacional determinada por las variaciones de temperatura y precipitación a lo largo del año 4.3.2.3, haciendo más fácil percepción del comportamiento de la enfermedad en el periodo de sequía que en el de lluvia.

La asociación más alta, significativa y altamente significativa para un  $\alpha \leq 0,05$  (Tablas N° 26, 27 y 28), entre severidad de la

enfermedad y las variables climáticas en la Zona del Sur de Lago de Maracaibo, se encontró con humedad relativa promedio para una amplitud de 6 días y un retraso de 17 días de ventana ( $r^2=0,6594$ ), punto de rocío promedio durante 2 días con un retraso de 99 días ( $r^2=0,6385$ ), velocidad del viento promedio durante 2 días para un retraso de 95 días ( $r^2=0,6145$ ) y la temperatura promedio durante 2 días con un retraso de 95 días ( $r^2=0,6143$ ). Una investigación realizada en el Sur del Lago de Maracaibo, demostró correlación entre la severidad de ataque de la enfermedad y la precipitación, temperatura del aire y humedad relativa acumulados 25 días previos a la evaluación de Sigatoka Negra 4.3.2.3.

## **CAPITULO VI - CONCLUSIONES Y RECOMENDACIONES**

Este trabajo constituye uno de los pocos estudios en sistemas de pronóstico realizados en el país mediante el cual se determinó y cuantificó, la relación existente entre la incidencia y severidad de la Sigatoka Negra y las condiciones meteorológicas para el cultivo de plátano en la zona Sur del Lago. Se proponen modelos predictivos de pronóstico que pueden ser usados para realizar pronósticos de alertas tempranas del desarrollo de la enfermedad a partir de la información meteorológica y biológica recopilada por la Estación Local Chama – INIA.

El modelo obtenido que mejor predice la severidad de la enfermedad es el modelo de la estación seca o de estío (Ecuación No. 25), está explica en un 79,22 %, basándose en los atributos de radiación solar, humedad relativa, precipitación acumulada y velocidad del viento, para amplitudes y retrasos de las ventanas temporales diferentes para cada promedio de las variables climáticas. En el caso de la temporada lluviosa, el modelo que mejor predice la severidad, Ecuación No. 26, (72,54 % de la varianza explicada por el modelo) está en función de la variables temperatura, evapotranspiración, humedad relativa y precipitación acumulada; mientras que el modelo obtenido con todos los datos, Ecuación No 27. (56,65 %) fueron la variables humedad relativa y precipitación acumulada.

Durante el proceso de verificación y depuración de los datos se rechazaron varios registros, no tanto por ser erróneos, sino por poseer datos faltantes. Como se explicó, la disposición de estos

datos faltantes en segmentos continuos evitó que pudiesen ser estimados, usando cualquiera de las técnicas estadísticas para este fin. Esto disminuyó notablemente el número de observaciones consideradas por el modelo (1825 registros esperados para cada variable climática a 1005 registros). Debido a ésto, para desarrollar los modelos se usó validación cruzada de diez pliegos y no un conjunto independiente de los datos para la generación del modelo y otro para su validación. Por esta razón, los modelos propuestos en este trabajo deben ser validados con conjuntos independientes de datos, para varios años, para poder establecer su validez definitiva.

Los modelos generados son lo suficientemente sencillos para que puedan ser evaluados, directamente en campo por los organismos involucrados (INIA, ULA, SASA). Para verificar que, en efecto, los modelos desarrollados pueden generalizarse para toda la zona del Sur del Lago, habría que replicar este análisis en varias fincas de productores de la zona. La disponibilidad actual de pequeñas estaciones meteorológicas de bajo costo y el hecho de que la evaluación biológica esta explícitamente diseñada para hacer sencilla su evaluación por parte de los productores, aumenta la factibilidad de esta propuesta.

A pesar de que los datos fueron sometidos a un importante número de técnicas de minería de datos, algunas bastante sofisticadas, sin embargo, la técnica de regresión permitió obtener modelos de menor complejidad y al mismo tiempo la más alta correlación, que las demás técnicas usadas. La implementación del criterio de sencillez tiene el objetivo de facilitar la posterior implementación

del modelo seleccionado en el campo. Mayormente la mejora en la correlación viene acompañada de una mejora en el error relativo el cual subyace como criterio para la selección de los modelos.

