



Universidad de Valladolid

FACULTAD DE CIENCIAS

TRABAJO FIN DE GRADO

Grado en Estadística

**DESARROLLO DE MODELOS PREDICTIVOS DE FLUCTUACIONES
DEMOGRÁFICAS DE TOPILLO CAMPESINO Y DEL RIESGO PARA LOS
CULTIVOS**

**Autor: Alex Cugnod Segovia
Tutores: Eusebio Arenal Gutiérrez
Constantino Caminero Saldaña
Año: 2024/2025**

Agradecimientos

En primer lugar, me gustaría dar las gracias al Instituto Tecnológico Agrario de Castilla y León (ITACYL), por haberme dado la oportunidad de desarrollar el presente TFG en sus instalaciones e integrarme dentro del equipo del Observatorio de plagas y enfermedades de Castilla y León, donde me he sentido más que a gusto gracias a todos sus integrantes.

A Eusebio, mi tutor por parte de la Universidad de Valladolid, por darme la oportunidad de llevar a cabo este trabajo y por aconsejarme y ayudarme correctamente cada vez que lo he necesitado.

A Constantino, mi tutor dentro del ITACYL, por haberme enseñado tanto y por haber hecho de este proyecto una experiencia tan enriquecedora.

A todos los profesores y profesoras de la Universidad de Valladolid con quienes he tenido la oportunidad de asistir a sus clases, y que han contribuido de manera fundamental a mi formación como estadístico.

A mis amigos y amigas, de dentro y fuera de la universidad, por estar ahí cuando hacía falta y por hacer más llevadero todo este proceso.

A mi hermano, por ser un modelo a seguir y el motivo de que haya estudiado esta carrera.

A mis padres, por animarme desde el primer momento y ayudarme a seguir adelante en los momentos en los que no me sentía capaz de ello.

Por último, este TFG no hubiese sido posible sin los datos que han sido puestos a mi disposición, producto del trabajo y esfuerzo del personal del ITACYL, de la Consejería de Agricultura, Ganadería y Desarrollo Rural de CyL y de los Servicios Territoriales que tiene en las nueve provincias castellanoleonesas.

Dichos trabajos se han ejecutado en el marco establecido por el ACUERDO 53/2009, de 14 de mayo, de la Junta de Castilla y León, por el que se aprueba el Plan Director de Lucha contra Plagas Agrícolas en Castilla y León y por la ORDEN AYG/96/2019, de 5 de febrero, por la que se establece la estrategia de gestión integrada de riesgos derivados de la presencia de topillo campesino en el territorio de Castilla y León.

Para su financiación, históricamente se ha contado con diversos proyectos, siendo el último el proyecto GESINTTOP, cofinanciado por el ITACYL y las Diputaciones Provinciales de Palencia y Valladolid y la participación en actividades investigadoras del ITACYL, de las Universidades de Salamanca y Valladolid y del Grupo para la Rehabilitación de la Fauna Autóctona y su Hábitat.

Código y conjunto de datos

Todos los cálculos y gráficos que se han realizado a lo largo de este proyecto se han desarrollado con el software estadístico R. Para cualquier cuestión al respecto, dirigirse al autor del presente TFG, Alex Cugnod Segovia en el siguiente correo electrónico: alex.cugnod@proton.me

Para cualquier cuestión relativa a los datos, contactar con el cotutor del presente TFG Constantino Caminero Saldaña, jefe del Área de Plagas del ITACYL, en el siguiente correo electrónico: camsalco@itacyl.es

Índice

Resumen	8
Abstract	10
1. Justificación	12
2. Objetivos	19
3. Materiales y métodos	20
3.1. Área de estudio	20
3.2. Datos y cálculo de variables	20
3.2.1. Variables denso-dependientes	23
3.2.2. Variables extrínsecas	23
3.3. Software	25
3.4. Metodología estadística	26
3.4.1. Ciclos	26
3.4.2. Series temporales	26
3.4.3. Regresión lineal múltiple	28
3.4.4. División entrenamiento-test	36
4. Resultados y discusión	37
4.1. Exploración de la base de datos	37
4.2. Profundización en la ciclicidad e identificación de la distribución geográfica . .	41
4.3. Identificación preliminar de factores extrínsecos implicados en la modulación de la abundancia	46
4.4. Comparación de distintas metodologías para la definición de algoritmos predic- tivos y selección de esta	50
4.4.1. Vías de dispersión	52
4.4.2. Cultivos	53
4.5. Selección y validación de la metodología y algoritmo que proporcione el mejor ajuste a la variable dependiente observada	54
4.5.1. Vías de dispersión	55
4.5.2. Cultivos	56
4.5.3. Comprobación de las suposiciones del modelo lineal en ambos modelos	57
5. Conclusiones	60
6. Bibliografía	62

7. Anexos	71
7.1. Anexo: Distribución de ZVs por provincia y comarca	71
7.2. Anexo: Descripción de variables	73
7.3. Anexo: Librerías de R utilizadas	74
7.4. Anexo: Mapas de calor de la abundancia y ciclicidad	75
7.5. Anexo: Información sobre la ciclicidad en cada ZV	77
7.6. Anexo: Gráficos de series temporales. periodogramas y autocorrelogramas para cada ZV, para cada tipo de hábitat	81
7.7. Anexo: Correlaciones de variables climáticas estacionales por periodo	87
7.8. Anexo: Signo de cada variable para cada uno de los modelos	88
7.9. Anexo: P-valores para los tests de Durbin-Watson para los residuos del modelo en cada zona	90
7.10. Anexo: P-valores para los tests de la I de Moran para los residuos del modelo en cada periodo	91

Índice de tablas

1.	Número de variables según tipo y hábitat.	25
3.	Información general sobre cada evento en vías de dispersión	38
4.	Información general sobre cada evento en cultivos	38
5.	Eventos registrados por zona y hábitat.	39
6.	Grupo asignado a cada ZV en base a la periodicidad/es detectada.	42
7.	Número y porcentaje de zonas por grupo	42
8.	Equivalencias entre periodicidades y color de su representación	44
9.	Comparación de periodicidades asignadas a cada ZV en el presente TFG y en (Caminero-Saldaña et al., 2024)	46
10.	Número de variables candidatas para los modelos según tipo y hábitat.	51
12.	RMSE de todos modelos de vías de dispersión sobre el conjunto total de los datos de entrenamiento y subconjuntos estacionales.	52
13.	Coeficiente de determinación ajustado de todos modelos de vías de dispersión sobre el conjunto total de los datos de entrenamiento y subconjuntos estacionales.	52
14.	RMSE de todos modelos de cultivos sobre el conjunto total de los datos de entrenamiento y subconjuntos estacionales.	53
15.	Coeficiente de determinación ajustado de todos modelos de cultivos sobre el conjunto total de los datos de entrenamiento y subconjuntos estacionales.	54
16.	RMSE de todos modelos de vías de dispersión sobre el conjunto total de los datos de TEST1 y subconjuntos estacionales.	55
17.	RMSE de todos modelos de vías de dispersión sobre el conjunto total de los datos de TEST2 y subconjuntos estacionales.	55
18.	RMSE de todos modelos de cultivos sobre el conjunto total de los datos de TEST1 y subconjuntos estacionales.	56
19.	RMSE de todos modelos de cultivos sobre el conjunto total de los datos de TEST2 y subconjuntos estacionales.	56
20.	Resultados de pruebas de heterocedasticidad para ambos modelos.	57
21.	Estimaciones y resultados estadísticos del modelo de vías de dispersión.	58
22.	Estimaciones y resultados estadísticos del modelo de cultivos.	59
23.	Valores de R^2 de cada modelo para cada conjunto de datos	59
24.	Distribución de ZVs por provincia y comarca	71
25.	Variables con sus tipos y descripciones	73
27.	Librerías de R utilizadas	74
28.	Número de eventos sufridos y datos relativos a la duración y asincronía de dichos eventos para cada ZV en cultivos	77
29.	Variables seleccionadas y su signo en cada modelo de vías de dispersión.	88
30.	Variables seleccionadas y su signo en cada modelo de cultivos.	89

31.	P-valores para los tests de Durbin-Watson para los residuos del modelo por en cada zona para Vías de dispersión y Cultivos	90
32.	P-valores para los tests de la I de Moran para los residuos del modelo por periodo para Vías de dispersión y Cultivos	91

Índice de figuras

1.	Mapa con las 40 ZVs agrupadas por colores en base a su periodicidad detectada	43
2.	Comparación de las ciclicidades encontradas en el presente TFG y en (Caminero-Saldaña et al., 2024)	44
3.	Correlaciones entre variables paisajísticas, agronómicas y edáficas con la abundancia en vías de dispersión y en cultivos y con la periodicidad encontrada para cada ZV	47
4.	Correlaciones entre variables climáticas con la abundancia en vías de dispersión y en cultivos y con la periodicidad encontrada para cada ZV	49
5.	QQ-plots de normalidad para los residuos de ambos modelos	58
6.	Correlaciones entre variables climáticas con la abundancia en vías de dispersión y en cultivos y con la periodicidad encontrada para cada ZV	72
7.	Mapa de calor para la abundancia (LNVAI) por ZV, periodo y ambos hábitats vías de dispersión (arriba) y cultivos (abajo).	75
8.	Mapa de calor para la existencia o no de ciclos por ZV, periodo y ambos hábitats vías de dispersión (arriba) y cultivos (abajo).	76
9.	Series temporales de LNVAI en vías de dispersión para las ZVs desde AV_01_AM hasta PA_09_EC	81
10.	Series temporales de LNVAI en vías de dispersión para las ZVs desde PA_10_EC hasta VA_05_SE	82
11.	Series temporales de LNVAI en vías de dispersión para las ZVs desde VA_06_SE hasta ZA_05_DB	83
12.	Series temporales de LNVAI en cultivos para las ZVs desde AV_01_AM hasta PA_09_EC	84
13.	Series temporales de LNVAI en cultivos para las ZVs desde PA_10_EC hasta VA_05_SE	85
14.	Series temporales de LNVAI en cultivos para las ZVs desde VA_06_SE hasta ZA_05_DB	86
15.	Correlaciones entre variables climáticas estacionales con la abundancia en vías de dispersión y en cultivos para cada periodo	87
16.	Correlaciones entre variables climáticas estacionales con la abundancia en vías de dispersión y en cultivos para cada periodo	87

Resumen

El topillo campesino (*Microtus arvalis* Pallas) es un pequeño roedor que experimenta fluctuaciones demográficas cíclicas que, dadas sus excepcionales características reproductoras, pueden provocar que se alcancen densidades de incluso miles de individuos por hectárea en un intervalo muy corto de tiempo, lo que motiva riesgos agrícolas y pérdidas significativas en las cosechas. Uno de los mayores problemas con que se encuentran los distintos agentes involucrados en la gestión integrada de este problema es que, dada la velocidad con que el topillo puede incrementar sus tamaños poblacionales, la ventana de reacción cuando se detecta el problema es muy corta, incluso inexistente. De esta forma, la principal herramienta necesaria para mejorar la lucha frente al topillo campesino es el desarrollo de modelos predictivos que permitan ampliar la ventana de reacción en un tiempo suficiente.

Los estudios precedentes refieren que su curva de abundancia poblacional depende de, y está modulada por, una compleja combinación de factores intrínsecos y extrínsecos y sus interacciones, no existiendo un consenso científico sobre la explicación final de este intrincado sistema en las distintas zonas donde el topillo está presente.

En este TFG se aborda esta problemática, explorando, en primer lugar, la variabilidad espacio-temporal, las componentes cíclicas y la detección de factores externos que pudieran estar involucrados en las fluctuaciones demográficas del topillo campesino en Castilla y León y, en segundo, proponiendo un primer modelo predictivo de la ratio de crecimiento poblacional a un año vista.

Para ello, se ha dispuesto de una base de datos de abundancia de topillo campesino, que incluye los resultados de su monitorización desde 2012 a 2024 en 40 zonas agrícolas de Castilla y León y dos tipos de hábitat (vías de dispersión y parcelas de cultivo) considerando una estructuración temporal que contempla tres períodos cuatrimestrales por año. También se ha dispuesto de la información relativa a variables relacionadas con factores externos potencialmente relacionados con las fluctuaciones demográficas, siendo dichas variables de varios tipos: paisajísticas, agronómicas, edafológicas y climatológicas.

La metodología estadística utilizada en los distintos pasos ha incluido, principalmente, la elaboración de periodogramas y autocorrelogramas, la obtención de correlaciones, la elaboración de modelos de regresión múltiple mediante diversas metodologías y la validación de los mismos.

Los principales resultados han mostrado un patrón muy complejo en cuanto a la expresión diferencial espacio-temporal de la abundancia, con significación tanto para los factores de zona, período como tipo de hábitat, así como para sus interacciones. Esto se ha traducido en la diferente expresión zonal y a nivel de hábitat, en cuanto a ocurrencia, asincronía y duración para cada uno de los seis eventos generales de incrementos demográficos que se han descrito en el periodo de estudio.

Se han detectado variables relacionadas con la expresión de la abundancia considerando sus componentes aditivas espaciales y temporales, destacando, por su novedad frente a estudios

precedentes: el interés de los pastizales arbolados y arbustivos como zonas de compensación ecológica con efecto tamponador de la abundancia; y la relación de esta con características texturales del suelo que favorezcan la cubierta vegetal y la estabilidad de las madrigueras.

Por último, se han desarrollado y comparado diversos modelos predictivos, tanto para las ratios de crecimiento de las abundancias en vías de dispersión como para la colonización de las parcelas de cultivo, en base a variables denso-dependientes y extrínsecas, utilizando para su ajuste un set de datos de entrenamiento y, para la validación de su bondad, otros dos sets de datos, uno geográfico y otro para su capacidad predictiva en lo temporal.

De esta forma, se han propuesto dos modelos, uno para el hábitat vías de dispersión y otro para las parcelas de cultivo:

- El de vías de dispersión ha mostrado un coeficiente de ajuste del 0.644 en el set de entrenamiento, del 0.684 en la bondad geográfica y del 0.559 en su capacidad predictiva temporal, que ha incluido 12 variables, de las cuales 5 están relacionadas con denso-dependencia, 1 con características paisajísticas y 6 con factores climatológicos.
- El de cultivos ha mostrado un coeficiente de ajuste del 0.670 en el set de entrenamiento, del 0.750 en la bondad geográfica y del 0.138 en su capacidad predictiva temporal, que ha incluido 12 variables, 6 relacionadas con denso-dependencia, 2 con características paisajísticas y 4 con factores climatológicos.

Abstract

The common vole (*Microtus arvalis* Pallas) is a small rodent that undergoes cyclical demographic fluctuations which, due to its exceptional reproductive characteristics, can lead to population densities of even thousands of individuals per hectare in a very short period of time, resulting in agricultural risks and significant crop losses.

One of the main challenges faced by the various stakeholders involved in the integrated management of this issue is that, given the speed at which the vole can increase its population size, the reaction window once the problem is detected is very short or even non-existent. Therefore, the main tool needed to improve the fight against the common vole is the development of predictive models that allow the reaction window to be extended sufficiently in time.

Previous studies report that the population abundance curve depends on, and is modulated by, a complex combination of intrinsic and extrinsic factors and their interactions, with no scientific consensus on the final explanation of this intricate system in the different regions where the vole is present.

This TFG addresses this problem by first exploring the spatio-temporal variability, the cyclical components, and the detection of external factors that may be involved in the demographic fluctuations of the common vole in Castilla y León, and secondly, by proposing an initial predictive model for the population growth rate one year ahead.

To this end, a vole abundance database has been used, including monitoring results from 2012 to 2024 across 40 agricultural zones in Castilla y León and two habitat types (dispersal corridors and crop plots), considering a temporal structure with three four-month periods per year. Additionally, data on variables potentially related to demographic fluctuations have been included, grouped into several categories: landscape, agronomic, soil, and climatic variables.

The statistical methodology used in the different stages mainly included the creation of periodograms and autocorrelograms, correlation analysis, multiple regression modeling using various approaches, and model validation.

The main results showed a very complex pattern in terms of the differential spatio-temporal expression of abundance, with significance for the factors of zone, period, and habitat type, as well as for their interactions. This has led to different zonal and habitat-specific expressions in terms of occurrence, asynchrony, and duration for each of the six general demographic increase events described during the study period.

Variables related to abundance expression were detected considering their additive spatial and temporal components. Notably, and in contrast to previous studies, the importance of wooded and shrubby grasslands as ecological compensation areas with a buffering effect on abundance was highlighted, along with the relationship between abundance and soil texture characteristics that promote vegetation cover and burrow stability.

Finally, several predictive models were developed and compared, both for growth rate predictions in dispersal corridors and for colonization of crop plots, based on density-dependent

and extrinsic variables. The models were fitted using a training dataset and validated using two other datasets: one for geographic validation and one for temporal predictive ability.

Two models were proposed, one for the dispersal habitat and one for the crop plots:

- The dispersal corridors habitat model showed a fit coefficient of 0.644 on the training set, 0.684 for geographic validation, and 0.559 for temporal predictive ability. It included 12 variables: 5 related to density dependence, 1 to landscape characteristics, and 6 to climatic factors.
- The crop habitat model showed a fit coefficient of 0.670 on the training set, 0.750 for geographic validation, and 0.138 for temporal predictive ability. It included 12 variables: 6 related to density dependence, 2 to landscape characteristics, and 4 to climatic factors.

1. Justificación

El topillo campesino, *Microtus arvalis* (Pallas, 1778) es un pequeño roedor perteneciente a la subfamilia de los arvicolinos. Cuenta con un cuerpo achatado y macizo, apenas cuello, orejas pequeñas y hocico ancho y corto. Su longitud, en los adultos, oscila desde los 94 mm hasta los 137 mm entre cuerpo y cabeza, a la que se suma una corta cola cuya longitud media es de aproximadamente 35 mm. De peso corporal medio entre 25 y 30 g, puede sobrepasar los 50 g en condiciones óptimas, habiéndose observado casos incluso de 77 g [1]. El pelaje es pardo-grisáceo en el dorso y más blanquecino en la zona ventral. No presenta dimorfismo sexual en el tamaño ni coloración [2].

Habita en hasta 41 países, principalmente en Europa, con una distribución que abarca desde la península ibérica hasta el centro de Rusia siendo la especie de micromamífero más abundante de este continente. Se han encontrado ejemplares en todo el rango desde 0 hasta 3000 m de altitud. En España está presente en la zona pirenaica y, de forma aislada al resto de la zona de distribución, en la meseta norte, principalmente en la comunidad autónoma de Castilla y León y sistemas montañosos circundantes.

En el plano ecológico, ocupa un nicho importante en la cadena trófica, siendo presa de gran cantidad de depredadores [3]. Asimismo, es consumidor y regulador de la fase vegetal, alimentándose de las hojas, tallos, raíces y granos de una gran variedad de plantas herbáceas, tanto monocotiledóneas como dicotiledóneas, tanto plantas silvestres como cultivadas [1]. Dentro de estas últimas, ataca a prácticamente todos los cultivos, desde cereales y leguminosas a colza, remolacha, patata, girasol u hortalizas, incluso pudiendo afectar a cultivos leñosos como el viñedo o árboles frutales [2].

Uno de los aspectos más interesantes de este roedor es la extraordinaria capacidad reproductiva que puede alcanzar: es la especie de mamífero con hembras más precoces a nivel reproductor, pudiendo llegar a ser fértiles en ocasiones con tan solo dos semanas de vida; el periodo de gestación también es corto, de tres semanas; las camadas resultantes pueden llegar a ser de hasta 13 crías; es una de las pocas especies que puede presentar estro postparto, pudiendo volver a quedar gestantes nada más parir [2] [3].

Como resultado de lo anterior, cuando expresan esa capacidad reproductiva son capaces de incrementar su tamaño poblacional de forma muy rápida, multiplicándose por varios cientos en un periodo muy corto de tiempo (apenas unos meses). Por citar un ejemplo con datos reales en Castilla y León [4], los valores de reclutamiento por natalidad primaverales (estación donde, en general, se dan las mayores ratios reproductivas) tienen un promedio que implica un potencial de multiplicación del tamaño poblacional por 196 durante este periodo. Pero, cuando se dan los condicionantes adecuados, se ha observado que esta cifra puede llegar a tomar valores superiores a 1100. De esta forma, pueden alcanzar de forma regular densidades superiores a los 2000 individuos por hectárea, un orden de magnitud superior al de cualquier otra especie de topillo [5].

La ocurrencia de este tipo de eventos de incrementos poblacionales drásticos junto a su no discriminación entre cultivos, han motivado históricamente que los brotes de topillo campesino hayan afectado considerablemente a la agricultura en Europa. Pudiendo llegar a producir pérdidas en parcelas de hasta el 100 %, los esfuerzos de control frente a esta especie son considerables en toda Europa, especialmente durante esos brotes poblacionales. De hecho, cuando se consideran tanto los daños y pérdidas ocasionados en los cultivos, como los costes de gestión asociados, el resultado es que el topillo campesino está considerado como la plaga de vertebrados más agresiva y dañina existente para la agricultura europea. [2].

A los conflictos agrícolas que ocasiona, se debe añadir la problemática sanitaria que puede llevar apareada. Así, durante los periodos de incrementos demográficos constituye un amplificador ambiental de patógenos de los que son hospedadores [6]. Lo que incluye agentes causales de enfermedades zoonóticas [7] [8], con los consiguientes riesgos que conlleva para la propia población humana. Un ejemplo son los brotes epidémicos de tularemia en humanos que se llevan registrando en Castilla y León desde 1997, únicos en España y sur de Europa, que están directamente relacionados con las fluctuaciones demográficas de topillo campesino [8] [9].

El fenómeno de los brotes de topillo, o explosiones demográficas, ha sido descrito ampliamente en cuanto a su ocurrencia, en prácticamente todas las zonas de su distribución, habiéndose detectado que suele suceder con un patrón cíclico que puede variar, en función de zonas, de 2 a 5 años [10]. Es durante estas explosiones cuando puede ocasionar los importantes conflictos con los intereses agrícolas comentados anteriormente [7].

El topillo campesino es una especie fosora (excavadora), que vive en guaridas subterráneas que él mismo construye. Estas madrigueras, también llamadas huras, de 10-40 cm de profundidad, están constituidas por uno o varios nidos, normalmente ubicados en las partes centrales de las madrigueras, y unas cavidades de almacenaje, conectados entre ellos mediante galerías subterráneas abiertas al exterior por las bocas o entradas [11]. En función del tiempo que lleven las madrigueras y del número de individuos que en ellas habiten, pueden ser muy sencillas, con tan sólo dos o tres bocas, o realmente muy complejas, con decenas de entradas ocupando en superficie varias decenas, incluso centenas, de metros cuadrados [2].

Estas madrigueras las excava en entidades paisajísticas, conocidas como zonas de reservorio, que tienen dos características principales: presencia continua de cubierta vegetal herbácea, lo que proporciona al topillo alimento y protección frente a parte de sus depredadores; y ausencia de actividad humana, o por otras causas, que implique remoción del terreno, lo que proporciona estabilidad en el tiempo para las madrigueras [12].

En territorio agrícola, en el caso de Castilla y León, existen distintos tipos de reservorio. En primer lugar, reservorios más o menos grandes, frecuentes o dispersos, representados por praderas, eriales, parcelas retiradas de cultivo, alfalfares y otros cultivos herbáceos plurianuales. Si bien con un hábitat no tan óptimo, las parcelas que se conducen en un régimen de siembra directa también son candidatas para suponer reservorio [7] [13].

Otro tipo de hábitat en territorio agrícola son las entidades “lineales” que forman las masas

herbáceas asociadas a cunetas de carreteras, caminos, acequias, arroyos, linderas, etc. Es decir, en general, zonas adyacentes a infraestructuras de comunicación y cauces de agua. Aunque en superficie representan un porcentaje del territorio muy bajo, son zonas de trascendental importancia para entender la dinámica poblacional del topillo como plaga: constituyen un entramado lineal de masas herbáceas interconectadas entre sí y distribuidas por todo el territorio que actúan como corredores, facilitando así tanto el mantenimiento como la dispersión de este animal. Dada esta última característica, en adelante, se hará referencia a este tipo de hábitat como “vías de dispersión”. En añadido, el hecho de ser contiguos y lindantes a las parcelas de cultivo hace que el topillo pueda llegar a ellas, suponiendo el frente de ataque directo a las mismas en los momentos de riesgo [7].

En periodos de baja densidad el topillo se halla recluido en los reservorios, resultando en una presencia casi imperceptible de la especie en el territorio y sin ocasionar afectaciones significativas a los cultivos. Sin embargo, cuando se dan las circunstancias de una explosión demográfica, los individuos se reproducen y se van expandiendo por la zona de reservorio hasta que, por la competencia por el espacio, el alimento o la reproducción, parte de la población comenzará a dispersarse a otras zonas, hasta llegar a un grado en el que esa competencia le fuerce a colonizar espacios o entidades paisajísticas que no ocuparía bajo otras circunstancias, entre ellos las parcelas de cultivo, a las que llegarán siendo el frente habitual directo de ataque las vías de dispersión [7]. Cuando llega la fase del colapso poblacional, se revierte este proceso, sucediendo la contracción y obteniendo como resultado que las poblaciones de topillo remanentes tiendan a volver a estar limitadas a los reservorios, lo que a su vez representa el “germen” para la próxima explosión demográfica.

Es decir, las fluctuaciones demográficas del topillo campesino implican la sucesión de procesos de expansión-contracción en el espacio. Debido a las extraordinarias características reproductivas del topillo ya señaladas, este proceso de expansión es realmente rápido: en muy pocos meses se puede producir la afectación a las parcelas de cultivo. Además, la explosión demográfica es prácticamente indetectable hasta que no salen de los reservorios, cuando, normalmente, ya es tarde para reaccionar y actuar.

El origen de la especie data de hace 0,5 millones de años en Europa Central, desde donde se fue expandiendo hasta alcanzar su zona de distribución actual. En el caso de España, penetró por la zona pirenaica, llegando a la Cordillera Cantábrica y a los Sistemas Ibérico y Central [14]. Por restos fósiles, se puede datar que esa llegada al interior peninsular ocurrió hace 90.000 años. Estas poblaciones evolucionaron de forma aislada a las del resto de Europa, constituyendo un clado genético claramente diferenciado que constituye la actual subespecie *Microtus arvalis asturianus* (Miller, 1908).

Sin embargo, al menos en la historia más reciente, en Castilla y León, hasta finales de la década de los 70 del siglo pasado, el topillo campesino estaba circunscrito a las zonas montañas que bordean Castilla y León. Pero, a partir de entonces comenzó un rápido proceso de colonización hacia el interior de la región, donde se encuentran las principales zonas agrarias

de la comunidad autónoma, de tal manera que en los años 90 ya afectaba a prácticamente todas las comarcas agrarias [15].

En paralelo a este proceso de colonización, y hasta la plaga de 2006/08, se describió la sucesión de fenómenos de fluctuaciones demográficas típicos de esta especie a escala regional, estimándose que sucedieron con una ciclicidad de cinco años [15]. Sin embargo, a partir del evento de 2006/08, esa ciclicidad dejó de ser explicativa, pues desde entonces se han detectado estos fenómenos, a mayor o menor escala geográfica, en los años 2009/10, 2011/12, 2013/14, 2016/17, 2019, 2021 y, descrito por primera vez en el presente TFG, 2023/24.

Todos estos episodios desde el evento de 2006/08 han motivado, en mayor o menor medida, el impulso y coordinación de actuaciones por parte de las administraciones y los agricultores. Así, y sin entrar en una relación exhaustiva y detallada de las mismas, cabe citar que, como actuaciones administrativas paralelas a estos eventos, por la Orden AYG/556/2007, de 19 de febrero, se declaró oficialmente la existencia de plaga de topillo campesino en el territorio de Castilla y León. Por su parte, el Ministerio de Agricultura, Pesca y Alimentación aprobó el Real Decreto 409/2008, de 28 de marzo, por el que se calificaba de utilidad pública la prevención y lucha contra las plagas del topillo campesino y otros microtininos y se establecía el programa nacional de control de los mismos.

Posteriormente, por Acuerdo 53/2009, de 14 de mayo, la Junta de Castilla y León aprobó su Plan Director de lucha contra las plagas agrícolas, incluyéndose el topillo campesino como una de las plagas objetivo en los trabajos desarrollados en su marco. Por Acuerdo de 13 de febrero de 2014, de la Junta de Castilla y León, se aprobaron una serie de medidas fitosanitarias para la prevención y control del desarrollo de poblaciones de topillo campesino, que fueron revisadas y actualizadas posteriormente en el Acuerdo 78/2016, de 1 de diciembre, de la Junta de Castilla y León.

Como consecuencia y fruto de la experiencia de todo lo anterior, se evolucionó hacia la Orden AYG/96/2019, de 5 de febrero, que continúa en vigor y en la que se establece la estrategia de gestión integrada de riesgos derivados de la presencia de topillo campesino en el territorio de Castilla y León [16].

Dicha Orden establece una serie de actuaciones a desarrollar para reducir los riesgos derivados de la presencia del topillo campesino. Es en aplicación del tipo de actuación 4 establecido en la orden (investigación) donde se enmarca el interés del presente Trabajo Fin de Grado. Y para ello se utilizará información proveniente de la actuación 1, programa de monitorización que se ejecuta desde la actual Consejería de Agricultura, Ganadería y Desarrollo Rural con la coordinación del Observatorio de Vigilancia y Control de Plagas Agrícolas del ITACYL, donde el autor de este TFG ha realizado su estudio.

Volviendo a la relación de eventos demográficos acontecidos desde 2006/08, todos ellos afectaron a zonas pertenecientes a la comarca natural comúnmente conocida como Tierra de Campos, si bien no siempre se limitaron a ella, no pudiéndose declarar ninguna comarca agraria de Castilla y León como exenta de riesgos de incidencia por el topillo campesino. Al hablar de

Tierra de Campos nos referimos al área conformada por las comarcas agrarias de Pisuergra en la provincia de Burgos, Esla-Campos y Sahagún en la provincia de León, Campos en la provincia de Palencia, Tierra de Campos en la provincia de Valladolid y Campos-Pan en la provincia de Zamora, según establece la Orden AYG/96/2019, de 5 de febrero [16].

Si bien es cierto que, como se mencionaba anteriormente, se ha encontrado una componente cíclica en los brotes poblacionales estudiados en diversas regiones de Europa [17], cuando se intenta realizar una predicción dentro de los esquemas de la Gestión Integrada de Plagas y en una zona tan amplia como es la comunidad autónoma de Castilla y León, el problema se convierte en algo mucho más complicado de lo que podría parecer a priori.

