

El análisis de clústeres jerárquicos aplicado en la organización territorial de los asentamientos medievales del siglo XI en el Baix Montseny (España)

Hierarchical Cluster Analysis in the Territorial Organization of Medieval Settlements of the 11th Century in Baix Montseny (Spain)

Dirección

Clara Martínez
Cantón

Gimena del Río
Riande

Francisco Barrón

Marc FERRER FERNÁNDEZ

Universidad de Barcelona

marcferrer@ub.edu

<https://orcid.org/0000-0003-1655-9741>

Editora asociada

Romina De León

RESUMEN

En este artículo detallaremos los principios teóricos y procedimentales para realizar un análisis de clústeres jerárquicos aplicado a los estudios históricos medievales utilizando el método de la distancia de Gower. Los análisis de clústeres son una herramienta usada frecuentemente en la Arqueología para clasificar materiales encontrados en los yacimientos, pero estos métodos estadísticos también pueden ser usados para definir estrategias de ocupación y explotación económica de un territorio. La distancia de Gower es una prueba no paramétrica muy útil cuando queremos saber cómo de diferentes son los distintos registros dentro de un grupo de datos mixtos. A partir del trabajo previo con las fuentes documentales y cartográficas mostraremos los resultados obtenidos al aplicar un análisis de clústeres para examinar posibles estrategias de explotación y ocupación en el conjunto de asentamientos medievales existentes a lo largo del siglo XI en la comarca del Baix Montseny de Cataluña.

PALABRAS CLAVE

Análisis de clústeres, Humanidades Digitales, análisis espacial, georreferenciación, técnicas de análisis.

ABSTRACT

In this article, we provide a detailed explanation of the theoretical and procedural principles for conducting a hierarchical cluster analysis in medieval historical studies using the Gower's distance method. Cluster analysis is a commonly used tool in Archaeology to classify materials found at sites, but these statistical methods can also be applied to understand occupation and economic exploitation strategies of a specific territory. Gower's distance is a non-parametric test that is particularly useful to know how different the records of a mixed data set are. Using a combination of documentary and cartographic sources, we present the results of a cluster analysis to define occupation and economic strategies in a set of medieval settlements in the Baix Montseny region of Catalonia (Spain) during the 11th century.

KEYWORDS

Cluster Analysis, Digital Humanities, Spatial Analysis, Georeferencing, Research Techniques.

1. INTRODUCCIÓN

La clasificación es una actividad humana que ha existido desde tiempos inmemoriales. Desde siempre se ha designado, agrupado y ordenado los objetos, hechos históricos o seres vivos en distintas categorías. De hecho, el mismo lenguaje no es sino una forma de clasificación. Las tareas de clasificación son una cuestión fundamental en diversas disciplinas, en la taxonomía biológica es evidente, pero también son cruciales en la Astronomía, con la clasificación de los cuerpos celestes; en la Arqueología, con la de los materiales cerámicos; o en la Archivística, con la catalogación de documentos. Por lo general, la clasificación de una muestra de datos se basa en una o varias variables elegidas por el investigador. Sin embargo, estas clasificaciones empíricas tienen un gran componente subjetivo, de tal modo que una misma muestra de datos puede ser clasificada de manera diferente dependiendo de qué investigador haga la clasificación. Con el fin de reducir la subjetividad y aportar una estabilidad a las clasificaciones, se han ido desarrollando desde las matemáticas diversos algoritmos estadísticos y sistemas de clasificación sin apriorismos. El principio básico de estos métodos es agrupar los valores con los perfiles más parecidos de una muestra de datos, con el objetivo de reducir al mínimo las diferencias intracategorías¹ y, a la vez, maximizar las diferencias intercategorías². El resultado final será una clasificación que no se encuentra predefinida, sino que depende solo de las características y de la variedad de los datos de la muestra estudiada.

A estos métodos de clasificación se les ha dado diversos nombres según su ámbito de desarrollo, pero el término más general y aceptado es el de análisis de conglomerados jerárquicos o de clústeres jerárquicos. Esta técnica estadística ofrece diversas posibilidades dentro los estudios históricos, como ayudar a desarrollar clasificaciones y tipologías, a identificar patrones y tendencias que podrían pasar desapercibidas, a escoger casos representativos, o a elaborar preguntas e hipótesis mediante el análisis de los datos. El análisis de clústeres jerárquicos se ha utilizado en los estudios arqueológicos e históricos a lo largo del tiempo especialmente en investigaciones relacionadas con la cerámica, por ejemplo, los trabajos de Bellanger (2021), Makridis y Daras (2012), Papageorgiou (2001) o Schwabe y Slusallek (1981), entre otros. Sin embargo, su aplicación en el campo del paisaje cultural histórico ha sido mucho menos frecuente (Bertoncello, 2006; Esquivel et al., 1999), en parte debido a la complejidad y laboriosidad de los cálculos, especialmente cuando se mezclan datos cuantitativos y cualitativos. A pesar de ello, estos métodos ofrecen muchas posibilidades para el estudio del territorio en períodos históricos donde la documentación y las evidencias arqueológicas son limitadas, como sería el caso de la región del Baix Montseny entre los siglos VIII-XI.

¹ Intracategorías hace referencia a las diferencias que existen entre los miembros de una misma agrupación. El objetivo del análisis de clústeres es reducir al mínimo estas diferencias para que los miembros dentro de cada agrupación sean lo más homogéneos posible en relación con las características que se están evaluando.

² Intercategorías hace referencia a las diferencias entre las agrupaciones formadas con el análisis de clústeres, con el fin de maximizar las diferencias entre estas para que sean lo más distintas posible entre sí.

La región del Baix Montseny es una comarca natural situada entre la provincia de Barcelona y Girona (Figura 1). Su geografía física está esencialmente compuesta por tres unidades de relieve: el macizo del Montseny, la depresión prelitoral catalana y el conjunto formado por el macizo del Montnegre y el Corredor. La orografía de la comarca presenta una evidente dicotomía entre el llano y la montaña, con un desnivel de más de 1500 metros entre ambos. La hidrografía está dominada por la cuenca del Tordera y sus afluentes, todos ellos de carácter torrencial. Las nieblas son frecuentes y de gran importancia cuando el balance hídrico es deficitario. Todo esto le confiere a la comarca una diversidad climática y una biodiversidad excepcional, podemos encontrar tanto flora típica mediterránea como también especies propias de los Pirineos y de paisajes de la Europa central húmeda (Salvat & Pié, 2014, p. 19). Con todo esto no es de extrañar que el macizo del Montseny esté catalogado por la UNESCO como Reserva de la Biosfera³.

El Baix Montseny está formado por veinte municipios y con un área aproximada de 630 km², tiene como núcleo neurálgico la población de Sant Celoni. Durante siglos, el poblamiento tradicional de carácter agropecuario ha tenido un gran impacto en el paisaje, modificándolo para adaptarlo a sus necesidades. Sin embargo, el sistema agropecuario dominante en la región empezó a cambiar en la década de 1920 con la llegada de la industria textil y, especialmente, con el gran desarrollo industrial de las décadas de 1950 y de 1970. La industrialización de las últimas décadas ha comportado importantes cambios en el paisaje, como una intensa urbanización de la llanura y un progresivo abandono de los campos de cultivo y las zonas de pasto (Pujantell et al., 2017).

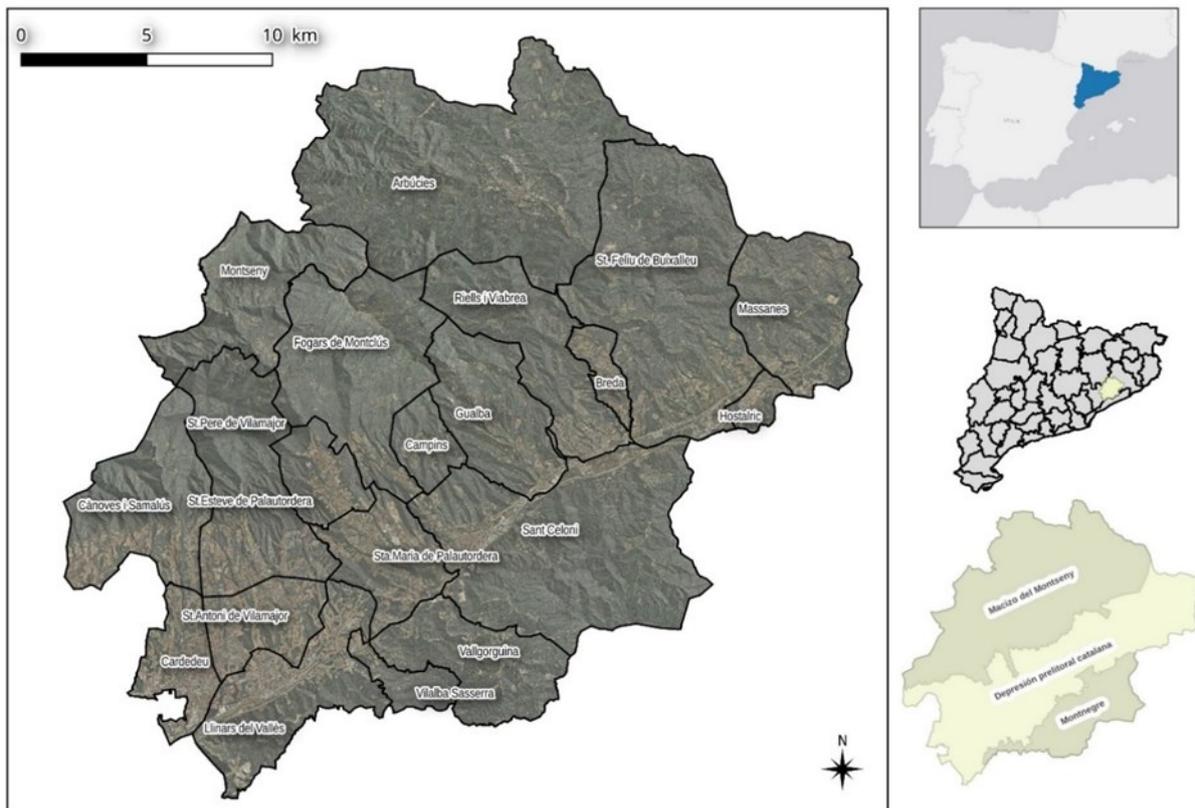


Figura 1. Ortofotomapa con la localización del Baix Montseny y su división municipal. Fuente: elaboración propia.

³ Accesible desde: <https://en.unesco.org/biosphere/eu-na/montseny>.

