

**ESTIMACIÓN DEL CRECIMIENTO POBLACIONAL DE
LEPTOPHARSA GIBBICARINA EN PALMA DE ACEITE
(CASO DE ESTUDIO).**

ALEJANDRO SALAZAR HOYOS

Proyecto de tesis para el programa: Maestría en ciencias de datos y analítica
(Profundización)

Asesor, docente

Ph.D Juan Felipe Restrepo Arias

UNIVERSIDAD EAFIT

Escuela de ciencias aplicadas e ingeniería

Maestría en ciencias de datos y analítica

Medellín

2024

Índice

1. Introducción	5
2. Planteamiento del Problema	5
3. Marco Teórico	6
4. Justificación	8
5. Objetivos	9
5.1. Objetivo General	9
5.2. Objetivos Específicos	9
6. Estado del Arte	9
6.1. Modelos poblacionales – Uniespecie.	9
6.2. Modelos poblacionales – Multiespecie	10
6.3. Leptopharsa gibbicularina y su efecto en la palma de aceite.	11
6.4. Recolección de datos para entrenamiento de modelos	12
6.5. Redes Neuronales en modelos poblacionales	13
7. Metodología	14
7.1. Identificación de métodos: Metodología PRISMA	14
7.1.1. Revisión de caso de estudio y preguntas de investigación	14
7.1.2. Criterios de búsqueda y selección	15
7.1.3. Iteración de algoritmo de búsqueda	15
7.2. Selección de métodos candidatos	17
7.3. Determinación del método	19
7.3.1. Construcción del dataset	19
7.3.2. Detección y eliminación de mediciones irregulares	21

7.3.3. Imputación de datos faltantes	22
8. Resultados	22
8.1. Identificación de métodos	22
8.2. Selección de métodos candidatos	25
8.3. Desempeño del modelo elegido	28
8.3.1. Desempeño de los modelos	28
8.3.2. Influencia de los vecinos inmediatamente adyacentes y vecinos exteriores (Vecinos de los vecinos)	29
9. Discusión	31
9.1. Comparación con interpolación geoespacial usando método de Kriging	31
9.2. Partición de los datos	32
10. Conclusión	33
11. Anexos	34
11.1. Repositorio en GitHub	34

Índice de figuras

1.	Aproximación al perímetro de la plantación con polígono cóncavo	18
2.	Intersección de polígono cóncavo y plano de hexágonos	19
3.	Mediciones señaladas por lote	21
4.	Distribución de temática referente a la Pregunta (1)	24
5.	Ejemplo ESQUEMÁTICO de modelo obtenible con pocos datos concentrados en etapa temprana en contraste con modelo teórico	26
6.	Comparación de modelo logístico y exponencial	27
7.	Coefficiente de determinación para el modelo ajustado en cada hexágono	28
8.	Distribución de coeficientes de determinación de los modelos	29
9.	Ejemplo de vecinos adyacentes y vecinos exteriores	30
10.	Histograma de coeficientes correspondientes a vecinos adyacentes y externos . . .	31
11.	Serie de interpolaciones con método de Kriging	32

Índice de tablas

1.	Esquema de dataset entregado por la plantación	20
2.	Ejemplos de dataset crudo	20

Resumen

Este trabajo de grado presenta una propuesta para modelar la población de *Leptopharsa gibbicularina* (*L. gibbicularina*) usando como caso de estudio una plantación de estudio ubicada en Cimitarra - Santander, Colombia. La propuesta incluye segmentar la plantación mediante un teselado hexagonal, teniendo en cuenta un comportamiento temporal y la influencia de poblaciones aledañas. Los modelos presentaron coeficientes de determinación cercanos a 0.9 y proponen una alternativa a la autocorrelación espacial con el método de Kriging, dado que tiene en cuenta el comportamiento histórico de la plaga en cada zona de la plantación y no únicamente un instante de tiempo.

Palabras clave: Estimación población, insectos, agricultura, machine learning, modelos poblacionales, autocorrelación geoespacial.

1. Introducción

El control de plagas es una parte fundamental en la agricultura para garantizar la productividad de los cultivos. Una de las plagas con más afectación en la palma de aceite colombiana es el insecto *Leptopharsa gibbicularina*. Actualmente, la estimación de su presencia, al igual que los protocolos de contingencia, son un proceso manual, empírico y costoso; en consecuencia, la respuesta de los agricultores suele ser tardía e inoportuna, produciendo altas pérdidas en productividad.

2. Planteamiento del Problema

El caso de estudio es una plantación ubicada en Cimitarra – Santander, Colombia quienes donaron los datos para el análisis junto con sus protocolos de sanidad para la contención de *L.gibbicularina*. Los datos son conteos de adultos acompañados de Latitud, Longitud, número de lote, número de línea, número de palma, entre otros datos. Se busca proponer un modelo que describa el comportamiento de estos datos para apoyar la toma de decisiones de los agricultores.

La estimación de la población de *L. gibbicularina* se usa actualmente para activar protocolos de contingencia en la plantación cuando se supera cierto límite de individuos; implicando costos de agentes pesticidas, labor humana para su aplicación, entre otros gastos. Esta estimación de individuos se hace asumiendo una distribución uniforme a lo largo de todas las hojas en todas las palmas; de esta forma, se toma una muestra de 1 palma en cada 5 palmas y 1 línea en cada 5 líneas, luego se estima cuantos individuos hay basado en el conteo de individuos en una hoja definida por el agrónomo en jefe según la edad de la palma; si esta estimación supera cierto valor umbral, se activa el protocolo. El principal problema de este método es que es laborioso y distribuye los recursos de monitoreo uniformemente en toda la plantación, cuando algunas zonas pueden ser más propensas a desarrollar la plaga que otras.

Esta necesidad de estimar la población y velocidad de crecimiento para reaccionar oportunamente genera el interrogante: ¿qué modelo matemático podría apoyar al agricultor en la estimación de la población y su crecimiento? En este trabajo se pretende proponer un modelo matemático para describir la población de esta peste de tal forma que el agricultor pueda tener alternativas basadas en datos para la distribución eficiente de recursos.

3. Marco Teórico

Los modelos de dinámica poblacional han sido un caso de estudio de la matemática por más de 300 años. En estos modelos se tratan las relaciones de una especie con su entorno y con sí misma; explicando el crecimiento poblacional como una condición de la disponibilidad de alimento, población de depredadores, capacidad del ecosistema de alojar individuos, etc. (Malthus, 1798).

Estudios pasados han mostrado cómo se pueden usar técnicas de aprendizaje automático para apoyar la toma de decisiones en agricultura; como por ejemplo, usar reconocimiento de imágenes para la clasificación de pestes en manzanas (Boniecki et al., 2015). También ha habido trabajos todavía más cercanos al propuesto usando redes neuronales para modelar poblaciones de

moscas en cultivos de aguacates (Ibrahim et al., 2022) o para modelar la cantidad de mosquitos en áreas urbanas (Lee et al., 2016). Otros trabajos incluso comparan múltiples métodos de modelación para la predicción de plagas en maní (Vennila et al., 2017).

El control de la *L. gibbicularina* ha sido documentado por su alto impacto en la producción de las plantaciones de palma de aceite (Aldana de La Torre, 2010; Barrios y Bustillo, 2014; Barrios-Trilleras et al., 2015); y a pesar de que se han propuesto muchas alternativas desde la biología para su control, son pocos los estudios desde la matemática y ciencia de datos para apoyar la implementación de estas alternativas.

Las alternativas propuestas desde la biología se separan en 2 grandes grupos:

- Controles químicos (Agentes basados en la intoxicación del insecto).
- Controles biológicos (Agentes basados en la introducción de un organismo vivo, usualmente depredadores naturales del insecto).

Los controles químicos son conocidos por su bajo costo y su alta eficacia inmediata. Pero por su gran número de desventajas y alta destructividad del medio ambiente, cada vez se han adoptado más métodos biológicos (Bustillo, 2014); los cuales han mostrado resultados de mortalidad de *L. gibbicularina* por encima del 79.2 % de los individuos en condiciones reales de plantación (Barrios T. et al., 2016).

De esta forma, las decisiones de intervención no pueden basarse únicamente en la toma de datos y experiencia empírica de los agricultores, que genera alertas únicamente cuando una medición de población ya superó cierto valor. Este valor puede estar errado o sesgado, produciendo una intervención prematura, desperdiciando agente biológico de control y requiriendo una segunda intervención o una intervención tardía luego de que los daños son considerables. No obstante, una intervención con alta tasa de mortalidad no es suficiente para un adecuado control de la plaga, ya que el daño generado por esta produce secuelas hasta 33 meses después de la intervención (Espriella, 1988). Es ahí donde cobra importancia la oportunidad de la intervención; no solo debe priorizarse acabar con la mayoría de los individuos de la plaga, sino hacerlo minimizando los

daños.

En este trabajo se buscó evaluar modelos tradicionales estadísticos como lo son los modelos logísticos, malthusianos y Lotka-Volterra (Lotka, 1925) y modelos de aprendizaje automático, buscando describir el crecimiento poblacional de *L. gibbicarina* con la mayor precisión posible. Para esto se realizó una revisión bibliográfica recolectando los modelos documentados hasta la fecha, se filtraron estos modelos según los datos disponibles y el comportamiento de la población, y se compararon usando métricas estadísticas como lo es el coeficiente de determinación (R^2).

