



Centro de Investigación en Matemáticas, A. C.

---

---

# **Modelación de poblaciones de cerambícidos en cultivos de agave**

TESIS

que para obtener el grado de

Maestro en Ciencias con Especialidad  
en Probabilidad y Estadística

presenta

**Lidia Ery Toledo Rodríguez**

---

Dr. Miguel Nakamura Savoy  
*Director de Tesis*

Dr. Fernando Ávila Murillo  
*Co – director de Tesis*

*Guanajuato, Gto.*

*Mayo de 2008*

# Índice general

---

<b>Agradecimientos</b>	<b>III</b>
<b>Resumen</b>	<b>IV</b>
<b>1. Marco contextual</b>	<b>1</b>
1.1. Plagas en el agave . . . . .	1
1.2. Cerambícido, <i>Acanthoderes funeraria</i> Bates . . . . .	2
1.2.1. Descripción . . . . .	2
1.2.2. Daños al agave . . . . .	2
1.2.3. Biología y hábitos de ataque . . . . .	3
1.3. Problemas de interés para la empresa . . . . .	4
<b>2. Análisis exploratorio</b>	<b>5</b>
2.1. Datos del primer muestreo, 2006 . . . . .	5
2.1.1. Descripción de los datos . . . . .	5
2.1.2. Depuración de datos . . . . .	6
2.1.3. Análisis exploratorio inicial . . . . .	6
2.1.3.1. Mapa de intensidades de grises . . . . .	6
2.1.3.2. Mapa de intensidades de grises con suavizamiento . . . . .	8
2.1.3.3. Gráfica de datos meteorológicos y variable conteo . . . . .	8
2.1.3.4. Análisis de momentos . . . . .	11
2.1.4. Diagnóstico . . . . .	14
2.1.4.1. Conclusiones preliminares . . . . .	14
2.1.4.2. Sugerencias . . . . .	15
2.2. Datos del segundo muestreo, 2007 . . . . .	15
2.2.1. Descripción de los datos . . . . .	15
2.2.2. Análisis exploratorio secundario . . . . .	16
2.2.2.1. Mapa de intensidades de grises . . . . .	16
2.2.2.2. Gráficas de la función de autocorrelación . . . . .	16
2.2.2.3. Gráficas de datos meteorológicos y variable conteo . . . . .	20

## ÍNDICE GENERAL

---

2.2.2.4.	Análisis de momentos y método gráfico de bondad de ajuste . . . . .	20
2.2.3.	Conclusiones . . . . .	23
<b>3.</b>	<b>Propuesta del modelo probabilístico</b>	<b>25</b>
3.1.	Modelo autorregresivo de orden 1 BN general . . . . .	25
3.1.1.	Interpretación de las variables aleatorias del Modelo AR(1) BN general. . . . .	27
3.1.1.1.	Variable aleatoria $N_n$ . . . . .	27
3.1.1.2.	Variable aleatoria $R_i$ . . . . .	27
3.1.1.3.	Variable aleatoria $W_{n,i}$ . . . . .	27
3.1.2.	Interpretación de los parámetros del Modelo AR(1) BN general	29
3.1.2.1.	Parámetro $\alpha_n$ . . . . .	29
3.1.2.2.	Parámetro $\theta_n$ . . . . .	29
3.1.2.3.	Parámetro $\mu_n$ . . . . .	30
3.1.3.	Caso particular del Modelo AR(1) BN general . . . . .	30
3.2.	Estimación de los parámetros del Modelo BN general vía la función generadora de probabilidades . . . . .	32
3.2.1.	Parámetros estimados y variables meteorológicas . . . . .	34
3.3.	Diagnóstico del Modelo AR(1) BN general . . . . .	38
<b>4.</b>	<b>Resumen de resultados</b>	<b>41</b>
4.1.	Conclusiones . . . . .	41
<b>Apéndices</b>		
<b>A.</b>	<b>Muestreo de muestreo</b>	<b>44</b>
<b>B.</b>	<b>Tablas</b>	<b>49</b>

# Agradecimientos

---

A toda mi familia. Especialmente a mi madre Reyna Rodríguez Coroy por ser el motor y la guía de mi vida, y a mis hermanos Liliana, Lorena y Mariano por el apoyo incondicional que siempre me han brindado.

A todo el pueblo mexicano que a través de CONACyT solventó mis estudios de maestría, con la beca no. 201912. De igual manera agradezco a CONCyTEG por apoyarme con una beca para seguir con este trabajo.

A todos los investigadores que a través de CIMAT, A. C. me apoyaron con una beca para terminar este proyecto de tesis.

A la Dra. Eloísa Díaz-Francés Murguía por todo el apoyo y consejos.

Al Dr. Miguel Nakamura Savoy por ser mi asesor y guiarme en todo este proceso.

Al Dr. Fernando Ávila Murillo por ser mi co-asesor y el vínculo para poder realizar esta tesis.

Al M. en C. Ramón Rubio Cortés y al Dr. José Ignacio del Real Laborde por proporcionar el problema de esta tesis y por ayudarme a entender los hábitos y biología de los insectos.

Al Dr. José Miguel Ponciano por las valiosas discusiones en torno a la génesis de modelos para conteos e interpretaciones biológicas.

Al Dr. Enrique Raúl Villa Diharce por aceptar ser mi sinodal.

A todos mis compañeros y amigos.

# Resumen

---

La materia prima para elaborar Tequila, piña del agave *Tequilana Weber* en su variedad azul, el único autorizado por la Norma Mexicana, tarda de 7 a 10 años en madurar. Durante toda su vida, el agave está expuesto a varias plagas y enfermedades, y muchas de ellas representan pérdidas importantes para los productores. De lo anterior, se deriva la importancia de estudiar a las plagas y enfermedades que atacan al agave. Este proyecto fue generado por la demanda de un cliente estratégico de CIMAT—Tequila Sauza—de donde surgió la motivación para esta tesis: entender mediante ciencia básica un fenómeno aleatorio de la naturaleza. El problema que se abordará es la descripción espacial y temporal de la ocupación del insecto *Acanthoderes funeraria* Bates en cultivos de agave, la plaga conocida comúnmente como cerambícido.

Los datos proporcionados consisten de conteos de cerambícidos sobre plantas en fechas equiespaciadas, en muestreos de 2006 y 2007. El principal objetivo de esta tesis es la obtención de un modelo probabilístico que describa el comportamiento de la plaga—espacial y temporalmente—así como investigar la relación que pueda tener con variables meteorológicas. Este entendimiento coadyuvará al control eficiente de cerambícidos por parte de la empresa.

En primer término, se implementó análisis exploratorio consistente de técnicas y gráficas especialmente propuestas para la situación. Con ellas se logró resumir los datos, buscar indicios de agregación espacio/temporal, averiguar sobre la distribución del número de insectos en el predio y a explorar correlaciones entre fechas rezagadas. Se concluyó que los insectos no exhiben un efecto de agregación espacial sobre el predio, que las variables climatológicas humedad, temperatura y precipitación pluvial sí tienen efecto en el número de insectos presentes en el predio y que el número de insectos en un día determinado sólo tiene fuerte correlación con el número correspondiente al día anterior.

El resultado más importante de la tesis es la construcción de un modelo probabilístico para describir el poblamiento de plantas por cerambícidos. Éste consiste de un modelo autorregresivo de orden uno, basado en variables de conteo. Se trata de un proceso aleatorio secuencial dotado de una interpretación biológica con gran afinidad al contexto. La formulación da pie a una distribución más general que la bi-

nomial negativa en marginales, la cual es una de las primeras opciones para describir poblaciones de insectos. Se desarrolla en la tesis un procedimiento de estimación de parámetros mediante la función generadora de probabilidades empírica, y se describe cómo podrían incorporarse variables meteorológicas de manera explícita cuando las hubiera con la calidad y cantidad suficientes. Se incluyeron algunos estudios de simulación para ilustrar algunas características esenciales del modelo probabilístico y del procedimiento de estimación.

Un producto estadístico desarrollado al margen de todo lo anterior lo fue un diseño muestral, con el cual se puede estimar de manera más económica la proporción de insectos hembras (u otra cualidad binaria de interés) en el predio. Lo anterior tendría implicaciones prácticas relacionadas con fumigaciones o controles biológicos.

---

# Capítulo 1

## Marco contextual

---

### 1.1. Plagas en el agave

El tipo de planta que se utiliza en la elaboración del tequila, agave *tequilana* Weber variedad azul, se cultiva en los municipios que conforman la DOT<sup>1</sup>. Como otras plantas, el agave está expuesto a enfermedades y plagas. Llamamos **plaga** a los organismos que se alimentan del cultivo, le causan daños y pueden llegar a destruirlo totalmente; las de mayor impacto, detectadas por la compañía Sauza son las comúnmente llamadas: picudo, cerambícido, escarabajo rinoceronte, barrenador de pencas y gusano blanco (Figura 1.1).



Figura 1.1: Plagas importantes en el agave: barrenador de pencas, escarabajo rinoceronte, picudo, gusano blanco y cerambícido, respectivamente.

El ciclo de cultivo del agave es largo, de 6 a 10 años, que es cuando se puede cosechar y enviar a fábrica para su transformación en tequila. Una vez que el agave

---

<sup>1</sup>Denominación de Origen del Tequila

madura se cosecha mediante lo que se conoce como la jima<sup>2</sup>. Debido a que el tiempo necesario para convertir el agave en tequila es muy grande, las plagas constituyen un factor de riesgo muy importante, dado que cualquier daño causado por éstas representa una pérdida cuantiosa para el productor. La plaga que nos ocupa en este trabajo es la de los cerambícidos, descritos en la siguiente sección.

## 1.2. Cerambícido, *Acanthoderes funeraria* Bates

### 1.2.1. Descripción

El cerambícido que ataca al agave tiene el nombre científico de *Acanthoderes funeraria* Bates y entre los primeros investigadores que lo reportó atacando plantaciones de agave fue el Dr. Héctor González (entomólogo graduado del Colegio de Posgraduados de Montecillo, Estado de México). Hay presencia de cerambícidos en la región oeste del Estado de México, Querétaro, Guanajuato, Oaxaca, San Luis Potosí y en las zonas agaveras de Jalisco, como: Arenal, Zapopán, Ameca, Tequila, Atotonilco, Tototlán, Tepatitlán y Amatitán. De acuerdo con Chemsak, Linsley y Noguera (1992) esta especie sólo se ha encontrado en México y aún no hay suficiente información acerca de sus hábitos y otras características; al parecer la primera información se está generando en Tequila Sauza.

El cerambícido adulto mide en promedio 2 cm de longitud y tiene un cuerpo color negro con manchas blancas (Figura 1.2b). Su larva (Figura 1.2a) es de color blanco y pasa por varios estadios, aunque aún no se ha determinado cuántos.

Esta especie puede volar y aunque no se ha evaluado su movilidad, al parecer (por la abundancia en predios) es muy reducida y generalmente se restringe al mismo predio o a predios muy cercanos, con algunos insectos que pueden volar largas distancias. Típicamente, los adultos no viven mucho tiempo; lo máximo que ha sobrevivido un adulto en condiciones de laboratorio son 83 días, pero normalmente no sobreviven más de 45 días.

### 1.2.2. Daños al agave

Los daños al agave se dividen en dos tipos, los que causan las larvas y los que causan los adultos. El adulto se alimenta principalmente en la base de las hojas donde deja una raspadura (Figura 1.3a); el daño no es grave pero la herida puede

---

<sup>2</sup>Consiste en separar la parte central del agave (corazón, piña o cabeza) de las pencas, ya que en la elaboración del tequila se utiliza únicamente la piña, que es donde se concentra la mayor cantidad de azúcares.

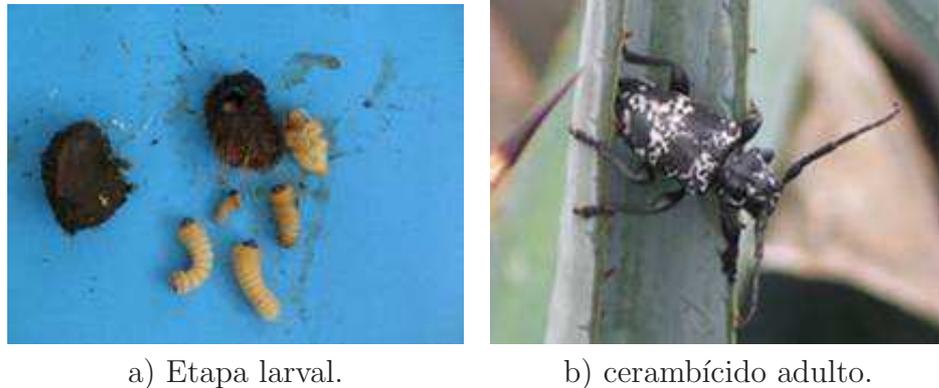


Figura 1.2: Etapas de crecimiento, cerambícidos.

funcionar como entrada de algún patógeno<sup>3</sup>. Sin embargo, los daños que causan sus larvas son:

1. Barrenan la piña y se alimentan de ella hasta por 10 meses; no matan al agave pero detienen su crecimiento.
2. Se alimentan de la zona de la corona eliminando las raíces (Figura 1.3b), provocando el síntoma de marchitez<sup>4</sup>.
3. Cuando el agave es pequeño, sobre todo si es micropropagado<sup>5</sup>, la larva se alimenta de toda la piña destruyéndola completamente.

### 1.2.3. Biología y hábitos de ataque

La compañía Tequila Sauza tiene evidencia empírica para creer que las poblaciones más grandes de adultos aparecen en la época de lluvias de cada año (aunque han encontrado algunos insectos en épocas diferentes). Se han observado insectos hasta por más de dos meses, ya que como una estrategia de supervivencia no todos los insectos emergen al mismo tiempo. Además, el manejo de las larvas es difícil por que se encuentran dentro de la piña y aún no se han encontrado resultados satisfactorios con insecticidas altamente tóxicos o sistémicos. Sin embargo, el adulto es muy sensible a cualquier insecticida, por lo que su control debe dirigirse a éste. Para lograr controlar al adulto es importante monitorear su aparición y entender su comportamiento, con lo cual se puede tomar medidas eficientes para su manejo.

<sup>3</sup>Microorganismo que puede causar o propagar enfermedades.

<sup>4</sup>Enfermedad del agave, que se debe a una deshidratación de los tejidos.

<sup>5</sup>Agave que se obtiene mediante la utilización de técnicas biotecnológicas.



a) Raspaduras en la base de las hojas de agave.



b) Daños a las raíces.

Figura 1.3: Daños al agave.

### 1.3. Problemas de interés para la empresa

Tequila Sauza necesita conocer aspectos de hábitos y biología de los cerambícidos adultos para implementar medidas de control, ya que como se mencionó en la sección anterior, las larvas son muy difíciles de manejar. Para ello es importante saber cómo se ubican en el predio de estudio: El Triángulo Amatitán, Jalisco.

El Dr. José Ignacio del Real Laborde y el M. C. Ramón Rubio Cortés representantes de Tequila Sauza, están interesados en responder a las siguientes preguntas:

1. ¿Forman los cerambícidos cúmulos en ciertas regiones del predio?
2. ¿Hay evidencia estadística que sustente el hecho de que los insectos emergen en temporada de lluvia?
3. Si la respuesta a la pregunta anterior es sí, entonces—con mayor precisión—¿cómo es que las variables meteorológicas intervienen en el número de insectos en el predio?
4. ¿Existe algún tipo de agregación temporal de los insectos?

Para contestar las preguntas anteriores Tequila Sauza cuenta con datos de dos muestreos, uno en el 2006 y otro en 2007, en el predio El Triángulo Amatitán Jalisco. En la siguiente sección se describen los datos y se exploran mediante herramientas ad-hoc.

---

## Capítulo 2

# Análisis exploratorio

---

## 2.1. Datos del primer muestreo, 2006

### 2.1.1. Descripción de los datos

El Dr. José Ignacio del Real Laborde y el M. en C. Ramón Rubio Cortés representantes de Tequila Sauza, proporcionaron la base de datos, consistente de conteos de cerambícidos tomados en distintas fechas equiespaciadas. Dichos datos fueron recolectados en Jalisco, a través de un muestreo en el predio El Triángulo Amatitán del 27 de mayo al 15 de agosto de 2006.

El muestreo se llevó de la siguiente manera:

- Se seleccionó una región en el centro del predio de 15 surcos con 150 plantas cada uno, donde la distancia entre agaves es de un metro y entre surcos es de tres metros.
- En cada surco se tomó una muestra aleatoria de diez plantas y se contó la cantidad de cerambícidos adultos presentes en cada una de éstas.
- El proceso se repitió cada 5 días.

El método de muestreo anterior se conoce en la literatura como muestreo sistemático. La base de datos está almacenada en un libro de Excel, la cual consta de los siguientes datos: número de surco, número de planta y cerambícidos encontrados; para cada fecha tienen una nueva hoja de cálculo (Ejemplo: Tabla B.1). Cabe mencionar que la forma de la base de datos es compleja y entorpece el análisis de datos mediante paquete especializado. Para ahorrar tiempo y poder trabajar en paquetes estadísticos especializados en la Sección 2.1.4.2 se hace una sugerencia para la recolección efectiva de los datos.

### 2.1.2. Depuración de datos

Un aspecto importante para el análisis estadístico eficiente es la forma de recopilar los datos. Los datos están en hojas de cálculo de Excel, y fácilmente pueden ser exportados a software estadístico. Lo ineficiente es que éstos estén separados por fecha en distintas hojas de cálculo. Por lo tanto se depuraron, es decir, se pasaron a una matriz (como en la Tabla 2.1) más fácil de leer y manipular (con la que trabajamos todo el estudio), y cuyas columnas están formadas por las siguientes variables:

- id \_ predio: Identificador del predio; sólo se tiene uno, pero es de importancia si el análisis se expande a otros predios.
- no \_ muestreo: Número de muestreo.
- Fecha: Fecha del muestreo.
- no \_ surco: Número de surco a muestrear.
- no \_ planta: Número de planta a muestrear correspondiente al surco dado por la variable “no \_ surco”.
- conteo: Número de cerambícidos encontrados en la posición dada por las variables “no \_ surco” y “no \_ planta”.

id _ predio	No _ muestreo	Fecha	No _ surco	No _ planta	conteo
1	1	5/27/06	1	19	1
1	1	5/27/06	1	54	1
1	1	5/27/06	1	59	1

Tabla 2.1: Matriz para la recopilación de información.

