



**UNIVERSIDAD DE PANAMÁ
FACULTAD DE CIENCIAS NATURALES, EXACTAS Y TECNOLOGÍA
ESCUELA DE BIOLOGÍA
DEPARTAMENTO DE CIENCIAS AMBIENTALES**

**MODELIZACIÓN DE LA DINÁMICA TEMPORAL DE *Euglossa imperialis*
(HYMENOPTERA: EUGLOSSINI) EN UN BOSQUE TROPICAL MEDIANTE
SERIES DE TIEMPO, ESTADÍSTICA CIRCULAR Y TÉCNICAS DE
APRENDIZAJE AUTOMÁTICO**

ELABORADO POR:

**YOSTIN JESÚS AÑINO RAMOS
2-730-1081**

**EMILIO ROMERO-ROMERO
8-835-1739**

Tesis presentada como uno
de los requisitos para optar
por la Licenciatura en
Biología con orientación en
Biología Ambiental

**PANAMÁ, REPÚBLICA DE PANAMÁ
2025**

**UNIVERSIDAD DE PANAMÁ
FACULTAD DE CIENCIAS NATURALES, EXACTAS Y TECNOLOGÍA
ESCUELA DE BIOLOGÍA
TRABAJO DE GRADUACIÓN**

Ciudad Universitaria Octavio Méndez Pereira, 14 de mayo de 2025

Programa: Licenciatura en Biología con orientación en Biología Ambiental

Línea de Investigación: Ecología, Biología Ambiental

Título: Modelización de la dinámica temporal de *Euglossa imperialis* (Hymenoptera: Euglossini) en un bosque tropical mediante series de tiempo, estadística circular y técnicas de aprendizaje automático

El trabajo fue elaborado por los estudiantes:

Yostin Jesús Añino
2-730-1081

Emilio Romero-Romero
8-835-1739

Dr. Enrique Medianero Segundo
Asesor Principal

Magister Clara Cruz
Co-asesor

Magister
Julio Trujillo
Co-asesor

DEDICATORIA Y AGRADECIMIENTO

Dedicatoria

Dedico esta investigación a Delia y Emilio, por su dedicación y apoyo incondicional en los momentos más importantes de mi vida. Gracias por todo lo que han invertido en mi formación, siempre guiados por el respeto y la sencillez.

A Sophia, por su comprensión infinita y esa caja llena de sonrisas que me motivaron a seguir adelante incluso en los días más difíciles.

Y a Xiomara por su paciencia, comprensión y apoyo inquebrantable. Gracias por caminar a mi lado en este recorrido.

Emilio Romero

Dedico este trabajo a mis padres, Sonia Ramos de Añino y Martin Eduardo Añino, por brindarme la oportunidad de estudiar, así como a mi abuela Elisa Gordon por ¡apoyarme siempre en todo!

Este trabajo también es una dedicatoria de una forma muy especial a Karla Jeannette Agrazal y Víctor Manuel Añino Ramos, por todos los momentos agradables que compartieron conmigo en laboratorio amenizando las largas jornadas de trabajo.

Yostin Jesús Añino Ramos

Agradecimiento

Primeramente, deseo expresar mi más sincero agradecimiento a nuestros asesores, los profesores Clara Cruz y Enrique Medianero, por proporcionarnos una visión crítica que permitió mejorar todos los aspectos del proyecto. Agradezco profundamente las horas que dedicaron, así como la atención cuidadosa que brindaron a cada uno de sus comentarios.

A mi compañero Yostin, por su constante disposición y apoyo a lo largo de toda la carrera. Siempre estuvo abierto al diálogo, a compartir ideas y a ofrecer su ayuda para el desarrollo de esta investigación.

A los docentes del Departamento de Fisiología y Comportamiento Animal, por su respaldo constante como unidad académica. Al profesor Roberto Cambra Torok, del Museo de Invertebrados de la Universidad de Panamá, por su generosidad al ofrecerme café y té durante mis largas jornadas de trabajo. A la Universidad de Panamá, por brindarme la oportunidad institucional de realizar este proyecto.

Agradezco a Xiomara Medina por su ánimo incondicional y su apoyo, ya sea ayudándome con diligencias, dándome un consejo, o simplemente ofreciéndome una nueva perspectiva ante las diversas situaciones. Su apoyo ha sido firme y constante, y ha sido un hombro en el que siempre pude apoyarme. Por último, a mis padres, Delia y Emilio, quienes, sin lugar a duda, me brindaron el mayor regalo: su tiempo. Mientras yo trabajaba, ellos cuidaban de mi hija, regalándole el tiempo que yo no podía darle. Ese gesto jamás dejará de conmoverme y agradecerlo.

Emilio Romero

Quiero agradecer a mis asesores Enrique Medianero, Clara Cruz, Julio Trujillo y Javier Sanchez Galan, así como a mi compañero Emilio Romero, por su disposición colaborativa para el desarrollo de esta tesis. Agradezco a mi familia y amigos, quienes han estado en diferentes momentos importantes de mi vida y por ende en decisiones decisivas. Agradezco de forma personal a Emilio Romero, Randhy Rodríguez y Roberto Cambra, por todos sus comentarios, opiniones y consejos en distintos tópicos. Finalmente quiero agradecer a todos aquellos colaboradores que alguna vez han depositado la confianza en mí.

Yostin Jesús Añino Ramos

Agradecemos también a la convocatoria de movilidad internacional contrato

DDCCT 02-2025 de SENACYT

ÍNDICES

Índice General y de Figuras

General

Dedicatoria	IV
Agradecimiento.....	V
Índice General y de Figuras	VII
Resumen y Abstract	X
Palabras claves:	X
Introducción	1
Marco Teorico.....	4
Objetivo General.....	11
Objetivos Específicos:	11
Hipótesis de Trabajo.....	11
Materiales y Métodos	12
Sitio de estudio	13
Bases de datos	13
Análisis de datos	15
Resultados y Discusión	22
Serie temporal de abundancia de <i>Euglossa imperialis</i> (Subset)	23
Serie temporal de abundancia de <i>Euglossa imperialis</i> (Base completa).....	27
Efecto de variables ambientales sobre la abundancia temporal de <i>Euglossa imperialis</i> en un bosque tropical- Análisis de series temporales mediante Machine Learning	27
Conclusiones	39
Referencias bibliográficas	41

Índice de Tablas

Tabla 1. Métricas de evaluación de modelos predictivos para series temporales.....	23
---	----

Tabla 2. Resultados de los Modelos de Machine Learning empleados para el análisis de serie temporal de <i>Euglossa imperialis</i> y factores ambientales.	35
---	----

Índice de Figuras

Figura 1. Variación en la abundancia anual de machos de <i>Euglossini</i> en un bosque tropical panameño.....	8
Figura 2. Individuos de <i>Euglossa imperialis</i> colectados de 1979 hasta 2018 en Pipeline Road (Base de datos).....	14
Figura 3. Serie de la variable abundancia de <i>Euglossa imperialis</i> (Original)	18
Figura 4. Serie de la variable abundancia de <i>Euglossa imperialis</i> (Completada)	19
Figura 5. Series de abundancia de <i>Euglossa imperialis</i> y pluviosidad en la zona completas y formateadas.	19
Figura 6. Modelo Lineal de serie temporal de <i>Euglossa imperialis</i>	24
Figura 7. Serie temporal de abundancia de <i>Euglossa imperialis</i> de 1980 a 1986.....	25
Figura 8. Salida del Modelo Autorregresivo de ARIMA en R.....	26
Figura 9. Variantes lineales y circulares de abundancia en la población de <i>Euglossa imperialis</i> durante los distintos años de estudio.	26
Figura 10. Serie temporal anual de la abundancia de <i>Euglossa imperialis</i> a lo largo de 40 años de estudio en un bosque tropical.	27
Figura 11. Curvas de entrenamiento y Prueba del Modelo 1 con redes LSTM.....	29
Figura 12. Resultados en la predicción de los últimos 8 años de la serie con el Modelo 1 con redes LSTM. En azul la serie real de abundancia, en naranja la serie de entrenamiento y en verde la curva de prueba predicha por el modelo.	29
Figura 13. Curvas de entrenamiento y prueba del Modelo 2 con redes LSTM-Bidireccionales.	30
Figura 14. Resultados en la predicción de los últimos 8 años de la serie con el modelo 2 con redes LSTM-Bidireccional. En azul la serie real de abundancia, en naranja la serie de entrenamiento y en verde la curva de prueba predicha por el modelo.	31
Figura 15. Modelo de predicción de abundancia mediante métodos de Random Forest.	32
Figura 16. Histogramas para nuevas variables meteorológicas.....	33
Figura 17. Curvas de entrenamiento y Prueba del Modelo 3 con redes LSTM, con más características meteorológicas.	34
Figura 18. Resultados en la predicción de un poco menos de los últimos 8 años de la serie con el modelo 4 con redes LSTM y los datos meteorológicos ampliados.	34

RESUMEN Y ABSTRACT

RESUMEN

El declive de las poblaciones de insectos, incluidos los polinizadores como las abejas, ha sido un tema central en la investigación ecológica debido a la creciente preocupación por los efectos del cambio climático y las perturbaciones antrópicas. En particular, las abejas de las orquídeas (Apidae: Euglossini), esenciales para la polinización en los ecosistemas tropicales, han sido objeto de estudios de monitoreo poblacional a lo largo de varias décadas. Sin embargo, en muchos casos se carece de modelos de series temporales que permitan analizar de manera detallada las dinámicas poblacionales de estas especies. Este estudio se centra en el análisis de la especie *Euglossa imperialis* en un bosque tropical panameño, utilizando modelos de series temporales, estadísticas circulares y técnicas de machine learning para evaluar su estacionalidad y tendencia poblacional durante 40 años de muestreos casi ininterrumpidos. Los resultados revelan una tendencia decreciente en la población, aunque con variabilidad interanual significativa influenciada por factores climáticos y la disponibilidad de recursos florales. El análisis de series temporales mediante la Media Móvil Centrada ($K=3$) y el modelo SARIMA mostró fluctuaciones estacionales y un declive moderado en la abundancia, mientras que el modelo de aprendizaje automático LSTM, incorporando variables ambientales como temperatura, pluviosidad y otros factores, permitió una predicción más precisa de la abundancia de la especie. Este estudio resalta la importancia de considerar múltiples enfoques y técnicas estadísticas para comprender la dinámica poblacional de especies clave como *Euglossa imperialis*, cuya estabilidad es esencial para la conservación de los ecosistemas tropicales. Los hallazgos sugieren que factores como eventos climáticos extremos y variables ambientales juegan un papel fundamental en la abundancia de las abejas, lo que subraya la necesidad de integrar estos elementos en estrategias de conservación.

Palabras Claves: Biología Ambiental, Biometeorología, Comportamiento estacional anual, Series de tiempo, Estadística avanzada.

ABSTRACT

The decline of insect populations, including pollinators such as bees, has been a central topic in ecological research due to growing concerns about the effects of climate change and anthropogenic disturbances. Orchid bees (Apidae: Euglossini), essential for pollination in tropical ecosystems, have been the subject of population monitoring studies over several decades. However, in many cases, there is a lack of time series models that allow for a detailed analysis of the population dynamics of these species. This study focuses on the analysis of *Euglossa imperialis* in a Panamanian tropical forest, using time series models, circular statistics, and machine learning techniques to assess its seasonality and population trends over 40 years of nearly continuous sampling. The results reveal a declining population trend, although with significant interannual variability influenced by climatic factors and the availability of floral resources. Time series analysis using Moving Average (K=3) and the SARIMA model showed seasonal fluctuations and a moderate decline in abundance, while the LSTM machine learning model, incorporating environmental variables such as temperature, rainfall, and other factors, allowed for more accurate predictions of the species' abundance. This study highlights the importance of considering multiple approaches and statistical techniques to understand the population dynamics of key species like *Euglossa imperialis*, whose stability is essential for the conservation of tropical ecosystems. The findings suggest that factors such as extreme climate events and environmental variables play a fundamental role in bee abundance, underscoring the need to integrate these elements into conservation strategies.

Keywords: Environmental Biology, Biometeorology, Annual Seasonal Behavior, Time Series, Advanced Statistics.

Capítulo I:

INTRODUCCIÓN

INTRODUCCIÓN

Los insectos, con su inmensa diversidad y vital importancia en los ecosistemas, han sido parte integral de nuestro planeta durante millones de años. En las últimas décadas, hemos sido testigos de una creciente producción de artículos relacionados con el declive en las poblaciones de insectos en todo el mundo (Wagner, 2020). Muchos de estos estudios buscan relacionar las cambiantes condiciones ambientales y fenómenos como el cambio climático, así como el efecto antrópico que presentan los ecosistemas (Dirzo *et al.*, 2014; Sánchez-Bayo & Wyckhuys, 2019). No obstante, pese a existir mucha literatura sobre este tema, no se puede generalizar el efecto de declive para todos los grupos de insectos (Wagner *et al.*, 2021), por ejemplo, Van Klink *et al.* (2020) encontraron mediante un metaanálisis que, si bien existía un declive en las poblaciones de insectos terrestres, en grupos de insectos acuáticos había un aumento en la abundancia.

En el caso de abejas generalmente se sugiere una disminución en las poblaciones asociada principalmente a perturbaciones antropogénicas (Biesmeijer *et al.*, 2006; Kosior *et al.*, 2007; Colla & Packer, 2008; Frankie *et al.*, 2009; Winfree *et al.*, 2009; Brown & Paxton, 2009; Potts *et al.*, 2010; Ollerton *et al.*, 2011; Burkle *et al.*, 2013; Senapathi *et al.*, 2015; Goulson *et al.*, 2015; Koh *et al.*, 2016; Ollerton, 2017) por lo que zonas protegidas o áreas conservadas usualmente presentan poblaciones relativamente estables (Roubik & Wolda, 2001; Roubik, 2001; Roubik & Villanueva-Gutiérrez, 2009; Archer, 2013; Herrera, 2019).

En Euglossini, la cual es una tribu de abejas neotropicales existen pocos estudios de monitoreos poblacionales extensos, entre los más citados podemos mencionar los monitoreos de Nemésio *et al.* (2015) en un remanente urbano de bosque en Brasil por 14 años y algunos realizados en la parte central de Panamá con muestreos de 21 (Roubik, 2001) y 40 años (Roubik *et al.*, 2021). De forma resumida podemos mencionar que la mayoría de estos estudios sugieren que las poblaciones de Euglossini se mantienen

estables. Sin embargo, también hay fuentes que indican declives en ciertas poblaciones, tal es el caso de Vega-Hidalgo et al. (2020) quienes documentan un declive en la abundancia de individuos de abejas de las orquídeas del Parque Nacional Darién durante 5 años de muestreo (2013-2017).

