

## Research Article

Zhen Shuo, Zhang Jingyu, Zhang Zhengxiang\*, and Zhao Jianjun

# Identifying the density of grassland fire points with kernel density estimation based on spatial distribution characteristics

<https://doi.org/10.1515/geo-2020-0265>

received September 15, 2020; accepted May 19, 2021

**Abstract:** Understanding the risk of grassland fire occurrence associated with historical fire point events is critical for implementing effective management of grasslands. This may require a model to convert the fire point records into continuous spatial distribution data. Kernel density estimation (KDE) can be used to represent the spatial distribution of grassland fire occurrences and decrease the influences historical records in point format with inaccurate positions. The bandwidth is the most important parameter because it dominates the amount of variation in the estimation of KDE. In this study, the spatial distribution characteristic of the points was considered to determine the bandwidth of KDE with the Ripley's K function method. With high, medium, and low concentration scenes of grassland fire points, kernel density surfaces were produced by using the kernel function with four bandwidth parameter selection methods. For acquiring the best maps, the estimated density surfaces were compared by mean integrated squared error methods. The results show that Ripley's K function method is the best bandwidth selection method for mapping and analyzing the risk of grassland fire occurrence with the dependent

or inaccurate point variable, considering the spatial distribution characteristics.

**Keywords:** grassland fire, risk, spatial cluster, Ripley's K

## 1 Introduction

Grassland fires are considered to be one of the major disturbances affecting natural ecosystems in grass-dominated regions with human activities, and the emissions of greenhouse gas is a main factor that influences the global climate change [1–4]. Annual fires are also problematic for soil erosion, grass resources, and biodiversity and diminish the production of agriculture and livestock [5]. In this sense, identifying and assessing the risk of grassland fire ignition occurrence are key issues for the adequate design and decision-making of fire management [6,7].

In wildland fire studies, the spatial pattern and distribution of fire events are essential to wildfire managers or scientists [1]. And continuous distribution surface of fire occurrence is widely used in the explanation for the spatial pattern of fire risk or the analysis of the influencing factors by decreasing the inaccuracies and substantial errors of fire history record as a point with x- and y-coordinates [1,8,9]. The methods that can convert the fire record point into continuous density are needed [8,10].

Kernel density estimate (KDE) is a nonparametric estimation of deriving a continuous surface with the underlying unknown density function [11]. And it was applied to effectively convert point-based events into grid-based distribution for analyses in various disciplines, such as animal home range estimation in wildlife ecology, traffic accident density estimation, analysis of central business districts, and seismic risk analysis in geology [12–14].

In wild fire studies, KDE is used extensively to convert fire historic point records into continuous fire occurrence density surfaces for defining the spatial distribution of fire ignition at different scales, analyzing its extent and

\* Corresponding author: Zhang Zhengxiang, Key Laboratory of Geographical Processes and Ecological Security in Changbai Mountains, Ministry of Education, School of Geographical Sciences, Northeast Normal University, Changchun 130024, China, e-mail: zhangzx040@nenu.edu.cn

Zhen Shuo: Key Laboratory of Geographical Processes and Ecological Security in Changbai Mountains, Ministry of Education, School of Geographical Sciences, Northeast Normal University, Changchun 130024, China

Zhang Jingyu: Faculty of Geographical Sciences, Beijing Normal University, Beijing 100857, China

Zhao Jianjun: Key Laboratory of Geographical Processes and Ecological Security in Changbai Mountains, Ministry of Education, School of Geographical Sciences, Northeast Normal University, Changchun 130024, China, e-mail: Zhaojj662@nenu.edu.cn

distribution, or representing the fire risk associated with the influencing factors [6,8,10,15–17]. KDE method has also been used in grassland fire analysis [10,17].

The kernel density function was defined [18] as follows:

$$\hat{f}(x) = \frac{1}{nh^2} \sum_{i=1}^n K \left\{ \frac{s - s_i}{h} \right\}, \quad (1)$$

$n$  is the total number of observed events,  $K$  is the kernel function,  $h$  is the bandwidth that defines the searching radius of the kernel function,  $s$  is the position where the estimation density value is being calculated, and  $s_i$  is the position of each observed event.

The bandwidth controls the smoothness of the KDE surface, and it is considered more critical than the selection of kernel functions [19]. A large value of  $h$  will produce a smooth distribution, but it may cause a loss of details on the resulting surfaces. In contrast, when using a narrow  $h$ , the estimated surface tends to show too many finer variations that often obscure the clustering characteristic [6].

Some methods were applied to set appropriate bandwidths that are widely used in many studies [20,21]. In one approach, the bandwidth is defined by the mean polygon size in the study area [20]. The mean polygon size is computed, and a theoretical square is created that has the same area as the mean polygon. Hence, the bandwidth of the square can be defined as the radius of the circle around the square, which also is half of the diagonal of the square as follows:

$$h = \frac{D}{2}, \quad (2)$$

where  $D$  is the diagonal of the theoretical square.

$RD_{mean}$  approach is given based on the calculation of local mean random distance [20]:

$$RD_{mean} = \frac{1}{2} \sqrt{\frac{A}{N}}, \quad (3)$$

where  $A$  is the mean size polygon and  $N$  is the average number of point events in a polygon. The final value of bandwidth is given by multiplying  $RD_{mean}$  by two [8].

Another bandwidth based on all of the observed points is defined [22]. Because it does not consider the area, the value of the bandwidth was altered according to the dimension of the area through multiplying it by the square root of the area, and it is named as Williamson function [23]:

$$h = 0.68N^{-0.2}\sqrt{A}, \quad (4)$$

where  $N$  is the number of all observed points, 0.68 is a constant, and  $A$  is the study area size.

The first method does not consider the number of observed points except the polygon of the study area. The last two methods consider the number of observed points and the size of the study area. And these bandwidths are used to calculate the concentrations of observed events in a specified area. With a high concentration of observed events, the value of bandwidth is small. And when the concentration is low, the value of bandwidths is big [21].

