

Metodología para el modelado y la estimación de parámetros del proceso de crecimiento de *Lobesia botrana*.

Estefania Aguirre-Zapata^{a,b,*}, Jose Garcia-Tirado^c, Humberto Morales^a, Fernando di Sciascio^a, Adriana N. Amicarelli^a

^aInstituto de Automática (INAUT)-Universidad Nacional de San Juan - Consejo Nacional de Investigaciones Científicas y Técnicas. San Juan, Argentina

^bUniversidad Nacional de Colombia, Grupo de investigación Kalman, Medellín, Colombia

^cCenter for Diabetes Technology, University of Virginia, Charlottesville, VA, USA

To cite this article: Aguirre-Zapata, E., Garcia-Tirado, J., Morales, H., di Sciascio, F., Amicarelli, A. 2023. Methodology for modeling and parameter estimation of the growth process of *Lobesia botrana*. Revista Iberoamericana de Automática e Informática Industrial 20, 68-79. <https://doi.org/10.4995/riai.2022.17746>

Resumen

Lobesia botrana (*L. botrana*), es una plaga cuarentenaria que provoca daños a la vid, y genera pérdidas económicas para la región de Cuyo en Argentina. Diferentes investigaciones han buscado salvaguardar la integridad de los viñedos, generando sistemas de alerta basados en modelos que permitan detectar los picos de ocurrencia de la plaga, y conocer el proceso de crecimiento de la polilla, de acuerdo a las condiciones ambientales de cada región. En este trabajo, se propone una metodología para la estimación de parámetros desconocidos en los modelos semi físicos basados en primeros principios (MSBPP), con una aplicación particular en el modelo de crecimiento de *L. botrana*, en condiciones de laboratorio. La principal contribución consiste en una metodología para la estimación de parámetros de un MSBPP, que considera un modelo matemático desarrollado por los autores en un trabajo previo, el análisis de identificabilidad estructural del modelo en cuestión y la estimación del conjunto de parámetros desconocidos que cumplen con la propiedad de identificabilidad estructural. En este trabajo se consideran, como herramientas principales para la estimación, el algoritmo de mínimos cuadrados no lineales, y un Filtro de Kalman Extendido. Se evidenció una mejoría en el ajuste del modelo matemático a los datos experimentales, con relación a los obtenidos previamente. Además, se estableció el grado de afinidad de cada estadio de crecimiento por el factor limitante del mismo, y se presentaron nuevos perfiles de mortalidad.

Palabras clave: Modelado e identificación de sistemas biológicos, Estimación paramétrica, Modelado de caja gris, *Lobesia botrana*, Mínimos cuadrados no lineales, Identificabilidad estructural.

Methodology for modeling and parameter estimation of the growth process of *Lobesia botrana*

Abstract

Lobesia botrana (*L. botrana*), is a quarantine pest that causes damage to grapevines and generates economic losses for the region of Cuyo in Argentina. Different researchers have sought to safeguard the integrity of the vineyards, generating alert systems based on models that allow detecting the peaks of occurrence of the pest, and knowing the growth process of the moth, according to the environmental conditions of each region. In this work, a methodology for estimating unknown parameters in semi-physical models based on first principles (MSBPP) is proposed, with a particular application in the growth model of *L. botrana* under laboratory conditions. The main contribution consists of a methodology for parameter estimation of an MSBPP, which considers a mathematical model developed by the authors in previous work, the structural identifiability analysis of the model in question, and the estimation of the set of unknown parameters that meet the structural identifiability property. In this work, the non-linear least squares algorithm and an Extended Kalman Filter are considered the main estimation tools. An improvement in the adjustment of the mathematical model to the experimental data was evidenced, in relation to those previously obtained. In addition, the degree of affinity of each growth stage for its limiting factor was established, and new mortality profiles were presented.

Keywords: Modeling and identification of biological systems, Parameter estimation, Gray box modeling, *Lobesia botrana*, Nonlinear least-squares, Structural identifiability.

*Autor para correspondencia: eaguirre@inaut.unsj.edu.ar

1. Introducción

La industria vitivinícola es de gran importancia económica para la Argentina, genera ingresos por aproximadamente 800 millones de dólares al año (de Estudios Económicos de Bodegas de Argentina, 2020). La producción de vino en este país, se concentra principalmente en las provincias de Mendoza y San Juan, con una representación del 94 % de la producción total, seguido de otras provincias con porcentajes de participación que van entre 0,05 % - 0,5 %. Por lo anterior, es prioridad nacional el diseño de estrategias que permitan salvaguardar la integridad de los cultivos frente a posibles amenazas como *L. botrana* que puedan afectar la producción de vinos. *L. botrana* es una plaga cuarentenaria en Argentina que produce daños a la vid en sus diferentes estados fenológicos (Torres Vila, 1995), que se manifiestan en la destrucción de las bayas y floriscencias, disminuyendo la cantidad y calidad de la cosecha y ocasionando mermas en la producción de vinos (Burgardt et al., 2010). Desde el punto de vista legislativo, las autoridades nacionales trabajan en acciones preventivas que obliguen a los productores a mejorar el control de la polilla de la vid mediante convocatorias estacionales adecuadas junto con normativas actualizadas (Carmona Torres, 2019). Para la región de Cuyo (Mendoza - San Juan - San Luis) en particular, el Servicio de Calidad (SENASA) declaró una alerta fitosanitaria que obliga al reporte de la presencia de *L. botrana* en las vides (SENASA, 2009).

Desde el punto de vista científico, recientes investigaciones han centrado sus esfuerzos en el desarrollo de modelos matemáticos para predecir el ciclo biológico de la polilla, basados en información climática, con el fin de facilitar el diseño de estrategias para el control de la misma. Así pues, se encuentran en la literatura modelos, como el presentado por Touzeau en el año 1981 (Touzeau, 1981), y que se ha constituido como uno de los modelos más utilizados para representar el proceso de crecimiento de *Lobesia botrana* con relación a la temperatura. Trabajos más recientes, como los presentados por Lessio y Alma en (Lessio and Alma, 2021), presentaron un panorama amplio con relación al modelado de *L. botrana* y otras plagas que afectan el desarrollo de la vid, destacando que, si bien los modelos de caja negra o derivados del ajuste de datos, son utilizados para predecir el comportamiento o aparición de la polilla para ciertas condiciones climáticas dadas, carecen de una base conceptual que facilite su uso por parte de viticultores, enólogos, profesionales de soporte técnico, servicios fitosanitarios y otras partes interesadas. Para dar solución a lo anterior, se encuentran en la literatura, modelos semi físicos basados en primeros principios (FPBSM - por sus siglas en inglés) como el presentado en (Aguirre-Zapata et al., 2022). Estos modelos, al heredar su estructura de los primeros principios o leyes físicas, se hallan dotados de buena interpretabilidad y en muchos casos con ello se asocia una gran capacidad de predecir el comportamiento del proceso, incluso por fuera del entorno en el que se ubican los puntos experimentales con los que se identifican sus parámetros empíricos (Lema-Perez, 2018). El proceso de ajuste o estimación de parámetros para este tipo de modelos es necesario para completar la estructura matemática de los mismos.

Durante el proceso de construcción de modelos de caja gris o MSBPP, se originan parámetros del sistema que tienen valo-

res desconocidos que deben ser determinados mediante datos provenientes, ya sea a partir de la experimentación o de fuentes bibliográficas (Ljung and Torkel, 1994). Sin embargo, el proceso de estimación de parámetros en este tipo de modelos suele presentar dificultades, que en la mayoría de los casos surgen por la falta de identificabilidad estructural de los mismos (Chis et al., 2011a). La propiedad de identificabilidad resulta indispensable tanto para el proceso de estimación de parámetros, como para el diseño óptimo de experimentos, tarea que solo es posible cuando se dispone de un conjunto de datos lo suficientemente informativo.

En los últimos años, se han publicado numerosos trabajos que exploran algoritmos y metodologías para la identificación de parámetros en modelos de caja gris. Aspiron y colaboradores, abordaron el problema del modelado de caja gris en procesos químicos, destacando las bondades de los modelos de caja gris para el monitoreo y control de procesos, y la descripción teórica de bioprocesos. En su trabajo, los autores propusieron una metodología que consideraba la selección y procesamiento de los datos como tarea previa al uso de algoritmos de optimización, para la estimación de parámetros. Sin embargo, no consideraron la propiedad de identificabilidad como un requisito previo (Aspiron et al., 2019). De manera similar, en el trabajo desarrollado por Baulmet y Dostál en el año 2020, si bien se proponen técnicas novedosas para la estimación de parámetros en modelos matemáticos de caja gris, los autores no consideran la propiedad de identificabilidad como requisito para la implementación de algoritmos de estimación de parámetros (Bäumelt and Dostál, 2020). En el año 2019, Brastein y colaboradores, propusieron el uso de los análisis de identificabilidad práctica e identificabilidad estructural, para determinar como el diseño óptimo de experimentos afecta la información dinámica contenida en los datos utilizados para la estimación de parámetros de un modelo de caja gris (Brastein et al., 2019). En el año 2021, Hotvedt y colaboradores, propusieron el uso de la regularización de parámetros como una estrategia para conservar la interpretabilidad física de un modelo, aun cuando sus parámetros no son identificables. En su trabajo, los autores destacaron la dificultad para garantizar la identificabilidad en modelos de caja gris debido a su elevado número de parámetros, y afirmaron que el rendimiento predictivo de los modelos y su interpretabilidad, está influenciada por los datos disponibles para el proceso de ajuste (Hotvedt et al., 2021). Los hallazgos anteriores dejan en evidencia que el concepto de identificabilidad paramétrica no es considerado de manera formal en las metodologías para la estimación de parámetros utilizadas comúnmente. Como lo explican García y Colaboradores en (García-Tirado et al., 2018), una posible explicación de por qué dicha propiedad es pasada por alto, tiene que ver con la disponibilidad de equipos de cómputo potentes capaces de resolver problemas complejos de optimización.