### 2.1.3. Análisis exploratorio inicial

#### 2.1.3.1. Mapa de intensidades de grises

Las figuras que se describen en esta sección, son para resumir gráficamente los datos y buscar indicios de agregación espacial/temporal. El predio de interés lo representamos por un rectángulo formado por subrectángulos, los cuales a su vez representan cada uno de los agaves en el predio. Para cada fecha de muestreo, dichos subrectángulos son coloreados con intensidades de grises, dependiendo del número de insectos presentes en cada agave y a la figura que resulta se le llama mapa de intensidades de grises.

En resumen se muestrearon 150 agaves cada 5 días. La Figura 2.1 muestra una resolución por planta (entendiendo por esto que cada subrectángulo representa sólo un agave). El color blanco significa que la planta no fue muestreada y entre más fuerte es el color gris, implica que la cantidad de insectos es mayor; para cada fecha se estimó la media de insectos por planta ( $\hat{\mu}$ ) que aparece en la parte inferior de la gráfica, para ver qué tan grande es la infestación.

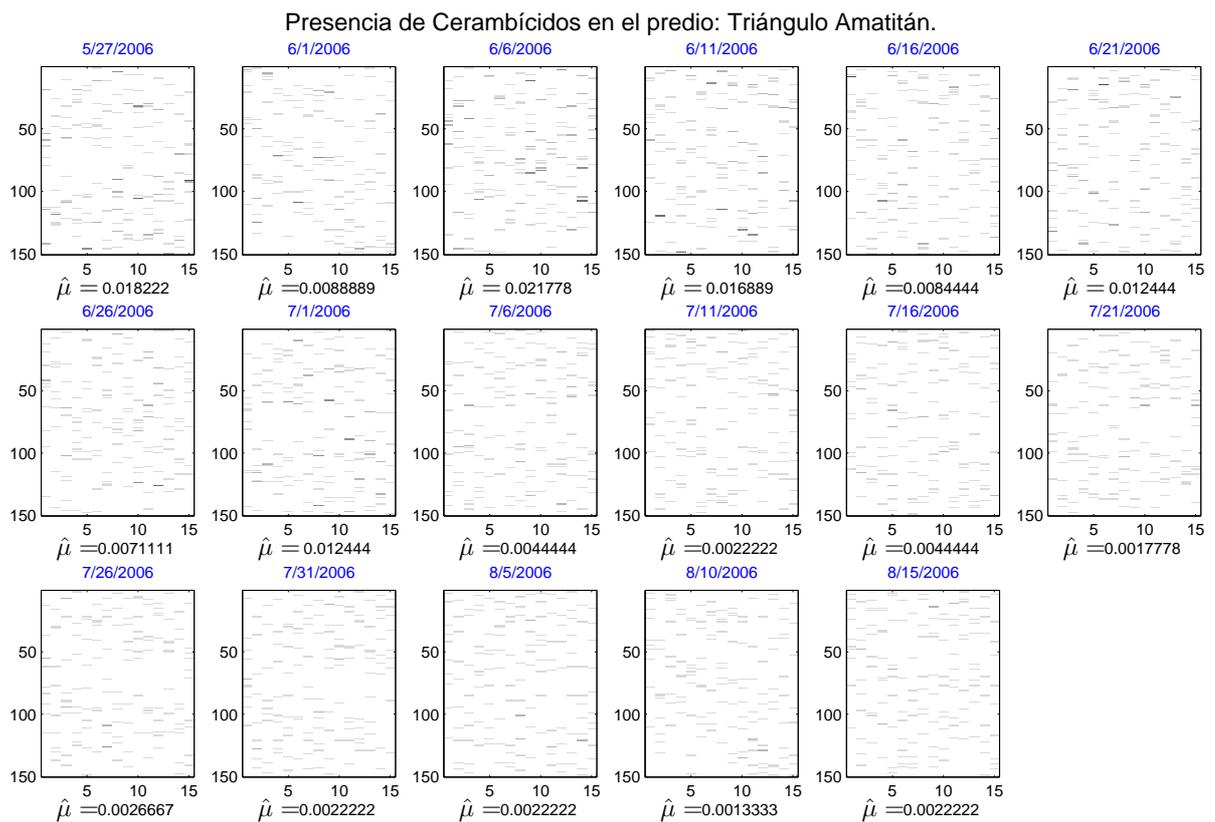


Figura 2.1: Presencia de insectos en el predio. El color blanco significa que la planta no fué muestreada, y entre más fuerte es el color gris, implica que la cantidad de insectos es mayor.

De lo anterior se tienen como resultados:

1. Si una planta tiene muchos insectos, las vecinas más cercanas muestreadas no necesariamente tienen muchos insectos; algunas ni siquiera tienen. En otras palabras, no se ven indicios de que exista agregación espacial.
2. Si una planta tiene muchos insectos para una fecha, después de 5 días ya no

tiene, es decir, las gráficas no sugieren que exista agregación temporal a largo plazo.

Dos explicaciones posibles a los resultados anteriores son que el lapso de tiempo entre muestreos es muy grande y que la resolución por planta de los mapas no es informativa. Es decir, cinco días pueden ser un período muy largo de tiempo para observar algún tipo de agregación y qué pasa si los mapas tienen una resolución mayor; esto es, ¿se observan indicios de agregación que no se ven con resolución planta? Para abordar este último punto se realizó un tipo de suavizamiento, descrito en la siguiente sección.

### 2.1.3.2. Mapa de intensidades de grises con suavizamiento

Las figuras descritas en esta sección tienen el mismo propósito que las presentadas en la sección anterior. Además, ayudaran a responder la pregunta ¿la resolución a una planta es informativa? Esto se logró a través de colorear rectángulos más grandes con intensidades de grises por niveles, es decir, se tomó el total de insectos encontrados en un rectángulo formado por un la unión de agaves que pueden pertenecer a diferentes surcos. La Figura 2.2 muestra un suavizamiento agrupando 10 plantas por surco y en la Figura 2.3 el suavizamiento consistió en tomar 3 surcos a la vez y 10 plantas en cada uno.

En conclusión, se observa lo mismo que con la gráfica a resolución planta, sólo que por zonas, sigue sin distinguirse agregación espacial/temporal. Con base en lo anterior, la resolución a planta sí es informativa.

### 2.1.3.3. Gráfica de datos meteorológicos y variable conteo

Tequila Sauza tiene evidencia empírica para pensar que los cerambícidos emergen inmediatamente después de una copiosa precipitación. Por tal motivo, se sospecha que el número de insectos presentes en el predio depende fuertemente de variables meteorológicas como: precipitación pluvial<sup>1</sup>, temperatura<sup>2</sup> y humedad relativa del medio ambiente<sup>3</sup>. La compañía tiene el equipo necesario para registrar dichas variables; sin embargo, a días transcurridos de haber iniciado el muestreo notaron que los instrumentos estaban mal calibrados y se decidió no utilizar éstos para el estudio. Dada la baja confiabilidad de los datos se empezó la tarea de conseguir datos climatológicos cercanos al predio en estudio. Los datos más viables fueron los registrados en la estación La Higuera, proporcionados por Herradura. Dicho lo anterior, la fluctuación de cada

---

<sup>1</sup>Precipitación pluvial: Un mm de precipitación significa que un recipiente abierto con paredes perfectamente verticales, acumula una columna de agua de 1 mm de altura.

<sup>2</sup>Temperatura: es tomada del medio ambiente con un termómetro.

<sup>3</sup>Humedad relativa: es el porcentaje de humedad que contiene el aire con respecto al total de humedad que es capaz de contener como función de su temperatura y su presión.

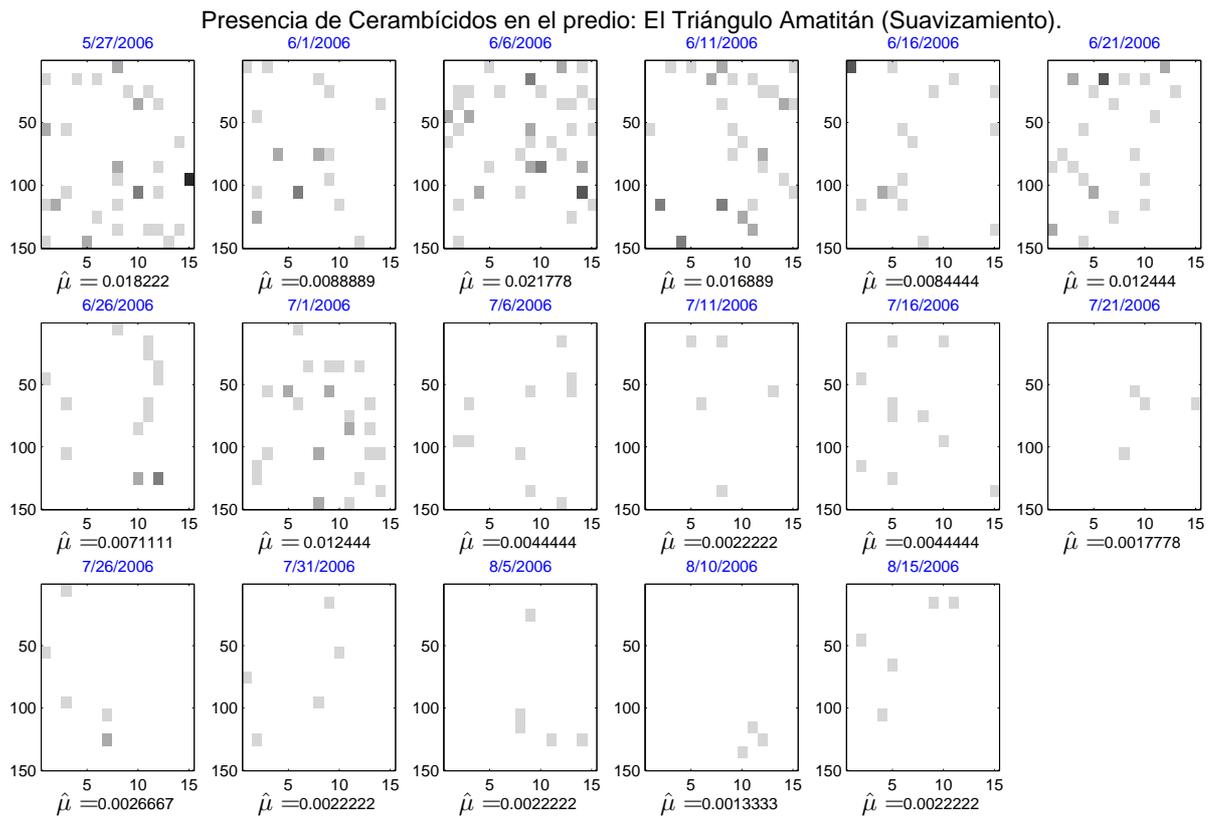


Figura 2.2: Presencia de insectos en el predio con suavizamiento. En cada surco se agruparon 10 plantas.

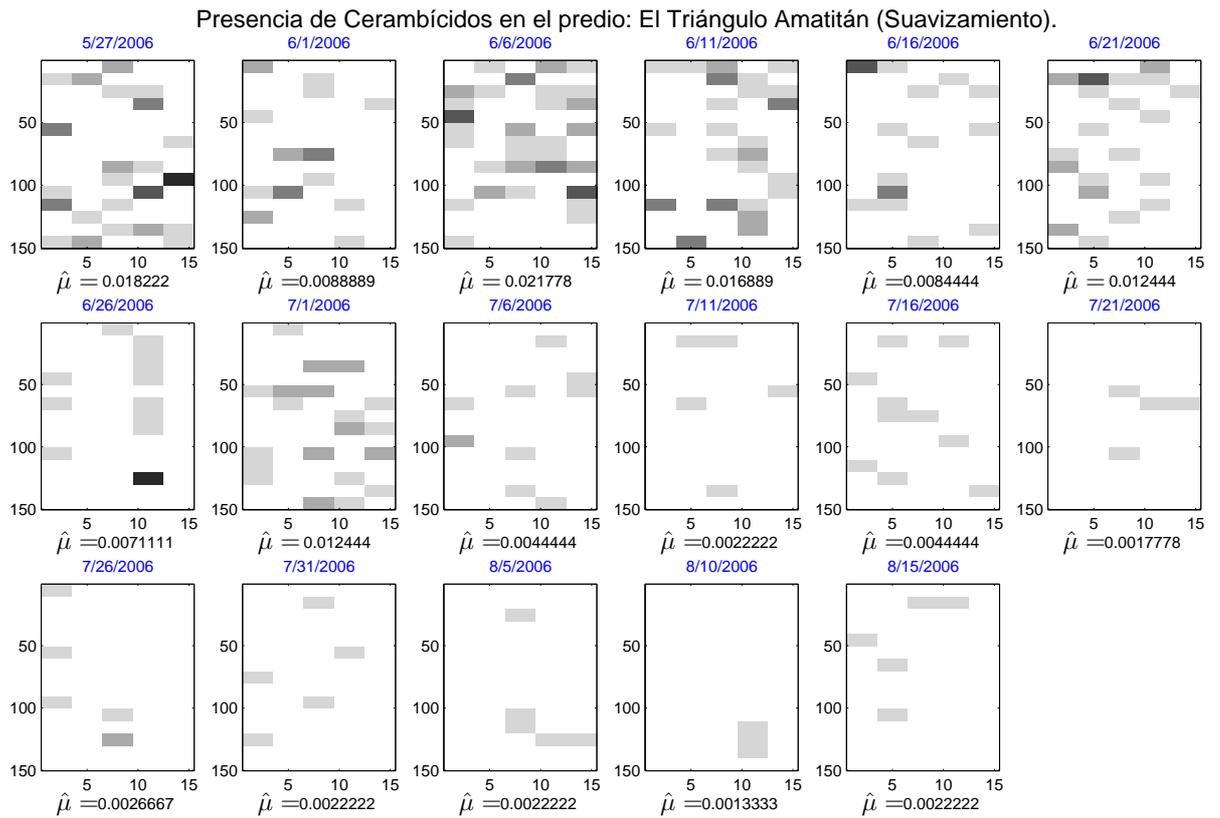


Figura 2.3: Presencia de insectos en el predio con suavizamiento. Se agruparon 3 surcos y 10 plantas en cada surco.

una de las variables meteorológicas se graficó y se comparó con los valores observados de la variable aleatoria conteo (número de insectos) para explorar una posible relación. En las Figuras 2.4 y 2.5 la línea continua representa la fluctuación de la variable meteorológica de interés (sus unidades en la ordenada izquierda), mientras los puntos representan los valores observados de la v.a. conteo (sus unidades en la ordenada derecha con el nombre de conteos). Se observa que aunque la humedad, la temperatura o precipitación aumente/disminuya no hay mayor/menor incidencia de insectos.

Los resultados que se obtienen son:

1. Las gráficas no dan evidencia que sugiera la influencia directa de las variables meteorológicas en el número de insectos.
2. Al no tener completa la base de los datos meteorológicos —sobre todo en fechas anteriores al inicio del muestreo— no se sabe si sucedió algún cambio radical en el clima que provocará la presencia de muchos insectos en el predio.

#### 2.1.3.4. Análisis de momentos

Los datos que se tienen, vistos en la Sección 2.1.1, son conteos de cerambícidos adultos y éstos deben describirse como una variable aleatoria discreta. Para indagar sobre la distribución del número de insectos en el predio y examinar la posibilidad de que se distribuyan como una variable aleatoria con distribución poisson, se comparó la media y varianza muestral de la variable conteo. La media y varianza muestral se resumen en la Figura 2.6.

Se observa que del primero de julio al quince de agosto del 2006, la media y la varianza muestral son bastante similares; esto podría dar evidencia para suponer que los datos provienen de una variable aleatoria Poisson. El significado de lo anterior es que los insectos llegan al predio y se posicionan uniformemente, sin preferencia por zonas del predio. Sin embargo, del 27 de mayo al 26 de junio la varianza es más grande y en general la varianza muestral siempre está por encima de la media muestral, es decir, parece haber sobre-dispersión. Para descartar que por fecha los datos se comportaran como una variable aleatoria Poisson, supusimos que sí lo hacían y se calculó la media, con ésta se simuló para cada fecha números aleatorios independientes Poisson y se comparó con los datos reales. El procedimiento anterior, dio evidencia contundente en contra de la distribución Poisson.

En resumen, se obtuvieron los siguientes resultados:

1. Hay sobre-dispersión, es decir, las varianzas son mayores que las medias.
2. Los datos no se distribuyen Poisson.

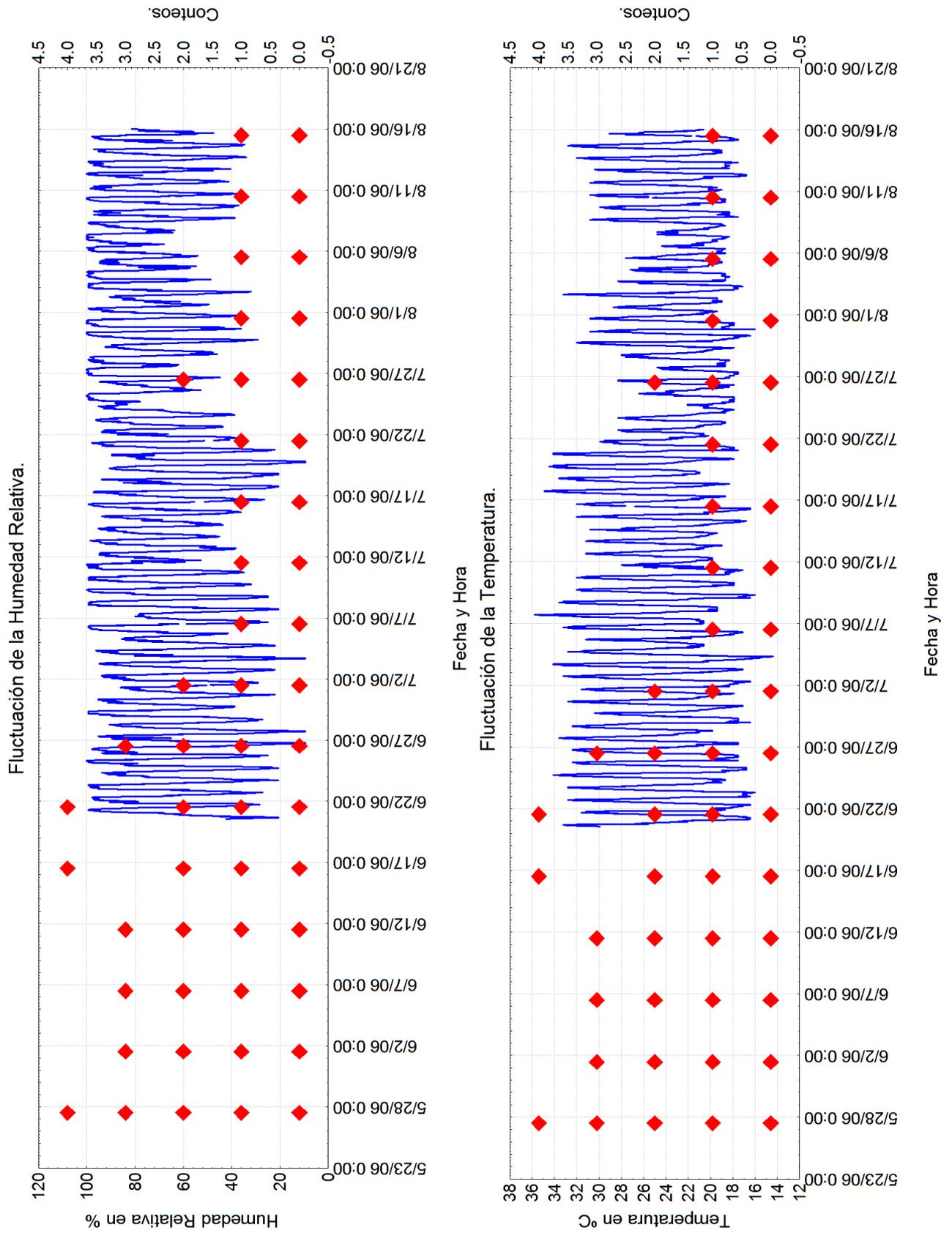


Figura 2.4: Fluctuación de variables meteorológicas.