4. Justificación

El problema de *L. gibbicarina* (Chinche de Encaje) en la palma de aceite en Colombia ha sido documentado por la Corporación Centro de investigación en palma de aceite (Cenipalma) con el apoyo de la Federación nacional de cultivadores de Palma (Fedepalma) (Aldana de La Torre, 2010). La *L. gibbicarina* tiene un ciclo de vida de 72.3 días aproximadamente repartidos en 3 etapas: huevo, ninfa y adulto (Barrios y Bustillo, 2014) el cual genera lesiones en las hojas de la palma permitiendo la infección por hongos de la familia *Pestalotiopsis*. Este hongo genera manchas en las hojas de la palma reduciendo su follaje hasta un 66 %; esto provoca una reducción en el peso de los racimos de hasta un 40 % (Barrios-Trilleras et al., 2015). Adicionalmente, los productos utilizados para su control implican una inversión en su adquisición y aplicación; la cual aumenta proporcionalmente a los lotes de la plantación en estado crítico. La producción de fruta de palma por hectárea se ha documentado cercana a las 20 ton/ha tomando de muestra la producción de países como Malasia, Indonesia, Tailandia y Costa Rica entre los años 1999 y 2004 (Simamora et al., 2018). Esto implica que una reducción del 40 % en el peso de los racimos equivale a una pérdida de productividad de 8 ton/ha.

5. Objetivos

5.1. Objetivo General

Formular un modelo para el crecimiento poblacional de *Leptopharsa gibbicularina* basado en datos recolectados en una plantación de palma de aceite en Cimitarra – Santander.

5.2. Objetivos Específicos

- Clasificar y mapear los métodos documentados en la literatura para estimar el crecimiento poblacional de insectos u otras plagas en agricultura.
- Seleccionar los métodos candidatos para la estimación basados en la caracterización de los datos poblacionales y comportamiento de *Leptopharsa gibbicularina*.
- Determinar el método más adecuado para la estimación de crecimiento poblacional de *Leptopharsa gibbicularina* basado en métricas estadísticas.

6. Estado del Arte

6.1. Modelos poblacionales – Uniespecie.

A lo largo de la historia se han propuesto múltiples modelos poblacionales; uno de los primeros precursores de este campo fue Thomas Robert Malthus. En su ensayo “Ensayo sobre la población humana” (Malthus, 1798), Malthus relaciona la población humana con factores como la comida disponible, la industrialización agrícola, entre otros. En este ensayo, Malthus plantea que, si una población no tiene ningún límite para crecer, su crecimiento seguirá un comportamiento exponencial (Malthus, 1798).

Los modelos exponenciales (o Malthusianos) son correctos únicamente en las primeras etapas del crecimiento de la población y están descritos por la siguiente ecuación diferencial:

$$\frac{dN}{dt} = rN \rightarrow N(t) = N_0 e^{rt} \quad (1)$$

Donde $N(t)$ es la población en un instante t , N_0 es la población inicial, r es la tasa de crecimiento poblacional y t es el tiempo. No obstante, si se incluyen variables como depredación o competencia por alimento, el problema se vuelve menos trivial (Tsoularis y Wallace, 2002). Ante esta problemática se propuso el modelo de crecimiento logístico; este modelo supone una eventual estabilización de la población cuando esta alcanza su punto de saturación, este punto está dado por características del entorno como depredación, competencia por alimento, sobrepoblación de los individuos, etc. Algunos de los predecesores de este caso fueron Verhulst-Pearl (Tsoularis y Wallace, 2002) basados en el trabajo de Lotka (Lotka, 1925) quienes proponen la siguiente ecuación diferencial.

$$\frac{dN(t)}{dt} = rN \left(1 - \frac{N}{K} \right) \rightarrow N(t) = \frac{KN_0}{(K - N_0)e^{-rt} + N_0} \quad (2)$$

Donde $N(t)$ es la población en un instante t , $\frac{dN(t)}{dt}$ es el crecimiento poblacional en un instante t , K es la capacidad de carga de la población, r es el crecimiento de la población y N_0 es la población inicial. Nótese que una población inicial de 0 no puede crecer, luego de un tiempo infinito la población converge a K y el punto de crecimiento máximo es exactamente cuando la población es $\frac{K}{2}$ (Tsoularis y Wallace, 2002). Otros modelos se han planteado en el último siglo teniendo en cuenta variables adicionales a la natalidad, como las muertes de los individuos e inmigración, demostrando que, si la tasa de natalidad no supera la tasa de mortalidad, la población se extingue (Ricciardi, 1986).

6.2. Modelos poblacionales – Multiespecie

Partiendo de la idea de que la población de una especie se puede ver afectada por la población de sus depredadores, Alfred J. Lotka en 1925 propone utilizar las siguientes ecuaciones diferenciales para plantear lo que se conoce como el sistema Lotka-Volterra de depredador y presa

(Lotka, 1925). En este sistema se describe la población de la presa como inversamente proporcional a la población del depredador y la población del depredador, proporcional a la población de la presa. Dando como resultado el siguiente sistema de ecuaciones diferenciales (Goel et al., 1971).

$$\frac{dN_1}{dt} = N_1(\alpha_1 - \lambda_1 N_2) \quad (3)$$

$$\frac{dN_2}{dt} = N_2(-\alpha_2 + \lambda_2 N_1) \quad (4)$$

Donde N_1 y N_2 describen las poblaciones de presa y depredador, respectivamente. Nótese que si N_2 (depredador) comienza a crecer desmesuradamente, por efecto de la ecuación 3, N_1 (presa) comenzará a decrecer. Y de forma similar, por efecto de la ecuación 4, si N_1 crece desmesuradamente, N_2 también lo hará. Los otros parámetros representan la interacción de las especies teniendo en cuenta la naturaleza de su convivencia (Goel et al., 1971).

6.3. *Leptopharsa gibbicularina* y su efecto en la palma de aceite.

A la hora de hablar específicamente de poblaciones de insectos, otros factores deben tomarse en cuenta; la etapa de la vida del insecto, por ejemplo. Suponiendo 3 etapas de vida (huevo, ninfa y adulto), no sería correcto suponer que todas las etapas aportan igual a la dinámica poblacional. Las ninfas y los huevos no se reproducen, consumen alimento ni tienen la misma mortalidad que los adultos; entonces no tiene sentido incluirlos en un modelo sin discriminación (Buffoni y Pasquali, 2007). Esta problemática fue abordada por Buffoni – Pasquali proponiendo un modelo donde se estima la población no solo en un instante temporal, sino también en una etapa de la vida (Buffoni y Pasquali, 2007).

Barrios en 2015 expone que *Leptopharsa gibbicularina* tiene una expectativa de vida de 72.1 ± 11.1 días distribuidos de la siguiente manera: huevo 16 ± 2 días, ninfa 18.9 ± 3.3 días y adulto 37.2 ± 5.9 días bajo unas condiciones de temperatura de 27.3 ± 3.8 °C y humedad relativa de 85 ± 15 % (Barrios-Trilleras et al., 2015). Estos parámetros servirán de base para el ajuste del modelo

de estimación poblacional. En esta investigación se hace una amplia documentación de las tasas de mortalidad en cada una de estas etapas, demostrando que es mayor en la etapa de huevo que en cualquier otra (Barrios-Trilleras et al., 2015).

Un método tradicional para el control de plagas es el control químico del insecto; este tipo de control es muy eficaz a nivel económico, pero tiene desventajas como contaminación de fuentes fluviales y del medio ambiente en general, el desarrollo de resistencia por parte de las plagas con el paso del tiempo, la destrucción de fauna benéfica como predadores de las plagas o polinizadores, entre otras (Bustillo, 2014). En el último siglo se han documentado alternativas biológicas al control químico en las que se destacan dos: la utilización de hormigas del género *Crematogaster*, conocidas por ser depredadores naturales de la plaga, y la utilización de hongos entomopatógenos: *Purpureocillium lilacinum*, *Beauveria bassiana*, *Isaria fumosorosea*, entre otros (Barrios-Trilleras et al., 2015).

La utilización de hongos entomopatógenos se ha documentado en condiciones controladas con mortalidades entre el 74.4 % y el 100 % (Barrios T. et al., 2016). Estos resultados no distan mucho de los obtenidos en plantaciones reales donde *Purpureocillium lilacinum* obtuvo la mayor mortalidad con una tasa del 80.7 % (Barrios T. et al., 2016)

6.4. Recolección de datos para entrenamiento de modelos

Para la toma de datos para estos modelos se han propuesto técnicas como trampas para la captura de individuos; buscando estimar la densidad poblacional a partir de los individuos capturados en cada trampa, la cantidad de trampas y su distribución (Petrovskii et al., 2012). Este método tiene limitantes e implicaciones relativas al comportamiento del insecto ya que la presencia de la trampa altera su libre movimiento; esto produce un sesgo en la densidad poblacional al ser mucho menor cerca de la trampa (Petrovskii et al., 2012). En el caso de *Leptopharsa gibbicarina* en la plantación de estudio, su detección se hace manualmente contando la cantidad de individuos en las hojas; por lo general, están ubicados en el tercio medio e inferior de la palma (Aldana de La Torre, 2010). La cantidad de individuos considerada crítica varía de plantación en plantación, como

expone Aldana en 2010.