En este trabajo se formaliza una metodología para la estimación de parámetros desconocidos en modelos semi físicos basados en primeros principios. En particular, se aplica la metodología propuesta para el ajuste de las tasas de mortalidad y las constantes de velocidad media de crecimiento de un modelo de crecimiento de *Lobesia botrana* desarrollado anteriormente

en (Aguirre-Zapata et al., 2022). En la Sección 2 se presenta la metodología propuesta que consta de tres pasos que se describen a lo largo de la sección. En la Sección 3 se presentan los resultados obtenidos, comparando el modelo matemático ajustado y con los parámetros determinados, con datos experimentales reales del proceso. Además, se presenta comparación de los parámetros ajustados siguiendo esta metodología y la empleada anteriormente.

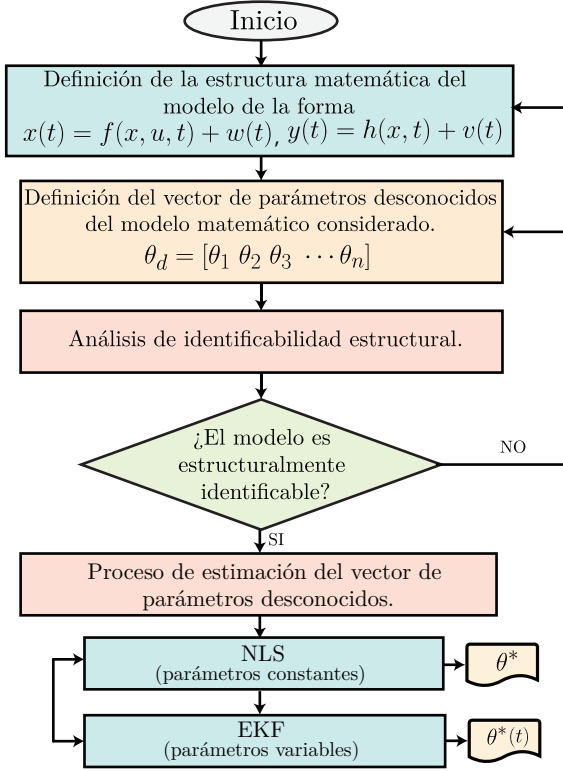


Figura 1: Metodología para la estimación de parámetros en modelos semi físicos basados en primeros principios.

2. Metodología

En la Figura 1 se presenta la metodología utilizada para el ajuste de los parámetros desconocidos del modelo de crecimiento de *Lobesia botrana*. La metodología propuesta parte de la selección de la estructura del modelo matemático a considerar (específica para el caso de aplicación), seguida de un análisis de identificabilidad estructural sobre dicho modelo. Una vez garantizada la identificabilidad estructural del modelo seleccionado, se propone realizar el proceso de estimación de parámetros en dos etapas (cuando corresponda): la estimación del conjunto de parámetros variables y la estimación del conjunto de parámetros constantes. Para el caso de estudio de proceso de crecimiento de *L. botrana* se consideraron como herramientas para la estimación un Filtro de Kalman Extendido y el algoritmo de mínimos cuadrados no lineales, respectivamente.

2.1. Paso 1: selección del modelo matemático

En este trabajo, se utilizó el modelo de crecimiento de *Lobesia botrana* desarrollado en un trabajo previo (Aguirre-Zapata

et al., 2022). Las Ecuaciones 1-4 constituyen la estructura básica del modelo en tiempo continuo para el proceso de crecimiento de *Lobesia botrana* en cada uno de sus estadios de desarrollo. En el modelo matemático seleccionado, $E_l(t)$, $L_l(t)$, $P_l(t)$ y $A_l(t)$ representan cantidades de huevos, larvas, pupas y adultos respectivamente, en un determinado instante de tiempo. Las ecuaciones constitutivas de modelo matemático se presentan en las Ecuaciones 5 - 16.

$$\frac{dE_l(t)}{dt} = \frac{\mu_{E_{max}(t)}GDA_E(t)}{K_{E_s} + GDA_E(t)}E_l(t) - K_{E_d}(t)E_l(t) \quad (1)$$

$$\frac{dL_l(t)}{dt} = \frac{\mu_{L_{max}(t)}GDA_L(t)}{K_{L_s} + GDA_L(t)}E_l(t) - K_{L_d}(t)L_l(t) \quad (2)$$

$$\frac{dP_l(t)}{dt} = \frac{\mu_{P_{max}(t)}GDA_P(t)}{K_{P_s} + GDA_P(t)}L_l(t) - K_{P_d}(t)P_l(t) \quad (3)$$

$$\frac{dA_l(t)}{dt} = \frac{\mu_{A_{max}(t)}GDA_A(t)}{K_{A_s} + GDA_A(t)}P_l(t) - K_{A_d}(t)A_l(t) \quad (4)$$

donde $K_{E_d}(t)$, $K_{L_d}(t)$, $K_{P_d}(t)$ y $K_{A_d}(t)$, son las tasas de mortalidad asociadas a cada uno de los estadios de crecimiento y K_{E_s} , K_{L_s} , K_{P_s} y K_{A_s} son las constantes de velocidad media de crecimiento. $\mu_{E_{max}}$, $\mu_{L_{max}}$, $\mu_{P_{max}}$ y $\mu_{A_{max}}$ corresponden a las tasas máximas de crecimiento que son calculadas con base en las tasas de mortalidad en cada instante de tiempo (ver Ecuaciones 5-8).

$$\mu_{E_{max}}(t) = 1 - K_{E_d}(t) \quad (5)$$

$$\mu_{L_{max}}(t) = 1 - K_{L_d}(t) \quad (6)$$

$$\mu_{P_{max}}(t) = 1 - K_{P_d}(t) \quad (7)$$

$$\mu_{A_{max}}(t) = 1 - K_{A_d}(t) \quad (8)$$

$$GDA_E(t) = \sum T_E(t) \text{ para } t = 1, 2, \dots, t_f \quad (9)$$

$$GDA_L(t) = \sum T_L(t) \text{ para } t = 1, 2, \dots, t_f \quad (10)$$

$$GDA_P(t) = \sum T_P(t) \text{ para } t = 1, 2, \dots, t_f \quad (11)$$

$$GDA_A(t) = \sum T_A(t) \text{ para } t = 1, 2, \dots, t_f \quad (12)$$

donde t_f es el tiempo final, $GDA_E(t)$, $GDA_L(t)$, $GDA_P(t)$ y $GDA_A(t)$ son los grados día acumulados, y $T_E(t)$, $T_L(t)$, $T_P(t)$ y $T_A(t)$ los grados día asociados a cada uno de los estadios de crecimiento. Estos últimos (ver Ecuaciones 13-16), se encuentran descritos mediante la ecuación de Touzeau que ha sido ampliamente utilizada en la literatura (Touzeau, 1981).

$$T_E(t) = \frac{T_{max}(t) - T_{min}(t)}{2} - T_{E_{th}} \quad (13)$$

$$T_L(t) = \frac{T_{max}(t) - T_{min}(t)}{2} - T_{L_{th}} \quad (14)$$

$$T_P(t) = \frac{T_{max}(t) - T_{min}(t)}{2} - T_{P_{th}} \quad (15)$$

$$T_A(t) = \frac{T_{max}(t) - T_{min}(t)}{2} - T_{A_{th}} \quad (16)$$

Los demás parámetros y constantes del modelo matemático se presentan en la Tabla 1.

2.2. Paso 2: análisis de identificabilidad estructural

La identificabilidad estructural es una propiedad binaria de los modelos matemáticos que se refiere a la posibilidad de calcular una solución única para los parámetros de un modelo, a

Tabla 1: Constantes y parámetros funcionales del modelo de crecimiento de *L. botrana*. Las unidades marcadas con – corresponden a aquellos parámetros que son adimensionales.

Símbolo	Unidades	Descripción	Ecuación o valor numérico	Referencia
T_{max}	°C	Temperatura máxima diaria	Valor variable	Datos experimentales
T_{min}	°C	Temperatura mínima diaria	Valor variable	
$T_{E_{th}}$	°C	Temperatura umbral huevos	11,6°C	(Rios Castro et al., 2016)
$T_{L_{th}}$	°C	Temperatura umbral larvas	8,8°C	
$T_{P_{th}}$	°C	Temperatura umbral pupas	10,35°C	
$T_{A_{th}}$	°C	Temperatura umbral adultos	10°C	
$K_{E_d}(t)$	d^{-1}	Tasa de mortalidad de huevos	Variable dependiente de la temperatura	(Herrera et al., 2016)
$K_{L_d}(t)$	d^{-1}	Tasa de mortalidad de larva		
$K_{P_d}(t)$	d^{-1}	Tasa de mortalidad de pupas		
$K_{A_d}(t)$	d^{-1}	Tasa de mortalidad de adultos		
K_{E_s}	–	Constante de velocidad media de crecimiento de huevos	Variable dependiente de la temperatura	Parámetros estimados
K_{L_s}	–	Constante de velocidad media de crecimiento de larvas		
K_{P_s}	–	Constante de velocidad media de crecimiento de pupas		
K_{A_s}	–	Constante de velocidad media de crecimiento de adultos		

partir de datos de entrada- salida libres de ruido. La propiedad de identificabilidad puede ser local o global dependiendo de si la propiedad se cumple dentro de una vecindad o para todo el espacio de parámetros como se muestra en las Ecuaciones 17 - 18 (Joubert et al., 2021).