No cabe duda de que para la debida y correcta interpretación sobre el estatus que presentan las poblaciones de los distintos grupos de insectos, se deben cumplir ciertas condiciones. Una de ellas es que se tenga registro de monitoreos extensos y que estos, en la medida de lo posible, se realicen de forma continua. También es importante el tipo de análisis de datos que se lleva a cabo, usualmente se sugieren modelos de series de tiempo (Ramírez *et al.*, 2015), considerando que evalúan cómo una variable respuesta (en este caso abundancia de individuos) varía a través del tiempo, utilizando diferentes propuestas de modelos que no solo incluyen un modelo lineal. A su vez, se ha propuesto el uso de enfoques complementarios como la estadística circular, especialmente útil cuando los patrones de actividad o captura siguen distribuciones cíclicas, por ejemplo, patrones diarios o estacionales (Fisher, 1995; Jammalamadaka & SenGupta, 2001). Este enfoque ha demostrado ser valioso en estudios de interacción planta-polinizador, como en el caso de ciertas especies de orquídeas polinizadas por abejas euglosinas (Benitez-Vieyra & Aizen, 2003).

Por otro lado, el aprendizaje automático ha emergido como una herramienta poderosa en la ecología moderna. Métodos como los árboles de decisión, los bosques aleatorios y las redes neuronales artificiales permiten detectar patrones complejos en grandes volúmenes de datos, lo cual es especialmente útil en estudios de biodiversidad y monitoreo de insectos (Cutler *et al.*, 2007; Olden *et al.*, 2008; Thessen, 2016). Estos enfoques pueden complementar los análisis tradicionales, mejorando la precisión y la capacidad predictiva al

estudiar los factores que determinan la abundancia o distribución de grupos como las abejas de las orquídeas.

El monitoreo continuo de poblaciones de Euglossini más extenso en la actualidad (Roubik *et al.*, 2021), pese a documentar el comportamiento fluctuante de múltiples especies de Euglossini, no emplea modelos de series temporales, los cuales podrían ayudar a mejorar la comprensión de la dinámica poblacional de algunas especies, como es el caso de *Euglossa imperialis*, la cual puede ser considerada una de las especies de abejas de las orquídeas más comunes en las matrices paisajísticas de bosques panameños (Com. Pers. Roubik, 2024). Entre los aportes que brindarían las series temporales están los análisis de la estacionalidad y tendencia que podrían presentar ciertas especies. El estudio de la estacionalidad también puede reforzarse con otras técnicas, como la aplicación de técnicas estadísticas circulares (Agostinelli *et al.*, 2017; Batschelet, 1981).

Considerando la importancia que poseen las abejas de las orquídeas en la dinámica ecosistémica de distintas matrices paisajísticas, comprender cómo varían sus poblaciones y si las mismas presentan alteraciones por factores externos como el cambio climático o efectos antrópicos es un tema fundamental para la biología de la conservación y ciencias ambientales. En este estudio se modeliza la dinámica temporal de una especie de Euglossini, siendo esta *Euglossa imperialis* mediante la aplicación de series de tiempo (estadísticas y de aprendizaje automático) y técnicas de estadística circular utilizando los datos de acceso abierto publicados por Roubik *et al.* (2021), con el fin de explorar y profundizar en aspectos relacionados a la estacionalidad y tendencia que presentaron estas poblaciones en un bosque tropical panameño durante 40 años de muestreos casi ininterrumpidos.

MARCO TEORICO

Abejas de las orquídeas (Hymenoptera: Apidae: Euglossini)

Las abejas de la tribu Euglossini, pertenecientes a la familia Apidae, son un grupo de himenópteros ampliamente distribuidos en el Neotrópico y conocidos por sus colores metálicos brillantes (Ackerman, 1983). Esta tribu comprende aproximadamente 250 especies agrupadas en cinco géneros: *Aglae*, *Eufriesea*, *Euglossa*, *Eulaema* y *Exaerete*. De estos, *Euglossa* es el más diverso, con especies de variados tamaños y una coloración típica de la tribu, pero que también incluye tonalidades no metálicas. Morfológicamente, las especies de *Euglossa* se distinguen por la presencia de un labro emblanquecido con marcas ovales, mandíbulas con dos o tres dientes y mesotibias de los machos con lunares afelpados (Arcos, 2020). En cuanto a su distribución, estas abejas pueden encontrarse desde el nivel del mar hasta los 2900 msnm. Desde una perspectiva ecológica, las abejas Euglossini desempeñan un papel fundamental en la polinización de más de 600 especies de orquídeas, de donde proviene su nombre común de "abejas de las orquídeas" (Ackerman, 1989). Las orquídeas emiten compuestos aromáticos específicos, como terpenos y fenoles volátiles, que atraen a los machos, quienes recolectan estas fragancias, mientras que las hembras visitan las flores en busca de polen y néctar (Roubik & Hanson, 2004).

Series temporales basada en modelos, estadística circular y aprendizaje automático

El análisis y modelado de series temporales es una herramienta clave en estadística y aprendizaje automático, con aplicaciones en diversas disciplinas, incluida la biología ambiental y la ecología de poblaciones. Modelos como ARIMA (Modelo Autoregresivo Integrado de Media Móvil), desarrollado por Box y Jenkins (1976), permiten analizar patrones en datos temporales, como fluctuaciones en poblaciones de especies, cambios en

parámetros ambientales y tendencias en eventos climáticos extremos. Su diferenciación integrada facilita la conversión de series no estacionarias en estacionarias, lo que es crucial para analizar procesos ecológicos sujetos a variabilidad natural. Sin embargo, cuando los datos presentan estacionalidad, SARIMA introduce un componente adicional que permite modelar ciclos anuales en dinámicas poblacionales o variaciones en la calidad del hábitat (Hyndman & Athanasopoulos, 2018). Alternativamente, los modelos de suavizado exponencial, como el desarrollado por Holt (1957) y extendido por Winters (1960), ofrecen una opción flexible para capturar tendencias y patrones recurrentes en datos ambientales, como la temperatura del agua, la cobertura vegetal o las tasas de reproducción de especies. La elección del modelo adecuado depende de la naturaleza de los datos y su estructura temporal, lo que requiere un análisis exhaustivo (Brockwell & Davis, 2016).

El avance del aprendizaje automático ha permitido desarrollar enfoques más sofisticados para modelar series temporales en biología ambiental y ecología, como las redes neuronales recurrentes, que pueden captar relaciones no lineales en la dinámica de poblaciones o en la respuesta de ecosistemas a perturbaciones (Goodfellow, Bengio, & Courville, 2016). Estas técnicas han demostrado utilidad en la predicción de fenómenos como la migración de especies en función del cambio climático, la evolución de brotes de enfermedades en poblaciones silvestres y la respuesta de los ecosistemas a eventos extremos. Sin embargo, su implementación requiere grandes volúmenes de datos y capacidad computacional. A pesar de la creciente sofisticación de estos métodos, la comprensión profunda de los datos y la selección de modelos apropiados sigue siendo esencial para garantizar la precisión y aplicabilidad de los resultados en el manejo de recursos naturales y la conservación de la biodiversidad (Makridakis, Spiliotis, & Assimakopoulos, 2018).

El análisis de series temporales en insectos, particularmente en abejas, ha permitido identificar patrones estacionales, circadianos y multianuales en su comportamiento y ecología. En este contexto, la estadística circular ofrece herramientas potentes para describir fenómenos cíclicos, como los ritmos de actividad diaria o las fechas de aparición estacional, aspectos fundamentales en insectos polinizadores como las abejas (Batschelet, 1981). Por ejemplo, la estadística circular ha sido aplicada para analizar la orientación de vuelos, las horas pico de actividad forrajera y la sincronización con la fenología floral (Fischer *et al.*, 2021). En el caso de abejas tropicales, donde la estacionalidad puede no ser marcada por temperaturas sino por lluvias, las variables circulares (como horas del día o días del año) permiten evaluar respuestas ecológicas sutiles a cambios climáticos, promoviendo un enfoque más robusto frente a la linealidad tradicional (Landim *et al.*, 2019).

Por otro lado, el aprendizaje automático ha emergido como una herramienta poderosa para modelar series temporales ecológicas complejas, permitiendo la detección de patrones no lineales y la predicción de eventos ecológicos en abejas. Algoritmos como Random Forest, redes neuronales recurrentes (RNN) y máquinas de soporte vectorial (SVM) han sido aplicados para predecir dinámicas poblacionales, abundancia relativa y distribución espacio-temporal de polinizadores en función de variables climáticas y antrópicas (Zhao *et al.*, 2020; Willcock *et al.*, 2018). La combinación de aprendizaje automático con datos derivados de sensores remotos y estaciones meteorológicas ha mejorado la resolución temporal y espacial de estos modelos, favoreciendo estrategias de conservación basadas en predicciones de cambios fenológicos y migratorios en comunidades de abejas (Goulson *et al.*, 2015). Integrar la estadística circular con técnicas de inteligencia artificial representa un enfoque prometedor para abordar la complejidad inherente a los sistemas ecológicos, especialmente en un contexto de cambio climático global.

Análisis estacional y temporal de poblaciones de Euglossini

El análisis de series temporales ha permitido comprender las dinámicas poblacionales de las abejas Euglossini a lo largo del tiempo, revelando variaciones estacionales y tendencias a largo plazo en diferentes hábitats. Estudios como el de Ramírez *et al.* (2015) han demostrado que la estructura de las comunidades de estas abejas fluctúa estacionalmente, influenciada por factores como la disponibilidad de recursos florales y las condiciones climáticas. De manera similar, Ackerman (1983, 1989) evidenció que la diversidad y la actividad de los machos de Euglossini varían a lo largo del año, con patrones de estacionalidad que reflejan cambios en la fenología de las plantas y la oferta de fragancias en el ambiente. A través de análisis temporales, se ha identificado que algunas especies presentan mayor abundancia en ciertas épocas del año (Fig. 1), lo que subraya la importancia de evaluar los cambios en la composición y diversidad de las poblaciones a lo largo del tiempo.

Los estudios de largo plazo han revelado tendencias preocupantes en las poblaciones de Euglossini, especialmente en áreas sometidas a alteraciones antropogénicas. Roubik *et al.* (2021), analizaron cuatro décadas de datos en Panamá y encontraron fluctuaciones en la abundancia de estas abejas dentro de áreas protegidas, lo que sugiere que factores como el cambio climático y la fragmentación del hábitat pueden estar afectando sus poblaciones. De manera similar, Vega-Hidalgo *et al.* (2020) documentaron un declive poblacional en un bosque tropical de Panamá, atribuyéndolo a la degradación del hábitat y la disminución de los recursos florales. Además, Nemésio *et al.* (2015) demostraron que en ambientes urbanos la dinámica poblacional de Euglossini se ve alterada por la fragmentación del paisaje y la reducción de la conectividad entre parches de vegetación. Estos hallazgos resaltan la necesidad de monitoreos continuos mediante análisis de series temporales para evaluar el impacto de las perturbaciones ambientales en la conservación de estas abejas.

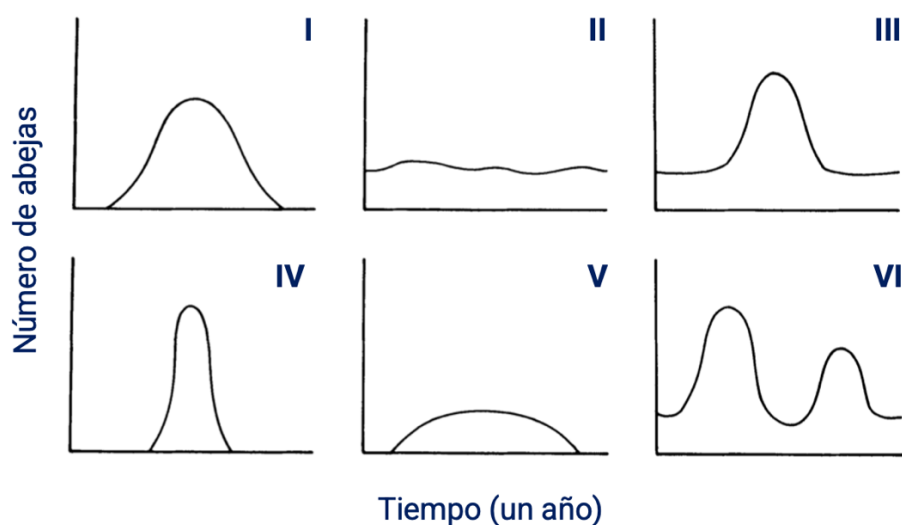


Figura 1. Variación en la abundancia anual de machos de Euglossini en un bosque tropical panameño. **Fuente:** Ackerman, 1983.

Importancia de los estudios de la dinamica temporal de Euglossini

El análisis de series temporales en las poblaciones de abejas Euglossini es fundamental para comprender las dinámicas ecológicas de estos polinizadores claves en los ecosistemas tropicales. Estas abejas desempeñan un papel crucial en la polinización de una amplia diversidad de plantas, incluidas orquídeas, por lo que cualquier fluctuación en sus poblaciones puede tener efectos significativos en la estabilidad y regeneración de los bosques tropicales (Ackerman, 1989; Ramírez *et al.*, 2015). Estudios de largo plazo han demostrado que las comunidades de Euglossini presentan patrones estacionales y tendencias a lo largo del tiempo, determinadas por factores como la disponibilidad de recursos florales, cambios climáticos y perturbaciones antropogénicas (Ackerman, 1983; Roubik *et al.*, 2021). Sin embargo, la falta de estudios sistemáticos que analicen estas variaciones mediante técnicas avanzadas de modelado temporal ha limitado la capacidad de prever cambios poblacionales y establecer estrategias de conservación efectivas.

En este contexto, la aplicación de modelos de series temporales permite evaluar de manera rigurosa las fluctuaciones en la abundancia y diversidad de *Euglossini* a lo largo del tiempo, proporcionando información clave sobre la estabilidad de sus poblaciones y las posibles amenazas que enfrentan. Investigaciones previas han documentado declives en sus poblaciones debido a la degradación del hábitat, el cambio climático y la fragmentación del paisaje, lo que subraya la necesidad de monitoreos continuos y análisis predictivos para orientar estrategias de conservación (Vega-Hidalgo *et al.*, 2020; Nemésio *et al.*, 2015). Al emplear herramientas estadísticas avanzadas, este estudio contribuirá a la identificación de patrones de variabilidad poblacional y permitirá diseñar modelos que ayuden a mitigar los efectos negativos de las alteraciones ambientales sobre estas abejas. Así, se generará un conocimiento esencial para la conservación de *Euglossini* y la preservación de la biodiversidad en los ecosistemas tropicales.

Capítulo 2:

OBJETIVOS E HIPÓTESIS

Objetivo General:

Evaluar la abundancia temporal de *Euglossa imperialis* (Hymenoptera: Euglossini) en un bosque tropical durante un periodo comprendido por 40 años de muestreo mediante series de tiempo, estadística circular y técnicas de aprendizaje automático.

