



# Using spatial point process models, clustering and space partitioning to reconfigure fire stations layout

Regina Bispo<sup>1</sup> · Francisca G. Vieira<sup>2</sup> · Clara Yokochi<sup>2</sup> · Filipe J. Marques<sup>1</sup> · Pedro Espadinha-Cruz<sup>3</sup> · Alexandre Penha<sup>4</sup> · António Grilo<sup>3</sup>

Received: 14 July 2023 / Accepted: 6 September 2023  
© The Author(s) 2023

## Abstract

Fire stations (FS) are typically non-uniformly distributed across space, and their service area is, in general, defined based on administrative boundaries. Since the location of FS may considerably influence the readiness and the effectiveness of the provided services, national and regional governments need research-based information to adequately plan where to establish firefighting facilities. In this study, we propose a method to reconfigure the fire stations layout using spatial point process models, clustering and space partitioning. First, modelling fire intensity variation across space through a point process model enables to replicate the process independently by simulation. Subsequently, for each simulation, the  $k$ -means algorithm is used to define a siting location, minimizing the total within distance between the fire occurrences and the new position. This method allows to obtain a set of locations from which the respective distribution is inferred. Assuming a bivariate normal spatial distribution, we further define confidence siting regions. Ultimately, new FS service areas are defined by Voronoi tessellation. To exemplify the application of the method, we apply it to reconfigure the fire station layout at Aveiro, Portugal.

**Keywords** Fire · Fire stations · Poisson point process ·  $k$ -means · Voronoi tessellation

## 1 Introduction

Fires are still today a worldwide leading cause of property loss, psychological effects, physical damage and death [1]. According to the data supplied by the Portuguese National Emergency and Civil Protection Authority (ANEPC, Autoridade Nacional de Emergência e Proteção Civil), the average number of urban fires in Portugal between 2012 and 2020 surpasses 8000 events/year [2]. In parallel, the incidence of forest fires in Portugal is still a matter of great concern.

Together, rural and forest fires have substantial societal and environmental impacts.

Fire stations (FS) location and resource allocation are key determinants of the efficiency and effectiveness of fire protection services [3, 4]. In addition, the efficacy of proper facility planning depends on choice of possible locations for fire stations [5]. Due to this importance, many studies have been conducted to analyse the location of FS. Generically, the expression *Location Analysis* encompasses modelling and finding siting solutions for facilities in some given space [6]. According to ReVelle and Eiselt [6], there are four components that characterize location problems, including *customers* with known location, *facilities* that need to be sited, a *space* occupied by both customers and facilities, and a *metric* that measures the cost allocation. The optimal location for facilities typically involves minimizing the given cost function, e.g., minimizing the time [7] or the distance [8–10]. There is a considerable number of literature reviews for the particular case of FS location problem [11, 12, e.g.]. Baçar et al. [12] present a taxonomic framework for the emergency service station location problem using an operations research perspective. These authors highlight the use of meta-heuristic approaches such as tabu search, genetic

✉ Regina Bispo  
r.bispo@fct.unl.pt

<sup>1</sup> Department of Mathematics, NOVAMATH Center for Mathematics and Applications, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Lisbon, Portugal

<sup>2</sup> NOVA School of Science and Technology, Universidade NOVA de Lisboa, Lisbon, Portugal

<sup>3</sup> Department of Mechanical and Industrial Engineering, UNIDEMI, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Lisbon, Portugal

<sup>4</sup> Comando Nacional de Emergência e Proteção Civil, ANEPC Autoridade Nacional de Emergência e Proteção Civil, Carnaxide, Portugal

algorithms, simulated annealing and ant colony optimization strategies. Aleisa [11] reviewed the approaches which include methods-based on operations research (e.g., maximal coverage location problem and maximal service area problem), fuzzy multi-objective optimization models [8, 10], the genetic algorithm [13], tabu search and simulating annealing [13], the ant algorithm and geospatial analytical methods based on Geographic Information Systems (GIS) [7].

In 2009, Bonneu and Thomas-Agnan [14] recognize that location-allocation traditional approaches ignored the potential inhomogeneous spatial distribution of customers and/or clients and stress that the few stochastic location decision methods that address the random nature of the process use, however, very naïve tools. As a consequence, these authors proposed to model *customers* as a realization of a spatial point process. By coupling simulations from a fitted point process model with standard optimization tools, they characterized the optimal solution using contour plots. However, according to the authors, two out of the three used optimization strategies presented the drawback of searching only within the subset of allocations restricted to the nearest facility and the used heuristic was limited for locating a single new position (although allegedly could be expanded to several).

More recently, Dey et al. [5] addressed fire stations location planning using machine learning models, based on Random Forest and Extreme Gradient Boosting, for demand prediction and utilize the models further to define a generalized index to measure quality of fire service in urban settings. These authors develop an optimization problem to select the best locations to install new fire stations and further develop a two-stage stochastic optimization model to characterize the confidence in the decision outcome.

Spatial methods have also been used to study fire clustering in space [15]. For that purpose, Ceyhan et al. [15] analysed the spatial distribution of fires in global and local scales using, respectively, first-order (intensity analysis using kernel density estimation) and second-order properties. In addition, clustering methods are commonly applied to deal with geographical referenced data and frequently used to address FS location problem [16–19, e.g.]. In particular, the *k*-means algorithm is a popular technique used in discovering groups that are similar to each other in respect of spatial attributes at any given instance of time [17, 19, 20].

In Portugal, the latest financial audit information on Portuguese FS [21] reports that the number of fire stations per municipality varies widely across the country which may lead to an unequal quality of the services provided. In addition, this report also underlines that the unbalanced distribution of resources may potentially compromising the territorial cohesion. These concerns alert for the need to restructure the organizational model of FS by rethinking the location and service areas of FS, including the possible extinction or grouping of firefighting facilities. Given this concern, this study aimed

at providing supported research-based information on local fire stations layout (FS locations and service areas) that may help the authorities in the process of decision-making.

We propose a method for public-service facilities siting and service area delimitation, based on spatial point process modelling, clustering and space partitioning. Conceptualizing observations as realizations of an underlying unknown spatial point process granted the framework to model and randomly simulate the process. Modelling the process allowed to infer the true underlying process and simulate independent patterns. This allowed to, instead of giving just one unique location as optimal, define a spatial distribution for locations and, consequently, define an optimal region for siting FS. Hence, modelling and simulating the process allowed to estimate the uncertainty associated with the optimal locations and, thus, define confidence siting regions. Finally, the service areas of the facilities were defined by Voronoi tessellation.

