Pattern Recognition Letters 32 (2011) 1894–1901

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

A method for modeling and visualizing the three-dimensional organization of neuron populations from replicated data: Properties, implementation and illustration

J. Burguet [a,b,*], Y. Maurin [a,b], P. Andrey [a,b,c]

[a] INRA, UR1197 Neurobiologie de l'Olfaction et Modélisation en Imagerie, F-78350 Jouy-en-Josas, France
[b] IFR 144, NeuroSud Paris, F-91190 Gif-Sur-Yvette, France
[c] UPMC, Univ Paris 06, France

ARTICLE INFO

Article history:
Received 4 March 2011
Available online 27 July 2011
Communicated by F. Tortorella

Keywords:
Cell density
*k*th nearest neighbor distance
Neuroanatomy
Point process

ABSTRACT

Understanding how the architecture of neuronal populations contributes to brain function requires three-dimensional representations and analyses. Neuroanatomical techniques are available to locate neurons in animal brains. Repeating an experiment in different individuals yields a collection of point patterns from which common organization principles are generally difficult to extract. We recently addressed the problem of generating statistical density maps to integrate replicated point pattern data into meaningful, interpretable representations. Applications to different neuroanatomical systems illustrated the ability of our method to reveal organization rules that cannot be perceived directly on raw data. To make the method practicable for further applications, the aim of the present paper is to establish general guidelines for appropriate parameter tuning, valid result interpretation as well as efficient implementation. Accordingly, we characterize the method by analyzing the role of its main parameter, by reporting results on its statistical properties and by demonstrating its robustness, using both simulated and real neuroanatomical data.

1. Introduction

Many biological studies entail the analysis of the three-dimensional (3D) position of structures that can be assimilated to points within cells, tissues or organs. Because experiments are repeated, typical datasets consist of sets of 3D point positions. A key issue is then to generate readable and meaningful statistical representations from these data.

Animal neuroanatomy is one of the many research fields that call for such developments. In the nervous system, neuronal populations are characterized by their morphological, biochemical and physiological properties as well as by their 3D organization. Understanding brain function therefore requires accurate representations and quantitative analyses of the spatial arrangement of these populations. Neurons of a population can be identified and located in an animal brain by using cell labeling techniques and by recording their 3D positions with microscopic imaging. Results from such experiments are affected by two sources of variability. The first one is biological: the actual number and positions of neurons in a population vary from one animal to the other. The second one is experimental: for a number of technical reasons, only a random subset of the population is generally labeled in one animal. Consequently, labeling the same population in several animals yields a collection of different 3D point sets. These fluctuations can be large enough to hide any organization rule in the collection of individual patterns.

To deal with variability, statistical methods are required. The theory of spatial point processes (Diggle, 2003) is the basis for most existing studies concerned with the spatial organization of neuron populations. In particular, summary statistics based on distance functions have been widely used for 2D and 3D analyses. However, whether they process point patterns independently (Armstrong, 2006; Bjaalie and Diggle, 1990; Bjaalie et al., 1991; Cotter et al., 2002; König et al., 1991; Schmitt et al., 2000) or integrate replicated data (Baddeley et al., 1993; Diggle et al., 1991, 2000; Landau et al., 2004; Reed and Howard, 1997; Webster et al., 2006), approaches based on summary statistics yield scalar descriptors and do not provide explicit 3D representations of spatial organizations. Such representations can be obtained by computing neuron intensity (mean number of cells by unit area or volume) maps. In 3D, prevailing approaches consist in generating 3D histograms of cell count (Nadasdy and Zaborszky, 2001; Odeh et al., 2005; Vibert et al., 1976; Zaborszky et al., 2005) or in computing kernel density estimates (Kopel et al., 2009). However, such methods generate density maps from single point patterns only.

* Corresponding author at: INRA, UR1197 Neurobiologie de l'Olfaction et Modélisation en Imagerie, F-78350 Jouy-en-Josas, France.
E-mail addresses: Jasmine.Burguet@jouy.inra.fr (J. Burguet), Yves.Maurin@jouy.inra.fr (Y. Maurin), Philippe.Andrey@jouy.inra.fr (P. Andrey).

0167-8655/$ - see front matter © 2011 Published by Elsevier B.V.
doi:10.1016/j.patrec.2011.07.008

In a recent neuroanatomical study, we proposed a solution based on the theory of spatial point processes and on distances to kth nearest cells to integrate replicated patterns and to generate statistical 3D intensity maps of neuronal distributions. In different neurobiological systems, this strategy revealed organization rules that were difficult, if not impossible, to detect on raw data (Burguet et al., 2009; Schwarz et al., 2010). The purpose of the present paper is to establish general guidelines for appropriate parameter tuning, valid result interpretation as well as efficient implementation, which are required to make the method practically useful for a large range of further applications. Therefore, we investigate and establish the meaning and the influence of the main parameter of the method, i.e. the rank k of the neighboring cells considered for intensity estimation. Further, we characterize the statistical properties of the method and evaluate its robustness to departures from the underlying parametric hypothesis (i.e., the assumption of locally, completely random distribution). Lastly, we address the implementation issue. Based on the distances to neighboring points, the method is indeed potentially costly and efficient algorithms are required.

The paper is organized as follows. The intensity mapping method is described in Section 2. Theoretical properties of the local intensity estimator are discussed. Computational aspects and visualization techniques are also described. Results obtained on simulated and real neuroanatomical data are reported in Section 3 and discussed in Section 4.