Este trabajo encontró modelos y bondades del ajuste diferentes para los datos considerados globalmente, los datos correspondientes a la estación lluviosa y los de la estación seca, lo cual evidencian el comportamiento estacional de la enfermedad reportado en la literatura. Dada la variabilidad climática interanual de la zona es importante considerar la definición de los periodos de lluvia y sequia, como criterios para discriminar los registros de los datos por los diferentes periodos climáticos que se pueden presentar de forma irregular en el transcurso de las mediciones.

Los modelos obtenidos para el preaviso del brote de la Sigatoka Negra, coincide con subrayar la importancia de las variables: humedad relativa, precipitación acumulada, temperatura del aire, radiación solar y temperatura de rocío (variable relacionada con la mojadura de las hojas), que otro trabajos (4.3.2.3, 4.3.2.3, 4.3.2.3, 4.3.2.3, 4.3.2.3, 4.3.2.3, 4.3.2.3, 4.3.2.3, 4.3.2.3, 4.3.2.3, 4.3.2.3, 4.3.2.3) han destacado como las variables que poseen mayor influencia sobre la enfermedad. Mayormente una de las técnicas más usadas en estos trabajos es la de considerar la influencia de la variable climáticas acumuladas por semanas consecutiva, y se escoge aquella ventana temporal que posea mayor correlación tal como es reseñado por Hernández 4.3.2.3 en su trabajo “Estudio preliminar de vigilancia y pronóstico para sigatoka negra (*Mycosphaerella fijiensis* Morelet) en el cultivo de

plátano (*Musa AAB cv Hartón*) en Macagua-Jurimiquire, estado Yaracuy”, el cual es el único trabajo de que se tiene conocimiento que haya abordado un modelo de pronóstico en el país. A diferencia del trabajo de Hernández este trabajo considera la variabilidad de la amplitud de las ventanas temporales (basado en el trabajo de Coalkley 4.3.2.3) al igual que el retraso de la misma, también, permitió estudiar las intercalaciones entre las variables climáticas como ventanas desfasadas en el lapso tiempo y las variables biológicas, es decir, se determina los efectos que pueden tener la suma de eventos temporales de las variables climáticas sobre las variables biológicas. La correlación obtenida por la metodología aplicada por Hernández para el promedio ponderado de infección es 0,333, en nuestro caso la severidad de enfermedad la correlación fue de 0,5662 para el modelo global. Lo cual mejora la predicción de la severidad. Esta mejora en la correlación puede deberse a que se hace más hincapié en los efectos que ejercen las variables climáticas en diferentes periodos de tiempo con amplitudes temporales más precisas, lo cual normalmente no se consideran importantes.

Basado en la discusión anterior, se recomienda:

1. Realizar esfuerzos por aumentar la completitud del registro meteorológico en la estación del INIA-Chama, de manera de contar con más registros para continuar con la validación y desarrollo de modelos de pronóstico
2. Establecer un proyecto que permita verificar la validez de los modelos propuestos temporal y espacialmente. Para esto hay que aplicar y validar el modelo para varios años y para varias fincas en el sector del Sur del Lago.



