ORIGINAL ARTICLE

Analysis of a spatial point pattern in relation to a reference point

Yukio Sadahiro[1] · Hidetaka Matsumoto[2]

© The Author(s) 2024

Abstract
This paper develops a new method for analyzing the relationship between a set of points and another single point, the latter of which we call a reference point. This relationship has been discussed in various academic fields, such as geography, criminology, and epidemiology. Analytical methods, however, have not yet been fully developed, which has motivated this paper. Our method reveals how the number of points varies by the distance from a reference point and by direction. It visualizes the spatial pattern of points in relation to a reference point, describes the point pattern using mathematical models, and statistically evaluates the difference between two sets of points. We applied the proposed method to analyze the spatial pattern of the climbers of Mt. Azuma, Japan. The result gave us useful and interesting findings, indicating the method's soundness.

Keywords Point pattern · Reference point · Directional variation · Distance · Visualization

JEL Classification C020 · C180 · C630

1 Introduction

This paper develops a new method for analyzing a set of points in relation to another single point, the latter of which we call a *reference point*. We aim to evaluate the spatial relationship between points and a reference point.

An example is the relationship between a retail store and its customer distribution, which has been discussed in retail geography and marketing science (Davies

✉ Yukio Sadahiro
sada@csis.u-tokyo.ac.jp

1 Interfaculty Initiative in Information Studies, The University of Tokyo, 7-3-1, Hongo, Bunkyo-Ku, Tokyo 113-8656, Japan

2 YAMAP Inc., 3-23-20, Hakataekimae, Hakata-ku, Fukuoka-shi, Fukuoka-ken 812-0011, Japan

🖄 Springer

2012; Scott 2017). Customers of convenience stores are tightly clustered around the stores, while those of shopping malls are widely spread. Epidemiology studies the relationship between air pollution and respiratory, cancers such as lung, bronchus, and larynx cancers (Filippini et al. 2019). Cancer cases are often clustered around air-polluting sites. Geographic profiling is an important topic in criminology (Kent et al. 2006; Trinidad et al. 2021). It analyzes the spatial relationship between crime locations and the offender's residence location.

The distance plays an important role in the above relationships. Customers generally decrease with the distance from retail stores. Cases of respiratory cancers often also decrease with the distance from air-polluting sites. Crime locations, on the other hand, exhibit different patterns. Offenders often avoid the neighborhood of their residence as crime locations, and consequently, crimes first increase and then decrease with the distance from their residences.

The direction is another important factor. Customers of retail stores are often more widely spread along highways due to easy accessibility. Air-polluting materials are conveyed by air currents, which leads to an anisotropic spatial pattern of cancer cases around air-polluting sites (Kurumatani and Kumagai 2008; Nakaya 2010). Directional variation is also found in crime locations since the spatial cognition of criminal offenders varies by direction (Kent and Leitner 2007; Frank et al. 2011; Mohler and Short 2012).

Studying the spatial relationship between points and a reference point helps us understand the properties of points and consider their underlying structures, i.e., how point patterns are formed. The relationship between retail stores and customer distributions tells us how the distance and direction affect the customers' store choice. Analysis of the relationship between cancer cases and air-polluting sites permits us to find air-polluting sites that are more likely to cause respiratory cancers, which may have to be closed or downsized (Brender et al. 2011; García-Pérez et al. 2015). Using past data to analyze the relationship between crime locations and offenders' residences permits us to detect offenders' unknown residences from their crime locations (Kent et al. 2006).

However, analytical methods of the spatial relationship between points and a reference point have not yet been fully established, as discussed in the next section. To fill the research gap, this paper proposes a new method for analyzing the spatial pattern of points in relation to a reference point. We aim to reveal how the number of points varies by the distance from a reference point and by direction. We consider the number of points in relation to the number of the potential locations of points, i.e., the locations where points can be located. Suppose the customers of a retail store in a region. All the residents in the region can be the store's customers; thus, their residence locations are represented as potential locations. Consideration of potential locations is important since their number affects the number of points. Potential locations are often referred to as inhomogeneous population in spatial statistics since the distribution of potential locations is generally inhomogeneous (Gatrell et al. 1996; Diggle 2013).

The rest of the paper is organized as follows. Section 2 reviews related works. Section 3 describes the proposed method in detail. Section 4 applies the method to analyze the spatial pattern of climbers of Mt. Azuma in Japan. Section 5 summarizes the conclusions with a discussion.

2 Related works

2.1 Analysis of a single set of points

Point pattern analysis has long been discussed extensively in geography, ecology, statistics, and other fields. The nearest-neighbor distance method is a basic but essential method to discuss clustered and dispersed point patterns (Clark and Evans 1954). The method evaluates the degree of point clustering within a statistical framework. The K-function and its standardized version, the L-function, consider point patterns with a scale parameter represented as the radius of circles randomly placed in a study region (Ripley 1976). These functions represent the degree of point clustering as a function of the geographic scale of analysis. The two-point correlation function is often used in astronomy (Peebles 1973, 1993). It aims to indicate the probability of finding an unknown galaxy as a function of the distance from known galaxies.

The above methods assume the complete spatial randomness (CSR) in the null hypothesis, i.e., point distribution follows a uniform distribution in a study region, which does not always hold in the real world. Cuzick and Edwards (1990) resolve this problem by a statistical method for evaluating point patterns where points can be located only at limited locations. Diggle and Chetwynd (1991) also discuss point clusters under inhomogeneous potential locations.

The above methods, unfortunately, do not meet our demand since they consider only a single set of points. We aim to discuss the relationship between a single set of points and another reference point, implying that we need to consider two different sets simultaneously.

2.2 Analysis of two sets of points

The nearest-neighbor spatial-association measure evaluates the spatial proximity between two sets of points (Lee 1979). The method provides the probability that the observed proximity occurs under the CSR. The cross K-function and L-function are more flexible and widely used in various academic fields (Ripley 1977). They can change the geographical scale of analysis to evaluate the spatial proximity and can handle cases where the location of one set of points is fixed under the null hypothesis. A drawback is that they also assume the CSR in the null hypothesis. The colocation quotient (CLQ) resolves this problem by introducing a randomization test (Leslie and Kronenfeld 2011). Cromley et al. (2014) generalizes the CLQ using the spatial weight function and proposes a local version of CLQ. Li et al. (2022) further extends the CLQs into the spatiotemporal dimension.

The above methods analyze the relationship between two sets of multiple points. They are effective for grasping the overall pattern of the relationship between points. Our interest, on the other hand, lies in the detailed relationship between a set of points and a reference point, which existing methods cannot

discuss. In addition, the above methods do not explicitly consider the directional variation in point patterns around reference points.

2.3 Analysis of the directional variation in a single set of points

Anisotropic *K*-function (Dale 2000; Rosser and Cheng 2019) considers the directional variation in point patterns. Extending the original *K*-function, it evaluates point clustering as a function of geographic scale and direction of analysis. The method, unfortunately, assumes a single set of points, and thus, it does not meet our objective.

Directional statistics is a useful tool for discussing the directional variation of spatial phenomena (Pewsey et al. 2013; Ley and Verdebout 2017). It has been widely used in biology, astronomy, climatology, etc. A drawback is that directional statistics does not consider the radial dimension, i.e., the distance from a reference point. Our study requires the consideration of both the directional and radial dimensions, which are not satisfied by directional statistics.

3 Method

3.1 Analysis of a single set of points 1: visualization

This subsection proposes a method for visualizing the relationship between a single set of points Ω and a reference point Z in the region Ξ. We first assume that the potential locations of points are uniformly distributed in Ξ. We describe our method under this situation and then proceed to the case where the potential locations of points are not uniformly distributed.

The method first divides the region Ξ into L sectors centered at Z, which are numbered clockwise from north $\{\Lambda_1, \Lambda_2, \ldots, \Lambda_L\}$ as shown in Fig. 1. The number of points in Λ_i is denoted by M_i. Let $\Omega_i = \{P_{i1}, P_{i2}, \ldots, P_{iMi}\}$ be the set of points in Λ_i arranged in increasing distance from Z. We use the polar coordinate system originating from Z to indicate the location of points. The location of P_{ij} is given by r_{ij} and θ_{ij}, the distance from Z and the angle measured clockwise from north, respectively. The following uses variables i and j to represent the ith sector and the jth point in each sector, respectively.