6.5. Redes Neuronales en modelos poblacionales

Las redes neuronales han mostrado un buen desempeño en el problema de estimación poblacional de plagas en agricultura. Algunos casos de aplicación han involucrado cultivos como aguacates, manzanas, tabaco, entre otras (Boniecki et al., 2015, Ibrahim et al., 2022, Vennila et al., 2017). También se han documentado aplicaciones en estimación de mosquitos en áreas urbanas (Ibrahim et al., 2022). Esos modelos se han usado en una amplia gama de variantes con diferentes objetivos; algunos buscan clasificar el tipo de plaga presente en la especie afectada usando clasificación de imágenes (Boniecki et al., 2015), otros más cercanos al objetivo de esta tesis buscan predecir la cantidad de adultos de una determinada peste en cultivos de tabaco usando redes neuronales polinomiales (PNN) con resultados $R^2 = 0,89$; prometiendo ser una alternativa viable para el campo de estudio (Vennila et al., 2017). Otros estudios comparan el desempeño de redes neuronales con métodos lineales como Multiple Linear Regression (MLR) donde mostraron que los métodos lineales pueden desempeñarse igual de bien que las redes en ciertas condiciones, por lo que no debe descartarse la utilización de métodos lineales (Lee et al., 2016)

El entrenamiento de estos modelos por lo general incluye variables climáticas como temperatura, humedad relativa y precipitaciones; pero eso no implica que no se hayan incluido otras variables relevantes para el problema como el estado actual de la especie invadida (Ibrahim et al., 2022). En aguacate, por ejemplo, relacionar las pestes con el estado de la planta (Dormida, floración, cosecha, etc. . .) puede ser de especial interés para determinar el momento más adecuado para la toma de decisiones (Ibrahim et al., 2022). Incluir estas variables cualitativas requiere de un proceso de transformación que puede acarrear otros problemas como dimensionalidad o escalabilidad de la variable. Una alternativa propuesta por Eric A. Ibrahim en 2022 es la utilización de lógica difusa en las variables de entrada de la red (Ibrahim et al., 2022).

7. Metodología

7.1. Identificación de métodos: Metodología PRISMA

La revisión de literatura se basó en los lineamientos de la metodología PRISMA expuestos en Page et al., 2021. Para esto, se plantearon dos preguntas de investigación basadas en el caso de estudio, se definieron criterios de exclusión, se iteró para refinar los criterios de búsqueda y, por último, se eligieron 33 artículos basados en su título y resumen para ser leídos en su totalidad.

7.1.1. Revisión de caso de estudio y preguntas de investigación

Los datos disponibles para el desarrollo de este proyecto son el conteo de adultos de *L. gibbicularina* en una plantación de palma de aceite cercana a Cimitarra – Santander, Colombia. El registro data de agosto de 2022 hasta la fecha actual, octubre de 2024. El esquema de datos entregado es el siguiente:

Los datos se recolectaron manualmente usando la aplicación CyberTracker (Cybertracker, s.f.) por un equipo de sanidad, siguiendo el siguiente protocolo:

1. Cada lote se muestrea cada 15 días aproximadamente.
2. Se realiza el conteo de individuos de la peste para una única hoja registrando su número.
3. Se toma de muestra una línea de cada 5 líneas y una palma de cada 5 palmas. Es decir, se toma de la línea 1 las palmas 1, 5, 10... luego se repite el proceso en las líneas 5, 10, 15...
4. El conteo se hace completo (Individuo por individuo) en media hoja si la cantidad de individuos es menor a 20 por foliolo y luego se multiplica por 2.
5. Si la cantidad de individuos por foliolo supera los 20, se eligen aleatoriamente 10 foliolos, se toma el promedio de individuos y se multiplica por 270 que es el promedio de foliolos por hoja.

6. Se procura que un mismo inspector no realice el control de un mismo lote repetidamente para evitar sesgos en la medición.

Dada la recolección de datos que se implementó, surge la necesidad de buscar modelos evaluados en condiciones reales poco controladas y entrenados con datos recolectados en condiciones similares con altos márgenes de error humano. De esta forma, se propusieron las siguientes preguntas de investigación para la revisión de literatura:

1. ¿Qué métodos han sido usados para recolectar variables y datos de una población de insectos?
2. ¿Que métodos se han usado específicamente en estimación poblacional de insectos?

7.1.2. Criterios de búsqueda y selección

Los criterios de selección de artículos fueron:

- Artículos publicados en las bibliotecas virtuales Springer, Scopus y JSTOR, o fuentes específicas tales como publicaciones de Cenipalma o Fedepalma.
- Artículos publicados después del año 2020 (inclusive); con el fin de priorizar artículos que incluyan nuevas técnicas de deep learning e inteligencia artificial generativa.
- Artículos en inglés o español.

7.1.3. Iteración de algoritmo de búsqueda

La búsqueda se realizó el día 28 de julio de 2024. Se refinó el algoritmo de búsqueda usando únicamente SCOPUS y luego se buscó con el algoritmo refinado en las librerías Springer (SpringerLink) y JSTOR. Se aplicó el algoritmo a búsqueda de palabras en el título, resumen y palabras clave.

1. **Primera búsqueda:** La primera búsqueda se realizó con el siguiente criterio "Population Dynamics"; luego de filtrar por año e idioma esta consulta devolvió 22,361 documentos.

2. **Segunda búsqueda:** Como se pretende encontrar modelos de dinámicas poblacionales, se limitó la búsqueda con la siguiente consulta:

"Population dynamics"

AND

("Model" OR "Modeling" OR "estimation")

Esto redujo la cantidad de documentos a 9793.

3. **Tercera búsqueda:** Para enfocar la consulta en crecimiento poblacional se agregó:

"Population dynamics"

AND

("Model" OR "Modeling" OR "estimation")

AND

"population growth"

generando 791 resultados

4. **Cuarta búsqueda:** Muchos artículos no se refieren al conteo de individuos específicamente sino a otros factores de la dinámica poblacional como expansión Urbana e impacto medioambiental o simplemente crecimiento pero de poblaciones humanas. Limitando la búsqueda de esta manera y obteniendo 495 resultados.

"Population dynamics"

AND

("Model" OR "Modeling" OR "estimation")

AND

"population growth"

AND NOT

("Global" OR "Urban" OR "Urbanization" OR "Human" OR "Humanity")

5. **Quinta búsqueda:** Ahora, se buscará enfocar la consulta en poblaciones de insectos exclu-

yendo estudios relativos al crecimiento de bacterias o enfermedades de la siguiente manera:

”Population dynamics”

AND

(”Model” OR ”Modeling” OR ”estimation”)

AND

”population growth”

AND

(”insect” OR ”insects” OR ”Colony”)

AND NOT

(”Global” OR ”Urban” OR ”Urbanization” OR ”Human” OR ”Humanity” OR ”disease”
OR ”bacteria”)

Reduciendo el número a solo 40 documentos en SCOPUS.

Por último, se replicó la búsqueda en Springer y JSTOR; luego de eliminar artículos duplicados y filtrar los que por su resumen o título no fueran pertinentes para la investigación, se eligieron 31 artículos para lectura completa.

7.2. Selección de métodos candidatos

Luego de la revisión de literatura, se filtraron los artículos que habían recolectado datos en condiciones similares al caso de estudio de la plantación de palma de aceite y se seleccionaron los métodos candidatos para evaluación, basado en el ajuste que tuvieron. El ajuste se midió basado en el coeficiente de determinación R^2 .

También, para la selección del método candidato, y teniendo en cuenta que debe apoyar la toma de decisiones del agricultor, se propuso subdividir la plantación en hexágonos de igual tamaño. Esto buscó ajustar un modelo a cada polígono por separado, describiendo el comportamiento que ha tenido la peste en esta sección de la plantación en particular. Esto se logró definiendo un polígono cóncavo que aproxime el perímetro exterior de la plantación y obteniendo la intersección con un plano infinito de hexágonos, obteniendo los resultados en la Figura 2. Plantear el problema

de esta forma tiene la ventaja de permitir ajustar un modelo en cada subdivisión, usando como variable respuesta su población y como variable de entrada el tiempo y las demás poblaciones. Es decir, el modelo seguirá una estructura como la que se describe en la ecuación 5.

$$N_{t,H_i} = f(\tau, H_1, H_2, \dots, H_{i-1}, H_{i+1}, \dots, H_n), \quad \tau \in \mathbb{R}, t \text{ es un instante específico en } \tau \quad (5)$$

Donde la población del i -ésimo hexágono H_i en un instante t se representa como N_{t,H_i} y es una función del tiempo τ y de las poblaciones de los demás hexágonos $H_1, H_2, \dots, H_{i-1}, H_{i+1}, \dots, H_n$. En este orden de ideas, se debería obtener un coeficiente de tiempo que indique cuánto permite la proliferación de la peste el hexágono por sí mismo, junto con coeficientes de los otros polígonos que indican cuánto afectan estas poblaciones la del hexágono en cuestión. De esta forma, el modelo candidato debe ser una regresión que permita ajustarse al planteamiento anterior.