En un primer momento, recordemos que los estudios realizados en la región revelaron una ciclicidad aparente de 5 años hasta la plaga de 2006/08 [15]. Sin embargo, estudios siguientes realizados a una escala más local, en Tierra de Campos, entre 2009 y 2017 encontraron un cambio a una periodicidad de tres años [18]. En añadido, los últimos estudios han probado que este resultado no es extrapolable a toda Castilla y León, detectándose variaciones en ciclicidad y periodo que dependen de la zona de estudio, con casuísticas que van de frecuencias de 2-3 años a frecuencias de 4-5 e incluso poblaciones que no mostraron comportamiento cíclico [19].

Si bien ocasionalmente a nivel europeo se ha observado cierta sincronización anual general en cuanto a brotes de topillo, por ejemplo, en 2019 [20], cuando se baja a una menor escala, incluso considerando zonas próximas dentro de Castilla y León con patrones cíclicos de frecuencias similares, la expresión de las fluctuaciones no tiene por qué guardar un patrón sincrónico. Tanto el inicio de la explosión como los máximos que puede alcanzar y su duración son distintos en tiempo y forma en las distintas zonas de Castilla y León. Esta complejidad en la expresión espacio-temporal de las tasas de crecimiento poblacionales del topillo campesino ha sido estudiada entre 2011 y 2017, encontrándose una explicación parcial mediante un modelo teórico/empírico complejo que implicaba un fenómeno de dos “ondas viajeras” con propagación radial a velocidades contrastantes [21]. Más tarde, estudios entre 2012 y 2022 usaron un enfoque diferente, basado en la detección y distribución en el espacio y el tiempo de “puntos calientes” y se identificaron evidencias de eventos claramente asincrónicos. Por ejemplo, mientras que en los episodios de 2013/14 y 2016/17 los principales puntos críticos se asociaron a zonas de Tierra de Campos en las provincias de Palencia, Valladolid y Zamora, en zonas periféricas de Burgos, Salamanca, Segovia o Soria sus puntos críticos se expresaron con posterioridad, en los años 2015 y 2018 respectivamente, cuando las abundancias en Tierra de Campos presentaban ya valores mínimos [22].

Si entramos ahora en el plano de la Gestión Integrada de este problema agrícola, lo primero a tener en cuenta es que no es una plaga agrícola “convencional”, en el sentido de que el agricultor pueda tener las herramientas necesarias para enfrentarse a ella. Esto ya está implícito en el Real Decreto 409/2008, de 28 de marzo [23], por el que se establece el programa nacional de control de las plagas del topillo campesino, en el que se indica que *“El control de las poblaciones de estos roedores reviste una especial dificultad debido a que hay que actuar preventivamente en*

situaciones y épocas en las que no existe una relación directa con los daños que aparecerán posteriormente, en predios sobre los cuales el agricultor no tiene directa responsabilidad . . . ”.

Esos predios a los que se refiere el Real Decreto 409/2008, sobre los que hay que actuar preventivamente y con antelación y sobre los que el agricultor no tiene competencia, son principalmente las vías de dispersión. Estudios recientes demuestran que la actuación en dichas vías asociadas a los márgenes de los campos de cultivo es fundamental a la hora de lidiar con esta plaga [7].

Las principales acciones que pueden considerarse en estas vías son su aprovechamiento para la instalación en estas entidades de estructuras relacionadas con el control biológico del topillo campesino, como cajas nido para rapaces (cernícalo vulgar, lechuza común, . . .) o posaderos. Si bien es indudable el interés de estas medidas relacionadas con el control biológico, su efecto es limitado y a largo plazo. Ante la amenaza de explosión demográfica se recomienda, mientras los cultivos adyacentes no estén desarrollados, la limpieza de su cubierta vegetal mediante desbroce, siendo la medida más adecuada el pase de cuchilla o motoniveladora, que complementa dicho desbroce con la destrucción mecánica de huras y galerías, siendo la época adecuada de plantear la actuación desde finales de verano hasta, a lo sumo, finales de invierno [24].

Llegados a este punto, surge uno de los grandes hándicaps a los que se enfrentan los agentes involucrados en la gestión del problema del topillo campesino: el impacto del topillo sólo puede reducirse a niveles tolerables cuando los gestores apliquen medidas preventivas y/o de control antes de que las poblaciones sean demasiado grandes; y esto en un escenario en el que, recordemos, debido a la explosividad en el crecimiento demográfico del topillo, la ventana de reacción es muy corta, cuando no inexistente.

Con independencia de las medidas que pueda tomar el agricultor en sus parcelas, a las que evidentemente también afecta esa ventana de reacción, sírvanos de ejemplo las actuaciones potenciales en las vías de dispersión. La responsabilidad y competencia de actuar en ellas no es del agricultor. La titularidad de estas vías suele recaer en organismos del tipo Ayuntamientos, Diputaciones Provinciales, Confederaciones Hidrográficas, responsables de carreteras, redes ferroviarias y similares.

A menudo, estos gestores tienen dificultades en aplicar las medidas por limitaciones de medios, de viabilidad en la ejecución en función del periodo del año o, incluso, ecológicas [25]. Siempre hace falta un periodo para la planificación, la implementación de los medios necesarios y el establecimiento de criterios de priorización. En añadido, cuando el agente responsable de la implementación de alguno de los medios es un organismo público, puede tener serias dificultades si no han sido previstos en las asignaciones presupuestarias correspondientes, normalmente cerradas antes de finales del año anterior.

Estas realidades, unidas a esa ventana de reacción habitualmente muy corta, conllevan que, para poder gestionar este problema, la principal herramienta necesaria es contar con mecanismos predictivos como apoyo esencial en los procesos de toma de decisiones. Mecanismos predictivos que solo serán realmente útiles cuando sean capaces de aumentar la ventana de reacción lo

suficiente como para cubrir los tiempos necesarios para la intervención (planificación, implementación de medios, . . .) [26]. Es decir, el objetivo real en un modelo predictivo no es sólo el buscar los mejores predictores, que también, si no que estos sean aplicables teniendo en cuenta los plazos de gestión. Y que, en añadido, permita tener un criterio de priorización sobre dónde y cuándo es realmente más necesario actuar.

Las fluctuaciones demográficas cíclicas del topillo campesino han sido objeto de diversos estudios previos, siendo descritas como fluctuaciones periódicas multianuales de densidad, caracterizadas por una denso-dependencia retardada en las tasas de crecimiento poblacional. Se han propuesto diferentes factores causales, conductores o moduladores que pueden influir en los brotes, incluyendo factores tanto intrínsecos a las poblaciones (genéticos, hormonales, estado sanitario, tasas reproductivas, . . .), como extrínsecos a las mismas (clima, cobertura vegetal, características paisajísticas, depredadores, disponibilidad de alimento, . . .) [26] [27] [28] [29] [30] [31].

Sin embargo, no existe un consenso científico sobre si los ciclos poblacionales en las distintas zonas donde suceden comparten o difieren en los mecanismos involucrados, o si implican interacciones entre distintos factores. Por lo tanto, para cada zona objetivo, en nuestro caso Castilla y León, es necesario proseguir con las investigaciones encaminadas a la identificación de factores impulsores y mecanismos moduladores involucrados en las fluctuaciones multianuales de abundancia de topillo campesino [17] [32] [33].

En el presente TFG abordaremos estas cuestiones para el caso de Castilla y León, mediante el estudio de variables/factores reportados como de interés potencial e involucrados con la denso-dependencia retardada y con factores extrínsecos poblacionales, como la distribución paisajística en tipos y cobertura del suelo, la edafología, la agronomía y las variaciones climatológicas [12] [13] [19] [26] [27] [34] [35] [36] [37].

En resumen, los procesos de toma de decisiones necesarios para la gestión integrada del topillo campesino pueden beneficiarse enormemente de estudios como el presente, orientados a la predicción de sus fluctuaciones demográficas permitiendo implementar medidas preventivas antes de que las poblaciones se vuelvan demasiado grandes o las enfermedades zoonóticas se propaguen demasiado.

Dado que se trata de un fenómeno complejo influido por múltiples factores, el desarrollo de un modelo predictivo para estas fluctuaciones demográficas supone un reto que requiere de un correcto uso de todos los métodos estadísticos pertinentes para lograr un resultado satisfactorio. Es importante, a su vez, una adecuada comprensión y manejo de los datos disponibles, tanto para obtener el mejor resultado posible como para lograr una interpretación acertada del mismo, que permita identificar los factores relacionados con la explosión poblacional del topillo. Será necesario estudiar la ciclicidad de los procesos de expansión-contracción descritos anteriormente, así como comprobar la relación entre las variables potencialmente predictoras de las que disponemos y estos procesos.

2. Objetivos

El objetivo global del presente TFG es contribuir a la mejora de las herramientas relacionadas con la Gestión Integrada de riesgos derivados de la presencia de topillo campesino en Castilla y León, en coherencia con lo establecido por la Orden AYG/96/2019, de 5 de febrero, en su capítulo II, apartado octavo (Investigación en alternativas aplicables a la gestión integrada de topillo campesino). Dentro de este marco, focalizando en los dos siguientes objetivos específicos:

1. Profundizar en el conocimiento de la variabilidad espacio-temporal de las fluctuaciones demográficas de topillo campesino en Castilla y León
2. Desarrollar una herramienta predictiva que permita una ventana temporal de reacción amplia, considerando como tal una anticipación de un año a la ocurrencia de fluctuaciones que desencadenen en explosiones demográficas con riesgo para los cultivos.

Relacionados con el objetivo 1, profundización en el conocimiento de la variabilidad espacio-temporal:

1. Exploración de la base de datos en cuanto a ocurrencia de eventos de explosiones demográficas significativas y obtención de evidencias empíricas sobre la existencia de variabilidad en la duración y magnitud de los eventos y la sincronía/asincronía entre zonas.
2. Profundización en la ciclicidad de la ocurrencia de explosiones demográficas y los diferenciales en la misma entre zonas, identificando la distribución geográfica de los patrones cíclicos encontrados.
3. Identificación preliminar de factores extrínsecos a las poblaciones de topillo, cuyos niveles de expresión estén potencialmente implicados en la modulación de su abundancia, considerando al respecto variables paisajísticas, agronómicas, edáficas y climáticas.

Relacionados con el objetivo 2, desarrollo de una herramienta predictiva:

1. Comparación de distintas metodologías para la definición de algoritmos predictivos de la variable dependiente, relacionada con las ratios de crecimiento, considerando como variables explicativas a los factores extrínsecos obtenidos, utilizando como entrenamiento un subset espacial que incluye el 80 % de las ZV en estudio (ZV Train) y un subset temporal (PER Train) que excluye al último año con datos disponibles (2024).
2. Selección y validación de la metodología y algoritmo que proporcione el mejor ajuste a la variable dependiente observada, considerando dos subset de prueba muestrales: el ajuste respecto a las ZV no utilizadas para el entrenamiento en los periodos de entrenamiento (TEST1) y el ajuste en cuanto a su aplicación predictiva considerando el conjunto de todas las ZV para los periodos no incluidos en el entrenamiento (TEST2).

3. Materiales y métodos

3.1. Área de estudio

Como breve contexto sobre la zona de estudio, la comunidad autónoma de Castilla y León está ubicada en la parte norte de la meseta Central de la Península Ibérica, y sus 9.422.600 ha de extensión la convierten en la comunidad más grande del país y en una de las mayores regiones administrativas de Europa [38]. Es de clima continental, inviernos fríos y veranos calurosos, con dos períodos generales húmedos en torno a la primavera y en la transición del otoño al invierno. El principal factor geomorfológico es el río Duero, que recorre toda la región de este a oeste por la zona central y crea una depresión principal donde prosperan cultivos de secano y de regadío en las terrazas fluviales. En cuanto al territorio agrícola, cuenta con una superficie arable (cultivable) de más de 3.500.000 ha (un 38 % de la superficie total de la comunidad), distribuida principalmente en la zona central, de las cuales en torno a 500.000 ha se cultivan en régimen de regadío. El 75 % de la superficie arable se dedica a los cultivos herbáceos, un 2,5 % a los cultivos leñosos y el resto a barbechos y tierras no ocupadas. Los cereales de invierno son el tipo de cultivo predominante, ocupando casi 2.000.000 de ha (el 71,6 % del terreno dedicado a los cultivos herbáceos), destacando la cebada y el trigo [39].

3.2. Datos y cálculo de variables

Los datos empleados en la realización de este Trabajo Fin de Grado han sido recogidos dentro del Programa de monitorización de riesgos derivados del topillo campesino [16] [40], que se ejecuta desde la Consejería de Agricultura, Ganadería y Desarrollo Rural de la JCyL, con la coordinación del Observatorio de Plagas del ITACYL. En líneas generales, una de las tareas que se realizan dentro de dicho programa es el seguimiento del estado de situación y evolución de las poblaciones de topillo campesino en el territorio agrícola de Castilla y León, mediante la evaluación espacio temporal de índices de abundancia de esta especie.

En concreto, la base de datos bruta utilizada tiene como una de sus variables el índice de abundancia de topillo campesino calculado para una unidad de inspección. Se considera unidad de inspección al resultado obtenido en una visita a un punto determinado, de forma que cada inspección está categorizada en función de la fecha y el recinto en que se realizó. Un recinto es la unidad mínima del territorio establecida según la división de este en función del SIGPAC (Sistema de Información Geográfico de la Política Agraria Comunitaria).

Cada inspección se asocia al tipo de hábitat presente en el recinto en el momento de la evaluación, categorizándose en general en cuatro tipos en función del uso y cobertura del suelo y su relación con la dinámica poblacional del topillo campesino:

- **Hábitat CULTIVO:** cultivos herbáceos anuales, que constituyen el principal hábitat a proteger de la incidencia del topillo.

- **Hábitat RESERVORIOS EN CULTIVO:** cultivos herbáceos plurianuales (en las zonas de estudio principalmente alfalfa), representativos de reservorios típicos de topillo desde los cuales pueden extenderse al resto del territorio, pero asimismo asociados directamente a parcelas cultivadas. La presencia en ellos de topillo puede suponer tanto afectación directa como posible fuente de dispersión futura.
- **Hábitat RESERVORIOS NO CULTIVO:** zonas no asociadas a cultivo que pueden constituir reservorios y fuente de dispersión futura, incluyéndose en esta categoría las praderas, perdidos, parcelas retiradas de cultivo y entidades asimilables.
- **Hábitat VÍAS DE DISPERSIÓN:** conjunto de entidades lineales, con características asimilables a reservorio e interconectadas entre sí, conformadas por cunetas, regatos, arroyos, linderas y entidades asimilables. En general, por las zonas adyacentes a infraestructuras de comunicación y de cauces de agua.

En nuestro estudio se han considerado las inspecciones realizadas en los hábitat Cultivo y Vías de Dispersión, este último por su importancia clave previamente reportada en la dinámica poblacional del topillo campesino [7]: proveen de un hábitat estable para sus colonias; representan los corredores que facilitan la dispersión del topillo; y constituyen el frente de ataque directo a las parcelas de cultivo en los momentos de explosión demográfica. Este tipo de hábitat ha sido considerado también en otros estudios previos realizados en Castilla y León sobre dinámica poblacional del topillo [19] [21] [22].

La escala temporal abarca un total de 13 años, desde 2012 hasta 2024. En la escala espacial se han considerado las evaluaciones en 40 zonas de vigilancia (ZV) establecidas en el mencionado Programa de Monitorización. Están distribuidas en 23 comarcas agrarias de las nueve provincias que forman Castilla y León (Tabla 24 y Figura 6 en el Anexo 7.1).

La variable del índice abundancia (Ia) de topillo campesino asociada a cada inspección se obtuvo mediante una metodología modificada a partir de la propuesta por Delattre y col. [41], habiendo sido calibrada y validada en distintos estudios [42] [43] [44]. Brevemente, consiste en la evaluación de transectos lineales de 99 metros de longitud, divididos en 33 unidades muestrales, cada una de $3 \times 1.5 \text{ m}^2$, en las se registra la presencia/ausencia de colonización con actividad reciente del topillo, entendiendo como tal: presencia de huras de topillo campesino asociadas a uno o varios de los indicios relacionados con presencia de excrementos frescos, materia vegetal fresca cortada en la boca de las huras o signos de actividad de excavación reciente.

El Ia para cada inspección se calcula como el porcentaje de unidades muestrales positivas respecto al total de unidades evaluadas, interpretándose como el porcentaje del terreno con colonización activa del topillo campesino.

A partir de la matriz bruta de datos se obtuvo la matriz de trabajo base para el presente estudio, concebida como una matriz de triple vía espacio-tiempo-hábitat, en el que el espacio está representado por las ZV. Para la estratificación temporal se considera la división de la

información en periodos cuatrimestrales, establecidos en conjunción de los estados de los cultivos más extendidos, en términos de provisión de cobertura vegetal, y con la dinámica reproductiva anual del topillo:

- El primer período, del 16 de febrero al 15 de junio, que abarca el final del invierno y la primavera, coincidiendo con el principal desarrollo vegetativo y reproductivo de los cultivos y con el aumento de las tasas de reproducción del topillo campesino. De aquí en adelante nos referiremos a este periodo como “primavera” (PRI en la codificación de determinadas variables).
- El segundo período, del 16 de junio al 15 de octubre, que abarca desde el verano hasta principios del otoño, coincidiendo con la cosecha de los cultivos y cuando la mayor parte del terreno está en rastrojo hasta la nueva siembra. De aquí en adelante nos referiremos a este periodo como “estío” (EST en la codificación de determinadas variables).
- El tercer período, del 16 de octubre al 15 de febrero, que abarca desde mediados del otoño hasta el invierno, coincidiendo con la época de siembra, el reposo vegetativo y la parada invernal de algunos cultivos, y los descensos en las tasas de reproducción del topillo. De aquí en adelante nos referiremos a este periodo como “frío” (FRI en la codificación de determinadas variables).

De esta forma, la matriz de trabajo consta de un total de 3120 celdas estructuradas en 40 ZV, 2 hábitat y series temporales de 39 periodos consecutivos (13 años * 3 periodos/año), desde la primavera de 2012 hasta el periodo frío de 2024.

Para cada celda se calculó la abundancia (VAI, Vole Abundance Index), como el promedio de los Ia asociados a cada ZV, hábitat y periodo. Posteriormente, de forma análoga a lo propuesto para estudios que trabajan con series temporales de este tipo de índices [26] [45] , el valor VAI se transformó por su logaritmo natural (LNVAI), siendo esta la variable a considerar en la matriz de abundancias final de trabajo. Para evitar valores de cero, sobre los que no se puede calcular el logaritmo, previo a la transformación se sumó al VAI la constante 3.03, correspondiente al valor no-cero más bajo que puede obtenerse a partir de la inspección de un transecto.

Sobre esta matriz se calculó una segunda, en este caso relativa a las ratios de crecimiento o variación temporal del LNVAI con los que identificar las fluctuaciones demográficas en las series temporales. En concreto, la tasa de crecimiento se ha calculado de forma análoga a trabajos previos [19] [21] tal y como se muestra a continuación:

$$GR_t = \log(VAI_t) - \log(VAI_{t-3})$$

Donde GR_t representa el growth rate o ratio de crecimiento en el periodo t, calculado respecto al valor de LNVAI de tres periodos previos, es decir, con una anterioridad de un año.

Por último, se dispone de una serie de variables concebidas en cuanto a su posible interés como conductores o moduladores de las fluctuaciones espacio-temporales demográficas del topillo, susceptibles de ser incluidas como covariables predictivas en los modelos objetivo. Estas variables se han obtenido para distintos ámbitos o tipos de información.

3.2.1. Variables denso-dependientes

Como variables denso-dependientes entendemos distintos rezagos del VAI (índice de abundancia) con una transformación logarítmica. Los retardos considerados son de 1, 2 y 3 unidades de periodo, además de retardos estacionales correspondientes a la primavera, estío y frío anteriores a cada periodo, dos estaciones atrás, y así sucesivamente hasta los cinco años, en función del máximo de periodicidad detectado en Castilla y León en estudios precedentes [4].

Además, en el caso de los cultivos, se han utilizado los retardos de la abundancia de las vías de dispersión además de los retardos de la propia.

3.2.2. Variables extrínsecas

Variables paisajísticas

Relacionadas con el tipo de cobertura del suelo. Se dispone de los mapas de cultivo y superficies naturales de Castilla y León (<https://mcsncyl.itacyl.es/descarga>) Estos mapas, desarrollados desde el ITACYL, han sido obtenidos por teledetección a través de imágenes de satélite con algoritmo de aprendizaje automatizado, y su objetivo es disponer de una cartografía anual de ocupación del suelo por los distintos cultivos y superficies de vegetación natural en Castilla y León.

Para el presente estudio se han considerado los mapas obtenidos desde 2017 a 2022, que usan imágenes Sentinel-2A y Sentinel-2B y mejoran la resolución de los años precedentes a un tamaño de pixel de 10m.

Sobre estos mapas se recortaron las ZV mediante herramientas SIG, utilizando el programa QGIS, considerando como superficie representativa de cada ZV un círculo de 5 km de radio y centro calculado como el promedio de las coordenadas X e Y de los centroides de los recintos inspeccionados incluidos en cada ZV. Se calculó la superficie dentro del círculo atribuida a cada tipo de cobertura incluida (los mapas identifican más de 120 clases de cobertura). Cada una de las clases se categorizó en función de su clasificación en condicionantes de hábitat diferenciales para el topillo campesino: cultivos herbáceos anuales, cultivos leñosos, reservorios asociados a cultivos herbáceos plurianuales, reservorios asociados a praderas, zonas arboladas y arbustivas, vías de dispersión, etc. Las covariables generadas fueron obtenidas como el promedio en los mapas considerados del porcentaje de superficie ocupado por cada categorización de las clases detectadas.

Para el caso de covariables relacionadas con caminos y cauces de agua, la resolución de pixel de 10 m es insuficiente para detectar adecuadamente sus márgenes, por lo que, en este caso, se

complementó con el recorte y solape sobre los mapas de cobertura de las capas relativas a usos CA (camino) y AG (masas de agua) disponibles en el SIGPAC.

Variables agronómicas

Son variables relacionadas con el modo de explotación de cada recinto de cultivo que, potencialmente, pueden influir en la dinámica poblacional del topillo. En este caso se han considerado dos: la relacionada con la forma de explotación entre secano y regadío y la relacionada con la preparación de la sementera, categorizando si se realizan labores de remoción de terreno (arado, gradeo, . . .) previas a la siembra o si se realiza considerando la siembra directa (sin laboreo previo).

Para el tipo de explotación secano-regadío, se ha utilizado el recorte de las ZV, de forma análoga a lo descrito para las covariables de cobertura del suelo, si bien esta vez sobre la capa SIGPAC de recintos de uso TA (tierra arable). La información disponible para cada recinto incluye el coeficiente de regadío (porcentaje de superficie regable), calculándose a partir de él la superficie regable de cada ZV.

En cuanto a las labores previas a la siembra, no existía una base de datos cartografiada susceptible de aportar esta información. En este caso, se ha calculado para cada ZV en función del porcentaje de recintos inspeccionados en los que el inspector, como parte de la categorización que acompaña a cada inspección, indicó en alguna ocasión condiciones de no laboreo o siembra directa.

Variables edáficas

Se dispone de mapas de clases de suelo en cuanto a variables edáficas, con ráster calculado por interpolación a partir de más de 28.000 muestras analizadas en Castilla y León (<https://suelos.itacyl.es/mapas>). Con metodología análoga a la realizada sobre los mapas de cobertura de suelo, se derivaron las covariables para las distintas ZV en función de la distribución por componentes texturales y propiedades hidráulicas.

Variables climáticas

Se cuenta con la información diaria provista por 149 estaciones climáticas, 52 pertenecientes a la red de Inforriego del ITACYL (<https://www.inforiego.org>) y 97 de la Agencia Estatal de Meteorología (AEMET, <https://www.aemet.es>). En concreto, se ha utilizado la información de precipitación y temperaturas media, mínima y máxima diarias. En función de ello, para cada ZV se seleccionaron entre cuatro y seis estaciones localizadas a una distancia máxima de 35 km respecto al centro de la ZV. Se estimaron los datos diarios atribuibles a cada ZV por la media ponderada de las distintas estaciones, considerando como ponderación la inversa de la distancia del centro de ZV a cada estación seleccionada. Con esta información, para cada combinación ZV * periodo se calcularon las covariables climáticas estacionales en función de los distintos factores calculados para distintas ventanas temporales. Se han calculado retardos

de estas variables de 1 a 6 periodos, además de retardos estacionales de forma análoga al caso de las variables denso-dependientes, correspondientes a la estación del año anterior y a la estación de dos años atrás.

En la Tabla 25 del Anexo 7.2 se puede consultar todas las variables y su respectiva descripción. En la Tabla 1 se puede ver el número de variables de cada tipo de las que se dispone.

Tabla 1: Número de variables según tipo y hábitat.

Tipo de covariable	Vías de dispersión	Cultivos
Denso-dependientes	18	36
Paisajísticas	5	
Agronómicas	2	
Edáficas	3	
Climáticas	32	
Total	60	78
Total (sin retardos)	14	

Estas variables han sido codificadas de forma que en algunos de los resultados aparecen con una letra como primer carácter, tal que se utiliza una D en el caso de las denso-dependientes, U para las paisajísticas, A para las agronómicas, E para las edáficas y C para las climáticas.

3.3. Software

Para la realización de la gran mayoría de los cálculos y gráficos se ha utilizado el software libre R [46] en su versión 4.2.2, con la excepción de algunas operaciones de manipulación de datos que fueron llevadas a cabo mediante el uso de hojas de cálculo. R es un lenguaje y entorno diseñado para la computación estadística y la generación de gráficos. Proporciona una amplia gama de herramientas para el manejo y análisis de datos, que, unido a la simplicidad y efectividad de su sintaxis [47], lo convierten en un instrumento ideal para la realización de trabajos como este. Además, es fácilmente extensible mediante la instalación de librerías específicas, razón por la cual es uno de los software más utilizados en el mundo de la estadística. Actualmente ocupa el puesto número 12 en el ranking de lenguajes de programación más populares, habiendo ascendido 12 posiciones en el último año [48].

Se ha hecho especial uso de los paquetes ggplot2 [49], dplyr [50], gridExtra [51], stringr [52], tidyr [53] y tidyverse [54], habiendo sido utilizados en la gran mayoría de los apartados. En la Tabla 27 del Anexo 7.3 se pueden consultar todo el resto de las librerías utilizadas a lo largo del trabajo.

3.4. Metodología estadística

3.4.1. Ciclos

A los efectos del presente trabajo, en la fase de exploración de la base de datos se ha considerado como ciclo a una serie de periodos consecutivos en la submatriz del hábitat vías de dispersión, que cumplan: se inicie en un cambio de un periodo con baja abundancia a uno con abundancia significativa, y debe prolongarse en el tiempo al menos otro periodo más por encima de dicho valor; finalice cuando ocurra un decrecimiento de la abundancia por debajo de la abundancia considerada como significativa y se prolongue al menos otro periodo por debajo de la misma. A los efectos, se ha considerado como significativo un nivel de abundancia con un VAI superior a 6.06, de acuerdo a estudios previos [44], equivalente a un LNVAI de 2.2172 en la matriz de trabajo.

3.4.2. Series temporales

Una serie temporal X_t es una secuencia ordenada de variables aleatorias, donde t es el índice de tiempo tal que $t \in T = \{\dots, -n, \dots, -1, 0, 1, \dots, n, \dots\}$ [55]. Los valores de la serie que han sido observados se denominan como muestra o realización y se representa como $\{x_{1:n}\} = \{x_1, x_2, \dots, x_n\} = \{x_t; t = 1, 2, \dots, n\}$. Las funciones de distribución para cada n en instantes $t_1, t_2, \dots, t_n$ se puede escribir como:

$$F_{1:n}(x_{1:n}) = F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) = \Pr(X_{t_1} \leq x_1, X_{t_2} \leq x_2, \dots, X_{t_n} \leq x_n)$$

Son de gran utilidad los momentos de primer y segundo orden de estas distribuciones, las funciones media, varianza, autocovarianza y autocorrelación:

- La función media de una serie temporal se define como el valor esperado en el instante t , es decir: $\mu_t = E(X_t)$.
- La función de varianza se define como la esperanza del cuadrado de las desviaciones respecto a la media: $\sigma_t^2 = Var(X_t) = E[(X_t - \mu_t)^2]$. A partir de esta función podemos obtener también la función de desviación estándar calculando su raíz cuadrada, tal que $\sigma_t = \sqrt{\sigma_t^2}$.
- La función de autocovarianza entre los instantes s y t , que representa la covarianza entre dichos instantes, se expresa como:

$$\gamma(s, t) = Cov(X_s, X_t) = E[(X_s - \mu_s)(X_t - \mu_t)]$$

- La función de autocorrelación, que se deriva de la de autocovarianza:

$$\rho(s, t) = Corr(X_s, X_t) = \frac{\gamma(s, t)}{\sigma_s \sigma_t} = \frac{Cov(X_s, X_t)}{\sqrt{Var(X_s) Var(X_t)}}$$

Estas dos últimas miden la correlación lineal entre los instantes s y t de la misma serie, con la diferencia de que la autocorrelación es adimensional, acotada en el intervalo $[-1, 1]$ y por tanto es más sencilla de interpretar y manejar.

Autocorrelograma

En la práctica, todas estas funciones son desconocidas y por tanto deben ser estimadas a partir de los datos disponibles. La función de autocovarianza muestral, que mide la relación lineal entre una serie temporal y sus respectivos retardos h , se define como:

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (x_t - \bar{x})(x_{t+h} - \bar{x}), \text{ para } h \geq 0$$

Donde $\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t$ es la media muestral. Partiendo de esta función, es sencillo calcular la función de autocorrelación muestral o ACF: $\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$.

La representación gráfica de esta función evaluada en los retardos $h = 1, 2, \dots, k$ frente a dichos retardos se conoce como autocorrelograma. Es una herramienta muy útil para detectar patrones de dependencia temporal.

Periodograma

El periodograma es una función que nos permite expresar una serie temporal como una combinación de ondas seno y coseno con diferentes periodos y amplitudes [56]. Entendemos por periodo el tiempo que toma la serie en completar un ciclo, y por amplitud la altura absoluta máxima de la serie. Sirve para detectar los periodos importantes de la serie y así comprender mejor el patrón oscilatorio de los datos observados.