Definición 1: *Identificabilidad estructural local* (Balsacanto et al., 2010). Un parámetro p_i es estructuralmente localmente identificable ¹ si para casi cualquier $p^* \in P$ existe un vecindario $V(p^*)$ tal que,

$$si p \in V(p^*) \rightarrow \Sigma(p) = \Sigma(p^*) \Rightarrow p_i = p_i^* \quad (17)$$

Esto es que, un parámetro es estructuralmente, localmente identificable, si es posible identificar su valor único, dentro de una vecindad en el espacio de todos los parámetros.

Definición 2: *Identificabilidad estructural global* (Balsacanto et al., 2010). Un parámetro p_i es estructuralmente globalmente identificable si para casi cualquier $p^* \in P$

$$\Sigma(p) = \Sigma(p^*) \Rightarrow p_i = p_i^* \quad (18)$$

Esto es que, un parámetro es estructuralmente, globalmente identificable, si es posible identificar de manera única el valor de dicho parámetro, en todo el espacio de los parámetros.

Definición 3: *No identificabilidad* (Balsacanto et al., 2010). Un parámetro p_i es estructuralmente no identificable si para casi cualquier $p^* \in P$ no existe un vecindario $V(p^*)$ tal que pueda encontrarse un valor único para dicho parámetro. La no identificabilidad del modelo implica que los algoritmos de estimación no podrán encontrar valores confiables para el conjunto

de parámetros, esto es, no podrán encontrar valores únicos ni local, ni globalmente.

	θ_1	θ_2	θ_3	θ_4	θ_j
C_1					x
C_2			x		
C_3	x	x			
C_4	x	x			
C_i					

Figura 2: Cuadro de identificabilidad para el análisis de identificabilidad estructural.

En la literatura, se han presentado diferentes herramientas que buscan facilitar el cálculo y análisis de la propiedad de identificabilidad estructural, como son el software SIAN en (Hong et al., 2019), el software DAISY presentado en (Bellu et al., 2007), el software STRIKE-GOLDD presentado en (Villaverde et al., 2016), o el software GenSSI presentando por Chis y colaboradores en (Chis et al., 2011a). Para este trabajo en particular, se usa el software GenSSI, el cual facilita el análisis de identificabilidad estructural mediante el método de generación de series. Para ello, se supone que los parámetros variantes en

¹El símbolo Σ hace referencia al modelo del sistema que depende del conjunto de parámetros p .

el tiempo (tasas de mortalidad) se mantienen constantes en cada tiempo de muestreo. Esta suposición es válida dado que éstos parámetros varían de forma suficientemente lenta. El resultado obtenido con el software GenSSI fue verificado mediante el uso del software STRIKE-GOLDD. El método de generación de series, consiste en generar un sistema no lineal de ecuaciones sobre los parámetros del modelo a partir del cálculo de las sucesivas derivadas de Lie sobre los parámetros del modelo matemático, y ofrece condiciones suficientes, más no necesarias para la identificabilidad estructural, por lo tanto, si la solución del sistema de ecuaciones es única, los parámetros son globalmente estructuralmente identificables. Teniendo en cuenta que la solución del conjunto de ecuaciones algebraicas resultante, es difícil de resolver analíticamente, Balsa Canto y colaboradores, introdujeron el uso de cuadros de identificabilidad para visualizar los posibles problemas de identificabilidad estructural de un modelo matemático (Balsa-Canto et al., 2010).

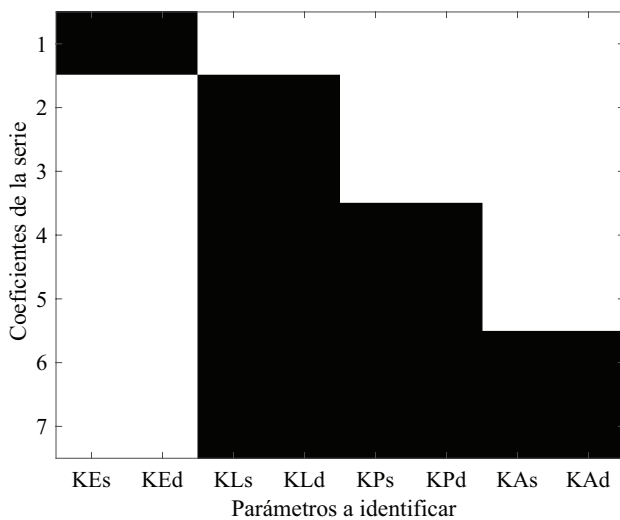


Figura 3: Cuadro de identificabilidad para el modelo de crecimiento de *Lobesia botrana* considerando todo el conjunto de parámetros desconocidos.

La Figura 2 (adaptada de (Balsa-Canto et al., 2010)) ilustra el método de cuadros de identificabilidad donde: una cruz en las coordenadas (i, j) indica que el coeficiente de la serie generadora depende del parámetro θ_j . Los círculos verdes representan aquellos escenarios en los que el parámetro correspondiente, puede ser calculado de manera única a partir de una sola ecuación del sistema (por ejemplo, el parámetro θ_3 que depende únicamente del coeficiente C_2). Los cuadros rojos representan aquellos escenarios en los que un conjunto de ecuaciones (dos o más), permiten identificar de manera única el parámetro correspondiente (por ejemplo, el parámetro θ_1 que depende de los coeficientes C_3 y C_4). Finalmente, la ausencia de una cruz en las coordenadas (i, j) indica que el coeficiente i no tiene dependencia del parámetro j . Un elemento único distinto de cero en una fila significa que el parámetro correspondiente es estructuralmente identificable (por ejemplo el parámetro θ_3). El rango del cuadro de identificabilidad (número de parámetros para los cuales se encontró una relación), en caso de ser completo, garantiza, al menos, la identificabilidad estructural local del modelo matemático (Balsa-Canto et al., 2010; Chis et al., 2011b).

Para el análisis de identificabilidad estructural se consideró un escenario típico de condiciones de laboratorio controladas, en el que es posible medir todos los estadios de la polilla ($y_1 = E_l, y_2 = L_l, y_3 = P_l, y_4 = A_l$). Además, se consideraron como variables de entrada las temperaturas mínima ($u_1 = T_{min}$) y máxima ($u_2 = T_{max}$) diarias. Para el caso particular del modelo de crecimiento de *Lobesia botrana*, los parámetros más difíciles de medir o cuantificar son las tasas de mortalidad asociadas a cada estadio ($K_{E_d}, K_{L_d}, K_{P_d}$ y K_{A_d}) (Pasquali and Sorensina, 2020). Lo anterior, dado que, si bien son parámetros con un significado físico, su valor está supeditado a las condiciones climáticas de la región en particular. Por la razón anterior, dichos parámetros fueron incluidos en el análisis de identificabilidad estructural. Además, se incluyeron en el análisis, los parámetros $K_{E_s}, K_{L_s}, K_{P_s}$ y K_{A_s} (constantes de velocidad media de crecimiento para cada estadio), que aparecen en las ecuaciones constitutivas de las tasas de crecimiento, y a los que, al ser parámetros empíricos, resulta difícil asignarles un valor a priori y deben ser estimados mediante algún algoritmo de estimación.

$$\theta_{ac} = [K_{E_s} \quad K_{E_d} \quad K_{L_s} \quad K_{L_d} \quad K_{P_s} \quad K_{P_d} \quad K_{A_s} \quad K_{A_d}] \quad (19)$$

$$x = [E_l \quad L_l \quad P_l \quad A_l] \quad (20)$$

$$x_0 = [30 \quad 0 \quad 0 \quad 0] \quad (21)$$

El análisis de identificabilidad estructural, se realizó considerando el vector de parámetros θ_{ac} (ver Ecuación 19) que incluye las tasas de mortalidad para cada estadio y las constantes de velocidad media de crecimiento, y el vector de condiciones iniciales presentado en la Ecuación 21.

Se realizó un primer análisis de identificabilidad estructural, considerando el conjunto de parámetros desconocidos que se muestra en la Ecuación 19. Dicho análisis, una vez explorado el resultado de los cuadros de identificabilidad hasta con 25 derivadas de Lie, evidenció que al considerar estos parámetros como desconocidos dentro de la estructura del modelo, el número de parámetros resultaba mayor que el número de relaciones, por lo que no era posible dar una solución explícita para el conjunto de parámetros desconocidos. En la Figura 3 se observa que existe un solo coeficiente (o relación) que depende de los parámetros K_{E_s} y K_{E_d} , por lo que se tienen más incógnitas que ecuaciones, es decir, el cuadro de identificabilidad presentado en la Figura 3 resultó ser de rango incompleto ($R \neq 7$) y por ende los parámetros son no identificables.