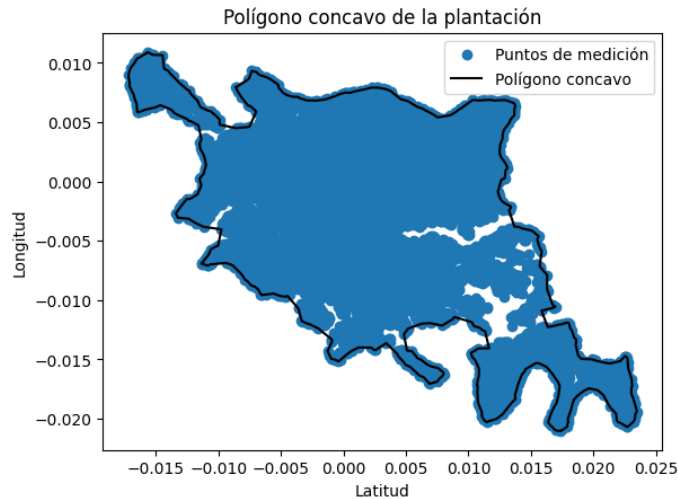


Figura 1: Aproximación al perímetro de la plantación con polígono cóncavo

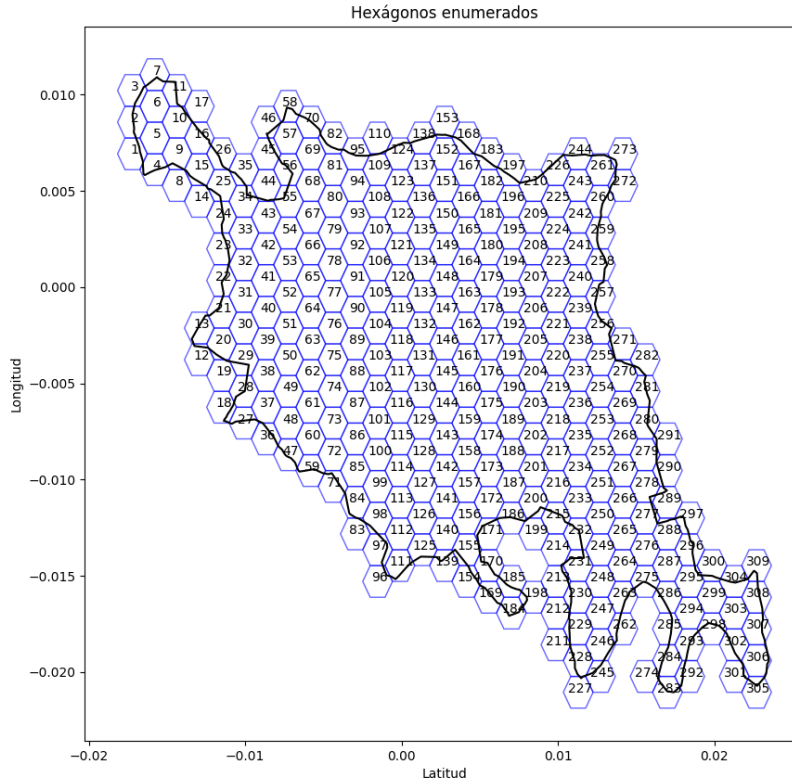


Figura 2: Intersección de polígono cóncavo y plano de hexágonos

7.3. Determinación del método

7.3.1. Construcción del dataset

Según el protocolo explicado en la sección 6.1.1, solo se tiene una medición por cada 5 palmas en ventanas de 15 días. De esta forma, partiendo del CSV entregado por la plantación (Tabla 1), se construyó el dataset tomando la máxima medición de Adultos etiquetados como *Lepthoparsa* para cada palma registrada en ventanas de 15 días, obteniendo 34 registros para cada palma medida.

Tabla 1: Esquema de dataset entregado por la plantación

Campo	Descripción
Date	Fecha
Time	Hora
Latitude	Latitud
Longitude	Longitud
Evaluador	Operario quien realizó la medición
Lotes	Lote
Enfermedad	La base de datos cuenta también con información de enfermedades encontradas
Línea	Número Línea
Palma	Número Palma
N° Hoja	Número de hoja
Plagas	En este campo es donde se reporta si la peste encontrada fue Leptopharsa u otra
Larva	Conteo de larvas. NOTA: Datos incompletos e inconsistentes
Estado	Estado de la larva: Mediano, pequeño. . .
Pupas	Conteo de pupas. NOTA: Datos incompletos e inconsistentes
Huevos	Conteo de huevos. NOTA: Datos incompletos e inconsistentes
Adultos	Adultos de las plagas

De esta forma, el dataset crudo con algunos ejemplos de registros fue el siguiente:

Date	Latitude	Longitude	Evaluador	Lotes	Linea	Palma	Adultos
2/2/2023	0.008713887	0.005137672	Ivan *****	656	30	26	8
2/4/2023	-0.01002148	-0.000121518	Ivan *****	358	45	5	4
2/24/2023	0.007166604	-0.006742528	Ivan *****	742	46	5	8
2/24/2023	-0.001730213	-0.004357698	Ivan *****	444	16	30	10
3/14/2023	0.00392107	0.002520202	Ivan *****	533	91	25	6

Tabla 2: Ejemplos de dataset crudo

7.3.2. Detección y eliminación de mediciones irregulares

La naturaleza del problema admite dos tipos de outliers no excluyentes: outliers espaciales (un conteo de adultos correcto en un lugar incorrecto) y outliers poblacionales (un conteo de adultos incorrecto). Los outliers espaciales se pueden apreciar en la Figura 3. El manejo de estos outliers se realizó, en primer lugar, utilizando un algoritmo KNN con distancia euclidiana para detectar marcas al interior de la plantación, pero etiquetadas en un lote incorrecto. De esta forma, se definió un espacio vectorial conformado por la longitud y latitud con el número de lote como etiqueta de los datos. La selección de la distancia euclidiana se basó en que los datos son coordenadas geoespaciales y esta distancia representa la distancia física entre los datos.

En segundo lugar, se implementó un algoritmo DBSCAN en este mismo espacio vectorial para detectar las mediciones por fuera de la plantación, posiblemente por errores del GPS durante la medición.

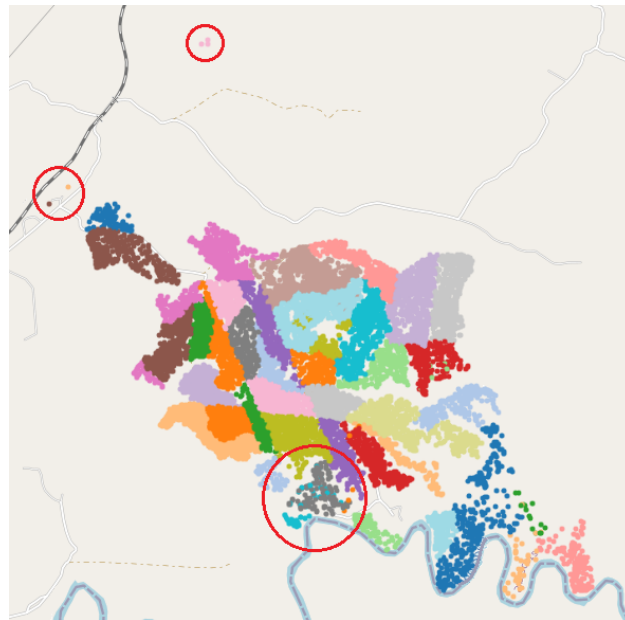


Figura 3: Mediciones señaladas por lote

Para el manejo de los outliers poblacionales, se debe tener en cuenta que cada hexágono puede contener una cantidad de palmas entre 1 y 167; cada palma de estas tienen una medición cada 15 días aproximadamente. En este orden de ideas, se utilizó el concepto de mediana de la siguiente

forma; para cada hexágono en una ventana de tiempo de 15 días, se tomó como su población la mediana de todas las mediciones de todas las palmas ubicadas en su interior. Es necesario resaltar que se definió un umbral mínimo para la cantidad de palmas que debe tener un hexágono para ser considerado. Se calculó el promedio de palmas (Con mediciones) por hexágono (60,729 palmas) y la desviación estándar (41,869); de esta forma se definió el umbral mínimo como promedio menos desviación estándar, dando como resultado 19 palmas. En otras palabras, un hexágono debe contener por lo menos 19 palmas con mediciones de plagas para ser considerado; posteriormente, se toma la mediana de estas mediciones en cada ventana de tiempo para definir la estimación de la población de ese hexágono en esa ventana. Este método, fundamentado en que palmas cercanas deben tener poblaciones similares, controla los outliers poblacionales.

7.3.3. Imputación de datos faltantes

En ocasiones, algunas zonas de la plantación no son medidas en plazos superiores a la ventana de 15 días, ocasionando datos faltantes. La imputación de estos datos se hizo en dos etapas, una primera etapa espacial, donde se aproximó el valor faltante usando la mediana de las poblaciones de los seis vecinos inmediatamente aledaños en esa misma ventana de tiempo. La segunda etapa se implementó en los casos en los que las poblaciones vecinas también estaban faltantes; estos casos eran raros y generalmente correspondían a hexágonos limítrofes de la plantación con pocos vecinos; esta etapa utilizó un enfoque temporal, aproximando el valor como el promedio de las ventanas posterior y anterior más cercanas no faltantes.

8. Resultados

8.1. Identificación de métodos

Los resultados de la pregunta (1) se exponen en la Figura 4. Se determinó que la mayoría de los artículos no usaban poblaciones de insectos; bien porque es un artículo completamente teórico y no hacen pruebas en poblaciones reales o porque las hacen, pero en poblaciones de otro tipo. En

segundo lugar, se usan datos de institutos dedicados al monitoreo de un insecto o trabajos anteriores del autor; por este motivo, no se describe en detalle la metodología utilizada para la obtención de estos datos. Empatado en segundo lugar están los estudios pertinentes para esta investigación que son los datos reales tomados en ambientes poco controlados de campo; el método más común en este tipo de prácticas consiste en la separación por lotes del área de estudio, generalmente de igual tamaño, el despliegue de un sistema de trampas y, pasado un tiempo, se cuentan manualmente los individuos capturados para obtener el estimado de densidad poblacional utilizando el área del lote; estos estudios suelen resaltar el sesgo que hay referente a la capacidad de contar los insectos en las trampas por parte de quien está tomando los datos. En cuarto lugar están los datos reales controlados que son datos tomados generalmente en invernaderos pequeños, con condiciones de humedad, temperatura y luminosidad controladas; también son datos tomados por un equipo de investigadores con equipos más sofisticados que un simple conteo manual. En quinto lugar están los estudios con datos simulados a partir de parámetros de natalidad, mortalidad, migración, etc. junto con redes neuronales, cadenas de Markov, procesos Monte Carlo u otros métodos estocásticos. Los cuatro artículos restantes no se incluyeron porque no se logró acceder a ellos por ningún recurso disponible.

