Environmental and Ecological Statistics
https://doi.org/10.1007/s10651-025-00645-x

Non-separable spatio-temporal Poisson point process models for fire occurrences

Nicoletta D'Angelo[1] · Alessandro Albano[1,2] · Andrea Gilardi[3] · Giada Adelfio[1]

Received: 31 July 2024 / Revised: 10 January 2025 / Accepted: 11 January 2025

Abstract
Our study addresses the analysis of environmental concerns through point process theory. Among those, Sicily faced an escalating issue of uncontrolled fires in recent years, necessitating a thorough investigation into their spatio-temporal dynamics. Each fire is treated as a unique point in both space and time, allowing us to assess the influence of environmental and anthropogenic factors. A non-separable spatio-temporal Poisson model is applied to investigate the influence of land use types on fire distribution, controlling for other environmental covariates. The results highlight the significant effect of human activities, altitude, and slope on spatio-temporal fire occurrences, also confirming their dependence on various environmental variables, including the maximum daily temperature, wind speed, surface pressure, and total precipitation. As a model with constant parameters in space and time may be too restrictive, a local version of the proposed model is also fitted. This allows us to obtain better performance and more valuable insight into the estimated effects of the different environmental covariates on the occurrence of fires, which we find to vary both in time and space. This research work's relevance lies in the analysis of an important environmental problem through complex point process models, yet easily interpretable, given their resemblance to regression-type models. We also provide reference to newly available open-source software for estimating such models. Finally, we contribute to the framework of spatio-temporal point process modelling by integrating data with different spatio-temporal resolutions from very diverse sources.

Keywords Fires · Land usage · Local analysis · Point processes · Spatial analysis · Intensity estimation

Handling Editor: Luiz Duczmal.

The paper has been presented at the 52nd Scientific Meeting of the Italian Statistical Society, University of Bari Aldo Moro, Bari, Italy, June 17–20, 2024.

Extended author information available on the last page of the article

Published online: 06 March 2025

1 Introduction

In recent years, Sicily has experienced an alarming increase in uncontrolled fires, establishing it as the Italian region with the highest frequency of events and, consequently, the largest burned area. The occurrence of these fires is primarily associated with ignition sources, forest fuels, and environmental conditions (Ganteaume et al. 2013; Hantson et al. 2015; Ricotta and Di Vito 2014). Ignition sources are typically categorized into natural causes, such as lightning and geological factors, and human causes, both accidental and intentional (Aldersley et al. 2011; Rodrigues and De la Riva 2014). Human-driven factors, particularly arson, often ignited for purposes like creating new pasture resources or burning stubble, stand out as the primary contributors to fires in Sicily, especially in areas where vegetation interfaces with urban structures (Ferrara et al. 2019).

Fires pose a significant threat to ecosystems, human settlements, and economic activities, demanding a comprehensive understanding of their spatio-temporal dynamics for effective mitigation and management. The island of Sicily, nestled in the heart of the Mediterranean Sea, has a rich history and diverse landscapes, making it vulnerable to the escalating impacts of climate change, including an increased frequency and intensity of fires. Recently, Sicily faced a pronounced fire season, underscoring the urgent need for advanced analytical methodologies to discover the underlying patterns and drivers of these destructive events.

Traditional approaches to fire analysis often rely on descriptive statistics and basic spatial visualization, providing limited insights into the effect of environmental, climatic, and anthropogenic factors that contribute to fire occurrences. This paper advocates for the application of point process methodology to explore jointly the spatial and temporal dynamics. In particular, we analyse the Sicilian fires occurred in 2023, which represent the most recent completely-available yearly dataset.

Point process methodology, rooted in statistical theory, enables the modelling of events occurring in space and time, making it an ideal tool for in-depth analysis of fire patterns. By treating each fire occurrence as a point in a spatial and temporal domain, we aim to discern potential predictors that may contribute to the ignition of fires across the Sicilian landscape.

Point process models have been successfully used in the literature to fit real fire data. For instance, Genton et al. (2006) employed spatial point process methods to analyse the clustering behaviour (using the spatial K and L functions) of fires that occurred in proximity of St Johns River (Florida) over 21 years (1981–2001). Other authors (Podur et al. 2003) focus on lighting fires (and detected clustering behaviour), whereas they focus on every type of fire ignition. The analyses are run yearly in a purely spatial context. They do not estimate any parametric model and clearly detect clustering behaviour for all four fire causes (arson, lightning, accident, and railroad).

In the spatio-temporal framework, Tonini et al. (2017) performed temporal clustering analysis for forest fires in Portugal and found a complex structure associated with different behaviour according to the size of the fire. Specifically, they state that "medium fires tend to aggregate around small fires, while large

fires aggregate at a larger distance and longer times, indicating that the return time following these events is longer than for small and medium fires". Defining a normalized empirical intensity ratio index, Gabriel et al. (2017) showed that inhibitive patterns between neighboring events can span several years. Some international studies considered first-order and second-order characteristics for identifying drivers and spatial distribution of fires (Raeisi et al. 2021).

Moreover, traditional analyses often fall short of capturing the relationships between environmental, climatic, and anthropogenic factors that contribute to fires. By exploiting spatio-temporal models, we aim to assess which variables influence the probability of fire occurrence. The external spatio-temporal variables considered in this work include land usage, Normalised Difference Vegetation Index (NDVI), elevation, slope, temperature, precipitation, and wind speed.

As noted by Butsic et al. (2015), fires and land usage are intrinsically connected, but research investigating this dynamic is limited. Specifically, the influence of land use on fire ignition is a well-documented phenomenon. Previous papers have shown that land use, particularly the presence of roads, can significantly impact the ignition of fires (Ricotta et al. 2018), with the influence of roads being much stronger in less dense land cover areas. Consequently, as human activities expand on natural landscapes, understanding the impact of land use is critical for effective wildfire management.

The proposed non-separable spatio-temporal Poisson point process model (D'Angelo and Adelfio 2024b) aims to provide a comprehensive yet accessible model for analysing fire occurrences, which can be applied across various types of fires. We focus our analysis on the fires that occurred in Sicily in 2023, as a specific case study. To this aim, the data and code that were used to build the spatio-temporal data object and estimate the (global and local) statistical models can be browsed at the following link: https://github.com/agila5/ppm-fire-occurrences. We first fit a global non-separable spatio-temporal Poisson model in order to investigate the influence of land use types on fire occurrence, controlling for other environmental covariates. Indeed, Schoenberg (2004), Díaz-Avalos et al. (2013), and Fuentes-Santos et al. (2018) show that the intensity of forest fire occurrences varies in space and time in a nonseparable way. Our results highlight the significant effect of human activities, altitude, and slope on spatial fire occurrence, confirming also their dependence on various environmental variables, including the maximum daily temperature, wind speed, surface pressure, and total precipitation.

However, a model with constant parameters in time and space may not represent detailed local variations in the data adequately, since the pattern may present temporal and spatial variation of the influence of covariates in the scale or spacing between points and the abundance of points (D'Angelo et al. 2022).

Recently, the inclusion and test of spatially (or spatio-temporally) varying covariates in intensity function has been of particular interest in Díaz-Avalos et al. (2014), Borrajo et al. (2017, 2020a, 2020b) and Myllymäki et al. (2021). Indeed, a different way of analysing a point pattern can be based on local techniques identifying specific and undiscovered local structures, for instance, sub-regions characterised by different interactions among points, intensity, and the influence of covariates.

Therefore, we proceed by fitting a localised version of the proposed spatio-temporal non-separable Poisson model. This allows us to obtain a better fitting to the data and to get further insight into the effects of the environmental covariates, which we find to vary both in time and space. We also carry out a comparison between the two models, showing the better fitting of the local model.

Throughout the paper, we shall distinguish between "global" models, in which the parameters are assumed to be constant in time and space, as in regular regression models, and "local" models, in which the parameters are allowed to vary with locations.

We believe this work's strength lies in the analysis of a relevant environmental problem through complex point process models, but easy to interpret. Indeed, we highlight throughout the paper their resemblance to regression-type models.

We further contribute to the framework of spatio-temporal point process modelling by integrating data with very different spatio-temporal resolutions from diverse sources. These include the Fire Information for Resource Management System (FIRMS) platform for the fires data, the CORINE Land Cover (CLC) dataset for the land usage, the Digital Elevation Model (DEM) data for the elevation and slope variables, the NDVI data derived from Sentinel-3 platform, and the Copernicus ERA5 for all the other environmental variables considered in our study.

All the analyses are carried out through the `R` statistical software (R Core Team 2024), making particular use of the package `stopp` (D'Angelo and Adelfio 2024b, c) for the model fitting, and `sf` (Pebesma 2018), `stars` (Pebesma and Bivand 2023), and `spatstat` (Baddeley and Turner 2005) for spatial and spatio-temporal data handing. The availability of open-source software for carrying out similar analyses to other environmental problems further highlights the relevance of the work.

The structure of the paper is as follows. Section 2 presents the data. Section 3 provides an overview of spatio-temporal point processes and the models employed in the paper, while Sect. 4 illustrates the actual model fitting through the global and local intensity estimation. Section 5 is devoted to the discussion and conclusions.

2 Data

This section focuses on the description of the data used in this paper, namely the fire data, which will represent the spatio-temporal point pattern under analysis, and some spatial and/or temporal variables which will serve as covariates, assumed to influence the overall occurrence of points, as is typical in point process analysis.

2.1 Fires point pattern

To conduct our analysis, we retrieve the data from the Fire Information for Resource Management System (FIRMS) platform, accessible for download at the following URL: https://firms.modaps.eosdis.nasa.gov/download/. This source provides near real-time active fire locations to natural resource managers derived from snapshots taken by a variety of satellites, such as VIIRS (Schroeder et al. 2014).

The variables considered in our study are:

- `Longitude` and `Latitude`: Center of 375 m × 375 m fire pixel, but not necessarily the actual location of the fire, as one or more fires can be detected within the pixel. Considering that fire pixels are extremely small compared to the size of the complete observation window (i.e. the Sicily island) and following a standard approach in the fire literature (see, for example, the pre-processing steps described in Koh et al. (2023) and references therein), we approximate the fire occurrences as point-level observations.
- `Acq_Date`: Date of acquisition.

In particular, we transformed `Longitude` and `Latitude` coordinates into a projected reference system (EPSG code 3003, https://epsg.io/3003) and used the resulting values as spatial coordinates of our point pattern, whereas `Acq_Date` is used for building fires time occurrence, taking the number of days from the 1st of January. The final point pattern of Sicilian fires in 2023 analysed throughout this paper contains 10,656 events.

Focusing exclusively on the spatial distribution of the events, Fig. 1 displays the yearly fire counts in Italy during 2023 at the regional level. Notably, this graphical representation emphasizes a pronounced concentration of fires in the Sicilian territory, with the second-highest incidence observed in the southern region of Puglia (*Apulia*).

Following this, Fig. 2 compares the fire counts that occurred in Italy and Sicily in 2023. It is evident that the summer months, particularly from July to October, stand out as critical periods with the highest number of fires. Specifically, July emerges as the most challenging month, recording 6942 fires in Italy, of which 3614 occurred exclusively in Sicily. Finally, Fig. 3 shows the spatial distribution of Sicilian fires in 2023. The inset map (which is reported in the top-left) displays the whole observation window that includes the mainland of the region and three smaller islands (namely Pantelleria, Lampedusa, and Linosa), located far South and reported in the bottom-left of the plot.

Figure 4 displays the spatio-temporal distribution of fires that were recorded in Sicily during 2023 divided by the month of occurrence. We can clearly notice the existence of space-time interactions (e.g. when comparing July and October), meaning that the intensity of the process in time is dependent on the space location, and vice versa. A formal test of separability is carried out as in Ghorbani et al. (2021), through the `kernstadapt` R package (González and Moraga 2022). With a p-value of 0.001, it indicates that the point pattern under analysis is not spatio-temporal separable. This motivates us to consider a possible non-separable specification of the spatio-temporal model we will fit to the data.

