Check for
updates

Classification of wildfires in relation to land cover types and associated variables by applying cluster analysis: a case study in the Iberian Peninsula

Laura Serra · Pablo Juan · Carlos Díaz-Avalos · Pau Aragó · Somnath Chaudhuri · Sergio Trilles

Abstract Wildfires are a major environmental problem that have both economic and ecological impacts. Wildfires typically spread in a particular pattern, determined by factors such as the elements on the ground that catch fire or their geographic location. This study reports and discusses how wildfires in the Valencian Community, Spain, have been spatially grouped in recent years (from 2016 to 2020). It also characterizes each cluster in terms of location and land cover. An exploratory analysis of the environmental variables associated with wildfires has been conducted using finite Gaussian mixture models in R (R package mclust). The primary findings can be used to better understand the types of wildfires that occur in individual spatial zones. Some interesting cluster patterns in specific geographical areas, such as river basins, have also been reported. The method can identify clusters of fires by detecting areas with similar characteristics at the land use level. It also allows for the implementation of measures aimed at reducing the impacts of wildfires and can help in the extinction of wildfires based on the characteristics of all the fires grouped using spatial and land cover dimensions.

L. Serra (✉) · P. Juan · S. Chaudhuri
Research Group on Statistics, Econometrics and Health (GRECS), University of Girona, 17004 Girona, Spain
e-mail: laura.serra@udg.edu

P. Juan
e-mail: juan@uji.es

S. Chaudhuri
e-mail: somnath.chaudhuri@soton.ac.uk

L. Serra
Center for Biomedical Research in Epidemiology and Public Health (CIBERESP), Madrid, Spain

P. Juan
IMAC, Department of Mathematics, Statistics Area, Universitat Jaume I, 12071 Castelló de La Plana, Spain

C. Díaz-Avalos
Departamento de Probabilidad y Estadística, IIMAS, UNAM, Mexico City, México
e-mail: carlos@sigma.iimas.unam.mx

P. Aragó
Servei Territorial de Medi Ambient. Conselleria de Medi Ambient d'infraestructures en Territori. Generalitat Valenciana, 46018 Valencia, Spain
e-mail: parago@uji.es

S. Chaudhuri
WorldPop Research Group, Department of Geography and Environmental Science, University of Southampton, Southampton, UK

S. Trilles
Institute of New Imaging Technologies (INIT), Universitat Jaume I, 12071 Castelló de La Plana, Spain
e-mail: strilles@uji.es

Introduction

Wildfires are one of the environmental hazards posing significant threats to wildlife and forests worldwide. According to Quílez Moraga (2019), the area covered by forests is growing, and the number of wildfires is decreasing. However, every year in the Valencian Community in Spain, a few wildfires account for most of the forest area burned. It is worth mentioning that even though the use of fire in agriculture is controlled and forest firefighters are quite efficient, in some cases, a few fires that started under adverse weather and land use conditions gave rise to significant wildfire incidences (de Rivera et al., 2020). Many scientific studies have analyzed the occurrence of wildfires in spatial domains aiming to understand the process and physical and biological factors that influence their incidence. For example, some studies have shown a strong relationship between the incidence of forest fires and climatic and meteorological variables in European countries (Hoinka et al., 2009; Moreno et al., 2014; Tedim et al., 2013; Venäläinen et al., 2014). A similar study by Aragó et al. (2016) investigated the incidence of forest fires in the province of Castelló, Spain, to identify risk factors associated with wildfires from 2001 to 2006.

Other studies (Carmo et al., 2011; Moreira et al., 2011; Nunes et al., 2016; Sebastián-López et al., 2007) have reported that human factors commonly linked to fire ignition are forest-agricultural or forest-urban interface constraints related to land use management. Thus, the characterization of wildfire events is necessary to model wildfire occurrence and to relate it to the spatial distribution of spatial varying factors, such as demographic indicators, land cover, and the presence of agricultural areas (Koutsias et al., 2010). Díaz-Avalos et al. (2016) explored and fitted statistical models using Bayesian statistics to identify relevant factors associated with the spatial variation of wildfire sizes using spatial marked point process methods.

On the other hand, recent research has shown that cluster analysis is a successful method for detecting regularities and can be used to design hybrid predictive models (Dong et al., 2016), as well as to study

existential frequencies (Parente et al., 2016; Strauss et al., 2013). Many existing studies in the broader literature have examined the application of clusterization algorithms to explore and analyze the cause and spatial distribution of wildfire incidences. A research study (Castro et al., 2020) applied cluster analysis to review the spatial characterization of the distribution of wildfire events in mainland Portugal between 1996 and 2015. The research examined the breakdown of the causes of these fires over this period. A similar study reviewed the use of clustering techniques to assess wildfire impacts in storm activity conditions, especially lightning (Nikolay et al., 2016). Along the same line, Parente et al. (2016) proposed space–time clustering analysis techniques that correctly identify wildfire clusters in terms of their number and location despite eventual splitting of the datasets according to space and/or time.

Although some authors have conducted studies that examine spatial patterns and possible clusters in the domain of wildfire instances, this problem is still insufficiently explored. Moreover, most of the research published to date rarely analyzes the spatial distribution of wildfires concerning the type of land cover (Koutsias et al., 2010) and other meteorological parameters such as temperature and precipitation. Good knowledge of the causes of wildfires and their spatial–temporal distribution is crucial to produce reliable predictions of the number of fires and where and why they are expected to occur (Castro et al., 2020). This work aims to explore these little-studied aspects to understand the strategic measures required for preventing wildfires under the hypothesis that they are concentrated in specific areas according to land use. Accordingly, the patterns produced by wildfire incidents in the Valencian Community are analyzed with the aim of classifying wildfires based on associated covariables, such as land cover, causes, altitude, temperature, precipitation, and the area burned. The main objective is to test whether cluster analysis could be used as a tool to classify and separate wildfires based on the above mentioned covariables, to provide information useful to forest fire prevention authorities. Graphically, a specific accumulation of fires in different areas of the Valencian Community has been observed, with cluster analysis being the ideal tool to understand why those groups are formed. The methodological

framework for analyzing wildfires in the Valencian Community consists of key stages: preparation (defining objectives and identifying the study area), data collection (wildfire incidents from 2016 to 2020 and environmental covariates such as land cover, causes, and meteorological data), and data processing (ensuring cleanliness and compatibility for analysis). Subsequently, cluster analysis is applied using finite Gaussian mixture models (GMM) through the R package mclust to identify and categorize wildfire patterns. Each cluster is then characterized based on its spatial distribution, land use type, and annual temporal variations. The analysis is validated with statistical tests to ensure the reliability of results. Finally, findings are interpreted within the context of wildfire management, with suggested future directions for research that include adapting the methodology for other regions and studying wildfire causes, including methods for detecting arsonists. The current study contributes to the relatively small amount of literature on cluster analysis methods to monitor wildfires in a broader picture. The proposed methodology

is dynamic and can be effectively adapted and applied to other locations worldwide. The current analysis is based on finite GMM, implemented using the open-source R package mclust.

Materials and methods

Study site

The study area consisted of the Valencian Community, located on the eastern coast of the Iberian Peninsula. The region is bordered by Catalonia to the north, the Iberian Mountain System range to the west, and the Mediterranean Sea to the east. It is a region with a surface area of 2,324,500 ha, representing 4.6% of the Spanish national territory. The Valencian Community can be divided into a coastal zone, where fires have become a common tool for the management of agricultural biomass waste, and an inland zone, characterized by the presence of natural forest and rural areas, where the abandonment of farming areas is leading to the increase in forest land (Fig. 1).
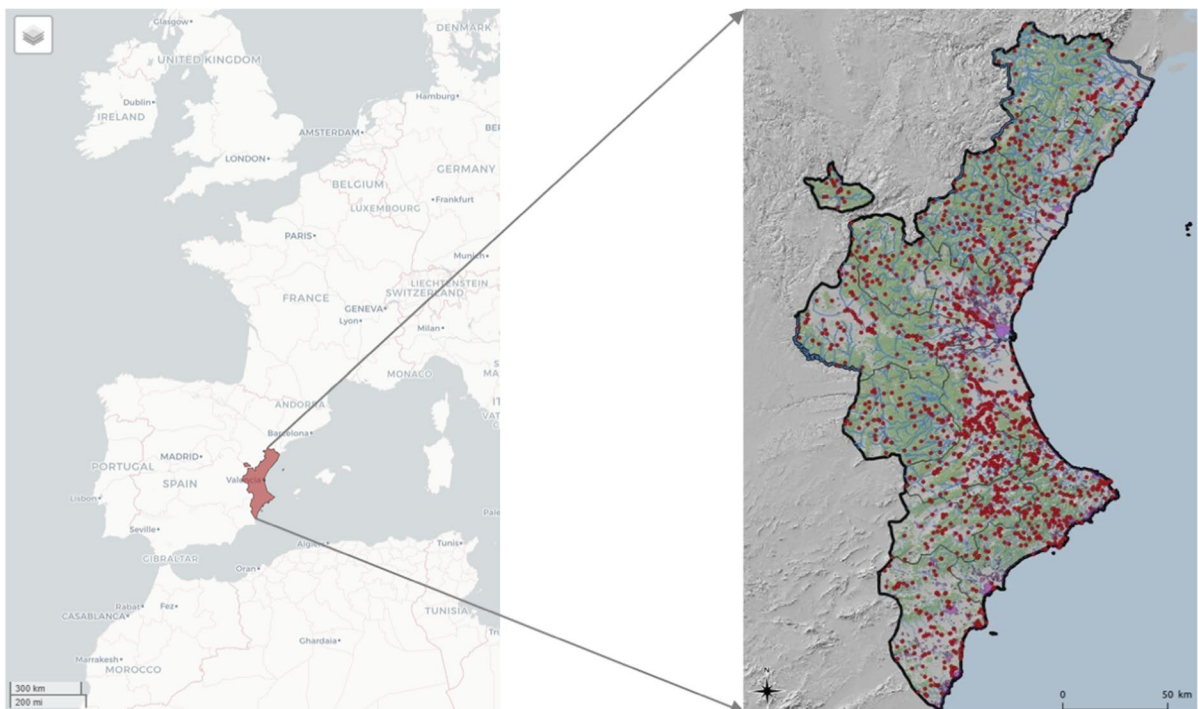

