

TRABAJO FIN DE GRADO

GRADO EN ESTADÍSTICA

UNIVERSIDAD DE VALLADOLID



ANÁLISIS DE LA EVOLUCIÓN Y CLASIFICACIÓN DE LA ABUNDANCIA DEL TOPILLO CAMPESINO (*Microtus arvalis* Pallas) EN CASTILLA Y LEÓN

Rubén Garduño Barbier

Tutores:

Agustín Mayo Íscar, Departamento de Estadística e
Investigación Operativa, Universidad de Valladolid

Constantino Caminero Saldaña, Área de plagas, Instituto
Tecnológico Agrario de Castilla y León

27 de junio de 2022

Agradecimientos

En primer lugar, me gustaría dar las gracias al Instituto Tecnológico Agrario de Castilla y León, por acogerme como un miembro más dentro del equipo del Área de Plagas y brindarme una magnífica oportunidad para aprender, colaborar y crecer como estudiante y como persona.

Agradezco de corazón el encomiable trato y atención que he recibido de manos de mi tutor, Constantino Caminero Saldaña, y de mi compañera, M^a Carmen García Ariza, así como su esfuerzo incondicional por enseñarme todo lo posible durante mi estancia y ayudarme en todo lo que necesitase, y su involucración en este proyecto.

También me gustaría dar las gracias a todos los profesores y profesoras que me han ayudado y guiado a lo largo de mi vida y trayectoria académica, puesto que gracias a su labor, hoy puedo estar donde estoy, dedicándoles estas palabras de agradecimiento.

Particularmente, doy las gracias a Agustín Mayo Íscar, por ofrecerme la oportunidad y supervisar mi trabajo a lo largo de este proyecto.

Gracias a mi pareja, Esther, por confiar en mí más que yo mismo, animarme en todo momento, y estar dispuesta a darme su apoyo de forma incondicional.

Y, sin ninguna duda, mi agradecimiento hacia mi familia y amigos es eterno y sin medida, por estar ahí siempre, y demostrarme el valor del esfuerzo y la constancia en cualquier ámbito de la vida.

¡ Muchas gracias !

Código y conjunto de datos

Todos los cálculos y visualizaciones que se han ido realizando lo largo de este proyecto se han desarrollado con el *software* estadístico *R*. Así mismo, los datos estudiados son propiedad del Instituto Tecnológico Agrario de Castilla y León.

Para cualquier cuestión o petición sobre el código utilizado, puede dirigirse al siguiente correo electrónico: ruben.garduno.barbier@gmail.com . Para más información sobre los datos, contacte con Constantino Caminero Saldaña (Tutor del presente *TFG*), al siguiente correo electrónico: camsalco@itacyl.es .

Prólogo

Si fuese posible elegir una palabra que resuma y explique el concepto de vida, una de las posibles candidatas sería evidente: equilibrio. Equilibrio entre especies, equilibrio demográfico, económico, social, personal... equilibrio de toda clase. Pues esta homeostasis que nos mantiene hoy por hoy como seres vivos, depende directamente de que la convivencia entre cada uno de nosotros sea equilibrada y estable. Este término está directamente vinculado con un fenómeno que, si bien tiene muchas y muy variadas interpretaciones, es de vital importancia en nuestro estudio, así como en nuestra propia vida. Y sí, la ley a la que estoy haciendo referencia, no podría ser otra que la Ley de Causa y Efecto.

A cada acción que sucede (denominada causa), le sigue una reacción (conocida como efecto). Simple en apariencia, pero... no nos engañemos, más que complejo en la realidad. Pues tras estas palabras se esconde un mundo de cuestiones dignas de atención, como podrían ser, por ejemplo, ¿qué efecto produce una determinada acción en un entorno concreto? ¿Es siempre el mismo? O bien, para una determinada reacción, ¿que acciones han podido ser las causantes? ¿Qué relación existe entre ellas?

A lo largo de los años, hemos podido ver como la ruptura de este equilibrio (ya enfocando nuestra atención en el ámbito biológico) ha sido la causante de problemas muy serios que afectaban a muchos ecosistemas diferentes, así como a la interacción entre su fauna y su flora.

Por ello, buscar el patrón al que obedece el comportamiento de este tipo de desequilibrios, encontrar las razones por las cuales se produce, y qué circunstancias agravan o moderan su efecto, o bien determinar la ciclicidad con la que se repite el fenómeno en cuestión, es crucial para poder entenderlo, y así poder tomar las medidas preventivas necesarias y anticiparse al futuro.

Pues ya lo dice la conocida frase:

*Aquellos que no pueden recordar su pasado
están condenados a repetirlo.*

(George Santayana, La vida de la razón,
1905)

Índice general

RESUMEN	9
ABSTRACT	11
1. Justificación	13
2. Objetivos	19
2.1. Objetivo global	19
2.2. Objetivos específicos	19
2.2.1. Exploración y racionalización espacio-temporal inicial de la base de datos	19
2.2.2. Detección inicial de expresiones diferenciales de la abundancia a nivel espacio-temporal	19
2.2.3. Agrupación espacial en función de la expresión diferencial de la evolución de la abundancia mediante el uso de métodos de aprendizaje no supervisado	20
2.2.4. Definición final de zonas diferenciales de riesgo y estudio comparativo en la evolución y ciclicidad	20
3. Materiales y métodos	21
3.1. Descripción de la metodología de datos en campo	21
3.1.1. Descripción del área de estudio	21
3.1.2. Descripción de los conjuntos de datos	25
3.2. Metodología estadística	29
3.2.1. Método de estimación del índice de abundancia	29
3.2.2. ANOVA de un factor	30
3.2.3. Test LSD de Fisher	32
3.2.4. Análisis cluster jerárquico	33
3.2.5. Series temporales	35
4. Resultados y discusión	39
4.1. Exploración y racionalización espacio-temporal inicial de la base de datos.	39
4.1.1. Revisión de los periodos de estudio	40
4.1.2. Revisión de las zonas de vigilancia	40
4.2. Detección inicial de expresiones diferenciales de la abundancia a nivel espacio-temporal.	42

4.3.	Agrupación espacial en función de la expresión diferencial de la evolución de la abundancia mediante el uso de métodos de aprendizaje no supervisado.	44
4.3.1.	Agrupación inicial de las zonas de vigilancia sobre la matriz de racionalización espacio-temporal	44
4.3.2.	Estimación de datos perdidos a partir del clustering inicial	53
4.3.3.	Agrupación final a partir del método Deformación Dinámica del Tiempo	55
4.4.	Definición final de zonas diferenciales de riesgo y estudio comparativo en la evolución y ciclicidad.	60
4.4.1.	Evolución temporal de las zonas diferenciales: gráficos y mapas de calor	60
4.4.2.	Análisis de la varianza y test LSD	63
4.4.3.	Correlogramas	67
4.5.	Pensando en estudios futuros...	68
5.	Conclusiones	71
6.	Anexos	75
6.1.	Anexo: Municipios de referencia de las zonas de vigilancia	75
6.2.	Anexo: Comarcas agrarias por provincias	77
6.3.	Anexo: Dendrograma la clasificación final de las zonas de vigilancia	78

Índice de figuras

1.1. Topillo Campesino (<i>Microtus arvalis</i>)	13
1.2. Distribución del topillo campesino en Europa y Asia [9]	15
1.3. Distribución de topillo campesino en España [3]	16
1.4. Proceso de colonización del topillo campesino en la región de Castilla y León desde 1973 hasta 2002 [10]	16
3.5. División del espacio en el estudio	21
3.6. Mapa de las comarcas agrarias de Castilla y León	22
3.7. Colocación de los transectos	23
3.8. Madriguera de topillo campesino	24
3.9. CartoDruid	25
3.10. Distancia D entre el recinto y el perímetro de 10 km alrededor del centroide de la zona de vigilancia	29
3.11. Análisis cluster aglomerativo	34
3.12. Deformación Dinámica del Tiempo	37
4.13. Mapa de calor para hábitat <i>Cultivos</i>	39
4.14. Mapa de calor para hábitat <i>Vías de dispersión</i>	40
4.15. Mapa de calor para el hábitat <i>Cultivos</i> desde el periodo 8 (Verano de 2013)	42
4.16. Mapa de calor para el hábitat <i>Vías de dispersión</i> desde el periodo 8 (Verano de 2013)	42
4.17. Dendrograma para hábitat <i>Cultivos</i>	45
4.18. Dendrograma para hábitat <i>Vías de dispersión</i>	46
4.19. Gráfico de codo para hábitat <i>Cultivos</i>	47
4.20. Gráfico de codo para hábitat <i>Vías de dispersión</i>	48
4.21. Agrupación inicial de las zonas de vigilancia para hábitat <i>Cultivos</i>	50
4.22. Distribución inicial de las zonas de vigilancia para hábitat <i>Vías de dispersión</i>	52
4.23. Mapa de calor tras la estimación para el hábitat <i>Cultivos</i>	54
4.24. Mapa de calor tras la estimación para el hábitat <i>Vías de dispersión</i>	54
4.25. Mapa de calor sin valores ausentes para el hábitat <i>Cultivos</i>	55
4.26. Mapa de calor sin valores ausentes para el hábitat <i>Vías de dispersión</i>	55
4.27. Dendrograma obtenido con la clasificación final	56
4.28. Gráfico de codo para la clasificación final	57
4.29. Agrupación final de las zonas de vigilancia en Castilla y León	59
4.30. Evolución de los grupos finales de zonas de vigilancia para el hábitat <i>Vías de dispersión</i>	61
4.31. Evolución de los grupos finales de zonas de vigilancia para el hábitat <i>Cultivos</i>	61

4.32. Mapas de calor de los grupos finales de zonas de vigilancia en Vías de dispersión y Cultivos	62
4.33. Propuesta de mapa para su posible exploración en estudios futuros . . .	69
6.34. Dendrograma obtenido con la clasificación final	79

Índice de tablas

3.1. Número de zonas de vigilancia por provincia	23
3.2. Número de inspecciones por hábitat	26
3.3. Correspondencia de los periodos de estudio	28
4.4. Porcentajes de valores ausentes por hábitat en los 7 primeros periodos .	40
4.5. Porcentaje de valores ausentes para las zonas de vigilancia eliminadas en cultivos	41
4.6. Porcentaje de valores ausentes para las zonas de vigilancia eliminadas en vías de dispersión	41
4.7. Porcentajes de valores ausentes para cultivos y vías de dispersión . . .	41
4.8. Aceptación o rechazo de la hipótesis de igualdad de medias (verde: re- chazo de la hipótesis de igualdad de medias; rojo: no rechazo de esta hipótesis)	43
4.9. Composición de los grupos iniciales en <i>Cultivos</i>	49
4.10. Composición de los grupos iniciales en <i>Vías de dispersión</i>	51
4.11. Zonas de vigilancia por grupos	57
4.12. Tabla resumen de los periodos significativos en vías de dispersión . . .	64
4.13. Tabla resumen de los periodos significativos en Cultivos	65
4.14. Correlogramas para los grupos en vías de dispersión y cultivos	67
6.16. Comarcas agrarias por provincia	78

RESUMEN

Las variaciones demográficas de carácter explosivo del topillo campesino pueden causar importantes daños y pérdidas en los cultivos en los ecosistemas agrarios donde está presente, siendo Castilla y León, el área de estudio donde se realiza el presente trabajo, una de las zonas afectadas por ello.

Tras un proceso de colonización que comenzó a finales de la década de los 70 del siglo pasado, desde las praderas asociadas a las zonas montanas que bordean Castilla y León hacia el interior de la meseta, donde se encuentran las principales zonas de producción agrícola, el topillo campesino está ya presente en la totalidad de la comunidad autónoma desde comienzos del presente siglo.

Así, hasta hace unos años, en Castilla y León se registraban episodios de explosiones demográficas de forma recurrente con una ciclicidad aparente de cinco años. Sin embargo, este patrón parece haber cambiado últimamente a frecuencias más cortas, al menos en algunas partes de Castilla y León.

En añadido, en todas las zonas el topillo campesino experimenta esos fenómenos asociados a los hábitat donde vive. Sin embargo, no en todas partes parece “traducirse” de la misma manera en cuanto a su expresión e impacto en las parcelas de cultivo y, por tanto en las consecuencias que este tipo de comportamiento biológico puede tener en cuanto a la concepción de “plaga”, no en un sentido meramente basado en los tamaños poblacionales, sino en cuanto al riesgo de afectación final a los cultivos.

Ante lo anteriormente expuesto, cuando se plantea una estrategia de Gestión Integrada frente a las plagas ocasionadas por el topillo campesino, resulta crucial el estudio espacio-temporal de zonas y momentos con distintos umbrales de riesgo final para los cultivos. Y es precisamente una primera aproximación a la racionalización espacio-temporal de la expresión diferencial de la evolución de la abundancia natural y del riesgo para los cultivos del topillo campesino en Castilla y León lo que se aborda en el presente *TFG*.

Se ha partido de la base de datos de la última década procedente del programa oficial de monitorización de abundancia de topillo campesino en Castilla y León, considerando la información obtenida en dos tipos de hábitat: “Vías de dispersión” y “Cultivos”. El primero representa un hábitat de tipo “reservorio”, donde habitan en periodos de baja densidad poblacional y desde los que se expanden en periodos de alta densidad, y en el que la evolución de la abundancia corresponde al efecto biológico natural sin la influencia del factor humano. El segundo representa un hábitat tipo a proteger en la

Gestión Integrada, incluyendo en este caso los datos en parcelas de cultivos herbáceos anuales, y en el que la evolución de la abundancia correspondería a la connotación de plaga en cuanto afectación potencial directa en los cultivos.

La analítica planteada ha estado orientada en un primer paso a una racionalización preliminar espacio-temporal del conjunto de los datos hacia una estructura matricial zonal (basada en las zonas originales preestablecidas de vigilancia del programa de monitorización) y estacional (en función de periodos trimestrales asociados a las estaciones del año). En este punto se abordó la problemática referente a datos perdidos y se hizo una primera aproximación en cuanto a pruebas de significancia para detectar comportamientos diferenciales entre zonas de vigilancia para los diferentes periodos estudiados, y se llevaron a cabo test *post-hoc* que revelasen más concretamente cuáles eran estas diferencias en periodos conflictivos.

Posteriormente se utilizó un método de clasificación no supervisado de tipo jerárquico como herramienta de apoyo para la estimación de los valores ausentes detectados. Una vez abordado este punto, se procedió a la realización de una clasificación final de las zonas de vigilancia en cuanto a comportamientos diferenciales de la expresión de la abundancia a partir del método *DTW* o Deformación Dinámica del Tiempo.

En base a los grupos de zonas de vigilancia obtenidos, se estudiaron a partir del análisis de varianza las diferencias en la expresión de la abundancia entre ellos, para los distintos periodos y hábitats considerados, incluyendo los correspondientes test *post-hoc* para especificar las diferencias. Asimismo, se estudió la ciclicidad en cada uno de los grupos a través de *autocorrelogramas*.

Con independencia de los resultados y conclusiones parciales obtenidos en cada paso, el resultado global final del trabajo ha conseguido responder al objetivo principal de mismo, estableciendo una primera Clasificación Zonal de Castilla y León en cuanto a expresión diferencial de umbrales de riesgo. Dicha clasificación ha establecido la existencia de tres zonas diferenciales.

Por último, se ha abierto la puerta a futuros estudios que profundicen en esta temática, proponiendo un último mapa en el que, a su vez, cada una de esas tres zonas podría estar incluyendo subzonas que pudieran implicar variaciones en esos comportamientos diferenciales que pudieran resultar de interés.

Palabras clave: *Microtus arvalis*, Variaciones demográficas, Variabilidad espacio-temporal, Análisis cluster, Series temporales

ABSTRACT

The explosive demographic outbreaks of the peasant vole can cause significant damage and losses to crops in the agricultural ecosystems where it is present, with Castilla y León, the study area where this work is carried out, being one of the affected areas.

After a process of colonization that began at the end of the 70's of the last century, from the landscapes associated with the mountainous areas that border Castilla y León towards the interior of the plateau, where the main areas of agricultural production are located, the peasant vole is already present in the entire autonomous community since the beginning of this century.

Thus, until a few years ago, episodes of population outbreaks were recorded in Castilla y León on a recurring basis with an apparent cyclicity of five years. However, this pattern seems to have changed lately to shorter frequencies, at least in some parts of Castilla y León.

In addition, in all areas the peasant vole experiences those phenomena associated with the habitats where it lives. However, it does not seem to be “translated” everywhere in the same way in terms of its expression and impact on farm plots and, therefore, in the consequences that this type of biological behavior can have in terms of the conception of “plague”, not in a sense merely based on population size, but in terms of the risk of final affectation to crops.

Given the above, when an Integrated Management strategy is proposed against the pests caused by the peasant vole, the space-time study of areas and moments with different final risk thresholds for crops is crucial. And it is precisely a first approximation to the space-time rationalization of the differential expression of the evolution of natural abundance and risk for the peasant vole crops in Castilla y León that is worked on this *TFG*.

It has been based on the database of the last decade from the official program for monitoring the abundance of the common vole in Castilla y León, considering the information obtained in two types of habitat: “Dispersion routes” and “Crops”. The first represents a “reservoir” type habitat, where they inhabit periods of low population density and from which they expand in periods of high density, and in which the evolution of abundance corresponds to the natural biological effect without the influence of the human factor. The second represents a typical habitat to be protected in Integrated Management, including in this case the data on plots of annual herbaceous crops, and in which the evolution of abundance would correspond to the connotation of a pest as

a potential direct affectation on crops.

The analysis proposed has been oriented in a first step to a preliminary space-time rationalization of the data set towards a zonal matrix structure (based on the original pre-established surveillance zones of the monitoring program) and seasonal (based on associated quarterly periods to the seasons of the year). At this point, the problem related to missing data was addressed and a first approximation was made in terms of significance tests to detect differential behavior between surveillance zones for the different periods studied, and *post-hoc* tests were carried out to reveal more concretely what these differences were in periods of conflict.

Subsequently, an unsupervised hierarchical classification method was used as a support tool for estimating the missing values detected. Once this point was addressed, a final classification of the monitoring zones was carried out in terms of differential behavior of the expression of abundance using the *DTW* or Dynamic Time Warping method.

Based on the groups of surveillance zones obtained, the differences in the expression of abundance between them, for the different periods and habitats considered, were studied from the analysis of variance, including the corresponding *post-hoc* tests for specify the differences. Likewise, the cyclicity in each of the groups was studied through *autocorrelograms*.

Regardless of the results and partial conclusions obtained in each step, the final global result of the work has managed to respond to its main objective, establishing a first Zonal Classification of Castilla y León in terms of differential expression of risk thresholds. This classification has established the existence of three differential zones.

Finally, the door has been opened to future studies that delve into this subject, proposing a final map in which, in turn, each of these three zones could be including subzones that could imply variations in these differential behaviors that could result of interest.

Key words: *Microtus arvalis*, Demographic variations, Space-time variability, Cluster analysis, Time series

1. Justificación

Las plagas producidas por roedores, especialmente en los casos que están relacionadas con incrementos demográficos súbitos de sus poblaciones, constituyen un problema histórico por haber sido causantes de grandes pérdidas y destrozos en cultivos de todo tipo y a lo largo de todo el mundo [1].

En el presente trabajo se consideran las plagas ocasionadas por uno de estos roedores, el topillo campesino (*Microtus arvalis* Pallas) (Figura 1.1) en el territorio agrícola de Castilla y León. El topillo campesino es el mamífero más abundante en Europa, pero también está considerado como la plaga agrícola más perjudicial que existe en el continente [2].



Figura 1.1: Topillo Campesino (*Microtus arvalis*)

Este pequeño roedor arvicolino, de aspecto macizo y redondeado, con cabeza achatada, cola y orejas cortas y pelaje pardo-amarillento, con la zona ventral blanquecino-grisácea [3], es una especie fosorial (habita en madrigueras que excava bajo tierra, en las que pasa gran parte de su tiempo) que representa un nicho fundamental en la cadena trófica de los ecosistemas en los que está presente: es un gran consumidor y regulador de la fase vegetal y constituye presa fundamental para muchas especies de depredadores.

Considerando a sus depredadores, se encuentran de muy distintos tipos y modos de acción en la captura [4]: aves rapaces que los capturan desde el aire, como el cernícalo vulgar con actividad cazadora diurna, o la lechuza campestre que actúa por las noches; otras aves que los capturan en esperas desde el suelo, como garzas o cigüeñas; pequeños mamíferos que penetran en sus madrigueras, como musarañas o comadrejas; mamíferos de mayor tamaño que los capturan cuando salen al exterior, como los gatos, o que incluso escarban las madrigueras en su búsqueda, como los zorros; o reptiles como la culebra de escalera o el lagarto ocelado.

De hecho, hay varias iniciativas investigadoras sobre la potenciación de la actividad de algunos de estos depredadores como medio de control biológico para la protección de los cultivos frente al topillo campesino, como por ejemplo, la colocación de nidales artificiales (en el caso de las rapaces) de forma estratégica en ubicaciones agrícolas donde la nidificación natural es escasa por la ausencia de sustratos apropiados [5].

En cuanto a su papel de consumidor/regulador de la fase vegetal, es de naturaleza eminentemente herbívora, alimentándose de hojas, tallos, raíces, frutos, semillas... en definitiva, de cualquier parte de la planta a la que tenga acceso, incluyendo desde vegetación silvestre hasta prácticamente cualquier cultivo (cereales, leguminosas, colzas, girasoles, patata, remolacha, etc.) [2]. Este hábito polifágico cuando se considera el espectro de materia vegetal potencial en su dieta, incluyendo cultivos, supone el primer factor explicativo a su problemática como plaga agrícola.

El topillo campesino habitualmente vive en un tipo de hábitat óptimo para él, que se caracteriza por la presencia continua de vegetación herbácea, que le proporciona alimento y protección frente a parte de sus depredadores, y cuyo suelo no sufre alteraciones físicas por la actividad humana, lo que proporciona estabilidad a sus madrigueras. A este tipo de hábitat se le conoce como reservorios. En el paisaje agrícola encontramos estos reservorios en praderas, zonas abandonadas de cultivo, alfalfares, cunetas de caminos, linderos entre parcelas y entidades similares [2]. Cuando el tamaño de sus poblaciones es bajo, tiende a no abandonar estos reservorios. En consecuencia, y pese a su hábito herbívoro, cuando su presencia se limita a esos reservorios, el impacto sobre la mayoría de los cultivos es mínimo.

Sin embargo, el topillo campesino experimenta, de forma natural e inevitable, un fenómeno demográfico de carácter cíclico por el cual es capaz de multiplicar exponencialmente su tamaño poblacional en un periodo muy corto de tiempo. Estos abruptos cambios en sus poblaciones, conocidos como explosiones demográficas, se deben en gran parte a la expresión de gran capacidad reproductora [4]: puede llegar a ser ininterrumpida a lo largo de todo el año; las hembras pueden llegar a ser fértiles con tan sólo dos semanas de edad; tras periodos de gestación de unos 21 días alumbran camadas de hasta once crías; no presentan estro postparto (pueden quedar de nuevo preñadas nada más parir, sin necesidad de que pase el periodo de lactación); apareamiento poligónico tanto en hembras como en machos. Cuando se da esta expresión reproductora, el tamaño poblacional se multiplica en varios cientos en apenas unos meses.

Esto ocasiona una situación de competencia por el espacio, el alimento y por la propia actividad reproductora, por la cual parte de la población debe abandonar esos reservorios originales y colonizar otros tipos de hábitat que habitualmente no ocuparía. Y esos tipos de hábitat incluyen a los cultivos. En consecuencia, si a su hábito herbívoro se le añaden las consecuencias de expansión hacia los cultivos motivada por los fenómenos de explosión demográfica, se completa la razón de porqué pueden suponer una plaga agrícola y estar considerada como la más perjudicial en Europa [3]).

En añadido, también estos fenómenos pueden presentar connotaciones importantes en otros apartados como es el sanitario. Son portadores potenciales de muchas enfermedades, algunas de ellas zoonóticas, siendo el ejemplo quizá más conocido el caso de la tularemia (enfermedad de origen bacteriano) que puede presentarse en roedores silvestres, así como en otros mamíferos, aves, reptiles y peces, y que puede afectar también a los humanos, causando dificultad respiratoria, fiebre, escalofríos y otros síntomas, y que, en función de la cepa presente, puede incluso ocasionar mortalidad en individuos

no tratados [6] [7]). Otro ejemplo se tiene en determinados tipos de virus, entre los que se encuentran los ortopoxvirus (género de virus ADN que abarcan varios tipos de virus de viruela, incluyendo la actualmente controvertida viruela del mono [8]).



Figura 1.2: Distribución del topillo campesino en Europa y Asia [9]

Este pequeño roedor se distribuye de forma continua a lo largo de Europa y parte del continente Asiático (figura 1.2), donde sufre explosiones demográficas más o menos regulares en intervalos de 2 a 5 años, superando los 2000 individuos por hectárea (orden de magnitud mucho mayor que en otros roedores) [2]; y cuenta con una población aislada en la Península Ibérica (figura 1.3). Con su llegada a la Península hace más de 90 000 años, su aparición en la Comunidad Autónoma de Castilla y León se centró en la ocupación de las zonas y parajes montanos de los sistemas Cantábrico, Ibérico y Central, rodeando así la mayor parte de la Meseta Norte.

