

Article

# Statistical Depth in Spatial Point Process

Xinyu Zhou  and Wei Wu \* 

Department of Statistics, Florida State University, Tallahassee, FL 32306, USA; xzhou3@fsu.edu

\* Correspondence: ww@stat.fsu.edu

**Abstract:** Statistical depth is widely used as a powerful tool to measure the center-outward rank of multivariate and functional data. Recent studies have introduced the notion of depth to the temporal point process, which exhibits randomness in the cardinality as well as distribution in the observed events. The proposed methods can well capture the rank of a point process in a given time interval, where a critical step is to measure the rank by using inter-arrival events. In this paper, we propose to extend the depth concept to multivariate spatial point process. In this case, the observed process is in a multi-dimensional location and there are no conventional inter-arrival events in the temporal process. We adopt the newly developed depth in metric space by defining two different metrics, namely the penalized metric and the smoothing metric, to fully explore the depth in the spatial point process. The mathematical properties and the large sample theory, as well as depth-based hypothesis testings, are thoroughly discussed. We then use several simulations to illustrate the effectiveness of the proposed depth method. Finally, we apply the new method in a real-world dataset and obtain desirable ranking performance.

**Keywords:** statistical depth; spatial point process; point process metric; hypothesis test; supervised classification

**MSC:** 62M30



**Citation:** Zhou, X.; Wu, W. Statistical Depth in Spatial Point Process. *Mathematics* **2024**, *12*, 595. <https://doi.org/10.3390/math12040595>

Academic Editor: Antonio Di Crescenzo

Received: 26 January 2024

Revised: 11 February 2024

Accepted: 14 February 2024

Published: 17 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Spatial point process is used to model and analyze patterns of a list of location points within a spatial domain; it has broad applications in various fields [1]. A lot of real-world data can be considered as realizations of spatial point processes such as spatial locations of an earthquake and its aftershocks, shooting positions of a basketball player in a single match, and car accident locations occurring in a city within a day. Common spatial models can be used to estimate the intensity function or the K-function of the point process [2], examine the nearest neighbors (NN) of any given point to build the NN distance distribution [3] and identify latent features [4,5] or investigate the cluster and inhibition phenomenon of point occurrence [6,7]. These methods mainly focus on representations and modelings of point patterns, but they have limited use in addressing statistical summaries and inferences in the space of the point process. For instance, given all shot positions of a basketball player, one can ask fundamental questions such as (1) “What is the typical or untypical shooting pattern of this player in a single match?” (2) “Does the shooting pattern show differences between the made and missed shots?” Statistical depth provides an ideal tool to answer those questions due to its ability to define a center-outward ranking for the shooting positions across all matches. In this paper, we aim to define the important notion of depth to the multi-dimensional spatial point process observations. For illustrative purposes, we only focus on two-dimensional spatial point process in a finite domain in this paper, whereas our approach can be naturally extended to higher dimensions. To our knowledge, this study is the first exploration to investigate the notion of statistical depth on the spatial point process.

Depth has been studied for decades to build a center-outward ranking for different types of data. Tukey [8] first introduced depth for multivariate data in a Euclidean space. From then on, a number of depth methods on multivariate data have been proposed, which include simplicial depth [9], Mahalanobis depth [10] and zonoid depth [11]. Over a decade ago, the research on depth started to focus on functional observations. López-Pintado and Romo [12] proposed the concept of functional depth for the first time. Nieto-Reyes [13] thoroughly examined the mathematical properties of functional depth. In recent years, depth was introduced to the data in a more complicated non-linear metric space. Dai et al. [14] extended the traditional Tukey's depth to Riemannian and general metric space, and Geenens et al. [15] introduced a new depth in any metric space with a more efficient computation. In addition, a lot of progress has been made in ranking observations from the temporal point process space. The generalized Mahalanobis depth [16] was the first depth method defined on the temporal point process. In Qi et al. [17], Dirichlet depth was proposed to overcome the boundary issue, and in Xu et al. [18] a smoothing approach was adopted to define depth using a functional depth on the smoothed process. In Zhou et al. [19], ILR depth was developed via the classical Isometric Log-Ratio (ILR) transformation to address the non-Euclidean issues in the point process space.

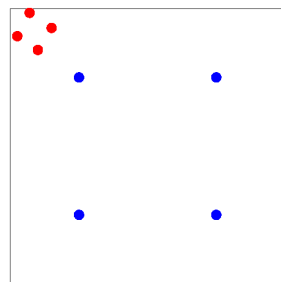
Despite the progress of statistical depth for temporal point process, the investigations on depth for the spatial point process are still under-explored. For the temporal point process, the depth on point locations can be defined based on the equivalent inter-event intervals [16,17,19]. However, this approach is not applicable in the spatial case due to the lack of point order and notion of inter-event. Therefore, in this paper, we consider a different approach and directly study the interaction among the entire spatial point process in the space. To achieve this goal, we propose two different proper metrics (namely the penalized metric and the smoothing metric) to measure the distance between any point processes. Then, we adopt a newly developed metric-based method [15] to define the depth on the spatial point process.

One significant advantage of this new method is that it is model-free. When computing the depth value, it is not necessary to first characterize the intensity function, whose estimation procedure is often demanding. In this case, the depth is only dependent on a metric between processes, obtained by point cardinality and point distribution directly. Another advantage is that by using the smoothing metric, the new depth exploits the cardinality and distribution under one framework, and a proper center-outward rank for a set of spatial data is naturally provided. This is in contrast to previous studies [17,19], where cardinality and distribution are combined in a weighted form and the weight coefficient may vary with respect to data.

We emphasize that the center-outward rank or importance of spatial point observations cannot be formulated by conventional likelihood methods [20]. We briefly illustrate this fact with the following example: We let  $x$  be a realization of homogeneous Poisson process on  $[0, 1]^2$ . Then, its likelihood is given as  $f(x) = \alpha\beta^{n(x)}$ , where  $\alpha, \beta > 0$  are two parameters and  $n(x)$  counts the cardinality in  $x$ . In this case, any realization with the same cardinality shares the same likelihood value regardless of point locations. Even though a process with uniformly distributed points should be considered as a typical, or important, example, it is not straightforward to quantify such importance with the likelihood function. A toy example is shown in Figure 1 to illustrate the comparison between the typical and potentially outlier pattern of a homogeneous Poisson process. Since the points are expected to be uniformly distributed within the domain, the realization in blue is naturally considered more typical compared to the one in red.

Once the center-outward rank is well defined, the corresponding depth-based analyses and inferences can be directly utilized. First, using the depth value, it is straightforward to check the typical and outlier patterns of the spatial data. This technique is useful for anomaly detection of spatial data. Next, depth has been widely adopted in conducting hypothesis testing for data samples. In this paper, we introduce a depth-based test approach to compare the distributions between spatial data groups. In addition, we generalize

the multivariate-based Depth–Depth classifier [21] to spatial point process to conduct supervised classification on simulations and real data.



**Figure 1.** Comparison of typical and outlier homogeneous Poisson processes with Cardinality 4. Blue and red dots represent the typical and outlier processes, respectively.

The rest of this paper is organized as follows. In Section 2, we provide a detailed construction of two metrics for the spatial point process. Then, a formal definition of spatial metric depth is given and the corresponding mathematical properties are examined. Furthermore, a depth-based hypothesis testing method is introduced to compare the distributions of two point process groups. In Section 3, simulation studies are conducted to illustrate the effectiveness of the newly developed depth. In Section 4, we adopt a real dataset to demonstrate the spatial metric depth in capturing typical patterns. Finally, the summary and future work are described in Section 5. All mathematical proofs are shown in Appendices A–E.

## 2. Methodology

In this section, we first introduce two proper metrics to measure the spatial point process distance. Then, we formally define a depth for the spatial point process. To make the new methods practically useful, we focus on observations from a simple, finite point process.

### 2.1. Penalized Metric

A realization of spatial point process can be viewed as a set of finite, non-overlapping points in a fixed domain. For simplicity, the domain is specified as  $[0, 1]^2$  in this paper. To measure the dissimilarity between two different sets, we first adopt the renowned Hausdorff distance to address the problem. We let  $s$  and  $t$  be two finite point processes in  $[0, 1]^2$ ; the Hausdorff metric  $d_H(s, t)$  between  $s$  and  $t$  is given as

$$d_H(s, t) = \max \left\{ \max_{s_i \in s} d_e(s_i, t), \max_{t_i \in t} d_e(t_i, s) \right\},$$

where  $d_e(s_i, t)$  measures the Euclidean distance between point  $s_i$  and the closest point in  $t$ , and similarly for  $d_e(t_i, s)$ . However, although the Hausdorff metric can capture the spatial point distances, it ignores the cardinalities of the processes. That is, as long as the point locations of the two sets are close to each other, their Hausdorff distance will be small. To compare events from two spatial point processes, we should compare not only their distributions but also their cardinalities, namely the numbers of events in both processes.

To overcome this problem, we introduce a penalized metric to take into account the importance on cardinality. The formal definition is given as follows:

**Definition 1.** Let  $s = \{s_1, s_2, \dots, s_m\}$  and  $t = \{t_1, t_2, \dots, t_n\}$  be two spatial point processes in domain  $[0, 1]^2$  with cardinalities of  $m$  and  $n$ , respectively. Then, the penalized metric  $d_{PH}(s, t)$  between  $s$  and  $t$  is defined as

$$d_{PH}(s, t) = \max \left\{ \sup_{s_i \in s} d_e(s_i, t), \sup_{t_i \in t} d_e(t_i, s) \right\} + \lambda |m - n|,$$

where  $d_e(s_i, \mathbf{t})$  measures the Euclidean distance between  $s_i$  and the closest point event in  $\mathbf{t}$ , and similarly for  $d_e(\mathbf{t}_i, \mathbf{s})$ .  $\lambda \geq 0$  is a hyper-parameter.

**Remark 1.** Compared with the conventional Hausdorff metric, the penalized version in Definition 1 includes a penalty term to emphasize the cardinality difference between two spatial point processes. Hyper-parameter  $\lambda$  controls the penalty effect. We emphasize that the computational cost of penalized distance is  $O(mn)$ . This cost is independent of the shape of the domain (i.e., same cost for any domain other than  $[0, 1]^2$ ).