Fig. 1 Geographical location of the study area—Valencian Community, Spain—and locations of wildfires in the study area from 2016 to 2020
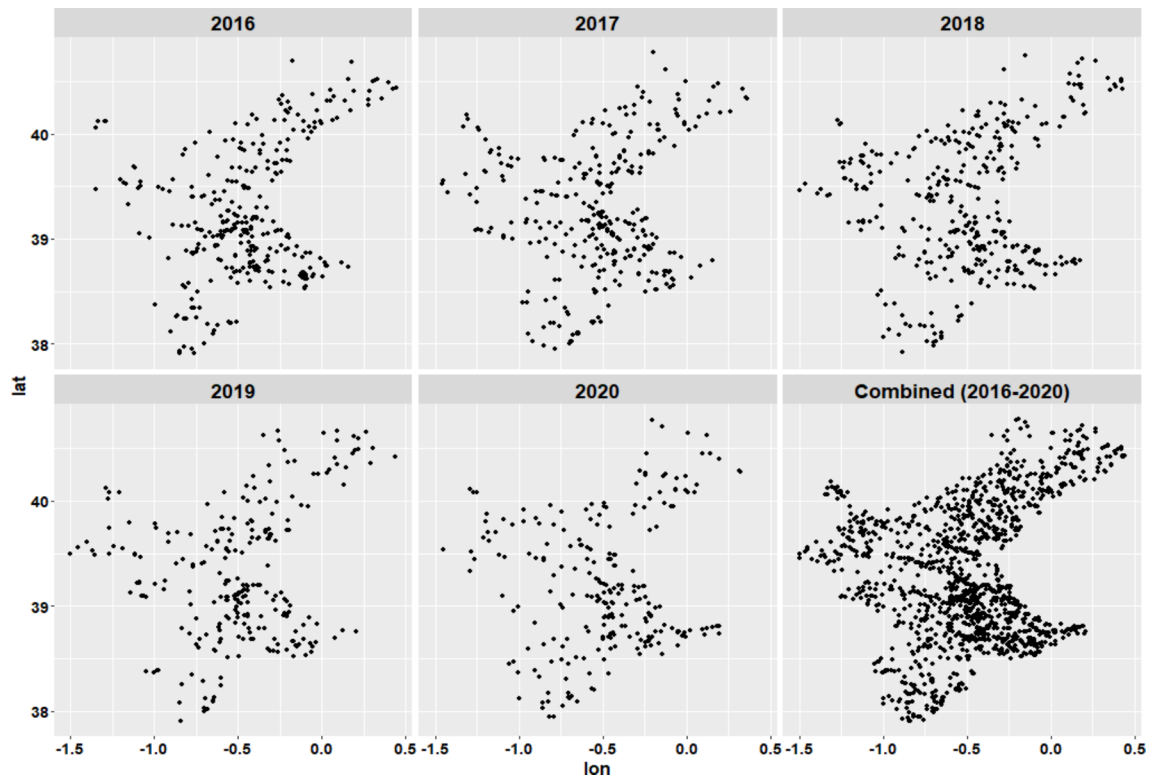
Fig. 2 Annual spatial distribution of fires from 2016 to 2020, and all 5 years together throughout the Valencian Community

Data settings

Data for this paper were obtained from the "Sistema integral de gestió d'incendis forestals" [Integrated wildfire management system] (Generalitat Valenciana, Spain). The data set includes the coordinates at which fire ignitions occurred, as well as information about the fire type, distinguishing between shrub and grassland fires, and wildfires. The elevation and meteorological variables values, such as annual average temperature and precipitation, were retrieved as raster files at a spatial resolution of 30 s using the R package *raster* (Hijmans & van Etten, 2012). These values were attached to the fire ignition data using GIS functions in R. The annual number of wildfire incidents in the study area was 341 in 2016, 346 in 2017, 375 in

Table 1 Statistics of the covariates

	Burned area (tree)	Burned area (not tree)	Burned area (total)	Altitude	Temperature (mean)	Temperature (max)	Temperature (min)	Precipitation
Minimum	0	0	0	− 2	0	0	− 2	0
Maximum	1425.33	6146	6146	1385	18.3	32.8	6.9	671
Quartile 1	0	0	0.05	66.5	14.3	28.4	2.5	433
Quartile 3	0.2	1	3	621	17.3	30.4	5.9	501
Mean	6.18	18.63	23.30	357.24	15.42	28.75	4.04	454.82
Median	0	0	0.36	261	16.1	29.4	4.6	456
Stdev	55.52	201.23	210.77	324.72	2.92	4.38	2.11	90.42

Fig. 3 Ellipses of isodensity for each of the 14 Gaussian models obtained by eigen-decomposition in the case of three groups in two dimensions, *mclust*
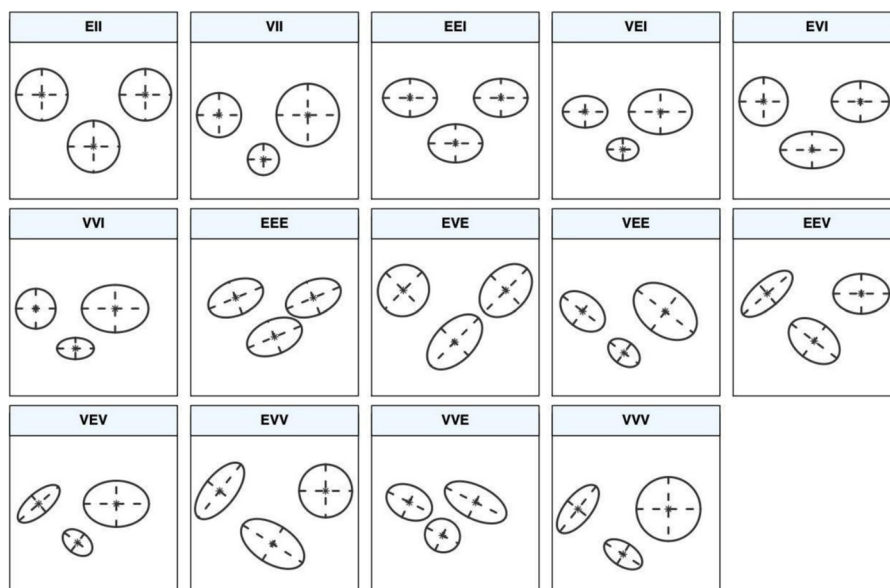
2018, 272 in 2019, and 251 in 2020. The spatial distribution of the wildfires in the data set is shown in Fig. 2. The last plot represents a total of 1587 wildfires for the entire study period (2016–2020), which will be analyzed as a whole set as well as in individual years.

The globally interpolated climate data were obtained as raster images from NASA's Shuttle Radar Topography Mission (SRTM) database (Farr et al., 2007). Table 1 reports the summary statistics of the covariates for the complete data set. It provides information about four variables: (1) the burned area, distinguishing between shrub and grassland wildfires (not trees), wildfires (trees), and the total burned area in hectares; (2) wildfire altitude; (3) the annual temperature, differentiating between its mean value, the maximum, and the minimum, all in degrees Celsius; and (4) the annual precipitation in millimeters. The statistics shown in this table are the minimum and maximum values, the first and the third quartiles, the mean, the median, the variance, and the standard deviation of each covariate.

Methodology

Background of the method

GMM is a probabilistic generative model that assumes the data is a mixture of multiple Gaussian distributions of the form:

$$p(x|\lambda) = \sum_{i=1}^{M} w_i g(x|\mu_i, \sum_i)$$

, each of which represents a latent cluster or group within the data. The mixture satisfies the restriction:

$$\sum_{i=1}^{M} w_i = 1$$

The model estimates the parameters of the Gaussian distributions, such as mean and covariance,

Table 2 Number of events by year

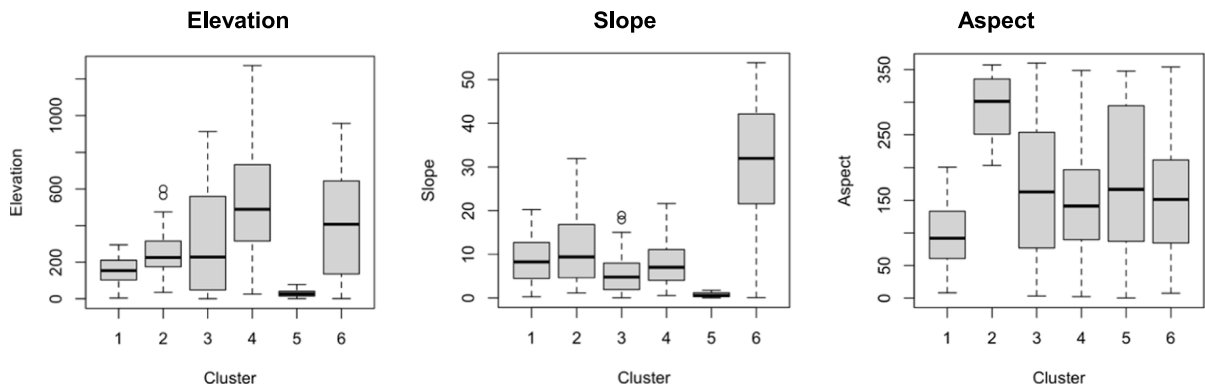| Year | Cluster | | | | | |
	1	2	3	4	5	6
2016	48	45	51	104	55	34
2017	52	138	32	89	31	-
2018	167	38	69	40	61	-
2019	60	39	46	76	31	19
2020	52	76	22	42	58	-

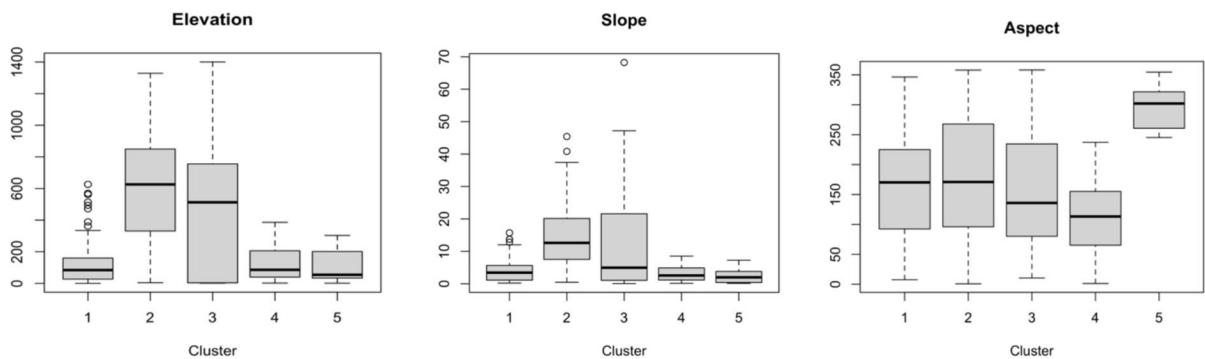**Fig. 4** Elevation, slope, and aspect by cluster, for 2016

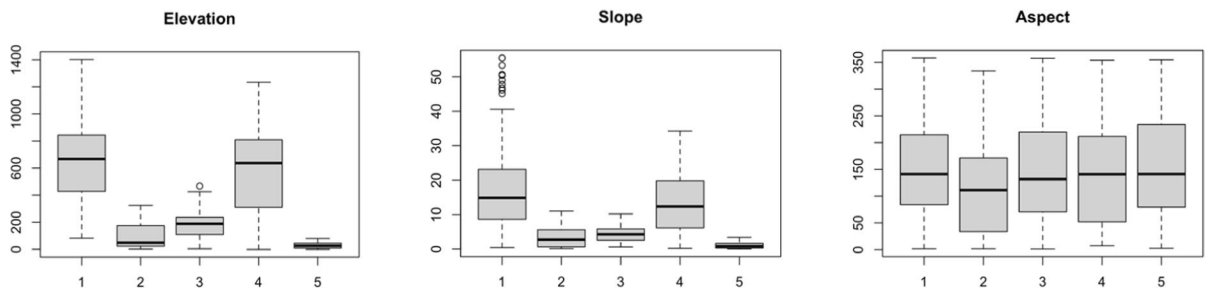**Fig. 5** Clusters 1 to 3 (elevation, slope, and aspect) for 2017

Fig. 6 Clusters 4 to 6 (elevation, slope, and aspect) for 2018

through maximum likelihood estimation or Bayesian inference (Reynolds, 2009). GMM is commonly used for cluster analysis and for other applications such as density estimation, anomaly detection, and dimensionality reduction. One of its advantages is its flexibility in modeling complex data distributions while offering a probabilistic interpretation of the results.