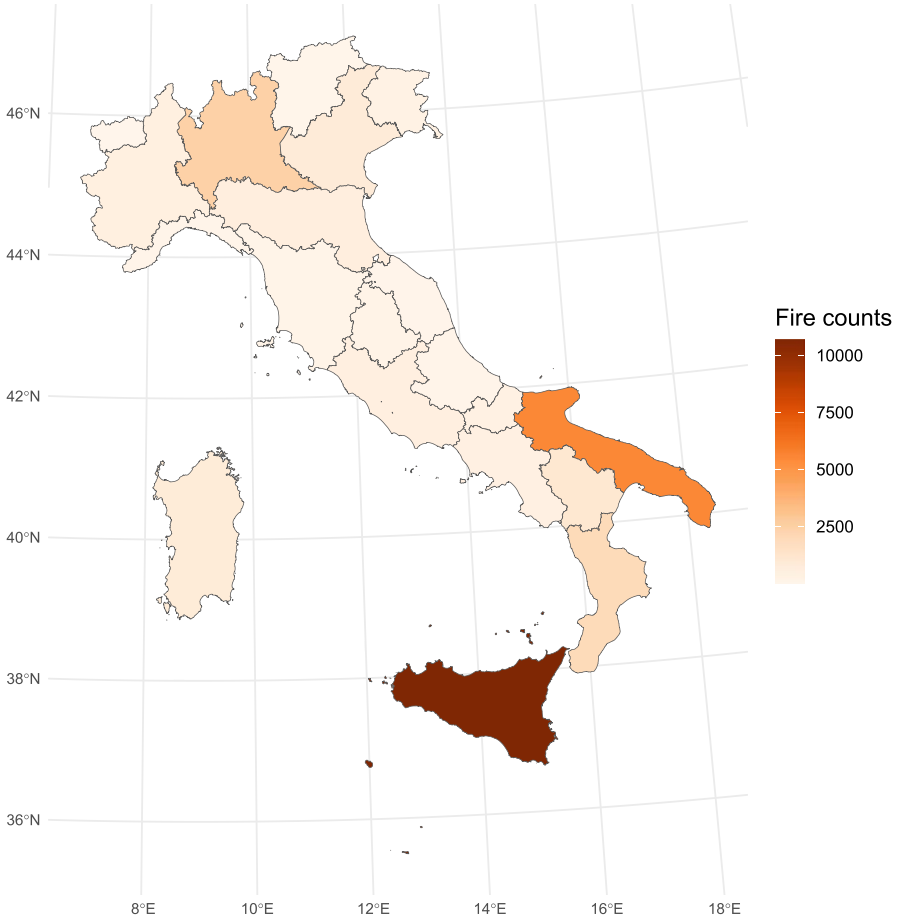

Fig. 1 Choropleth map displaying the counts of the fires that occurred in each of the 20 Italian regions during 2023

2.2 Land use

The land use data come from the CORINE Land Cover (CLC) dataset,[1] which is a comprehensive land-cover and land-use database created by the European Environment Agency (EEA) to facilitate environmental monitoring and assessment across Europe. It is usually employed to provide insights into the changing landscapes of Europe, aiding in several projects such as environmental management and spatial planning, in addition to fire analysis as exemplified in this paper. Combining satellite imagery (obtained from Sentinel-2 and Landsat-8 projects) and ground-based information, the CLC classifies land cover into a hierarchical raster system of over

[1] URL: https://land.copernicus.eu/en/cart-downloads; data downloaded during 2023-11.

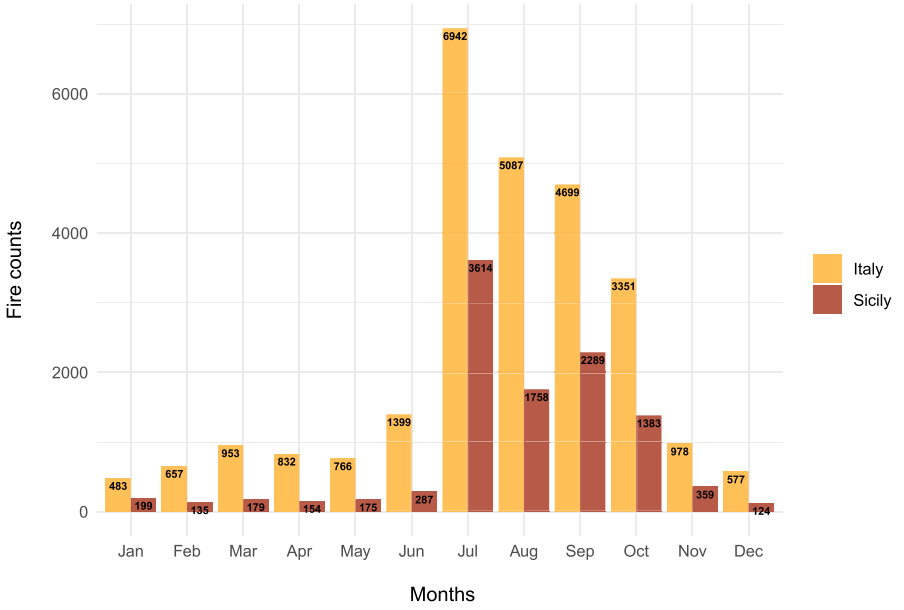

Fig. 2 Barplot comparing the number of fires that occurred in the whole country (yellow bars) and the Sicily region only (red bars)

40 classes, including urban areas, forests, water bodies, and agricultural lands. The detailed classification according to CLC comes in Table 1, whereas the product documentation and the nomenclature guidelines can be browsed on the project's website.[2]

The data regarding the land use in Sicily are depicted in Fig. 5 where each raster tile is coloured according to level 1 of the CLC Legend (in Table 1, after merging the 4th and 5th categories, i.e. wetlands and water bodies, "waterbodies" from now on). In particular, we represent the macro land usage classification, whose categories are:

- Artificial surfaces;
- Agricultural areas;
- Forest and semi-natural areas;
- Waterbodies.

As we can see, the majority of the Sicilian territory is constituted by agricultural areas, mainly in the eastern and southern parts of the region, followed by forests and semi-natural areas, typically located in proximity to some mountain chains in the North and North-East. Artificial surfaces, commonly found nearby the most

[2] URL: https://land.copernicus.eu/content/corine-land-cover-nomenclature-guidelines/html/. Last access: Feb. 2024.

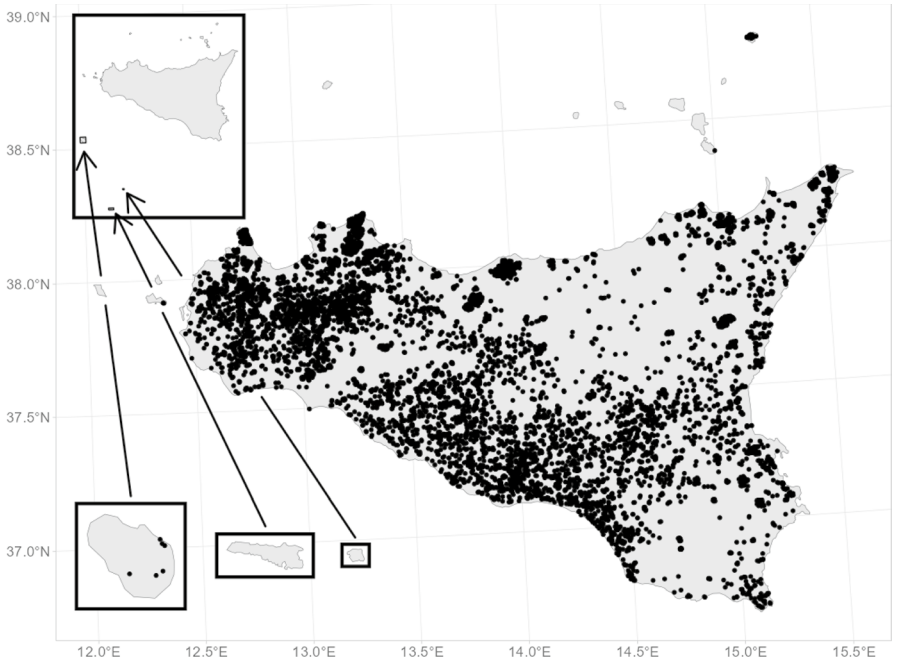

Fig. 3 Spatial distribution of the 10,656 fires recorded in the Sicily region during 2023. The inset map reported in the top-left displays the whole observation window which includes the mainland and three smaller islands (namely Pantelleria, Lampedusa, and Linosa) located far South and reported in the bottom-left as inset maps

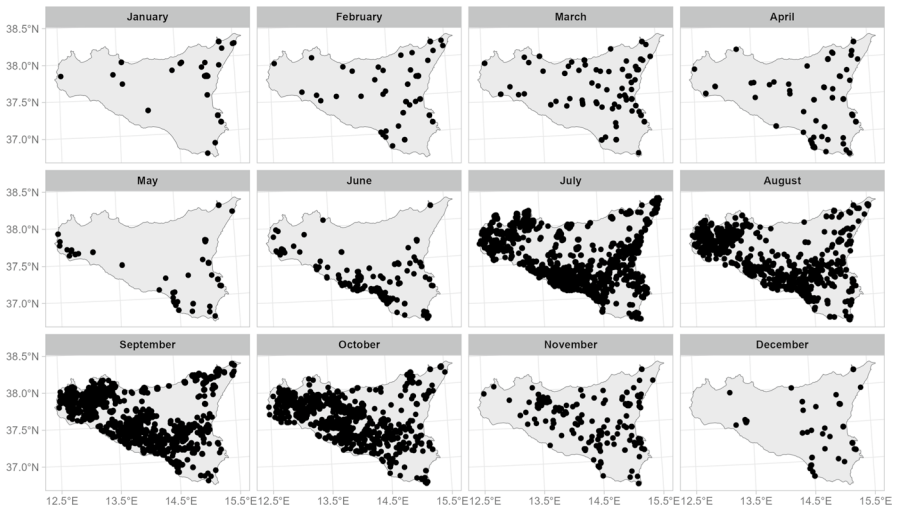

Fig. 4 Spatial distribution of fires recorded in the Sicily region during 2023 by month. To simplify the graphical representation, we focused only on the fires that occurred in the mainland

Table 1 Land use classification according to Corine Land Cover Legend

Level 1	Level 2	**Level 3**
Artificial surfaces	Urban fabric	Continuous urban fabric
		Discontinuous urban fabric
	Industrial, commercial and transport units and transport units	Industrial or commercial units
		Road and rail networks and associated land
		Port areas
		Airports
	Mine, dump and construction sites	Mineral extraction sites
		Dump sites
		Construction sites
	Artificial, non-agricultural vegetated areas	Green urban areas
		Sport and leisure facilities
Agricultural areas	Arable Land	Non-irrigated arable land
		Permanently irrigated land
	Permanent crops	Vineyards
		Fruit trees and berry plantations
		Olive groves
	Heterogeneous agricultural areas	Annual crops associated with permanent crops
		Complex cultivation patterns
		Land principally occupied by agriculture with significant areas of natural vegetation
Forest and semi-natural areas	Forests	Broad-leaved forest
		Coniferous forest
		Mixed forest
	Scrub and/or herbaceous vegetation associations	Natural grasslands
		Moors and heathland
		Sclerophyllous vegetation

Table 1 (continued)

Level 1	Level 2	**Level 3**
		Transitional woodland-shrub
	Open spaces with little or no vegetation	Beaches dunes sands
		Bare rocks
		Sparsely vegetated areas
		Burnt areas
Wetlands	Inland wetlands	Inland marshes
	Marine wetlands	Salt marshes
		Salines
Water bodies	Inland waters	Water courses
		Water bodies
	Marine waters	Coastal lagoons
		Sea and ocean

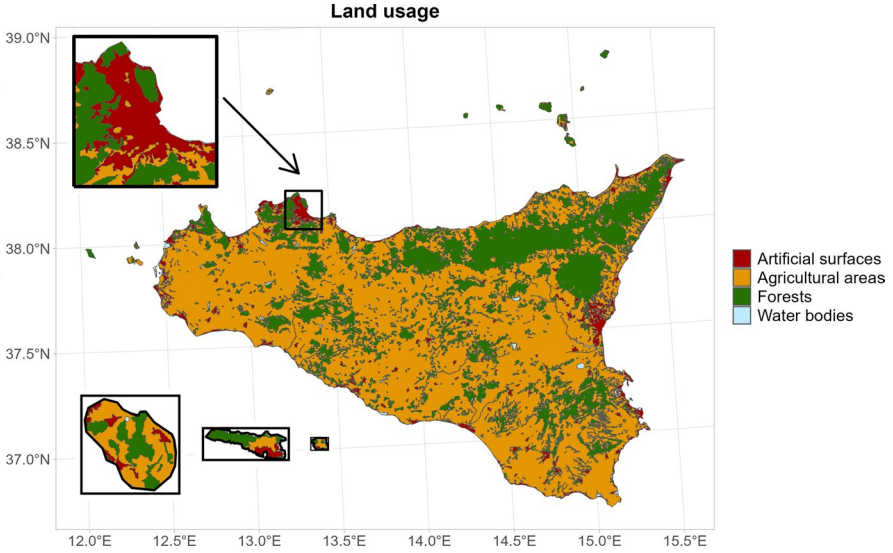

Fig. 5 Land usage in Sicily as reported in the CORINE Land Cover dataset. The inset map included in the top-left zooms around the area of Palermo (which is the capital of the region), whereas the three inset maps displayed in the bottom-left showcase the land usage for three islands located far South from the mainland (see Fig. 3)

important metropolitan areas, and water bodies represent the smallest portion. These findings are summarised in Table 2, which also contains a set of summary statistics regarding the spatial distribution of fires per land use type. More than 90% of fires have occurred in areas classified as forest or agricultural territories. Nevertheless, if we consider the intensity of the events (i.e. number of points divided by the area of the corresponding land type in square metres), we can actually notice that artificial surfaces can be regarded as the riskiest land type in terms of fires.