3. Cambiar la definición de las temporadas de sequía y lluvia para reflejar mejor la variabilidad climática interanual.

## CAPITULO VII - REFERENCIAS

- [1]Bureau, E., Marín, D. y Guzmán, J. 1992. El sistema de preaviso para el combate de la Sigatoka Negra en banano y plátano. Unión de Países Exportadores de Banano (UPEB). Panamá. 40 pp.
- [2]Chuang, T.Y., Jeger, M. 1987. Predicting the rate of development of black Sigatoka (*Mycosphaera fijiensis* var. *difformis*) disease southern Taiwan. *Phytopathology* 77: 1542-1547.
- [3]Cancelado, R.E., Espinoza M., A.S. Intriago, M. 1991. "Manejo de Sigatoka amarilla en Ecuador con flusilazol, benomilo y aceite agrícola". ACORBAT 89, Mérida (VEN), 1989/09/24-29.
- [4]Coakley, S. M., McDaniel, L. R., and Shaner, G. 1985. Model for predicting severity of *Septoria tritici blotch* on winter wheat. *Phytopathology*. Vol 72 75: 1245-1251.
- [5]Coakley, S. M., Boyd W. A., and Line R. F. 1982. Statistical Models for Predicicting Stripe Rust on Winter Wheat in the Pacific Northwes. *Phytopathology*. Vol 72 N°12 .1539-1542
- [6]Consejo Zuliano de Planificación. 1994. Plan integral control de la Sigatoka Negra en las zonas plataneras de Maracaibo. Maracaibo, Venezuela. Mimeografiado.
- [7]Cornejo, J. M. 1988. Técnicas de investigación Social: El Análisis de Correspondencia. PPU. Barcelona.
- [8]Crouch, J. H., V Vuylsteke D., Ortiz R. 1998. Perspectives on the application of biotechnology to assist the genetic enhancement of plantain and banana (*Musa* spp.), EJB Electronic Journal of Biotechnology. Vol.1 No.1.
- [9]Delgado, M., García, A., Nava, J., Carroz, G., Barbosa, F., 1998. Diagnóstico técnico-socioeconómico de la zona platanera

- subregión Sur del Lago de Maracaibo. Informe Técnico CORPOZULIA-FUSAGRI. Sp
- [10]Díaz, F. 1995. Manejo integrado de la Sigatoka Negra. MAC. SASA. Comité Regional contra la Sigatoka Negra. zona sur del Lago de Maracaibo. Venezuela. Mimeografiado.
- [11]Douglas M & Ronald. R, 1992. El combate de la Sigatoka Negra. Boletín No 4. Departamento de investigaciones. CORBANA. 22p
- [12]Etebu E., Pasberg-Gauhl C., Gauhl F. y Ayibo Daniel-Kalio L. 2005. Efecto de la luz y aeración sobre la esporulación y crecimiento lineal de *Mycosphaerella fijiensis*. InfoMusa - Vol. 14 - No.1. pag 24-25
- [13]Fondo Nacional de Investigaciones Agropecuarias (FONAIAP), 1992. Aprendiendo a convivir con la Sigatoka Negra. Maracay. 6 p.
- [14]Fouré. E, 1994. Leaf spot disease of banana and plantain cause by *Mycosphaerella fijiensis* and *Mycosphaerella musicola*. 37-46p in the improvement and testing of musa: A global partnership. Proceedings of the first global conference of the international musa testing. Progranme (D. Jones, ed) INIBAP, Montpellier, France.
- [15]Ganry, J., Y Meyer, J. P. 1972a. La lutte controlée contre le Cercospora aux antilles. Bases climatiques de l' avertissement. Fruits 27: 665 - 676.
- [16]García, A., Sosa, L. 2001. Caracterización agronómica del híbrido de plátano FHIA-21 (Musa AAAB) resistente a Sigatoka Negra (*Mycosphaerella fijiensis* M.), Rev. fac. Agron. (LUZ). 18: 117-123.