Aunque la comarca tiene una ocupación ininterrumpida desde el neolítico hasta nuestros días, en este estudio, nos centraremos en el período medieval, con un énfasis especial en el siglo XI. El fin del imperio romano conllevó a una dilatada desaparición en el tiempo de la organización fiscal, del comercio a gran escala y del sistema de producción de las *villae* (Nolla, 2010, p. 33). Aún con todos los cambios, la estructura del mundo rural siguió siendo continuista tanto en las técnicas como en las especies cultivadas, destacando los cultivos de cereales, viñas y oliveras (Palol, 2004, p. 499). No disponemos de fuentes arqueológicas para el Baix Montseny que lo corroboren, pero las distintas menciones a la viña en los primeros documentos que hacen mención explícita a la región en el siglo IX serían un indicador del potencial agrícola de esta zona. Durante los siglos VI-VIII Cataluña fue dominada primero por los visigodos y luego los musulmanes, pero ninguno de ellos fue capaz de consolidar un poder real en todo el territorio. Esta inestabilidad se refleja con una gran disminución del número de yacimientos existentes en la comarca. Aun así, no podemos hablar de una despoblación total ni de un abandono de las regiones de la llanura, pues la ocupación siguió en los mismos sectores de Cànovas i Samalús, Llinars del Vallés y Sant Esteve de Palautordera.

A partir de los siglos IX y X, con la estabilización de las fronteras con la conquista carolingia, el territorio del Baix Montseny ya quedó incluido en su mayor parte dentro del condado de Barcelona y la parte oriental dentro del condado de Girona. El conjunto de movimientos bélicos y revueltas de los siglos VIII-IX habrían dejado el territorio catalán poco poblado, lo que motivó la necesidad a los nuevos condados catalanes llevar una repoblación de sus tierras. No obstante, las fuentes escritas que hacen referencia a nuestra comarca de estudio no sugieren que estemos delante de un territorio despoblado en el que hay la necesidad de repoblar (Vilaginés, 1987); todo lo contrario, parecen plasmar la realidad de una población ya existente desde inicios del siglo IX, sino anterior. Los cambios producidos en el período carolingio llevaron a configurar una etapa de crecimiento sostenido que sirvió para asentar las bases de la expansión de los siglos posteriores. Este crecimiento se ve reflejado en nuestra comarca por el aumento gradual en el número de asentamientos existentes; mientras que en siglo IX se han podido geolocalizar 21 asentamientos, en el siglo X aumenta a 53, y en el siglo XI la cifra sitúa en 92 (puede comprobarse más adelante, en la Figura 4). Es importante mencionar que ya en el siglo IX ocupación territorial supera los límites de los periodos anteriores como el romano, tanto en cantidad de asentamientos como en extensión geográfica, ya que ahora se extiende a todas las regiones de la comarca.

En líneas generales el siglo XI puede considerarse como un período en el que se finaliza la implementación del sistema feudal en toda la sociedad catalana. La acumulación de bienes por parte de las élites se acentuó a partir del primer cuarto de siglo comportó un empobrecimiento progresivo del campesinado libre, que acabó siendo sometido por la aristocracia local. En el Baix Montseny hay un par de casos que ejemplifican este proceso, como sería la construcción de los castillos del Montnegre (Sant Celoni) y del Montclús (Sant Esteve de Palautordera). Estos castillos se asientan en áreas que ya pobladas en siglos anteriores, de tal modo que los territorios circundantes a estos quedaron adscritos a las nuevas circunscripciones señoriales.

La implementación del feudalismo también supuso cambios en el paisaje rural del Baix Montseny, puesto que se produjo una recuperación y una ampliación de nuevas zonas de cultivo. Estas ampliaciones parecen estar destinadas a satisfacer los estamentos privilegiados, ya que los análisis bioarqueológicos indican un aumento de los cultivos de cereales, viñedos y olivos, los productos más demandados por la nobleza y el clero (Burjachs & Rueda, 1999, pp. 117-120; Fortó et al., 2009, pp. 263-274). Además, también se observan cambios en la ganadería, donde se detecta cierta reducción de los restos del binomio ovino-caprino y un aumento de los restos de suidos y aves de corral (Novella, 2014, p. 114; Vilaginés, 2001, p. 260).

Durante el siglo XI en Cataluña, la ocupación del territorio se caracterizó por la presencia de las masías –la unidad de producción agraria tradicional de épocas bajomedieval y moderna– mencionadas frecuentemente en la documentación; y de las primeras villas mercado como estructura intermedia entre la ciudad y el campo (Furió, 2004, p. 267; Salrach, 2004, p. 37). En el caso del Baix Montseny, encontramos las primeras menciones a masías desde el año 923⁴ y a lo largo del siglo XI ya son mencionadas con regularidad en la documentación. La distribución territorial de los distintos asentamientos del siglo XI muestra un modelo de explotación del territorio y de ocupación que quedará fijado hasta bien entrada la época contemporánea. Por contra, no hay suficientes evidencias que permitan confirmar en esta fecha la prominencia de Sant Celoni como núcleo neurálgico de la región, pues no fue hasta la llegada de los hospitalarios cuando dotaron al núcleo de dinamismo, tal y como demostraría la primera mención de un mercado en la población en el año 1157⁵.

2. METODOLOGÍA Y OBJETIVOS

En este artículo se presentan los principios teóricos y procedimentales para llevar a cabo un análisis de clústeres jerárquicos utilizando el coeficiente de distancia de Gower. Se describe la aplicación de este método en un conjunto de asentamientos medievales del siglo XI ubicados en la comarca del Baix Montseny (Cataluña, España).

El estudio tiene como objetivo principal realizar una contribución a la comprensión de la organización territorial y la explotación del territorio de la comarca durante este período histórico específico. Además, se busca evaluar la idoneidad de la metodología de análisis de clústeres jerárquicos en el estudio del territorio medieval.

Para alcanzar estos objetivos, se ha empleado una metodología que combina el análisis exhaustivo de fuentes documentales con el uso de Sistemas de Información Geográfica (SIG). Este enfoque tiene el objetivo de identificar y recopilar la información sobre los diferentes espacios habitados en el Baix Montseny entre los siglos IX y XI, así como caracterizar el espacio a su alrededor utilizando las capacidades de los SIG, incluyendo el análisis de la altitud, la pendiente, el trazado de caminos o las cuencas visuales, entre otras posibilidades (Ferrer, 2019). Los datos

⁴ La mención aparece en el documento número 189 sobre la consagración de la iglesia de Sant Quirze a Arbúcies en el año 923 (Sobrequés et al., 2003, pp. 190-191).

⁵ La primera mención al mercado de Sant Celoni aparece en un testamento de 1157 (Pons, 1974, p. 170).

recopilados de las fuentes documentales y del SIG se han volcado en una base de datos MySQL, con el fin de facilitar el análisis de los datos y la ampliación en investigaciones posteriores, tanto en lo que se refiere al marco temporal como al territorial.

2. 1. Recopilación y tratamiento de los datos

Para realizar este estudio se han utilizado diferentes fuentes documentales, cartográficas e inventarios del patrimonio arqueológico y arquitectónico. La cartografía es una fuente esencial para localizar topónimos fosilizados en el territorio, así como para extraer datos geográficos empleando Sistemas de Información Geográfica (SIG). Las referencias cartográficas más antiguas del Baix Montseny suelen aparecer en mapas poco detallados e inexactos de toda Cataluña, por lo que no nos aportan utilidad en esta investigación. El primer mapa práctico, para nuestros intereses, data de 1716 y es un atlas realizado por el conde Darnius⁶ con los diferentes corregimientos de Cataluña. Además de la cartografía antigua, también se ha utilizado cartografía actual, más exacta y variada, lo que permite aplicar diferentes algoritmos incluidos en los SIG. Se han utilizado principalmente los diferentes mapas y servicios facilitados por el *Institut Cartogràfic i Geològic de Catalunya* (ICGC)⁷. Se han empleado ortofotomapas y mapas topográficos de Cataluña en escalas, desde la escala 1:1.000 a 1:250.000 mapas topográficos y años diferentes, desde las fotografías aéreas del llamado vuelo americano de 1945-46 hasta los ortofotomapas vía satélite del 2022. Un mapa fundamental es el Modelo de Elevación de Terreno (MET), un tipo de mapa que nos proporciona la altitud exacta de una parcela en un territorio dado. El MET suele ser un requisito habitual en muchos de los algoritmos de los SIG para calcular áreas de captación de recursos, cuencas de visibilidad, etc. En este estudio hemos usado el MET proporcionado por el ICGC con una resolución de 15 x 15 m al ofrecer resultados precisos en un tiempo y tamaño adecuados. Asimismo, se han usado los mapas con el sistema hidrológico completo de Cataluña proporcionados por la *Agència Catalana de l'Aigua*⁸. Hay que tener en cuenta que estos mapas tienen un evidente problema de actualismo, es decir, al reflejar la realidad actual o un pasado próximo debemos tener claro las limitaciones que conllevan al ser utilizados para estudiar un pasado remoto como el medieval.

Por lo que se refiere a las fuentes documentales, las primeras menciones documentales referentes a lugares concretos del Baix Montseny las encontramos a partir del siglo IX. La documentación de los siglos IX al XI es relativamente escasa y, en líneas generales, de carácter administrativo y legal. Las principales tipologías de este conjunto documental son donaciones, testamentos y compraventas, un tipo de documentación útil para el estudio del paisaje cultural al describir las afrontaciones. El primer cuerpo documental utilizado han sido las transcripciones del cartulario del monasterio de Sant Cugat del Vallès (Rius, 1945) junto a su índice toponímico (Roca, 1981). El cenobio de Sant Cugat se convirtió en un gran tenedor de las tierras del Baix Montseny

⁶ Registro RM.250110, accesible desde: <http://cartotecadigital.icgc.cat>.

⁷ Accesible desde: <https://openicgc.github.io> y <https://www.icgc.cat/Descarregues>.

⁸ Accesibles desde: <https://aca.gencat.cat/ca/laigua/consulta-de-dades/descarrega-cartografica/index.html>.

durante los siglos X y XI. El segundo grupo de fuentes consultadas ha sido el diplomatario de la catedral de Barcelona empleando la recopilación con la datación de Mn. Josep Baucells (Fàbrega, 1995). La tercera fuente de documentación analizada ha sido los volúmenes originales de *Catalunya Carolíngia* (Abadal, 2007), referentes a los diplomas reales, y el quinto volumen referente al condado de Girona (Sobrequés et al., 2003).

El tercer tipo de fuentes utilizadas han sido diversos inventarios arqueológicos y arquitectónicos. El *Inventari del Patrimoni Arqueològic i Paleontològic de Catalunya* es una fuente primordial en trabajos que tratan el paisaje cultural. Se encuentra accesible en varios formatos: una versión impresa entre 1988 y 1993 en varios volúmenes⁹, una versión en fichas ya desactualizada que puede consultarse en la sede *Servei d'Arqueologia de Catalunya*, en Barcelona, y una versión telemática¹⁰ más práctica y actualizada. El segundo inventario utilizado es el *Inventari del Patrimoni Arquitectònic de Catalunya* que, al igual que el anterior, tiene una versión impresa entre 1985 y 1999, y también se encuentra accesible en línea¹¹. Aunque estos dos inventarios oficiales son una fuente muy útil, es importante tener en cuenta que la calidad de la información puede variar de una ficha a otra. El último inventario empleado son los volúmenes de *Catalunya Romànica* (Pladevall & Vigué, 1984). Se trata de una obra exhaustiva y de gran utilidad para localizar estructuras de época medieval hoy en día perdidas pero que aún podemos ubicarlas en un espacio concreto.