Sea A la amplitud, que determina la máxima altura absoluta de la serie, ω la frecuencia, que controla cómo de rápido oscila la curva y ϕ la fase, que determina el punto de partida en grados para la función coseno, ajustamos una onda coseno a la serie temporal:

$$x_t = A \cdot \cos(2\pi\omega t + \phi)$$

Teniendo en cuenta la siguiente identidad trigonométrica, donde $\beta_1 = A \cdot \cos(\phi)$ y $\beta_2 = -A \cdot \sin(\phi)$:

$$A \cdot \cos(2\pi\omega t + \phi) = \beta_1 \cdot \cos(2\pi\omega t) + \beta_2 \sin(2\pi\omega t)$$

Sean ω_j las frecuencias armónicas, tal que $\omega_j = j/n$ para $j = 1, 2, \dots, n/2$, la serie temporal se puede escribir como:

$$x_t = \sum_{j=1}^{n/2} \left[\beta_1 \left(\frac{j}{n} \right) \cos(2\pi\omega_j t) + \beta_2 \left(\frac{j}{n} \right) \sin(2\pi\omega_j t) \right]$$

Podemos entender $\beta_1 \left(\frac{j}{n} \right)$ y $\beta_2 \left(\frac{j}{n} \right)$ como parámetros de regresión, que se calculan mediante la Transformada rápida de Fourier (FFT). Una vez los parámetros han sido estimados, definimos:

$$P \left(\frac{j}{n} \right) = \hat{\beta}_1^2 \left(\frac{j}{n} \right) + \hat{\beta}_2^2 \left(\frac{j}{n} \right)$$

Esta es la función que nos proporciona el valor del periodograma en la frecuencia j/n . Es decir, el periodograma es un gráfico de $P \left(\frac{j}{n} \right)$ evaluado en las frecuencias armónicas, es decir, con $j = 1, 2, \dots, n/2$ frente a dichas frecuencias armónicas. Valores altos de esta función indican una gran importancia de la frecuencia en la que se encuentra dicho valor a la hora de explicar la oscilación de la serie temporal.

En el presente TFG se ha utilizado como principal criterio para la decisión de que periodo/ciclicidad se atribuirá a cada ZV la información de los periodogramas, en concreto considerando el pico más acusado, dando prioridad al de vías de dispersión, al ser el hábitat donde más se expresan las posibles fluctuaciones. La información de los autocorrelogramas y la propia serie se han utilizado como apoyo en casos como la aparición de un segundo pico que pudiera generar dudas al tener en cuenta la información en uno de los hábitats o la conjunta entre ambos. La existencia de estos segundos picos puede generar cierta incertidumbre a la hora de clasificar, y serán comentados en algunos casos particulares, si bien para el resto del estudio se proseguirá considerando el criterio basado en la selección del pico más alto definido en el periodograma.

3.4.3. Regresión lineal múltiple

La regresión lineal es un método estadístico que se utiliza para modelar la relación entre una variable dependiente y una o más variables independientes. Fue desarrollada por Sir Francis Galton a finales del siglo XIX mientras estudiaba la relación entre las alturas de padres y sus respectivos hijos [57]. El objetivo de este modelo es describir como varía la variable dependiente en función de las variables independientes mediante una ecuación lineal con la siguiente forma [58]:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

Donde:

- Y es la variable dependiente o respuesta.
- $X_1, X_2, \dots, X_p$ son las variables independientes o explicativas.
- β_0 es el intercepto o término independiente.
- $\beta_1, \beta_2, \dots, \beta_p$ son los coeficientes correspondientes a cada variable independiente.
- ε es el error aleatorio, con media $E(\varepsilon) = 0$ y varianza $Var(\varepsilon) = \sigma^2$. Los errores para dos observaciones distintas, ε_i y ε_j , son incorrelados, luego su covarianza es 0, $Cov(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i, j / i \neq j$

- p es el número de variables independientes. Cuando $p > 1$, estamos ante una regresión múltiple.

Esto implica que la variable respuesta Y viene de una distribución de probabilidad tal que:

$$\begin{aligned} E(Y) &= E(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon) \\ &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + E(\varepsilon) \\ &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p, \end{aligned}$$

$$Var(Y) = Var(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon) = Var(\varepsilon) = \sigma^2,$$

$$y_i \text{ e } y_j \text{ son incorrelados } \forall i, j / i \neq j.$$

Nuestro objetivo es obtener estimadores $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ para los coeficientes del modelo de regresión de forma que se ajuste lo mejor posible a los datos de los que disponemos, tal que $y_i \approx \hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \hat{\beta}_2 x_{2,i} + \cdots + \hat{\beta}_p x_{p,i}$ para $i = 1, \dots, n$, donde n es el número de observaciones. Aunque hay varias formas de medir cómo de bueno es este ajuste, la más habitual y la que usaremos en el presente trabajo pasa por minimizar el criterio de mínimos cuadrados. Sea $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \hat{\beta}_2 x_{2,i} + \cdots + \hat{\beta}_p x_{p,i}$ la predicción de Y para la i -ésima observación de nuestro conjunto de datos, definimos el residuo de dicha observación como la diferencia entre el valor real y el valor predicho, es decir, $e_i = y_i - \hat{y}_i$. Por tanto, la suma de cuadrados residual se calcula como $RSS = e_1^2 + e_2^2 + \cdots + e_n^2 = \sum_{i=1}^n e_i^2$, que es el valor que intentamos minimizar al calcular los estimadores de los coeficientes. Sea X la matriz de diseño del modelo, es decir:

$$X = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,p} \end{pmatrix}$$

y $\hat{\beta}$ el vector con todos los coeficientes: $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)$, la estimación de los coeficiente mediante este método de puede escribir de forma matricial de la siguiente manera:

$$\hat{\beta} = (X^t X)^{-1} X^t Y$$

Además, podemos calcular la varianza de los coeficientes tal que: $Var(\hat{\beta}) = \sigma^2 (X^t X)^{-1}$

Suposiciones

Hasta el momento se ha hablado de un modelo de regresión lineal múltiple sin especificar la distribución de los errores. Sin embargo, si queremos construir intervalos y realizar contrastes de hipótesis, es necesario asumir la normalidad de los errores [57], tal que:

$$\varepsilon_i \sim N(0, \sigma^2), \quad \varepsilon_i \text{ independiente de } \varepsilon_j \quad \forall i, j / i \neq j$$

Gracias a la normalidad, la asunción del modelo anterior de incorrelación entre los errores se convierte en una de independencia. Esto tiene como consecuencia que la variable respuesta Y también sigue una distribución normal:

$$y_i \sim N(\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \dots + \beta_p x_{p,i}, \sigma^2), \quad y_i \text{ independiente de } y_j \forall i, j / i \neq j$$

Esta asunción de normalidad se justifica considerando los errores como los efectos de factores externos al modelo, que varían con independencia de las variables explicativas. Además, puede haber errores aleatorios en la medición de la variable respuesta Y que contribuyan a estos errores. Dado que los errores del modelo representan la influencia agregada de múltiples factores aleatorios no observados e independientes, es razonable suponer que, según el Teorema Central del Límite, su distribución tienda a aproximarse a una distribución normal cuando el número de efectos considerados es grande.

Las suposiciones realizadas sobre este modelo se pueden resumir en las siguientes:

- Linealidad de la relación entre la variable dependiente y las independientes.
- Independencia entre errores.
- Homocedasticidad o igualdad de la varianzas.
- Normalidad de los errores.

Aunque en la práctica es común que una o varias de estas suposiciones no se cumplan, es necesario tenerlas en cuenta para una correcta interpretación de los resultados arrojados por el ajuste del modelo.

Hay distintos métodos de evaluación de dichas suposiciones, a destacar:

- Para la normalidad de los errores, tests de normalidad como el de Lilliefors [59], con hipótesis nula H_0 : los datos provienen de una población con distribución normal, y gráficos QQ de normales [60].
- Para la homocedasticidad, los tests de Breusch-Pagan [61] y de varianza no constante [62], que comparten hipótesis nula de varianza constante.
- Para la independencia entre errores en datos temporales, el test de Durbin-Watson [63], con las hipótesis:

H_0 : No hay autocorrelación de primer orden en los errores

H_1 : Hay autocorrelación de primer orden positiva o negativa

- Para la independencia espacial entre errores, el test de la I de Moran [64], cuya hipótesis nula se puede escribir como:

$$H_0 : \text{No hay autocorrelación espacial}$$

Intervalos de confianza y contrastes de hipótesis sobre los parámetros

Al hacer inferencia sobre los parámetros del modelo es de gran interés, como al realizar cualquier estimación, proporcionar un intervalo de confianza que acompañe a la estimación puntual. Recordamos el cálculo de la varianza de los estimadores de los coeficientes: $Var(\hat{\beta}) = \sigma^2(X^t X)^{-1}$

Es importante recalcar que para que la fórmula anterior sea estrictamente válida, la homocedasticidad e independencia de los errores son necesarias [58]. Aun así, en el caso de que no se den estas condiciones puede seguir siendo una buena aproximación. Además, se puede aplicar la corrección de White si se sospecha heterocedasticidad, para ajustar las estimaciones de la varianza de los coeficientes y obtener estimaciones válidas [65]. En concreto, con la especificación HC3:

$$\widehat{Var}_{HC3}(\hat{\beta}) = (X^t X)^{-1} \left(\sum_{i=1}^n \frac{e_i^2}{(1 - h_{ii})^2} x_i x_i' \right) (X^t X)^{-1}$$

En nuestro caso, dado que los datos están agrupados por zonas, resulta especialmente pertinente el uso de métodos para la estimación de errores estándar robustos por conglomerados, conocidos como CRVE (Cluster-Robust Variance Estimators o errores estándar robustos por agrupamiento). Estos métodos permiten corregir tanto la no independencia de los errores como la posible heterocedasticidad, y existen distintas variantes disponibles en la literatura [66]. En particular, se emplea la variante CR2, que corresponde a una forma de reducción del sesgo en la linealización, conocida como BRL (Bias-Reduced Linearization).

A partir de esto y después de estimar σ^2 con s^2 , podemos obtener el error estándar, tal que:

$$SE(\hat{\beta}_j) = \sqrt{s^2 \cdot [(\mathbf{X}^\top \mathbf{X})^{-1}]_{jj}}$$

Un intervalo de confianza del $100 \cdot (1 - \alpha) \%$ se calcula como:

$$\hat{\beta}_j \pm t_{n-p-1, 1-\frac{\alpha}{2}} \cdot SE(\hat{\beta}_j)$$

Donde $t_{n-p-1, 1-\frac{\alpha}{2}}$ es el cuantil $1 - \frac{\alpha}{2}$ de la distribución t de Student con $n - p - 1$ grados de libertad. Este intervalo indica que, con probabilidad $100 \cdot (1 - \alpha) \%$, el valor real del coeficiente $\hat{\beta}_j$ se encuentra dentro de los límites calculados. Si los errores no siguen una distribución normal, este intervalo no será del todo preciso. En caso de corregir los errores utilizando errores estándar robustos, los grados de libertad $n - p - 1$ no reflejan bien la incertidumbre real, y por tanto es conveniente usar el ajuste de Satterthwaite [67].

Hay dos tipos de contrastes de hipótesis que nos interesa realizar sobre los parámetros de nuestro modelo. En primer lugar, contrastes individuales para los coeficientes del modelo,

correspondientes a los intervalos de confianza ya mencionados. Son conocidos como test t. En estos contrastes, las hipótesis se definen como sigue [58]:

$$H_0 : \beta_j = 0 \quad \text{vs.} \quad H_1 : \beta_j \neq 0$$

La hipótesis nula podría transcribirse como “La variable predictora x_j no tiene efecto sobre la variable respuesta una vez controladas las demás variables independientes”. El estadístico de contraste se calcula como:

$$t = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \sim t_{n-p-1} \text{ bajo } H_0$$

Rechazamos la hipótesis nula a un nivel de confianza del $100 \cdot (1 - \alpha) \%$ cuando:

$$|t| > t_{n-p-1, 1-\frac{\alpha}{2}}$$

Al igual que en el caso de los intervalos de confianza, estos contrastes dependen de las suposiciones del modelo lineal normal. Además de contrastes individuales, podemos realizar contrastes conjuntos para varios coeficientes, conocido como test F. En este caso las hipótesis del modelo son:

$$H_0 : \beta_{j_1} = \beta_{j_2} = \dots = \beta_{j_q} = 0 \quad \text{vs.} \quad H_1 : \exists k / \beta_{j_k} \neq 0$$

Este contraste se realiza ajustando un nuevo modelo que usa todas las variables menos las especificadas en la hipótesis anterior, siendo q este número de variables. Sea RSS_0 la suma de cuadrados de los residuos de este nuevo modelo, el estadístico F que utilizamos para el contraste se calcula como:

$$F = \frac{(RSS_0 - RSS)/q}{RSS/(n-p-1)} \sim F_{q, n-p-1}$$

Cuando $q=1$, este test es equivalente al test t. Otro caso concreto de este contraste ocurre cuando se desea comprobar si al menos un predictor tiene un efecto significativo, dando lugar al conocido como contraste de la regresión o test F. Las hipótesis del contraste pasan a ser:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0 \quad \text{vs.} \quad H_1 : \exists j / \beta_j \neq 0$$

Y el estadístico F:

$$F = \frac{(TSS - RSS)/p}{RSS/(n-p-1)} \sim F_{p, n-p-1}$$

Si las asunciones del modelo lineal normal se cumplen, se puede demostrar que:

$$E \left[\frac{RSS}{n-p-1} \right] = \sigma^2 \quad \text{y} \quad E \left[\frac{TSS - RSS}{p} \right] = \sigma^2 \quad \text{bajo } H_0$$

Por tanto, cuando no existe relación entre la respuesta y los predictores, se espera que

el estadístico F tome un valor próximo a 1. Incluso en el caso de que los errores no estén normalmente distribuidos, este estadístico sigue aproximadamente una distribución F, siempre que el número de observaciones n sea grande.

Coefficiente de determinación y coeficiente de determinación ajustado

El estadístico R^2 , o coeficiente de determinación, es una medida que indica la proporción de varianza total de la variable respuesta que explica un modelo [58]. Es uno de los criterios más usados para evaluar el ajuste en modelos de regresión lineal, debido a la sencillez tanto de su cálculo como de su interpretación, ya que su valor siempre se encuentra entre 0 y 1, de tal forma que cuanto más próximo a 1, mejor es el ajuste, pues explica un mayor porcentaje de la variabilidad de la variable respuesta. Se calcula de la siguiente mediante la siguiente fórmula:

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

Donde $TSS = \sum_{i=1}^n (y_i - \bar{y})^2$, es la suma total de cuadrados, que representa la varianza total en la variable dependiente. RSS, anteriormente definida, se puede entender como la varianza que queda sin explicar tras la regresión, por lo que a menor varianza sin explicar, mayor coeficiente de determinación. Dado que este coeficiente incrementa su valor al incluir variables en el ajuste, no es un buen criterio de comparación entre modelos en el caso de que no todos tengan el mismo número de variables. Es por ello por lo que surge la necesidad de realizar un pequeño ajuste a este estadístico, penalizando por el número de variables, resultando en el conocido como coeficiente de determinación ajustado. Recordando que p es el número de variables independientes, se define como:

$$R_a^2 = 1 - \frac{RSS/(n - p - 1)}{TSS/(n - 1)}$$

Tiene la misma interpretación que el estadístico R^2 : cuanto más próximo a 1, mejor es el ajuste, con la diferencia de que puede llegar a ser negativo. Es importante recalcar que, para que tanto el coeficiente de determinación como el coeficiente de determinación ajustado tengan esta interpretación, debe haber un intercepto o término independiente en el modelo.

No obstante, hay que ser consciente de las limitaciones de este estadístico, ya que un coeficiente alto no significa que se puedan realizar predicciones útiles o precisas, ni que la regresión estimada sea un buen ajuste, en el sentido de que tal vez la relación no sea lineal sino que tenga otra forma [57]. Además, es por este mismo motivo que un R^2 próximo a cero no indica que no haya relación entre las variables.

MSE y RMSE

El MSE o error cuadrático medio es otra medida utilizada para evaluar el ajuste del modelo a

los datos con los que se ha desarrollado [41]. Se calcula como:

$$s^2 = MSE = \frac{SSE}{n - p - 1} = \frac{1}{n - p - 1} \sum_{i=1}^n (y - \hat{y}_i)^2 = \frac{1}{n - p - 1} \sum_{i=1}^n e_i^2$$

Donde s^2 es la varianza muestral de los errores del modelo. La corrección de dividir entre $n - p - 1$ en vez de simplemente entre n , se debe al hecho de que en regresión, el error cuadrático medio es utilizado para estimar la varianza de los errores, σ^2 , y es necesario restar los grados de libertad que se pierden al estimar los $p + 1$ parámetros [57]. Se puede demostrar que s^2 es un estimador insesgado de σ^2 , es decir:

$$E(s^2) = E(MSE) = \sigma^2$$

Y por tanto un estimador de σ será:

$$s = \sqrt{MSE} = RMSE$$

A pesar de que el estimador $RMSE$ no es insesgado para σ , sí es consistente, lo que significa que converge en probabilidad a σ a medida que $n \rightarrow \infty$.

Factor de inflación de la varianza

La multicolinealidad en un modelo de regresión lineal ocurre cuando algunas de las variables independientes están relacionadas entre sí [58]. Esto puede suponer un problema para determinar la importancia relativa de cada uno de los predictores y para comprobar cuánto afectan a la variable respuesta, ya que causan un crecimiento en el error estándar de los coeficientes. Por tanto es necesario buscar una solución a este problema cuando se ajustan modelos lineales.

El VIF (variance inflation factor) o factor de inflación de la varianza es una medida de la colinealidad. Se calcula como:

$$VIF(\hat{\beta}_j) = \frac{1}{1 - R_j^2}$$

Donde R_j^2 es el coeficiente de determinación resultante del modelo de regresión de la variable X_j sobre el resto de predictores. Como regla general, se suele considerar que un VIF mayor de 5 o 10 es una señal clara de multicolinealidad.

Algoritmos automáticos de selección de variables

A la hora de ajustar modelos en los que intervienen muchas variables predictoras, uno de los principales desafíos consiste en hacer una selección de las variables que son relevantes en el ajuste, especialmente cuando el número de variables es elevado, ya que actuar puramente en base a los p-valores individuales obtenidos de los test t puede ser incorrecto [58]. De esta necesidad y ante la imposibilidad de probar todos los modelos posibles con las variables disponibles surgen

los algoritmos automáticos de selección de variables, destacando los métodos de regresión stepwise o escalonada.

La regresión stepwise consiste en añadir o eliminar variables predictoras al modelo secuencialmente en base a algún criterio estadístico, como puede ser el AIC (criterio de información de Akaike), BIC (criterio de información de Bayes), R_a^2 , etc. En nuestro caso usaremos los p-valores de los test F, eligiendo previamente un p-valor de entrada y/u uno de salida. Existen principalmente tres versiones, que se describen con sus correspondientes pasos:

1. Selección hacia adelante o *forward*.

1.1. Se ajustan todos los modelos posibles con una sola variable.

1.2. Se identifica aquella variable con un menor p-valor en el contraste F:

$$H_0 : \beta_k = 0 \quad vs. \quad H_1 : \beta_k \neq 0$$

1.3. Si dicho p-valor es menor que el p-valor de entrada, se reajusta el modelo añadiendo la variable correspondiente y se vuelve a 1.2.

1.4. Si el p-valor es mayor, finaliza el algoritmo.

2. Selección hacia atrás o *backward*.

2.1. Se ajusta el modelo con todas las variables.

2.2. Se identifica aquella variable con un mayor p-valor en el contraste F:

$$H_0 : \beta_k = 0 \quad vs. \quad H_1 : \beta_k \neq 0$$

2.3. Si dicho p-valor es mayor que el p-valor de salida, se reajusta el modelo eliminando la variable correspondiente y se vuelve a 2.2.

2.4. Si el p-valor es menor, finaliza el algoritmo.

3. Selección híbrida, paso a paso o *stepwise*.

3.1. Se ajustan todos los modelos posibles con una sola variable.

3.2. Se identifica aquella variable con un menor p-valor en el contraste F:

$$H_0 : \beta_k = 0 \quad vs. \quad H_1 : \beta_k \neq 0$$

3.3. Si dicho p-valor es menor que el p-valor de entrada, se reajusta el modelo añadiendo la variable correspondiente.

3.1. Se identifica aquella variable con un mayor p-valor en el contraste F:

$$H_0 : \beta_k = 0 \quad vs. \quad H_1 : \beta_k \neq 0$$

- 3.2. Si dicho p-valor es mayor que el p-valor de salida, se reajusta el modelo eliminando la variable correspondiente y se vuelve a 3.1.
- 3.3. Si el p-valor es mayor se vuelve a 3.2.
- 3.4. Si el p-valor es mayor finaliza el algoritmo.

3.4.4. División entrenamiento-test

El sobreajuste, también conocido como overfitting, ocurre cuando un modelo no generaliza bien de los datos observados los no observados, es decir, tiene un buen rendimiento sobre el conjunto de datos con el que se ha entrenado, pero resulta en un ajuste deficiente cuando se prueba sobre un nuevo conjunto de observaciones [68]. Esto se debe a que el modelo puede llegar a entrenarse de una manera extremadamente compleja dando como resultado una varianza alta y un sesgo bajo [69]. Este es uno de los problemas más fundamentales y recurrentes del ajuste de modelos lineales, ya que aparte del ya mencionado inconveniente de la excesiva complejidad, también conlleva un aumento de coste computacional e inestabilidad [70].

Esto resulta en la necesidad de adoptar una estrategia para hacer frente a esta situación, que pasa por desarrollar algoritmos que penalicen el exceso de variables y elegir el mejor en base a su rendimiento sobre conjuntos de datos externos al conjunto con el que se ha creado el modelo. En el presente estudio se ha considerado la siguiente división de los datos en conjuntos de entrenamiento y test:

- Conjunto TRAIN: son los datos que se van a utilizar para ajustar (entrenar) los modelos de regresión. Está formado por un conjunto del 80 % de las ZV en estudio para los periodos 4 a 36 (excluyendo los tres periodos de 2024), lo que resulta en un total de 32 zonas y 33 periodos para entrenar los algoritmos.
- Conjunto TEST1: es un subconjunto de los datos originales compuesto por el 20 % de las ZV no utilizadas en el entrenamiento, contemplando los periodos de entrenamiento, para cotejar la capacidad de los algoritmos sobre datos zonales no observados en el intervalo de tiempo entrenado, resultando en un total de 8 zonas y 33 periodos.
- Conjunto TEST2: es el subconjunto de datos orientado a cotejar la capacidad de ajuste en cuanto a predicción temporal no observada, el verdadero objetivo del modelo a seleccionar, contemplando todas las ZV para los periodos no utilizados para el entrenamiento (año 2024), resultando en un total de 40 ZV y 3 periodos., total de las ZV un subconjunto de los datos originales compuesto por los datos correspondientes a los tres últimos periodos, correspondientes al año 2024.

La elección de las ZV que forman parte del conjunto TEST1 se realizó de acuerdo a un muestreo aleatorizado con el único requisito de que no contuviese más de una ZV por comarca agraria.

4. Resultados y discusión

4.1. Exploración de la base de datos

La exploración que se plantea en este apartado tiene como objetivo la detección de generalidades en la relación entre las abundancias entre hábitats, la racionalización de eventos de fluctuaciones de abundancia detectados y su comparación básica en cuanto a sincronía/asincronía en el inicio y su duración.

En la Figura 7 del Anexo 7.4 se representan como mapas de calor las abundancias (LNVAI) de la matriz de partida de triple vía zona/periodo/hábitat.

En una primera inspección meramente visual parece intuirse que la expresión de la abundancia tiende a ser mayor en el hábitat vías de dispersión que en los cultivos, lo cual se corrobora cuando se comprueba que de las 1560 combinaciones posibles ZV*periodo, en 1421 casos (91.09 %) el valor de abundancia es mayor en las vías de dispersión. En 111 casos (7.12 %) los valores fueron iguales, si bien todos ellos son combinaciones en las cuales tanto vías de dispersión como cultivos tienen valor de cero en la abundancia. Tan sólo en 28 casos (1.79 %) el valor de las vías es menor que el de cultivo, pero siempre involucrando valores de abundancia por debajo del considerado como identificativo de colonización significativa, en función de la calibración del índice de abundancia utilizado [44]. Esto corrobora lo expuesto en trabajos anteriores en cuanto a que las vías de dispersión constituyen un hábitat reservorio para el topillo en que, y con independencia de la expresión de explosiones demográficas, las abundancias en él tienden a ser superiores que en las parcelas adyacentes, con independencia del tipo de cultivo en dichas parcelas [7].

En la Figura 7 del Anexo 7.4, visualmente parece intuirse la ocurrencia de eventos de fluctuaciones demográficas con posibles diferenciales en la expresión dependiendo periodos, ZVs y hábitats. Esto queda más claramente identificado cuando se infieren los ciclos con los criterios descritos en el apartado correspondiente de materiales y métodos, cuyos resultados aparecen reflejados en la Figura 8 (Anexo 7.4) y en la Tabla 28 (Anexo 7.5).

De esta forma, se han detectado un total de 6 eventos generales de incrementos demográficos desde 2013. Cabe mencionar que los datos de 2012 también encierran compatibilidad con ocurrencia de eventos, si bien estos habían comenzado en algunos casos con anterioridad, por lo que no se han utilizado para este apartado. Para facilitar la comprensión, se ha considerado una denominación para los eventos generales en función del año en que se detectó mayor número de inicios de su expresión en las vías de dispersión de las distintas ZV involucradas para cada caso. De esta forma, con independencia de su duración o posible expresión anterior o posterior en alguna zona, los eventos generales se denominarán 2013, 2015, 2016, 2019, 2021 y 2023.

En las Tablas 3, para el hábitat vías de dispersión, y 4, para cultivos, se resume la información básica referente a estos ciclos. En la Tabla 3 (vías) para cada evento general se incluye: su inicio, como el primer periodo en que se detectaron incrementos demográficos en al menos una ZV; en

cuantas ZV se detectaron los incrementos; la duración media, mínima y máxima, en número de periodos, del evento en función de lo observado en las ZV involucradas; la asincronía promedio y máxima detectada, en cuanto a periodos de retardo con que se expresó en las ZV involucradas respecto al inicio del evento general. En la Tabla 4 (cultivos), se incluye la misma información, si bien en este caso la asincronía incluye la mínima, media y máxima respecto al momento en que los incrementos se expresaron en las vías para las ZV involucradas.

En la Tabla 5 se incluye la información básica en cuanto a número de eventos detectados en cada ZV particular, pudiendo consultarse la información detallada referente a inicio en cada una, hábitats en que se detecta, duración o asincronía en la Tabla 28 (Anexo 7.5).

Tabla 3: Información general sobre cada evento en vías de dispersión

Evento	Inicio	Nº ZVs	Duración			Asincronía	
			Media	Mínima	Máxima	Media	Máxima
2013	13_PRI	38	4.76	2	7	1.42	4
2015	15_EST	1	2.00	2	2	0.00	0
2016	16_PRI	28	4.07	2	6	1.36	4
2019	18_FRI	36	5.19	2	11	2.17	5
2021	21_PRI	9	2.29	2	3	0.86	1
2023	23_PRI	37	3.53	2	5	2.32	5

Tabla 4: Información general sobre cada evento en cultivos

Evento	Inicio	Nº ZVs	Duración			Asincronía		
			Media	Mínima	Máxima	Media	Mínima	Máxima
2013	13_PRI	18	1.83	1	4	1.28	0	3
2015	15_EST	0	-	-	-	-	-	-
2016	16_PRI	10	2.40	1	4	1.10	0	3
2019	18_FRI	18	1.22	1	2	1.44	0	4
2021	21_PRI	4	1.25	1	2	0.50	0	1
2023	23_PRI	23	1.43	1	3	1.39	0	3

De los seis eventos mencionados, lo primero a reseñar es que el de 2015 tan sólo se detectó en una de las ZV (SE.02_CU), teniendo además una duración muy corta (dos periodos) y sin expresarse en las parcelas de cultivo. Teniendo en cuenta esto, es probable que el evento 2015 no debiese ser considerado como tal, en cuanto a producto de variaciones interanuales, sino que sea simplemente una sobreexpresión de la variabilidad estacional intraanual típica del topillo campesino. En consecuencia, en los siguientes comentarios sobre los eventos, no será considerado.

Del resto de eventos, ninguno se detectó en el total de las 40 ZVs de estudio, si bien los eventos 2013, 2019 y 2023 afectaron a la práctica totalidad (38, 36 y 37 casos respectivamente). Exceptuando al de 2015, el evento que afectó a menos ZV fue el 2021 con tan sólo 9 ZVs. Si consideramos la expresión en cultivos, en todos los casos el número de ZVs en que se detectaron fue igual o inferior al 50 % del número de ZVs en que se detectó en vías, salvo el evento 2023,

Tabla 5: Eventos registrados por zona y hábitat.

Nº eventos			Nº eventos			Nº eventos		
ZV	V	C	ZV	V	C	ZV	V	C
AV_01_AM	4	1	PA_08_CA	5	5	VA_01_CE	3	2
AV_02_AM	4	0	PA_09_EC	3	2	VA_02_CE	2	0
BU_01_AR	3	1	PA_10_EC	4	2	VA_03_CE	4	2
BU_03_BU	3	1	PA_11_SV	2	0	VA_04_SU	4	1
BU_04_PI	5	1	SA_02_PB	4	2	VA_05_SE	3	0
LE_01_EC	4	1	SA_03_SA	4	1	VA_06_SE	4	1
LE_02_EC	4	1	SA_04_PB	4	1	VA_07_TC	5	5
LE_03_SA	4	3	SE_01_CU	5	1	VA_08_TC	5	4
PA_01_BO	2	2	SE_02_CU	5	0	ZA_01_BV	4	4
PA_02_CA	4	2	SE_03_CU	2	1	ZA_02_CP	4	3
PA_03_CA	4	3	SE_04_SP	4	1	ZA_03_CP	4	4
PA_04_CA	5	4	SO_01_BO	2	1	ZA_04_CP	4	3
PA_05_CA	5	2	SO_03_CG	2	1	ZA_05_DB	2	0
PA_06_CA	4	4						

V.: Vías de dispersión; C: Cultivos.

en el que en más zonas se expresó en cultivos (23 ZVs, el 57.5 % de las zonas de estudio, el 62.2 % de las ZVs en que se expresó en vías de dispersión). Si utilizásemos como indicador del riesgo potencial para la agricultura castellanoleonesa el número de ZVs en que se detectó colonización en los cultivos, el evento 2023 podría identificarse como el de mayores riesgos en la última década.

En cuanto a la duración de los eventos, se observa que los valores mínimos en vías de dispersión son de dos periodos, con una duración promedio dependiente del evento considerado, pero que va de 2.29 periodos en el caso del evento 2021 a superar los 3 en todos los demás casos (4.76, 4.07, 5.19 y 3.53), lo que implica que los periodos continuos de colonización significativa de topillo puedan superar fácilmente un año, siendo de hecho lo más habitual.