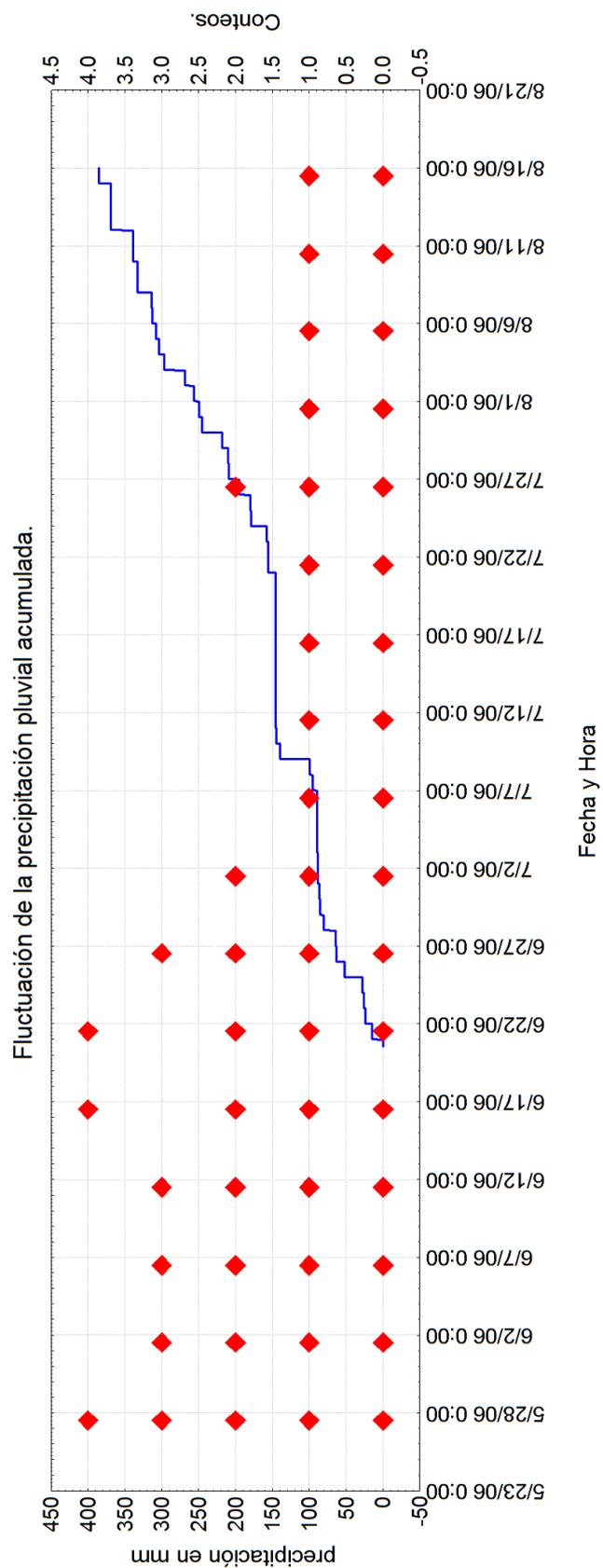


Figura 2.5: Fluctuación de variables meteorológicas (continuación).

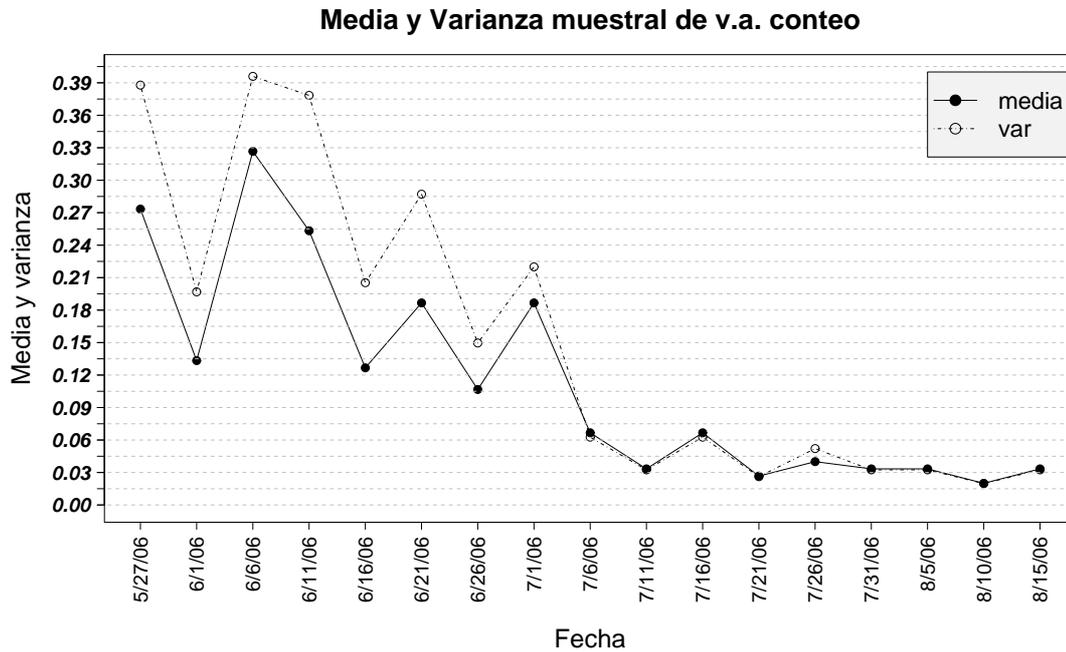


Figura 2.6: Media contra varianza muestral.

## 2.1.4. Diagnóstico

### 2.1.4.1. Conclusiones preliminares

Con base en las secciones anteriores concluimos lo siguiente:

1. El muestreo a 5 días es no informativo, es decir, este periodo de tiempo es muy grande para observar algún tipo de agregación.
2. No existe evidencia contundente para concluir la existencia de agregación espacial/temporal.
3. Los datos meteorológicos no son confiables y no se observa ningún indicio de cómo éstos afectan el número de insectos presentes en el predio.

En el marco contextual se vio que los daños más graves al agave son causados por las larvas de los cerambícidos. Sin embargo, tratar de hacer un muestreo de larvas es bastante difícil ya que están dentro de la piña. Para poder observar larvas es necesario partir la piña, cosa que de ninguna manera se puede llevar al cabo, ya que esto implicaría la destrucción total del agave. Por lo tanto, en la Sección 2.1.4.2 se propone un análisis para hacer inferencia estadística de las larvas a través de las hembras.

#### 2.1.4.2. Sugerencias

Se tienen las siguientes sugerencias:

1. Realizar el muestreo diariamente. Como se vio, realizarlo cada 5 días es insuficiente, ya que en este tiempo no se ve alguna estructura de agregación espacial/temporal. Además, los representantes de Tequila Sauza mencionan que no es difícil contar los cerambícidos adultos en campo, lo cual hace esta tarea mucho más fácil.
2. Cambiar el muestreo que estaban llevando al cabo, por un muestreo completamente aleatorio<sup>4</sup>, lo que permitiría un análisis sencillo para calcular la proporción de alguna característica binaria de interés para la Tequila Sauza, especificando la función de verosimilitud. Sin embargo, Tequila Sauza decidió disminuir la zona a muestrear y hacer un censo. Es decir, muestrear todas las plantas.
3. Cambiar el formato de recopilación de datos, vía una matriz, siguiendo la idea de la Sección 2.1.2.
4. Realizar un muestreo para estimar la proporción de hembras presentes en el predio, el cual se describe completamente en Apéndice A. Aunque no se implementó por razones económicas y biológicas, argumentando que no tenían el personal necesario para sexar a los insectos y que no se sabe que reacción tendrían éstos al quitar algunos del predio.

## 2.2. Datos del segundo muestreo, 2007

### 2.2.1. Descripción de los datos

El segundo muestreo fue diseñado tras conclusiones de primer muestreo, detalladas en el capítulo anterior. El Dr. José Ignacio del Real Laborde y el M. en C. Ramón Rubio Cortés decidieron realizar un censo diariamente, pero ahora se redujo la zona del predio a muestrear; empezando el 9 de junio de 2007. El censo se llevó de la siguiente manera:

- Se seleccionó una región en el centro del predio de 10 surcos con 100 plantas cada uno, donde la distancia entre agaves es de un metro y entre surcos es de tres metros.
- En todas las plantas (1000) se contó la cantidad de cerambícidos adultos presentes.
- El proceso se repitió diariamente.

---

<sup>4</sup>Consiste en tomar las 150 plantas de manera uniforme sobre todo el predio.

## 2.2.2. Análisis exploratorio secundario

### 2.2.2.1. Mapa de intensidades de grises

En este capítulo se resumen los datos mediante las mismas gráficas descritas en la Sección 2.1.3.1 y 2.1.3.2, las cuales ayudarán para ver si hay agregación espacio/temporal, Figura 2.7 y 2.8.

Los mapas de intensidades de grises sugieren agregación temporal. Se observa que si en una fecha la planta tiene presencia de insectos para una o dos fechas posteriores sigue teniendo, es decir, hay correlación temporal. El número de insectos presentes en el predio para una fecha determinada, depende del número de insectos presentes en fechas anteriores; en la siguiente sección se presenta una técnica para determinar cuántos días anteriores a la aparición de insectos son estadísticamente significativos. Sin embargo, la agregación espacial no se nota a simple vista (se esperaría que si una planta está infestada con muchos individuos entonces las vecinas más cercanas también deberían de tener insectos), es decir, parece que el número de insectos por planta son independientes. Esto se verá con detalle más adelante vía simulaciones.

### 2.2.2.2. Gráficas de la función de autocorrelación

Esta herramienta se utiliza en teoría de series de tiempo para obtener el orden de un proceso autorregresivo o de promedios móviles. Con dicha herramienta se obtendrán las correlaciones entre fechas rezagadas, es decir, cuántos días anteriores a una fecha influyen significativamente en el número de insectos presentes en una agrupación (diez plantas en cada surco)<sup>5</sup>. Para cada agave, se tomó el número de insectos en cada fecha creando una serie de tiempo y se calculó la función de autocorrelación parcial. La Figura 2.9 muestra 25 ejemplos de lo que se observa en general. Los puntos son la correlación estimada entre fechas y las líneas horizontales perpendiculares al eje de las abscisas representan el intervalo de confianza para la correlación cero; en la mayoría de todas las agrupaciones sólo el rezago de orden 1 es significativo.

Con base en las gráficas de la función de autocorrelación, se tienen las siguientes conclusiones:

1. Claramente los conteos no son estacionarios.
2. El número de insectos presentes para una fecha determinada, depende significativamente sólo del día anterior.

---

<sup>5</sup>Se tomaran los datos por agrupaciones para disminuir el número de gráficas a presentar y por que son más claras.

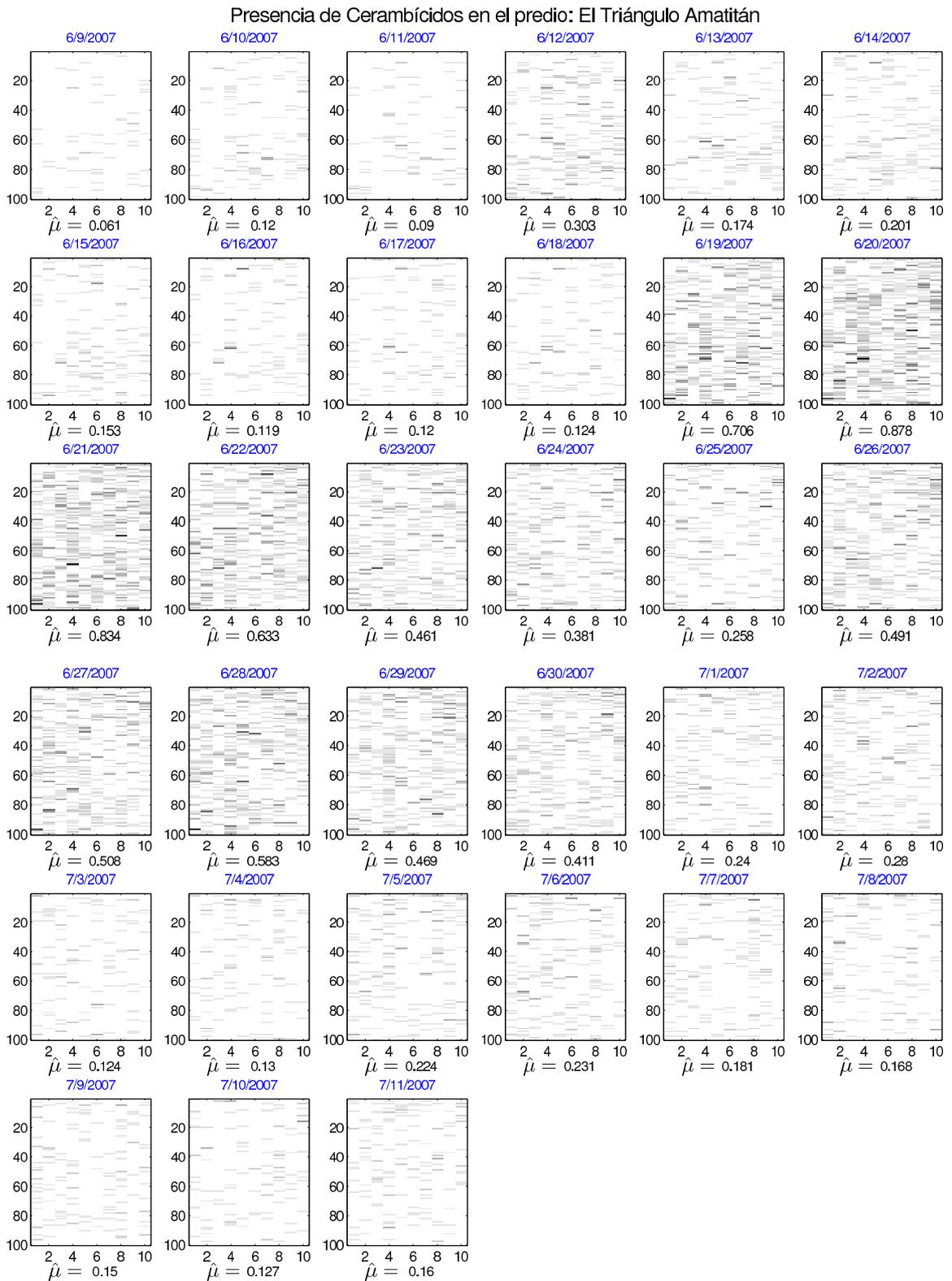


Figura 2.7: Presencia de insectos en el predio El Triángulo Amatitán.

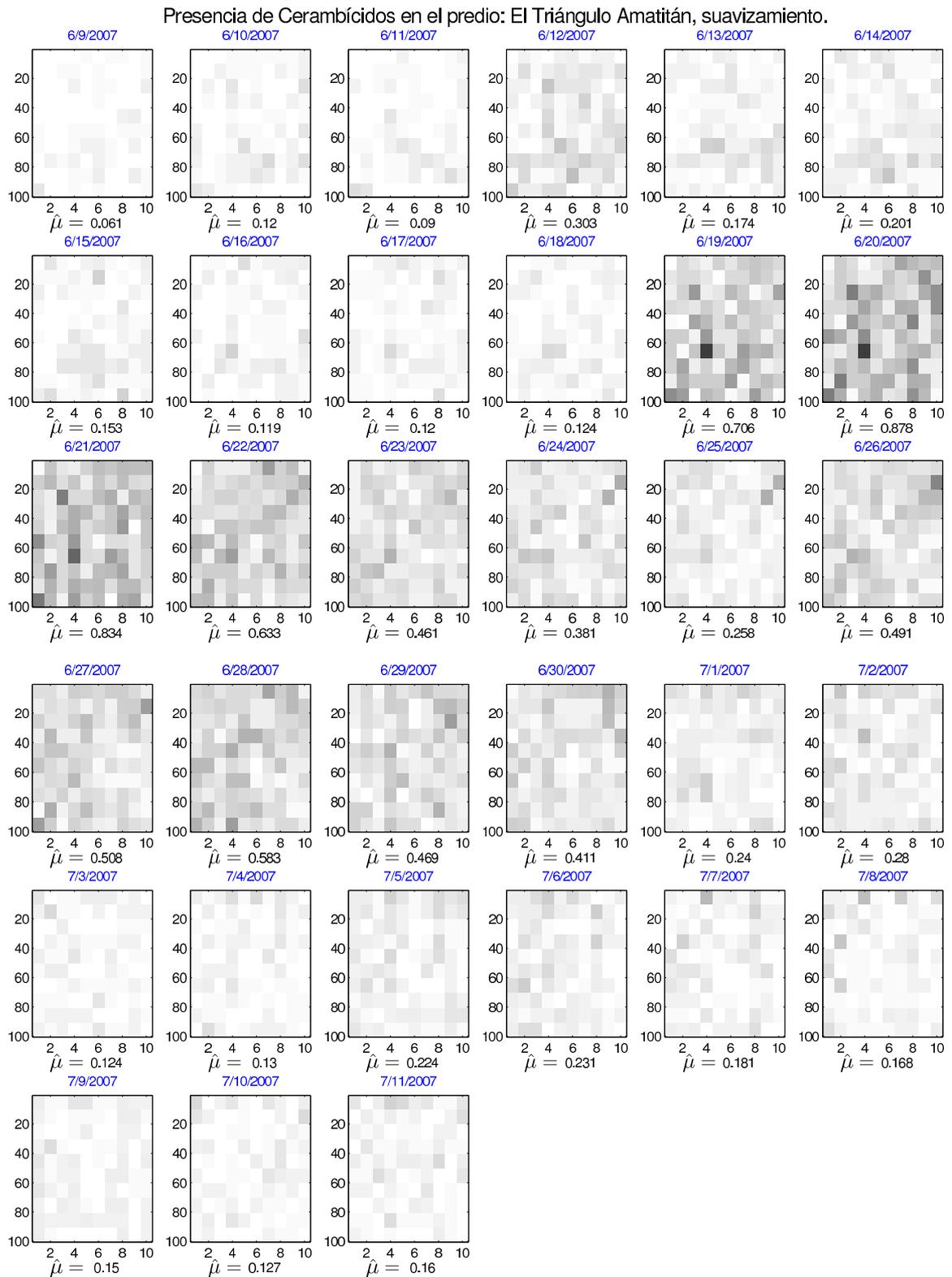


Figura 2.8: Presencia de insectos en el predio El Triángulo Amatitán con suavizamiento, con agrupación de 10 plantas en cada surco.