El contenido vaciado de las diferentes fuentes comentadas se ha volcado en una base de datos relacional, utilizando *MySQL* como gestor de base de datos y *LibreOffice* para crear el formulario (Figura 2), basada en el modelo de Unidad Topográfica (UT). Se entiende como Unidad Topográfica el indicador de una acción o situación que se puede ubicar en el tiempo y en el espacio, independientemente de la especificidad de la fuente de información y de su origen biótico, abiótico o antrópico (Mauri, 2006, p. 43). Asimismo, hemos aprovechado las funcionalidades proporcionadas por los Sistemas de Información Geográfica (SIG) para geolocalizar los asentamientos mencionados en la documentación y para obtener datos de naturaleza geográfica usando el programa *QGIS*. Estos datos de carácter geográfico nos han sido de gran utilidad para posteriormente emplearlos como variables para caracterizar el entorno que rodea cada Unidad Topográfica. El proceso de vaciado de las distintas fuentes junto con el uso del SIG, nos ha permitido geolocalizar y describir un total de 157 UT del Baix Montseny existentes entre el siglo VIII y XI.

⁹ De esta primera versión se han encontrado 11 volúmenes. Al faltar varias comarcas, todo parece indicar que nunca se terminó y que fue una tirada muy limitada, ya que no se encuentra completa en ninguna biblioteca.

¹⁰ Accesible desde: <http://invarque.cultura.gencat.cat>.

¹¹ Accesible desde: <http://invarquit.cultura.gencat.cat>.

The form is organized into several sections:

- FONT (blue):** idFont: 331, TipusFor: Documental, Nom: Cartulari de Sant Cugat, Registre: 20, Descripció: Venda que Servand i les seves germanes fan al comte Sunyer per 1000 sous.
- ESPAI (green):** idEspai: 394, ETRS89-31: 448289,4615427, Nom: Can Bruques, Municipi: Sant Pere de Vilamajor, Protecció: Titularitat Privac, Propietar: [empty]
- ACCIÓ (blue):** idAccio: 460, Conservació: [empty], Nom: Villa Brucarias, Tipologia: Lloc d'habitació, Cronologia: Alt Medieval, 941, 941. Observacions: Un altre argument que pot donar versemblança a la teoria és que brugarias prové de la paraula francesa bruyère la qual va passar de França a la Llombardia on es coneguda amb el nom de brughiera i que feia referència a terra inculta, erma, plena de plantes invasives i no fructíferes. La forma llatina de brugueria és brucaria, essent el sufix -aria una expressió de col·lectivitat. Un document de l'any 942 fa menció a la vila de brucarias en el terme actual de Vilamajor.
- CRONOLOGIA (orange):** Període: Alt Medieval (941-1000), Ple Medieval (1001-1100).
- ACTORS (grey):** idActor: 116, Nom: Sunyer, Sexe: Home, Ocupac: Comte.
- PAPER (grey):** Tipus: Compravenda, Motius: Venda que Servand i les seves germanes fan al comte Sunyer per 1000 sous, Data Inic: 30/10/941, Data F: 30/10/941.

Buttons at the bottom: Anàlisi, Primer, Anterior, Següent, Últim, Nou, Guardar, Eliminar.

Figura 2. Formulario de entrada de la UT de Can Bruques. Fuente: elaboración propia.

2. 2. Los análisis de clústeres jerárquicos

Los análisis de clústeres o de conglomerados son un conjunto de técnicas que agrupan los datos basados en su similitud y permiten la creación de una jerarquía de grupos. Para realizar un análisis de clúster es necesario previamente crear una tabla de datos que contenga la caracterización completa de los elementos que forman la muestra. La elección de las variables para la caracterización es una tarea crucial, ya que de ellas dependen las agrupaciones posteriores. El tipo de variables usadas depende de la naturaleza del objeto a estudiar, así como de la información disponible para la región y el marco temporal escogidos.

Para lograr el objetivo de generar una serie de agrupaciones sin apriorismos, es esencial no dar un peso extra a ninguna de las variables ni que existan variables fuertemente correlacionadas, pues de lo contrario se daría una importancia extra a variables muy parecidas. Como trabajamos con un conjunto de datos mixtos, es decir que incluyen tanto variables cuantitativas como cualitativas, ha sido necesario utilizar una combinación de pruebas estadísticas apropiadas para cada tipo de variable. Por ejemplo, para analizar la correlación entre una variable cuantitativa y una variable categórica se puede emplear la prueba *t* de Student o la prueba no paramétrica de Kruskal-Wallis (Barceló, 2007, pp. 205-221; Fernández, 2015, pp. 129-132), mientras que para analizar la correlación entre dos variables cuantitativas se puede utilizar la prueba de correlación de Pearson o la prueba no paramétrica de Spearman (Barceló, 2007, pp. 195-200; Fernández, 2015, pp. 88-92). Así mismo, es importante tener en cuenta que, aunque exista una correlación entre variables, esta no implica necesariamente una relación causal entre ellas.

Las técnicas de análisis de clústeres no son un método desconocido en la Arqueología y la Historia y existen numerosos trabajos que los aplican, normalmente para clasificar el material

cerámico o lítico. Sin embargo, el objeto de estudio de nuestro trabajo hace que no tenga sentido usar las variables habituales en otros estudios con análisis de conglomerados, ya que nuestro objetivo no es clasificar restos materiales. Para analizar un conjunto de asentamientos medievales junto a su espacio ecológico inmediato, necesitamos que las variables describan el asentamiento, el espacio temporal, el medio antrópico y el medio físico cercano a este; esta caracterización resultante mostrará una instantánea que refleja un momento álgido de cada asentamiento (Bertoncello, 2006, p. 59). Los escasos estudios que tratan el análisis de clústeres jerárquicos aplicados a conjuntos de yacimientos suelen basarse en un importante trabajo previo arqueológico, por lo que las estructuras y los restos materiales son una parte fundamental en la caracterización de sus fichas y métodos (Fiches, 2006). Nuestro enfoque basado en las unidades topográficas supone que muchas veces no disponemos de yacimientos ni de restos materiales, por lo que estas experiencias previas no nos resultan del todo útiles. Por ejemplo, no podemos usar el tamaño de los asentamientos como variable pues, ya sea porque solo tenemos una mención documental o por la falta de intervenciones arqueológicas, no se ha podido cuantificar en muchos casos. Las peculiaridades de nuestra muestra de datos hacen indispensable diseñar una caracterización adaptada a nuestra información disponible y a nuestro objeto de estudio.

A continuación, explicaremos brevemente las distintas variables usadas para caracterizar cada asentamiento. Algunas de las variables se basan en metodologías geoestadísticas avanzadas, como el cálculo de cuencas visuales o la delimitación del Área de Captación de Recursos (ACR); sin embargo, no entraremos en la explicación detallada sobre la obtención de cada una de las variables (Ferrer, 2019), ya que nos apartaríamos del objetivo de este artículo. El mapa de la Figura 3 muestra, a modo de ejemplo, la distribución territorial de las distintas unidades topográficas existentes en el siglo XI junto algunas variables usadas como su Función, los principales cursos fluviales, los recursos minerales y las tres grandes vías de comunicación medievales de la región.

- **Área de Captación de Recursos (ACR):** variable cuantitativa en kilómetros cuadrados que mide la superficie total de la ACR calculada con el algoritmo *r.walk*¹². La superficie delimitada equivale al espacio que se puede recorrer a pie en una hora. A mayor superficie mayor será la capacidad teórica de influencia o control en su espacio cercano.

- **Preponderancia Topográfica:** variable cuantitativa continua con un valor entre 0 y 1 que mide la altitud relativa de un asentamiento respecto a su entorno inmediato. Esta variable sirve para evaluar el papel de la altitud en la organización del territorio, ya sea para evaluar posibles estrategias defensivas o de accesibilidad de un asentamiento.

- **Distancia a la fuente de agua:** variable cuantitativa en metros que indica la distancia hasta el recurso hidrológico principal o secundario más cercano del

¹² Documentación del algoritmo *r.walk* accesible desde: <https://grass.osgeo.org/grass82/manuals/r.walk.html>.

asentamiento.

- **Tierra cultivable:** variable cuantitativa en porcentaje. Indica el porcentaje solo de aquel terreno que sería ideal para la agricultura, aquel con un pendiente igual o inferior a los 7° (García, 2005, p. 220; Sheng, 1992, p. 171), dentro del espacio que se puede recorrer a pie en quince minutos.

- **Minerales:** variable cuantitativa discreta que contabiliza el número de minas y/o de filones existentes cercanos dentro de la ACR de cada asentamiento. Junto con la variable de Tierra cultivable ayuda a determinar un posible modelo de explotación del territorio.

- **Visibilidad:** variable cuantitativa en kilómetros cuadrados con la superficie de la cuenca visual, mapa con las zonas visibles y las que no, de cada asentamiento en un radio de diez kilómetros. Esta variable se ha calculado con el *plugin* QGIS Visibility Analysis¹³ (Cuckovic, 2016).

- **Función:** variable cualitativa de tipo nominal que contiene una clasificación tipológica general según la función del asentamiento. Los valores utilizados son: Religioso, Militar, Agropecuario, Obra Pública, Poblamiento, Producción y Almacenaje.

- **Menciones Documentales:** variable cuantitativa discreta que contabiliza el número de menciones documentales de un asentamiento entre los siglos IX y XI. Un mayor número de menciones indicaría una posible mayor relevancia.

- **Distancia a la red viaria:** variable cuantitativa en metros que contiene la distancia hasta la vía de comunicación principal más cercana existente en el siglo XI. Es un indicador de la capacidad de comunicación a media distancia de un asentamiento. El principal eje que cruzaría la región en este período sería la *strata francisca*, la antigua *Vía Augusta*, que transcurriría en horizontal por la depresión prelitoral catalana. También hay dos caminos principales más y que cruzan en perpendicular la *strata francisca*, el camino del Congost en el extremo este de la comarca y el eje Vic-Blanes en el oeste.

- **Relaciones:** variable cuantitativa discreta que contabiliza el número de asentamientos existentes dentro del ACR de cada yacimiento. Indicaría la capacidad de comunicación en el ámbito local y nos sirve, junto a las menciones documentales, como indicador de relevancia.

- **Ocupación anterior:** variable binaria que indica la presencia o ausencia de fases ocupacionales anteriores en cada ubicación, utilizada para evaluar la continuidad histórica del asentamiento a lo largo del tiempo.

- **Abandono:** variable cualitativa de tipo ordinal con el período histórico en el que se abandona el asentamiento, aunque su función haya cambiado a lo largo del tiempo. Sirve

¹³ Accesible desde: <https://landscapearchaeology.org/qgis-visibility-analysis/>.

como indicador del grado de éxito a lo largo del tiempo de un asentamiento.