We then determine the area inside which the α percentage points around Z are contained in each sector. The number of points inside this location in sector Λ_i is given by

$$j = \frac{\alpha}{100} M_i, \tag{1}$$

where M_i is the number of points in Λ_i. If j is an integer, we take P_{ij} as the location inside which the α percentage points around Z are contained. If j is not an integer, we consider the two adjacent integers j_1 and j_2 between which j is contained. Using
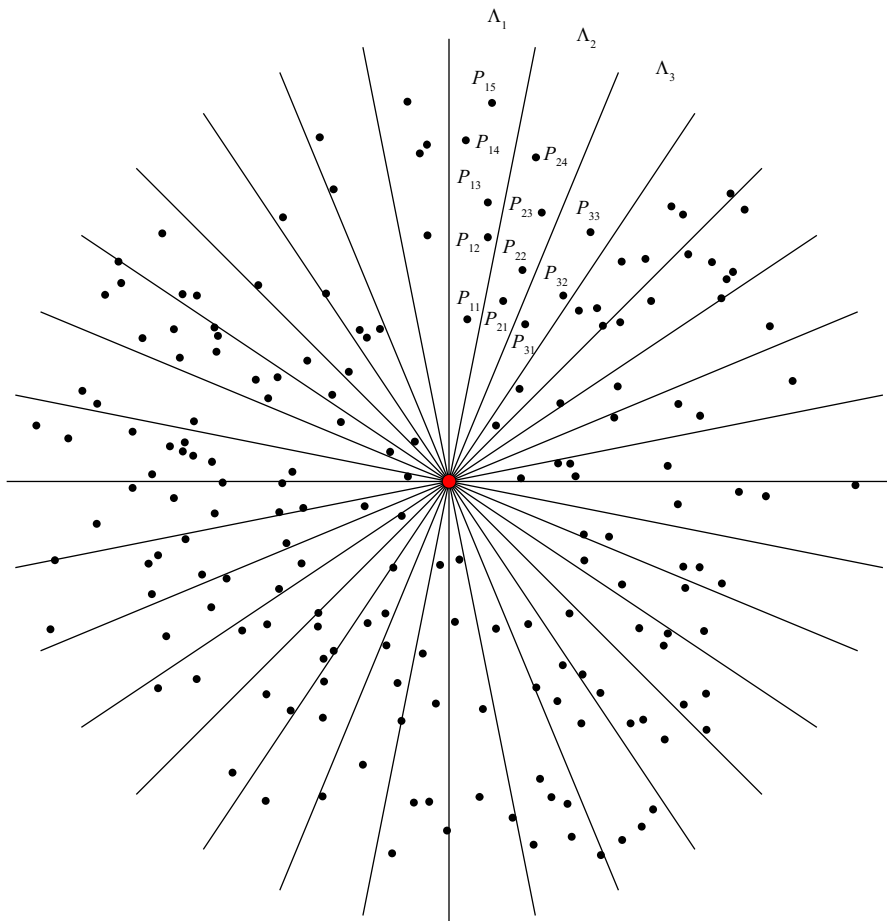
Fig. 1 Radial sectors centered at the reference point Z indicated as the red point. Labels of the points in Λ_1, Λ_2, and Λ_3 are indicated as examples (color figure online)

points P_{ij1} and P_{ij2}, we define U_i, the representative location of the α-covering area in sector Λ_i. Its radial and angular coordinates are given by

$$R_i(\alpha) = \frac{r_{ij_1} + r_{ij_2}}{2} \tag{2}$$

and

$$\theta_i = \frac{2\pi}{L}\left(i - \frac{1}{2}\right), \tag{3}$$

respectively, where L is the number of sectors. The red points in Fig. 2 indicate the representative locations of the points shown in Fig. 1, where $\alpha = 50$.
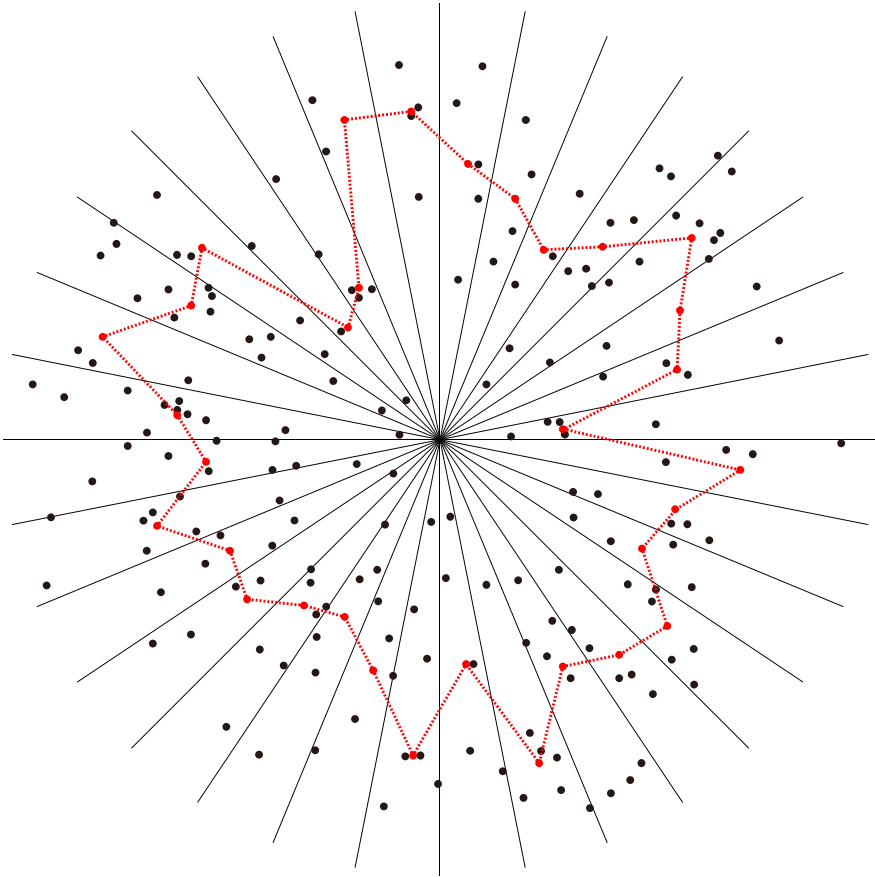
Fig. 2 Representative locations where $\alpha = 50$ (red points) and the 50-covering area (red dotted lines) shown in Fig. 1 (color figure online)

We finally connect the representative points from U_1 to U_L to generate a polygon containing the α percentage points. We call it the *α-covering area*. The red dotted line in Fig. 2 indicates the 50-covering area of points shown in Fig. 1. Unlike existing methods, this area explicitly assumes that the point pattern varies by direction, thus permitting us to grasp the anisotropic pattern of points in relation to reference point *Z*.

In the above procedure, we must determine the number of radial sectors *L*. A large *L* is desirable if there are enough points since it permits us to discuss the point pattern in detail. A concern is that the α-covering area can drastically fluctuate along the angular axis, preventing us from grasping the overall directional pattern of points. Spatial smoothing, such as the moving average and kernel smoothing along the angular axis resolves this problem (Silverman 1986; Scott 2015). Smoothing conceals complicated details and increases the interpretability of overall point patterns.