It is straightforward to verify that the penalized metric is indeed a proper metric. That is, it satisfies positiveness, symmetry, and triangle inequality. The formal proof is provided in Appendix A. Therefore, the penalized metric provides an appropriate criterion to measure the spatial point process distance and can be further used to define the notion of depth.

Although the penalized metric provides a proper distance measure between spatial point processes with efficient computation, there are still two apparent drawbacks that may affect its performance. (1) In Definition 1, the distribution of spatial points and their cardinality are considered separately and their contribution to the distance is balanced by hyper-parameter  $\lambda$ . Hence, an appropriate value of  $\lambda$  has to be precisely determined in practical use. (2) The penalized metric is sensitive to extreme outliers. A single distant outlier may dominate metric measurement. In the next subsection, we introduce an alternative approach to define the metric between spatial point processes to overcome these issues.

## 2.2. Smoothing Metric

In this subsection, we introduce a new metric for the spatial point process which is less sensitive to outliers. Furthermore, the point distribution and cardinality of the point process are integrated under one framework.

### 2.2.1. Mapping between Spatial Point Process and Bivariate Function

For each spatial point process, we first transform it to a multivariate function by a smoothing kernel, and then adopt the  $\mathbb{L}_2$  functional metric to define the distance. In this paper, the transformed function is called the smoothed function or the smoothed point process. Xu et al. [18] first adopted a Gaussian kernel function for the temporal point process within a given domain, and then used the conventional  $\mathbb{L}_2$  metric on the smoothed functions. However, this distance involves numerical integration as there is no closed-form expression available in general. The computational cost is manageable for a one-dimensional temporal domain, whereas the cost can be highly demanding for multivariate spatial point processes due to the curse of dimensionality.

To address this problem, we adopt a different approach when transforming a spatial point process to a bivariate function. Given finite domain  $[0, 1]^2$ , we first adopt the inverse of the sigmoid function to bijectively transform the point processes from the finite domain to  $\mathbb{R}^2$ . In this case, given any spatial point process  $\mathbf{s} = \{s_1, s_2, \dots, s_n\}$ , where  $s_i = (x_i, y_i)$  for  $i = 1, \dots, n$ , the transformed point process  $\mathbf{s}^* = \{s_1^*, s_2^*, \dots, s_n^*\}$  is in  $\mathbb{R}^2$ , where  $s_i^* = (x_i^*, y_i^*) = \left( \log\left(\frac{x_i}{1-x_i}\right), \log\left(\frac{y_i}{1-y_i}\right) \right)$  for  $i = 1, \dots, n$ . Here, we ignore the points on the boundary lines by assuming that the realization points are within  $(0, 1)^2$  almost surely (this is true for commonly used point processes). Next, a proper kernel function needs to be defined on the transformed point process in the infinite domain. We propose to adopt the conventional Gaussian kernel function. Using the same notation as above, the Gaussian kernel function is applied on each point event of the transformed process  $\mathbf{s}^*$  as follows:

$$K(s_i^*) = K((x_i^*, y_i^*)) = c_1 e^{-c_2(x_i^{*2} + y_i^{*2})}, \quad i = 1, \dots, n \quad (1)$$

where  $c_1$  and  $c_2$  are two positive hyper-parameters that control the kernel scale and width, respectively.

Next, we introduce the mapping between the spatial point process and a bivariate function via the Gaussian kernel function in Equation (1). First, we denote  $S_k$  as the space of the spatial point process with cardinality  $k$  in domain  $[0, 1]^2$ , that is,  $S_k = \{s = \{s_1, s_2, \dots, s_k\} = \{(x_1, y_1), \dots, (x_k, y_k)\} \mid (x_i, y_i) \in [0, 1]^2, i = 1, \dots, k\}$ . Therefore,  $S = \cup_{k=0}^{\infty} S_k$  is the space of all spatial point processes. We note that if cardinality  $k = 0$ , there is no event in the certain point process. Next, we denote  $S_k^*$  as the space of the transformed point processes with cardinality  $k$  by the inverse of the Sigmoid function, which is  $S_k^* = \{s^* = \{s_1^*, \dots, s_k^*\} = \{(x_1^*, y_1^*), \dots, (x_k^*, y_k^*)\} \mid (x_i^*, y_i^*) \in \mathbb{R}^2, i = 1, \dots, k\}$ . Then,  $S^* = \cup_{k=0}^{\infty} S_k^*$  is the space of all transformed processes.

Based on the kernel function, the smoothing function can be formally introduced in the following definition:

**Definition 2.** For any spatial point process  $s$  in domain  $[0, 1]^2$  with cardinality  $k$ , we denote  $s^*$  as its transformed process in an infinite domain by the inverse of the Sigmoid function. Smoothing function  $f_s : \mathbb{R}^2 \rightarrow \mathbb{R}$  is given in the following form:

$$f_s(x, y) = \sum_{i=1}^k K((x, y) - s_i^*),$$

where  $K(\cdot)$  is the Gaussian kernel function in Equation (1).

In the remaining part of this paper, we call  $f_s$  the smoothed process of  $s$ . The space of the smoothed processes with cardinality  $k$  can be defined as  $F_k = \{f_s : \mathbb{R}^2 \rightarrow \mathbb{R} \mid s \in S_k\}$ . Thus,  $F = \cup_{k=0}^{\infty} F_k$  is the space of the smoothed process with any cardinality. We note that if  $k = 0$ , then  $f_s(x, y) = 0$  is a constant function in  $F_0$ . Similarly to the result on the temporal point process in [18], we can show that the mapping from the spatial point process space  $S$  to the space of the smoothed process  $F$  is a bijection. Mathematical details are given in Appendix B.

### 2.2.2. Definition of Smoothing Metric

In this subsection, we define the smoothing metric for the spatial point process. We propose to adopt the conventional  $\mathbb{L}^2$  distance and directly apply it on the smoothed processes. The definition is given as follows.

**Definition 3.** Given two spatial point processes  $s$  and  $t$  in domain  $[0, 1]^2$ , we denote the smoothed processes of  $s$  and  $t$  as  $f_s$  and  $f_t$ , respectively. The smoothing distance between  $s$  and  $t$  is given in the following form:

$$d_F(s, t) = \|f_s - f_t\|_2,$$

where  $\|\cdot\|_2$  denotes the conventional  $\mathbb{L}^2$  distance.

With the Gaussian kernel in Equation (1), the distance can be given in a closed form in the following proposition, where the mathematical proof is shown in Appendix C.

**Proposition 1.** For point processes  $s = \{(x_1, y_1), \dots, (x_m, y_m)\}$  and  $t = \{(u_1, v_1), \dots, (u_n, v_n)\}$  in domain  $[0, 1]^2$ , we denote the transformed processes as  $s^* = \{(x_1^*, y_1^*), \dots, (x_m^*, y_m^*)\}$  and  $t^* = \{(u_1^*, v_1^*), \dots, (u_n^*, v_n^*)\}$ , where  $m$  and  $n$  are the cardinalities of  $s$  and  $t$ , respectively. The smoothing distance in Definition 3 can be given in the following closed form:

$$d_F(\mathbf{s}, \mathbf{t}) = \left( \frac{c_1^2 \pi}{2c_2} \cdot \left( \sum_{i=1}^m \sum_{j=1}^m e^{-\frac{c_2}{2} [(x_i^* - x_j^*)^2 + (y_i^* - y_j^*)^2]} + \sum_{i=1}^n \sum_{j=1}^n e^{-\frac{c_2}{2} [(u_i^* - u_j^*)^2 + (v_i^* - v_j^*)^2]} - 2 \sum_{i=1}^m \sum_{j=1}^n e^{-\frac{c_2}{2} [(x_i^* - u_j^*)^2 + (y_i^* - v_j^*)^2]} \right) \right)^{\frac{1}{2}},$$

where  $c_1$  and  $c_2$  are two hyper-parameters in Equation (1).

**Remark 2.** The closed-form metric in Proposition 1 is a significant advantage in terms of computational efficiency in a spatial domain. Compared with numerical integration, computational cost is reduced from  $O(N^2mn)$  to  $O(mn)$ , where  $N$  is the grid size for numerical estimation. Hyper-parameter  $c_1$  controls the overall magnitude and has the same impact on any process.  $c_2$  plays a more important role in each individual process. If  $c_2$  is too large or small, then the point locations become less influential when determining distance. It makes the point locations more meaningful when  $c_2$  takes an appropriate value. Optimization approaches such as a cross-validation may be applied to find suitable values for  $c_2$  in practical use.

Based on the bijective mapping between point process and its smoothed function, the smoothing metric is a proper metric and can be directly applied to conduct the depth method. Compared with the penalized metric, this option provides a more robust metric measurement. However, one disadvantage is that it has a higher requirement for the shape of the domain. When the domain is a general rectangle,  $[a, b] \times [c, d]$ , it is still convenient to conduct domain transformation. If the domain is not rectangle-shaped, then there is no straightforward transformation to expand the bounded domain to  $\mathbb{R}^2$ . In this case, a numerical method with grids has to be implemented to approximate the functional integral, which significantly increases computational cost.

### 2.3. Spatial Metric Depth for Spatial Point Process

In this section, we introduce the definition of metric-based depth for the spatial point process and study its mathematical properties. Geenens et al. [15] introduced the notion of depth for any abstract metric space. This depth can be applied to measure the center-outward rank of any object sample as long as there is a proper metric for the object space. Therefore, with two proper metrics for the spatial point process, we are able to formally define the metric-based depth for the spatial point process as follows.

**Definition 4.** We denote  $(S, d)$  as the metric space for all spatial point processes in domain  $[0, 1]^2$  with respect to probability measure  $P \in \mathcal{P}$ , where  $d$  is the metric for the point process and  $\mathcal{P}$  is the space of all probability measures for the spatial point process in  $[0, 1]^2$ . Given any  $\mathbf{s} \in S$ , the spatial metric depth of  $\mathbf{s}$  with respect to  $(S, d)$  is defined as

$$D(\mathbf{s}) = D(\mathbf{s}, P) = P\left(d(\mathbf{p}, \mathbf{q}) > \max \{d(\mathbf{s}, \mathbf{p}), d(\mathbf{s}, \mathbf{q})\}\right). \tag{2}$$

By definition, the depth value of each process varies with respect to the selected metric. This provides more flexibility to build the center-outward ranking. Since different metrics focus on distinct aspects of the process, one can create the ranking framework by adopting the most appropriate metric based on specific goals. In this paper, we adopt both the penalized metric and the smoothing metric to evaluate the depth value and compare their performances.