Table 2 Summary statistics of fire occurrences per land use type

	Area perc.	# of points	Perc. of points	Intensity ($\times 10^{-7}$)
Artificial surfaces	0.050	858	0.081	6.591
Agricultural areas	0.683	6529	0.613	3.700
Forest and semi-natural areas	0.259	3224	0.303	4.813
Water bodies	0.007	45	0.004	2.476

The four columns contain: (1) percentage of area; (2) raw number of fires; (3) percentage of fires; (4) intensity of events (number of points divided by the area of the corresponding land type in square metres)

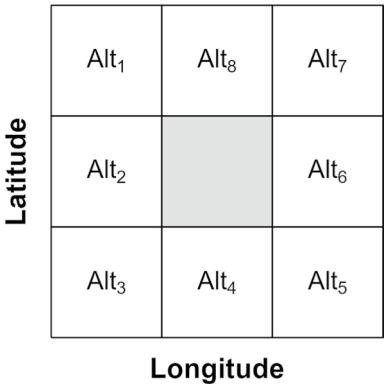

Fig. 6 3×3 grid map summarising the notation adopted in Eq. 2 to approximate the altitude variations along the two axis directions with respect to the central cell of the scheme (which is coloured in grey)

2.3 Elevation and slope

We downloaded the Digital Elevation Model (DEM) data for the Sicily region from the webpage[3] of the National Institute of Geophysics and Vulcanology (Tarquini et al. 2023), obtaining a raster object with 10 m resolution that represents the altitude in the area of interest. Then, we used Horn's formula (1981) through the GDAL DEM utility command-line tool (GDAL/OGR contributors 2023) to estimate the slope angle in Sicily and therefore build an additional variable to use in our analysis.

More precisely, let $f(\text{lon}, \text{lat})$ denote a 2D surface representing the altitude (henceforth denoted as "Alt") in a spatial region as a function of longitude ("lon") and latitude ("lat"). Furthermore, let $\nabla f = \left(\frac{\partial \text{Alt}}{\partial \text{lon}}, \frac{\partial \text{Alt}}{\partial \text{lat}} \right)'$ be its gradient vector. According to standard mathematical theory, ∇f represents the direction of the steepest ascent of the 2D surface and the corresponding magnitude is given by

$$|\nabla f| = \sqrt{\left(\frac{\partial \text{Alt}}{\partial \text{lon}} \right)^2 + \left(\frac{\partial \text{Alt}}{\partial \text{lat}} \right)^2}.$$

The slope, say ϕ, of a 2D surface at a given point represents the angle (in radians from the horizontal axis) of the rise in the direction of the gradient and it is equal to

$$\phi = \arctan(|\nabla f|) = \arctan\left(\sqrt{\left(\frac{\partial \text{Alt}}{\partial \text{lon}} \right)^2 + \left(\frac{\partial \text{Alt}}{\partial \text{lat}} \right)^2} \right). \tag{1}$$

However, real-world surfaces (like the altitude in the Sicily region) can be rarely represented by analytic functions such as the one described above. Therefore, given a DEM, we approximate the two partial derivatives in Eq. (1) using grid computations (see Fig. 6) and finite-order methods. For example, Horn's formula (Horn 1981) estimates the rate of altitude variation along the East–West (Longitude) and North–South (Latitude) directions as

[3] URL: https://tinitaly.pi.ingv.it/Download_Area1_1.html. Last access: Nov. 2023.

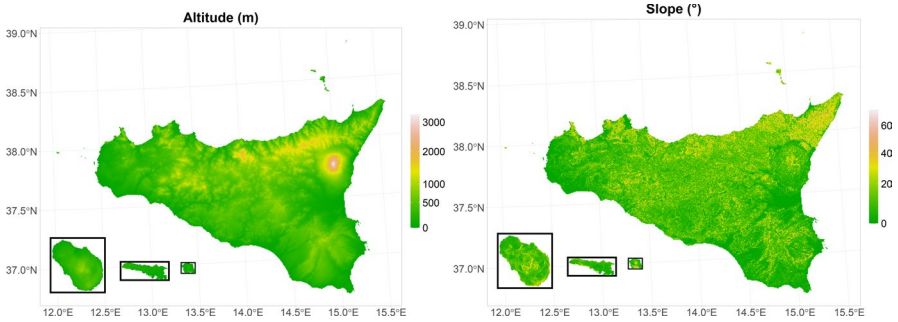

Fig. 7 Spatial variables representing the altitude (left) and slope (right) in Sicily

$$\begin{aligned} \frac{\partial \text{Alt}}{\partial \text{lon}} &\approx \frac{(\text{Alt}_1 + 2\text{Alt}_2 + \text{Alt}_3) - (\text{Alt}_7 + 2\text{Alt}_6 + \text{Alt}_5)}{8\Delta\text{lon}}, \\ \frac{\partial \text{Alt}}{\partial \text{lat}} &\approx \frac{(\text{Alt}_1 + 2\text{Alt}_8 + \text{Alt}_7) - (\text{Alt}_3 + 2\text{Alt}_4 + \text{Alt}_5)}{8\Delta\text{lat}}, \end{aligned} \tag{2}$$

where $\text{Alt}_1, \ldots, \text{Alt}_8$ denote the altitude values in a 3×3 grid surrounding a cell under analysis following the scheme detailed in Fig. 6, whereas Δlon and Δlat represent, respectively, the spacing between DEM cells in the horizontal and vertical directions. We refer to Weih and Mattson (2004) and the GDAL DEM webpage[4] for alternative approximation methods and more details regarding the computations.

The resulting altitude and slope spatial variables are displayed in Fig. 7. As we can see, the first map clearly highlights the Etna volcano (located in the eastern part of the region), the mountain chains in the North and northeast areas (Monti Nebrodi), and the Catania valley surrounding the homonymous city. The slope surface is generally flat or almost flat, apart from some regions in the northeast (near Messina).

2.4 Normalised Difference Vegetation Index (NDVI)

The Normalised Difference Vegetation Index (NDVI) is a dimensionless indicator of the greenness of the biomes. It provides effective means for assessing vegetation health and density over large areas and its application spans a variety of research areas such as environmental monitoring, agricultural management or, as in our paper, fires analysis and prevention (Ba et al. 2022). NDVI values are calculated by comparing the visible and the near-infrared (NIR) sunlight reflected from the land surface. In formula:

$$\text{NDVI} = \frac{\text{meanNIR} - \text{meanRED}}{\text{meanNIR} + \text{meanRED}} \tag{3}$$

[4] URL: https://gdal.org/programs/gdaldem.html.

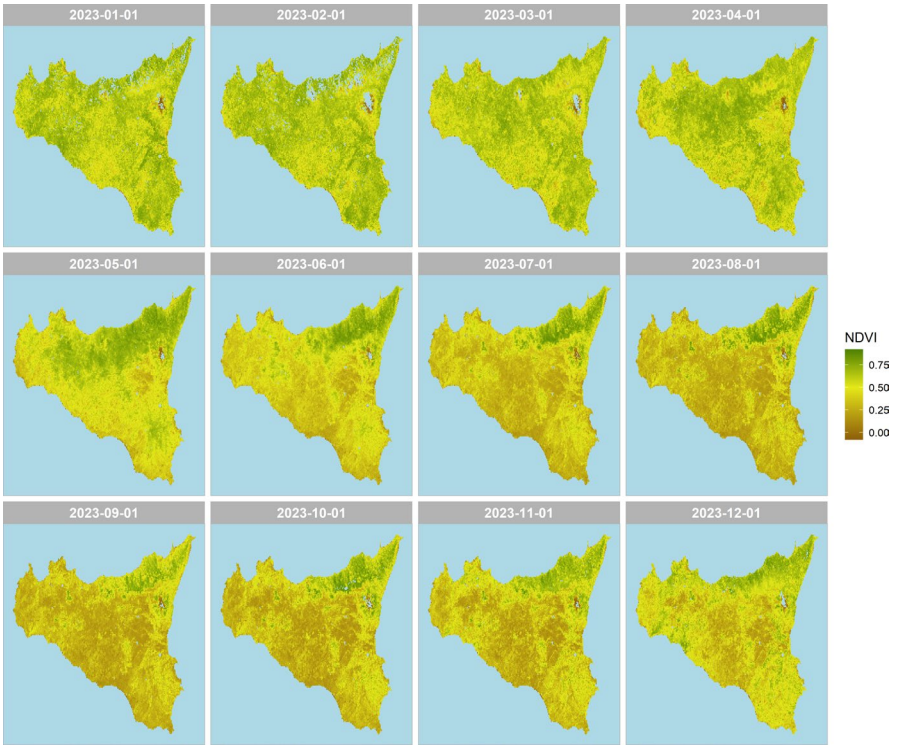

Fig. 8 Monthly estimates of NDVI values for the Sicily region during 2023. To simplify the graphical representation, we displayed only the pixels pertaining to the mainland

where meanNIR and meanRED are, respectively, the averages of the 4 NIR and RED bands recorded by satellites orbiting the Earth, such as Sentinel-3A and -3B OLCI.

Global NDVI data for the year 2023 were sourced from a publicly available repository[5] managed by the Copernicus Land Monitoring Service (CLMS) project. Our dataset comprises 36 rasters, with three observations per month, each capturing approximately 10 days of NDVI estimates at a spatial resolution of 300 ms. Post-download, each raster was cropped to retain only the pixels corresponding to the Sicily region. It is worth noting that the CLMS service rescales the NDVI values [as defined in Eq. (3)] from -0.08 to 0.92.

Figure 8 displayes the monthly NDVI values for the Sicily region during 2023. Each figure was derived taking the pixel-wise average of the rasters recorded during a given month. As we can see, the colors range from brown (lower NDVI, indicating arid land and dry biomes) to green (higher NDVI, indicating lush, healthy biomes). The vegetation density increases and peaks during early spring (March and April), and then declines during the hot and dry summer months

[5] URL: https://globalland.vito.be/download/manifest/ndvi_300m_v2_10daily_netcdf/manifest_clms_global_ndvi_300m_v2_10daily_netcdf_latest.txt. Accessed in September 2024.

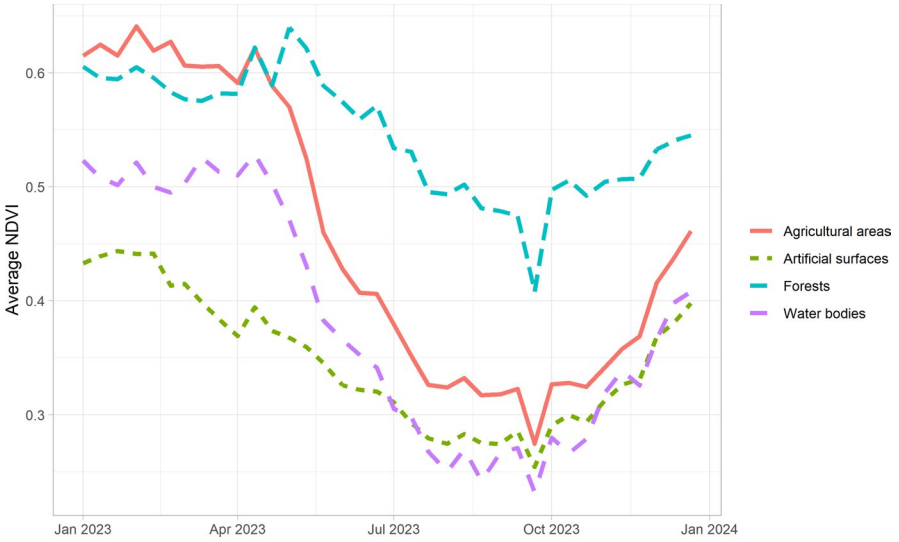

Fig. 9 Time series of 10-days average NDVI values in Sicily during 2023 for each land use type

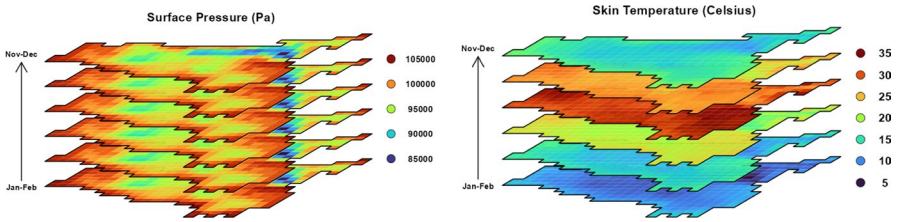

Fig. 10 Spatio-temporal variables representing the daily means of the surface pressure and skin temperature in Sicily in 2023

(from July to September). Some recovery is visible in the autumn (October and November).