If there are not enough points, we have to decrease sectors. The minimum number of points necessary in each sector depends on the value of α. Let us assume $\alpha = 50$, where we aim to find the area containing half the points. We consider two cases where each sector contains two points and two hundred points, respectively. We obtain the same 50-covering area in both cases, but the result obtained in the latter is more reliable since it is based on more information. To evaluate the reliability, we introduce a statistical framework that is used for discussing the necessary sample size in binomial distribution (Lwanga et al. 1991; Desu 2012). We assume that the probability that a point is located inside the 50-covering area is 50 percent. In the first case, where there are two points in each sector, the probability that one point is located inside the 50-covering area and the other is outside the area is

$$\left(\frac{1}{2}\right)^2 {}_2C_1 = 0.5. \tag{4}$$

If we set the significance level to five percent, this situation is insignificant because it can happen by chance. In the second case, where there are two hundred points in each sector, the probability that half of the points are located inside the 50-covering area is

$$\left(\frac{1}{2}\right)^{200} {}_{200}C_{100} = 0.056. \tag{5}$$

Though it is still insignificant at the five percent level, the result is more reliable than the first case's. An increase of points decreases the probability and thus increases the reliability. Table 1 shows $M_{\min}(\alpha, s)$, the minimum number of points in each sector required at significance level s. As seen in the table, more points are necessary as α goes away from 50. When points are insufficient, we have to decrease sectors or use α value close to 50.

So far, we have assumed that the potential locations of points are uniformly distributed in the region Ξ. We then proceed to the cases where the potential locations of points are not uniformly distributed. Let us assume that the region Ξ consists of Y regions, in each of which the number of potential locations is reported. A typical example is the population data aggregated across spatial units such as

Table 1 The minimum number of points in each sector

α	5	10	15	20	25	30	35	40	45	50
$M_{\min}(\alpha, 0.05)$	1338	707	499	398	339	302	277	264	257	255
$M_{\min}(\alpha, 0.10)$	335	177	125	99	85	73	68	66	62	62
$M_{\min}(\alpha, 0.20)$	83	45	31	24	21	19	17	16	15	15
α	55	60	65	70	75	80	85	90	95	
$M_{\min}(\alpha, 0.05)$	257	264	277	302	339	398	499	707	1338	
$M_{\min}(\alpha, 0.10)$	62	66	68	73	85	99	125	177	335	
$M_{\min}(\alpha, 0.20)$	15	16	17	19	21	24	31	45	83	

census districts and zip code areas. The numbers of points and potential locations are denoted by $\{p_1, p_2, \ldots, p_Y\}$ and $\{q_1, q_2, \ldots, q_Y\}$, respectively. The average of the latter is

$$\overline{q} = \frac{\sum\limits_i q_i}{Y}. \tag{6}$$

We standardize the number of points in each region to eliminate the effect of the variation in the number of potential locations. The standardized number of points in the ith region is given by

$$p_i' = \frac{\overline{q}}{q_i} p_i. \tag{7}$$

This value works as the ratio of the number of points to the number of potential locations. If q_i is larger than its average, we reduce the number of points by randomly choosing p_i' points from the p_i points. If q_i is smaller than its average, we randomly locate p_i'-p_i points in the ith region. This eliminates the effect of the variation in the number of potential locations. After the standardization, we construct the α-covering area.

The computational complexity of the above processes is given as follows. We rearrange the points in increasing distance from Z in each sector. The calculation order of sorting algorithms such as heap sort and merge sort in the ith sector is

$$\mathrm{O}\big(M_i \log\big(M_i\big)\big) \tag{8}$$

(Tridgell 2005; Cormen et al. 2022). The total complexity is thus given by

$$\begin{aligned}
\sum_i \mathrm{O}\big(M_i \log\big(M_i\big)\big) &= L\mathrm{O}\Big(\frac{M}{L} \log\Big(\frac{M}{L}\Big)\Big) \\
&= \mathrm{O}\Big(M \log\Big(\frac{M}{L}\Big)\Big) \\
&= \mathrm{O}(M \log M),
\end{aligned} \tag{9}$$

where

$$M = \sum_i M_i. \tag{10}$$

Equation (9) assures that the computation ends in an acceptable time (Papadimitriou 2003; Arora and Barak 2009).

3.2 Analysis of a single set of points 2: point pattern modeling

Visualization is effective at an early stage of analysis. Visual analysis, however, is inevitably subjective, as mentioned in Sect. 2. This subsection proposes a more objective method to complement the visual analysis.

We introduce mathematical models to describe each sector's decreasing pattern of points. The models are applied after the standardization mentioned earlier. The negative power function is used in gravity models (Colwell 1982; Fotheringham and O'Kelly 1989):

$$f_i(r) = \frac{\beta_i}{r^{-\gamma_i}}, \tag{11}$$

where r represents the distance from Z. The parameters β_i and γ_i represent the number of points at the reference point and the decreasing speed of points with the distance from the reference point, respectively.

A drawback of this function is that it is undefined at $r=0$. The negative exponential model is often used instead that can be defined at $r=0$ (Haggett 1966; Haynes 1975).

$$f_i(r) = \beta_i \exp\left[-\gamma_i r\right]. \tag{12}$$

The logistic model is another option:

$$f_i(r) = 1 - \frac{\beta_i}{1 + \exp\left[-\gamma_i(r - \mu_i)\right]}, \tag{13}$$

Where μ_i represents the distance between the inflection point and the reference point.

Criminology often uses more complicated models for describing crime locations since crimes first increase and then decrease with the distance from the offender's residence. Normal and lognormal models describe the spatial pattern of crimes (Snook et al. 2005; Paulsen 2006; O'Leary 2009, 2011):

$$f_i(r) = \frac{\beta_i}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{r - \mu_i}{\gamma_i}\right)^2\right] \tag{14}$$

and

$$f_i(r) = \frac{\beta_i}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\log r - \mu_i}{\gamma_i}\right)^2\right]. \tag{15}$$

The functions are maximized at $r=\mu_i$ and $r=\exp[\mu_i]$ in Eqs. (14) and (15), respectively.

Choosing a mathematical model, we estimate the model from the observed data. We consider two cases that are different in the possible location of points. The first case is that points can be located anywhere in the region Ξ. In this case, the function $f_i(r)$ is the probability density function of points. We estimate the model by using the maximum likelihood method, where the likelihood is given by

$$L(f_i(r)) = \prod_j f_i(r_{ij}). \tag{16}$$

The second case is that points can be located at limited discrete locations in Ξ. This happens, for instance, when points represent retail stores that can be located

only at the sites in commercial areas. Sites are classified into two groups, i.e., inside and outside commercial areas. We denote the latter in sector Λ_i and their distances from Z as $\Omega'_i = \{P'_{i1}, P'_{i2}, \ldots, P'_{iMi}\}$ and $\{r'_{i1}, r'_{i2}, \ldots, r'_{imi}\}$, respectively. They represent the locations where points cannot be placed. We consider a stochastic model where the probability that a point is located at a given location is given by the function $f_i(r)$. The likelihood of this model is

$$L'\left(f_i(r)\right) = \prod_j f_i\left(r_{ij}\right) \prod_j \left\{ 1 - f_i\left(r'_{ij}\right) \right\}. \tag{17}$$

We estimate the model again based on the maximum likelihood method.

Having estimated models, we statistically compare them to choose the best model. If models have the same number of parameters, we can compare them by their likelihoods. If models contain a different number of parameters, information criteria such as AIC and BIC are effective for evaluating the model fitness (Kuha 2004; Chakrabarti and Ghosh 2011).

The choice of L depends on the analysis objective and the number of points. A detailed discussion of the spatial pattern of the models requires a large L, while the global pattern can be captured by a small L. Appropriate model estimation requires a certain number of points in each sector. The sample size problem has long been discussed in statistics (Cohen 2013; Kraemer 1974). The necessary sample size depends on the model structure, the number of parameters to be estimated, the sample distribution, and so forth. Therefore, discussing sample size problems in a general setting is difficult. In regression modeling, 100–200 samples are often said to be the minimum size for obtaining significant results (Comrey 1978; Green 1991; Tosteson et al. 2003). Since the structures of the models mentioned above are similar to those of simple regression models, it sounds desirable to determine L so that each sector contains at least 100–200 points.

