



Universidad Nacional del Comahue Centro Regional Universitario Bariloche

Modelos jerárquicos predictivos para la dinámica poblacional de pentatómidos en cultivos de soja

Trabajo Final para optar por el título de Licenciado/a en Ciencias Biológicas

Francis Felici

Director: Juan Manuel Morales

Codirector: Juan Manuel Gurevitz

INIBIOMA

2021

Resumen. En las últimas décadas, la intensificación de la producción agrícola, acompañada de una mayor presión por plagas, resultó en un aumento sustancial del uso de pesticidas sintéticos. El manejo integrado de plagas (MIP) brinda un marco para el desarrollo y uso de estrategias de control que sean efectivas y a la vez sustentables. Promueve el uso de procesos biológicos renovables para el control de plagas, el monitoreo de su dinámica poblacional y el uso de herramientas de decisión de control, tales como umbrales de daño y modelos predictivos. En este trabajo, se propone un enfoque de modelado empírico basado en un modelo bayesiano jerárquico con estructura espacio-estado para aumentar la eficacia y la eficiencia de los controles basados en el MIP y aumentar así su adopción por el sector productivo. Utilizando datos de abundancia de chinches (Pentatomidae) y de fenología del cultivo de soja, junto con datos meteorológicos para 8 localidades de la región pampeana, (1) se estimó el tamaño poblacional *real* de chinches través del tiempo, (2) se realizaron predicciones a corto plazo del tamaño poblacional, (3) se creó un método para estimar el daño acumulado por parte de las chinches que presentaban los cultivos, (4) se realizó un análisis de incertidumbre en las predicciones en función del tamaño muestral y (5) se evaluó la capacidad predictiva de distintas variantes del modelo con el criterio de información WAIC y el método de validación cruzada LOO-CV. Las predicciones mostraron un grado de exactitud razonable y la estimación del daño acumulado permitió aumentar la información disponible para los productores acerca de cuándo realizar un control. Por último, el análisis de incertidumbre sugirió un tamaño muestral de 60 para obtener un buen balance entre la precisión y el esfuerzo de muestreo.

Palabras clave: *ecología de poblaciones, espacio-estado, estadística bayesiana, manejo integrado de plagas.*

Abstract. In the last few decades, the intensification of agricultural production accompanied by an increasing pressure from pests in various crops, is resulting in a substantial increase in the use of synthetic pesticides. Integrated pest management (IPM) provides a framework for the development and use of sustainable yet effective control strategies. This framework promotes the use of renewable biological processes for the control of pests, the monitoring of pest population dynamics, and the use of control decision tools, such as action thresholds and predictive models. In this work, an empirical modeling approach based on a hierarchical Bayesian model with a space-state structure is developed with the aim of increasing the efficacy and efficiency of MIP-based pest control tactics and thus favouring their adoption by stakeholders. Using data of stinkbug (Pentatomidae) abundance and crop phenology, along with meteorological data from 8 different locations in Argentina we, (1) estimated the true population size of stink bugs through time, (2) made short-term forecasts of population size, (3) developed a method to estimate the cumulative damage on crops caused by stink bug populations, (4) we analyzed how the uncertainty of the predictions vary as a function of sample size and (5) we evaluated the predictive capacity of different variants of the model using WAIC information criterion and LOO-CV cross-validation. The forecasts showed a reasonable degree of accuracy and the cumulative damage estimation can increase the information available to stakeholders about when to carry out a given control measure. Finally, the uncertainty analysis suggests a sample size of 60 to obtain a good balance between precision and sampling effort.

Keywords: *bayesian statistics, population ecology, integrated pest management, state-space.*

ÍNDICE

INTRODUCCIÓN	4
Impactos negativos de pesticidas	6
Manejo Integrado de Plagas	6
Problemas de aplicación y de adopción del MIP	7
OBJETIVOS	10
METODOLOGÍA	11
Datos	11
Modelo espacio-estado	14
Estructura jerárquica	19
Ajuste	20
Simulaciones predictivas previa y posterior	20
Variantes del modelo y comparación en términos de poder predictivo	21
Predicciones	23
Estimaciones de daño acumulado sobre los cultivos	23
Incertidumbre en las predicciones en función del tamaño muestral	25
RESULTADOS	25
DISCUSIÓN	38
CONCLUSIÓN	42
BIBLIOGRAFÍA	43
APÉNDICE	43
Simulación predictiva previa	47
Simulación predictiva posterior	49
Posterior de parámetros	51

INTRODUCCIÓN

En las últimas décadas, ha ocurrido un proceso de intensificación en la producción agrícola. Un resultado de esta intensificación, es el uso generalizado de cultivos genéticamente homogéneos, la utilización a grandes escalas de monocultivos, la utilización de variedades altamente productivas y el incremento de la densidad de plantas, entre otras (Dent and Binks 2020). Esta intensificación vino acompañada de una mayor presión por parte de diversas plagas, lo que ha dado lugar a un mayor uso de pesticidas químicos de origen sintético. A nivel mundial, entre 1990 y 2009 ha habido un incremento del 1,8% en el uso de pesticidas por cada 1% de incremento en producción agrícola por hectárea, con un efecto más acentuado en los países menos desarrollados (Schreinemachers and Tipraqsa 2012). Esto ha hecho que el panorama fitosanitario de los cultivos se torne insumo-dependiente (Gamundi and Sosa 2008).

En Argentina, el cultivo de soja (*Glycine max*) se ha ido expandiendo considerablemente a partir de la década del 2000 y ahora es el cultivo con mayor superficie sembrada en el país (<http://datosestimaciones.magyp.gob.ar>). La soja es una leguminosa que se siembra entre septiembre y enero (según el ambiente en donde se realiza el cultivo y el período de lluvias) y es de gran interés económico por su contenido de aceite y de proteína. Una de las principales plagas que afectan a este cultivo son los hemípteros fitófagos de la familia Pentatomidae (Hemiptera), comúnmente conocidos como chinches. Estos insectos son hemimetábolos, poseen alas en su estadio adulto y pasan el invierno debajo de la cubierta del suelo o de la hojarasca. Al comienzo de la primavera, los individuos que sobrevivieron al invierno suelen encontrarse en especies arbóreas en los bordes de los cultivos y migran a los cultivos a medida que se van desarrollando (Swanson 2012). En los cultivos, las hembras depositan cúmulos de huevos debajo de las hojas. Las ninfas de estadio I-III poseen un hábito gregario y una vez que alcanzan el estadio IV se cree que pueden desplazarse hasta 7-10 m del lugar de oviposición (Panizzi *et al.* 1980).

Si bien existen numerosas especies de pentatómidos en el cultivo de soja, solo un reducido subconjunto es de marcada relevancia en la producción agrícola argentina debido a su frecuencia e impacto en el rendimiento. Estas especies son *Piezodorus guildinii*, *Nezara viridula*, *Edessa meditabunda* y *Dichelops furcatus* (Gamundi 2007).

Tanto ninfas como adultos de las chinches, obtienen su alimento mediante una acción picadora-suctora del tejido vegetal, inyectando enzimas digestivas que licúan las porciones sólidas de la célula. Si bien pueden alimentarse de distintas partes de la planta (tallos, follaje, flores, vainas y semillas), muestran una marcada preferencia por vainas y semillas una vez que el cultivo alcanzó la etapa reproductiva. Entre los impactos que tienen las chinches sobre los cultivos de soja se encuentran: el aborto o deformación de los granos; mayor vulnerabilidad al ataque de microorganismos patógenos; madurez demorada del cultivo; granos con menor tasa de germinación, emergencia y supervivencia (Vyavhare *et al.* 2015); retención de hojas más allá del estado fenológico de madurez completa; alteración de la relación proteína/lípido de los granos (aumenta el contenido de proteína y disminuye el contenido de aceite) (Gamundi y Sosa 2008), y deterioro de la calidad del aceite (Todd *et al.* 1973), entre otros.

El daño causado al cultivo varía según la especie de chinche, su estadio y la fenología del cultivo. Por ejemplo, entre las especies que causan mayor impacto en la soja en la Argentina, *P. guildinii* presenta el mayor daño, mientras que *D. furcatus* representa el menor (Vicentini and Jimenez 1977). Esto se atribuye a diferencias en las enzimas salivares (Depieri and Panizzi 2011). El daño que provocan las ninfas de cuarto y quinto estadio es similar al de los adultos (aproximadamente 10 picaduras por día), mientras que el primer estadio ninfal no se alimenta (Bowling 1980).

En lo que respecta a la fenología del cultivo, existen algunas discrepancias entre autores respecto al periodo crítico de daño. Si bien todos concuerdan que este periodo se encuentra en la etapa reproductiva, algunos indican que las fenologías vegetales R3-R4 (R3: comienzo de fructificación. R4: plena fructificación) son las que presentan mayor vulnerabilidad al daño de chinches (Yeargan 1977; Iannone 2004), mientras que otros sugieren las fenologías R4-R5.5 (R5.5: vainas con 50% de granos) (Correa-Ferreira and Panizzi 1999; Brier and Rogers 1991).

Particularmente en la Argentina, en las últimas décadas ha habido cambios en la dinámica poblacional de las plagas de la soja, principalmente debido a la utilización de cultivares de ciclo corto (aquellos con un umbral fotoperiódico mayor) y crecimiento indeterminado (el crecimiento de los nudos continúa una vez que comienza la floración), siembras más tempranas, menores espaciamientos entre líneas de

siembra y un marcado incremento del área sembrada que creó un corredor continuo de soja de madurez escalonada a lo largo del país. Estos cambios en las técnicas de producción agrícola han provocado que los cultivos sean más sensibles al ataque de las chinches y que la calidad de la aplicación de los pesticidas sea menor (Gamundi 2007), lo que resulta en un uso aún mayor de estos compuestos.

Impactos negativos de pesticidas

El uso indiscriminado de muchos de los pesticidas sintéticos está asociado a impactos negativos como son la degradación del ambiente, la generación de resistencia a estos compuestos, impactos negativos sobre enemigos naturales de las plagas y menor protección en términos de salud para los trabajadores que realizan las aplicaciones de pesticidas (Altieri and Nicholls 2000; Boller *et al.* 1999). Estas técnicas entran en conflicto con una actividad agrícola sustentable que supone que estos sistemas deben satisfacer las demandas productivas sin comprometer la capacidad de las generaciones futuras de satisfacer las suyas (Lichtfouse *et al.* 2009). La preocupación por las consecuencias de una producción agrícola no sustentable dio lugar a un mayor interés por medidas alternativas para el control de plagas que causen poco o ningún impacto en el ambiente, en los organismos benéficos para los cultivos y en los humanos (Dent and Binks 2020; Boller *et al.* 1999). Esto ha dado lugar al comienzo de una transición hacia un modelo de manejo de plagas denominado Manejo Integrado de Plagas (MIP; Lefebvre *et al.* 2015; Altieri and Nicholls 2000).

Manejo Integrado de Plagas

El MIP es una forma de abordar la producción agrícola que brinda un marco para el desarrollo y uso de técnicas de control que involucren el uso de múltiples estrategias coordinadas para optimizar el manejo de plagas, de manera que se contemplen los factores ecológicos, económicos y sociales involucrados, minimizando los efectos secundarios y no deseados para proteger el medioambiente y la salud humana (Bajwa and Kogan 2002). A su vez, el MIP promueve un enfoque diversificado, limitando la excesiva dependencia en cualquier tecnología específica de control (e.g., pesticidas

sintéticos) y promoviendo una mayor dependencia en los procesos biológicos renovables, tal como el uso de enemigos naturales.

Este enfoque incluye elementos clave como: (1) el monitoreo del complejo de plagas del cultivo en cuanto a su dinámica poblacional y a los daños que le provocan al cultivo, y (2) la utilización de herramientas de decisión, tales como el uso de modelos predictivos y el uso de umbrales de control (UC) por especie o multiespecie (Flint and Van den Bosch 2012; Lewis *et al.* 1997; Duggal and Siddiqi 2008). Estos umbrales son valores prácticos o reglas operacionales; por ejemplo, la densidad poblacional a la que deben iniciarse medidas de control para evitar que se alcance un nivel de daño económico, definido como “la densidad poblacional mínima que puede causar un daño económico” (Stern *et al.* 1959; Cancelado and Radcliffe 1979; Pedigo *et al.* 1986). A su vez, los UC pueden variar según la fenología del cultivo y la especie de plaga.

Problemas de aplicación y de adopción del MIP

A pesar de que el manejo integrado de plagas es un enfoque deseable desde varios aspectos, por lo general los productores no siguen sus recomendaciones. En la práctica, muchos programas de manejo no respetan los umbrales de control y solo realizan controles programados en función de la fenología del cultivo o, con el objetivo de simplificar las labores operacionales, realizan aplicaciones preventivas de insecticidas para hacer coincidir su aplicación con la de herbicidas y/o fungicidas (Gamundi and Sosa 2008). Esto da lugar a controles poco efectivos y poco eficientes. La efectividad de algunos pesticidas depende de que sean aplicados en el momento indicado. La baja eficiencia es debida a que si las aplicaciones se realizan cuando el cultivo alcanza una fenología determinada, puede que en este momento la densidad poblacional de la plaga sea aún baja. Además, si bien el uso de modelos predictivos es uno de los elementos básicos del manejo integrado de plagas, en la práctica son poco utilizados por los productores en la Argentina.

Incluso en los cultivos en los cuales se realizan monitoreos periódicos y se utilizan umbrales para determinar en qué momento realizar un control, muchos productores observan reducciones en el rinde mayores a las esperadas. Esto puede deberse a

varias razones. Una de ellas es que para que los umbrales de control den buenos resultados deben ser acompañados de un sistema de muestreo y monitoreo que asegure una muestra representativa del tamaño poblacional de chinches en los cultivos. Entre los problemas que dificultan la obtención de esta muestra, independientemente del método de muestreo que se utilice, se encuentran: (1) el patrón espacial agregado que presentan las chinches en los estadios reproductivos anteriores a R5.5 (vainas con el 50% del total de granos; luego de este estadio, la distribución de chinches es más uniforme) (Kuss *et al.* 2007), y (2) el alto costo de realizar los monitoreos, que resulta en un bajo número de muestras (según los productores, 10 por cultivo en promedio) y en que no estén distribuidas uniformemente a través del cultivo (Massaro 2005).

Otra razón por la que se observan reducciones en el rinde mayores a las esperadas, incluso cuando se utilizan umbrales de control, es la falta de una estimación del daño acumulado sobre un cultivo a través del tiempo. Se ha observado que tamaños poblacionales por debajo del umbral de control y durante un tiempo prolongado pueden provocar mermas en el rendimiento de los cultivos (McPherson *et al.* 1979), pero no existe un método que permita cuantificar este daño.

En resumen, la falta de uso de monitoreos y umbrales de control, las reducciones en el rinde mayores a las esperadas incluso cuando se utilizan umbrales y la falta del uso de modelos predictivos del crecimiento poblacional de chinches reducen la eficacia de los planes de control basados en el MIP. Esto, a su vez, dificulta la adopción del MIP por parte del sector productivo que entonces suele optar por una estrategia de control basada en el uso excesivo de pesticidas sintéticos.