Considerando el resultado anterior, fue necesario volver al paso uno de la metodología presentada en la Figura 1, evaluar nuevamente la estructura del modelo matemático y elegir fijar a priori uno de los parámetros asociados al estadio de crecimiento huevo. Se fijó la tasa de mortalidad de huevos K_{E_d} como un parámetro conocido, considerando que por la interpretabilidad contextualizada de dicho parámetro, era posible reconstruirlo a partir de datos experimentales o a partir de datos disponibles en la literatura. En este sentido, se realizó un nuevo análisis de identificabilidad estructural considerando el nuevo vector de parámetros desconocidos presentado en la Ecuación 22. Fueron necesarias 7 derivadas de Lie para completar el análisis de identificabilidad estructural del modelo presentado en las Ecuaciones 19 y 21.

ciones 1-4. El cuadro de identificabilidad que se muestra en la Figura 4 resultó ser de rango completo ($R = 7$) por lo que la identificabilidad local está garantizada.

$$\theta_{an} = [K_{E_s} \quad K_{L_s} \quad K_{L_d} \quad K_{P_s} \quad K_{P_d} \quad K_{A_s} \quad K_{A_d}] \quad (22)$$

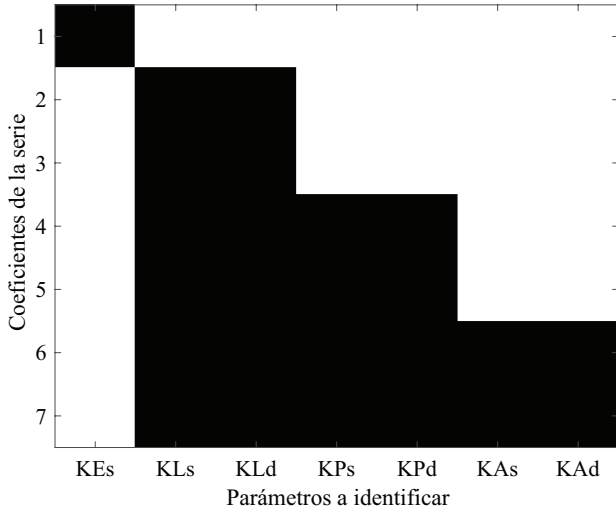


Figura 4: Cuadro de identificabilidad para el modelo de crecimiento de *Lobesia botrana* considerando el parámetro K_{E_d} conocido.

En el cuadro de identificabilidad presentado en la Figura 4 se observa como el parámetro K_{E_s} depende únicamente del coeficiente 1, lo que habla de la identificabilidad global del mismo. Por otra parte, los parámetros restantes, si bien requieren de más de una ecuación para su identificación, son también globalmente estructuralmente identificables. Del análisis anterior, es posible concluir que el modelo de crecimiento de *Lobesia botrana* es estructuralmente, globalmente identificable, una vez fijado el parámetro de mortalidad de huevos (K_{E_d}).

El análisis de identificabilidad estructural fue realizado también con la herramienta STRIKE-GOLDD que permite analizar la identificabilidad local de los modelos matemáticos. El resultado obtenido fue igual en cada caso. Al considerar el vector de parámetros completo, presentado en la Ecuación 19, el análisis concluyó que el vector de parámetros $[K_{L_s} \quad K_{L_d} \quad K_{P_s} \quad K_{P_d} \quad K_{A_s} \quad K_{A_d}]$ era identificable, y que los parámetros K_{E_s} y K_{E_d} eran no identificables. De manera similar, al fijar la tasa de mortalidad de huevos y considerar el vector presentado en la Ecuación 22, el resultado del análisis concluyó que el nuevo vector de parámetros era localmente, estructuralmente identificable.

2.3. Paso 3: estimación de parámetros

Para abordar el problema de estimación de parámetros, se consideraron dos tipos de parámetros dentro del modelo matemático. Esto es, aquellos parámetros cuyos valores son constantes y aquellos parámetros que por su significado dentro del contexto de aplicación del modelo, se conoce que son variables en el tiempo. En ambos casos, se usaron datos experimentales a escala de laboratorio para el ajuste de los mismos. Herrera y colaboradores (Herrera et al., 2016), reportaron un

método práctico para el desarrollo de la cría de *L. botrana* en condiciones de laboratorio. En dicho experimento, los autores hicieron seguimiento de proceso de crecimiento de un pie de cría obtenido de racimos infestados con *L. botrana* en la región vitícola de la localidad Cruz de Piedra en Maipú, Mendoza. A partir de los experimentos realizados a diferentes temperaturas ($15^\circ C$, $20^\circ C$ y $25^\circ C$), se obtuvieron datos relevantes sobre el proceso de crecimiento de *L. botrana*, como la cantidad de días por estadio, temperaturas umbrales para el crecimiento por estadio, entre otros. Para cada temperatura se realizaron tres experimentos (E_1 - E_3). En el ajuste de los parámetros se consideró el experimento E_3 . Los demás datos experimentales fueron utilizados para validar la respuesta del modelo matemático. En la Figura 5 se presentan las curvas de crecimiento de *L. botrana* en condiciones de laboratorio a $20^\circ C$ usados para el proceso de estimación de parámetros.

2.3.1. Estimación de parámetros constantes

Las constantes de velocidad media de crecimiento de la polilla en sus diferentes estadios de crecimiento (K_{E_s} , K_{L_s} , K_{P_s} , K_{A_s}), son parámetros constantes del modelo presentado en las Ecuaciones 1- 4 que aparecen en el denominador de las cinéticas de crecimiento, y que, al no tener interpretabilidad son los parámetros con un rango de operación más amplio, y por ende los más difíciles de ajustar en el modelo de crecimiento de *L. botrana*. La estimación de este conjunto de parámetros (ver Ecuación 27), se realizó mediante el uso de un algoritmo de optimización por mínimos cuadrados no lineales usando el método de búsqueda de Levenberg-Marquardt.

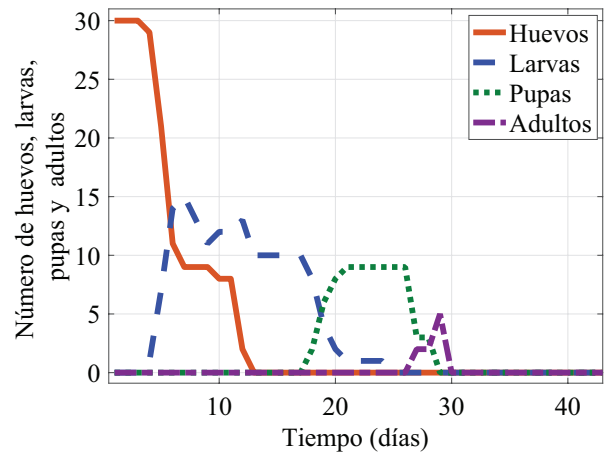


Figura 5: Curvas de crecimiento de *Lobesia botrana* en condiciones de laboratorio a $20^\circ C$.

Durante la etapa de estimación de parámetros constantes, los parámetros variables en el tiempo fueron fijados en los perfiles utilizados en un trabajo previo como se muestran en la Figura 6. En (Aguirre-Zapata et al., 2022) se construyeron perfiles de mortalidad experimentales, tomando como base los datos de crecimiento de *Lobesia botrana* en condiciones de laboratorio obtenidos por Herrera y colaboradores en (Herrera et al., 2016). Esto se realizó tabulando los datos de crecimiento para calcular la cantidad de huevos, larvas pupas y adultos muertos por día (E_{dd} , L_{dd} , P_{dd} y A_{dd}). Lo anterior, sin considerar aquellos huevos, larvas o pupas que trascendían a la siguiente etapa de cre-

cimiento. Las tasas de mortalidad diarias se calcularon como se muestra en las Ecuaciones 23-24.

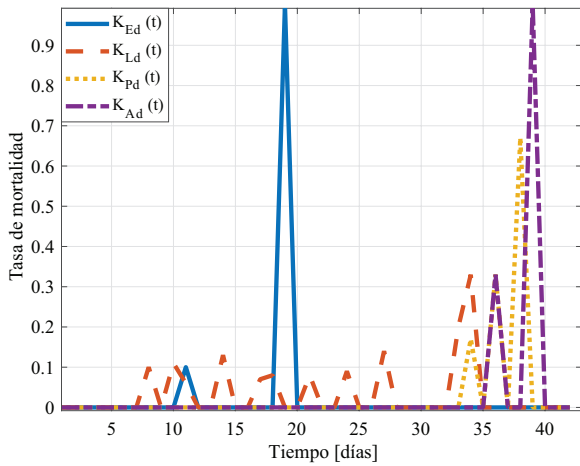


Figura 6: Perfiles de mortalidad diseñados a partir de los datos experimentales.

$$K_{E_d}(1) = \frac{E_{dd}(1)}{E_{l_{max}}} \tag{23}$$

Para $t > 1$

$$K_{E_d} = \frac{E_{dd}(t)}{E_l(t-1)} \tag{24}$$

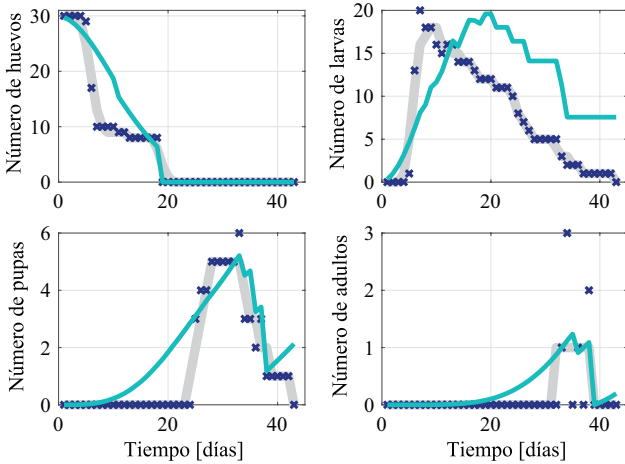


Figura 7: Respuesta del modelo con perfiles de mortalidad experimentales y constantes de velocidad media de crecimiento ajustadas de manera heurística, en contraste con los datos experimentales a 20°C. La línea gris (línea gruesa) corresponde a los datos experimentales suavizados, las marcas (x) azules corresponden a los datos experimentales originales, y la curva celeste (-) corresponde a la respuesta de modelo de crecimiento previa al uso de la metodología propuesta.