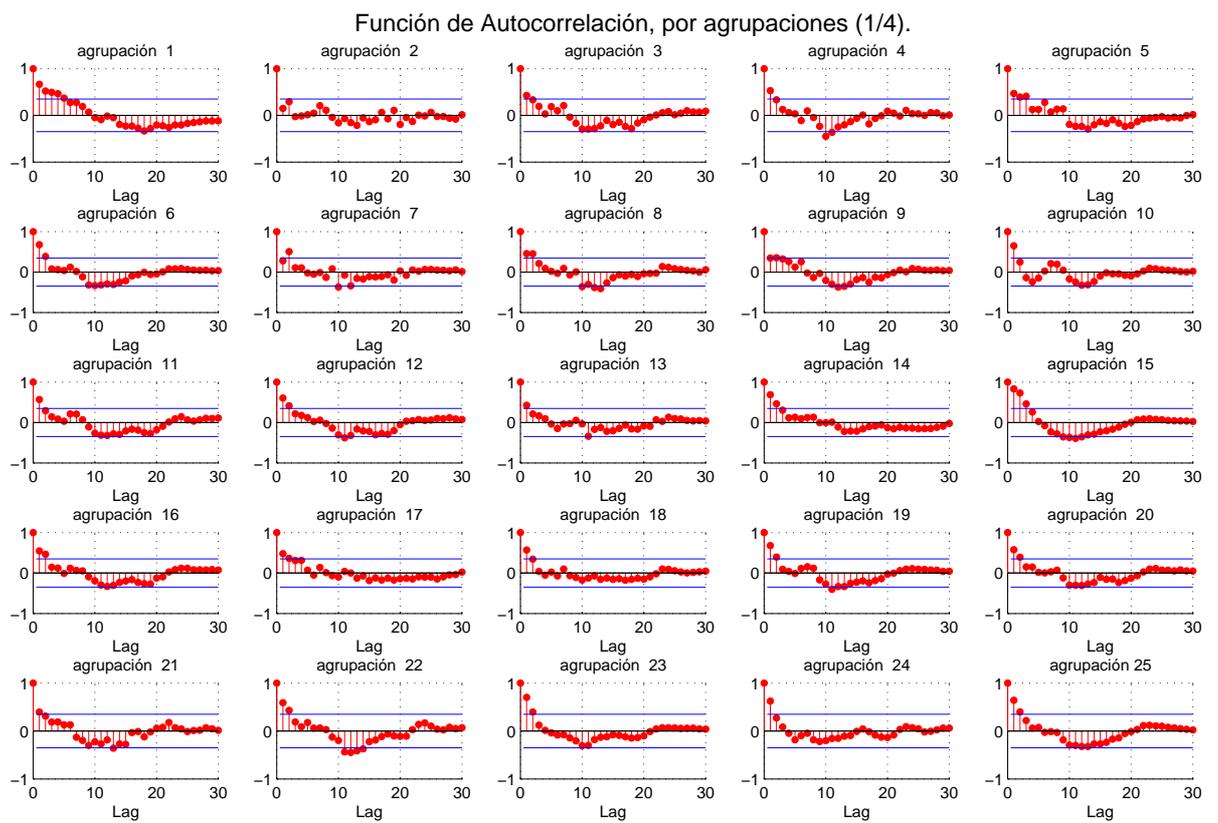


Figura 2.9: Función de autocorrelación de la agrupación 1 a la 25.

### 2.2.2.3. Gráficas de datos meteorológicos y variable conteo

En la Sección 2.1.3.3 se mencionó que Tequila Sauza, tiene evidencia empírica para pensar que los cerambícidos emergen después de una copiosa precipitación. Es decir, que el número de insectos en el predio depende de las variables meteorológicas: precipitación pluvial, temperatura y humedad relativa del medio ambiente. A diferencia del análisis del primer muestreo, los datos meteorológicos de este análisis sí son del predio, fueron tomados con instrumentos de Tequila Sauza y tienen una buena resolución. En esta sección al igual que en la Sección 2.1.3.3 se presenta la fluctuación de las variables meteorológicas de interés contra los valores observados de la variable aleatoria conteo para explorar una posible relación, Figuras (2.10) y (2.11). La variable meteorológica que da mayor información es la humedad relativa; si ésta aumenta hay mayor número de insectos. La fluctuación de la temperatura y la precipitación pluvial no son muy informativas. Sin embargo, recordemos que las variables meteorológicas están correlacionadas entre sí.

En conclusión, hay correlación entre las variables meteorológicas y el número de insectos.

### 2.2.2.4. Análisis de momentos y método gráfico de bondad de ajuste

Para indagar sobre la distribución de la variable aleatoria conteo, se comparó la media y la varianza muestral, lo cual se muestra en la Figura 2.12. La varianza muestral está siempre por encima de la media muestral, es decir, hay sobre-dispersión. Autores como Anscombe (1949), Ross y Preece (1985) sugieren que la distribución binomial negativa se ajusta bien a datos de conteo con sobre-dispersión, proporcionando una interpretación física para el caso de poblaciones de insectos. Siguiendo estas ideas se decidió explorar el ajuste de esta distribución a los datos, vía la función generadora de probabilidades empírica, el cual es un método gráfico de bondad de ajuste (se explica a detalle en la Sección 3.2). La Figura 2.13 muestra el ejemplo de un ajuste bueno y otro malo.

Se obtienen las siguientes conclusiones:

1. Hay sobre-dispersión.
2. El ajuste de la distribución binomial negativa es muy bueno para algunos casos.
3. Aproximadamente para el 40% de las fechas el ajuste no es aceptable.

Con base en las observaciones anteriores y como ejercicio académico, se supuso que los datos observados provienen de una distribución binomial negativa, se estimaron sus parámetros y se simularon variables aleatorias independientes para cada fecha. Hecho lo anterior, los resultados de la simulación se compararon con los datos originales, lo cual nos llevó a las conclusiones de que las distribuciones marginales de los datos

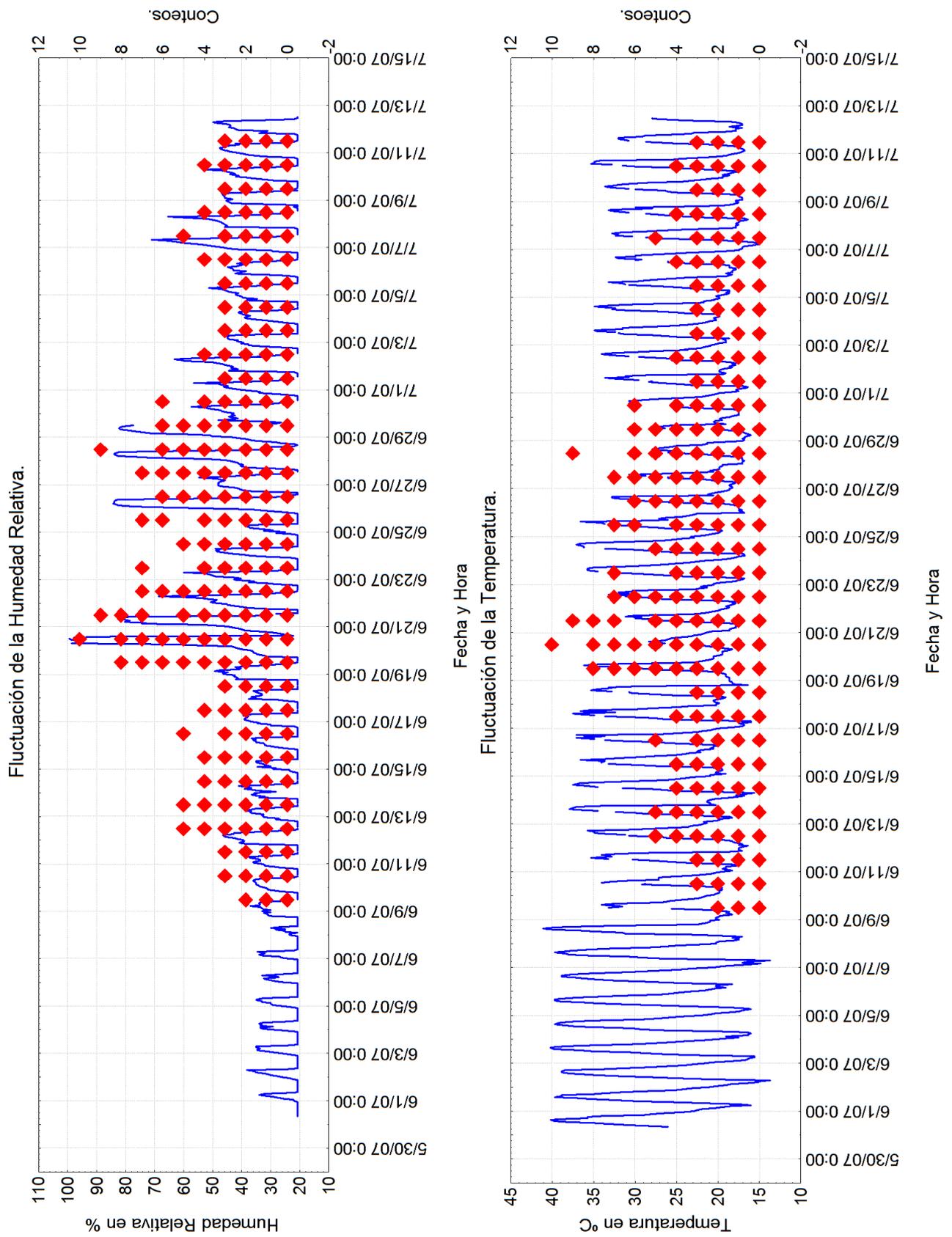


Figura 2.10: Variables meteorológicas y conteos.

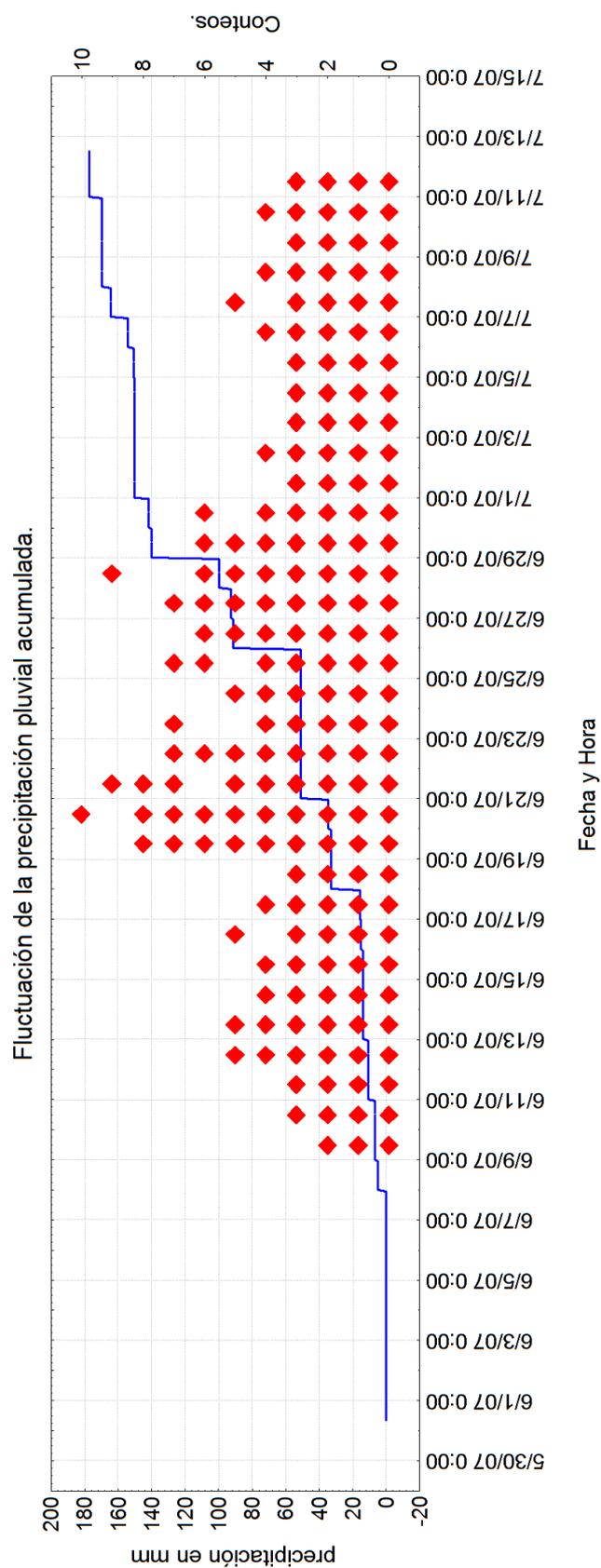


Figura 2.11: Variables meteorológicas y conteos (continuación).

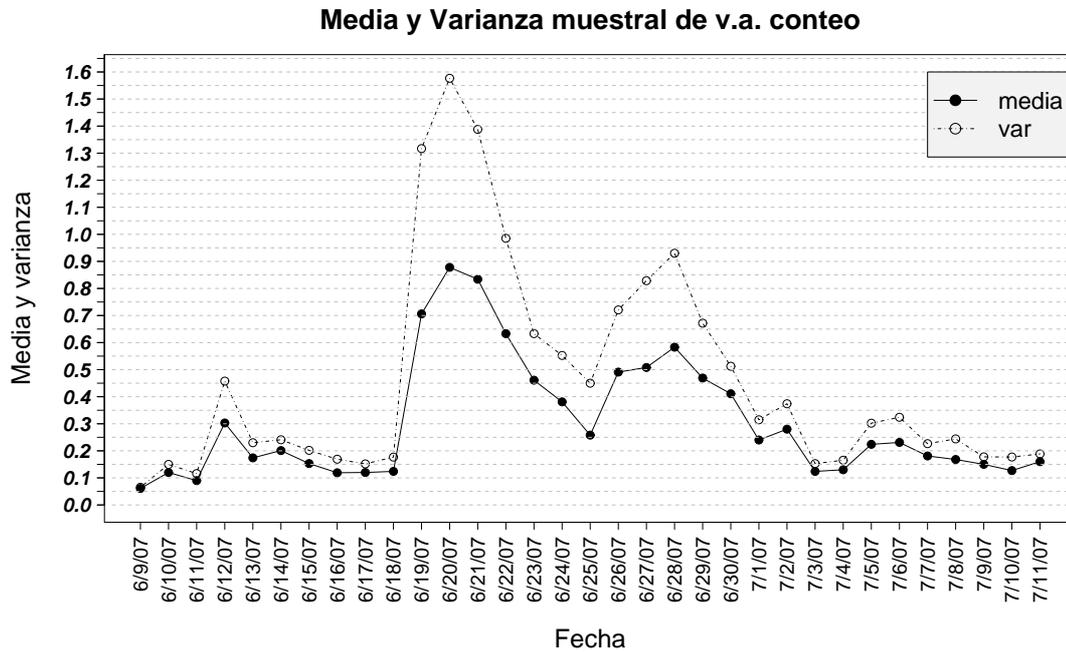


Figura 2.12: Media contra varianza muestral.

podrían ser ajustadas con esta distribución y que no hay agregación espacial. Sin embargo, es importante notar que esta distribución no recoge la correlación temporal, es decir, es necesario especificar la distribución conjunta.

### 2.2.3. Conclusiones

En esta sección se resumen las conclusiones del análisis exploratorio secundario.

1. Hay correlación temporal.
2. No hay agregación espacial.
3. Hay sobre-dispersión.
4. Las variables meteorológicas: humedad relativa del medio ambiente, temperatura y precipitación pluvial influyen en el número de insectos presentes en el predio (variable aleatoria conteo).
5. Para cada fecha los conteos podrían ser ajustados aproximadamente por la distribución binomial negativa, sin embargo, esta no recoge la correlación temporal, si se suponen conteos independientes por fecha.