As an application example, the method was used to reconfigure FS layout at Aveiro, Portugal. This study partially followed Bonneu and Thomas-Agnan [14] approach. We have used clustering to solve the drawbacks of searching only within the subset of allocations restricted to the nearest facility and the limitation of locating only a single new position, mentioned by the authors. In addition, this study fully addresses the problem of charactering the optimal location distribution using nonparametric and parametric approaches and defines a method for service area delimitation.

In Sect. 2, we describe the proposed method. The application example is presented in Sect. 3, including the data and the obtained results. In Sect. 4, we highlight the main conclusions of this study.

## 2 Methods

Our method includes the following steps:

1. *Exploring*: perform an exploratory analysis on spatial fire data to define a modelling strategy (Sect. 2.1);
2. *Modelling*: infer the underlying point process by modelling fire intensity variation across space (Sect. 2.2);
3. *Characterizing spatial optimal location distribution*: generate random realizations of the fitted model, perform a spatial clustering on each simulated point pattern and define optimal regions for FS location (Sect. 2.3);
4. *Allocating*: define service areas by tessellation (Sect. 2.4).

In the following subsections, we describe in detail these steps. All statistical analyses were carried out in R environment [22]. The package SPSVERBc1 [23] was used to read shapefiles and deal with geometry indexed objects. Spatial point patterns were modelled and simulated using the

package SPSVERBc2 [24]. Voronoi diagrams were defined with the aid of the package SPSVERBc3 [25]. To avoid common difficulties in implementing and/or reproducing of the proposed method, we provide the R code and the used data on GitHub repository <https://github.com/rb1970/FireStationsLayout>, making it ready to apply the method in other regions.

## 2.1 Exploratory data analysis

A point pattern is a collection of points of the form  $\{(\mathbf{u}_i, t)\}_{i=1}^{n_t}$  (with  $\mathbf{u}_i \equiv (x_i, y_i)$  being the spatial coordinates, longitude  $x_i$  and latitude  $y_i$ , for the  $i$ th event at time  $t \in T$  and  $n_t$  the number of observed events at time  $t$ ) observed in  $W \times T$  such that  $W \subseteq \mathbb{R}^2$  and  $T \subseteq \mathbb{R}$ . Typically, spatial processes, which produce point patterns, are firstly described by its intensity (i.e., the expected number of points per unit area,  $\lambda$  ( $\lambda > 0$ )). If the intensity of the point process is constant for all locations inside the region  $W$ , which is known as the *homogeneous intensity assumption*, the number of points falling in  $W$  can be assumed to follow a Poisson distribution with mean simply given by  $\lambda |W|$  (with  $|W|$  being the area of  $W$ ). In addition, the *interpoint independence and uniformity assumption* implies there are no interactions amongst the events and that the points distribute uniformly across the space. These assumptions define an *homogeneous Poisson process* that provides a benchmark of complete spatial randomness against which various kinds of patterns can be compared [27].

If the intensity of the point process depends on the spatial location, the number of events in the bounded region  $W$  follows a Poisson distribution with mean  $\lambda(\mathbf{u})$ , which is a function of the location  $\mathbf{u}$ . Under these circumstances, the expected number of points falling in  $W$  is given by  $I = \int_W \lambda(\mathbf{v}) d\mathbf{v}$  and the independent points are no longer uniformly distributed, with common probability density function (pdf) given by

$$f(\mathbf{u}) = \frac{\lambda(\mathbf{u})}{I}. \quad (1)$$

Under these circumstances, the Poisson process is said to be *inhomogeneous*.

Homogeneity and interpoint independence are key assumptions to validate in the context of a point pattern analysis. Inhomogeneity is important to detect and model as it reflects the intensity dependence on spatially varying external factors. In this case, the analysis aims at inferring this spatial dependence from the data [26]. The presence of interpoint dependence means that the existence of an event either encourages or inhibits the occurrence of other events in the neighbourhood. Thus, if the spatial pattern suggest associations between the points at short distances, the models must be corrected to explicitly incorporate and capture interpoint

interactions. In these cases, cluster or Cox processes can be applied for positive correlations and Gibbs models can be used for negative dependences between points [26].

The homogeneity assumption is typically studied characterizing the spatial density (which is proportional to the intensity). This is considered a *first-order property* as it allows to describe the global distribution of the points. Spatial density is often assessed nonparametrically by kernel estimation. Kernel estimators depend on a *kernel function* and a *bandwidth parameter* (which determines the amount of smoothing). Generically, the kernel estimator of a bivariate pdf,  $f(x)$ , based on a sample with  $n$  points, takes the form

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n k_h(x - x_i) \quad (2)$$

with  $h$  being the bandwidth, i.e., the distance from the point  $x$  and  $k_h$  the kernel function (satisfying symmetry, non-negativity for all  $x \geq 0$  and  $\int k(v)dv = 1$ ). A common choice to define  $k_h$  is the Gaussian probability density. This function together with the standard deviation of the kernel were, respectively, taken as the kernel function and the smoothing bandwidth.

Interpoint independence is considered a *second-order property* as it characterizes the local spatial distribution. The most common method for a second-order analysis is the  $K(r)$  function [28] originally proposed by Ripley [29]. This function is the cumulative average number of data points lying within a distance  $r$  ( $r \geq 0$ ) of a data point. Under a homogeneous Poisson process, which assumes interpoint independence,  $K_{\text{pois}}(r) = \pi r^2$ , which serves as benchmark for *no correlation* [27]. Values of  $K(r)$  above  $\pi r^2$  reveal a clustered pattern (more neighbours than would be expected under interpoint independence) and below  $\pi r^2$  indicate a regular pattern (fewer neighbours than would be expected under interpoint independence). The variant of the  $K$ -function,  $L(r) = \sqrt{K(r)/\pi}$ , which basically represents the same information, has graphical and statistical analytical advantages being, as such, preferred in practice. When the events are clustered  $L(r) > r$  and when the events are regular  $L(r) < r$ . The events are clustered if  $L(r) > r$ , and regular, if  $L(r) < r$ .