GMM can be used to analyze and predict the occurrence of wildfires. It can also be used to analyze the spatiotemporal patterns of wildfire occurrence, as well as the frequency, intensity, and size of wildfires (Chunyu et al., 2009; Han et al., 2017; Qian et al., 2018; Torabian et al., 2021; Yoon & Min, 2013; Zhao et al., 2011). By modeling the probability distribution of these patterns (Velizarova & Alexandrov, 2021; Zhang et al., 2017),

Fig. 7 Boxplot of annual precipitation (2016–2020) per individual cluster. From clusters 1–6 (2016), from clusters 7–11 (2017), from clusters 12–16 (2018), from clusters 17–22 (2019), and from clusters 23–27 (2020)

Fig. 8 Boxplot of annual temperature (2016–2020) per individual cluster. From clusters 1–6 (2016), from clusters 7–11 (2017), from clusters 12–16 (2018), from clusters 17–22 (2019), and from clusters 23–27 (2020)
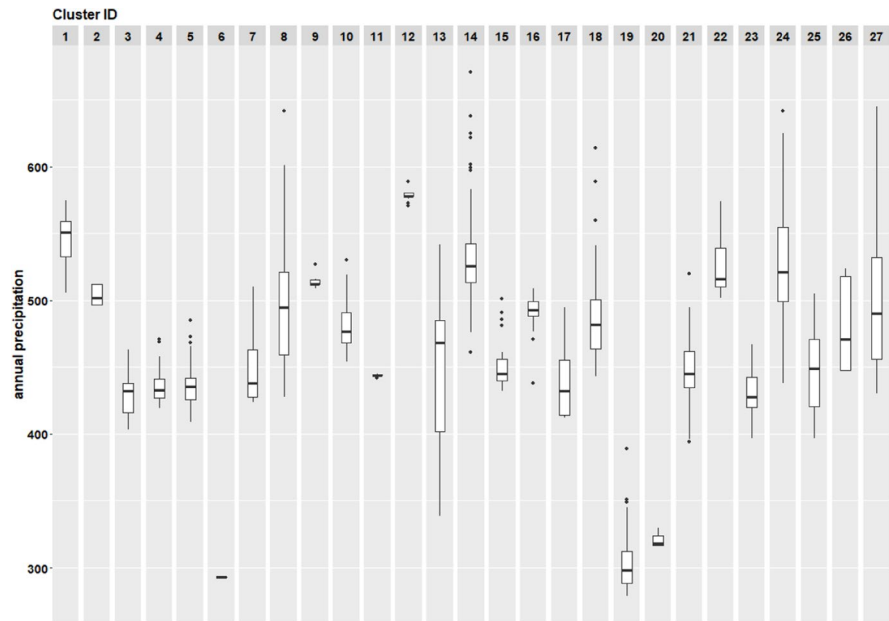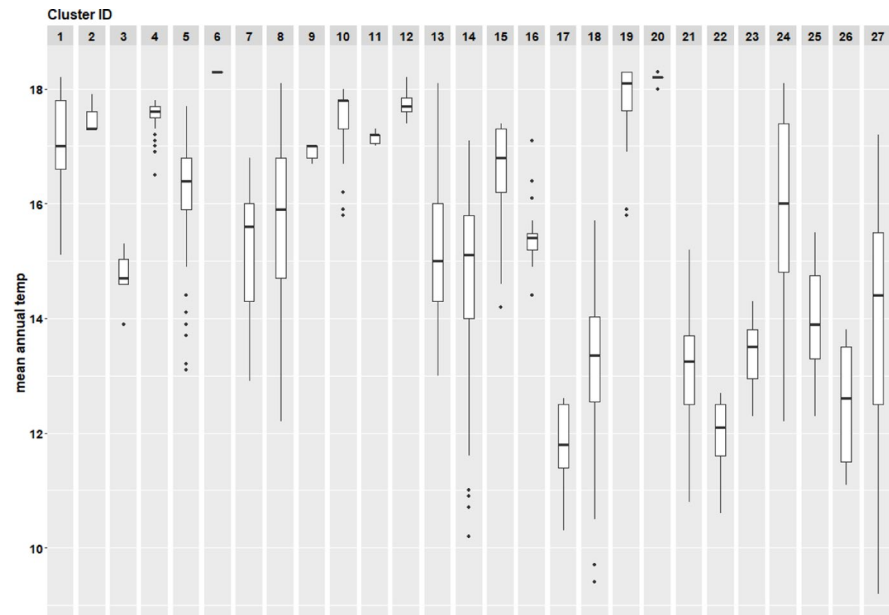

GMM can be used to predict the likelihood of future wildfires in a particular region. Moreover, it can be used to identify the factors contributing to the occurrence of wildfires, such as weather conditions, topography, and human activities (Camarero et al., 2018; Shenoy et al., 2011; Wang et al., 2021; Ying et al., 2018). The insights from GMM analysis can be used to develop more effective wildfire prevention and management strategies.

The R-package mclust 5 (Fraley & Raftery, 2003; Scrucca et al., 2016) has been utilized in the current study, with the GMM approach being implemented. The results obtained allow to detect forest fire clusters based on input variables such as elevation, slope, and aspect. Little research has been conducted in which *mclust 5* has been used in forest fire prediction and analysis. Nonetheless, a study by Saxe et al. (2018)
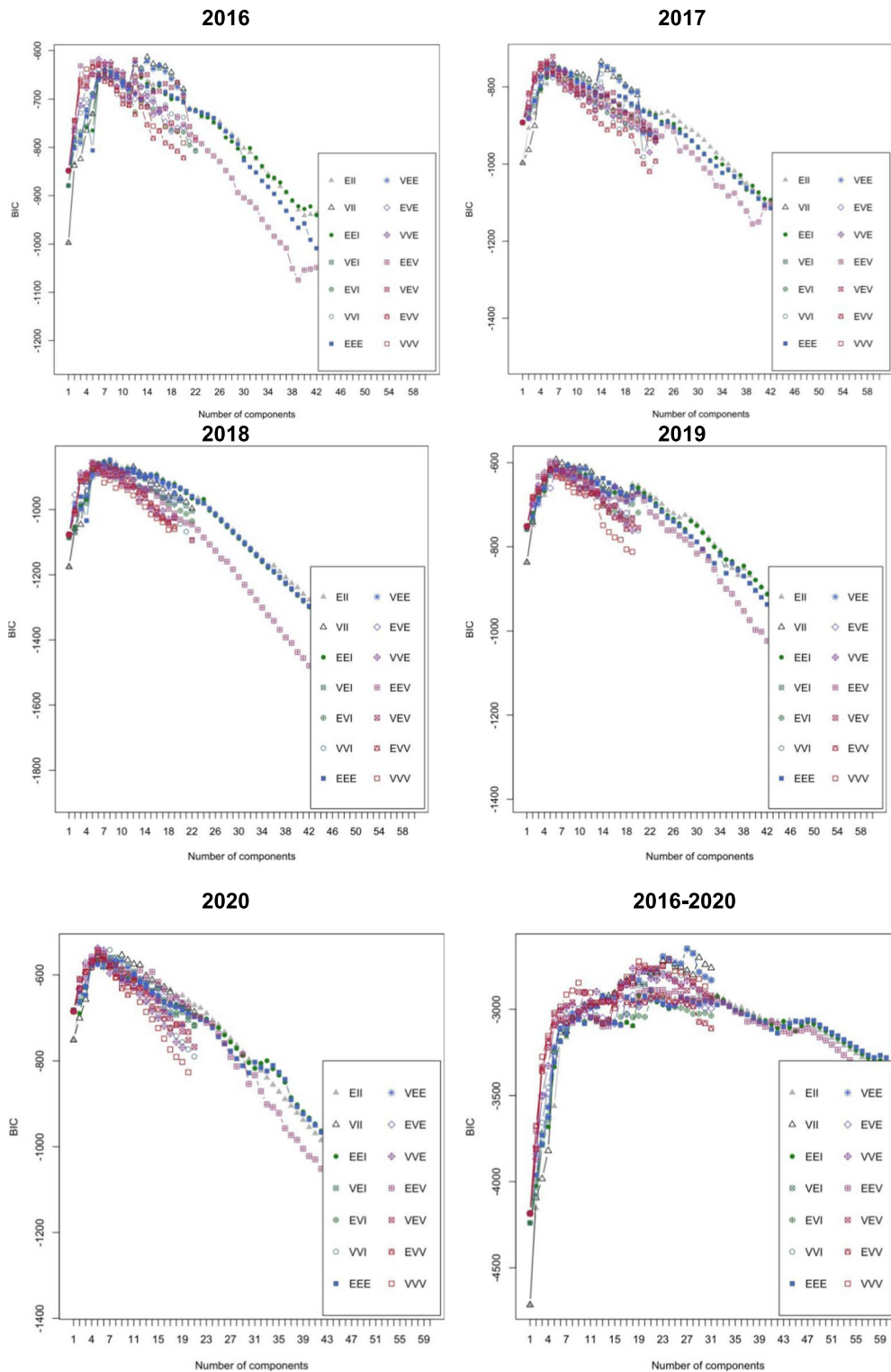
Fig. 9 BIC values for the different models according to the model applied and the number of clusters considered from 2016 to 2020, and all 5 years together, using the function *densityMclust*. The different symbols represent events within each individual cluster detected

characterized and evaluated controls on post-fire streamflow response across western US watersheds. The mclust 5 package has been selected for application in the present study to predict the likelihood of fire ignition in specific areas based on the input variables, which facilitates proactive wildfire risk management and mitigation through cluster identification.