Objetivos Específicos:

- Describir el comportamiento estacional y tendencial de *Euglossa imperialis* en un bosque tropical.
- Proponer un modelo de serie temporal optimizado con estadística circular y técnicas de aprendizaje automático adecuado para el análisis de la dinámica temporal de *Euglossa imperialis* en un bosque tropical.
- Determinar el grado de efecto de las estaciones (seca y lluviosa) y factores ambientales asociados sobre la abundancia temporal de *Euglossa imperialis* en un bosque tropical.

HIPÓTESIS DEL TRABAJO

La abundancia poblacional de *Euglossa imperialis* ha disminuido durante los últimos 40 años. Los patrones estacionales tienen una relación con variables ambientales como la temperatura y la pluviosidad. Mediante el uso combinado de modelos de series temporales (como ARIMA y Media Móvil Centrada), estadísticas circulares y redes neuronales LSTM, es posible detectar estas tendencias estacionalidades, permitiendo una mejor comprensión no solo de los patrones estacionales sino de la dinámica ecológica poblacional de estas abejas.

Capítulo 3:

MATERIALES Y MÉTODOS

Sitio de Estudio

Los datos provienen de muestreos realizados en Pipeline Road (Fig. 1S), un sitio ubicado dentro del Parque Nacional Soberanía en la provincia de Panamá. Pipeline Road, es un corredor de bosque tropical húmedo de gran relevancia para la investigación ecológica en Panamá. Originalmente construido durante la Segunda Guerra Mundial como una vía de acceso para un oleoducto, este sendero se ha transformado en un importante sitio para estudios sobre la flora y fauna tropical. El ecosistema está compuesto por un bosque tropical húmedo, con características bioclimáticas que incluyen precipitaciones anuales superiores a 2,500 mm y temperaturas promedio de 25-27°C. En esta área se encuentran diversas especies de árboles, arbustos y plantas epífitas que forman una compleja red de interacciones ecológicas. La estructura vertical del bosque permite la coexistencia de diversas especies a diferentes niveles, desde el dosel hasta el suelo forestal, contribuyendo a la complejidad del ecosistema (Chavarria *et al.*, 2007). Este corredor natural es clave para estudios sobre dinámica de comunidades vegetales, relaciones ecológicas entre especies y efectos del cambio climático en los bosques tropicales. Su accesibilidad y la riqueza florística de la zona lo convierten en un sitio privilegiado para investigaciones a largo plazo sobre los impactos de la deforestación y la conservación de los ecosistemas tropicales (Gentry, 1993).

Bases de datos

Muestreos de Euglossini Pipeline Road (1979-2021)

La base de datos consiste en un set de datos sobre el número de individuos por especies de abejas de las orquídeas (Euglossini) colectadas mensualmente en Pipeline Road durante cuarenta años (1979-2021) de forma casi ininterrumpida, dichos datos fueron analizados por Roubik *et al.* (2021), quienes compartieron la base de datos incluida en el material suplementario del artículo publicado en formato de acceso abierto. La base datos, contiene 21305 registros e incluye datos de la mayoría de los años que comprende el rango

de muestreo, sin embargo, no se registraron colectas en los años 1993 y 1995 y tampoco se incluyen datos para todos los meses durante los 40 años de colectas. Considerando que la base de datos solo presentaba los meses y años en los que se habían registrado colectas, fue necesario emplear la herramienta *Kutools* en Excel para agregar las celdas que contuvieran los meses y años faltantes de forma continua, para luego ordenar los datos mediante el uso de tablas dinámicas. Para la imputación de datos faltantes se seleccionó el periodo de tiempo con la menor cantidad de celdas vacías, siendo este el periodo que comprendía de enero de 1980 hasta diciembre de 1985. Este periodo de tiempo contenía menos del 10% de las celdas (meses) vacías, por lo que dichas celdas fueron rellenas promediando los valores correspondientes al mes del año anterior y el mes del año posterior a la celda faltante. Se aplicó esta misma metodología para imputación de datos faltantes de los meses de febrero (serie continua 1980-2018) y mayo (1980-2016) (Fig. 2).

ANOS	enero	febrero	marzo	abril	mayo	junio	julio	agosto	septiembre	octubre	noviembre	diciembre	Total
1979										164	175	165	504
1980	160	132	239	184	123	128	170	214	238	200	106	144	2038
1981	138	93	76	190	151	80	110	110	51	117	69	43	1228
1982	100	100	129	154	162	161	272	170	86	102	42	140	1618
1983	242	156	242	156	208	122	222	115	123	78	38	156	1357
1984	194	117	105	124	175	106	171	252	123	119	111	171	1345
1985	146	64	103	177	139	91	254	166	160	212	66	117	1578
1986		65			95				140				300
1987		73			130			86		44		76	409
1988	90	58	231	126	132				95	72	63	150	1017
1989	184	72		138	115	189	130	120	21	101	35		990
1990		41	91	221	100	106	140					51	750
1991	87	78		143	146	200				110			764
1992		118		178	231	137							664
1993		86			196								
1994	59	53			196								112
1995		78			196								
1996		78		160	196								160
1997		78			160								160
1998		104			82	49			21				256
1999		63			187								250
2000		43			145								188
2001		41			70								111
2002		72			121								193
2003		63			140				94				297
2004		63			129				86				278
2005		30	85		52	81			117				365
2006		63			95								158
2007		52		151	124	56		97	130		58	73	617
2008		40			152	104	141				19		456
2009		36		87	70	136		50	95	53	10	58	595
2010	96	40	113		57	31	60						357
2011		45			60	79			63				187
2012		46			64								110
2013		51			121				51				223
2014		20			100	99							119
2015		13			81	63			86				243
2016		57			115				81				253
2017		64					77						141
2018		22		169					85				276

Figura 2. Individuos de *Euglossa imperialis* colectados de 1979 hasta 2018 en Pipeline Road (Base de datos). En celeste se muestra la imputación de datos mensuales faltantes de 1980 a 1985. En Rojo se muestra la imputación de datos faltantes para los meses de

febrero y mayo, se imputo el mismo valor en aquellas celadas que no contenían el valor del mes previo o anterior. **Fuente:** <https://doi.org/10.25573/data.16598810.v1>

Análisis de datos

Considerando que el periodo de tiempo que comprende los años de 1980 a 1985, presenta menos del 10% de datos imputados, se seleccionó este periodo de 70 meses continuos para para realizar el análisis de series temporales mensuales. Se realizó un análisis exploratorio empleando modelos no formales de series temporales como Media Móvil Centrada K=3 (MMC), Media Móvil No Centrada K=2 (MMNC), Suavizamiento Exponencial (SE) y Modelos Lineales (ML). Para cada uno de estos modelos se compararon los índices residuales: Desviación absoluta Media (DAM), Error cuadrático Medio (EMC), Porcentaje de Error Medio Absoluto (PEMA), Porcentaje de Error Medio (PME) y Raíz del Error cuadrático Medio (RECM). Esta parte de los análisis fue realizada en el software Excel. La medición del error en el pronóstico de series de tiempo es fundamental para evaluar la precisión de un modelo de pronóstico y determinar su efectividad en la predicción futura (Hanke & Wichern, 2005). Existen varias medidas comunes para evaluar el rendimiento de un modelo de pronóstico de series de tiempo. Algunas de las medidas más utilizadas incluyen:

- **Error:** error o residual: es la diferencia entre un valor real y su valor de pronóstico.

$$e_t = Y_t - \hat{Y}_t$$

- **DAM:** Calcula la diferencia promedio entre las observaciones reales y las predicciones del modelo, ignorando la dirección de la diferencia.

$$\frac{\sum_{t=1}^n |Y_t - \hat{Y}_t|}{n}$$

- **ECM:** Error cuadrático medio: Calcula la media de los cuadrados de las diferencias entre las observaciones reales y las predicciones del modelo. MSE da más peso a las grandes diferencias.

$$\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}$$

- **PEMA:** Porcentaje de error medio absoluto: Calcula el porcentaje promedio de diferencia entre las observaciones reales y las predicciones del modelo.

$$\frac{\sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{Y_t}}{n}$$

- **PME:** Porcentaje de error medio; indica cuan desviada esta la técnica de pronóstico usada.

$$\frac{\sum_{t=1}^n \frac{(y_t - \hat{y}_t)}{Y_t}}{n}$$

Se evaluaron aspectos generales y estructurales de la serie de datos, dando como resultado una serie de tiempo con una tendencia y sin estacionalidad, la cual se ajusta a un modelo multiplicativo Holt. La tendencia fue analizada mediante la interpretación de un gráfico lineal y la estacionalidad fue analizada de una forma más profunda con estadística circular. El análisis de estacionalidad circular se realizó utilizando el paquete circular (Agostinelli & Lund, 2017) en R, el cual proporciona herramientas robustas para el tratamiento de datos angulares, como meses transformados en radianes para representar ciclos anuales. Se aplicó al conjunto de series (1980-1985) una prueba de Rayleigh para verificar si existía una dirección predominante hacia alguno de los meses, luego se empleó la prueba no paramétrica de Mardia-Watson-Wheeler para comparar las distribuciones circulares de la abundancia entre las series temporales.

También se corroboró la irregularidad de la serie mediante gráficos de Boxplot y la prueba Dickey-Fuller Test, la cual dio como resultado que la serie no es estacionaria (p-value = 0.56), por lo que se procedió a realizar las respectivas transformaciones de varianza mediante funciones matemáticas $f(x)=X^\lambda$ como: función inversa ($\lambda=-1$), logarítmica ($\lambda=0$),

idéntica ($\lambda=1$), raíz inversa ($\lambda=-0.5$), raíz cuadrada ($\lambda=0.5$) y cuadrática ($\lambda=2$). Mediante el uso de Excel se calcularon los promedios mensuales para la serie completa, así como la desviación estándar. Posteriormente se calcularon mediante las funciones y los valores lambda (λ), los valores correspondientes a cada función y a los mismos se les calcularon la media y coeficiente de variación. La función con menor coeficiente de variación fue la función idéntica ($f(x)=X^1$), por lo que solo se procedió a realizar dos procesos de diferenciación y corroborar si las transformaciones dieron resultado, utilizando nuevamente la prueba Dickey-Fuller Test, arrojando esta vez que la serie de tiempo con las respectivas transformaciones es estacionaria (p-value = 0.01). Después de realizar todo el análisis exploratorio de la serie y ajustada la estacionariedad se seleccionó el Modelo Autorregresivo Integrado de Media Móvil (ARIMA) como el modelo más adecuado. Para la aplicación de dicho modelo y las estimaciones de los valores p d y q del modelo se utilizó una función de optimización (auto-arima) del paquete Forecast en R, que emplea el Criterio de Información de AKAIKE (AIC) y el Criterio de Información Bayesiano (BIC) de Schwarz. Finalmente se modelaron los 40 años de muestreo empleando los valores de los meses de febrero y mayo por separado, mediante un modelo de Media Móvil Centrada $K=3$, el cual según la validación de supuestos y residuales es el modelo más adecuado para explicar el comportamiento de este conjunto de datos anuales.

Machine Learning para análisis de series temporales

El presente análisis de series temporales utilizó aprendizaje profundo para modelar y predecir la abundancia de *Euglossa imperilis* en función del tiempo. En primer lugar, se instalaron e importaron librerías esenciales como pandas, numpy, scikit-learn, TensorFlow/Keras y matplotlib. Estas librerías permitieron gestionar datos, realizar transformaciones y crear modelos de aprendizaje automático, incluyendo redes neuronales

recurrentes basadas en LSTM (Long Short-Term Memory) y modelos LSTM bidireccionales, un enfoque avanzado para trabajar con secuencias temporales.

Los datos fueron procesados mediante la biblioteca pandas y numpy (McKinney, 2011) para estructurar y manipular la información extraída de un archivo en formato Excel. Se notó que la serie original tenía algunos valores faltantes para la variable de abundancia, como se puede ver en la Figura 3. Por lo tanto, se utilizó una interpolación polinómica de orden 2 para completar estos valores, manteniendo las demás columnas que contienen la información temporal (años, meses y días). En la Figura 4, se muestra la serie una vez completada con la interpolación polinómica. Además, se combinó la información de año, mes y día en una sola columna de fecha, facilitando el manejo y análisis de la serie temporal. Finalmente, en la Figura 5, se muestran los gráficos para la serie interpolada de abundancia, sumada a la serie de precipitación, organizada por fechas. Esto permitiendo verificar que el proceso de tratamiento de datos ha sido exitoso.

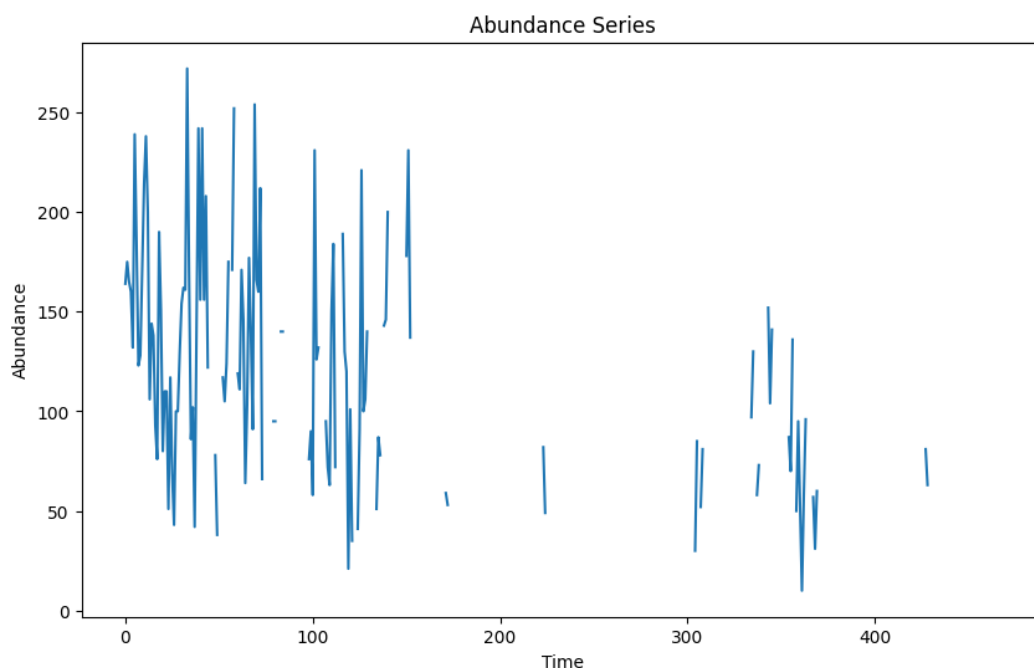


Figura 3. Serie de la variable abundancia de *Euglossa imperialis* (Original)