No se observa lo mismo en las parcelas de cultivo, en los que la duración promedio, salvo en el evento de 2016 (2.40 periodos), resultó inferior a 2 (1.83, 1.22, 1.25 y 1.43). Si bien se han detectado máximos de 4 periodos en los eventos 2013 y 2016, en todos los casos en que se han expresado eventos en los cultivos, incluidos esos máximos, la duración ha sido siempre menor que en vías de dispersión, siendo de hecho el valor más habitual para los cultivos la duración de un solo periodo.

En cuanto a la asincronía, se ha detectado que, salvo en el evento 2021 con una asincronía máxima de un periodo, en el resto de los casos la expresión de los eventos en vías de dispersión ha llegado a diferir entre ZVs en más de un año, con asincronías máximas que incluso pueden llegar a los 5 periodos (eventos de 2019 y 2023). Esto puede ir en coherencia con trabajos anteriores en que la variabilidad espacio-temporal de los incrementos demográficos puede ser parcialmente explicada por el efecto de dos ondas viajeras [21].

En relación a los cultivos, en ningún caso se ha detectado expresión de eventos anterior a

su expresión en vías. Si bien se han detectado mínimos de expresión coincidente con el inicio en vías de dispersión, lo más habitual es su expresión retardada respecto a dicho momento, pudiendo alcanzar máximos de hasta un año de retardo (eventos 2013, 2016 y 2023) o incluso superiores (2019). En ningún caso se ha detectado inicio en los cultivos y duración que implicase un fin de evento en cultivos posterior a que finalizase en vías.

En los eventos 2013, 2016 y 2023 no se detecta ningún solape temporal con ningún otro evento (ver Figura 8 del Anexo 7.4). Sin embargo, sí se detecta entre los eventos 2019 y 2021 para el caso de las vías de dispersión. Así, hay algunas ZVs en las que se han detectado ambos eventos de forma independiente (ejemplo VA_08_CA), mientras que en otras tan sólo se ha detectado el de 2019, pero con duración más alta de lo que podría ser habitual, incluso de 11 periodos (prácticamente tres años), que conlleva que su final solape con los periodos establecidos para el evento 2021. Esta definición de inicio de evento correspondiente a 2019, pero que por duración llega a solaparse con el de 2021, ha ocurrido en 17 ZVs, el 47.2 % de las ZVs en que se describió evento asociado a 2019. Todas estas ZV están ubicadas en la zona centro-sur de la comunidad. Tan sólo en cinco de ellas se observó expresión del evento en los cultivos y, en estos casos, salvo para VA_01_CE con un retardo de un periodo respecto a vías, para el resto el retardo fue de mínimo un año.

El caso más acentuado para este solape se aprecia en SE_02_CU, con un inicio en verano de 2019 y una duración de 11 periodos, prácticamente cuatro años, finalizando en el periodo frío de 2022, lo cual se sale, definitivamente, fuera de toda la normalidad. En añadido, no se detectó evento asociado en sus cultivos. Esto abre a la reflexión y, si bien no se puede decir con certeza, estos solapes podrían estar realmente enmascarando la detección y definición parcial de ciclos involucrados en esas zonas y periodos, pues pudieran estar afectados no sólo por fluctuaciones interanuales sino por un mantenimiento continuado por encima de los niveles mínimos considerados debido a condicionantes externos.

Así, y considerando que la cubierta vegetal es un factor que favorece la alimentación y la protección de las poblaciones, cabe indicar que 2020 y 2021, donde se encuentran los periodos involucrados en el solape, fueron años tremendamente favorables para la producción vegetal, lo que podría haber favorecido el mantenimiento de las poblaciones en unos parámetros superiores a lo habitual. Esto podría apoyarse en función de los datos de producción de cultivos, en que, por ejemplo, para trigo y cebada, los dos cultivos mayoritarios en las zonas de estudio, 2020 y 2021 fueron dos años consecutivos por encima de las producciones medias (no hay más casos de años consecutivos en el periodo de estudio), siendo de hecho 2020 el año más productivo, no sólo en el periodo de este estudio, sino en los últimos 30 años [71].

Si bien en ninguno de los eventos se ha detectado su expresión en todas las ZV, en todas ellas se han detectado al menos dos eventos en vías de dispersión, con un promedio de 3.7. Exceptuando el caso particular del evento 2015 expuesto previamente detectado en una sola ZV, el mínimo ha sido de dos, en 7 ZVs, en los casos de PA_01_BO, PA_11_SV, SE_03_CU, SO_01_BO, SO_03_CG, VA_02_CE y ZA-05-DB. El máximo de cinco se ha detectado en 6 ZVs,

para los casos de BU_04_Pi, PA_04_CA, PA_05_CA, PA_08_CA, VA_07_TC y VA_08_TC, todas ellas encuadradas dentro de la Tierra de Campos.

Sin embargo, en cultivos ha habido seis ZVs en que no se ha expresado ningún evento: AV_02_AM, PA_11_SV, SE_02_CU, VA_02_CE, VA_05_SE, ZA_05-DB. Si bien alguna de ellas está incluida dentro de las que menores eventos se detectaron en vías, cabe reseñar que, en una de ellas, SE_02_CU, se expresaron un total de cinco eventos en vías, si se considera esta vez el posiblemente anómalo de 2015, en ninguno de los cuales hubo expresión en los cultivos. En añadido, esta ZV ha sido la más representativa en cuanto al solape entre los eventos 2019 y 2021 comentados anteriormente.

En cuanto al máximo número de eventos en cultivos, se ha detectado para los casos de PA_08_CA y VA_07_TC, con cinco eventos. Si el indicador de riesgos agrícolas se estableciese en función del número de eventos afectando a los cultivos, estas serían las zonas con mayor riesgo potencial. Con cuatro se ha detectado en los casos de PA_04_CA, PA_06_CA, VA_08_TC, ZA_01_BV y ZA_03_CP, todas ellas, salvo ZA_01_BV, enmarcadas en Tierra de Campos.

Por último, sólo han sido cinco las ZV en que se ha detectado evento asociado a cultivos siempre que se ha detectado en ellas en vías: PA_01_BO, con tan sólo dos eventos durante el periodo de estudio, PA_06_CA y ZA_03_CP con cuatro y PA_08_CA y VA_07_TC con cinco.

4.2. Profundización en la ciclicidad e identificación de la distribución geográfica

El objetivo de este apartado consiste en expandir, completar y contrastar los resultados obtenidos en estudios anteriores, realizados en la misma área de trabajo, sobre la ciclicidad de las fluctuaciones en las distintas zonas de Castilla y León [19] [72]. Para ello, se han representado tres gráficos para cada una de las zonas de trabajo, tanto para las vías de dispersión como para los cultivos, como se puede ver en el Anexo 7.6: el gráfico de la serie, periodograma y autocorrelograma, en ese orden.

Así, se han definido ciclicidades asociadas principalmente a las frecuencias 0.05, 0.075, 0.1 y 0.125, correspondientes a 20, 13.33, 10 y 8 periodos respectivamente. A modo de excepción, se han detectado dos frecuencias particulares candidatas a ser tenidas en cuenta, de 0.15 (6.66 periodos) y 0.175 (~ 5.71 periodos), asociada la primera a vías de dispersión de las zonas PA_03_CA y BU_03_BU y la también en vías de dispersión en SE_04_SP y SE_02_CU. Para los casos de PA_03_CA y BU_03_BU y SE_04_SP estas frecuencias no resultaron ser los picos más prominentes en el periodograma de vías y no se detectaron en cultivos, por lo que, si bien no debe olvidarse su detección por su posible influencia en estudios futuros, no se han seguido considerando en este estudio.

Sin embargo, para el caso de la frecuencia de 5.71 periodos en SE_02_CU, si resultó ser de forma clara el pico más acusado en vías de dispersión. Una frecuencia de 5.71 periodos, es decir, menor de dos años, representaría el único caso descrito en bibliografía para el topillo

campesino, al menos hasta donde se tiene constancia. En añadido, SE_02_CU es la única ZV donde se detectó el evento de 2015 en vías de dispersión, que ya se ha comentado quizá no fue realmente fruto de fluctuación interanual, y es una de las ZV donde se detectó el solape entre los eventos 2019 y 2020, siendo el caso más acusado de dicho solape. Para esta ZV, en el periodograma de cultivos no se detectó ninguna evidencia de esta frecuencia, detectándose en él una frecuencia de 20 periodos con un posible segundo pico en 10. Con todo esto en cuenta, se ha descartado la frecuencia de 5.71, considerando para SE_02_CU la frecuencia de 20 definida en el hábitat cultivo.

En la Tabla 6 se representan las ciclicidades definidas para cada ZV a ser consideradas en este estudio. En algunos casos se aportan dos periodicidades, siendo la primera que aparece la que se utilizará en este estudio y la segunda representa un posible segundo pico que, por su prominencia, podría ser tenido en cuenta en estudios futuros orientados en profundizar en la ciclicidad. En la Tabla 7 se incluye el resumen en cuanto a número de zonas en que se ha definido cada frecuencia final. En la Figura 1 se representa la distribución geográfica de las distintas frecuencias en función de las ZV.

Tabla 6: Grupo asignado a cada ZV en base a la periodicidad/es detectada.

Zona	Per.	G	Zona	Per.	G	Zona	Per.	G
AV_01_AM	10	2	PA_08_CA	8	1	VA_01_CE	20; 10	4
AV_02_AM	10	2	PA_09_EC	20; 13.33	4	VA_02_CE	20, 10	4
BU_01_AR	10	2	PA_10_EC	10; 20	2	VA_03_CE	10	2
BU_03_BU	13.33	3	PA_11_SV	13.33; 20	3	VA_04_SU	10	2
BU_04_PI	8	1	SA_02_PB	10	2	VA_05_SE	20	4
LE_01_EC	10; 13.33	2	SA_03_SA	10	2	VA_06_SE	10	2
LE_02_EC	13.33; 8	3	SA_04_PB	10	2	VA_07_TC	8	1
LE_03_SA	8	1	SE_01_CU	10	2	VA_08_TC	8	1
PA_01_BO	13.33	3	SE_02_CU	20; 10	4	ZA_01_BV	8; 20	1
PA_02_CA	10	2	SE_03_CU	20; 10	4	ZA_02_CP	8	1
PA_03_CA	8; 6.66	1	SE_04_SP	20; 5.71	4	ZA_03_CP	8	1
PA_04_CA	8	1	SO_01_BO	10; 20	2	ZA_04_CP	10	2
PA_05_CA	8	1	SO_03_CG	10; 13.33	2	ZA_05_DB	10	2
PA_06_CA	8	1						

Per.: Periodo/s; G: Grupo asignado.

Tabla 7: Número y porcentaje de zonas por grupo

Grupo	Periodo	Número de zonas	Porcentaje de zonas
1	8	12	30 %
2	10	17	42.5 %
3	13.33	4	10 %
4	20	7	17.5 %

Se observa, si bien no perfecta, cierta tendencia geográfica en cuanto a la distribución espacial de las frecuencias. Así, en general, las ZV incluidas en el grupo 1 (ciclicidad 8 periodos), están

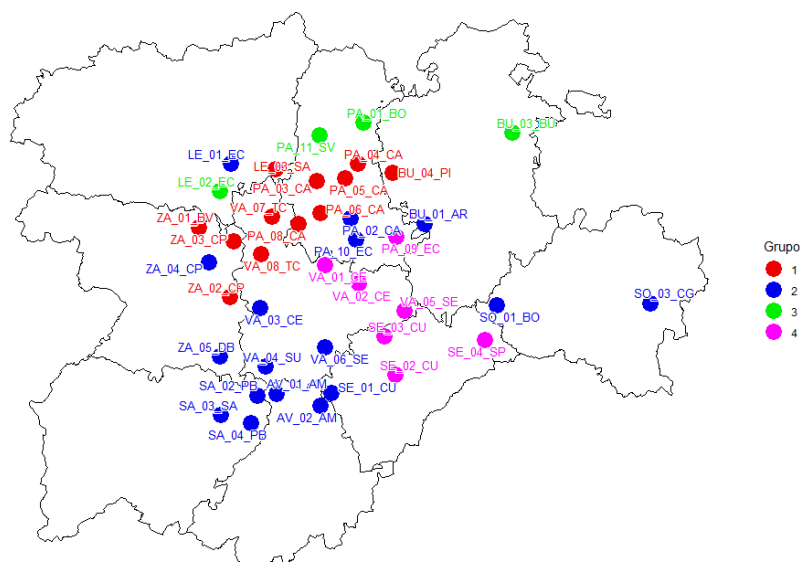


Figura 1: Mapa con las 40 ZVs agrupadas por colores en base a su periodicidad detectada. Fuente: elaboración propia.

distribuidas por Tierra de Campos. El suroeste de la comunidad está identificado por el grupo 2 (10 periodos), si bien la distribución de este grupo no se limita ahí, siendo el que mayor dispersión geográfica ha mostrado. El grupo 3 (13.33 periodos), incluye las ZV que delimitan el norte del área de trabajo. Por último, en la zona sureste de la región se han distribuido principalmente las ZV del grupo 4 (20 periodos), siendo la excepción las dos ZV de la provincia de Soria que se asignaron al grupo 2.

En referencia a estas dos ZV de Soria, en ambos casos, junto al pico en 10 periodos se detectó un posible segundo pico, con frecuencia 20 para el caso de SO_01_BO y con frecuencia 13.33 para SO_03_CG. La existencia de este segundo pico, si fuese considerado, implicaría que SO_01_BO quedaría asignada al grupo 4 y SO_03_CG al 3, lo que resultaría en una distribución geográfica más racional: el grupo 3 ocuparía las zonas más próximas a los sistemas montaños cantábrico e ibérico, a excepción de LE_01_EC (aunque también se ha detectado un segundo pico de periodo 13.33 en esta ZV), quedando el grupo 4 representando de mejor manera al sureste de la comunidad. Comentar que la circunstancia de un posible segundo pico también se ha detectado, por ejemplo, para el caso de PA_10_EC, con un segundo pico en la periodicidad de 20, pudiendo formar parte del grupo 4, lo que aclararía aún más la distribución, reduciendo las zonas del grupo 2 que parecen ubicarse como “frontera” entre Tierra de Campos y el sureste. En cualquier caso, pese a estas observaciones, que no deben ser obviadas en estudios futuros, en el presente se continuará con la clasificación obtenida en función del pico más prominente.

Llegado este punto, procede realizar una comparativa de los resultados obtenidos con el precedente publicado más inmediato al presente trabajo, realizado con las mismas ZV de trabajo e incluyendo también una parte relacionada con la ciclicidad. Así, en (Caminero-Saldaña et al., 2024) [19] se habla de la existencia en Castilla y León de ciclos de 2 a 3 años hasta de 4 a 5,

además de identificar zonas en las que no se encontró ciclicidad.

La equivalencia parcial de nuestros resultados con los de dicho trabajo es inmediata si consideramos transformar estos ciclos de años a periodos: una ciclicidad precedente de 2-3 (6 a 9 periodos), sería asimilable a la periodicidad de 8 definida en nuestro trabajo, de 3-4 años se asimilaría al 10 y la de 4-5 al 13.33.

Los primeros comentarios importantes en la comparativa son, en primer lugar, que en nuestro estudio se ha definido una frecuencia (20 periodos, 6-7 años) que no fue detectada en el estudio precedente. En segundo, en nuestro estudio se ha conseguido establecer una frecuencia en todas las ZV, mientras que en el anterior hubo siete ZVs donde no se detectó ciclicidad.

Ambas casuísticas, detección de una frecuencia más larga y mayor rango de ZVs resueltas, pueden ser debidas, al menos parcialmente, a que el estudio precedente fue realizado adoleciendo de los seis últimos periodos aquí utilizados (años 2023 y 2024, al no estar disponibles en aquel momento), lo que implica no sólo una serie temporal más corta sino que se abordó con un evento general (2023) menos que el presente trabajo.

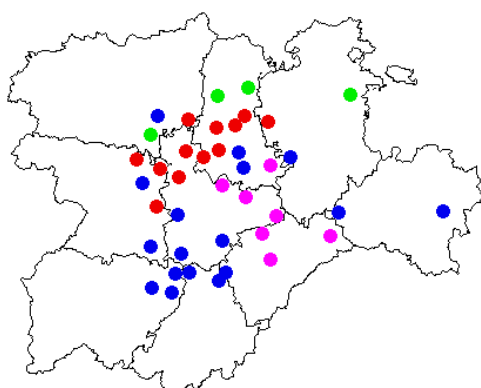
En la Tabla 8 se resumen los datos en cuanto a periodicidad detectada en ambos trabajos. En la Figura 2 se representa la comparativa en la distribución geográfica de las frecuencias.

Tabla 8: Equivalencias entre periodicidades y color de su representación

Grupo	Per. A	Per. B	Color en los mapas
1	8	2-3	Rojo
2	10	3-4	Azul
3	13.33	4-5	Verde
4	20	-	Magenta
5	-	1 (no ciclicidad detectada)	Dorado

Per. A: Periodicidad en el presente TFG; Per. B: Periodicidad en (Caminero-Saldaña et al., 2024)

Periodicidad presente estudio



(Caminero Saldaña et al., 2024)

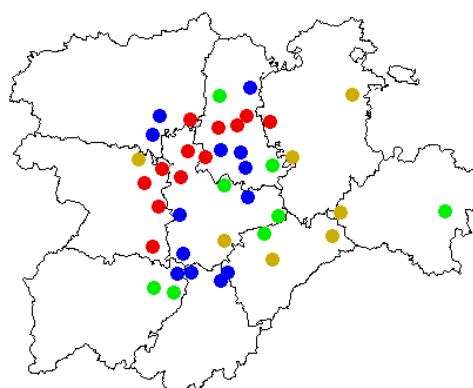


Figura 2: Comparación de las ciclicidades encontradas en el presente TFG y en (Caminero-Saldaña et al., 2024). Fuente: elaboración propia.

Para las 7 ZV no informadas previamente, se han detectado los cuatro tipos de frecuencia: un caso para periodicidad de 8 (ZA_01_BV), tres para 10 (BU_01_AR, SO_01_BO y VA_06_SE),

uno para 13.33 (BU_03_BU) y dos para 20 (SE_02_CU y SE_04_SP). De las 33 ZVs restantes, 20 de ellas se han clasificado con las mismas periodicidades equivalentes en ambos trabajos. Los mayores cambios han ocurrido en la periodicidad más larga (4-5) del estudio previo (8 ZVs en dicho estudio), en la que en tan sólo una ZV (PA_11_SV) ha mantenido dicha frecuencia. Cuatro ZVs (PA_09_EC, SE_03_CU, VA_01_CE y VA_05_SE) pasan ahora a definirse con una frecuencia aún más larga (20 periodos). En tres ZVs (SA_03_SA, SA_04_PB y SO_03_CG) se ha encontrado una periodicidad algo más corta (10, equiparable al 3-4 precedente).

En nuestro estudio se han detectado dos ZVs previamente clasificadas con periodicidad menor y que ahora se definen con una frecuencia de 13.33, equiparable al 4-5 previo: LE_02_EC y PA_01_BO. Es reseñable que LE_02_EC es una de las ZVs leonesas englobadas en Tierra de Campos, caracterizada, como se verá en el párrafo siguiente, por las periodicidades más cortas. Esta ZV tampoco había sido incluida previamente en dicho nivel, sino en el de 3-4, habiendo sido ahora definida con una periodicidad aún más larga.

La menor frecuencia (8 en el trabajo actual, 2-3 en el precedente), había sido detectada previamente principalmente asociada a Tierra de Campos, salvo el caso de ZA_05_DB. En el presente trabajo Tierra de Campos continúa siendo la zona representativa de las periodicidades más cortas, incluso en mayor medida, pues ZA_05_DB ha salido de esta categoría, pasando al grupo de frecuencia 10, mientras que PA_06_CA, perteneciente a Tierra de Campos, pasa ahora a clasificarse con la periodicidad más corta frente al 3-4 del trabajo precedente.

De hecho, sin tener en cuenta la ZV leonesa comentada anteriormente, con la clasificación actual, todas las ZV asociadas a Tierra de Campos, ya han sido categorizadas con la menor periodicidad, salvo la excepción de ZA_04_CP, clasificada ahora con una periodicidad de 10, algo más larga. Comentar que esta ZV sí estaba incluida en las frecuencias más cortas en el estudio anterior. Sobre la periodicidad 3-4, estaba integrada por 13 ZVs en el trabajo precedente, siendo la frecuencia detectada en un mayor número de ZVs. En dicho estudio no podía siquiera intuirse ningún patrón en su distribución geográfica. Con la clasificación actual se identificaría la zona suroeste de la comunidad, con algunas excepciones como se ha indicado y comentado ya en párrafos anteriores.

Ahora continúa siendo la frecuencia mayoritaria, aumentando a 17 el número de ZVs implicadas. De ellas: nueve han conservado la frecuencia en ambas clasificaciones. Tres no habían sido clasificadas previamente (BU_01_AR, SO_01_BO y VA_06_SE). Dos (ya comentadas anteriormente, ZA_04_CP y ZA_05_DB) han sido clasificadas con una frecuencia mayor; en oposición, tres (SA_03_SA, SA_04_PB y SO_03_CG) han sido categorizadas con una frecuencia menor.

Sobre la nueva frecuencia detectada (20 periodos), se ha asignado a siete ZVs, de las que sólo SE_02_CU y SE_04_SP no habían sido clasificadas previamente. Cuatro (PA_09_EC, E_03_CU, VA_01_CE y VA_05_SE) habían sido clasificadas con la frecuencia más larga de 4-5 detectada en el estudio anterior, habiendo detectado ahora una periodicidad incluso mayor.

Sobre la séptima integrante de esta nueva frecuencia, el caso de VA_02_CE representa la ZV que mayor variación presenta en la comparativa, pasando a una periodicidad de 20 (6-7 años)

cuando previamente fue clasificada de 3-4. Sin embargo, se debe mencionar que para esta ZV se detectó en el periodograma de vías de dispersión un segundo pico en 10 periodos, de hecho, casi tan alto como el de 20 finalmente considerado en nuestra clasificación. Esto podría implicar, como incluso pudiera intuirse visualmente de su serie temporal particular, que realmente tuviese una periodicidad de 10 (conservando la categoría de 3-4 del trabajo precedente), pero con una sobreexpresión cada 20, lo que se aconseja ser tenido en cuenta en trabajos futuros.

La comparación completa de ambas clasificaciones se puede consultar en la Tabla 9.

Tabla 9: Comparación de periodicidades asignadas a cada ZV en el presente TFG y en (Caminero-Saldaña et al., 2024)

Zona	Per. A	Per. B	Zona	Per. A	Per. B
AV_01_AM	10	3-4	SA_04_PB	10	4-5
AV_02_AM	10	3-4	SE_01_CU	10	3-4
BU_01_AR	10	1	SE_02_CU	20	1
BU_03_BU	13.33	1	SE_03_CU	20	4-5
BU_04_PI	8	2-3	SE_04_SP	20	1
LE_01_EC	10	3-4	SO_01_BO	10	1
LE_02_EC	13.33	3-4	SO_03_CG	10	4-5
LE_03_SA	8	2-3	VA_01_CE	20	4-5
PA_01_BO	13.33	3-4	VA_02_CE	20	3-4
PA_02_CA	10	3-4	VA_03_CE	10	3-4
PA_03_CA	8	2-3	VA_04_SU	10	3-4
PA_04_CA	8	2-3	VA_05_SE	20	4-5
PA_05_CA	8	2-3	VA_06_SE	10	1
PA_06_CA	8	3-4	VA_07_TC	8	2-3
PA_08_CA	8	2-3	VA_08_TC	8	2-3
PA_09_EC	20	4-5	ZA_01_BV	8	1
PA_10_EC	10	3-4	ZA_02_CP	8	2-3
PA_11_SV	13.33	4-5	ZA_03_CP	8	2-3
SA_02_PB	10	3-4	ZA_04_CP	10	2-3
SA_03_SA	10	4-5	ZA_05_DB	10	2-3

Per. A: Periodicidad en el presente TFG; Per. B: Periodicidad en (Caminero-Saldaña et al., 2024)

4.3. Identificación preliminar de factores extrínsecos implicados en la modulación de la abundancia

El objetivo de este apartado es realizar una primera aproximación a la posible relación de las variables consideradas para factores extrínsecos con la abundancia y la ciclicidad (periodicidad calculada en el apartado anterior).

Para ello, se han calculado las correlaciones entre dichas variables y la variable a estudiar, en nuestro caso LNVAI, por una parte considerando los promedios zonales por ZV y por otra los estacionales.

En la Figura 3 se han representado las correlaciones obtenidas mediante el test de correlación de Pearson, tachando con una cruz aquellas que no resultasen significativas a nivel 0.05.

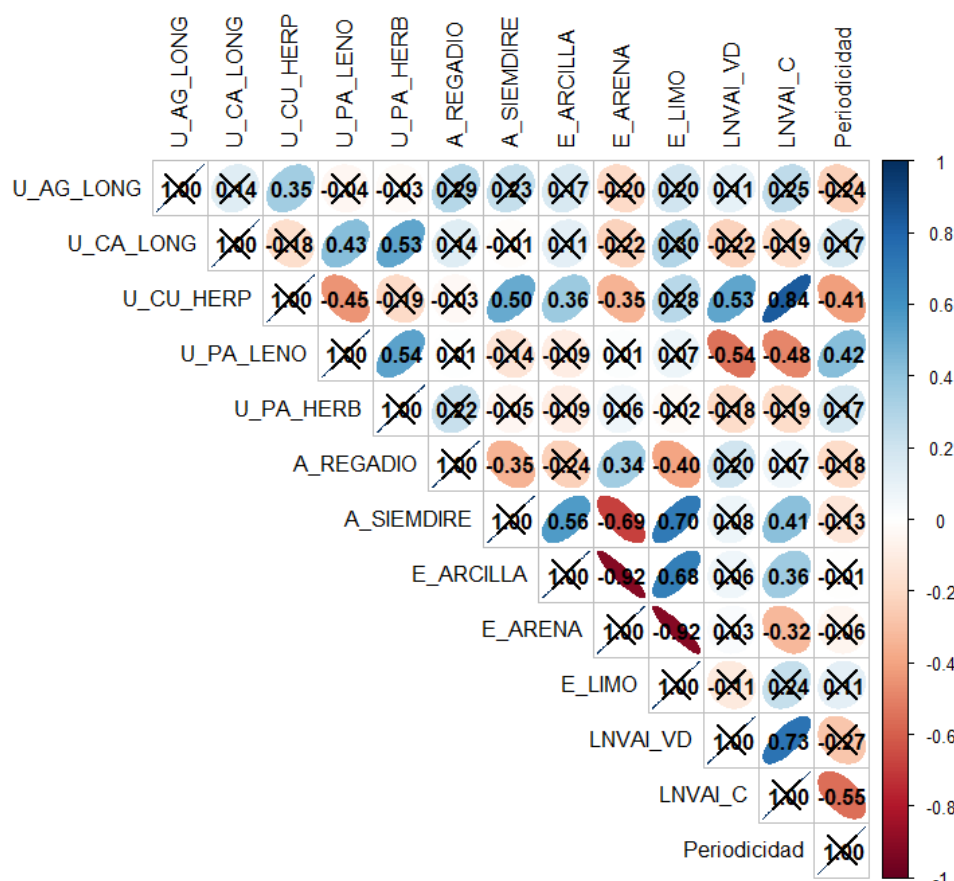


Figura 3: Correlaciones entre variables paisajísticas, agronómicas y edáficas con la abundancia en vías de dispersión y en cultivos y con la periodicidad encontrada para cada ZV. Fuente: elaboración propia.

Respecto a LNVAI en vías, lo primero a mencionar es que no se ha detectado una correlación significativa con la periodicidad, indicando, en contra de lo que pudiera esperarse, que la mayor frecuencia en la sucesión de eventos demográficos no parece influir, al menos de forma significativa, en la mayor o menor expresión de LNVAI en las vías de dispersión. En referencia a los factores extrínsecos, para LNVAI en vías de dispersión no se han detectado correlaciones para las variables relacionadas con factores agronómicos ni edafológicos. Respecto a las paisajísticas, tan sólo se ha detectado para CU_HERP y PA_LEN0, la primera con signo positivo y la segunda negativo. Son también los dos únicos casos en que se han detectado correlaciones significativas con la periodicidad, en este caso de signo contrario al obtenido con LNVAI en vías.

La correlación con CU_HERP está en coherencia con trabajos anteriores en cuanto a la importancia de este tipo de hábitat en la dinámica poblacional del topillo [7] en cuanto a constituir uno de sus reservorios típicos. En cuanto a PA_LEN0, el presente es el primer estudio, hasta donde somos conscientes, donde se ha detectado su posible influencia tanto en los valores de LNVAI en vías como en la periodicidad. Se abre la hipótesis, a testar en estudios futuros, de que los pastizales arbolados y arbustivos puedan suponer zonas de compensación ecológica en el territorio agrícola que puedan actuar como tamponadores en la dinámica poblacional del topillo, por ejemplo, porque, considerando la biodiversidad en ellos incluida, puedan guarecer

parte de la población de depredadores que no encuentra cobijo en otros tipos de hábitat.

En cuanto a LNVAI en cultivos, lo primero reseñable es la alta correlación positiva detectada con LNVAI en vías (0.73). Pese a las importantes diferencias en su expresión detectadas en el capítulo de exploración de la base de datos, en cuanto a no expresión significativa siempre que se da en vías, o la asincronía retardada en su expresión, esta alta correlación corrobora la importancia en la dinámica del topillo de las vías sobre los riesgos posteriores agrícolas [7] [22], siendo la colonización a los cultivos no sólo dependiente de ello, sino cuantitativamente relacionada con la abundancia que se alcance en las vías.

En el caso de LNVAI en cultivos, sí se ha detectado una correlación negativa con la periodicidad. Si bien para LNVAI en vías la periodicidad no parecía influir de forma significativa, sí parece hacerlo en cuanto a la colonización en los cultivos, pudiendo indicarse que, en general, los riesgos agrícolas vinculados a dicha colonización son mayores cuanto menor es la periodicidad en la ocurrencia de eventos. De hecho, y sin entrar en demasiados detalles, la simple visualización de los mapas de calor de la Figura 7 (Anexo 7.4) ya identifica que los mayores valores recurrentes de LNVAI en cultivos se identifican en ZVs vinculadas a Tierra de Campos, donde se han definido los menores valores de periodicidad en el apartado anterior.