To further explore the depth framework, we first examine the depth mathematical properties of spatial metric depth. Details are given below, in Proposition 2. Based on the result in the general metric space Geenens et al. [15], we present four mathematical properties specifically for the point process. A more detailed interpretation of the properties is given in Appendix D.

**Proposition 2.** *The spatial metric depth in Definition 4 satisfies the following properties:*

- (P<sub>1</sub>) *Linear invariance:* For point processes in any general rectangular domain, Definition 4 can still be adopted to define the depth value via the two metrics. We let  $\mathbf{s} = \{s_1, s_2, \dots, s_n\} = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$  be an arbitrary point process in  $[0, 1]^2$  without overlapping points. We suppose  $a$  and  $c$  are any positive numbers, and  $b$  and  $d$  are any real numbers. We denote  $U = \begin{pmatrix} a & 0 \\ 0 & c \end{pmatrix}$ ,  $w = [b, d]^T$  and  $\mathbf{s}' = \{Us_i + w\}_{i=1}^n$  in domain  $[b, a + b] \times [d, c + d]$ . Then,  $D(\mathbf{s}, P) = D(\mathbf{s}', UP + w)$ .
- (P<sub>2</sub>) *Vanishing at infinity:* For any  $\mathbf{s} \in (S, d)$ , depth value  $D(\mathbf{s}) \rightarrow 0$  if the cardinality of  $\mathbf{s}$  rises to infinity.
- (P<sub>3</sub>) *Continuity in  $\mathbf{s}$ :* For any  $P \in \mathcal{P}$ ,  $\mathbf{s} \in (S, d)$  and  $\epsilon > 0$ , there exists  $\delta > 0$  such that  $\sup_{\mathbf{t}: d(\mathbf{s}, \mathbf{t}) < \delta} |D(\mathbf{t}, P) - D(\mathbf{s}, P)| < \epsilon$ .
- (P<sub>4</sub>) *Continuity in  $P$ :* For any  $P \in \mathcal{P}$ ,  $\mathbf{s} \in (S, d)$  and  $\epsilon > 0$ , there exists  $\delta > 0$  such that  $|D(\mathbf{s}, P) - D(\mathbf{s}, Q)| < \epsilon$ —almost surely for all  $Q \in \mathcal{P}$  with  $d_{\mathcal{P}}(P, Q) < \delta$ —almost surely, where  $d_{\mathcal{P}}$  measures the topology of weak convergence on  $\mathcal{P}$ .

The four properties given above have clear correspondents in the multivariate depth [22].  $P_1$  corresponds to “affine invariance”, which illustrates that the depth value should be invariant under linear transformation.  $P_2$  corresponds to “vanishing at infinity”. In this paper, the point process approaches infinity under two conditions: (1) cardinality tends to infinity; (2) point locations move close to the domain boundary. In Appendix D, we show that the depth value becomes 0 only when the first condition holds.  $P_3$  and  $P_4$  correspond to the “continuity” property. Since the metric is designed and there exists proper probability measure on point processes [1], these two properties are naturally established.

Moreover, as illustrated in Geenens et al. [15], the important depth properties “maximality at center” and “monotonicity relative to the deepest point” may not hold for metric depth. In general, the definition of symmetry is unclear if there is no concrete assumption on the space structure. In the case of the spatial point process, the randomness exists in both cardinality and location. It is not straightforward to find a general “center” for processes. However, if cardinality is given, the notion of symmetry in multivariate data may be further explored to define a “conditional center”.

Before spatial metric depth can be adopted in practice, large sample theory should be discussed. Definition 4 is given for population depth. In most cases, the population probability measure of the spatial point process sample is unknown, and an empirical probability measure is used to substitute the population one. We suppose there is an independent sample of point processes  $\{s_i; i = 1, \dots, n\}$ . We denote  $\hat{P}_n$  as the empirical probability measure corresponding to these observations. Then,  $\hat{P}_n$  is the collection of  $\frac{1}{n}$  weighted point masses at  $s_1, \dots, s_n$ . Thus, given any point process  $\mathbf{s}$ , the empirical depth is given as

$$D(\mathbf{s}, \hat{P}_n) = \frac{1}{\binom{n}{2}} \sum_{i < j} I\{d(\mathbf{s}_i, \mathbf{s}_j) > \max\{d(\mathbf{s}, \mathbf{s}_i), d(\mathbf{s}, \mathbf{s}_j)\}\}. \tag{3}$$

For each process  $\mathbf{s}$ , we have its population depth in Equation (2) and sample depth in Equation (3). One natural and important question is on consistency—does the sample depth converge to the population one in a large sample? Our answer to this question is a “Yes” and this result is summarized in the following proposition, where the proof is shown in Appendix E.

**Proposition 3.** *Sample depth  $D(\mathbf{s}, \hat{P}_n)$  in Equation (3) converges to population depth  $D(\mathbf{s}, P)$  in Equation (2) almost surely. That is,*

$$\lim_{n \rightarrow \infty} D(\mathbf{s}, \hat{P}_n) = D(\mathbf{s}, P) \quad a.s.$$

When we illustrate the proposed depth using simulations and real data in the next section, we adopt the empirical version to compute the depth value. It is worth mentioning that the metric depth built by Geenens et al. [15] is not the only metric-based depth framework so far. Dai et al. [14] also introduced Tukey's depth for the general metric space. This depth framework adopts the idea of the classical Tukey's halfspace depth [8] to construct a metric-based depth formula. The mathematical theory for Tukey's metric depth is well formulated, whereas its computational cost is much more expensive. For this reason, we adopt sample depth Equation (3) with the metric depth in Geenens et al. [15] to rank spatial point process observations.

#### 2.4. Depth-Based Hypothesis Testing

In this section, we introduce a depth-based hypothesis testing method to compare two point processes. Hypothesis testing has been an important application of depth on multivariate data. Liu and Singh [10] introduced a distribution-dependent depth-based hypothesis test to compare two groups of multivariate observations. This method is only applicable for multivariate data with specific distribution (mainly for normal distribution and its extended forms). Wilcox [23] proposed two distribution-free test approaches for two-group multivariate data as an extension. There were also previous studies focusing on the hypothesis test on the point process. Berman [24] introduced a test approach to check whether there exists association between a point process and other stochastic processes based on the Poisson assumption. Schoenberg [25] conducted a non-parametric test to investigate the separability of a spatial-temporal marked point process. Guan [26] proposed a formal method to test the stationarity of the spatial point process. In a recent study, Fuentes-Santos et al. [27] introduced a non-parametric test to compare patterns between two groups of processes by estimating their intensity functions.

To our knowledge, the testing method in this paper is the first study examining the spatial point process using a depth framework. We adopt a nonparametric permutation approach to test whether two groups of point process observations are from the same distribution. This approach only depends on the depth values in both groups, which can be obtained by the spatial metric depth in this paper. That is, we consider the following hypothesis test for two groups ( $g_1$  and  $g_2$  with sample sizes  $m$  and  $n$ , respectively) of point processes:

- $H_0$ : The two groups of point process realizations follow the same distribution;
- $H_1$ : The two groups of point process realizations do not follow the same distribution.

Our testing algorithm is given in Algorithm 1. This testing approach is based on the newly defined depth on the point process with a standard permutation test framework. It utilizes the common testing procedure comparing multivariate data and generalizes the testing objects to point process data by using spatial metric depth.

If the testing result rejects the null hypothesis, then a follow-up classification method can be conducted to distinguish the point process groups. The studies on depth-based classification have been extensively conducted for decades. Liu [9] first introduced a simple maximum-depth classifier. Then, Li et al. [21] improved it and designed the well-known Depth-Depth (DD) classifier by finding an optimal boundary function in the DD plot [28]. A follow-up study [29] boosted the DD classifier by considering the second-order interaction of two groups' depth values and proposed the  $DD\alpha$  classifier. In a recent study, Zhou and Wu [30] further improved the DD classifier by restricting monotonicity of the boundary function and first applied it on the classification of temporal point processes. In the following sections, we use simulation and real data to demonstrate the effectiveness of the proposed testing method and evaluate the classification performance by the improved DD classifier.



**Algorithm 1** Hypothesis testing algorithm based on spatial metric depth

---

**Input:**  $m$  point process realizations from Group 1, denoted as  $g_1$ ;  $n$  realizations from Group 2, denoted as  $g_2$ ; hyper-parameter  $B$  for permutation test.

- Let  $g_1$  be the sample for the empirical distribution  $\hat{F}_n$  in Equation (3). Compute the spatial metric depth value of each realization in  $g_1$ , and denote the set of depth values as  $d_1$ . Then, compute the depth value of each realization in  $g_2$  and denote them as  $d_2$ ;
- Calculate the Kolmogorov–Smirnov (KS) statistic between  $d_1$  and  $d_2$  and denote it as  $K$ ;
- Combine  $g_1$  and  $g_2$  as one point process group  $g$  with sample size  $m + n$ . Initialize a counting index  $c = 0$ ;
- for**  $i = 1$  to  $B$  **do**
  - Randomly resample (without replacement)  $m$  realizations from  $g$  and denote them as  $g_1$ , and then denote the remaining  $n$  realizations as  $g_2$ ;
  - Recalculate depth values  $d_1$  and  $d_2$  of  $g_1$  and  $g_2$  based on the sample distribution of  $g_1$ , and then obtain the KS statistic between  $d_1$  and  $d_2$  as  $K'$ ;
  - if**  $K < K'$  **then**
    - $c = c + 1$ ;
  - end if**
- end for**
- Obtain the first  $p$ -value  $p_1 = \frac{c}{B}$ ;
- Repeat all previous steps by swapping the roles of  $g_1$  and  $g_2$  and obtain the second  $p$ -value  $p_2$ ;
- Compute  $p$  by conducting the Benjamini–Hochberg correction between  $p_1$  and  $p_2$ ;

**Output:**  $p$  is the final  $p$ -value.

---

**3. Simulation Illustrations**

In this section, we conduct simulation studies to illustrate the spatial metric depth via various types of spatial point processes. We examine and compare data from the Cox process, the Poisson process, the hard core process, and the Strauss process in two examples.