While defining  $K$ - and  $L$ -functions, it is important to consider the edge-effects. These effects arise frequently in spatial point pattern analysis because, in most cases, the study area (on which the pattern is observed) is part of a larger region on which the underlying process operates. This phenomenon is important to consider as the unobserved events outside the study area may interact with the events within it but, as the former are not observed, they are not accounted for [27]. To prevent these effects, it is usual to correct the  $K$ - and  $L$ -functions. The *border correction* restricts the functions to cases where the circle of radius distance  $r$  lies entirely inside

the window of observation so that the bias due to the edge effect does not occur. This prevents the density sharp decline close to the boundary of the observation window [26].

Lastly, because applying these measures when the process is inhomogeneous can overestimate the dependence between events, a generalized version of both functions,  $K_{\text{inhom}}(r)$  and  $L_{\text{inhom}}(r)$ , is usually computed to avoid this bias [30].  $K_{\text{inhom}}(r)$  function may be defined as the expected total weight of all random points within a distance  $r$  of the point  $\mathbf{u}$ , where the weight of a point  $\mathbf{u}$  is  $1/\lambda(\mathbf{u})$  [26]. For an inhomogeneous Poisson process with intensity  $\lambda(\mathbf{u})$ , that also assumes interpoint independence, the inhomogeneous  $K$  function is also given by  $\pi r^2$ , which again serves as a benchmark.  $L_{\text{inhom}}(r)$  is accordingly defined as  $L_{\text{inhom}}(r) = \sqrt{K_{\text{inhom}}(r)}/\pi$ .

## 2.2 Point process model

Modelling a point process implies searching for a predictor of the intensity of the process inside the observed region. Log-linear Poisson point process models are typically used to model the process intensity as these have a specially convenient structure in which the log intensity is a linear function of the parameters, encompassing a very wide class of models [26]. The most simple model for a Poisson point process is the constant intensity model, i.e., the homogeneous model, equivalent to complete spatial randomness, which can be simply represented by  $\log \lambda(\mathbf{u}) = \beta_0$ .

Point process models may include an *offset*. An offset is a term in the linear predictor which does not involve any parameters of the model, being used to include an effect which is already known to occur. An inhomogeneous Poisson model with intensity proportional to a particular covariate ( $Z$ ) can be represented by  $\log \lambda(\mathbf{u}) = \beta_0 + \log(Z(\mathbf{u}))$  and is called a *baseline model* [26].

When available, other variables may be added to the linear predictor generically including terms which are space dependent. When there are no covariates available, it is common to include the Cartesian coordinates  $x$  and  $y$  as spatial covariates. These are particularly useful for investigating spatial trends as they inherently encapsulate/capture the spatial effects of unavailable covariates.

A wide class of models can be built up by combining the Cartesian coordinates in convenient ways, namely through the use of polynomials. Interactions between the Cartesian coordinates, up to the polynomial degree, may also be considered to reflect that the effect of  $x$  over the intensity of the process may depend on  $y$  value and *vice versa*.

In this study, the intensity of the point process was modelled using complete log-polynomial models on the events coordinates up to the third-degree. These models may be

represented by equations

$$\log \lambda(\mathbf{u}) = \beta_0 + \log(Z(\mathbf{u})) \quad (3)$$

$$\log \lambda(\mathbf{u}) = \beta_0 + \log(Z(\mathbf{u})) + \beta_1 x + \beta_2 y + \beta_3 xy \quad (4)$$

$$\log \lambda(\mathbf{u}) = \beta_0 + \log(Z(\mathbf{u})) + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 xy + \beta_5 y^2 \quad (5)$$

$$\log \lambda(\mathbf{u}) = \beta_0 + \log(Z(\mathbf{u})) + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 xy + \beta_5 y^2 + \beta_6 x^3 + \beta_7 x^2 y + \beta_8 x y^2 + \beta_9 y^3 \quad (6)$$

where  $\beta_i$  ( $i = 1, \dots, 9$ ) denotes the unknown model parameters,  $Z$  the offset covariate and  $\mathbf{u} = (x, y)$  the vector of the spatial coordinates, longitude and latitude, respectively.

Model (3) served as the baseline model for comparison purposes. Population density, which is known to affected fire occurrence [2], was included as an offset term (defined by a real-valued pixel image) to effectively adjust the prediction for population inhomogeneity across space. The subsequent models [Eqs. (4)–(6)] were built progressively adding Cartesian coordinate based terms, including interactions, up to a third-degree polynomial.

The goodness-of-fit for the different models was assessed by employing the Akaike Information Criterion (AIC) [31, 32] and the model with the lowest value of AIC was chosen. In particular, AIC differences (for the  $j$ th model  $\Delta \text{AIC}_j = \text{AIC}_j - \text{AIC}_{\min}$ , where  $\text{AIC}_{\min}$  is the lowest AIC value for the fitted models) were computed. As the best fit was taken according to the lowest AIC value, AIC differences express the loss of information if the fitted model was used instead of the best adjusted model. Backward stepwise using AIC was used to perform the model terms selection.

## 2.3 Optimal location distribution

To characterize the optimal location distribution, we have simulated random realizations of the fitted model, performed a spatial clustering on each simulated point pattern and defined confidence siting regions for FS location.

**Simulation** Based on the best model, we have generated independent realisations of the point process model. The generated point patterns were then analysed by spatial clustering and the respective cluster centroids taken as independent realizations of the random optimal location.

**Clustering** Clustering methods when applied to spatially indexed observations allow to identify geographically bounded groups of occurrences. The simulated fire events were clustered using  $k$ -means algorithm. Formally, given the two-dimensional objects composed by the spatial coordinates, this algorithm grouped the data into  $k$  spatial clusters

$C_j$  ( $j = 1, \dots, k$ ) such that  $\{C_j\}_{j=1}^k$  formed a spatial partition of the  $n$  occurrences (with  $C_i \cap C_j = \emptyset$ ,  $\forall i \neq j$ ). Each cluster  $C_j$  was represented by the respective centroid  $\mu_j$  which minimizes the total within-cluster distances between each event and its centroid, *i.e.*,

$$\min_{\mu_1, \dots, \mu_k} \sum_{j=1}^k \sum_{i: x_i \in C_j} \|x_i - \mu_j\|. \quad (7)$$

The number of spatial clusters ( $k$ ) was taken by the number of existent FS, to allow a direct comparison between the actual and the new FS layouts.