There is a clear parallelism between NDVI values and the spatio-temporal distribution of fires reported in Fig. 4. Furthermore, due to their nature and definition, there exists also a relationship between NDVI and land use that we explored by computing the time series of average NDVI values for each 10-day interval and for each land use category. The results are displayed in Fig. 9 which reveals distinct seasonal patterns, with forests generally maintaining higher NDVI values throughout the year compared to agricultural areas and artificial surfaces. The lowest NDVI values for agricultural areas coincide with the summer months, reflecting the impact of seasonal crop cycles and potential drought stress, while forested areas exhibit a less pronounced decline during the same period.

2.5 Environmental variables from Copernicus ERA5

Finally, Copernicus ERA5 data contains additional environmental spatio-temporal variables that were downloaded from https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview. All of them are obtained in a 38×34 spatial grid with $10\,\text{km} \times 10\,\text{km}$ raster cells and hourly resolution for each day of the year.

Figure 10 contains an example of the spatio-temporal resolution of such variables, averaged temporally to aid visualization, highlighting the more evident spatial variability of the surface pressure if compared to its own temporal variability. On the contrary, the skin temperature varies more temporally than spatially. The complete description of all the available environmental variables comes as follows.

The `10u` and `10v` variables are the eastward and northward components, respectively, of the 10 m wind. They represent the horizontal speed of air moving towards East and North, at a height of ten metres above the surface of the Earth, in metres per second. In particular, the U wind component is parallel to the x-axis (i.e. longitude) while the V wind component is parallel to the y-axis (i.e. latitude). This implies that a positive U wind comes from the West, and a positive V wind comes from the South.

The surface pressure `sp` is the pressure (force per unit area) of the atmosphere on the surface of the land, sea, and in-land water. It is a measure of the weight of all the air in a column vertically above the area of the Earth's surface represented at a fixed point. The units of this variable are Pascals (Pa).

The total precipitation `tp` is the accumulated liquid and frozen water, comprising rain and snow, that falls to the Earth's surface, not including fog, dew, or the precipitation that evaporates in the atmosphere before it lands at the surface of the Earth. This variable is the total amount of water accumulated over a particular time period, which depends on the data extracted. The units of this variable are depth in metres of water equivalent.

The dew point temperature `d2m` is the temperature to which the air, at 2 ms above the surface of the Earth, would have to be cooled for saturation to occur. The air temperature at 2 m above the surface of land, sea, or in-land waters is indicated by `t2m`. The skin temperature `skt` is the temperature of the surface of the Earth, that is, the theoretical temperature that is required to satisfy the surface energy balance. It represents the temperature of the uppermost surface layer, which has no heat capacity and so can respond instantaneously to changes in surface fluxes. All the soil temperatures `stl1`, `stl2`, `stl3`, and `stl4` are the temperature of the soil in the middle of that given layer. The Advancing Global NWP defines four soil layers through the International Collaboration (ECMWF) Integrated Forecasting System (IFS), as follows: Layer 1 (0–7 cm), Layer 2 (7–28 cm), Layer 3 (28–100 cm), and Layer 4 (100–289 cm). Soil temperature is set at the middle of each layer, and heat transfer is calculated at the interfaces between them. It is assumed that there is no heat transfer out of the bottom of the lowest layer. In this paper, all the variables representing temperatures have units of kelvin (K) but have been converted to degrees Celsius (C) by subtracting 273.15. They are calculated by interpolating between

the lowest model level and the Earth's surface, taking account of the atmospheric conditions.

3 Spatio-temporal point processes

This section presents the statistical modelling methodology employed in the paper. We consider a spatio-temporal point process with no multiple points as a random countable subset X of $\mathbb{R}^2 \times \mathbb{R}$, where a point $(u, t) \in X$ corresponds to an event at location $u \in \mathbb{R}^2$ occurring at time $t \in \mathbb{R}$. A typical realisation of a spatio-temporal point process X on $\mathbb{R}^2 \times \mathbb{R}$ is a finite set $\{(u_i, t_i)\}_{i=1}^n$ of distinct points within a bounded spatio-temporal region $W \times T \subset \mathbb{R}^2 \times \mathbb{R}$, where $n \geq 0$ is not fixed in advance. Denoting by ℓ the Lebesgue measure, the bounded region will have area $\ell(W) > 0$ and length $\ell(T) > 0$. In this context, $N(A \times B)$ denotes the number of points of a set $(A \times B) \cap X$, where $A \subseteq W$ and $B \subseteq T$. When $N(W \times T) < \infty$ with probability 1, which holds e.g. if X is defined on a bounded set, we call X a finite spatio-temporal point process (Daley and Vere-Jones 2007).

For a given event (u, t), the events that are close to (u, t) in both space and time, for each spatial distance r and time lag h, are given by the corresponding spatio-temporal cylindrical neighbourhood of the event (u, t), which can be expressed by the Cartesian product as

$$b((u, t), r, h) = \{(v, s) : ||u - v|| \leq r, |t - s| \leq h\}, \qquad (u, t), (v, s) \in W \times T,$$

where $||\cdot||$ denotes the Euclidean distance in $\mathbb{R}^2$ and $|\cdot|$ is the absolute value. Note that $b((u, t), r, h)$ is a cylinder with centre (u, t), radius r, and height $2h$.

Product densities $\lambda^{(k)}$, $k \in \mathbb{N}$ and $k \geq 1$, arguably the main tools in the statistical analysis of point processes, may be defined through the Campbell Theorem (see Daley and Vere-Jones 2007), that constitutes an essential result in spatio-temporal point process theory. It states that, given a spatio-temporal point process X, and denoting by ζ_i the spatio-temporal location (u_i, t_i) of the ith point of the pattern, for any non-negative function f on $(\mathbb{R}^2 \times \mathbb{R})^k$ the following holds

$$\mathbb{E}\left[\sum_{\zeta_1,\ldots,\zeta_k \in X}^{\neq} f(\zeta_1, \ldots, \zeta_k)\right] = \int_{\mathbb{R}^2\times\mathbb{R}} \cdots \int_{\mathbb{R}^2\times\mathbb{R}} f(\zeta_1, \ldots, \zeta_k)\lambda^{(k)}(\zeta_1, \ldots, \zeta_k) \prod_{i=1}^{k} \mathrm{d}\zeta_i,$$

where $\neq$ indicates that the sum is over distinct values. In particular, for $k = 1$ these functions are called the *intensity function* λ. Broadly speaking, the intensity function describes the rate at which the events occur in the given spatio-temporal region, representing the point process analogues of the mean function of a real-valued process. Then, the first-order intensity function is formally defined as

$$\lambda(u, t) = \lim_{\ell(\mathrm{d}u\times\mathrm{d}t)\to 0} \frac{\mathbb{E}[N(\mathrm{d}u \times \mathrm{d}t)]}{\ell(\mathrm{d}u \times \mathrm{d}t)},$$

where $\mathrm{d}u \times \mathrm{d}t$ defines a small region around the point (u, t), and therefore $\ell(\mathrm{d}u \times \mathrm{d}t)$ is its volume.

3.1 Spatio-temporal Poisson point processes

The description of the observed point pattern intensity is a crucial issue when dealing with spatio-temporal point pattern data, and specifying a statistical model is a very effective way compared to analyzing data by calculating summary statistics. Formulating and adapting a statistical model to the data allows taking into account effects that otherwise could introduce distortion in the analysis (Baddeley et al. 2015).

When dealing with intensity estimation for spatio-temporal point processes, it is quite common to assume that the intensity function $\lambda(u,t)$ is separable in space and time (Diggle 2013; Gabriel and Diggle 2009). Under this assumption, the spatio-temporal intensity function is given by the product $\lambda(u,t) = \lambda(u)\lambda(t)$ where $\lambda(u)$ and $\lambda(t)$ are non-negative functions on W and T, respectively. When separability of the spatial and temporal component is not plausible for the data, a non-separable specification of the intensity function is more advisable.

Under non-separability, we may assume that the template model is a Poisson process, with a parametric intensity or rate function

$$\lambda(u,t;\boldsymbol{\theta}), \qquad u \in W, \quad t \in T, \quad \boldsymbol{\theta} \in \Theta,$$

where $\boldsymbol{\theta}$ are some unknown parameters.

The log-likelihood is

$$\log L(\boldsymbol{\theta}) = \sum_i \log \lambda(u_i, t_i; \boldsymbol{\theta}) - \int_W \int_T \lambda(u,t;\boldsymbol{\theta}) \mathrm{d}t \mathrm{d}u \tag{4}$$

up to an additive constant, where the sum is over all points (u_i, t_i) in the point pattern $\mathbf{x}$ (Daley and Vere-Jones 2007).

Throughout the paper, we shall consider intensity models of log-linear form

$$\lambda(u,t;\boldsymbol{\theta}) = \exp(\boldsymbol{\theta}^\top \mathbf{Z}(u,t)), \quad u \in W, \quad t \in T$$

where $\mathbf{Z}(u,t) = \{Z_1(u,t), \ldots, Z_p(u,t)\}$ are known spatio-temporal covariate functions.

The non-separable specification has the advantage of giving the possibility to include purely spatial and purely temporal covariates, denoted by $\mathbf{Z}_S(u)$ and $\mathbf{Z}_T(t)$, in addition the spatio-temporal ones, with the following general formulation

$$\lambda(u,t) = \exp(\theta_0 + \boldsymbol{\theta}_S^\top \mathbf{Z}_S(u) + \boldsymbol{\theta}_T^\top \mathbf{Z}_T(t) + \boldsymbol{\theta}^\top \mathbf{Z}(u,t)).$$

3.2 Spatio-temporal Poisson point process model estimation

For model estimation, the challenge is typically to compute the integral in the log-likelihood (4). To solve this problem, we employ the finite cubature approximation to the log-likelihood of D'Angelo et al. (2023) and D'Angelo and Adelfio (2024a),

called *cubature scheme*. Such methodology is implemented in the `stopp R` package (D'Angelo and Adelfio 2024b, c).

The *cubature scheme* procedure begins by renaming the data points as $x_1, \dots, x_n$ with $(u_i, t_i) = x_i$ for $i = 1, \dots, n$, then generate m additional "dummy points" $(u_{n+1}, t_{n+1}) \dots, (u_{m+n}, t_{m+n})$ to form a set of $n + m$ cubature points (where $m > n$). Then we determine cubature weights $a_1, \dots, a_m$ so that integrals in (4) can be approximated by a Riemann sum

$$\int_W \int_T \lambda(u, t; \boldsymbol{\theta}) \mathrm{d}t \mathrm{d}u \approx \sum_{k=1}^{n+m} a_k \lambda(u_k, t_k; \boldsymbol{\theta})$$

where a_k are the cubature weights such that $\sum_{k=1}^{n+m} a_k = \ell(W \times T)$.

The log-likelihood (4) of the template model can be approximated by

$$\begin{aligned} \log L(\boldsymbol{\theta}) &\approx \sum_k \log \lambda(u_k, t_k; \boldsymbol{\theta}) + \sum_k (1 - \lambda(u_k, t_k; \boldsymbol{\theta})) a_k \\ &= \sum_{k=1}^{n+m} \left(e_k \log \lambda(u_k, t_k; \boldsymbol{\theta}) + (1 - \lambda(u_k, t_k; \boldsymbol{\theta})) a_k \right) \end{aligned} \tag{5}$$

where e_k is the indicator that equals 1 if (u_k, t_k) is a data point.