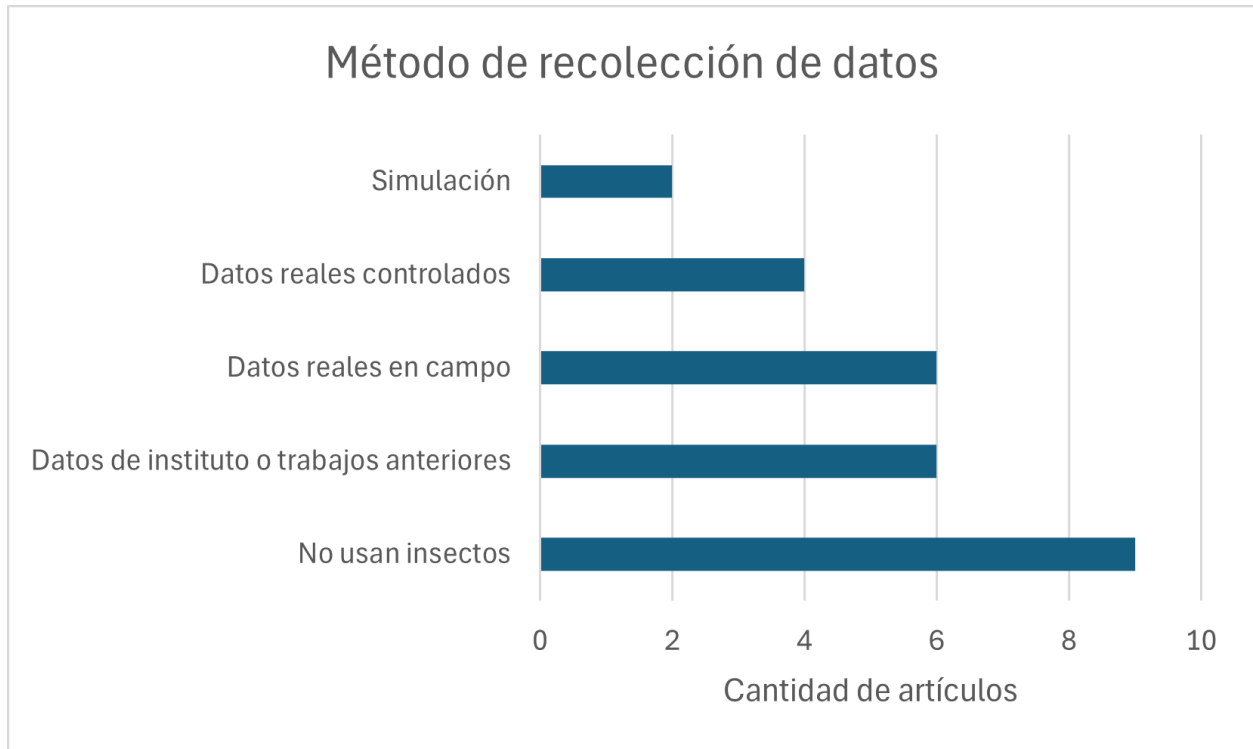


Figura 4: Distribución de temática referente a la Pregunta (1)

Respecto a la pregunta (2), se encontraron seis tipos de modelos para la descripción de poblaciones:

- Modelos basados en Ecuaciones diferenciales
- Modelos basados en Regresión logística
- Modelos basados en Funciones lineales
- Modelos basados en Matrices
- Modelos basados en Métodos estocásticos
- Modelos basados en Series de tiempo

Pero si se tienen en cuenta únicamente los artículos que utilizan datos reales de campo, solamente se incluirían Regresión logística y Funciones lineales. Algunos trabajos con Regresión

logística alcanzan unos coeficientes de determinación de hasta 0.86 en condiciones reales de campo y hasta 0.92 en condiciones controladas. Mientras que los artículos con regresión lineal no entregan información del ajuste del modelo en condiciones reales, pero con datos de instituto o condiciones controladas logran un ajuste de tan solo 0.286 en el mejor de los casos. Incluso, algunos de estos estudios resaltan la relación no lineal entre las variables de entrada y la población.

8.2. Selección de métodos candidatos

La revisión de literatura determinó que el método con mejor desempeño en condiciones similares fue la regresión logística, por lo que se tomó como punto de partida la ecuación 2. Esta ecuación se ajustó para ser alimentada con las 234 variables poblacionales de los hexágonos y la variable temporal de la siguiente forma:

$$N_{t,H_i} = \frac{KN_0}{(K - N_0)e^{\vec{H} \cdot \vec{A} + \beta} + N_0} \quad (6)$$

Donde N_{t,H_i} es la población del hexágono i en un instante t , N_0 es la población inicial, K es la saturación del entorno, que se traduce como el valor teórico de individuos que una palma puede alojar, \vec{H} es el vector de variables $(\tau, H_1, H_2, \dots, H_{i-1}, H_{i+1}, \dots, H_n)$, \vec{A} es el vector de coeficientes $(\alpha_\tau, \alpha_{H_1}, \dots, \alpha_{H_n})$ y β es un término independiente.

El modelo descrito por la ecuación 6 presenta varios inconvenientes. El primero de ellos es que el dataset solo cuenta con 34 registros que representan cerca de un año y medio de mediciones, pero tiene 237 coeficientes (234 poblaciones de hexágonos, un coeficiente de tiempo, un término independiente β y la saturación del entorno desconocida K), por lo que no existe solución única. Por este motivo, se decantó por una alternativa más simple usando únicamente los seis vecinos inmediatos, los doce vecinos adyacentes a estos y la variable tiempo. De esta forma, el modelo requiere ajustar como máximo 21 parámetros.

El segundo inconveniente está relacionado con el ajuste del modelo en etapas tempranas de la población. Dado que el agricultor no permite que la población supere cierto umbral, todos los

datos corresponden a etapas tempranas de la población; en esta etapa, la curva no sigue el mismo comportamiento que tendría una población no intervenida por tiempo infinito. Para ejemplificar mejor, al suponer por un momento que se requiere ajustar una única población utilizando únicamente el tiempo como variable de entrada, se obtendrá un resultado como el de la Figura 5, en el que la falta de datos en etapas posteriores del desarrollo de la población hace que la curva que mejor se ajusta diste mucho del modelo teórico. En consecuencia, la curva ajustada sugiere una población de saturación poco realista cuando debería estar por el orden de los miles.

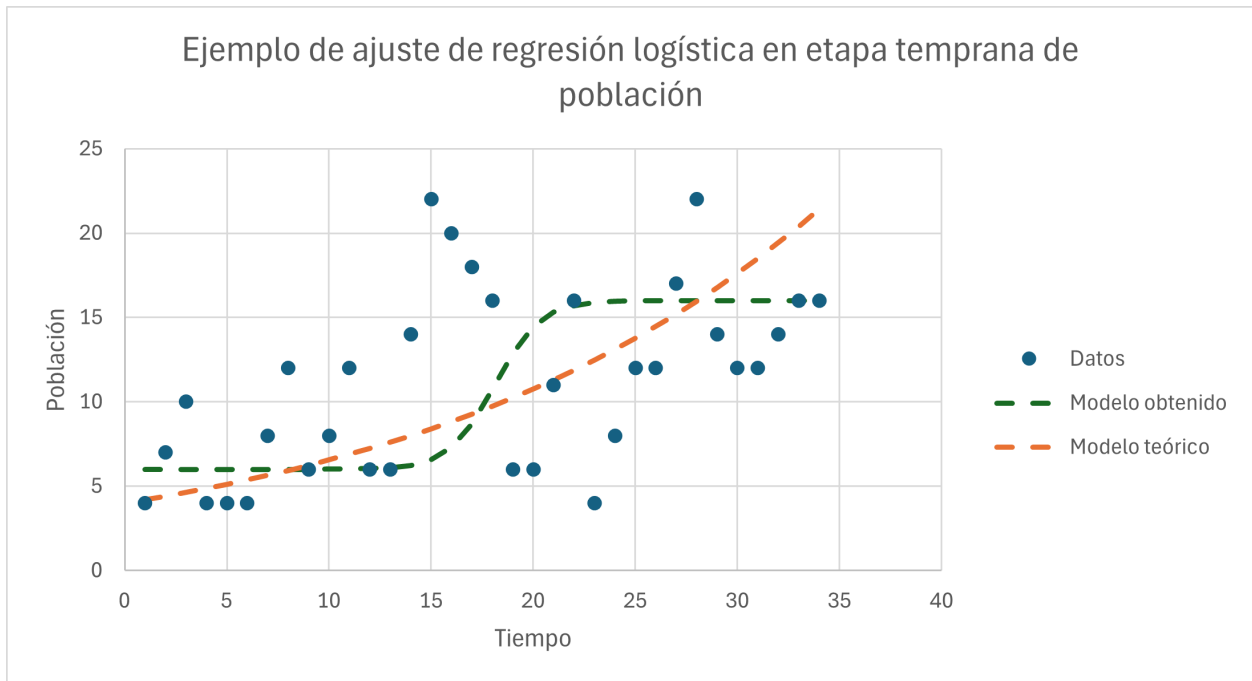


Figura 5: Ejemplo ESQUEMÁTICO de modelo obtenible con pocos datos concentrados en etapa temprana en contraste con modelo teórico

Por estos motivos se prefirió utilizar el modelo descrito en la Ecuación 1. Este modelo en etapas tempranas del desarrollo de la población se comporta de forma muy aproximada a la regresión logística, como se puede ver en la Figura 6.