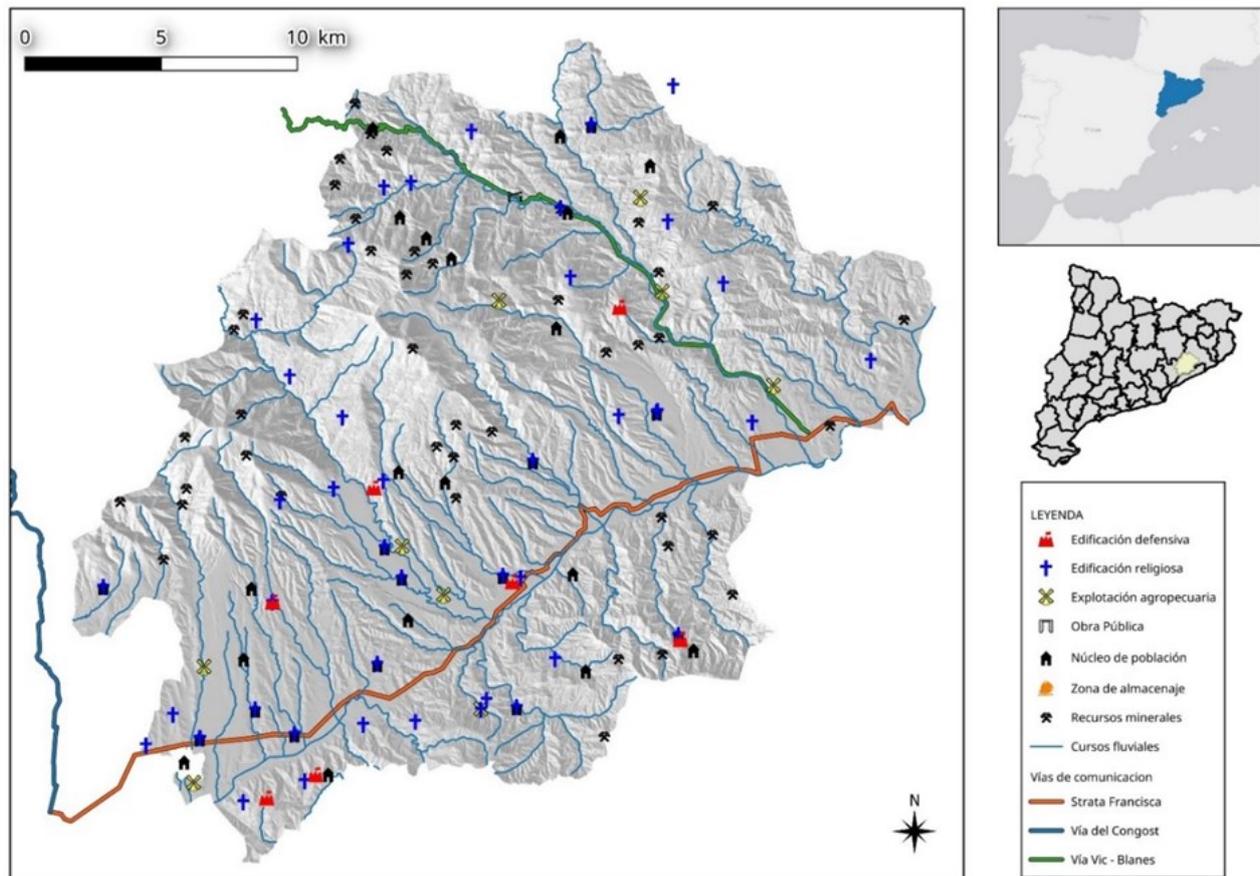


Figura 3. Mapa con las unidades topográficas existentes en el siglo XI según la variable Función, los cursos fluviales, los recursos minerales y las principales vías de comunicación de la región. Fuente: elaboración propia.

El paso siguiente es realizar un análisis multivariante de la tabla de datos con la finalidad de clarificar la estructura de la tabla y de disponer de la información esencial para clasificar los elementos, y así evitar formular un modelo causal que condicione previamente la interpretación histórica (Batista & Sureda, 1987, p. 172; Navarro, 1983, pp. 33-34). Los estudios de análisis de clústeres que trabajan con restos materiales suelen utilizar variables de tipo cuantitativo y, en cierta manera, estandarizadas ya que para ellos el método recomendado suele ser el Análisis de Correspondencias Múltiples (ACM). Este es una versión del análisis factorial de correspondencias, o simplemente análisis de correspondencias, que permite manejar un conjunto grande de variables. El ACM es una técnica multivariable de reducción de datos que tiene como objetivo extraer los principales factores de una tabla de contingencia de grandes dimensiones¹⁴ para luego representarlos gráficamente en una nube de puntos de dos o tres dimensiones (Benzecri, 1973; Greenacre, 2008, p. 193).

La muestra de nuestro estudio está formada por variables que definen un conjunto de asentamientos y su espacio cercano, por lo tanto, es una muestra con variables cualitativas y variables cuantitativas con unidades muy diferentes. El Análisis de Correspondencias Múltiples

¹⁴ El objetivo es encontrar los valores propios (*eigenvalues*) y los vectores propios (*eigenvectors*) de una matriz que contiene las distancias de todas las filas o columnas utilizando la prueba χ^2 al cuadrado. Los *eigenvalues* serán las varianzas de las nuevas variables no correlacionadas, mientras que los *eigenvectors* son los valores que multiplican las antiguas variables para obtener las nuevas. El resultado será una reducción de una tabla de n dimensiones a dos o tres dimensiones (Benzecri, 1973; Fernández, 2015, p. 155; Hammer, 2022, p. 87).

puede trabajar con datos mixtos si se realiza una serie de transformaciones previas al conjunto de datos¹⁵, pero este proceso, junto a la aplicación del propio ACM, conlleva destruir parte de la información contenida en la variable original. Esta pérdida de información supone una inevitable aparición de anomalías en los resultados que pueden afectar la posterior interpretación histórica. En muestras que contengan datos mixtos resulta más idóneo utilizar otros análisis multivariantes que no supongan una transformación de las variables ni una pérdida de la información, como es el caso de la matriz de disimilitudes utilizando el coeficiente de Gower. Una matriz de disimilitudes es una tabla¹⁶ que contiene los valores de la distancia¹⁷ entre cada par de observaciones en un conjunto de datos (Everitt & Hothorn, 2011, p. 166). El coeficiente de distancia de Gower (1971) es una prueba no paramétrica que calcula la distancia entre las observaciones de cada variable¹⁸ y después son escaladas con su media para obtener un valor entre 0 y 1. El coeficiente de Gower también nos permite ponderar las distintas variables para dar más o menos peso en las distintas variables, aunque como nuestra idea es hacer agrupaciones sin apriorismos no hemos usado esta posibilidad. El resultado final de aplicar este estadístico muestral es una tabla en la que cada objeto, en este caso asentamientos, se compara numéricamente con el resto con un valor entre 0, los objetos son iguales, y 1, los objetos son completamente distintos.

El tercer y último paso en el análisis de clústeres consiste en representar gráficamente la información de los resultados obtenidos en el paso anterior. Hay varios métodos clasificatorios disponibles, como el análisis factorial, el análisis discriminante o el análisis de conglomerados. Dentro de los análisis de conglomerados, se utilizan diferentes métodos, como la clasificación piramidal (Capdevila et al., 1995), la agrupación vecinal (Hammer, 2022, p. 94) o el clúster jerárquico (Cuadras, 2014, pp. 187-210). En este estudio, utilizaremos el método del clúster jerárquico, que es el más común. Una ventaja de este método en comparación con otros es que no implica una reducción de las dimensiones totales de la muestra.

Al realizar un análisis de clústeres jerárquico tenemos que escoger una medida que cuantifique la distancia entre los casos individualmente, así como una métrica o criterio de similitud que exprese numéricamente la semejanza entre unos y otros grupos. Probablemente, este es el principal inconveniente, ya que existen distintas distancias y métricas por lo que escoger una u otra comportará obtener diferentes agrupaciones. Además, como suele ser habitual en estadística, las diferencias se hacen más evidentes cuanto más pequeña sea la muestra con la que trabajemos. Es importante tener en cuenta que no hay una distancia o métrica mejor que otra, y la elección dependerá del conjunto de datos específico con el que se esté trabajando. La distancia más utilizada en datos cuantitativos es la euclidiana. Como el conjunto de datos de este estudio es

¹⁵ El proceso consiste en realizar una discretización de todas las variables, lo que convierte las variables cuantitativas en variables de tipo cualitativo ordinal, y luego convertir la matriz resultante en una tabla de datos binarios.

¹⁶ La matriz de disimilitudes se representa con una tabla cuadrada, simétrica y con ceros en la diagonal principal.

¹⁷ En estadística la distancia es una medida que cuantifica la similitud o la diferencia entre los individuos de un conjunto de datos.

¹⁸ En las variables cuantitativas se procede a una estandarización de los datos para trabajar con unidades distintas. Las variables cualitativas ordinales se ordenan en valores numéricos y luego se calcula la distancia.

mixto, emplearemos como distancia la matriz resultante de aplicar la distancia de Gower. De las métricas existentes, las que ofrecen unos resultados más «extremos» son las de enlace simple y de enlace completo¹⁹, entre medio de estas dos existen otras métricas como la del enlace medio o la de Ward (Everitt & Hothorn, 2011, p. 167). Nos hemos decidido a emplear la métrica de Ward²⁰, basada en la mediana, ya que resulta más diferenciadora al definir los distintos niveles de agrupación sin llegar a distorsionar los datos originales (Ward, 1963). El resultado final será un dendrograma (diagrama-árbol) que contendrá las diferentes observaciones en agrupaciones ordenadas según el nivel de similitud (Batista & Sureda, 1987, p. 182). Las agrupaciones no siempre son perfectas y es habitual encontrarse con algún miembro con valores atípicos respecto al conjunto del grupo. Si se detecta una gran cantidad de miembros discordantes, suele ser un indicativo de que las variables seleccionadas no son las más adecuadas para nuestro conjunto de datos y, por lo tanto, sería necesario replantear el análisis.

Para realizar una correcta interpretación del dendrograma, es importante determinar el número total de agrupaciones que tiene un clúster. A nivel teórico es una cuestión que no está completamente resuelta, por lo que existen diversas formas para determinarlo: desde la elección del número por parte del propio investigador o, métodos más formales, como los basados en la representación gráfica como el método *Dindex* (Charrad et al., 2014, p. 14) o las distintas técnicas basadas en índices estadísticos. Existen multitud de índices, algunos de los más comunes son el índice de *Krzanowski-Lai* (KL), *Silhouette*, *Gap*, *Calinski-Harabasz* (CH) o el de *Hartigan* (Baidari & Patil, 2020; Charrad et al., 2014, pp. 10-18). Cada índice resulta más o menos adecuado según como sea el conjunto de datos, por lo que se recomienda siempre utilizar varios índices y comparar los resultados para determinar el número óptimo de agrupaciones. Para este estudio hemos computado los cinco índices mencionados y el método gráfico de *Dindex* para determinar la media del número de agrupaciones adecuado para el dendrograma.