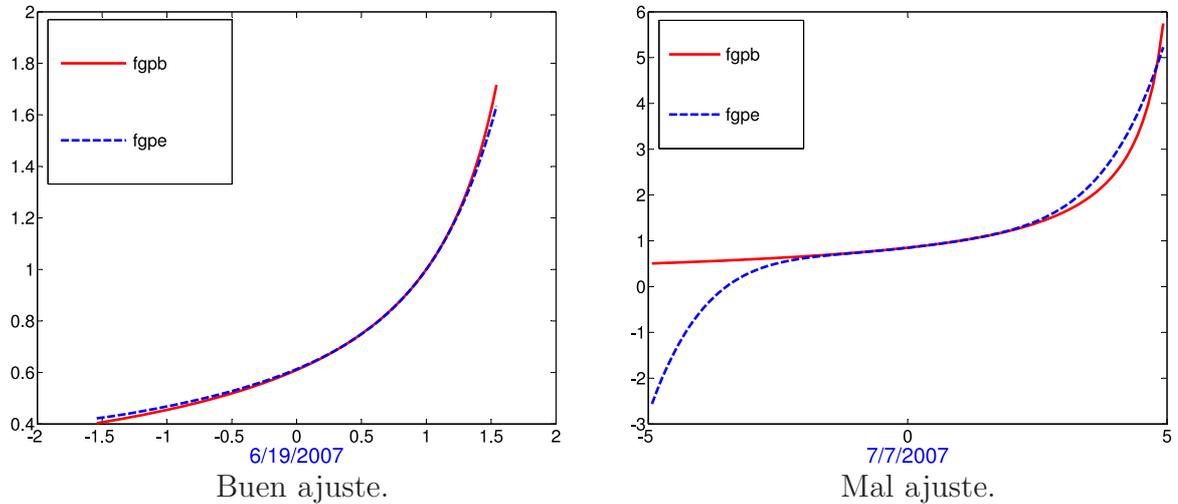


Figura 2.13: Método gráfico de bondad de ajuste: función generadora de probabilidades empírica (fgpe) contra la función generadora de probabilidades estimada de una variable aleatoria binomial negativa (fgpb).

6. Los conteos no son estacionarios.
7. El número de insectos presentes para una fecha determinada, depende significativamente sólo del día anterior.

Estas conclusiones del análisis exploratorio secundario, ayudaron a proponer y obtener características para la modelación, como son:

1. Se propone un modelo autorregresivo de orden 1, con una distribución más general que la binomial negativa.
2. Como las variables meteorológicas influyen en el número de insectos, entonces los parámetros del modelo deben depender de éstas.

---

## Capítulo 3

# Propuesta del modelo probabilístico

---

### 3.1. Modelo autorregresivo de orden 1 BN general

El modelo probabilístico que se propone en esta sección, es una propuesta de solución a la pregunta ¿existe agregación espacio/temporal? Explicando el comportamiento de los cerambícidos en el predio, a través del tiempo y espacio. La modelación tomó en cuenta los siguientes tres aspectos: 1) El análisis secundario (la no estacionaridad de los conteos, rezagos temporales de orden un día, los conteos no son necesariamente binomiales negativos, la no correlación espacial y la posibilidad de incorporar meteorología en los parámetros del modelo), 2) características de cerambícidos (posicionamiento del predio, su necesidad de humedad, el difícil manejo de larvas y la movilidad de los insectos) y 3) antecedentes bibliográficos (de la distribución binomial negativa y de los modelos autorregresivos de orden 1, AR(1)).

El desarrollo del siguiente modelo probabilístico, fue un proceso que consistió de investigación, creatividad y, ensayo y error. El artículo de Anscombe (1949), describe un análisis estadístico de conteos de insectos con la distribución binomial negativa; en el cual observa que los datos de conteos de insectos u otras poblaciones generalmente tienen sobre-dispersión, y por lo tanto pueden ser bien ajustados con la distribución binomial negativa. Además, da interesantes interpretaciones biológicas a los parámetros de dicha distribución. El artículo de McKenzie (1988), introduce a la una familia de procesos estocásticos que tienen distribuciones discretas de cierto tipo, conocidas como “self-decomposable” en marginales. Estos procesos tienen muchas características en común con los modelos ARMA; como la forma de expresar el proceso, por una relación recursiva y la forma de la función de autocorrelación. Finalmente, el artículo de McCormick (1992) define el proceso autorregresivo de orden 1 de la binomial negativa con interesantes propiedades asintóticas. Estos tres artículos son clave importante para la propuesta del siguiente modelo.

Para cada uno de los 1000 agaves que conforman el predio de estudio, se propone un modelo autorregresivo de orden 1 no estacionario con distribuciones marginales más generales que la distribución binomial negativa (BinNeg), que llamaremos: “Modelo AR(1) BN general”. Dicho modelo está definido implícitamente por la ecuación recursiva

$$X_{n,i} = \alpha_n * X_{n-1,i} + Z_n, \quad n \geq 1 \quad \& \quad i = 1, \dots, 1000,$$

$$X_{0,i} \sim \text{BinNeg}(\theta_0, \beta_0), \quad (3.1)$$

donde,

- La variable aleatoria  $X_{n,i}$  representa el número de insectos presentes al tiempo  $n$  en la planta  $i$ -ésima.
- La variable aleatoria  $\alpha_n * X_{n-1,i}$  se define como  $\sum_{j=1}^{X_{n-1,i}} B_j(\alpha_n)$  donde  $\{B_j(\alpha_n), j \geq 1\}$  es una secuencia de variables aleatorias binarias independientes e idénticamente distribuidas (v.a.i.i.d) con  $P[B_j(\alpha_n) = 1] = \alpha_n = 1 - P[B_j(\alpha_n) = 0]$ . El  $*$ -operador de  $\alpha_n$  sobre  $X_{n-1,i}$  describe un adelgazamiento binomial de la v.a.  $X_{n-1,i}$ .
- La variable aleatoria  $Z_n$  es función de tres variables aleatorias, definida por:

$$Z_n = \sum_{j=1}^{N_n} (\alpha_n^{R_j}) * W_{n,j} = \sum_{j=1}^{N_n} \sum_{k=1}^{W_{n,j}} B_k(\alpha_n^{R_j}), \quad n \geq 1, \quad (3.2)$$

donde  $N_n, R_j, W_{n,j}, j \geq 1$  son variables independientes tales que

$$\begin{aligned} N_n &\sim \text{Poisson}(\mu_n = -\beta_n \ln(\alpha_n)) \\ R_j &\sim \text{Uniforme}(0, 1) \\ W_{n,j} &\sim \text{Geométrica}(\theta_n). \end{aligned} \quad (3.3)$$

Además,  $Z_n$  es independiente de la variable aleatoria  $\alpha_n * X_{n-1,i}$ . Notemos que para cada fecha, este modelo tiene tres parámetros:  $\alpha_n, \theta_n$  y  $\mu_n$  los cuales pueden hacerse depender de variables meteorológicas (humedad relativa, temperatura y precipitación pluvial). El Modelo AR(1) BN general se interpreta como:

*El número de insectos en una planta al tiempo  $n$ , es igual a algunos que se quedaron del tiempo  $n - 1$  más otros insectos que emergen aleatoriamente.*

### 3.1.1. Interpretación de las variables aleatorias del Modelo AR(1) BN general.

#### 3.1.1.1. Variable aleatoria $N_n$

Primero se define el concepto “oleada”. Una **oleada** se define como la emergencia de nuevos individuos en la planta  $j$ -ésima provocada por el ambiente, que consiste de adultos que emergen en otras plantas y se fijaron en ésta, y de adultos que emergen sobre la misma planta.

La cantidad  $N_n$  representa el número aleatorio de eventos que causan oleadas en un día determinado  $n$ ; ésta depende de factores meteorológicos y estrategias de sobrevivencia de los insectos a través de sus parámetros. Suponiendo que en un día  $n$  los fenómenos climatológicos como lloviznas, aguaceros, cambio de temperaturas, *etc.* suceden aleatoriamente y uniformes sobre el predio, la distribución que sigue la v.a.  $N_n$  está dada por

$$P(N_n = \eta) = \frac{\mu_n^\eta e^{-\mu_n}}{\eta!}, \quad \eta > 0. \quad (3.4)$$

El parámetro  $\mu_n$ , siempre es positivo y se define como:

$$\mu_n = -\beta_n \ln(\alpha_n). \quad (3.5)$$

#### 3.1.1.2. Variable aleatoria $R_i$

Para un índice fijo,  $R_i$  determina la probabilidad de que insectos surgidos de las oleadas permanezcan en la planta. Otra razón de tener esta variable aleatoria es un razón puramente matemática; para que el modelo definido en la Ecuación (3.1) se pueda generalizar a la distribución de una v.a. binomial negativa en marginales, como se describirá en la Sección 3.1.3. La distribución que sigue  $R_i$  es,

$$P[R_i \leq r] = rI_{(0,1)}(r) + I_{[1,\infty)}(r). \quad (3.6)$$

#### 3.1.1.3. Variable aleatoria $W_{n,i}$

Para un tiempo  $n$  y un índice fijo,  $W_{n,i}$  representa el número aleatorio de insectos producidos en una oleada, es decir,  $W_{n,i}$  es a su vez la suma de dos variables aleatorias. La primera cuenta insectos locales (los que emergieron de la planta) en una oleada y la segunda cuenta insectos foráneos (que llegaron de otras plantas). Esta v.a. sigue la ley de una distribución geométrica con parámetro  $\theta_n$ , es decir

$$P[W_{n,i} = w] = \theta_n(1 - \theta_n)^w, \quad w = 0, 1, \dots \quad \& \quad \theta_n \in (0, 1). \quad (3.7)$$

Para justificar que  $W_{n,i}$  es una v.a. geométrica, se prueba que en realidad es una v.a. poisson con parámetro aleatorio, como lo demuestra el corolario del siguiente

teorema.

**Teorema 1** Si la variable aleatoria  $X|\Lambda \sim \text{Poisson}(\Lambda)$  y  $\Lambda \sim \text{Gamma}(a, b)$  con  $a, b > 0$ , entonces  $X \sim \text{BinNeg}(p, k)$  donde  $p = \frac{a}{a+1}$  y  $k = b$ .

**Demostración**

Recordemos que para  $a, b > 0$ ,

$$f_{\Lambda}(\lambda) = \frac{a^b}{\Gamma(b)} \lambda^{b-1} e^{-a\lambda}; \quad (3.8)$$

$$f_{X|\Lambda=\lambda}(x|\lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \text{ si } \lambda > 0; \text{ y} \quad (3.9)$$

$$\Gamma(b) = \int_0^{\infty} x^{b-1} e^{-x} dx = (b-1)! \text{ si } b \text{ es un entero.} \quad (3.10)$$

Utilizando las hipótesis del teorema, se tiene

$$\begin{aligned} f_X(x) &= \int_0^{\infty} f_{X|\Lambda=\lambda}(x|\lambda) f_{\Lambda}(\lambda) d\lambda \\ &= \int_0^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} \frac{a^b}{\Gamma(b)} \lambda^{b-1} e^{-a\lambda} d\lambda \\ &= \frac{a^b}{\Gamma(b)x!} \int_0^{\infty} \lambda^{b-1+x} e^{-\lambda(a+1)} d\lambda. \end{aligned}$$

Utilizando el cambio de variable  $u = \lambda(a+1)$ , se obtiene

$$\begin{aligned} f_X(x) &= \frac{a^b}{\Gamma(b)x!} \frac{1}{a+1} \int_0^{\infty} \left( \frac{u}{a+1} \right)^{b-1+x} e^{-u} du \\ &= \frac{a^b}{\Gamma(b)x!} \frac{1}{(a+1)^{b+x}} \int_0^{\infty} u^{b-1+x} e^{-u} du \\ &= \frac{a^b}{\Gamma(b)x!} \frac{1}{(a+1)^{b+x}} \Gamma(b+x) \\ &= \frac{a^b}{(a+1)^{b+x}} \frac{\Gamma(b+x)}{\Gamma(b)\Gamma(x+1)} \\ &= \frac{a^b}{(a+1)^b} \frac{1}{(a+1)^x} \frac{\Gamma(b+x)}{\Gamma(b)\Gamma(x+1)} \\ &= \left( \frac{a}{a+1} \right)^b \left( 1 - \frac{a}{a+1} \right)^x \frac{\Gamma(b+x)}{\Gamma(b)\Gamma(x+1)} \\ &= p^k (1-p)^x \frac{\Gamma(b+x)}{\Gamma(b)\Gamma(x+1)}. \end{aligned} \quad (3.11)$$

Tomando  $p = \frac{a}{a+1}$  y  $k = b$  se concluye que la variable aleatoria  $X \sim \text{BinNeg}(p, k)$ . ■

**Corolario 1** Si la variable aleatoria  $X|\Lambda \sim \text{Poisson}(\Lambda)$  y  $\Lambda \sim \text{Exponencial}(a)$  con  $a > 0$ , entonces  $X \sim \text{Geométrica}(p)$  donde  $p = \frac{a}{a+1}$ .

Este tipo de variables, especificadas por un mecanismo de condicionamiento, da lugar a una clase de modelos conocida como modelos jerárquicos. Es un recurso muy utilizado—aunque no exclusivamente—en la literatura de inferencia bayesiana. Supongamos que inicialmente se tienen  $Y_1, \dots, Y_m$  variables aleatorias independientes con  $Y_i \sim f_j(y; \theta)$ , donde  $\theta$  es un parámetro escalar. Supongamos además que existen factores variables que influyen sobre la distribución de observaciones. Un ejemplo de un posible factor es que las observaciones se toman en diferentes regiones geográficas o zonas heterogéneas. Una manera de representar esto es concibiendo que el parámetro  $\theta$  es a su vez de una observación de una variable aleatoria  $\Theta$ . En el predio de estudio existen una multiplicidad de factores bióticos y abióticos, que sugieren que el parámetro de la variable aleatoria Poisson (compuesta por la suma de insectos locales y foráneos) sea aleatorio. Para que se obtenga como resultado la variable aleatoria geométrica prescrita en el modelo, una suposición congruente es que la distribución para  $\Theta$  es exponencial. Así, este término tiene la interpretación de ser consecuencia de una jerarquización exponencial-Poisson, y es una forma de explicar el fenómeno de sobre-dispersión que muestran los datos. Para mayor detalle ver el artículo de Cox (1983).

### 3.1.2. Interpretación de los parámetros del Modelo AR(1) BN general

#### 3.1.2.1. Parámetro $\alpha_n$

Para un día  $n$ , el parámetro  $\alpha_n$  se interpreta como la probabilidad de que un insecto que está en el tiempo  $n - 1$  en una planta, permanezca ahí al tiempo  $n$ . El parámetro  $\alpha_n$  toma valores en el intervalo abierto  $(0, 1)$  y se encuentra en los dos sumandos de la Ecuación (3.1).

#### 3.1.2.2. Parámetro $\theta_n$

Para un día  $n$ , el parámetro  $\theta_n$  tiene la interpretación de ser un modulador de la distribución del tamaño de las poblaciones que emergen en una oleada; éste toma valores en el intervalo abierto  $(0, 1)$ . El tamaño medio de dicha población está dado por

$$t_m = \frac{1 - \theta_n}{\theta_n}. \quad (3.12)$$

### 3.1.2.3. Parámetro $\mu_n$

Para un día  $n$ , el parámetro  $\mu_n$  se define por la ecuación

$$\mu_n = -\beta_n \ln(\alpha_n), \quad \beta_n > 0. \quad (3.13)$$

Este parámetro se interpreta como la media de las oleadas y es no negativo. Además, es función de dos parámetros  $\alpha_n$  (cuya interpretación se dio en la Sección 3.1.2.1) y  $\beta_n$ , éste último se interpreta como un parámetro moderador de la media de oleadas. En realidad el parámetro  $\mu_n$  es un estira y afloja entre  $\beta_n$  y  $\alpha_n$ , es decir, entre meteorología y estrategias de supervivencia de los insectos, por ejemplo:

1. Si  $\alpha_n$  es aproximadamente uno, la media de oleadas es pequeña.
2. Si  $\beta_n$  es aproximadamente cero, la media es pequeña, *etc.*

### 3.1.3. Caso particular del Modelo AR(1) BN general

Si  $\beta_n = \beta$ ,  $\theta_n = \theta$  y  $\alpha_n = \alpha$  para todo  $n$ , es decir, los parámetros se mantienen constantes para toda fecha. El Modelo AR(1) BN general se particulariza a un modelo autorregresivo estacionario con marginales que se distribuyen como una v.a. binomial negativa. En esta sección se verá la bibliografía y teoría relacionada con este modelo. En los artículos de McKenzie (1988) y, Al-Osh y Alzaid (1987) se introducen una familia de procesos estocásticos que tienen distribuciones discretas de cierto tipo conocido como DSD (por sus siglas en inglés “discrete self-decomposable”) como marginales. La diferencia esencial entre dichos procesos y los modelos ARMA es que la multiplicación ha sido reemplazada con una operación de adelgazamiento.

Supongamos que  $X$  es una variable aleatoria que toma valores en  $N_0 = \{0, 1, \dots\}$  con función generadora de probabilidad (fgp)  $\phi_X(t) = E[t^X]$ . Entonces, la distribución de  $X$  es DSD si

$$\phi_X(t) = \phi_X(1 - \alpha + \alpha t)\phi_\alpha(t), \quad |t| \leq 1, \quad \text{y} \quad \alpha \in (0, 1), \quad (3.14)$$

donde  $\phi_\alpha(t)$  es una función generadora de probabilidad.

**Teorema 2** Sea  $B_i, i \geq 1$  una secuencia de v.a.i.i.d Bernoulli con  $P(B_i = 1) = \alpha$ . Entonces, la v.a.  $\sum_{i=1}^X B_i$ , donde la v.a.  $X$  es independiente de  $\{B_i, i \geq 1\}$ , tiene función generadora de probabilidad  $\phi_X(1 - \alpha + \alpha t)$ .