Writing $y_k = e_k / a_k$, (5) becomes

$$\log L(\boldsymbol{\theta}) \approx \sum_k a_k (y_k \log \lambda(u_k, t_k; \boldsymbol{\theta}) - \lambda(u_k, t_k; \boldsymbol{\theta})) + \sum_k a_k. \tag{6}$$

Apart from the constant $\sum_k a_k$, this expression is formally equivalent to the weighted log-likelihood of a Poisson regression model with responses y_k and and weights a_k. This can be maximised using standard Generalized Linear Models (GLM) software.

To fit such models, the spatio-temporal cubature scheme is obtained by defining a spatio-temporal partition of $W \times T$ into cubes C_k of equal volume v, assigning the weight $a_k = v / n_k$ to each cubature point (dummy or data), where n_k is the number of points that lie in the same cube as the point u_k. The number of dummy points should be sufficient for an accurate estimate of the likelihood, and numerical experiments are currently under development to give guidelines on this aspect (D'Angelo and Adelfio 2024a).

Given the above, it is easy to understand that to fit spatio-temporal Poisson point process models with first-order intensity depending on external covariates, their values must be known in every data and dummy point location.

This can require additional effort in the very common case in which the locations of the covariates do not correspond to the locations of points. This may happen when the sampling of the covariates is somehow burdensome.

For instance, spatial smoothing of the numeric values observed at the covariate locations can be employed through the inverse-distance weighting (Shepard 1968). Alternatives include other spatio-temporal smoothings using a Gaussian kernel weighting, which would lead to the Nadaraya-Watson smoother (Nadaraya

1964, 1989; Watson 1964), in addition to the most used kriging (Matheron 1963) and nearest neighbors interpolation.

For our application, the spatial resolution of the available covariates is so detailed that it was sufficient just to take the closest value to the observed point.

3.3 Local spatio-temporal Poisson point processes

In local likelihood estimation of Poisson processes (Loader 1999), the estimated intensity at (u, t) is taken to be the plug-in value

$$\hat{\lambda}(u,t) = \lambda(u,t;\hat{\boldsymbol{\theta}}(u,t))$$

associated with the fitted parameter vector at (u, t).

Inference in local models is performed through the fitting of a GLM using a localised version of the cubature scheme, firstly introduced in the spatio-temporal framework by D'Angelo et al. (2023).

The local log-likelihood associated with the spatio-temporal location (v, s) is given by

$$\begin{aligned}\log L((v,s);\boldsymbol{\theta}) = &\sum_i w_{\sigma_s}(u_i - v) w_{\sigma_t}(t_i - s) \log \lambda(u_i, t_i;\boldsymbol{\theta}) \\ &- \int_W \int_T \lambda(u,t;\boldsymbol{\theta}) w_{\sigma_s}(u_i - v) w_{\sigma_t}(t_i - s) \mathrm{d}t \mathrm{d}u\end{aligned} \tag{7}$$

where w_{σ_s} and w_{σ_t} are weight functions, and $\sigma_s, \sigma_t > 0$ are the smoothing bandwidths. It is not necessary to assume that w_{σ_s} and w_{σ_t} are probability densities. For simplicity, we shall consider only kernels of fixed bandwidth, even though spatially adaptive kernels could also be used. Note that if the template model is the homogeneous Poisson process with intensity λ, then the local likelihood estimate $\hat{\lambda}(v, s)$ reduces to the kernel estimator of the point process intensity (Diggle 2013) with kernel proportional to $w_{\sigma_s} w_{\sigma_t}$.

We now use in (7) a similar approximation as in (6) for the local log-likelihood associated with each desired location $(v, s) \in W \times T$

$$\log L((v,s);\boldsymbol{\theta}) \approx \sum_k w_k(v,s) a_k (y_k \log \lambda(u_k, t_k;\boldsymbol{\theta}) - \lambda(u_k, t_k;\boldsymbol{\theta})) + \sum_k w_k(v,s) a_k.$$

Basically, for each desired location (v, s), we replace the vector of quadrature weights a_k by $a_k(v, s) = w_k(v, s)a_k$ and use the GLM software to fit the Poisson regression.

In this paper, $w_k(v,s) = w_k(x,y,t) = w_{\sigma_x}(x - x_k) w_{\sigma_y}(y - y_k) w_{\sigma_t}(t - t_k)$ are three Gaussian kernels, whose bandwidths $\{\sigma_x, \sigma_y, \sigma_t\}$ are selected following Silverman Silverman (2018)'s rule.

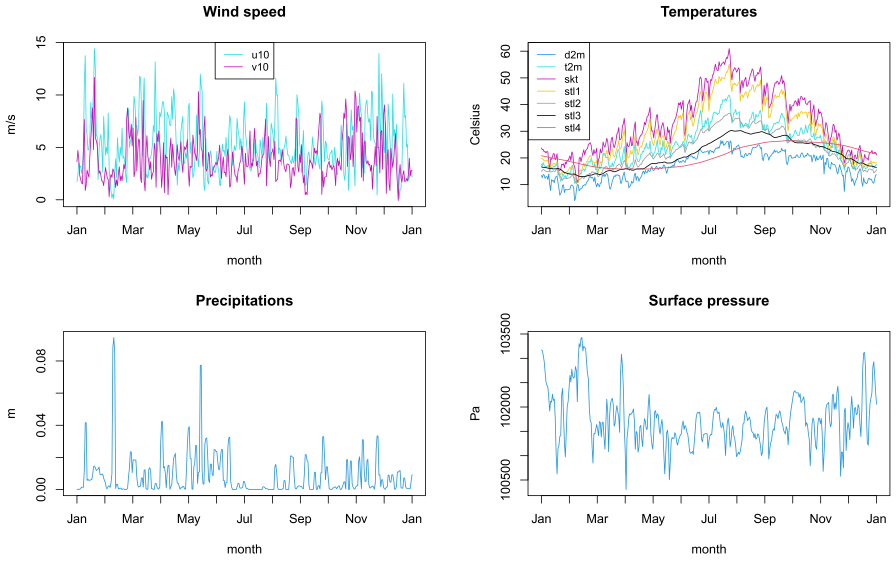

Fig. 11 Some temporal covariates obtained from the ERA5 environmental variables: daily maxima of the wind components, temperatures, precipitation and surface pressure

4 Analysis of the fires data and results

This section is devoted to the presentation of the spatio-temporal analysis and the modelling results. First, we fit a non-separable spatio-temporal Poisson model, allowing us to assess the significance of the potential spatio-temporal covariates and interpret their effects on the spatio-temporal intensity of fires. Moreover, we continue by fitting a local version of the first proposed model to obtain a better fitting to the data and get further insights into the effects of the environmental covariates, which we find to vary both in time and space. Finally, we also report a comparison between the two models, proving the better fitting of the local model.

4.1 Global model

In this section, we present the results obtained fitting the non-separable spatio-temporal intensity function presented in Sect. 3 to the data introduced in Sect. 2. Starting from the model containing all the spatio-temporal covariates previously introduced and ontinuing with a backward procedure based on the comparison of the AIC values, the chosen model has a linear predictor that includes a parametric expression for some of the spatio-temporal covariates, as follows:

$$\begin{aligned}\hat{\lambda}(u,t) = \exp(&\hat{\theta}_0 + \hat{\theta}_1 \cdot \texttt{LandUse}(u) + \hat{\theta}_2 \cdot \texttt{NDVI}(u,t) + \hat{\theta}_3 \cdot \texttt{Elevation}(u) \\ &+ \hat{\theta}_4 \cdot \texttt{Slope}(u) + \hat{\theta}_5 \cdot \texttt{WindSpeedFromSouth}(u,t) \\ &+ \hat{\theta}_6 \cdot \texttt{SkinTemperature}(u,t) + \hat{\theta}_7 \cdot \texttt{TotalPrecipitation}(u,t)).\end{aligned} \tag{8}$$

Note that Skin Temperature has been selected among other temperature-related variables, and Surface Pressure and Wind Speed from the East have been excluded due to lack of statistical significance.

In order to fit this model, the temporal resolution of the ERA5 environmental covariates, originally recorded on an hourly level, should match that of the fire data. For this reason, the hourly data have been summarised taking the daily maxima. The resulting covariate values are shown in Fig. 11 in their purely temporal dimension. As evident, the seven temperatures appear highly correlated among themselves and strongly depend on the period of the year. In general, the deeper the surface layers of detection, the smoother the temperature variability. Figure 12 further illustrates such high correlations, especially among temperature variables.

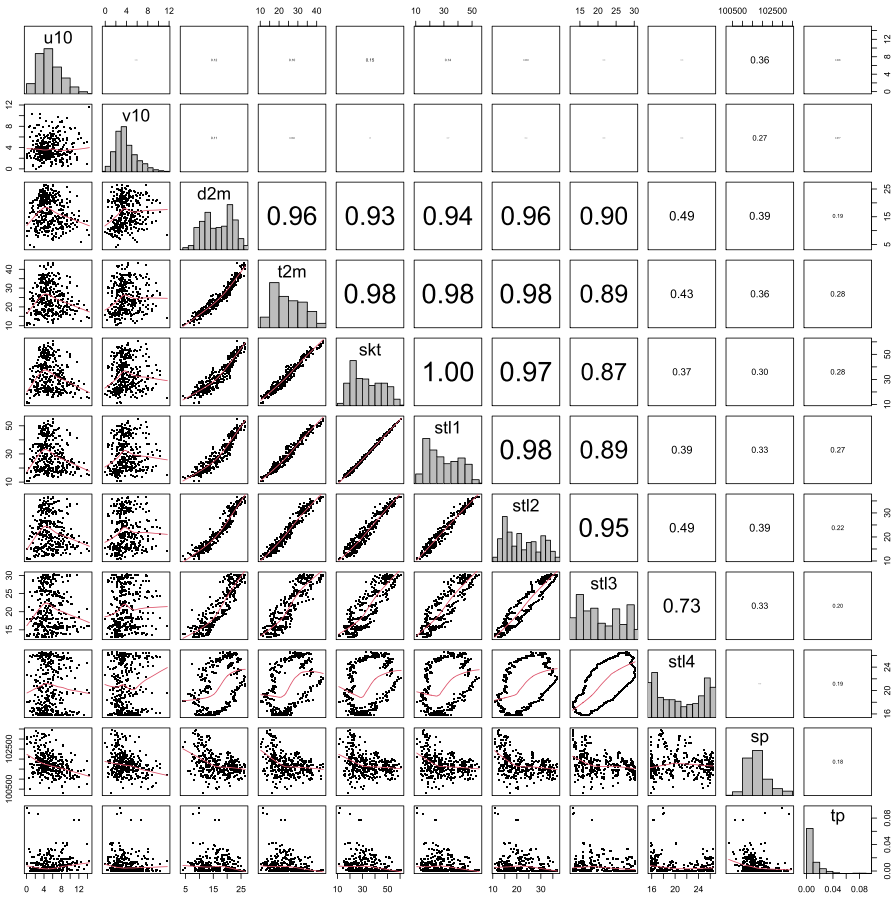

Fig. 12 Correlations among the temporal covariates obtained from the ERA5 environmental variables. *Main diagonal*: the univariate distributions of the variables. *Lower panels*: the scatterplots of each pair of variables with a smoothing function overlapped in red. *Upper panels*: the correlations among the variables

Table 3 reports the estimated coefficients of the fitted spatio-temporal Poisson model (8). In addition to the intercept, the first three rows correspond to the categories of the *land use* variable, with *Artificial surfaces* serving as the baseline. Also note that the elevation has been converted from meters to kilometres, to obtain estimated coefficients on the same magnitude as the other ones.

Considering the different land usages, only *Forest and semi-natural areas* display a significant and negative effect, implying a considerable decrease in the fire occurrence intensity moving from *Artificial Surfaces* to woodland areas. On the other hand, *Agricultural areas* and *Waterbodies* are not significantly different from the baseline. The coefficient obtained for the NDVI variable is negative and strongly significant, indicating that areas with higher NDVI values, which correspond to denser and healthier vegetation, are associated with a lower intensity of fire occurrences, probably because of their higher resistance to fires. These results are coherent with the descriptive analyses included in Table 2 and Sect. 2.4.