The computational complexity of model estimation is given as follows. We estimate the models using a numerical optimization algorithm such as the EM algorithm. Each iterative process treats M/L points, whose calculation order is $O(M/L)$. We repeat the model estimation in L sectors, and thus, the total complexity of model estimation is

$$L \times O\left(\frac{M}{L}\right) = O(M). \tag{18}$$

3.3 Analysis of two sets of points

So far, we have discussed the relationship between a single set of points Ω and a reference point Z. This subsection compares the two sets of points that share the same reference point. Suppose another set of points Θ in the region Ξ. Let $\Theta_i = \{Q_{i1}, Q_{i2}, \ldots, Q_{iNi}\}$ be the set of points in sector Λ_i arranged in ascending order of the distance from Z. The distance of Q_{ij} from Z is denoted by s_{ij}.

Our interest lies in the difference in the spatial patterns of points around Z, or more precisely, whether points of Ω are more widely spread than those of Θ,

or vice versa. We first compare the two point sets in each sector. To this end, we adopt a statistical procedure that compares the two sets of distances $\{r_{i1}, r_{i2}, …, r_{iMi}\}$ and $\{s_{i1}, s_{i2}, …, s_{iNi}\}$. We combine and compare the two sets by rearranging the distances in ascending order. This is a typical two-sample problem where many statistical tests have been proposed, such as the Mann–Whitney and the Kolmogorov–Smirnov tests (Corder and Foreman 2014; Deshpande et al. 2017). The Mann–Whitney test is effective when two sets share an equal variance (Sokal and Rolf 1981; Siegel and John Jr 1988). The Kolmogorov–Smirnov test is applicable under unequal variances. A drawback of the Kolmogorov–Smirnov test is that it is less statistically powerful than the Mann–Whitney test due to its wide flexibility. Runs tests are also available for this two-sample problem (O'Brien 1976; O'Brien and Dyck 1985; Maritz 1995). Tests have been developed based on the number and the maximum run length.

The above statistics permit us to test the statistical difference between Ω_i and Θ_i in each sector. Let ω_i be the statistic calculated in sector Λ_i. Using ω_i, we evaluate the overall difference between Ω and Θ, where the null hypothesis is that there is no statistical difference between the two sets. Let ω_i be the statistic calculated in sector Λ_i. Its summation is given by

$$\omega = \sum_i \omega_i. \tag{19}$$

The static ω shows a large positive value if the points of Ω are more widely distributed around Z than those of Θ. If the points of Θ are more widely distributed, ω shows a small negative value. The static ω is close to zero if no significant difference exists between Ω and Θ.

When each sector contains enough points (at least 50, but 100 is desirable (Zhang and Wu 2002; Happ et al. 2019; Uttley 2019)), the probability distribution of ω_i under the null hypothesis independently follows the same distribution. The Mann–Whitney statistic, for instance, follows the standard normal distribution under the null hypothesis. The same distribution also approximates the Kolmogorov–Smirnov statistic.

If the number of radial sectors L is large enough, the central limit theorem allows us to easily calculate the probability distribution of ω under the null hypothesis. Let φ and δ^2 be the mean and variance of the probability density distribution of ω_i under the null hypothesis. We define the statistic ω_0 as

$$\omega_0 = \frac{\omega - L\varphi}{\sqrt{L}\delta}. \tag{20}$$

The central limit theorem assures that ω_0 follows the standard normal distribution under the null hypothesis, and thus we can evaluate the statistical significance of ω, i.e., the overall difference between Ω and Θ.

Even if L is not large enough, the probability density distribution of ω is approximated by a normal distribution when ω_i approximately follows a normal distribution under the null hypothesis. The Mann–Whitney statistic is a case where this requirement is satisfied. Some statistics used in runs tests also follow normal distributions.

The reproductive property of the normal distribution assures that ω follows a normal distribution whose mean and variance can be calculated.

The above discussion assumes that each sector contains enough points. If points are not enough, we calculate the probability distribution of ω under the null hypothesis by using the Monte Carlo simulation. We randomize the location of two sets of points in each sector, calculate ω_i, and sum up all the ω_i's. We repeat this process many times (generally 10,000 times) to obtain the probability distribution of ω.

The computational complexity of the above processes is given by the same discussion in Sect. 3.1, i.e.,

$$\sum_i \mathrm{O}\big((M_i + N_i)\log(M_i + N_i)\big) = \mathrm{O}((M + N)\log(M + N)), \tag{21}$$

where

$$N = \sum_i N_i. \tag{22}$$

Similar to Eq. (9), the calculation order shown in Eq. (21) is acceptable.

4 Application

This section applies the method proposed in the previous section to the analysis of the spatial pattern of climbers of Mt. Azuma in Fukushima prefecture, Japan. Mt. Azuma is a popular mountain that gathers many climbers from all over Japan. Its elevation is 2035 m, and a wide variety of climbers, from beginners to experts, enjoy hiking, trekking, and mountaineering at Mt. Azuma. We aim to analyze climbers' spatial and temporal patterns and reveal their underlying structure.

YAMAP Inc. developed a smartphone software called YAMAP that records the trajectories of climbers. Thirty percent of all the climbers in Japan have installed this software. YAMAP Inc. kindly provided us with the location data of YAMAP users and those of the climbers of Mt. Azuma from January 2019 to December 2021. Mt. Azuma and its climbers serve as a reference point and a set of points in the proposed method, respectively. The location data of YAMAP users work as the potential locations of points mentioned in Sect. 3.1, which are aggregated at the city level to conceal individual information.

Figure 3 shows the location of Mt. Azuma and its competing mountains, whose elevation is between 1500 and 2500 m. Competing mountains are primarily located in the southwest of Mt. Azuma, while they are sparse in the northern area. Navy shades indicate the number of Mt. Azuma climbs per YAMAP user from 2019 to 2021. The spatial pattern of navy shades reflects the number of competing mountains, i.e., the value is lower in the southwest and higher in the north.

We divided the whole region into $L = 96$ radial sectors centered at Mt. Azuma since this value was large enough to analyze the detailed variation in the spatial pattern of climbers, and each sector contains enough climbers for significant

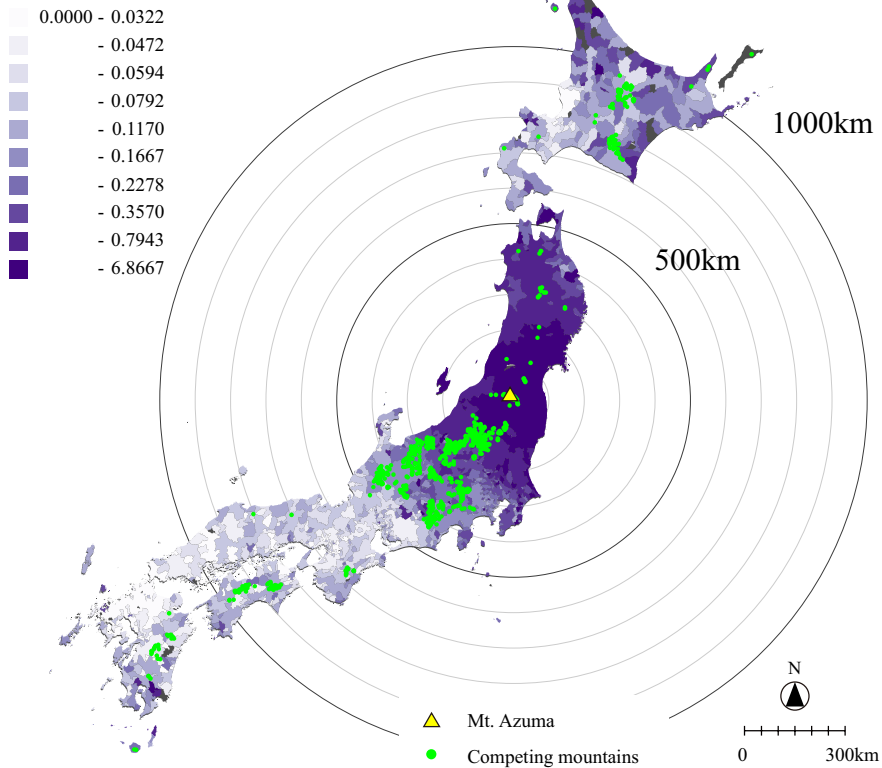Number of Mt. Azuma climbs per YAMAP user

0.0000 - 0.0322
       - 0.0472
       - 0.0594
       - 0.0792
       - 0.1170
       - 0.1667
       - 0.2278
       - 0.3570
       - 0.7943
       - 6.8667

1000km

500km

N

△ Mt. Azuma

● Competing mountains

0   300km

Fig. 3 Mt. Azuma and its competing mountains, whose elevation is between 1500 and 2500 m. Navy shades indicate the number of Mt. Azuma climbs per YAMAP user from 2019 to 2021. The data are aggregated at the city level (color figure online)

analysis. We performed the analysis using programs written in C++, which ran on an i9-9900U, CPU 3.60 GHz, RAM 32 GB computer.