En relación a las correlaciones con los factores extrínsecos, respecto a las variables paisajísticas se han detectado de nuevo con CU_HERP y PA_LEN0, y con los mismos signos que para LNVAI en vías. Las reflexiones acerca de PA_LEN0 serían las mismas que anteriormente, si bien para el caso de CU_HERP resulta reseñable el alto valor de correlación detectado (0.84), incluso mayor que el detectado entre LNVAI en cultivos con LNVAI en vías. Esto realza la importancia previamente descrita [7] [36] de este tipo de superficies reservorio en el riesgo para los cultivos anuales, siendo un factor determinante, a efectos cuantitativos, de los niveles de colonización que puedan acontecer en los mismos.

En el caso de LNVAI en cultivo, se ha detectado una correlación positiva implicando variables agronómicas, en concreto con A_SIEMDIRE, vinculada al porcentaje de parcelas de cultivo que se explotan en régimen de siembra directa. Este tipo de parcelas han sido previamente descritas [13] como reservorios potenciales que implican riesgos no sólo para ellas mismas, sino también para las parcelas colindantes. Para ellas mismas, dado que la siembra directa es una técnica de cultivo en la que no se altera el suelo, lo esperable es que a mayor porcentaje de cultivos trabajados mediante esta técnica, mayor abundancia de topillo contendrán. La correlación detectada corrobora esto, y de forma independiente en apariencia con el posible efecto que pudiera tener sobre las vías, pues, de hecho, la correlación de A_SIEMDIRE con LNVAI en vías fue no sólo no significativa sino prácticamente nula (0.08). Con esto se pone de manifiesto el realce de los riesgos agrícolas en zonas donde se realiza esta práctica, siendo candidatas las parcelas de cultivo a cobijar mayores abundancias.

En cuanto a las variables edáficas, se ha hallado correlación positiva con E_ARCILLA y negativa con E_arena. Este resultado resulta de gran interés, pues estas componentes edáficas identifican elementos finos y gruesos, respectivamente, en la textura de los suelos. Y esto guarda

relación, por ejemplo, con la mayor capacidad de retención de agua. Los suelos arcillosos, tienden a ser más productivos que los arenosos, donde el agua tiene más problemas de percolación y, por tanto, menor disponibilidad para las plantas.

Esta posible mayor productividad, en cuanto a traducción en cubierta vegetal, podría ser un factor explicativo de la tendencia a detectar mayores abundancias cuanto mayor es el porcentaje de elementos finos y menor el de gruesos. Pero, asimismo, un mayor porcentaje de elementos finos se traduce en suelos más compactos, el de gruesos en más sueltos. Siendo sus madrigueras subterráneas un factor clave en la presencia del topillo en el territorio, se ha descrito la textura del suelo como uno de los factores que las afecta [11], de forma que una mayor capacidad de compactación aporta a las madrigueras una mayor estabilidad, mientras que en suelos muy arenosos la tendencia al derrumbe es mayor. Las correlaciones detectadas en este TFG con estas covariables relacionadas con la textura del suelo apoyan una vía de trabajos futuros que profundicen sobre ello.

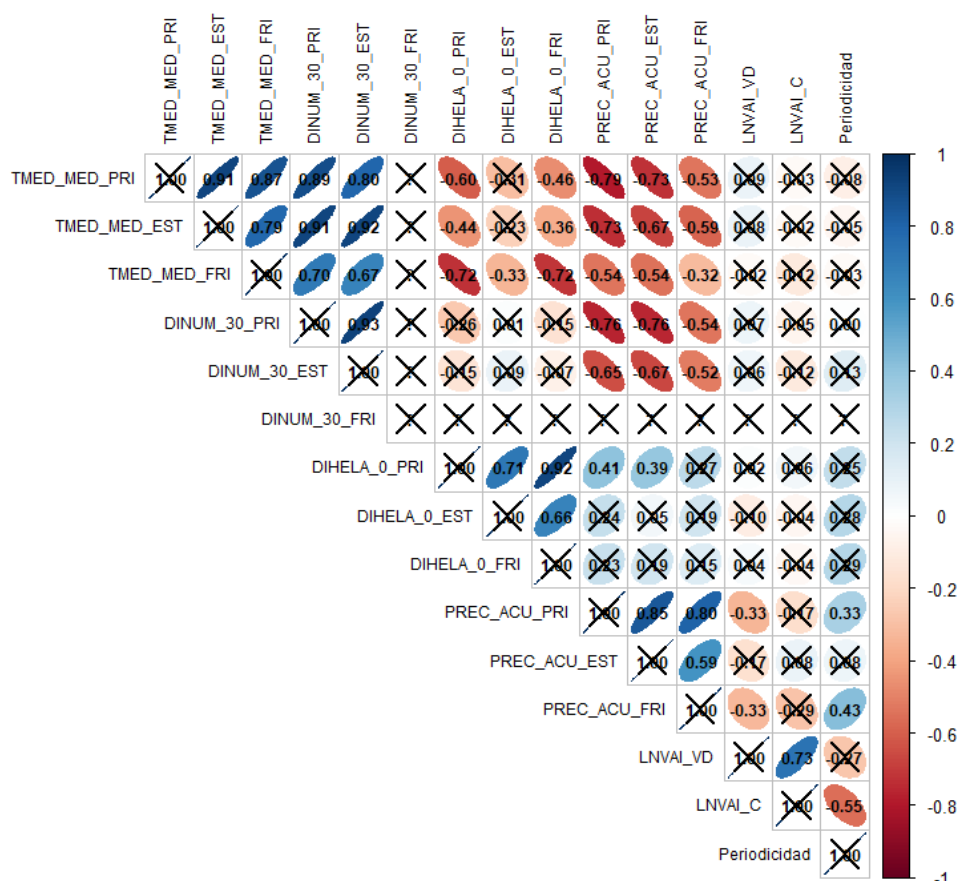


Figura 4: Correlaciones entre variables climáticas con la abundancia en vías de dispersión y en cultivos y con la periodicidad encontrada para cada ZV. Fuente: elaboración propia.

Respecto a factores climáticos, de las variables consideradas tan sólo PREC_ACU_PRI y PREC_ACU_FRI han resultado significativas con el LNVAI en vías, si bien no con el LNVAI en cultivos (ver Figura 4). Ambas correlaciones han sido negativas, detectándose también con la periodicidad, con signo opuesto. Por el momento no se pueden establecer demasiadas reflexiones al respecto, más allá de que las zonas donde mayores abundancias se dan en vías de dispersión

tienden a ser zonas con menores precipitaciones en los periodos de primavera y frío.

Tras el análisis a nivel espacial de los factores extrínsecos a través de la expresión general de los LNVAI en las zonas de vigilancia, también se han obtenido las correlaciones a nivel temporal, esta vez considerando la expresión de los LNVAI generales en los periodos de estudio. En este caso, se han calculado las correlaciones considerando los promedios de varios retardos de las variables climáticas para cada periodo (Figuras 15 y 16 del Anexo 7.7), en concreto respecto a los periodos primaveral, estival y de frío inmediatamente anteriores al periodo considerado (nos referiremos a ellos como del mismo año) y los mismos periodos en un año antes.

Respecto a las del mismo año, han sido tres las correlaciones detectadas con LNVAI en vías. Las tres también han sido detectadas, con el mismo signo, con LNVAI en cultivos. Dos de ellas (C_PREC_ACU_EST_0 y C_PREC_ACU_FRI_0) refieren correlación positivas referentes a precipitación acumulada, en concreto en los periodos estival y de frío. La tercera correlación, con C_DIHELA_0_FRI, que refleja influencia de la temperatura en cuanto a número de días de helada durante el periodo frío, ha resultado negativa. Para LNVAI en cultivos se ha encontrado una correlación más, C_TMED_MED_FRI_0, también relacionada con la temperatura en el periodo frío, en este caso referente a la temperatura media detectada.

De las correlaciones con estaciones en el año anterior, no se ha detectado ninguna que implique variables relacionadas con la precipitación. En vías se ha detectado influencia de la temperatura en primavera en cuanto a las correlaciones con C_TMED_MED_PRI_1 y C_DIHELA_0_PRI_1, de signo negativo la primera y positivo la segunda. Esta segunda, con el mismo signo, también se ha detectado para LNVAI en cultivos. También en cultivos se ha detectado otra correlación positiva con otra variable relacionada con temperatura (C_TMED_MED_EST_1), en este caso en el periodo estival.

En función de todas estas correlaciones, se puede resumir, en cuanto a temperaturas, que los precedentes de periodos fríos y primaveras suaves (no excesivamente fríos los primeros y no excesivamente cálidas los segundos), son adecuados para favorecer la abundancia posterior. Esto también puede verse favorecido cuando el periodo de estío del año anterior es relativamente cálido. Respecto a las precipitaciones, las mayores lluvias durante los periodos estival y frío del mismo año favorecerán las mayores abundancias posteriores.

Estas características climatológicas en distintas estaciones precedentes que pudieran favorecer la abundancia posterior de topillo campesino resultan coherentes en su generalidad con lo descrito en trabajos anteriores [34] [26] [25] [19], no habiendo detectado en nuestro estudio ninguna relación adicional de interés o algún caso que pudiera suponer un efecto opuesto.

4.4. Comparación de distintas metodologías para la definición de algoritmos predictivos y selección de esta

El objetivo de este apartado es la comparación de diversas metodologías para el ajuste de un modelo predictivo espacio-temporal de las fluctuaciones demográficas de topillo campesino

en Castilla y León en cuanto a ratios de crecimiento zonales entre dos periodos. En función de las necesidades de la estrategia de Gestión Integrada de esta especie, en el presente TFG se ha planteado cotejar hasta donde se puede llegar a predecir con un año de antelación.

En consecuencia, en función de la estructuración temporal de los periodos utilizados en este trabajo, la variable a explicar es el ratio de crecimiento de la abundancia respecto a tres periodos de antelación, como se ha especificado anteriormente, $GR_t = \log(VAI_t) - \log(VAI_{t-3})$.

Por tanto, las variables con retardos inferiores a tres periodos (un año) no pueden ser tenidas en cuenta a la hora de ajustar el modelo, reduciéndose así el número de variables tal y como se puede apreciar en la Tabla 10.

Tabla 10: Número de variables candidatas para los modelos según tipo y hábitat.

Tipo de covariable	Vías de dispersión	Cultivos
Denso-dependientes	13	26
Paisajísticas	5	
Agronómicas	2	
Edáficas	3	
Climáticas	14	
Total	37	50

Además, las variables denso-dependientes causan que haya periodos con datos faltantes, ya que, por ejemplo, para una variable con un retardo de tres periodos no hay datos hasta la cuarta observación. Esto ocurre en cada zona, y por tanto el número de observaciones en el conjunto de entrenamiento se ve bastante disminuido, pasando de 1152 (36 periodos * 32 zonas) observaciones a 672 (21 periodos * 32 zonas), es decir, se ven reducidas en aproximadamente un 41.67 %.

Utilizando el set de datos de entrenamiento (TRAIN) se han implementado los tres algoritmos automáticos de selección de variables (Forward, Backward e Híbrida), considerando cuatro subgrupos de los datos, en función de la estacionalidad, tal que:

- Modelo GLOBAL. Ajustado con todos los datos disponibles (672 observaciones).
- Modelo PRI. Ajustado con los datos de los periodos de tipo primavera (224 observaciones).
- Modelo EST. Ajustado con los datos de los periodos de tipo estío (224 observaciones).
- Modelo FRI. Ajustado con los datos de los periodos de tipo frío (224 observaciones).

En la implementación de los algoritmos se ha usado la corrección de White para el test de la F por sospecha de heterocedasticidad, con el estimador HC3 (Heteroskedasticity-Consistent estimator 3). Como p-valor de entrada y de salida se ha escogido el valor 0.01. Además, una vez finalizado el algoritmo, se procedió a corregir la posible multicolinealidad eliminando, en cada iteración, la variable con el mayor factor de inflación de la varianza (VIF), siempre que este

superara el valor umbral de 5. El modelo se reajustó sucesivamente hasta que todas las variables incluidas presentaron un VIF inferior a 5. Esto resulta en un total de 12 modelos distintos para cada tipo de hábitat.

4.4.1. Vías de dispersión

Antes de continuar, hay que indicar que al utilizar el método de selección Híbrida se ha llegado exactamente a los mismos resultados que con la selección forward, por lo que de aquí en adelante solo se hablará de los modelos Forward y Backward.

Se han calculado los RMSE y los coeficientes de determinación ajustados con cada uno de los cuatro modelos sobre cuatro subconjuntos de los datos de entrenamiento: uno con todos los datos (TODOS), otro sólo con las primaveras (PRI), otro sólo con los estíos (EST) y por último otro sólo con los fríos (FRI). En el caso de los modelos estacionales, para el conjunto con todos los datos el RMSE se ha obtenido combinando los residuos de cada modelo estacional aplicado exclusivamente a los datos de su estación correspondiente. Los resultados aparecen en las Tablas 12 y 13.

Tabla 12: RMSE de todos modelos de vías de dispersión sobre el conjunto total de los datos de entrenamiento y subconjuntos estacionales.

	Modelo			
	Forward		Backward	
Datos	GLOBAL	ESTACIONAL	GLOBAL	ESTACIONAL
TODOS	0.635	0.590	0.623	0.597
PRI	0.564	0.612	0.554	0.613
EST	0.606	0.603	0.633	0.577
FRI	0.716	0.544	0.676	0.597

Tabla 13: Coeficiente de determinación ajustado de todos modelos de vías de dispersión sobre el conjunto total de los datos de entrenamiento y subconjuntos estacionales.

	Modelo			
	Forward		Backward	
Datos	GLOBAL	ESTACIONAL	GLOBAL	ESTACIONAL
TODOS	0.644	–	0.649	–
PRI	0.634	0.622	0.627	0.621
EST	0.712	0.721	0.679	0.765
FRI	0.534	0.736	0.577	0.683

Tanto en base a los valores RMSE como a los coeficientes de determinación ajustados para los distintos modelos, se observa lo siguiente:

- Los modelos estacionales por separado parecen ajustar mejor que el global, salvo para los datos de las primaveras.
- No hay una diferencia clara entre el rendimiento de los modelos Forward y Backward.

En la Tabla 29 (Anexo 7.8) se muestran las variables incluidas en cada uno de los modelos. La única variable presente en todos los modelos considerados se trata de D_LNVAI_T3, siempre con signo negativo. Considerando la globalidad de los modelos, se han incluido en alguno de ellos otras 10 variables denso-dependientes, destacando las relacionadas con los periodos de estío y primavera, siempre negativas en este último caso menos la relacionada con el periodo de 4 años atrás. De entre las del estío, la abundancia de dos estaciones anteriores ha sido la que más ha aparecido, siempre con signo positivo. También han sido relevantes las relacionadas con los periodos frío relativas a 2 y 3 años atrás, siempre con signo positivo.

De las variables paisajísticas, U_PA_LEN0 es aquella que ha participado en un mayor número de modelos (3), siempre con signo negativo, confirmando lo comentado en apartados anteriores sobre el posible interés de las superficies de pastizal arbolado o arbustivo como tamponadoras de las abundancias de topillo.

En cuanto a las variables agronómicas, A.REGADIO ha sido la que más veces ha sido incluida, en concreto en cuatro de los modelos, en todos los casos con signo positivo, lo que puede ser un indicativo que el porcentaje de superficie en régimen de regadío puede influir en incrementos positivos de la abundancia. De las variables climáticas, se han llegado a incluir hasta 7 en un mismo modelo, en el caso del modelo Backward Global. C_DIHELA_0_PRI_1 y C_DINUM_30_PRI_1 son de las que están presente en más modelos (3 y 5 respectivamente), la primera con signo positivo en todos los casos y segunda con signo negativo también en todos los casos. Las variables relacionadas con las precipitaciones apenas han sido incluidas en los modelos.

4.4.2. Cultivos

Análogamente a los resultados en vías de dispersión, al utilizar el método de selección híbrida se han obtenido idénticos resultados que con la selección forward, por lo que, de nuevo, de aquí en adelante solo se hablará de los modelos Forward y Backward.

Tabla 14: RMSE de todos modelos de cultivos sobre el conjunto total de los datos de entrenamiento y subconjuntos estacionales.

	Modelo			
	Forward		Backward	
Datos	GLOBAL	ESTACIONAL	GLOBAL	ESTACIONAL
TODOS	0.318	0.326	0.315	0.301
PRI	0.278	0.269	0.275	0.254
EST	0.365	0.374	0.362	0.346
FRI	0.304	0.284	0.300	0.276

En las Tablas 14 y 15 se muestran los resultados para los valores RMSE y R2, respectivamente, para cada uno de los modelos en el hábitat cultivos.

En esta ocasión, se observa que, por lo general, los modelos Globales parecen haber ajustado peor que los estacionales y que, en todas las categorías, los modelos Backward han arrojado

Tabla 15: Coeficiente de determinación ajustado de todos modelos de cultivos sobre el conjunto total de los datos de entrenamiento y subconjuntos estacionales.

Datos	Modelo			
	Forward		Backward	
	GLOBAL	ESTACIONAL	GLOBAL	ESTACIONAL
TODOS	0.670	–	0.675	–
PRI	0.621	0.653	0.624	0.686
EST	0.678	0.673	0.680	0.714
FRI	0.655	0.708	0.660	0.723

mejores resultados.

Entre las variables denso-dependientes, D_LNVAI_T3 vuelve a ser la variable más a destacar, apareciendo en todos los modelos (ver Tabla 30). De forma interesante, en alguno de los modelos se han incluido 7 de las variables denso-dependientes asociadas al hábitat vías de dispersión, casi tantas como las del propio hábitat en análisis, lo que pone de manifiesto, una vez más, la importancia de las vías en la colonización posterior de los cultivos. De ellas, han destacado D_LNVAI_FRI_1_VD y D_LNVAI_PRI_2_VD, apareciendo 3 veces siempre con signo negativo y D_LNVAI_PRI_2, apareciendo en los mismos tres modelos que D_LNVAI_PRI_2_VD pero con signo positivo.

En cuanto a las variables paisajísticas, la variable U_CU_HERP es la que ha sido incluida en la mayor cantidad de modelos, en concreto en seis, y siempre con signo positivo, poniendo de manifiesto el riesgo para los cultivos próximos del hábitat reservorio del tipo alfalfas y otros herbáceos plurianuales.

De las agronómicas, A_REGADIO ha vuelto a estar presente en tres de los modelos, dos de ellos referentes a las primaveras, en los tres con signo positivo.

Respecto a las variables climáticas, C_DINUM_30_PRI_1 vuelve a ser la relacionada con temperaturas que más se ha incluido, un total de cuatro veces y, de forma análoga a los modelos en vías, con signo negativo. Siendo la variable indicadora el número de días por encima de 30° C, vuelve a ponerse de manifiesto que el exceso de calor durante la primavera no parece favorecer los incrementos de topillo. En este caso, una variable relacionada con precipitación si parece más representativa, C_PREC_ACU_3, apareciendo en cuatro de los modelos, los dos globales y los dos estacionales de los periodos frío, con signo positivo, sugiriendo que una alta cantidad de precipitación acumulada en los periodos frío aumenta la abundancia.

4.5. Selección y validación de la metodología y algoritmo que proporcione el mejor ajuste a la variable dependiente observada

El objetivo de este apartado es la selección de uno de los modelos previamente ajustados como modelo final a ser propuesto, en base a su rendimiento en un subconjunto de test espacial (TEST1) y otro subconjunto de test temporal (TEST2), tal y como se ha especificado anteriormente. Para

ello, se analizará la raíz cuadrada del error cuadrático medio (RMSE) de cada modelo sobre dichos subconjuntos, para el total de los datos y para cada uno de los periodos (primavera, estío y frío).

4.5.1. Vías de dispersión

Tabla 16: RMSE de todos modelos de vías de dispersión sobre el conjunto total de los datos de TEST1 y subconjuntos estacionales.

	Modelo			
	Forward		Backward	
	GLOBAL	ESTACIONAL	GLOBAL	ESTACIONAL
Datos				
TODOS	0.724	0.682	0.694	0.671
PRI	0.681	0.702	0.662	0.673
EST	0.718	0.750	0.686	0.693
FRI	0.765	0.573	0.732	0.642

Tabla 17: RMSE de todos modelos de vías de dispersión sobre el conjunto total de los datos de TEST2 y subconjuntos estacionales.

	Modelo			
	Forward		Backward	
	GLOBAL	ESTACIONAL	GLOBAL	ESTACIONAL
Datos				
TODOS	0.686	0.776	0.958	0.857
PRI	0.745	0.772	1.261	0.715
EST	0.613	0.796	0.759	1.036
FRI	0.692	0.759	0.767	0.788

En las Tablas 16 y 17 aparecen los resultados para los subconjuntos de test espacial y temporal respectivamente. En función de ello se puede comentar:

- La metodología Backward parece haber ajustado mejor en general para TEST1, salvo para el caso estacional del periodo frío. El modelo global aplicado a las estaciones parece mejorar incluso a las propias predicciones desarrolladas con sus modelos específicos, salvo para el caso del periodo frío.
- Para el caso del TEST2, la metodología Forward ha resultado la que mejor ha ajustado para todos los casos y de forma relativamente clara. En este caso, considerando la metodología Forward, el modelo global aporta los mejores ajustes, tanto para el conjunto de todos los datos como para cada una de las estaciones.

A la vista de los resultados tanto para el subconjunto TRAIN como para los subconjuntos de TEST, no se puede establecer con consistencia cuál de las metodologías Forward o Backward ha aportado mejores ajustes generales, ni tampoco cuál de los modelos, global o estacionales.

En consecuencia, y si bien se matiza que es a efectos del presente TFG, pero que debe seguir profundizándose en la bondad de cada una de las combinaciones presentadas, se ha tomado el

criterio de considerar aquel modelo que más se ciñe a la necesidad y uso previsto, es decir, la predicción futura de eventos. Dado que el conjunto de test que mejor simula una situación real de predicción es el TEST2, por el hecho de que son datos provenientes de periodos desconocidos, se le ha considerado con un mayor peso a la hora de tomar la decisión sobre modelo finalmente propuesto. Considerando que para TEST2 el modelo que mejor ha ajustado en todos los casos es el Forward Global, será este el finalmente propuesto como resultado del presente estudio para el caso de las vías de dispersión.

4.5.2. Cultivos

Tabla 18: RMSE de todos modelos de cultivos sobre el conjunto total de los datos de TEST1 y subconjuntos estacionales.

	Modelo			
	Forward		Backward	
Datos	GLOBAL	ESTACIONAL	GLOBAL	ESTACIONAL
TODOS	0.307	0.320	0.307	0.324
PRI	0.262	0.279	0.261	0.256
EST	0.335	0.340	0.351	0.403
FRI	0.321	0.338	0.304	0.315

Tabla 19: RMSE de todos modelos de cultivos sobre el conjunto total de los datos de TEST2 y subconjuntos estacionales.

	Modelo			
	Forward		Backward	
Datos	GLOBAL	ESTACIONAL	GLOBAL	ESTACIONAL
TODOS	0.606	0.672	0.669	0.670
PRI	0.604	0.601	0.627	0.561
EST	0.601	0.699	0.704	0.743
FRI	0.612	0.711	0.675	0.694

En las Tablas 18 y 19 aparecen los resultados para los dos subconjuntos de test. En función de ello se puede comentar:

- En el caso del TEST1, no hay grandes diferencias entre las dos metodologías, aunque los modelos Forward parecen ajustar mejor los estíos y los Backward las primaveras y los fríos. Por lo general, los modelos globales funcionan mejor que los estacionales.
- Para el TEST2, sí encontramos diferencias notables en el rendimiento del modelo Forward Global, muy superior al resto, destacando su desempeño en los estíos y en los fríos. En cuanto a las primaveras, el modelo Backward Estacional es el que mejor resultados proporciona.

Por las mismas razones que las explicadas anteriormente para el caso del modelo de las vías de dispersión, se le ha asignado una mayor importancia a la hora de tomar la decisión sobre

el modelo final a escoger a los resultados sobre el conjunto TEST2. Debido a esto, y teniendo en cuenta la mejoría de rendimiento que supone el modelo Forward Estacional respecto a los demás sobre este conjunto de datos, dicho modelo es el propuesto para el hábitat de parcelas de cultivo, coincidiendo la metodología con el caso de las vías de dispersión.

4.5.3. Comprobación de las suposiciones del modelo lineal en ambos modelos

En la Tabla 20 se pueden apreciar los p-valores para los test de heterocedasticidad de Breusch-Pagan y de Varianza No Constante.

Tabla 20: Resultados de pruebas de heterocedasticidad para ambos modelos.

	Breusch-Pagan	Varianza No Constante
Vías de dispersión	6.2E-11	4.1282E-05
Cultivos	2.051E-15	2.4456E-05

Todos los p-valores son significativos a nivel 0.05, es decir, se rechaza la hipótesis nula de homocedasticidad y se concluye que sí hay heterocedasticidad.

Además, se han realizado tests de Durbin-Watson para detectar la presencia de autocorrelación en los residuos del modelo para cada una de las zonas (Tabla 31 del Anexo 7.9), obteniéndose para las vías de dispersión en 23 de los casos p-valores inferiores a 0.05 y para los cultivos 11 de las 32 con p-valores inferiores a 0.05. Por tanto se rechazan las hipótesis nulas de independencia de los errores entre zonas.

También se han realizado tests de la I de Moran para cada periodo y hábitat, cuyos p-valores se pueden consultar en la Tabla 32 del Anexo 7.10. En el caso de las vías de dispersión, ninguno de los p-valores ha resultado significativo a nivel 0.05. Para los cultivos, de los 21 periodos analizados, en tres (periodos 20, 34 y 38) se obtuvieron p-valores inferiores a 0.05, lo que indica la presencia de autocorrelación espacial en los residuos. En el resto de periodos no se rechaza la hipótesis nula de ausencia de autocorrelación espacial. Por tanto, dado que la mayoría de los periodos no presentan dicha autocorrelación, se considera que ambos modelos captan adecuadamente la estructura espacial en términos generales.

Se ha comprobado la normalidad del modelo estadísticamente mediante la prueba de Lilliefors y visualmente mediante la representación de los cuantiles de los residuos con los cuantiles de una distribución normal teórica, gráfico conocido como QQ-plot.

En cuanto a los resultados del test, se han obtenido p-valores de 6.693E-06 y 2.2E-16 para los residuos de los modelos de vías de dispersión y cultivos respectivamente, rechazando rotundamente la normalidad, conclusión que apoyan los QQ-plots correspondientes (ver Figura 5).

Para el cálculo de los errores estándar se utilizaron errores estándar robustos por conglomerados con ajuste BRL (Bias-Reduced Linearization o reducción del sesgo en la linealización), con el fin de corregir la heterocedasticidad detectada y la falta de independencia en los residuos. Se ha utilizado el ajuste de Satterthwaite para los grados de libertad.

En base a dichas correcciones, se obtienen los resultados mostrados en las Tablas 21 y 22.

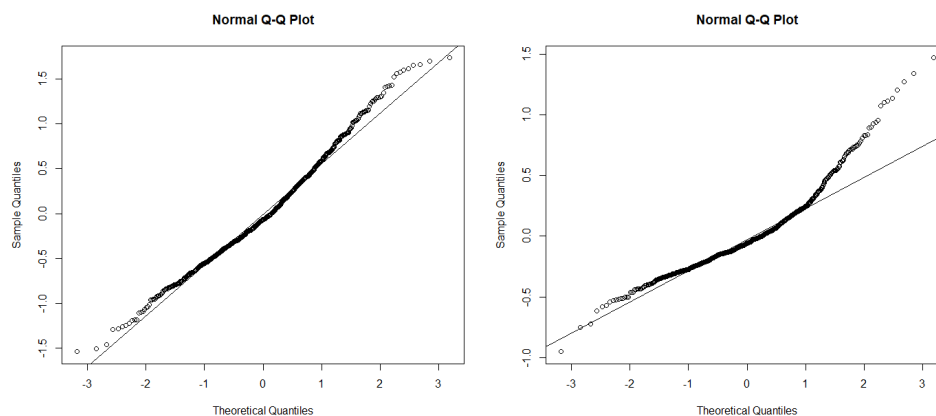


Figura 5: QQ-plots de normalidad para los residuos de ambos modelos. Fuente: elaboración propia.

Tabla 21: Estimaciones y resultados estadísticos del modelo de vías de dispersión.

Coefficiente	Est.	Error est.	Estadístico t	G. l.	P-valor	SS
(Intercept)	-7.886	0.959	-8.22	26.18	<0.001	31.869
D_LNVAI_EST_3	-0.145	0.038	-3.83	23.32	<0.001	7.039
D_LNVAI_FRI_2	0.260	0.040	6.55	25.54	<0.001	21.773
D_LNVAI_FRI_3	0.190	0.039	4.83	25.95	<0.001	9.009
D_LNVAI_PRI_4	0.197	0.034	5.72	26.10	<0.001	10.071
D_LNVAI_T3	-1.031	0.057	-18.24	27.15	<0.001	262.173
C_DIHELA_0_FRI_1	0.013	0.003	4.07	25.93	<0.001	7.131
C_DIHELA_0_PRI_1	0.013	0.005	2.47	25.10	0.0208	4.476
C_DINUM_30_EST_1	-0.024	0.005	-4.71	25.42	<0.001	14.400
C_DINUM_30_PRI_1	-0.052	0.005	-10.42	23.67	<0.001	27.927
C_TMED_MED_EST_1	0.369	0.062	5.96	27.36	<0.001	22.848
C_TMED_MED_FRI_1	0.320	0.067	4.78	23.54	<0.001	9.749
U_AG_LONG	5.775	2.228	2.59	7.87	0.0324	4.262

Est.: estimación del coeficiente; Error est.: estimación del error estándar; G. l.: Grados de libertad; SS: suma de cuadrados

En el modelo de las vías de dispersión, la variable con una mayor suma de cuadrados asociada es D_LNVAI, con una diferencia muy grande respecto al resto. Además, aparecen otras variables denso-dependientes estacionales referentes a periodos aproximadamente desde 2 hasta 4 años atrás, destacando D_LNVAI_FRI_2 en cuanto a relevancia.

Ninguna de las variables climáticas asociadas a las precipitaciones forma parte de este modelo, al revés que en el caso de las relacionadas con la temperatura, estando 6 de ellas en el modelo. Destacan C_DINUM_30_PRI_1 y C_TMED_MED_EST_1, pues solo D_LNVAI_T3 tiene una suma de cuadrados asociada mayor.

Del resto de variables, solo aparece U_AG_LONG. Esta variable representa la longitud de vías dispersión asociadas a cauces de agua y ha sido incluida con signo positivo, denotando que este tipo particular de vías de dispersión puede ser especialmente relevante en la dinámica poblacional en cuanto a favorecer los incrementos de abundancia.

En el caso de las parcelas de cultivo, se da el mismo caso con D_LNVAI_T3 que en el

Tabla 22: Estimaciones y resultados estadísticos del modelo de cultivos.