A partir de los años 70, ha ido colonizando de forma progresiva e ininterrumpida los distintos paisajes agrícolas de la cuenca del Duero, reflejándose así en brotes cíclicos de daños en los cultivos en intervalos de 5 años, coincidiendo con la detección de explosiones demográficas de los mismos [10]. En la campaña año 2007/08, se produjo el fenómeno demográfico en la Castilla y León de mayor magnitud, tanto en tamaños poblacionales alcanzados o superficie afectada como otros factores como la propia repercusión mediática que generó. En ese momento, el topillo ya había colonizado la totalidad de Castilla y León, incluyendo sus las principales zonas productoras agrícolas en la meseta. [10] .

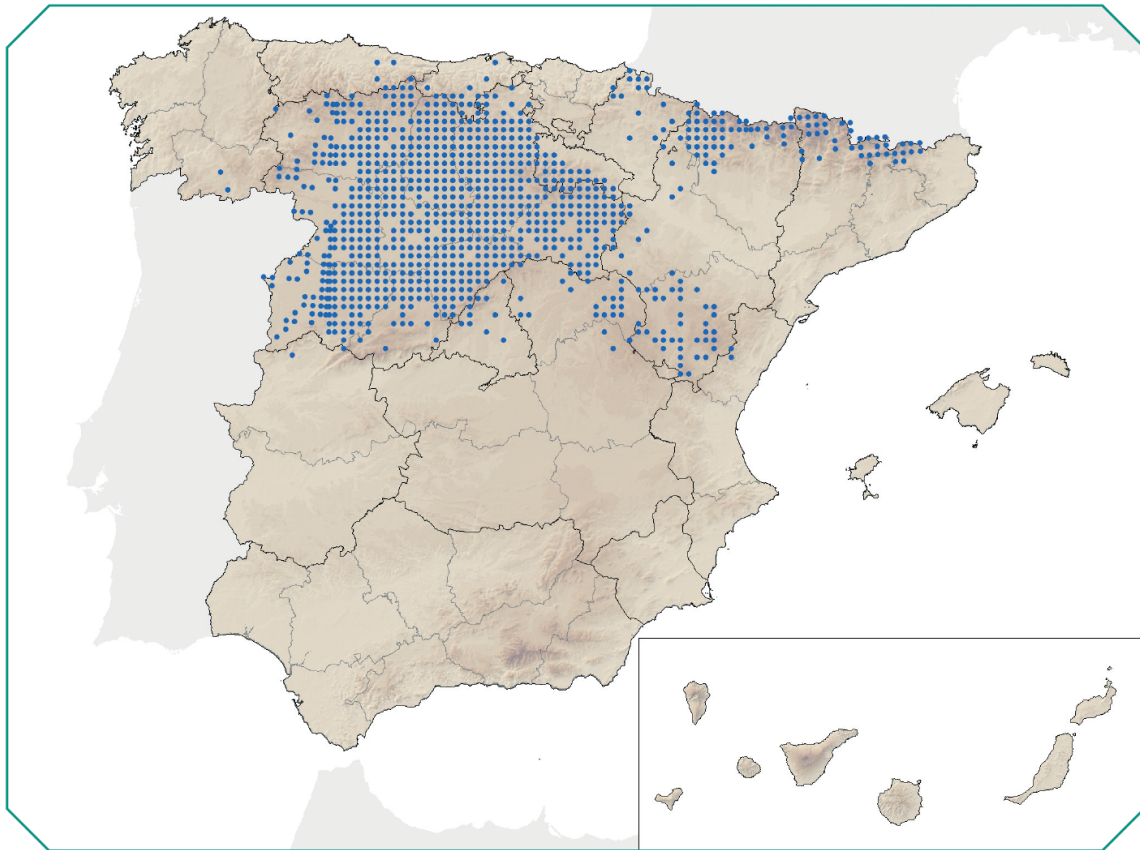


Figura 1.3: Distribución de topillo campesino en España [3]

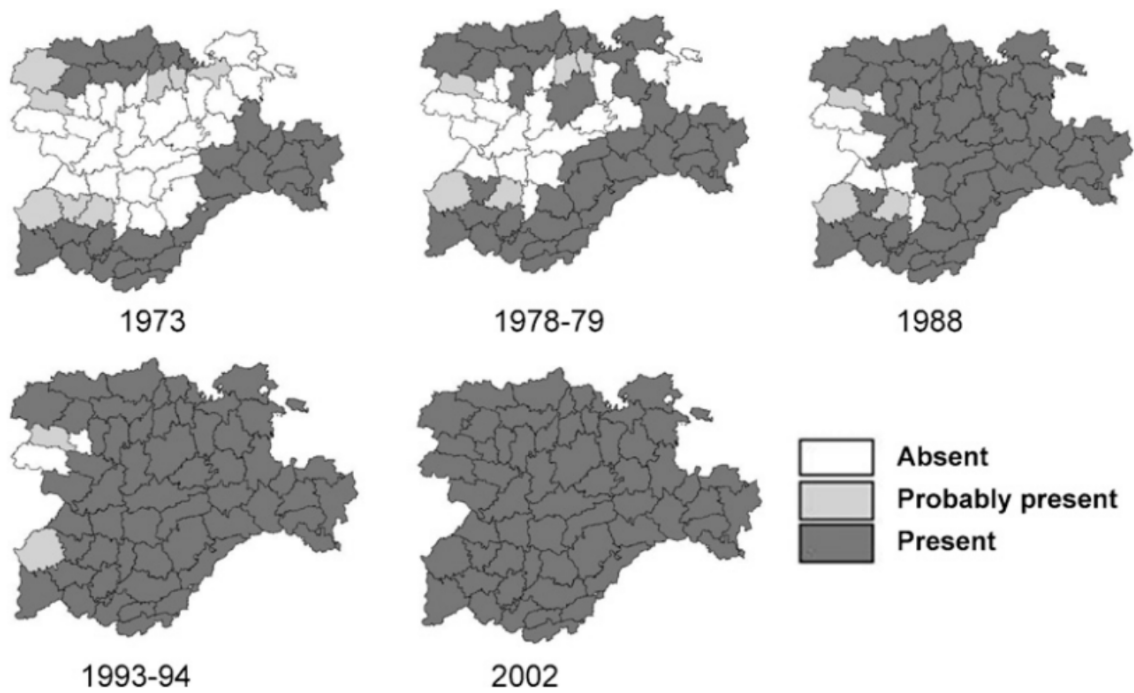


Figura 1.4: Proceso de colonización del topillo campesino en la región de Castilla y León desde 1973 hasta 2002 [10]

En consecuencia de todo lo anterior, mantener una monitorización periódica del estado de sus poblaciones en zonas agrícolas se torna en una tarea necesaria, suponiendo un gran reto a nivel organizativo, científico y económico. Para llevar a cabo esta monitorización existen varios métodos de extracción y recopilación de información como pueden ser los métodos basados en el trapeo (capturas de única sesión, métodos de captura-marcado-recaptura, ...), o bien el estudio a partir de índices indirectos de actividad. Ha sido uno de estos últimos el método con el que se ha trabajado a lo largo de este estudio. Estos índices indirectos de actividad persiguen inferir la presencia de topillo campesino y cuantificarla en una medida relacionada con su abundancia, a partir de indicios/evidencias observadas, en el caso de este trabajo, asociadas a las madrigueras en las que habita.

Para poder llevar a cabo todas estas tareas, la Consejería de Agricultura, Ganadería y Desarrollo Rural (en el marco del Plan Director de Lucha contra Plagas Agrícolas en Castilla y León y la Orden AYG/96/2019 de gestión integrada de riesgos derivados de la presencia de topillo campesino en el territorio de Castilla y León [11]) tiene establecido un programa de prevención, control y minimización de riesgos, siendo coordinado por el Observatorio de Vigilancia y Control de Plagas Agrícolas de Castilla y León (perteneciente al ITACyL, Instituto Tecnológico Agrario de Castilla y León) y ejecutado en conjunto con los Servicios Territoriales de Agricultura, Ganadería y Desarrollo Rural de las distintas provincias, y bajo la asesoría científica de la Comisión de roedores y otros vertebrados del Comité Científico de Lucha Contra Plagas Agrícolas en Castilla y León.

El correcto uso de la estadística en este ámbito será fundamental para conocer los distintos aspectos que pueden estar relacionados con la actividad de estos característicos roedores. Necesitaremos conocer el origen de los datos manejados, su forma de recogida, así como su posterior tratamiento y depuración, para situarnos correctamente en el contexto del problema y lograr nuestros objetivos planteados. Será necesario tener una visión general de lo sucedido a lo largo del tiempo de estudio, y elegir correctamente aquellos métodos que permitan hacer un estudio correcto, consistente y fiable.

Uno de los problemas que más acostumbra a surgir en lo referente al análisis de datos funcionales es la facilidad para encontrarse con situaciones de datos perdidos o ausentes, ya que no siempre es posible recoger información en una zona concreta, bien sea por problemas económicos (falta de recursos para asistir al lugar), climáticos (encontrarse con situaciones adversas en las que no es posible hacer la recogida de los datos) o de cualquier otro tipo.

La búsqueda de similitudes entre diferentes zonas de estudio puede ser una vía válida de trabajo, que cumplimente esta ausencia de información original, y permita al personal encargado del mantenimiento y supervisión del entorno agrícola tomar las mejores decisiones de prevención y poder tomar las medidas que consideren oportunas a tiempo, consiguiendo así un menor impacto en la producción agrícola.

Por otro lado, el uso de diferentes metodologías estadísticas de estimación de valores ausentes será muy útil para completar esta información que no se pudo recoger en la fase de toma de datos. Estos métodos han de ser cuidadosamente elegidos, ya que un abuso

en esta estimación puede causar una suposición de los hechos muy distinta a la realidad.

El lenguaje *R* es un lenguaje de programación orientado al análisis estadístico. A día de hoy es muy usado en investigación científica, siendo muy popular en campos como el *machine learning* (o aprendizaje automático), economía, biología o bioestadística, entre otros. Proporciona un amplio abanico de librerías y paquetes especializados en distintos ámbitos que facilitan la extracción, depuración, análisis y visualización de los datos, otorgando al usuario una flexibilidad de trabajo muy interesante.

Dentro del presente trabajo fin de grado (de aquí en adelante, *TFG*), se pretende contribuir en la problemática expuesta mediante la racionalización espacio-temporal de la evolución de la abundancia natural y de los riesgos para los cultivos del topillo campesino en Castilla y León, orientada a la detección de distintas zonas con expresión diferencial y que, por tanto, deberían ser consideradas como tal en todos los procesos relacionados con la Gestión/Lucha Integrada frente a los riesgos derivados del topillo campesino. Todo ello, ayudándonos de las diferentes herramientas y métodos que permiten trabajar con información ausente. Para conseguirlo, se plantearon varios métodos de clasificación de datos funcionales, se hizo un estudio de los diferentes grupos resultantes, y se analizó la evolución temporal de cada grupo, incluyendo el estudio de la ciclicidad.

2. Objetivos

Considerando los datos de los últimos diez años obtenidos en el programa oficial de monitorización de la abundancia de topillo campesino en Castilla y León, en la elaboración del presente *TFG* se ha planteado un objetivo global, relativo a la contribución a la mejora de la Gestión Integrada de las plagas ocasionadas por esta especie en Castilla y León, desglosado en una serie de objetivos específicos, considerados en función de las distintas fases de análisis estadístico abordadas. En concreto:

2.1. Objetivo global

Detección de zonas diferenciales de evolución de la abundancia (tanto natural como en su expresión en los cultivos) del topillo campesino en Castilla y León, orientada a la definición de zonas con diferentes umbrales de riesgo a ser consideradas en los procesos de Gestión/Lucha Integrada frente a las plagas agrícolas ocasionadas por esta especie.

2.2. Objetivos específicos

2.2.1. Exploración y racionalización espacio-temporal inicial de la base de datos

El primer objetivo se centró en la exploración de la base de datos y en la racionalización espacio-temporal de la misma, orientada a su reestructuración matricial de dos vías (espacio en zonas y tiempo en estaciones) en cuanto a expresión del índice de abundancia promedio del topillo campesino. En este objetivo se han abordado, por ejemplo, problemas relacionados con datos perdidos o sistemas de ponderación en cuanto a atribución de datos en zonas.

2.2.2. Detección inicial de expresiones diferenciales de la abundancia a nivel espacio-temporal

En el segundo objetivo se ha abordado el estudio de la matriz obtenida en el objetivo anterior para la detección de los momentos temporales en los cuales se produjo expresión diferencial de la abundancia en función de la categorización espacial, a partir de la metodología del Análisis de Varianza (ANOVA).

2.2.3. Agrupación espacial en función de la expresión diferencial de la evolución de la abundancia mediante el uso de métodos de aprendizaje no supervisado

El tercer objetivo focalizó los esfuerzos en la agrupación de las zonas de vigilancia por similitud en su evolución temporal. Para ello se han tratado aspectos como las medidas de similitud o disimilitud y la aplicación de métodos de agrupación o *clustering*. Respecto a esto último, se realizó una primera jerarquización aportando varias comparaciones gráficas de la composición de los grupos obtenidos según los tipos de hábitat de estudio. Posteriormente, esta jerarquización inicial se optimizó mediante la aplicación del método *DTW* o Deformación Dinámica del Tiempo. El resultado de este objetivo representaría las zonas diferenciales de riesgo perseguidas en el objetivo global del estudio.

2.2.4. Definición final de zonas diferenciales de riesgo y estudio comparativo en la evolución y ciclicidad

En el cuarto objetivo se procedió a la detección de las diferencias entre las zonas diferenciales de riesgo. Para ello se estudiaron las series temporales de las zonas de riesgo obtenidas, se utilizó de nuevo el ANOVA, así como pruebas/test *post-hoc* para comprender el comportamiento diferencial en cada periodo estudiado y se incluyó la comparativa de la ciclicidad en cada zona a través de la obtención de correlogramas.

Como resultado final del cumplimiento gradual de las tareas relacionadas con cada uno de estos objetivos específicos, se habrá conseguido la detección de zonas de riesgo con expresión diferencial de la abundancia del topillo, así como el conocimiento de las diferencias entre ellas. Esta información resulta de indudable interés para su consideración y aplicación en las tareas futuras relacionadas con la Gestión Integrada de las plagas de topillo campesino. Y, por tanto, los resultados de los distintos objetivos específicos rematan con el cumplimiento del objetivo global que se ha planteado en el presente *TFG*.

3. Materiales y métodos

Dedicaremos esta sección a conocer con detalle las diferentes metodologías usadas para llevar a cabo la recogida de información, análisis y cálculos que justifican las posteriores conclusiones del *TFG*.

3.1. Descripción de la metodología de datos en campo

La información manejada a lo largo del proyecto procede de años de estudio y seguimiento de la actividad del roedor en el ámbito rural del área geográfica de Castilla y León. Más concretamente, estos datos se recogieron a lo largo de 10 años consecutivos (desde el mes de noviembre de 2011, hasta el mes de julio de 2021).

Para una mejor organización de esta monitorización, se dividió la región de trabajo en lo que se conoce como *comarcas agrarias*, que a su vez contendrán múltiples *zonas de vigilancia*. Cada *zona de vigilancia* contará con los recintos que le hayan sido asignados (figura 3.5), donde se realizaron las inspecciones pertinentes para obtener la información y poder así hacer el seguimiento.



Figura 3.5: División del espacio en el estudio

3.1.1. Descripción del área de estudio

Como es lógico, el control de esta actividad no se hace para toda la extensión de terreno (sería una tarea inabarcable, así como inviable económicamente), sino que se

llevan a cabo muestreos controlados en determinadas parcelas.

Comarcas agrarias

El territorio denominado *comarca agraria* es aquel que se define (por parte del Ministerio de Agricultura, Alimentación y Medio Ambiente [12]) como las superficies dentro de cada provincia que integran un número determinado de términos municipales con una homogeneidad notable en aspectos como su potencial productivo agrario, su sistema de cultivo y aprovechamiento agrario y su desarrollo económico.

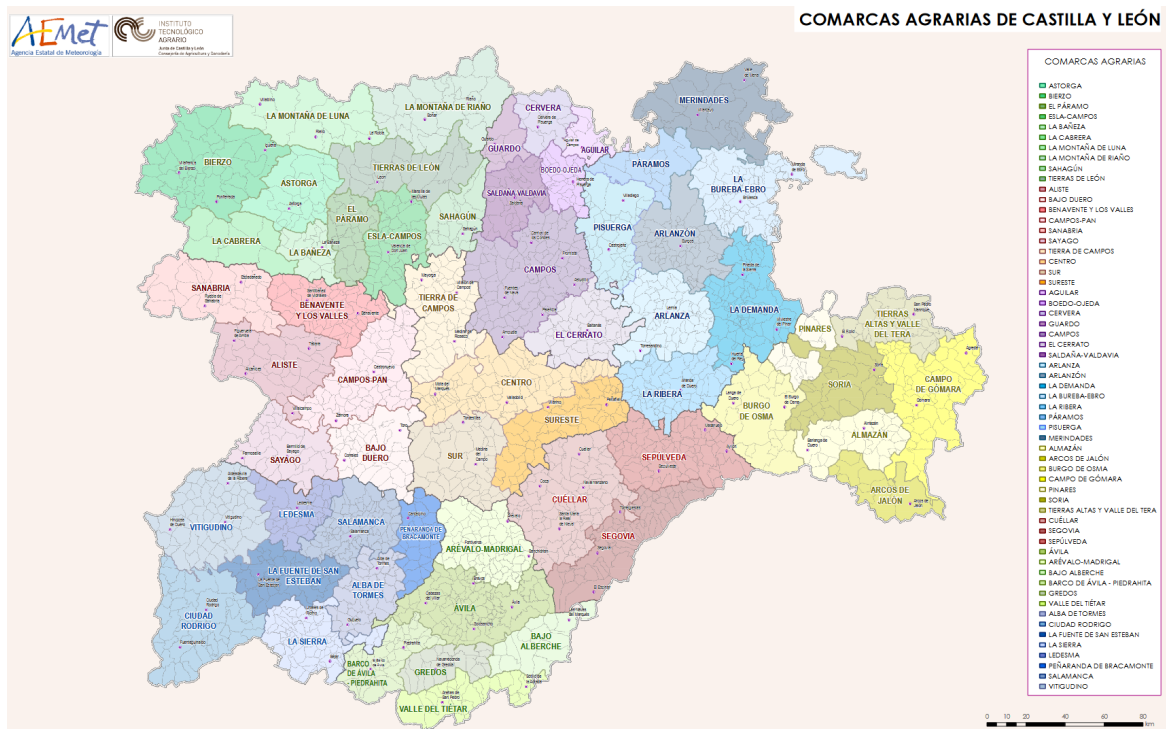


Figura 3.6: Mapa de las comarcas agrarias de Castilla y León

El conjunto de datos disponía de información de 40 de las 59 comarcas agrarias presentadas en la figura 3.6. Estas comarcas se presentan en el anexo 6.2, en función de la provincia a la que pertenecen.

Zonas de vigilancia

Las zonas de vigilancia preliminares en este trabajo constituían áreas del territorio con características diferenciales en cuanto a criterios geográficos, ecológicos, agronómicos y de la información histórica sobre la incidencia del topillo campesino. En cada una de esas zonas de vigilancia se realizan las inspecciones periódicas en un conjunto de recintos representativos. Por recinto se considera la unidad mínima del territorio establecida según la división del mismo en función del SIGPAC, Sistema de Información Geográfico de la Política Agraria Comunitaria (a partir de ahora, por agilizar la redacción del documento y la lectura, se usarán indistintamente las palabras “recinto” y “parcela”, aunque esta última corresponde realmente a una categoría superior en la nomenclatura del SIGPAC).

A efectos de nomenclatura para el presente estudio, cada zona de vigilancia se identifica preliminarmente concatenando las 2 primeras letras de la provincia a la cual pertenece, un guión, y el número que le haya sido asignado. El número de zonas de vigilancia asignadas por cada provincia se resume en la tabla 3.1:

Tabla 3.1: Número de zonas de vigilancia por provincia

Provincia	Nº Zonas de vigilancia
Ávila	4
Burgos	4
León	3
Palencia	11
Salamanca	4
Segovia	4
Soria	4
Valladolid	8
Zamora	5

León es la provincia con menos zonas de vigilancia asignadas (con un total de 3), mientras que Palencia con 11 zonas, y Valladolid con 8, encabezan la lista.

Las inspecciones en los recintos se llevan a cabo de la siguiente forma: se marca un transecto de 99 metros a lo largo de la vía de dispersión adyacente, y otros dos transectos en el interior del recinto, perpendiculares al primero y paralelos entre sí, cada uno de ellos también con 99 metros de largo.

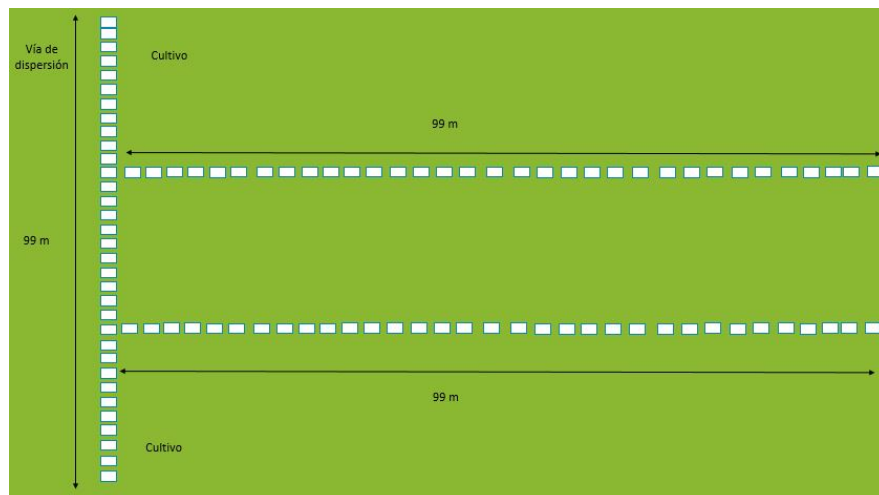


Figura 3.7: Colocación de los transectos

Tal y como aparece en la figura 3.7, cada transecto se divide en rectángulos de 3m x 1.5m formando 33 unidades muestrales de $4.5 m^2$. Por lo tanto, en cada inspección, se tendrá en cuenta el estado de 99 unidades muestrales, de las cuales 66 pertenecen al interior del recinto en cuestión, y 33 pertenecen a la vía o camino que linda con él.

Posteriormente se categoriza cada unidad muestral en función de la presencia o ausencia de indicios que identifiquen colonización activa reciente del roedor, considerando para ello:

- Presencia de madrigueras (figura 3.8). Esta condición es indispensable para que la evaluación defina evidencia de colonización. Para la identificación de evidencias de actividad reciente, el requerimiento es que esta presencia de madrigueras esté acompañada por al menos uno de los siguientes indicadores:
 - Presencia de excrementos frescos en las bocas de las madrigueras.
 - Acumulación de vegetación cortada fresca en las bocas de las madrigueras.
 - Evidencias de excavación reciente (presencia de tierra desmenuzada fresca).



Figura 3.8: Madriguera de topillo campesino

El cálculo del índice de actividad/abundancia viene dado por la siguiente expresión:

$$I_{act} = \sum_{i=1}^n I_{presencia}(i), \quad (3.1)$$

Con $I_{presencia}(i)$ el indicador en la unidad muestral i -ésima de evidencias de presencia u ausencia de topillo campesino, y n el número de unidades muestrales evaluadas. Este índice de abundancia es interpretado como la proporción (en porcentaje) de superficie de terreno con evidencias de presencia en esa determinada inspección, lo que representa el porcentaje de superficie con colonización activa, habiendo sido calibrado para su relación con la densidad real de topillo campesino.

Parte de los trabajos orientados a dicho calibrado fueron el objetivo de un Trabajo Fin de Grado en Estadística de la Universidad de Valladolid [13] en el que se definió que la relación entre la densidad y el índice de abundancia respondía a la siguiente relación, con un ajuste de la curva resultante de $R^2 = 0,9$.

$$\sqrt{D} = 0,57 + 0,47 * I_{act} \quad (3.2)$$

Siendo D la densidad de topillo campesino asociada a la superficie colonizada e I_{act}

el índice de abundancia.

La detección de estas evidencias será anotada por el técnico especialista que esté revisando el recinto en cuestión, a través de la App CartoDruid.

CartoDruid (Figura 3.9) es una herramienta GIS (Geographic Information System) creada y desarrollada por el Instituto Tecnológico Agrario de Castilla y León, que sirve como método de apoyo para determinadas tareas realizadas en el campo.



Figura 3.9: CartoDruid

El uso de esta aplicación resuelve numerosos problemas de utilización y manejo de información en lugares que pueden no tener conexión a Internet (puesto que es bien conocido que muchas zonas apartadas del entorno urbano carecen de cobertura, siendo este un gran impedimento para el uso de mapas conectados a la red), y donde se llevan a cabo las labores y mediciones de las que surgirán nuestros datos de partida.