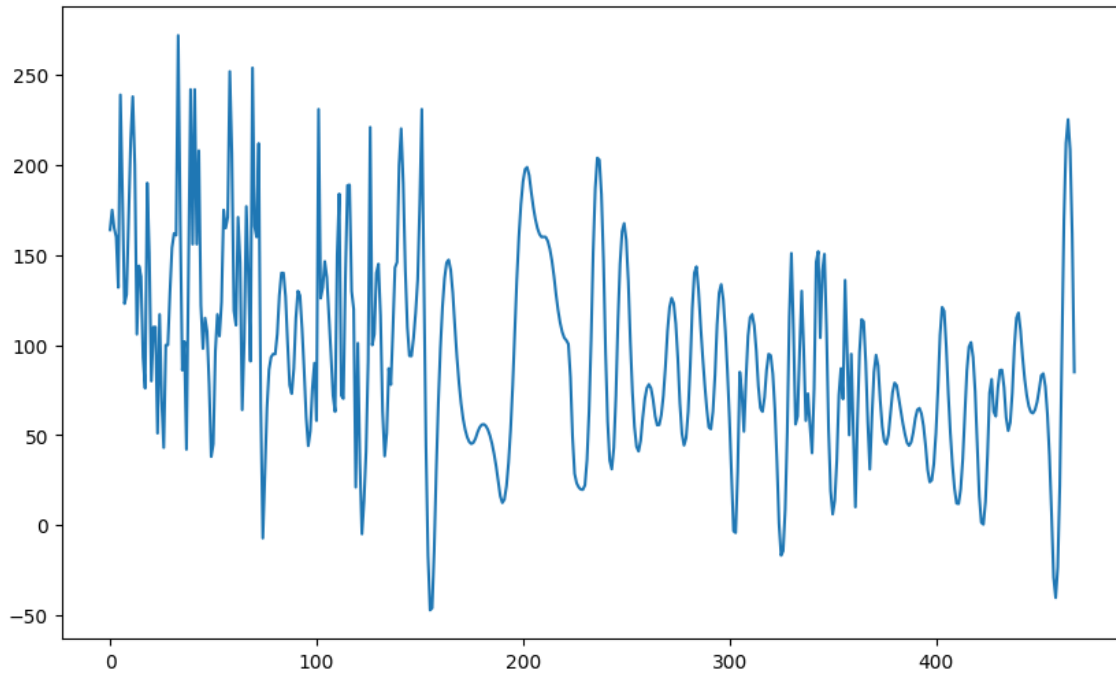


Figura 4. Serie de la variable abundancia de *Euglossa imperialis* (Completada)

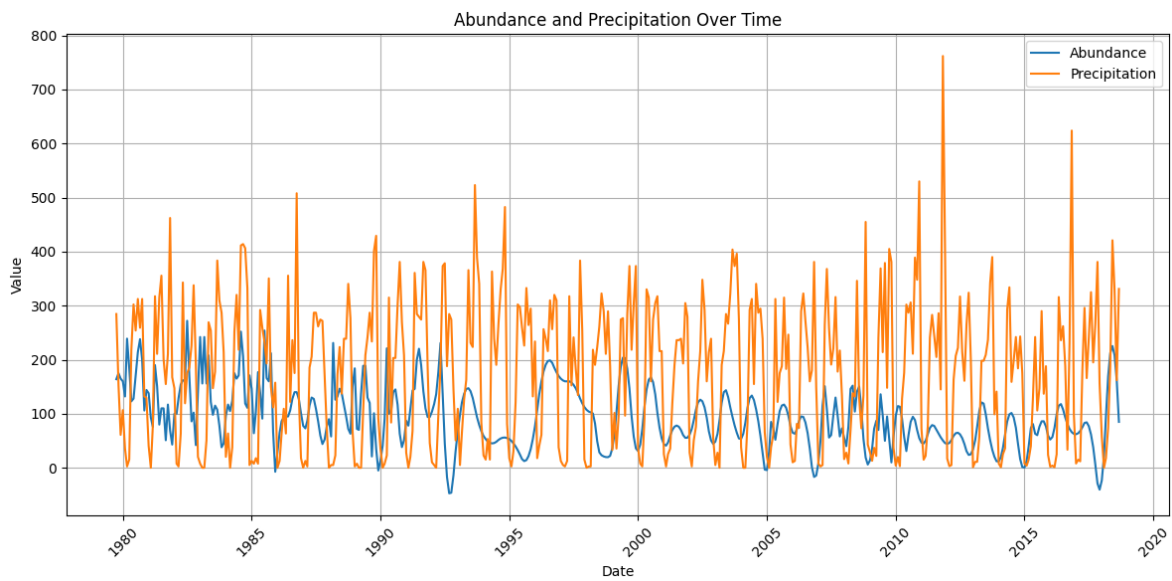


Figura 5. Series de abundancia de *Euglossa imperialis* y pluviosidad en la zona completas y formateadas.

Se implementaron técnicas de normalización utilizando scikit-learn y numpy (Pedregosa et al., 2011), específicamente la transformación Min-Max Scaling, con el objetivo de mejorar

la estabilidad numérica del modelo y facilitar la convergencia durante el entrenamiento. Para explorar las relaciones temporales y las variables ambientales, se realizaron visualizaciones con matplotlib (Hunter, 2007), incluyendo la representación gráfica de la serie de abundancia y su comparación con los valores de precipitación y la estación (seca y lluviosa), permitiendo identificar posibles correlaciones en la dinámica poblacional.

Con el fin de modelar patrones temporales en la abundancia, se construyó una red neuronal recurrente utilizando la biblioteca TensorFlow/Keras (Abadi *et al.*, 2016), incluyendo modelos LSTM bidireccionales. La arquitectura del modelo incluyó dos capas LSTM con 50 neuronas cada una, activadas mediante la función ReLU, seguidas de una capa densa encargada de la predicción del valor futuro de abundancia. La capacidad de las redes LSTM para retener información a largo plazo y mitigar el problema del desvanecimiento del gradiente resultó fundamental para capturar patrones complejos en los datos.

La optimización del modelo se realizó con el algoritmo Adam, una variante del descenso de gradiente adaptativo ampliamente utilizada en modelos de series temporales. Para evitar el sobreajuste, se implementó Early Stopping, que detiene el entrenamiento cuando la pérdida en la validación no mejora tras un número determinado de iteraciones. Finalmente, el modelo fue entrenado y evaluado utilizando un conjunto de datos dividido en entrenamiento (80%) y prueba (20%) mediante la función `train_test_split` de scikit-learn. Durante el entrenamiento, se monitoreó la pérdida (Mean Squared Error, MSE) y la métrica de error absoluto medio (Mean Absolute Error, MAE), cuyos valores fueron representados gráficamente para evaluar el desempeño del modelo.

El análisis permitió capturar patrones en la variabilidad temporal de la abundancia de *Euglossa imperilis*, proporcionando una herramienta útil para la predicción de tendencias ecológicas. Este enfoque demuestra la aplicabilidad del aprendizaje profundo en el estudio

de series temporales ecológicas, lo que podría contribuir a una mejor comprensión de las dinámicas poblacionales en ecosistemas tropicales. Esta sección de la tesis fue realizada en Python y todos los datos de las variables ambientales y meteorológicas (Temperatura, Humedad, Radiación solar y Precipitación) fueron suministrados por la Autoridad del Canal de Panamá (ACP), a través de su Dirección de Meteorología e Hidrología, que se puede acceder a través del portal de datos de STRI.

Capítulo 4:

RESULTADOS Y DISCUSIÓN

RESULTADOS

Serie temporal de abundancia de *Euglossa imperialis* (Subset)

Aproximadamente 450 meses entre 1979 y 2018 fueron analizados mediante un modelo lineal de regresión, la tendencia en este modelo sugiere una leve pendiente negativa lo que sugiere asumir que las poblaciones de *Euglossa imperialis* podrían estar disminuyendo (Fig. 6A); no obstante, al revisar los datos obtenidos a lo largo del muestreo se puede observar que existe información faltante lo implica que las técnicas de imputación consideren aspectos relacionados a la dinámica poblacional de la especie de estudio, por lo que el análisis de un fragmento de la base con pocos datos faltantes (1980-1985) es una decisión correcta en términos de robustez ante un análisis completo que no contemple los espacios o datos faltantes.

Tras la estimación y comparación de los índices residuales de cada modelo no formal obtenido al realizar los análisis de series temporales a este subset de datos (Tabla 1), el Modelo de Media Móvil Centrada K=3 resultó ser el mejor modelo. Este modelo posee valores bajos en la Diferencia Absoluta Media (DAM = 16.89), Error Medio Cuadrático (EMC = 434), y demás métricas para la evaluación de modelos predictivos. Los Modelos de Suavizamiento Exponencial y Media Móvil no Centrada, presentaron valores similares en las diferentes métricas y resultaron ser mejores modelos que el modelo lineal.

Tabla 1. Métricas de evaluación de modelos predictivos para series temporales

Modelo	DAM	EMC	PEMA (%)	PME (%)	RECM
MMNC K=2	28.00	1167.02	22.80	-7.23	28.00
SE ($y=-0.4372^{0.2}+151.11$)	28.00	1167.02	22.80	-7.23	28.00
Lineal	43.13	3552.59	47.20	-25.24	48.13
MMC K=3	16.89	434.10	14.97	-4.79	16.89

Nota: MMNC= Media Móvil No Centrada, SE = Suavizamiento exponencial, MMC = Media Móvil Centrada, DAM = Diferencia Absoluta Media, EMC= Error Medio Cuadrático, PEMA (%) = Porcentaje de Error Medio Absoluto, PME (%) = Porcentaje de Error Medio y RECM = Raíz del Error Cuadrático Medio.

El Modelo de Media Móvil Centrada presenta un ajuste bastante preciso al conjunto de datos originales, lo que muestra que a lo largo de los 70 meses que comprende el subconjunto de datos, la tendencia de la serie es casi uniforme a lo largo del tiempo, con leves variaciones en los años 1980 y 1981 (Fig. 6B). Un análisis mediante estadística circular de este fragmento de los datos también sugiere que no existe una estacionalidad mensual marcada a lo largo de la serie (Fig. 6C), lo que se aprecia cuando no existe un patrón pronunciado (picos) hacia algún mes o conjunto de ellos. Esta serie posee tendencia y no tiene estacionalidad (Fig. 7A), se ajusta a un modelo multiplicativo Holt. En cuanto al análisis de la irregularidad (Fig. 7B) que fue realizado mediante una prueba de Dickey-Fuller, se aprecia que esta serie no es estacionaria ($p\text{-value} = 0.56$), por lo que, al ajustar la estacionariedad, se podría implementar el uso de un Modelo ARIMA.

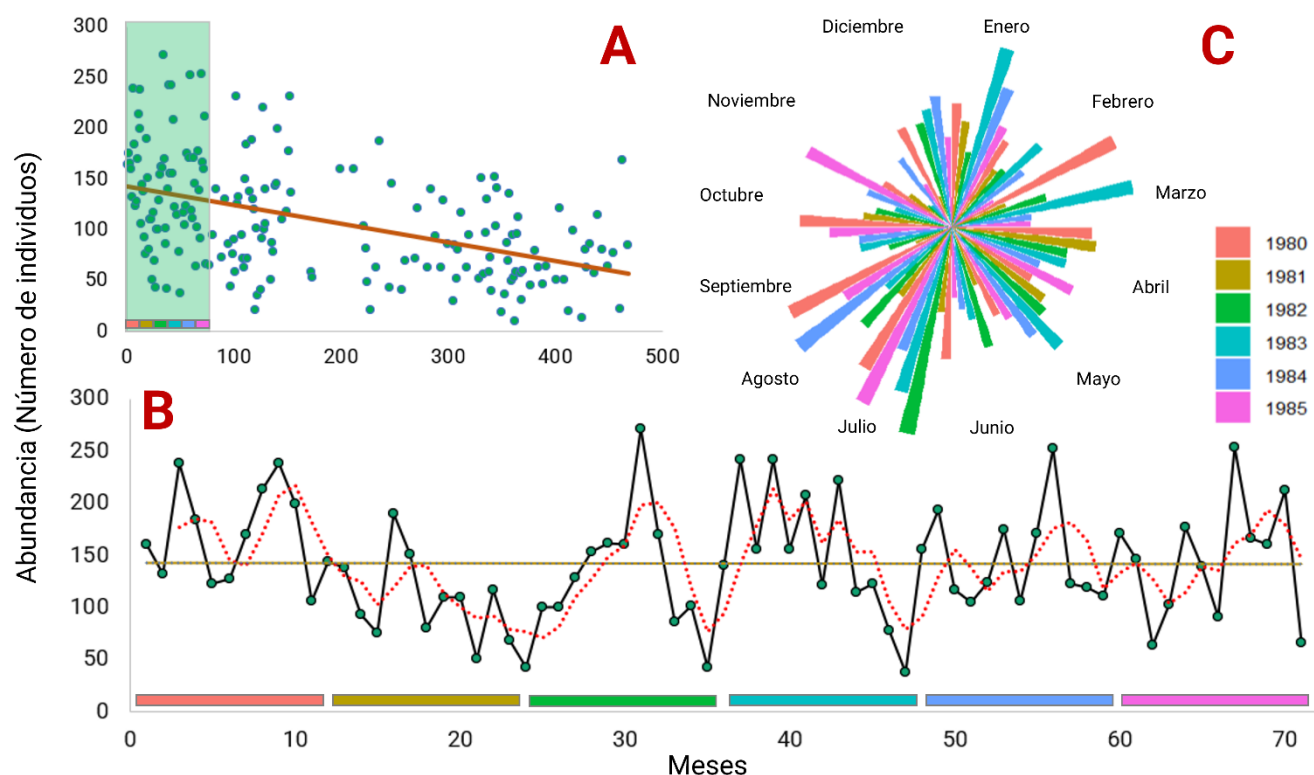


Figura 6. A) Modelo Lineal de serie temporal de *Euglossa imperialis*, B) Modelo de Media Móvil Centrada en subconjunto de datos, C) Abundancia mensual de *Euglossa imperialis* en el subconjunto analizado.