- [17]Gauhl F. 1990. Epidemiología y Ecología de la Sigatoka Negra (*Mycosphaerella Fijiensis Morelet* en Plátano (*Musa* sp.), en Costa Rica). Unión de Países Exportadores de Banano (UPEB).
- [18]Gauhl, F.; Pasberg-Gauhl, C. 1994. Epidemiology of black Sigatoka disease on plantain in Nigeria. *Phytopathology* (USA), 84, (10), 1080.
- [19]Gómez, C., S. Sáez, V. y M. Puche. 1999. Análisis de temperatura y humedad relativa en un cultivo de plátano y su relación con el desarrollo de la sigatoka negra *Mycosphaerella fijiensis* en el estado Zulia durante 1996-1998. Resúmenes XVI Congreso Venezolano de Fitopatología, Barquisimeto 14 al 18, Noviembre de 1999. *Fitopatología Venezolana* (SVF). 12(2):57.
- [20]Gómez, C., Bracho, T., Andressen R. 2001. "SIBiLA: Sistema de información Bioclimática"; disponible en [www.saber.ula.ve/postgrados](http://www.saber.ula.ve/postgrados); Internet; consultado Junio de 2005.
- [21]García J. 2004. 30 razones contra los cultivos transgénicos en Costa Rica, *AMBIENTICO*, ISS1409-214X, N° 132, 9 -10.
- [22]Hernández G. L, Hidalgo W., Linares B, Hernández J., Romero N. y Fernández S. 2005. Estudio preliminar de vigilancia y pronóstico para sigatoka negra (*Mycosphaerella fijiensis Morelet*) en el cultivo de plátano (*Musa AAB cv Hartón*) en Macagua-Jurimiquire, estado Yaracuy. *Rev. Fac. Agron. (LUZ)*. 2005, 22: 324-337
- [23]Hernández O., José y Ramírez Q., M<sup>a</sup> José. 2004. Introducción a la minería de datos. Edit. Prentice Hall. Person Educación S.A. Madrid.
- [24]Ibanez F., Conversi A. 2002. Prediction of missing values and detection of exceptional events in a chronological planktonic

- series: a single algorithm. Ecological Modelling. vol 154, pag 9-23.
- [25]Jacome, L.H., Schuh, W. y Steverson, R.E. Effect of temperature and relative humidity on germination and germ tube development of *Mycosphaerella fijiensis* var. *difformis*. Phytopathology 12:1480-1485. 1991.
- [26]Jiménez, F., Tapia, A., Gribuis, N., Escalant, J. 1995. Relación entre la duración de la lluvia y el desarrollo de la Sigatoka Negra en plátano. Centro Agronómico Tropical de Investigación y Enseñanza CATIE, Turrialba Costa Rica.
- [27]Klein, H. H. 1960. Control of Cercospora leaf spot of bananas with applications of oil spray based on the disease cycle. Phytopathology 50: 488-490.
- [28]Marín, D., Romero, R. A. 1992. El combate de la Sigatoka Negra. Boletín N° 4 Departamento de Investigaciones. CORBANA. Costa Rica.
- [29]Martínez, G., Pargas, Rafael, D., Muño, D. "La Sigatoka Negra y su avance en el territorio venezolano: implicaciones socioeconómicas". CENIAP; disponible en <http://www.fonaiap.gov.ve/publica/divulga/fd59/sigatok.html>. Internet, consultado en Mayo de 2005.
- [30]Marulanda F., Mabel C y Cordona A., Omar D., 2006. Análisis del impacto de desastres menores y moderados a nivel local en Colombia. Informe final de proyecto. <http://www.desinventar.org/sp/proyectos/articulos/index.html>
- [31]Méndez A. 2001. Estimación robusta: una aplicación informática con fines didácticos. Estadística Española. Vol. 43, Núm. 147, págs. 105 a 123

- [32]Meredith, D.S. and Lawrence, J.S. 1969. Black leaf streak Disease of bananas (*Mycosphaerella fijiensis*): symptoms of disease in Hawaii and notes on the conidial state of the causal fungus. Trans. Br. Mycol. Soc. 52: 459-476.
- [33]Mourichon, X.; Carlier, J.; Fouré, E. 1997. Enfermedades de Sigatoka. INIBAP, Montpellier (FRA), 4 p.
- [34]Mourichon, X.; Zapater, M.F.1990 .Obtention in vitro du stade *Mycosphaerella fijiensis*, forme parfaite de *Cercospora fijiensis*. Fruits (FRA), 45, (6), p. 553-557.
- [35]Nava, C., Villarreal, E. 2000. Aplicación de nitrógeno, potasio, boro, magnesio y zinc a plantaciones de plátano, Musa AAB cv. Hartón en presencia de la Sigatoka Negra. Rev. Fac. Agron. (LUZ). 17: 20-35
- [36]Núñez, R. 1989. El Cultivo del Banano. Ministerio de Agricultura y Ganadería. Programa Nacional del Banano.
- [37]Ordosgoitti F., A. 1999. Enfermedades del banano y el plátano en Venezuela. Medidas de control en Venezuela. Maracay, Fondo Nacional de Investigaciones Agropecuarias. 71 p. (Serie B, No. 37).
- [38]Orjeda G., 1998, "Evaluation of Musa germplasm for resistance to Sigatoka diseases and Fusarium wilt", Technical Guidelines 3. Intenational Planta Genetic Resources Intitute.
- [39]Orozco-Santos, M. 1998. Manejo integrado de la Sigatoka Negra del plátano. SAGAR, INIFAP, CIPAC. Campo Experimental Tecomán. Tecomán, Colima, México. Folleto técnico No. 1. 95 p.
- [40]Pérez V., L. F.; Mauri M., F.; Hernández M., A.; Abreu A., E.; Porras G., A. 2004. Epidemiología de la Sigatoka Negra (*Mycosphaerella fijiensis* Morelet) en Cuba. I. Pronóstico bioclimático de los tratamientos de fungicidas en bananos