3.1.2. Descripción de los conjuntos de datos

El conjunto de datos con el que se trabajó contenía información referente a los resultados de las inspecciones llevadas a cabo durante los últimos 10 años en la comunidad autónoma de Castilla y León. Cada entrada de esta base de datos contiene:

- Un identificador de inspección, único para cada entrada.
- La comarca agraria asociada a ese recinto concreto.
- La provincia en la que se encuentra.
- El municipio al que pertenece el recinto donde se ha llevado a cabo la inspección.
- Las coordenadas de los centroides asociados a dicha parcela.
- El hábitat al que pertenece el recinto.
- La fecha en la que se llevó a cabo la inspección.
- El valor del índice de actividad registrado.
- El nivel de riesgo con el que se cataloga.
- Zona de vigilancia asignada.

- Distancia al centroide.
- Valores de la ponderación por comarca, y por distancia al centroide.
- El valor total de la ponderación espacial.

Así, contábamos con 47 zonas de vigilancia repartidas por la comunidad, y con un total de 112414 entradas en la base de datos. De ese total de entradas, 2770 pertenecían a inspecciones de recintos sin zona de vigilancia asignada, que no se tuvieron en cuenta para la realización del estudio. Esto nos dejó con un total de 109644 inspecciones.

Hábitats de estudio

Los 4 hábitats en los que se podía catalogar cada inspección son:

- **Cultivos**, que representan las parcelas dedicadas a las rotaciones típicas cerealistas.
- **Reservorios en cultivo**, representando aquellos cultivos de alfalfa (o bien otros cultivos herbáceos plurianuales) que se asocian directamente a parcelas cultivadas.
- **Reservorios no cultivo**, que son aquellas para las cuales la acción del hombre es mínima o nula (perdidos, parcelas retiradas de cultivo, parcelas...). Estas no están asociadas directamente a un cultivo, pero en un futuro pueden constituir un reservorio o fuente de dispersión.
- **Vías de dispersión**, representando las zonas de territorio con condiciones similares a los reservorios, pero que también pueden constituir las vías de dispersión desde las zonas habituales de presencia (considerándose cunetas, lindes, regatos y arroyos).

El número de observaciones por cada uno de los hábitats viene dado en la tabla 3.2:

Tabla 3.2: Número de inspecciones por hábitat

Hábitat	Número de inspecciones
Cultivo	32310
Reservorios en cultivo	12360
Reservorios no cultivo	11415
Vías de dispersión	56329

Esta estructura básica de distribución espacial original, especialmente en lo concerniente a zonas de vigilancia y recintos representativos a lo largo del tiempo, ha ido sufriendo una serie de modificaciones desde sus inicios, en función de la propia evolución de los procedimientos de trabajo, de los propios eventos de explosiones demográficas

del topillo habidos y de la intensidad de los esfuerzos que a efectos prácticos eran necesarios en cada momento. En consecuencia, y para unificar a efectos del presente estudio las posibles variaciones espaciales habidas por los procesos indicados y asignar de forma correcta las ubicaciones de los distintos recintos considerados a lo largo de los años, se redefinieron las zonas de vigilancia y asignaron los distintos recintos de la siguiente forma:

A partir de las coordenadas de los recintos incluidos en el conjunto de municipios que componían cada zona de vigilancia inicial, se calculó el centroide de la superficie geográfica en que estaban distribuidos, considerándolo como el centroide teórico de la zona de vigilancia en sí. Sobre este centroide, para cada zona de vigilancia, se trazó un radio de 10 km, de forma que aquellos recintos que se encontrasen dentro de este rango, serían asignados a la zona de vigilancia en cuestión.

En los casos en que por el radio de 10 km pudiera darse un solapamiento entre zonas, un determinado recinto se asignó a aquella zona cuya distancia entre centroide y recinto fue menor, siempre y cuando las zonas de vigilancia pertenezcan a la misma comarca agraria. En caso de que las zonas de vigilancia estaban en comarcas diferentes, la asignación se hará a la zona que se encuentre en la misma comarca agraria al que pertenece dicho recinto.

A efectos de recintos susceptibles de aportar información a una determinada zona de vigilancia se incluyeron también, siguiendo el mismo criterio de asignación, recintos en radio máximo desde el centroide de hasta 30 kilómetros, en el caso de coincidencia de comarca recinto/zona de vigilancia, y de máximo de 20 en caso de no coincidir la comarca del recinto con la de la zona de vigilancia. En estos casos, como se verá en el apartado siguiente, los recintos se incluyeron con un criterio de ponderación de su información sobre la zona de vigilancia en función de la distancia al perímetro establecido de 10 km en el paso primero.

Para la nomenclatura final en cuanto a zonas de vigilancia se estableció un código que identificaba el código original de la zona de vigilancia (dos primeras letras de la provincia seguidas de un guión y un número) seguido de un guión y dos letras identificativas de la comarca agraria en que se sitúa cada una. A efectos de facilitar la interpretación de resultados o para facilitar la ubicación por un lector conocedor de la geografía castellano y leonesa, el código se completó con un nuevo guión seguido del nombre de un municipio de referencia, considerando como tal a alguno de los que tenían recintos incluidos en la zona de vigilancia y que resultase más conocido dentro del conjunto de municipios que la componen (para facilitar así el posterior manejo de estas zonas). Estos municipios de referencia aparecen reflejados en el anexo 6.1.

Periodos de estudio

Para el desarrollo del presente estudio, la racionalización de la distribución temporal de las inspecciones en los distintos recintos se ha seguido un criterio de año y estacionalidad, de forma que se tomó cada trimestre como referencia para agrupar las

diferentes observaciones. Así, los periodos de tiempo considerados aparecen reflejados en la tabla 3.3:

Tabla 3.3: Correspondencia de los periodos de estudio

Periodo	Año	Trimestre considerado	Estación atribuida
1	2011	De octubre a diciembre	Otoño
2	2012	De enero a marzo	Inverno
3	2012	De abril a junio	Primavera
4	2012	De julio a septiembre	Verano
5	2012	De octubre a diciembre	Otoño
6	2013	De enero a marzo	Inverno
7	2013	De abril a junio	Primavera
8	2013	De julio a septiembre	Verano
9	2013	De octubre a diciembre	Otoño
10	2014	De enero a marzo	Inverno
11	2014	De abril a junio	Primavera
12	2014	De julio a septiembre	Verano
13	2014	De octubre a diciembre	Otoño
14	2015	De enero a marzo	Inverno
15	2015	De abril a junio	Primavera
16	2015	De julio a septiembre	Verano
17	2015	De octubre a diciembre	Otoño
18	2016	De enero a marzo	Inverno
19	2016	De abril a junio	Primavera
20	2016	De julio a septiembre	Verano
21	2016	De octubre a diciembre	Otoño
22	2017	De enero a marzo	Inverno
23	2017	De abril a junio	Primavera
24	2017	De julio a septiembre	Verano
25	2017	De octubre a diciembre	Otoño
26	2018	De enero a marzo	Inverno
27	2018	De abril a junio	Primavera
28	2018	De julio a septiembre	Verano
29	2018	De octubre a diciembre	Otoño
30	2019	De enero a marzo	Inverno
31	2019	De abril a junio	Primavera
32	2019	De julio a septiembre	Verano
33	2019	De octubre a diciembre	Otoño
34	2020	De enero a marzo	Inverno
35	2020	De abril a junio	Primavera
36	2020	De julio a septiembre	Verano
37	2020	De octubre a diciembre	Otoño
38	2021	De enero a marzo	Inverno
39	2021	De abril a junio	Primavera

De esta forma, cada periodo comenzaba el 1 del mes que le correspondiese (enero, abril, julio u octubre) y terminaba el 30 o 31 del último mes del trimestre. A efectos de nomenclatura más comprensible, a cada periodo se le atribuyó una estación, siendo la que la mayor parte del tiempo transcurriese en ese periodo de estudio.

3.2. Metodología estadística

3.2.1. Método de estimación del índice de abundancia

El cálculo del índice de actividad en cada zona de vigilancia para un periodo concreto, fue calculado como la media ponderada de cada una de las inspecciones realizadas en los recintos correspondientes a esa zona de vigilancia, atendiendo a la distancia entre el centroide de dicha zona, y las coordenadas del recinto en cuestión.

Así, se ponderó teniendo en cuenta la distancia entre el recinto y un perímetro de 10 km alrededor del centroide de la zona de vigilancia (figura 3.10), con lo que se consideró la influencia de aquellas inspecciones realizadas en un radio máximo de 30 km (20 km en el caso de que se encuentren en distinta comarca agraria). De esta forma, las medias obtenidas para cada zona de vigilancia en un periodo concreto, tendrán influencia de los recintos que tengan asignados, así como de aquellos más próximos que no lo estén.

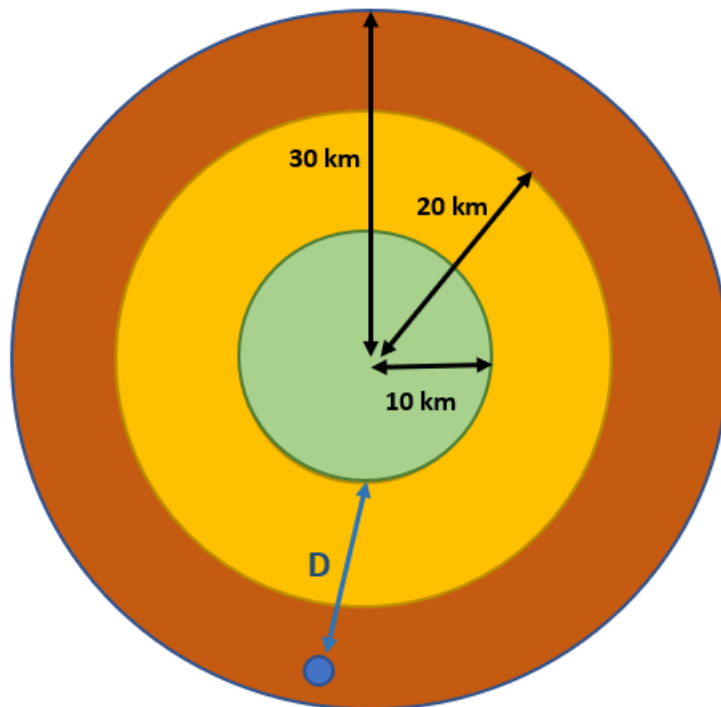


Figura 3.10: Distancia D entre el recinto y el perímetro de 10 km alrededor del centroide de la zona de vigilancia

A mayores, se ponderó con 1 si el recinto pertenecía a la misma comarca agraria que la zona de vigilancia, y con 0.5 en el caso contrario.

La expresión del índice de actividad medio $I_{medio}(i, j)$ en la zona i y el periodo j vendrá dado por la siguiente expresión:

$$I_{medio}(i, j) = \sum_{k=1}^P \frac{I(k, j) \times W(k)}{P} \quad (3.3)$$

Con:

- P el número de inspecciones consideradas para la zona i en el periodo j .
- $I(k, j)$ el índice de actividad obtenido en la inspección k –ésima.
- $W(k)$ la ponderación otorgada a dicha inspección, obtenida a partir del producto de la ponderación por comarca, y por distancia.

$$W(k) = C(k) \times D(k) \quad (3.4)$$

Como paso final antes de construir las matrices cruzadas espacio-temporales, se contabilizó el número de inspecciones con las que se había calculado el índice de actividad medio para cada zona de vigilancia y periodo.

Se observó que había un cierto número de índices calculados a partir de un número muy pequeño de inspecciones. Esto podría ser una situación en la que la muestra no es representativa, lo cual influiría negativamente a la hora de plantear inferencias sobre los resultados.

Por ello, se tomó como medida cautelar la decisión de no tener en cuenta aquellas índices obtenidos con menos de 5 inspecciones, tomando aquellos valores que cumplieren esta condición como valores ausentes.

3.2.2. ANOVA de un factor

Como medio para detectar aquellos periodos en los cuales existían diferencias significativas entre zonas de vigilancia, se utilizó el conocido método ANOVA.

El ANOVA (o análisis de la varianza) de un factor trata de buscar el efecto de una variable independiente categórica sobre las diferentes medias de una variable de tipo continuo, cuando este factor dispone de, al menos, 2 o más niveles. Esta técnica fue desarrollada por el matemático y biólogo Ronald Aylmer Fisher en 1930 [14].

En el análisis de la varianza [15], se parte de la hipótesis nula de igualdad de medias para los diferentes niveles de estos factores, de la siguiente manera:

- H_0 : No existen diferencias significativas entre las medias de los p niveles del factor: $\mu_1 = \mu_2 = \dots = \mu_p$
- H_1 : Al menos hay diferencia significativa entre 2 medias: $\exists i, j, i \neq j$, tal que $\mu_i \neq \mu_j$

Esta hipótesis nula implicaría que los miembros de los diferentes grupos proceden de una misma población, con igual media y varianza. Así, estas medias por grupos se compararán haciendo un estudio de sus respectivas varianzas, siendo necesaria una descomposición de la variabilidad en *intervarianza* (varianza entre los niveles del factor) e *intravarianza* (varianza dentro de cada nivel del factor) de esta forma:

Variabilidad total = Variabilidad explicada por el factor + Variabilidad residual o no explicada por el factor

Lo que se traducirá como

Variabilidad total = Variabilidad entre niveles + Variabilidad dentro de los niveles

O bien

Variabilidad total = Intervarianza + Intravarianza

Llegados a este punto, será necesario comprender el concepto de *Sumas de Cuadrados* para hacer el cálculo del estadístico:

- **Total Sum of Squares o TSS:** Representa el total de la variabilidad de los datos, y se calcula como la suma de los cuadrados en las diferencias entre el valor observado, y la media global de los datos.
- **Sum of Squares of Treatment o SST:** Representa la variabilidad producida por el factor sobre la media de los datos, y se calcula como la suma de los cuadrados de la diferencia entre las medias de cada grupo, y la media global de la variable.
- **Sum of Squares of Error o SSE:** Representa la variabilidad dentro de cada nivel del factor, y se obtiene sumando los cuadrados de las desviaciones de cada instancia con la media relativa al nivel del factor que le corresponde.

La descomposición de la variabilidad se puede expresar con respecto a las sumas de cuadrados de la siguiente manera:

$$TSS = SST + SSE \quad (3.5)$$

Será necesario dividir estas sumas de cuadrados entre sus respectivos grados de libertad, para así poder conseguir la expresión de las sumas de cuadrados medios y construir el estadístico con el que se comprobará la igualdad de medias:

- Variabilidad total = Cuasivarianza total = Cuadrados medios totales = $\frac{TSS}{N-1} = \hat{S}_T^2$
- Intervarianza = Cuadrados medios del factor = $\frac{SST}{k-1} = \hat{S}_t^2$
- Intravarianza = Cuadrados medios del error = $\frac{SSE}{N-k} = \hat{S}_E^2$

Haciendo uso de las equivalencias anteriores, podemos llegar a la expresión del estadístico F_{ratio} :

$$F_{ratio} = \frac{\hat{S}_t^2}{\hat{S}_E^2} = \frac{intervarianza}{intravarianza} \sim F_{k-1, N-k} \quad (3.6)$$

Que se usará para detectar la diferencia entre las medias de las distintas categorías a comparar. El indicador utilizado para validar estas diferencias es el conocido como *p-valor*, calculado a partir de la siguiente expresión:

$$p - valor = 1 - P(F_{ratio} > F_{k-1, N-k}) \quad (3.7)$$

Valores pequeños de este *p-valor* indicarán la posibilidad de rechazar la hipótesis de igualdad de medias para los diferentes niveles del factor estudiado.

3.2.3. Test LSD de Fisher

Los test post-hoc son necesarios para determinar cuales son las medias para las que se detectan diferencias significativas, tras rechazar la hipótesis nula de igualdad de medias en el procedimiento ANOVA. En 1935, Fisher sugirió la que una de las técnicas más antiguas y populares de comparaciones múltiples: el contraste de la mínima diferencia significativa[15].

Para cada pareja de medias de los k niveles del factor de estudio, se realizará un contraste de hipótesis con la siguiente forma:

- $H_0 : \mu_i = \mu_j$
- $H_1 : \mu_i \neq \mu_j$

Así, se determina la expresión del estadístico t :

$$t = \frac{\bar{y}_i - \bar{y}_j}{\sqrt{\hat{S}_E^2(\frac{1}{n_i} + \frac{1}{n_j})}} \sim t_{N-k} \quad (3.8)$$

donde

- \bar{y}_i, \bar{y}_j representan las medias de la variable numérica en cuestión, para los niveles i -ésimo y j -ésimo del factor.
- \hat{S}_E^2 representa la varianza residual.
- n_i, n_j representan los tamaños muestrales para los niveles i -ésimo y j -ésimo del factor.
- N representa el número total de observaciones en el conjunto de datos.
- t_{N-k} representa una distribución t de *Student* con $N - k$ grados de libertad.

A partir de la construcción de este estadístico, se calcula un valor umbral o frontera denominado LSD o *Less Significant Difference*, usando la siguiente expresión:

$$LSD = t_{N-k(\frac{\alpha}{2})} \sqrt{\hat{S}_E^2(\frac{1}{n_i} + \frac{1}{n_j})} \quad (3.9)$$

con $t_{N-k(\frac{\alpha}{2})}$ el valor crítico de una distribución t de *Student* con $N - k$ grados de libertad.

Con este valor calculado, la comparación que se lleva a cabo es la diferencia (en valor absoluto) de ambas medias, frente a este valor:

$$|\bar{y}_i - \bar{y}_j| > LSD \quad (3.10)$$

Si esta condición se cumple, la pareja de medias es significativamente distinta a nivel $1 - \alpha$. Este test *post-hoc* no conlleva a priori ninguna corrección de significancia, lo que puede suponer una inflación en el error de tipo I (rechazos incorrectos (falsos positivos) de las hipótesis nulas) a medida que el número de comparaciones a realizar

aumenta.

Es por ello que se utilizó la corrección de Bonferroni, que divide el nivel de significancia α entre el número de comparaciones dos a dos realizadas, y asegura que la probabilidad de obtener al menos un falso positivo entre todas las comparaciones sea menor o igual que α , el valor de la significancia [16]. Así, esta corrección se aplica como sigue:

$$\alpha_{\text{corregido}} = \frac{\alpha}{N} \quad (3.11)$$

Con

- $\alpha_{\text{corregido}}$ la significancia corregida.
- α el valor original de la significancia.
- N el número de comparaciones dos a dos realizadas.

Estos test post-hoc nos sirvieron como complemento a los análisis de la varianza realizados en los distintos periodos, obteniendo así más detalles sobre lo que sucede en un periodo concreto.

Aunque existen otros test *post-hoc* que podrían obtener resultados distintos, se decidió realizar el mencionado, ya que los grupos no eran balanceados.

3.2.4. Análisis cluster jerárquico

Se conoce como *análisis cluster* al conjunto de técnicas estadísticas multivariantes pertenecientes a la minería de datos que, a partir de un conjunto de n variables X_j ($j=1, 2.. n$), tratan de construir un número k de grupos o conglomerados de observaciones lo más homogéneos posible [17]. Dado que no se conocen los grupos a priori, el análisis cluster puede clasificarse dentro del conjunto de métodos de aprendizaje no supervisado.

En el caso del *Hierarchical Clustering* (o agrupación jerárquica), la creación de los *clusters* mantiene un orden y una jerarquía concreta, atendiendo a un criterio que depende del método de construcción elegido. Estos algoritmos de clasificación se agrupan en 2 categorías: aglomerativo y divisivo, siendo el de tipo aglomerativo el utilizado durante el estudio.

Análisis cluster aglomerativo

En este tipo de clasificación, se parte inicialmente de un número de clusters igual al número de observaciones (cada observación se asigna a su propio cluster), se calculan las similitudes o distancias entre cada uno de los grupos, y se juntan los dos clusters con menor distancia o mayor similitud, repitiéndose este proceso hasta conseguir el número de grupos estipulado (figura 3.11).

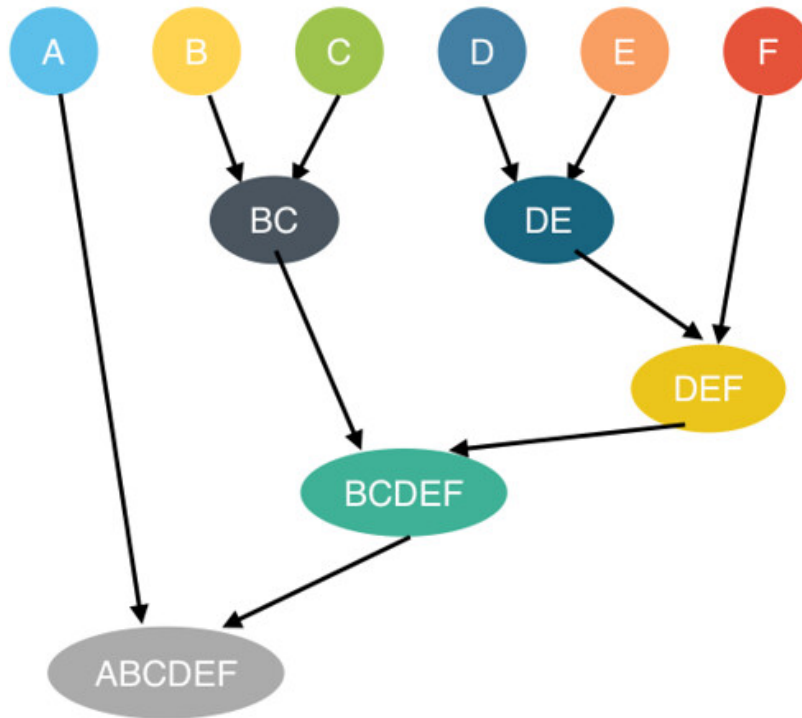


Figura 3.11: Análisis cluster aglomerativo

Este fue el método utilizado para construir la primera clasificación del proyecto, que sirvió como base para la estimación de datos ausentes.

Medidas de distancia y similitud

Para construir la agrupación, será necesario buscar una medida de distancia o similitud con la que sea posible hacer una comparación entre las observaciones del estudio. Se conoce como *distancia* o *disimilitud* entre 2 individuos i y j a una medida (habitualmente denotada como $d(i,j)$ que mide el grado de desemejanza entre estos individuos, en función de ciertos atributos o características medibles. Este valor siempre será positivo y, a mayor valor de la distancia, mayor será la disparidad entre las observaciones.

Las propiedades que deben verificar estas distancias son las siguientes:

- $d(i,j) \geq 0$ (propiedad de no negatividad).
- $d(i,j) = 0$ (la distancia de una observación a sí misma es cero).
- $d(i,j) = d(j,i)$ (propiedad de simetría)

Así, la distancia utilizada para el desarrollo de este trabajo es la conocida *Distancia Euclídea*.

Distancia Euclídea

La definición de distancia euclídea coincide con el concepto más común de distancia que se usa en el día a día Estadístico. Así, la distancia euclídea entre 2 puntos de N dimensiones A y B se ajusta a la siguiente expresión:

$$d(A, B) = \sqrt{\sum_{i=1}^N (b_i - a_i)^2} = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \dots + (b_N - a_N)^2} \quad (3.12)$$

Métodos de agrupación jerárquica

Los métodos de agrupación jerárquica determinan el criterio con el cual se van formando los grupos a medida que se itera en el algoritmo de clasificación jerárquica. En el presente estudio se ha utilizado el método de *Ward* [18], que se basa en la maximización de la homogeneidad dentro de cada conglomerado, en cuanto a la minimización de la suma de cuadrados aportada en cada caso.

Los pasos a seguir en cada iteración son los que siguen:

1. Se realiza el cálculo de los centroides de los grupos formados en dicha iteración, encontrando la media de de todas las variables dentro del grupo o conglomerado.
2. Se calcula la distancia euclídea entre el centroide y todas las observaciones del grupo, y se hace la suma total de estas distancias.
3. Se unen o agrupan los 2 conglomerados que generen un menor aumento en la suma de cuadrados de cada conglomerado.

3.2.5. Series temporales

Es necesario conocer el concepto de serie temporal para comprender la estructura y el contexto del problema. Así, se conoce como serie temporal univariante a una sucesión de observaciones medidas sobre una variable, ordenadas cronológicamente en intervalos de tiempo regulares. La periodicidad de estas series puede ser diaria, semanal, mensual, trimestral... dependiendo de la frecuencia con la que se hayan hecho las diferentes mediciones. En nuestro caso, esta periodicidad será trimestral.

Las series temporales son muy utilizadas en campos como la economía, la climatología, la biología o la medicina, entre muchos otros, y su correcto análisis y estudio es necesario para conseguir predicciones fiables. Así, en el presente *TFG* se construyó una serie temporal para cada una de las zonas de vigilancia consideradas, que fueron posteriormente clasificadas usando el método de la Deformación Dinámica del Tiempo.

Deformación Dinámica del Tiempo

Dentro del estudio y análisis de series temporales, existe un algoritmo destinado a medir la similitud entre dos secuencias de tiempo, que permite mejorar los resultados cuando existen desfases temporales [19]. Este es el conocido como *Dynamic Time Warping (DTW)* o Deformación Dinámica del Tiempo.