### Demostración

Notemos que  $\sum_{i=1}^X B_i$  dada la v.a.  $X$  tiene distribución Binomial con parámetros  $X$  y  $\alpha$ , por lo tanto

$$E \left[ E(t^{\sum_{i=1}^X B_i} | X) \right] = E \left[ (1 - \alpha + \alpha t)^X \right] = \phi_X(1 - \alpha + \alpha t). \blacksquare \quad (3.15)$$

Con el teorema anterior, notemos que la descomposición en la Ecuación (3.14) es equivalente a

$$X \stackrel{d}{=} \alpha * X + Z_\alpha, \quad (3.16)$$

donde  $\alpha * X$  denota  $\sum_{i=1}^X B_i$  y  $Z_\alpha$  es una variable independiente de  $\alpha * X$  que tiene pgf  $\phi_\alpha(t)$ . Para mayores detalles de distribuciones DSD, ver Steutel y van Harn (1979). Entonces, un proceso AR(1) de  $X_n$  se define por

$$X_n = \alpha * X_{n-1} + Z_n, \quad n \geq 1, \quad (3.17)$$

cuando  $X_0$  es DSD y las variables aleatorias  $Z_n$  son i.i.d. con fgp  $\phi(t)$  dada en la Ecuación (3.14). Entonces, el modelo dado por la Ecuación (3.17) definido de manera precisa es como sigue. Sea  $\{B_{in}(\alpha), i \geq 1, n \geq 0\}$  un arreglo de variables aleatorias i.i.d. Bernoulli con probabilidad de éxito  $\alpha$ . Entonces

$$X_n = \sum_{i=1}^{X_{n-1}} B_{in}(\alpha) + Z_n, \quad n \geq 1, \quad (3.18)$$

donde las v.a.  $(Z_n, n \geq 1)$  son i.i.d. e independientes de  $(B_{in}(\alpha))$  y  $X_0$ . Dado que la distribución binomial negativa con parámetros  $\beta$  y  $\theta$  dada por

$$P(X = n) = \binom{\beta + n - 1}{n} (1 - \theta)^\beta \theta^n, \quad n = 0, 1, \dots, \quad \beta = 1, 2, \dots, \quad 0 < \theta < 1 \quad (3.19)$$

es DSD, se define el proceso estacionario Binomial Negativo AR(1) como

$$X_n = \alpha * X_{n-1} + Z_n, \quad n \geq 1, \quad (3.20)$$

donde  $X_0 \sim \text{BinNeg}(\beta, \theta)$  y  $Z_n$  tiene fgp

$$\phi_\alpha(t) = \frac{\phi_X(t)}{\phi_X(1 - \alpha + \alpha s)} = \left( \frac{1 - \theta[1 - \alpha + \alpha s]}{1 - \theta s} \right)^\beta. \quad (3.21)$$

Una representación para la variable aleatoria con la fgp anterior esta dada por

$$Z = \sum_{i=1}^N (\alpha^{R_i}) * W_i = \sum_{i=1}^N \sum_{k=1}^{W_i} B_k(\alpha^{R_i}), \quad (3.22)$$

donde  $N$ ,  $R_i$ ,  $W_i$ ,  $i \geq 1$  son variables independientes con  $N$  distribuida como una variable aleatoria poisson con media  $\mu = \beta \ln(\alpha)$ ,  $R_i$  es uniforme sobre  $(0, 1)$  y  $W_i$  se distribuye binomial negativa con parámetros 1 y  $\theta$ , es decir una variable aleatoria geométrica con parámetro  $\theta$ .

### 3.2. Estimación de los parámetros del Modelo BN general vía la función generadora de probabilidades

La estimación de parámetros vía la función generadora de probabilidades es una técnica estadística, que como su nombre lo indica, sirve para estimar los parámetros de familias de distribuciones discretas con soporte en los enteros no negativos. Si  $X$  es una variable aleatoria que pertenece a alguna de estas familias, su función generadora de probabilidad (fgp),  $\phi(t) = E[t^X]$  está siempre definida para  $-1 \leq t \leq 1$  y caracteriza completamente a la distribución de  $X$ .

Sea  $x_1, x_2, \dots, x_n$  una muestra aleatoria de tamaño  $n$  de una distribución  $F$  con fgp  $\phi$ , la función generadora de probabilidades empírica (fgpe) se define como

$$\phi_n(t) = \frac{1}{n} \sum_{i=1}^n t^{x_i}, \quad (3.23)$$

y tiene gran importancia en inferencia estadística. Los métodos estadísticos que utilizan la fgp y fgpe están basados en el hecho de que  $\phi_n(t)$  converge en varios modos a  $\phi(t)$  cuando  $n \rightarrow \infty$ ; para más detalle ver Feueverger y McDunnough (1984) y, Marques y Pérez-Abreu (1989). En esta sección utilizaremos a la fgpe para estimar los parámetros del Modelo AR(1) BN general; ya que no se conoce explícitamente a la función de densidad o distribución de las variables aleatorias  $X_{j,i}$ , pero sí se puede obtener fácilmente su función generadora de probabilidades. Primero se calcula la fgp de la v.a.  $Z_n$ , condicionando sobre las variables que la componen. El resultado es

$$\phi_{Z_n} = \left[ \frac{1 - \theta_n(1 - \alpha_n + \alpha_n t)}{1 - \theta_n t} \right]^{\beta_n}. \quad (3.24)$$

Como observación, una ventaja de utilizar estimación máxima verosimilitud sería el hecho de hacer selección de modelos mediante razón de verosimilitudes o criterios de información como BIC (“Bayesian Information Criterion”).

Sea  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_m)^T \in \Theta$  un parámetro  $m$  dimensional y considere la familia

de fgp,  $\{\phi(t, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ . Kemp y Kemp (1988) propusieron varios métodos para estimar  $\boldsymbol{\theta}$ ; el más simple de estos es definir al parámetro estimado  $\hat{\boldsymbol{\theta}}_n$  como una solución del conjunto de  $m$  ecuaciones, dadas por

$$\phi_n(t_j) = \phi(t_j, \boldsymbol{\theta}), \quad 1 \leq j \leq m, \quad (3.25)$$

con  $t = (t_1, t_2, \dots, t_m)$  fijo y  $-1 \leq t_1 < t_2 < \dots < t_m \leq 1$ . Además, existe teoría asintótica para elegir  $t$  de manera óptima; para mayor detalle ver Dowling and Nakamura (1997).

Recordando que el modelo por planta es  $X_{n,i} = \alpha_n * X_{n-1,i} + Z_n(\mu_n, \theta_n, \alpha_n)$ ,  $i = 1, 2, \dots, 1000$  y  $n = 1, 2, \dots$  donde  $i$  y  $n$  representan el número de planta y el tiempo respectivamente. Los parámetros cambian día a día; entonces, éstos se tendrán que estimar con los datos por día que se tienen. Como  $\mu_n = -\beta_n \ln(\alpha_n)$ , nótese que estimar al parámetro  $\mu_n$  es equivalente a estimar  $\beta_n$ . Por lo tanto,

$$\begin{aligned} X_{0,i} &\sim \text{BinNeg}(\beta_0, \theta_0), \\ X_{1,i} &= \alpha_1 * X_{0,i} + Z_1(\beta_1, \alpha_1, \theta_1), \\ X_{2,i} &= \alpha_2 * X_{1,i} + Z_2(\beta_2, \alpha_2, \theta_2), \end{aligned}$$

y así sucesivamente.

Para estimar el parámetro  $\boldsymbol{\varphi}_n = (\beta_n, \theta_n, \alpha_n)$ , debemos resolver el sistema de ecuaciones

$$\begin{cases} \phi_{X_{n,i}}(t_1) = \frac{1}{m} \sum_{i=1}^m t_1^{x_{n,i}} \\ \phi_{X_{n,i}}(t_2) = \frac{1}{m} \sum_{i=1}^m t_2^{x_{n,i}} \\ \phi_{X_{n,i}}(t_3) = \frac{1}{m} \sum_{i=1}^m t_3^{x_{n,i}} \end{cases}, \quad (3.26)$$

con  $-1 < t_1, t_2, t_3 < 1$  fijos. Para ejemplificar, la fgp para la variable aleatoria  $X_{1,i}$  está dada por

$$\begin{aligned} \phi_{X_{1,i}}(t) &= E(t^{X_{1,i}}) \\ &= E\left(t^{\sum_{k=1}^{X_{0,i}} B_k(\alpha_1) + \sum_{j=1}^{N(\beta_1, \alpha_1)} \sum_{k=1}^{W_j(\theta_1)} B_k(\alpha_1^{R_i})}\right) \\ &= E\left(t^{\sum_{k=1}^{X_{0,i}} B_k(\alpha_1)}\right) E\left(t^{\sum_{j=1}^{N(\beta_1, \alpha_1)} \sum_{k=1}^{W_j(\theta_1)} B_k(\alpha_1^{R_i})}\right). \end{aligned} \quad (3.27)$$

Como  $\sum_{k=1}^{X_{0,i}} B_k(\alpha_1) | X_{0,i}$  es binomial con parámetros  $X_0$  y  $\alpha_1$ , entonces

$$E\left(t^{\sum_{k=1}^{X_{0,i}} B_k(\alpha_1)} | X_{0,i}\right) = (\alpha_1 t + (1 - \alpha_1))^{X_{0,i}}. \quad (3.28)$$

La v.a.  $X_{0,i} \sim \text{BinNeg}(\beta_0, \theta_0)$ , por lo tanto

$$\begin{aligned} E(t^{\sum_{k=1}^{X_{0,i}} B_k(\alpha_1)}) &= E(E(t^{\sum_{k=1}^{X_{0,i}} B_k(\alpha_1)} | X_{0,i})) \\ &= \phi_{X_{0,i}}(1 - \alpha_1 + \alpha_1 t) \\ &= \left[ \frac{1 - \theta_0}{1 - \theta_0(1 - \alpha_1 + \alpha_1 t)} \right]^{\beta_0}. \end{aligned} \quad (3.29)$$

Además de la Ecuación (3.24) la fgp para la v.a.  $Z_1$  esta dada por

$$E(t^{Z_1}) = \left[ \frac{1 - \theta_1(1 - \alpha_1 + \alpha_1 t)}{1 - \theta_1 t} \right]^{\beta_1}. \quad (3.30)$$

Por lo tanto, sustituyendo (3.29) y (3.30) en (3.27), se tiene

$$\phi_{X_{1,i}}(t) = \left[ \frac{1 - \theta_0}{1 - \theta_0(1 - \alpha_1 + \alpha_1 t)} \right]^{\beta_0} \left[ \frac{1 - \theta_1(1 - \alpha_1 + \alpha_1 t)}{1 - \theta_1 t} \right]^{\beta_1}. \quad (3.31)$$

Análogamente a los cálculos anteriores, se tiene que

$$\phi_{X_{n,i}}(t) = \phi_{X_{n-1,i}}(\alpha_n t + 1 - \alpha_n) \left[ \frac{1 - \theta_n(1 - \alpha_n + \alpha_n t)}{1 - \theta_n t} \right]^{\beta_n}. \quad (3.32)$$

Con la Ecuación (3.32) se procede a estimar los parámetros mediante la fgp, resolviendo el sistema de ecuaciones dado en la Ecuación (3.26) con  $t_1, t_2$  y  $t_3$  lo suficientemente separados para evitar multiplicidad en la solución. Como no existe una forma cerrada para resolver dicho sistema de ecuaciones, éste se resolvió mediante métodos numéricos. El método numérico se implementó en una función de Matlab que utiliza el método “trust-region dogled”<sup>1</sup>. Hay que tomar en cuenta, que si existen más de 2 raíces en el sistema, hay que elegir la que mejor se ajuste al modelo. En algunos casos este método tardó entre 30 y 50 minutos para resolver un sistema, pero para la mayoría la convergencia fue en promedio de 5 minutos. Las Figuras 3.1, 3.2 y 3.3 nos resumen los valores de los parámetros estimados.

### 3.2.1. Parámetros estimados y variables meteorológicas

En esta sección se presentan las gráficas que comparan los parámetros estimados y las variables meteorológicas, para indagar sobre una posible relación. Para ejemplificar los resultados, sólo mostraremos las gráficas de los parámetros estimados y la variable humedad (que es en donde mejor se observa la correlación que existe entre ambos).

---

<sup>1</sup>Como referencia ver el libro: Numerical Optimization Jorge Nocedal & Stephen J. Wright, Springer, pág. 71

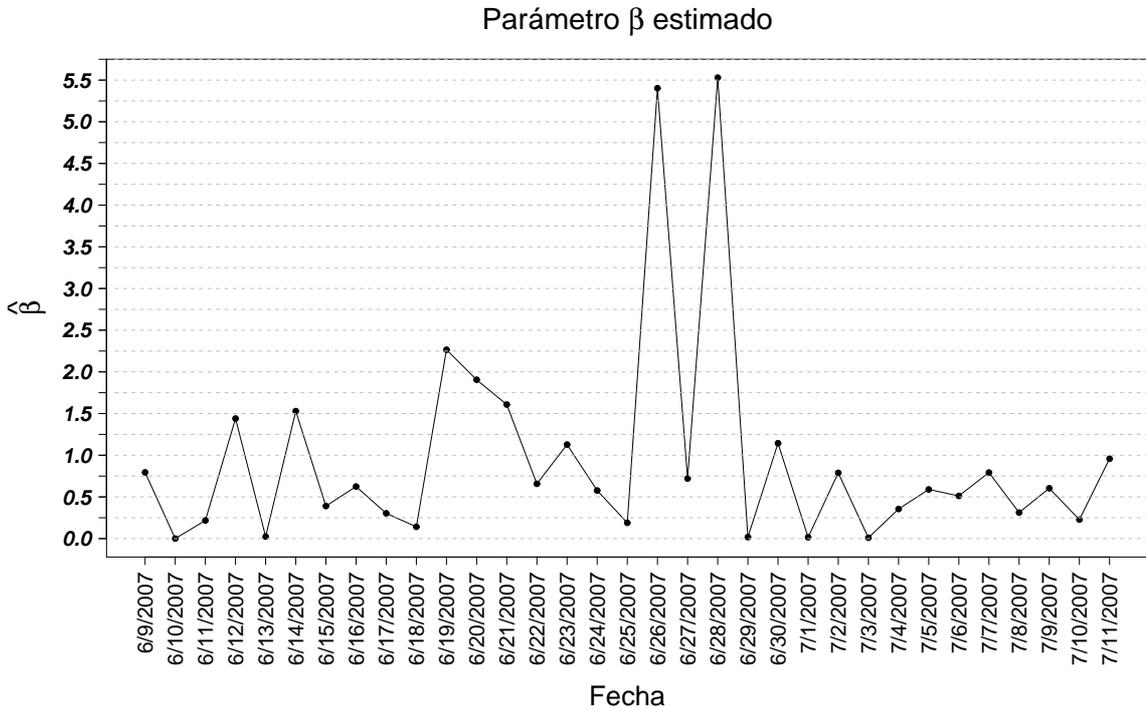


Figura 3.1: Valores estimados del parámetro  $\beta$  a través del tiempo.

Con base en las gráficas de la Figura 3.4 concluimos que sí hay una relación, ya que el cambio en los parámetros es muy parecido a la fluctuación de la variable humedad, es decir, cuando la humedad aumenta/disminuye también lo hacen los parámetros. Con las variables temperatura y precipitación no se ve muy clara la relación, pero las variables meteorológicas están correlacionadas. Por lo tanto, concluimos que sí existe una relación entre los parámetros del modelo y meteorología; sin embargo, no es explícita. Con lo anterior, se quiere decir que no sabemos si la relación entre los parámetros del modelo y la meteorología, es un polinomio de orden 3, 4 o algo más complejo. La inferencia formal para averiguar dicha relación, no se hace en esta tesis. Sin embargo, este análisis descriptivo de los datos ayuda a responder de manera cualitativa a la pregunta que planteó Tequila Sauza ¿cómo intervienen las variables meteorológicas en el número de insectos en el predio? La respuesta a dicha pregunta, es que efectivamente el clima es un factor importante en la emergencia de cerambícidos y que esta interviene en los parámetros del modelo que se esta proponiendo. Para incorporar implícitamente la meteorología al Modelo AR(1) BN general es necesario un estudio más profundo, con datos meteorológicos de mayor resolución y calidad.

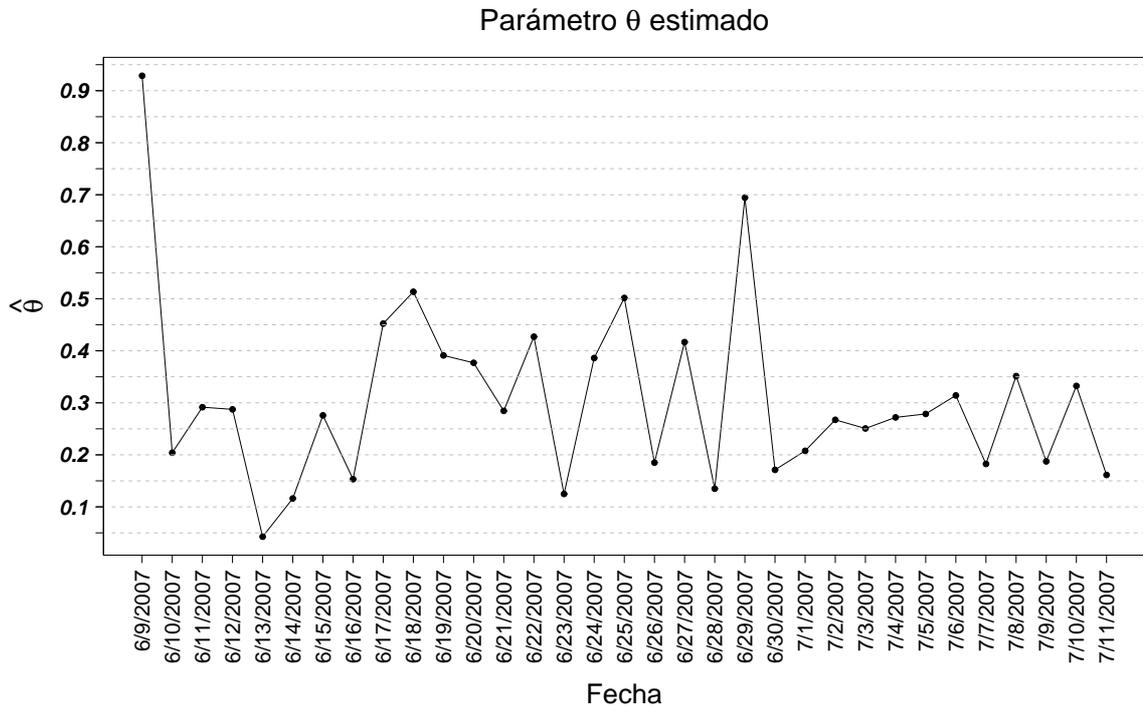


Figura 3.2: Valores estimados del parámetro  $\theta$  a través del tiempo.