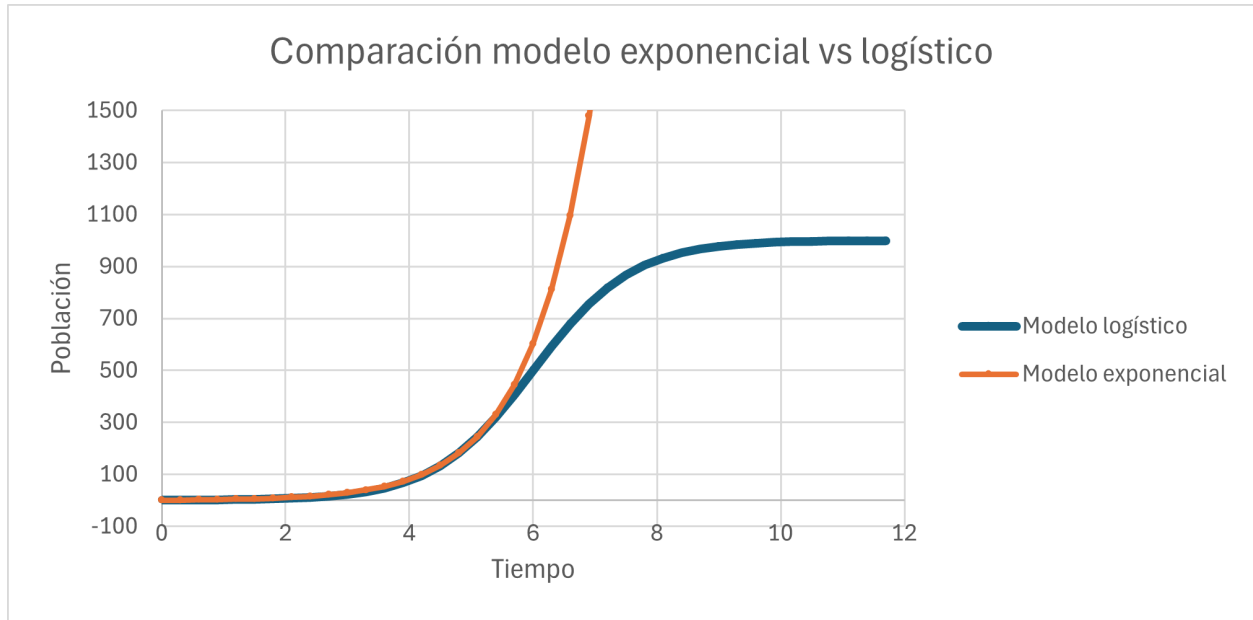


Figura 6: Comparación de modelo logístico y exponencial

De esta forma, se ajustó el modelo exponencial como se muestra en la Ecuación 7 y simplificándolo en la Ecuación 8.

$$N_{t,H_i} = e^{\vec{H} \cdot \vec{A} + \beta} \quad (7)$$

$$\ln(N_{t,H_i}) = \vec{H} \cdot \vec{A} + \beta \quad (8)$$

Donde N_{t,H_i} es la población del hexágono i en un instante t , \vec{H} es el vector de variables incluyendo únicamente los seis vecinos inmediatos y los doce vecinos adyacentes a estos $(\tau, H_1, H_2, \dots, H_{i-1}, H_{i+1}, \dots, H_n)$, \vec{A} es el vector de coeficientes $(\alpha_\tau, \alpha_{H_1}, \dots, \alpha_{H_n})$ y β es un término independiente. De esta forma, el modelo elegido es simplemente una regresión lineal ajustada al logaritmo natural de las poblaciones en la variable respuesta.

8.3. Desempeño del modelo elegido

8.3.1. Desempeño de los modelos

En la Figura 7 se aprecia el coeficiente de determinación de cada uno de los modelos individuales ajustados a cada hexágono. Se destaca un ajuste particularmente bueno en la zona suroriental de la plantación que linda con una fuente fluvial. Además, se puede observar en la figura 8 que la mayoría de los modelos se ajustan con coeficientes de determinación por encima de 0.8. Eso implica que esta metodología podría ser comparable con los modelos encontrados en el estado del arte.