**3.1. Example 1: Log Gaussian Cox Process and Homogeneous Poisson Process**

First, we illustrate the depth ranking result on simulations from a Log Gaussian Cox process (LGCP) group and a homogeneous Poisson process (HPP) group. To simulate LGCP realizations, the first step is to create a Gaussian random field in the given domain. We design a random field,  $Y \in [0, 1]^2$ , such that close locations have relatively higher correlations, while far-away ones have relatively lower correlations. In this example, the mean is given as a constant,  $m(\xi) = \mathbb{E}(Y(\xi)) = 1.7$ , and the covariance is defined as a Laplacian kernel,  $C(\xi, \eta) = \text{Cov}(Y(\xi), Y(\eta)) = 3e^{-5\|\xi - \eta\|}$ ,  $\xi, \eta \in [0, 1]^2$ . Here,  $\|\cdot\|$  indicates Euclidean distance in  $\mathbb{R}^2$ .

Once the mean and covariance are obtained, it is straightforward to generate random field  $Y$ . Next, a Poisson process  $s$  driven by the intensity function  $\mu^* = \exp(Y)$  is the anticipated LGCP realization. In this case, the log-intensity varies in different realizations. Two example heatmaps of the log-intensity functions are shown in Figure 2. From the heatmaps, we can find that both larger and smaller values occur in clear clusters. This coincides with the covariance design such that points close to each other have higher correlations.

According to the mean  $m(\cdot)$  and covariance  $C(\cdot, \cdot)$  of the LGCP [31], the population mean of the intensity function is given as  $\mu(\xi) = \mathbb{E}(\exp(Y(\xi))) = e^{m(\xi) + \frac{C(\xi, \xi)}{2}}$ . That is, the expected intensity  $\mu(\xi)$  is a constant function  $e^{3.2} \approx 24.53$ . In this study, we propose to adopt this constant to simulate a sample of a homogeneous Poisson process (HPP) as comparison. This can be treated as a first-order approximation to the LGCP. That is, two groups of point processes are simulated as follows:

- Group 1 (LGCP): 1000 independent LGCP realizations on  $[0, 1]^2$  with the Gaussian random field given above;

- Group 2 (HPP): 1000 independent HPP realizations in  $[0, 1]^2$  with constant intensity function  $\mu = e^{3.2}$ .

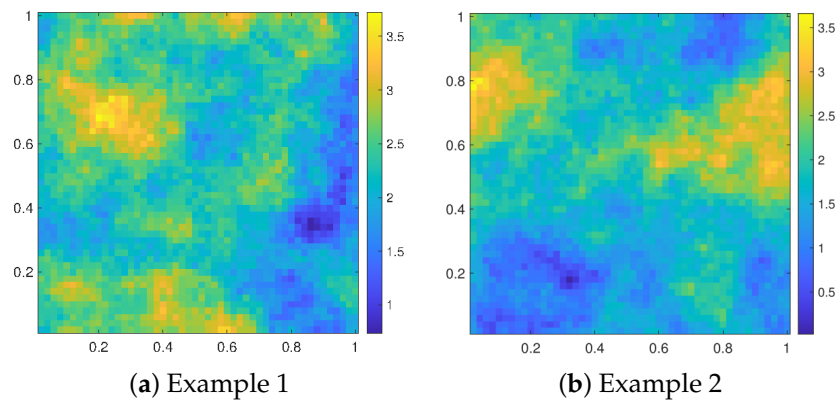


Figure 2. Heatmap examples of log-intensities from different LGCP realizations.

Next, the proposed spatial metric depth is applied to provide a center-outward ranking for the two groups. Hyper-parameters  $\lambda$  and  $c_2$  are chosen as 0.05 and 1, respectively (this is based on a cross-validation procedure and details are provided later in this section). Since  $c_1$  has no impact on the depth value, it is fixed as a constant one throughout this paper. For the LGCP Group, the histogram of the cardinalities is given in Figure 3a with mean 24 and median 18, respectively. The cardinality distribution is right-skewed with extreme outliers. Figure 3b,c shows the typical and outlier patterns based on the penalized metric and the smoothing metric, respectively. The typical patterns exhibit a distinct clustering phenomenon: if there exists one point in a certain area, then it is more likely to have multiple points alongside with it. The cardinalities of the typical patterns are around 15–20, which follows median cardinality. The outlier patterns are straightforward to distinguish with apparently more or less points.

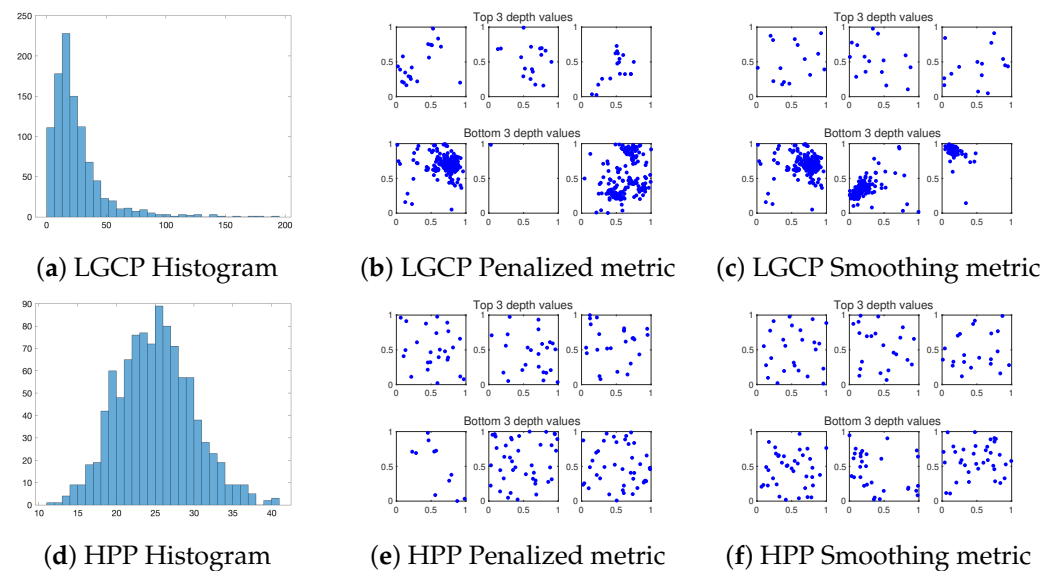


Figure 3. Simulation results of the LGCP Group and the HPP Group. (a) Histogram of the cardinalities of the 1000 simulated processes in the LGCP Group. (b) Typical and outlier patterns with top and bottom 3 depth values using the penalized metric on the first and second rows, respectively. (c) Same as (b) except for the smoothing metric. (d–f) Same as (a–c) except for the HPP Group.

For the HPP Group, the histogram of the cardinalities is given in Figure 3d. Based on the definition of HPP, the cardinality follows a Poisson distribution. Since the sample size is large, the distribution is nearly symmetric, the bell shape with both sample mean

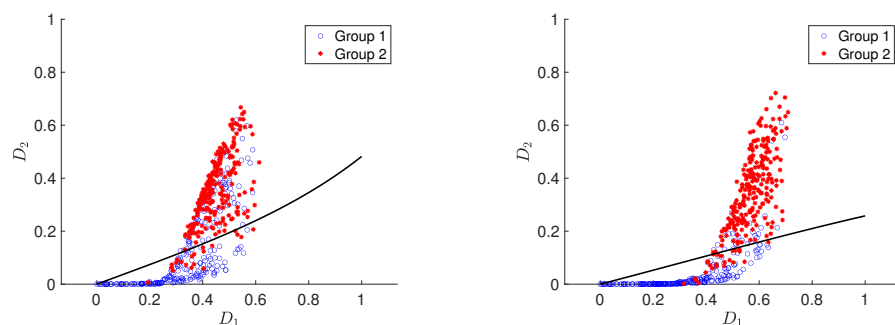
and median close to 24. The typical and outlier patterns are shown in Figure 3e,f for the two metrics, respectively. Compared with the result of the LGCP Group, the typical patterns are more uniformly distributed within the domain with less clusters. The points are able to cover most of the region of the domain. The outliers exhibit significantly different cardinalities.

We then conduct the proposed hypothesis tests in Section 2.4 to evaluate whether the spatial metric depth is capable of capturing the distribution information of the two groups. Here, three types of comparisons are conducted, where the first two types are for within-group comparison, and the third one is for across-group comparison.

1. A uniformly random subsample with size 100 from Group 1 vs. another uniformly random subsample with size 100 from Group 1.
2. A uniformly random subsample with size 100 from Group 2 vs. another uniformly random subsample with size 100 from Group 2.
3. A uniformly random subsample with size 100 from Group 1 vs. a uniformly random subsample with size 100 from Group 2.

For each of the above three types, we repeat the testing procedures in Algorithm 1 50 times with a significance level of 0.05. For Type 1, 47 and 50 experiments show non-significant results for the penalized metric and the smoothing metric, respectively. Similarly for Type 2, 48 and 46  $p$ -values are greater than 0.05 for both metrics. In general, around 5% of the total experiments show a false positive result. This coincides with the pre-specified significance level of 0.05. To further examine the capability of the depth function, Type 3 is conducted to evaluate statistical power. In this case, none of the  $p$ -values from the 50 repetitions are higher than 0.05 for both metrics. Therefore, spatial metric depth can capture the distribution information of point processes appropriately and demonstrate significant efficacy in distinguishing processes between different distributions.