En este trabajo se propone un enfoque de modelado para obtener estimaciones robustas del tamaño poblacional de chinches en cultivos de soja y para realizar predicciones de tamaños poblacionales. El propósito último es brindar una nueva herramienta que asista al proceso de decisión acerca de cuál es el momento más indicado para realizar un control y con ello aumentar la eficacia en el manejo integrado de plagas. Varias características del sistema de estudio fueron tenidas en cuenta a la hora de elegir el enfoque de modelado de este trabajo:

1. **Datos disponibles del sistema.**

El tipo de datos que recolectan los productores durante los monitoreos de los cultivos (abundancia de chinches y fenología del cultivo) y los que pueden obtenerse de bases de datos meteorológicas (temperatura) brindan un conocimiento empírico de cómo funciona el sistema más que un conocimiento basado en principios biológicos mecanísticos. Un enfoque de modelado empírico (top-down) se ajusta muy bien a esta situación y tiene la ventaja de que no requiere de experimentos adicionales que midan los parámetros fisiológicos y sus variaciones en función de la meteorología que requeriría un enfoque mecanístico (bottom-up) para modelar la dinámica de chinches en cultivos de soja.

2. **Presencia de un error de observación considerable y sesgado.**

Las mediciones de abundancia de chinches realizadas con el método del paño vertical (ver metodología) son en sí poco precisas y tienden a subestimar considerablemente la abundancia, a pesar de ser el mejor método disponible (Gamundi 1995). Esto hace conveniente el uso de alguna herramienta que permita considerar este error para obtener una estimación del tamaño poblacional *real* a partir del tamaño poblacional *observado*, evitando que este error se propague en las predicciones. Los modelos espacio-estado logran este objetivo modelando de forma separada el proceso de dinámica poblacional y la observación del sistema. Sumado a esto, un enfoque bayesiano permite reducir la incertidumbre en los parámetros que regulan la dinámica del sistema (y por lo tanto en las predicciones) al incorporar la información disponible antes del ajuste del modelo a través de las distribuciones previas.

3. **Variabilidad inherente del sistema biológico.**

Existen varias fuentes de variabilidad en los sistemas biológicos que se ven reflejadas en los datos recolectados y que van más allá del error de observación (ya sea error aleatorio o sistemático). Estas fuentes pueden deberse a: heterogeneidades espaciales o temporales (error de parámetros), aproximar el funcionamiento del sistema a través de un modelo determinado (error estructural de modelo), u otros procesos que afectan al sistema que no

han sido contemplados porque no se conocen lo suficiente o porque es muy costoso medirlos o porque no es viable por la capacidad de cómputo que sería necesaria para estimar todos los parámetros relevantes. Todas estas fuentes de variación, conjuntamente llamadas *error de proceso*, pueden ser incorporadas al modelo incluso en los casos donde no sea posible explicar qué las genera. Su incorporación permite realizar predicciones con mayor exactitud y evitar un exceso de seguridad en la estimación de la incertidumbre (Michael C. Dietze 2017). Un enfoque bayesiano permite incorporar estas fuentes de variabilidad de forma directa a través del uso de distribuciones posteriores en la variable respuesta (en nuestro caso, abundancia de chinches) y en los efectos de las covariables.

A partir de estas consideraciones, se optó por un enfoque de modelado empírico basado en un modelo bayesiano jerárquico con una estructura espacio-estado (Wikle *et al.* 1998; Dietze 2017).

OBJETIVOS

1. Desarrollar modelos predictivos de la dinámica poblacional de pentatómidos de los géneros *Nezara*, *Piezodorus*, *Edessa* y *Dichelops* en cultivos de soja de la Argentina.
2. Estimar el daño acumulado que causan las chinches en los cultivos de soja.
3. Evaluar el impacto que tiene el tamaño muestral en la incertidumbre de las predicciones del modelo.
4. Comparar el poder predictivo de distintos modelos sobre la abundancia de pentatómidos mediante técnicas de validación cruzada y criterios de información.

METODOLOGÍA

Datos

Para ajustar el modelo se utilizaron datos de fenología del cultivo y de abundancia de chinches de las especies *Nezara viridula*, *Piezodorus guildinii*, *Edessa mediatubunda* y *Dichelops furcatus* recolectados por la Asociación Argentina de Protección Profesional de Cultivos Extensivos (AAPPCE. <https://aappce.org/>) en 8 localidades (Figura 1). Los datos fueron recolectados entre enero y abril de 2017 y, en cada localidad, el muestreo se llevó a cabo durante al menos 8 semanas consecutivas. El diseño de muestreo consistió en seleccionar, para cada localidad, un lote de 150 m x 300 m donde no se hubiera realizado un control de chinches y el cultivo se encontrara en un estado vegetativo con al menos tres nudos por planta.

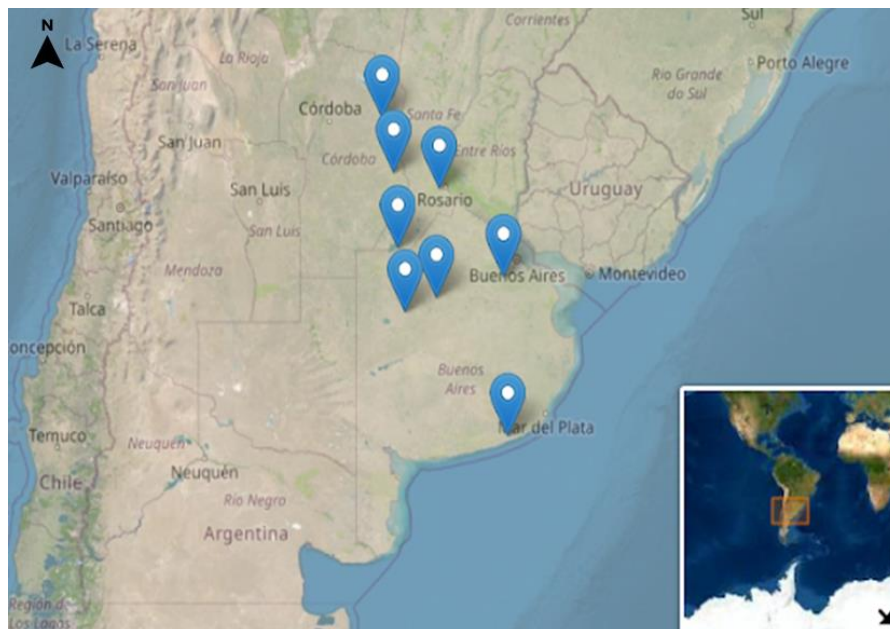


Figura 1. Localidades donde se realizaron los relevamientos de chinches de la soja: América, Diego de Alvear, La Francia, Leones, Necochea, Pehuajó, San Vicente, Zavalla.

Semanalmente y dentro de cada lote, se tomó un total de 180 muestras de 1 m lineal utilizando el método del paño vertical (Gamundi 1995) (Figura 2), siempre siguiendo transectas paralelas y de forma tal que la distancia entre muestras fuera de al menos 10 m. El método del paño consiste en golpear las plantas de soja contra un paño de tela que contiene un recolector en la parte inferior para poder registrar la cantidad de individuos encontrados. Sobre el paño se golpean tantas plantas como entren a lo

largo del metro lineal del paño y la cantidad de individuos recolectados de todas estas plantas representan una muestra.



Figura 2. Muestreo con el método de paño vertical.

Con este método se registró la cantidad de ninfas del 4^{to} y 5^{to} estadio y de adultos presentes de cada una de las cuatro especies consideradas. Las ninfas de 4^{to} y 5^{to} estadio se agruparon como “chinchas grandes”. Las distintas especies se equipararon según las relaciones de daño al cultivo utilizadas por INTA (Instituto Nacional de Tecnología Agropecuaria. <https://www.argentina.gob.ar/inta>). Para esto se creó una variable llamada Equivalente Nezara (EqNv) que se construye sumando todos los individuos de cada especie ponderando cada valor por el nivel de daño de cada especie relativo a *N. viridula*. (*P. guildinii* = 1.6, *N. viridula* = 1, *E. meditabunda* = 0.6 y *D. furcatus* = 0.3) (Figura 3).

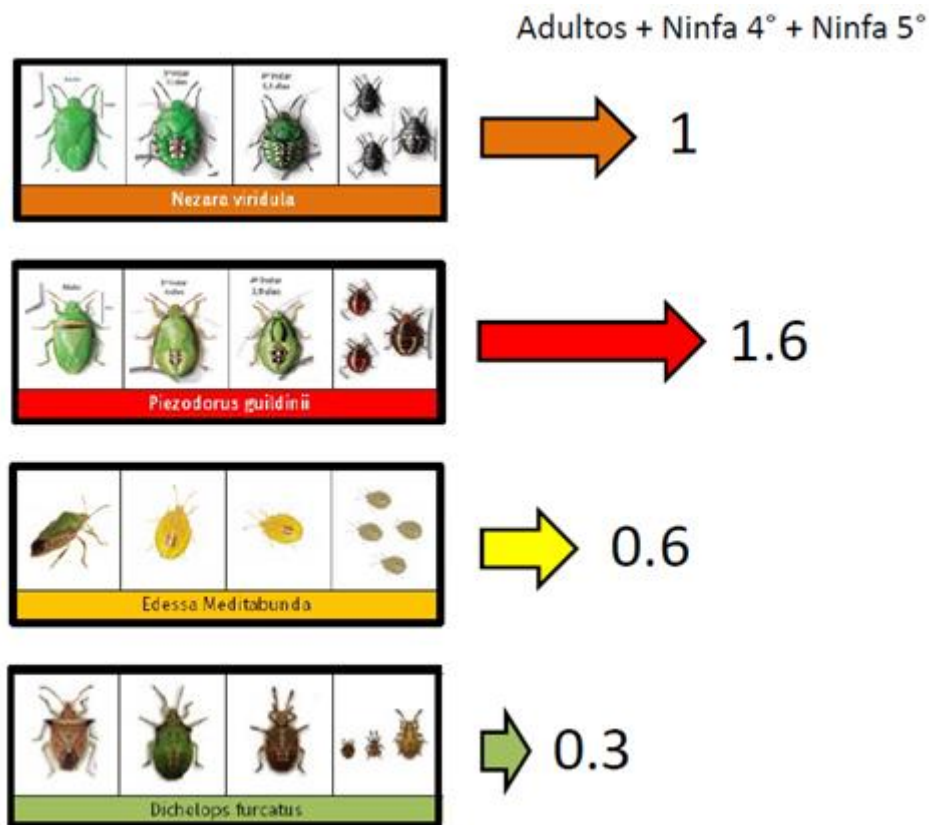


Figura 3. Relaciones de daño que dan lugar a la variable Equivalente Nezara.

A su vez, en estos muestreos se registró el estadio fenológico del cultivo según la siguiente clasificación (V = Vegetativo, R = Reproductivo) (Fehr and Caviness 1977) (Figura 4):

- V1: Primer Nudo.
- V2: Segundo Nudo.
- Vn: N° Nudos.
- R1: Comienzo de Floración.
- R2: Plena Floración.
- R3: Comienzo de Fructificación.
- R4: Plena Fructificación.
- R5: Comienzo de llenado de granos.
- R6: Llenado máximo de granos.
- R7: Inicio de Madurez.
- R8: Plena Madurez.

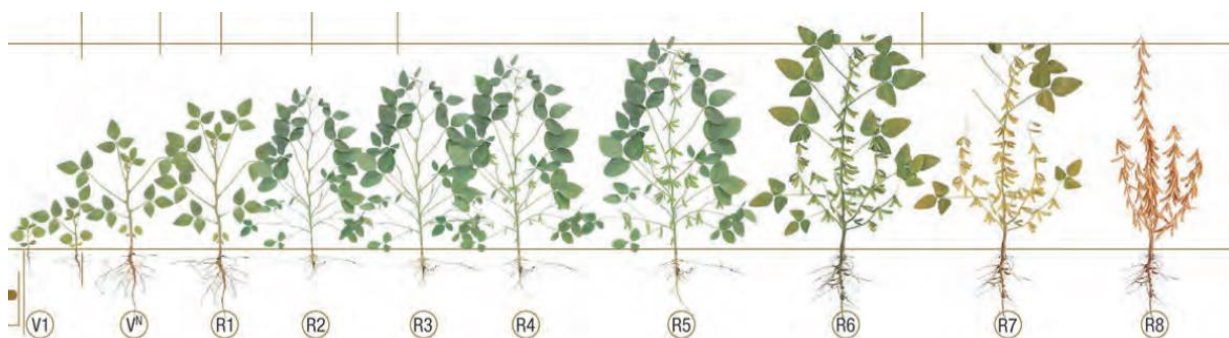


Figura 4. Estadios fenológicos de la soja.

Por último, se recopilaron datos de temperatura media diaria de cada una de las localidades a partir de estaciones meteorológicas (NOAA, National Oceanic and Atmospheric Administration. <https://www.noaa.gov/>) y se calculó la temperatura media semanal por localidad (Figura 5). Se optó por obviar el uso de datos de precipitación por ser demasiado escasos para las localidades y fechas de las que provenían los datos de abundancia de chinches.

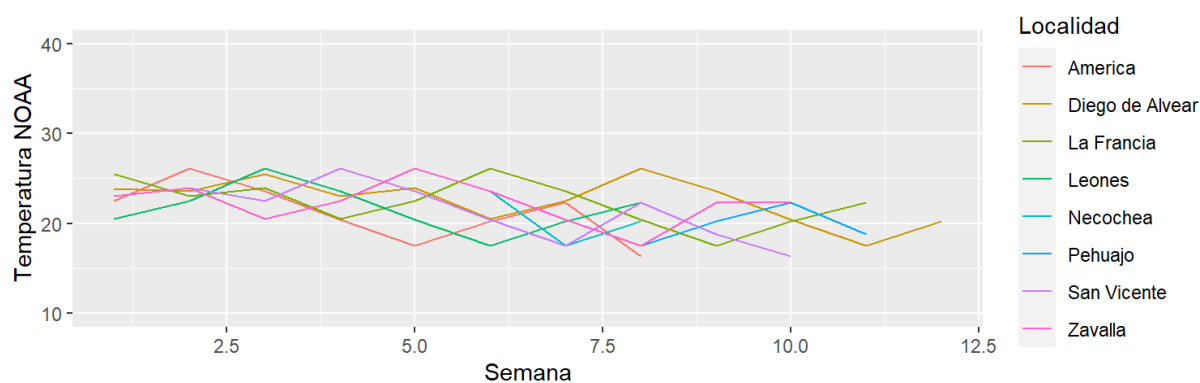


Figura 5. Temperatura media semanal (°C) en cada localidad.

Modelo espacio-estado

Los modelos espacio-estado son modelos de aprendizaje no supervisado (Baum and Eagon 1967; Baum and Petrie 1966). Constan de dos submodelos interconectados. Por un lado, se modela el estado o valor “real” de una variable de interés que no es observada directamente (llamada variable “latente”) de forma dinámica, generalmente

utilizando un proceso Markoviano de primer orden. Este proceso implica que el estado X de la variable latente en tiempo t depende únicamente de su estado en el tiempo anterior $t - 1$ tal que $X_t = f(X_{t-1})$. A este submodelo se lo denomina “modelo de proceso”. Por otro lado, se modela el proceso de observación de esta variable latente en un tiempo dado $Y_t = g(X_t)$, contemplando el error de observación presente en este proceso (es decir, teniendo en cuenta cómo puede variar el valor observado respecto al valor real). A este submodelo se lo llama “modelo de datos” y usarlo permite una mejor estimación de la dinámica del sistema de estudio (particularmente en los casos donde la observación es muy “ruidosa”), ya que separa al proceso biológico de su observación. Esta separación permite contemplar que la observación de un sistema es imperfecta y que la dinámica del sistema depende del valor *real* de su estado y no de nuestras observaciones. En síntesis, el objetivo de la inferencia estadística con los modelos espacio-estado es estimar el estado o valor de una variable latente en un tiempo determinado a partir de una serie temporal de observaciones. Con un enfoque bayesiano esto equivale a estimar la distribución de probabilidad posterior $p(X_t | Y_t, Y_{t-1}, \dots, Y_{t=0})$.