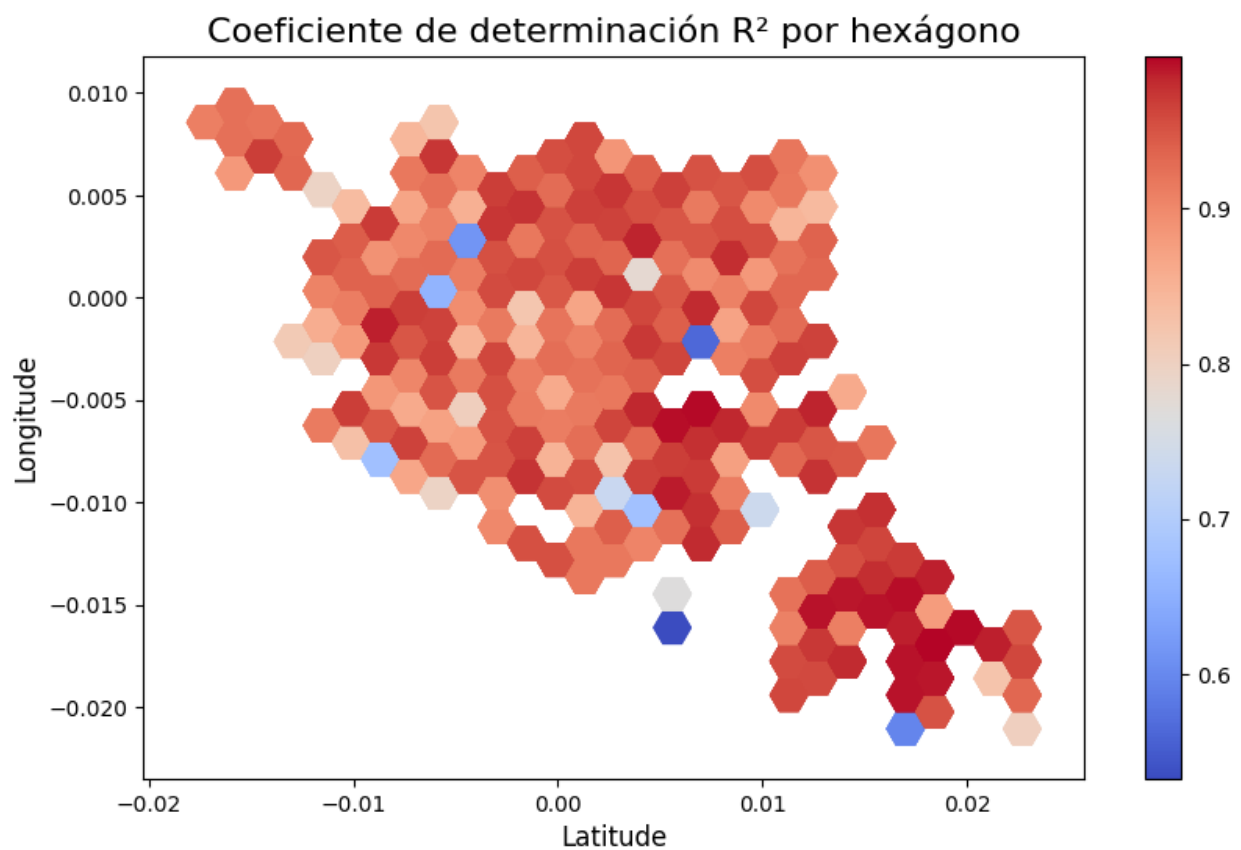


Figura 7: Coeficiente de determinación para el modelo ajustado en cada hexágono

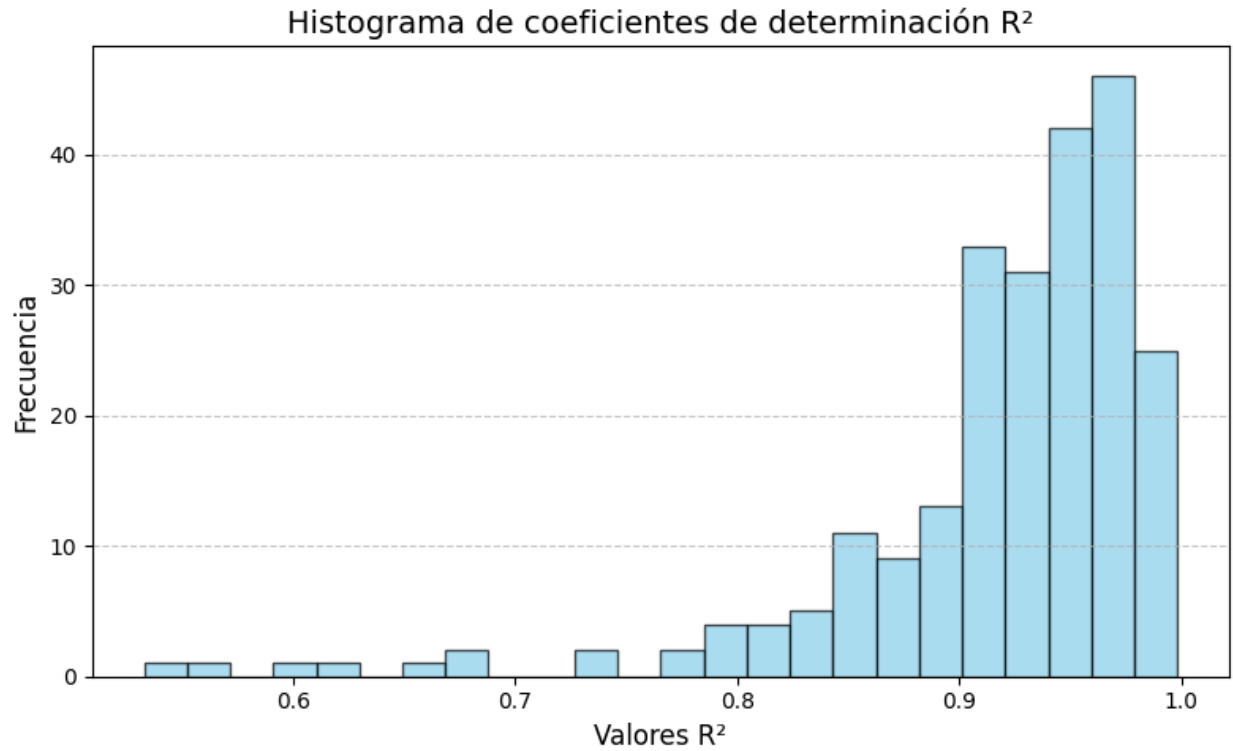


Figura 8: Distribución de coeficientes de determinación de los modelos

8.3.2. Influencia de los vecinos inmediatamente adyacentes y vecinos exteriores (Vecinos de los vecinos)

Los vecinos adyacentes son aquellos que limitan inmediatamente con un hexágono cualquiera; los vecinos exteriores son todos aquellos que limitan con los vecinos adyacentes sin ser uno y sin ser el hexágono en cuestión. Ejemplo en la figura 9.

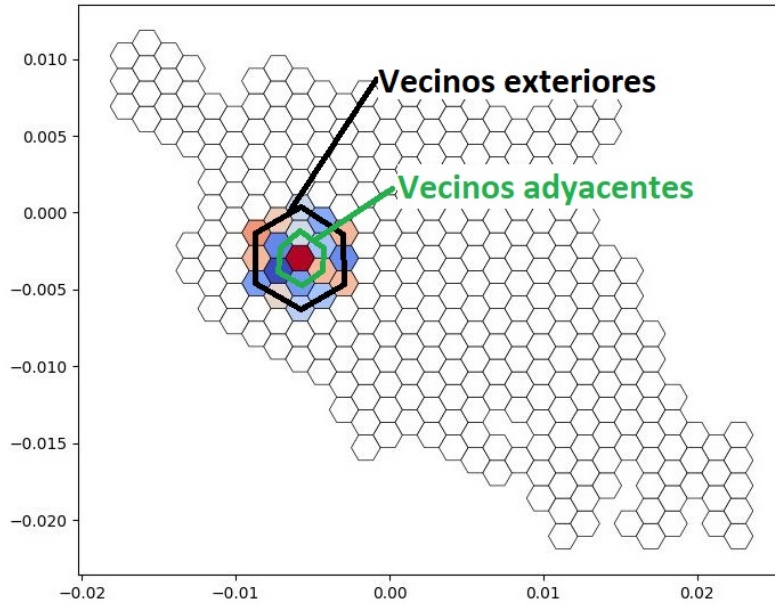


Figura 9: Ejemplo de vecinos adyacentes y vecinos exteriores

Al tomar los coeficientes correspondientes a los vecinos adyacentes en todos los modelos, se debe obtener una tendencia a tener más influencia sobre la variable respuesta que los vecinos exteriores. Eso se comprobó por medio de una prueba de hipótesis usando el método de Mann-Whitney (Dodge, 2008) donde se evaluó la hipótesis en la ecuación 9 usando un nivel de significancia de 0.05.

$$H_0 : S_{ady} = S_{ext} \quad (9)$$

$$H_a : S_{ady} \neq S_{ext}$$

Donde S_{ady} es la muestra de todos los coeficientes correspondientes a un vecino adyacente en todos los modelos y, de forma similar, S_{ext} es la muestra de los coeficientes para los vecinos exteriores.

Se obtuvo un valor P de $8,2926 * 10^{-17}$ rechazando la hipótesis nula, donde se evaluaba la igualdad en las muestras. De esta forma, se concluye que sí hay una diferencia entre ambas distribuciones.

Por último, se realizó un análisis descriptivo de los histogramas de ambas muestras, elimi-

nando el 10 % de los valores superiores e inferiores de cada serie por separado. Los resultados se ven en la figura 10 donde se aprecia un sesgo por parte de los vecinos adyacentes hacia valores más positivos, mientras que los vecinos exteriores tienden a estar centrados en 0.

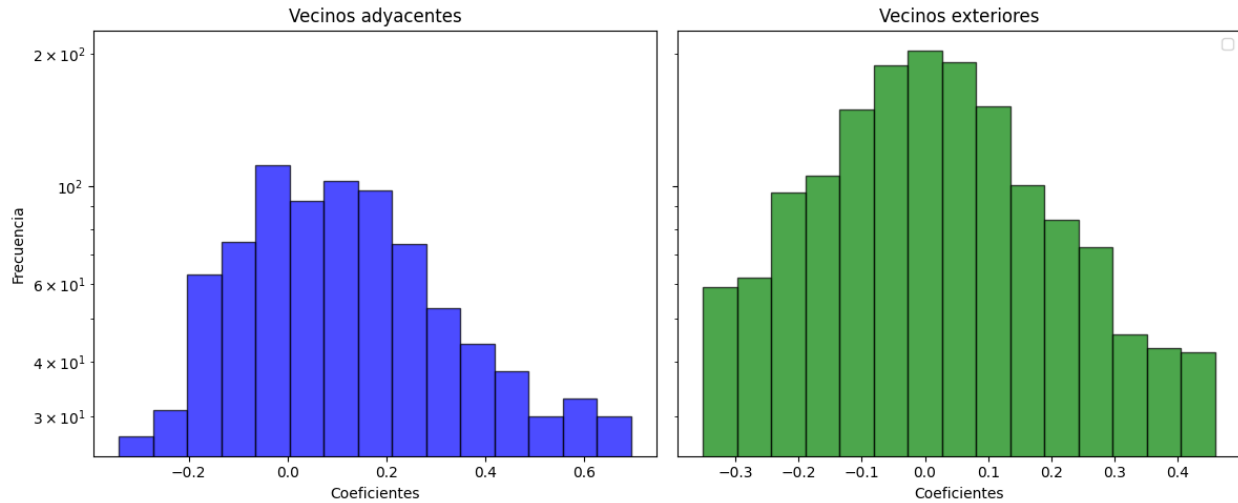


Figura 10: Histograma de coeficientes correspondientes a vecinos adyacentes y externos

9. Discusión

9.1. Comparación con interpolación geoespacial usando método de Kriging

Durante el desarrollo del proyecto, se tomaron los puntos centroides de cada hexágono con su población para plantear la interpolación de Kriging. Este método admite únicamente analizar una ventana de tiempo a la vez, así que se iteró a lo largo de las 34 ventanas para obtener el desarrollo de la plaga a lo largo del tiempo. Particularmente, en las últimas seis ventanas (lo que comprende un período desde el 16 de julio de 2024 hasta el 1 de octubre de 2024) se puede apreciar el desarrollo de varios focos de la peste, como se puede apreciar en la figura 11.