Since there exists significant difference between the distributions of Groups 1 and 2, a classification with the DD classifier is conducted for them. For each group, 75% realizations are randomly selected as training data and the remaining 25% are used as test data. Then, a five-fold cross-validation is applied inside the training data with classification accuracy as the metric to determine hyper-parameter values. Both  $\lambda$  and  $c_2$  vary in a large range  $\{0.005, 0.01, 0.05, 0.1, 0.5, 1, 5, 10\}$ . This leads to the optimal values of  $\lambda = 0.05$  and  $c_2 = 1$ . Next, the DD classifier can be built on the whole training data. The test results are shown in Figure 4 in the DD plot. The test accuracies are 79.4% and 89.8% for the penalized metric and the smoothing metric, respectively. These high accuracies indicate the practicability of this newly proposed depth framework.



(a) Test result with the penalized metric (b) Test result with the smoothing metric

**Figure 4.** Test result of the DD classifier with two metrics, where the x-axis and y-axis are for the depth value in the LGCP Group and the HPP Group, respectively (denoted as  $D_1$  and  $D_2$ ). Blue circle indicates the realization in the LGCP Group, and the red star is for the HPP Group. The black curve represents the trained boundary of the DD classifier. (a) Test result with the penalized metric. (b) Same as (a) except for the smoothing metric.

### 3.2. Example 2: Hard Core Process and Strauss Process

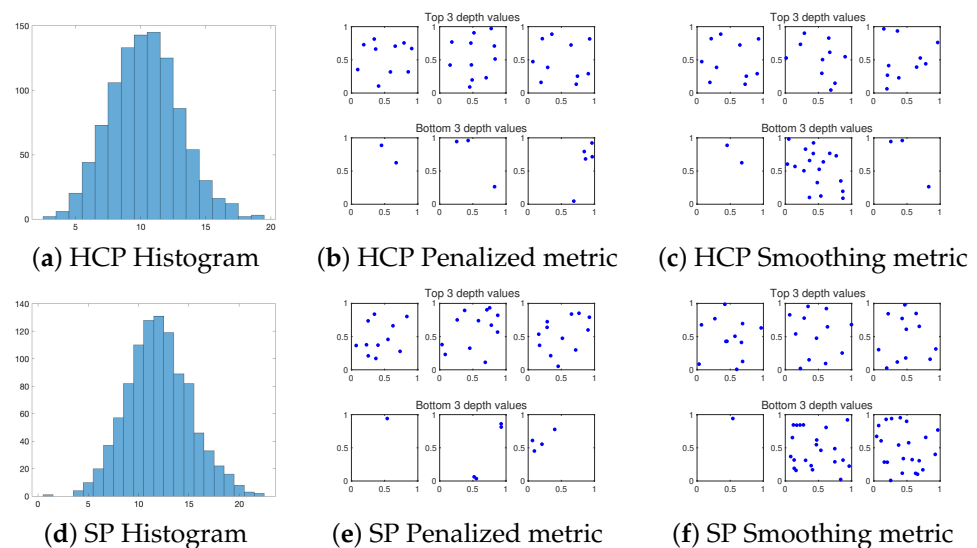
In this second example, we generate realizations from the Hard core process (HCP) and the Strauss process (SP). The Hard core process is similar to the homogeneous Poisson process except that it has one more parameter  $r$  that prohibits any two points within distance  $r$ . For a finite Hard core process  $s$ , the density function is given as  $f(s) = \exp(\sum_{u \in s} V_1(u) + \sum_{u,v \in s} V_2(u,v))$ , where

$$V_1(u) = \log \beta \quad V_2(u,v) = \begin{cases} 0 & \text{if } \|u - v\| > r \\ -\infty & \text{if } \|u - v\| \leq r \end{cases}$$

The Strauss process has “soft inhibition” between neighbouring pairs of points [1] by changing the  $V_2$  term as  $V_2(u,v) = (\log \gamma)\mathbf{1}(\|u - v\| \leq r)$ , where  $0 \leq \gamma \leq 1$  and  $\mathbf{1}(\cdot)$  is the indicator function. In this case, if  $\gamma = 0$ , then the Strauss process is equivalent to the Hard core process. If  $\gamma = 1$ , then the Strauss process has identical distribution with the homogeneous Poisson process with intensity  $\beta$ . The simulation groups are given as below.

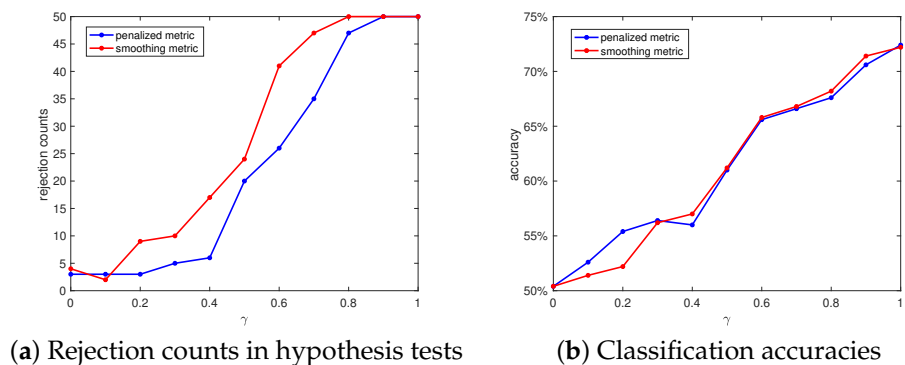
- Group 3 (HCP): 1000 independent Hard core processes in domain  $[0, 1]^2$  with  $\beta = 15$  and  $r = 0.1$ .
- Group 4 (SP): 1000 independent Strauss processes in domain  $[0, 1]^2$  with  $\beta = 15$ ,  $r = 0.1$  and  $\gamma = 0.5$ .

Analogous to the previous example, a cross-validation is applied to first determine the hyper-parameters with the same CV range, and the optimal result is  $\lambda = 0.05$  and  $c_2 = 0.05$ . Then, the ranking results are shown in Figure 5. The cardinality of the Hard core process shows a nearly symmetric distribution centered at about 10 and 11, which is well captured by the typical patterns. Comparing those typical patterns with the result of HPP, the points are again located uniformly within the domain in both cases. However, unlike in HPP, there are no points close to each other, which follows the property of the Hard core process. The outlier processes differ mainly in cardinalities for both metrics. Similar to the Hard core process, the cardinality of the Strauss process shows symmetric distribution centered at around 12. The typical patterns show points uniformly distributed within the domain, albeit with points close to each other. This result follows the definition of the Strauss process that it relaxes the restriction of the occurrence of neighboring points. On the other hand, the cardinalities of the outlier patterns are significantly different from the typical ones.



**Figure 5.** Simulation result of the HCP Group and the SP Group. (a) Histogram of the cardinalities of the 1000 simulated processes in the HCP Group. (b) Typical and outlier patterns with top and bottom 3 depth values using the penalized metric on the first and second row, respectively. (c) Same as (b) except for the smoothing metric. (d–f) Same as (a–c) except for the result of the SP Group.

Next, several hypothesis tests and classifications are conducted to distinguish the Hard core process and the Strauss process with different  $\gamma$  values. The two hyper-parameter values are kept the same as before for convenience. The result is shown in Figure 6. The difference between the Hard core process and the Strauss process becomes more distinguishable when  $\gamma$  is increasing from 0 to 1. Both metrics exhibit more testing power when  $\gamma$  is increasing and achieve a 70% testing power when  $\gamma$  is beyond 0.7. The classification shows better performance as  $\gamma$  becomes larger. When  $\gamma$  is close to zero, the two groups show high similarity with each other, which introduces confusion to the classifier. When  $\gamma$  approaches one, the accuracy achieves 70%. The performance demonstrates a decrement compared with the previous classification example due to the significant relationship between the two groups of processes.



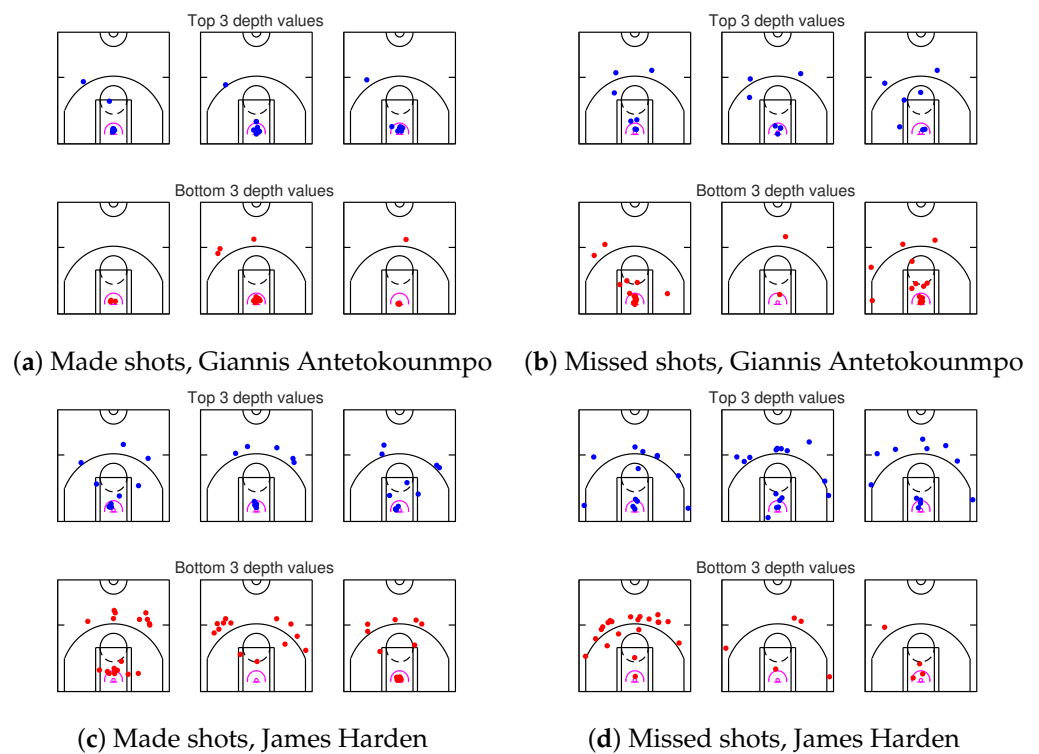
**Figure 6.** Hypothesis test and classification result between the Hard core process and the Strauss process. (a) The counts of rejection time among 50 tests with the variation in the value of  $\gamma$ . The blue and red curve represent the count results with the penalized metric and the smoothing metric, respectively. (b) The classification accuracies in both metrics with the value of  $\gamma$  varying from 0 to 1.