In a model-based clustering approach, each component of a finite mixture density is usually associated with a group or a cluster. Most applications assume that all component densities arise from the same parametric distribution family, although this does not always have to be the case. For data x_i, …., x_n in a D-dimensional space, a popular model is the GMM, a type of clustering algorithm that assumes a D-dimensional multivariate Gaussian distribution:

$$g\left(x|\mu_i, \sum_i\right) = \frac{1}{(2\pi)^{\frac{D}{s}} \sum^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x - \mu_i)\prime \sum^{-1}(x - \mu_i)\right\}$$

for each component i, giving ellipsoidal clusters (Banfield & Raftery, 1993; Celeux & Govaert, 1995). It could be used for different possibilities, like the analysis included in Androniceanu et al. (2020). As its name implies, each cluster is modeled according to another Gaussian distribution. GMM is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. These are used to represent normally distributed subpopulations within an overall population. The advantage of mixture models is that they do not require knowledge of which subpopulation a data point belongs to, but instead they allow the model to learn the subpopulations automatically (McLachlan et al., 2019). There are just two models in one dimension: *E* for equal variance and *V* for varying variance. The volume, shape, and orientation of the covariances in the multivariate setting can be constrained to be similar or variable across groups. Hence, 14 possible models with different geometric characteristics can be specified and are presented in Fig. 3. The dimensions considered in the analysis include elevation, slope, aspect, and combustible class. The use of GMMs to detect wildfires is a novelty from the last few years (Munshi, 2021) that can be further developed in future work with new methodologies. An integrated approach to finite mixture models is provided, with functions that combine model-based hierarchical clustering, the

expectation–maximization (EM) algorithm for mixture estimation, and several tools for model selection. In particular, the EM algorithm is an affordable and numerically stable algorithm to implement, which has reliable global convergence under general conditions. However, the likelihood surface in mixture models tends to have multiple modes. Thus, initialization of EM is crucial because it usually produces sensible results based on reasonable starting values (Wu, 1983). This can be carried out using model-based hierarchical agglomerative clustering (MBHAC), which allows the underlying probabilistic model to be shared by both the initialization and model fitting steps. In addition, MBHAC is computationally efficient, as a single run provides the basis for initializing the EM algorithm for any number of mixture components and component-covariance parameterizations. There are different R packages to analyze clusters, for instance, *Mixtools* (Benaglia et al., 2009) or *Flexmix* (Leisch, 2004). The current analysis has been based on finite GMMs performed using the R package mclust (Fraley et al., 2014), which provides many functions to discover clustering. This software is a popular R package for model-based clustering, classification, and density estimation based on finite GMM.

Methodology followed

In this study, the *mclust* package was used to apply GMM for wildfire prediction and analysis as it provides a comprehensive strategy for clustering, density estimation, and discriminant analysis, which is the most important objective of the present study and

Table 3 Best BIC values from 2016 to 2020

Year	VII,14	VVE,6	VEI,12
2016	− 613.08	− 617.93	− 618.95
2017	VEV,6	VEV,5	VII,14
	− 720.71	− 733.24	− 734.49
2018	EII,8	EEE,8	EII,7
	− 846.57	− 848.09	− 851.45
2019	VII,6	EEV,5	VVE,6
	− 592.87	− 597.47	− 600.22
2020	VVE,5	VVI,7	VEV,5
	− 537.80	− 541.59	− 541.79
2016–2020	VEI,27	VEE,27	VEI,28
	− 2645.31	− 2651.23	− 2671.84

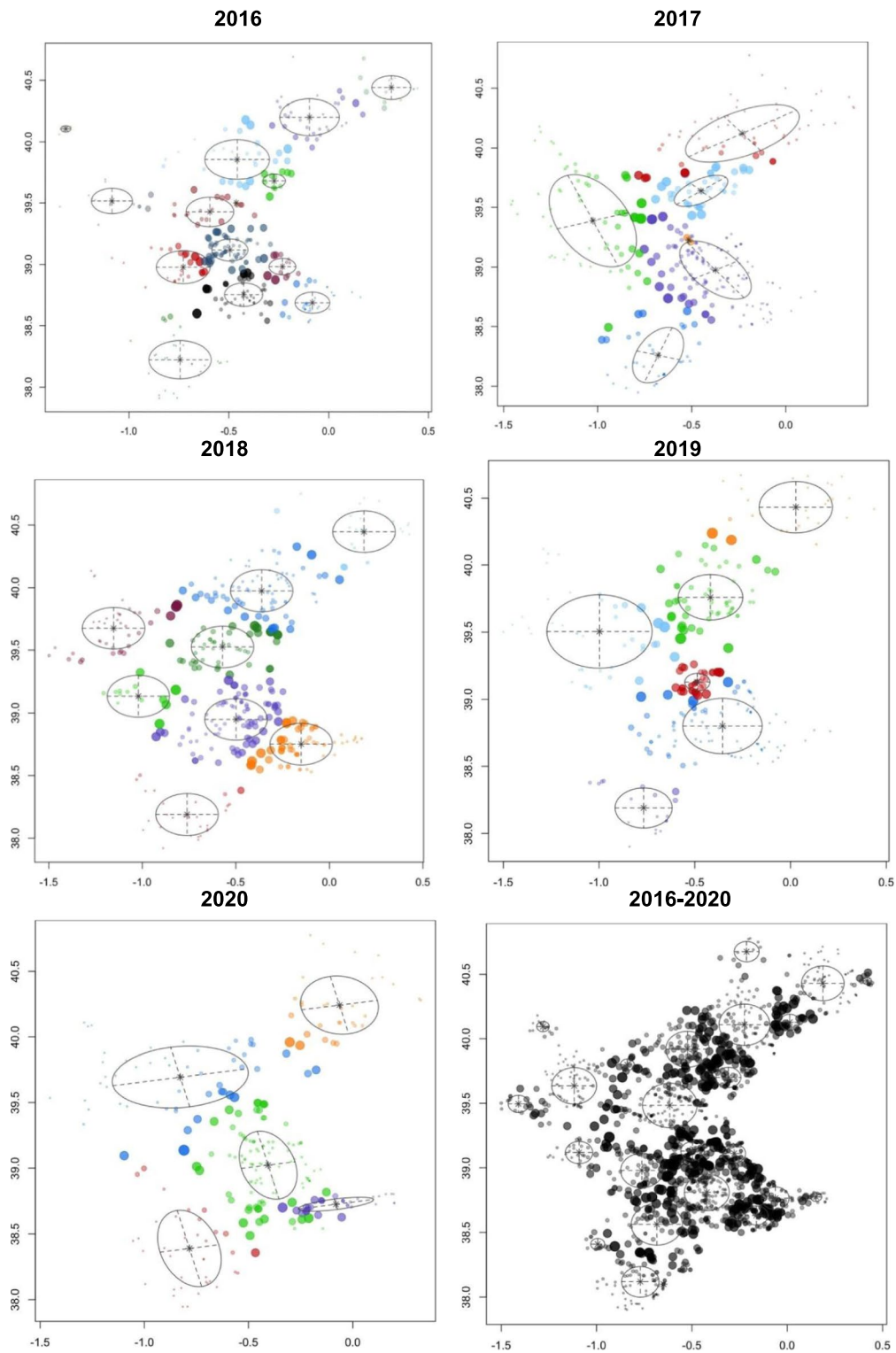**Fig. 10** Cluster dispersion from 2016 to 2020, and all 5 years together, using the function *Mclust* over the surface area of the Valencian Community (different symbols represent events within individual clusters detected and each color represents unique clusters)
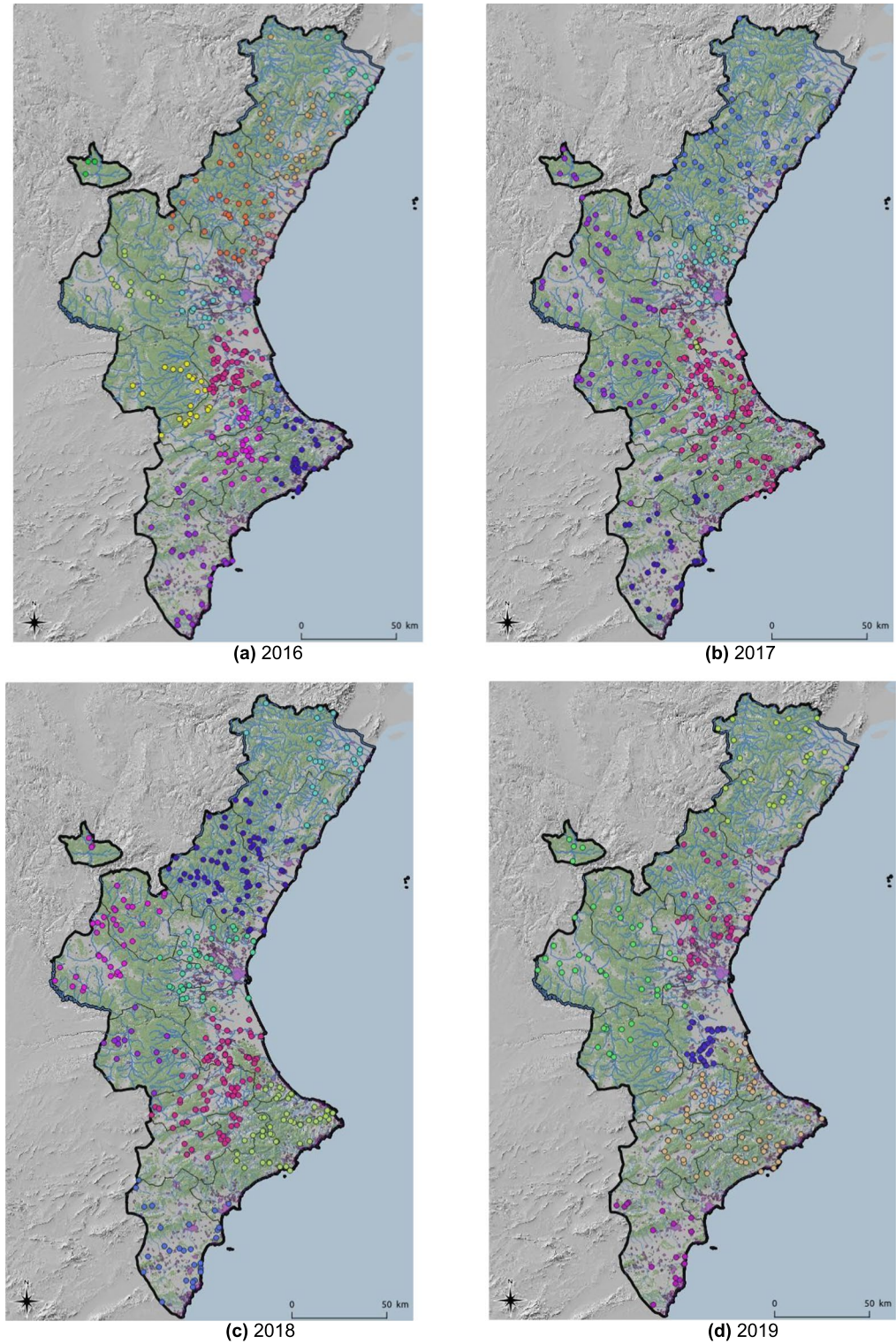
Fig. 11 Distribution of wildfires from 2016 to 2020 and 2016–2020 according to cluster analysis (each color represents unique clusters)
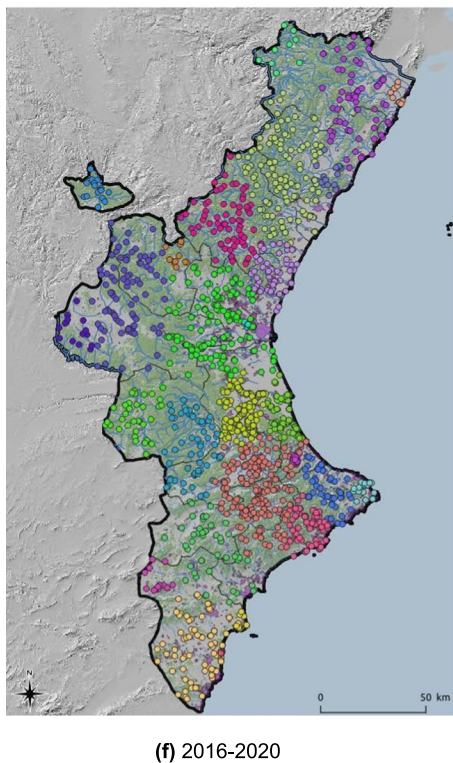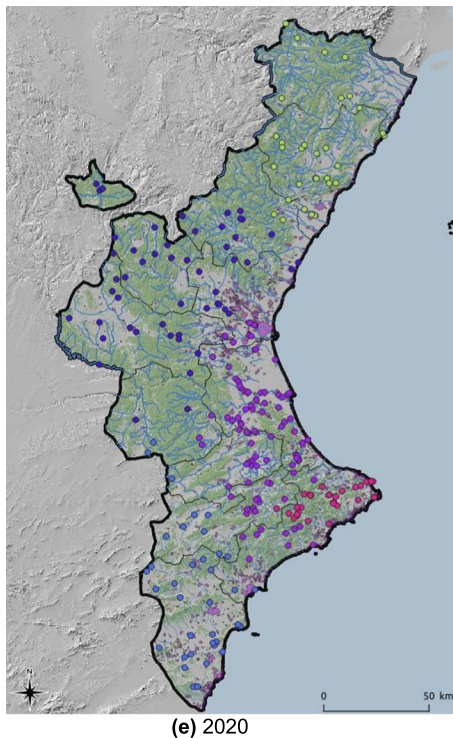
(e) 2020