En la Figura 7, se presentan las curvas de crecimiento de huevos, larvas, pupas y adultos obtenidas a partir de la simulación del modelo matemático, considerando los perfiles de mortalidad experimentales presentados en la Figura 6 y considerando el vector de constantes de velocidad media de crecimiento (θ_h) ajustado mediante ensayo y error como se muestra en la Ecuación 26.

$$\theta_h = [K_{E_s} \quad K_{L_s} \quad K_{P_s} \quad K_{A_s}] \tag{25}$$

$$\theta_h = [0,95e3 \quad 1e3 \quad 1,55e4 \quad 1,35e4] \tag{26}$$

En la Tabla 2 se presentan los resultados cuantitativos correspondientes al error cuadrático medio (RMSE) resultante de comparar la respuesta del modelo matemático sin ajustar (mediante la metodología propuesta), con los datos experimentales pre procesados y los datos experimentales originales. El problema de estimación de parámetros constantes se planteó como sigue:

- Considerando el vector de parámetros desconocidos

$$\theta = [K_{E_s} \quad K_{L_s} \quad K_{P_s} \quad K_{A_s}] \tag{27}$$

- Se planteó el siguiente problema de optimización

$$\min_{\theta} f(x, \theta) = \sum_{i=1}^m |f_i(x, \theta) - y_i| \tag{28}$$

suje to a

$$lb_{\theta} = [0,001 \quad 0,001 \quad 0,001 \quad 0,001] \tag{29}$$

$$ub_{\theta} = [1e5 \quad 1e5 \quad 1e5 \quad 1e5] \tag{30}$$

donde lb_{θ} es el vector de límites inferiores sobre el vector θ y ub_{θ} es el vector de límites superiores sobre el vector θ . Las restricciones, tanto superiores como inferiores para el vector de parámetros, se definieron con base en el conocimiento del rango de variación de cada estado. Los parámetros $K_{E_s}, K_{L_s}, K_{P_s}, K_{A_s} > 0$ dado su significado físico y por condicionamiento matemático, debido a que se encuentran en el denominador de las ecuaciones que describen las tasas de crecimiento de la polilla en cada uno de sus estadios. Los límites superiores de los mismos parámetros fueron determinados partiendo de valores aleatorios que fueron modificados de manera iterativa hasta reducir el error en el ajuste.

El proceso de estimación de parámetros se realizó para el rango de temperatura de $20^{\circ}C \pm 2^{\circ}C$. Se eligieron los datos experimentales a dicha temperatura, dado que por ser una de las temperaturas ideales para el crecimiento de la polilla, permiten tener comportamientos dinámicos en todos los estadios de crecimiento, esto es, que, para temperaturas muy altas o muy bajas, se inhibe el crecimiento de la polilla en el estadio adulto. En la Ecuación 31 se presentan los valores óptimos encontrados mediante el método de búsqueda por mínimos cuadrados de Levenberg-Marquardt, para el vector de parámetros que se muestra en la Ecuación 27.

$$\theta^* = [0,98e3 \quad 0,28e3 \quad 0,55e3 \quad 0,85e3] \tag{31}$$

Los parámetros $K_{E_s}, K_{L_s}, K_{P_s}$ y K_{A_s} (constantes de velocidad media de crecimiento), representan la afinidad de la polilla (en cada estadio) por el factor limitante de crecimiento (la temperatura acumulada). En este sentido, los valores encontrados mediante el algoritmo de mínimos cuadrados presentados en la Ecuación 31, muestran como los estadios larva y pupa (con valores K_{L_s} y K_{P_s} más pequeños), son más afines a la temperatura acumulada y por ende se ven más influenciados por ella. Por otra parte, los estadios huevo y adulto (con valores K_{E_s} y K_{A_s} más grandes), tienen menor afinidad por la temperatura acumulada y por ende se ven influenciados en menor medida por ella.

Tabla 2: Raíz del error cuadrático medio entre el modelo matemático sin ajuste, los datos experimentales originales y los datos experimentales pre procesados para el rango de temperatura de $20 \pm 2^\circ C$.

Estadio	RMSE datos pre-procesados	RMSE datos originales	Error porcentual (%)	Unidades
E_l	4,01	3,84	12,73	Número de huevos
L_l	6,65	6,47	32,5	Número de larvas
P_l	1,11	1,056	17,67	Número de pupas
A_l	0,48	0,234	16	Número de adultos

2.3.2. Estimación de parámetros variables

Si bien, en la etapa de estimación de parámetros constantes, las tasas de mortalidad fueron fijadas en los perfiles reconstruidos a partir de los datos experimentales, se evidenció la necesidad de estimar perfiles de mortalidad variables en el tiempo, pero con transiciones suaves (sin ser constantes), con el fin de mejorar el ajuste del modelo en los estados transitorios. Para el ajuste de las tasas de mortalidad, se consideró el uso del Filtro de Kalman Extendido (EKF), que es una variación del filtro de Kalman, que permite abordar el problema de la estimación de estados y parámetros variantes en el tiempo en sistemas no lineales (Simon, 2006). Considere un sistema no lineal de tiempo discreto que depende de forma no lineal de un vector de parámetros desconocidos θ_k , como se muestra en la Ecuación 32.

$$\begin{aligned} x_k &= f(x_{k-1}, \theta_k, u_{k-1}, w_{k-1}) \\ y_k &= g_k(x_k, \theta_k, u_k, v_k) \end{aligned} \quad (32)$$

donde $f(x_{k-1}, \theta_k, u_{k-1}, w_{k-1})$ es la función de transición de estado del sistema dinámico; $g(x_k)$ es la función de medición; $x_k \in R^n$ es el vector de estados del sistema con condición inicial $x(0)$; u_k es el vector de entradas del sistema dinámico. Considerando el sistema anterior, se encuentran en la literatura diferentes aproximaciones para la estimación de estados y parámetros usando un EKF, tales como los algoritmos de estimación conjunta, estimación dual o los escenarios de estimación de parámetros cuando el estado del sistema es conocido (Simon, 2006). En este sentido, el Filtro de Kalman Extendido se convierte en un estimador de parámetros al combinar los elementos del vector de estado del sistema y el vector de parámetros θ_k , lo que se conoce como estimación conjunta o de estado aumentado. Para ello, se considera la dinámica de estado aumentado como se muestra en la Ecuación 33.

$$\begin{aligned} \begin{bmatrix} x_k \\ \theta_k \end{bmatrix} &= \begin{bmatrix} f_{k-1}(x_{k-1}, u_{k-1}, \theta_{k-1}, w_{k-1}) \\ \theta_{k-1} + r_{k-1} \end{bmatrix} \\ y_k &= g_k(x_k, u_k, \theta_k, v_k) \end{aligned} \quad (33)$$

Para simplificar la notación puede escribirse el sistema como se muestra en la Ecuación 34.

$$\begin{aligned} X_k &= F_{k-1}(X_{k-1}, u_{k-1}, W_{k-1}) \\ y_k &= g_k(X_k, u_k, v_k) \end{aligned} \quad (34)$$

donde X_k es el nuevo vector de estado aumentado, que comprende tanto el estado como los parámetros, W_k es el vector que comprende el ruido del proceso y el ruido de los parámetros, F es la función que combina la dinámica del estado y la dinámica de los parámetros, y r_{k-1} es un pequeño término de ruido artificial que permite al EKF cambiar su estimación de θ_k . Con

el modelo aumentado, se aplican las ecuaciones del EKF como se muestra en la Tabla 3, cuya implementación se da en cuatro momentos: la definición del sistema dinámico no lineal, la inicialización del filtro, la linealización punto a punto del sistema dinámico no lineal definido inicialmente, y por último la actualización de la estimación como se detalla en (Simon, 2006).

Tabla 3: Pseudocódigo de implementación del Filtro de Kalman Extendido

Calcular para $k = 1, 2, \dots, n$

Actualización de la estimación a priori

$$\hat{X}_k^- = F_{k-1}(\hat{X}_{k-1}^+, u_{k-1})$$

Actualización de la covarianza del error a priori

$$P_k^- = A_{k-1} P_{k-1}^+ A_{k-1}^T + Q_{k-1}$$

Cálculo de la ganancia de Kalman

$$K_k = P_k^- C_k^T (C_k P_k^- C_k^T + R_k)^{-1}$$

Actualización de la estimación a posteriori

$$\hat{X}_k^+ = \hat{X}_k^- + K_k [y_k - g_k(\hat{X}_k^-)]$$

Actualización de la covarianza del error a posteriori

$$P_k^+ = (I - K_k C_k) P_k^-$$

donde Q y R son matrices de covarianza definidas positivas, A corresponden a la matriz de derivadas parciales del sistema F respecto al estado X y C corresponde a la matriz de derivadas parciales de la salida del sistema g respecto al estado X . El modelo de espacio de estados utilizado en este trabajo (ver la Ecuación 35), es una versión en tiempo discreto del modelo de tiempo continuo (sin ruido) desarrollado en Aguirre-Zapata et al. (2022) mediante la aproximación de Euler hacia adelante $\dot{x}[x_{k+1} - x_k]/T_s$, considerando $T_s = 1$.