Based on the two simulation results, it can be concluded that the proposed depth approach provides an effective way to rank and separate commonly used spatial point processes. In the next section, a real-world dataset example is analyzed to further exhibit the applicability of this new method.

#### 4. Real Data Analysis

In this section, we apply the spatial metric depth framework on a real dataset. We collect the data of the shot positions of NBA players in each match of the Season 2018–2019, and each shot is recorded as “made” or “missed”. In this case, all made shots are by one player in one match form a single spatial point process. This is also the case for missed shots. The spatial domain is constrained as the standard half basketball court. For illustrative purposes, we select two well-known NBA players with different court positions and playing styles, Giannis Antetokounmpo and James Harden, to evaluate whether their made and missed shots exhibit different patterns. For simplicity, we only demonstrate the results with a smoothing metric.

Giannis Antetokounmpo and James Harden played 72 and 78 matches in that season, respectively, which leads to a sample size of 72 for both made and missed groups of Giannis Antetokounmpo, and 78 for both groups of James Harden. Similar to the previous simulation examples, a five-fold cross-validation is first conducted to determine the value of  $c_2$  from the range of  $\{0.001k \mid k = 1, 2, \dots, 1000\}$  and leads to  $c_2 = 0.782$  for Giannis Antetokounmpo and 0.736 for James Harden. The ranking result is shown in Figure 7.

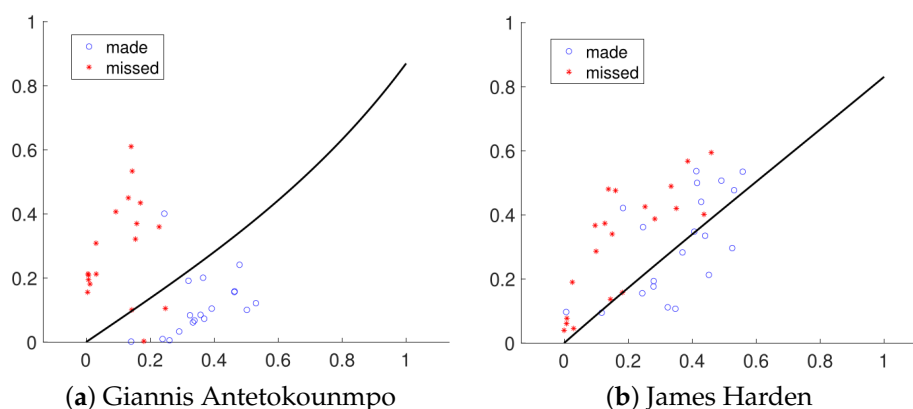


**Figure 7.** The typical and outlier patterns for made and missed shots of Giannis Antetokounmpo and James Harden. (a) The made shot positions exhibition of Giannis Antetokounmpo. The first row shows the shot positions (in blue) of the three typical matches, the second row shows the shot positions (in red) of the three outlier matches. (b) Same as (a) except for the missed shots. (c,d) Same as (a,b) except for the result of James Harden.

From typical patterns (with top three depth values) in Panels (a) and (b), we can see that Giannis Antetokounmpo’s shot positions exhibit clear difference between made and missed groups. It shows that he is more successful when shooting under the basket and within the three-second zone. Although it is not common, he may make some attempts outside the three-point line in a single match. If a three-point ball is attempted, he is more confident to shoot from the head (slightly towards to right wing) than other positions. Giannis Antetokounmpo may also shoot outside the three-second zone and within the three-point line, but it is more likely to result in a missed shot. In contrast, as shown in Panels (c) and (d), it is not straightforward to summarize the difference between James Harden’s made shot positions and missed ones. He prefers to shoot from the head of the key to the position around the two corners. It appears that he lacks proficiency in shooting from the two corners since it usually leads to a missed shot there. If James Harden enters the three-point line, he seldom shoots outside the three-second zone. Instead, he takes the ball to enter the restricted area near the basket and attempts to finish a layup. By comparing the typical patterns between these two players, we can conclude that Giannis Antetokounmpo prefers to attack under the basket while James Harden is more active outside the three-point line. Moreover, James Harden attempts more shots in a single match than Giannis Antetokounmpo, which indicates that basketball may be predominantly led by guards rather than forwards or centers.

Next, a hypothesis test is conducted to examine whether the made and missed shot positions come from the same point process. Unlike the previous simulation examples, the made and missed shot positions are paired instead of independent data. Thus, it is necessary to modify the test procedures in Algorithm 1 to make it appropriate for these data. In this case, when resampling the observations from the original two groups in each repetition, we just randomly swap (with 50% probability) the made and missed processes from one match to reform the two groups. All other steps remain the same. The experiment

shows the  $p$ -value equal to zero for both Giannis Antetokounmpo and James Harden, which shows that both players have their preferable shot positions. Since shot position distribution varies between the two groups, the DD classifier is built to separate them. For each player, 75% matches are randomly selected as training data and the remaining matches are test data. The test result is shown in Figure 8. The test accuracies are 92% and 75% for Giannis Antetokounmpo and James Harden, respectively. The classifier shows better performance for Giannis Antetokounmpo since his made and missed shot positions exhibit greater distinction.



**Figure 8.** Test classification result between made and missed shot positions. The x-axis and y-axis are for the depth value in the made and missed group, respectively. Blue circle indicates the realization in the made group, and the red star is for the missed group. The black curve represents the trained boundary of the DD classifier. (a) Test result for shot positions of Giannis Antetokounmpo. (b) Same as (a) except for James Harden.

The above analysis demonstrates the similarity and difference between made and missed shot positions for the two NBA players. More shooting patterns can be examined when more information is available; for example, we can collect the shot positions of a team in a season to study the team's offensive style or collect the shot positions of an opponent team to work on defense preparation.

## 5. Summary and Future Work

In this paper, we introduced a new framework to define depth for the spatial point process. The definition can be divided into two parts: (1) definition of a proper metric between the spatial point processes and (2) definition of the depth based on the proper metric. We proposed two types of proper metrics, the penalized metric and the smoothing metric, to measure the process distance. The metric properties and computational issues were extensively discussed. Simulations and a real dataset were applied to illustrate the effectiveness of the novel depth. We also compared similarities and differences between the two metrics.

To our knowledge, the spatial metric depth is the first attempt to define depth for the spatial point process. The entire framework is model-free and performs with high flexibility and efficiency to deal with different types of processes. Moreover, unlike the previous inter-arrival-event-based studies on temporal point process, the spatial metric depth regards the cardinality and event distribution as a whole under one unified framework to define the depth value. The spatial metric depth also provides a powerful method to conduct outlier detection of the spatial data by its natural center-outward ranking. The proposed depth-based hypothesis test provides a new tool to examine the similarity among point process groups. If the difference is identified, a DD classifier can be adopted as a powerful classification tool.

There are clear topics to further investigate in the future. First, the mathematical properties of the spatial metric depth are still incomplete. There is no clear symmetry in the

process space to define a “center of maximum”. Second, the model-free depth framework is flexible but lacks power in capturing the data pattern. A more specific depth framework, i.e., depth for the Poisson process, LGCP, etc., can be explored to extract spatial information for specific data groups. Finally, the current depth focuses on the spatial point process, but not the temporal information of each point. Future research will be conducted to study both spatial and temporal variabilities for more complex point processes, such as earthquake data, in real applications.

**Author Contributions:** Conceptualization, X.Z. and W.W.; methodology, X.Z. and W.W.; formal analysis, X.Z. and W.W.; writing—original draft preparation, X.Z. and W.W.; writing—review and editing, X.Z. and W.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Appendix A. Proof of the Properties of the Penalized Metric

The proof of the four properties is shown below:

- Nonnegativity: Trivial.
- Zero distance of a point process with itself: Trivial.
- Symmetry: Trivial.
- The Triangle Inequality: We suppose there are three point processes  $a, b, c$  with cardinality  $k, m, n$ , respectively. Then, based on the fact that the conventional Hausdorff metric is a proper metric, we have

$$\begin{aligned}
 d_{PH}(a, b) &= d_H(a, b) + \lambda|k - m| \\
 &\leq d_H(a, c) + d_H(b, c) + \lambda|k - m| \\
 &= d_H(a, c) + d_H(b, c) + \lambda(|k - n + n - m|) \\
 &\leq d_H(a, c) + d_H(b, c) + \lambda(|k - n| + |m - n|) \\
 &= d_{PH}(a, c) + d_{PH}(b, c)
 \end{aligned}$$

### Appendix B. Proof of the Bijection Mapping between Point Process and Its Smoothed Process

Given any point process  $s = \{s_1, s_2, \dots, s_n\}$  in  $[0, 1]^2$ , it is trivial to show that the transformation via the inverse of the Sigmoid function is bijective. Thus, we focus on the proof that the mapping between transformed process  $s^* = \{s_1^*, s_2^*, \dots, s_n^*\} = \{(x_1^*, y_1^*), \dots, (x_n^*, y_n^*)\}$  and its smoothed process  $f_s(x, y) = \sum_{i=1}^n K((x, y) - s_i^*)$  is bijective. Before the proof of bijection, the prerequisite shown below is necessary to verify:

- Kernel function  $K(s_i^*)$  in Equation (1) is linearly independent: for any  $n \in \mathbb{N}^+$ ,  $\sum_{i=1}^n \alpha_i K((x, y) - s_i^*) \doteq 0$  for  $(x, y) \in \mathbb{R}^2 \iff \alpha_1 = \dots = \alpha_n = 0$ .