- (*Musa acuminata* AAA) .Revista Mexicana de Fitopatología (MEX), Vol. 1, p. 15-26.
- [41]Pérez V., L. F.; Mauri M., F.; Hernández M., A.; Abreu A., E.; Porras G., A. 2006, Relaciones entre los factores climáticos y la velocidad de evolución de la Sigatoka Negra en bananos y plátanos .Memorias de XVII Reunión Internacional de ARCORBAT, Vol. 2. p. 702-709.
- [42]Petravalle S., Shaw M. W., Paker S.R., and van de Bosck F., 2003, "Modeling of Relationships Between Veather and *Septoria tritici* Epidemics on Winter Wheat: A Critical Approach", Epidemiology, Vol. 93 N° 10, 1329-1339.
- [43]Porras, A.; Perez, L. 1997. Efecto de la temperatura sobre el crecimiento de los tubos germinativos de las ascósporas de *Mycosphaerella spp.* causantes de Sigatoka en banano .Infomusa (FRA), 6, (2), p. 27-32.
- [44]Romero, R. A. Y Sutton, T. B. 1995. Dynamics of fungicide-resistant populations of *Mycosphaere fljiensis* and epidemiology of black Sigatoka of bananas. PhD thesis. Department of Plant Pathology, NCSU, Raleigh. 113 p.
- [45]Sáez, V. y Rivas, C. 1992. Zonificación agroclimática de cultivos en la región Sur del Lago de Maracaibo (Estado Zulia). MARNR. 108 pp.
- [46]Stover R., H. 1984. Las manchas producidas por la sigatoka en hojas de plátanos y bananos. Curso internacional de reconocimiento, diagnóstico y control de Sigatoka Negra del plátano y banano. Mayo 14-18p. Tulenapa-Colombia 15p.
- [47]Stover, R., H. 1990. Sigatoka leaf spot: Thirty years of changing control strategies. Pages 66-74 in: Sigatoka Leaf Spot Diseases of Bananas. R. A: Fullerton and R. H. Stover, eds.

Proceedings of an international workshop held at San Jose, Costa Rica. INIBAP. France.

- [48]Stover, R. H., Simmonds, N. W. 1987. Bananas. Tropical Agriculture Series. Longman Group UK 468 pp.
- [49]W. Klösgen, J.M. Zytkow. Handbook of Minería de datos and Knowledge Discovery. Oxford University Press. 2002.
- [50]Zabala, M. E Bermudez, A. 1999. El costo del control de la Sigatoka Negra (*Mycosphaerella fijiensis*, Morelet) y su efecto en el beneficio económico del productor de plátano (*Musa AAB* cv. Hartón) del sur del lago de Maracaibo. Revista de la Facultad de Agronomía-LUZ, vol.16, no.1, p.107-119. ISSN 0378-7818.