Both elevation and slope exhibit a significant effect, with their increase being associated with higher intensity. Wind speed demonstrates a notable positive influence, implying that higher wind speed from the South correlates with increased fire counts, highlighting the role of the *Scirocco* wind in fire dynamics. *Scirocco* is a Mediterranean wind that comes from the Sahara and brings heat and dust from African coastal regions. As expected, temperature also shows a significant positive relationship with fire counts. Precipitation reasonably reveals a substantial negative effect, indicating fewer fires with higher daily precipitation, highlighting its mitigating effect. This model has allowed us to obtain interesting results on the explanation of the fire occurrences in Sicily in 2023 thanks to the interpretation of the covariates' effects. However, these effects may not be constant throughout the year and in different regions of Sicily. Indeed, fire occurrence is a complex phenomenon that might not be totally captured by a model with constant parameters in space and time. This leads us to consider a localized version of the proposed model, which we report in the following section.

Table 3 Estimated coefficients of the fitted spatial Poisson point process model

	Estimate	Std. Error	z value	p value
Intercept	− 43.274	0.865	− 50.039	< 0.001
Agricultural areas	− 0.022	0.048	− 0.463	0.643
Forest and semi-natural areas	− 0.205	0.059	− 3.485	< 0.001
Wetlands and water bodies	0.210	0.182	1.151	0.250
NDVI	− 4.790	0.106	− 45.000	< 0.001
Elevation	0.250	0.032	7.610	< 0.001
Slope	0.021	0.001	14.372	< 0.001
Wind speed from South	0.034	0.006	5.262	< 0.001
Skin temperature	0.128	45.005	79.375	< 0.001
Total precipitation	− 54.544	7.470	7.302	< 0.001

The reference category for the land use dummy variable (included in the second, third, and fourth rows) is *Artificial Surfaces*

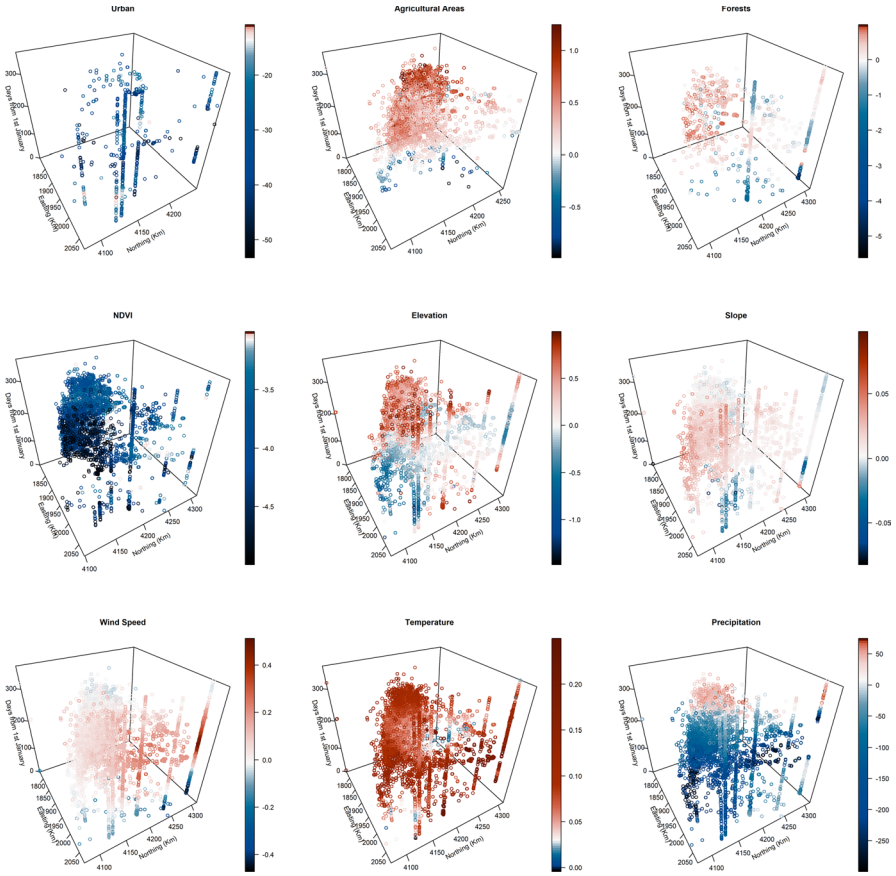

Fig. 13 Local estimates of the local Poisson model described in Eq. (9). The first three plots show the local estimates related to the land use variable (excluding the *Waterbodies* level), whereas the other plots focus on the remaining covariates

4.2 Local model

In this section, we report the result obtained when fitting the localized version of the model (8). This is expressed as:

$$\begin{aligned}\hat{\lambda}(u,t) = \exp(&\hat{\theta}_0(u,t) + \hat{\theta}_1(u,t) \cdot \texttt{LandUse}(u) + \hat{\theta}_2(u,t) \cdot \texttt{NDVI}(u,t) \\ &+ \hat{\theta}_3(u,t) \cdot \texttt{Elevation}(u) + \hat{\theta}_4(u,t) \cdot \texttt{Slope}(u) \\ &+ \hat{\theta}_5(u,t) \cdot \texttt{WindSpeedFromSouth}(u,t) \\ &+ \hat{\theta}_6(u,t) \cdot \texttt{SkinTemperature}(u,t) \\ &+ \hat{\theta}_7(u,t) \cdot \texttt{TotalPrecipitation}(u,t)).\end{aligned} \tag{9}$$

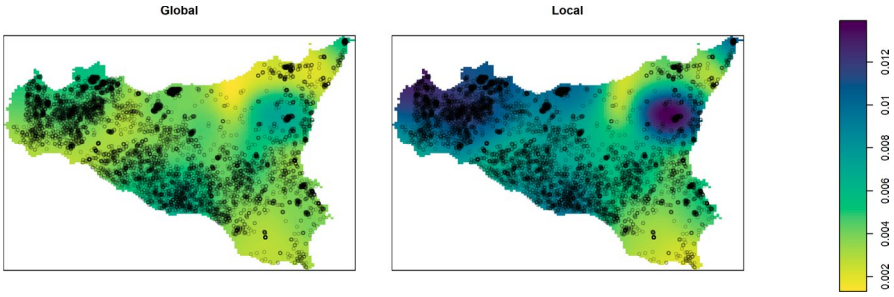

Fig. 14 Intensity fitted according to model (8), and its local version (9), displayed in space through Kernel smoothing

The local Poisson model fitting procedure provides a set of parameters for each point in the analysed pattern. Therefore, Fig. 13 shows such local estimates, displayed in space and time, rotated to focus on the East coast, to aid visual inspection. The scale colour is constructed in such a way that white values indicate coefficients estimated equal to zero, red colour devoted to positive values of the estimates, and blue colour to negative ones.

The first three plots show the local estimates for three dummy variables related to land usage. We decided to exclude the fourth category level from the graph, namely *Waterbodies*, since that's the least frequent land use type and we wanted to preserve the 3×3 plot structure to simplify the visualisation. Of course, for each of them, we only report the points occurring exactly on that specific type of land use. The remaining plots display the local estimates for each point of the analysed pattern.

From this visual inspection, we gain interesting information on the effects of the environmental covariates employed to explain the spatio-temporal occurrence of fires. The most important result is surely the great variability of the estimates both in space and time, signifying that most of the covariates' effects actually do change with location and the period of the year. This highlights the advantage of using a local model as it allows us to detect these spatio-temporal dynamics in a way that a global model cannot.

Being in an forestal area has a negative effect on fire occurrence, consistent with the findings from the global model, but this effect is mostly observed in specific spatial regions and temporal periods of the year. This pattern suggests that while artificial land use contributes to higher fire occurrences in general, the risk shifts towards forestal areas moving West of Sicily. This is possibly due to seasonal changes in vegetation and land management practices. The NDVI plot shows that dry and arid vegetation is consistently more susceptible to fire ignition across the region throughout the year, highlighting the influence of vegetation condition on fire risk.

The effects of elevation and slope show more complex behaviours, with differences also in space. Indeed, Fig. 13 demonstrates that positive coefficients of both elevation and slope (as in the global model) only occur in the northwest part of Sicily and only in the second part of the year. Conversely, the southern region is characterized by a negative effect of the elevation and a positive effect of the slope. A

positive effect of elevation and a negative effect of the slope is instead found along the East coast. This result is not surprising given that literature has proven elevation and slope to greatly change depending on the characteristics of the analysed region (Ganteaume et al. 2013; Gabriel et al. 2017; Opitz et al. 2020).

Another major point of the local model's result is that we have been able to obtain local estimates varying both in space and time, even though the land usage, elevation, and slope covariates are purely spatial.

Finally, also the wind speed from the South, the temperature, and the precipitation show great variability, but always within a reasonable magnitude of the parameters. For instance, the temperature effects vary among only positive values, and the precipitation effects are mostly negative. The only exception is given in the extreme northeast region at the very end of the year, which shows both a negative effect of the wind speed from the South and a positive effect of the precipitation. This result could be further investigated in the future, but we note that it only refers to an extremely short period, not particularly affected by the occurrence of fires.

4.3 Diagnostics

Figure 14 shows the intensity predicted according to the fitted model in Eq. (8), and its local version (9), marginalised with respect to the temporal component.

We may notice a different displacement of the higher intensity values, always corresponding to the places with higher concentrations of fires, for both models. The major difference is that the global model displays an overall lower intensity if compared to that of the local model, which in contrast estimates the highest intensity in the urban areas of North and West Sicily, namely around the cities of Palermo and Cefalù, and the Etna region.

We further proceed by carrying out a residual analysis to understand which of the two models best fits the data.

Residual analysis typically involves examining the residuals between the observed variable and that predicted by the model. For spatio-temporal point processes, this helps in identifying model inadequacies or areas where the model could be improved.

As stated in some previous papers (Adelfio et al. 2020; Adelfio and Schoenberg 2009), the main problem when dealing with residual analysis for point processes is to find a correct definition of residuals since the one used in dependence models cannot be used for point processes. Two of the most used methods for diagnostics of spatial point processes are the inhomogeneous K-function and the (smoothed) raw residuals employed in the following.

Raw residuals for spatio-temporal point processes are calculated as the difference between the observed number of events and the expected number of events in a given region of space-time.

For an inhomogeneous Poisson process model, with fitted intensity $\hat{\lambda}(u,t)$, the predicted number of points falling in any region $W \times T$ is $\int_W \int_T \hat{\lambda}(u,t)\mathrm{d}t\mathrm{d}u$. Hence, the residual in each region $W \times T \subset \mathbb{R}^2 \times \mathbb{R}$ is the 'observed minus predicted' number of points falling in $W \times T$, that is

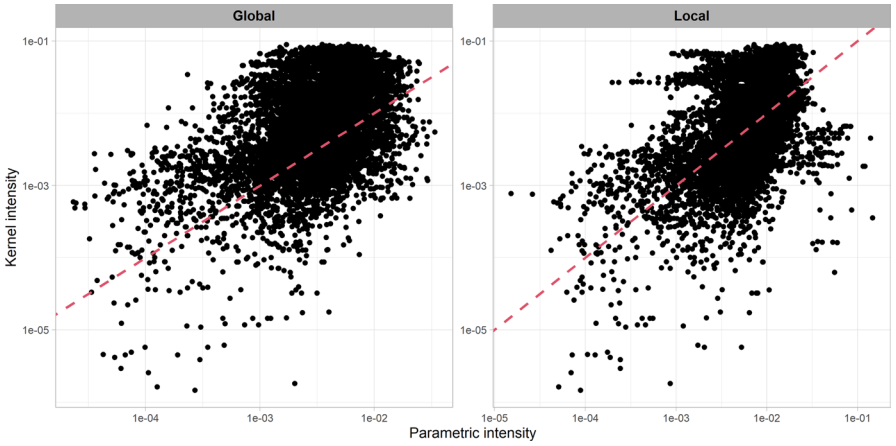

Fig. 15 Relationship between the fitted kernel intensity and the global (left panel) and local (right panel) estimated intensity, on the $\log_{10}$ scale. The red dashed line represents the 1:1 relationship, i.e. the line of perfect agreement between $\hat{\lambda}(\cdot,\cdot)$ and the non-parametric kernel estimate of the intensity

$$R(W \times T) = n(\mathbf{x} \cap (W \times T)) - \int_W \int_T \hat{\lambda}(u,t)\mathrm{d}t\mathrm{d}u,$$

where $\mathbf{x}$ is the observed point pattern, $n(\mathbf{x} \cap (W \times T))$ the number of points of $\mathbf{x}$ in the region $W \times T$, and $\hat{\lambda}(u,t)$ is the intensity of the fitted model to diagnose.