(f) 2016-2020

Fig. 11 (continued)

has been applied in other studies on topics other than wildfires (Garcia-Rudolph et al., 2020; Zhang & Di, 2020). A variety of covariance structures obtained through eigenvalue decomposition are available. Functions for performing single E and M steps and for simulating data for each available model are also included. The functions included in the *mclust* package provide a data matrix from which the optimal number of components and the covariance parameterization are selected according to the Bayesian information criterion (BIC) for all the models and years considered (Scrucca et al., 2016). Hence, a good way to determine the final number of clusters is to consider the BIC value, which rises as the complexity of the model increases and decreases as the probability increases. So, the model with the lowest BIC is preferred. In the current study, the BIC value has been calculated for parameterized GMMs fitted by the EM algorithm initialized by model-based hierarchical clustering.

Additionally, within the *mclust* package, it is possible to visualize the cluster structure and the geometric characteristics (elliptic, spherical, etc.) induced by the estimated finite GMM. The data can also be projected onto a suitable dimension reduction subspace using the function *MclustDR()*, which implements the methodology introduced in Scrucca (2010).

Finally, *mclust* also provides GMMs with a simple interface for univariate and multivariate density estimation, which is very important as density estimation plays an important role in applied statistical data analysis and theoretical research. Finite mixture models provide a flexible semi-parametric model-based approach to density estimation, thereby making it possible to accurately approximate any given probability distribution using the *densityMclust()* function (Scrucca et al., 2016). Of the 14 potential models, as shown in Fig. 3, the standard models applied to analyze the data are as follows:

- VII,m: spherical and varying volume type of model with m clusters
- VEI,n: diagonal and equal shape type of model with n clusters
- VEE,p: ellipsoidal, equal shape and orientation type of model with p clusters
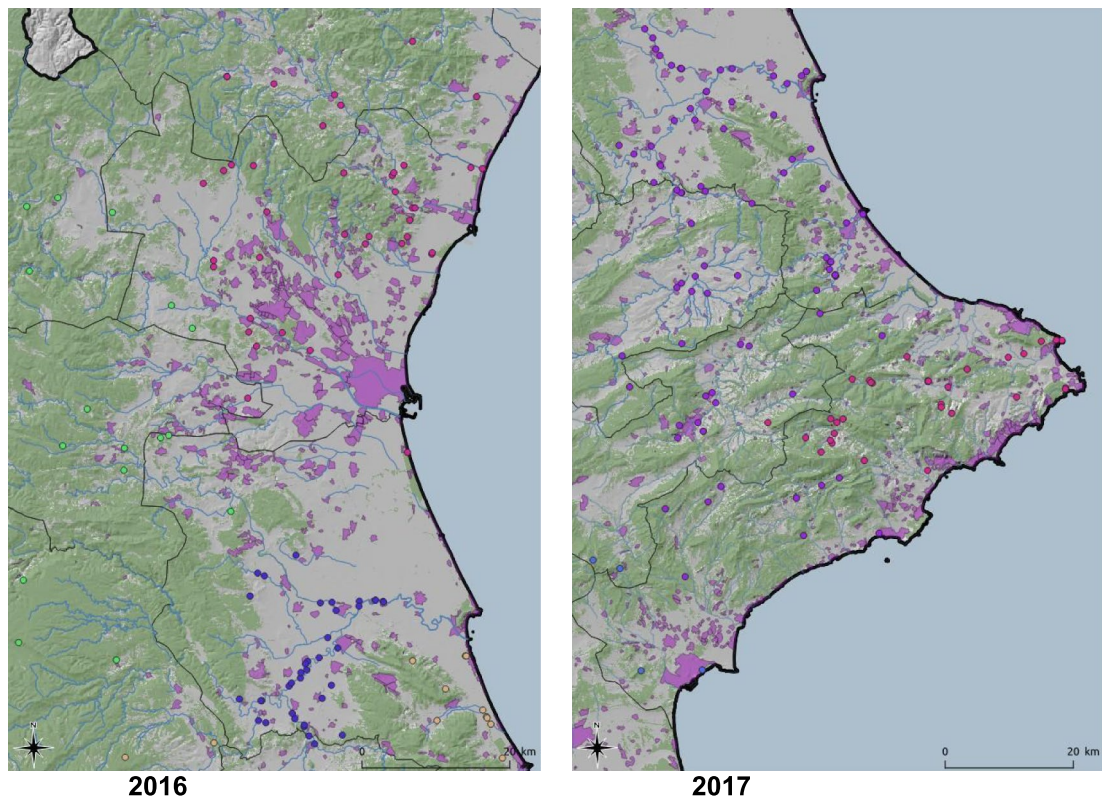- VEV,q: ellipsoidal and equal shape type of model with q clusters

Fig. 12 Distribution of clusters in zoomed zone for 2016 and 2017 (each color represents unique clusters)

- EEV,r: ellipsoidal, equal volume and shape type of model with r clusters
- EVI,s: diagonal, equal volume and varying shape type of model with s clusters
- EII,t: spherical, equal volume type of model with t clusters
- EEI,x: diagonal, equal volume and shape type of model with x clusters

Results

Cluster analysis groups spatially occurring events based on similarity or distance measures. GMM classify observations by likelihood, assuming clusters exist a priori. Table 2 shows the number of fires per cluster from 2016 to 2020. Figure 1 illustrates the geographical location of wildfire occurrences in the Valencian Community. In our analysis, clusters are formed in covariate space, meaning they do not necessarily exhibit a spatially clustered pattern. Cluster labels are arbitrary, and therefore, they are not consistent across years; for example, conditions for Cluster 1 in 2016 may differ from Cluster 1 in 2018. Nevertheless, our interest is on the characteristics of the clusters formed for each year, as they provide insight regarding the conditions for wildfire occurrence.

Figures 4, 5, 6, 7 and 8 present boxplots of cluster characteristics (elevation, slope, and aspect) for 2016, 2017, and 2018, respectively. Fires in Clusters 1, 2, and 5 occurred at low elevations and low to moderate slopes, with cluster 5 at the lowest elevations. Cluster 1 fires occurred on east-facing slopes, Cluster 2 on west-northwest-facing slopes, and Cluster 5 on slopes with a wider aspect range (east, southeast, south, and southwest). The wide range of aspect values for Cluster 5 suggests that those fires occurred on flat areas.
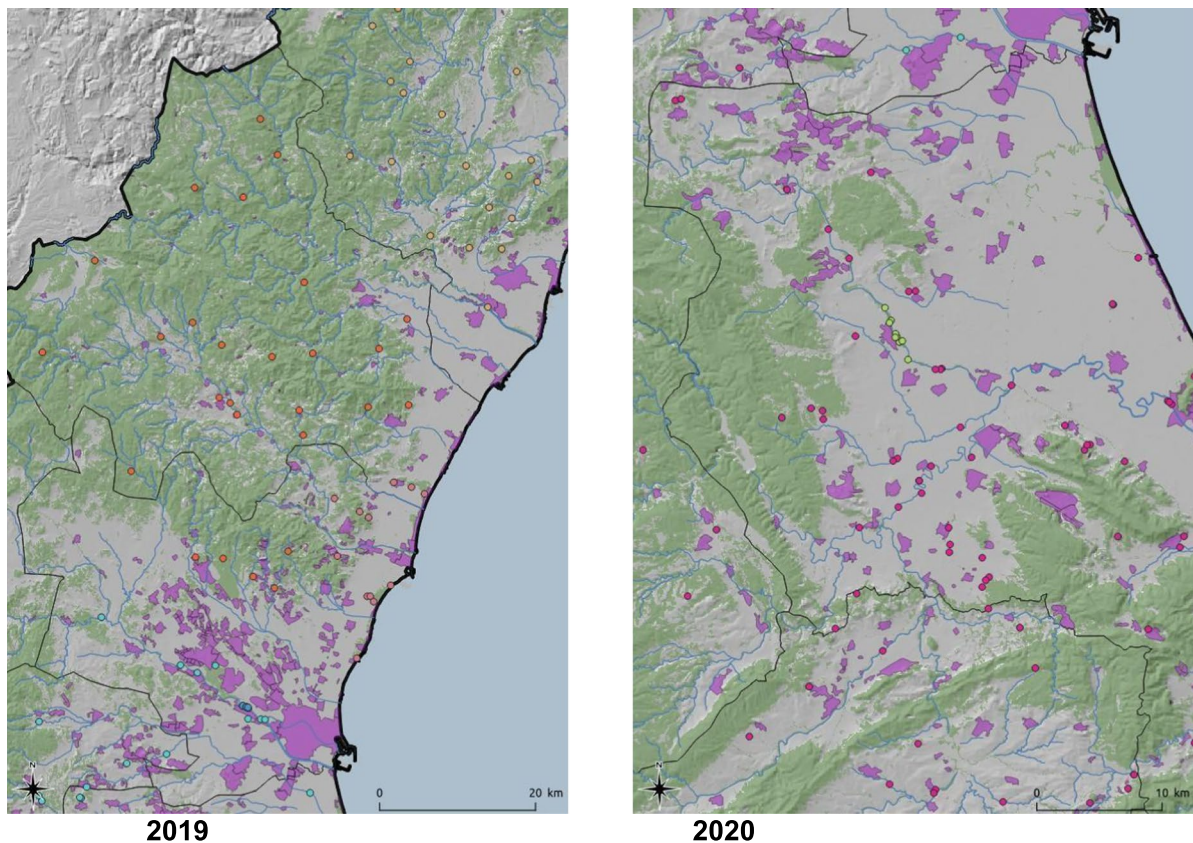
2019

2020

Fig. 13 Distribution of clusters in zoomed zone for 2019 and 2020 (each color represents unique clusters)

Most fires occurred in areas with Combustion Model Category 2, corresponding to grass and shrub stands, prevalent in Mediterranean regions during summer or drought periods. These areas are often found in agro-forestry systems or uncultivated fields.