4.1 Visualization and comparison of point patterns

The analysis started with visualizing the spatial pattern of the climbers of Mt. Azuma. We calculated the α-covering area of the climbers defined in Sect. 3.1. The moving average window of width 5 smoothed the areas. Figure 4 shows the 50- and 75-covering areas from 2019 to 2021. The areas widely spread from north to southwest, suggesting that Mt. Azuma attracted climbers from all over Japan. The noncircular shape of the areas indicates the directional variation in the spatial pattern of climbers. The areas were largest in 2019, before the outbreak of COVID-19. The areas drastically shrank in 2020 since climbers avoided traveling to distant places for mountaineering. The areas expanded in 2021, though they are smaller than in 2019.
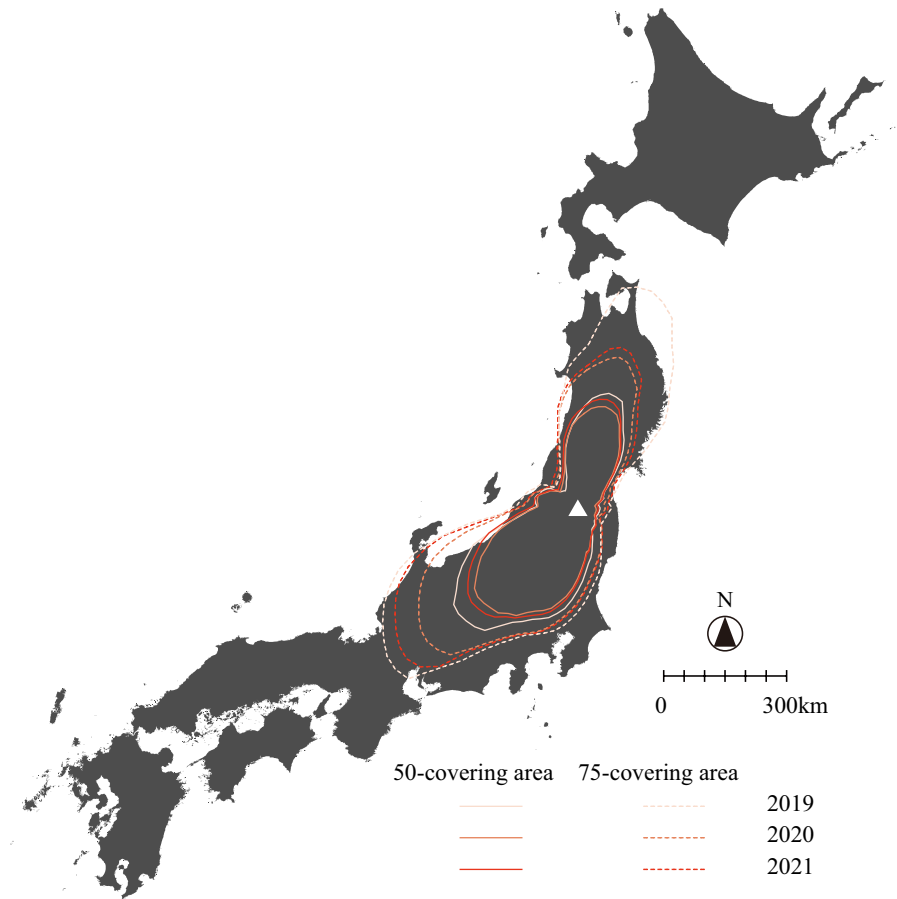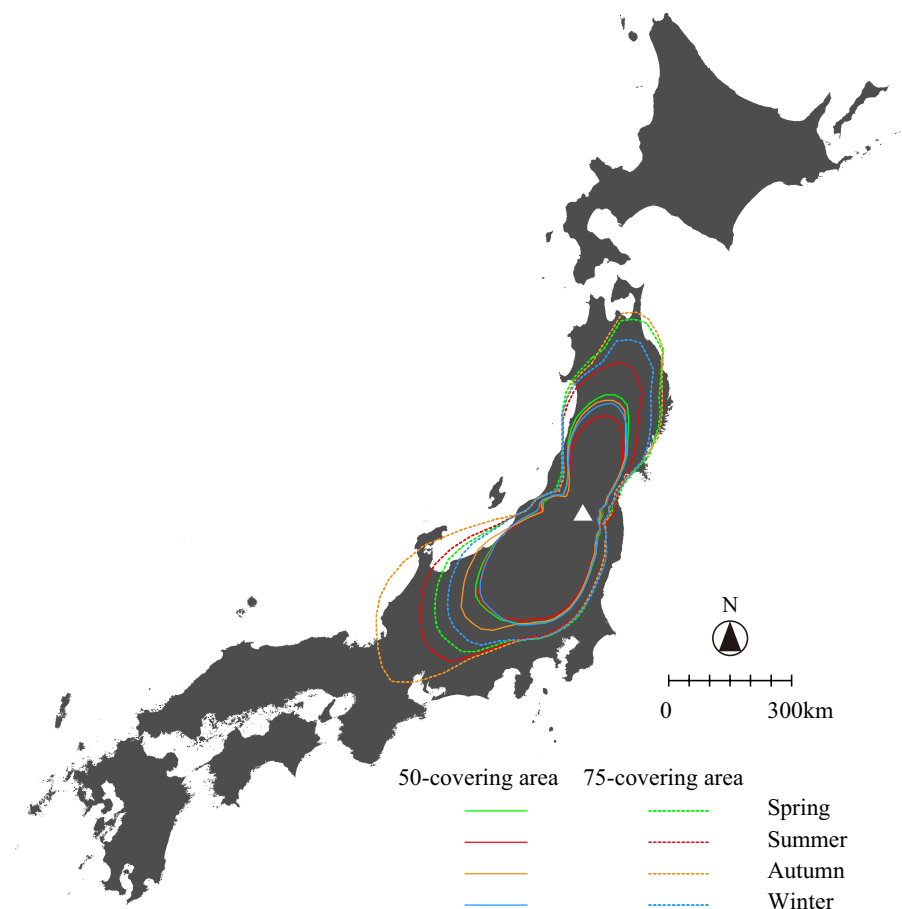
Fig. 4 The 50- and 75-covering areas of climbers of Mt. Azuma from 2019 to 2021. The white triangle indicates Mt. Azuma

Table 2 The statistic ω_0 of the Kolmogorov–Smirnov test that evaluates the difference in the patterns of climbers between different years

	2019	2020	2021
2019		2.016*	0.816
2020	− 2.016*		− 0.739
2021	− 0.816	0.739	

Positive values indicate that the climbers in the year shown in the left column are more widely spread than in the year shown in the top row. Asterisks indicate statistical significance at a five percent level

Table 2 shows the statistic ω_0 of the Kolmogorov–Smirnov test that compares the spatial pattern of climbers between different years. The absolute value of ω_0 is large if the patterns of two point sets differ, while it is small if the difference is insignificant. As seen in the table, the climbers in 2019 are more widely spread

than in 2020 at a significance level of five percent, while no significant difference was found between other pairs of years. This indicates that COVID-19 has drastically narrowed the catchment area of Mt. Azuma.

Figure 5 shows the seasonal variation in the 50- and 75-covering areas. We considered the four seasons: spring (March to May), summer (June to August), autumn (September to November), and winter (December to February). The figure indicates that the areas are generally the largest in autumn. This sounds reasonable since the autumn leaves of Mt. Azuma attract many climbers from wider areas. In contrast, the areas are relatively small in summer, especially in the northern area of Mt. Azuma. Summer is the best season for climbing higher mountains, requiring sophisticated skills and techniques that are not necessary for climbing Mt. Azuma. People prefer difficult mountains in summer, which shrank the covering areas of Mt. Azuma.