In our previous works,  $RD_{mean}$  is used to determine the bandwidth of the KDE method for providing the grassland fire regime studies by the fire density maps with MODIS fire active production in the eastern Inner Mongolia of China [10,17]. Although the results have appropriately expressed the characteristic of grassland fire distribution in the whole study area, when analyzing the fire occurrence risk in the local area, the calculated fire density varies markedly with the same influence factors. At or near the observed fire points, the density is higher. Around the points, the density is lower. Hence, the risk in the area with the observed points will be overestimated and the risk in the area without or near the observed points will be underestimated. These bandwidth selection methods don't consider the spatial correlation of locations of the observed points. Mostly, the characteristic of the point distribution pattern was not considered in these methods for defining the bandwidth. In a specified area with the same number of points, the characteristic of the distribution can be clustered, random or disperse, but the concentration characteristic of points in the area can be the same or similar if don't consider the spatial distribution of events, as shown in Figure S1. Then, only using the area and the number of points in the methods can't express the characteristic of spatial distribution. In the field, once a location is targeted by a fire, nearby locations face an elevated risk of experiencing the same event shortly. Thus, spatial point pattern analysis has been intensively investigated to reveal the scale, extent, and dynamics of point events and to test potential patterns related to spatial mechanisms of fire occurrence. Due to the distribution of wildland fire events is a cluster, the characteristic of the spatial distribution pattern should be considered when defining the bandwidth of KDE [24].

The main purpose of this study is to identify the spatial continuous distribution surface of grassland fire occurrence with improving the bandwidth selection method of KDE by considering the spatial distribution patterns of observed point events and to provide guidance on which method most accurately determines the underlying density. The study area is in eastern Inner Mongolia, China, which has suffered from a high grassland fire incident rate, using

KDE methods with MODIS fire active products (for the period from 2001 to 2014). This method demonstrates that such a technique, which considers the spatial distribution characteristics, can be extended to different spatial data analyses. To achieve this objective, several methods of choosing the bandwidth were applied. Then, the KDE maps were compared to reveal the effects of different bandwidths. A case study is expressed by figuring out the distribution pattern of the spatial point data, and the suitable bandwidth selection can easily be implemented for geographical areas. The results of the current research can also be a valuable reference for policy-making in fire management.

## 2 Study area and methods

### 2.1 Study area and data

Hulunbuir is located in the northeast Inner Mongolia Autonomous Region of China, from  $47.08^{\circ}\text{N}$  to  $53.23^{\circ}\text{N}$  and  $115.22^{\circ}\text{E}$  to  $126.06^{\circ}\text{E}$  (Figure 1). The study area is approximately 681 km long and 703 km wide, covering a total area of  $252,948 \text{ km}^2$ . The climate in the region is a typical temperate continental monsoon climate, with low, irregular rainfall and extreme temperature variations in summer and winter. The annual mean air temperature and precipitation are about  $-2.3^{\circ}\text{C}$  and 320 mm, respectively. The vegetation in the eastern Inner Mongolia grassland region consists of diverse plant communities, which

are dominated by *Stipa baicalensis*, *Filifolium sibiricum*, and *Leymus chinensis*. The elevation ranges from 200 to 1500 m. The Greater Khingan Mountains is located in the middle region, stretching from the center of the region toward the east and west. The elevation gradually decreases, and the topography gradually becomes flatter. The activities of the inhabitants lead to accidental fires. Approximately 600 wild fires occur annually in this region. In previous studies, KDE method was used to access the risk of grassland fire in this area, which is why it was chosen as the study area to analyze the bandwidth selection method [10,17].

In the study area, the MODIS Active Fire Data contain daily fire pixel coordinate positions and they are most suitable to determine the spatial and temporal distributions of fire points. The MODIS Terra and Aqua daily active fire product data are available (MOD14A1 and MYD14A1). The fire product data set from 2001 to 2014 was downloaded from the Land Processes Distributed Active Archive Centre (LP-DAAC) using a web-based interface known as Reverb, which is a replacement for the Warehouse Inventory Search Tool (<http://reverb.cs.washington.edu/>). The resolution of the fire product is 1 km pixels, in which burning was detected at the times of Terra and Aqua satellite overpass under the condition of less cloudiness. Calculated from 4 and 11  $\mu\text{m}$  channels, the brightness temperature of fire pixels was obtained based on an algorithm that enhances the detection sensitivity of smaller, cooler fires and decreases the emergence of false alerts [25–28].

As the major component of the MODIS active fire product, the fire mask is stored as an 8-bit unsigned integer Scientific Data Set (SDS). In the data set, the value of each pixel is assigned to 1–9 classes [29]. These classes are listed in Table 1. According to the meaning of data code, 7, 8, and 9 are, respectively, low confidence fire point, medium confidence fire point, and high confidence fire point. Meanwhile, previous studies have shown that

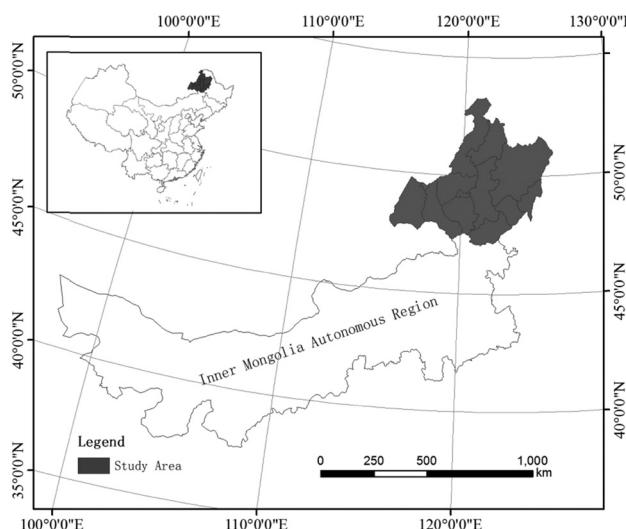


Figure 1: The location of the study area in the northeast Inner Mongolia Autonomous Region of China.