Puesto que no se conoce el resultado antes de hacer inferencias sobre los datos, podemos clasificar al algoritmo dentro de los métodos de aprendizaje no supervisados.

Las aplicaciones actuales de este algoritmo son variadas, pasando por el análisis de secuencias de vídeo o audio, clasificación de patrones faciales, y muchas otras aplicaciones relacionadas con la minería de datos [19]. Fue el método utilizado para la construcción final de los grupos de zonas de vigilancia, ya que permite considerar cada una de las series temporales como una observación única.

El algoritmo *DTW* busca el punto más cercano entre cada punto de ambas series, lo que permitiría diferenciar formas similares que hayan sufrido un desajuste en el tiempo. Así, se siguen los dos pasos que se describen a continuación.

1. Cálculo de la matriz de costes

El cálculo de la matriz de coste $D \in \mathbb{R}^{(n+1) \times (m+1)}$ se realiza como sigue. Sean X e Y dos series temporales univariantes ($X = \{x_1, x_2 \dots x_n\}$ e $Y = \{y_1, y_2 \dots y_m\}$) con n y m observaciones, respectivamente. Inicialmente, partimos de la siguiente condición:

- Para todo $i \in 1, 2 \dots n$, $D(i, 0) = \infty$
- Para todo $j \in 1, 2 \dots m$, $D(0, j) = \infty$
- $D(0, 0) = 0$

Así, el coste $D(i, j)$ entre 2 observaciones i y j viene dado por la siguiente expresión:

$$D(i, j) = \text{dist}(x_i, y_j) + \min \begin{cases} D(i-1, j) \\ D(i, j-1) \\ D(i-1, j-1) \end{cases} \quad (3.13)$$

Con $\text{dist}(x_i, y_j)$ la métrica de distancia escogida para el problema en cuestión entre las observaciones x_i e y_j .

2. Búsqueda del “camino óptimo”

Ya con la matriz de coste D construida, el siguiente paso será buscar la distancia más corta entre los puntos, logrando así lo que se conoce como el “camino óptimo”.

Para ello, partiendo de $D(n, m)$, se recorre la matriz hasta $D(0, 0)$, buscando en cada iteración el mínimo entre $D(i-1, j)$, $D(i, j-1)$ y $D(i-1, j-1)$, lo que da como resultado este camino, y la alineación entre los puntos correspondiente.

Se puede apreciar más visualmente la diferencia entre la distancia euclídea y el *DTW* en la Figura 3.12:

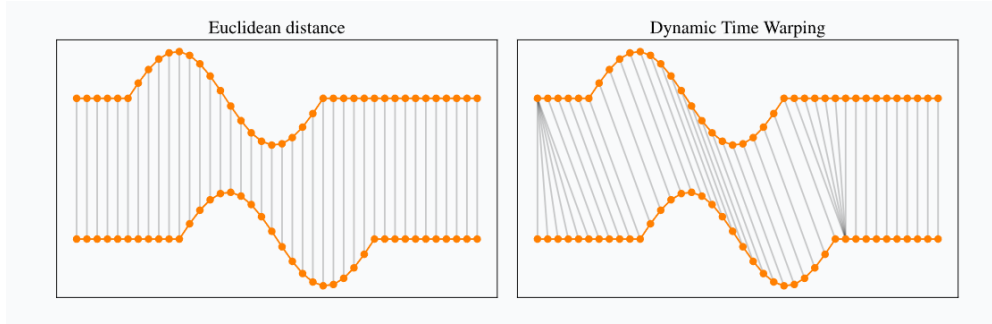


Figura 3.12: Deformación Dinámica del Tiempo

La diferencia principal es la alineación entre los puntos: mientras que el método euclídeo relaciona los puntos uno a uno, esta métrica optimiza el camino y busca las similitudes entre ambas secuencias de una manera más óptima.

Función de Autocorrelación o *ACF* y Correlogramas

La relación de variación conjunta entre dos variables viene dada por un estadístico conocido como *Covarianza*. Y, dado que en el ámbito de las series temporales nuestras observaciones vienen dadas por una misma variable medida en diferentes instantes de tiempo, usaremos el término *Autocovarianza* para denotar la covarianza entre observaciones dentro de la serie.

La expresión de la autocovarianza entre dos observaciones separadas k instantes de tiempo X_t y X_{t+k} se reflejará de la forma que sigue:

$$\gamma_k = E[(X_t - \mu)(X_{t+k} - \mu)] \quad (3.14)$$

Siendo μ la media muestral calculada sobre el conjunto de observaciones de la serie temporal [20]. Una de las herramientas más usadas para estudiar la relación entre los elementos de la serie temporal es la conocida como *función de autocorrelación*.

Se conoce como Función de Autocorrelación o *ACF* (*Auto Correlation Function*) la función matemática que nos permite determinar la dependencia entre una observación con otra obtenida k periodos antes [21].

Esta autocorrelación para un retardo de orden k viene dada por la siguiente expresión:

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^T [(x_t - \hat{\mu})(x_{t-k} - \hat{\mu})]}{\sum_{t=1}^T (x_t - \hat{\mu})^2} = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \quad (3.15)$$

donde:

- $\hat{\rho}_k$ representa la autocorrelación de orden k .

- T indica el número máximo de retardos.
- x_t representa la observación en el instante t de la serie temporal.
- $\hat{\mu}$ es la media de las observaciones en la serie temporal.
- $\hat{\gamma}_k$ es la autovarianza de orden k .
- $\hat{\gamma}_0$ equivaldrá a la varianza de la observación x_t con respecto de la media poblacional.

Su estudio y análisis es muy usado dentro del ámbito de los datos funcionales, ya que permite encontrar dependencias que revelarán si existe periodicidad en la serie, así como su valor. Habitualmente, esta función de autocorrelación se estudia gráficamente por medio de un *correlograma*.

Un correlograma (o autocorrelograma) en el ámbito de las series temporales representará de forma gráfica el valor de las autocorrelaciones de esta serie en función del retraso con el cual se ha calculado esta autocorrelación.

4. Resultados y discusión

4.1. Exploración y racionalización espacio-temporal inicial de la base de datos.

Con el fin de tener una “fotografía” que explique y resuma la evolución en las diferentes regiones de la comunidad, se construyó una matriz de información que contuviese el índice de actividad medio, según el momento temporal y la región correspondiente.

Se utilizó la información disponible en *Vías de dispersión*, en cuanto a hábitat representativo de la evolución natural del topillo. Asimismo, se utilizó la información disponible para *Cultivos* en cuanto a su abundancia en el interior de las parcelas de cultivo y, por tanto, del concepto de plaga en cuanto a afección potencial en las mismas.

Ambos tipos de hábitat de trabajo confluirán en secciones posteriores, y servirán como punto de comparación de su comportamiento en diferentes situaciones.

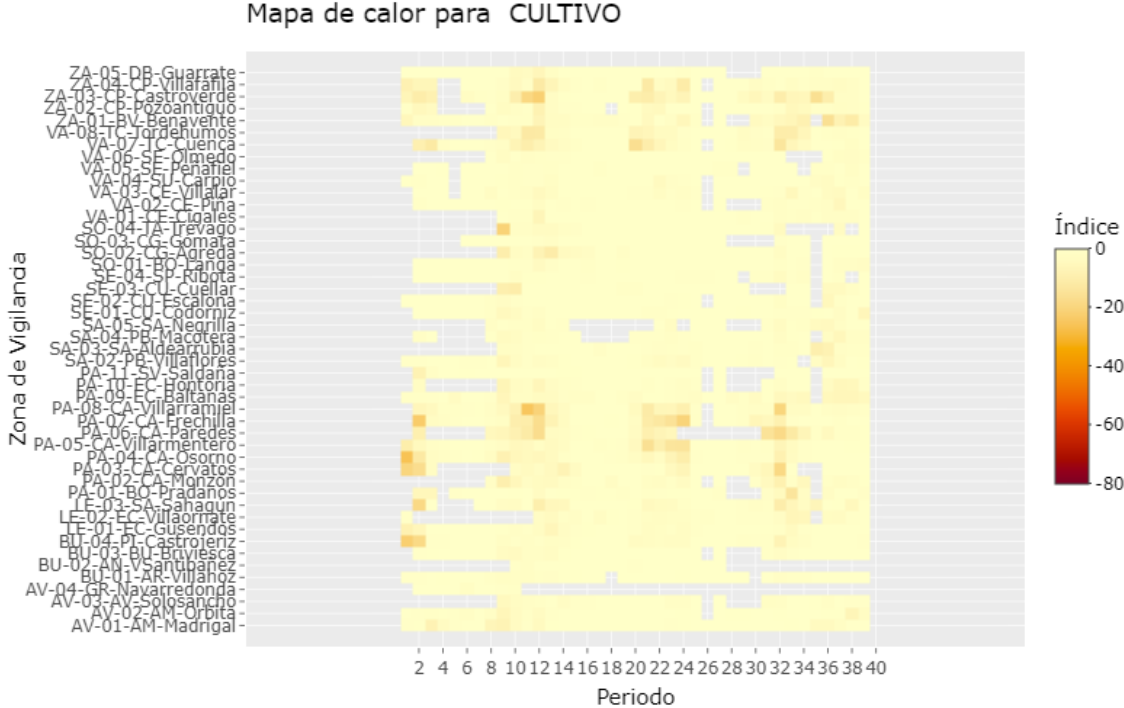


Figura 4.13: Mapa de calor para hábitat Cultivos

Mapa de calor para VÍAS DE DISPERSIÓN

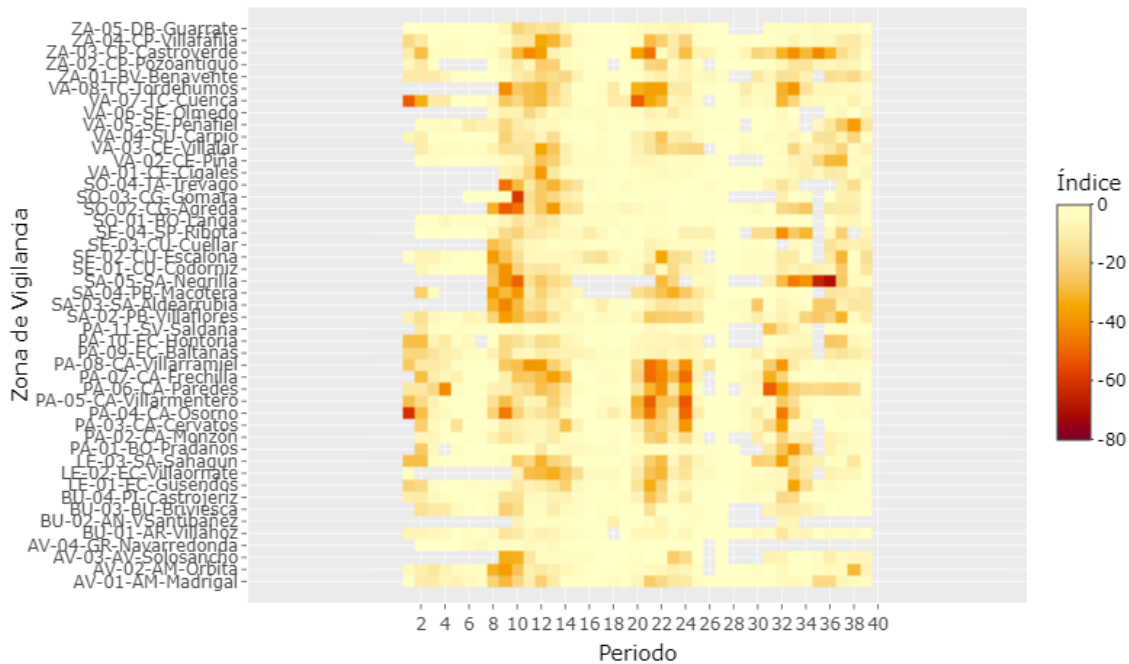


Figura 4.14: Mapa de calor para hábitat Vías de dispersión

4.1.1. Revisión de los periodos de estudio

Se detectó que la ausencia de información en los periodos iniciales (del 1 al 7) era notable (tabla 4.4), con un 41.95 % de valores ausentes en el caso de los cultivos, y un 31.91 % en el caso de las vías de dispersión. Por lo tanto, se decidió no considerar los 7 primeros periodos para llevar a cabo la clasificación, por el riesgo de que acarreasen ruido en las posibles clasificaciones posteriores. De esta forma, el nuevo conjunto de datos comienza en el periodo correspondiente al verano de 2013 (periodo 8) y contando con un total de 32 periodos de estudio.

Tabla 4.4: Porcentajes de valores ausentes por hábitat en los 7 primeros periodos

	% No ausentes	% Ausentes
Cultivos	58.05	41.95
Vías de dispersión	68.09	31.91

Esta ausencia de información (representada de color gris) también se puede observar en la figura 4.13 para los cultivos y en la figura 4.14 para las vías de dispersión.

4.1.2. Revisión de las zonas de vigilancia

Otro aspecto que se valoró antes de comenzar con la clasificación fue la posible eliminación de determinadas zonas de vigilancia para las cuales la información disponible era tan escasa, que incluirlas en el análisis podría resultar perjudicial. Esto parece ocurrir en las zonas de vigilancia AV-04, BU-02, SA-04 y SA-05.

La tabla 4.5 refleja el porcentaje en estas zonas de vigilancia de valores ausentes para cultivo.

Tabla 4.5: *Porcentaje de valores ausentes para las zonas de vigilancia eliminadas en cultivos*

	% No ausentes	% Ausentes
AV-04-GR-Navarredonda	23.08	76.92
BU-02-AN-VSantibañez	46.15	53.85
SA-04-PB-Macotera	74.36	25.64
SA-05-SA-Negrilla	53.85	46.15

Se puede apreciar como estas 4 zonas de vigilancia contienen una cantidad muy elevada de datos ausentes. Y esto mismo ocurre para el caso de las vías de dispersión (tabla 4.6)

Tabla 4.6: *Porcentaje de valores ausentes para las zonas de vigilancia eliminadas en vías de dispersión*

	% No ausentes	% Ausentes
AV-04-GR-Navarredonda	64.10	35.90
BU-02-AN-VSantibañez	51.28	48.72
SA-05-SA-Negrilla	53.85	46.15
SA-04-PB-Macotera	74.36	25.64

Por lo tanto, esta fase de depuración de los datos se decidió eliminar estas cuatro zonas de vigilancia, dejando un total de 43 zonas de para proseguir con el estudio.

De esta manera, el estudio prosigue considerando un total de 43 zonas de vigilancia y 32 periodos, cuyos mapas de calor se reflejan en las figuras 4.15 en el caso de los cultivos y 4.16 para las vías de dispersión.

En función de estos mapas de calor, se aprecia que sigue existiendo un problema con los datos perdidos, tal y como se ve en la tabla 4.7, con un 7.74 % en el caso de los cultivos, y un 5.16 % en las vías de dispersión.

Tabla 4.7: *Porcentajes de valores ausentes para cultivos y vías de dispersión*

	% No ausentes	% Ausentes
Cultivos	92.66	7.34
V.D.	94.84	5.16

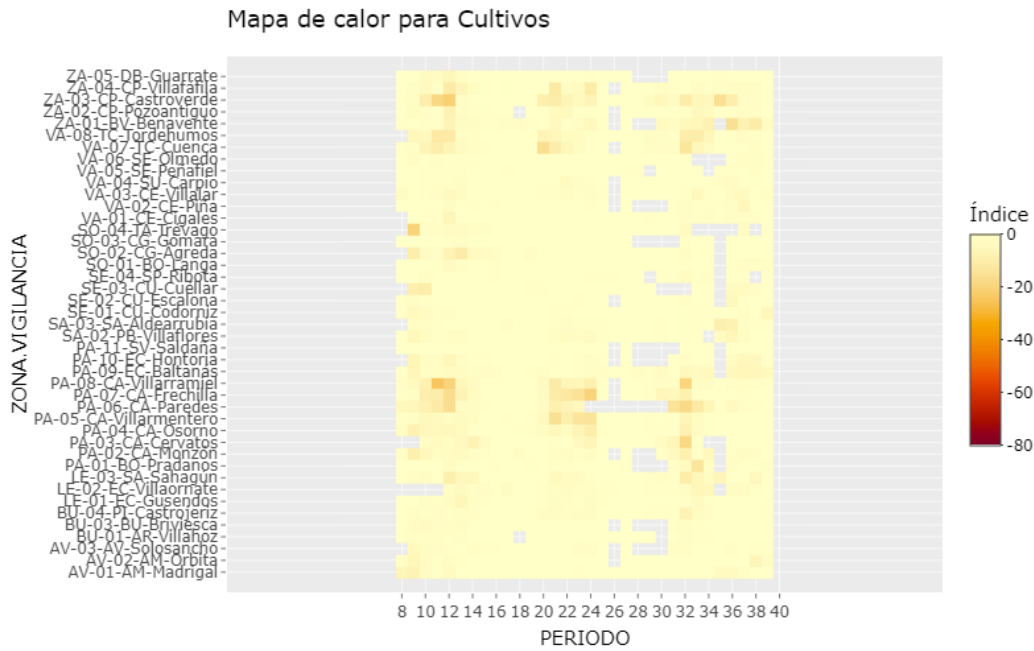


Figura 4.15: Mapa de calor para el hábitat Cultivos desde el periodo 8 (Verano de 2013)

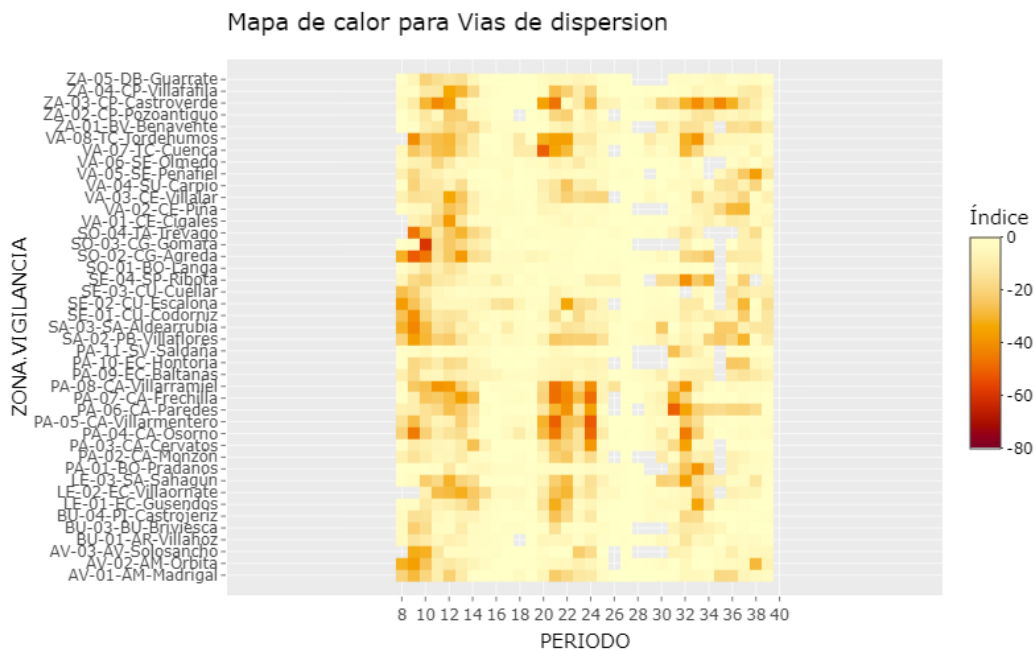


Figura 4.16: Mapa de calor para el hábitat Vías de dispersión desde el periodo 8 (Verano de 2013)

4.2. Detección inicial de expresiones diferenciales de la abundancia a nivel espacio-temporal.

Para comprobar en qué periodos existían diferencias significativas entre zonas de vigilancia, se realizó un análisis ANOVA en cada uno de los periodos, tanto para ins-

pecciones en cultivos, como en vías de dispersión.

Se representaron en verde aquellos periodos para los que sí que se detectan diferencias significativas a nivel 0,05 (es decir, los periodos en los que el p – *valor* del test sea menor a 0,05) y en rojo, aquellos para los que no existen estas diferencias.

La significancia del test ANOVA realizado para ambos hábitat en los distintos periodos aparece reflejada en la tabla 4.8.

Tabla 4.8: Aceptación o rechazo de la hipótesis de igualdad de medias (verde: rechazo de la hipótesis de igualdad de medias; rojo: no rechazo de esta hipótesis)

Periodo	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39			
Cultivos	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	
V.D.	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█

Tal y como se aprecia en la tabla 4.8, en casi todos los periodos de estudio se encontraron diferencias significativas, a excepción de los periodos 15, 17, 19, 26 y 35, en los que no se detectan estas diferencias entre periodos (en el caso de los cultivos). En cuanto a las vías de dispersión, el procedimiento revela que en el periodo 35 (correspondiente a la primavera de 2020) no se detectan diferencias significativas, al igual que en el periodo 26 (invierno de 2018), coincidiendo estos resultados para estos dos periodos en ambos hábitat de estudio. Para el resto, sí que se detectan estas diferencias.

Ciertos de los resultados obtenidos son esperables, pues el periodo 35, por ejemplo, coincide con el inicio del conocido “confinamiento por COVID-19”, lo que produjo que muchas inspecciones no se pudieran realizar.

4.3. Agrupación espacial en función de la expresión diferencial de la evolución de la abundancia mediante el uso de métodos de aprendizaje no supervisado.

La siguiente etapa del *TFG* consistió en aplicar diversas técnicas de análisis cluster jerárquico para encontrar zonas de vigilancia cuya evolución a lo largo del periodo de estudio fuese similar. Esta clasificación se realizó tanto para el hábitat *Cultivos* como para *Vías de dispersión*.

4.3.1. Agrupación inicial de las zonas de vigilancia sobre la matriz de racionalización espacio-temporal

Los pasos realizados para la obtención de los grupos de zonas de vigilancia fueron los que siguen.

Construcción de la matriz de distancias media

Para ello, se decidió optar por construir una matriz de distancias euclídeas entre zonas de vigilancia para cada periodo estudiado, obteniendo así tantas matrices como periodos de estudio se incluyesen en el análisis. Una vez obtenidas estas matrices, se calculó la *matriz de distancias media*, haciendo la suma de todas estas matrices, y dividiendo esta suma total por el número de veces que esa distancia no había resultado nula. Así, los valores de las distancias medias entre dos individuos i y j se calcularon como sigue:

$$d_{\mu}(i, j) = \sum_{k=1}^P \frac{d_k(i, j)}{n_{i,j}} \quad (4.16)$$

con:

- $d_{\mu}(i, j)$ la distancia media entre las observaciones i y j .
- P el número total de periodos incluidos en el análisis.
- $d_k(i, j)$ la distancia k -ésima entre los individuos i y j .
- $n_{i,j}$ el número de veces en las que la distancia obtenida entre los individuos i y j no es nula.

Ya con la matriz de distancias calculada, el siguiente paso consistió en plantear el análisis cluster aglomerativo, haciendo uso de la función *hclust* de *R*. Para llevar a cabo la clasificación, fue necesario decidir el tipo de distancia a utilizar (en nuestro caso, la distancia euclídea), el método de enlace para formar los grupos (para el que se decidió utilizar el *método de Ward*) y el número de grupos o clusters deseados.

El dendrograma correspondiente a la clasificación para el hábitat cultivos se refleja en la figura 4.17. De igual manera, la figura 4.18 refleja este dendrograma en el caso de las vías de dispersión.