2. Intensity estimation and mapping

2.1. Definitions and problem statement

A spatial point process is a stochastic process generating sets of points in $\mathbb{R}^d$. A realization of such a process is a spatial point pattern and a sample of the process is a set of spatial point patterns. In the following, "points" will refer to elements of the spatial patterns and "positions" to arbitrary elements in $\mathbb{R}^d$. The intensity $\lambda(p)$ is the average number of points per unit volume at position $p \in \mathbb{R}^d$. The process is homogeneous if λ is constant throughout space and inhomogeneous otherwise.

In the spatial point process theory, the homogeneous Poisson process is a reference model corresponding to complete spatial randomness (CSR). In this model, the intensity λ does not vary with position; the number $N(R)$ of points in any space region R of volume $|R|$ follows a Poisson distribution with parameter $\lambda|R|$; the $N(R)$ points in R are independently and uniformly distributed over R.

We will focus here on the 3D case ($d = 3$) because of its relevance for biological studies in general, but the method can be applied to other dimensions as well. We are given a sample of 3D point patterns generated by a given spatial point process. The objective is to estimate the intensity of the process and to build a 3D map of intensity variations with space position. We assume that the process may be inhomogeneous, but that the intensity in the neighborhood of any position varies slowly enough to be considered as locally constant. Thus, points are locally distributed according to the CSR model.

2.2. Local intensity estimation

Let n be the sample size (Fig. 1A) and p a position (Fig. 1B). To simplify notations, the local intensity $\lambda(p)$ will be noted λ in this subsection. Let X_{ik} be the distance between p and its kth nearest point in the ith set. Under CSR, the probability density function of X_{ik} depends on λ and is:

$$f(x_{ik}) = \frac{(4\lambda\pi)^k}{3^{k-1}(k-1)!}\, x_{ik}^{3k-1}\, \exp\left(-\frac{4\lambda\pi}{3}x_{ik}^3\right)$$

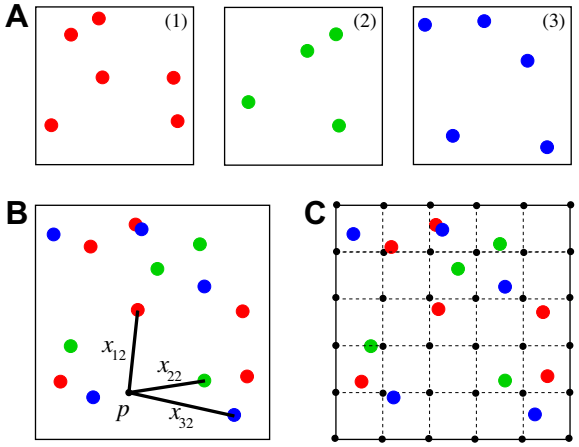

Fig. 1. Intensity mapping. (A) Sample of point patterns ($n = 3$). Bounding box: common volume of interest. (B) Local intensity estimation at position p as a function of the distances to the kth nearest neighbors in all patterns ($k = 2$). (C) Intensity map generation by estimating the local intensity at every node of a grid covering the volume.

The log-likelihood of the distances $X_{1k},\ldots,X_{nk}$ is:

$$\ln \prod_{i=1}^{n} f(x_{ik}) = nk \ln\lambda - \frac{4\pi}{3}\lambda \sum_{i=1}^{n} x_{ik}^3 + C \tag{1}$$

where C is independent of λ. Maximizing Eq. (1) with respect to λ yields the maximum likelihood estimator $\tilde{\lambda}$ of λ:

$$\tilde{\lambda} = \frac{nk}{\frac{4\pi}{3}\sum_{i=1}^{n} x_{ik}^3}$$

As previously shown (Burguet et al., 2009), $\tilde{\lambda}$ is biased (the expected value of $\tilde{\lambda}$ is not equal to the true value λ), but a simple correction yields the following unbiased estimator:

$$\hat{\lambda} = \frac{nk-1}{\frac{4\pi}{3}\sum_{i=1}^{n} x_{ik}^3} \tag{2}$$

of variance (Burguet et al., 2009):

$$V(\hat{\lambda}) = \lambda^2/(nk-2) \tag{3}$$

2.3. Local intensity estimator properties

Since it is unbiased and since $V(\hat{\lambda}) \to 0$ as $n \to \infty$, the estimator $\hat{\lambda}$ is consistent (its distribution concentrates more and more around λ as the sample size increases). Moreover, it can be shown from Eq. (1) that the Cramér-Rao bound (the theoretical lower bound of the estimator variance) is λ^2/nk. Thus, the efficiency of $\hat{\lambda}$ (the ratio between this bound and the estimator variance) is equal to $(nk-2)/nk$ and converges towards 1 when n grows. Hence, $\hat{\lambda}$ is asymptotically efficient. When k increases, the estimator variance converges toward 0 and its efficiency toward 1. Thus, k can be adjusted to control the degree of smoothing of the intensity estimate. Besides, larger values for k are expected to provide more accurate estimations.

Finally, in most practical situations (in particular when dealing with neuron populations), it is reasonable to assume that no two points of the same pattern can be located at the same position. Hence, imposing $k \geqslant 2$ guarantees that $x_{ik} > 0$, for all i, and thus that $\hat{\lambda}$ is finite. Moreover, this ensures that its variance is positive.

2.4. Intensity map generation and visualization

To generate a 3D intensity map, the volume of interest is discretized using a regular grid whose resolution is controlled by