$$F(X_k, u_k) = \begin{bmatrix} X_k^1 + T_s \left[\frac{\mu_{Emaxk} \frac{u_k^1 - u_k^2}{2} - T_{Eth}}{K_{Es} + \frac{u_k^1 - u_k^2}{2} - T_{Eth}} - K_{Ed} \right] X_k^1 \\ X_k^2 + T_s \left[\frac{\mu_{Lmaxk} \frac{u_k^1 - u_k^2}{2} - T_{Lth}}{K_{Ls} + \frac{u_k^1 - u_k^2}{2} - T_{Lth}} X_k^1 - X_k^5 X_k^2 \right] \\ X_k^3 + T_s \left[\frac{\mu_{Pmaxk} \frac{u_k^1 - u_k^2}{2} - T_{Pth}}{K_{Ps} + \frac{u_k^1 - u_k^2}{2} - T_{Pth}} X_k^2 - X_k^6 X_k^3 \right] \\ X_k^4 + T_s \left[\frac{\mu_{Amaxk} \frac{u_k^1 - u_k^2}{2} - T_{Ath}}{K_{As} + \frac{u_k^1 - u_k^2}{2} - T_{Ath}} X_k^3 - X_k^7 X_k^4 \right] \\ 0_{3 \times 1} \end{bmatrix} \quad (35)$$

Se consideró un sistema no lineal de la forma que se presentó en 34, donde se combinaron los estados del sistema dinámico y las tasas de mortalidad asociadas a cada estadio de crecimiento de la polilla (parámetros desconocidos), para la estimación

de estos últimos. El vector de transición de estados utilizado en este trabajo está dado por la Ecuación 35.

El vector de estados $X \in R^{+(7 \times 1)}$ y el vector de entrada $u \in R^{+(2 \times 1)}$ son

$$X_k = [X_k^1 \ X_k^2 \ X_k^3 \ X_k^4 \ X_k^5 \ X_k^6 \ X_k^7]^T \quad (36)$$

$$X_k = [E_{lk} \ L_{lk} \ P_{lk} \ A_{lk} \ K_{L_{dk}} \ K_{P_{dk}} \ K_{A_{dk}}]^T$$

Considerando el sistema anterior, el EKF fue programado y simulado en MATLAB® & Simulink®, considerando el esquema presentado en la Figura 9 y las condiciones iniciales que se presentan en las Ecuaciones 37 y 38.

$$x_0 = [30 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0] \quad (37)$$

$$P_0^+ = 150I_7 \quad (38)$$

donde I_7 corresponde a la matriz identidad. Las matrices de covarianza Q y R fueron ajustadas empíricamente mediante simulaciones, obteniendo como resultado las matrices que se muestran en las Ecuaciones 39 Y 40.

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{bmatrix} \quad (39)$$

$$R = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 20 & 0 & 0 \\ 0 & 0 & 20 & 0 \\ 0 & 0 & 0 & 20 \end{bmatrix} \quad (40)$$

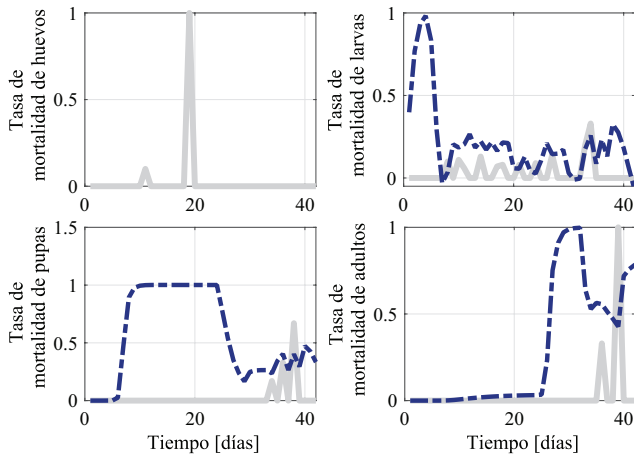


Figura 8: Tasas de mortalidad de *Lobesia botrana* estimadas mediante el EKF comparadas con las tasas de mortalidad construidas con las Ecuaciones 23-24. La línea gris (-) corresponde a las tasas de mortalidad experimentales (Ecuaciones 23-24) a 20°C. La línea azul (-.-) corresponde a las tasas de mortalidad estimadas mediante el EKF

En la Figura 8 se presentan los perfiles de mortalidad estimados mediante el EKF, en contraste con los perfiles de mortalidad experimentales. Se observan algunas similitudes entre

los perfiles experimentales y los perfiles estimados. Sin embargo, los perfiles estimados permiten obtener una respuesta más suave del modelo como se muestra en la Sección 3.

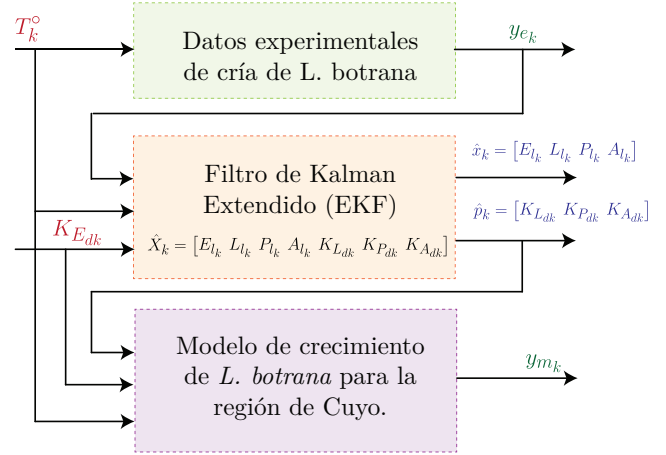
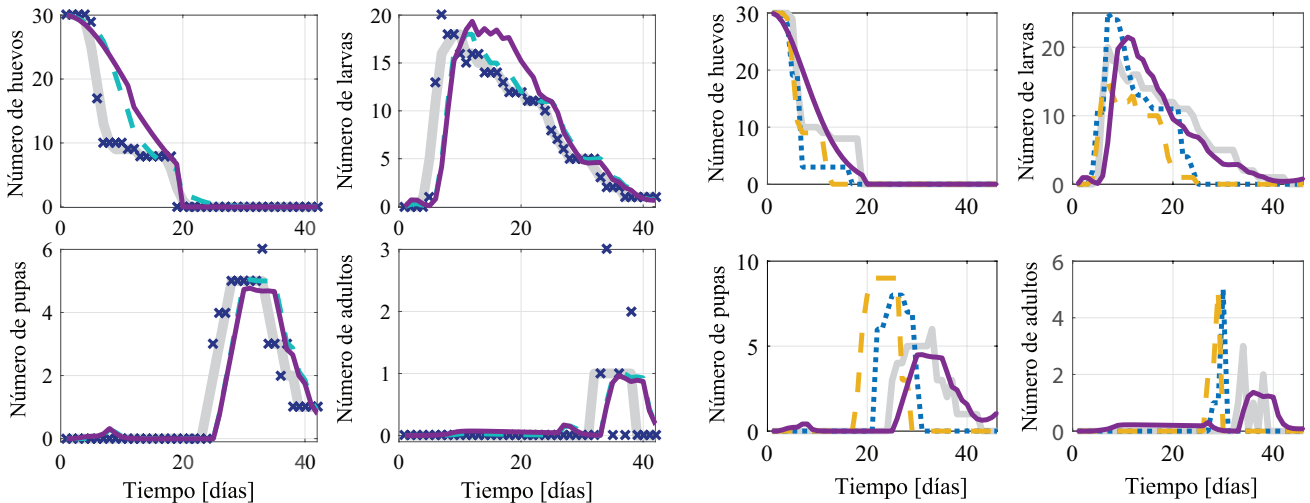


Figura 9: Esquema de estimación de parámetros variables para el modelo de crecimiento de *L. botrana*, donde y_e representa los datos experimentales medidos e y_m representa la salida de modelo matemático ajustado.

3. Resultados

En la Figura 10 se presentan las curvas de evolución del modelo luego del ajuste de los parámetros, y se validan contrastando con los datos experimentales obtenidos en laboratorio. Para ello, se compara la respuesta del modelo matemático con el conjunto de datos utilizados para validación (R_1) (ver Figura 10a), y con dos conjuntos de datos adicionales, obtenidos de manera experimental bajo las mismas condiciones de laboratorio (ver Figura 10b). En Figura 10a, se observan los datos experimentales originales (marcados con x) y un preprocesamiento de los mismos (línea gris gruesa). El preprocesamiento de los datos fue necesario, dado que algunas observaciones experimentales se encontraban por fuera de la tendencia de crecimiento y dificultaban el proceso de estimación de los parámetros. El hallazgo de datos por fuera de la tendencia de crecimiento, se explica por el hecho de que el registro durante el proceso experimental presentado en (Herrera et al., 2016), se realizaba día a día mediante la observación por parte de profesionales calificados para dicha tarea, sin embargo, los datos experimentales tienen asociado el error humano propio de este tipo de experimentos. Con relación al estadio de crecimiento E_l (huevo), se observa que el modelo matemático sigue de forma adecuada la tendencia de crecimiento en el estadio en cuestión y se reporta un porcentaje de ajuste entre el modelo ajustado y los datos experimentales de 84,4 %. En el estado larval L_l se aprecia un máximo un poco más elevado al presentado por los datos experimentales pero con la tendencia adecuada y un porcentaje de ajuste aceptable del 82,4 %, siendo el estadio con menor porcentaje de ajuste. El estadio pupa (P_l), presenta un buen ajuste a la tendencia de los datos. Sin embargo, queda en evidencia la necesidad de mejorar el ajuste del perfil de mortalidad asociado a dicho estadio ($K_{L_{dk}}$) con el fin de evitar los cambios abruptos de comportamiento como el observado en el día ~ 38 . El porcentaje de ajuste para el estadio de pupa fue del 86,3 %. Finalmente, el estadio adulto también evidenció un correcto ajuste a



(a) Curva de respuesta de modelo en contraste con los datos usados para el ajuste. (b) Curva de respuesta de modelo en contraste con datos experimentales no usados para el ajuste.