En nuestro sistema biológico, la variable latente representa el tamaño poblacional de chinches en un momento determinado y con el modelo de proceso se representa su dinámica poblacional a través del tiempo, mientras que el modelo de datos representa la observación de esta dinámica a través de monitoreos con paño vertical (Figura 6).

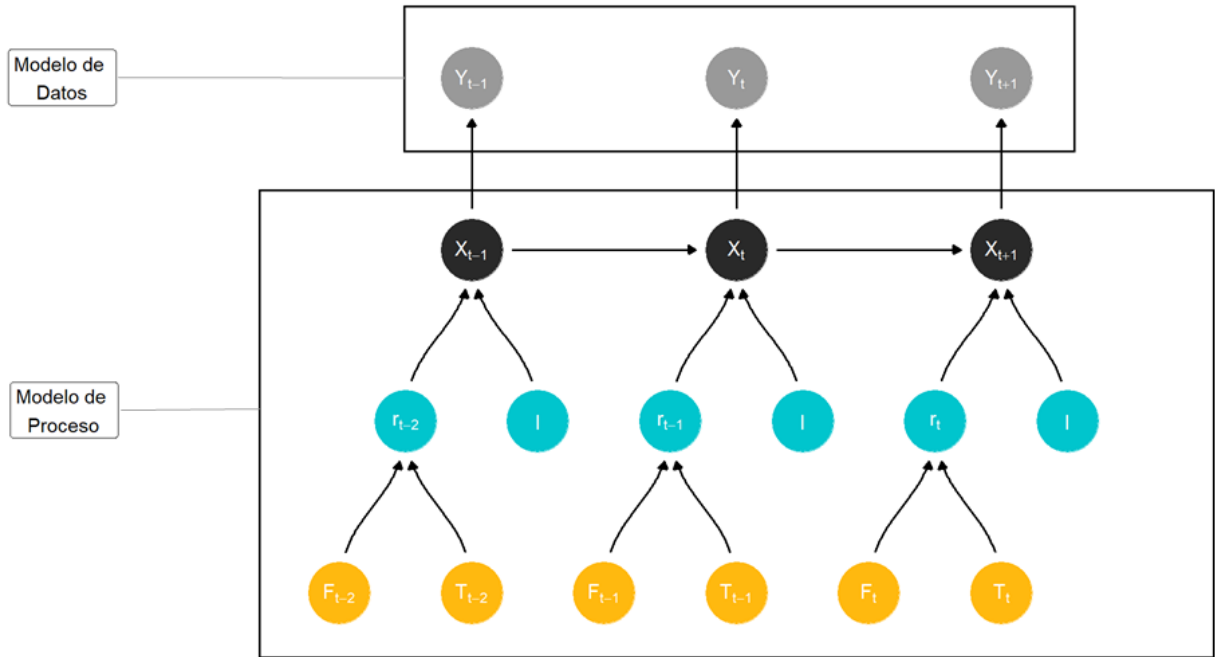


Figura 6. Esquema de modelo espacio-estado. Y : tamaño poblacional *observado*; X : tamaño poblacional *real* estimado; r : tasa de crecimiento poblacional; I : inmigración de chinches grandes desde cultivos cercanos; F : fenología del cultivo; T : temperatura del aire. Las flechas indican la relación de dependencia.

En este modelo, el tamaño poblacional en un tiempo t (X_t) depende de: el tamaño poblacional en el tiempo previo (X_{t-1}), la tasa de crecimiento poblacional en el tiempo previo (r_{t-1}) y la inmigración por semana de chinches grandes (I). A su vez, la tasa de crecimiento poblacional (r_{t-1}) depende de la temperatura (T_{t-1} , variable numérica escalada a media cero y desvío estándar igual a 1) y de la fenología del cultivo (F_{t-1} , variable categórica numérica y ordenada, donde todos los estadios fenológicos vegetativos son iguales a cero y los estadios reproductivos toman valores mayores a uno, enteros y ascendentes). Por último, este tamaño poblacional (X_t) es observado con cierta probabilidad (P_{obs}) y resulta en un tamaño poblacional observado (Y_t). Esto da lugar al siguiente modelo determinístico:

$$Y_t = P_{obs} \cdot (X_t)$$

$$X_t = r_{t-1} \cdot (X_{t-1}) + I$$

$$\log(r_{t-1}) = \beta_T \cdot (T_{t-1}) + \beta_F \cdot (F_{t-1})$$

Por un lado, la elección de las covariables F y T como moduladores de la tasa de crecimiento poblacional (r) fue en base a trabajos donde, por ejemplo, se ha encontrado que variaciones en la temperatura tienen un efecto sobre el tamaño de puesta de huevos, la tasa de desarrollo, la longevidad y la mortalidad; y que la dinámica poblacional de chinches está sincronizada con la fenología del cultivo, donde la mayor abundancia se encuentra durante el llenado de vainas o durante la madurez temprana (Todd 1989; Panizzi and Slansky 1985). Por otro lado, la decisión de agregar un parámetro de inmigración semanal de chinches (I) se debe a que estos insectos migran desde las zonas circundantes a medida que los cultivos se van desarrollando (Swanson 2012).

A este modelo determinístico se le incluyeron distribuciones de probabilidad para contemplar la estocasticidad propia de la dinámica poblacional y del proceso de observación, junto con la variabilidad no explicada por las covariables elegidas. Esto resultó en el siguiente modelo probabilístico:

$$Y_t \sim \text{Binomial}(X_t, P_{obs})$$

$$X_t \sim \text{Poisson}(\lambda_t)$$

$$\lambda_t = r_{t-1} \cdot (X_{t-1}) + I$$

En lo que respecta al modelo de datos, existen varias formas en las que se podría modelar la observación del tamaño de una población en un tiempo determinado. Por ejemplo, se podría utilizar una distribución *normal*, suponiendo que el error de observación es simétrico y está centrado en el estado latente (es decir, el valor *real* del tamaño poblacional). Pero en este caso las mediciones a campo del tamaño poblacional de chinches dan valores que son enteros y que tienden a ser menores al tamaño *real*. Por lo tanto, una distribución *binomial* es una elección más apropiada porque es una distribución discreta que permite contemplar este sesgo a través del valor de su parámetro p (probabilidad de éxito en los ensayos, P_{obs} en nuestro modelo). En este contexto esta probabilidad representa una medida de cuánto se subestima el valor *real* del tamaño poblacional. Una vez obtenidas las estimaciones

de tamaño poblacional de chinches, estos valores fueron transformados a densidades por metro lineal (utilizando la cantidad de muestras utilizadas) para que tuvieran la misma unidad que los umbrales de control.

En base al conocimiento brindado por la AAPPCE, se estableció la siguiente distribución previa para la probabilidad de observación (P_{obs}) (Figura 7):

$$P_{obs} \sim Beta(8, 2.5)$$

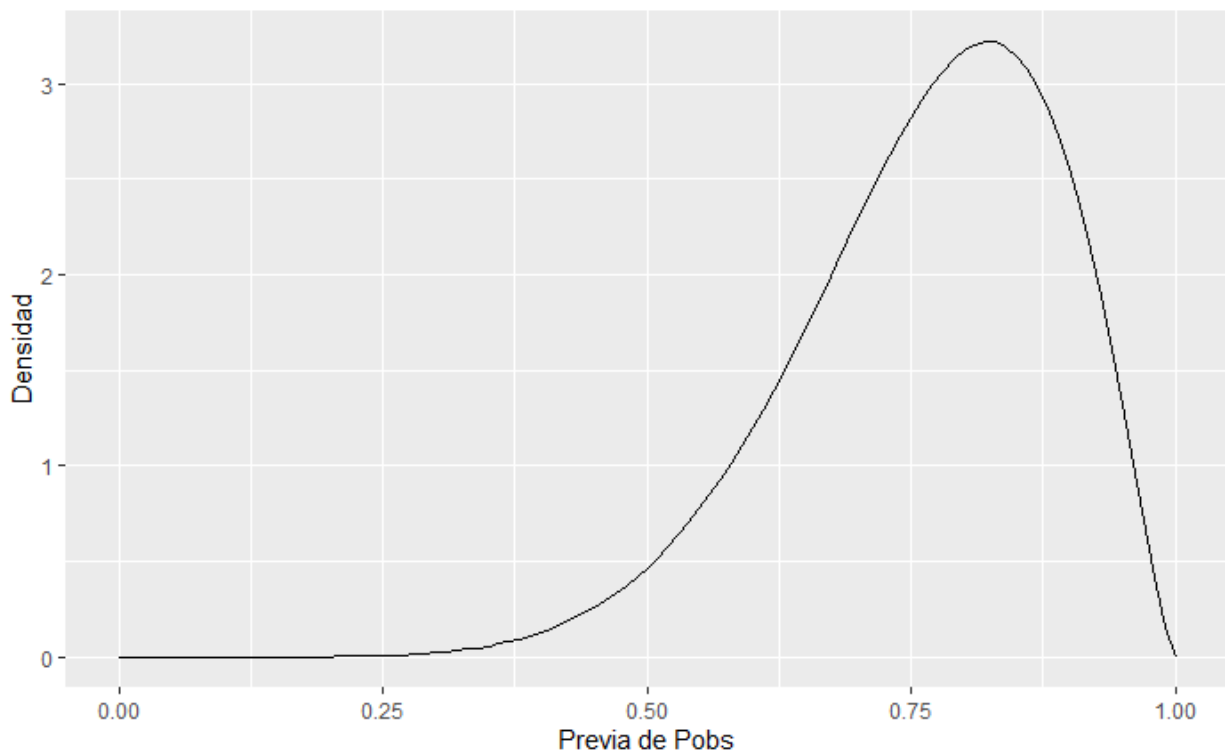


Figura 7. Distribución Beta(8, 2.5), previa establecida para la probabilidad de observación (P_{obs} . Parámetro p de la distribución binomial).

En cuanto al modelo de proceso, se optó por utilizar una distribución de *Poisson*, la cual representa la cantidad de eventos (en este caso, chinches) que ocurren en un tiempo determinado y al tener un solo parámetro, es el modelo más sencillo que se puede tomar como punto de partida.

Estructura jerárquica

Este modelo se amplió a todas las localidades (j) utilizando una estructura jerárquica para los parámetros que afectan la tasa de crecimiento poblacional y para el parámetro de inmigración.

$$\beta_{Tj} \sim Normal(\mu_T, \sigma_T)$$

$$\beta_{Fj} \sim Normal(\mu_F, \sigma_F)$$

$$I_j \sim HalfNormal(\mu_I, \sigma_I)$$

Esta estructura permite que los datos de una localidad determinada (nuestra unidad experimental) no solo informen sobre los parámetros que dominan la dinámica poblacional de ese lugar, sino que también informen sobre los parámetros de las demás localidades. Integrar la información disponible en distintos niveles (localidad particular y conjunto de localidades) es lo que diferencia a la estructura jerárquica de los enfoques “no pooling” o “complete pooling”, en los cuales las unidades experimentales (localidades) se tratan como completamente independientes o como exactamente iguales, respectivamente. Esto es lo que les permite a los modelos jerárquicos realizar estimaciones más robustas, incluso en las unidades experimentales donde hay poca información.

Esta integración de la información disponible en distintos niveles se logra gracias a que en los modelos jerárquicos los parámetros específicos de cada localidad (β_{Tj} , β_{Fj} , e I_j) son muestreados de una distribución “global” (caracterizada por los parámetros μ , σ) que representa a la “población” de localidades. Esto permite modelar tanto variaciones individuales (a nivel de localidad) como la variabilidad a nivel grupal (al nivel de la “población” de localidades) (Shiffrin *et al.* 2008). A diferencia de los parámetros específicos de cada localidad (β_{Tj} , β_{Fj} , e I_j), los parámetros que caracterizan la distribución global (μ , σ) requieren de distribuciones previas que

describan dónde está centrado el efecto global (μ) y cuál es la variabilidad entre localidades (σ). Para esto, se establecieron las siguientes distribuciones previas:

$$\mu_F \sim \text{Normal}(0, 0.5)$$

$$\mu_T \sim \text{Normal}(0, 0.5)$$

$$\mu_I \sim \text{Normal}(0, 1)$$

$$\sigma_F \sim \text{Exponential}(5)$$

$$\sigma_T \sim \text{Exponential}(5)$$

$$\sigma_I \sim \text{Exponential}(1)$$

Ajuste

Para el análisis general del trabajo se utilizó el lenguaje de programación R 4.0.3 (Ihaka and Gentleman 1996) y el ambiente integrado de desarrollo RStudio 1.4.1103. (RStudio Team 2020). Para realizar el ajuste del modelo se utilizó un método de Markov Chain Monte Carlo a través del programa JAGS 4.3.0 (Just Another Gibbs Sampler) (Plummer 2003).

Simulaciones predictivas previa y posterior

Para corroborar que la elección de distribuciones previas resultara en un modelo generativo (es decir, que fuera capaz de generar datos) que se correspondiera con lo que podría ocurrir en el sistema biológico según nuestro conocimiento previo se realizó una simulación predictiva previa (Apéndice). Para corroborar que el modelo una vez ajustado a los datos fuera capaz de representar esos datos se realizó una simulación predictiva posterior (Apéndice).

Variantes del modelo y comparación en términos de poder predictivo

Se realizaron variantes del modelo antes descrito (de ahora en más, “modelo base”) con el fin de evaluar qué variante presenta el mayor poder predictivo fuera de la muestra (esto es, la capacidad de predecir datos que no han sido utilizados para el ajuste del modelo).

La primera variante consistió en quitar el efecto de la temperatura sobre la tasa de crecimiento poblacional de chinches, con el resto del modelo (incluyendo las previas) exactamente igual:

$$\log(r_{t-1}) = \beta_0 + \beta_F \cdot (F_{t-1})$$

La segunda variante consistió en agregar un sub-modelo para la inmigración basado en un decrecimiento exponencial a través del tiempo, bajo el supuesto de que, del total de chinches que aparecen semanalmente en el cultivo, la reproducción de la población dentro del cultivo toma más predominancia por sobre la inmigración desde zonas aledañas a medida que transcurre el tiempo. El resto del modelo (incluyendo el resto de las previas) se mantiene igual:

$$\lambda_t = r_{t-1} \cdot (X_{t-1}) + I_{tj}$$

$$I_{tj} = I_{inicial\ j} \cdot (1 - I_{decaimiento})^t$$

$$I_{inicial\ j} \sim Normal(\mu_{I\ inicial}, \sigma_{I\ inicial})$$

$$\mu_{I\ inicial} \sim HalfNormal(5, 5)$$

$$\sigma_{I\ inicial} \sim HalfNormal(5, 2)$$

$$I_{decaimiento} \sim Beta(3.5, 10)$$

En este modelo, la inmigración de chinches grandes que ocurre en cada semana en una localidad determinada (I_t) depende de la inmigración que ocurrió en la primer semana de los datos ($I_{inicial,j}$, parámetro jerárquico por localidad) y de un factor de decaimiento en el tiempo de la tasa de inmigración ($I_{decaimiento}$).