Fig. 5 The 50- and 75-covering areas of climbers of Mt. Azuma calculated for each season. The white triangle indicates Mt. Azuma

Table 3 The statistic ω_0 of the Kolmogorov–Smirnov test that evaluates the difference in the patterns of climbers between different seasons

	Spring	Summer	Autumn	Winter
Spring		0.939	− 1.887	1.412
Summer	− 0.939		− 2.351*	− 0.018
Autumn	1.887	2.351*		2.293*
Winter	− 1.412	0.018	− 2.293*	

Positive values indicate that the climbers in the season shown in the left column are more widely spread than in the season in the top row. Asterisks indicate statistical significance at a five percent level

Table 3 shows the statistic ω_0 that compares the spatial pattern of climbers between different seasons. Large positive values of autumn indicate that the climbers are more widely spread than in other seasons, especially summer and winter. No significant difference was observed between other pairs of seasons. Mt. Azuma gathers its climbers from the widest area in autumn, while the spatial pattern of climbers is not so different between other seasons. Autumn seems to be considered the best season for climbing Mt. Azuma.

4.2 Modelling point patterns

This subsection builds mathematical models to describe the spatial pattern of climbers of Mt. Azuma in detail. One of our primary interests lies in the directional variation in the distance deterrence of climbers. Figure 6 shows the 24 directions in which we estimated the models. Directions are numbered clockwise from the north. The radial lines connect Mt. Azuma and the farthest YAMAP users in individual directions.

We estimated the models represented by Eqs. (12) and (13) by using the likelihood defined by Eq. (17). Comparing the models using AIC, we found that the model represented by Eq. (13) is better in all directions. The following thus focuses on the result obtained based on Eq. (13).

The vertical axis in Fig. 7 indicates the probability density function $f_i(r)$. Gray lines correspond to the areas where YAMAP users do not exist. We classified the 24 directions by the slope of functions in Fig. 7: $G_1 = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, $G_2 = \{11, 12\}$, $G_3 = \{13, 14, 15, 16, 17, 18\}$, and $G_4 = \{19, 20, 21, 22, 23\}$. Figure 7b shows that the slopes of Group $G_1 = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ are the most gentle among the four groups. The slopes of Group $G_4 = \{19, 20, 21, 22, 23\}$ are also gentle but relatively steeper than those of Group G_1. One reason for this difference is the difference in train accessibility between these groups. Figure 8 shows the railway network around Mt. Azuma. We checked the timetable of the railway lines and found that the trains of Tohoku Shinkansen and Tohoku Lines indicated by thick white lines are faster and more frequent than those of other lines. This increases the accessibility of Group G_1 compared with that of Group G_4.

The distance where $f_i(r)$ begins to decrease varies among Group $G_4 = \{19, 20, 21, 22, 23\}$. They decrease from their maxima β to their half at r from 60 to 150 km. It is consistent with Figs. 4 and 5, where the 50-covering areas show
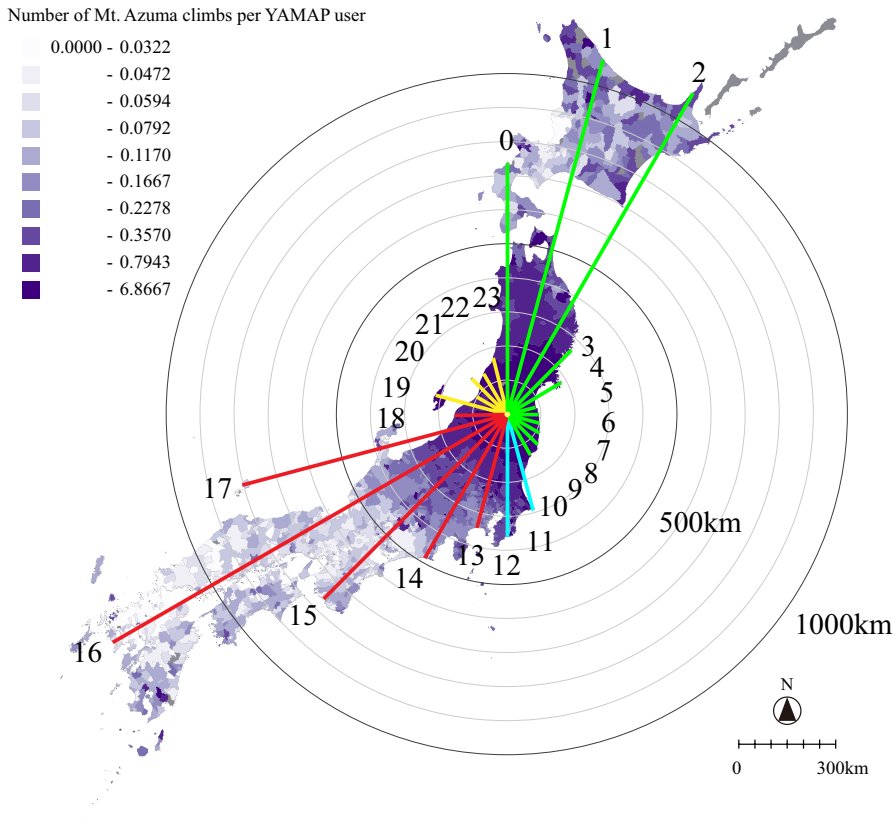
Fig. 6 The 24 directions in which the models were estimated. The radial lines connect Mt. Azuma and the farthest YAMAP users in individual directions. The numbers and colors of radial lines correspond to those in Fig. 7. Navy shades indicate the number of Mt. Azuma climbs per YAMAP user from 2019 to 2021 (color figure online)

clear anisotropic patterns. For instance, the climbers of direction 19 rapidly decrease with the distance from Mt. Azuma in Fig. 8, while the climbers of direction 23 gradually decrease. The traffic network is weak in these directions, and the variation of accessibility causes the variation in estimated models.

Groups $G_2 = \{11, 12\}$ and $G_3 = \{13, 14, 15, 16, 17, 18\}$ have many competing mountains in their directions. A difference in their patterns in Fig. 7 lies in that $f_i(r)$ begins to decrease at a farther distance in Group G_2. This is due to the train accessibility of the Tohoku Shinkansen and Tohoku Lines mentioned earlier. Easy accessibility attracts distant climbers of Group G_2. The slopes of Group G_3 are the steepest among the four groups. The probability remains maximum until 200 km in directions $\{14, 15, 16, 17\}$, and suddenly decreases to zero. This result is consistent with Fig. 8, where navy shades are dark at least within 200 km of Mt. Azuma. Competing mountains cause the steepest slopes in these directions.
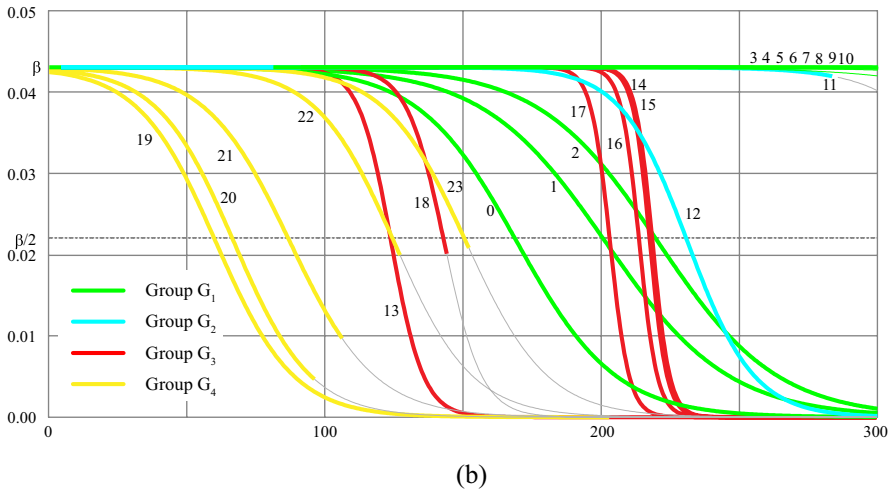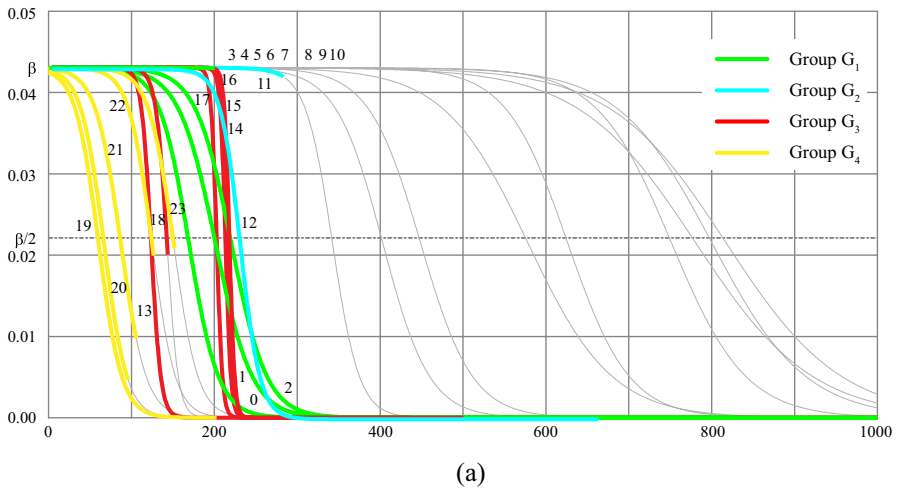
(a)