**Confidence siting regions** Spatial clustering was performed on each simulated realization of the fitted model to find simulated optimal locations for fire stations. This action resulted in  $s \times t \times k$  realizations of the optimal FS locations (with  $s$ ,  $t$  and  $k$  being the number of simulations, the number of replicated observed point patterns and the number of clusters, respectively). From this set of points, we characterized the optimal location distribution and defined confidence siting regions, using both parametric and nonparametric approaches. The parametric approach was based on a bivariate Gaussian distribution assumption with location and scale parameters, respectively, estimated by the coordinates mean vector and the covariance matrix. Under the nonparametric approach, we resorted to kernel density estimation with top high-density regions centres given by local density maxima. These points were found through focal analysis, *i.e.*, the maximum density was determined using a moving window that swept the entire density surface. The moving window, defined by a matrix, specifies a local neighbourhood. Then, the maximum density within each neighbourhood is found and the correspondent Cartesian coordinates determined.

## 2.4 Space partitioning

In the last step of our method, the study region was divided into non-overlapping areas by Voronoi tessellation, which is a widely used method to deal with spatial partitioning problems, especially in delimiting the service area of facilities [33]. Voronoi tessellation divides the plane in adjacent cells/regions. A region  $R_j$  is defined by the set of points that are closer to the region centre  $\mathbf{c}_j$  than to any other centre  $\mathbf{c}_k$  ( $j \neq k$ ), *i.e.*, a point  $\mathbf{u}$  is included in region  $R_j$  instead of  $R_k$  if  $\|\mathbf{u} - \mathbf{c}_j\| < \|\mathbf{u} - \mathbf{c}_k\|$ ,  $\forall j \neq k$  [34]. As a consequence, the regions borders lie exactly on the perpendicular bisectors of the line segment linking  $\mathbf{c}_j$  to  $\mathbf{c}_k$ .

True events were allocated to the defined tiles and the average distances between true incidents and both the true and the new FS locations were computed. To proxy true road distances, we have calculated the Euclidean distances. This was done, to reduce the computational burden, after performing

an exploratory analysis that showed a strong linear correlation (see Fig. 6, appendix 1) between Euclidean and by road distances (Google map based). Furthermore, this procedure is in accordance with the method adopted in previous studies in which the Euclidean distance (or the squared Euclidean distance) was taken to proxy the cost of reallocating fire stations (e.g., [14]).

## 3 Application

In this section, we describe the available data (Sect. 3.1) and the results obtained from applying the method detailed in Sect. 2 to the Portuguese district of Aveiro. Section 3.2 summarizes first- and second-order analysis of fire events in the study region. While in Sect. 3.3 we illustrate the use of the proposed method in reconfiguring the FS layout, in Sect. 3.4 we exemplify the definition of a confidence region for a particular fire station.

### 3.1 Data description

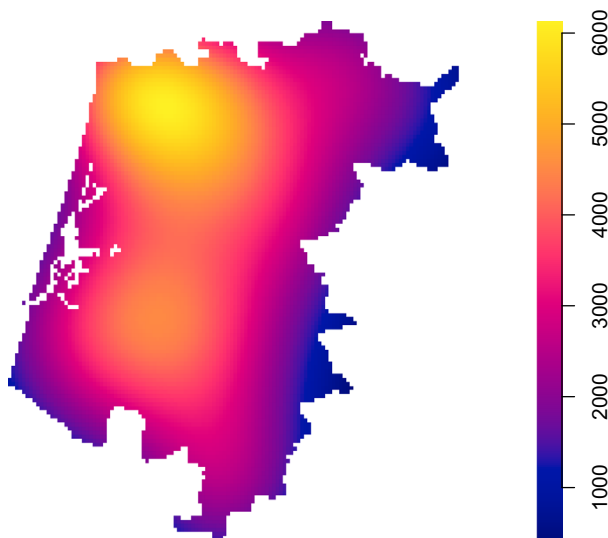
A dataset containing a total of 233 204 fire occurrences (urban and rural) in mainland Portugal, between 2012 and 2020, was supplied by ANEPC. Table 1 summarizes the observed fire intensity at a municipality level. The asymmetry across the country is clear with average annual intensities ranging from 2.11 to 0.05 fires/km<sup>2</sup>.

**Table 1** Number of fire stations ( $n$ ) and observed fire intensity for each district

District	$n$	Fires/year	Fires/year/km <sup>2</sup>
Porto	47	4920.89	2.11
Lisboa	57	3040.78	1.10
Braga	21	2243.11	0.83
Aveiro	27	2138.44	0.78
Viseu	34	1704.56	0.34
Setúbal	25	1449.67	0.29
Vila Real	28	1341.00	0.31
Viana Do Castelo	12	1307.11	0.59
Santarém	29	1220.11	0.18
Leiria	25	1218.67	0.35
Coimbra	24	1032.67	0.26
Guarda	23	842.44	0.15
Faro	17	827.44	0.17
Bragança	15	791.56	0.12
Castelo Branco	12	626.89	0.09
Beja	15	486.33	0.05
Évora	14	383.56	0.05
Portalegre	15	336.33	0.06

Each record in the dataset included the date, time and incident location (latitude and longitude). The available data also included the main responsible FS for each event and the respective location of the facilities. In addition, the population size by parish was retrieved from the last available Portuguese Census (2021) and downloaded from Statistics Portugal website ([www.ine.pt](http://www.ine.pt)). Data regarding the geographical delimitation of the administrative units of Portugal were obtained from shapefiles downloaded from the open data public certified service <https://dados.gov.pt/> that contains the geometry of the administrative divisions of the country, namely the official boundaries of districts, municipalities, and parishes. These geometry data were restricted to mainland Portugal to allow the rigorous mapping of the events. The shapefiles are made available online by the Agência para a Modernização Administrativa, I.P. (AMA), a Portuguese public institute entrusted with the task of fostering and advancing administrative modernization in Portugal, which operates under the supervision of the Secretary of State for Digitalization and Administrative Modernization.

To exemplify the application of the method, we have applied it to reconfigure the fire stations layout at Aveiro, Portugal. However, the provided dataset encompasses information for all the 18 districts of mainland Portugal which, together with the provided R code, allows it to readily implement the method in any other district.