For 2017, Clusters 1, 2, and 4 grouped most fires, primarily in Combustion Model Category 2 (grass height > 1 m). Cluster 2 also included fires in Categories 6 and 7. Clusters were geographically distinct, with Cluster 3 in the north and Cluster 4 in the south. Fires in Clusters 1 and 4 occurred at low elevations, low slopes, and east, southeast, west, and southwest-facing hills (Fig. 11).

For 2018, fires were classified into five clusters, with Clusters 1, 3, and 5 having the most memberships. Cluster 1 fires occurred in Combustion Model Categories 2, 6, 7, and 8 (grass and shrubs) at elevations of 500–800 m, mid-slopes, and south-facing hills. Clusters 3 and 5 corresponded to low-elevation, low-slope areas with east-to-south-facing hills.

For 2019, six clusters were identified, with most fires in Clusters 1 and 4. Fires in Clusters 4 and 5 ignited in Combustion Model Categories 6, 7, and 8, while others were primarily in Category 2 (Fig. 11).

For 2020, five clusters were identified, with most fires in Combustion Model Categories 2, 6, 7, and 8. Cluster 3 fires occurred in Categories 2, 3, and 5 (grass and slash). Fires in Clusters 1–4 occurred at elevations of 100–900 m, while Cluster 5 fires were at low elevations and nearly flat terrains (Fig. 11). For this year, there were no significant differences regarding aspect.

The spatial distribution can be observed by examining all wildfires that occurred during the study period (Fig. 1) and by analyzing the annual ignition points of these wildfires (Fig. 2). More specifically, looking at Fig. 2, although there are differences in the distribution of wildfires according to the year of occurrence, a certain grouping pattern is observed in specific parts of the study area. Some outbreaks with

Table 4　CORINE Land Cover wildfire distribution by cluster in 2016

Type of land cover	Clusters														Grand total	Percentage
	1	2	3	4	5	6	7	8	9	10	11	12	13	14		
Rice fields	1						1								2	1
Mixed forest		1						1							2	1
Coniferous forest	11	1	4	3	1		5	3		3	2	9		12	54	16
Broad-leaved forest			3	1		2	1								7	2
Sparsely vegetated areas	1	1													2	1
Fruit trees and berry plantations	8	5	2	42		1	1			2	3	3		1	68	20
Inland marshes				1											1	0
Marshlands							1				3				4	1
Transitional woodland-shrub	1		3		2	1	2								9	3
Complex cultivation patterns	1		4			4	7	1		2		2		3	24	7
Olive groves							2			6		1			9	3
Natural grasslands	4	1	8	6	2	2	10	1		3	1	5	2	5	50	15
Beaches, dunes, sand				1											1	0
Grasslands			1	1			2								4	1
Salt marshes			1												1	0
Continuous urban fabric			1												1	0
Discontinuous urban fabric	2		1		2	1		3						1	10	3
Land mainly devote to agriculture, with significant areas of natural vegetation	2	2	2	2	1	2	3	1	2	1		2		1	21	6
Permanently irrigated land							2				1		2		5	1
Non-irrigated arable land					2										2	1
Sclerophyllous vegetation	5	1		4	3	7	5	5	5	7	2	9		6	59	17
Vineyards					3		2								5	1
Total fires	36	12	30	61	16	20	44	15	7	24	12	31	4	29	341	100

Table 5 CORINE Land Cover wildfire distribution by cluster in 2017

Type of land cover	Clusters						Grand total	Percentage
	1	2	3	4	5	6		
Rice fields				5			5	1
Coniferous forest	18	15	26	19		1	79	23
Broad-leaved forest		4					4	1
Sparsely vegetated areas		1		2			3	1
Fruit trees and berry plantations		7	4	44	1	10	66	19
Marshlands				1			1	0
Transitional woodland-shrub	4	2	3	4		2	15	4
Complex cultivation patterns		3	3	4		4	14	4
Olive groves		1	1	5		1	8	2
Natural grasslands	9	4	8	18	6	6	51	15
Beaches, dunes, sand		1			3	1	5	1
Pastures				1		2	3	1
Salt marshes	1						1	0
Continuous urban fabric	3						3	1
Discontinuous urban fabric		2	1	1			4	1
Land mainly devoted to agriculture, with significant areas of natural vegetation	1	1	6	14		9	31	9
Permanently irrigated land	2						2	1
Non-irrigated arable land		1					1	0
Sclerophyllous vegetation	1	15	6	18		5	45	13
Vineyards			2				2	1
Industrial or commercial units				1		1	2	1
Urban green zones	1						1	0
Total fires	40	57	60	137	10	42	346	100

higher wildfire density are observed in the southern part of the Valencian Community, which has less forestry area but more borders between agriculture and forest. However, wildfires are primarily concentrated in the northern and central-western parts of the study area. This phenomenon could shed light on determining which factors related to the type of terrain or the cause of ignition are relevant for wildfires. The BIC influences the choice of cluster numbers and models by balancing model fit (how well the model explains the data) and model complexity (number of parameters). Lower BIC values indicate a better trade-off, guiding the selection of the optimal number of clusters and model type (e.g., VII, VVE, VEI). Variations in optimal models across years occur due to changes in data distribution, complexity, sample size, and noise levels. BIC values for the different models applied are shown in Fig. 9 for each year, according to the number of clusters considered. This figure

shows that the best model is the one that reaches the maximum of the curve. In addition, Table 3 provides information about the BIC value, specifying the three models that perform the best every year and the optimum number of clusters, based on the BIC values (the minimum value being an absolute value). For example, simpler models like VII may suffice for years with less complex data, while more flexible models like VEI are preferred for years with intricate patterns. These variations reflect the dynamic nature of the data and its alignment with different model assumptions. Therefore, the VEI,12 model emerges as the best overall option for the period from 2016 to 2020, as it has the lowest cumulative BIC value (-2671.84). This indicates that it provides the best balance of fit and complexity across the 5-year dataset, effectively capturing the underlying structure of the data while avoiding overfitting. While the optimal model varies by year depending on data

Table 6 CORINE Land Cover wildfire distribution by cluster in 2018

Type of land cover	Clusters								Grand total	Percentage
	1	2	3	4	5	6	7	8		
Rice fields				1					1	0
Mixed forest						1	1		2	1
Coniferous forest	16	8	6	12	12	1	14	8	77	21
Broad-leaved forest	2			1					3	1
Sparsely vegetated areas	1								1	0
Fruit trees and berry plantations	3	3		32	8	2	1	10	59	16
Inland marshes					1				1	0
Marshlands		3				1			4	1
Transitional woodland-shrub	5	1	2	2	3	1	2	1	17	5
Complex cultivation patterns	3	1		3	4	3		7	21	6
Olive groves	3			5		2			10	3
Natural grasslands	16	4	3	11	10	5	3	2	54	14
Beaches, dunes, sand				1		1			2	1
Pastures		2			1			3	6	2
Bare rocks					1				1	0
Continuous urban fabric								1	1	0
Discontinuous urban fabric		1		1	2	2	1	2	9	2
Land mainly devoted to agriculture, with significant areas of natural vegetation	8	2	1	8	8	2	1	4	34	9
Permanently irrigated land	1	1					4		6	2
Non-irrigated arable land				1	1				2	1
Sclerophyllous vegetation	12	2	1	13	6	6	3	9	52	14
Vineyards							10		10	3
Industrial or commercial units		1							1	0
Burned areas				1					1	0
Total fires	70	29	13	92	57	27	40	47	375	100

characteristics, VEI,12 is the most robust choice for the entire period. The selected models differ according to the year, which means that the clusters and the shapes of these clusters in the Valencian Community during this period, 2016 to 2020, are different. This is the expected behavior due to the structure of wildfires and the burned zones. The best model and the optimal number of clusters for each year are reported in the first column of Table 3. The number of clusters in 2020 is lower than in all the previous years. This is because of the nationwide lockdown situation brought about by the ongoing global pandemic.

Figure 10 illustrates cluster density and dispersion, with circle size indicating ignition point dispersion. Cluster shapes and structures varied annually, reflecting diverse fire behaviors and characteristics.

As already depicted in the graphs in Fig. 2, the highest density is observed in the southeastern part of the study area. The last figure represents the spatial clusters for all five study years together. As mentioned in the Introduction section and represented in Fig. 11, fires are located from south to north throughout the study area. The maps in Fig. 11 show the same distribution as Fig. 2, but now consider the cluster analysis. Therefore, each color represents the fires that belong to the same cluster. In addition, to better understand these maps concerning the type of land use, forestry areas are represented in green, rivers are identified with blue lines, and urban areas are shown in purple (Fig. 12).

Regarding the distribution of wildfires in clusters, it is observed that, except for 2016, wildfires are clearly distributed in groups throughout the entire territory,

Table 7 CORINE Land Cover wildfire distribution by cluster in 2019

Type of land cover	Clusters						Grand total	Percentage
	1	2	3	4	5	6		
Rice fields	1						1	0
Coniferous forest	22	1	11	5	4	6	49	18
Broad-leaved forest			2		1	2	5	2
Sparsely vegetated areas	1						1	0
Fruit trees and berry plantations	20	30	7	1	3	4	65	24
Inland marshes	1						1	0
Coastal lagoons	2						2	1
Transitional woodland-shrub	3		1	3		4	11	4
Complex cultivation patterns	5		4	1	2	6	18	7
Olive groves	6			1			7	3
Natural grasslands	5	1	10	1	8	5	30	11
Beaches, dunes, sand			1	1			2	1
Pastures	1		1	1	1	1	5	2
Bare rocks	2						2	1
Continuous urban fabric	1	2		3			6	2
Discontinuous urban fabric	2		1				3	1
Land mainly devoted to agriculture, with significant areas of natural vegetation	9	1	5	1		1	17	6
Permanently irrigated land			2	1		1	4	1
Non-irrigated arable land	1					1	2	1
Sclerophyllous vegetation	8	2	8	1	9	5	33	12
Vineyards				1		7	8	3
Total fires	90	37	53	19	30	43	272	100