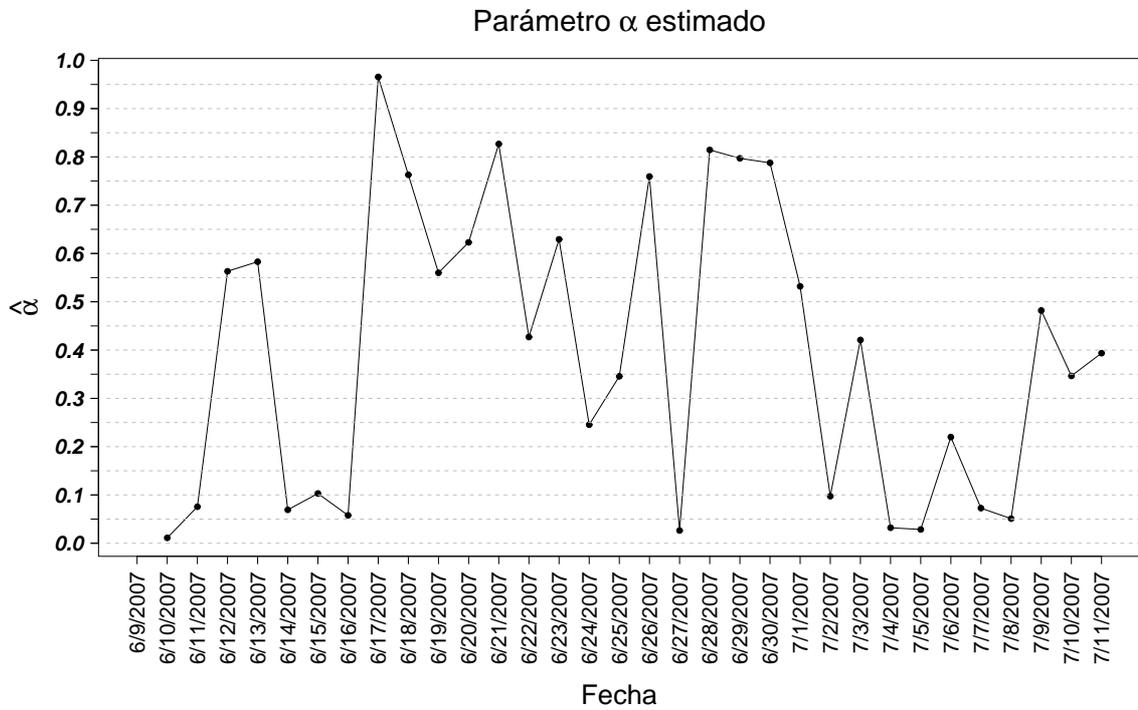


Figura 3.3: Valores estimados del parámetro  $\alpha$  a través del tiempo.

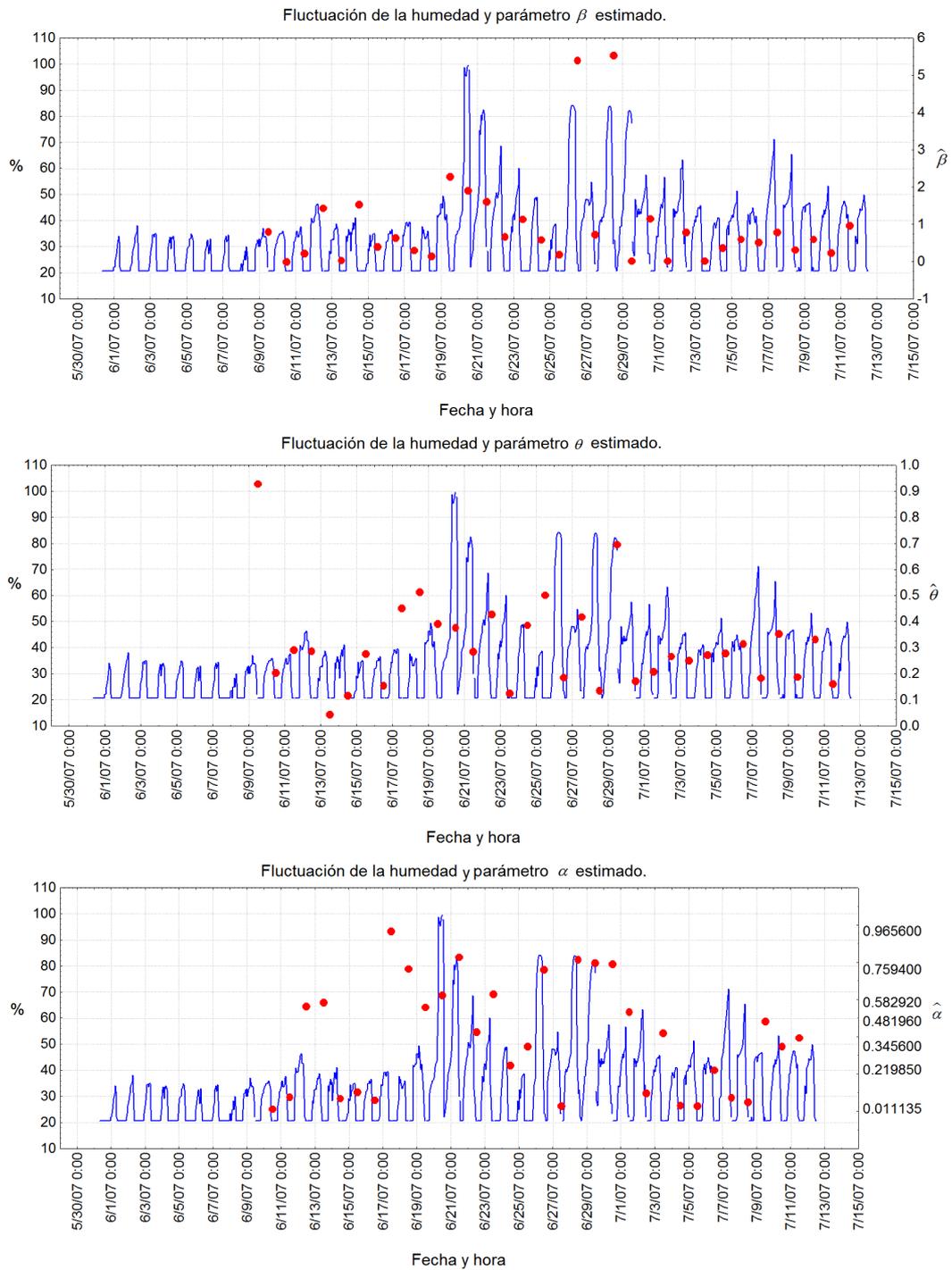


Figura 3.4: Fluctuación de la humedad y parámetros estimados.

### 3.3. Diagnóstico del Modelo AR(1) BN general

El instrumento de diagnóstico, que ayudara a ver el ajuste marginal del Modelo AR(1) BN general a los datos, es la comparación de la gráfica de la fgp estimada y la fgpe. El procedimiento anterior, es un método gráfico de bondad de ajuste que consiste en ver si la fgp y la fgpe (funciones de  $t$ ) se empalman en el intervalo  $(0, 1)$ ; si las gráficas de éstas funciones de  $t$  están muy distantes, entonces el ajuste no es bueno.

Este instrumento fue validado al observar qué sucedía si los datos por día provienen de una v.a. con distribución poisson o binomial negativa, por día estimamos sus parámetros y se simularon valores de éstas distribuciones, después se graficó la fgp estimada y la fgpe, observando que el instrumento da evidencia contundente para decir que los datos no provienen de una v.a. Poisson y que en algunos casos la distribución binomial negativa se ajusta bien, aunque esto ya se había observado en secciones anteriores.

Con la Figura 3.5 se concluye que el ajuste de los datos al Modelo AR(1) BN general es bastante bueno. Finalmente, para el diagnóstico en forma completa se simularon valores con este modelo y se compararon con los observados en campo. La simulación anterior recoge bien lo que se observa en campo. El ejemplo se muestra en la Figura 3.6.

Bondad de ajuste.

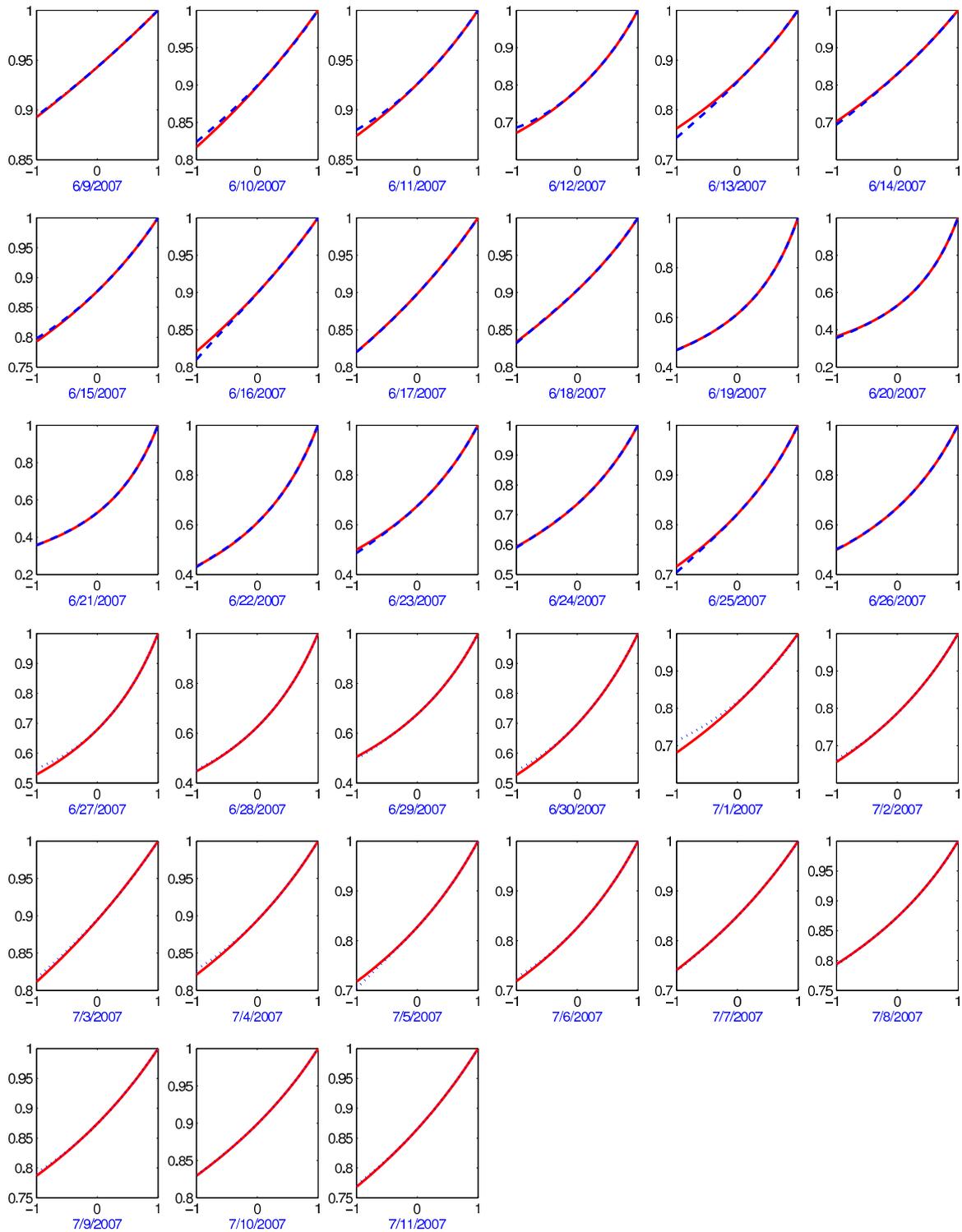
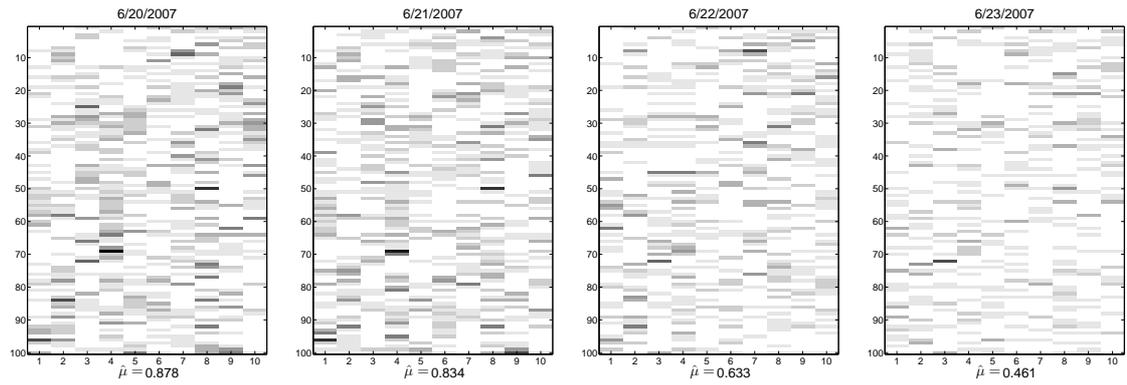


Figura 3.5: Bondad de Ajuste del Modelo AR(1) BN General a los datos del muestreo 2007, fgp estimada (—) y fgpe (---).

Datos observados:



Datos simulados:

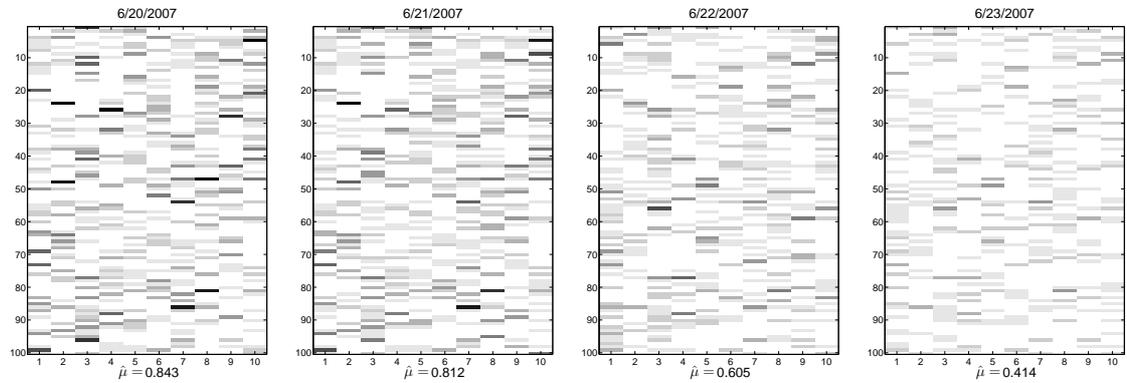


Figura 3.6: Ejemplo: Comparación de datos observados con simulados del modelo AR(1) BN General.

---

## Capítulo 4

# Resumen de resultados

---

### 4.1. Conclusiones

Mediante una batería de técnicas, incluyendo depuración y organización de datos, análisis de datos y modelación, implementaciones gráficas y numéricas, se fundamentaron respuestas a las preguntas planteadas por Tequila Sauza. Estas preguntas tienen que ver con entender cómo la plaga denominada cerambícido *Acanthoderes funeraria* Bates ocupa el predio en tiempo y espacio. Las preguntas que planteó Tequila Sauza por medio del Dr. del Real Laborde y el M.C. Rubio Cortés, fueron:

1. ¿Forman los cerambícidos cúmulos en ciertas regiones del predio?
2. ¿Hay evidencia estadística que sustente el hecho de que los insectos emergen en temporada de lluvia?
3. Si la respuesta a la pregunta anterior es sí, entonces—con mayor precisión—¿cómo es que las variables meteorológicas intervienen en el número de insectos en el predio?
4. ¿Existe algún tipo de agregación temporal de los insectos?

Los principales productos de esta tesis son los siguientes:

- La propuesta de un diseño de muestreo para estimar la proporción de hembras (o cualquier otra cualidad binaria de interés en el predio por día). La relevancia de este resultado es que se podría relacionar con el número de larvas presentes, con una logística de muestreo que economiza el número de evaluaciones de la cualidad binaria dado que se realiza muestreo aleatorio simple de plantas en el predio. Las implicaciones en términos de control de plagas son inmediatas: Teniendo un estimado del número de larvas, se estimaría el número de adultos plausibles para la siguiente emergencia de insectos. Lo anterior ayudaría a estimar el material necesario para el control de este insecto, ya sea material de fumigación o biológico, para evitar que los nuevos cerambícidos adultos logren

ovipositar; ya que las larvas de éste insecto son las causantes de los daños más graves al agave. Además, se podría investigar la fluctuación de la proporción de hembras con respecto a las variables meteorológicas para determinar una posible relación, con lo cual se puede predecir el número de hembras y huevecillos para algunas fechas posteriores.

- Análisis exploratorio de datos. Como consecuencia de diversas técnicas gráficas y numéricas de exploración de datos, se obtuvieron varias conclusiones de tipo cualitativo respecto al fenómeno cerambícidos. Por ejemplo, las gráficas del segundo muestreo donde se comparan a las variables meteorológicas: humedad relativa, temperatura y precipitación pluvial con los datos observados del número de cerambícidos en el predio (v.a. llamada conteo), descritas en la Sección 2.2.2.3 dan evidencia contundente para afirmar que la emergencia de insectos se da en temporada de lluvias. La variable meteorológica que aporta más información es la humedad relativa, pero a su vez ésta está correlacionada con la temperatura y precipitación pluvial.
- Modelación probabilística del número de insectos. El resultado más exhaustivo desde un punto de vista técnico es la propuesta de un modelo probabilístico llamado Modelo AR(1) BN general, el cual sirve para describir los datos temporales y espaciales sobre todo el predio. El modelo se describe con detalle en la Sección 3.1. Para dicho Modelo AR(1) BN general, se da la interpretación biológica/ambiental de cada variable aleatoria y parámetro involucrado. Consecuente con los resultados del análisis exploratorio, el modelo sí prescribe agregación temporal y no lo hace en cuanto a agregación espacial. En lo relativo a variables meteorológicas, el modelo especifica cómo y dónde tienen efecto. Así, prepara el camino para estudios posteriores en los que las variables meteorológicas se registren con mayor resolución y calidad. El trabajo incluyó metodología para estimar los parámetros del Modelo AR(1) BN general vía la función generadora de probabilidades empírica. Teniendo la relación entre variables meteorológicas y parámetros del modelo se podría predecir la cantidad de insectos en el predio para fechas posteriores. Además, extrapolar los resultados a otros predios. Como observación, la modelación depende de la ventana de observación de los datos. Es decir, si solamente se hubieran tenido los datos de primer muestreo (con ventana de 5 días), entonces la modelación hubiese sido más simple (no correlación espacial, ni correlación temporal).

Las consecuencias que el desarrollo de este trabajo tienen para Tequila Sauza son las siguientes:

1. El desarrollo de un diseño de muestreo específico para calcular la proporción de hembras u otra cualidad binaria de interés para Tequila Sauza, llamado muestreo de muestreo. Incluye logística y metodología de análisis de datos.