Table 1: MOD14/MYD14 fire mask pixel classes

Class	Meaning
0	Not processed (missing input data)
2	Not processed (other reason)
3	Water
4	Clouds
5	Non-fire clear land
6	Unknown
7	Low-confidence fire
8	Nominal-confidence fire
9	High-confidence fire

the combination of 8 and 9 has the best accuracy of MODIS active fires [30]. The pixels with values of 8 and 9 were extracted from the MOD14A1 and MYD14A1 downloaded collections and converted from raster to vector format as active fire points [10].

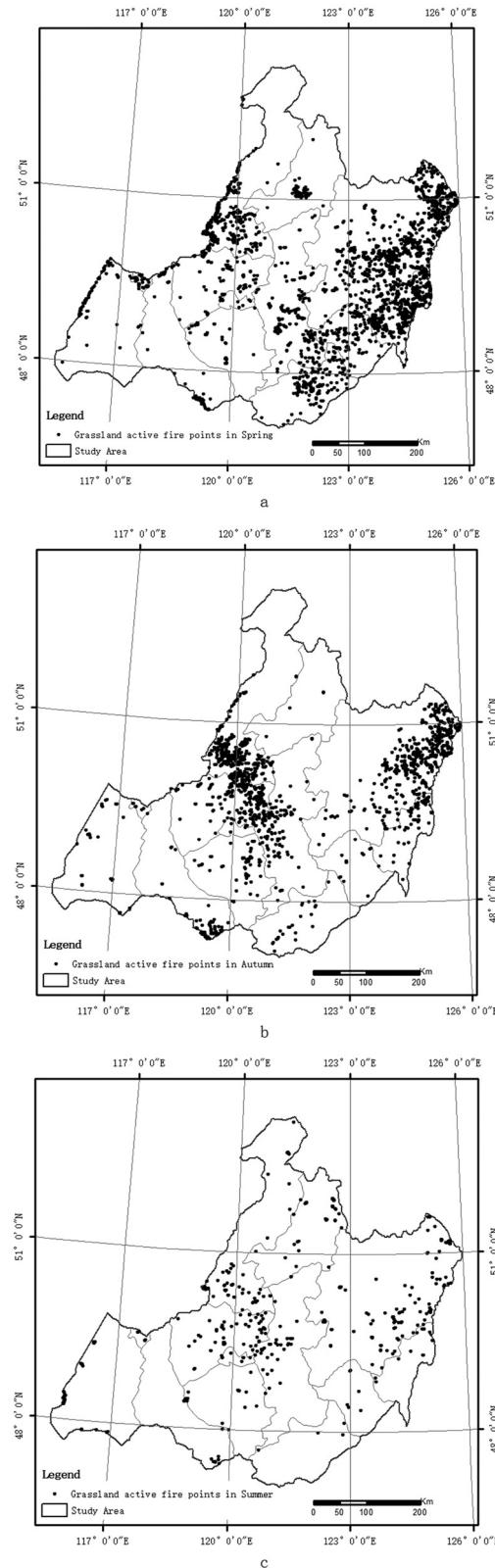
The land use data set of the study area was offered by the Data Centre for Resources and Environmental Sciences at the Chinese Academy of Sciences (RESDC) (<http://www.resdc.cn>). These data were obtained from the interpretation of Landsat TM/ETM from 2010 at the scale of 1:100,000. The format of the land use data set was raster. The grid with values representing the grassland type was selected and extracted as grassland data. Then, the active fire data sets from 2001 to 2014 were overlaid to a whole layer and then clipped by grasslands to erase the non-grassland fire events. The occurrence times and coordinates were saved in an attribute table [10]. The total number of grassland fire events in the study area was 4628 over the 14 years (2001–2014). Annually, the peak months were April, March, September, and May, and the months with the fewest events were January, December, and July. The time series reveals seasonal patterns: a large number of events occurred between March and May, with a maximum in April. A secondary large number occurred between August and October, with a peak in September. The probabilities of grassland fire occurrence in the study area in decreasing order are spring, autumn, summer, and winter, and the total number of fire events in each season was 2639, 1505, 406, and 78, respectively. Because there were nearly no events in winter, the season was excluded from this research. Three different mean density categories based on area divided by the numbers of seasonal fire points were set to analyze the bandwidth selection. Figure 2 shows the distribution of active grassland fire events in spring, autumn, and summer.

The standard map of the study area in digital format and grassland fire data obtained from MODIS were all set to a Transverse Mercator projection and D\_WGS\_1984 datum.

### 3 Methods

#### 3.1 Kernel bandwidths

As a nonparametric statistical approach, KDE has been extensively applied in some research fields to estimate the probability densities of events [6,8,16,31,32]. There



**Figure 2:** Distribution of the grassland active fire points from 2001 to 2014 in spring (a), autumn (b), and summer (c) observed by MODIS.

are several kernel functions, and the Epanechnikov kernel function was used in this study [19,20].

The choice of the bandwidth value is the most critical step in the analysis of KDE. When calculating the density of the Kernel surface, the bandwidth parameter determines the distance to find neighbors and affects the smoothness of the output density surfaces [20]. At present, the commonly used bandwidth selection methods are mainly based on these three methods: theoretical bandwidth,  $RD_{mean}$ , and Williamson function [8,20–23]. Four different bandwidth selection methods were used to identify the bandwidth parameter of the KDE model for getting the appropriate continuous surface. The first three methods were calculated according to the previous functions (2)–(4) with the active grassland fires in spring, autumn, and summer.

To characterize the spatial distribution patterns of historical fire point events, Ripley's K function was used as it determines the events that show a statistically significant spatial clustering or dispersion by cumulating the distributional frequency of the distances among the events with the changing of neighborhood size [12,33,34]. Commonly, Ripley's K function is transformed as a linear  $L$  function for facile interpretation [35]. The  $L$  function is implemented as follows:

$$L(d) = \sqrt{\frac{A \sum_{i=1}^n \sum_{j=1, j \neq i}^n k_{i,j}}{\pi n(n-1)}}, \quad (5)$$

where  $d$  is the distance,  $n$  is equal to the total number of point events,  $A$  is the total area,  $k_{i,j}$  is a weight, which (if there is no boundary correction) is 1 when the distance between  $i$  and  $j$  is less than or equal to  $d$  and 0 when the distance between  $i$  and  $j$  is greater than  $d$ .