Luego, se comparó el poder predictivo *fuera de la muestra* del modelo base y estas variantes utilizando el criterio de información WAIC (Widely Applicable Information Criterion) (Watanabe and Opper 2010) y el método de validación cruzada LOO-CV (Leave-One-Out Cross Validation) (Vehtari *et al.* 2017).

WAIC es un criterio de información completamente bayesiano que posee mejoras sustanciales respecto a los criterios de información clásicos como AIC o DIC (Vehtari *et al.* 2017). Tal como el resto de los criterios de información, con el WAIC se obtiene una *estimación* del poder predictivo fuera de la muestra. Esta estimación resulta de evaluar el grado de ajuste del modelo al conjunto de datos completo pero añadiendo una penalización por estar utilizando el mismo conjunto de datos tanto para ajustar el modelo como para evaluar su capacidad predictiva.

En cambio, los métodos de validación cruzada separan el conjunto de datos completo en un subconjunto de ajuste y un subconjunto de prueba. Esto permite realmente evaluar el poder predictivo fuera de la muestra en vez de obtener solo una estimación, pero con la penalidad de un costo de computación considerablemente mayor. LOO-CV consiste en formar un subconjunto de ajuste con $n-1$ datos y formar un subconjunto de prueba con el dato restante y luego se evalúa la capacidad de predecir el subconjunto de prueba. Estos pasos se repiten hasta que se haya evaluado la capacidad de predecir cada uno de los datos del conjunto completo. Si bien el WAIC es asintóticamente igual a la LOO-CV, este último brinda un valor del poder predictivo fuera de la muestra más robusto cuando el conjunto de datos es pequeño, las distribuciones previas son poco informativas o existen datos influyentes (Vehtari *et al.* 2017).

Para computar el WAIC y la LOO-CV se utilizó el paquete de R “loo” (Vehtari *et al.* 2019) que aproxima el resultado de LOO-CV utilizando Pareto-Smoothed Importance Sampling (PSIS-LOO-CV: Pareto-Smoothed Importance Sampling Leave One Out Cross Validation) para reducir considerablemente el tiempo de cómputo y obtener un resultado en un tiempo similar al que requiere el WAIC.

A partir de los resultados del WAIC y la LOO-CV, se eligió el modelo con mejor poder predictivo fuera de la muestra. Se utilizó este modelo para realizar las estimaciones de la densidad poblacional real y daño acumulado, para realizar las predicciones de éstas últimas y para realizar el análisis de incertidumbre en función del tamaño muestral.

Predicciones

Para cada localidad, se realizaron predicciones de la densidad de chinches por metro lineal para la semana siguiente a la disponible en los datos en cada una de las localidades, tal como se usaría el modelo en tiempo real con los productores. Para esto, se utilizaron datos meteorológicos recopilados y se supuso que la fenología en la semana de la predicción era igual a la fenología de la última semana de los datos. Para verificar que las predicciones que realizara el modelo tuvieran una exactitud razonable, se realizó un ajuste con todos los datos de densidad de chinches por metro lineal observada de la serie temporal excepto las últimas dos semanas y se compararon las predicciones del modelo para estas dos semanas con los datos observados. Al realizar estas predicciones, se utilizaron los datos de fenología y temperatura observados, lo que representa un escenario de predicción utópico donde no existe incertidumbre en el valor futuro de las covariables.

Estimaciones de daño acumulado sobre los cultivos

Los productores de la AAPPCE habían expresado interés en contar con una medida del daño acumulado por el ataque chinches que presentan los cultivos. Hasta ahora, esta medida ha sido muy poco explorada y casi no se tiene en cuenta al momento del control. Para esto se creó un método para estimar el daño acumulado que causan las chinches en los cultivos estimado en función del Equivalente Nezara. El método consistió en agregar al modelo un módulo que monitoree la densidad poblacional de chinches acumulada a través del tiempo y luego en crear dos variables:

- Unidades-umbral de daño
- Unidades-umbral de daño acumulado

La primera variable es un estimador del daño que genera una cantidad de chinches durante un período de una semana. La segunda variable estima este mismo daño, pero de forma acumulada en el tiempo. Ambas variables suponen que el daño que producen las chinches al cultivo es igual independientemente de la fenología vegetal y suponen que el daño es aditivo.

Para estimar la primera variable (“unidades-umbral de daño”), primero se fija un umbral que se decide previamente por el productor o el encargado de tomar la decisión de cuándo hacer un control. Segundo, se estima la densidad de chinches por metro lineal durante una semana. Por último, se calcula qué proporción del umbral representa esta densidad. Por ejemplo, si se establece que el umbral es 2 chinches/m lineal y en una semana determinada el modelo estima que hay 0,5 chinches/m lineal, entonces la variable “unidades-umbral de daño” toma el valor 0,25. Esto implica que, durante esta semana, las chinches provocaron un daño igual al 25% del umbral previamente definido. A la semana siguiente, se realiza el mismo procedimiento y se suman los valores de “unidades-umbral de daño” de ambas semanas para obtener el valor de la variable “unidades-umbral de daño acumulado”. Por ejemplo, si en esta semana hubo un incremento poblacional respecto a la semana previa y ahora el cultivo tiene una densidad poblacional estimada de 1 chinches/m lineal, esto significa que la variable “unidades-umbral de daño” de esta semana toma el valor de 0,5 (ya que 1 chinche/m lineal representa el 50% del umbral previamente definido). Luego, para calcular el valor de la variable “unidades-umbral de daño acumulado”, se suma el valor de unidades-umbral de daño de esta semana y de la semana previa, obteniendo un valor de 0,75 unidades-umbral de daño acumulado (0,25 de la semana anterior + 0,5 de esta semana).

De esta forma, semana a semana se va estimando el daño acumulado que presenta el cultivo plasmado en un indicador de cuándo sería necesario realizar un control. Estas dos variables se calculan para cada fenología de forma separada, lo que resulta en una variable distinta según en qué estadio se encuentre el cultivo en un tiempo determinado (por ejemplo, unidades-umbral de daño acumulado en R5, unidades-umbral de daño acumulado en R6, etc). Adicionalmente, se realizaron predicciones del daño acumulado para la semana siguiente a los datos disponibles utilizando las predicciones de densidad poblacional de EqNv.

Incertidumbre en las predicciones en función del tamaño muestral

Para evaluar el impacto del tamaño muestral en las predicciones del modelo se realizó un análisis de incertidumbre en función del tamaño muestral. Este análisis consistió en tomar, para cada una de las localidades, una cantidad determinada de muestras de forma aleatoria a partir de las 180 muestras que se midieron por semana y ajustar el modelo solo con estas muestras. Este proceso se realizó para distintos tamaños muestrales (10, 20, 40, 60, 80, 100, 120, 140, 180 muestras) con 9 réplicas por cada tamaño. Siempre se utilizaron los datos de todas las semanas del muestreo excepto la última con el fin de comparar las predicciones del modelo con los datos observados en esta última semana. Una vez ajustado el modelo, se simuló dinámicas poblacionales de chinches en toda la serie temporal (incluyendo la semana sobre la que se hacía la predicción). Estos resultados se compararon con los obtenidos en las demás réplicas y tamaños muestrales.

RESULTADOS

El modelo con estructura espacio-estado permitió obtener estimaciones de la densidad poblacional *real* (X) para los tiempos de los datos (las series temporales utilizadas para ajustar el modelo, $Y_t, Y_{t-1}, \dots, Y_{t=0}$) y para tiempos futuros (predicciones, Y_{t+1}, Y_{t+2}). Las estimaciones para los tiempos de los datos de las densidades poblacionales *reales* difirieron considerablemente de las densidades poblacionales *observadas* (Figura 8).

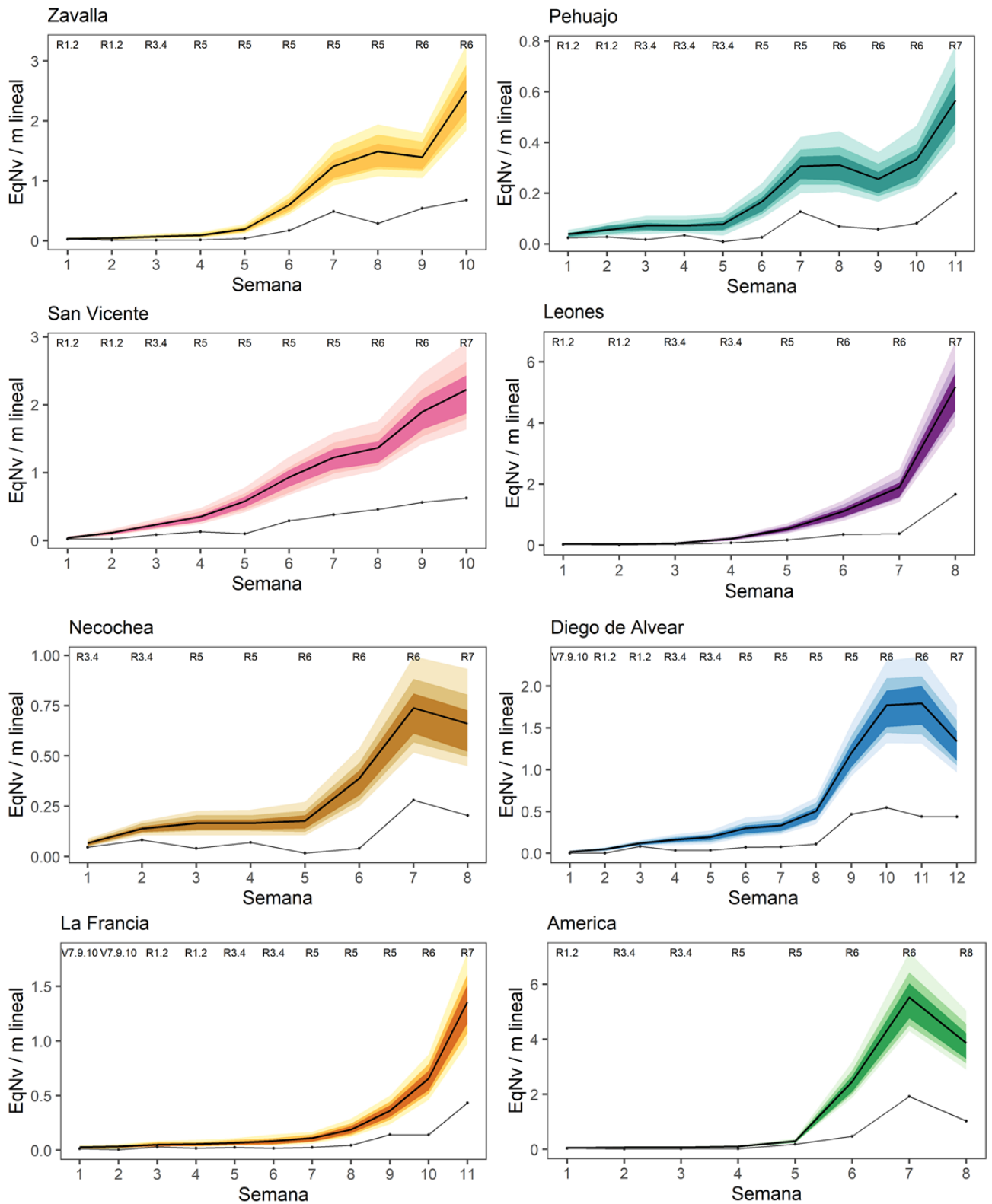


Figura 8. Estimación de densidad de Equivalente Nezar (EqNv) por metro lineal por medio de un modelo jerárquico para las 8 localidades estudiadas. La curva gruesa indica la mediana y los sombreados indican los intervalos de credibilidad (IC) 60, 80 y 95%. La curva fina indica los valores observados. La fenología del cultivo correspondiente a cada semana se indica en la parte superior del gráfico.

Las estimaciones de la densidad poblacional a tiempo futuro (para las dos semanas siguientes a los datos utilizados) mostraron en general un grado de exactitud aceptable, ya que varias veces los puntos (valores observados) se encontraron dentro de los intervalos de credibilidad de la predicción y en varios casos cercanos a la mediana predicha (Figura 9).

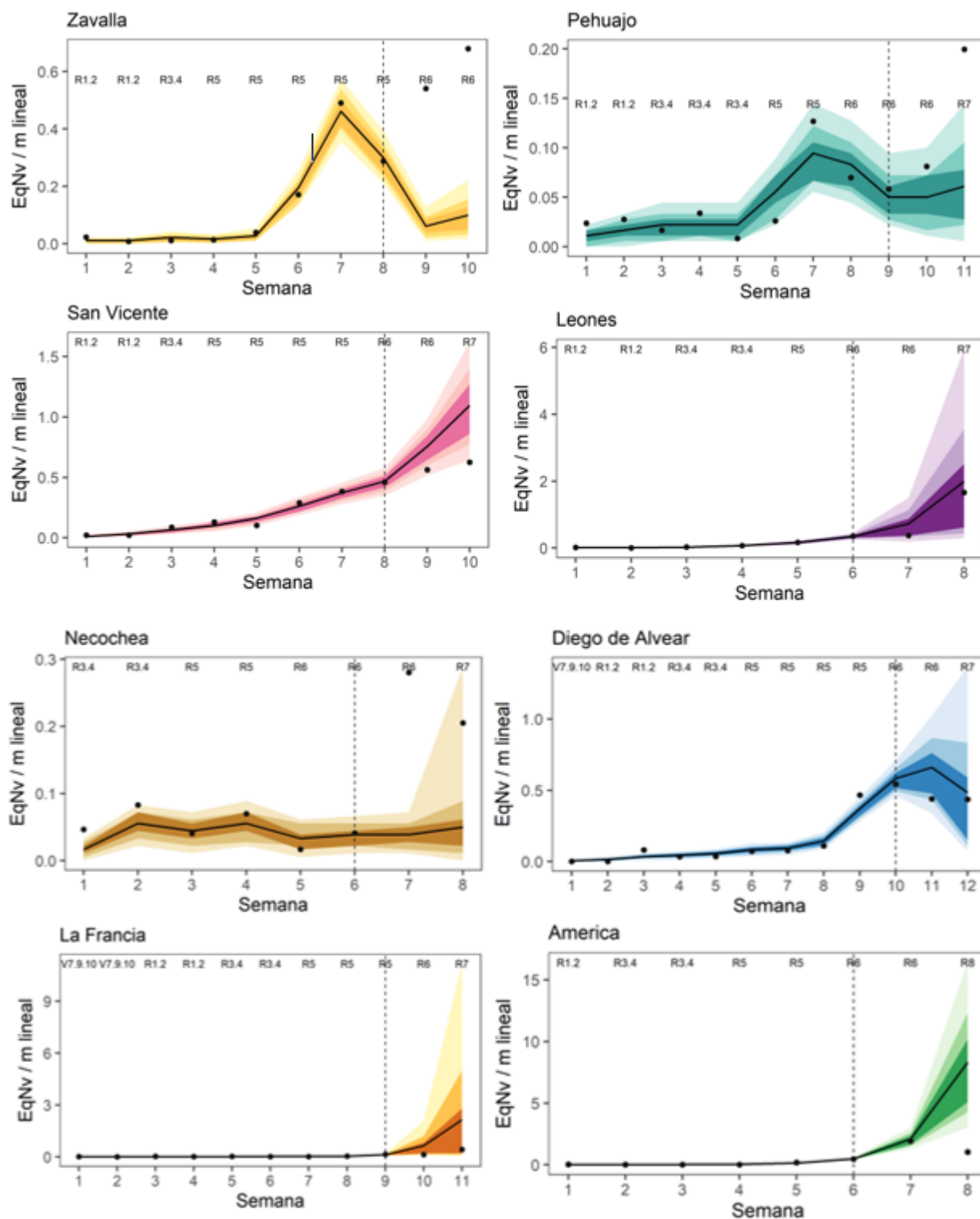


Figura 9. Predicciones de la densidad *observada* de EqNv/m lineal por medio de un modelo jerárquico para todas las localidades. La línea vertical indica el comienzo de la predicción realizada para las dos semanas siguientes a los datos utilizados para el ajuste y tomando los datos de fenología y temperatura observados en ese momento. La línea negra indica la mediana estimada y los sombreados indican los IC 60, 80 y 95% estimados. Los puntos indican los valores observados dejados fuera del conjunto de datos utilizado para el ajuste del modelo. La fenología del cultivo correspondiente a cada semana se indica en la parte superior del gráfico.