DENDROGRAMA EN CULTIVOS

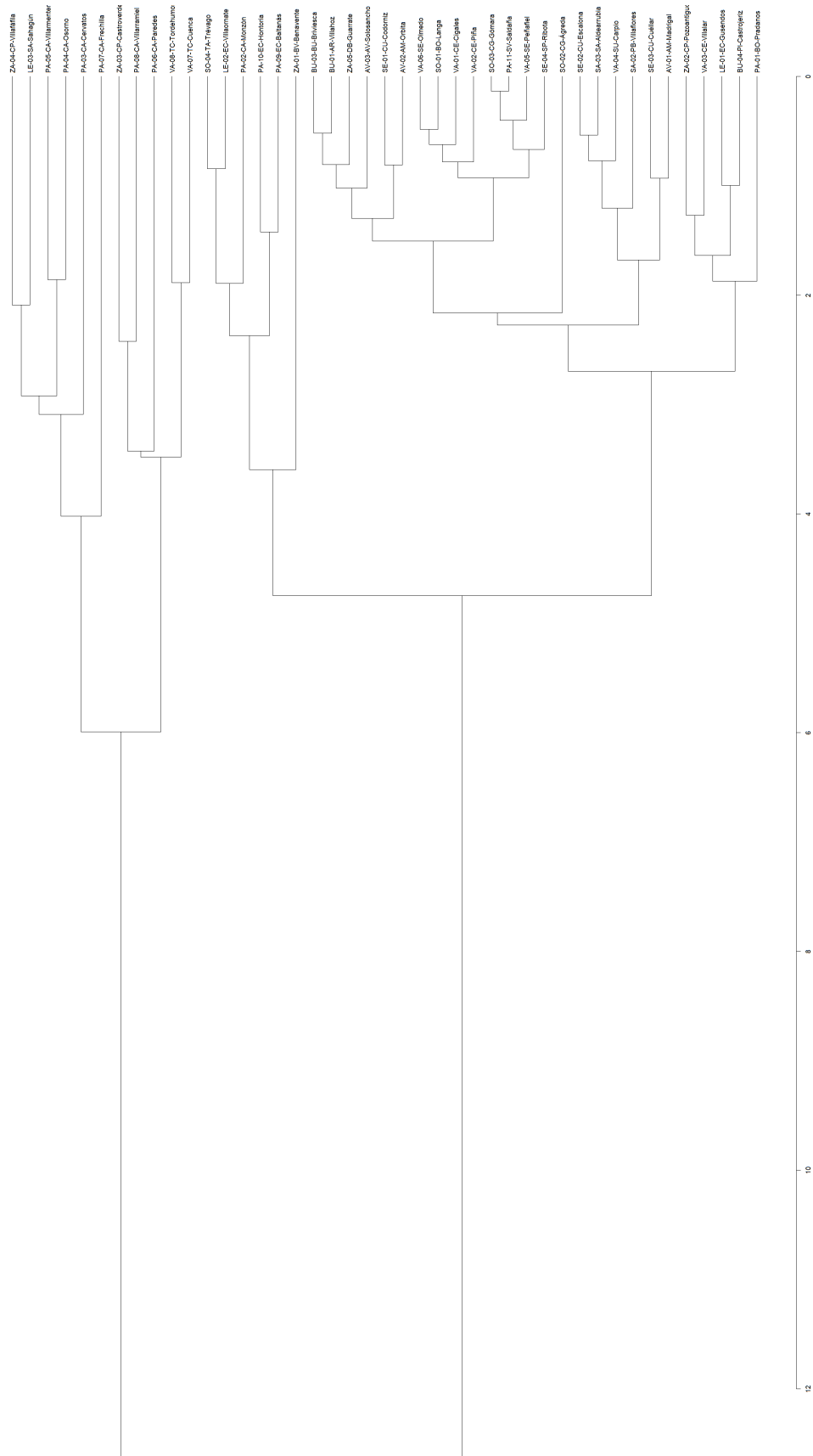


Figura 4.17: Dendrograma para hábitat Cultivos

DENDROGRAMA EN VIAS DE DISPERSION

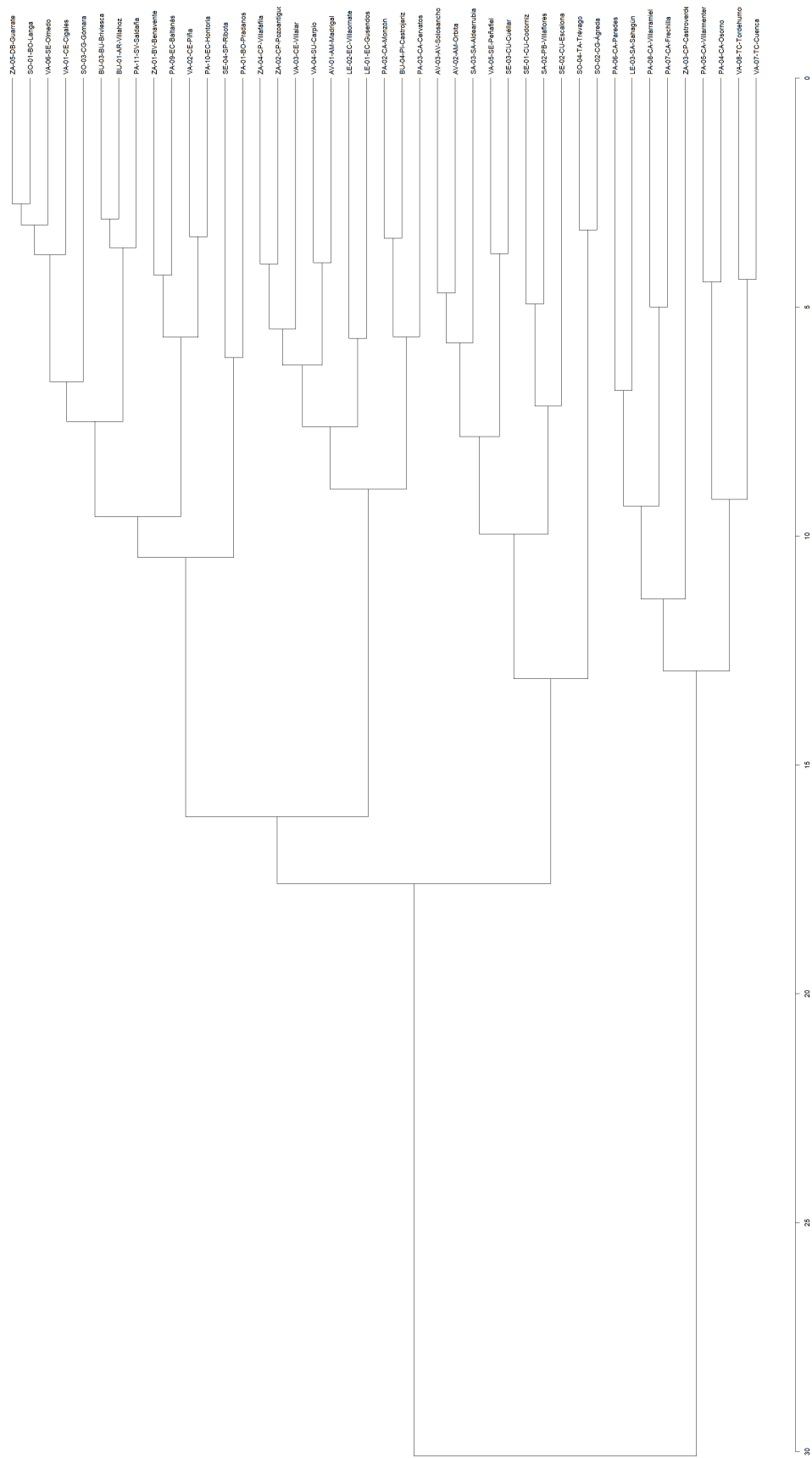


Figura 4.18: Dendrograma para hábitat. Vías de dispersión

Obtención del número óptimo de clusters

Para la obtención de un número óptimo de grupos, se recurrió al *método del codo*. Con este método, se calculan las sumas de cuadrados para un determinado rango de valores de k (siendo k el número de grupos construidos con el análisis), se representan estas sumas de cuadrados frente a los valores de k , y se estudia el comportamiento de la gráfica buscando una forma de codo, es decir, cambios bruscos en la pendiente que indiquen que aumentar el número de grupos no cambia significativamente la variabilidad explicada.

En el caso de los cultivos, esta representación se reflejan en la figura 4.19.

Método del codo para las sumas de cuadrados

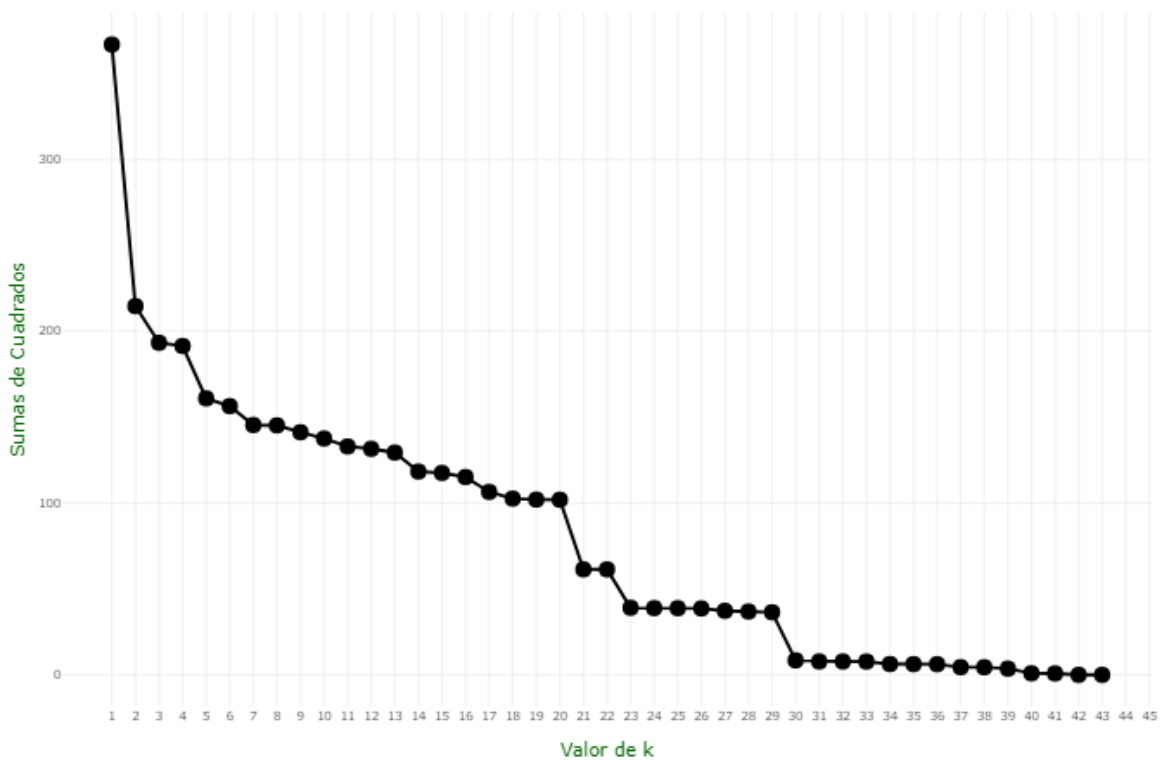


Figura 4.19: Gráfico de codo para hábitat Cultivos

Se detectan cambios bruscos en la pendiente para los valores de k 3, 5, 21, 23 y 30, ya que para el resto de los periodos, este cambio no es tan pronunciado. Por otro lado, la figura 4.20

Método del codo para las sumas de cuadrados

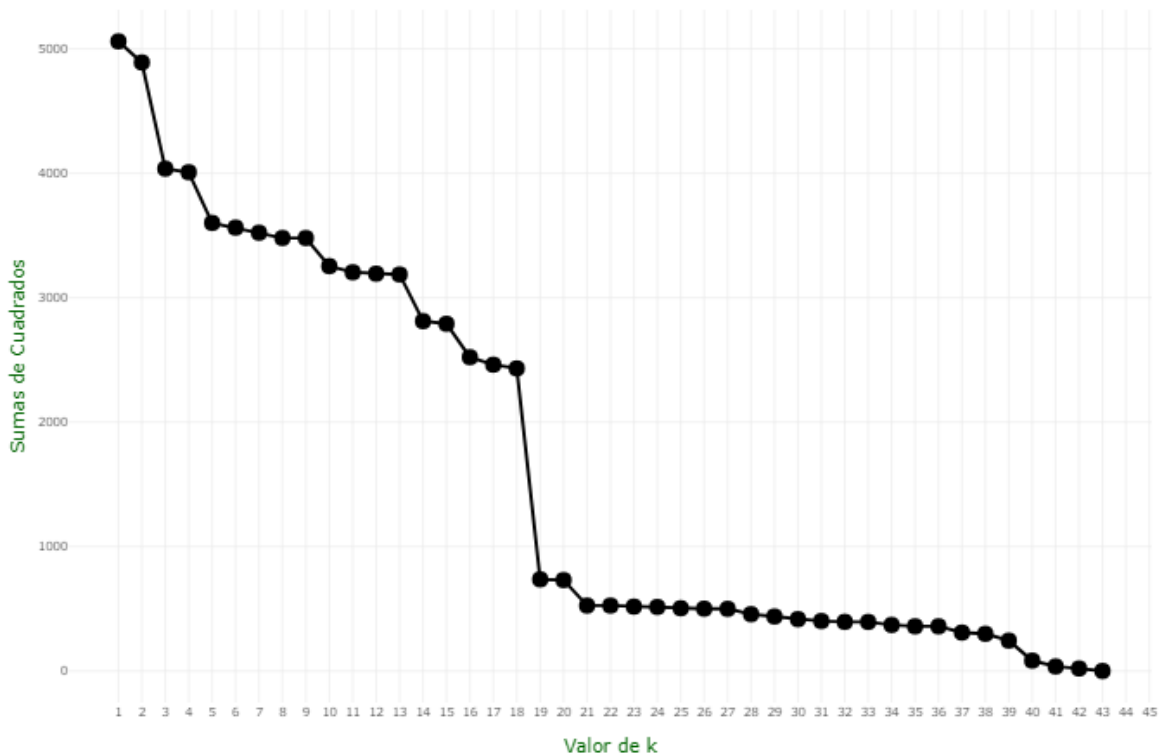


Figura 4.20: Gráfico de codo para hábitat Vías de dispersión

En este caso, los cambios pronunciado se detectan para los valores 3, 5, 14, 16 y 19 de k , siendo este último valor para el cual se produce un cambio muy grande en esta pendiente.

Comprobados ambos hábitat y estudiadas las gráficas, se decidió fijar el número de clusters a construir en 5, ya que ambas gráficas revelaban saltos grandes para este valor, y un valor más alto del número de grupos entorpecería el posterior análisis de la evolución.

Estudio y visualización de los grupos obtenidos

Fijado el número de grupos, se obtuvo una división de las zonas de vigilancia que se presentarán a continuación. Estos grupos se visualizarán en forma de mapa, además de exponer la composición de cada grupo formado. La composición detallada de estos grupos se presenta en la tabla 4.9.

Se puede apreciar un gran grupo, compuesto por 26 zonas de vigilancia, un grupo con 6, otros 2 con 5 y finalmente un grupo compuesto por una única zona de vigilancia (correspondiente a la zona de Frechilla). La distribución espacial de estos grupos se expone en la figura 4.21.

Tabla 4.9: Composición de los grupos iniciales en Cultivos

Zona de Vigilancia	Grupo
AV-01-AM-Madrigal	1
AV-02-AM-Orbita	1
AV-03-AV-Solosancho	1
BU-01-AR-Villahoz	1
BU-03-BU-Briviesca	1
BU-04-PI-Castrojeriz	1
LE-01-EC-Gusendos	1
PA-01-BO-Pradanos	1
PA-11-SV-Saldaña	1
SA-02-PB-Villaflores	1
SA-03-SA-Aldearrubia	1
SE-01-CU-Codorniz	1
SE-02-CU-Escalona	1
SE-03-CU-Cuéllar	1
SE-04-SP-Ribota	1
SO-01-BO-Langa	1
SO-02-CG-Ágreda	1
SO-03-CG-Gómara	1
VA-01-CE-Cigales	1
VA-02-CE-Piña	1
VA-03-CE-Villalar	1
VA-04-SU-Carpio	1
VA-05-SE-Peñafiel	1
VA-06-SE-Olmedo	1
ZA-02-CP-Pozoantiguo	1
ZA-05-DB-Guarrate	1
LE-02-EC-Villaornate	2
PA-02-CA-Monzón	2
PA-09-EC-Baltanás	2
PA-10-EC-Hontoria	2
SO-04-TA-Trévago	2
ZA-01-BV-Benavente	2
LE-03-SA-Sahagún	3
PA-03-CA-Cervatos	3
PA-04-CA-Osorno	3
PA-05-CA-Villarmentero	3
ZA-04-CP-Villafáfila	3
PA-06-CA-Paredes	4
PA-08-CA-Villarramiel	4
VA-07-TC-Cuenca	4
VA-08-TC-Tordehumos	4
ZA-03-CP-Castroverde	4
PA-07-CA-Frechilla	5

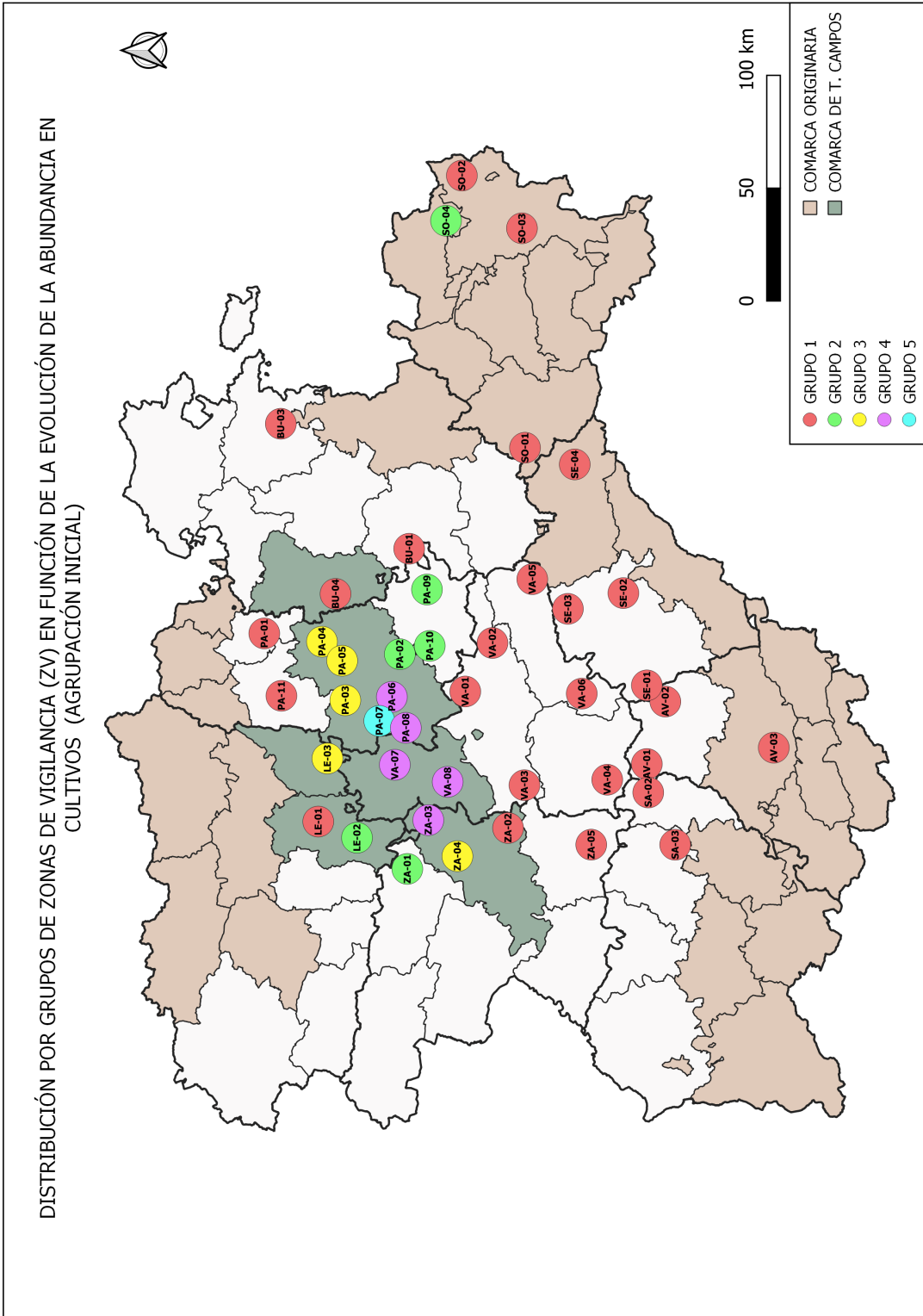


Figura 4.21: Agrupación inicial de las zonas de vigilancia para hábitat Cultivos

Tabla 4.10: Composición de los grupos iniciales en Vías de dispersión

Zona de Vigilancia	Grupo
AV-01-AM-Madrigal	1
BU-04-PI-Castrojeriz	1
LE-01-EC-Gusendos	1
LE-02-EC-Villaornate	1
PA-02-CA-Monzón	1
PA-03-CA-Cervatos	1
VA-03-CE-Villalar	1
VA-04-SU-Carpio	1
ZA-02-CP-Pozoantiguo	1
ZA-04-CP-Villafáfila	1
AV-02-AM-Orbita	2
AV-03-AV-Solosancho	2
SA-02-PB-Villaflores	2
SA-03-SA-Aldearrubia	2
SE-01-CU-Codorniz	2
SE-02-CU-Escalona	2
SE-03-CU-Cuéllar	2
VA-05-SE-Peñafilel	2
BU-01-AR-Villahoz	3
BU-03-BU-Briviesca	3
PA-01-BO-Pradanos	3
PA-09-EC-Baltanás	3
PA-10-EC-Hontoria	3
PA-11-SV-Saldaña	3
SE-04-SP-Ribota	3
SO-01-BO-Langa	3
SO-03-CG-Gómara	3
VA-01-CE-Cigales	3
VA-02-CE-Piña	3
VA-06-SE-Olmedo	3
ZA-01-BV-Benavente	3
ZA-05-DB-Guarrate	3
LE-03-SA-Sahagún	4
PA-04-CA-Osorno	4
PA-05-CA-Villarmentero	4
PA-06-CA-Paredes	4
PA-07-CA-Frechilla	4
PA-08-CA-Villarramiel	4
VA-07-TC-Cuenca	4
VA-08-TC-Tordehumos	4
ZA-03-CP-Castroverde	4
SO-02-CG-Ágreda	5
SO-04-TA-Trévago	5

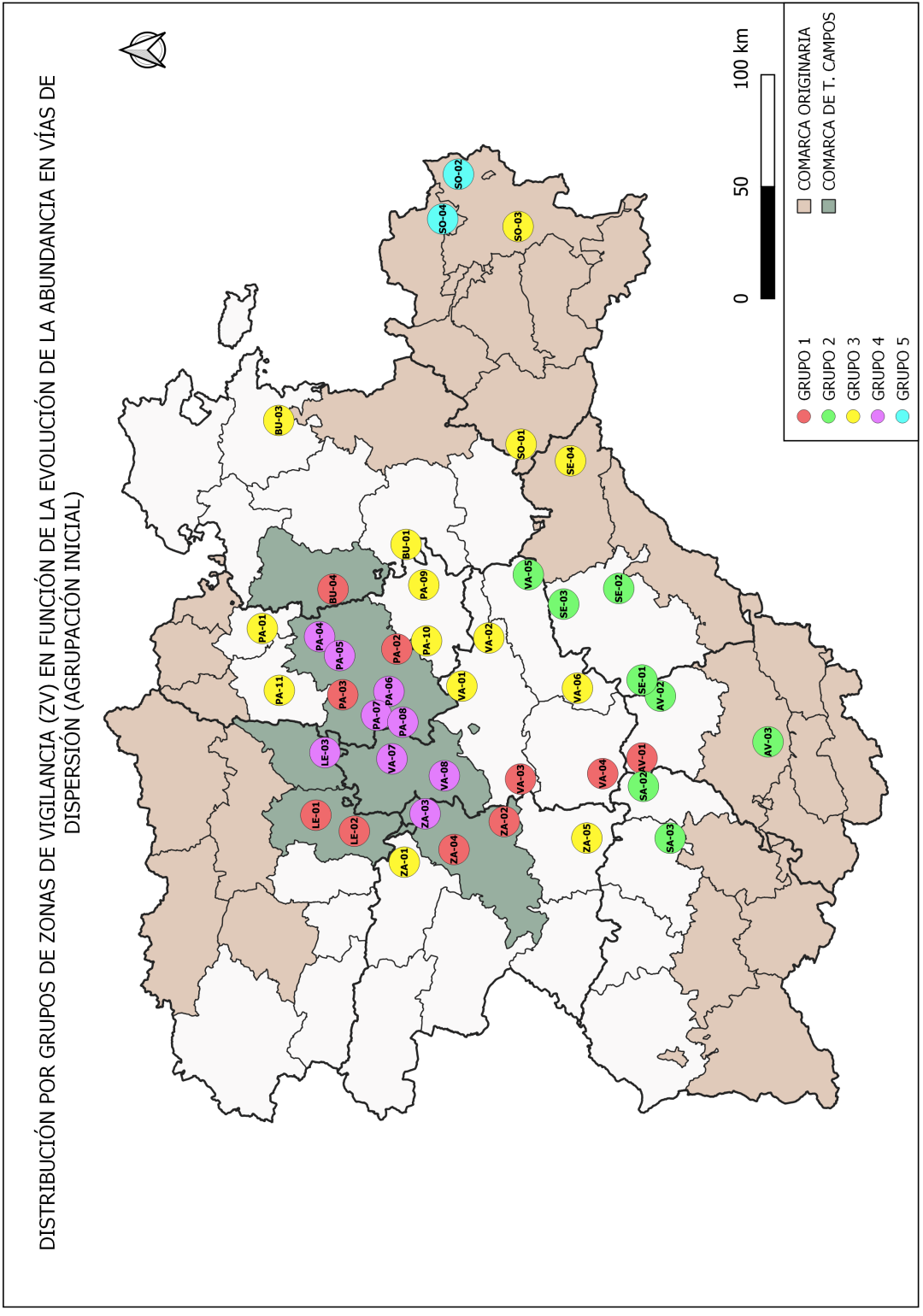


Figura 4.22: Distribución inicial de las zonas de vigilancia para hábitat Vías de dispersión

El grupo visto anteriormente con 26 observaciones ocuparía la mayor parte del territorio sur, además de algunas zonas situadas más al norte. Por otro lado, el resto de grupos se repartirán mayoritariamente en la zona conocida por los expertos en la materia como “Tierra de campos”, zona conflictiva y con mucha actividad conocida.