The Ripley's K tool that was used is a built-in function of ArcGIS software (<https://desktop.arcgis.com/zh-cn/arcmap/latest/tools/spatial-statistics-toolbox/multi-distance-spatial-cluster-analysis.htm>). The output of Ripley's K tool includes the values of ExpectedK, ObservedK, DiffK, LwConfEnv, and HiConfEnv. The ObservedK value refers to the actual density of points in different distances, and the ExpectedK value refers to the expected distribution in the case of random distribution. LwConfEnv and HiConfEnv contain confidence interval information for each iteration of the tool. If the ObservedK of the specific distance is greater than the ExpectedK, the distribution is more clustering than that of the random distribution. If the ObservedK is less than the ExpectedK, the dispersion of the distribution is higher than that of the random distribution. If the ObservedK is greater than the HiConfEnv, the spatial cluster characteristic for the

distance is statistically significant. If the ObservedK is less than LwConfEnv, the spatial dispersion characteristic for the distance is statistically significant. The DiffK contains the difference between the ObservedK and the ExpectedK. The maximum DiffK will determine the most obvious distance where the spatial clustering process is most pronounced.

If the spatial distribution of points is a cluster, half of the maximum DiffK value is applied as the recommended value of the bandwidth that represents the clustering characteristic of the spatial distribution pattern of active grassland fires. If the distribution of points is not a cluster, there is no use for this DiffK value and it cannot be used as a bandwidth. In this circumstance, other methods, such as Theoretical bandwidth,  $RD_{mean}$ , and Williamson, should be used to calculate the KDE surface.

### 3.1.1 KDE calculations and comparison

For estimating the kernel densities of active grassland fire, Ripley's K and kernel density tools for spatial analysis of points were provided by ArcGIS software (V10.2). The calculation of KDE was conducted using a grid of  $1000 \times 1000$  m resolution with bandwidths defined by the four methods.

For reflecting the effect of the bandwidth selection method, the grassland active fire points with different concentrations according to spring, summer, and autumn were used to calculate the density values. A raster mask layer was used to control the boundary of results in the study area map.

The best bandwidth was selected by observing the variability of the estimated density surface according to the characteristic of the spatial distribution, which avoided too spiky or excessive smoothing surfaces [8]. The selected bandwidth is the one that provides an acceptable medium between these two extremes.

In theory, the accuracy of the density estimation is accessed by comparing the mean-square error (MSE) between the estimated density and the true density [36]. The mean integrated squared error (MISE) of different bandwidths of kernel estimates is estimated as follows to determine which one best fits the true distribution [37]:

$$MISE = \frac{1}{n} \sum_{i=1}^n \frac{[\hat{f}(x) - f(x)]^2}{f(x)}, \quad (6)$$

where  $n$  is the total amount of active fire point events,  $\hat{f}(x)$  is the estimated kernel density, and  $f(x)$  is the true density at the location of active fire points. By calculating

the density estimate of the error on the same point for four density estimation methods, MISE provides a useful comparison among the methods of bandwidth selections.

To acquire the true density of each active fire point, an adaptive procedure is defined:

$$D_{\text{tr}} = \frac{A_{(r)i,k}}{k}, \quad (7)$$

where  $D_{\text{tr}}$  is the true density for an active fire point  $i$ ,  $k$  is the number of adaptive nearest neighbor points,  $r$  is the distance to the  $k$ th nearest neighbor point at the sample point  $i$ , and  $A_{(r)i,k}$  is the circular area with the radius  $r$ . This procedure emphasizes the estimation in the vicinity of each active fire position, rather than estimating the density of the whole samples in the study area. According to the number of observed points in spring, summer, and autumn,  $k$  is defined as 10, 20, and 30, respectively. The mean true density value of each active fire point is calculated by function (7) at these three neighbor numbers. Combined with equations (6) and (7), the minimum value of MISE shows the best bandwidth of KDE for active grassland fire points.

## 4 Results

Using the analysis tool of Ripley's K, the spatial distribution pattern of grassland fire events was investigated in spring, autumn, and summer. The same distribution characteristic of the fire points is found to be spatially clustered in these three seasons (Figure 3). For the values of ExpectedK greater than those of HiconfEnv, the statistical tests show that the cluster distributions are significant. The largest DiffK values, which indicate the most pronounced distance of spatial processes promoting clustering, are 99, 82, and 80 km in spring, autumn, and summer, respectively. In this circumstance, half of the value of maximum DiffK can be applied as the bandwidth of KDE that represents the clustering characteristic of the spatial distribution pattern of active grassland fires in the three seasons.