**Fig. 1** First-order analysis: estimated annual fire intensity at Aveiro, Portugal

### 3.2 Exploratory data analysis

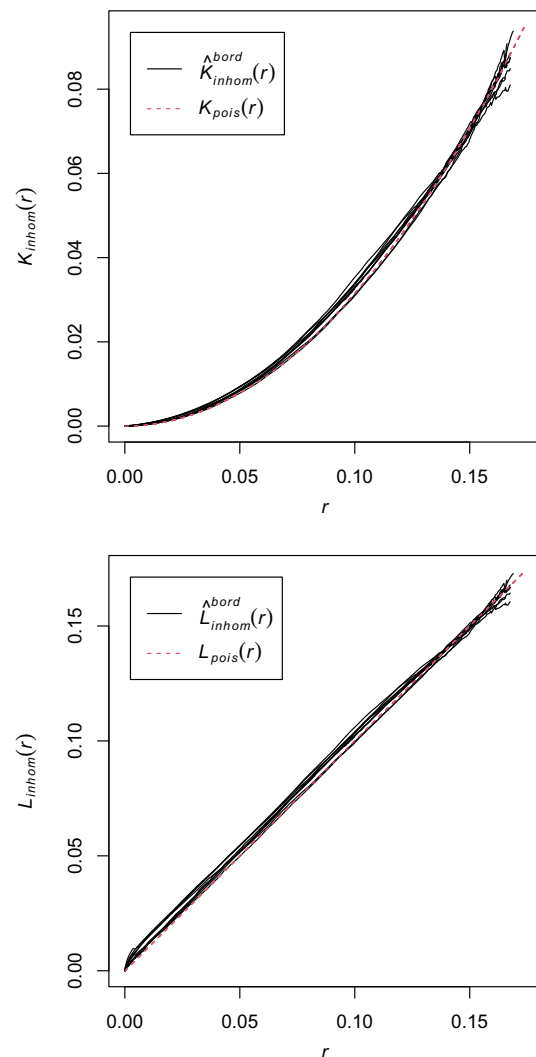
This section summarizes the obtained results regarding homogeneity and interpoint independence assumptions which guided the modelling strategy.

#### Homogeneity assumption

Fig. 1 shows the estimated annual fire intensity at Aveiro. This figure shows a clear increase in the intensity from east to west and from south to the north of the region, i.e., it shows that the homogeneous intensity assumption does not hold, pointing towards an inhomogeneous Poisson point process.

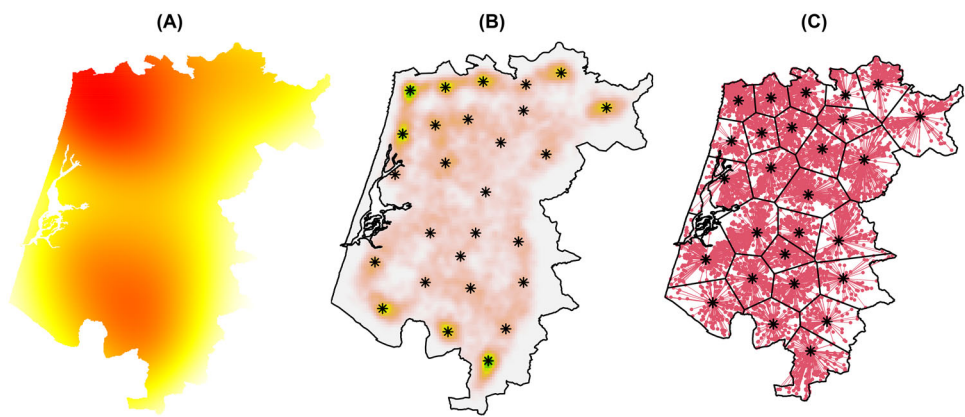
#### Interpoint independence assumption

Figure 2 depicts the border corrected inhomogeneous functions  $\hat{K}_{\text{inhom}}^{\text{bord}}(r)$  and  $\hat{L}_{\text{inhom}}^{\text{bord}}(r)$  of the observed point



**Fig. 2** Second-order analysis: edge-corrected (border method) estimates of the inhomogeneous  $\hat{K}_{\text{inhom}}^{\text{bord}}(r)$  (top plot) and  $\hat{L}_{\text{inhom}}^{\text{bord}}(r)$  (bottom plot) functions of the point pattern observed at Aveiro, Portugal.  $K_{\text{pois}}(r)$  and  $L_{\text{pois}}(r)$  represent the theoretical functions for a Poisson process

**Fig. 3** **A** Fire intensity model, **B** density of optimal locations and respective local maxima (black stars), **C** events allocation to new FS locations (black stars) based on the service area delimitation defined by the Voronoi tessellation



pattern at the study region. These curves overlap the theoretical functions for an inhomogeneous Poisson process, indicating that there is no correlation between the points, i.e., suggesting the absence of clustering (positive dependence between points) or regularity (negative dependence between points).

In summary, the exploratory data analysis supported the existence of an inhomogeneous process with no interaction between the events.

### 3.3 Fire stations layout reconfiguration

Figure 3 illustrates graphically the method. Figure 3A shows the fire intensity model from which new point patterns were simulated. This model was selected from the set of models summarized in Table 2. The results point to Model (6) as the best fit. Backward stepwise terms selection kept all the terms defined in Eq. (6).

Based on the fitted model, we have generated 50 independent simulated realisations of the point process, resulting in a total of 50 (simulations)  $\times$  9 (years) = 450 generated point patterns. Each set of simulated ignition points was divided into  $k = 27$  clusters (recall that the number of spatial clusters was taken by the number of existent FS). The centroids of these clusters were then used to define optimal FS locations. Figure 3B shows the optimal locations density reflected by the spatial variation of the clusters centroids after running all the simulations. This density was characterized nonparametrically by kernel estimation and the respective local maxima (black stars) were taken from the top high-density areas. These points served as centres of the Voronoi tiles (Fig. 3C) which jointly established a new spatial reconfiguration of the FS, both in terms of FS locations and service areas, as if the system could be reconstructed from scratch.