Los procedimientos matemáticos necesarios para realizar el análisis de clústeres jerárquicos son extremadamente complejos y laboriosos como para realizarlos manualmente en un conjunto de datos mínimamente complejos. De ahí surge la necesidad de la utilización de un lenguaje de programación estadística, como R (R Core Team, 2021) que nos ayude llevar a cabo todo este proceso matemático de manera fiable y rápida. Hemos creado diversos *scripts* en R utilizando el paquete de funciones *daisy*²¹ para calcular la matriz de distancias con el coeficiente de Gower, *NbClust*²² para determinar el número de agrupaciones, *hclust*²³ para efectuar el análisis de

Los datos cualitativos nominales se convierten en binarios, y en las variables binarias se calcula la distancia a partir de una tabla de validación.

¹⁹ El método de enlace simple, o mínimo, busca la distancia mínima entre los grupos de un clúster. El método de enlace completo, o máximo, busca la distancia máxima entre los grupos de un clúster. El del enlace medio busca la distancia media entre los grupos de un clúster.

²⁰ El criterio de varianza mínima de Ward consiste en unir en cada paso los dos clústeres que tengan el menor incremento en el valor total de la suma de los cuadrados de las distancias, dentro de cada clúster y de cada individuo del centroide del clúster.

²¹ Documentación del paquete *daisy*: <https://www.rdocumentation.org/packages/cluster/versions/2.0.7-1/topics/daisy>

²² Documentación del paquete *Nbclust*: <https://www.rdocumentation.org/packages/NbClust/versions/3.0.1/topics/NbClust>

²³ Documentación del paquete *hclus*: <https://www.rdocumentation.org/packages/stats/versions/3.5.1/topics/hclus>

clústeres y *fviz_dend*²⁴ para la visualización gráfica en forma de dendrograma.

Una vez completado todo el proceso, hemos procedido a integrar la información de las distintas agrupaciones en un programa de Sistema de Información Geográfica (SIG), para obtener un mapa con la distribución territorial de las distintas agrupaciones. Finalmente, procederemos a la interpretación de los resultados del análisis de clústeres jerárquicos y del mapa con el objetivo de buscar una interpretación histórica que explique las posibles estrategias de asentamiento y explotación del territorio. La premisa del análisis de los datos será encontrar aquello que tienen en común los miembros de un mismo grupo y aquello que les diferencia como grupo del resto de agrupaciones. Para lograrlo, nos fijaremos en las medias de las variables de cada grupo y las compararemos con las del resto.

3. RESULTADOS

El estudio de las diversas fuentes y la utilización de tecnologías SIG han permitido geolocalizar y describir 157 unidades topográficas en el Baix Montseny comprendidas entre los siglos VIII y XI. Durante este período, se puede notar claramente una tendencia ascendente en el número de asentamientos: en el siglo VIII se han registrado 6, en el siglo IX se han identificado 21, en el siglo X se han contabilizado 53 y en el siglo XI se han localizado 92 (Figura 4). El salto tan significativo en el número de asentamientos identificados entre el siglo VIII y el siglo IX se debe a que con el dominio carolingio aparecen las primeras menciones documentales a lugares específicos de la comarca. Los siglos X y XI sigue la tendencia de conseguir una mayor ocupación y explotación de toda la comarca, tanto de la llanura como de la montaña, lo que indicaría una articulación del paisaje cultural similar a la encontrada en siglos posteriores antes de la industrialización (Boada, 2002, pp. 44-54; Portals, 2019, pp. 171-173).

²⁴ Documentación del paquete *fviz_dend*: https://www.rdocumentation.org/packages/factoextra/versions/1.0.7/topics/fviz_dend.

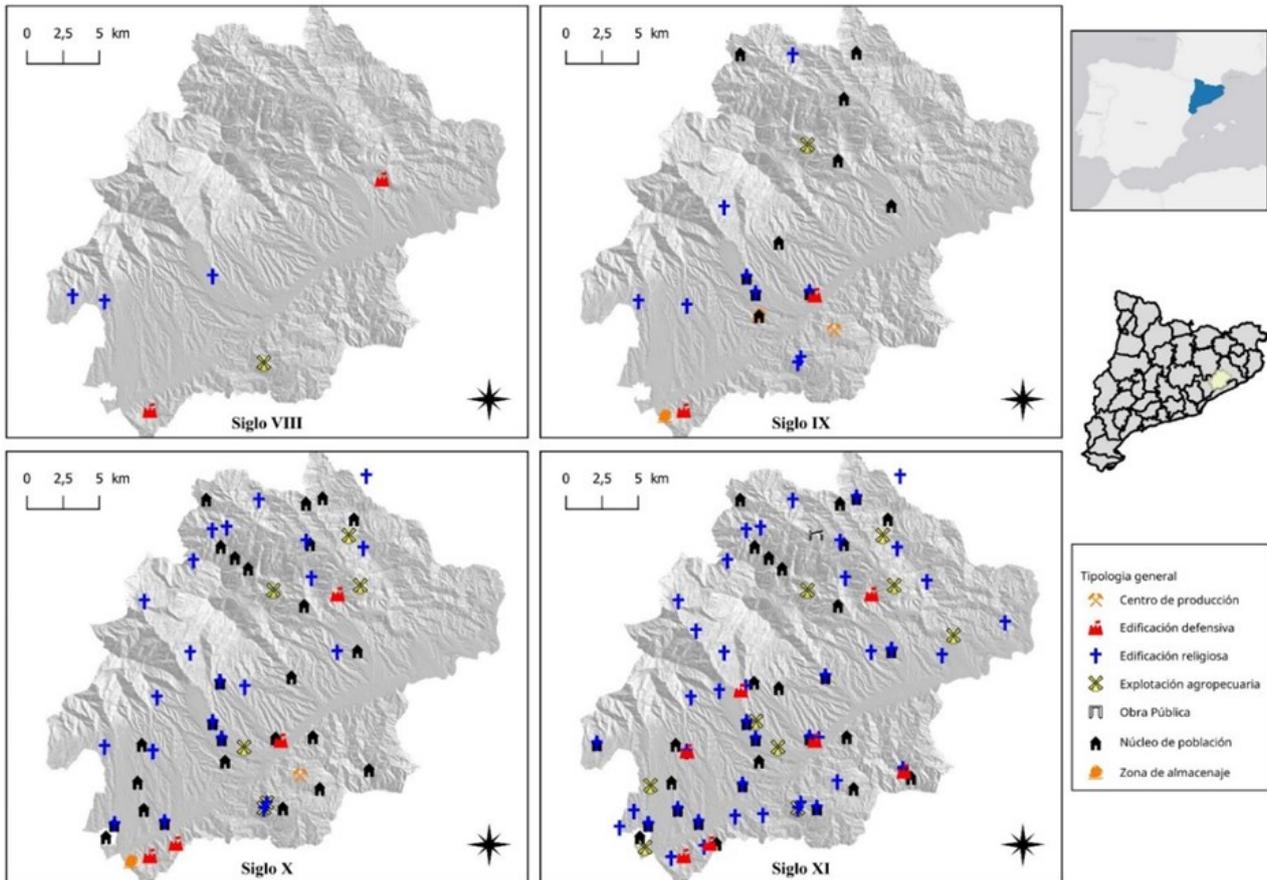


Figura 4. Evolución de los asentamientos entre los siglos VIII y XI en el Baix Montseny. Fuente: elaboración propia.

El análisis de clústeres jerárquico se ha realizado a partir de una tabla de datos con las variables expuestas anteriormente para todos los asentamientos del siglo XI. Se ha calculado una matriz de distancias usando el coeficiente de Gower con el paquete de R *daisy*. A continuación, hemos determinado el número óptimo de agrupaciones utilizando el paquete *NbClust*, obteniendo un valor promedio de seis agrupaciones, por lo que procederemos a recortar el dendrograma en seis grupos distintos. El tercer paso ha sido realizar el clúster en sí mismo con el método de *Ward.D2* empleando el paquete *hclust*. El resultado de este proceso es el dendrograma que hay a continuación con la clasificación de los distintos asentamientos (Figura 5).

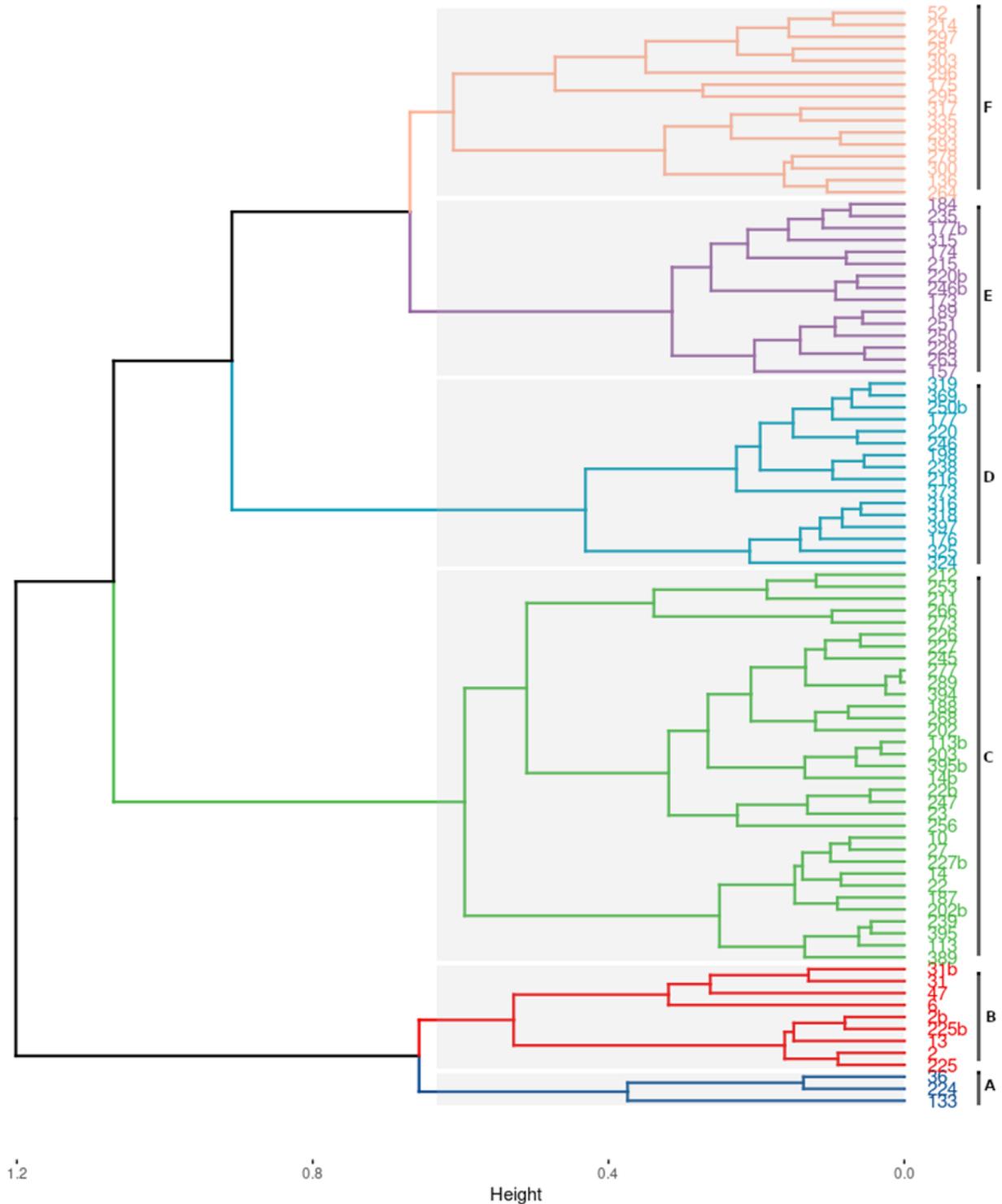


Figura 5. Dendrograma con los clústeres de los asentamientos del Baix Montseny en el siglo XI. Fuente: elaboración propia.