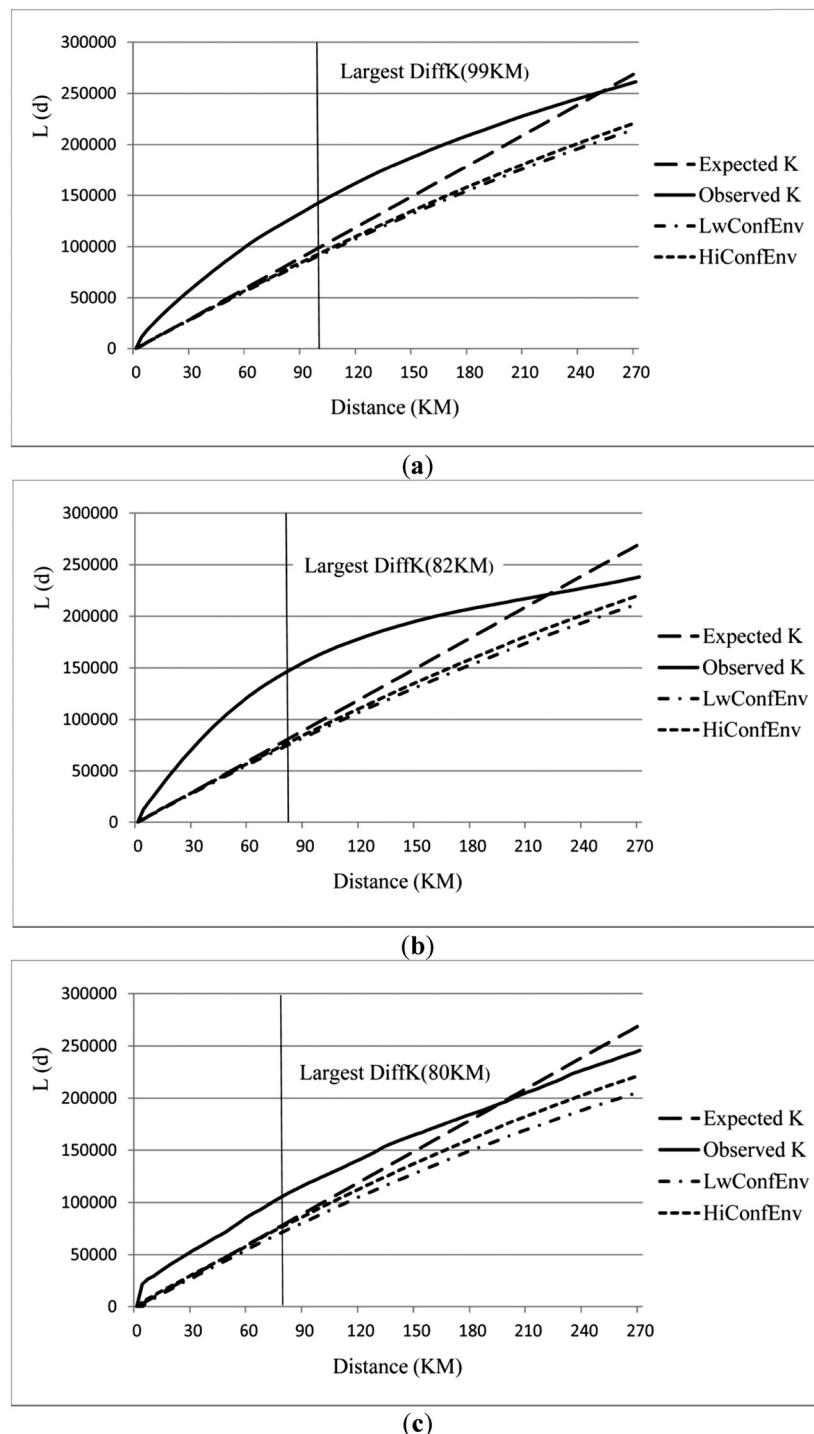
Parameters related to the bandwidth calculations and the values of the bandwidth calculated according to the previous functions (2)–(4) with the grassland active fires are presented in detail in Tables 2 and 3. The differences among these bandwidth selection methods are obvious. In all seasons, the bandwidths of the theoretical bandwidth method are largest, and the values of

$RD_{\text{mean}}$  are the smallest. The bandwidths of the other two methods are mid-range. In high, medium, and low concentration scenarios, the differences of bandwidth values of  $RD_{\text{mean}}$  and Williamson methods are greater than those of the other two methods.

Using the bandwidth values of different methods, seasonal KDE maps of continuous surfaces of fire events were obtained (Figure 4). According to the seasonal KDE maps, the densities of active grassland fire points show different spatial distribution patterns, although all of them show the clustering characteristic (Figure 4). In all seasons, the density surfaces calculated by the  $RD_{\text{mean}}$  method have spikier than those calculated with other methods. In contrast, the results calculated with the theoretical bandwidth method are the smoothest surfaces. With Williamson and Ripley's K methods, the degrees of surface smoothness are medium. The difference of maximum density values of different methods is substantial in all seasons. The method with the biggest density value is  $RD_{\text{mean}}$ , followed by Ripley's K, Williamson, and Theoretical bandwidths. The estimated density value negatively correlates to the value of bandwidth, and the surface smoothness is positively correlated with the bandwidth value. Thus, with increasing values of bandwidth, the density surfaces become smoother, and the maximum density values become smaller.

An accuracy assessment approach is added to select the surface which better expresses the fire occurrence distribution. In the measurements of the density estimator using MISE, the true density of each active grassland fire point was calculated by averaging the value with function (7) at 10, 20, and 30 neighbor distances of each fire event. The density estimation of each active grassland fire point was derived from the density surfaces of the four methods in spring, autumn, and summer. The MISE values of the four methods in all seasons are shown in Table 4. The differences in MISE among the various bandwidths of the four methods are quite large. Ripley's K method for calculating bandwidth has the lowest MISE values among the methods in spring, autumn, and summer. The MISE shows that the kernel density with Ripley's K method gives the best density estimation of grassland fires in high, medium, and low concentration scenes.

Therefore, the continuous spatial density surfaces of grassland fire in different concentration scenes were obtained as shown in Figure 4(F5). These density surfaces can be used in grassland fire management practice and risk research.



**Figure 3:** The curves of Ripley's K function for active grassland fire events in spring (a), autumn (b), and summer (c) in the study area.

## 5 Discussion

Choosing the suitable bandwidth is very critical because it determines the range of variation of the estimation in KDE [21]. These four bandwidth selection methods show different abilities in calculating the density values of

KDE. We found that the bandwidth selection of Ripley's K method performed better than the other methods in high, medium, and low concentration scenes of our samples. The previous methods that considered the sample and region size are reasonable because the amount of the observed points has a relationship with the information

**Table 2:** Parameters related to bandwidth calculations for different methods

Parameter	Value
Total size of the study area ( $A$ )	252,948 km <sup>2</sup>
Total number of polygons	13
Mean polygon size	19,594 km <sup>2</sup>
Total number of active fire events in Spring ( $N$ )	2639
Total number of active fire events in Autumn ( $N$ )	1505
Total number of active fire events in summer ( $N$ )	406

**Table 3:** The values of bandwidths for different methods

Method	Spring (km)	Autumn (km)	Summer (km)
Theoretical bandwidth (function 2)	197	197	197
$RD_{mean}$ (function 3)	10	13	25
Williamson (function 4)	71	79	103
Ripley's K (function 5)	49.5	41	40

content. If the amount of sample points is more, which represents a large informative dataset, a smaller bandwidth is more suitable for avoiding over smooth and loss of the variability in the estimation. In contrast, if the amount of sample points is small, which represents a minor informative dataset, a larger bandwidth is more proper because a smaller bandwidth will cause the estimated density that has little contact with neighbor points [32]. In essence, these bandwidth selection methods represent the concentration of the samples in a specific area, and the difference in spatial distribution pattern is neglected. In spatial datasets, especially for wildland fire samples, the clustering characteristic is the main distribution pattern, and these samples have the characteristic of spatial autocorrelation. The bandwidth selection method of KDE by Ripley's K can express the character of the distribution pattern.