2. Técnicas gráficas y numéricas para explorar la relación cualitativa entre variables y emergencia de insectos.
3. Descripción probabilística del fenómeno aleatorio de infestación de la plaga denominada cerambícidos en cultivos de agave, llamado Modelo AR(1) BN general para entender mejor a este insecto y poder aplicar técnicas de control que les sean adecuadas.
4. Con base en el modelo anterior, el número de cerambícidos presentes en el predio para un día determinado, depende del número de insectos del día anterior. Además, el número de cerambícidos por planta es independiente de los que tienen las demás.
5. Los cerambícidos no tienen preferencias por zonas del predio.
6. La sugerencia de integrar los datos mediante la matriz propuesta en la Sección 2.1.2 para la recolección de datos en campo, para estudios futuros.
7. La sugerencia de aplicar el modo de muestreo dado en el apéndice A con su plantilla de recolección.
8. El apoyo para realizar un estudio futuro para determinar explícitamente o estimar la relación que existe entre los parámetros del Modelo AR(1) BN general y las variables meteorológicas.

---

## Apéndice A

# Muestreo de muestreo

---

El propósito de esta sección es proponer un diseño de muestreo que permita estimar la proporción  $p$  de hembras en el predio, o cualquier otra cualidad binaria de interés para Tequila Sauza en un día determinado. Para economizar el número de evaluaciones de la cualidad binaria, se propone cambiar la forma de muestreo que se estaba llevando al cabo (muestreo sistemático), por uno completamente aleatorio sin reemplazo; cada agave tiene la misma probabilidad de ser seleccionado, para permitir el análisis que a continuación se describe. El procedimiento de selección de este diseño de muestreo es:

1. Elegir una planta del predio mediante muestreo completamente aleatorio sin reemplazo.
2. Si la planta tiene individuos, se realizará un ensayo de Bernoulli con probabilidad de éxito igual a  $q$ . Si el resultado del ensayo es éxito, se recolectarán los insectos para sexarlos y determinar el número de hembras. De lo contrario, no se realiza la determinación de sexo de los insectos encontrados.

El efecto que tiene el parámetro  $q$  es controlar el número de determinaciones de sexo que se deben realizar, lo cual se asume es mucho más difícil que contar el número de insectos en total que tiene una planta. La idea es que los datos de conteo (que de cualquier forma se obtendrán aunque no se tenga la intención de estimar la proporción de hembras) se utilicen para estimar el número de insectos, y que los datos de sexo se utilicen para estimar la cantidad  $p$ .

Para obtener los datos en campo se tienen dos opciones:

- (I) Realizar los ensayos de Bernoulli en campo. Es decir, si el agave en observación tiene individuos se realiza el ensayo de Bernoulli con ayuda de una calculadora/computadora, para saber si hay o no que sexar a los insectos.
- (II) Realizar los ensayos de Bernoulli antes de salir a campo. Es decir, llevar una planilla de recolección que contenga los agaves a mostrar, y además si debemos o no recolectar a los insectos para sexar; sin saber de antemano si el agave tiene o no insectos.

Las dos formas anteriores de recolección de datos son equivalentes. Dada la incomodidad de llevar una calculadora/computadora a campo, la segunda opción es más viable; ésta se implementó en una hoja de cálculo en Excel (la hoja para la logística se muestra en la Tabla B.2). Para demostrar la equivalencia entre los procedimientos (I) y (II), se verifica que la probabilidad de éxito del ensayo de Bernoulli dado que la planta tiene individuos, es la misma en ambos casos. Se definen las variables aleatorias

$Z$  = resultado del ensayo de Bernoulli ( $Z = 1$  es éxito, y  $Z = 0$  es fracaso),

y

$X$  = el número de individuos en la planta.

Bajo el esquema (I), se tiene que

$$q = P(Z = 1 \mid X > 0). \quad (\text{A.1})$$

La probabilidad  $P(Z = 1 \mid X > 0)$  bajo el esquema (II) está dada por

$$\begin{aligned} P(Z = 1 \mid X > 0) &= \\ &= \frac{P(Z = 1)P(X > 0 \mid Z = 1)}{P(Z = 1)P(X > 0 \mid Z = 1) + P(Z = 0)P(X > 0 \mid Z = 0)} \\ &= \frac{qP(X > 0)}{qP(X > 0) + (1 - q)P(X > 0)} \\ &= \frac{qP(X > 0)}{qP(X > 0) + P(X > 0) - qP(X > 0)} \\ &= q. \end{aligned} \quad (\text{A.2})$$

Por el contexto, la variable aleatoria  $X$  es independiente de la variable aleatoria  $Z$ ; esto se utilizó en la segunda igualdad de la Ecuación (A.2). Con el resultado anterior, queda demostrada la aseveración de que los procedimientos (I) y (II) son equivalentes.

En cuanto al planteamiento del problema estadístico de estimación del parámetro  $p$ , la manera de proceder es determinando la función de verosimilitud con los datos que se obtienen del muestreo. La estimación del parámetro puede basarse en esta verosimilitud. Los datos consisten de observaciones de las siguientes variables aleatorias:

$X_{n,j}$  = número de insectos en la planta  $j$  al tiempo  $n$ , y

$Y_{n,j}$  = número de hembras en la planta  $j$  al tiempo  $n$ .

Debe notarse que en caso de que  $X_{n,j} = 0$ , entonces  $Y_{n,j}$  no es observado. Para ayudar a determinar la función de verosimilitud de la muestra de la variable aleatoria  $Y_{n,j}$ , se obtuvo el árbol de probabilidades de la Figura A.

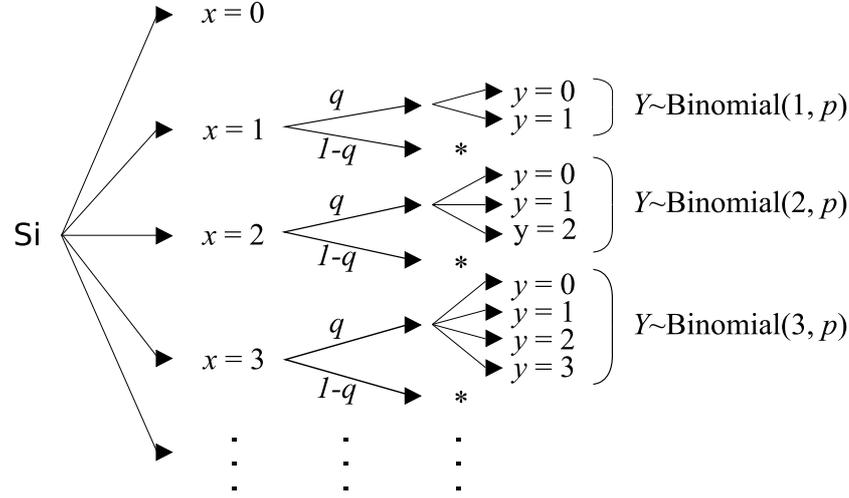


Figura A.1: Árbol de probabilidades.

Con base en los resultados del análisis exploratorio del segundo muestreo y el Modelo AR(1) BN general, existen dos formas de aproximar  $P(X_{n,j} = i)$ , las cuales son las siguientes:

- Adoptar la aproximación binomial negativa por día. Es decir, aproximar  $p_{n,i} = P(X_{n,j} = i)$  con una variable aleatoria binomial negativa.
- Adoptar el Modelo AR(1) BN general. Como en este modelo la distribución está especificada por vía de funciones generatrices de probabilidad, las densidades necesarias para escribir la función de verosimilitud se obtienen como

$$p_{n,i} = P(X_{n,j} = i) = \frac{\phi_{X_{n,j}}^{(i)}(0)}{i!}, \quad (\text{A.3})$$

donde  $\phi_{X_{n,j}}^{(i)}(0)$  representa la  $i$ -ésima derivada de la fgp de  $X_{n,j}$  evaluada en 0.

Los dos procedimientos anteriores para obtener  $p_{n,i}$ , deben dar resultados parecidos; sin embargo, el segundo es más laborioso computacionalmente. Para evitar excesiva notación se escribirá a continuación  $X = X_{n,j}$  y  $Y = Y_{n,j}$ .

Bajo este esquema sea  $Y_1, Y_2, \dots, Y_k$  una muestra aleatoria de conteos de hembras en un día  $n$ . Entonces

$$P(Y = *) = 1 - \sum_{y=0}^{\infty} P(Y = y). \quad (\text{A.4})$$

Observando que la variable aleatoria  $Y|X = x \sim \text{Binomial}(x, p)$ , y utilizando el árbol de probabilidades, se tiene que

$$P(Y = y) = \sum_{x=y}^{\infty} \binom{x}{y} p^y (1-p)^{x-y} q p_{n,x}. \quad (\text{A.5})$$

Por lo tanto la verosimilitud está dada por

$$L(p; \{y_i\}_{i=1}^n) = \left\{ 1 - \sum_{y=0}^{\infty} P(Y = y) \right\}^m \prod_{y_j \text{ observado}} P(Y = y_j), \quad (\text{A.6})$$

donde  $m$  es el número de datos no observados. Mediante la Ecuación (A.6) puede realizarse estimación del parámetro de interés  $p$ , por el principio de máxima verosimilitud.

El análisis anterior se basa solamente en las observaciones de la variable aleatoria  $Y$ , pero debido a la logística propuesta de muestreo también se conocen las observaciones de la variable aleatoria  $X$ . Una mejora al análisis anterior es considerar que la muestra aleatoria es de la siguiente forma:

$$(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n), \text{ donde } Y_i = \begin{cases} (*) \text{ faltante} \\ \text{observado} \end{cases}.$$

Haciendo uso nuevamente del árbol de probabilidades, en la siguiente tabla se observa el espacio muestral de la nueva variable aleatoria  $(X, Y)$ , así como la probabilidad de cada evento.

$(0, *)$ $p_{n,0}$			
$(1, *)$ $p_{n,1}(1-q)$	$(1, 0)$ $p_{n,1}qb_0$	$(1, 1)$ $p_{n,1}qb_1$	
$(2, *)$ $p_{n,2}(1-q)$	$(2, 0)$ $p_{n,2}qb_0$	$(2, 1)$ $p_{n,2}qb_1$	$(2, 2)$ $p_{n,2}qb_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Tabla A.1: Espacio muestral con probabilidades de la variable aleatoria  $(X, Y)$ .

En esta Tabla A.1,  $b_y = \binom{x}{y} p^y (1-p)^{x-y}$ ,  $y = 0, 1, 2, \dots, x$ . Bajo este esquema la

función de verosimilitud está dada por

$$L(p; \{(x_i, y_i)\}_{i=1}^n) = \prod_{\text{faltantes}} p_{n,x}(1 - q) \prod_{\text{observados}} p_{n,x}qb_y. \quad (\text{A.7})$$

Al igual que en el caso anterior, con la Ecuación (A.7) se estimaría al parámetro  $p$ . La proporción de hembras en el predio es de interés para predecir el número de huevecillos ovipositados y cuántos adultos habrá en el siguiente período. Con ello que se pueden determinar controles (biológicos o fumigaciones) adecuados, como el material necesario para éstos.

## Apéndice B

# Tablas

Fecha: 5/27/2006		Monitoreo de fluctuación poblacional en el predio: El Triángulo de Amatitán, Jal.					
SURCO 1	No. Individuos	SURCO 2	No. Individuos	SURCO 3	No. Individuos	SURCO 4	No. Individuos
19	1	26	0	11	0	19	1
54	1	40	0	32	0	23	0
59	1	48	0	41	0	27	0
63	0	99	0	50	0	29	0
64	0	112	0	57	1	47	0
73	0	116	0	58	0	103	0
79	0	118	2	88	0	113	0
112	0	124	0	107	0	122	0
114	1	125	0	108	1	125	0
141	1	126	0	121	0	135	0
<b>Total</b>	<b>5</b>		<b>2</b>		<b>2</b>		<b>1</b>

SURCO 5	No. Individuos	SURCO 6	No. Individuos	SURCO 7	No. Individuos	SURCO 8	No. Individuos
16	0	2	0	12	0	5	2
20	0	12	1	31	0	23	0
46	0	30	0	36	0	37	0
49	0	42	0	50	0	90	2
57	0	47	0	57	0	100	1
102	0	48	0	65	0	104	0
104	0	58	0	97	0	114	1
117	0	63	0	115	0	125	0
124	0	93	0	138	0	131	1
145	2	124	1	145	0	144	0
<b>Total</b>	<b>2</b>		<b>2</b>		<b>0</b>		<b>7</b>

SURCO 9	No. Individuos	SURCO 10	No. Individuos	SURCO 11	No. Individuos	SURCO 12	No. Individuos
7	0	7	0	30	1	10	0
19	0	20	0	32	0	35	1
30	1	32	2	48	0	45	0
56	0	33	0	69	0	60	0
79	0	37	0	87	0	88	1
81	0	49	0	104	0	102	1
117	0	63	0	113	0	107	0
122	0	103	0	118	0	123	0
134	0	105	3	134	1	130	0
137	0	138	0	137	0	133	1
<b>Total</b>	<b>1</b>		<b>5</b>		<b>2</b>		<b>4</b>

SURCO 13	No. Individuos	SURCO 14	No. Individuos	SURCO 15	No. Individuos
15	0	24	0	21	0
30	0	38	0	44	0
65	0	39	0	62	0
92	0	70	1	63	0
97	0	85	0	70	0
98	0	109	0	86	0
102	0	123	0	91	4
107	0	135	0	92	1
141	0	138	1	100	0
149	1	144	0	122	0
<b>Total</b>	<b>1</b>		<b>2</b>		<b>5</b>

**TOTAL** 41

Plantas muestreadas 150

Nota: Cuatro días antes del muestreo llovió

Tabla B.1: Ejemplo de recopilación de datos, Tequila Sauza.

Fecha (mes/día/año):  Hora:  No. de Muestreo:   $q :=$

Surco	Planta	D	No. de Individuos	Surco	Planta	D	No. de Individuos	Surco	Planta	D	No. de Individuos	Surco	Planta	D	No. de Individuos
1	30	No		3	23	No		4	134	No		5	104	No	
1	41	No		3	28	Si		4	137	Si		5	120	No	
1	44	Si		3	44	No		4	142	Si		5	127	Si	
1	45	Si		3	66	No		4	148	Si		5	132	No	
1	53	No		3	81	Si		5	13	Si		5	137	No	
1	56	Si		3	103	No		5	16	Si		6	2	No	
1	75	No		4	24	Si		5	25	Si		6	4	No	
1	111	No		4	37	No		5	40	Si		6	20	No	
1	132	Si		4	48	No		5	43	Si		6	33	Si	
2	20	Si		4	52	Si		5	51	No		6	37	Si	
2	31	No		4	62	No		5	53	Si		6	52	No	
2	106	Si		4	64	Si		5	64	No		6	54	No	
2	138	No		4	65	No		5	96	No		6	58	No	
2	145	No		4	85	Si		5	98	No		6	63	No	
3	16	No		4	98	No		5	101	Si		6	130	No	

Surco	Planta	D	No. de Individuos	Surco	Planta	D	No. de Individuos	Surco	Planta	D	No. de Individuos	Surco	Planta	D	No. de Individuos
8	33	No		10	1	No		11	112	Si		12	94	No	
8	36	Si		10	12	No		11	125	No		12	107	Si	
8	47	No		10	19	No		11	132	Si		12	109	Si	
8	75	Si		10	46	No		11	142	Si		12	140	Si	
8	89	Si		10	55	Si		12	30	No		12	149	Si	
8	97	No		10	82	Si		12	45	Si		13	30	No	
8	99	No		10	113	No		12	55	Si		13	46	Si	
8	128	Si		10	115	No		12	64	No		13	71	Si	
8	131	No		10	134	No		12	74	Si		13	74	No	
9	43	Si		11	42	Si		12	77	Si		13	106	Si	
9	67	Si		11	51	Si		12	79	No		13	121	No	
9	89	No		11	55	No		12	80	Si		13	131	No	
9	110	No		11	56	Si		12	85	Si		14	2	Si	
9	118	No		11	96	No		12	87	Si		14	9	Si	
9	133	No		11	104	Si		12	88	No		14	33	No	

Nota.  $q$  es la probabilidad de censar una submuestra positiva, para identificar el número de hembras.  
 Si el No. de Individuos en el Agave es una cantidad positiva y en la casilla D aparece un si, entonces se procederá a llevar al laboratorio los individuos para determinar su sexo, en otro caso ignorar la casilla D.

Tabla B.2: Hoja para recolectar los datos en campo, implemento del tema “Muestreo de muestreo”.

# Bibliografía

---

- [1] Chemsak, J. A., Linsley, E. G., Noguera, F. A.(1992). Listados faunísticos de México. II. Los Cerambycidae y Disteniidae de Norteamérica, Centroamérica y las Indias Occidentales (Coleóptera). *Universidad Nacional Autónoma de México*.
- [2] Héctor González H., José Ignacio del Real L., Juan F. Solis A. (2004). Manejo de Plagas del agave Tequilero. *Colegio de Postgraduados Tequila Sauza, S.A. de C.V.*
- [3] Consejo Regulador del Tequila, A.C., Comité Técnico Agronómico (2005). Plagas y Enfermedades del **agave Tequilana** Weber var. azul. (*Subcomité de investigación*)
- [4] Anscombe, F. J. (1949). The statistical analysis of insect counts based on the negative binomial distribution. *Biometrics* 5, 165-173.
- [5] Waters, W. E. (1959). A quantitative measure of aggregation in insects. *Journal of Economic Entomology* 52, 1180-1184
- [6] Clapham, A. R. (1936). Over-dispersion in grassland communities and the use of statistical methods in plant ecology. *Journal of Ecology*, 24, pp.232-251.
- [7] Ross, G. J. S., Preece, D. A. The Negative Binomial Distribution. *The Statistician*, Vol. 34, No. 3. (1985), pp. 323-335.
- [8] Mary M. Dowling and Miguel Nakamura (1997). Estimating parameters for discrete distributions via the empirical probability generating function. *Commun. Statist.-Simula.*, 26(1), pp.301-313.
- [9] Kemp, C. D., and Kemp, A. W. (1988), Rapid estimation for discrete distributions. *The Statistician* 37, pp. 243-255.

## BIBLIOGRAFÍA

---

- [10] Marques, M. and Pérez-Abreu, V. (1989), Law of large numbers and central limit theorem of the empirical probability generating function of stationary random sequences and processes. *Aportaciones Matemáticas, Sociedad Matemática Mexicana* 4, 100-109.
- [11] Feuerverger, A. and McDunnough, P. (1984), On statistical transform methods and their efficiency. *Canadian Journal of Statistics* 12, 303-317.
- [12] Cox, D. R. (1983). Some Remarks on Overdispersion. *Biometrika* 70, 269-274.
- [13] Steutel, F. W. and Van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *Ann. Prob.* 7, 893-899.