A continuación, se muestra un ejemplo de cómo funcionaría el modelo en tiempo real. Se informaría a los productores sobre la estimación de las densidades poblacionales *reales* a través del tiempo hasta el último monitoreo, junto con la densidad predicha para las próximas dos semanas (Figura 10) y la probabilidad de superar distintos umbrales de densidad de Equivalente Nezara que justificarían un control. (Tabla 1. Se muestra ejemplo para la localidad Zavalla).

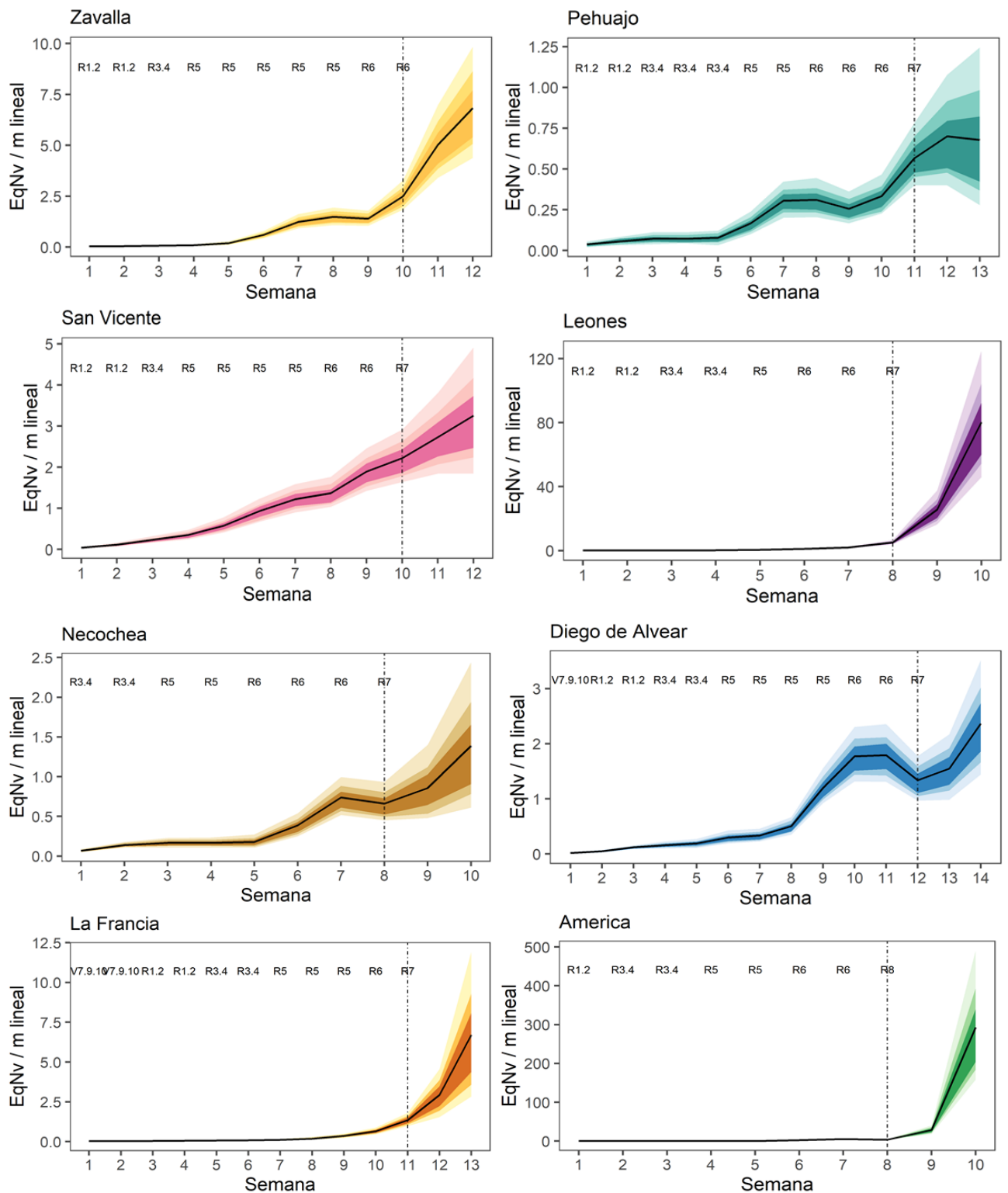


Figura 10. Estimación de densidad *real* de Equivalente Neza (EqNv) por metro y predicciones por medio de un modelo jerárquico para todas las localidades. La línea vertical indica el comienzo de la predicción, la cual se realiza para la semana siguiente a los datos y suponiendo que la fenología del cultivo es igual a la de la última semana de los datos. La línea negra indica la mediana y los sombreados indican los IC 60, 80 y 95%. La fenología del cultivo correspondiente a cada semana se indica en la parte superior del gráfico.

Tabla 1. Probabilidad de superar distintos umbrales de Equivalente Nezara para la Localidad Zavalla

Localidad	Umbral EqNv	Probabilidad de superar umbral EqNv (%)
Zavalla	6.0	16.43
Zavalla	5.5	30.83
Zavalla	5.0	50.44
Zavalla	4.5	72.72
Zavalla	4.0	90.04
Zavalla	3.5	98.31
Zavalla	3.0	99.87
Zavalla	2.5	100.00

Además de las estimaciones de densidad poblacional *real*, se realizaron estimaciones de daño acumulado de los tiempos de los datos y de tiempo futuro. Estas estimaciones fueron sobre la variable “unidades-umbral de daño acumulado”, por lo que el modelo no solo estima la probabilidad de superar un umbral que justifique un control, sino que también estima la probabilidad de superar el umbral de daño acumulado previamente definido por el productor.

Estas dos estimaciones (tiempo de datos y tiempo futuro) permiten ver si en algún momento se superó el umbral de daño acumulado preestablecido para alguna fenología determinada hasta el día del último dato recolectado y la probabilidad de superar un umbral determinado en la semana siguiente (Figura 11, Tabla 2. Ejemplo localidad Zavalla). En el caso de la localidad Zavalla, se puede observar que dentro de las semanas de los muestreos (semanas 1 a 10) se superó el umbral de daño acumulado de la fenología R5 en la semana 8 (mediana e IC mayores a 1) y que el modelo predice que hay una alta probabilidad de superar, en la semana 11, el umbral de daño acumulado de la fenología R6.

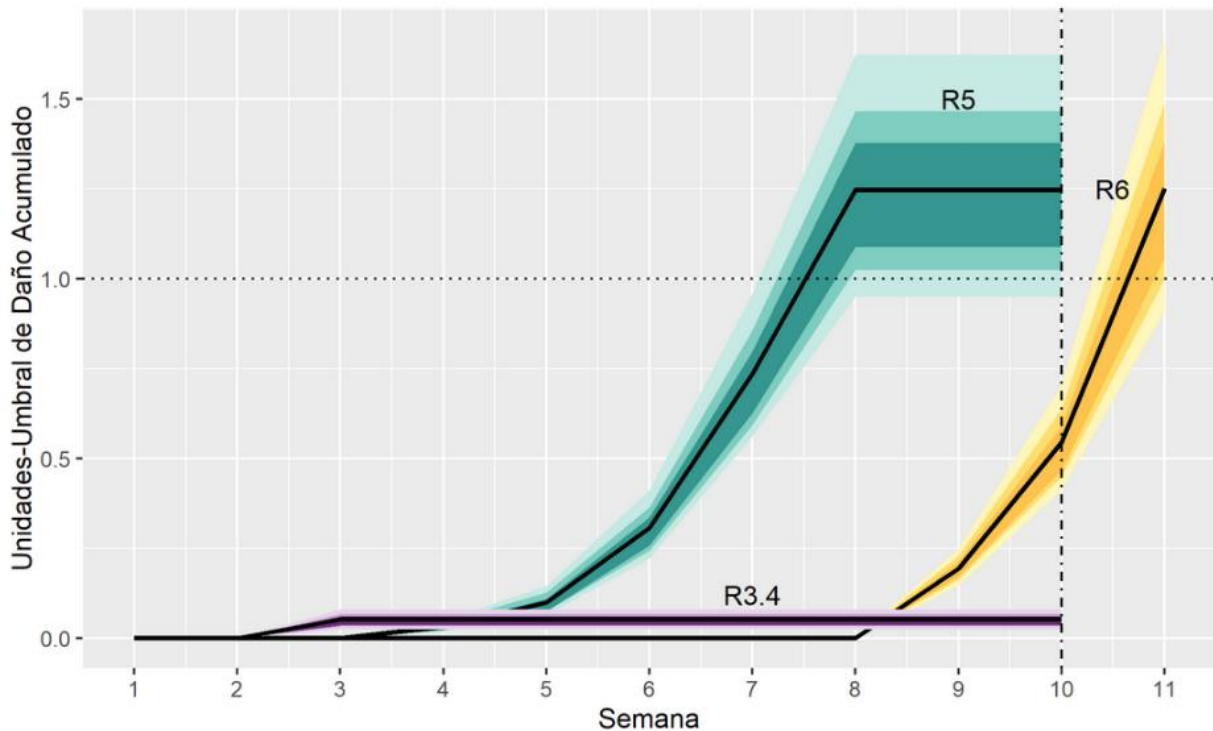


Figura 11. Unidades-umbral de daño acumulado según la fenología para la localidad Zavalla. La línea vertical indica el inicio de la predicción y la línea horizontal indica cuándo se alcanza, de forma acumulada, el umbral para cada fenología (Umbral utilizado de EqNv / m lineal: R3.4 = 1.4, R5 = 2.9, R6 = 7.1). Cada color indica la variable estimada para una fenología distinta, las líneas negras indican las medianas y los sombreados indican los IC 60, 80 y 95%.

Las predicciones de la localidad Zavalla (semana 11) para distintas unidades de la variable “unidades-umbral de daño acumulado” muestran cuán probable es que haya un daño acumulado mayor al umbral establecido (2 unidades de esta variable implican que el daño acumulado es equivalente a dos veces el umbral predefinido) (Tabla 2).

Tabla 2. Predicciones de distintos valores de la variable unidades-umbral de daño acumulado en R6 para la localidad Zavalla.

Localidad	Umbral-Unidad de Daño Acumulado en R6 utilizado	Probabilidad de superar el Umbral-Unidad de Daño Acumulado en R6 (%)
Zavalla	1.0	93.73
Zavalla	1.5	13.23
Zavalla	2.0	0.25
Zavalla	2.5	0.00

El análisis de incertidumbre en función del tamaño muestral para toda la serie temporal de datos mostró que, a medida que disminuye el tamaño muestral, las dinámicas poblacionales son más variables y ruidosas, lo que disminuye cuán confiables son las estimaciones realizadas (Figura 12. Ejemplo localidad San Vicente). Sumado a esto, el análisis de incertidumbre de las predicciones mostró que los modelos ajustados con tamaños muestrales pequeños tienden a tener predicciones con mucha incertidumbre (distribuciones con colas largas) y a subestimar los posibles valores de EqNv/m lineal (distribución a la izquierda de la línea discontinua que representa el valor observado) (Figura 13). La incertidumbre o varianza (promediada entre réplicas y relativa al mayor valor obtenido) en función del tamaño muestral mostró una relación negativa a medida que se utilizó un tamaño muestral mayor, empezando a saturar la reducción en la incertidumbre a partir de 60 muestras. Esto sugiere un tamaño muestral de aproximadamente 60 para obtener un buen balance entre esfuerzo de muestreo y precisión en las predicciones (Figura 14).

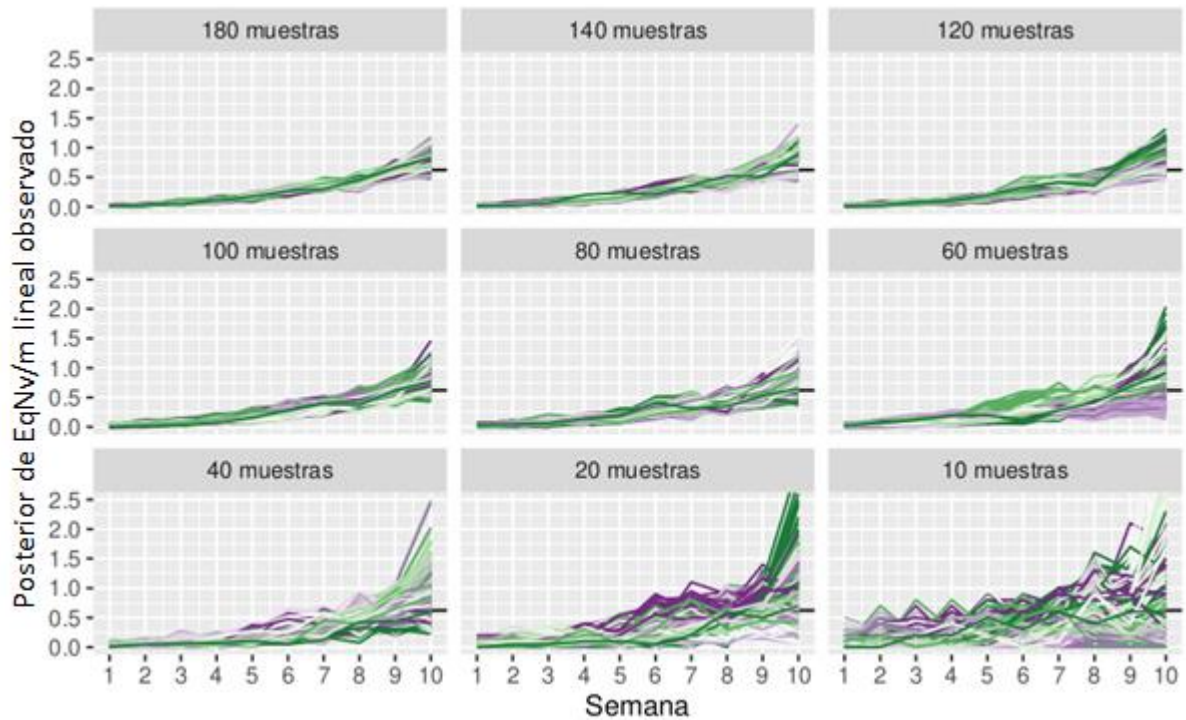


Figura 12. Simulaciones de la dinámica poblacional de EqNv/m lineal utilizando distintos tamaños muestrales. Cada panel representa un tamaño muestral distinto. Cada color representa una réplica distinta. Se grafican 80 simulaciones (líneas) por cada réplica (esto es, 80 líneas por cada color).