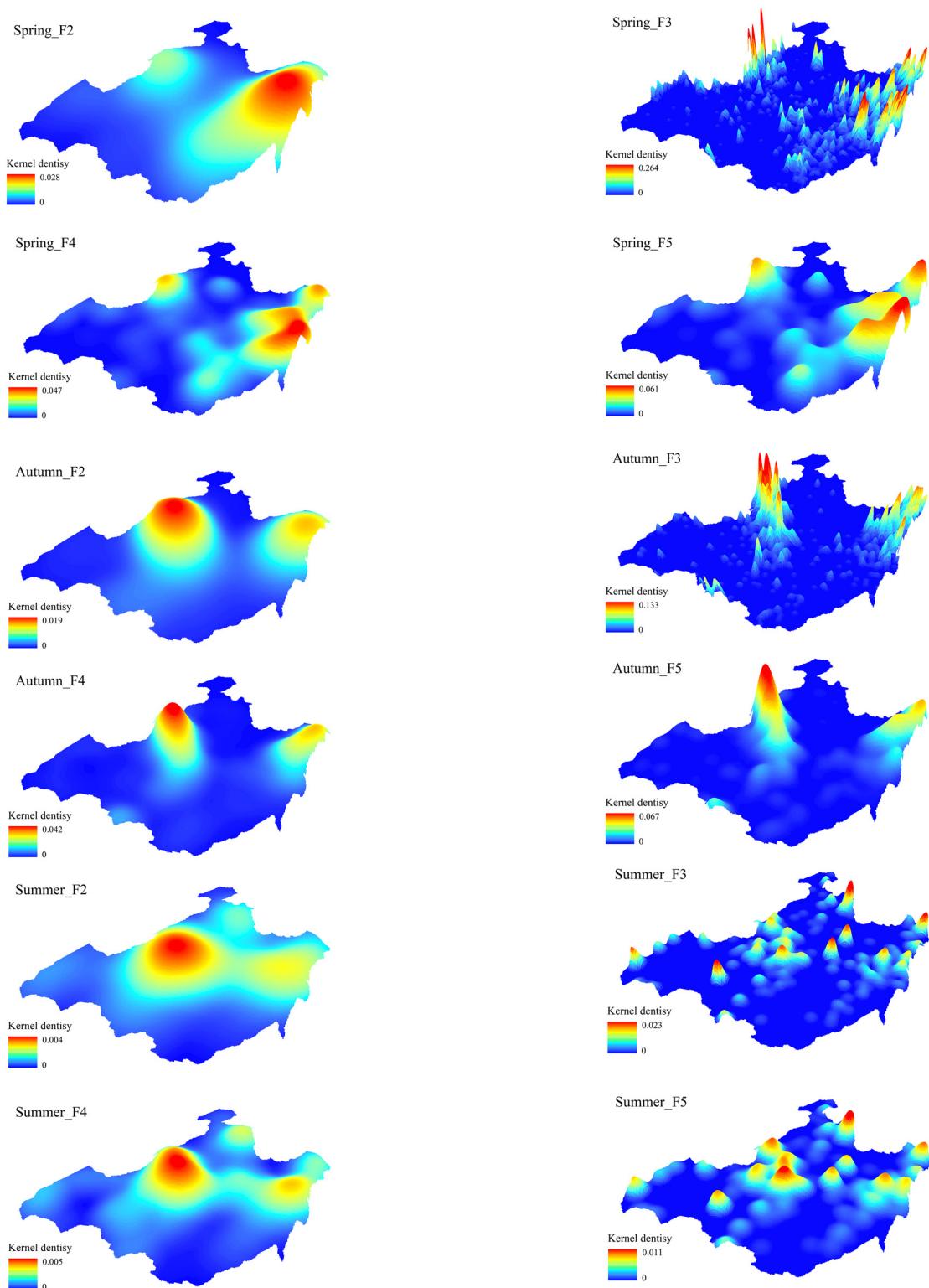
In this study, the boundary correction method in process of applying Ripley's K was not considered. In this case, the neighbor number of samples near the edge was underestimated. Increasing the number of samples outside the study area is one method to rectify this situation. Another method is to reduce the analyzed area and make some samples that occur outside of the reduced area. In this research, the previous three methods of bandwidth selection do not have the requirement of boundary correction. When applying Ripley's K function, the boundary correction was neglected to maintain consistency of the

area and the sample size with the other methods. When using the bandwidth selection method of Ripley's K in other studies, the correction method should be considered to increase the accuracy of the estimated density surface, especially at the edge of the study area.

The best bandwidth selection by the MISE method is affected by the true density of samples. These two things cannot be acquired exactly. For the expert, these two things can also be presented by intuitive graphics in Figure 4. Thus, the three-dimensional graph is an effective method to achieve the objective. In the measurements of the accuracy of KDE using MISE, the neighbor number or the radius size is a critical factor in the identification of the true density of the samples. To decrease the deviation of true densities of active grassland fire points, the mean densities of each sample at 10, 20, and 30 neighbor distances, which are also subjective, were applied in this study. The true density of samples cannot be known, so expert knowledge of the study area is needed.