Coefficiente	Est.	Error est.	Estadístico t	G. l.	P-valor	SS
(Intercept)	-1.1658	0.427328	-2.73	20.69	0.01271	1.317
D_LNVAI_EST_3	-0.14762	0.02431	-6.07	17.35	<0.001	2.822
D_LNVAI_FRI_2	0.1896	0.041928	4.52	18.98	<0.001	2.754
D_LNVAI_FRI_3_VD	0.06748	0.023528	2.87	27.16	0.00789	1.389
D_LNVAI_PRI_2	0.23133	0.057834	4.00	21.17	<0.001	1.918
D_LNVAI_PRI_2_VD	-0.07463	0.024872	-3.00	25.79	0.00591	0.907
D_LNVAI_T3	-1.09769	0.054753	-20.05	19.98	<0.001	99.170
C_DINUM_30_3	0.00434	0.000871	4.98	27.89	<0.001	3.132
C_DINUM_30_PRI_1	-0.02226	0.002916	-7.63	24.49	<0.001	5.395
C_PREC_ACU_3	0.00111	0.000279	3.97	27.51	<0.001	1.842
C_TMED_MED_EST_1	0.10277	0.021815	4.71	21.54	<0.001	4.657
U_AG_LONG	3.23478	1.190973	2.72	7.26	0.02892	1.474
U_CU_HERP	0.01838	0.004826	3.81	12.88	0.00221	4.182

Est.: estimación del coeficiente; Error est.: estimación del error estándar; G. l.: Grados de libertad; SS: suma de cuadrados

anterior modelo, siendo esta la variable más importante para el ajuste. El resto de las variables denso-dependientes no son especialmente relevantes, aunque es de interés remarcar que dos de las variables asociadas a la abundancia en vías de dispersión aparecen en el modelo, mostrando la relación previamente comentada.

Sí aparece en este caso (aunque con una suma de cuadrados pequeña) una variable asociada a las precipitaciones (C_PREC_ACU_3). Respecto a las relativas a las temperaturas, aparecen 3 de ellas: C_DINUM_30_3, C_DINUM_30_PRI_1 y C_TMED_MED_EST_1.

Repite U_AG_LONG como variable paisajística, denotando su importancia para la colonización posterior de los cultivos. En esta ocasión U_CU_HERP sí ha entrado en el modelo, de hecho con la tercera mayor suma de cuadrados de entre todas las variables, lo que significa que es especialmente relevante para el ajuste, concordando con lo ya comentado en otros apartados. Los coeficientes de determinación de los modelos finales sobre cada conjunto de test se pueden apreciar en la Tabla 23.

Tabla 23: Valores de R^2 de cada modelo para cada conjunto de datos

	Vías de dispersión	Cultivos
TRAIN	0.644	0.670
TEST1	0.684	0.750
TEST2	0.559	0.138

5. Conclusiones

1. Tanto las componentes aditivas de los tres factores (ZV, periodo y hábitat) considerados, como las interacciones entre cada uno de ellos, han resultado significativas en cuanto a su contribución a la variabilidad espacio-temporal de la abundancia del topillo campesino, lo que pone de manifiesto la complejidad encerrada en la expresión de sus fluctuaciones demográficas en Castilla y León.
2. En una descripción preliminar de dicha complejidad, se han detectado seis eventos generales de incrementos demográficos en el periodo de estudio, con diferenciales de expresión entre ellos en cuanto a duración y número de zonas y hábitat afectados. Asimismo, en cada uno de ellos se han detectado diferencias en su duración entre zonas, así como asincronía entre ellas en el inicio de su expresión, que puede llegar a ser mayor de un año.
3. La colonización en los cultivos no se ha detectado en todos los casos en que los incrementos se han expresado en las vías de dispersión y, cuando lo ha hecho, ha ocurrido generalmente con posterioridad y siempre con valores menores. En consecuencia, la colonización de los cultivos muestra tener como precursor a los incrementos de abundancia en vías de dispersión, pero no siempre que estos ocurren se traducen en afectación agrícola.
4. Se ha profundizado en los patrones cíclicos presentes en Castilla y León, habiéndose completado su descripción en zonas que previamente no se habían identificado, confirmado la existencia de periodicidades en los rangos de 2-3, 3-4 y 4-5 años y detectado una nueva periodicidad de 6-7 años no descrita previamente en la zona de estudio.
5. La distribución geográfica de los rangos de periodicidad, si bien no se resuelve de forma perfecta, ha mostrado una tendencia a la ocurrencia de las periodicidades más bajas vinculadas a Tierra de Campos, las más altas al sureste de la comunidad, las de 4-5 años a la zona norte y las de 3-4 años al suroeste.
6. La componente aditiva de la abundancia en ambos hábitats se ha correlacionado con el porcentaje del territorio cubierto por alfalfa y con el cubierto por pastizal arbolado o arbustivo, en el primer caso de forma positiva y en el segundo negativa. Si bien la primera resultaba esperable, la segunda abre una interesante hipótesis sobre si este tipo de pastizales pudieran constituir zonas de compensación ecológica con efecto tamponador sobre los riesgos agrícolas.
7. La colonización en las parcelas de cultivo se ha correlacionado positivamente con el porcentaje de parcelas en que se realiza siembra directa, corroborando la problemática que puede conllevar esta práctica agronómica en lo referente a riesgos agrícolas por topillo campesino.

8. Se han encontrado interesantes correlaciones entre la colonización en los cultivos con factores texturales del suelo, de forma que los riesgos agrícolas parecen ser mayores cuanto más finos son dichos elementos, lo que favorece una mejor captación del agua, y en consecuencia un mejor desarrollo de la cubierta vegetal, y, en paralelo, aportaría mejores condiciones para la estabilidad de las madrigueras.
9. Las condiciones generales climatológicas que pudieran modular la expresión de la abundancia refieren tendencia a mayor abundancia en precedentes otoño-invernales y primaverales suaves (no excesivamente fríos los primeros y no excesivamente cálidas las segundas) y favorecido por mayores precipitaciones durante los periodos estival y otoño-invernal precedentes.
10. Considerando las metodologías aplicadas en los modelos globales y estacionales no se puede desprender cual pudiera ser más acorde para su propuesta final, sin embargo, bajo el criterio de su aplicación prevista de predicción de eventos futuros, ha sido la metodología forward sin separar por estaciones la que mejor ajuste proporciona, tanto para el hábitat vías de dispersión como para el de cultivos.
11. El modelo propuesto para el hábitat vías de dispersión incluye 12 variables, de las cuales 5 están relacionadas con denso-dependencia, 1 con características paisajísticas y 6 con factores climatológicos.
12. El modelo propuesto para el hábitat cultivos ha incluido 12 variables, 6 relacionadas con denso-dependencia, 2 con características paisajísticas y 4 con factores climatológicos.

6. Bibliografía

- [1] M. C. H. González, B. Sánchez-González, Á. Navarro-Castilla e I. Barja. “Topillo campesino - *Microtus arvalis* (Pallas, 1779).” Accedido el 2 de mayo de 2025. (mar. de 2017), dirección: <https://www.vertebradosibericos.org/mamiferos/micarv.html>.
- [2] J. Jacob, P. Manson, R. Barfknecht y T. Fredricks, “Common vole (*Microtus arvalis*) ecology and management: implications for risk assessment of plant protection products: Common voles in the risk assessment of plant protection products,” *Pest Management Science*, vol. 70, n.º 6, págs. 869-878, ene. de 2014, ISSN: 1526-498X. DOI: 10.1002/ps.3695. dirección: <http://dx.doi.org/10.1002/ps.3695>.
- [3] J. González-Esteban e I. Villate, “*Microtus arvalis* (Pallas, 1778),” en *Atlas y Libro Rojo de los Mamíferos Terrestres de España*, L. J. Palomo, J. Gisbert y J. C. Blanco, eds., Madrid: Dirección General para la Biodiversidad -SECEM-SECEMU, 2007, págs. 426-428.
- [4] C. Caminero-Saldaña, J. J. Luque-Larena, M. Fernández-Villán et al., “Can population structure derived variables be useful for predictive modelling of demographic fluctuations of common voles?” es, en *Proceedings of the 17th Rodent Spat - International Conference on Rodent Biology*, S. Herrero-Cófreces y J. J. Luque-Larena, eds., Poster presentado en la conferencia, Valladolid, Spain: University of Valladolid, 2023, pág. 37, ISBN: 978-84-09-51624-7. dirección: <http://hdl.handle.net/10261/145911>.
- [5] J. Jacob y E. Tkadlec, “Rodent outbreaks in Europe: dynamics and damage,” en *Rodent Outbreaks: Ecology and Impacts*, G. Singleton, S. Belmain, P. Brown y B. Hardy, eds., International Rice Research Institute (IRRI), 2010, págs. 207-223, ISBN: 978-971-22-0257-5.
- [6] F. Keesing y R. S. Ostfeld, “Emerging patterns in rodent-borne zoonotic diseases,” *Science*, vol. 385, n.º 6715, págs. 1305-1310, sep. de 2024, ISSN: 1095-9203. DOI: 10.1126/science.adq7993. dirección: <http://dx.doi.org/10.1126/science.adq7993>.
- [7] R. Rodríguez-Pastor, J. J. Luque-Larena, X. Lambin y F. Mougeot, ““Living on the edge”: The role of field margins for common vole (*Microtus arvalis*) populations in recently colonised Mediterranean farmland,” *Agriculture, Ecosystems & Environment*, vol. 231, págs. 206-217, sep. de 2016, ISSN: 0167-8809. DOI: 10.1016/j.agee.2016.06.041. dirección: <http://dx.doi.org/10.1016/j.agee.2016.06.041>.
- [8] S. Herrero-Cófreces, F. Mougeot, X. Lambin y J. J. Luque-Larena, “Linking Zoonosis Emergence to Farmland Invasion by Fluctuating Herbivores: Common Vole Populations and Tularemia Outbreaks in NW Spain,” *Frontiers in Veterinary Science*, vol. 8, ago. de 2021, ISSN: 2297-1769. DOI: 10.3389/fvets.2021.698454. dirección: <http://dx.doi.org/10.3389/fvets.2021.698454>.

- [9] D. Jareño, J. Viñuela, J. J. Luque-Larena, L. Arroyo, B. Arroyo y F. Mougeot, “Factors associated with the colonization of agricultural areas by common voles *Microtus arvalis* in NW Spain,” *Biological Invasions*, vol. 17, n.º 8, págs. 2315-2327, mar. de 2015, ISSN: 1573-1464. DOI: 10.1007/s10530-015-0877-4. dirección: <http://dx.doi.org/10.1007/s10530-015-0877-4>.
- [10] E. Tkadlec y N. C. Stenseth, “A new geographical gradient in vole population dynamics,” *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 268, n.º 1476, págs. 1547-1552, ago. de 2001, ISSN: 1471-2954. DOI: 10.1098/rspb.2001.1694. dirección: <http://dx.doi.org/10.1098/rspb.2001.1694>.
- [11] A. Brügger, W. Nentwig y J.-P. Airoldi, “The burrow system of the common vole (*M. arvalis*, Rodentia) in Switzerland,” *mamm*, vol. 74, n.º 3, págs. 311-315, ago. de 2010, ISSN: 0025-1461. DOI: 10.1515/mamm.2010.035. dirección: <http://dx.doi.org/10.1515/mamm.2010.035>.
- [12] M. Heroldová, J. Šipoš, J. Suchomel y J. Zejda, “Influence of crop type on common vole abundance in Central European agroecosystems,” *Agriculture, Ecosystems & Environment*, vol. 315, pág. 107 443, ago. de 2021, ISSN: 0167-8809. DOI: 10.1016/j.agee.2021.107443. dirección: <http://dx.doi.org/10.1016/j.agee.2021.107443>.
- [13] D. Roos, C. Caminero Saldaña, B. Arroyo, F. Mougeot, J. J. Luque-Larena y X. Lambin, “Unintentional effects of environmentally-friendly farming practices: Arising conflicts between zero-tillage and a crop pest, the common vole (*Microtus arvalis*),” *Agriculture, Ecosystems & Environment*, vol. 272, págs. 105-113, feb. de 2019, ISSN: 0167-8809. DOI: 10.1016/j.agee.2018.11.013. dirección: <http://dx.doi.org/10.1016/j.agee.2018.11.013>.
- [14] C. Tougaard, E. Renvoisé, A. Petitjean y J.-P. Quéré, “New Insight into the Colonization Processes of Common Voles: Inferences from Molecular and Fossil Evidence,” *PLoS ONE*, vol. 3, n.º 10, M. Hofreiter, ed., e3532, oct. de 2008, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0003532. dirección: <http://dx.doi.org/10.1371/journal.pone.0003532>.
- [15] J. J. Luque-Larena, F. Mougeot, J. Viñuela et al., “Recent large-scale range expansion and outbreaks of the common vole (*Microtus arvalis*) in NW Spain,” *Basic and Applied Ecology*, vol. 14, n.º 5, págs. 432-441, ago. de 2013, ISSN: 1439-1791. DOI: 10.1016/j.baae.2013.04.006. dirección: <http://dx.doi.org/10.1016/j.baae.2013.04.006>.
- [16] Consejería de Agricultura y Ganadería, *Orden AYG/96/2019, de 5 de febrero, por la que se establece la estrategia de gestión integrada de riesgos derivados de la presencia de topillo campesino - *Microtus arvalis* (Pallas) en el territorio de Castilla y León*, *Boletín Oficial de Castilla y León*, núm. 27, 8 de febrero de 2019, pp. 5679–5689, Disponible en: <https://bocyl.jcyl.es/> [Consulta: 11 de junio de 2025], 2019.

- [17] X. LAMBIN, V. BRETAGNOLLE y N. G. YOCCOZ, “Vole population cycles in northern and southern Europe: Is there a need for different explanations for single pattern?” *Journal of Animal Ecology*, vol. 75, n.º 2, págs. 340-349, mar. de 2006, ISSN: 1365-2656. DOI: 10.1111/j.1365-2656.2006.01051.x. dirección: <http://dx.doi.org/10.1111/j.1365-2656.2006.01051.x>.
- [18] F. Mougeot, X. Lambin, R. Rodríguez-Pastor, J. Romairone y J.-J. Luque-Larena, “Numerical response of a mammalian specialist predator to multiple prey dynamics in Mediterranean farmlands,” *Ecology*, vol. 100, n.º 9, ago. de 2019, ISSN: 1939-9170. DOI: 10.1002/ecy.2776. dirección: <http://dx.doi.org/10.1002/ecy.2776>.
- [19] C. Caminero-Saldaña, J. P. Correa-Cuadros, A. Baños-Herrero et al., “Exploring the influence of density-dependence and weather on the spatial and temporal variation in common vole (*Microtus arvalis*) abundance in Castilla y León, NW Spain,” *Pest Management Science*, vol. 80, n.º 11, págs. 5527-5536, ene. de 2024, ISSN: 1526-4998. DOI: 10.1002/ps.7954. dirección: <http://dx.doi.org/10.1002/ps.7954>.
- [20] J. Jacob, C. Imholt, C. Caminero-Saldaña et al., “Europe-wide outbreaks of common voles in 2019,” *Journal of Pest Science*, vol. 93, n.º 2, págs. 703-709, ene. de 2020, ISSN: 1612-4766. DOI: 10.1007/s10340-020-01200-2. dirección: <http://dx.doi.org/10.1007/s10340-020-01200-2>.
- [21] D. Roos, C. Caminero-Saldaña, D. Elston et al., “From pattern to process? Dual travelling waves, with contrasting propagation speeds, best describe a self-organised spatio-temporal pattern in population growth of a cyclic rodent,” *Ecology Letters*, vol. 25, n.º 9, págs. 1986-1998, jul. de 2022, ISSN: 1461-0248. DOI: 10.1111/ele.14074. dirección: <http://dx.doi.org/10.1111/ele.14074>.
- [22] N. Sánchez, J. Plaza, C. García-Ariza et al., “Analyzing the spatiotemporal habitat colonization dynamics of the common vole (*Microtus arvalis* Pallas) in Castilla y León, northwest Spain, using a hotspot-based approach,” *Pest Management Science*, vol. 80, n.º 11, págs. 5584-5596, jul. de 2024, ISSN: 1526-4998. DOI: 10.1002/ps.8344. dirección: <http://dx.doi.org/10.1002/ps.8344>.
- [23] Ministerio de Agricultura, Pesca y Alimentación, *Real Decreto 409/2008, de 28 de marzo, por el que se establece el programa nacional de control de las plagas del topillo de campo, «Microtus arvalis» (Pallas), y otros microtinios*, *Boletín Oficial del Estado*, núm. 86, 9 de abril de 2008, Referencia: BOE-A-2008-6244, 2008. dirección: <https://www.boe.es/eli/es/rd/2008/03/28/409>.
- [24] Observatorio de Plagas y Enfermedades Agrícolas de Castilla y León, *Recomendaciones de manejo: periodo otoño-invernal*, En *Fichas de apoyo de plagas y enfermedades: Topillo campesino*, Instituto Tecnológico Agrario de Castilla y León (ed), Código: TO-GI-10, 2023. dirección: <https://plagas.itacyl.es/topillo-campesino>.

- [25] R. Bernardo-Madrid, "Population Dynamics in Voles: Characterization and Modelling of Global Spatio-Temporal Patterns," PhD Thesis, University of Seville, Seville, Spain, 2021.
- [26] A. Esther, C. Imholt, J. Perner, J. Schumacher y J. Jacob, "Correlations between weather conditions and common vole (*Microtus arvalis*) densities identified by regression tree analysis," *Basic and Applied Ecology*, vol. 15, n.º 1, págs. 75-84, ene. de 2014, ISSN: 1439-1791. DOI: 10.1016/j.baae.2013.11.003. dirección: <http://dx.doi.org/10.1016/j.baae.2013.11.003>.
- [27] P. Delattre, P. Giraudoux, J. Baudry, P. Musard, M. Toussaint, D. Truchetet et al., "Land use patterns and types of common vole (*Microtus arvalis*) population kinetics," *Agriculture, Ecosystems & Environment*, vol. 39, n.º 2, págs. 153-168, 1992.
- [28] X. Lambin, S. J. Petty y J. L. Mackinnon, "Cyclic dynamics in field vole populations and generalist predation," *Journal of Animal Ecology*, vol. 69, n.º 1, págs. 106-119, 2000.
- [29] P. Turchin y G. O. Batzli, "AVAILABILITY OF FOOD AND THE POPULATION DYNAMICS OF ARVICOLINE RODENTS," *Ecology*, vol. 82, n.º 6, págs. 1521-1534, jun. de 2001, ISSN: 0012-9658. DOI: 10.1890/0012-9658(2001)082[1521:aofatp]2.0.co;2. dirección: [http://dx.doi.org/10.1890/0012-9658\(2001\)082%5B1521:AOFATP%5D2.0.CO;2](http://dx.doi.org/10.1890/0012-9658(2001)082%5B1521:AOFATP%5D2.0.CO;2).
- [30] B. F. Blank, J. Jacob, A. Petri y A. Esther, "Topography and soil properties contribute to regional outbreak risk variability of common voles (*Microtus arvalis*)," *Wildlife Research*, vol. 38, n.º 7, pág. 541, 2011, ISSN: 1035-3712. DOI: 10.1071/wr10192. dirección: <http://dx.doi.org/10.1071/WR10192>.
- [31] P. Giraudoux, P. Villette, J.-P. Quéré, J.-P. Damange y P. Delattre, "Weather influences *M. arvalis* reproduction but not population dynamics in a 17-year time series," *Scientific Reports*, vol. 9, n.º 1, sep. de 2019, ISSN: 2045-2322. DOI: 10.1038/s41598-019-50438-z. dirección: <http://dx.doi.org/10.1038/s41598-019-50438-z>.
- [32] M. K. Oli, "Population cycles in voles and lemmings: state of the science and future directions," *Mammal Review*, vol. 49, n.º 3, págs. 226-239, mayo de 2019, ISSN: 1365-2907. DOI: 10.1111/mam.12156. dirección: <http://dx.doi.org/10.1111/mam.12156>.
- [33] H. P. Andreassen, J. Sundell, F. Ecke et al., "Population cycles and outbreaks of small rodents: ten essential questions we still need to solve," *Oecologia*, vol. 195, n.º 3, págs. 601-622, dic. de 2020, ISSN: 1432-1939. DOI: 10.1007/s00442-020-04810-w. dirección: <http://dx.doi.org/10.1007/s00442-020-04810-w>.
- [34] C. Imholt, A. Esther, J. Perner y J. Jacob, "Identification of weather parameters related to regional population outbreak risk of common voles (*Microtus arvalis*) in Eastern Germany," *Wildlife Research*, vol. 38, n.º 7, pág. 551, 2011, ISSN: 1035-3712. DOI: 10.1071/wr10196. dirección: <http://dx.doi.org/10.1071/WR10196>.

- [35] P. Delattre, P. Giraudoux, J. Baudry, J. P. Quéré y E. Fichet, “Effect of landscape structure on Common Vole (*Microtus arvalis*) distribution and abundance at several space scales,” *Landscape Ecology*, vol. 11, n.º 5, págs. 279-288, oct. de 1996, ISSN: 1572-9761. DOI: 10.1007/bf02059855. dirección: <http://dx.doi.org/10.1007/BF02059855>.
- [36] A. Baños-Herrero, Y. A. Pallavicini-Fernández, M. C. García-Ariza et al., “Prospecting the effects of the landscape structure on the spatio-temporal evolution of the abundance of the common vole in Castilla y León, Spain,” es, en *Proceedings of the 17th Rodens Spat - International Conference on Rodent Biology*, S. Herrero-Cófreces y J. J. Luque-Larena, eds., Valladolid, Spain: University of Valladolid, 2023, pág. 26, ISBN: 978-84-09-51624-7.
- [37] J. Fotheringham, K. Rijdsdijk e Y. Artzy-Randrup, “How landscape heterogeneity influences the outbreak dynamics of the common vole (*Microtus arvalis*) in European farmlands: A review,” oct. de 2024. DOI: 10.3897/arphapreprints.e139043. dirección: <http://dx.doi.org/10.3897/arphapreprints.e139043>.
- [38] Instituto Geográfico Nacional. “Datos de población por comunidad autónoma.” Consulta: 11 de junio de 2025. (2025), dirección: <https://www.ign.es/web/ane-datos-geograficos/-/datos-geograficos/datosPoblacion?tipoBusqueda=CCAA>.
- [39] D. A. N. García, N. G. del Pozo, M. V. Á. Arias et al., *Atlas Agroclimático de Castilla y León*, 1.ª ed. Junta de Castilla y León: Instituto Tecnológico Agrario de Castilla y León (ITACyL), sep. de 2013, Colaboración con la Agencia Estatal de Meteorología (AEMET). Disponible en línea bajo licencia Creative Commons BY-SA 3.0 ES, ISBN: NIPO digital 281-13-008-5. dirección: <http://atlas.itacyl.es>.
- [40] Junta de Castilla y León, *ACUERDO 53/2009, de 14 de mayo, de la Junta de Castilla y León, por el que se aprueba el Plan Director de Lucha contra Plagas Agrícolas en Castilla y León*, <https://bocyl.jcyl.es/>, Publicado en el Boletín Oficial de Castilla y León, n.º 103, Miércoles 3 de junio de 2009, 2009.
- [41] P. Delattre, P. Giraudoux, J.-P. Damange y J.-P. Quéré, “Recherche d’un indicateur de la cinétique démographique des populations du Campagnol des champs (*Microtus arvalis*),” *Revue d’Écologie (La Terre et La Vie)*, vol. 45, n.º 4, págs. 375-384, 1990, ISSN: 0249-7395. DOI: 10.3406/revec.1990.2013. dirección: <http://dx.doi.org/10.3406/revec.1990.2013>.
- [42] C. C. Saldaña, X. Lambin, F. Mougeot y B. Arroyo, *A contribution for the calibration of an abundance index useful for large-scale monitoring of common vole*, Poster presentado en la 10th European Vertebrate Pest Management Conference (EVPMC), Sevilla, España, Consultado el 17 de junio de 2025, sep. de 2015. dirección: <http://hdl.handle.net/10261/145911>.

- [43] J. Romairone, J. Jiménez, J. J. Luque-Larena y F. Mougeot, “Spatial capture-recapture design and modelling for the study of small mammals,” *PLOS ONE*, vol. 13, n.º 6, A. Margalida, ed., e0198766, jun. de 2018, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0198766. dirección: <http://dx.doi.org/10.1371/journal.pone.0198766>.
- [44] A. del Hierro Diez, *Comparación y calibrado de dos métodos de estimación de la abundancia del topillo campesino (Microtus arvalis Pallas)*, es, Trabajo Fin de Grado, Grado en Estadística, Universidad de Valladolid, Facultad de Ciencias, Director: Agustín Mayo Íscar. Co-director: Constantino Caminero Saldaña, 2020. dirección: <http://uvadoc.uva.es/handle/10324/43851>.
- [45] T. Royama, *Analytical Population Dynamics*. Springer Netherlands, 1992, ISBN: 9789401129169. DOI: 10.1007/978-94-011-2916-9. dirección: <http://dx.doi.org/10.1007/978-94-011-2916-9>.
- [46] R Core Team, “R: A Language and Environment for Statistical Computing,” 2022. dirección: <https://www.R-project.org/>.
- [47] R Core Team, *The R Project for Statistical Computing: About R*, Consultado el 10 de junio de 2025, 2024. dirección: <https://www.r-project.org/about.html>.
- [48] TIOBE Software BV, *TIOBE Index for the Popularity of Programming Languages*, Consultado el 4 de junio de 2025, 2025. dirección: <https://www.tiobe.com/tiobe-index/>.
- [49] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016, ISBN: 978-3-319-24277-4. dirección: <https://ggplot2.tidyverse.org>.
- [50] H. Wickham, R. François, L. Henry, K. Müller y D. Vaughan, *dplyr: A Grammar of Data Manipulation*, R package version 1.1.4, 2023. dirección: <https://CRAN.R-project.org/package=dplyr>.
- [51] B. Auguie, *gridExtra: Miscellaneous Functions for "Grid" Graphics*, R package version 2.3, 2017. dirección: <https://CRAN.R-project.org/package=gridExtra>.
- [52] H. Wickham, *stringr: Simple, Consistent Wrappers for Common String Operations*, R package version 1.5.1, 2023. dirección: <https://CRAN.R-project.org/package=stringr>.
- [53] H. Wickham, D. Vaughan y M. Girlich, *tidyr: Tidy Messy Data*, R package version 1.3.1, 2024. dirección: <https://CRAN.R-project.org/package=tidyr>.
- [54] H. Wickham, M. Averick, J. Bryan et al., “Welcome to the tidyverse,” *Journal of Open Source Software*, vol. 4, n.º 43, pág. 1686, 2019. DOI: 10.21105/joss.01686.
- [55] C. Huang y A. Petukhina, *Applied Time Series Analysis and Forecasting with Python*. Springer International Publishing, 2022, ISBN: 9783031135842. DOI: 10.1007/978-3-031-13584-2. dirección: <http://dx.doi.org/10.1007/978-3-031-13584-2>.
- [56] M. Romer, *Lesson 6: The Periodogram*, Disponible en línea, Consultado el 17 de junio de 2025, 2023. dirección: <https://online.stat.psu.edu/stat510/lesson/6>.

- [57] M. H. Kutner, C. J. Nachtsheim, J. Neter y W. Li, *Applied Linear Statistical Models*, 5.^a ed. McGraw-Hill, 2004, Consultado el 17 de junio de 2025, ISBN: 9780071122214.
- [58] G. James, D. Witten, T. Hastie y R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*. Springer US, 2021, ISBN: 9781071614181. DOI: 10.1007/978-1-0716-1418-1. dirección: <http://dx.doi.org/10.1007/978-1-0716-1418-1>.
- [59] Wikimedia Foundation, *Prueba de Lilliefors*, Disponible en línea, Consultado el 17 de junio de 2025, 2025. dirección: https://es.wikipedia.org/wiki/Prueba_de_Lilliefors.
- [60] Wikimedia Foundation, *Q-Q plot*, Disponible en línea, Consultado el 17 de junio de 2025, 2025. dirección: https://en.wikipedia.org/wiki/Q%E2%80%93Q_plot.
- [61] Wikimedia Foundation, *Test de Breusch–Pagan*, Disponible en línea, Consultado el 17 de junio de 2025, 2025. dirección: https://es.wikipedia.org/wiki/Test_de_Breusch-Pagan.
- [62] R. D. COOK y S. WEISBERG, “Diagnostics for heteroscedasticity in regression,” *Biometrika*, vol. 70, n.º 1, págs. 1-10, 1983, ISSN: 1464-3510. DOI: 10.1093/biomet/70.1.1. dirección: <http://dx.doi.org/10.1093/BIOMET/70.1.1>.
- [63] Wikimedia Foundation, *Durbin–Watson statistic*, Disponible en línea, Consultado el 17 de junio de 2025, 2025. dirección: https://en.wikipedia.org/wiki/Durbin%E2%80%93Watson_statistic.
- [64] H. Li, C. A. Calder y N. Cressie, “Beyond Moran’sI: Testing for Spatial Dependence Based on the Spatial Autoregressive Model,” *Geographical Analysis*, vol. 39, n.º 4, págs. 357-375, sep. de 2007, ISSN: 1538-4632. DOI: 10.1111/j.1538-4632.2007.00708.x. dirección: <http://dx.doi.org/10.1111/j.1538-4632.2007.00708.x>.
- [65] Wikimedia Foundation, *Errores estándar consistentes con heterocedasticidad*, Disponible en línea, Consultado el 17 de junio de 2025, 2025. dirección: https://es.wikipedia.org/wiki/Heteroskedasticity-consistent_standard_errors.
- [66] J. E. Pustejovsky, *Cluster-robust standard errors and hypothesis tests in panel data models*, Trabajo en progreso o prepublicación, 31 de marzo de 2025, mar. de 2025.
- [67] Wikimedia Foundation, *Ecuación de Welch–Satterthwaite*, Disponible en línea, Consultado el 17 de junio de 2025, 2025. dirección: https://es.wikipedia.org/wiki/Ecuaci%C3%B3n_de_Welch%E2%80%93Satterthwaite.
- [68] X. Ying, “An Overview of Overfitting and its Solutions,” *Journal of Physics: Conference Series*, vol. 1168, pág. 022 022, feb. de 2019, ISSN: 1742-6596. DOI: 10.1088/1742-6596/1168/2/022022. dirección: <http://dx.doi.org/10.1088/1742-6596/1168/2/022022>.