Figura 10: Respuesta del modelo ajustado en contraste con los datos experimentales a 20°C para tres corridas experimentales diferentes. En la Figura (a), la línea gris (línea gruesa) corresponde a los datos experimentales suavizados utilizados para realizar el ajuste (R_1), las marcas (x) corresponden a los datos experimentales originales, la línea azul (- -) corresponde al estado estimado mediante el EKF, y la curva violeta (-) corresponde a la respuesta de modelo de crecimiento cuyos parámetros fueron estimados. En la Figura (b), la línea gris (-) corresponde a los datos utilizados para el ajuste, las líneas (:) y (-), corresponden a los datos de dos corridas experimentales a la misma temperatura (R_2 y R_3), y la curva violeta (-) corresponde a la respuesta de modelo de crecimiento cuyos parámetros fueron estimados.

los datos experimentales con un ajuste del 89%, siendo este el más alto de todos.

se presentan los resultados obtenidos al evaluar la respuesta del modelo matemático ajustado en contraste con datos experimentales a una temperatura de $15 \pm 2^\circ\text{C}$.

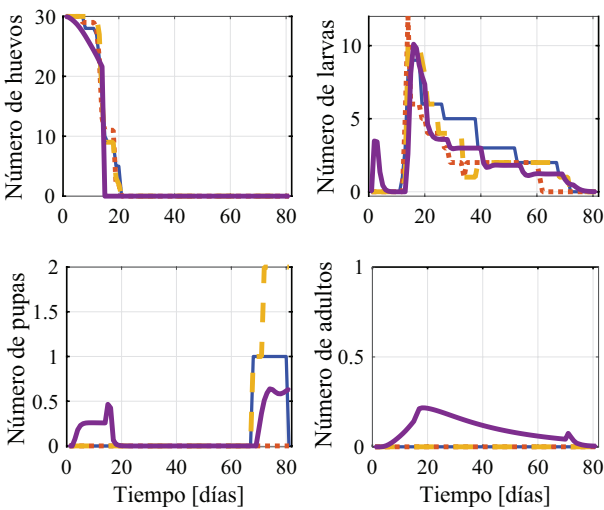


Figura 11: Respuesta del modelo ajustado con la metodología propuesta en contraste con tres corridas experimentales a una temperatura de $15 \pm 2^\circ\text{C}$. La línea azul (-) corresponde a los datos experimentales R_1 , la línea naranja (-) corresponde a los datos experimentales R_2 , la curva amarilla (- -) corresponde a los datos experimentales R_3 , y la curva violeta (- -) corresponde a la respuesta del modelo matemático ajustado para estas condiciones de temperatura.

Con relación a la Figura 10b, con el fin de validar la metodología propuesta, se contrastó la respuesta del modelo matemático ajustado, con dos conjuntos de datos experimentales a la misma temperatura. En este caso, en los estadios huevo y larva, se observan porcentajes de ajuste similares a los obtenidos con el conjunto de datos R_1 . Para el estadio pupa, se observa que existe cierta disparidad entre los experimentos R_1 , R_2 y R_3 , por lo que el ajuste en este caso es del 60%. En la Figura 11

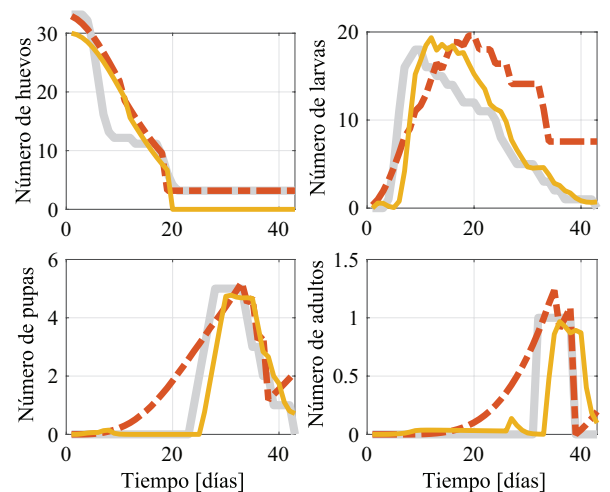


Figura 12: Respuesta del modelo ajustado con la metodología propuesta en contraste con la respuesta del modelo no ajustado, a $20 \pm 2^\circ\text{C}$. La línea gris (línea gruesa) corresponde a los datos experimentales suavizados, la línea naranja (- -) corresponde al estado estimado mediante el EKF, y la curva amarilla (-) corresponde a la respuesta de modelo de crecimiento cuyos parámetros fueron estimados.

Los resultados obtenidos evidencian un ajuste superior al 90% en los estadios de huevo y larva, y un ajuste superior al 70% en los estadios pupa y adulto. Es importante destacar, que el rango de temperatura $15 \pm 2^\circ\text{C}$, no es óptimo para el crecimiento de la polilla, lo que explica la ausencia de dinámica en el estadio adulto.

Tabla 4: Raíz del error cuadrático medio entre el modelo matemático ajustado, los datos experimentales utilizados para el ajuste de parámetros (R_1), y dos corridas experimentales adicionales, para el rango de temperatura de $20 \pm 2^\circ\text{C}$.

Estadio	RMSE	% de	RMSE	% de	RMSE	% de	Unidades
	R_1	ajuste R_1	R_2	ajuste R_2	R_3	ajuste R_3	
E_l	4,69	84,4 %	5,12	83,9 %	3,64	87,8 %	Número de huevos
L_l	3,53	82,4 %	5,28	78 %	5,41	63,9 %	Número de larvas
P_l	0,82	86,3 %	3,17	60,4 %	4,01	55 %	Número de pupas
A_l	0,31	89,7 %	0,93	81,4 %	0,96	80,8 %	Número de adultos

En la Tabla 4, se presentan los resultados cuantitativos para la evaluación del ajuste del modelo matemático a los datos experimentales. Se presenta el error cuadrático medio (RMSE) tanto para los datos experimentales pre procesados como para los datos experimentales originales, en contraste con la respuesta de modelo matemático. En la Figura 12, se presentan los resultados cualitativos obtenidos al comparar el modelo matemático ajustado de manera heurística en contraste con el modelo matemático ajustado mediante la metodología propuesta en este trabajo. En la figura se observa la notable mejoría presentada al ajustar el modelo por la metodología presentada en este trabajo. La dinámica del proceso es modelada con más precisión y se captura con mayor exactitud la tendencia de los datos experimentales. En la Tabla 5 se presentan los resultados cuantitativos, obtenidos al comparar el modelo matemático ajustado de manera heurística en contraste con el modelo matemático ajustado mediante la metodología propuesta en este trabajo. Como se observa en la Tabla citada, tres de cuatro estados presentaron una mejoría en el ajuste igual o superior al 25 % y solo uno de ellos un error más alto.

4. Conclusiones y discusión

En este trabajo se presentó una metodología para la estimación de parámetros desconocidos en modelos semi físicos basados en primeros principios, que poseen parámetros constantes y parámetros variables en el tiempo. La metodología, además de considerar métodos de estimación de parámetros conocidos, propone realizar como paso previo a la implementación de dichos métodos, un análisis de identificabilidad estructural que permita garantizar la unicidad en la solución de los mismos. La metodología fue aplicada para el caso de estudio del proceso de crecimiento de *L. botrana* abordado desde el punto de vista del modelado en un trabajo previo.

A partir de los resultados obtenidos, fue posible determinar la necesidad de fijar la tasa de mortalidad de huevos en el perfil experimental, con el fin de garantizar la identificabilidad estructural del modelo matemático utilizado. Lo anterior tiene que ver con el hecho, de que el modelo matemático no es anticipativo y, por lo tanto, requiere una condición inicial en el parámetro de mortalidad para garantizar que puedan reconstruirse los demás perfiles (mortalidad de larvas, pupas y adultos). En este sentido, a partir del análisis de identificabilidad realizado, se constató que la estructura matemática del modelo desarrollado en un trabajo previo permite determinar de manera única el conjunto de parámetros desconocidos de la misma. Lo anterior, indica que el modelo utilizado no se encuentra sobre parametrizado y que, por lo tanto, su complejidad es aceptable. Esto último, da vía libre a la formalización de un modelo de crecimiento y voltinismo de *L. botrana* de mayor complejidad (como los

propuestos en (Ortega-Lopez et al., 2014; Castex et al., 2020)), basado en las ecuaciones presentadas y ajustadas en este trabajo, que pueda ser extendido para una temporada completa (365 días), y validado con datos de captura de la polilla en campo. La metodología propuesta será útil en un trabajo futuro, para el ajuste y validación con datos de captura en campo, de un modelo de voltinismo de *L. botrana* que facilite la toma de decisiones de control por parte de personal experto en la región de Cuyo en Argentina. Es importante destacar que, la identificabilidad estructural global analizada en este trabajo para el modelo de crecimiento de *L. botrana* está supeditada a la posibilidad de medir todos los estados del modelo, dado que para este caso en particular, se disponía de datos experimentales para cada uno de los estadios de crecimiento. Se propone como trabajo futuro, analizar la identificabilidad estructural del modelo, considerando un escenario a escala piloto en el que algunos de los estados no sean medibles.