También se ha añadido en el SIG la información de las distintas agrupaciones con el objetivo de obtener información sobre la distribución espacial de estos grupos en la región del estudio. El mapa siguiente, representado en la Figura 6, proporciona una visualización de los grupos en el espacio²⁵.

²⁵ Debido a las dimensiones de la región y la proximidad de ciertas ubicaciones, se producen algunas superposiciones de los puntos, lo que implica que algunos no son visibles en los mapas regionales fuera del SIG.

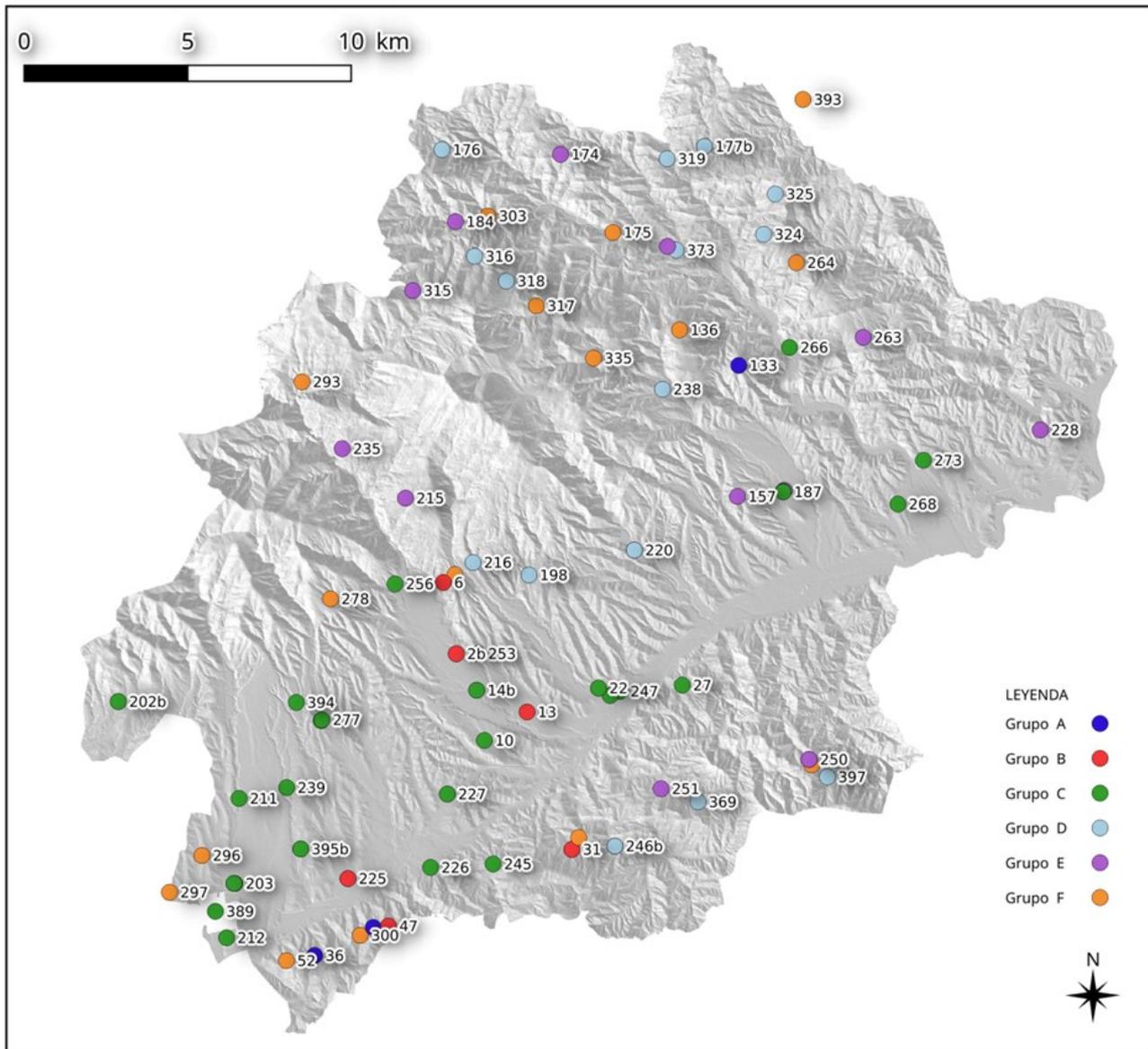


Figura 6. Asentamientos del Baix Montseny en el siglo XI según las agrupaciones del análisis de clústeres jerárquico. Fuente: elaboración propia.

Una vez obtenida la clasificación nos hemos fijado en la composición de los grupos del clúster para encontrar aquellas variables que definen a cada agrupación y determinar la coherencia interna de los grupos. Para ellos hemos calculado para cada grupo la media, la mediana y la desviación estándar de cada variable (Tablas 1 y 2).

	Menciones documentales	Relaciones	Abandono	Ocupación anterior	Función	Distancia red viaria
Grupo A	2,3	7,7	Contemporánea	Si	Militar	1743 m
Grupo B	2,9	7,8	Contemporánea	Si	Diversa	2597 m
Grupo C	2,3	5,6	Actualidad	No	Diversa	1924 m
Grupo D	1,9	3,2	Actualidad	No	Poblamiento	3109 m
Grupo E	2,1	2,2	Actualidad	No	Religioso	3507 m
Grupo F	1,3	4,6	Moderna	No	Religioso	6316 m

Tabla 1. Tablas con las medias de las variables de las distintas agrupaciones del dendrograma. Fuente: elaboración propia.

	ACR	Minerales	Distancia fuente de agua	Tierra cultivable	Preponderancia Topográfica	Visibilidad
Grupo A	31 km ²	1,6	385 m	10%	0,92	87 km ²
Grupo B	27 km ²	0	170 m	53%	0,60	28 km ²
Grupo C	29 km ²	0,3	261 m	62%	0,60	30 km ²
Grupo D	20 km ²	1,7	383 m	12%	0,64	20 km ²
Grupo E	21 km ²	1,1	370 m	12%	0,62	21 km ²
Grupo F	23 km ²	1,2	352 m	14%	0,70	31 km ²

Tabla 2. Tablas con las medias de las variables de las distintas agrupaciones del dendrograma. Fuente: elaboración propia.

El algoritmo tiende a agrupar aquellos elementos que comparten las variables que los diferencian del resto, por lo que en los grupos grandes encontramos algún elemento que, si fuera por el resto de las variables secundarias a la hora de formar la agrupación, bien podrían pertenecer a otro grupo. Sin embargo, los casos con valores extremos no son muy comunes, por lo que el algoritmo de clúster utilizando la distancia de Gower ha sido eficiente. A continuación, procedemos a interpretar los resultados valorando para cada uno de los seis grupos una posible estrategia de ocupación y explotación del territorio.

El grupo A está compuesto por tres ubicaciones (Figura 5) de las que destacamos el conocido castillo de Montsoriu (Font et al., 2016; Rueda et al., 1983), el punto 133 de la Figura 6. Este grupo tiene la media más alta en el Coeficiente de Preponderancia, la mayor superficie de Área de Captación de Recursos y la mayor superficie de Cuenca Visual²⁶ casi triplicando la media del segundo grupo con mayor superficie en esta variable. La variable Función corresponde en los tres casos a edificaciones de carácter defensivo, aunque no todas las estructuras de esta tipología se encuentran en este grupo. Al observar su distribución en el mapa, se puede ver como estos asentamientos, aunque se ubican fuera de los núcleos urbanos, se encuentran muy cerca de las principales vías de comunicación y cuentan con un gran número de asentamientos en su alrededor²⁷. Estas variables nos permiten caracterizar el grupo A como aquellos establecimientos defensivos, con una evidente vocación de control sobre un territorio extenso y sobre las principales vías de comunicación de la comarca.

El grupo B está formado por nueve asentamientos (Figura 5), destacando entre ellos los alrededores de la iglesia de Sant Esteve de Palautordera (Olivé, 2009), punto 2b de la Figura 6, y Santa María en Llinars del Vallés (Pladevall & Vigué, 1984), punto 225. Los miembros de este grupo comparten características como tener la mayor media en las variables de Relaciones y Menciones Documentales²⁸. Además, se ubicarían, normalmente, en zonas de baja altitud, muy cerca de las fuentes de agua y con un porcentaje elevado de terreno potencialmente cultivable²⁹,

²⁶ La primera agrupación tiene una media en el coeficiente de preponderancia de 0,92. La superficie media del área de captación de recursos es de 31 km² y la media de la cuenca visual es de 87 km².

²⁷ El primer grupo se ubica de media a unos 1.743 metros de las principales vías de comunicación. El promedio de la variable Relaciones es de 7,7 asentamientos por ubicación, el segundo más elevado del clúster.

²⁸ El segundo grupo tiene una media de 7,8 Relaciones. La media de la variable de Menciones Documentales es de 2,9 menciones por cada ubicación.

²⁹ El Coeficiente de Preponderancia medio es de 0,6. La distancia media hasta los principales ríos y rieras

estas tres variables indican, por lo tanto, un evidente interés agrícola. Al analizar su localización sobre el mapa, vemos como se encuentran en el sector de la depresión prelitoral catalana entre Llinars del Vallès y Palautordera. La variable de función resulta muy diversa, por lo que no ha sido clave para establecer esta agrupación. Así pues, podemos caracterizar este grupo como aquellos núcleos de población, con fases de ocupación anteriores al periodo medieval, situados cerca del llano, muy bien documentados y comunicados y con una producción económica enfocada a la agricultura. Es interesante ver que el algoritmo ha agrupado en la misma rama el grupo A y B, unos asentamientos muy diferentes, pero que tienen en común tener las mayores medias en Relaciones y tener fases de ocupación anteriores a la época medieval.