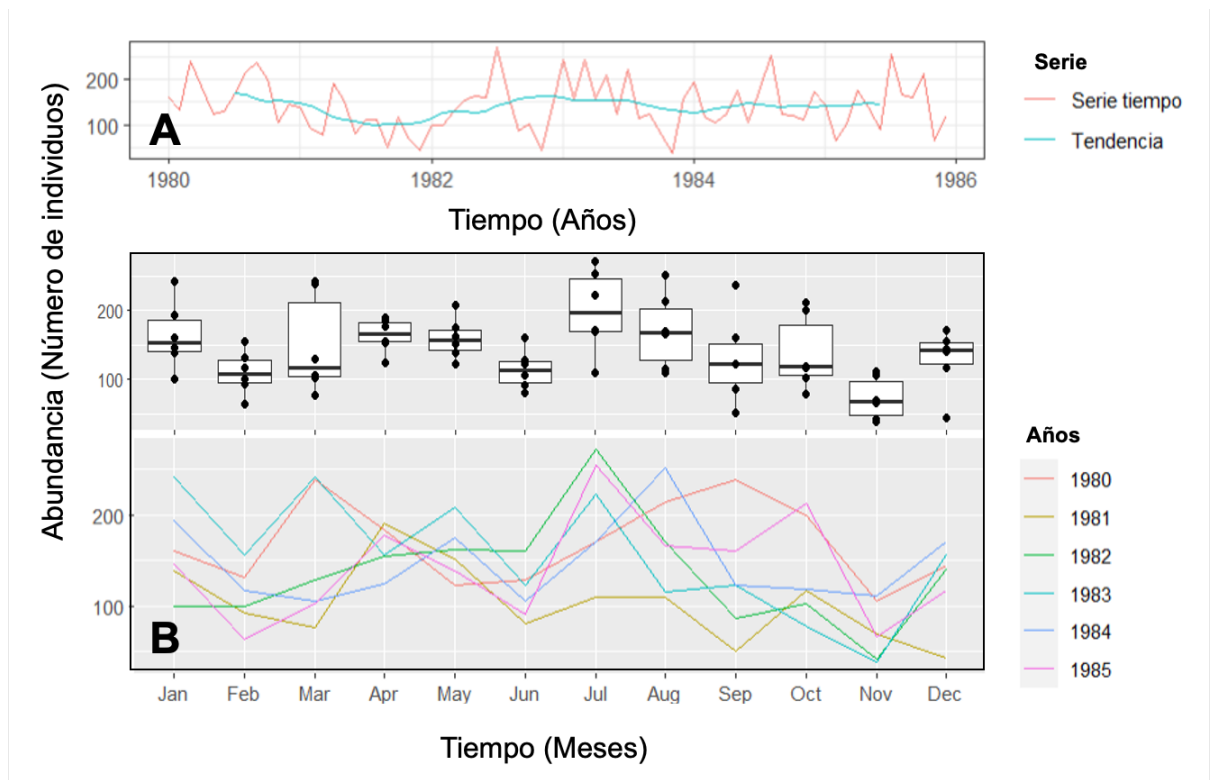


Figura 7. Serie temporal de abundancia de *Euglossa imperialis* de 1980 a 1986. A) Tendencia, B) Estacionariedad.

Al ajustar la estacionariedad y seleccionado el Modelo Autorregresivo Integrado de Media Móvil (ARIMA) mediante un algoritmo de optimización, se estimaron los valores p d y q sugiriendo el modelo SARIMA (1,0,0) (1,0,0) como el más adecuado (Fig. 8). Un aspecto importante que destacar es que, aunque el algoritmo de sugerencia indica que para esta serie temporal el mejor modelo es un SARIMA (1,0,0) (1,0,0), este modelo plantea que la serie es estable en media y en varianza, lo que necesariamente no se cumple, ya que esta serie presenta variabilidad (Figura 7B) y tendencia a lo largo del tiempo.

En cuanto a la estacionalidad analizada desde una perspectiva de estadística circular, en general las pruebas de Rayleigh dieron resultado no significativo ($p > 0.05$) es decir de forma circular los datos no presentan estacionalidad. No obstante, la prueba no paramétrica de Mardia-Watson-Wheeler si resultó significativa ($p < 0.001$), lo que indica que las variantes mensuales de los 6 años son distintas (Fig. 9). Las distintas series lineales anuales

mostraron formas distintas en cuanto a la clasificación de Ackerman (1983), por ejemplo, los años 1980 y 1981 presentaron una forma tipo I (Fig. 9A, B), 1982 tipo V (Fig. 9C) y los años 1983, 1984 y 1985 y comportamiento tipo VI (Fig. 9 D, E, F).

```
Series: data_serie
ARIMA(1,0,0)(1,0,0)[12] with non-zero mean

Coefficients:
      ar1      sar1      mean
    0.2836  0.2484  143.0082
s.e.  0.1144  0.1261  10.6468

sigma^2 = 2753: log likelihood = -386.19
AIC=780.39  AICc=780.98  BIC=789.49

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -0.9928785  51.36565  40.7788 -18.26396  37.2108  0.7389694 -0.0006875542
```

Figura 8. Salida del Modelo Autorregresivo de ARIMA en R.

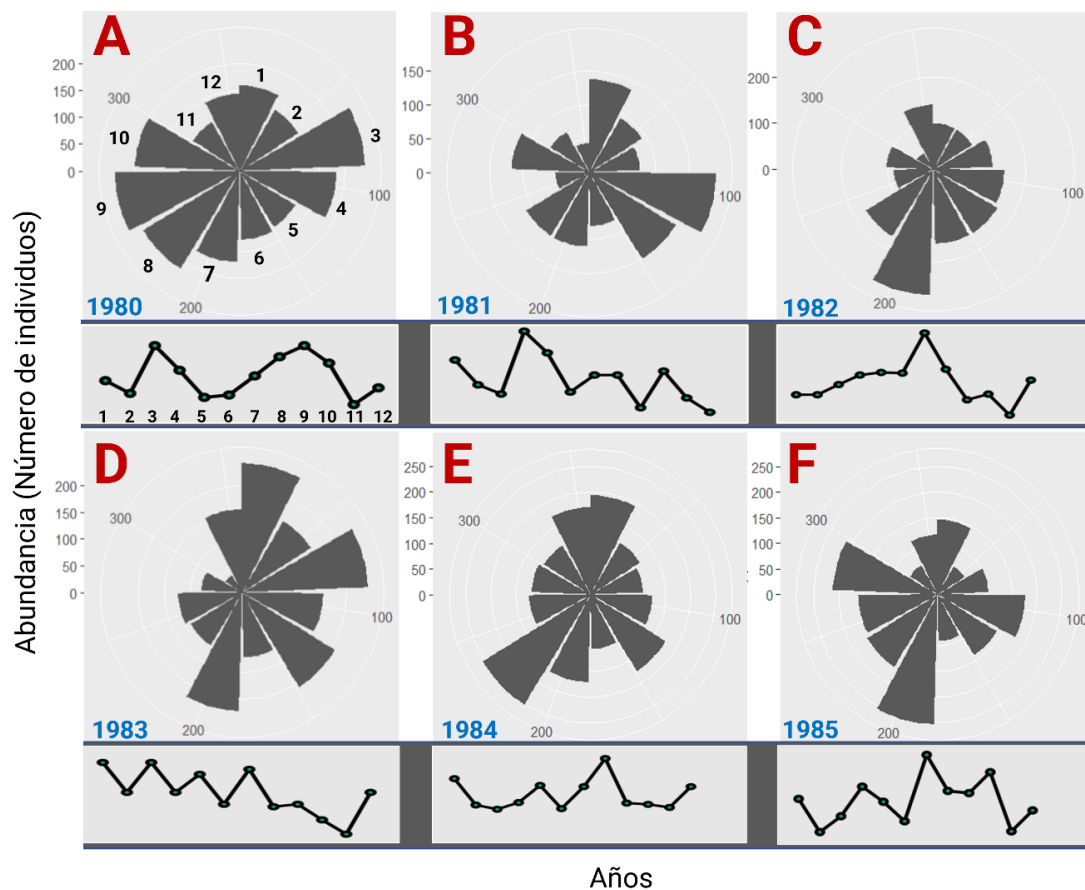


Figura 9. Variantes lineales y circulares de abundancia en la población de *Euglossa imperialis* durante los distintos años de estudio.

Serie temporal de abundancia de *Euglossa imperialis* (Base completa)

Considerando la tendencia de la serie temporal y la baja variabilidad que posee la misma en la estación seca y parte inicial de la lluviosa en cuanto a la abundancia relativa, se eligieron los meses de febrero y mayo al tratarse de meses con la menor cantidad de datos faltantes. Al emplear el modelo de serie temporal no formal de Media Móvil Centrada K=3, se puede observar que la tendencia a disminuir la abundancia de *Euglossa imperialis* a través de los años, no es tan marcada como lo sugieren los modelos lineales. Si bien existe una disminución principalmente en el mes de febrero, la misma es oscilatoria y se aprecian secciones en donde esta disminución es compensada con incremento en el mes de mayo, mes el cual posee mayor abundancia de individuos en 38 de los 40 años de estudio (Fig. 10).

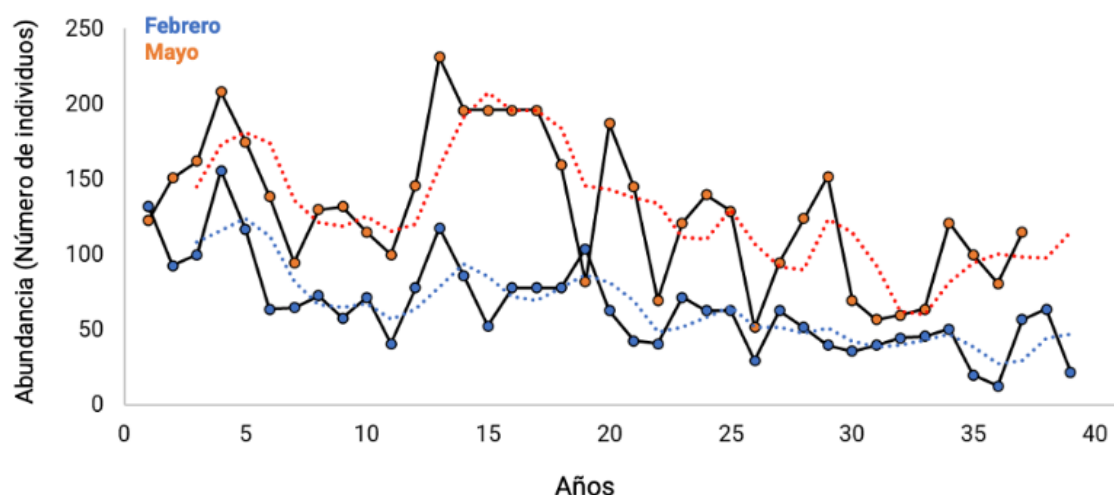


Figura 10. Serie temporal anual de la abundancia de *Euglossa imperialis* a lo largo de 40 años de estudio en un bosque tropical.

Efecto de variables ambientales sobre la abundancia temporal de *Euglossa imperialis* en un bosque tropical- Análisis de series temporales mediante Machine Learning

En esta sección se trabajó con la serie completada que incluía la pluviosidad y la abundancia, una vez ambas series fueron normalizadas para tener varianza unitaria y

promedio de cero (escalonamiento estándar min-max entre 0 y 1) fueron introducidas a una red neuronal del tipo LSTM. Se definió una función para crear secuencias a partir de los datos, estableciendo una longitud de secuencia determinada, lo que es esencial para entrenar modelos de redes neuronales recurrentes. A continuación, se muestran los resultados más relevantes de los modelos empleados:

A. Modelo 1: Redes Neuronales LSTM

La arquitectura de la red sigue: 1 capa de entrada, 2 capas LSTM de 50 neuronas con activación ReLU y una capa densa de salida con 1 unidad. El optimizador utilizado fue Adam con un *learning rate* de 0.001. La red se entrenó sobre 100 épocas, y se entrenó con batches de 32 secuencias. El conjunto de datos se divide en partes de entrenamiento y prueba en proporción 80% y 20% respectivamente. Se construyó un modelo secuencial que incluye capas LSTM bidireccionales y una capa densa para la predicción final. Además, se implementa una estrategia de *early stopping* para evitar el sobreajuste durante el entrenamiento. Los gráficos presentados en la Figura 11 se muestra la evolución de la pérdida (Loss), tanto en entrenamiento como en validación y se calculan métricas de desempeño, como el error cuadrático medio (RMSE), para evaluar la precisión de las predicciones.

Se puede notar que para ambos conjuntos la red aprendió sin mayores problemas, obteniendo una pérdida y un MSE cercanos a 0. Los resultados de predicción de este modelo se pueden ver en la Figura 12. Como resultado este modelo obtuvo un error RMSE de entrenamiento de 69.47 y un RMSE en los datos de prueba de 74.23.

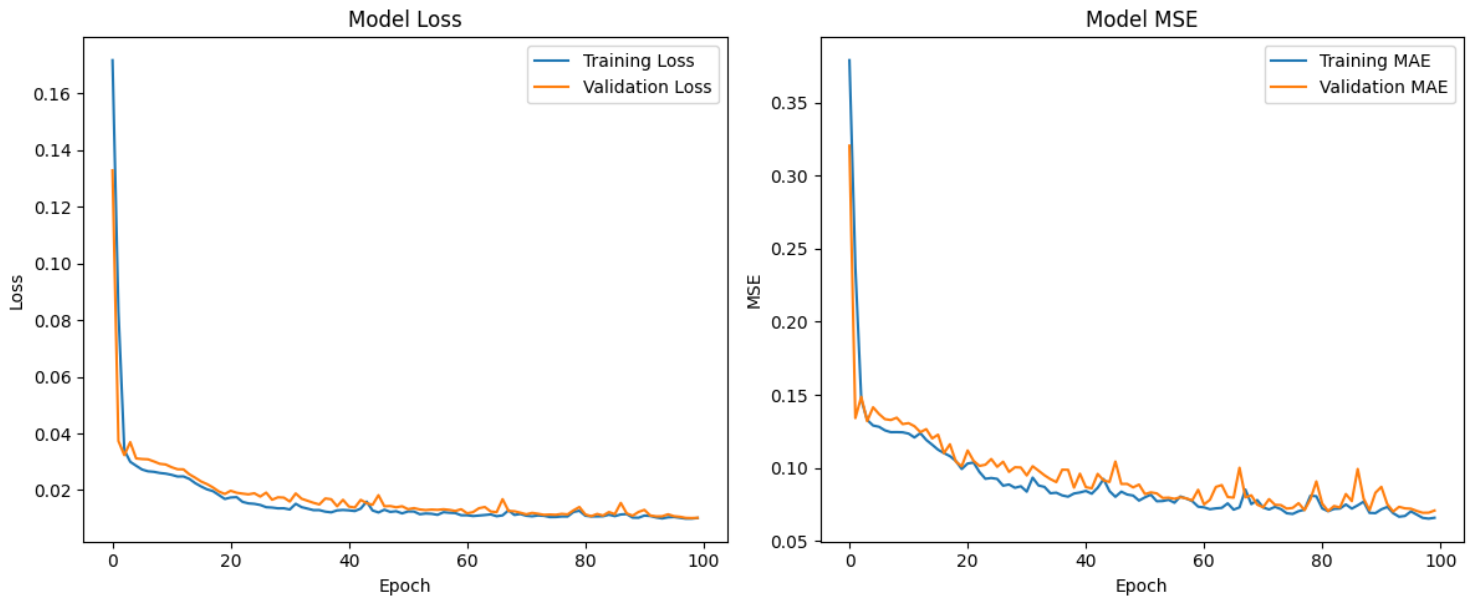


Figura 11. Curvas de entrenamiento y Prueba del Modelo 1 con redes LSTM.