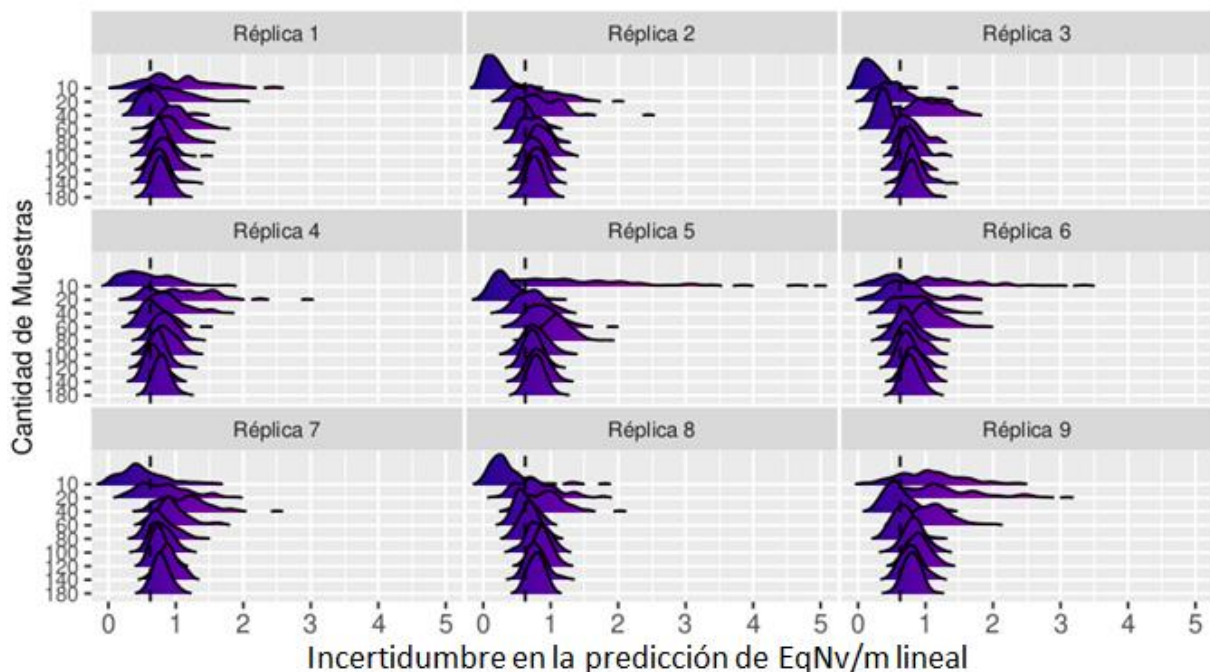


Figura 13. Incertidumbre en la predicción de EqNv/m lineal en la última semana de los datos para la localidad San Vicente. La línea vertical discontinua muestra el valor observado en la última semana y cada panel representa el resultado obtenido en una réplica distinta.

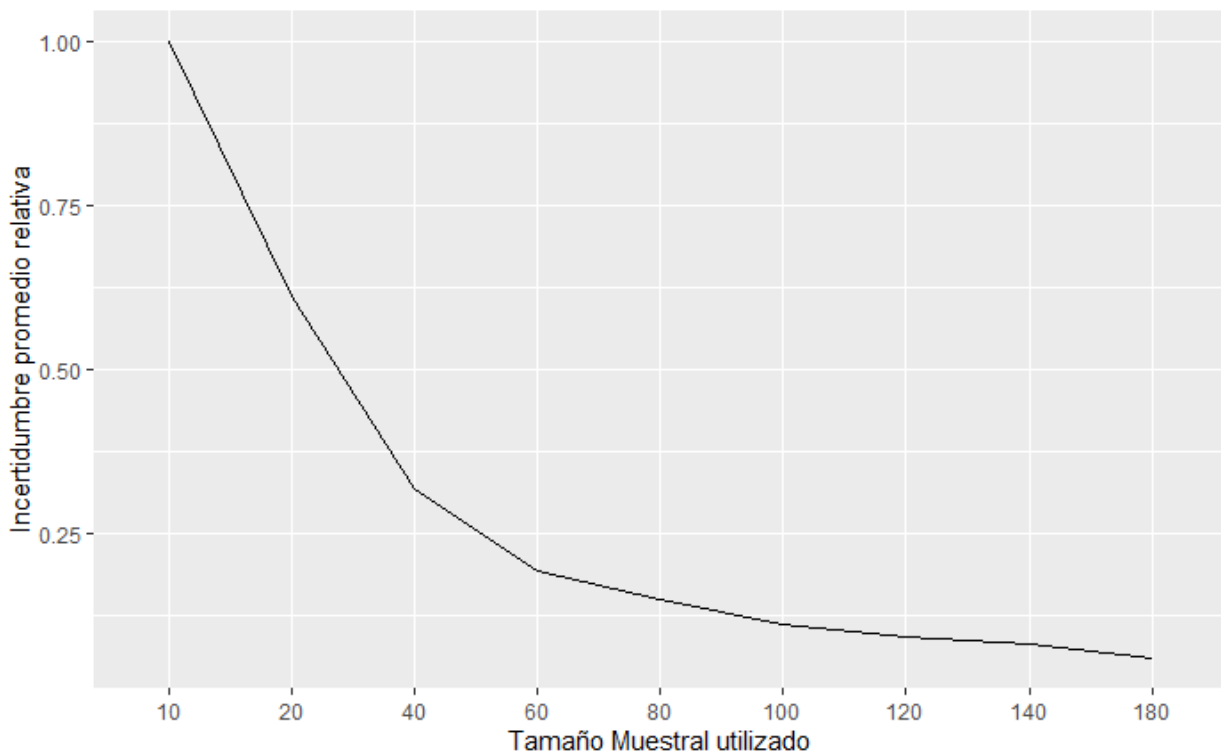


Figura 14. Incertidumbre promedio (entre réplicas) relativa (al mayor valor) en función del tamaño muestral para la localidad San Vicente.

En lo que respecta al ajuste del modelo, las posteriores de las distribuciones globales (aquellas caracterizadas por los parámetros globales μ , σ) de los principales parámetros del modelo (β_F , β_T , l) están centradas en 0, lo que sugiere la ausencia de un efecto generalizable a todas las localidades (Figura 15). A su vez, los parámetros por localidad que representan el efecto de las covariables ($\beta_{F,j}$, $\beta_{T,j}$) muestran una gran variabilidad inter-localidad (Figura 15). Por otro lado, la distribución posterior de la probabilidad de observación tomó valores considerablemente menores a la distribución previa elegida (Figura 16).

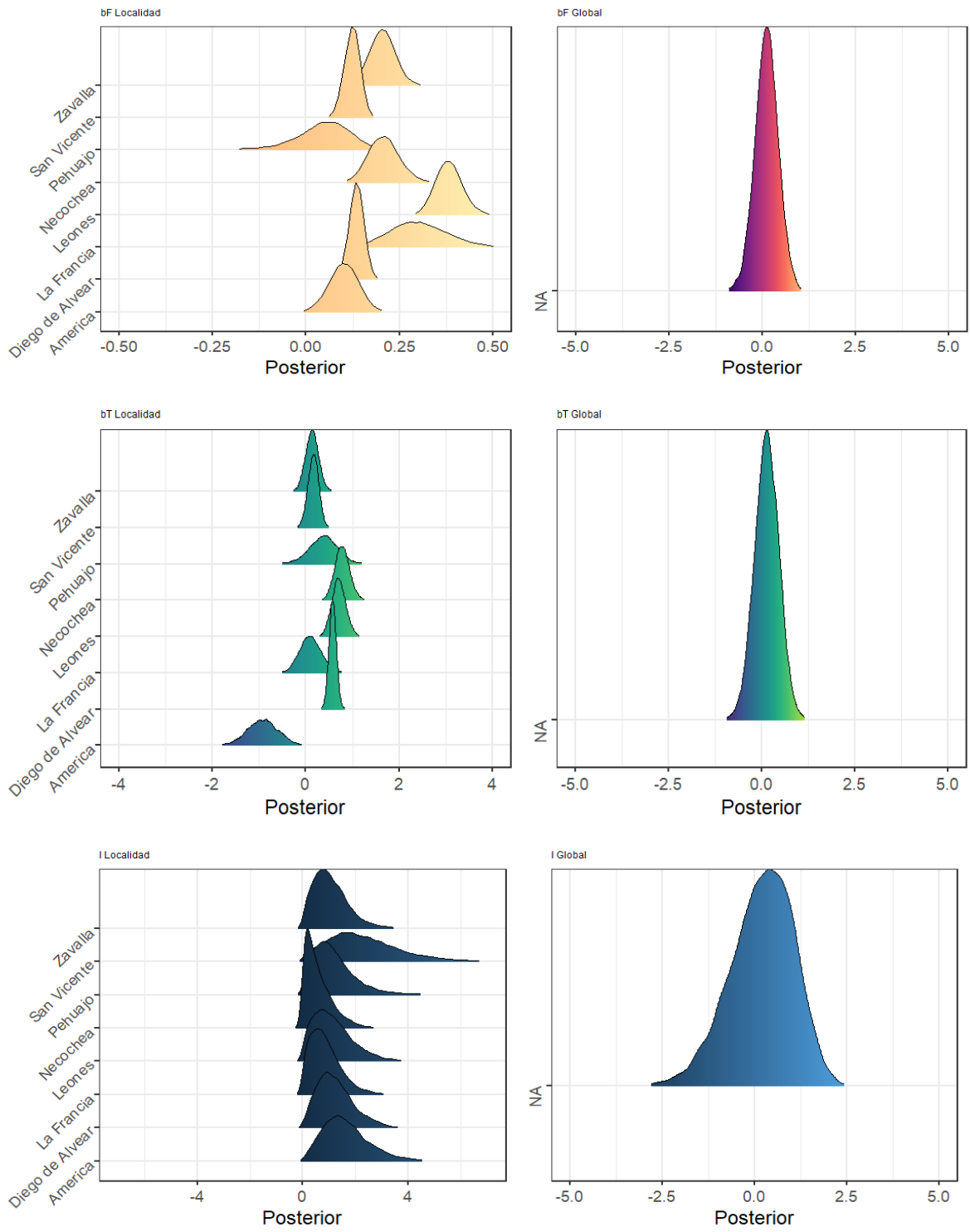


Figura 15. Distribuciones de probabilidad posteriores de los parámetros que representan el efecto de las covariables (β_F , β_T) y de la inmigración semanal de chinches grandes (β). Las figuras de la izquierda muestran las distribuciones posteriores por localidad; las de la derecha muestran la distribución posterior de la media (μ) que caracteriza la distribución global.

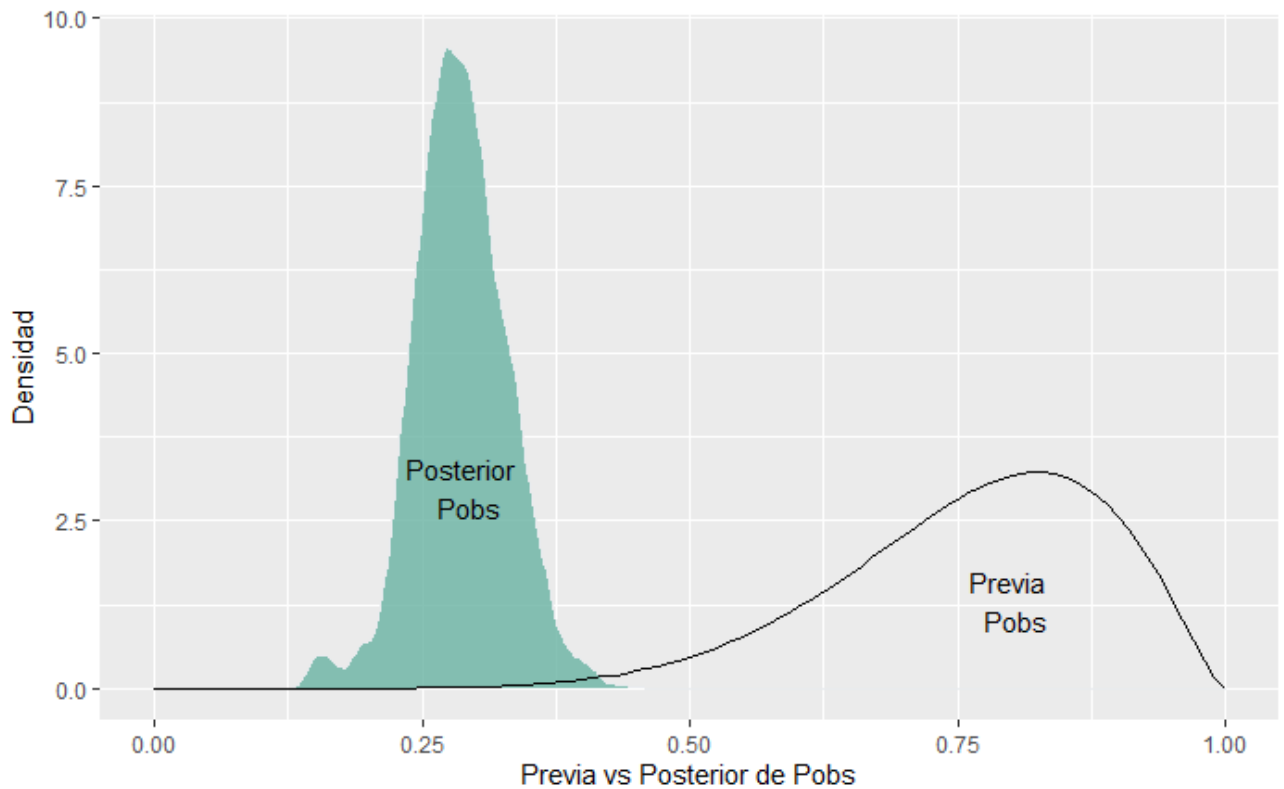


Figura 16. Posterior de la Probabilidad de observar una chinche con el método del paño vertical (P_{obs}) en contraste con la distribución previa elegida.

Por último, la comparación de modelos en términos del poder predictivo fuera de la muestra (WAIC y LOO-CV) mostró que el modelo base fue el mejor (Tabla 3. Modelo 1, $elpd$ con valor más cercano a 0), pero es muy cercano al poder predictivo del modelo con submodelo de inmigración. De todas formas, se utilizó el modelo base para realizar las estimaciones de la densidad poblacional real y las del daño acumulado, para realizar las predicciones de éstas y para realizar el análisis de incertidumbre en función del tamaño muestral.