Con relación al proceso de estimación de parámetros constantes, a partir de la implementación de algoritmo de mínimos cuadrados no lineales, fue posible obtener valores dimensionalmente consistentes para todas las constantes de velocidad media de crecimiento. Se observó que al ajustar dicho conjunto de parámetros de manera heurística, y considerando los perfiles de mortalidad experimentales, se obtenían valores con dimensiones considerablemente diferentes (ver Ecuación 26), que fueron mejorados de manera considerable una vez implementada la metodología propuesta, lo que se evidenció de manera cualitativa en una respuesta más suave del modelo matemático, y cuantitativamente mediante una reducción en el porcentaje de error. Con base en los valores óptimos encontrados (ver Ecuación 31), fue posible analizar el grado de influencia de la temperatura sobre cada estadio de crecimiento de la polilla. En este sentido, al realizar tareas de estimación de parámetros surge un cuestionamiento relevante relacionado con el significado de los parámetros ajustados. Esto es, ¿los valores encontrados tienen coherencia en el contexto de aplicación del modelo matemático? Para el caso puntual de este trabajo, los parámetros K_{E_s} , K_{L_s} , K_{P_s} y K_{A_s} , los cuales fueron estimados mediante un algoritmo de búsqueda por mínimos cuadrados, aparecen en una ecuación constitutiva similar estructuralmente a la ecuación de Monod, que fue utilizada por analogía para el modelado del proceso de crecimiento de *L. botrana* partiendo de la hipótesis de que el proceso de crecimiento de la polilla podría asemejarse al proceso de crecimiento de algunos microorganismos. El parámetro K_s de la ecuación de Monod, si bien es un parámetro empírico, posee interpretabilidad local o contextualizada, dado que representa la afinidad del microorganismo cuyo crecimiento se está modelando, por el factor limitante de crecimiento (o

Tabla 5: Raíz del error cuadrático medio entre el modelo matemático ajustado y sin ajuste, y los datos experimentales preprocesados para el rango de temperatura de $20 \pm 2^\circ\text{C}$.

Estadio	RMSE sin ajuste datos preprocesados	RMSE ajustado datos preprocesados	Porcentaje de mejoría	Unidades
E_l	3,82	4,69	-18 %	Número de huevos
L_l	6,47	3,53	83 %	Número de larvas
P_l	1,06	0,82	29,4 %	Número de pupas
A_l	0,223	0,31	25 %	Número de adultos

sustrato). Dicha afinidad se encuentra bien determinada en la literatura para diversos microorganismos y sustratos. Sin embargo, al provenir de una hipótesis de modelado, estos parámetros resultan novedosos en el contexto de aplicación de proceso de crecimiento de *Lobesia botrana*. Finalmente, con relación al proceso de estimación de parámetros variables en el tiempo (tasas de mortalidad), a partir de la implementación de un EKF con estado extendido, fue posible reconstruir los perfiles de mortalidad asociados a los estadios de crecimiento larva, pupa y adulto. Algunos de los perfiles obtenidos (mortalidad de larvas y mortalidad de adultos) resultaron ser similares a los perfiles experimentales, con sutiles variaciones que permitieron obtener una respuesta más suave y con menor porcentaje de error de ajuste.

Agradecimientos

Estefanía Aguirre-Zapata está financiada por una beca doctoral latinoamericana del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) de Argentina, y cofinanciada por el programa ENLAZAMUNDOS de la Agencia de Educación Postsecundaria (SAPIENCIA) de Medellín, Colombia. Humberto Morales tiene una beca doctoral del Servicio de Intercambio Académico Alemán (DAAD). Los datos experimentales utilizados para el proceso de ajuste y validación del modelo fueron proveídos por el Instituto Nacional de Tecnología Agropecuaria (INTA) - Mendoza, Argentina.

Referencias

- Aguirre-Zapata, E., Morales, H., Dagatti, C. V., Di Sciascio, F., Amicarelli, A. N., 2022. Semi physical growth model of lobesia botrana under laboratory conditions for argentina's cuyo region. *Ecological Modelling* 464, 109803.
- Asprion, N., Böttcher, R., Pack, R., Stavrou, M.-E., Höller, J., Schwientek, J., Bortz, M., 2019. Gray-box modeling for the optimization of chemical processes. *Chemie Ingenieur Technik* 91 (3), 305–313.
- Balsa-Canto, E., Alonso, A. A., Banga, J. R., 2010. An iterative identification procedure for dynamic modeling of biochemical networks. *BMC systems biology* 4 (1), 1–18.
- Bäumelt, T., Dostál, J., 2020. Distributed agent-based building grey-box model identification. *Control Engineering Practice* 101, 104427.
- Bellu, G., Saccomani, M. P., Audoly, S., D'Angiò, L., 2007. Daisy: A new software tool to test global identifiability of biological and physiological systems. *Computer methods and programs in biomedicine* 88 (1), 52–61.
- Brastein, O. M., Sharma, R., Skeie, N.-O., 2019. Sensor placement and parameter identifiability in grey-box models of building thermal behaviour.
- Burgardt, M. I., da Silva, S. M., Ohaniana, G., San Blas, G., McLeod, J. N., Burána, M., 2010. Datos preliminares sobre el estudio de la "polilla de la vid", *Lobesia botrana* (lepidoptera: tortricidae), en mendoza.
- Carmona Torres, H., sep 2019. Pheromones for *Lobesia* pest control are available (in spanish).
- Castex, V., Cortázar-atauri, I. G. D., Calanca, P., Beniston, M., Moreau, J., 2020. Assembling and testing a generic phenological model to predict *Lobesia botrana* voltinism for impact studies. *Ecological Modelling* 420 (December 2019), 108946. DOI: 10.1016/j.ecolmodel.2020.108946
- Chis, O., Banga, J. R., Balsa-Canto, E., 2011a. Genssi: a software toolbox for structural identifiability analysis of biological models. *Bioinformatics* 27 (18), 2610–2611. DOI: 10.1093/bioinformatics/btr431
- Chis, O.-T., Banga, J. R., Balsa-Canto, E., 2011b. Structural identifiability of systems biology models: a critical comparison of methods. *PloS one* 6 (11), e27755.
- de Estudios Económicos de Bodegas de Argentina, C., 2020. Informe de actividad de la industria segundo trimestre de 2020.
- García-Tirado, J., Zuluaga-Bedoya, C., Breton, M. D., 2018. Identifiability analysis of three control-oriented models for use in artificial pancreas systems. *Journal of diabetes science and technology* 12 (5), 937–952.
- Herrera, M. E., Dagatti, C. V., Becerra, V. C., 2016. A practical rearing method for *Lobesia botrana* den. schiff. (lepidoptera: Tortricidae) under laboratory conditions (in spanish). *Magazine of the Argentine Entomological Society* 75, 160–164.
- Hong, H., Ovchinnikov, A., Pogudin, G., Yap, C., 8 2019. Sian: Software for structural identifiability analysis of ode models. *Bioinformatics* 35, 2873–2874. DOI: 10.1093/bioinformatics/bty1069
- Hotvedt, M., Grimstad, B., Imsland, L., 2021. Identifiability and physical interpretability of hybrid, gray-box models-a case study. *IFAC-PapersOnLine* 54 (3), 389–394.
- Joubert, D., Stigter, J. D., Molenaar, J., 12 2021. Assessing the role of initial conditions in the local structural identifiability of large dynamic models. *Scientific Reports* 11. DOI: 10.1038/s41598-021-96293-9
- Lema-Perez, L., 2018. Parameters Interpretability in Phenomenological-Based Semiphysical Models . A Human Glucose Homeostasis Model. Doctoral thesis, Universidad Nacional de Colombia.
- Lessio, F., Alma, A., 2021. Models applied to grapevine pests: A review. *insects* 2021, 12, 169. DOI: 10.3390/insects12020169
- Ljung, L., Torkel, G., 1994. Modeling of dynamic systems. Prentice Hall information and system sciences series, New Jersey 07632.
- Ortega-Lopez, V., Amo-Salas, M., Ortiz-Barredo, A., Diez-Navajas, A., 2014. Male flight phenology of the european grapevine moth lobesia botrana (lepidoptera: Tortricidae) in different wine-growing regions in spain. *Bulletin of entomological research* 104 (5), 566–575.
- Pasquali, S., Soresina, C., 2020. Estimation of the mortality rate functions from time series field data in a stage-structured demographic model for *Lobesia botrana*, 0–2.
- Rios Castro, F. A., et al., 2016. Calculo de temperatura umbral y requerimientos termicos pa *Lobesia botrana* a den. y schiff bajo condiciones de laboratorio. Repositorio Académico de la Universidad de Chile.
- SENASA, 2009. Resolución 362-2009.
- Simon, D., 2006. Optimal state estimation: Kalman, H infinity, and nonlinear approaches. John Wiley & Sons.
- Torres Vila, L. M., 1995. Regulatory factors of biotic potential and polyandry in the grape cluster moth *Lobesia botrana* Den. and Schiff (in spanish). Ph.D. thesis, Universidad Politécnica of Madrid.
- Touzeau, J., 1981. Modeling of the evolution of the eudemis of the vine for the midi-pyrenées region (in french). *Bol Zool Agr Bach II* 16, 26–8.
- Villaverde, A. F., Barreiro, A., Papachristodoulou, A., 2016. Structural identifiability of dynamic systems biology models. *PLoS computational biology* 12 (10), e1005153.