Para el caso de las vías de dispersión, esta composición de los grupos cambiará. En este caso, se contará con grupos más repartidos, teniendo 5 grupos con 10 zonas de vigilancia en el grupo 1, 8 en el grupo 2, 14 en el 3, 9 en el 4 y, por último, 2 en el 5. La composición de estos grupos se presenta en la tabla 4.10. Y se presenta esta composición en forma de mapa en la figura 4.22.

4.3.2. Estimación de datos perdidos a partir del clustering inicial

El uso principal de la clasificación presentada en la sección anterior fue la obtención de matrices de datos más completas, para poder así estimar los valores perdidos y poder realizar una clasificación con datos más completos.

Para solventarlo, se procedió como sigue: para cada zona de vigilancia, se estimaron los datos de aquellos periodos con datos perdidos a partir de el valor medio en dicho periodo del grupo al que esa zona de vigilancia estaba asignada.

Tras imputar estos valores ausentes, la información disponible se refleja en la figura 4.23 en el caso de los cultivos.

En este caso, se aprecia que sigue faltando información, tanto en el periodo 26 (Invierno de 2018) en la zona de Frechilla, como en el periodo 35 (Primavera de 2020) en las zonas de Villaornate (León), Monzón, Baltanás y Hontoria (Palencia), Trévago (Soria) y Benavente (Zamora).

En el caso de las vías de dispersión (figura 4.24) la cantidad de información ausente es menor, teniendo de nuevo para el periodo 35 información ausente en las zonas de Ágreda y Trévago, en la provincia de Soria.

Mapa de calor para Cultivos tras estimar

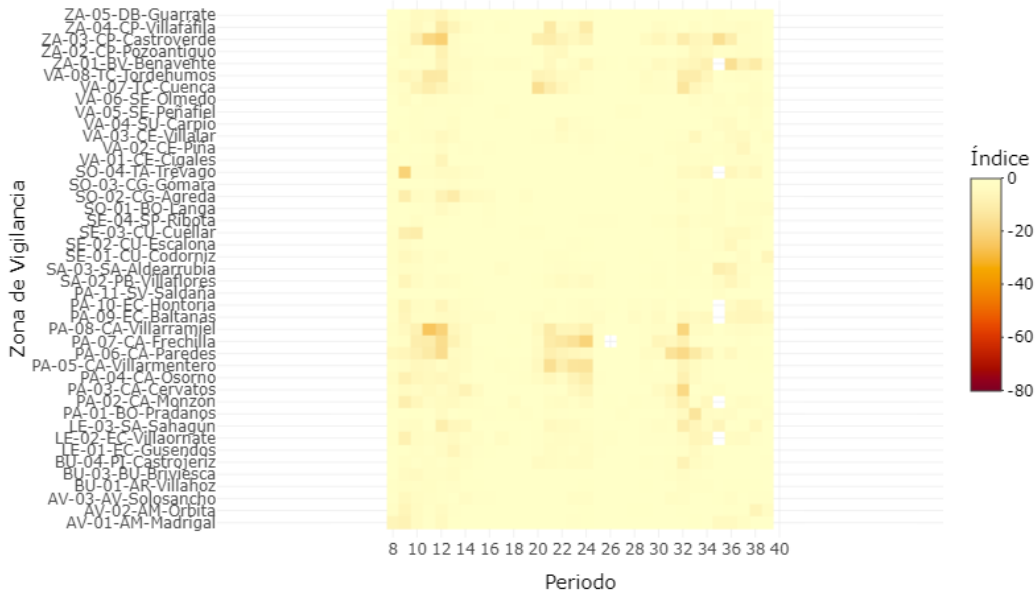


Figura 4.23: Mapa de calor tras la estimación para el hábitat Cultivos

Mapa de calor para Vías de Dispersión tras estimar

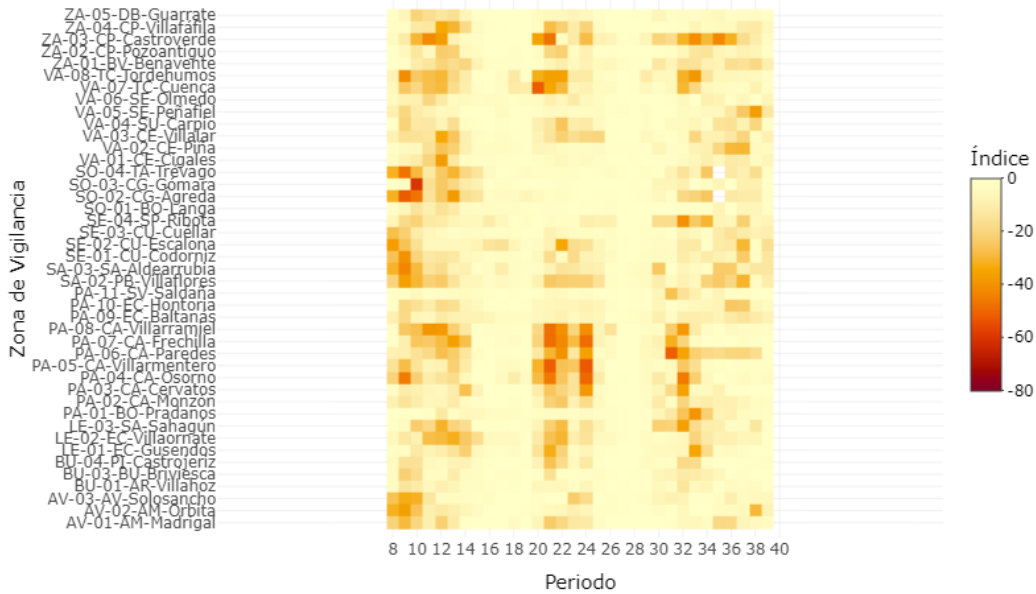


Figura 4.24: Mapa de calor tras la estimación para el hábitat Vías de dispersión

Como solución a esta nueva falta de información, y tras analizarlo con los profesionales especializados en la materia, se decidió interpolar estos valores ausentes en cada zona de vigilancia a partir de los valores de los índices en los periodos colindantes (anterior y posterior), de forma que el nuevo valor del índice de actividad I_{ij} en la zona i y el periodo j será:

$$I_{i,j} = \frac{I_{i,j-1} + I_{i,j+1}}{2} \tag{4.17}$$

Así, esta información completa se representará en las figuras 4.25 y 4.26 para cultivos y vías de dispersión, respectivamente.



Figura 4.25: Mapa de calor sin valores ausentes para el hábitat Cultivos

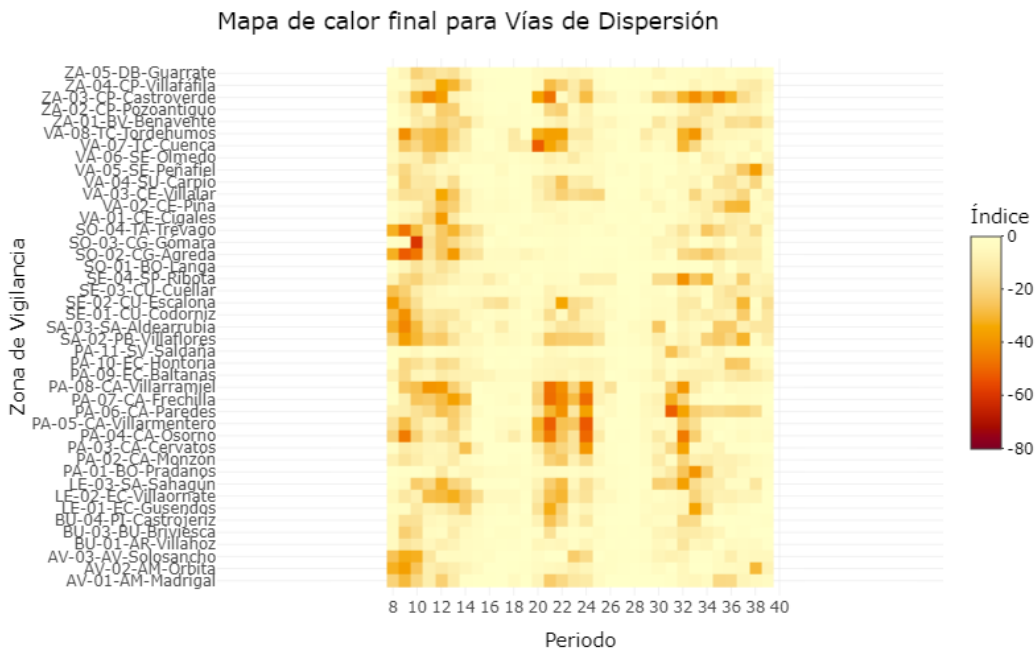


Figura 4.26: Mapa de calor sin valores ausentes para el hábitat Vías de dispersión

4.3.3. Agrupación final a partir del método Deformación Dinámica del Tiempo

Se aplicó el método de clasificación DTW para obtener la clasificación final de las zonas de vigilancia, tras haber estimado los valores ausentes a partir del método

descrito en la sección 4.3.2. En esta ocasión, se realizó una única clasificación para ambos hábitat. La figura 4.27 muestra el dendrograma obtenido tras la clasificación.

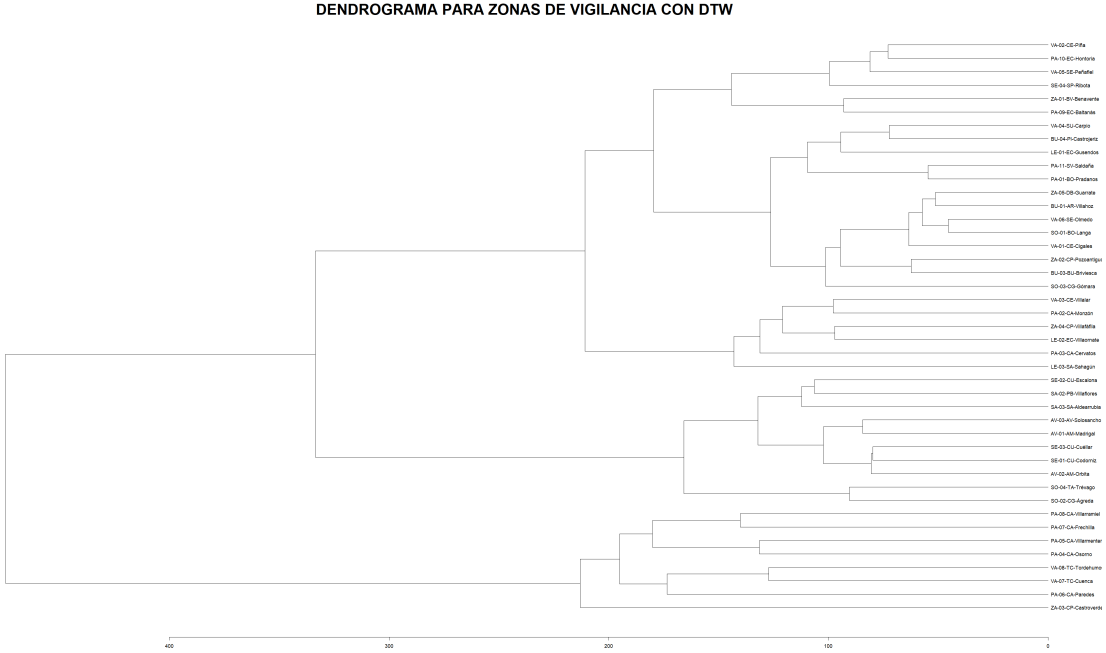


Figura 4.27: Dendrograma obtenido con la clasificación final

Para una mejor visualización del dendrograma, se aporta también en el anexo 6.3.

Haciendo uso del *método del codo* para seleccionar el número de grupos, se representaron las sumas de cuadrados frente a un rango de valores de k (figura 4.28). En función de ello, se observa que para un valor de $k = 3$, se produce el primer cambio de pendiente pronunciado.

El salto más pronunciado se produce para un valor de $k = 24$, produciendo un número de grupos excesivamente grande para facilitar una racionalización de la información. Por tanto, el número de grupos finalmente considerado fue de tres.

La composición de los grupos finales de zonas de vigilancia se refleja en la tabla 4.11. La distribución geográfica final se representa en la figura 4.29.

Método del codo para sumas de cuadrados

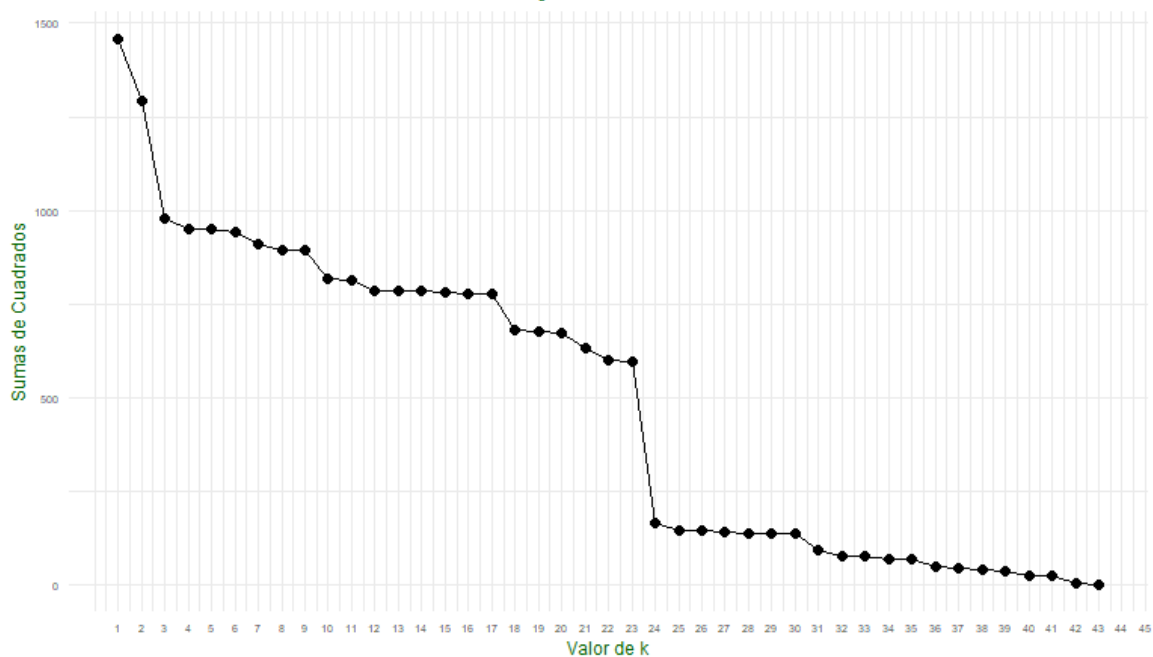


Figura 4.28: Gráfico de codo para la clasificación final

Tabla 4.11: Zonas de vigilancia por grupos

Zona de Vigilancia	Grupo
AV-01-AM-Madrigal	1
AV-02-AM-Orbita	1
AV-03-AV-Solosancho	1
SA-02-PB-Villaflores	1
SA-03-SA-Aldearrubia	1
SE-01-CU-Codorniz	1
SE-02-CU-Escalona	1
SE-03-CU-Cuéllar	1
SO-02-CG-Ágreda	1
SO-04-TA-Trévago	1
BU-01-AR-Villahoz	2
BU-03-BU-Briviesca	2
BU-04-PI-Castrojeriz	2
LE-01-EC-Gusendos	2
LE-02-EC-Villaornate	2
LE-03-SA-Sahagún	2
PA-01-BO-Pradanos	2
PA-02-CA-Monzón	2
PA-03-CA-Cervatos	2
PA-09-EC-Baltanás	2
PA-10-EC-Hontoria	2
PA-11-SV-Saldaña	2

Continúa en la página siguiente

Zona de Vigilancia	Grupo
SE-04-SP-Ribota	2
SO-01-BO-Langa	2
SO-03-CG-Gómara	2
VA-01-CE-Cigales	2
VA-02-CE-Piña	2
VA-03-CE-Villalar	2
VA-04-SU-Carpio	2
VA-05-SE-Peñafiel	2
VA-06-SE-Olmedo	2
ZA-01-BV-Benavente	2
ZA-02-CP-Pozoantiguo	2
ZA-04-CP-Villafáfila	2
ZA-05-DB-Guarrate	2
PA-04-CA-Osorno	3
PA-05-CA-Villarmentero	3
PA-06-CA-Paredes	3
PA-07-CA-Frechilla	3
PA-08-CA-Villarramiel	3
VA-07-TC-Cuenca	3
VA-08-TC-Tordehumos	3
ZA-03-CP-Castroverde	3

El grupo 1 está compuesto por 10 zonas de vigilancia. Ocho de ellas se sitúan al sur de la comunidad autónoma, dentro o en las proximidades de las comarcas originarias asociadas al Sistema Central, en las provincias de Salamanca, Ávila y Segovia. Las otras dos se sitúan en la provincia de Soria (zonas de Ágreda y Trévago), también en comarcas originarias, en este caso asociadas al Sistema Ibérico. Este grupo parece en cierta forma determinado por su proximidad a las zonas originarias del sur y oeste, no habiéndose incluido en él tan sólo tres zonas de este tipo (Ribota en Segovia y Langa y Gómara en Soria, asociadas a la confluencia de ambos Sistemas montañosos).

El grupo 3, compuesto por ocho zonas de vigilancia, está claramente asociado a la macrocomarca conocida como Tierra de Campos, al tener todas sus zonas de vigilancia incluidas en su zona más central, en concreto desde el este de la comarca de Campos-Pan en Zamora hasta el este de la comarca de Campos palentina, e incluyendo las zonas de vigilancia ubicadas en la comarca de Tierra de Campos de Valladolid. No se han incluido en este grupo otras zonas de vigilancia también de Tierra de Campos, concretamente aquellas que rodean a las incluidas en este grupo.

Por último, el grupo 2 incluye a todo el resto de zonas de vigilancia, sin observarse un patrón geográfico especial en él, siendo éste el más numeroso con 26 zonas de vigilancia.

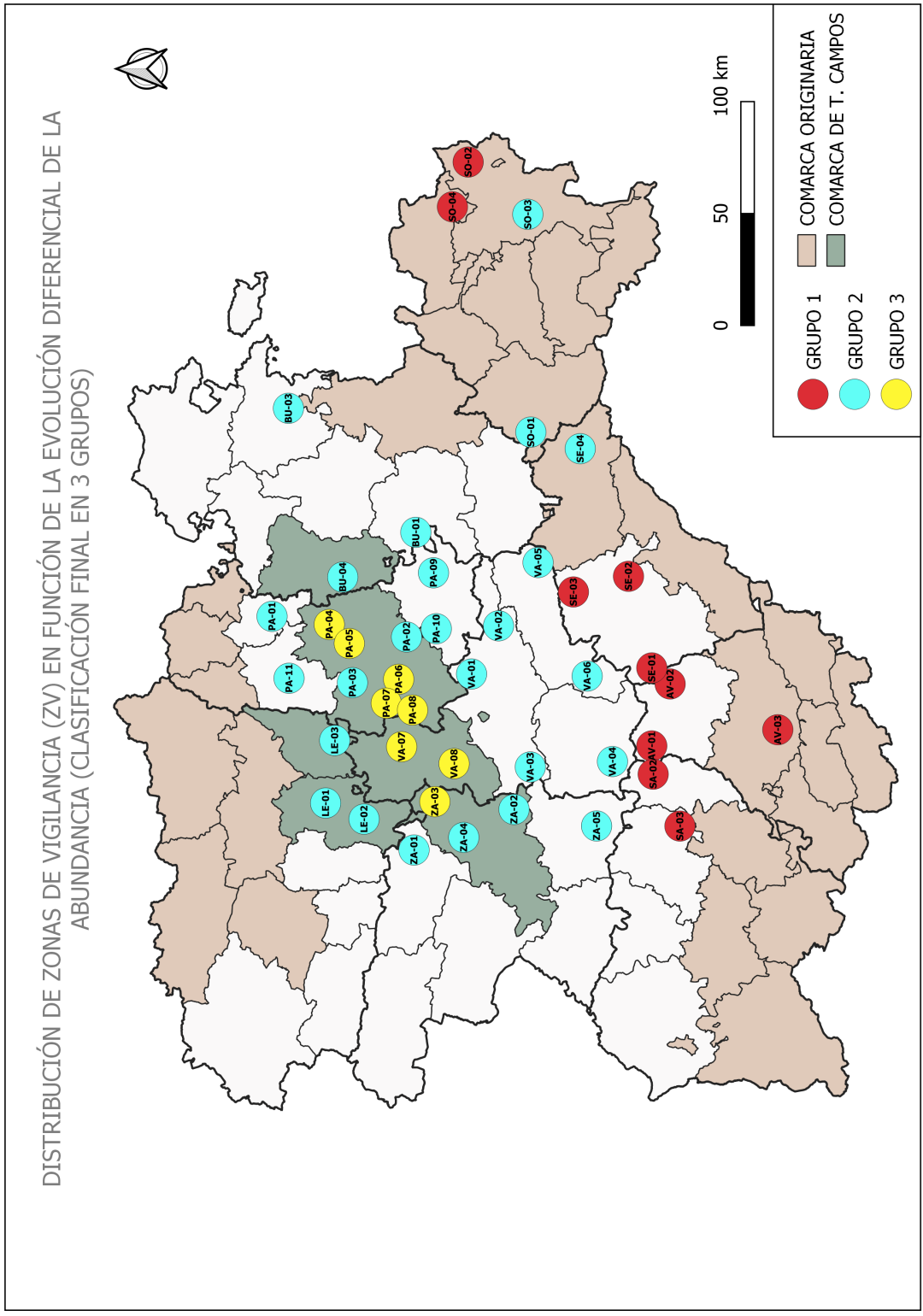


Figura 4.29: Agrupación final de las zonas de vigilancia en Castilla y León

4.4. Definición final de zonas diferenciales de riesgo y estudio comparativo en la evolución y ciclicidad.

Los tres grupos de zonas de vigilancia definidos en el paso anterior, habiendo sido obtenidos a través de una jerarquización final conjunta de la expresión diferencial de la evolución de la abundancia natural (a partir de del hábitat *vías de dispersión*) y de su expresión en los cultivos (a partir del hábitat *cultivos*), definirán la propuesta de de zonas diferenciales de riesgo buscada como parte del objetivo global del presente estudio. El objetivo de los apartados que se abordan a continuación consiste en la detección de las diferencias entre ellos, de forma que puedan definirse los umbrales/cualificación del riesgo de plaga en cada uno de ellos.

Para ello se contará con el apoyo de técnicas y métodos relativos a representación gráfica de sus evoluciones comparativas, ANOVA y test LSD de Fisher (con la corrección de Bonferroni) en los distintos periodos o la construcción de correlogramas.

4.4.1. Evolución temporal de las zonas diferenciales: gráficos y mapas de calor

Las figuras 4.30 y 4.31 contienen la evolución temporal en forma de gráficos de líneas de los grupos en los hábitat *vías de dispersión* y *cultivos* respectivamente. Como complemento, en la figura 4.32 se representa la misma información en forma de mapas de calor.

De la observación conjunta de ambas figuras, tanto considerando la imagen global como lo sucedido en cada uno de los grupos, en primer lugar se desprende que en todos los casos la abundancia de topillo campesino en las vías de dispersión es siempre superior a la que se observa en los cultivos. Por otra parte, se deduce que el periodo de estudio ha abarcado total o parcialmente tres ciclos de explosión demográfica.

En general, considerando la globalidad de Castilla y León, se podría intuir que el primero de esos ciclos parece haber comenzado ya en la representación gráfica. Sin embargo, si se hace una retrospectiva en este mismo estudio hacia los mapas de calor expuestos en las figuras 4.13 y 4.14, y se observa el periodo 7 (uno de los eliminados en la primera parte del proceso por exceso de datos perdidos), en aquellos casos que se tiene información para los periodos 7 y 8 las abundancias eran prácticamente nulas en el primero de estos periodos, detectándose los incrementos en el segundo. Por tanto, puede considerarse que el primer ciclo observado realmente comenzó en el periodo 8 incluido en los gráficos.

Análogamente, podría intuirse que el último de los ciclos aún no habría concluido en el periodo de estudio a a la vista de las figuras 4.30 y 4.31. En este caso, no se cuenta con datos posteriores al mismo, por lo que no se puede deducir más allá que la percepción de que ese ciclo parece estar en fase de declive.