El grupo C, conformado por treinta y tres asentamientos (Figura 5), resaltamos el Molí de Can Maspons en Cardedeu (Comas, 1982), punto 211 de la Figura 6, la masía de Can Domènec en Sant Feliu de Buixalleu (Vila & Sagarra, 2002), punto 273, la iglesia de Sant Andreu de Samalús, punto 202b, y la de Santa Maria de Cardedeu (Pladevall & Vigué, 1984), punto 203. La agrupación se caracteriza por tener la segunda mayor superficie de Área de Captación de Recursos, por estar muy cerca de los recursos hídricos y por situarse en las cotas bajas³⁰, con todo esto no es de extrañar que sea el grupo con el mayor porcentaje de Terreno Cultivable³¹. La variable de Función es variada, aunque las más habituales son las de hábitat, los edificios religiosos y, en menor medida, las explotaciones agropecuarias. Como se puede ver en el mapa, principalmente se ubican a lo largo de la depresión prelitoral catalana. Podemos considerar los miembros del grupo como aquellos núcleos de población y parroquias de fundación medieval, situados a lo largo de toda la depresión del Vallès y destinados a la explotación agrícola.

El grupo D está formado por dieciséis ubicaciones (Figura 5), y contiene pequeños núcleos de población como Riells (Pérez, 2003), punto 238 de la Figura 6 y masías como el Boidaxó en Arbúcies (Pladevall & Vigué, 1984), punto 325. Las variables con información geográfica indican que los miembros de este grupo se ubican en sectores de relieve complejo y alejados de los recursos hídricos³², pero con una destacada presencia de recursos minerales a su alrededor³³. La distribución del grupo en el mapa de la comarca muestra como la mayoría se encuentran en las zonas montañosas del Montseny y del Montnegre, lo que concuerda con lo anterior. La agrupación tiene una tipología general mayoritaria de hábitat, pero resulta ser de las peor documentadas y de las más aisladas respecto a otros asentamientos³⁴. El porcentaje de terreno cultivable no ha sido una variable clave para la agrupación, aunque la media es baja existe bastante

es de 170 metros. Es el segundo grupo con un mayor porcentaje de Terreno Cultivable, un 53%, aunque dos de sus miembros están muy por debajo, lo que afecta a la media al ser grupo pequeño.

³⁰ La tercera agrupación tiene una media de la superficie del Área de Captación de Recursos es de 29 km². La distancia media hasta los principales ríos y rieras es de 261 metros. El Coeficiente de Preponderancia medio es de 0,6, el más pequeño de la muestra.

³¹ Este tercer grupo tiene la mayor media de porcentaje de Terreno Cultivable con un 62%

³² El cuarto grupo tiene una media de superficie del ACR es de 20 km², la más baja de toda la muestra. La distancia media hasta los recursos hídricos es de 383 m. siendo el segundo grupo más alejado de estos.

³³ El número de Recursos Minerales es de 1,7 por ubicación lo que supone la media más alta de todos los grupos.

³⁴ El grupo tiene una media en la variable Relaciones de 3,2 la segunda más baja de todos los grupos.

variabilidad³⁵ entre sus miembros. Se observan cinco casos, como los núcleos de Arbúcies y Gualba, ubicados en pequeños llanos y con un porcentaje de suelo agrícola que duplica o triplica la media del grupo, por lo que el desarrollo de esta actividad sería posible. La caracterización sería la de asentamientos en zonas montañosas, de origen medieval, poco documentados en este periodo y con una explotación del territorio diversificada, los pequeños núcleos tendrían cierta capacidad agrícola y los asentamientos de carácter disperso tendrían una clara vocación hacia la explotación forestal y ganadera.

El grupo E contiene quince ubicaciones (Figura 5), destacando entre ellas la iglesia de Sant Martí de Montnegre en Sant Celoni (Pladevall & Vigué, 1984), punto 250, o la ermita de Sant Marçal del Montseny (Pou, 2014), punto 315. Las variables que caracterizarían la agrupación son la de Relaciones, ya que es el grupo con un menor número de relaciones con el resto de los asentamientos³⁶ y la variable *Función* es muy homogénea pues todos sus miembros son edificios religiosos. También se observa como es el segundo grupo más alejado de las vías de comunicación³⁷. Como se puede ver en el mapa y por las variables geográficas, principalmente se ubican en cotas altas y en terrenos escarpados³⁸. A pesar del teórico aislamiento, tal y como se percibe por la variable *Abandono*, la mayoría de los miembros del grupo han mantenido su uso como espacio de culto hasta la actualidad o al menos hasta las primeras décadas del siglo XX. Como definición, podemos decir que esta quinta agrupación estaría compuesta por aquellas iglesias y ermitas de montaña que han llegado hasta nuestros tiempos y que brindaban servicio a las pequeñas comunidades de su alrededor o que servían como lugar de reclusión ascética.

El grupo F está formado por dieciséis ubicaciones (Figura 5), entre las que seleccionamos la celda de Santa Eulàlia de Tapiolas en Vallgorguina (Rodrigo, 2003), punto 31, o la capilla de Sant Cristòfol de Can Bordoï (Pladevall & Vigué, 1984), punto 300. La agrupación parece basarse principalmente en agrupar todos aquellos asentamientos de origen medieval y con una fase de abandono situada entre el período medieval y, en el mejor de los casos, a inicios del contemporáneo. Otras variables clave para establecer la agrupación son la de *Menciones documentales*, ya que es el grupo menos documentado en el período analizado³⁹, y la variable *Distancia a la red viaria* al ser el grupo a más distancia de los caminos principales⁴⁰. La variable *Función* indica que la mayoría eran edificaciones religiosas, pero también se incluye en el grupo una estructura defensiva y asentamientos de hábitat y explotación agropecuaria. Al analizar su ubicación en el mapa y las variables geográficas, se observa que se encuentran en zonas de montaña elevadas y que destacarían por tener un dominio visual relativamente bueno⁴¹. A nivel

³⁵ La media del porcentaje de Terreno Cultivable es de 12%, pero con una elevada desviación estándar, el valor mínimo de este grupo es de un 1% y el máximo de 32%

³⁶ El quinto grupo tiene un número de Relaciones medio de 2,2 asentamientos dentro del área delimitada por su ACR.

³⁷ La distancia media hasta las vías de comunicación principales del periodo medieval del grupo es de unos 3,5 km.

³⁸ El coeficiente de preponderancia del quinto grupo es de 0,62 y la superficie del ACR es de 21 km².

³⁹ El sexto grupo tiene un número medio de menciones documentales de 1,3 por ubicación.

⁴⁰ El sexto grupo tiene una distancia media hasta la red de comunicación principal es de 6.316 metros.

⁴¹ El grupo tiene la segunda media más elevada en el coeficiente de preponderancia, un 0,70. Tiene

productivo, estos asentamientos podrían estar destinados a la explotación forestal y a la ganadería⁴². Concluimos que este último grupo estaría compuesto por establecimientos de origen medieval, ubicados en zonas de montaña, mal documentados y con una fase de abandono temprana. Estas características hacen pensar que, al menos una parte del grupo, podrían ser edificios religiosos surgidos por iniciativa popular.

4. CONCLUSIONES

El análisis de clústeres jerárquicos ha demostrado ser un método eficaz en las ciencias sociales y también para ayudar a identificar diferentes patrones de estrategias de asentamiento y explotación del territorio en un conjunto de establecimientos medievales. Entre las ventajas de esta técnica se encuentra su capacidad de realizar agrupaciones sin apriorismos, basándose en las similitudes y las diferencias de todas las variables de una muestra. De este modo podemos agregar nuevas perspectivas y revelar patrones adicionales que pueden no haber sido detectados anteriormente. Además, el análisis de clústeres jerárquicos también nos sirve para seleccionar casos representativos entre todo el conjunto, así como ayudar a formular nuevas preguntas e hipótesis basadas en el análisis de los datos resultantes. Al trabajar con unidades topográficas es esencial pensar en unas variables que permitan llevar a cabo una caracterización del asentamiento y de su espacio inmediato. En este estudio, se utilizó un conjunto de variables que incluye datos históricos, arqueológicos y geográficos, cuantitativos y cualitativos, provenientes de diversas fuentes, un conjunto de variables que ha resultado útil cuando los restos materiales son limitados. El enfoque del análisis de clústeres jerárquicos basado en la métrica de Gower, ha resultado especialmente adecuado para manejar este conjunto de datos mixtos, ya que, al evitar la pérdida de información comparada con otras técnicas, permite formar clústeres de alta coherencia interna con pocos elementos discordantes.

En el Baix Montseny, la aplicación del análisis de clústeres jerárquicos en los asentamientos medievales del siglo XI ha permitido identificar seis agrupaciones con distintas estrategias de ocupación y explotación del territorio. Estas agrupaciones proporcionan una instantánea del poblamiento rural del siglo XI y muestran una clara diferenciación entre el modelo seguido en la llanura y el de las zonas de montaña. Las agrupaciones D, E y F cuelgan de una misma rama del dendrograma y describen los asentamientos en las zonas de montaña. Estas agrupaciones estarían formadas por pequeños núcleos o asentamientos de carácter disperso, tendrían un origen medieval y seguirían un modelo económico basado en la explotación forestal, ganadera y, en menor medida, agrícola. También se observa la importancia de la Iglesia en la organización de la población en estas zonas de montaña, ya que a menudo los edificios religiosos son el único indicador de la presencia humana en estas zonas. Los grupos B y C representan núcleos en las

también la segunda mayor media de Visibilidad con 31 km², aunque es casi tres veces menor que la de la primera agrupación.

⁴² El porcentaje de terreno cultivable tiene una media de 14%, aunque la mediana es tan solo de un 8%, y la distancia hacia la fuente de agua más cercana tiene una media de 352 metros.

zonas llanas alrededor de la depresión prelitoral catalana, unos asentamientos con una clara vocación agrícola tal y como se desprende de las medias de las variables geográficas. Los miembros de estos grupos estarían mejor relacionados con otros asentamientos, mejor documentados y más cercanos a las vías de comunicación; todo esto nos haría pensar que se trataban de núcleos de mayor importancia. Detectamos que el poder laico es muy presente en la depresión del Vallès, especialmente en los miembros del grupo A, pues se observa una evidente vocación de control sobre los asentamientos y vías de comunicación de este sector.

En conclusión, durante el siglo XI se identificaron diversas características en la organización territorial que perduraron como pautas generales del poblamiento tradicional hasta mediados del siglo XX. Estas características incluyen un modelo dual en la explotación económica de la comarca, la importancia de la Iglesia en la articulación de los núcleos de población y una posible mayor concentración de población en la depresión prelitoral catalana. A pesar de sus limitaciones, la metodología utilizada basada en el análisis de clústeres jerárquicos ha demostrado su eficacia, especialmente al abordar conjuntos de datos mixtos de gran magnitud, y puede facilitar la realización e interconexión de estudios sobre el paisaje histórico a una escala territorial más amplia.