(b)

Fig. 7 The estimated logistic models defined by Eq. (13). The vertical axis indicates the probability density function $f_i(r)$. Gray lines correspond to the areas where YAMAP users do not exist. Graphs show the models between **a** [0, 1000] meters and **b** [0, 300] meters (color figure online)

5 Concluding discussion

This paper has developed a new method for analyzing the relationship between points and a reference point. We aim to reveal how the number of points varies by the distance from a reference point and by direction. The α-covering area visualizes the directional variation in the spatial pattern of points, while mathematical models statistically describe the relationship between points and a reference point. The statistics ω and ω_0 permit us to evaluate whether a set of points are more widely spread

Number of Mt. Azuma climbs per YAMAP user
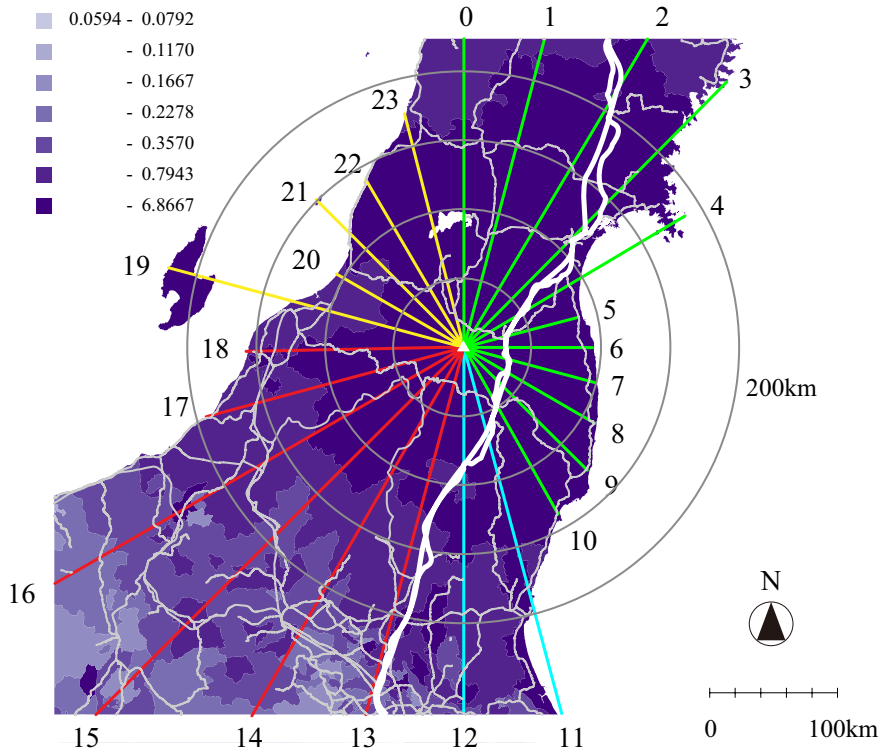
Fig. 8 Railway lines around Mt. Azuma. Thick white lines indicate Tohoku Shinkansen and Tohoku Lines, whose frequencies are relatively higher in this region. The numbers and colors of radial lines correspond to those in Fig. 7. Navy shades indicate the number of Mt. Azuma climbs per YAMAP user from 2019 to 2021 (color figure online)

around a reference point than another set of points. Section 4 applied the proposed method to analyze the spatial pattern of the climbers of Mt. Azuma, Japan. The results gave us useful and interesting empirical findings, which support the soundness of the method.

The strength of the proposed method is summarized as follows. Firstly, the proposed method is effective for analyzing the relationship between a set of points in relation to a reference point, which cannot be fully investigated by existing methods discussed in Sect. 2. Particularly the α-covering area is a useful tool for an intuitive understanding of the overall spatial pattern of points. Secondly, the method is flexible in that various models are applicable, as shown in Sect. 3.2. Thirdly, the method lets us compare the spatial spread of two sets of points within a statistical framework.

The method, however, has several limitations. We finally discuss its future extensions. Firstly, the method neglects the temporal dimension in point patterns. Empirical application discusses the difference between different periods. However, this

application involves data aggregation along the temporal dimension, reducing the original information's temporal resolution. An analytical method should be developed to consider the spatial and temporal dimensions without aggregation.

Secondly, we should introduce a statistical framework for evaluating the α-covering area. This permits us to evaluate the statistical significance of the obtained areas. Statistical framework, however, needs careful and extensive discussion. We need to consider null hypotheses, statistics, their probability distributions, and so forth. Though this is an important topic, we leave it for future research.

Thirdly, further applications are necessary to evaluate the effectiveness of the proposed method. An empirical study in this paper gave us new findings useful for understanding the factors determining the climbers' behavior, such as train accessibility and competing mountain locations. This, however, does not assure that the proposed method always works successfully in other academic fields. Applications in epidemiology, criminology, ecology, and other fields of spatial information science are necessary.

Fourthly, the definition of distance needs reconsideration. The proposed method evaluates the distance between points by the Euclidean distance. The Euclidean distance, however, is not always appropriate for describing the trip behavior of individuals (Fortney et al. 2000; Okabe and Sugihara 2012). If we introduce the time distance in Sect. 4, the directional variation observed in the distribution of climbers may be relaxed. It will also be better to include the waiting time for trains. Time, network, and mental distances are possible alternatives that can improve the model description.

Fifthly, the method neglects the attributes of points. Climbers of Mt. Azuma vary in their attributes, such as age, income, expertise level, etc. Older climbers may tend to choose nearby mountains than younger ones. One option is to classify climbers by age and compare their α-covering areas. This approach, however, depends on the classification scheme, which can be subjective, especially when treating numerical attributes. Further discussion is necessary to consider the point attributes in the analysis.

Data availability The data (the location data of YAMAP users and the climbers of Mt. Azuma) are unavailable due to commercial restrictions. However, the codes and sample datasets that support the present study's findings are available on Figshare at https://figshare.com/articles/dataset/_strong_Analysis_of_a_spatial_point_pattern_in_relation_to_a_reference_point_strong_Program_main_cpp_c_Data_sample_csv/23500326.