The true events were then allocated to the nearest new location according to the service area delimitation defined by the Voronoi tiles (Fig. 3C) and the average distance between events and the new locations was computed. The results

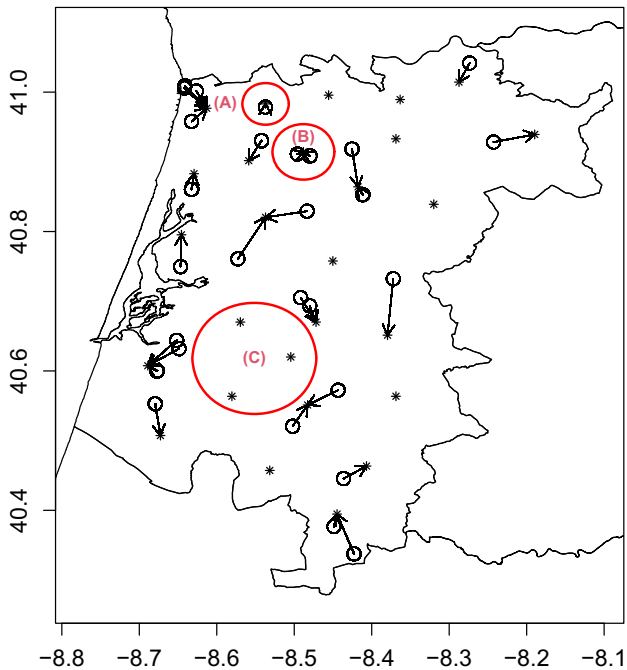
showed that the average distance between the ignition points and the FS locations decreased from 4.28 to 3.85 km.

Figure 4 shows the actual and the new FS positions at Aveiro region. In this figure, the arrows point from the actual FS (represented by open circles) to the closest new location (represented by black stars). Our results suggest two possible strategies to reconfigure FS in space, either by merging

**Table 2** Models summary

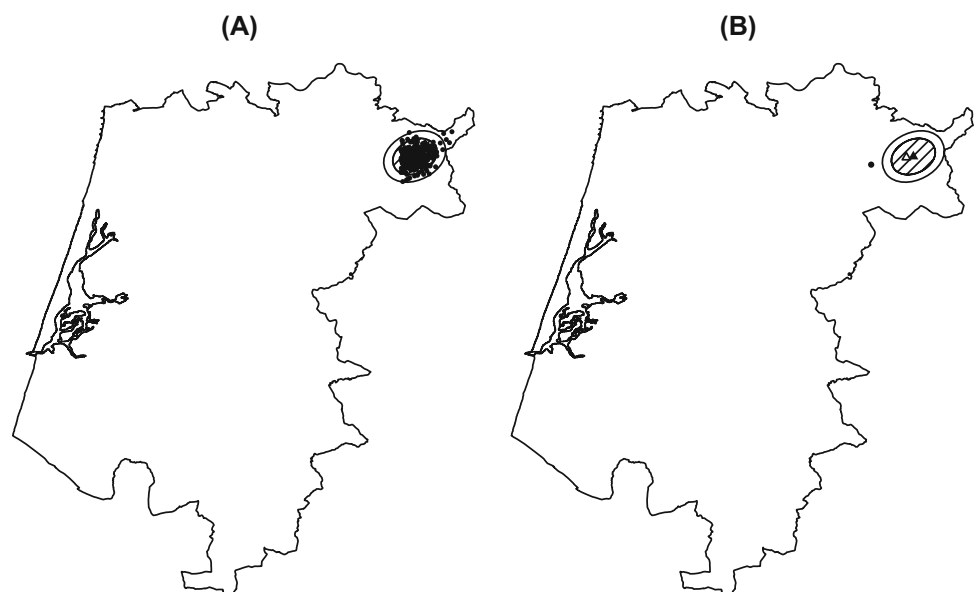
	Estimate	Std. error	AIC	$\Delta$ AIC
Model (3):			-121842.8	1692.3
$\beta_0$	-0.293	0.019		
Model (4):			-122315.7	1219.5
$\beta_0$	-1864.164	292.619		
$\beta_1$	-217.181	34.527		
$\beta_2$	46.066	7.177		
$\beta_3$	5.368	0.847		
Model (5):			-123040.6	494.6
$\beta_0$	11154.061	985.930		
$\beta_1$	-386.651	51.804		
$\beta_2$	-628.787	44.677		
$\beta_3$	-5.177	1.119		
$\beta_4$	7.357	0.947		
$\beta_5$	8.494	0.533		
Model (6):			-123535.2	0
$\beta_0$	-2343319.525	225185.807		
$\beta_1$	-168017.950	15108.736		
$\beta_2$	137916.904	15165.968		
$\beta_3$	-3563.127	537.823		
$\beta_4$	6754.696	581.425		
$\beta_5$	-2691.513	352.960		
$\beta_6$	-9.710	7.671		
$\beta_7$	81.208	9.329		
$\beta_8$	-65.908	6.084		
$\beta_9$	17.530	2.815		

several FS into a single one (Fig. 4, section (B)) or by setting new FS locations in less well covered areas (Fig. 4, section (C)). Some new stations appear in a very narrow vicinity (or even at the same location) of existing ones (Fig. 4, section (A)), which suggests the need to define homologous stations and keep the existent locations.



**Fig. 4** Actual (black open circles) and new (black stars) fire station (FS) layout at Aveiro. Regions A, B and C suggest possible strategies to reconfigure fire station layout: A keeping existent, B merging, and C adding FS in poorly covered areas

**Fig. 5** Gaussian based 90% and 99% confidence ellipsoids for the optimal siting of a FS at Arouca, Aveiro: **A** The dashed area represents the 90% confidence region and the points correspond to the 450 optimal locations obtained by simulation and clustering; **B** the black dot represents the actual location of the local FS ( $-8.242602$  lon,  $40.928327$  lat). Open and filled triangles correspond, respectively, to non-parametric ( $-8.189591$  lon,  $40.939293$  lat) and parametric ( $-8.179116$  lon,  $40.940552$  lat) estimates for the FS optimal location positions



### 3.4 Optimal siting region

Figure 5 shows the 90% and 99% confidence ellipsoids for the optimal siting of a FS at Arouca (Aveiro), assuming bivariate normally distributed data. These ellipsoids define confidence regions within which it is expected to find the true optimal position given the chosen confidence.

Figure 5B depicts the relative positions of the local FS and the nonparametric and parametric estimated locations. In this example, these two approaches give approximately the same solution which is clearly distant from the actual FS location. As expected, the nonparametric site is less influenced by more extreme positions.

## 4 Conclusion

Throughout time, the number and locations of FS has grown organically, depending on circumstantial factors such as, e.g., population size, urban growth, territorial occupation and other time dependent changing variables. Never before was possible to think and plan FS layout actually as a consequence of emergency events, namely fires occurrences, accounting both for their number and location. Nowadays, due to the capacity to easily store and process big amounts of information is finally possible to rethink spatial FS layout based on a data-driven perspective. Our study started by analysing the replicated point pattern based on first- and second-order properties to decide on the modelling strategy. Given the space inhomogeneity and the influence of population density, the intensity of the process was modelled by a complete log-cubic predictor on the spatial coordinates, including population density as an offset. This model allowed to simulate

the process. Based on each simulation, the optimal locations were found by spatial clustering, minimizing the distance between fire ignition points and FS locations. As fires do not distribute uniformly across space but, instead, tend to form geographically well-defined clusters, these groups were identified and FS locations were taken by the respective centroids as if there were no prior FS and the system could be reset.