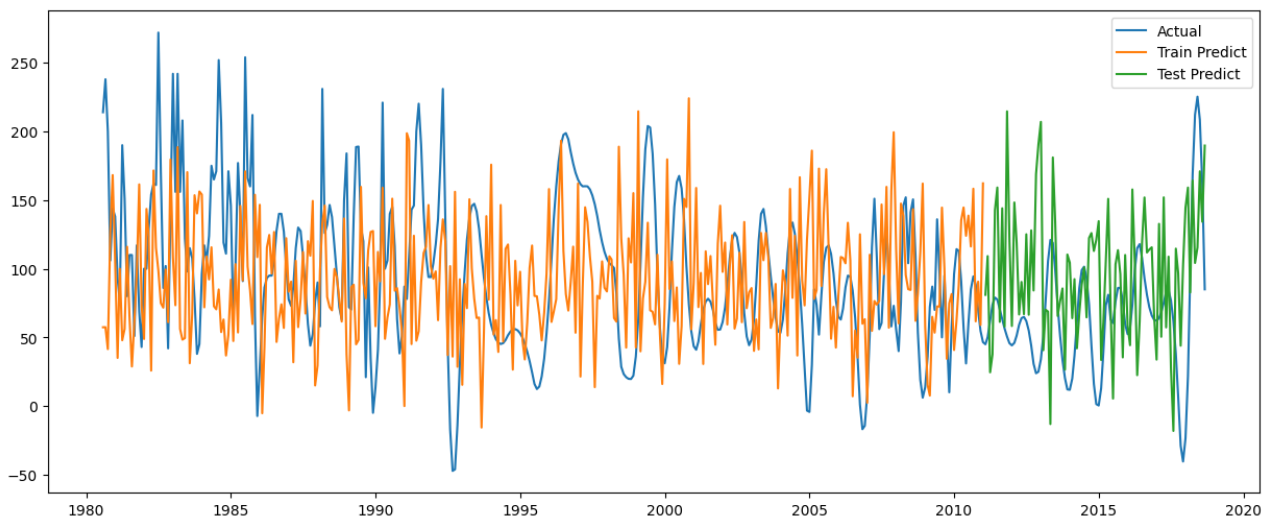


Figura 12. Resultados en la predicción de los últimos 8 años de la serie con el Modelo 1 con redes LSTM. En azul la serie real de abundancia, en naranja la serie de entrenamiento y en verde la curva de prueba predicha por el modelo.

B. Modelo 2: Redes Neuronales LSTM-Bidireccionales

Siguiendo con la investigación del modelo más adecuado, se trabajó con modelos LSTM-bidireccionales. La arquitectura de la red varía de la primera teniendo 1 capa de entrada, 2 capas LSTM-bidireccionales de 50 neuronas con activación ReLU, una capa densa de 25

neuronas con activación ReLU y una capa densa de salida con 1 unidad. El optimizador utilizado fue Adam con un *learning rate* de 0.001. La red se entrenó sobre 100 épocas, y el mismo tamaño de batches de 32 secuencias. Los gráficos presentados en la Figura 13 se muestra la evolución de la pérdida (Loss), tanto en entrenamiento como en validación) y se calculan métricas de desempeño, como el error cuadrático medio (RMSE), para evaluar la precisión de las predicciones con el modelo LSTM-bidireccional.

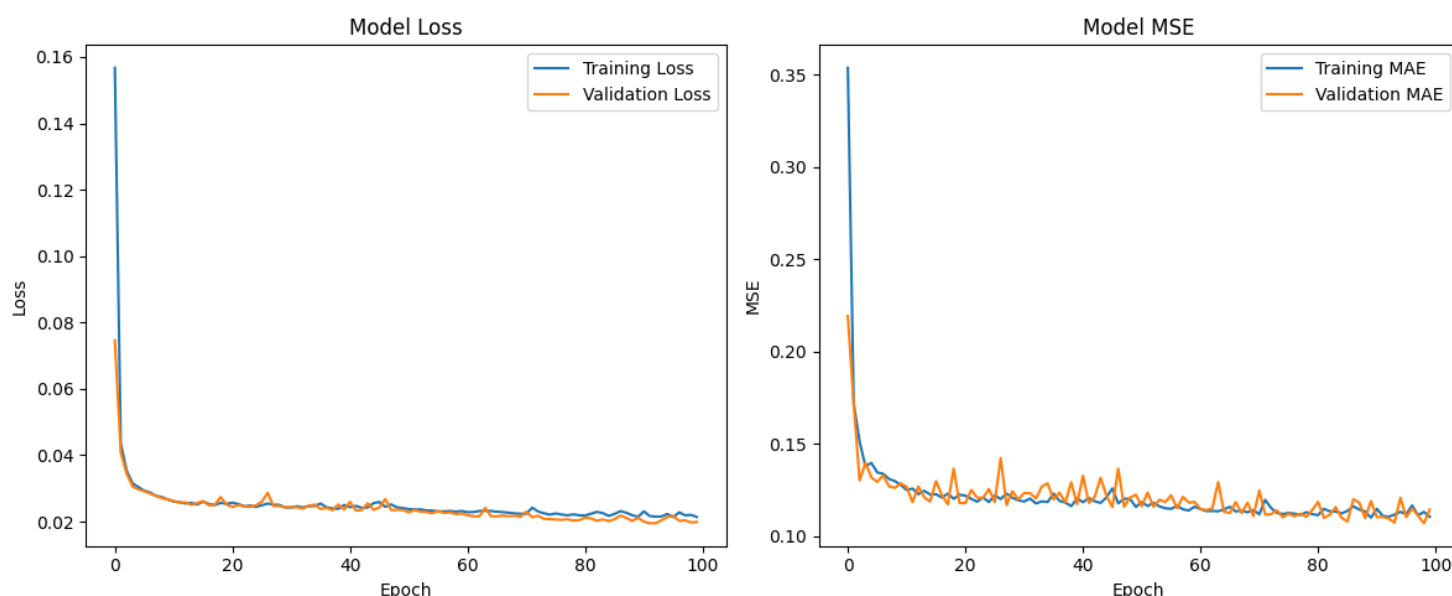


Figura 13. Curvas de entrenamiento y prueba del Modelo 2 con redes LSTM-Bidireccionales.

Se puede notar que para ambos conjuntos la red aprendió sin mayores problemas, obteniendo una perdida y un MSE cercanos a 0. Además de esto, si los comparamos con los resultados de la Figura 14, se pueden decir que estos resultados son aún más ‘suaves’ lo que se puede interpretar el optimizador está funcionando correctamente, además que el modelo se comporta bien en entramiento y prueba. Este segundo modelo obtuvo resultado RMSE de entrenamiento más bajo que el anterior con 46.51 y un RMSE en los datos de prueba de 51.29, respectivamente.

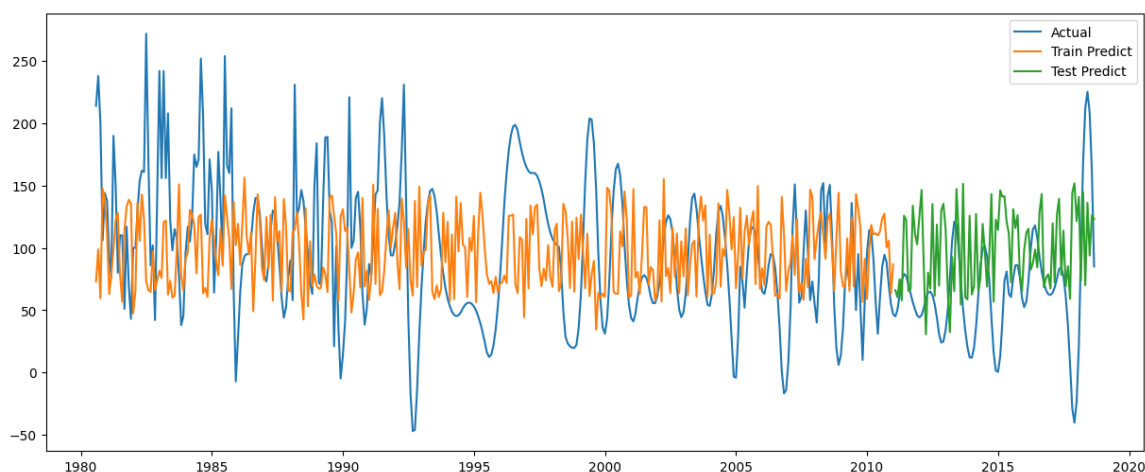


Figura 14. Resultados en la predicción de los últimos 8 años de la serie con el modelo 2 con redes LSTM-Bidireccional. En azul la serie real de abundancia, en naranja la serie de entrenamiento y en verde la curva de prueba predicha por el modelo.

C. Modelo 3: Árboles Aleatorios

Se entrenó también un modelo de árboles aleatorios (Random Forest - RF) sencillo para ver la fiabilidad de los modelos y la relevancia de los datos. Para este modelo se utilizaron los datos de precipitación y abundancia. Se entrenó un modelo de regresión basado en RF, tomando como base 100 árboles estimadores, al mejor de los árboles se evalúan su rendimiento mediante métricas de RMSE y el coeficiente de determinación (R^2). La idea es que estos resultados permitan comparar y El modelo de RF obtiene un RMSE en entrenamiento de 36.33 y de prueba de 74.91. Además, se obtiene un R^2 en entrenamiento de 0.53 y en prueba de 0.37. La Figura 15, presenta los resultados del modelo entrenado, que se utiliza para predecir la abundancia. El modelo tiene un pobre ajuste, indicando que es muy probable que estos datos sean no lineales, o hasta quizás bimodales y se deban ajustar con 2 modelos distintos.

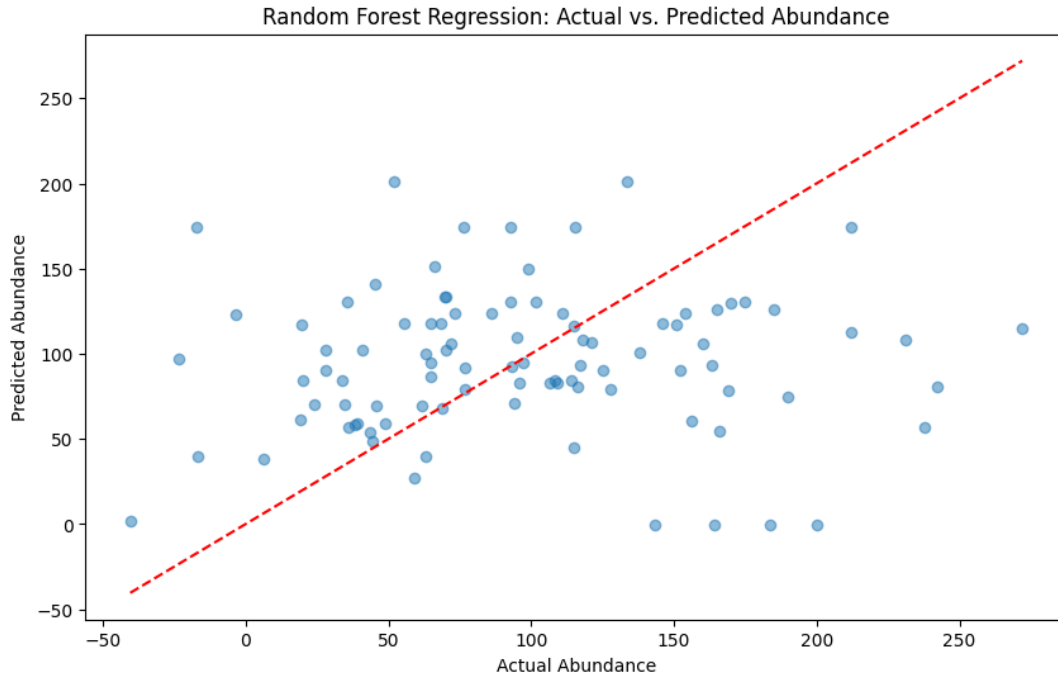


Figura 15. Modelo de predicción de abundancia mediante métodos de Random Forest.

D. Modelo 4: Redes Neuronales LSTM con otras variables meteorológicas

Se integraron y procesaron datos meteorológicos provenientes de varios archivos que contienen variables como temperatura (at), humedad relativa (rh) y radiación solar (sr). Sobre estos nuevos datos también se realizó una limpieza adicional de los datos, identificando valores negativos en las series y sustituyéndolos por valores faltantes, los cuales se completan nuevamente mediante interpolación polinómica. Se calculó el promedio mensual a partir de promedios diarios utilizando el método de re-muestreo, y se combinan estos datos en un único dataframe, ajustado a las fechas de colecta de abundancia de *Euglossa imperialis*. Los dataframes procesados se exportan a archivos CSV o Excel, y se generan visualizaciones (como histogramas) para evaluar la distribución de los datos y confirmar la calidad del proceso de limpieza, como se ve en la Figura 16.

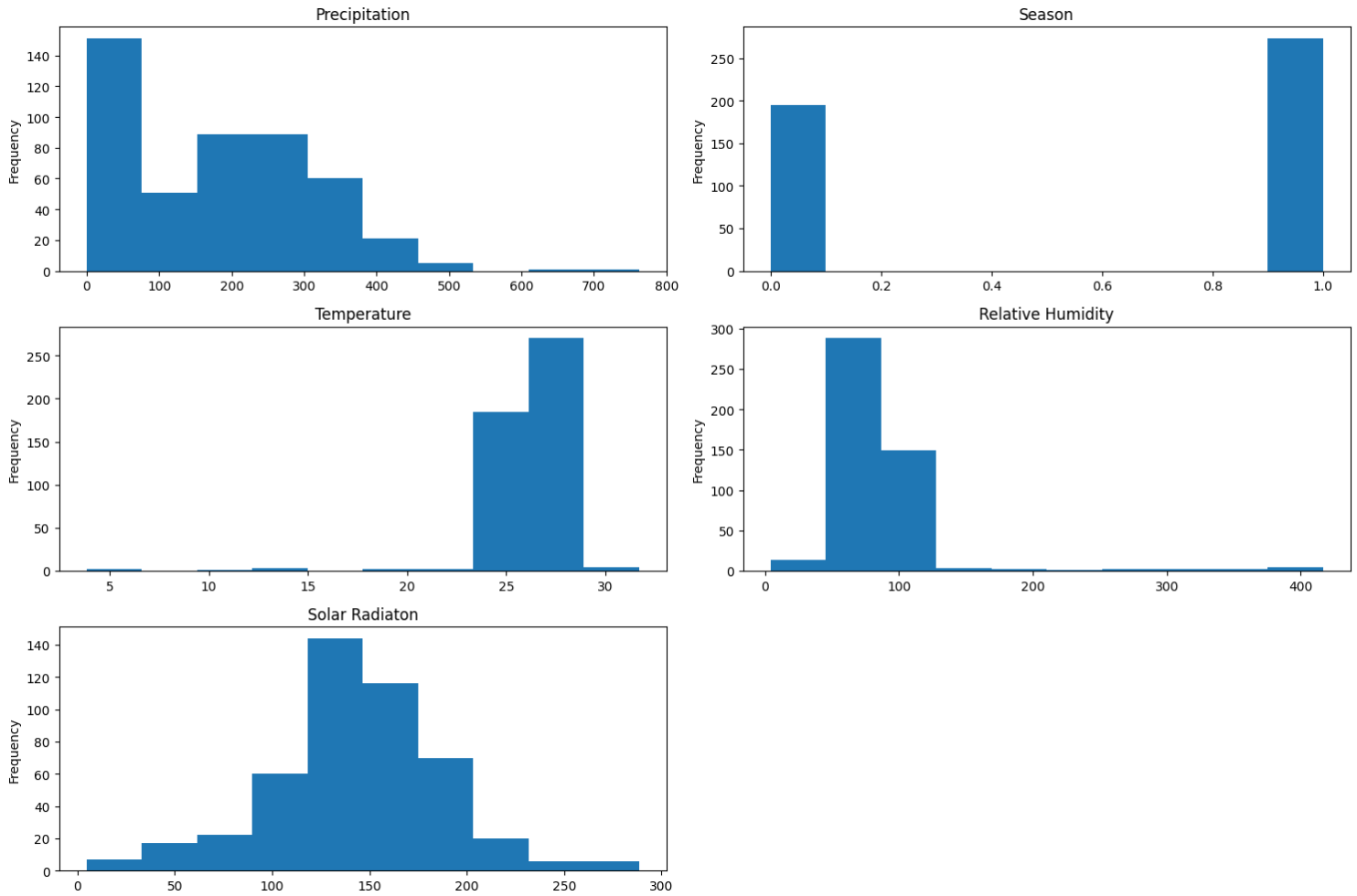


Figura 16. Histogramas para nuevas variables meteorológicas.