- [69] B. Ghojogh y M. Crowley, *The Theory Behind Overfitting, Cross Validation, Regularization, Bagging, and Boosting: Tutorial*, ver. v2, Lecture by Benyamin Ghojogh: <https://www.youtube.com/watch?v=wds4KdXQJIA>; Lecture by Prof. Ali Ghodsi: <https://www.youtube.com/watch?v=21jL0I6wbns>, mayo de 2023. arXiv: 1905.12787 [stat.ML]. dirección: <https://arxiv.org/abs/1905.12787v2>.
- [70] Z. Zhu, “Systematic Optimization of Overfitting Problem in Machine Learning,” *Highlights in Science, Engineering and Technology*, vol. 111, págs. 353-359, ago. de 2024, ISSN: 2791-0210. DOI: 10.54097/3tkzrj84. dirección: <http://dx.doi.org/10.54097/3tkzrj84>.
- [71] I. T. A. de Castilla y León (ITACyL), *Boletines de predicción de cosechas*, <https://cosechas.itacyl.es/boletines>, Accedido el 12 de junio de 2025, 2025. dirección: <https://cosechas.itacyl.es/boletines>.
- [72] R. G. Barbier, *Análisis de la evolución y clasificación de la abundancia del topillo campesino (Microtus arvalis Pallas) en Castilla y León*, es, Trabajo Fin de Grado, Grado en Estadística, Universidad de Valladolid, Facultad de Ciencias, Director: Agustín Mayo Íscar. Co-director: Constantino Caminero Saldaña, jun. de 2022. dirección: <https://uvadoc.uva.es/handle/10324/57948>.
- [73] D. Hernangómez, *mapSpain: Administrative Boundaries of Spain*, ver. 0.10.0, 2025. DOI: 10.5281/zenodo.5366622. dirección: <https://ropenspain.github.io/mapSpain/>.
- [74] T. Wei y V. Simko, *R package 'corrplot': Visualization of a Correlation Matrix*, (Version 0.95), 2024. dirección: <https://github.com/taiyun/corrplot>.
- [75] G. Yu, *shadowtext: Shadow Text Grob and Layer*, R package version 0.1.4, 2024. dirección: <https://CRAN.R-project.org/package=shadowtext>.
- [76] J. Fox y S. Weisberg, *An R Companion to Applied Regression*, Third. Thousand Oaks CA: Sage, 2019. dirección: <https://www.john-fox.ca/Companion/>.
- [77] A. Zeileis y T. Hothorn, “Diagnostic Checking in Regression Relationships,” *R News*, vol. 2, n.º 3, págs. 7-10, 2002. dirección: <https://CRAN.R-project.org/doc/Rnews/>.
- [78] A. Zeileis, S. Köll y N. Graham, “Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in R,” *Journal of Statistical Software*, vol. 95, n.º 1, págs. 1-36, 2020. DOI: 10.18637/jss.v095.i01.
- [79] A. Zeileis, “Econometric Computing with HC and HAC Covariance Matrix Estimators,” *Journal of Statistical Software*, vol. 11, n.º 10, págs. 1-17, 2004. DOI: 10.18637/jss.v011.i10.
- [80] E. Pebesma y R. Bivand, *Spatial Data Science: With applications in R*. Chapman and Hall/CRC, 2023. DOI: 10.1201/9780429459016. dirección: <https://r-spatial.org/book/>.
- [81] H. Wickham y J. Bryan, *readxl: Read Excel Files*, R package version 1.4.3, 2023. dirección: <https://CRAN.R-project.org/package=readxl>.

- [82] R. Kolde, *pheatmap: Pretty Heatmaps*, R package version 1.0.13, 2025. dirección: <https://CRAN.R-project.org/package=pheatmap>.
- [83] A. Kassambara, *ggpubr: 'ggplot2' Based Publication Ready Plots*, R package version 0.6.0, 2023. dirección: <https://CRAN.R-project.org/package=ggpubr>.
- [84] J. Ooms, *magick: Advanced Graphics and Image-Processing in R*, R package version 2.8.3, 2024. dirección: <https://CRAN.R-project.org/package=magick>.
- [85] J. Ooms, *pdftools: Text Extraction, Rendering and Converting of PDF Documents*, R package version 3.4.0, 2023. dirección: <https://CRAN.R-project.org/package=pdftools>.
- [86] K. Slowikowski, *ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'*, R package version 0.9.5, 2024. dirección: <https://CRAN.R-project.org/package=ggrepel>.

7. Anexos

7.1. Anexo: Distribución de ZVs por provincia y comarca

Tabla 24: Distribución de ZVs por provincia y comarca

ZV	Provincia	Comarca
AV_01_AM	Ávila	Arévalo-madrigal
AV_02_AM	Ávila	Arévalo-madrigal
BU_01_AR	Burgos	Arlanza
BU_03_BU	Burgos	Bureba-Ebro
BU_04_PI	Burgos	Pisuerga
LE_01_EC	León	Esla-campos
LE_02_EC	León	Esla-campos
LE_03_SA	León	Sahagún
PA_01_BO	Palencia	Boedo-Ojeda
PA_02_CA	Palencia	Campos
PA_03_CA	Palencia	Campos
PA_04_CA	Palencia	Campos
PA_05_CA	Palencia	Campos
PA_06_CA	Palencia	Campos
PA_08_CA	Palencia	Campos
PA_09_EC	Palencia	El Cerrato
PA_10_EC	Palencia	El Cerrato
PA_11_SV	Palencia	Saldaña-Valdavia
SA_02_PB	Salamanca	Peñaranda de bracamonte
SA_03_SA	Salamanca	Salamanca
SA_04_PB	Salamanca	Peñaranda de bracamonte
SE_01_CU	Segovia	Cuellar
SE_02_CU	Segovia	Cuellar
SE_03_CU	Segovia	Cuellar
SE_04_SP	Segovia	Sepúlveda
SO_01_BO	Soria	Burgo de Osma
SO_03_CG	Soria	Campo de Gómara
VA_01_CE	Valladolid	Centro
VA_02_CE	Valladolid	Centro
VA_03_CE	Valladolid	Centro
VA_04_SU	Valladolid	Sur
VA_05_SE	Valladolid	Sureste

Continúa en la siguiente página

Tabla 24 – continuación

ZV	Provincia	Comarca
VA_06_SE	Valladolid	Sureste
VA_07_TC	Valladolid	Tierra de campos
VA_08_TC	Valladolid	Tierra de campos
ZA_01_BV	Zamora	Benavente y los valles
ZA_02_CP	Zamora	Campos-pan
ZA_03_CP	Zamora	Campos-pan
ZA_04_CP	Zamora	Campos-pan
ZA_05_DB	Zamora	Duero bajo

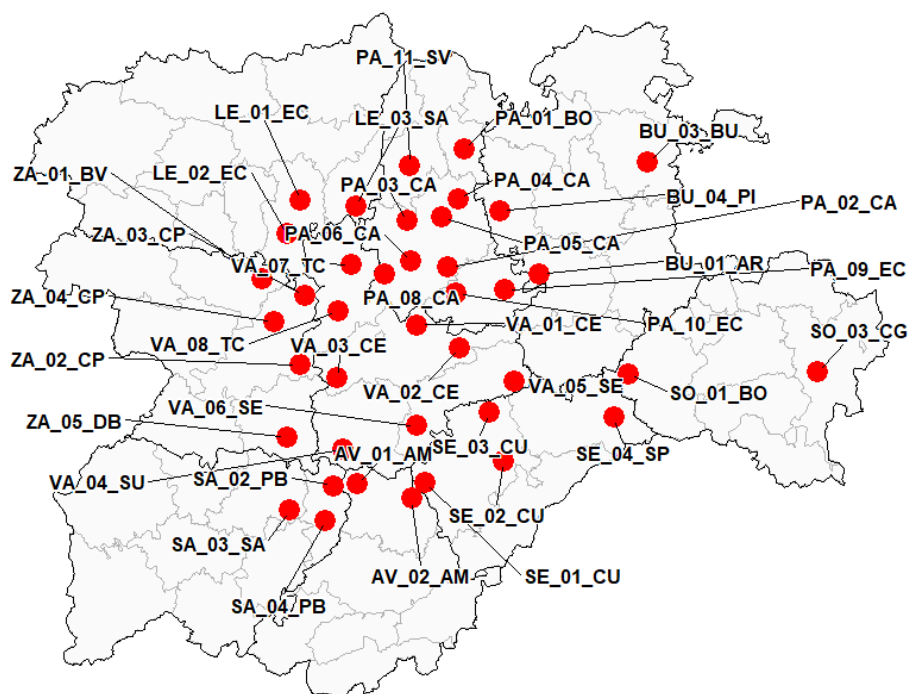


Figura 6: Correlaciones entre variables climáticas con la abundancia en vías de dispersión y en cultivos y con la periodicidad encontrada para cada ZV. Fuente: elaboración propia.

7.2. Anexo: Descripción de variables

Tabla 25: Variables con sus tipos y descripciones

Nombre	Tipo	Descripción
AG_LONG	Paisajística	Longitud de las vías de dispersión asociadas a cauces de agua
CA_LONG	Paisajística	Longitud de las vías de dispersión asociadas a caminos e infraestructuras de comunicación
CU_HERP	Paisajística	Porcentaje de la superficie cubierta por cultivos herbáceos plurianuales
PA_LENO	Paisajística	Porcentaje de la superficie cubierta por pastizal leñoso, arbolado y/o arbustivo
PA_HERB	Paisajística	Porcentaje de la superficie cubierta por pastizal herbáceo
REGADIO	Agronómica	Porcentaje de la superficie que se explota en condiciones de regadío
SIEMDIRE	Agronómica	Porcentaje de la superficie que se explota en condiciones de siembra directa
ARCILLA	Edáfica	Porcentaje de arcilla en el suelo
LIMO	Edáfica	Porcentaje de limo en el suelo
ARENA	Edáfica	Porcentaje de arena en el suelo
TMED_MED	Climática	Temperatura media de medias
DINUM_30	Climática	Número días por encima de 30 grados
DIHELA_0	Climática	Número días de helada por debajo de 0 grados
PREC_ACU	Climática	Precipitación acumulada

7.3. Anexo: Librerías de R utilizadas

Tabla 27: Librerías de R utilizadas

Paquete	Referencia
mapSpain	[73]
corrplot	[74]
shadowtext	[75]
car	[76]
lmtest	[77]
sandwich	[78] [79]
sf	[80]
readxl	[81]
pheatmap	[82]
ggpubr	[83]
magick	[84]
pdftools	[85]
ggrepel	[86]

7.4. Anexo: Mapas de calor de la abundancia y ciclicidad

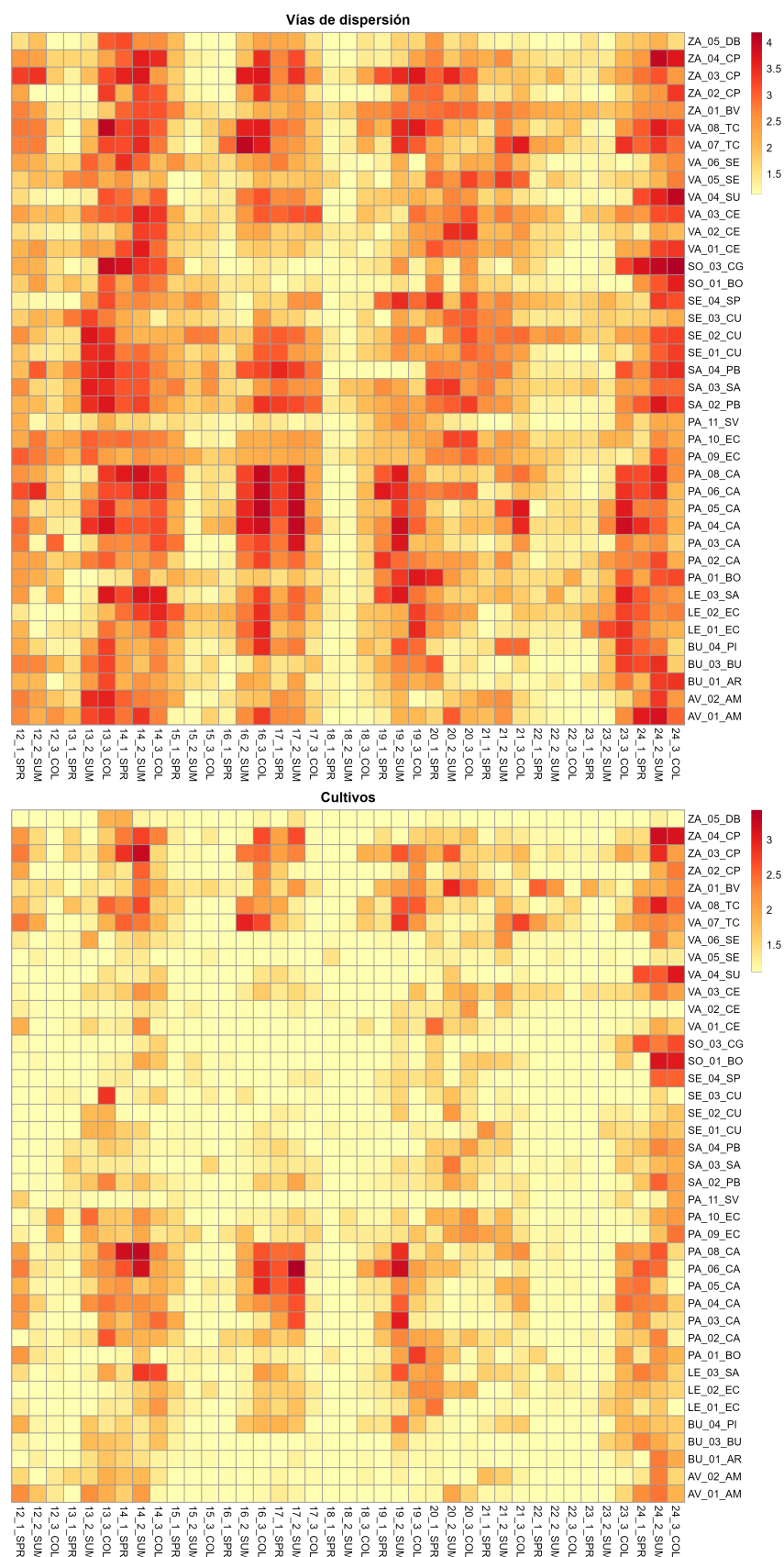


Figura 7: Mapa de calor para la abundancia (LNVAI) por ZV, periodo y ambos hábitats vías de dispersión (arriba) y cultivos (abajo). Fuente: elaboración propia.

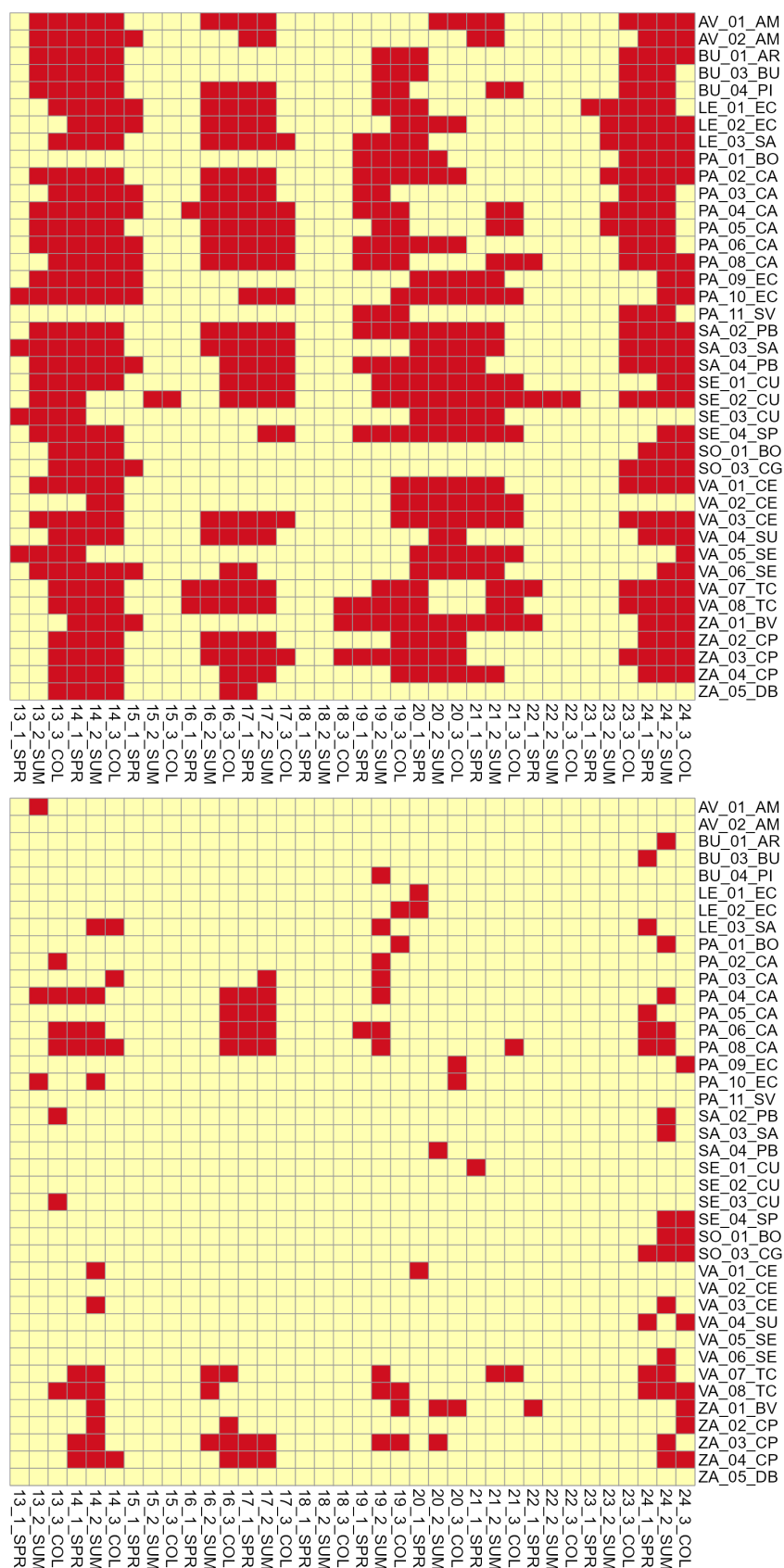


Figura 8: Mapa de calor para la existencia o no de ciclos por ZV, periodo y ambos hábitats vías de dispersión (arriba) y cultivos (abajo). Fuente: elaboración propia.

7.5. Anexo: Información sobre la ciclicidad en cada ZV

Tabla 28: Número de eventos sufridos y datos relativos a la duración y asincronía de dichos eventos para cada ZV en cultivos

ZV	Evento	Inicio	Vías de dispersión				Cultivos		
			Inicio en la ZV	Asincronía	Duración	Final en la ZV	Inicio en la ZV	Asincronía (respecto a vías)	Duración
AV_01_AM	2013	4	5	1	5	9	5	0	1
AV_02_AM	2013	4	5	1	6	10	-	-	-
BU_01_AR	2013	4	5	1	5	9	-	-	-
BU_03_BU	2013	4	5	1	5	9	-	-	-
BU_04_PI	2013	4	5	1	5	9	-	-	-
LE_01_EC	2013	4	6	2	5	10	-	-	-
LE_02_EC	2013	4	7	3	4	10	-	-	-
LE_03_SA	2013	4	6	2	4	9	8	2	2
PA_02_CA	2013	4	5	1	5	9	6	1	1
PA_03_CA	2013	4	6	2	5	10	9	3	1
PA_04_CA	2013	4	5	1	6	10	5	0	4
PA_05_CA	2013	4	5	1	5	9	-	-	-
PA_06_CA	2013	4	5	1	6	10	6	1	3
PA_08_CA	2013	4	6	2	5	10	6	0	4
PA_09_EC	2013	4	5	1	6	10	-	-	-
PA_10_EC	2013	4	4	0	7	10	5	1	1
SA_02_PB	2013	4	5	1	5	9	6	1	1
SA_03_SA	2013	4	5	1	6	10	-	-	-
SA_04_PB	2013	4	4	0	6	9	-	-	-
SE_01_CU	2013	4	5	1	5	9	-	-	-
SE_02_CU	2013	4	5	1	3	7	-	-	-
SE_03_CU	2013	4	4	0	4	7	6	2	1
SE_04_SP	2013	4	5	1	5	9	-	-	-
SO_01_BO	2013	4	6	2	4	9	-	-	-
SO_03_CG	2013	4	6	2	5	10	-	-	-
VA_01_CE	2013	4	5	1	5	9	8	3	1
VA_02_CE	2013	4	8	4	2	9	-	-	-
VA_03_CE	2013	4	5	1	5	9	8	3	1
VA_04_SU	2013	4	6	2	4	9	-	-	-
VA_05_SE	2013	4	4	0	4	7	-	-	-
VA_06_SE	2013	4	5	1	6	10	-	-	-
VA_07_TC	2013	4	6	2	4	9	7	1	2
VA_08_TC	2013	4	6	2	4	9	6	0	3
ZA_01_BV	2013	4	7	3	4	10	8	1	1
ZA_02_CP	2013	4	6	2	4	9	8	2	1
ZA_03_CP	2013	4	6	2	4	9	7	1	2
ZA_04_CP	2013	4	6	2	4	9	7	1	3
ZA_05_DB	2013	4	6	2	4	9	-	-	-

Continúa en la siguiente página

Tabla 28 – continuación

ZV	Evento	Inicio	Vías de dispersión				Cultivos		
			Inicio en la ZV	Asincronía	Duración	Final en la ZV	Inicio en la ZV	Asincronía (respecto a vías)	Duración
SE.02_CU	2015	11	11	0	2	12	-	-	-
AV.01_AM	2016	13	14	1	4	17	-	-	-
AV.02_AM	2016	13	16	3	2	17	-	-	-
BU.04_PI	2016	13	14	1	4	17	-	-	-
LE.01_EC	2016	13	14	1	4	17	-	-	-
LE.02_EC	2016	13	14	1	4	17	-	-	-
LE.03_SA	2016	13	14	1	5	18	-	-	-
PA.02_CA	2016	13	14	1	4	17	-	-	-
PA.03_CA	2016	13	14	1	4	17	17	3	1
PA.04_CA	2016	13	13	0	6	18	15	2	3
PA.05_CA	2016	13	14	1	5	18	15	1	3
PA.06_CA	2016	13	14	1	5	18	15	1	3
PA.08_CA	2016	13	14	1	5	18	15	1	3
PA.10_EC	2016	13	16	3	3	18	-	-	-
SA.02_PB	2016	13	14	1	5	18	-	-	-
SA.03_SA	2016	13	15	2	4	18	-	-	-
SA.04_PB	2016	13	14	1	5	18	-	-	-
SE.01_CU	2016	13	15	2	4	18	-	-	-
SE.02_CU	2016	13	15	2	4	18	-	-	-
SE.04_SP	2016	13	17	4	2	18	-	-	-
VA.03_CE	2016	13	14	1	5	18	-	-	-
VA.04_SU	2016	13	14	1	4	17	-	-	-
VA.06_SE	2016	13	15	2	2	16	-	-	-
VA.07_TC	2016	13	13	0	5	17	14	1	2
VA.08_TC	2016	13	13	0	5	17	14	1	1
ZA.02_CP	2016	13	14	1	4	17	15	1	1
ZA.03_CP	2016	13	14	1	5	18	14	0	4
ZA.04_CP	2016	13	15	2	3	17	15	0	3
ZA.05_DB	2016	13	15	2	2	16	-	-	-
AV.01_AM	2019	21	26	5	4	29	-	-	-
BU.01_AR	2019	21	23	2	3	25	-	-	-
BU.03_BU	2019	21	23	2	3	25	-	-	-
BU.04_PI	2019	21	23	2	2	24	23	0	1
LE.01_EC	2019	21	23	2	3	25	25	2	1
LE.02_EC	2019	21	24	3	4	27	24	0	2
LE.03_SA	2019	21	22	1	4	25	23	1	1
PA.01_BO	2019	21	22	1	5	26	24	2	1
PA.02_CA	2019	21	22	1	6	27	23	1	1
PA.03_CA	2019	21	22	1	2	23	23	1	1
PA.04_CA	2019	21	22	1	3	24	23	1	1
PA.05_CA	2019	21	23	2	2	24	-	-	-

Continúa en la siguiente página

Tabla 28 – continuación

ZV	Evento	Inicio	Vías de dispersión				Cultivos		
			Inicio en la ZV	Asincronía	Duración	Final en la ZV	Inicio en la ZV	Asincronía (respecto a vías)	Duración
PA_06_CA	2019	21	22	1	6	27	22	0	2
PA_08_CA	2019	21	22	1	3	24	23	1	1
PA_09_EC	2019	21	25	4	5	29	27	2	1
PA_10_EC	2019	21	24	3	7	30	27	3	1
PA_11_SV	2019	21	22	1	3	24	-	-	-
SA_02_PB	2019	21	22	1	8	29	-	-	-
SA_03_SA	2019	21	22	1	7	28	26	4	1
SA_04_PB	2019	21	25	4	5	29	-	-	-
SE_01_CU	2019	21	23	2	8	30	-	-	-
SE_02_CU	2019	21	23	2	11	33	-	-	-
SE_03_CU	2019	21	25	4	5	29	-	-	-
SE_04_SP	2019	21	22	1	9	30	-	-	-
VA_01_CE	2019	21	24	3	6	29	25	1	1
VA_02_CE	2019	21	24	3	7	30	-	-	-
VA_03_CE	2019	21	24	3	7	30	-	-	-
VA_04_SU	2019	21	26	5	2	27	-	-	-
VA_05_SE	2019	21	25	4	6	30	-	-	-
VA_06_SE	2019	21	25	4	5	29	-	-	-
VA_07_TC	2019	21	23	2	3	25	23	0	1
VA_08_TC	2019	21	21	0	5	25	23	2	2
ZA_01_BV	2019	21	21	0	11	31	24	3	1
ZA_02_CP	2019	21	24	3	4	27	-	-	-
ZA_03_CP	2019	21	21	0	7	27	23	2	2
ZA_04_CP	2019	21	24	3	6	29	-	-	-
AV_02_AM	2021	28	28	0	2	29	-	-	-
BU_04_PI	2021	28	29	1	2	30	-	-	-
PA_04_CA	2021	28	29	1	2	30	-	-	-
PA_05_CA	2021	28	29	1	2	30	-	-	-
PA_08_CA	2021	28	29	1	3	31	30	1	1
SE_01_CU	2021	28	-	-	-	-	28	-	1
VA_07_TC	2021	28	29	1	3	31	29	0	2
VA_08_TC	2021	28	29	1	2	30	-	-	-
ZA_01_BV	2021	28	-	-	-	-	31	-	1
AV_01_AM	2023	34	36	2	4	39	-	-	-
AV_02_AM	2023	34	37	3	3	39	-	-	-
BU_01_AR	2023	34	36	2	4	39	38	2	1
BU_03_BU	2023	34	36	2	3	38	37	1	1
BU_04_PI	2023	34	36	2	3	38	-	-	-
LE_01_EC	2023	34	34	0	5	38	-	-	-
LE_02_EC	2023	34	35	1	5	39	-	-	-
LE_03_SA	2023	34	35	1	5	39	37	2	1

Continúa en la siguiente página

Tabla 28 – continuación

ZV	Evento	Inicio	Vías de dispersión				Cultivos		
			Inicio en la ZV	Asincronía	Duración	Final en la ZV	Inicio en la ZV	Asincronía (respecto a vías)	Duración
PA_01_BO	2023	34	36	2	4	39	38	2	1
PA_02_CA	2023	34	35	1	5	39	-	-	-
PA_03_CA	2023	34	36	2	3	38	-	-	-
PA_04_CA	2023	34	35	1	4	38	38	3	1
PA_05_CA	2023	34	35	1	4	38	37	2	1
PA_06_CA	2023	34	36	2	3	38	37	1	2
PA_08_CA	2023	34	36	2	4	39	37	1	2
PA_09_EC	2023	34	38	4	2	39	39	1	1
PA_10_EC	2023	34	38	4	2	39	-	-	-
PA_11_SV	2023	34	36	2	3	38	-	-	-
SA_02_PB	2023	34	36	2	4	39	38	2	1
SA_03_SA	2023	34	36	2	4	39	-	-	-
SA_04_PB	2023	34	36	2	4	39	38	2	1
SE_01_CU	2023	34	38	4	2	39	-	-	-
SE_02_CU	2023	34	36	2	4	-	-	-	-
SE_04_SP	2023	34	38	4	2	39	38	0	2
SO_01_BO	2023	34	37	3	3	-	38	1	2
SO_03_CG	2023	34	36	2	4	39	37	1	3
VA_01_CE	2023	34	36	2	4	39	-	-	-
VA_03_CE	2023	34	36	2	4	-	38	2	1
VA_04_SU	2023	34	37	3	3	39	37	0	1
VA_05_SE	2023	34	39	5	-	-	-	-	-
VA_06_SE	2023	34	38	4	2	-	38	0	1
VA_07_TC	2023	34	36	2	4	39	37	1	2
VA_08_TC	2023	34	36	2	4	39	37	1	3
ZA_01_BV	2023	34	37	3	3	-	39	2	1
ZA_02_CP	2023	34	37	3	3	39	39	2	1
ZA_03_CP	2023	34	36	2	4	-	38	2	1
ZA_04_CP	2023	34	37	3	3	39	38	1	2

7.6. Anexo: Gráficos de series temporales. periodogramas y autocorrelogramas para cada ZV, para cada tipo de hábitat

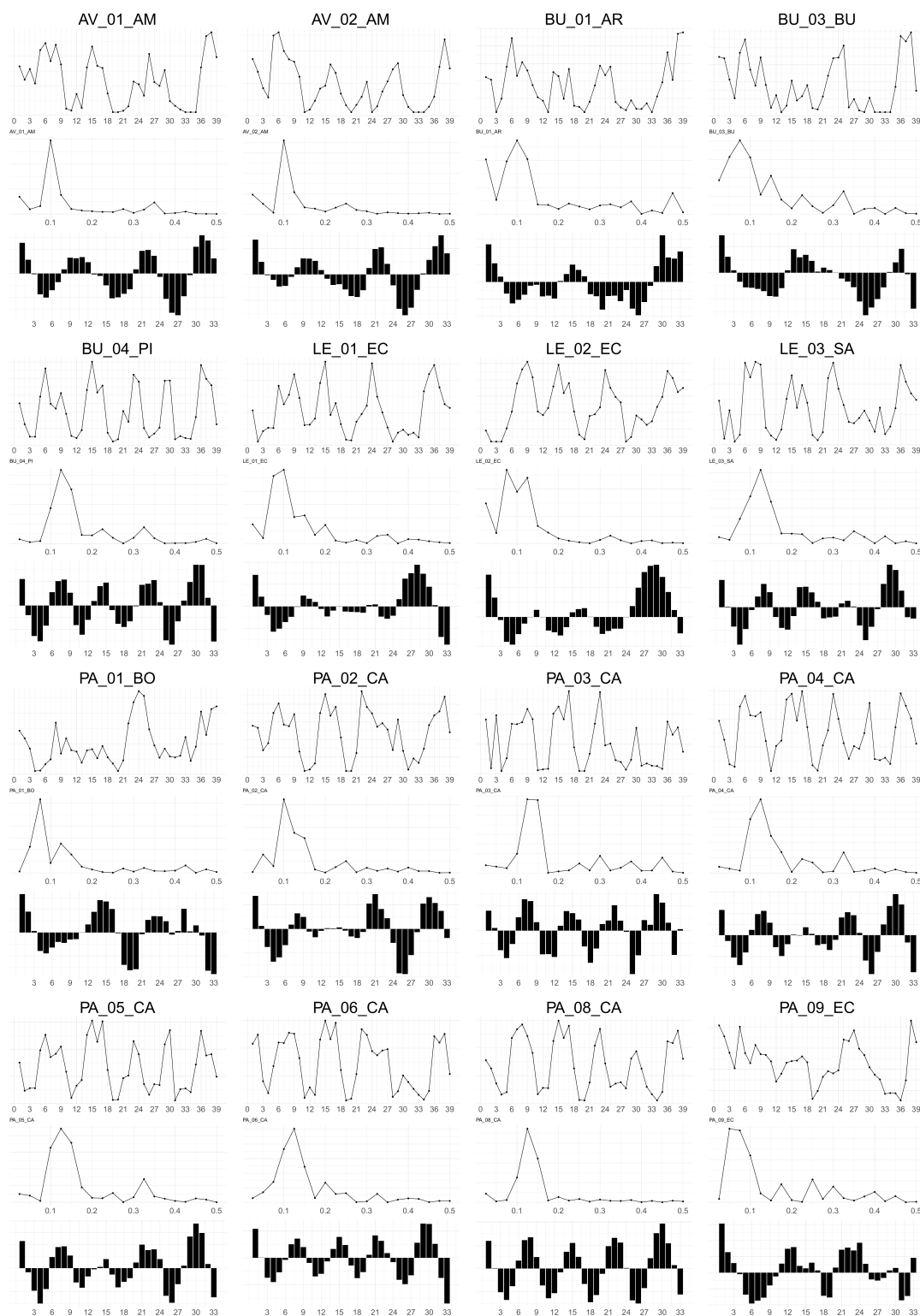


Figura 9: Series temporales de LNVAI en vías de dispersión para las ZVs desde AV_01_AM hasta PA_09_EC. Fuente: elaboración propia.