REFERENCIAS BIBLIOGRÁFICAS

- Abadal, R. (2007). *Catalunya carolíngia. Volum II. Els diplomes carolingis a Catalunya*. Institut d'Estudis Catalans.
- Baidari, I., & Patil, C. (2020). A Criterion for Deciding the Number of Clusters in a Dataset Based on Data Depth. *Vietnam Journal of Computer Science*, 07(04), 417-431. <https://doi.org/10.1142/S2196888820500232>
- Barceló, J. (2007). *Introducción al estudio de la variabilidad de las evidencias arqueológicas*. Servei de publicacions Universitat Autònoma de Barcelona.
- Batista, J. M., & Sureda, J. (1987). Análisis de correspondencias y técnicas de clasificación: Su interés para la investigación en las ciencias sociales y del comportamiento. *Infancia y Aprendizaje*, 10(39-40), 171-186. <https://doi.org/10.1080/02103702.1987.10822184>
- Bellanger, L., Coulon, A., & Husi, P. (2021). Determination of cultural areas based on medieval pottery using an Original Divisive Hierarchical Clustering Method with Geographical Constraint (MapClust). *Journal of Archaeological Science*, 132, 105431. <https://doi.org/10.1016/j.jas.2021.105431>
- Benzecri, J.-P. (1973). *L'analyse des données. Tome 2: L'analyse des correspondances*. Dunod.
- Bertoncello, F. (2006). De la trama de puntos al sistema de poblamiento. En *Catastros, habitats y vía romana: Paisajes históricos de Europa* (pp. 47-70). Generalitat Valenciana.
- Boada, M. (2002). *El Montseny, cinquanta anys d'evolució dels paisatges*. L'Abadía de Montserrat.
- Burjachs, F., & Rueda, J. M. (1999). El paisatge vegetal a la Vall d'Arbúcies en època ibèrica (Can Pons) i a l'edat mitjana (Castell de Montsoriu). *Monografies del Montseny*, 14, 111-122.
- Capdevila, C., Gil, A. J., & Arcas, A. (1995). Algunas propiedades de un algoritmo de clasificación

- piramidal. *Estadística Española*, 37(138), 101-126.
- Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClus: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, 61(6). <https://doi.org/10.18637/jss.v061.i06>
- Comas, P. (1982). *Cases de Pagès. Cardedeu i rodalia*. Casal de Cultura Dr. Daurella de Cardedeu.
- Cuadras, C. (2014). *Nuevos métodos de análisis multivariante*. CMC Editions.
- Cuckovic, Z. (2016). Advanced Viewshed Analysis: A Quantum GIS Plug-In for the Analysis of Visual Landscapes. *Journal of Open Source Software*, 1(4), 32. <https://doi.org/10.21105/joss.00032>
- Esquivel, J., Peña, J., & Rodríguez-Ariza, M. O. (1999). Multivariate Statistic Analysis of the Relationship between Archaeological Sites and the Geographical Data of their Surroundings. A Quantitative Model. En *Archaeology in the Age of the Internet. CAA 97: Computer Applications and Quantitative Methods in Archaeology: Proceedings of the 25th Anniversary Conference*. University of Birmingham.
- Everitt, B., & Hothorn, T. (2011). *An Introduction to Applied Multivariate Analysis with R*. Springer New York. <https://doi.org/10.1007/978-1-4419-9650-3>
- Fàbrega, À. (1995). *Diplomatari de la Catedral de Barcelona. Documents dels anys 844-1260*. Arxiu Capítular de la Catedral de Barcelona.
- Fernández, V. (2015). *Arqueo estadística. Métodos cuantitativos en arqueología*. Alianza Editorial.
- Ferrer, M. (2019). *Arqueologia del paisatge altmedieval al Baix Montseny, segles VI-X. Una demostració pràctica dels sistemes d'informació geogràfica* [Universitat de Barcelona]. <http://hdl.handle.net/10803/666872>
- Fiches, J.-L. (2006). De la prospección a la caracterización de los yacimientos de época romana. En *Catastros, habitats y vía romana: Paisajes históricos de Europa* (pp. 19-30). Generalitat Valenciana, Conselleria de cultura, educació i esport.
- Font, G., Llorens, J. M., Mateu, J., Pujadas, S., Tura, J., & Codolà, C. (2016). *Montsoriu, el castell gòtic de Catalunya*. Museu Etnològic del Montseny.
- Fortó, A., Martínez, P., & Muñoz, V. (2009). Los yacimientos del Pla del Serrador (Les Franqueses del Vallès) y de Ca l'Estrada (Canovelles), dos ejemplos de asentamiento agrícola entre los siglos V y XIII d.C. en el entorno de Granollers (Vallès Oriental, Barcelona). En *The archaeology of early medieval villages in Europe*. Universidad del País Vasco.
- Furió, A. (2004). L'organització del territori: L'espai i el poblament. En *Història agrària dels països catalans*. Fundació Catalana per a la Recerca.
- García, L. (2005). *Introducción al reconocimiento y análisis arqueológico del territorio*. Editorial Ariel.
- Gower, J. C. (1971). A General Coefficient of Similarity and Some of Its Properties. *Biometrics*, 27(4), Article 4. <https://doi.org/10.2307/2528823>
- Greenacre, M. (2008). *La práctica del análisis de correspondencias*. Fundación BBVA.
- Hammer, Ø. (2022). *PAST. Paleontological Statistics Reference manual*. Natural History Museum, University of Oslo. <https://www.nhm.uio.no/english/research/resources/past/downloads/>

[past4manual.pdf](#)

- Makridis, M., & Daras, P. (2012). Automatic Classification of Archaeological Pottery Sherds. *Journal on Computing and Cultural Heritage*, 3(4), Article 4.
- Mauri, A. (2006). *Configuració del paisatge medieval: El comtat de Barcelona fins al segle XI*. Universitat de Barcelona.
- Navarro, M. L. (1983). Aspectos teóricos y una aplicación práctica del análisis factorial de correspondencias. *Estadística Española*, 99, 33-59.
- Nolla, J. M. (2010). La Catalunya vella: De la caiguda de l'Imperi romà a la fi del domini carolingi. Una visió des de l'arqueologia. En *Actes del IV congrés d'arqueologia Medieval i moderna a Catalunya* (pp. 27-45). Ajuntament de Tarragona.
- Novella, V. (2014). *Estudi de les pautes d'accés i consum dels recursos animals a partir de l'arqueozoologia l'exemple del Castell de Montsoriu* [Tesis doctoral Universitat Autònoma de Barcelona]. Repositorio institucional de la Universitat Autònoma de Barcelona. <http://hdl.handle.net/10803/285035>
- Olivé, M. (2009). *Intervenció a la plaça de l'església de Sant Esteve de Palautordera*. (Memòria Científica N.º 9438; Número 9438). Arxiu Servei Arqueologia i Paleontologia.
- Palol, P. (2004). Economia i Societat del Baix Imperi i L'Època Visigoda. En *Història agrària dels països catalans* (Vol. 1). Universitat de Barcelona.
- Papageorgiou, I., Baxter, M. J., & Cau, M. A. (2001). Model-based Cluster Analysis of Artefact Compositional Data. *Archeometry*, 4(43), Article 43.
- Pérez, R. (2003). *9 focs. Un tomb per la història de Riells i Viabrea*. Ajuntament de Riells i Viabrea.
- Pladevall, A., & Vigué, J. (1984). *Catalunya romànica*. Enciclopèdia Catalana.
- Pons, J. M. (1974). La successió de Guillem Umbert de Basella. *Annals de l'Institut d'Estudis Gironins*, 22, 153-176.
- Portals, J. (2019). La Costa del Montseny i Santa Susanna. Dos exemples de l'evolució de la població al Montseny des de mitjan segle XIX fins avui. En *Monografies del Montseny 34* (pp. 171-184). Amics del Montseny.
- Pou, J. (2014). Anàlisi de les esglésies que figuren en l'acta de consagració de l'any 923 de la parròquia de Sant Quirze d'Arbúcies. En *VIII Monografies del Montseny* (pp. 408-416). Diputació de Barcelona.
- Pujantell, J., Boada, M., & Varga, D. (2017). Història ambiental al massís del Montnegre: El cas d'estudi del municipi de Sant Celoni. En *III Trobada d'Estudiosos de la Serralada Litoral Central i VII del Montnegre i el Corredor: 19 i 20 de novembre de 2015* (pp. 102-117). Diputació de Barcelona.
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org>
- Rius, J. (1945). *Cartulario de Sant Cugat del Vallés*. C.S.I.C. Sección de estudios medievales de Barcelona.
- Roca, P. (1981). *Índex toponímic del Cartulari de Sant Cugat del Vallés*. Museu d'Història de

Sabadell.

- Rodrigo, E. (2003). *Memòria de l'excavació arqueològica al jaciment de Sant Genís de Tapioles. Vallgorguina*. (Memòria Científica N.º 6635). Generalitat de Catalunya.
- Rueda, J. M., Llorens, J. M., & Mataró, M. (1983). *Memòria. Castell de Montsoriu (Arbúcies – Sant Feliu de Buixalleu). Campanya 1983*. [Memòria Científica]. Generalitat de Catalunya.
- Salrach, J. M. (2004). Època prefeudal (Segles vii-x) i canvi feudal (segles xi-xii). En *Història agrària dels països catalans* (Vol. 2). Fundació Catalana per a la Recerca.
- Salvat, A., & Pié, G. (2014). Pla de conservació del Parc Natural. Reserva de la Biosfera del Montseny: Diagnosi de la flora i els hàbitats. En *VIII monografies del Montseny* (pp. 16-31). Diputació de Barcelona.
- Schwabe, R. S., & Slusallek, K. (1981). Application of the Cluster Analysis on Element Concentrations of Archaeological Bronzes, Ceramics and Glass. *Revue d'Archéométrie*, 6, Article 6.
- Sheng, T. C. (1992). *Manual de campo para la ordenacion de cuencas hidrograficas: Estudio y planificacion de cuencas hidrograficas*. Food & Agriculture Org.
- Sobrequés, S., Ordeig, R., Riera, S., & Rovira, M. (2003). *Catalunya Carolíngia. Els comtats de Girona, Besalú, Empúries i Peralada* (Vol. 5). Institut d'Estudis Catalans.
- Vila, M.-A., & Sagarra, M. (2002). *La casa rural a Catalunya. Cases aïllades i cases de poble*. Edicions 62.
- Vilaginés, J. (1987). *La transició al feudalisme. Un cas original. El Vallès Oriental*. Ajuntament de Granollers.
- Vilaginés, J. (2001). *El paisatge, la societat i l'alimentació al Vallès Oriental (segles X-XII)*. Publicacions de l'abadia de Montserrat.
- Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58, Article 58.