Finalmente se entrenó un modelo de una red LSTM, con la misma arquitectura del modelo 1, una con memoria y margen de predicción de 10 puntos. Como en otros casos la red obtuvo una pérdida y un MSE cercanos a 0. Además, se ve el uso del Early Stopping, ya que el entrenamiento termina solo con 14 épocas. Este segundo modelo obtuvo resultado RMSE de entrenamiento más bajo que el anterior con 50.27 y un RMSE en los datos de prueba de 44.07, respectivamente.

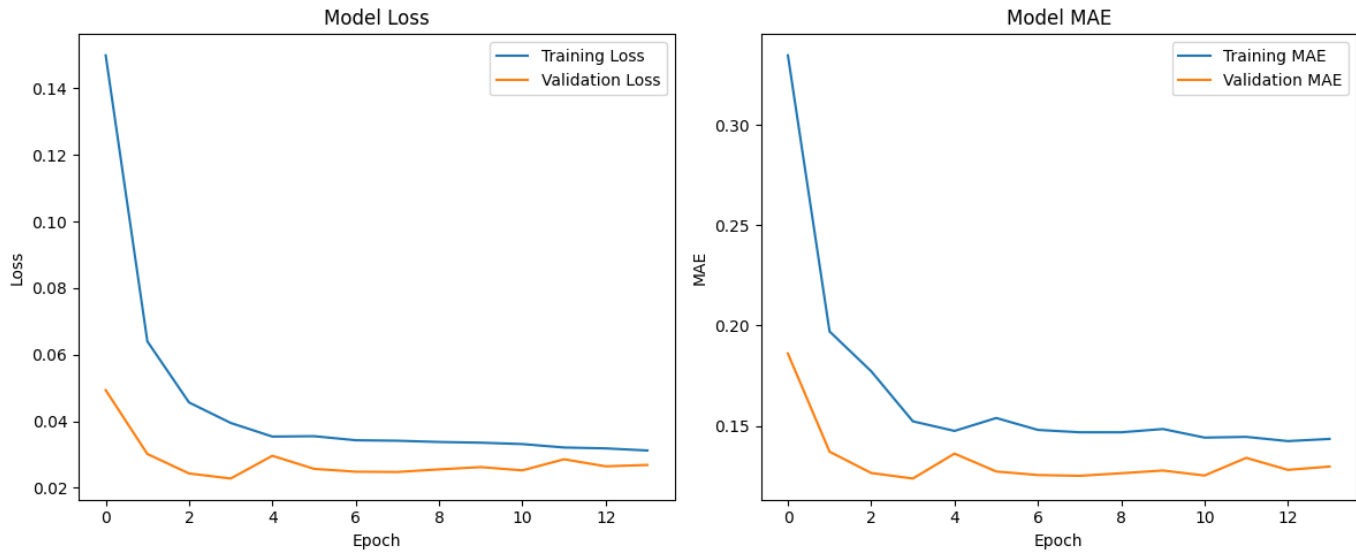


Figura 17. Curvas de entrenamiento y Prueba del Modelo 3 con redes LSTM, con más características meteorológicas.

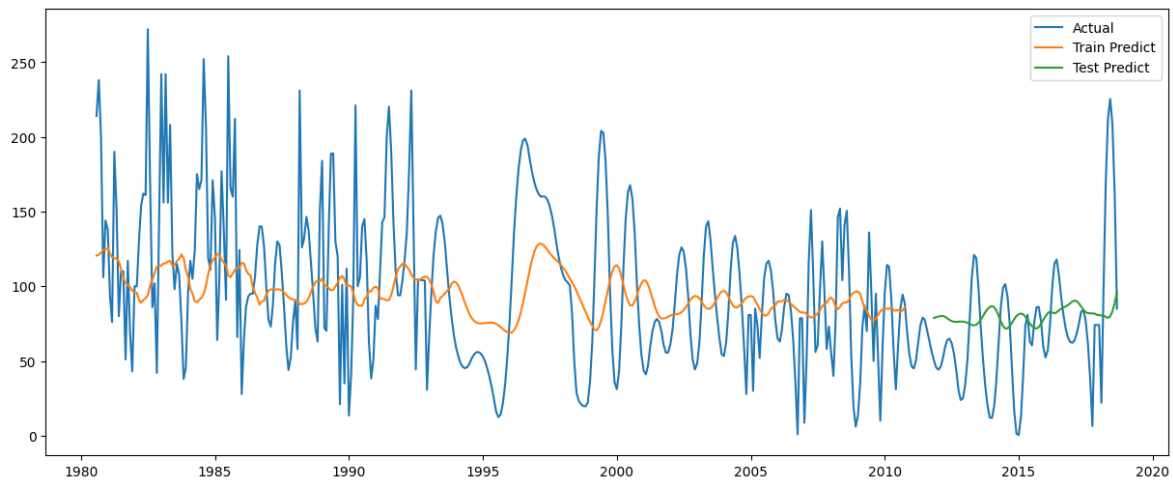


Figura 18. Resultados en la predicción de un poco menos de los últimos 8 años de la serie con el modelo 4 con redes LSTM y los datos meteorológicos ampliados. En azul la serie real de abundancia, en naranja la serie de entrenamiento y en verde la curva de prueba predicha por el modelo.

La tabla 2 se resume los resultados de los modelos para la predicción de la variable abundancia. Esta tabla muestra los resultados de la métrica de error RMSE para los

modelos en datos de entrenamiento y prueba, resultando como mejor modelo, el modelo 4 (Red LSTM con series de Precipitación, Temporada, Temperatura, Humedad Relativa, Radiación Solar). Se prevé que el modelo 4 puede ser mejorado, por ejemplo, utilizando otros modelos basados en múltiples series, como son los que provee Prophet (<https://facebook.github.io/prophet/>) o PyCaret para manejo de series temporales (https://pycaret.readthedocs.io/en/latest/api/time_series.html). Otra ruta que se debe explorar en el análisis de estas series es la introducción de retrasos y alineamiento de las series (Valencia *et al.*, 2021). También en un futuro se recomienda el uso de Transformadores para series de tiempo, como lo son Temporal Fusion Transformers (Lim & Zohren, 2021).

Tabla 2. Resultados de los Modelos de Machine Learning empleados para el análisis de serie temporal de *Euglossa imperialis* y factores ambientales.

Modelo	Datos de Entrenamiento	RMSE (Entrenamiento)	RMSE (Prueba)
Modelo 1 - Red LSTM	Serie: Precipitación	69.47	74.23
Modelo 2 - Red LSTM-Bidireccional	Serie: Precipitación	46.51	51.29
Modelo 3 - Random Forest	Serie: Precipitación	36.33	74.91
Modelo 4 - Red LSTM	Series: Precipitación, Temporada, Temperatura, Humedad Relativa, Radiación Solar	50.27	44.07

DISCUSIÓN

El análisis de la serie temporal de *Euglossa imperialis* revela una tendencia general decreciente a lo largo de los 40 años de estudio, aunque dicha disminución no es uniforme, es decir, existen secciones de la serie que no prestan disminución. La pendiente negativa observada en el modelo lineal sugiere un declive poblacional, pero los modelos de series temporales más sofisticados, como la Media Móvil Centrada K=3 y el modelo ARIMA,

indican que la variabilidad interanual y la influencia de factores ambientales pueden modular esta tendencia. Estos resultados concuerdan con estudios previos que han documentado fluctuaciones estacionales y cambios en la abundancia de abejas Euglossini en respuesta a la disponibilidad de recursos florales y la fragmentación del hábitat (Ackerman, 1989; Ramírez *et al.*, 2015). Específicamente, la menor abundancia en febrero y la relativa estabilidad en mayo pueden estar relacionadas con cambios en la fenología de las plantas hospedadoras o con dinámicas propias de la especie, lo que resalta la importancia de considerar múltiples escalas temporales al interpretar tendencias poblacionales (Janzen *et al.*, 1982).

El monitoreo químico de estas abejas macho realizado por Roubik *et al.* (2021) en dos sitios reveló que el 75 % de las especies estaban estables o en aumento. Las abejas del bosque de 33 especies, con avistamiento en vivo en Pipe Line Road (PR) y trampas en la isla de Barro Colorado (BCI), incluyeron 132,000 individuos. No se produjeron cambios de especies o comunidades en los fuertes eventos climáticos de El Niño Oscilación del Sur (ENOS), que duraron 145 meses en total durante más de 70 generaciones de abejas. Las abejas parásitas y anfitrionas fluctuaron al unísono. Unas pocas especies muy comunes, adaptadas a un hábitat relativamente perturbado, disminuyeron con el tiempo en PR, mientras que los avistamientos y las trampas revelaron rangos de abundancia estables en *Euglossa* durante 40 años. La abundancia de abejas y la biomasa se mantuvieron estables, pero el 50% de las especies tuvieron pocos registros. Las abejas de las orquídeas parecen más estables en los bosques más viejos, evidentemente aisladas de la perturbación del ENSO y probablemente beneficiándose de la creciente abundancia de lianas.

Desde una perspectiva metodológica, el uso de modelos de suavizamiento como la Media Móvil Centrada permitió detectar patrones más complejos que el modelo lineal, reduciendo el impacto de los datos faltantes y proporcionando una visión más precisa de la tendencia

poblacional. Además, el análisis mediante estadística circular sugiere la ausencia de estacionalidad marcada, lo que se alinea con estudios previos que indican que la actividad de estas abejas está influenciada por múltiples factores ambientales y no sigue estrictamente ciclos anuales (Roubik *et al.*, 2021). Sin embargo, la significancia de la prueba de Mardia-Watson-Wheeler indica variaciones interanuales en la abundancia de individuos, lo que podría ser resultado de eventos climáticos extremos o cambios en la composición floral a lo largo de los años. Estudios recientes han demostrado que la variabilidad en la fenología de la floración, impulsada por cambios en la irradiancia y la estacionalidad climática, afecta directamente la disponibilidad de recursos para los polinizadores en los trópicos (Abrahamczyk *et al.*, 2011; Wright & Van Schaik, 1994). Esta fluctuación puede explicar la variabilidad observada en la presencia de *Euglossa imperialis* en diferentes momentos del año y en distintos periodos del estudio.

El análisis de la influencia de variables ambientales sobre la abundancia temporal de *Euglossa imperialis* mediante Machine Learning aportó resultados complementarios. La red neuronal LSTM con arquitectura de dos capas de 50 neuronas y activación ReLU mostró una pérdida y un MSE cercanos a 0, con un RMSE de entrenamiento de 69.47 y de prueba de 74.23. La versión bidireccional del modelo mejoró estos resultados, con un RMSE de entrenamiento de 46.51 y de prueba de 51.29, sugiriendo una mejor capacidad de generalización. Los modelos de Random Forest, por otro lado, presentaron un rendimiento menor, con un RMSE de prueba de 74.91 y un R^2 de 0.37, lo que podría indicar la naturaleza no lineal o bimodal de los datos.

Finalmente, la inclusión de variables adicionales como temperatura, humedad relativa y radiación solar en un modelo LSTM mostró aún mejores resultados, con un RMSE de entrenamiento de 50.27 y de prueba de 44.07, demostrando el valor de integrar múltiples variables ambientales. La Tabla 2 resume el desempeño de todos los modelos probados,

destacando el modelo LSTM con múltiples variables como el más adecuado para predecir la abundancia de *Euglossa imperialis*.

La clasificación de las series lineales anuales según la tipología de Ackerman (1983) sugiere que las dinámicas poblacionales de *E. imperialis* pueden responder a múltiples factores ecológicos y antropogénicos. La interacción entre disponibilidad de recursos, presión de competencia y variabilidad climática debe ser considerada en futuros estudios. Se ha documentado que las poblaciones de polinizadores pueden responder de manera diferencial a la fragmentación del hábitat, con algunos grupos siendo más sensibles que otros a la alteración ambiental (Kerr *et al.*, 2015; Steffan-Dewenter & Tschamntke, 1999). La implementación de modelos más avanzados que integren datos climáticos y de uso del suelo podría mejorar la capacidad de predicción de estos patrones, proporcionando herramientas valiosas para la conservación de estas abejas, cuyo rol en la polinización es clave para la estabilidad de los ecosistemas tropicales (Colwell *et al.*, 2012; Phillips *et al.*, 2008).

CONCLUSIONES

1. El análisis de la serie temporal muestra una tendencia decreciente en la abundancia de *Euglossa imperialis* durante los últimos 40 años, aunque con fluctuaciones interanuales y una fuerte variabilidad en las estaciones, lo que sugiere la influencia de factores ambientales y los eventos climáticos extremos. La tendencia decreciente se ve de forma marcada en los modelos lineales.
2. Los modelos de series temporales avanzados, como la Media Móvil Centrada y ARIMA, así como los enfoques de Machine Learning como el modelo LSTM, ofrecen una visión más precisa de la dinámica poblacional de *Euglossa imperialis*, revelando patrones complejos en la abundancia que no son captados por modelos lineales simples. En si el uso integrado de distintas técnicas permite explorar patrones estacionales con mayor profundidad.
3. Resultados sugieren que las poblaciones de *Euglossa imperialis* se mantienen más estables pese a que ciertos modelos sugieren disminución, lo que refuerza la importancia de conservar los tipos de matrices paisajísticas en los que habitan estas abejas y actúan como refugios clave para la biodiversidad de polinizadores tropicales.
4. En cuanto a la relación de las variables climáticas el modelo LTSM, sugiere que la Precipitación, Temporada, Temperatura, Humedad Relativa y Radiación Solar, influyen en los patrones estacionales observados. De estas variables la Precipitación parece ser la variable más influyente, generando usualmente picos de abundancia más elevados cuando las precipitaciones son más bajas (estación seca).
5. Este estudio resalta la necesidad de implementar modelos más avanzados que integren datos climáticos y de variabilidad biológica, para mejorar las predicciones sobre las dinámicas poblacionales de polinizadores y facilitar la toma de decisiones en la conservación y gestión de los ecosistemas tropicales.