References

Arora S, Barak B (2009) Computational complexity: a modern approach. Cambridge University Press, Cambridge

Brender JD, Maantay JA, Chakraborty J (2011) Residential proximity to environmental hazards and adverse health outcomes. Am J Public Health 101:S37–S52

Chakrabarti A, Ghosh JK (2011) AIC, BIC and recent advances in model selection. Philos. Stat. 2011:583–605

Clark PJ, Evans FC (1954) Distance to nearest neighbor as a measure of spatial relationships in populations. Ecology 35:445–453

Cohen J (2013) Statistical power analysis for the behavioral sciences. Academic Press, Cambridge

Colwell PF (1982) Central place theory and the simple economic foundations of the gravity model. J Reg Sci 22:541–546

Comrey AL (1978) Common methodological problems in factor analytic studies. J Consult Clin Psychol 46:648

Corder GW, Foreman DI (2014) Nonparametric statistics: a step-by-step approach. John Wiley & Sons, Hoboken

Cormen TH, Leiserson CE, Rivest RL, Stein C (2022) Introduction to algorithms. MIT Press, Cambridge

Cromley RG, Hanink DM, Bentley GC (2014) Geographically weighted colocation quotients: specification and application. Prof Geogr 66:138–148

Cuzick J, Edwards R (1990) Spatial clustering for inhomogeneous populations. J Roy Stat Soc: Ser B (Methodol) 52(1):73–104

Dale MR (2000) Spatial pattern analysis in plant ecology. Cambridge University Press, Cambridge

Davies R (2012) Marketing geography. Routledge, London

Deshpande JV, Naik-Nimbalkar U, Dewan I (2017) Nonparametric statistics: theory and methods. World Scientific, Singapore

Desu M (2012) Sample size methodology. Elsevier, Amsterdam

Diggle PJ (2013) Statistical analysis of spatial and spatio-temporal point patterns. Chapman and Hall/CRC, Boca Raton

Diggle PJ, Chetwynd AG (1991) Second-order analysis of spatial clustering for inhomogeneous populations. Biometrics 47:1155–1163

Filippini T, Hatch EE, Rothman KJ, Heck JE, Park AS, Crippa A, Orsini N, Vinceti M (2019) Association between outdoor air pollution and childhood leukemia: a systematic review and dose–response meta-analysis. Environ Health Perspect 127:046002

Fortney J, Rost K, Warren J (2000) Comparing alternative methods of measuring geographic access to health services. Health Serv Outcomes Res Method 1:173–184

Fotheringham AS, O'Kelly ME (1989) Spatial interaction models: formulations and applications. Kluwer Academic Publishers, Dordrecht

Frank R, Andresen MA, Cheng C, Brantingham P (2011) Finding criminal attractors based on offenders' directionality of crimes. In: 2011 European intelligence and security informatics conference. IEEE, Athens, Greece

García-Pérez J, López-Abente G, Gómez-Barroso D, Morales-Piga A, Romaguera EP, Tamayo I, Fernández-Navarro P, Ramis R (2015) Childhood leukemia and residential proximity to industrial and urban sites. Environ Res 140:542–553

Gatrell AC, Bailey TC, Diggle PJ, Rowlingson BS (1996) Spatial point pattern analysis and its application in geographical epidemiology. Trans Inst Br Geogr 21:256–274

Green SB (1991) How many subjects does it take to do a regression analysis. Multivar Behav Res 26:499–510

Haggett P (1966) Locational analysis in human geography

Happ M, Bathke AC, Brunner E (2019) Optimal sample size planning for the Wilcoxon–Mann–Whitney test. Stat Med 38:363–375

Haynes RM (1975) Dimensional analysis: some applications in human geography. Geogr Anal 7:51–68

Kent J, Leitner M (2007) Efficacy of standard deviational ellipses in the application of criminal geographic profiling. J Investig Psychol Offender Profiling 4:147–165

Kent J, Leitner M, Curtis A (2006) Evaluating the usefulness of functional distance measures when calibrating journey-to-crime distance decay functions. Comput Environ Urban Syst 30:181–200

Kraemer HC (1974) The non-null distribution of the Spearman rank correlation coefficient. J Am Stat Assoc 69:114–117

Kuha J (2004) AIC and BIC: comparisons of assumptions and performance. Sociol Methods Res 33:188–229

Kurumatani N, Kumagai S (2008) Mapping the risk of mesothelioma due to neighborhood asbestos exposure. Am J Respir Crit Care Med 178:624–629

Lee Y (1979) A nearest-neighbor spatial-association measure for the analysis of firm interdependence. Environ Plan A 11:169–176

Leslie TF, Kronenfeld BJ (2011) The colocation quotient: a new measure of spatial association between categorical subsets of points. Geogr Anal 43:306–326

Ley C, Verdebout T (2017) Modern directional statistics. Chapman and Hall/CRC, Boca Raton

Li L, Cheng J, Bannister J, Mai X (2022) Geographically and temporally weighted co-location quotient: an analysis of spatiotemporal crime patterns in greater manchester. Int J Geogr Inf Sci 36:918–942

Lwanga SK, Lemeshow S, Organization WH (1991) Sample size determination in health studies: a practical manual. World Health Organization, Geneva

Maritz JS (1995) Distribution-free statistical methods. CRC Press, Boca Raton

Mohler GO, Short MB (2012) Geographic profiling from kinetic models of criminal behavior. SIAM J Appl Math 72:163–180

Nakaya T (2010) Exploring the geography of deaths from methotelioma in Japan using spatial data analysis and geovisualisation for spatial epidemiology. In: Proceedings of the association of Japanese geographers, p 148

O'Brien PC (1976) A test for randomness. Biometrics 32:391–401

O'Brien PC, Dyck PJ (1985) A runs test based on run lengths. Biometrics 41:237–244

O'Leary M (2009) The mathematics of geographic profiling. J Investig Psychol Offender Profiling 6:253–265

O'Leary M (2011) Modeling criminal distance decay. Cityscape 13:161–198

Okabe A, Sugihara K (2012) Spatial analysis along networks: statistical and computational methods. John Wiley & Sons, Chichester

Papadimitriou CH (2003) Computational complexity. In: Abrams RT (ed) Encyclopedia of computer science. Nova Science Pub Inc, New York, pp 260–265

Paulsen DJ (2006) Connecting the dots: assessing the accuracy of geographic profiling software. Policing Int J Police Strateg Manag 29:306–334

Peebles P (1973) Statistical analysis of catalogs of extragalactic objects. I. Theory. Astrophys J 185:413–440

Peebles PJE (1993) Principles of physical cosmology. Princeton University Press, Princeton

Pewsey A, Neuhäuser M, Ruxton GD (2013) Circular statistics in R. Oxford University Press, Oxford

Ripley BD (1976) The second-order analysis of stationary point processes. J Appl Probab 13:255–266

Ripley BD (1977) Modelling spatial patterns. J R Stat Soc Ser B (methodol) 39:172–212

Rosser G, Cheng T (2019) Improving the robustness and accuracy of crime prediction with the self-exciting point process through isotropic triggering. Appl Spat Anal Policy 12:5–25

Scott DW (2015) Multivariate density estimation: theory, practice, and visualization. John Wiley & Sons, New York

Scott P (2017) Geography and retailing. Routledge, London

Siegel S, John N Jr (1988) Nonparametric statistics for the behavioral sciences. McGrawHill, New York

Silverman BW (1986) Density estimation for statistics and data analysis. CRC Press, Boca Raton

Snook B, Zito M, Bennell C, Taylor PJ (2005) On the complexity and accuracy of geographic profiling strategies. J Quant Criminol 21:1–26

Sokal R, Rolf F (1981) Biometry, 2nd edn, chapter 9. WH Freeman, San Francisco

Tosteson TD, Buzas JS, Demidenko E, Karagas M (2003) Power and sample size calculations for generalized regression models with covariate measurement error. Stat Med 22:1069–1082

Tridgell A (2005) Efficient algorithms for sorting and synchronization

Trinidad A, Vozmediano L, Ocáriz E, San-Juan C (2021) "Taking a walk on the wild side": exploring residence-to-crime in juveniles. Crime Delinq 67:58–81

Uttley J (2019) Power analysis, sample size, and assessment of statistical assumptions—improving the evidential value of lighting research. Leukos 15:143–162

Zhang J, Wu Y (2002) Beta approximation to the distribution of Kolmogorov–Smirnov statistic. Ann Inst Stat Math 54:577–584

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