It is a fact that the number of samples and the distribution pattern affect the degree of smoothing. In our study, the sample points of fire filtered by grassland represent a discontinued part of the area, and the distribution characteristic is a cluster. If the samples are generated for the whole area, Ripley's K function must be used first for examining the distribution characteristic. When the ObservedK of the specific distance is greater than the ExpectedK and HiConfEnv, the distribution characteristic is a significant cluster and the density of these samples can be estimated by our method. If the ObservedK values are less than the confidential envelopes (LwconfEnv, HiconfEnv), a bias is generated between the samples and our methods because the sample distribution is dispersed or random. In this situation, the bandwidth selection method of Ripley's K may not be better for calculating the kernel density surface. The previous bandwidth selection methods, especially  $RD_{mean}$  and Williamson, may also be used. However, it is important to apply this new bandwidth selection method in many other studies because clustering is one of the important characteristics of spatial elements.

The effect of KDE is to calculate the density of samples for converting point data into continuous data. When calculating the densities with our method, the type of spatial distribution of points should be calculated first. If the spatial distribution type of points is a cluster, the optimal bandwidth (half of the DiffK) for the dataset will be calculated with Ripley's K. However, other methods should be used to calculate the bandwidth of the KDE map.



**Figure 4:** The KDE maps in spring, autumn, and summer, according to different bandwidth selection methods (F2: Theoretical bandwidth, function 2; F3: RD<sub>mean</sub>, function 3; F4: Williamson, function 4; F5: Ripley's K, function 5).

**Table 4:** Mean integrated squared error (MISE) for different bandwidths of kernel estimates in all seasons

Method	MISE		
	Spring	Autumn	Summer
Theoretical bandwidth (function 2)	108.30	34.67	2.83
RD <sub>mean</sub> (function 3)	92.06	32.40	5.49
Williamson (function 4)	81.32	20.10	2.60
Ripley's K (function 5)	68.17	11.23	2.47

## 6 Conclusion

In this study, the eastern Inner Mongolia Autonomous Region of China, one of the main regions in the Asian grassland that was significantly disturbed by grassland fires, was selected as the study area. The MODIS Terra and Aqua daily active fire product data (MOD14A1 and MYD14A1) belonging to this region were downloaded from LP-DAAC. As a nonparametric method for obtaining continuous surfaces from point observations, KDE has been widely used to convert point-based data into density maps and to decrease the positional inaccuracy at the same time. To find an adaptive bandwidth for fitting the grassland fire events and the study area, a bandwidth selection method on the basis of Ripley's K function, which considered the clustering characteristic of the grassland fire events distribution pattern, was developed. It demonstrated that the developed bandwidth selection method is better than the previous methods in high, medium, and low concentration scenes. Applying this method can promote wildland fire management in areas that are more prone to fire damage. Our approach provides a general bandwidth selection method for KDE in different studies as long as the samples have the clustering characteristics of a distribution pattern.

**Funding information:** This work was supported by the National Natural Science Foundation of China under Grant [number 41977407 and 41571489]; Jilin Provincial Science and Technology Development Project (20190101025JH).

**Author contributions:** Z.X.Z. proposed the idea, J.Y.Z. and S.Z. conducted data collection and experimental analysis, and J.J.Z. supervised the experiment. S.Z. wrote the manuscript. All the authors had read the manuscript and agreed to submit it.

**Conflict of interest:** The authors do not have any possible conflicts of interest.

**Data availability statement:** Only freely and publicly available datasets were used in this study. Data sources were described in the paper.

## References

- [1] Hao WM, Liu MH. Spatial and temporal distribution of tropical biomass burning. *Global Biogeochem Cy.* 1994;8(4):495–503. doi: 10.1029/94gb02086.
- [2] Noymeir I. Interactive effects of fire and grazing on structure and diversity of mediterranean grasslands. *J Veg Sci.* 1995;6(5):701–10. doi: 10.2307/3236441.
- [3] Ojima DS, Schimel DS, Parton WJ, Owensby CE. Long-term and short-term effects of fire on nitrogen cycling in tallgrass prairie. *Biogeochemistry.* 1994;24(2):67–84. doi: 10.1007/bf02390180.
- [4] Oom D, Pereira JMC. Exploratory spatial data analysis of global MODIS active fire data. *Int J Appl Earth Obs.* 2013;21:326–40. doi: 10.1016/j.jag.2012.07.018.
- [5] Chuvieco E, Aguado I, Yebra M, Nieto H, Salas J, Pilar Martin M, et al. Development of a framework for fire risk assessment using remote sensing and geographic information system technologies. *Ecol Model.* 2010;221(1):46–58. doi: 10.1016/j.ecolmodel.2008.11.017.
- [6] Amatulli G, Perez-Cabello F, de la Riva J. Mapping lightning/human-caused wildfires occurrence under ignition point location uncertainty. *Ecol Model.* 2007;200(3–4):321–33. doi: 10.1016/j.ecolmodel.2006.08.001.
- [7] Pew KL, Larsen CPS. GIS analysis of spatial and temporal patterns of human-caused wildfires in the temperate rain forest of Vancouver Island, Canada. *Forest Ecol Manag.* 2001;140(1):1–18. doi: 10.1016/s0378-1127(00)00271-1.
- [8] Koutsias N, Kalabokidis KD, AllgÖwer B. Fire occurrence patterns at landscape level: beyond positional accuracy of ignition points with kernel density estimation methods. *Nat Resour Model.* 2004;17(4):359–75. doi: 10.1111/j.1939-7445.2004.tb00141.x.
- [9] Martinez-Fernandez J, Chuvieco E, Koutsias N. Modelling long-term fire occurrence factors in Spain by accounting for local variations with geographically weighted regression. *Nat Hazard Earth Syst.* 2013;13(2):311–27. doi: 10.5194/nhess-13-311-2013.
- [10] Zhang ZX, Feng ZQ, Zhang HY, Zhao JJ, Yu S, Du W. Spatial distribution of grassland fires at the regional scale based on the MODIS active fire products. *Int J Wildland Fire.* 2017;26(3):209–18. doi: 10.1071/wf16026.
- [11] Waller LA, Gotway CA. Applied spatial statistics for public health data. Hoboken, New Jersey, USA: John Wiley & Sons, Inc.; 2004.
- [12] Gatrell AC, Bailey TC, Diggle PJ, Rowlingson BS. Spatial point pattern analysis and its application in geographical epidemiology. *T I Brit Geogr.* 1996;21(1):256–74. doi: 10.2307/622936.
- [13] Steiniger S, Hunter AJS. A scaled line-based kernel density estimator for the retrieval of utilization distributions and home

ranges from GPS movement tracks. *Ecol Inform.* 2013;13:1–8. doi: 10.1016/j.ecoinf.2012.10.002.