**Proof of  $\implies$ :** We suppose  $\sum_{i=1}^n \alpha_i K((x, y) - s_i^*) = c_1 \sum_{i=1}^n \alpha_i e^{-c_2[(x-x_i^*)^2+(y-y_i^*)^2]} \doteq 0$  and there is no overlapping points. Then, we have

$$\begin{aligned}
 0 &\doteq c_1 \sum_{i=1}^n \alpha_i e^{-c_2[x^2+y^2-2xx_i^*-2yy_i^*+x_i^{*2}+y_i^{*2}]} \\
 &= \sum_{i=1}^n \alpha_i e^{-c_2(x_i^{*2}+y_i^{*2})} e^{2c_2x_i^*x} e^{2c_2y_i^*y} \\
 &= \sum_{i=1}^n p_i q_i^x h_i^y
 \end{aligned}$$



where  $p_i = \alpha_i e^{-c_2(x_i^{*2} + y_i^{*2})}$ ,  $q_i = e^{2c_2 x_i^*}$  and  $h_i = e^{2c_2 y_i^*}$  for  $i = 1, 2, \dots, n$ . We denote  $g(x, y) = \sum_{i=1}^n p_i q_i^x h_i^y$ ; since  $g(x, y) \doteq 0$ , we have

$$g(x, y) = p_1 q_1^x h_1^y + p_2 q_2^x h_2^y + \dots + p_n q_n^x h_n^y \doteq 0$$

We suppose there exists subset  $\{p_{(1)}, p_{(2)}, \dots, p_{(r)}\} \subset \{p_i\}_{i=1}^n$  such that  $p_{(j)} \neq 0$  for  $j = 1, 2, \dots, r$ . Then, we have

$$g(x, y) = p_{(1)} q_{(1)}^x h_{(1)}^y + p_{(2)} q_{(2)}^x h_{(2)}^y + \dots + p_{(r)} q_{(r)}^x h_{(r)}^y \doteq 0$$

Here,  $\{q_{(j)}\}_{j=1}^r$  are permuted in a decreasing order. If there exist  $a$  and  $b$  ( $a < b$ ) such that  $q_{(a)} = q_{(a+1)} = \dots = q_{(b)}$ , then  $\{h_{(j)}\}_{j=a}^b$  will be permuted in a decreasing order.

Next, two different cases are considered separately. (1) We assume  $q_{(1)}$  is the unique maximum one among  $\{q_{(j)}\}_{j=1}^r$ ; since  $q_{(j)} > 0$  for  $j = 1, 2, \dots, r$ , we have

$$p_{(1)} \left(\frac{q_{(1)}}{q_{(1)}}\right)^x h_{(1)}^y + p_{(2)} \left(\frac{q_{(2)}}{q_{(1)}}\right)^x h_{(2)}^y + \dots + p_{(r)} \left(\frac{q_{(r)}}{q_{(1)}}\right)^x h_{(r)}^y \doteq 0 \tag{A1}$$

We let  $x \rightarrow \infty$ , then  $p_{(1)} h_{(1)}^y \doteq 0$ . Since  $h_{(j)} > 0$  for  $j = 1, 2, \dots, r$ , then  $p_{(1)} = 0$ . This contradicts the assumption that  $p_{(j)} \neq 0$  for  $j = 1, 2, \dots, r$ . Therefore,  $p_i = 0$  for  $i = 1, 2, \dots, n$ , which leads to the conclusion that  $\alpha_i = 0$  for  $i = 1, 2, \dots, n$ .

(2) We assume  $q_{(1)}, q_{(2)}, \dots, q_{(k)}$  shares the same maximum value among  $\{q_{(j)}\}_{j=1}^r$ , then  $h_{(1)}$  must be the unique maximum value among  $\{h_{(j)}\}_{j=1}^k$  since there are no overlapping points. Next, based on Equation (A1), we let  $x \rightarrow \infty$ ; we have  $p_{(1)} h_{(1)}^y + p_{(2)} h_{(2)}^y + \dots + p_{(k)} h_{(k)}^y \doteq 0$ . We divide both sides with  $h_{(1)}^y$  and let  $y \rightarrow \infty$ ; we have  $p_{(1)} = 0$ . Again, this contradicts the initial assumption and leads to the conclusion that  $\alpha_i = 0$  for  $i = 1, 2, \dots, n$  holds.

**Proof of  $\Leftarrow$ :** If  $\alpha_i = 0$  for  $i = 1, 2, \dots, n$ , then it is straightforward to verify that for any  $n \in \mathbb{N}^+$ ,  $\sum_{i=1}^n \alpha_i K((x, y) - s_i^*) \doteq 0$  for  $(x, y) \in \mathbb{R}^2$ .

Next, we move on to prove the bijection mapping between the transformed process and its smoothed process.

**Surjection:** Based on the definition of the space of the smoothed process, if there exists a non-constant function  $f \in F$ , then there must be a transformed process  $s^* = \{s_1^*, \dots, s_k^*\}$  such that  $f = \sum_{i=1}^k K((x, y) - s_i^*)$ . Therefore, the surjection is verified.

**Injection:** We suppose there are two transformed processes  $s^*$  and  $t^*$  with cardinality  $m$  and  $n$ , respectively; their corresponding smoothed processes are  $f_s(x, y) = \sum_{i=1}^m K((x, y) - s_i^*)$  and  $f_t(x, y) = \sum_{j=1}^n K((x, y) - t_j^*)$ . We suppose  $f_s(x, y) \doteq f_t(x, y)$ ; then, we have

$$\begin{aligned} 0 &\doteq f_s(x, y) - f_t(x, y) \\ &= \sum_{i=1}^m K((x, y) - s_i^*) - \sum_{j=1}^n K((x, y) - t_j^*) \end{aligned}$$

Since the kernel is linearly independent and none of the coefficients are zero, then the equation holds if  $m = n$  and  $\{s_i^*\}_{i=1}^m = \{t_j^*\}_{j=1}^n$ . Thus, the two transformed process are identical and the injection is verified.

Therefore, the mapping process between the original point process and its smoothed process is bijective.

### Appendix C. Proof of Proposition 1

Based on Equation (1) and Definition 2, for the given spatial point processes  $s$  and  $t$  and their transformed point processes  $s^*$  and  $t^*$ , respectively, the square of the  $\mathbb{L}^2$  distance of their smoothed processes is

$$\begin{aligned}
 d_F(s, t)^2 &= \|f_s - f_t\|_2^2 \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( \sum_{i=1}^m c_1 e^{-c_2[(x-x_i^*)^2+(y-y_i^*)^2]} - \sum_{j=1}^n c_1 e^{-c_2[(x-u_j^*)^2+(y-v_j^*)^2]} \right)^2 dx dy \\
 &= \sum_{i=1}^m \sum_{j=1}^m \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c_1^2 e^{-c_2[(x-x_i^*)^2+(y-y_i^*)^2+(x-x_j^*)^2+(y-y_j^*)^2]} dx dy \\
 &\quad + \sum_{i=1}^n \sum_{j=1}^n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c_1^2 e^{-c_2[(x-u_i^*)^2+(y-v_i^*)^2+(x-u_j^*)^2+(y-v_j^*)^2]} dx dy \\
 &\quad - 2 \sum_{i=1}^m \sum_{j=1}^n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c_1^2 e^{-c_2[(x-x_i^*)^2+(y-y_i^*)^2+(x-u_j^*)^2+(y-v_j^*)^2]} dx dy \\
 &= c_1^2 \sum_{i=1}^m \sum_{j=1}^m \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-2c_2[(x-\frac{x_i^*+x_j^*}{2})^2+\frac{1}{4}(x_i^*-x_j^*)^2+(y-\frac{y_i^*+y_j^*}{2})^2+\frac{1}{4}(y_i^*-y_j^*)^2]} dx dy \\
 &\quad + c_1^2 \sum_{i=1}^n \sum_{j=1}^n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-2c_2[(x-\frac{u_i^*+u_j^*}{2})^2+\frac{1}{4}(u_i^*-u_j^*)^2+(y-\frac{v_i^*+v_j^*}{2})^2+\frac{1}{4}(v_i^*-v_j^*)^2]} dx dy \\
 &\quad - 2c_1^2 \sum_{i=1}^m \sum_{j=1}^n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-2c_2[(x-\frac{x_i^*+u_j^*}{2})^2+\frac{1}{4}(x_i^*-u_j^*)^2+(y-\frac{y_i^*+v_j^*}{2})^2+\frac{1}{4}(y_i^*-v_j^*)^2]} dx dy \\
 &= \frac{c_1^2 \pi}{2c_2} \cdot \left( \sum_{i=1}^m \sum_{j=1}^m e^{-\frac{c_2}{2}[(x_i^*-x_j^*)^2+(y_i^*-y_j^*)^2]} + \sum_{i=1}^n \sum_{j=1}^n e^{-\frac{c_2}{2}[(u_i^*-u_j^*)^2+(v_i^*-v_j^*)^2]} \right. \\
 &\quad \left. - 2 \sum_{i=1}^m \sum_{j=1}^n e^{-\frac{c_2}{2}[(x_i^*-u_j^*)^2+(y_i^*-v_j^*)^2]} \right)
 \end{aligned}$$

Then, the result in Proposition 1 is easy to obtain.

### Appendix D. Interpretation of the Properties in Proposition 2

The properties are discussed one by one.

( $P_1$ ): In this paper, the domain of the spatial point process is fixed as  $[0, 1]^2$  in order to simplify illustration. However, the depth formula can be generalized to any rectangular domain with an invariant result. Given any point process  $s$  in  $[0, 1]^2$ ,  $s' = Us + w$  is the transformed result via translation and scaling. In this case, the entire point process population is updated by the same parameters. Based on the definition of the two metrics, for any four processes  $s, t, p$  and  $q$ , the relationship between  $d(s, t)$  and  $d(p, q)$  is preserved after transformation, which leads to the linear invariance result. It is noteworthy to point out that the value of hyper-parameter  $\lambda$  is necessary to change in order to preserve the relationship for the penalized metric. In real applications, if the domain is not  $[0, 1]^2$ , then we can first normalize the domain before computing the depth value.

In fact, linear invariance holds in a more general case. We suppose  $U$  is any two-by-two non-singular matrix; then, for any point process  $s$ ,  $Us$  leads to the rotation and stretch result and the square domain becomes a parallelogram. Then, the inner distance relationship of any paired points within individual process are still preserved, which results in the same distance relationship among the processes in the metric space. However, in the simulation studies and the real data experiment of this paper, we only consider the point processes in horizontal rectangular domains. Further details of the linear invariance property in a more general domain are not discussed.