This method allowed to obtain not one but, instead, as many “optimal locations” as the number of ran simulations. From this set of coordinates, we have estimated nonparametrically the respective optimal locations density with top high-density points defined by local density maxima. In parallel, assuming a bivariate normal distribution, we defined a confidence region with centre given by the respective mean vector.

By linking the actual with the closest new FS locations, we found two possible strategies to reconfigure FS in space, either by merging FS or setting new ones in areas poorly covered. In addition, as relocating fire stations that are within a narrow vicinity to where they should be seems unreasonable and impractical, we suggest adopting the concept of homologous facilities and, consequently, not relocating these stations.

Finally, we defined service areas by space partitioning. This procedure allowed to ensure that each new site had the minimal possible distance of arrival from any emergency event that took place within the respective delimited area.

The *k*-means method, although very efficient at clustering data, tends to produce spherical clusters of similar size and give solutions strongly influenced by outliers. If clusters have different sizes and densities, which may happen in this case, this algorithm may not perform ideally. Other spatial clustering techniques and/or optimization procedures can be used. In addition, to improve the estimation of the distance and travel time between events and fire stations facilities, traffic congestion and/or alternative road directions could be

considered. Further, differences between available resources between FS should also be taken into account.

In summary, this study establishes a method to rethink FS layout at a regional scale. It was applied to a particular region to reconfigure the FS layout, both in terms of FS locations and service areas, as if the entire system was being reconstructed anew. Acknowledging the limitations of achieving a complete system reconstruction, the study’s purpose is to establish an essential groundwork for shaping forthcoming decisions concerning the strategic positioning of fire stations. Particularly noteworthy is its possibility to identify confidence siting locations transcending the constraints of a fixed, deterministic, location.

Recognizing the intricate interplay of political, social, and economic considerations in the process of fire station layout reconfiguration, we hope that this study may offer guidance to decision-makers in the future, considering the multifaceted implications inherent to the process of siting fire stations.

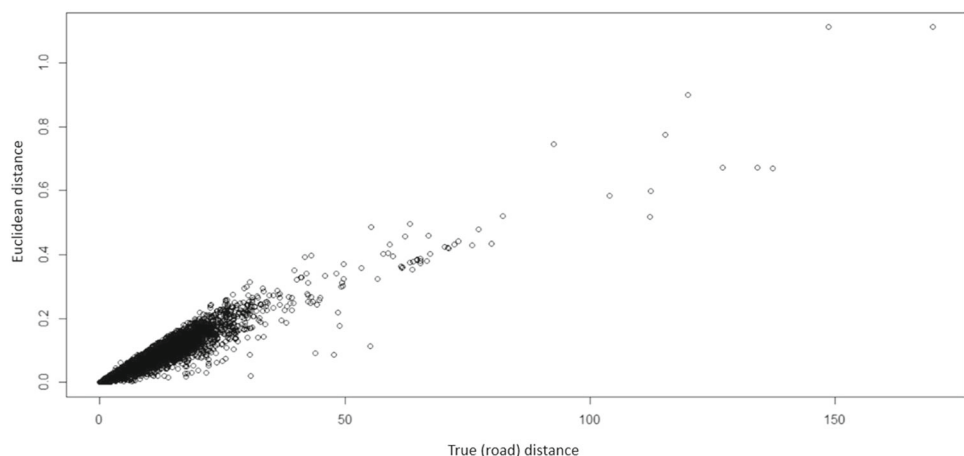
**Funding** Open access funding provided by FCTIFCCN (b-on). This work is funded by national funds through the FCT—Fundação para a Ciência e a Tecnologia, I.P., under the scope of the project DSAIPA/DS/0088/2019 and research and development units UNIDEMI (project UIDB/00667/2020) and NOVAMATH (projects UIDB/00297/2020 and UIDP/00297/2020).

## Declarations

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the

**Fig. 6** Relation between true (by road) and Euclidean distances from the incidents to the fire station’s locations



permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix A: Euclidean distances

To screen the spatial configuration of FS, we have calculated the Euclidean distances between the incidents and the FS locations (both actual and predicted) as a proxy to true road distance. This was done, to reduce the computational burden, following an exploratory analysis that showed a strong linear correlation between Euclidean and by road distances ( $r = 0.96$ , Fig. 6).