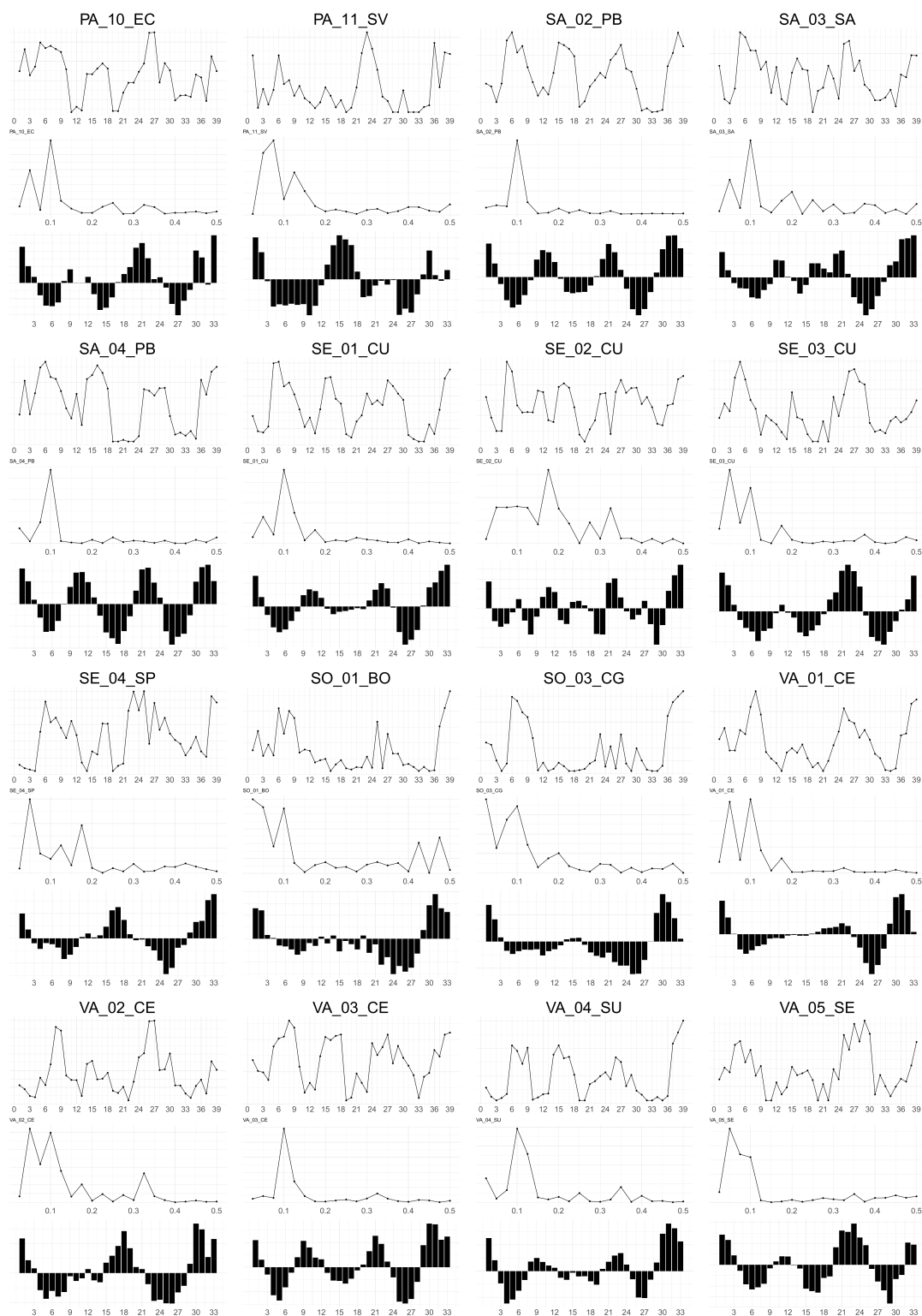


Figura 10: Series temporales de LNVAI en vías de dispersión para las ZVs desde PA_10_EC hasta VA_05_SE. Fuente: elaboración propia.

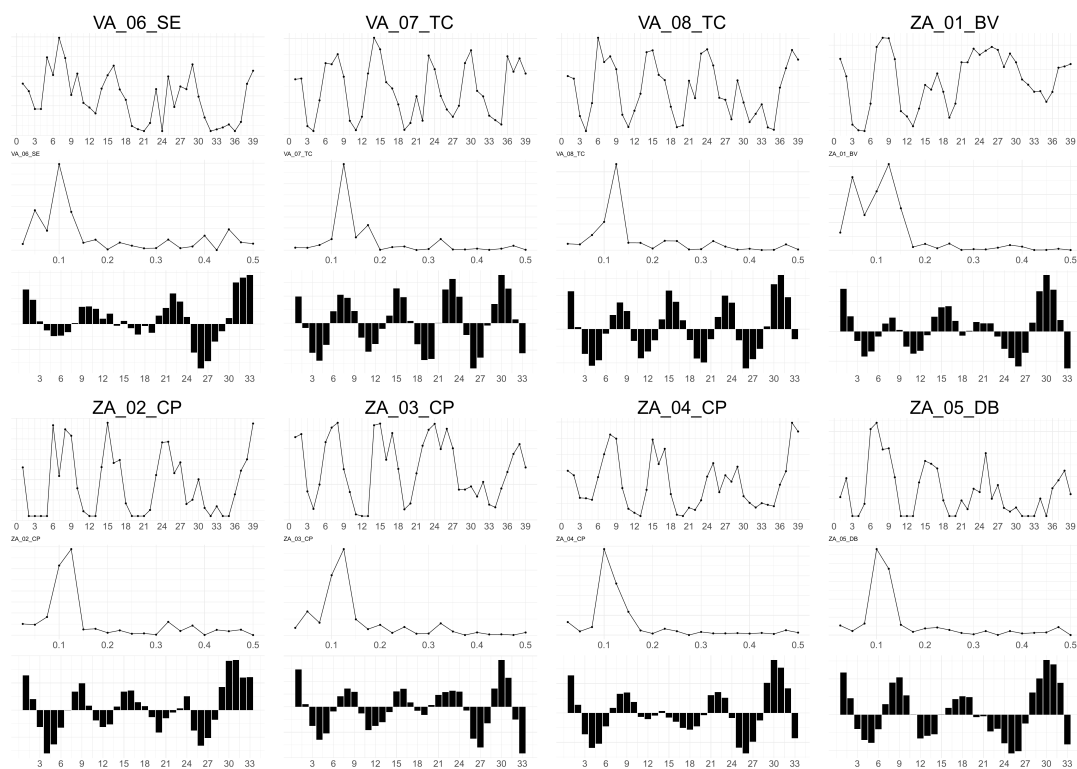


Figura 11: Series temporales de LNVAI en vías de dispersión para las ZVs desde VA_06_SE hasta ZA_05_DB. Fuente: elaboración propia.

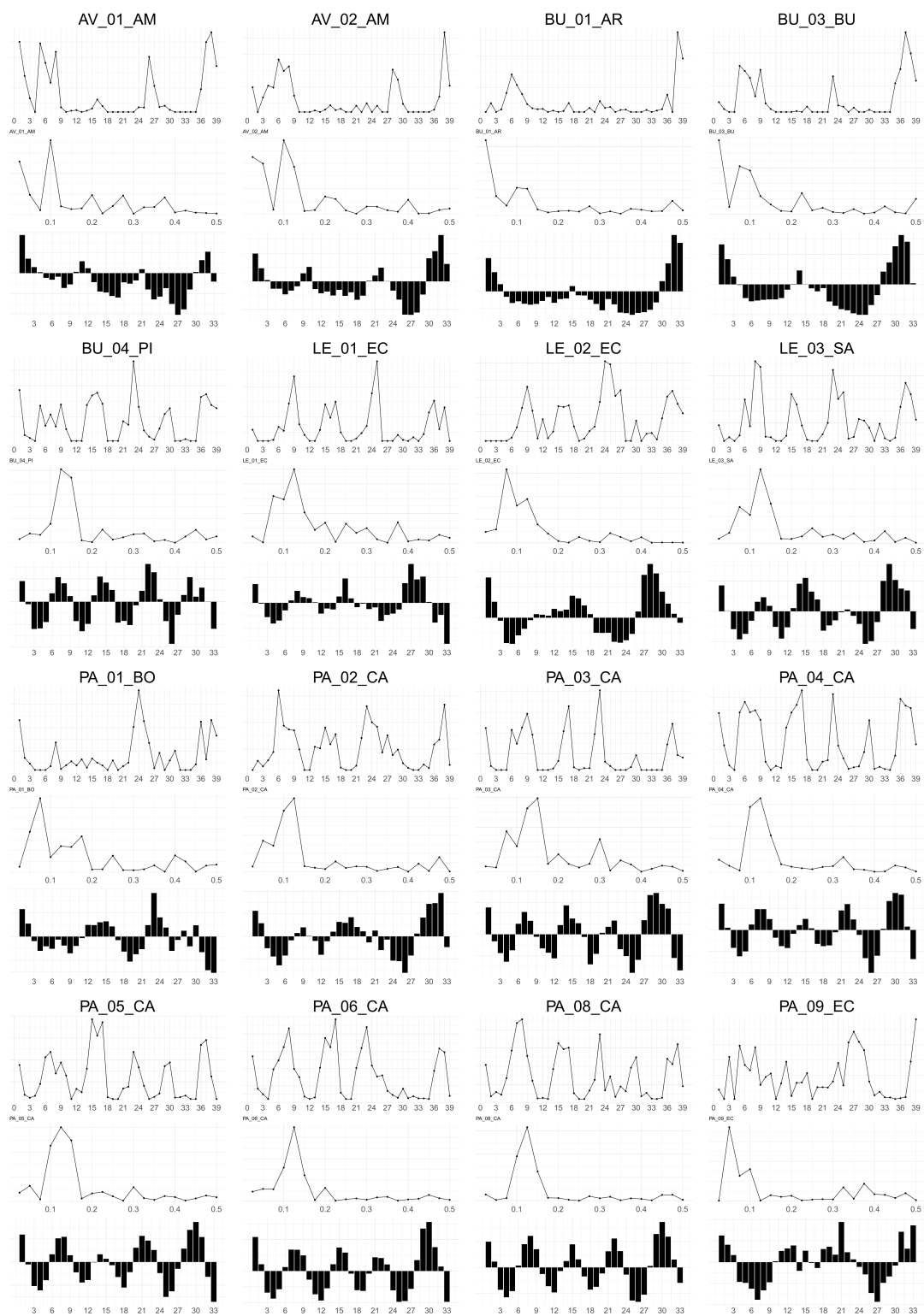


Figura 12: Series temporales de LNVAI en cultivos para las ZVs desde AV_01_AM hasta PA_09_EC. Fuente: elaboración propia.

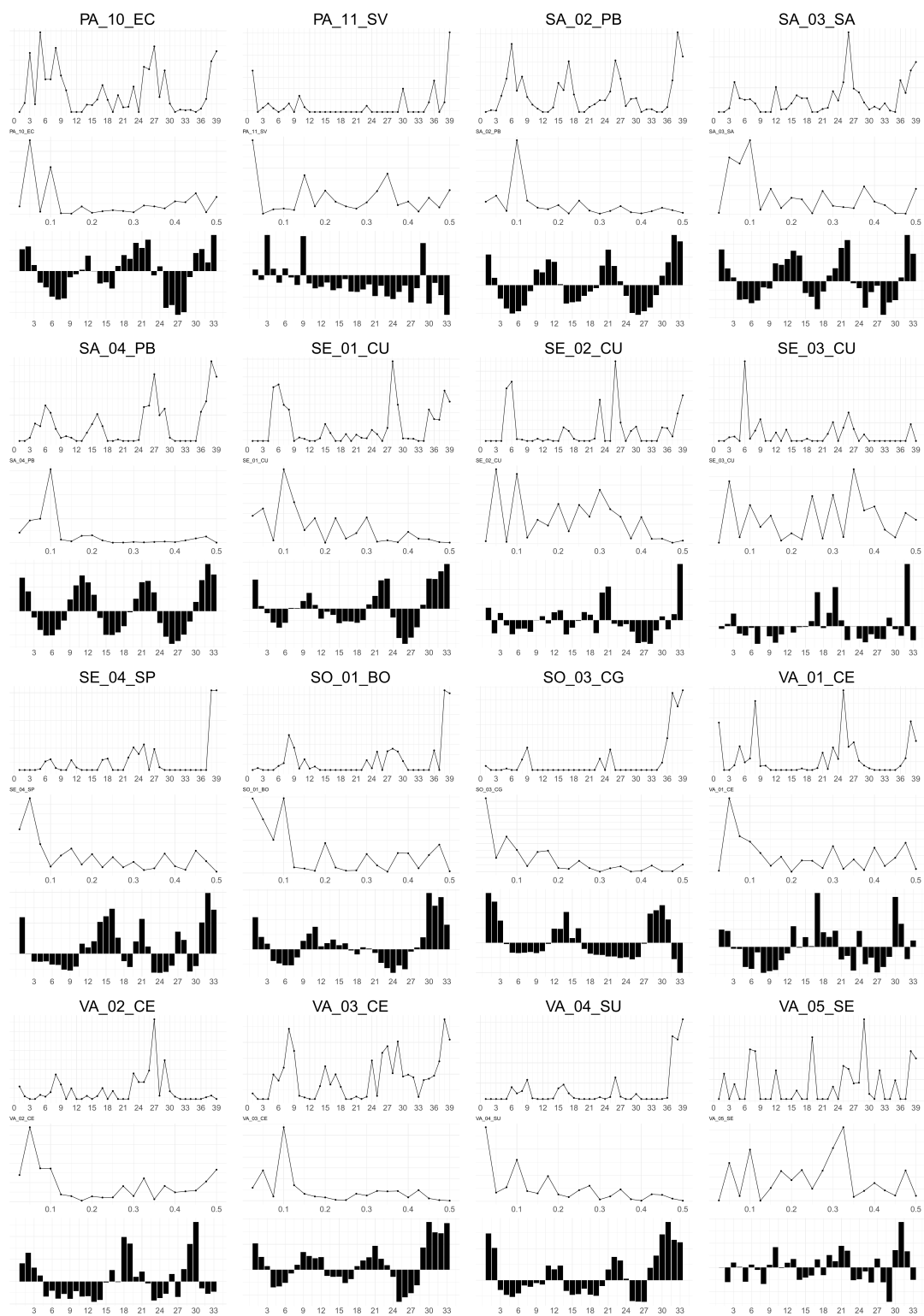


Figura 13: Series temporales de LNVAI en cultivos para las ZVs desde PA_10_EC hasta VA_05_SE. Fuente: elaboración propia.

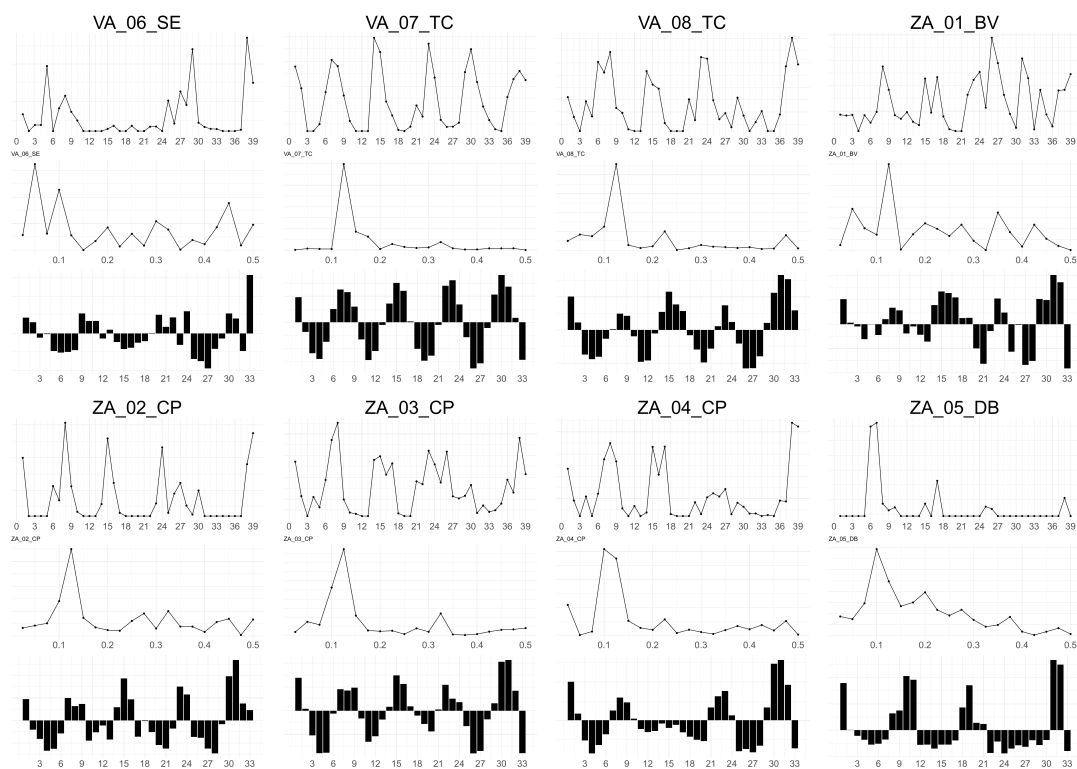


Figura 14: Series temporales de LNVAI en cultivos para las ZVs desde VA_06_SE hasta ZA_05_DB. Fuente: elaboración propia.

7.7. Anexo: Correlaciones de variables climáticas estacionales por periodo

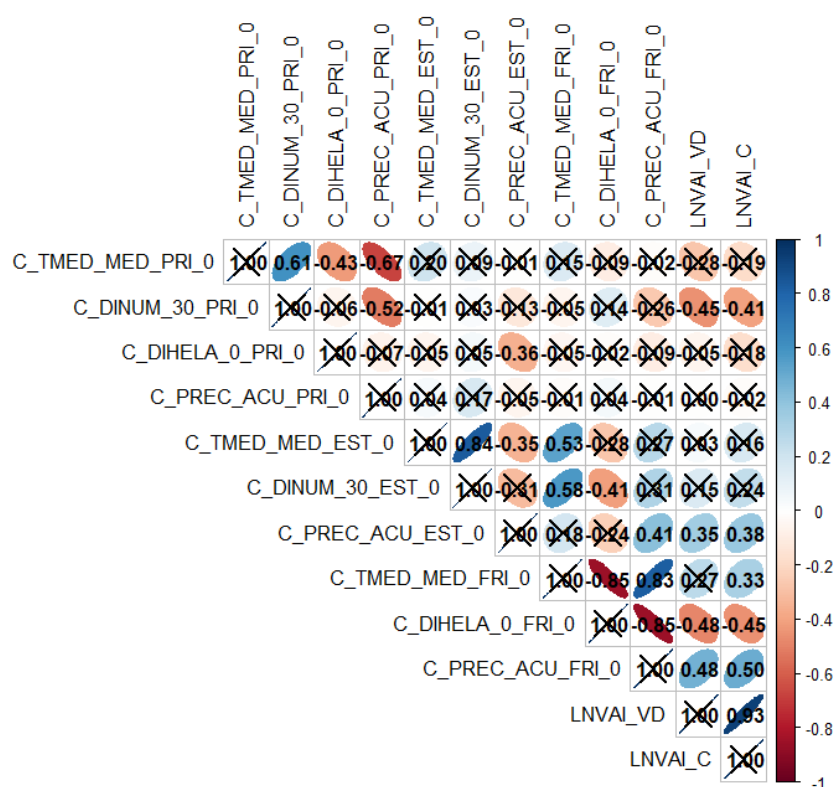


Figura 15: Correlaciones entre variables climáticas estacionales con la abundancia en vías de dispersión y en cultivos para cada periodo. Fuente: elaboración propia.

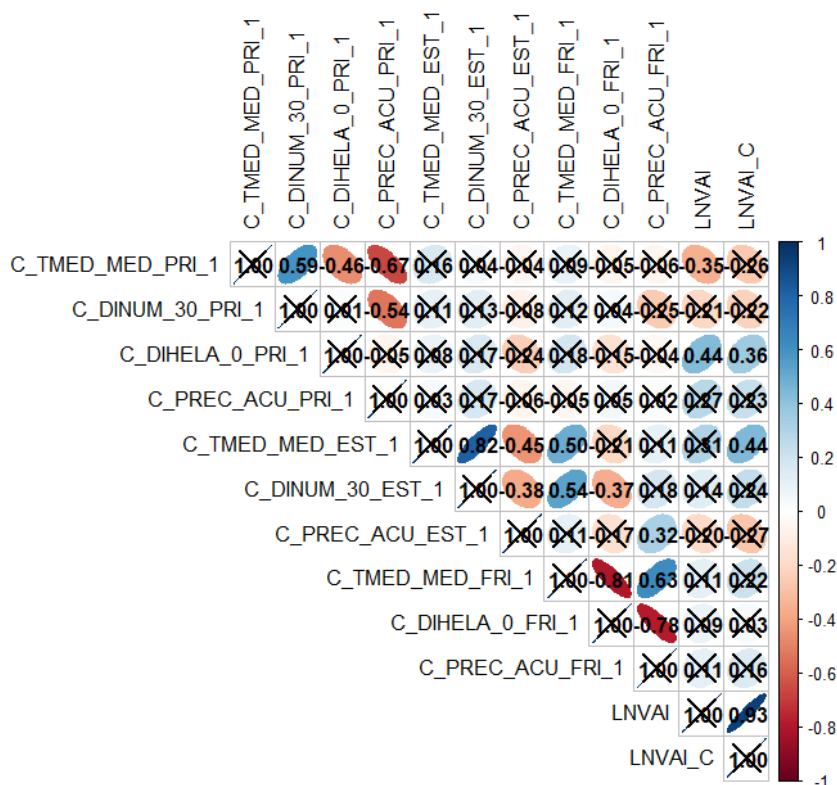


Figura 16: Correlaciones entre variables climáticas estacionales con la abundancia en vías de dispersión y en cultivos para cada periodo. Fuente: elaboración propia.

7.8. Anexo: Signo de cada variable para cada uno de los modelos

Tabla 29: Variables seleccionadas y su signo en cada modelo de vías de dispersión.

Variable	Forward				Backward				Nº de Modelos
	G	P	E	F	G	P	E	F	
D_LNVAI_EST_1				-					1
D_LNVAI_EST_2				+	+			+	3
D_LNVAI_EST_3	-								1
D_LNVAI_EST_4								+	1
D_LNVAI_FRI_2	+		+		+		+	+	5
D_LNVAI_FRI_3	+		+		+				3
D_LNVAI_PRI_1					-				1
D_LNVAI_PRI_2					-			-	2
D_LNVAI_PRI_3					-				1
D_LNVAI_PRI_4	+			+					2
D_LNVAI_T3	-	-	-	-	-	-	-	-	8
A_REGADIO		+			+	+	+		4
A_SIEMDIRE					-				1
C_DIHELA_3		+	+	-	+	+			5
C_DIHELA_FRI_1	+			+					2
C_DIHELA_PRI_1	+			+			+		3
C_DINUM_3_3					+	-			2
C_DINUM_3_EST_1	-	-							2
C_DINUM_3_PRI_1	-		-		-		-	-	5
C_PREC_ACU_3					+				1
C_PREC_ACU_FRI_1					-				1
C_PREC_ACU_PRI_1					+		+		2
C_TMED_MED_3		-	+					+	3
C_TMED_MED_EST_1	+	+							2
C_TMED_MED_FRI_1	+				+				2
U_AG_LONG	+					-			2
U_CU_HERP					+				1
U_PA_LENO				-		-	-		3
Nº de variables	13	7	7	9	18	7	8	8	

G.: Modelo global; P: Modelo de las primaveras; E: Modelo de los estíos; F: Modelo de los fríos.

Tabla 30: Variables seleccionadas y su signo en cada modelo de cultivos.

Variable	Forward				Backward				Nº de Modelos
	G	P	E	F	G	P	E	F	
D_LNVAI_EST_1				-					1
D_LNVAI_EST_2				+	+			+	3
D_LNVAI_EST_3	-								1
D_LNVAI_EST_4							+		1
D_LNVAI_FRI_2	+		+		+		+	+	5
D_LNVAI_FRI_3	+		+		+				3
D_LNVAI_PRI_1					-				1
D_LNVAI_PRI_2					-			-	2
D_LNVAI_PRI_3					-				1
D_LNVAI_PRI_4	+			+					2
D_LNVAI_T3	-	-	-	-	-	-	-	-	8
A_REGADIO		+			+	+	+		4
A_SIEMDIRE					-				1
C_DIHELA_3		+	+	-	+	+			5
C_DIHELA_FRI_1	+			+					2
C_DIHELA_PRI_1	+			+			+		3
C_DINUM_3_3					+	-			2
C_DINUM_3_EST_1	-	-							2
C_DINUM_3_PRI_1	-		-		-		-	-	5
C_PREC_ACU_3					+				1
C_PREC_ACU_FRI_1					-				1
C_PREC_ACU_PRI_1					+		+		2
C_TMED_MED_3		-	+					+	3
C_TMED_MED_EST_1	+	+							2
C_TMED_MED_FRI_1	+				+				2
U_AG_LONG	+					-			2
U_CU_HERP					+				1
U_PA_LENO				-		-	-		3
Nº de variables	13	7	7	9	18	7	8	8	

G.: Modelo global; P: Modelo de las primaveras; E: Modelo de los estíos; F: Modelo de los fríos.

7.9. Anexo: P-valores para los tests de Durbin-Watson para los residuos del modelo en cada zona

Tabla 31: P-valores para los tests de Durbin-Watson para los residuos del modelo por en cada zona para Vías de dispersión y Cultivos

Periodo	P-valor para Vías de dispersión	P-valor para Cultivos
AV_02_AM	0.02351	0.13322
BU_01_AR	0.00505	0.00343
BU_03_BU	0.00113	0.00075
BU_04_PI	0.03486	0.17387
LE_01_EC	0.00006	0.14006
LE_02_EC	0.00208	0.00149
LE_03_SA	0.03431	0.09207
PA_01_BO	0.00253	0.00226
PA_02_CA	0.06527	0.00134
PA_03_CA	0.15035	0.17467
PA_04_CA	0.02517	0.06866
PA_05_CA	0.03214	0.02768
PA_08_CA	0.01374	0.38525
PA_09_EC	0.00115	0.00020
PA_10_EC	0.03738	0.04411
SA_02_PB	0.00039	0.00737
SA_03_SA	0.40638	0.06838
SE_01_CU	0.00349	0.32117
SE_02_CU	0.01670	0.66725
SE_04_SP	0.14939	0.33174
SO_01_BO	0.13657	0.27270
VA_01_CE	0.00005	0.27914
VA_03_CE	0.24540	0.26341
VA_04_SU	0.01151	0.10637
VA_05_SE	0.01396	0.13873
VA_06_SE	0.05167	0.38766
VA_07_TC	0.13912	0.04333
VA_08_TC	0.03826	0.07411
ZA_01_BV	0.00024	0.09672
ZA_03_CP	0.00081	0.05446
ZA_04_CP	0.00964	0.08566
ZA_05_DB	0.09162	0.00338

7.10. Anexo: P-valores para los tests de la I de Moran para los residuos del modelo en cada periodo

Tabla 32: P-valores para los tests de la I de Moran para los residuos del modelo por periodo para Vías de dispersión y Cultivos

Periodo	P-valor para Vías de dispersión	P-valor para Cultivos
19	0.2756	0.602
20	0.6196	3.58E+04
21	0.6703	0.3767
22	0.5897	0.2203
23	0.1123	0.3142
24	0.4373	0.854
25	0.3447	0.9354
26	0.6878	0.6734
27	0.5984	0.2678
28	0.9102	0.8823
29	0.2611	0.1926
30	0.6939	0.5103
31	0.2726	0.4436
32	0.0658	0.5481
33	0.5017	0.9998
34	0.063	0.029
35	0.1943	0.7734
36	0.5972	0.2842
37	0.1328	0.2952
38	0.0744	1.00E-04
39	0.4994	0.5537