Table 8 Types of land use in CORINE Land Cover

Code CLC 09	Land use
10	Urban and suburban areas
20	Temporary rainfed crops
30	Temporary irrigated crops
40	Forest
50	Heathlands and sclerophyll vegetation
60	Prairies
70	Areas of sparse vegetation
80	Areas of low vegetation and soil regularly flooded
90	Bare areas
100	Complex crops
110	Agro-forestry systems
120	Mosaic of natural vegetation (herbaceous, bushes and trees)
200	Water bodies

especially in 2018 and 2019. For 2016, wildfires do not seem to share the same pattern, as no well-defined group distribution appears throughout the study area. The southern forested area, therefore, has a very well-represented group. Moving to the central part of the study area, from 2017 to 2019 wildfires went from east
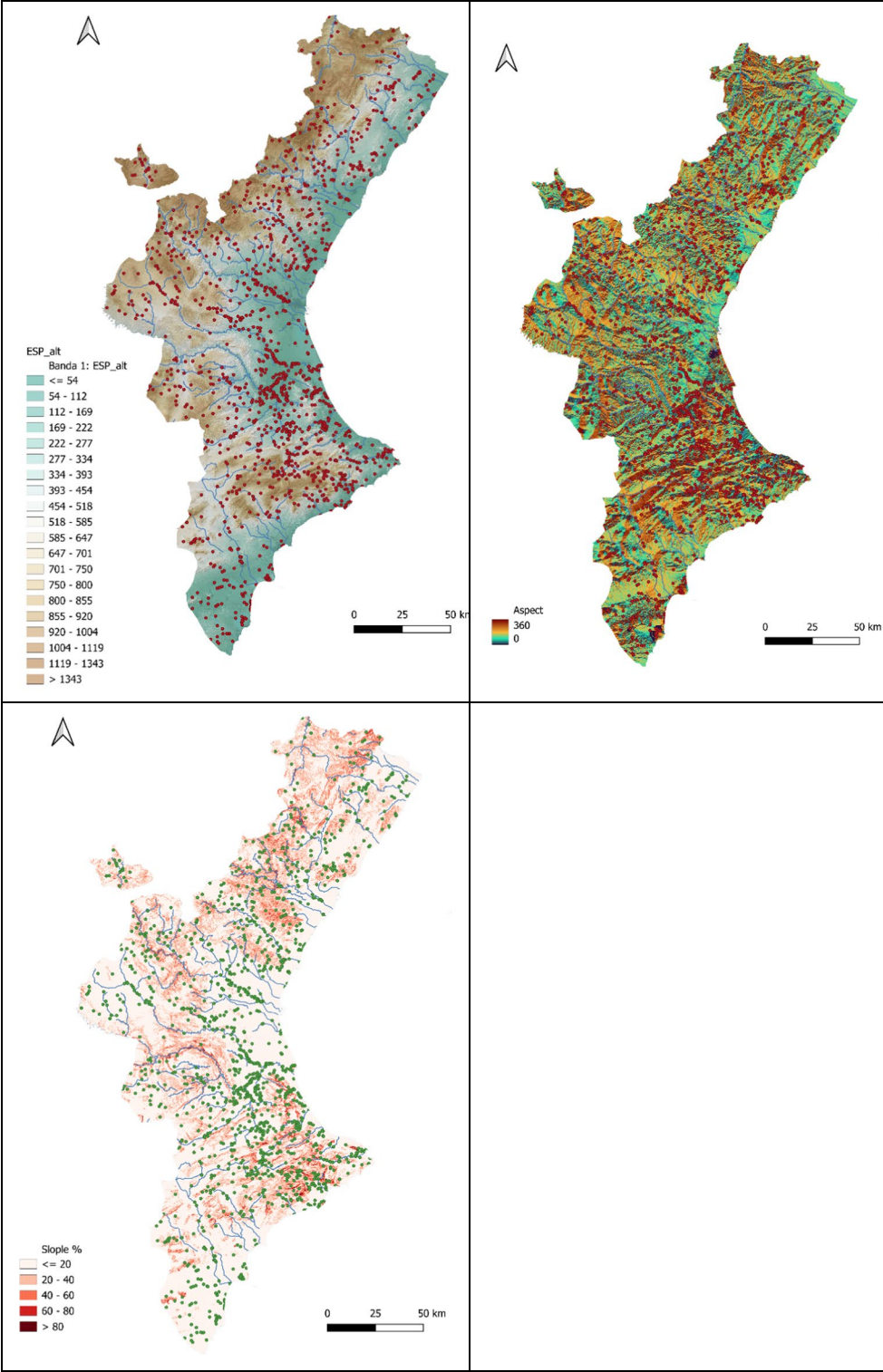
Fig. 14 Altitude, aspect, and slope figures

to west. By looking at the center of the map in the province of Valencia, reveals the following locations of clusters: one in the southeast, another in the northeast, and finally another in the western area (inland). In the north, basically two clusters (2018–2019) divide this area into two parts, one further north and one further south of this northern part of the study area. In contrast, for this same area (the north), in 2017, the wildfires divided this area between coastline and inland. Records for 2020 show similar wildfire patterns to those of 2019, but with fewer wildfires in each cluster (Fig. 13). This clustering division changes from year to year, and therefore, a further analysis should be performed to have a relation between spatial distribution, weather and causes.

Table 1 summarizes covariate distributions, showing wide variation in burned area, with many small fires and few large ones. Precipitation (Fig. 7) and temperature (Fig. 8) varied across clusters (Fasullo et al., 2018). Supplementary Figures S1 and S2 show precipitation and temperature trends for 2016 and 2020.

Land use analysis (Tables 4, 5, 6, 7 and 8) revealed that wildfires primarily burned scrub/herbaceous vegetation, permanent crops, and forests. In 2019, permanent crops slightly outnumbered scrub/herbaceous vegetation. Heterogeneous agricultural areas accounted for ~13% of burned land. Clusters differed between natural and agricultural land uses, with some clusters (e.g., 2017 Clusters 8 and 12) burning no forest land.

Discussion

Cluster analysis confirmed that wildfires with shared characteristics group together, providing robust insights into spatial and environmental patterns. The methodology validated intuitive observations and highlighted the heterogeneity of wildfires over time. Clusters were strongly influenced by orographic features, such as mountains and river basins, dividing the region into flat and mountainous areas (Júnior et al., 2022; Sarala et al., 2022; Saxe et al., 2021).

The distribution of wildfires varied annually, with 2016 showing less defined clustering compared to 2018 and 2019. Southern forested areas exhibited clear clustering, while central areas showed east-to-west wildfire spread. Northern areas were divided into coastal and inland clusters in 2017 and further subdivided in 2018–2019. The 2020 patterns resembled 2019 but with fewer fires.

Land use played a significant role in wildfire distribution, with scrub/herbaceous vegetation and permanent crops being the most frequently burned. Clusters near rivers, particularly in areas with Arundo donax L., highlighted the continued use of fire for land management despite its illegal status and associated risks (Jiménez-Ruiz et al., 2021).

The study underscores the importance of considering environmental covariates (e.g., elevation, slope, aspect, precipitation, temperature) and land use in wildfire analysis. The findings provide a foundation for targeted wildfire management strategies, particularly in high-risk areas identified through cluster analysis. The BIC values and model selections (Zhang & Di, 2020) further support the robustness of the clustering methodology.

Human management of land cover and the traditional use of fire are closely related (Badia et al., 2019). In this study, we explore the identification of land cover across wildfire clusters. While factors such as aspect, altitude, and slope can help explain wildfire behavior, land cover—particularly human-influenced land cover—offers valuable insights into the social dynamics surrounding wildfire events (Fig. 14).

The use of GMM in this study was driven by their flexibility and ability to model complex, multi-modal data distributions. GMMs allow for the estimation of clusters based on probabilistic distributions, making them particularly well-suited for identifying hidden structures within the wildfire data. The application of the mclust package for clustering was chosen due to its robust set of tools for model-based clustering, density estimation, and discriminant analysis, which align with the study's objective of understanding the spatial distribution and characteristics of wildfires over time. Moreover, GMMs were validated through the BIC, which provides a statistical measure to assess the trade-off between model fit and complexity. By selecting the model that minimizes the BIC, we ensure that the clustering solution reflects both the underlying data distribution and avoids overfitting. The iterative process of selecting the optimal number of clusters and the appropriate model (e.g., VII, VEI) across multiple years highlights the dynamic nature of the data and ensures that the clustering methodology remains adaptable to the changing characteristics of wildfires.

A limitation of the current study is the exclusive focus on GMMs for clustering. In future research, it would be valuable to expand upon the findings by conducting a comparative analysis with other clustering

techniques, such as *k*-means or hierarchical clustering, to better understand their relative strengths and limitations in the context of wildfire prediction. This study focused on exploring the applicability of GMMs as a robust and well-established method for identifying spatial and temporal wildfire patterns. The use of GMMs, supported by the mclust package and BIC, serves as an initial step in demonstrating the feasibility of this approach. A more comprehensive comparison with alternative methods could provide additional insights and refine the clustering process in future work, building on the foundation established here.

Conclusions

This work shows the results obtained thanks to the application of statistical techniques that have been little used up until now in the literature in the field of wildfires. The package of functions used, the *mclust* library implemented in R, has made it possible to characterize and analyze the spatial distribution of wildfires, creating clusters that share common characteristics and, therefore, can help better understand their wildfires. Thus, in this paper, the finite GMM methodology has been presented to analyze the behavior of fires in terms of clustering in order to know how fires are grouped in the Valencian Community according to their spatial arrangement. The results show that fire occurrences in the study area are closely linked to the same set of covariates along years but that the way those covariates influence fire occurrences varies along years. The most relevant covariates associated to clustering of fire occurrences were topography and land cover. Temperature and precipitation did not play a significant role in clustering, probably because most of the fires occur during hot summer months. Cluster analysis proved to be useful to understand the main particularities of fire occurrences in areas with similarities in terms of land cover, providing key information for improvements of wildfire management policies. Our results are shown graphically in two maps of spatial coordinates to give a clear and comprehensive message to fire managers. In the specific case of the Valencian Community, located in the Mediterranean area of Spain, a clear clustering pattern is observed and helps to understand which areas are potentially prone to occurrences of scrub fires, wildfires, or fires related to agriculture.

The current study shows a clear pattern of wildfire clusters near the paths followed by rivers. This can be an interesting domain for exploration by future spatial researchers. It is worth noting that fewer wildfire clusters were reported in 2020 because of the pandemic lockdown.

The methodology can be used in future studies to understand which variables favor ignition in specific areas and can also shed light on a smaller scale, that is, to work on elements of the fires in municipalities or districts where fire outbreaks occur. The study helps to depict how wildfires are grouped, whether due to negligence or arson. All this will be addressed in future studies in which other variables of interest are available that help understand the behavior of these environmental phenomena. Moreover, if the analysis includes the cause of the wildfire, it can help detect regions in which an arsonist may act. In this way, it can become possible to detect areas with similar characteristics of forest ignition, allowing the implementation of improvements in wildfire prevention strategies, and provides a baseline to help researchers understand wildfire behavior with emphasis on the type of land cover and other covariates using cluster analysis techniques.

Another aspect to consider is that, while this study primarily applied GMM as a robust methodology for wildfire analysis, future research will explore comparisons between GMMs and other clustering techniques to further refine and validate the findings.

Declarations

Ethical responsibilities of authors All authors have read, understood, and have complied as applicable with the statement on "Ethical responsibilities of Authors" as found in the Instructions for Authors.

References

Androniceanu, A., Kinnunen, J., & Georgescu, I. (2020). E-government clusters in the EU based on the Gaussian mixture models. *Administratie Si Management Public, 35*, 6–20.

Aragó, P., Juan, P., Díaz-Avalos, C., & Salvador, P. (2016). Spatial point process modeling applied to the assessment of risk factors associated with forest wildfires incidence in Castellón, Spain. *European Journal of Forest Research, 135*, 451–464.

Badia, A., Pallares-Barbera, M., Valldeperas, N., & Gisbert, M. (2019). Wildfires in the wildland-urban interface in Catalonia: Vulnerability analysis based on land use and land cover change. *Science of the Total Environment, 673*, 184–196.

Banfield, J. D., & Raftery, A. E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics, 49*(3), 803–821.