## CAPITULO VIII - APÉNDICE A

Tabla N° 27. Resúmenes estadísticos de variables climatológicas por temporada de Sequía

	Temperatura			ET	Radiación Solar	Energía Solar	Velocidad del viento	Precipitación Acumulada	Humedad	Punto de Rocío
	Promedio	Alta	Baja							
Min.	17,37	17,63	17,15	5,80E-03	0	0	17,37	0	69,38	15,88
Max.	42,33	43,1	41,53	5,17E-01	506,1	44,612	42,33	104,4	100	37,2
Mediana	26,15	26,52	25,73	1,73E-01	204,6	12,877	26,14	0	83,42	22,43
Mean	27,65	28,05	27,24	1,66E-01	200,0	13,846	27,65	4,303	84,25	23,22
1stQu.	25,4	25,75	25,04	1,41E-01	170,1	8,885	25,4	0	79,88	21,71
3rdQu.	27,2	27,66	26,81	1,95E-01	232,6	18,902	27,19	2,4	87,39	23,14
NA's	127	127	127	207	127	127	127	127	207	207

Tabla N° 28. Resúmenes estadísticos de variables climatológicas por temporada de lluvia

	Temperatura			ET	Radiación Solar	Energía Solar	Velocidad del viento	Precipitación Acumulada	Humedad	Punto de Rocío
	Promedio	Alta	Baja							
Min.	16,43	18,11	16,13	2,10E-03	0	0	16,43	0	62,64	13,12
Max.	40,5	41,1	39,8	7,00E+	875,7	58,16	40,5	62,6	100	37,79

Media n	26,87	27,2 9	26,4 6	1,89E- 01	236,4	17,41	26,86	0	82,83	22,8
Mean	28,7	29,2	28,2 3	1,43E+ 03	240,7	16,86	28,69	2,966	84,86	24,29
1stQu.	25,96	26,3 2	25,6 1	1,61E- 01	202,5	10,48	25,94	0	79,21	22,3
3rdQu.	30,66	31,3 1	30,1 3	2,18E- 01	267,8	22,16	30,66	1,6	88,14	23,54
NA's	156	156	156	236	152	152	156	151	238	238

## APÉNDICE B

Tabla N° 29. Análisis multivariantes de relación de la severidad y las variables climáticas para un lapso de exploración de 140 días.

	Estimado	Error Estándar	t- valor	Pr(> t )	código
Intercepción	-13,2308	5,5787	-2,3720	1,9662E-02	*
P.A <sub>(12;41)</sub>	0,3625	0,1004	3,6120	4,8100E-04	***
Hr <sub>(9;50)</sub>	0,3906	0,0672	5,8100	7,7900E-08	***

Tabla N° 30. Análisis multivariantes de relación de la severidad y las variables climáticas para un lapso de exploración de 140 días, periodo de lluvia.

	Estimado	Error Estándar	t- valor	Pr(> t )	código
Intercepción	-0,2757	9,2205	-0,0300	9,7626E-01	
Tp <sub>(2;122)</sub>	0,5825	0,2713	2,1470	3,6185E-02	*
ET <sub>(9;93)</sub>	-75,5733	19,7231	-3,8320	3,2900E-04	***
Hr <sub>(3;71)</sub>	0,18668	0,07335	2,545	0,013764	*
P.A <sub>(13;57)</sub>	0,53505	0,11038	4,847	1,06E-05	***
P.A <sub>(11;43)</sub>	0,26708	0,12837	2,081	0,042144	*

Tabla N° 31. Análisis multivariantes de relación de la severidad y las variables climáticas para un lapso de exploración de 140 días, periodo de sequía.

	Estimado	Error Estándar	t- valor	Pr(> t )	código
Intercepción	-41,1931	10,6889	-3,8540	4,3500E-04	***
Hr <sub>(14;29)</sub>	0,4095	0,0960	4,2630	1,2800E-04	***
P.A <sub>(13;41)</sub>	0,7443	0,1809	4,1140	2,0100E-04	***
R.S <sub>(9;78)</sub>	0,06538	0,01464	4,467	6,91E-05	***

$V_{V(2;95)}$	0,46545	0,24906	1,869	6,94E-02	.
---------------	---------	---------	-------	----------	---

Significado de los códigos:

Muy altamente significativo	0 - 0.001	' ***
	0.001 -	'
Altamente significativo	0.001	**'
Significativo	0.001 - 0.05	'*'
Poco Significativo	0.05 - 0.1	'.'
No significativo	0.1 - 1	' '