In purely spatial statistics, a common practice is to smooth such residual, therefore named 'smoothed residual fields', by replacing the number of points with a non-parametric, kernel estimate of the fitted intensity, which serves as a proxy of the 'true intensity', and by replacing the predicted number of points by a smoothed version of the (typically parametric) estimate of the intensity of the fitted model. Of course, the residual field should be approximately zero when the fitted model is close to the real one. Therefore, the best model is the one with the lowest values of the smoothed raw residuals.

In this paper, we propose to compute raw residuals as

$$r(u_i,t_i) = \tilde{\lambda}(u_i,t_i) - \hat{\lambda}(u_i,t_i), \tag{10}$$

where $\tilde{\lambda}(\cdot,\cdot)$ is a non-parametric kernel estimate of the fitted intensity $\hat{\lambda}(\cdot,\cdot)$, computing them only at observed locations. Notice that we do not employ the most classical 'smoothed residual fields' for multiple reasons. First, since the residuals of this particular application do not vary much, smoothing them would result in a further reduction of their variability, with a consequent impossibility of interpreting them. Secondly, it would be difficult to represent and therefore to grasp differences in space and time. Furthermore, we want to investigate the individual behaviour of points in terms of goodness-of-fit and therefore, using the non-smoothed residuals represents the best choice.

Figure 15 shows the relationship between the fitted Kernel intensity and the global (left panel) and local (right panel) estimated intensity, on the $\log_{10}$ scale. The

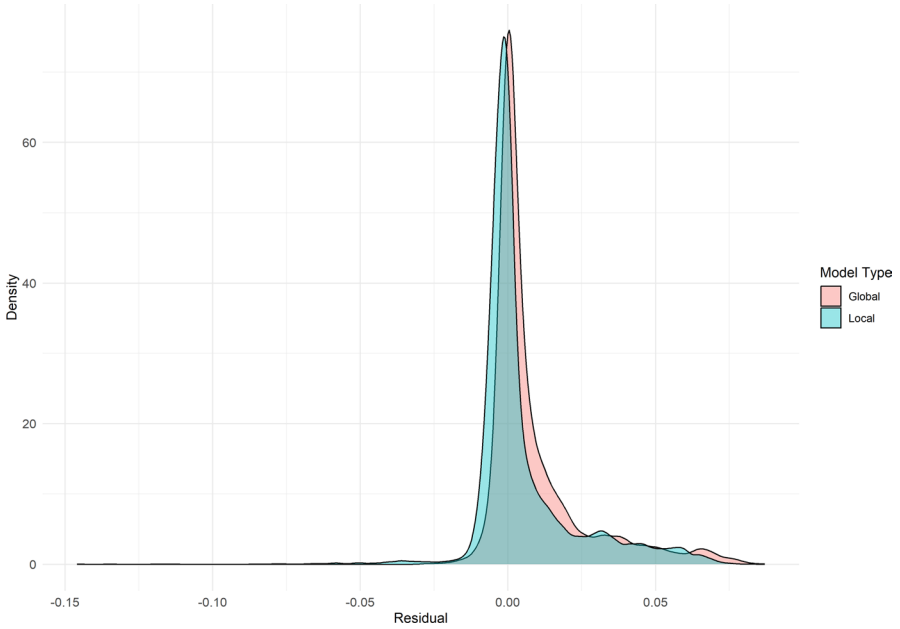

Fig. 16 Density distribution of the residuals obtained from the global and local model applying the formula described in Eq. (10)

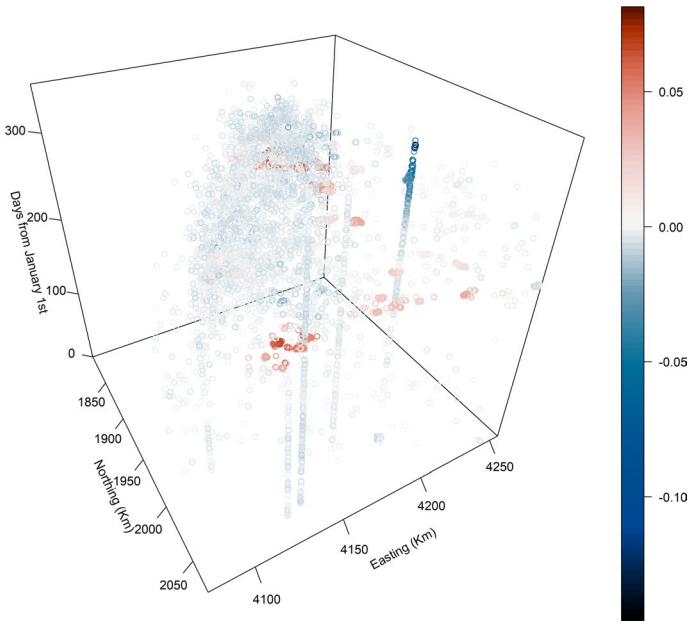

Fig. 17 Residuals of the local model displayed in space and time

red dashed line represents the 1:1 relationship, i.e. the line of perfect agreement between $\hat{\lambda}(\cdot,\cdot)$ and the non-parametric kernel estimate of the intensity.

These graphs help to understand the overall goodness-of-fit of the models. Furthermore, we may note that the correlation between the Kernel intensity and the one fitted by the global model is 0.241, while it increases when computing the correlation between the Kernel intensity and the intensity of the local model, that is, 0.351.

Figure 16 displays the density of the residual distributions for both the global and the local model, showing little differences between the two. In particular, the residual of the local model tends to be slightly more concentrated around zero. Moreover, the Mean Square Error, computed as $\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\tilde{\lambda}(u_i,t_i) - \hat{\lambda}(u_i,t_i))^2$, is lower for the local model.

These results demonstrate a good ability for both models to capture the spatio-temporal variability of fire counts, with additional advantages considering the local model.

Furthermore, looking at the residuals of the local model in Fig. 17, we do not observe regions with particular overestimation of the intensity function, except for the Etna region.

Still, some spots present underestimated intensity values, mostly in some cities along the West coast. These also correspond to some spatial clusters of fires, meaning that there is possibly room for improvement in the model specification, with particular reference to point process models for clustered point patterns.

5 Discussion and conclusions

The proposed analysis of the spatial and temporal distribution of fire occurrences in Sicily during 2023 has provided insights into the factors influencing these events, with a particular emphasis on the role of land usage and environmental variables.

To do this, we have integrated data with different spatio-temporal resolutions from very diverse sources, which represents one of the contributions of this work.

Most importantly, we have addressed the analysis of an environmental issue using complex yet easily interpretable point process models. Throughout the paper, we have emphasised that these models are interpretable, resembling regression-type models thanks to the estimation of linear effects of covariates. Unlike many traditional spatio-temporal point process models, our approach specifically accounts for the influence of external environmental spatial and spatio-temporal covariates on the point pattern under analysis in a non-separable manner. Additionally, we enhanced the possibility of conducting similar analyses on other environmental challenges, linking the proposed models to the new availability of open-source software. These constitute one of the first attempts in the literature to deal with the analysis of space-time fire occurrences via inhomogeneous point processes, representing a useful tool to better understand and predict the fire spread phenomenon in other regions.

As already mentioned, the proposed methodology integrates the effects of external covariates into the modelling approach, providing deeper insights into fire dynamics and potentially contributing to fire prevention and risk reduction.

These insights might be particularly relevant for predictive modelling, which can support proactive measures to mitigate fire risks.

Based on the results, land use has a significant effect on fire prevalence, suggesting a potential link to human activities. However, further research would be needed to explore the possible specific economic and social factors, such as connectivity to agricultural areas, that may influence the incidence of fires. The identification of artificial surfaces as a key contributor to the increased probability of fire occurrence, in some specific periods, highlights the urgent need for targeted intervention and policy measures. These results have broader implications for regional planning, resource allocation, and the development of proactive measures to manage fire risks effectively. The proposed research could suggest some directions for further insights, identifying the possible nature of fire occurrences. This leads us to argue for sustainable land-use practices and the implementation of policies aimed at minimizing the risk of fires in areas characterized by artificial surfaces.

Our analysis accounts for further environmental covariates that significantly influenced the occurrence of fires in Sicily during 2023. Our findings reveal a significant role of these environmental factors over the spatio-temporal event occurrences. Particularly, wind speed from the South and temperature emerge as crucial variables, emphasizing the importance of climatic conditions in amplifying fire risks. Conversely, our study identifies total precipitation as an environmental covariate that negatively influences fire occurrences as a mitigating factor. Additionally, our research highlights the significance of terrain characteristics in explaining the spatio-temporal distribution of fire points. In particular, both elevation and slope exhibit a positive effect on the occurrence of fires, indicating that areas with higher elevations and steeper slopes are more susceptible. As most of these results may seem intuitive, we believe that confirming their significance within the specific context of Sicily contributes to a deeper understanding of regional fire behavior. These help to localising known patterns and validating the consistency of such variables in different geographic settings, which is important for creating effective region-specific fire mitigation strategies.

In addition, the fitted local models allowed us to get further insights into the effect of these significant variables on the intensity of fire occurrences. In particular, we have found that all of them actually have different effects across Sicily and throughout 2023. This means that the effect of the variables estimated by the global model still holds true but in specific regions of Sicily and periods of the analysed year. Specifically, being in an agricultural area, rather than an urban area, has a negative effect on fire occurrences only up to spring. Elevation and slope effects show more complex behaviors, with positive estimates observed in the northwest of Sicily and in the latter half of the year, while the South exhibits a negative effect of elevation and a positive effect of slope. In the eastern coast, there is a positive effect of elevation and a negative effect of slope. Wind speed from the South, temperature, and precipitation also vary in space and time, with temperature consistently having a positive effect throughout the year, and precipitation mostly negative, except in the extreme northeast at the end of the year.

From a methodological point of view, the richness of the fires data, including the covariates, pushed us to develop and implement both the global and local spatio-temporal Poisson point process models. While these models are theoretically well-established, their practical implementation in open-source environments (such as the R software) has been limited. By addressing this gap and providing companion code, we offer a methodological contribution that extends the applicability of these models to broader contexts. This development has potential utility for experts in various scientific domains, offering a versatile tool for analysing spatio-temporal phenomena beyond fire occurrences in a local and non-separable way.

Our research comes with certain limitations. To begin with, it seems somewhat reductive not to take the size of the fires into account. This is of course due to the unavailability of such information in our dataset. Indeed, fires that burn larger areas typically have a totally different processing and history than small and local fires. Nevertheless, our non-separable approach also comes in the multitype version (see D'Angelo and Adelfio 2024b). For instance, a categorization in classes indicating the size of the fire could use as a categorical mark, and therefore the random effects of the environmental covariates associated with the size of the fire could be estimated. This could identify different effects of the covariates for fires of different size.

Secondly, it is important to note that the specific causes of the fires are not directly addressed in our study. The available data do not provide detailed information on fire ignition sources, and as such, we cannot conclusively determine the underlying causes.

However, considering the land use characteristics of Sicily, it is plausible that agricultural practices may play a role in influencing fire occurrence, especially given the dominance of agricultural land in the region. Similar to other Mediterranean areas, where fires are often linked to land management practices, the higher frequency of fires during the summer months may also be related to seasonal agricultural activities.

That said, these are speculative observations based on general patterns seen in Mediterranean regions, and we caution against drawing definitive conclusions without further specific investigation into fire causes. Future studies with more detailed data on fire ignition sources would be needed to fully explore this aspect.

Another limitation of this study is the reliance on a single year of data for the analysis. Therefore, the findings presented here may reflect specific conditions of that year and might not fully generalize to other time periods.

However, the main focus of this study is to demonstrate the applicability and effectiveness of the non-separable spatio-temporal Poisson model in analysing fire data. Future research could expand on this work by incorporating data from multiple years to better assess temporal variability. Such an extension would likely enhance the robustness of the model and provide a more comprehensive understanding of the long-term dynamics of fires in Sicily. Nonetheless, we believe that the current analysis serves as a valuable foundation for applying advanced statistical models to fire data and contributes to the broader field of spatio-temporal modelling.

Further information could be added to fire locations, such as the burned areas or the cause of ignition, and can be treated as marks. Indeed, many national institutes

have a much more detailed database with fires, including their sizes, causes, duration, spread and local conditions. Having access to such more detailed information would allow to make more precise analyses, including comparisons of fire characteristics, the influence of local conditions, and potentially even the effectiveness of different fire management strategies. For instance, Schoenberg (2004) found evidence of a lack of separability between fire occurrences and sizes due to small-scale clustering. Hence, the marks may not be separable from the points. However, this aspect is not explored in our paper due to the lack of relevant marks and because it would require the development of ad-hoc marked point pattern methodologies, which are outside the scope of our paper. Moreover, the literature on the so-called marked point process methodologies is rather specific to particular physical phenomena, like the seismic ones with the ETAS models (Ogata 1988), and quite limited for cases in the presence of multiple marks. This is the reason we chose not to include marks in our analysis, despite the fact that we believe it is an extremely interesting topic, both methodologically and practically.