## References

1. Agbola, S.B., Falola, O.J.: Seasonal and locational variations in fire disasters in Ibadan, Nigeria. *Int. J. Disaster Risk Reduct.* **54**, 102035 (2021). <https://doi.org/10.1016/j.ijdrr.2021.102035>
2. Bispo, R., Vieira, F.G., Bachir, N., Espadinha-Cruz, P., Lopes, J.P., Penha, A., Marques, F.J., Grilo, A.: Spatial modelling and mapping of urban fire occurrence in Portugal. *Fire Saf. J.* **66**, 103802 (2023). <https://doi.org/10.1016/j.firesaf.2023.103802>
3. Chevalier, P., Thomas, I., Geraets, D., Goetghebeur, E., Janssens, O., Peeters, D., Plastria, F.: Locating fire stations: an integrated approach for Belgium. *Socio-Econ. Plan. Sci.* **46**(2), 173–182 (2012). <https://doi.org/10.1016/j.seps.2012.02.003>
4. Eslamzadeh, M.K., Grilo, A., Espadinha-Cruz, P., Rodrigues, J.P.C., Lopes, J.P.: A framework for fire departments' performance assessment: a systematic literature review. *Int. J. Public Sect. Manag.* **35**(3), 349–369 (2022). <https://doi.org/10.1108/IJPSM-06-2021-0149>
5. Dey, A., Heger, A., England, D.: Urban fire station location planning using predicted demand and service quality index. *Int. J. Data Sci. Anal.* **15**(1), 33–48 (2022). <https://doi.org/10.1007/s41060-022-00328-x>
6. ReVelle, C.S., Eiselt, H.A.: Location analysis: a synthesis and survey. *Eur. J. Oper. Res.* **165**(1), 1–19 (2005)
7. Huang, B., Liu, N., Chandramouli, M.: A GIS supported Ant algorithm for the linear feature covering problem with distance constraints. *Decis. Support Syst.* **42**(2), 1063–1075 (2006). <https://doi.org/10.1016/j.dss.2005.09.002>
8. Yang, L., Jones, B.F., Yang, S.-H.: A fuzzy multi-objective programming for optimization of fire station locations through genetic algorithms. *Eur. J. Oper. Res.* **181**(2), 903–915 (2007). <https://doi.org/10.1016/j.ejor.2006.07.003>
9. Badri, M.A., Mortagy, A.K., Alsayed, C.A.: A multi-objective model for locating fire stations. *Eur. J. Oper. Res.* **110**(2), 243–260 (1998). [https://doi.org/10.1016/S0377-2217\(97\)00247-6](https://doi.org/10.1016/S0377-2217(97)00247-6). EURO Best Applied Paper Competition
10. Tzeng, G., Chen, Y.: The optimal location of airport fire stations: a fuzzy multi objective programming and revised genetic algorithm approach. *Transp. Plan. Technol.* **23**(1), 37–55 (1999). <https://doi.org/10.1080/03081069908717638>
11. Aleisa, E.: The fire station location problem: a literature survey. *Int. J. Emerg. Manag.* **14**(3), 291–302 (2018). <https://doi.org/10.1504/IJEM.2018.094239>
12. Başar, A., Çatay, B., Ünlüyurt, T.A.: A taxonomy for emergency service station location problem. *Optim. Lett.* **6**, 1147–1160 (2012). <https://doi.org/10.1007/s11590-011-0376-1>
13. Indriasari, V., Mahmud, A.R., Ahmad, N., Shariff, A.R.M.: Maximal service area problem for optimal siting of emergency facilities. *Int. J. Geograph. Inf. Sci.* **24**(2), 213–230 (2010). <https://doi.org/10.1080/13658810802549162>
14. Bonneu, F., Thomas-Agnan, C.: Spatial point process models for location-allocation problems. *Comput. Stat. Data Anal.* **53**(8), 3070–3081 (2009)
15. Ceyhan, E., Ertuğay, K., Düzgün, Ş.: Exploratory and inferential methods for spatio-temporal analysis of residential fire clustering in urban areas. *Fire Saf. J.* **58**, 226–239 (2013). <https://doi.org/10.1016/j.firesaf.2013.01.024>
16. Wang, Y.: Optimization on fire station location selection for fire emergency vehicles using k-means algorithm. In: 2018 3rd International Conference on Advances in Materials, Mechatronics and Civil Engineering (ICAMMCE 2018), pp. 323–333. Atlantis Press (2018)
17. de Domingo, M., Ortigosa, N., Sevilla, J., Roger, S.: Cluster-based relocation of stations for efficient forest fire management in the Province of Valencia (Spain). *Sensors* **21**(3), 797 (2021)
18. Dey, A., Heger, A., England, D.: Urban fire station location planning using predicted demand and service quality index. *Int. J. Data Sci. Anal.* **66**, 1–16 (2022)
19. Khairani, N.A., Sutoyo, E.: Application of k-means clustering algorithm for determination of fire-prone areas utilizing hotspots in West Kalimantan province. *Int. J. Adv. Data Inf. Syst.* **1**(1), 9–16 (2020). <https://doi.org/10.25008/ijadis.v1i1.13>
20. Hamfelt, A., Karlsson, M., Thierfelder, T., Valkovsky, V.: Beyond K-means: clusters Identification for GIS. In: Popovich, V., Claramunt, C., Devogele, T., Schrenk, M., Korolenko, K. (Eds.) *Information Fusion and Geographic Information Systems. Lecture Notes in Geoinformation and Cartography*, pp. 93–105. Springer, Berlin (2011). [https://doi.org/10.1007/978-3-642-19766-6\\_8](https://doi.org/10.1007/978-3-642-19766-6_8)
21. Tribunal de Contas: Auditoria ao financiamento pelos municípios de corpos e associações de bombeiros. Relatório N° 5/2022 (2022)
22. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna (2021). R Foundation for Statistical Computing. <https://www.R-project.org/>
23. Pebesma, E.: Simple features for R: standardized support for spatial vector data. *R J.* **10**(1), 439–446 (2018). <https://doi.org/10.32614/RJ-2018-009>
24. Baddeley, A., Turner, R.: spatstat: an R package for analyzing spatial point patterns. *J. Stat. Softw.* **12**(6), 1–42 (2005). <https://doi.org/10.18637/jss.v012.i06>
25. Turner, R.: Deldir: Delaunay Triangulation and Dirichlet (Voronoi) Tessellation. R package version 0.1-29 (2020). <https://CRAN.R-project.org/package=deldir>
26. Baddeley, A., Rubak, E., Turner, R.: *Spatial Point Patterns: Methodology and Applications with R*. CRC Press, NY (2015)
27. Diggle, P.: *The Statistical Analysis of Spatial Point Patterns* (2003)
28. Bivand, R.S., Pebesma, E., Gomez-Rubio, V.: *Applied Spatial Data Analysis with R*, 2nd Edn. Springer, NY (2013). <https://asdar-book.org/>
29. Ripley, B.D.: Modelling spatial patterns. *J. R. Stat. Soc. Ser. B Methodol.* **39**(2), 172–192 (1977)
30. Baddeley, A.J., Møller, J., Waagepetersen, R.: Non-and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* **54**(3), 329–350 (2000)
31. Akaike, H.: *Information Theory and an Extension of the Maximum Ory*. by BN Petrov & F. Csaki, Eds (1973)
32. Akaike, H.: A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**(6), 716–723 (1974). <https://doi.org/10.1109/tac.1974.1100705>
33. Renkas, A., Popovych, V., Rudenko, D.: Optimization of fire station locations to increase the efficiency of firefighting in natural ecosystems. *Environ. Res. Eng. Manag.* **78**(1), 97–104 (2022). <https://doi.org/10.5755/j01.ere.m.78.1.25581>

34. Koivistoinen, H., Ruuska, M., Elomaa, T.: A Voronoi diagram approach to autonomous clustering. In: Discovery Science, pp. 149–160. Springer, Berlin (2006). [https://doi.org/10.1007/11893318\\_17](https://doi.org/10.1007/11893318_17)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.