Benaglia, T., Chauveau, D., Hunter, D. R., & Young, D. (2009). mixtools: An R package for analyzing finite mixture models. *Journal of Statistical Software, 32*, 1–29.

Camarero, J. J., Sangüesa-Barreda, G., Montiel-Molina, C., Seijo, F., & López-Sáez, J. A. (2018). Past growth suppressions as proxies of fire incidence in relict Mediterranean black pine forests. *Forest Ecology and Management, 413*, 9–20.

Carmo, M., Moreira, F., Casimiro, P., & Vaz, P. (2011). Land use and topography influences on wildfire occurrence in northern Portugal. *Landscape and Urban Planning, 100*, 169–176. https://doi.org/10.1016/j.landurbplan.2010.11.017

Castro, A. C. M., Nunes, A., Sousa, A., & Lourenço, L. (2020). Mapping the causes of wildfires in Portugal by clustering analysis. *Geosciences, 10*, 53. https://doi.org/10.3390/geosciences10020053

Celeux, G., & Govaert, G. (1995). Gaussian parsimonious clustering models. *Pattern Recognition, 28*, 781–793.

Chunyu, Y., Yongming, Z., Jun, F., & Jinjun, W. (2009, October). Texture analysis of smoke for real-time fire detection. In *2009 second international workshop on computer science and engineering* (Vol. 2, pp. 511–515). IEEE.

de Rivera, O. R., López-Quílez, A., Blangiardo, M., & Wasilewska, M. (2020). A spatio-temporal model to understand wildfire causality in Europe. *Environmental and Ecological Statistics, 27*(3), 431–454. https://doi.org/10.1007/s10651-020-00453-5

Díaz-Avalos, C., Juan, P., & Serra-Saurina, L. (2016). Modeling fire size of wildfires in Castellon (Spain), using spatiotemporal marked point processes. *Forest Ecology Management, 381*, 360–369.

Dong, L., Wang, L., Khahro, S. F., Gao, S., & Liao, X. (2016). Wind power day-ahead prediction with cluster analysis of NWP. *Renewable and Sustainable Energy Reviews, 60*, 1206–1212. https://doi.org/10.1016/j.rser.2016.01.106

Farr, T. G. et al. (2007). The shuttle radar topography mission. *Reviews of Geophysics, 45*. https://doi.org/10.1029/2005rg000183

Fasullo, J. T., Otto-Bliesner, B. L., & Stevenson, S. (2018). ENSO's changing influence on temperature, precipitation, and wildfire in a warming climate. *Geophysical Research Letters, 45*(17), 9216–9225.

Fraley, C., Raftery, A. E., & Scrucca, L. (2014). mclust: Normal mixture modeling for model-based clustering, classification, and density estimation (Version 4) [Computer software]. https://CRAN.R-project.org/package=mclust

Fraley, C., & Raftery, A. E. (2003). Enhanced model-based clustering, density estimation, and discriminant analysis software: MCLUST. *Journal of Classification, 20*(2), 263–286.

Garcia-Rudolph, A., Garcia-Molina, A., Opisso, E., & Muñoz, J. T. (2020). Personalized web-based cognitive rehabilitation treatments for patients with traumatic brain injury: Cluster analysis. *JMIR Medical Informatics, 8*(10), e16077.

Han, X. F., Jin, J. S., Wang, M. J., Jiang, W., Gao, L., & Xiao, L. P. (2017). Video fire detection based on Gaussian Mixture Model and multi-color features. *Signal, Image and Video Processing, 11*, 1419–1425.

Hijmans, R. J., & van Etten, J. (2012). *r*aster: Geographic analysis and modeling with raster data (Version 2.0-12) [R package]. https://CRAN.R-project.org/package=raster

Hoinka, K. P., Carvalho, A., & Miranda, A. I. (2009). Regional-scale weather patterns and wildland fires in central Portugal. *International Journal of Wildland Fire, 18*, 36. https://doi.org/10.1071/wf07045

Jiménez-Ruiz, J., Hardion, L., Del Monte, J. P., Vila, B., & Santín-Montanyá, M. I. (2021). Monographs on invasive plants in Europe N° 4: Arundo donax L. *Botany Letters, 168*(1), 131–151.

Júnior, J. S., Paulo, J. R., Mendes, J., Alves, D., Ribeiro, L. M., & Viegas, C. (2022). Automatic forest fire danger rating calibration: Exploring clustering techniques for regionally customizable fire danger classification. *Expert Systems with Applications, 193*, 116380.

Koutsias, N., Martínez-Fernández, J., & Allgöwer, B. (2010). Do factors causing wildfires vary in space? evidence from geographically weighted regression. *GIScience & Remote Sensing, 47*, 221–240. https://doi.org/10.2747/1548-1603.47.2.221

Leisch, F. (2004). FlexMix: A general framework for finite mixture models and latent class regression in R. *Journal of Statistical Software, 11*, 1–18. https://doi.org/10.18637/jss.v011.i08

McLachlan, G. J., Lee, S. X., & Rathnayake, S. I. (2019). Finite mixture models. *Annual Review of Statistics and Its Application, 6*, 355–378.

Moreira, F., et al. (2011). Landscape–wildfire interactions in southern Europe: Implications for landscape management. *Journal of Environmental Management, 92*, 2389–2402. https://doi.org/10.1016/j.jenvman.2011.06.028

Moreno, M. V., Conedera, M., Chuvieco, E., & Pezzatti, G. B. (2014). Fire regime changes and major driving forces in Spain from 1968 to 2010. *Environmental Science & Policy, 37*, 11–22. https://doi.org/10.1016/j.envsci.2013.08.005

Munshi, A. (2021). Fire detection methods based on various color spaces and gaussian mixture models. *Advances in Science and Technology Research Journal, 15*(3), 209–219.

Nikolay, B., Svetlana, K., Marina, B., & Anton, P. (2016). WWLLN data cluster analysis methods for lightning-caused wildfires monitoring. *International Journal of Electrical and Computer Engineering (IJECE), 6*, 3112. https://doi.org/10.11591/ijece.v6i6.12780

Nunes, A., Lourenço, L., & Meira, A. C. (2016). Exploring spatial patterns and drivers of wildfires in Portugal (1980–2014). *Science of The Total Environment, 573*, 1190–1202. https://doi.org/10.1016/j.scitotenv.2016.03.121

Parente, J., Pereira, M. G., & Tonini, M. (2016). Space-time clustering analysis of wildfires: The influence of dataset characteristics, fire prevention policy decisions, weather and climate. *Science of The Total Environment, 559*, 151–165.

Qian, J., Fu, J., Qian, J., Yang, W., Wang, K., & Cao, P. (2018). Automatic early forest fire detection based on gaussian mixture model. In *2018 IEEE 18th International Conference on Communication Technology (ICCT)* (pp. 1192–1196). IEEE.

Quílez Moraga, R. (2019). Los incendios forestales en la comunitat valenciana en el contexto del cambio climático. In *III Congreso forestal de la CV: Gestión de incendios forestales en el contexto del cambio climático* (p. 8). Universitat de València.

Reynolds, D. A. (2009). Gaussian mixture models. In S. Z. Li, & A. Jain (Eds.), *Encyclopedia of biometrics* (pp. 659–663). Springer. https://doi.org/10.1007/978-1-4899-7488-4_196

Sarala, P., Lunkka, J. P., Sarajärvi, V., Sarala, O., & Filzmoser, P. (2022). Timing of glacial-non-glacial stages in Finland: An exploratory analysis of the OSL data. *Arctic, Antarctic, and Alpine Research, 54*(1), 428–442.

Saxe, S., Hogue, T. S., & Hay, L. (2018). Characterization and evaluation of controls on post-fire streamflow response across western US watersheds. *Hydrology and Earth System Sciences, 22*(2), 1221–1237.

Saxe, S., Hogue, T. S., & Hay, L. (2021). Characterization of post-fire streamflow response across western US watersheds. *Hydrological Processes, 35*(10), e14339. https://doi.org/10.1002/hyp.14339

Scrucca, L. (2010). Dimension reduction for model-based clustering. *Statistics and Computing, 20*, 471–484.

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: Clustering, classification and density estimation using gaussian finite mixture models. *The R Journal, 8*, 289.

Sebastián-López, A., Salvador-Civil, R., Gonzalo-Jiménez, J., & SanMiguel-Ayanz, J. (2007). Integration of socio-economic and environmental variables for modelling long-term fire danger in southern europe. *European Journal of Forest Research, 127*, 149–163. https://doi.org/10.1007/s10342-007-0191-5

Shenoy, A., Johnstone, J. F., Kasischke, E. S., & Kielland, K. (2011). Persistent effects of fire severity on early successional forests in interior Alaska. *Forest Ecology and Management, 261*(3), 381–390.

Strauss, C., Rosa, M. B., & Stephany, S. (2013). Spatio-temporal clustering and density estimation of lightning data for the tracking of convective events. *Atmospheric Research, 134*, 87–99. https://doi.org/10.1016/j.atmosres.2013.07.008

Tedim, F., Remelgado, R., Borges, C., Carvalho, S., & Martins, J. (2013). Exploring the occurrence of mega-fires in Portugal. *Forest Ecology and Management, 294*, 86–96. https://doi.org/10.1016/j.foreco.2012.07.031

Torabian, M., Pourghassem, H., & Mahdavi-Nasab, H. (2021). Fire detection based on fractal analysis and spatio-temporal features. *Fire Technology, 57*(5), 2583–2614.

Velizarova, E., & Alexandrov, A. (2021). Informatics approaches for forest fire spread prediction. *Research in Computer Science in the Bulgarian Academy of Sciences*, 493–501.

Venäläinen, A., et al. (2014). Temporal variations and change in forest fire danger in Europe for 1960–2012. *Natural Hazards and Earth Systems Sciences, 14*, 1477–1490. https://doi.org/10.5194/nhess-14-1477-2014

Wang, M., Gao, G., Huang, H., Heidari, A. A., Zhang, Q., Chen, H., & Tang, W. (2021). A principal component analysis-boosted dynamic Gaussian mixture clustering model for ignition factors of Brazil's rainforests. *IEEE Access, 9*, 145748–145762.

Wu, C. J. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics, 11*(1), 95–103. https://doi.org/10.1214/aos/1176346060

Ying, L., Han, J., Du, Y., & Shen, Z. (2018). Forest fire characteristics in China: Spatial patterns and determinants with thresholds. *Forest Ecology and Management, 424*, 345–354.

Yoon, S. H., & Min, J. (2013). An intelligent automatic early detection system of forest fire smoke signatures using Gaussian mixture model. *Journal of Information Processing Systems, 9*(4), 621–632.

Zhang, W., & Di, Y. (2020). Model-based clustering with measurement or estimation errors. *Genes, 11*(2), 185.

Zhang, R. T., Ma, T., Lin, J., Huang, Y., & Li, Y. Q. (2017). Analysis of fire incidents and characteristics of spatio-temporal distributions for serious fires from 2007 to 2016 in China. *J Xi'an Univ Sci Technol, 6*, 829–836.

Zhao, J., Zhang, Z., Han, S., Qu, C., Yuan, Z., & Zhang, D. (2011). SVM based forest fire detection using static and dynamic features. *Computer Science and Information Systems, 8*(3), 821–841.