Índice de actividad por grupos en Vías de dispersión

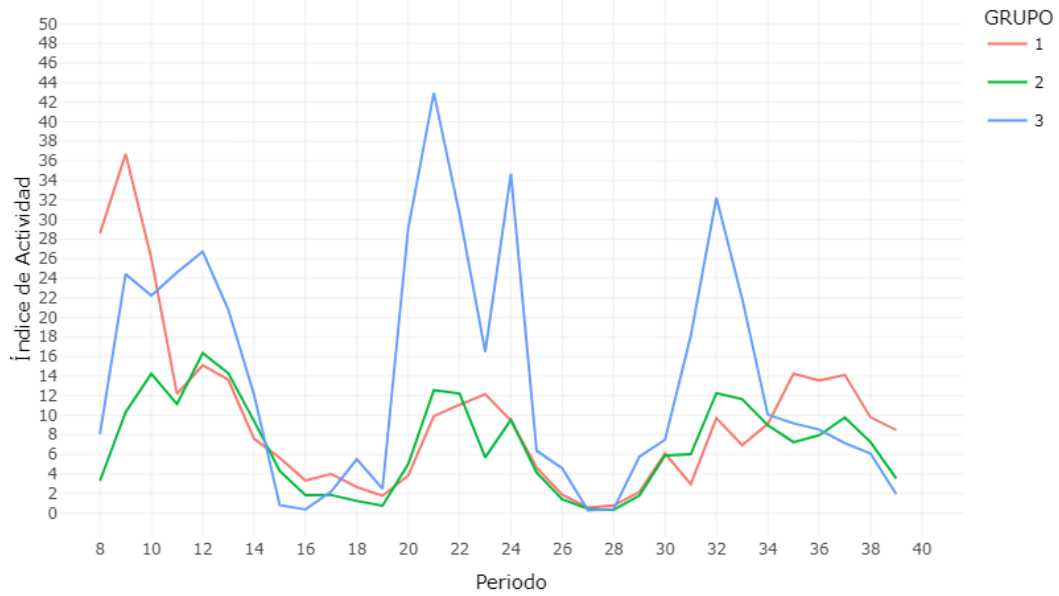


Figura 4.30: Evolución de los grupos finales de zonas de vigilancia para el hábitat Vías de dispersión

Índice de actividad por grupos en Cultivos

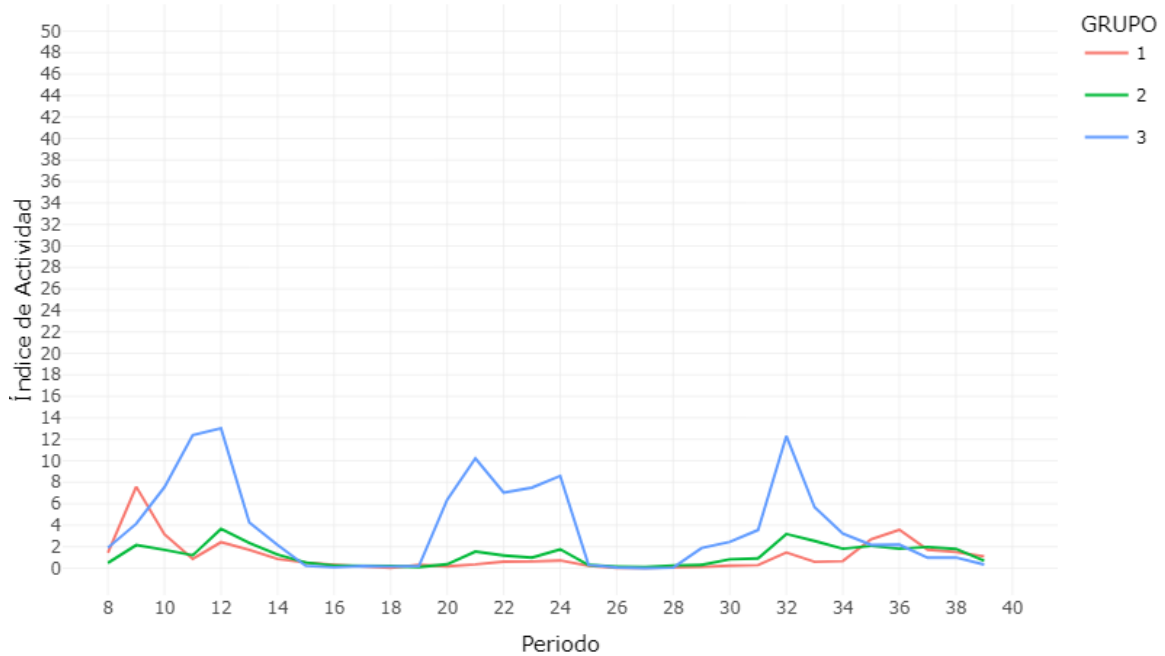


Figura 4.31: Evolución de los grupos finales de zonas de vigilancia para el hábitat Cultivos

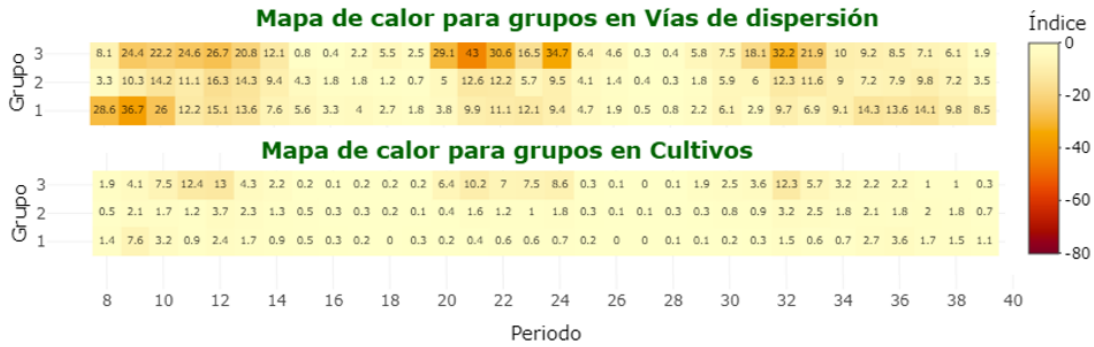


Figura 4.32: Mapas de calor de los grupos finales de zonas de vigilancia en Vías de dispersión y Cultivos

En consecuencia, el primer ciclo abarcaría desde el inicio del estudio hasta el periodo 14 (duración de 6 estaciones, equivalente a 1.5 años), el segundo desde el 20 hasta el 25 (5 estaciones, 1,25 años) y el tercero comenzaría en el 29, ignorándose el final (siendo el más largo, con una duración mínima de 10 estaciones, 2,5 años).

Cuando se observa el comportamiento de los grupos, lo primero que resalta es que, durante los ciclos de explosión demográfica, la tendencia general en los periodos implicados es la observación de valores superiores de abundancia asociados al grupo 3. Esto se expresa también en cuanto a los picos de abundancia máximos asociados a cada ciclo, que se detectan siempre en el grupo 3, salvo en el primer ciclo para el caso de las vías de dispersión, que aparecen en el grupo 1.

En añadido, en el primer ciclo, el pico máximo para el grupo 1 se detecta en el periodo 9 (otoño de 2013), para pasar posteriormente a un declive progresivo. Por su parte, los grupos 2 y 3, tras el periodo 9 tuvieron un ligero descenso en el invierno siguiente, pero alcanzaron el pico máximo en el periodo 12 (verano de 2014), con un retraso de tres estaciones respecto al 1.

En el segundo ciclo, en apariencia la duración es similar en los tres grupos para el caso de las vías de dispersión. Sin embargo, para los grupos 2 y 3 el pico máximo se detecta en el periodo 21 (otoño de 2016), con un declive hasta el 23 (primavera de 2017) y un rebote en el verano siguiente. Para el grupo 1 el máximo se detecta en ese periodo 23, con un retraso de dos estaciones respecto a los otros dos grupos.

En el caso de los cultivos, se observa un patrón similar para el grupo 3, con un máximo en el periodo 21 y un rebote, en este caso de magnitud similar, en el 24. Para el grupo 2 parece observarse un patrón similar al del 3, pero con una importante diferencia en las magnitudes alcanzadas. En el grupo 1 no se aprecia ningún tipo de expresión del ciclo.

En el último ciclo, para el caso de las vías de dispersión, los grupos 2 y 3 alcanzan su máximo en el periodo 32 (verano de 2019), para entrar el grupo 3 a continuación en declive. El grupo 2 parece entrar también en ese declive pero presenta un rebote que se expresa en el periodo 37 (otoño de 2020). Por su parte, el grupo 1 alcanza su máximo

en el periodo 35 (primavera de 2020), con tres estaciones de retraso respecto a los otros dos grupos, manteniéndose hasta el periodo 37 antes de comenzar en apariencia con su declive.

En los cultivos, el tercer ciclo parece expresarse similar para los grupos 2 y 3, aunque con una evidente diferencia en las magnitudes. Sin embargo, si bien en el 3 parece comenzar en el periodo 29 con el máximo en el 32, en el 2 se alcanza el máximo en el mismo momento, pero con un comienzo en apariencia ligeramente retrasado. En añadido, el declive en el grupo 3 es relativamente pronunciado desde el momento del máximo, mientras que en el 32 parece ser mucho más atenuado hasta el periodo 36. El retraso en la expresión del ciclo parece más evidente en el grupo 1, pues el ascenso comienza más claramente en el periodo 35, seis estaciones posteriores a la detección en el 3, con máximo en el 36, cuatro estaciones (un año) más tarde del máximo en los otros dos grupos.

Por último, en cuanto a las magnitudes de los máximos por ciclo en cada grupo, para el caso de las vías de dispersión, en el grupo 1 resultan claramente diferentes, con una mayor expresión en el primer ciclo respecto a los otros dos. En el grupo 2 la magnitud del máximo parece más uniforme, ligeramente superior en el primer ciclo. En el grupo 3 parece existir también cierta uniformidad pero, a diferencia de los otros dos grupos, en el primer ciclo se detecta la menor magnitud alcanzada, observándose en el segundo ciclo el mayor valor.

En cultivos, en el grupo 1 se observa de nuevo la mayor magnitud en el primer ciclo, sin observarse expresión del segundo. En los grupos 2 y 3 la expresión de máximos parece relativamente uniforme, en este caso observándose la mayor magnitud asociada al primer ciclo.

4.4.2. Análisis de la varianza y test LSD

En las tablas 4.12 y 4.13 se sintetizan, para los hábitat *vías de dispersión* y *cultivos* respectivamente, de forma conjunta los resultados obtenidos en el análisis de varianza para detección de diferencias significativas entre grupos y del test LSD de Fisher (con la corrección de Bonferroni) como apoyo para la explicación de las posibles diferencias detectadas.

Asociado a las vías de dispersión, el análisis de varianza ha detectado diferencias significativas entre los grupos en un total de 18 periodos, frente a los 30 en que se detectaban diferencias significativas entre zonas de vigilancia en la tabla 4.8.

Los periodos en los que no se han detectado diferencias significativas entre grupos, y que sí se detectaron entre zonas de vigilancia al inicio, han sido: 13 y 14, coincidentes con el final del declive del primer ciclo; 16, 17 y 19, incluidos en momentos de bajas densidades entre el primer y el segundo ciclo; 25, al final del declive del segundo ciclo; 27 y 28, periodos de baja densidad entre el segundo y el tercer ciclo; 30, periodo en fase de ascenso de densidades poblacionales del tercer ciclo, siendo este el

único periodo de este tipo donde no se han identificado diferencias; y 34, 36, 37 y 38, todos ellos incluidos en el periodo de declive poblacional del tercer ciclo para el grupo 1.

Tabla 4.12: *Tabla resumen de los periodos significativos en vías de dispersión*

Periodo	8	9	10	11	12	15	18	20	21
Año	2013	2013	2014	2014	2014	2015	2016	2016	2016
Estación¹	VER	OTO	INV	PRI	VER	PRI	INV	VER	OTO
R²²	0.85	0.61	0.22	0.33	0.15	0.17	0.29	0.73	0.65
Pr.³	**	**	**	**	*	*	**	**	**
Grupo 1⁴	a	a	a	b	a	a	ab	b	b
Grupo 2	c	c	b	b	a	ab	b	b	b
Grupo 3	b	b	ab	a	a	b	a	a	a

Periodo	22	23	24	26	29	31	32	33	39
Año	2017	2017	2017	2018	2018	2019	2019	2019	2021
Estación	INV	PRI	VER	INV	OTO	PRI	VER	OTO	PRI
R²	0.33	0.36	0.49	0.17	0.46	0.26	0.43	0.21	0.29
Pr.	**	**	**	*	**	**	**	*	**
Grupo 1	b	a	b	ab	b	b	b	b	a
Grupo 2	b	b	b	b	b	b	b	b	b
Grupo 3	a	a	a	a	a	a	a	a	b

En cuanto a los cultivos, se han detectado diferencias significativas entre los grupos en un total de 15 periodos, frente a los 27 en los que se detectaron diferencia significativas entre zonas de vigilancia (4.8).

Los periodos en los que no se han detectado diferencias y que sí lo habían sido entre zonas, son: 13 y 14 (análogos a los detectados en vías de dispersión, final declive primer ciclo); 16 y 18 (bajas densidades entre primer y segundo ciclo); 25 (final del declive del segundo ciclo, análogo a lo detectado en vías de dispersión), 27 y 28 (análogos a vías de dispersión, bajas densidades entre segundo y tercer ciclo); y 34, 36, 37, 38 y 39 (todos análogos, salvo el 39, a lo observado en vías de dispersión, correspondiendo el 39 a una fase de declive en ambos hábitat).

¹Estación del año, donde INV representa el invierno, PRI representa la primavera, VER representa el verano y OTO representa el otoño

²Coefficiente de Determinación Parcial

³Significatividad del test ANOVA, * a nivel 0.05, ** a nivel 0.01

⁴Clasificación obtenida en el test LSD de Fisher con la corrección de Bonferroni

En consecuencia, la clasificación en tres grupos obtenida sí parece contemplar en general la capacidad de definir diferencias en momentos importantes a tener en cuenta de cara a la Gestión Integrada: inicios de ciclos, con especial importancia en tareas predictivas y preventivas, y de momentos de máxima expresión de los mismos, de especial importancia en función de la posible intensidad de implementación de medidas.

Tabla 4.13: Tabla resumen de los periodos significativos en Cultivos

Periodo	8	9	10	11	12	20	21	22
Año	2013	2013	2014	2014	2014	2016	2016	2017
Estación ⁵	VER	OTO	INV	PRI	VER	VER	OTO	INV
R^2 ⁶	0.20	0.29	0.56	0.70	0.47	0.57	0.68	0.62
Pr. ⁷	*	**	**	**	**	**	**	**
Grupo ⁸ 1	ab	a	b	b	b	b	b	b
Grupo 2	b	b	b	b	b	b	b	b
Grupo 3	a	ab	a	a	a	a	a	a

Periodo	23	24	29	30	31	32	33
Año	2017	2017	2018	2019	2019	2019	2019
Estación	PRI	VER	OTO	INV	PRI	VER	OTO
R^2	0.59	0.40	0.63	0.24	0.20	0.50	0.23
Pr.	**	**	**	**	*	**	**
Grupo 1	b	b	b	b	b	b	b
Grupo 2	b	b	b	b	b	b	ab
Grupo 3	a	a	a	a	a	a	a

En cuanto a los coeficientes de determinación parcial explicados por las diferencias entre grupos, se han encontrado valores entre 0.15 (periodo 12) y 0.85 (periodo 8) en el caso de las vías de dispersión. Para los cultivos han sido entre 0.20 (para los periodos 8 y 31) y 0.70 (periodo 11). Considerando ambos ambientes 9 de los coeficientes de determinación parcial han resultado inferiores a 0.25, 13 mayores de 0.25 hasta 0.5, 10 mayores de 0.5 hasta 0.75 y 1 superior a 0.75. En general, parece observarse que los coeficientes pueden tender a ser relativamente bajos. Esto implica que puede quedar una componente de variabilidad encerrada en las diferencias entre zonas de vigilancia dentro de cada grupo, que pudiera ser interesante contemplar, o intentar ahondar en

⁵Estación del año, donde INV representa el invierno, PRI representa la primavera, VER representa el verano y OTO representa el otoño

⁶Coefficiente de Determinación Parcial

⁷Significatividad del test ANOVA, * a nivel 0.05, ** a nivel 0.01

⁸Clasificación obtenida en el test LSD de Fisher con la corrección de Bonferroni

ella, en estudios próximos.

En referencia a los test LSD, se han obtenido clasificaciones en todos los casos, salvo para el periodo 12 (verano de 2014) en el ambiente de las vías de dispersión. Puede resultar llamativo el no haber encontrado una clasificación de los grupos para este caso, puesto que a la vista de la figura 4.30, e incluso de la figura 4.32, si parece existir una diferencia en las magnitudes de las abundancias entre el grupo 3 respecto a los otros dos, siendo además un periodo en que se dieron los valores máximos para el primer ciclo en los grupos 2 y 3. Esto puede deberse a la existencia de alguna, o algunas, zonas de vigilancia dentro de alguno de los grupos que no se esté comportando de acuerdo a lo esperado en el grupo, y que, en función de ello, el test LSD con la corrección de Bonferroni resulte demasiado restrictivo como para definir una clasificación pese a que ANOVA sí detecte diferencias.

En cuanto a las clasificaciones en sí, para el caso de las vías de dispersión el grupo 3 ha quedado casi siempre clasificado en la categoría de mayores abundancias, en un total de 14 periodos, con las excepciones de los periodos 8 y 9, en los que también se ha detectado una categoría de abundancia inferior a él, y los periodos 15 y 39. En el caso del grupo 2 lo ha sido siempre en el de categoría más baja, con la única duda del periodo 15 en que no se discrimina entre ninguna de las categorías para este grupo. Respecto al grupo 1, ha sido clasificado en la categoría de mayor abundancia en siete de los periodos (8, 9, 10, 15, 18, 23 y 39), resultando en la categoría inferior en once de los restantes, apareciendo en el periodo 26 como no discriminado en cuanto a posible categoría.

Para los cultivos, en este caso el grupo 3 resulta en la categoría de mayor abundancia en todos los casos, salvo en el periodo 9 en que no discrimina. El grupo 2, de nuevo, aparece como el de densidad inferior, salvo en el caso del periodo 33 en el que no aparece discriminado. Respecto al grupo 1, en este caso siempre aparece asociado a la categoría de menor abundancia, salvo en el periodo 9, que aparece en el de alta, resultando como no claramente clasificado en el 8.

Puede deducirse que el grupo 3 representa zonas en las que los valores de abundancia natural (representadas en el hábitat *vías de dispersión*) tienden a ser los más altos, y en ningún caso han sido los más bajos (salvo en periodo 39, de declive poblacional en todos los casos). Asimismo, también resulta, y de forma muy consistente, como el de mayores abundancias asociadas al posible efecto en los cultivos (representadas en el hábitat *cultivos*). El grupo 2 representa la situación más opuesta, con las menores abundancias de forma consistente. El grupo 1 resulta intermedio, pues si bien tiende a tener menores abundancias naturales que el grupo 1, de forma puntual puede llegar a tenerlas incluso superiores. En cualquier caso, en este grupo 1 el efecto posible sobre los cultivos ha resultado similar al grupo 2, con las menores influencias.

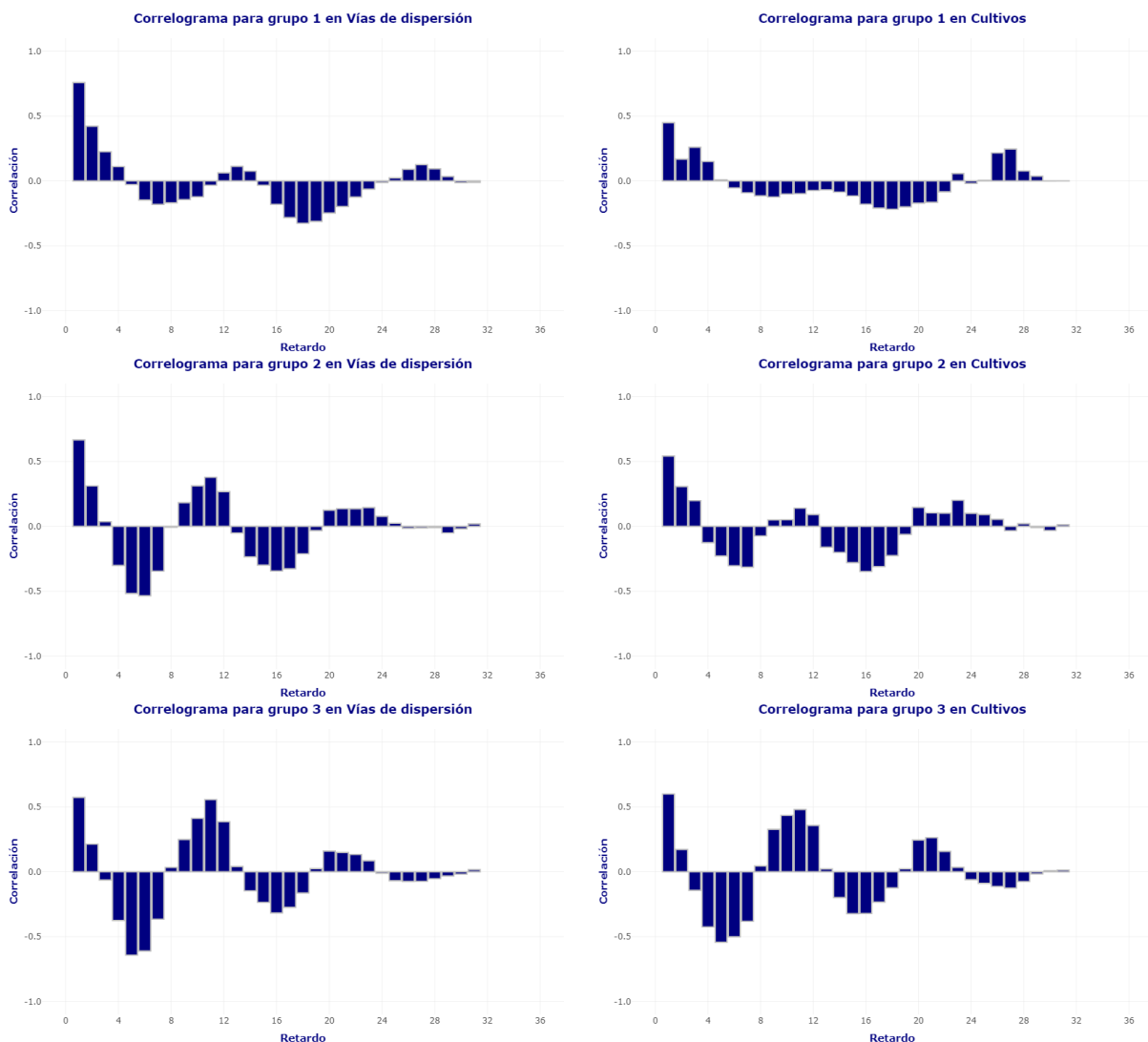
A la hora del posible establecimiento final de zonas de riesgo buscado en el estudio, y considerando que el apoyo para su decisión se realiza en función de los tres grupos utilizados en este capítulo, se considera para las atribuciones una nomenclatura sencilla de "zona de Riesgo alto", "Medio." "Bajo".

En función de ello:

- **Grupo 1:** incluye zonas de vigilancia clasificadas como Zonas de Riesgo Medio.
- **Grupo 2:** incluye zonas de vigilancia clasificadas como Zonas de Riesgo Bajo.
- **Grupo 3:** incluye zonas de vigilancia clasificadas como Zonas de Riesgo Alto.

4.4.3. Correlogramas

Tabla 4.14: Correlogramas para los grupos en vías de dispersión y cultivos



En la tabla 4.14 se representan los correlogramas obtenidos para los tres grupos que establecen las zonas de riesgo definidas, tanto para la abundancia natural como para el riesgo que puedan aparear para los cultivos.

Así, comenzando en este caso por el grupo 3, Riesgo Alto, se observa en ambos casos un patrón muy claro de ciclicidad entre 10 y 12 estaciones (2.5 o 3 años), viéndose cierta uniformidad en los valores de las correlaciones sucesivas en función del paso de ese número de estaciones, lo que puede ser un reflejo de valores de abundancia con magnitudes relativamente similares entre mismos momentos de los distintos ciclos. Esto parecía intuirse ya a la vista de los gráficos que ilustraban las figuras 4.30 y 4.31, quedando confirmado con el resultado de sus correlogramas. En este caso, recordar que esas magnitudes están clasificadas casi siempre como altas.

En el grupo 2 también se observan patrones claros de ciclicidad entre 10 y 12 estaciones, con la diferencia del grupo anterior de que dichas correlaciones no parecen tan uniformes, especialmente en el caso de los cultivos que tienden a tener menores valores. Esto puede ser reflejo de que, existiendo esa ciclicidad, el efecto en los cultivos pueda traducirse en diferentes magnitudes en los ciclos sucesivos. En cualquier caso, se debe recordar que, con independencia de la ciclicidad o las diferentes magnitudes en que pueda darse el fenómeno, las magnitudes siempre son bajas.

En el grupo 1 los patrones de ciclicidad son diferentes. Así, en cuanto a la abundancia natural la ciclicidad que pudiera interpretarse es más larga, entre 12 y 14 estaciones (que ya se intuía en los gráficos al observar cierto adelanto en la expresión de primer ciclo y retraso en la del tercero). En añadido, los valores de las correlaciones parecen no ser uniformes en su sucesión, interpretándose como diferencias en las magnitudes entre periodos sucesivos, en este claro muy visualizable con el apoyo de la figura 4.30, en la que se observan magnitudes de abundancia asociadas al primer ciclo claramente superiores a los otros dos.