Tabla 3. Comparación del poder predictivo fuera de la muestra. Arriba: WAIC; Abajo: LOO-CV. Modelo 1 = Modelo base; Modelo 2 = Modelo sin temperatura; Modelo 3 = Modelo con submodelo de inmigración. elpd: expected log predictive density, estimador bayesiano de LOO. DE: Desvío Estándar. p = estimación del número efectivo de parámetros. $WAIC = -2 * (elpd WAIC)$. $LOOIC = -2 * (elpd LOO)$

Modelo	Diferencia de elpd	DE de diferencia	WAIC elpd	DE de elpd WAIC	p WAIC	DE de p WAIC	WAIC	DE de WAIC
model1	0	0	-510.78	75.67	218.88	58.1	1021.57	151.33
model3	-1.11	11.07	-511.89	79.19	210.8	57.17	1023.78	158.37
model2	-60.24	30.26	-571.03	90.22	211.78	53.88	1142.06	180.43

Modelo	Diferencia de elpd	DE de diferencia	LOO elpd	DE de elpd LOO	p LOO	DE de p LOO	LOOIC	DE de LOOIC
model1	0	0	-523.45	65.82	231.55	47.62	1046.9	131.65
model3	-2.03	13.05	-525.48	72.11	224.39	49.76	1050.96	144.23
model2	-58.39	35.05	-581.84	86	222.59	49.58	1163.68	172

DISCUSIÓN

En este trabajo se apuntó a resolver dos problemas centrales a través del uso de un modelo con estructura espacio-estado: (1) la falta de un modelo que tenga en cuenta el error de observación de los muestreos con paño vertical para poder estimar la densidad poblacional *real* a partir de los valores observados; y (2) la falta de un modelo que realice predicciones sobre la densidad poblacional *real* evitando que el error de observación se propague en el tiempo.

El primero tiene importancia tanto biológica como productiva. Por un lado, la importancia biológica reside en brindar una herramienta que permita estudiar este sistema biológico sin sesgos y contemplando la variabilidad que poseen las observaciones (tal como tener un termómetro que esté calibrado y que se conozca su error para realizar un estudio). Por el otro, la importancia productiva reside en que la falta de una estimación de la densidad poblacional *real* genera pérdidas en el rinde de los cultivos, incluso en aquellos donde se realizan monitoreos periódicos y se utilizan umbrales para determinar en qué momento realizar un control. La razón de esto es que los monitoreos con paño vertical ofrecen una estimación de la

observación del proceso de dinámica poblacional de las chinches, pero no una estimación del proceso en sí (es decir, de la densidad poblacional *real* en un tiempo determinado). Esto es particularmente problemático en este sistema porque los monitoreos (que son realizados con paños verticales) tienden a dar valores poco precisos y sesgados, por más que hasta este momento el paño vertical es el método más preciso y eficiente que se encuentra disponible para el monitoreo de hemípteros fitófagos (Gamundi 1995). En la práctica, utilizar solo los valores observados podría estar dando lugar a decisiones erradas en cuanto al momento en el cual se debe realizar un control, por más que los umbrales se hayan definido en base a las densidades de chinches *observadas* en experimentos a campo o en jaulas (e.g., Iannone and Leiva 2004; Gamundi *et al.* 2003). Esto se debe a que no existe un trabajo que haya cuantificado de forma robusta el nivel de sesgo del paño vertical y que haya analizado si la magnitud del sesgo varía en función de alguna característica del cultivo (tal como estadio fenológico del cultivo o espaciamiento entre líneas de siembra), de alguna variable meteorológica o según el diseño experimental (a campo o en jaula). De ser la magnitud de este sesgo variable, los umbrales de control definidos en base a chinches observadas no son generalizables a todos los escenarios. Consideramos necesario tener un modelo de observación que permita estimar la densidad de chinches real y redefinir los umbrales de control en base a densidades reales, con el fin de reducir esta fuente de variabilidad y aumentar la eficacia de los umbrales de control.

El segundo problema que apuntó a resolver nuestro modelo, la falta de un modelo predictivo adecuado, tiene grandes implicancias productivas ya que brinda a los productores una herramienta de decisión que permite definir previamente qué densidad de chinches están dispuestos a tolerar (es decir, elegir un umbral) y, con la predicción del modelo, decidir si realizar un control tempranamente para evitar que se provoque una pérdida en el rinde. Se mostró un ejemplo de cómo funcionaría en tiempo real para los productores (Figura 10) y el modelo mostró un grado de exactitud aceptable en las predicciones (Figura 9), con la salvedad de que para realizarlas se utilizaron los valores observados de las covariables y en la realidad habría un mayor nivel de incertidumbre por tener que estimar el valor futuro de éstas. A pesar de que existen localidades donde el modelo no hizo buenas predicciones, este es un proceso iterativo que requiere de varias instancias de ajustar o plantear un distinto modelo y

volver a contrastar las predicciones contra los datos futuros para ir incrementando paulatinamente el nivel de exactitud y precisión.

Sumado a esto, si bien la capacidad predictiva brinda grandes ventajas en términos productivos, tener modelos que realicen predicciones a corto plazo y poder contrastarlas con los datos que se vayan adquiriendo también contribuye a avanzar el conocimiento sobre este sistema biológico. Este enfoque permite que exista un ciclo iterativo donde se formaliza una hipótesis sobre cómo funciona el sistema (es decir, se plantea un modelo), se hacen predicciones cuantitativas a corto plazo, se las contrasta con los nuevos datos adquiridos y luego se refina la hipótesis inicial o se plantea una alternativa. Esta alternativa será nuevamente contrastada con los datos para determinar si nuestra nueva hipótesis es más cercana a cómo funciona el sistema biológico realmente (Dietze *et al.* 2018).

En lo que respecta al daño acumulado que presentan los cultivos, se presentó el resultado de una primera aproximación a cuantificarlo (Figura 11). Si bien esta es solo una forma de monitorear la densidad poblacional acumulada a través del tiempo y relativizada a un umbral previamente definido, es una herramienta flexible que permite aumentar la información disponible para los productores sobre cuándo es el momento idóneo para realizar un control. De esta forma, los productores pueden no solo decidir realizar una aplicación cuando el umbral de control fue superado, sino también optar por realizar una aplicación si el umbral no fue superado pero el daño acumulado es alto. Tener esta información disponible es importante ya que se ha observado que densidades de chinches por debajo de los umbrales de control pero durante un tiempo prolongado pueden provocar mermas en el rendimiento de los cultivos (McPherson *et al.* 1979).

A pesar de que esta primera aproximación requirió supuestos fuertes (el daño por parte de las chinches es aditivo y su intensidad es independiente de la fenología del cultivo), sirve como una base para extraer la mayor cantidad de información posible de los monitoreos. De todas formas, es necesario realizar diseños experimentales que permitan poner estos supuestos a prueba y para generar medidas cuantitativas que permitan complejizar la estimación del daño acumulado.

En base al análisis de incertidumbre de las predicciones en función del tamaño muestral, se propone un tamaño muestral de 60 para obtener un buen balance entre el esfuerzo de muestreo y la precisión (Figura 14). Este resultado contrasta fuertemente con las 10 muestras que toman, en promedio, los productores y evidencia la necesidad de aumentar el esfuerzo de muestreo para tener un nivel de precisión aceptable.

En cuanto a la distribución posterior de la probabilidad de observación (P_{obs}), el hecho de que haya tomado valores tan bajos (Media = 0,31, DE = 0,04. Ver Tabla 4 en Apéndice) a pesar de la distribución previa establecida (Figura 7) sugiere que el error de observación que tiene el paño vertical es mayor al estimado por la AAPPCE y refleja la necesidad de realizar experimentos a campo que cuantifiquen este error.

Entre las limitaciones que presenta este modelo se encuentra la falta de datos meteorológicos detallados y la falta de un modelo que prediga la fenología vegetal. La primera se debe a que los datos meteorológicos que utilizamos del NOAA para las localidades y fechas de interés se limitaron a datos de temperatura media, mínima y máxima diarias, y los datos de precipitaciones eran demasiado escasos para ser utilizados. Sería conveniente tener datos de temperatura por hora para poder informar mejor al modelo de cómo los cambios en la temperatura afectan la dinámica poblacional de chinches, más allá de las tendencias generales que se pueden obtener con la temperatura media, y obtener datos completos de precipitaciones para evaluar si esta covariable incrementa el poder predictivo del modelo. La segunda limitación surge de que la dinámica poblacional de chinches depende fuertemente de la fenología vegetal. Esto hace que las predicciones de la densidad poblacional estén fuertemente condicionadas por la fenología vegetal que se supone que estará presente, pudiendo dar lugar a predicciones erradas cuando no se cumple el supuesto de que la fenología vegetal es igual a la de la semana anterior.

Por último, si bien los modelos espacio-estado han sido utilizados con éxito para realizar predicciones en un diverso rango de campos, este es el primer trabajo en utilizarlos para modelar la dinámica poblacional de hemípteros fitófagos en cultivos de soja. Existen muchos modelos de insectos que han sido desarrollados, pero en su mayoría son modelos mecanísticos y frecuentistas de desarrollo de insectos basados en temperatura para determinar el momento de transición de estadio (Ver: Damos y

Savopoulou-Soultani 2011). Este trabajo se distingue de los previos en que presenta un enfoque empírico y bayesiano para modelar la dinámica poblacional de chinches en cultivos de soja contemplando el error de observación del método de muestreo.

CONCLUSIÓN

En este trabajo, hemos presentado un método para obtener una estimación robusta y no sesgada de la densidad poblacional de chinches en cultivos de soja en un momento determinado, para obtener una estimación del daño acumulado que presenta un cultivo, para realizar predicciones sobre este daño y para analizar la incertidumbre en función del tamaño muestral. Esto permite a los productores tomar una decisión mejor informada acerca de cuál es el momento adecuado para realizar un control, aumentando así la eficacia y la eficiencia de los programas basados en el manejo integrado de plagas. Esto, a su vez, contribuirá a lograr una mayor adopción de este enfoque por parte del sector productivo, dando lugar al uso racional de pesticidas sintéticos, al cuidado de los recursos naturales y a una producción agrícola sustentable en general.

BIBLIOGRAFÍA

Altieri M. & Nicholls C. (2000) Agroecology in action. Indigenous and modern approaches to IPM in Latin America. ESPM Division of Insect Biology. *Univ. Calif. Berkeley USA*.

Bajwa W. I. & Kogan M. (2002) Compendium of IPM definitions (CID). What is IPM and how is it defined in the worldwide literatura? IPPC Publication No. 998, Integrated Plant Protection Center (IPPC), *Oregon State University, Corvallis, OR 97331, USA*.

Baum L. E. & Eagon J. A. (1967) An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. *Bull. Am. Math. Soc.* **73** , 360–363.

Baum L. E. & Petrie T. (1966) Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *Ann. Math. Stat.*

Boller E., El Titi A., Gendrier J., Avilla J., Jörg E. & Malavolta C. (1999) Integrated production: principles and technical guidelines. *Bull. OILB SROP Fr.*

Bowling C. (1980) The stylet sheath as an indicator of feeding activity by the southern green stink bug on soybeans. *J. Econ. Entomol.* **73** , 1–3.

Brier H. & Rogers D. (1991) Susceptibility of soybeans to damage by *Nezara viridula* (L.) (Hemiptera: Pentatomidae) and *Riptortus serripes* (F.) (Hemiptera: Alydidae) during three stages of pod development. *Aust. J. Entomol.* **30** , 123–128.

Cancelado R. E. & Radcliffe E. B. (1979) Action thresholds for green peach aphid on potatoes in Minnesota. *J. Econ. Entomol.* **72** , 606–609.

Correa-Ferreira B. S. & Panizzi A. R. (1999) Percevejos da soja e seu manejo. *Embrapa Soja-Circ. Téc. INFOTECA-E*.

Damos P. & Savopoulou-Soultani M. (2011) Temperature-Driven Models for Insect Development and Vital Thermal Requirements. *Psyche (Stuttg.)*. **2012**, e123405.

Dent D. & Binks R. H. (2020) *Insect Pest Management, 3rd Edition*. CABI.

Depieri R. & Panizzi A. (2011) Duration of feeding and superficial and in-depth damage to soybean seed by selected species of stink bugs (Heteroptera: Pentatomidae). *Neotrop. Entomol.* **40** , 197–203.

Dietze Michael C. (2017) Prediction in ecology: a first-principles framework. *Ecol. Appl.* **27** , 2048–2060.

Dietze Michael C (2017) *Ecological Forecasting*. Princeton University Press.

Dietze M. C., Fox A., Beck-Johnson L. M. *et al.* (2018) Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proc. Natl. Acad. Sci.* **115** , 1424–1432.

Duggal N. & Siddiqi Z. (2008) Providing decision making analytical tools to IPM managers through web based: electronic pest monitoring, and pesticide use reporting system. In: pp. 275–280 6th International Conference on Urban Pests, Budapest, Hungary, 13-16 July 2008. International Conference on Urban Pests (ICUP).

Fehr W. & Caviness C. (1977) Stages of soybean development. *Spec. Rep.*

Flint M. L. & Van den Bosch R. (2012) *Introduction to integrated pest management*. Springer Science & Business Media.

Gamundi J. (2007) Control de hemípteros fitófagos en el cultivo de soja. *En EV Trumper JD Edelstein Eds Chinchas Fitófagas En Soja Revisión Av. En El Estud. Su Ecol. Manejo.*

Gamundi J. C. (1995) Evaluación de técnicas de muestreo de insectos plaga y depredadores en cultivos de soja con diferentes sistemas de siembra y labranza. In: Primer Congreso Nacional de Soja, Segunda Reunión Nacional de Oleaginosos. AIANBA, Pergamino (Bs. As.).

Gamundi J. C., Andrian M., Bacigualuppo D. *et al.* (2003) Incidencia del complejo de chinches en el cultivo de soja con diferentes espaciamientos entre líneas. *Soja. Para mejorar la producción de soja*. INTA, Estación Experimental Agropecuaria Oliveros. Publicaciones Regionales, 24:79-86.

Gamundi J. C. & Sosa M. A. (2008) Caracterización de daños de chinches en soja y criterios para la toma de decisiones de manejo. *Chinchas Fitófagas En Soja Revisión Av. En El Estud. Su Ecol. Manejo.*

Iannone N. (2004) Niveles de daños de chinches en soja. *Día Campo Jorn. Prof. Agríc. 2004 2004 02 26 26 Febrero 2004 Gen. Villegas AR.*

Iannone N. & Leiva P. D. (2004) Daños, toma de decisiones y control cultural de chinches en soja. *Carpeta de producción vegetal*. XIII: 4p

Ihaka R. & Gentleman R. (1996) R: A Language for Data Analysis and Graphics. *J. Comput. Graph. Stat.* **5** , 299–314.

Kuss R. C. R. M., Guedes J. V. C., Fiorin R. A., Rodríguez R. B., Moser G. R. Z. & Stecca C. dos S. (2007) Levantamento georreferenciado da ocorrência de *Piezodorus guildinii* na cultura da soja.

Lefebvre M., Langrell S. R. H. & Gomez-y-Paloma S. (2015) Incentives and policies for integrated pest management in Europe: a review. *Agron. Sustain. Dev.* **35** , 27–45.