Moving to other possible future research paths, a major topic that could be addressed is the definition and application of tailored diagnostic procedures to assess the need for more complex models in cases of residual clustering behaviour of points not taken into account by the already considered factors. Indeed, the employment of most known diagnostic tools based on second-order summary statistics (Adelfio and Schoenberg 2009; Adelfio et al. 2020), both global and local, becomes computationally hard as the number of points increases.

Future work could include the exploration of more complex models based on the Poisson processes employed in our paper, like the log-Gaussian Cox processes and self-exciting point process models (Siino et al. 2018; Adelfio and Chiodi 2021).

An even more interesting research path concerns adapting the aforementioned models to their local version. Nevertheless, we point out that the development of local clustered and self-exciting point process models is still an open research area (D'Angelo et al. 2023; Zhuang 2015).

Acknowledgements This publication has been prepared using European Union's Copernicus Land Monitoring Service information.

Author contributions The authors contributed equally to the work.

Funding Open access funding provided by Università degli Studi di Palermo within the CRUI-CARE Agreement. The research work of Nicoletta D'Angelo and Giada Adelfio was supported by: Targeted Research Funds 2025 (FFR 2025) of the University of Palermo (Italy); European Union - NextGenerationEU, Mission 4, Component 2, in the framework of the GRINS -Growing Resilient, INclusive and Sustainable project (GRINS PE00000018 - CUP C93C22005270001); PRIN 2022: Spatio-temporal Functional Marked Point Processes for Probabilistic Forecasting of Earthquakes 2022BN7CJP P. I. Giada Adelfio. CUP B53C24006340006. The research work of Giada Adelfio has been supported by National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.4 - Call for tender No. 3138 of 16 December 2021, rectified by Decree n.3175 of 18 December 2021 of Italian Ministry of University and Research funded by the European Union - NextGenerationEU; Award Number: Project code CN-00000033, Concession Decree No. 1034 of 17 June 2022 adopted by the Italian Ministry of University and Research, CUP UNIPA B73C22000790001, Project title "National Biodiversity Future Center - NBFC". Andrea Gilardi acknowledges the support by MUR, grant Dipartimento di Eccellenza 2023-2027. His research work is funded by the European Union - NextGenerationEU, Mission 4, Component 2, in the framework of the GRINS -Growing Resilient, INclusive and Sustainable project (GRINS

PE00000018 - CUP D43C22003110001). The research work of Alessandro Albano has been supported by the European Union - NextGenerationEU - National Sustainable Mobility Center CN00000023, Italian Ministry of University and Research Decree n. 1033- 17/06/2022, Spoke 2, CUP B73C2200076000 and Targeted Research Funds 2025 (FFR 2025) of the University of Palermo (Italy). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors do not have any conflict of interest.

References

Adelfio G, Chiodi M (2021) Including covariates in a space-time point process with application to seismicity. Stat Methods Appl 30:947–971

Adelfio G, Schoenberg FP (2009) Point process diagnostics based on weighted second-order statistics and their asymptotic properties. Ann Inst Stat Math 61(4):929–948

Adelfio G, Siino M, Mateu J, Rodríguez-Cortés FJ (2020) Some properties of local weighted second-order statistics for spatio-temporal point processes. Stoch Environ Res Risk Assess 34(1):149–168

Aldersley A, Murray SJ, Cornell SE (2011) Global and regional analysis of climate and human drivers of wildfire. Sci Total Environ 409(18):3472–3481

Ba R, Song W, Lovallo M, Zhang H, Telesca L (2022) Informational analysis of MODIS NDVI and EVI time series of sites affected and unaffected by wildfires. Phys A Stat Mech Appl 604:127911

Baddeley A, Turner R (2005) Spatstat: An R package for analyzing spatial point patterns. J Stat Softw 12(6):1–42

Baddeley A, Rubak E, Turner R (2015) Spatial point patterns: methodology and applications with R. Chapman and Hall/CRC, London

Borrajo MI, González-Manteiga W, Martínez-Miranda M (2020) Bootstrapping kernel intensity estimation for inhomogeneous point processes with spatial covariates. Comput Stat Data Anal 144:106875

Borrajo M, González-Manteiga W, Martínez-Miranda M (2020) Testing for significant differences between two spatial patterns using covariates. Spatial Stat 40:100379

Borrajo M, González-Manteiga W, Martínez-Miranda M (2017) Testing first-order intensity model in non-homogeneous Poisson point processes with covariates. arXiv preprint arXiv:1709.07716

Butsic V, Kelly M, Moritz MA (2015) Land use and wildfire: a review of local interactions and teleconnections. Land 4(1):140–156

Daley DJ, Vere-Jones D (2007) An introduction to the theory of point processes. Volume II: General theory and structure, 2nd edn. Springer, New York

D'Angelo N, Adelfio G (2024a) Cubature scheme for spatio-temporal Poisson point processes estimation. Preprint

D'Angelo N, Adelfio G (2024b) stopp: An R package for spatio-temporal point pattern analysis. J Stat Softw. To Appear. arXiv preprint arXiv:2408.15052

D'Angelo N, Adelfio G (2024c) stopp: Spatio-temporal point pattern methods, model fitting, diagnostics, simulation, local tests. R package version 0.2.4
D'Angelo N, Siino M, D'Alessandro A, Adelfio G (2022) Local spatial log-gaussian cox processes for seismic data. AStA Adv Stat Anal 106(4):633–671
D'Angelo N, Adelfio G, Mateu J (2023) Locally weighted minimum contrast estimation for spatio-temporal log-gaussian cox processes. Comput Stat Data Anal 180:107679
Díaz-Avalos C, Juan P, Mateu J (2013) Similarity measures of conditional intensity functions to test separability in multidimensional point processes. Stoch Environ Res Risk Assess 27:1193–1205
Díaz-Avalos C, Juan P, Mateu J (2014) Significance tests for covariate-dependent trends in inhomogeneous spatio-temporal point processes. Stoch Environ Res Risk Assess 28:593–609
Diggle PJ (2013) Statistical analysis of spatial and spatio-temporal point patterns. Chapman and Hall/CRC, London
Ferrara C, Salvati L, Corona P, Romano R, Marchi M (2019) The background context matters: local-scale socioeconomic conditions and the spatial distribution of wildfires in Italy. Sci Total Environ 654:43–52
Fuentes-Santos I, González-Manteiga W, Mateu J (2018) A first-order, ratio-based nonparametric separability test for spatiotemporal point processes. Environmetrics 29(1):e2482
Gabriel E, Diggle PJ (2009) Second-order analysis of inhomogeneous spatio-temporal point process data. Stat Neerl 63(1):43–51
Gabriel E, Opitz T, Bonneu F (2017) Detecting and modeling multi-scale space-time structures: the case of wildfire occurrences. J de la Société Française de Statistique 158(3):86–105
Ganteaume A, Camia A, Jappiot M, San-Miguel-Ayanz J, Long-Fournel M, Lampin C (2013) A review of the main driving factors of forest fire ignition over Europe. Environ Manag 51:651–662
GDAL/OGR contributors (2023) GDAL/OGR geospatial data abstraction software library. Open Source Geospatial Foundation
Genton MG, Butry DT, Gumpertz ML, Prestemon JP (2006) Spatio-temporal analysis of wildfire ignitions in the St Johns river water management district, Florida. Int J Wildland Fire 15(1):87–97
Ghorbani M, Vafaei N, Dvořák J, Myllymäki M (2021) Testing the first-order separability hypothesis for spatio-temporal point patterns. Comput Stat Data Anal 161:107245
González JA, Moraga P (2022) Kernstadapt: spatio-temporal adaptive kernel estimators for intensities. R package version 0.0.2
Hantson S, Pueyo S, Chuvieco E (2015) Global fire size distribution is driven by human impact and climate. Glob Ecol Biogeogr 24(1):77–86
Horn BK (1981) Hill shading and the reflectance map. Proc IEEE 69(1):14–47
Koh J, Pimont F, Dupuy J-L, Opitz T (2023) Spatiotemporal wildfire modeling through point processes with moderate and extreme marks. Ann Appl Stat 17(1):560–582
Loader CR et al (1999) Bandwidth selection: classical or plug-in? Ann Stat 27(2):415–438
Matheron G (1963) Principles of geostatistics. Econ Geol 58(8):1246–1266
Myllymäki M, Kuronen M, Mrkvička T (2021) Testing global and local dependence of point patterns on covariates in parametric models. Spatial Stat 42:100436
Nadaraya EA (1964) On estimating regression. Theory Probab Appl 9(1):141–142
Nadaraya EA (1989) Nonparametric estimation of probability densities and regression curves. Springer, Berlin
Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point processes. J Am Stat Assoc 83(401):9–27
Opitz T, Bonneu F, Gabriel E (2020) Point-process based Bayesian modeling of space-time structures of forest fire occurrences in Mediterranean France. Spatial Stat 40:100429
Pebesma E (2018) Simple features for R: standardized support for spatial vector data. R Journal 10(1):439–446
Pebesma E, Bivand R (2023) Spatial data science: with applications in R. Chapman and Hall/CRC, London
Podur J, Martell DL, Csillag F (2003) Spatial patterns of lightning-caused forest fires in Ontario, 1976–1998. Ecol Model 164(1):1–20
R Core Team (2024) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
Raeisi M, Bonneu F, Gabriel E (2021) A spatio-temporal multi-scale model for Geyer saturation point process: application to forest fire occurrences. Spatial Stat 41:100492

Ricotta C, Di Vito S (2014) Modeling the landscape drivers of fire recurrence in Sardinia (Italy). Environ Manag 53(6):1077–1084
Ricotta C, Bajocco S, Guglietta D, Conedera M (2018) Assessing the influence of roads on fire ignition: does land cover matter? Fire 1(2):24
Rodrigues M, De la Riva J (2014) An insight into machine-learning algorithms to model human-caused wildfire occurrence. Environ Model Softw 57:192–201
Schoenberg FP (2004) Testing separability in spatial-temporal marked point processes. Biometrics 471–481
Schroeder W, Oliva P, Giglio L, Csiszar IA (2014) The new VIIRS 375 m active fire detection data product: algorithm description and initial assessment. Remote Sens Environ 143:85–96
Shepard D (1968) A two-dimensional interpolation function for irregularly-spaced data. In: Proceedings of the 1968 23rd ACM national conference, pp 517–524
Siino M, Adelfio G, Mateu J (2018) Joint second-order parameter estimation for spatio-temporal log-Gaussian cox processes. Stoch Environ Res Risk Assess 32(12):3525–3539
Silverman BW (2018) Density estimation for statistics and data analysis. Routledge, London
Tarquini S, Isola I, Favalli M, Battistini A, Dotta G (2023) TINITALY, a digital elevation model of Italy with a 10 meters cell size, 1.1 edition. Istituto Nazionale di Geofisica e Vulcanologia (INGV). https://doi.org/10.13127/tinitaly/1.1
Tonini M, Pereira MG, Parente J, Vega Orozco C (2017) Evolution of forest fires in Portugal: from spatio-temporal point events to smoothed density maps. Nat Hazards 85:1489–1510
Watson GS (1964) Smooth regression analysis. Sankhyā Indian J Stat Ser A 359–372
Weih RC Jr, Mattson TL (2004) Modeling slope in a geographic information system. J Arkansas Acad Sci 58(1):100–108
Zhuang J (2015) Weighted likelihood estimators for point processes. Spatial Stat 14:166–178. Spatio-Temporal Stochastic Modelling of Environmental Hazards

Authors and Affiliations

Nicoletta D'Angelo[1] · Alessandro Albano[1,2] · Andrea Gilardi[3] · Giada Adelfio[1]

✉ Nicoletta D'Angelo
nicoletta.dangelo@unipa.it

Alessandro Albano
alessandro.albano@unipa.it

Andrea Gilardi
andrea.gilardi@polimi.it

Giada Adelfio
giada.adelfio@unipa.it

[1] Department of Economics, Business and Statistics, University of Palermo, Palermo, Italy

[2] Sustainable Mobility Center (Centro Nazionale per la Mobilità Sostenibile—CNMS), Milano, Italy

[3] MOX - Department of Mathematics, Politecnico di Milano, Milano, Italy