( $P_2$ ): Geenens et al. [15] proved that in any metric space, the depth value of an object becomes zero if the distance between itself and any fixed object approaches infinity. In this paper, the primary question to address is: If point process  $t$  is fixed with cardinality  $n$ , under which conditions of another process  $s$  will distance  $d(s, t)$  approach infinity? The answer can be discussed in two aspects.

1. Fix the cardinality of  $s$  as a finite integer  $m$ ; suppose there exists at least one individual point approaching the boundary of the point process domain. If the smoothing metric is applied, then  $d_F(s, t)$  is always finite based on the result in Proposition 1. If the penalized metric is applied, then  $d_{PH}(s, t)$  is finite if the point process domain is bounded. Thus,  $d(s, t)$  is always finite for any process in a bounded domain.
2. The cardinality of  $s$  (denoted as  $m$ ) approaches infinity. In this case, the penalty term of the penalized metric is infinite, which leads to the infinite value of the distance. For the smoothing metric, the value is  $O(m^2)$ , which approaches infinity with respect to  $m$ .

Therefore, for any process  $s$ , the spatial metric depth value approaches zero if its cardinality increases to infinity.

( $P_3$ ): Similar to  $P_2$ , the continuity property was verified by Geenens et al. [15]. The remaining question is: Under which conditions will the distance between two processes  $s$  and  $t$  be zero? When the smoothing metric is applied, based on the proof in Appendix B, the smoothed functions of  $s$  and  $t$  are equivalent if and only if  $s$  and  $t$  are identical. Thus,  $d_F(s, t) = 0$  is under the condition that  $s$  and  $t$  have the same cardinality and point locations. If the penalized metric is applied, then  $d_{PH}(s, t) \neq 0$  if they own different cardinalities due to the penalty term. Based on the same cardinality, if there exists one point in  $t$  (denoted as  $t_j$ ) such that the Euclidean distance between  $t_j$  and all points in  $s$  are not zero, then the first term of the penalized metric is greater than zero. Therefore, for both of the two metrics, the distance between  $s$  and  $t$  is zero if and only if  $s$  and  $t$  share identical point clouds.

( $P_4$ ): For any spatial point process  $s$ , its counting measure  $N_s$  is a measurable map from a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  to the outcome space  $(\mathbb{N}, \mathcal{N})$ . Then, its distribution can be formally defined as probability measure  $P$  on the outcome space as  $P(A) = \mathbb{P}(N_s \in A) = \mathbb{P}(\omega \in \Omega : N_s(\omega) \in A)$ , where  $A \in \mathcal{N}$  [1]. This property emphasizes the continuity on this probability measure.

### Appendix E. Proof of Proposition 3

Given any probability measure  $P$ , the proof of convergence between sample probability to population probability is straightforward and given as follows: For a metric space  $(\mathcal{M}, d)$  with respect to probability measure  $P \in \mathcal{P}$ , given any subset  $A \in \mathcal{M}$ , the population probability of  $A$  is defined as  $P(A) = \int_A dP = \int I(X \in A) dP$ . Thus, for an empirical probability of  $A$  defined as  $\hat{P}_n(A) = \frac{1}{n} \sum_{i=1}^n I(X_i \in A)$ , from the strong Law of Large Number, we have  $\hat{P}_n(A) \rightarrow P(A)$  almost surely if  $n \rightarrow \infty$ .

Based on the result above, the convergence of the empirical spatial metric depth to the population one is shown as follows: For any  $s \in S$  and a sample of point process  $s_i; i = 1, \dots, n \in S$ , the empirical depth is given as  $D(s, \hat{P}_n) = \frac{1}{\binom{n}{2}} \sum_{i < j} I\{d(s_i, s_j) > \max\{d(s_i, s), d(s_j, s)\}\}$ . Thus, based on the strong Law of Large Number and the definition of population spatial metric depth,  $D(s, \hat{P}_n)$  converges to  $D(s, P)$  almost surely.

### References

1. Baddeley, A.; Bárány, I.; Schneider, R. Spatial point processes and their applications. In *Stochastic Geometry: Lectures Given at the CIME Summer School Held in Martina Franca, Italy, 13–18 September 2004*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 1–75.
2. Waagepetersen, R.; Guan, Y. Two-step estimation for inhomogeneous spatial point processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2009**, *71*, 685–702. [[CrossRef](#)]
3. Talgat, A.; Kishk, M.A.; Alouini, M.S. Nearest neighbor and contact distance distribution for binomial point process on spherical surfaces. *IEEE Commun. Lett.* **2020**, *24*, 2659–2663. [[CrossRef](#)]

4. Byers, S.; Raftery, A.E. Nearest-neighbor clutter removal for estimating features in spatial point processes. *J. Am. Stat. Assoc.* **1998**, *93*, 577–584. [[CrossRef](#)]
5. Pei, T.; Zhu, A.X.; Zhou, C.; Li, B.; Qin, C. Detecting feature from spatial point processes using Collective Nearest Neighbor. *Comput. Environ. Urban Syst.* **2009**, *33*, 435–447. [[CrossRef](#)]
6. Bar-Hen, A.; Emily, M.; Picard, N. Spatial cluster detection using nearest neighbor distance. *Spat. Stat.* **2015**, *14*, 400–411. [[CrossRef](#)]
7. Al-Hourani, A.; Evans, R.J.; Kandeepan, S. Nearest neighbor distance distribution in hard-core point processes. *IEEE Commun. Lett.* **2016**, *20*, 1872–1875. [[CrossRef](#)]
8. Tukey, J.W. Mathematics and the picturing of data. In Proceedings of the International Congress of Mathematicians, Vancouver, BC, Canada, 21–29 August 1974; Volume 2, pp. 523–531.
9. Liu, R.Y. On a notion of data depth based on random simplices. *Ann. Stat.* **1990**, *18*, 405–414. [[CrossRef](#)]
10. Liu, R.Y.; Singh, K. A quality index based on data depth and multivariate rank tests. *J. Am. Stat. Assoc.* **1993**, *88*, 252–260.
11. Dyckerhoff, R.; Mosler, K.; Koshevoy, G. Zonoid data depth: Theory and computation. In *Proceedings of the COMPSTAT*; Springer: Berlin/Heidelberg, Germany, 1996; pp. 235–240.
12. López-Pintado, S.; Romo, J. On the concept of depth for functional data. *J. Am. Stat. Assoc.* **2009**, *104*, 718–734. [[CrossRef](#)]
13. Nieto-Reyes, A. On the properties of functional depth. In *Recent Advances in Functional Data Analysis and Related Topics*; Physica: Heidelberg, Germany, 2011; pp. 239–244.
14. Dai, X.; Lopez-Pintado, S.; Initiative, A.D.N. Tukey’s depth for object data. *J. Am. Stat. Assoc.* **2023**, *118*, 1760–1772. [[CrossRef](#)]
15. Geenens, G.; Nieto-Reyes, A.; Francisci, G. Statistical depth in abstract metric spaces. *Stat. Comput.* **2023**, *33*, 46. [[CrossRef](#)]
16. Liu, S.; Wu, W. Generalized mahalanobis depth in point process and its application in neural coding. *Ann. Appl. Stat.* **2017**, *11*, 992–1010. [[CrossRef](#)]
17. Qi, K.; Chen, Y.; Wu, W. Dirichlet depths for point process. *Electron. J. Stat.* **2021**, *15*, 3574–3610. [[CrossRef](#)]
18. Xu, Z.; Wang, C.; Wu, W. A unified framework on defining depth for point process using function smoothing. *Comput. Stat. Data Anal.* **2022**, *175*, 107545. [[CrossRef](#)]
19. Zhou, X.; Ma, Y.; Wu, W. Statistical depth for point process via the isometric log-ratio transformation. *Comput. Stat. Data Anal.* **2023**, *187*, 107813. [[CrossRef](#)]
20. Illian, J.; Penttinen, A.; Stoyan, H.; Stoyan, D. *Statistical Analysis and Modelling of Spatial Point Patterns*; John Wiley & Sons: Hoboken, NJ, USA, 2008.
21. Li, J.; Cuesta-Albertos, J.A.; Liu, R.Y. DD-classifier: Nonparametric classification procedure based on DD-plot. *J. Am. Stat. Assoc.* **2012**, *107*, 737–753. [[CrossRef](#)]
22. Zuo, Y.; Serfling, R. General notions of statistical depth function. *Ann. Stat.* **2000**, *28*, 461–482.
23. Wilcox, R.R. Two-Sample, Bivariate Hypothesis Testing Methods Based on Tukey’s Depth. *Multivar. Behav. Res.* **2003**, *38*, 225–246. [[CrossRef](#)]
24. Berman, M. Testing for spatial association between a point process and another stochastic process. *J. R. Stat. Soc. Ser. C Appl. Stat.* **1986**, *35*, 54–62. [[CrossRef](#)]
25. Schoenberg, F.P. Testing separability in spatial-temporal marked point processes. *Biometrics* **2004**, *60*, 471–481. [[CrossRef](#)] [[PubMed](#)]
26. Guan, Y. A KPSS test for stationarity for spatial point processes. *Biometrics* **2008**, *64*, 800–806. [[CrossRef](#)] [[PubMed](#)]
27. Fuentes-Santos, I.; González-Manteiga, W.; Mateu, J. A nonparametric test for the comparison of first-order structures of spatial point processes. *Spat. Stat.* **2017**, *22*, 240–260. [[CrossRef](#)]
28. Liu, R.Y.; Parelius, J.M.; Singh, K. Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *Ann. Stat.* **1999**, *27*, 783–858. [[CrossRef](#)]
29. Lange, T.; Mosler, K.; Mozharovskiy, P. Fast nonparametric classification based on data depth. *Stat. Pap.* **2014**, *55*, 49–69. [[CrossRef](#)]
30. Zhou, X.; Wu, W. Depth-Based Statistical Inferences in the Spike Train Space. *arXiv* **2023**, arXiv:2311.13676.
31. Daley, D.J.; Vere-Jones, D. *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*; Springer: New York, NY, USA, 2003.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.