- [14] Yu W, Ai T, Shao S. The analysis and delimitation of central business district using network kernel density estimation. *J Transp Geogr.* 2015;45:32–47. doi: 10.1016/j.jtrangeo.2015.04.008.
- [15] Gonzalez-Olabarria JR, Mola-Yudego B, Coll L. Different factors for different causes: analysis of the spatial aggregations of fire ignitions in catalonia (Spain). *Risk Anal.* 2015;35(7):1197–209. doi: 10.1111/risa.12339.
- [16] Kuter N, Yenilmez F, Kuter S. Forest fire risk mapping by kernel density estimation. *Croat J For Eng.* 2011;32(2):599–610.
- [17] Li YP, Zhao JJ, Guo XY, Zhang ZX, Tan G, Yang JH. The influence of land use on the grassland fire occurrence in the Northeastern Inner Mongolia autonomous region, China. *Sensors (Basel).* 2017;17(3):437. doi: 10.3390/s17030437.
- [18] Silverman BW. Density estimation for statistics and data analysis. London, UK: Chapman & Hall; 1986.
- [19] Kuter S, Usul N, Kuter N. Bandwidth determination for kernel density analysis of wildfire events at forest sub-district scale. *Ecol Model.* 2011;222(17):3033–40. doi: 10.1016/j.ecolmodel.2011.06.006.
- [20] de la Riva J, Perez-Cabello F, Lana-Renault N, Koutsias N. Mapping wildfire occurrence at regional scale. *Remote Sens Environ.* 2004;92(3):363–9. doi: 10.1016/j.rse.2004.06.022.
- [21] Worton BJ. Kernel methods for estimating the utilization distribution in home-range studies. *Ecology.* 1989;70(1):164–8. doi: 10.2307/1938423.
- [22] Bailey TC, Gatrell AC. Interactive spatial data analysis. Essex: Longman; p. 413.
- [23] Williamson D, McLafferty S, Goldsmith V, Mollenkopf J, McGuire PA. Better method to smooth crime incident data. *ESRI ArcUser Magazine.* California, USA: RedLands; 1999 Jan–Mar.
- [24] Telesca L, Amatulli G, Lasaponara R, Lovallo M, Santulli A. Time-scaling properties in forest-fire sequences observed in Gargano area (southern Italy). *Ecol Model.* 2005;185(2–4):531–44. doi: 10.1016/j.ecolmodel.2005.01.009.
- [25] de Klerk H. A pragmatic assessment of the usefulness of the MODIS (Terra and Aqua) 1 km active fire (MOD14A2 and MYD14A2) products for mapping fires in the fynbos biome. *Int J Wildland Fire.* 2008;17(2):166–78. doi: 10.1071/wf06040.
- [26] Giglio L, Descloitres J, Justice CO, Kaufman YJ. An enhanced contextual fire detection algorithm for MODIS. *Remote Sens Environ.* 2003;87(2–3):273–82. doi: 10.1016/s0034-4257(03)00184-6.
- [27] Giglio L, van der Werf GR, Randerson JT, Collatz GJ, Kasibhatla P. Global estimation of burned area using MODIS active fire observations. *Atmos Chem Phys.* 2006;6:957–74. doi: 10.5194/acp-6-957-2006.
- [28] Morisette JT, Giglio L, Csiszar I, Setzer A, Schroeder W, Morton D, et al. Validation of MODIS active fire detection products derived from two algorithms. *Earth Interact.* 2005;9:1–25.
- [29] Giglio L. MODIS collection 5 active fire product user's guide version 2.5. College Park, MD: University of Maryland; 2013. p. 1–61.
- [30] He C, Gong YX, Zhang SY, He TF, Chen F, Sun Y, et al. Forest fire division by using MODIS data based on the temporal-spatial variation law. *Spectrosc Spectr Anal.* 2013;33(9):2472–7. doi: 10.3964/j.issn.1000-0593(2013)09-2472-06.
- [31] Boer MM, Sadler RJ, Wittkuhn RS, McCaw L, Grierson PF. Long-term impacts of prescribed burning on regional extent and incidence of wildfires—Evidence from 50 years of active fire management in SW Australian forests. *Forest Ecol Manag.* 2009;259(1):132–42. doi: 10.1016/j.foreco.2009.10.005.
- [32] Koutsias N, Balatsos P, Kalabokidis K. Fire occurrence zones: kernel density estimation of historical wildfire ignitions at the national level, Greece. *J Maps.* 2014;10(4):630–9. doi: 10.1080/17445647.2014.908750.
- [33] Isham V, Northrop P. Statistical analysis of spatial point patterns. Devon, England: Exeter EX4 4QJ; 2003.
- [34] Ripley BD. The second-order analysis of stationary point processes. *J Appl Probab.* 1976;13(2):255–66. doi: 10.2307/3212829.
- [35] Besag J. Efficiency of pseudolikelihood estimation for simple gaussian fields. *Biometrika.* 1977;64(3):616–8. doi: 10.2307/2345341.
- [36] Katkovnik V, Shmulevich I. Kernel density estimation with adaptive varying window size. *Pattern Recogn Lett.* 2002;23(14):1641–8. doi: 10.1016/s0167-8655(02)00127-7.
- [37] Seaman DE, Powell RA. An evaluation of the accuracy of kernel density estimators for home range analysis. *Ecology.* 1996;77(7):2075–85. doi: 10.2307/2265701.