Lewis W. J., Van Lenteren J., Phatak S. C. & Tumlinson J. (1997) A total system approach to sustainable pest management. *Proc. Natl. Acad. Sci.* **94** , 12243–12248.

Lichtfouse E., Navarrete M., Debaeke P., Souchère V., Alberola C. & Ménassieu J. (2009) Agronomy for sustainable agriculture. A review. *Agron. Sustain. Dev.* **29** , 1–6.

- Massaro R. A. (2005) Instrumentación del control integrado de plagas. Una historia de 25 años en el área pampeana. INTA - Estación Experimental Agropecuaria Rafaela. Información técnica cultivos de verano. Publicación miscelánea N° 104
- McPherson R. M., Newsom L. D. & Farthing B. F. (1979) Evaluation of Four Stink Bug Species From Three Genera Affecting Soybean Yield and Quality in Louisiana. *J. Econ. Entomol.* **72** , 188–194.
- Panizzi A., Galileo M., Gastal H., Toledo J. & Wild C. (1980) Dispersal of *Nezara viridula* and *Piezodorus guildinii* nymphs in soybeans. *Environ. Entomol.*
- Panizzi A. R. & Slansky F. (1985) Review of Phytophagous Pentatomids (Hemiptera: Pentatomidae) Associated with Soybean in the Americas. *Fla. Entomol.* **68** , 184–214.
- Pedigo L. P., Hutchins S. H. & Higley L. G. (1986) Economic injury levels in theory and practice. *Annu. Rev. Entomol.* **31** , 341–368.
- Plummer M. (2003) JAGS: A Program for Analysis of Bayesian Graphical Models using Gibbs Sampling. *3rd Int. Workshop Distrib. Stat. Comput. DSC 2003 Vienna Austria.*
- RStudio Team (2020) RStudio: Integrated Development for R. [online]. Available from: <http://www.rstudio.com/>.
- Schreinemachers P. & Tipraqsa P. (2012) Agricultural pesticides and land use intensification in high, middle and low income countries. *Food Policy* **37** , 616–626.
- Shiffrin R. M., Lee M. D., Kim W. & Wagenmakers E.-J. (2008) A Survey of Model Evaluation Approaches With a Tutorial on Hierarchical Bayesian Methods. *Cogn. Sci.* **32**, 1248–1284.
- Stern V., Smith R., van den Bosch R. & Hagen K. (1959) The integration of chemical and biological control of the spotted alfalfa aphid: The integrated control concept. *Hilgardia* **29** , 81–101.
- Swanson D. (2012) An updated synopsis of the Pentatomoidea (Heteroptera) of Michigan. *Gt. Lakes Entomol.*
- Todd J. W. (1989) Ecology and Behavior of *Nezara Viridula*. *Annu. Rev. Entomol.* **34** , 273–292.
- Todd J. W., Jellum M. D. & Leuck D. B. (1973) Effects of Southern Green Stink Bug 1 Damage on Fatty Acid Composition of Soybean Oil 2 , 3. *Environ. Entomol.* **2** , 685–690.
- Vehtari A., Gelman A. & Gabri J. (2019) loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models.
- Vehtari A., Gelman A. & Gabry J. (2017) Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27** , 1413–1432.
- Vicentini R. & Jimenez H. A. (1977) *El vaneó de los frutos en soja*. INTA.

Vyavhare S. S., Way M. O. & Medina R. F. (2015) Determination of Growth Stage-Specific Response of Soybean to Redbanded Stink Bug (Hemiptera: Pentatomidae) and its Relationship to the Development of Flat Pods. *J. Econ. Entomol.* **108** , 1770–1778.

Watanabe S. & Opper M. (2010) Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.* **11**.

Wikle C. K., Berliner L. M. & Cressie N. (1998) Hierarchical Bayesian space-time models. *Environ. Ecol. Stat.* **5** , 117–154.

Yeargan K. V. (1977) Effects of green stink bug damage on yield and quality of soybeans. *J. Econ. Entomol.* **70** , 619–622.

APÉNDICE

Simulación predictiva previa

El modelo basado únicamente en las distribuciones previas pudo generar datos similares a lo que podría ocurrir en este sistema biológico (los puntos se encuentran dentro de los IC), validando así la elección de distribuciones previas (Figura 17).

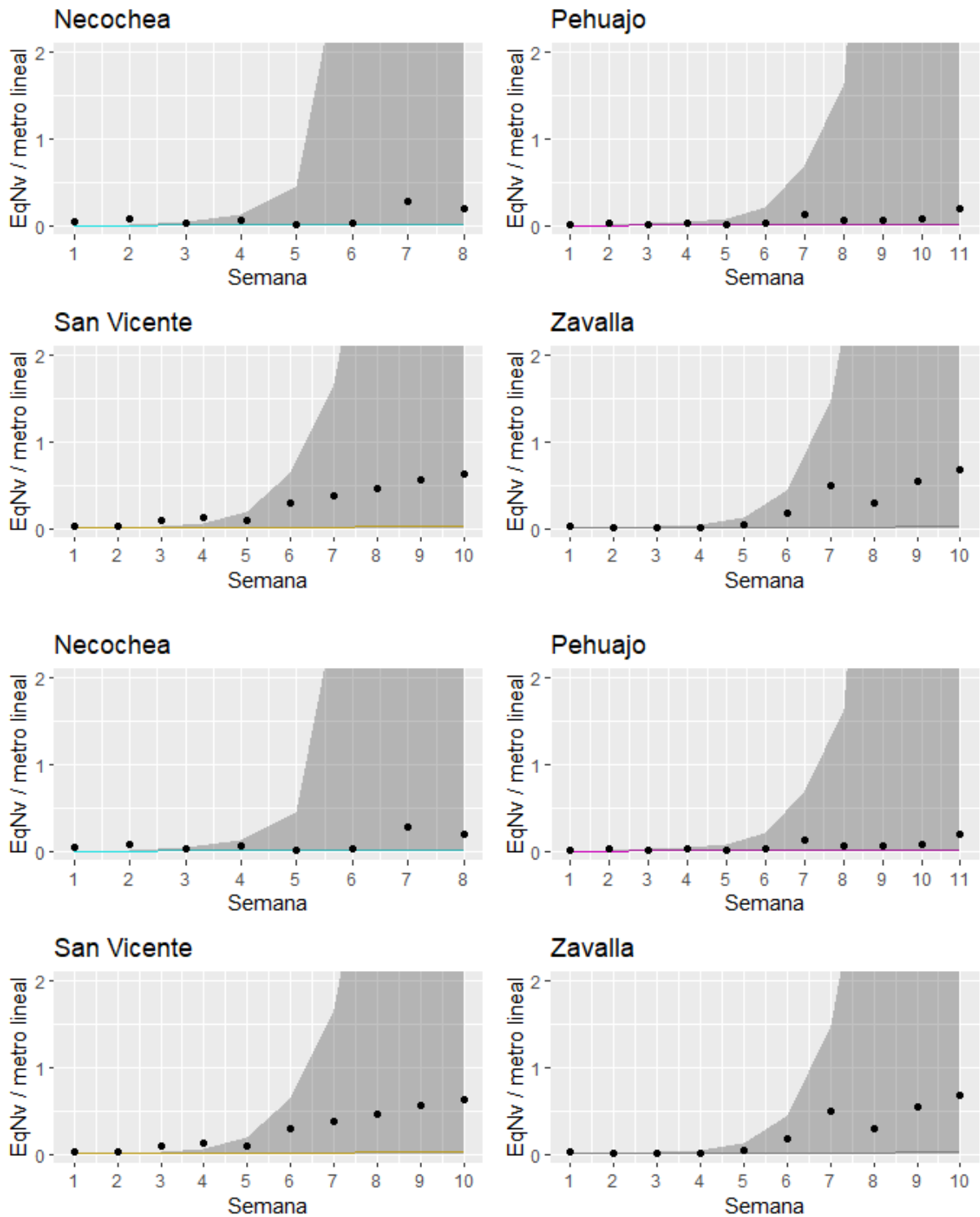


Figura 17. Resultados de las simulaciones predictivas previas para cada una de las localidades. Las líneas indican la mediana y el sombreado indica el IC 89%. Los puntos indican los valores observados.

Simulación predictiva posterior

El modelo ajustado con los datos pudo reproducir de forma bastante acertada los valores observados (los puntos se encuentran cerca de la mediana y general dentro del IC 89%), lo que indica que el modelo incorporó correctamente la información proveniente de los datos (Figura 18).

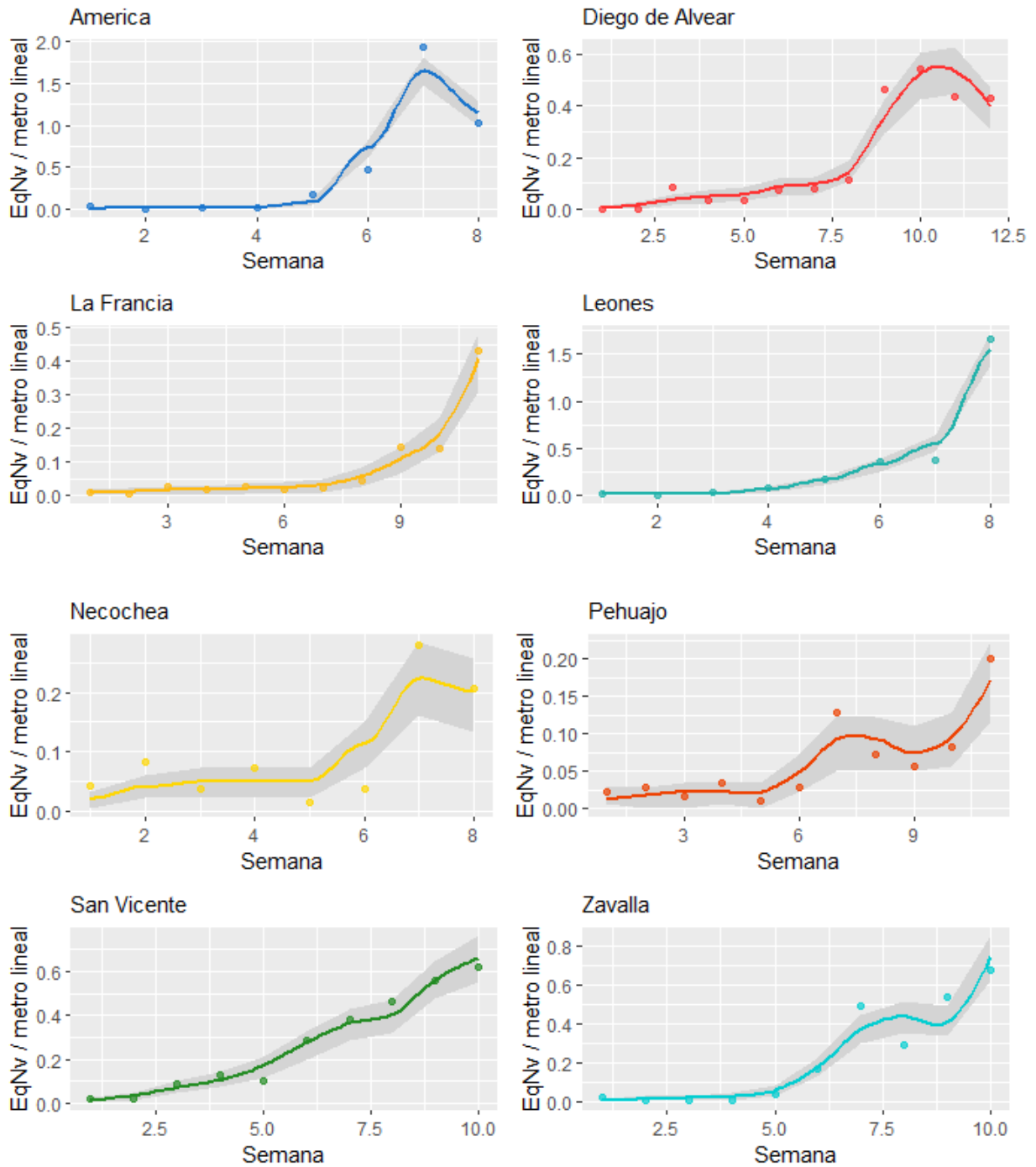


Figura 18. Resultados de las simulaciones predictivas posteriores para cada una de las localidades. Las líneas indican la mediana y el sombreado indica el IC 89%. Los puntos indican los valores observados.

Posterior de parámetros

Tabla 4. Resumen del ajuste del modelo brindado por JAGS. Rhat: diagnóstico de convergencia de cadenas (valores cercanos a 1.0 indican buena mezcla de cadenas). n efectivo: medida del número efectivo de muestras de la posterior de cada parámetro. DE: Desvío Estándar. f: proporción de la posterior con el mismo signo que la media. Los números entre corchetes indican la localidad, en orden ascendente estas son: América, Diego de Alvear, La Francia, Leones, Necochea, Pehuajo, San Vicente, Zavalla.

Parámetros	Media	DE	Rhat	n efectivo	f
pobs.EqNv.dat	0.31	0.04	1.02	120	1
I.EqNv.dat[1]	10.35	2.98	1	535	1
I.EqNv.dat[2]	4.97	2.23	1	2280	1
I.EqNv.dat[3]	1.6	1.22	1	6477	1
I.EqNv.dat[4]	1.42	1.24	1	3010	1
I.EqNv.dat[5]	2.58	2.2	1	2444	1
I.EqNv.dat[6]	1.35	1.2	1	2858	1
I.EqNv.dat[7]	11.14	4.74	1	1294	1
I.EqNv.dat[8]	3.01	1.92	1	2938	1
bF.EqNv.dat[1]	0.07	0.01	1	33149	1
bF.EqNv.dat[2]	0.09	0.01	1	8033	1
bF.EqNv.dat[3]	0.18	0.04	1	45000	1
bF.EqNv.dat[4]	0.32	0.02	1	3133	1
bF.EqNv.dat[5]	0.09	0.02	1	34442	1
bF.EqNv.dat[6]	0.1	0.03	1	24342	1
bF.EqNv.dat[7]	0.07	0.01	1	3161	1
bF.EqNv.dat[8]	0.16	0.02	1	15937	1
bT.A.EqNv.dat[1]	-1.66	0.1	1	1290	1
bT.A.EqNv.dat[2]	0.35	0.05	1	5495	1
bT.A.EqNv.dat[3]	-0.01	0.12	1	45000	0.52
bT.A.EqNv.dat[4]	0.46	0.1	1	3708	1
bT.A.EqNv.dat[5]	0.3	0.1	1	45000	1
bT.A.EqNv.dat[6]	0.36	0.12	1	20916	1
bT.A.EqNv.dat[7]	0.14	0.08	1	17076	0.96
bT.A.EqNv.dat[8]	0.31	0.06	1	31806	1
muF.EqNv.dat	0.09	0.31	1	45000	0.62
muT.EqNv.dat	0.02	0.33	1	24088	0.52
mul.EqNv.dat	0.08	1	1	45000	0.53
sigmaF.EqNv.dat	0.93	0.42	1	26545	1
sigmaT.EqNv.dat	0.7	0.32	1	36432	1
sigmal.EqNv.dat	0.04	0.04	1.01	2462	1