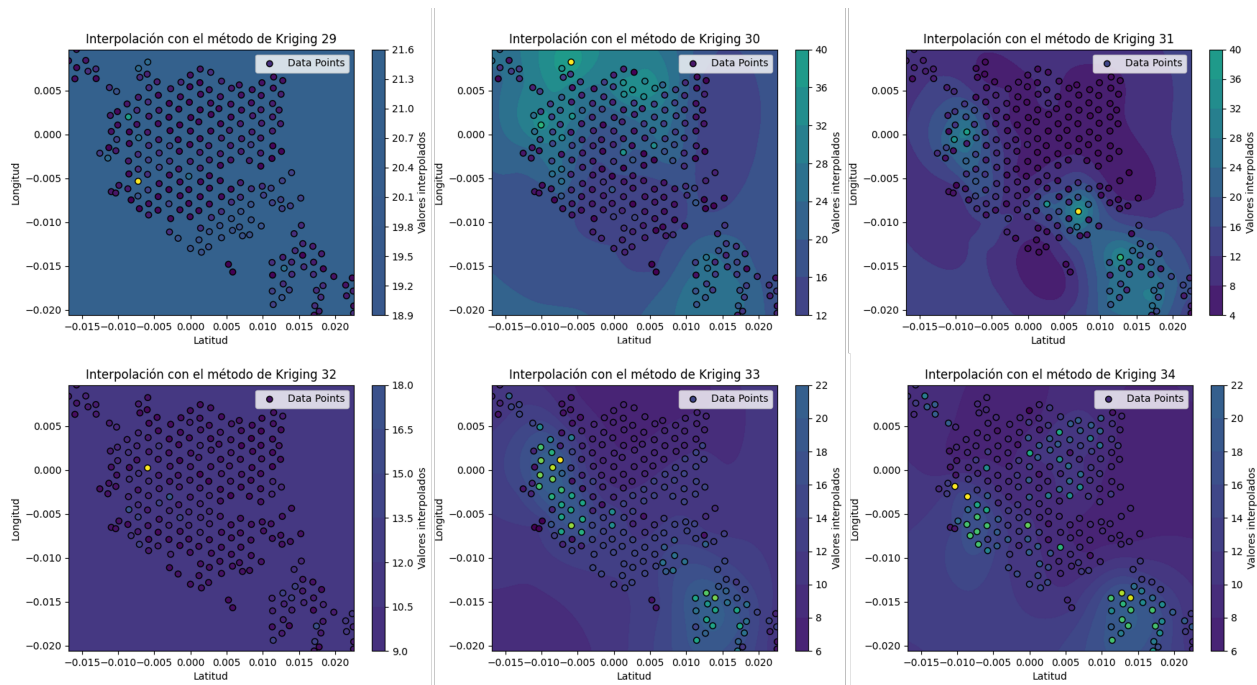


Figura 11: Serie de interpolaciones con método de Kriging

Este método únicamente tiene en cuenta las mediciones geospaciales en un instante de tiempo, lo que es inconveniente en sistemas poblacionales donde hay fenómenos de migración, ciclos reproductivos e intervención humana. Esto representa una desventaja sobre el método propuesto, donde también se considera la variable temporal e influencia histórica de los vecinos sobre la variable respuesta. Además, al ser una interpolación, tiende a presentar problemas de sobreajuste cuando la cantidad de mediciones es elevada.

9.2. Partición de los datos

Un proceso tradicional de machine learning debería requerir partir los datos en segmentos de entrenamiento y testeo; paso que se omitió en el proyecto. A pesar de que esto es importante para detectar sobreentrenamiento, el dataset disponible solo contaba con 34 registros; esto implica que segmentar los datos generaría dos conjuntos estadísticamente insignificantes. En otras palabras, suponiendo que se separan los datos en una proporción 80-20 de tal forma que se obtengan veintisiete datos para entrenamiento y siete datos para testeo de forma aleatoria. Calcular una métrica

de testeo con tan solo siete datos elegidos aleatoriamente es sensible a outliers y puede arrojar un resultado irreal del desempeño del modelo. Alternativamente, se podría usar una proporción menor de entrenamiento, suponer por ejemplo, 60-40; de esta forma se tendrían veinte datos para entrenamiento y catorce para testeo. Este caso es todavía menos preferible que el anterior dado que al entrenar el modelo con tan pocos datos omitirá información relevante del sistema. De esta forma, particionar los datos no genera un análisis más confiable que simplemente entrenar el modelo con la totalidad de los datos y esperar datos futuros para hacer un análisis OOT.

10. Conclusión

Se logró modelar el desarrollo poblacional de la peste leptopharsa gibbicarina utilizando una metodología de teselado hexagonal para la modelación individual de áreas iguales y adyacentes de la plantación. La implementación de modelos lineales logró un ajuste similar a algunos modelos encontrados en el estado del arte, medido con coeficiente de determinación superior a 0.8.

Se superaron limitantes referentes a la cantidad de datos utilizando modelos tradicionales de ML y restringiendo la cantidad de variables. A pesar de que los datos no lograron ajustarse a los modelos teóricos, se propuso un modelo alternativo que se ajustó suficientemente bien. En futuros trabajos se podría evaluar la posibilidad de usar técnicas de reducción de dimensionalidad como lo es PCA en lugar de arbitrariamente elegir un criterio de cercanía.

Se demostró la influencia positiva que tienden a tener los hexágonos aledaños sobre la población de un hexágono cualquiera por encima de la influencia de los hexágonos exteriores. Esto demuestra que dos hexágonos con la misma población pueden no verse afectados de la misma forma y, por ende, deben ser priorizados por el agricultor teniendo en cuenta sus poblaciones vecinas y no solo su población per se. Este enfoque implementando influencia histórica en lugar de un instante de tiempo sugiere una alternativa más robusta a métodos tradicionales como interpolación de Kriging.

11. Anexos

11.1. Repositorio en GitHub

El repositorio se puede encontrar en el link "https://github.com/Ales9914/MCDA_Final_project". Si bien la plantación accedió a compartir los datos con la comunidad de la universidad EAFIT o cualquier miembro académico, actualmente se encuentran en una base de datos relacional en el servicio de Google Cloud Plataform: BigQuery. Para acceder, se requiere una llave de acceso la cual debe ser solicitada al correo asalazarh@eafit.edu.co. Con la llave, puede consultar la base por medio del cliente para Python («Python client library — Google Cloud — cloud.google.com», s.f.).

Referencias

- Aldana de La Torre, R. C. (2010). *Manual de plagas de la palma de aceite en Colombia*. Cenipalma.
- Barrios, C., & Bustillo, A. E. (2014). Biología de la chinche de encaje leptopharsa gibbicarina y su control con hongos entomopatógenos. *Ceniavances*, (180).
- Barrios T., C. E., Bustillo P., A. E., Ocampo R., K. L., Reina C., M. A., & Alvarado M., H. L. (2016). Eficacia de hongos entomopatógenos en el control de Leptopharsa gibbicarina (Hemiptera: Tingidae) en palma de aceite. *Revista Colombiana de Entomología*, 42(1). <https://doi.org/10.25100/socolen.v42i1.6665>
- Barrios-Trilleras, C. E., Cuchimba-Triana, M. S., & Bustillo-Pardey, A. E. (2015). Parámetros poblacionales de Leptopharsa gibbicarina (Hemiptera: Tingidae) plaga de la palma de aceite. *Revista Colombiana de Entomología*, 41(1).
- Boniecki, P., Koszela, K., Piekarska-Boniecka, H., Weres, J., Zaborowicz, M., Kujawa, S., Majewski, A., & Raba, B. (2015). Neural identification of selected apple pests. *Computers and Electronics in Agriculture*, 110, 9-16. <https://doi.org/10.1016/j.compag.2014.09.013>

- Buffoni, G., & Pasquali, S. (2007). Structured population dynamics: Continuous size and discontinuous stage structures. *Journal of Mathematical Biology*, 54(4), 555-595. <https://doi.org/10.1007/s00285-006-0058-2>
- Bustillo, A. (2014). Manejo de insectos-plaga de la palma de aceite con énfasis en el control biológico y su relación con el cambio climático. *Revista Palmas*, 35(4).
- Cybertracker. (s.f.). Cybertracker website [Accessed on 2/12/2024]. <https://cybertracker.org/>
- Dodge, Y. (2008). Mann–Whitney Test. En *The Concise Encyclopedia of Statistics* (pp. 327-329). Springer New York. https://doi.org/10.1007/978-0-387-32833-1_243
- Espriella, A. G. (1988). Intervencion del director ejecutivo de fedepalma en el acto de clausura del xv congreso nacional de cultivadores de palma africana. *Congreso nacional de cultivadores de palma africana*, P33-34.
- Goel, N. S., Maitra, S. C., & Montroll, E. W. (1971). On the volterra and other nonlinear models of interacting populations. *Reviews of Modern Physics*, 43(2). <https://doi.org/10.1103/RevModPhys.43.231>
- Ibrahim, E. A., Salifu, D., Mwalili, S., Dubois, T., Collins, R., & Tonnang, H. E. Z. (2022). An expert system for insect pest population dynamics prediction. *Computers and Electronics in Agriculture*, 198. <https://doi.org/10.1016/j.compag.2022.107124>
- Lee, K. Y., Chung, N., & Hwang, S. (2016). Application of an artificial neural network (ANN) model for predicting mosquito abundances in urban areas. *Ecological Informatics*, 36, 172-180. <https://doi.org/10.1016/j.ecoinf.2015.08.011>
- Lotka, A. J. (1925). *Elements of physical biology*. Williams & Wilkins Company.
- Malthus, T. R. (1798). *An essay on the principle of population*. J. Johnson.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *The BMJ*, 372. <https://doi.org/10.1136/bmj.n71>

- Petrovskii, S., Bearup, D., Ahmed, D. A., & Blackshaw, R. (2012). Estimating insect population density from trap counts. *Ecological Complexity*, *10*, 69-82. <https://doi.org/10.1016/j.ecocom.2011.10.002>
- Python client library — Google Cloud — cloud.google.com [[Accessed 25-02-2025]]. (s.f.).
- Ricciardi, L. M. (1986). Stochastic Population Theory: Diffusion Processes. En T. G. Hallam & S. A. Levin (Eds.), *Mathematical Ecology: An Introduction* (pp. 191-238). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-69888-0_9
- Simamora, M., Rachma, N., & Sugiyono, S. (2018). Science and Technology Network in the Innovation System of Up-Stream Oil Palm Industry in Indonesia. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2729226>
- Tsoularis, A., & Wallace, J. (2002). Analysis of logistic growth models. *Mathematical Biosciences*, *179*(1), 21-55. [https://doi.org/10.1016/S0025-5564\(02\)00096-2](https://doi.org/10.1016/S0025-5564(02)00096-2)
- Vennila, S., Singh, G., Jha, G. K., Rao, M. S., Panwar, H., & Hegde, M. (2017). Artificial neural network techniques for predicting severity of *Spodoptera litura* (Fabricius) on groundnut. *Journal of Environmental Biology*, *38*(3), 449-456. <https://doi.org/10.22438/jeb/38/3/MS-163>