REFERENCIAS BIBLIOGRÁFICAS

- Abrahamczyk, S., Kluge, J., Gareca, Y., Reichle, S., & Kessler, M. 2011.** The influence of climatic seasonality on the diversity of different tropical pollinator groups. *PLoS ONE*, 6(11), e27115. <https://doi.org/10.1371/journal.pone.0027115>
- Ackerman, J. D. 1983.** Diversity and seasonality of male euglossine bees (Hymenoptera: Apidae) in Central Panama. *Ecology*, 64(2), 274-283.
- Ackerman, J. D. 1986.** Sistemas en orquídeas. *Lindleyana*, 1 (2), 108-113.
- Ackerman, J. D. 1989.** Geographic and seasonal variation in fragrance choices and preferences of male euglossine bees. *Biotropica*, 340-347.
- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. 2016.** TensorFlow: A system for large-scale machine learning. 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), 265-283.
- Agostinelli, C., & Lund, U. 2017.** R Package Circular: Circular Statistics (version 0.4-93). Available online: <https://r-forge.r-project.org/projects/circular/>.
- Archer, M.E. 2013.** The solitary wasps and bees of a suburban garden in Leicester, England, over 27 years. *Entomol. Mon. Mag.* 149 (1948), 93–121.
- Arcos, J. F. 2020.** Diversidad de abejas de las orquídeas (Apidae: Euglossini) en el vaso del embalse del proyecto hidroeléctrico “Hidroituango”, Antioquia Colombia.
- Batschelet, E. 1981.** Circular Statistics in Biology. London: Academic Press.
- Biesmeijer, J., Roberts, S., Reemer, M., Ohlemuller, R., Edwards, M., Peeters, T., Schaffers, A., Potts, S., Kleukers, R., Thomas, C., Settele, J. & Kunin, W. 2006.** Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. *Science*, 313(5785): 351–354.
- Benitez-Vieyra, S., Aizen, M.A. 2003.** Ecological significance of pollinator timing in a tropical orchid. *Plant Syst. Evol.* 241 (1–2), 77–85.
- Box, G. E. P., & Jenkins, G. M. 1976.** Time series analysis: forecasting and control. San Francisco: Holden-Day.
- Brown, M., Paxton, R. 2009.** The conservation of bees: a global perspective. *Apidologie* 40 (3), 410-416.
- Brockwell, P. J., & Davis, R. A. 2016.** Introduction to time series and forecasting. Springer. <https://doi.org/10.1007/b97391>
- Burkle, L.A., Marlin, J.C. & Knight, T.M. 2013.** Plant-pollinator interactions over 120 years: loss of species, co-occurrence and function. *Science* 339(6127), 1611-1615.
- Colla, S.R. & Packer, L. 2008.** Evidence for decline in eastern North American bumblebees (Hymenoptera: Apidae), with special focus on *Bombus affinis* Cresson. *Biodivers. Conserv.*, 17: 1379-1391.

- Condit, R., Watts, K., Bohlman, S. A., Pérez, R., Foster, R. B., & Hubbell, S. P. 2001.** Quantifying the deciduousness of tropical forest canopies under varying climates. *Journal of Vegetation Science*, 12(4), 649-658.
- Croat, T. B. (1978).** *Flora of Barro Colorado Island*. Stanford University Press.
- Colwell, R. K., Brehm, G., Cardelús, C. L., Gilman, A. C., & Longino, J. T. 2012.** Global warming, elevational range shifts, and lowland biotic attrition in the wet tropics. *Science*, 322(5899), 258–261. <https://doi.org/10.1126/science.1162547>
- Chavarria, G., Gómez, A., & Rodríguez, G. 2007.** Vegetación y flora del Parque Nacional Soberanía, Panamá. Instituto Smithsonian de Investigaciones Tropicales.
- Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J. 2007.** Random forests for classification in ecology. *Ecology* 88 (11), 2783–2792.
- Dirzo, R., Young, H.S., Galetti, M., Ceballos, G., Isaac, N. & Collen, B. 2014.** Defaunation in the Anthropocene. *Science*, 345(6195): 401-406. <https://doi.org/10.1126/science.1251817>
- Frankie, G., Rizzardi, M., Vinson, S., Griswold, T. 2009.** Decline in bee diversity and abundance from 1972-2004 on a flowering leguminous tree, *Andira inermis* in Costa Rica at the interface of disturbed dry forest and the urban environment. *J. Kansas Entomol. Soc.* 82 (1), 1–21.
- Fisher, N.I. 1995.** Statistical analysis of circular data. Cambridge University Press, Cambridge.
- Fischer, D., Smit, J., & Samways, M. J. 2021.** Temporal niche partitioning among pollinators: A key to biodiversity maintenance. *Ecological Entomology*, 46(1), 134–144. <https://doi.org/10.1111/een.12937>
- Gentry, A. H. 1993.** Floristic analysis of the tree flora of the Parque Nacional Soberanía. *Revista de Biología Tropical*, 41(3), 161-178.
- Goodfellow, I., Bengio, Y., & Courville, A. 2016.** Deep Learning. MIT Press. <https://doi.org/10.4258/hir.2016.22.4.351>
- Goulson, D., Nicholls, E., Botías, C. & Rotheray, E.L. 2015.** Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science* 347 (6229), 1255957.
- Herrera, C. 2019.** Complex long-term dynamics of pollinator abundance in undisturbed Mediterranean montane habitats over two decades. *Ecol. Monogr.* 89(1), e01338.
- Holt, C. C. 1957.** Forecasting seasonals and trends by exponentially weighted moving averages. Carnegie Institute of Technology, Office of Naval Research.
- Hunter, J. D. 2007.** Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90-95.
- Hyndman, R. J., & Athanasopoulos, G. 2018.** Forecasting: principles and practice. OTexts. <https://doi.org/10.1016/j.ijforecast.2003.09.015>

- Janzen, D. H., DeVries, P. J., Higgins, M. L., & Kimsey, L. S. 1982.** Seasonal and site variation in Costa Rican euglossine bees at chemical baits in lowland deciduous and evergreen forests. *Ecology*, 63(1), 66–74. <https://doi.org/10.2307/1937032>
- Jammalamadaka, S.R., SenGupta, A. 2001.** Topics in circular statistics. World Scientific Publishing, Singapore.
- Kerr, J. T., Pindar, A., Galpern, P., Packer, L., Potts, S. G., Roberts, S. M., Rasmont, P., Schweiger, O., Colla, S. R., Richardson, L. L., Wagner, D. L., Inouye, D. W., & Fontaine, C. 2015.** Climate change impacts on bumblebees converge across continents. *Science*, 349(6244), 177–180. <https://doi.org/10.1126/science.aaa7031>
- Kosior, A., Celary, W., Olejniczak, P., Fijał, J., Król, W., Solarz, W. & Płonka, P. 2007.** The decline of the bumble bees and cuckoo bees (Hymenoptera: Apidae: Bombini) of Western and Central Europe. *Oryx*, 41(1): 79-88.
- Koh, I., Lonsdorf, E. V., Williams, N.M., Brittain, C., Isaacs, R., Gibbs, J., Ricketts, T.H. 2016.** Modeling the status, trends, and impacts of wild bee abundance in the United States. *Proc. Natl. Acad. Sci. U. S. A.* 113(1): 140-145.
- Leigh, E. G. 1999.** Tropical forest ecology: A view from Barro Colorado Island. Oxford University Press.
- Leigh, E. G., Rand, A. S., & Windsor, D. M. 1996.** The ecology of a tropical forest: Seasonal rhythms and long-term changes. Smithsonian Institution Press.
- Lim, B., & Zohren, S. 2021.** Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A*, 379(2194), 20200209.
- Landim, L. P., Guimarães, E., & Ferreira, R. P. 2019.** Circular statistical methods applied to the phenology of orchid bees. *Ecological Modelling*, 400, 25–32. <https://doi.org/10.1016/j.ecolmodel.2019.03.005>
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. 2018.** Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS one*, 13(3), e0194889.
- McKinney, W. 2011.** pandas: A foundational Python library for data analysis. *Python for High Performance and Scientific Computing*, 14(9), 1-9.
- Nemésio, A., Santos, L.M., Vasconcelos, H.L. 2015.** Long-term ecology of orchid bees in an urban forest remnant. *Apidologie* 46 (3), 359–368.
- Olden, J.D., Lawler, J.J., Poff, N.L. 2008.** Machine learning methods without tears: a primer for ecologists. *Q. Rev. Biol.* 83 (2), 171–193.
- Ollerton, J., Winfree, R., Tarrant, S. 2011.** How many flowering plants are pollinated by animals? *Oikos* 120 (3), 321-326.
- Ollerton, J. 2017.** Pollinator Diversity: Distribution, Ecological Function, and Conservation. *Annu. Rev. Ecol. Evol. Syst.* 48 (1), 353-376.
- Phillips, O. L., Aragão, L. E. O. C., Lewis, S. L., Fisher, J. B., Lloyd, J., López-González, G., Malhi, Y., Monteagudo, A., Peacock, J., Quesada, C. A., van der Heijden, G.,**

- Vásquez, R., & Almeida, S. 2008.** Drought sensitivity of the Amazon rainforest. *Science*, 323(5919), 1344–1347. <https://doi.org/10.1126/science.1164033>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. 2011.** Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
- Potts, S., Biesmeijer, J., Kremen, C., Neumann, P., Schweiger, O., Kunin, W. 2010.** Global pollinator declines trends, impacts and drivers. *Trends Ecol. Evol.* 25(6): 345-353.
- Ramírez, S., Hernández, C., Link, A. & López-Urbe, M. 2015.** Seasonal cycles, phylogenetic assembly, and functional diversity of orchid bee communities. *Ecology and Evolution*; 5(9), 1896–1907. <https://doi.org/10.1002/ece3.1466>
- Roubik, D.W., Villanueva-Gutiérrez, R. 2009.** Invasive Africanized honey bee impact on native solitary bees: A pollen resource and trap nest analysis. *Biol. J. Linn. Soc.* 98 (1): 152-160.
- Roubik, D.W. 2001.** Ups and downs in pollinator populations: when is there a decline? *Conserv. Ecol.* 5 (1),2.
- Roubik, D., Wolda, H. 2001.** Do competing honey bees matter? dynamics and abundance of native bees before and after honey bee invasion. *Popul. Ecol.* 43 (1), 53–62.
- Roubik, D., Basset, Y., Lopez, Y., Bobadilla, R., Perez, F. & Ramírez, J. 2021.** Long-term (1979–2019) dynamics of protected orchid bees in Panama. *Conservation Science and Practice*, 3:e543.
- Sánchez-Bayo, F. & Wyckhuys, K.A.G. 2019.** Worldwide decline of the entomofauna: A review of its drivers. *Biological Conservation*, 232: 8-27. <https://doi.org/10.1016/j.biocon.2019.01.020>
- Senapathi, D., Carnevalheiro, L., Biesmeijer, J., Dodson, C., Evans, R., McKerchar, M., Morton, R., Moss, E., Roberts, S. & Kunin, W. 2015.** The impact of over 80 years of land cover changes on bee and wasp pollinator communities in England. *Proc. R. Soc. B Biol. Sci.*, 282(1806), 20150294.
- Steffan-Dewenter, I., & Tschardt, T. 1999.** Effects of habitat isolation on pollinator communities and seed set. *Oecologia*, 121, 432–440. <https://doi.org/10.1007/s004420050949>
- Thessen, A.E. 2016.** Adoption of machine learning techniques in ecology and earth science. *One Ecosystem* 1, e8621.
- Valencia, V., Díaz, Y., Pascale, J. M., Boni, M. F., & Sanchez-Galan, J. E. 2021.** Assessing the effect of climate variables on the incidence of dengue cases in the metropolitan region of Panama City. *International journal of environmental research and public health*, 18(22), 12108.
- Van Klink, R., Bowler, D. E., Gongalsky, K. B., Swengel, A. B., Gentile, A. & Chase, J. M. 2020.** Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. *Science*, 368(6489): 417–420. <https://doi.org/doi:10.1126/science.aax9931>

- Vega-Hidalgo, A., Añino, Y., Krichilsky, E., Smith, A. R., Santos-Murgas, A., & Gálvez, D. 2020.** Decline of native bees (Apidae: Euglossa) in a tropical forest of Panama. *Apidologie*, 51, 1038–1050. <https://doi.org/10.1007/s13592-020-00781-2>
- Wagner, D.L. 2020.** Insect declines in the Anthropocene. *Annual Review of Entomology*, 65: 457-480. <https://doi.org/10.1146/annurev-ento-011019-025151>
- Wagner, D.L., Grames, E.M., Forister, M.L., May R. Berenbaum, M.R. & Stopak, D. 2021.** Insects decline in the Anthropocene: Death by a thousand cut. *Proceedings of the National Academy of Sciences*, 118(2): 1-10. <https://doi.org/10.1073/pnas.2023989118>
- Winfree, R., Aguilar, R., Vasquez, D., LeBuhn, G. & Aizen, M. 2009.** A meta-analysis of bees' responses to anthropogenic disturbance. *Ecology*, 90(8), 2068-2076.
- Windsor, D. M. 1990.** Climate and moisture variability in a tropical forest: Long-term records from Barro Colorado Island, Panama. *Smithsonian Contributions to the Earth Sciences*, 29, 1-145.
- Winters, P. R. 1960.** Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3), 324-342.
- Wright, S. J., & Van Schaik, C. P. 1994.** Light and the phenology of tropical trees. *The American Naturalist*, 143(1), 192–199. <https://www.journals.uchicago.edu/doi/10.1086/285600>
- Willcock, S., Martínez-López, J., Hooftman, D. A., Bagstad, K. J., Balbi, S., Marzo, A., & Dawson, T. P. 2018.** Machine learning for ecosystem services. *Ecosystem Services*, 33, 165–174. <https://doi.org/10.1016/j.ecoser.2018.04.004>
- Zhao, Q., Xu, H., Ma, Y., He, Y., & Yu, H. 2020.** Monitoring spatiotemporal patterns of pollinator abundance using deep learning: A case study on honeybee activity. *Remote Sensing*, 12(2), 286. <https://doi.org/10.3390/rs12020286>