Para el caso de los cultivos, en el grupo 1 no se observa ninguna de las ciclicidades observadas hasta ahora, sino que, caso de existir, recordaría a un patrón de 26-27 estaciones, prácticamente siete años. Esto implica que, en estas zonas, no siempre que suceden eventos naturales terminen expresándose en los cultivos, resultando en apariencia que la expresión sucede cada dos ciclos naturales. Si bien la probabilidad de ocurrencia en los cultivos parece ser menor que en los dos otros grupos, no se debe olvidar que, en este caso, las magnitudes que pueden alcanzarse, caso de suceder, pueden ser altas (ejemplo de ciclo 1 en 4.31).

4.5. Pensando en estudios futuros...

Si bien en el apartado anterior se considera que ya se ha dado cumplimiento al objetivo global del presente *TFG*, y se considera adecuada la clasificación obtenida en zonas de riesgo y la descripción de cada una, se somete una propuesta para abrir la posibilidad en estudios venideros. Esta se basa en que, pese a todo, se han detectado periodos en los que no se han observado diferencias entre grupos (aunque en los que se ha detectado puedan ser los más importantes. Y de las que se han detectado ha habido casos en los que se sospecha que pueda existir una variabilidad interesante encerrada en las diferencias entre zonas de vigilancia dentro de los grupos.

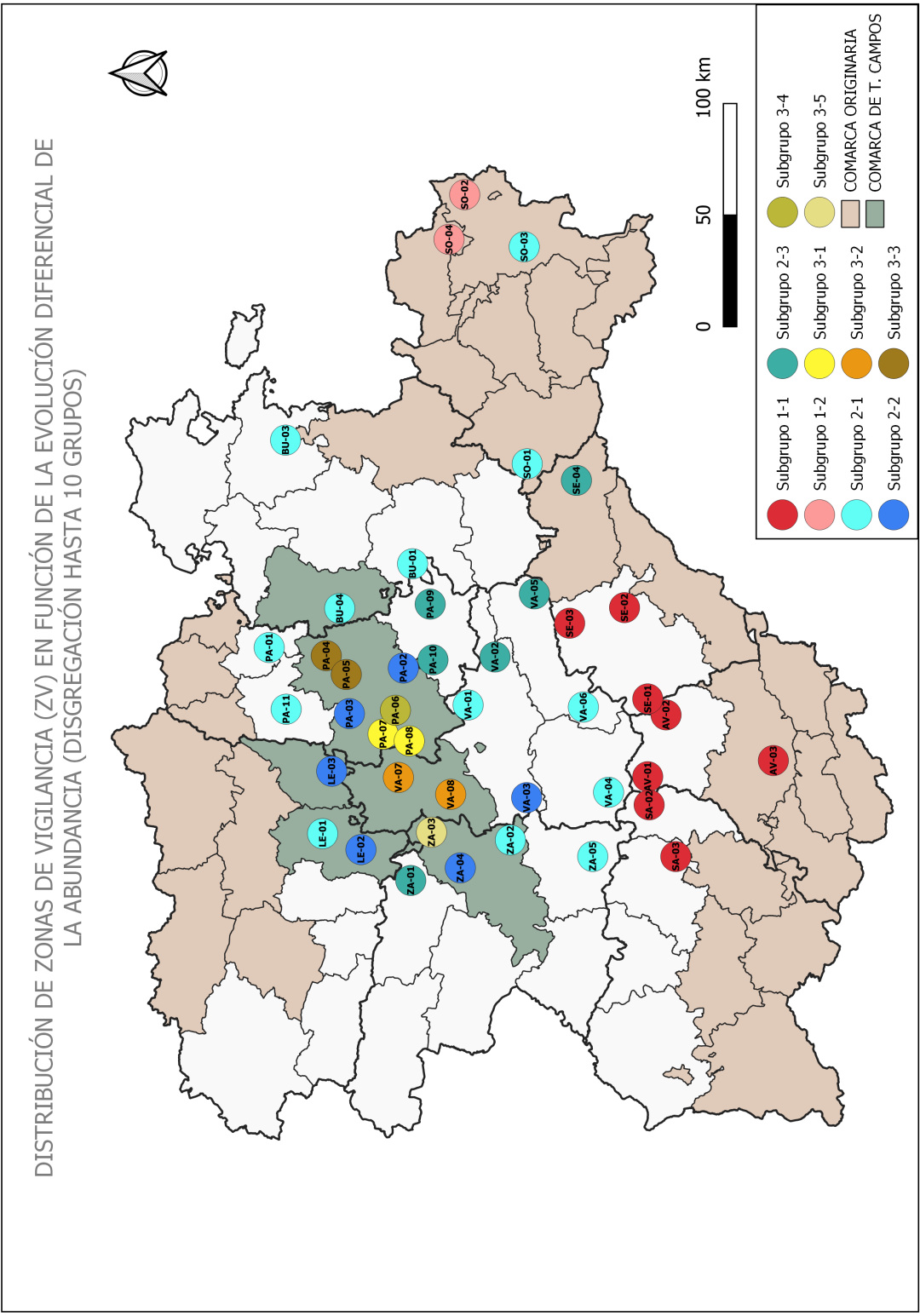


Figura 4.33: Propuesta de mapa para su posible exploración en estudios futuros

Por ello, se propone, sin profundizar más en ello ni someterlo a análisis especiales, que es lo que se juzga de interés para ser abordado en estudios futuros, únicamente un mapa con la distribución geográfica de las zonas de vigilancia agrupadas esta vez considerando una hipotética decisión de establecimiento de diez grupos a partir del dendrograma de la figura 4.27 en función del gráfico de codo de la figura 4.28. El mapa resultante aparece en la figura 4.33, en el que lo que se ilustra es la descomposición de los grupos de trabajo en subgrupos en función de la nueva agrupación.

Así, el grupo 1, de Riesgo Medio, compuesto por inicialmente diez zonas de vigilancia, aparentemente relacionadas con su inclusión o proximidad a comarcas originarias del sureste de Castilla y León, queda disgregado en dos subgrupos: uno, con ocho zonas de vigilancia, asociado a las comarcas próximas a las originarias vinculadas al Sistema Central en Salamanca, Ávila y Segovia; el segundo incluyendo las dos comarcas asociadas al Sistema Ibérico en la provincia de Soria.

El grupo 2, de riesgo bajo, compuesto por 25 zonas de vigilancia, y que en apariencia no presentaba ningún patrón de distribución geográfico especial, se divide en tres subzonas. Una de ellos (Subgrupo 2-2), incluye seis zonas de vigilancia, todas ellas en la macrocomarca de Tierra de Campos y rodeando a las zonas de esta macrocomarca incluidas en el grupo 3, salvo la zona de vigilancia VA-03 (Villalar) situada un poco más al sur.

El subgrupo 2-3 queda formado por seis zonas de vigilancia que incluye las del Cerrato palentino y se extiende aparentemente de forma lineal hacia el sureste hasta la zona en comarca originaria SE-04 (Ribota). En él se ha incluido, en este caso sin una aparente relación geográfica, la zona de vigilancia ZA-01 (Benavente) en la comarca de Benavente y los Valles.

El subgrupo 2-1, formado por 13 zonas de vigilancia, no muestra ningún patrón de distribución aparente más allá que distribuirse por las áreas no ocupadas por el resto.

Por último, el grupo 3, de riesgo alto, compuesto inicialmente por el menor número de zona de vigilancia, ocho, es el que resulta en esta ocasión como más disgregado, detectándose hasta cinco subgrupos. Tres de dichos subgrupos se forman de dos zonas de vigilancia, e incluso aparecen dos subgrupos con tan sólo una, en los casos de ZA-03 (Castroverde en Campos-Pan de Zamora) y PA-06 (Paredes en Palencia).

Esta disgregación que se observa en el grupo 3 parece revelar no sólo que se trate de aquel en el que se incluyen las zonas de vigilancia con mayores riesgos, sino que entre ellas pudiera existir una variabilidad suficiente como para detectarse nuevos patrones de expresión susceptibles de interés en estudios futuros, considerando que esto ocurre en una superficie pequeña cuando se compara con la totalidad de la zona de estudio.

5. Conclusiones

1. A partir de los datos del programa de monitorización se ha obtenido una primera matriz de dos vías con 47 zonas de vigilancia y 39 periodos que, tras una primera revisión de datos perdidos, definió una matriz resultante final de trabajo con una estructura de 43 zonas de vigilancia y 32 periodos.
2. Se han encontrado diferencias en la expresión de la abundancia de topillo campesino entre zonas de vigilancia para 30 periodos en el caso de la evolución natural de la especie, y en 27 para el caso de su expresión en los cultivos.
3. La aproximación a la jerarquización de las zonas de vigilancia a través del análisis clúster aglomerativo ajustado a las series temporales ha resultado efectiva, utilizando como medida de disimilitud la distancia euclídea y el método de agrupación de Ward, detectándose, tanto para la evolución de la abundancia natural como para su expresión en los cultivos, un total de 5 grupos de zonas de vigilancia.
4. Se ha utilizado la Deformación Dinámica del Tiempo como apoyo para la clasificación de las series temporales, obteniéndose una clasificación final de las zonas de vigilancia en 3 grupos, siendo estos los utilizados para definir la distribución de zonas con distintos potenciales de riesgo frente a la plaga de topillo campesino.
5. Se han definido tres umbrales de riesgo diferenciales en Castilla y León frente a la incidencia de la plaga del topillo campesino:
 - Zonas de Riesgo Alto: distribuidas en la parte central de la comarca de Tierra de Campos, con valores de abundancia, tanto natural como de expresión en cultivos, consistentemente altos, y ciclicidad observada de 10-12 estaciones.
 - Zonas de Riesgo Medio: distribuidas en las proximidades de comarcas originarias asociadas a los Sistemas Central e Ibérico, con valores de abundancia natural ocasionalmente altos y asociados a una ciclicidad de 12-14 estaciones, y con una expresión en los cultivos que no siempre parece traducirse y apuntando a una ciclicidad de 26-27 estaciones, pero que puede ser también ocasionalmente alta.
 - Zonas de Riesgo Bajo: sin ninguna distribución geográfica aparente más allá de su distribución en la zona de estudio en áreas no ocupadas por las zonas de los otros grupos, con valores de abundancia tanto natural como de expresión en los cultivos bajos y ciclicidad de 10-12 años.
6. Se ha propuesto un nuevo mapa de disgregación de zonas de vigilancia a partir de la clasificación obtenida en el presente estudio, como posible punto de reflexión para el planteamiento de estudios futuros.

Bibliografía

- [1] Grant R. Singleton, Steve R. Belmain, Peter R. Brown, and Bill Hardy, editors. (2010). *Rodent outbreaks: ecology and impacts*. Los Baños (Philippines): International Rice Research Institute. 289 p.
- [2] Barfknecht, Ralf, Frederiricks, Timothy, Jacob, Jens, and Manson, Phil. 2014. *Common vole (Microtus arvalis) ecology and management: implications for risk assessment of plant protection products*. (wileyonlinelibrary.com) DOI 10.1002/ps.3695
- [3] González-Esteban, Jorge y Villate, Idoia. (s.f.) *Microtus arvalis* en el Atlas de la biodiversidad del Ministerio de Medio Ambiente de España (MMA). Recuperado de https://www.miteco.gob.es/es/biodiversidad/temas/inventarios-nacionales/ieet_mami_microtus_arvalis_tcm30-99825.pdf
- [4] Hernández, M. C., Sánchez-González, B., Navarro-Castilla, A., Barja, I. (2017). *Topillo campesino – Microtus arvalis*. En: *Enciclopedia Virtual de los Vertebrados Españoles*. Salvador, A., Barja, I. (Eds.). Museo Nacional de Ciencias Naturales, Madrid. Recuperado de <http://www.vertebradosibericos.org/>
- [5] Hernández, José Alfredo *El topillo campesino y sus depredadores* Recuperado de <https://zamorabiodiversa.wordpress.com/2022/01/25/el-topillo-campesino-y-sus-depredadores/>
- [6] Rodríguez Ferri, Elías-Fernando y colaboradores (2017). *Tularemia: Una aproximación a su estudio integral en Castilla y León*. (pp 20-21). Junta de Castilla y León, Academia Ciencias Veterinarias de Castilla y León.
- [7] Luque-Larena JJ, Mougeot F, Vidal Roig D, Lambin X, Rodríguez-Pastor R, Rodríguez-Valín E, et al. *Tularemia outbreaks and common vole (Microtus arvalis) irruptive population dynamics in Northwestern Spain, 1997–2014*. *Vector-Borne Zoonotic Dis.* (2015) 15:568–70. 10.1089/vbz.2015.1770
- [8] Herrero-Cófreces, S., Mougeot, F., Sironen, T., Meyer, H., Rodríguez-Pastor, R., Luque-Larena, J. (2022). Viral Zoonoses in Small Wild Mammals and Detection of Hantavirus, Spain. *Emerging Infectious Diseases*, 28(6), 1294-1296. Recuperado de <https://doi.org/10.3201/eid2806.212508>.
- [9] Yigit, N., Hutterer, R., Kryštufek, B. Amori, G. (2016). *Microtus arvalis*. *The IUCN Red List of Threatened Species* 2016: e.T13488A22351133. Recuperado de <https://dx.doi.org/10.2305/IUCN.UK.2016-2.RLTS.T13488A22351133>. en Accedido el 20 de Mayo de 2022.

- [10] Luque-Larena, Juan, Mougeot, Francois, Viñuela, Javier, Arroyo, Beatriz, Jareño, Daniel, Arroyo, Leticia y Lambin, Xavier. (2013). *Recent large-scale range expansion and outbreaks of the common vole (*Microtus Arvalis*) in NW Spain*. Basic and Applied Ecology. in press. 10.1016/j.baae.2013.04.006.
- [11] Orden AYG/96/2019 de gestión integrada de riesgos derivados de la presencia de topillo campesino en el territorio de Castilla y León, *Boletín Oficial de Castilla y León*, 27, del 8 de febrero de 2019. <https://bocyl.jcyl.es/boletines/2019/02/08/pdf/B0CYL-D-08022019-14.pdf>
- [12] Agencia Estatal de Meteorología (AEMET), Instituto Tecnológico Agrario de Castilla y León (ITACyL) (2014) *Mapa de las comarcas agrarias de Castilla y León*. Recuperado de http://ftp.itacyl.es/Atlas_Agroclimatico/03_ActividadAgraria/Comarcalizacion/02_059_1_AtlasAC_Comarcalizacion_Comarcas_Agrarias.pdf
- [13] del Hierro, A. (2020) *Comparación y calibrado de dos métodos de estimación de la abundancia del topillo campesino (*Microtus arvalis* Pallas)*. Universidad de Valladolid. Recuperado de <http://uvadoc.uva.es/handle/10324/43851>
- [14] Fernández, Tomás y Tamaro, Elena. (2004) Biografía de Ronald Aylmer Fisher. En Biografías y Vidas. *La enciclopedia biográfica en línea [Internet]*. Barcelona, España. Recuperado de <https://www.biografiasyvidas.com/biografia/f/fisher.htm>
- [15] Amat Rodrigo, Joaquín. (2016) *ANOVA análisis de varianza para comparar múltiples medias*. En *cienciadedatos.net*. Recuperado de <https://www.ugr.es/~bioestad/guiaspss/practica7/ArchivosAdjuntos/ComparacionesMultiples.pdf>
- [16] Amat Rodrigo, Joaquín. (2016) *Comparaciones múltiples: corrección de p-value y FDR*. En *cienciadedatos.net*. Recuperado de https://www.cienciadedatos.net/documentos/19b_comparaciones_multiples_correccion_p-value_fdr
- [17] M. A. J. Cuadrillero (2018) *Clustering Jerárquico en R*. Recuperado de <https://rpubs.com/mjimcua/clustering-jerarquico-en-r>
- [18] Joe H. Ward Jr. (1963) Hierarchical Grouping to Optimize an Objective Function, *Journal of the American Statistical Association*, 58:301, 236-244, DOI: 10.1080/01621459.1963.10500845
- [19] Gançarski, Pierre, Ketterlin, Alain and Petitjean, François (2011). A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, Vol.44, Issue 3 (pp. 678-693) Recuperado de <https://doi.org/10.1016/j.patcog.2010.09.013>.
- [20] Rodó, Paula. *Autocovarianza*. En *Economipedia.com*, 2022. Recuperado de <https://economipedia.com/definiciones/autocovarianza.html#:~:text=La%20autocovarianza%20es%20un%20estad%3%ADstico,en%20diferentes%20momentos%20del%20tiempo.>

[21] Rodó, Paula. *Función de autocorrelación simple*. En Economipedia.com, 2020. Recuperado de <https://economipedia.com/definiciones/funcion-de-autocorrelacion-simple.html>

6. Anexos

6.1. Anexo: Municipios de referencia de las zonas de vigilancia

El presente anexo muestra los municipios de referencia de cada zona de vigilancia.

Municipio de referencia	Zona de vigilancia
Madrigal de las Altas Torres	AV-01-AM-Madrigal
Orbita	AV-02-AM-Orbita
Solosancho	AV-03-AV-Solosancho
Navarredonda de Gredos	AV-04-GR-Navarredonda
Villahoz	BU-01-AR-Villahoz
Valle de Santibáñez	BU-02-AN-VSantibáñez
Briviesca	BU-03-BU-Briviesca
Castrojeriz	BU-04-PI-Castrojeriz
Gusendos de los Oteros	LE-01-EC-Gusendos
Villaornate y Castro	LE-02-EC-Villaornate
Sahagún	LE-03-SA-Sahagún
Prádanos de Ojeda	PA-01-BO-Pradanos
Monzón de Campos	PA-02-CA-Monzón
Cervatos de la Cueva	PA-03-CA-Cervatos
Osorno la Mayor	PA-04-CA-Osorno
Villarmentero de Campos	PA-05-CA-Villarmentero
Paredes de Nava	PA-06-CA-Paredes
Frechilla	PA-07-CA-Frechilla
Villarramiel	PA-08-CA-Villarramiel
Baltanás	PA-09-EC-Baltanás
Hontoria de Cerrato	PA-10-EC-Hontoria
Saldaña	PA-11-SV-Saldaña
Villaflores	SA-02-PB-Villaflores
Macotera	SA-04-PB-Macotera
Aldearrubia	SA-03-SA-Aldearrubia
Negrilla de Palencia	SA-05-SA-Negrilla
Codorniz	SE-01-CU-Codorniz
Escalona del Prado	SE-02-CU-Escalona
Cuéllar	SE-03-CU-Cuéllar

Continúa en la página siguiente

Municipio de referencia	Zona de vigilancia
Ribota	SE-04-SP-Ribota
Langa de Duero	SO-01-BO-Langa
Ágreda	SO-02-CG-Ágreda
Gómara	SO-03-CG-Gómara
Trévago	SO-04-TA-Trévago
Cigales	VA-01-CE-Cigales
Piña de Esgueva	VA-02-CE-Piña
Villalar de los Comuneros	VA-03-CE-Villalar
Carpio	VA-04-SU-Carpio
Peñafiel	VA-05-SE-Peñafiel
Olmedo	VA-06-SE-Olmedo
Cuenca de Campos	VA-07-TC-Cuenca
Tordehumos	VA-08-TC-Tordehumos
Benavente	ZA-01-BV-Benavente
Pozoantiguo	ZA-02-CP-Pozoantiguo
Castroverde de Campos	ZA-03-CP-Castroverde
Villafáfila	ZA-04-CP-Villafáfila
Guarrate	ZA-05-DB-Guarrate

6.2. Anexo: Comarcas agrarias por provincias

Este anexo muestra las comarcas agrarias utilizadas en el estudio, distribuidas por provincias

Provincia	Comarca agraria
Ávila	Arévalo-Madrigal
Ávila	Ávila
Ávila	Gredos y Valle del Tiétar
Burgos	Pisuerga
Burgos	Arlanza
Burgos	La Ribera
Burgos	Arlanzón
Burgos	Bureba-Ebro
Burgos	Demanda
León	Tierras de León
León	Elsa-Campos
León	Sahagún
León	Bierzo
León	La Bañeza
Palencia	Campos
Palencia	El Cerrato
Palencia	Saldaña-Valdavia
Palencia	Boedo-Ojeda
Palencia	Guardo
Salamanca	Alba de Tormes
Salamanca	Peñaranda de Bracamonte
Salamanca	Salamanca
Salamanca	Fuente de San Esteban
Salamanca	Ledesma
Segovia	Segovia
Segovia	Sepúlveda
Segovia	Cuéllar
Soria	Almazán
Soria	Soria
Soria	Campo de Gómara
Soria	Burgo de Osma
Soria	Tierras Altas y Valle del Tera
Valladolid	Sur
Valladolid	Tierra de Campos
Valladolid	Sureste
Valladolid	Centro
Zamora	Duero Bajo
Zamora	Benavente y los Valles
Zamora	Campos-Pan

Continúa en la página siguiente

Provincia	Comarca agraria
Zamora	Sayago

Tabla 6.16: Comarcas agrarias por provincia

6.3. Anexo: Dendrograma la clasificación final de las zonas de vigilancia

Este anexo muestra el dendrograma correspondiente a la última clasificación realizada.

DENDROGRAMA PARA ZONAS DE VIGILANCIA CON DTW

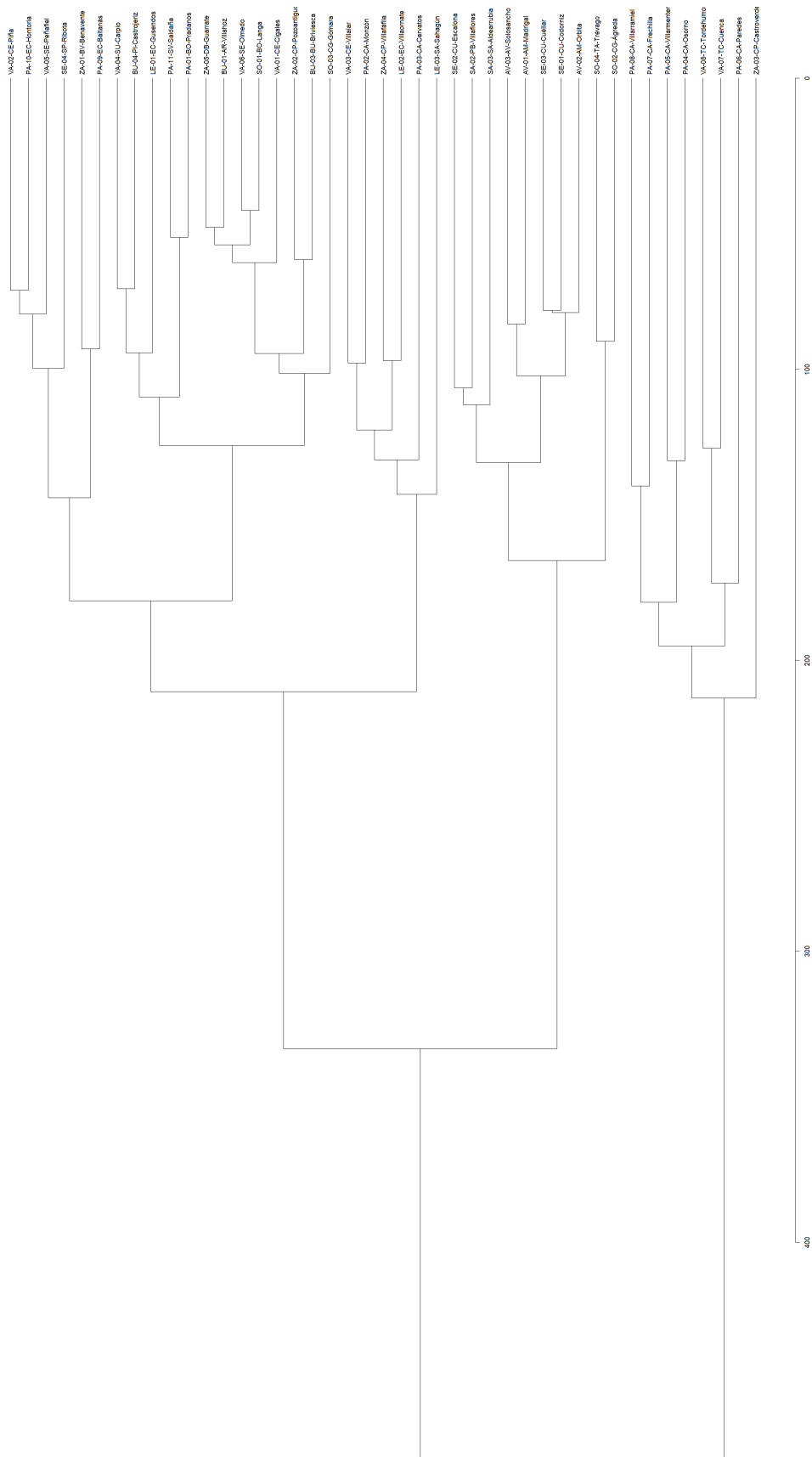


Figura 6.34: Dendrograma obtenido con la clasificación final