Published in final edited form as:

Spat Stat. 2015 August ; 13: 76–89. doi:10.1016/j.spasta.2015.05.004.

Measuring Aggregation of Events about a Mass Using Spatial Point Pattern Methods

Michael O. Smith^{a,*}, Jackson Ball^{b,c}, Benjamin B. Holloway^{b,c}, Ferenc Erdelyi^d, Gabor Szabo^d, Emily Stone^{a,b}, Jonathan Graham^a, and J. Josh Lawrence^{b,c}

^aDepartment of Mathematical and Statistical Sciences, University of Montana, Missoula, MT 59812

^bCenter for Structural and Functional Neuroscience, University of Montana, Missoula, MT 59812

^cDepartment of Biomedical and Pharmaceutical Sciences, University of Montana, Missoula, MT 59812

^dLaboratory of Molecular Biology and Genetics, Institute of Experimental Medicine, Hungarian Academy of Sciences, H-1450 Budapest, Hungary

Abstract

We present a methodology that detects event aggregation about a mass surface using 3-dimensional study regions with a point pattern and a mass present. The Aggregation about a Mass function determines aggregation, randomness, or repulsion of events with respect to the mass surface. Our method closely resembles Ripley's K function but is modified to discern the pattern about the mass surface. We briefly state the definition and derivation of Ripley's K function and explain how the Aggregation about a Mass function is different. We develop the novel function according to the definition: the Aggregation about a Mass function times the intensity is the expected number of events within a distance h of a mass. Special consideration of edge effects is taken in order to make the function invariant to the location of the mass within the study region. Significance of aggregation or repulsion is determined using simulation envelopes. A simulation study is performed to inform researchers how the Aggregation about a Mass function performs under different types of aggregation. Finally, we apply the Aggregation about a Mass function to neuroscience as a novel analysis tool by examining the spatial pattern of neurotransmitter release sites as events about a neuron.

Keywords

3-Dimensions; Clustering; Point Process; Spatial Patterns; Spatial Statistics

*corresponding author at: Department of Mathematical and Statistical Sciences, University of Montana, Missoula, MT, USA 59812
Tel.: +1.406.243.6712.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Ripley's K function is a tool used to characterize the strength of spatial dependence at multiple scales in a spatial point pattern. The K function has been widely used to identify clustering, randomness, or regularity among events in a spatial point pattern (Ripley, 1977). Recent applications of Ripley's K function have been to the field of biology, where the spatial organization of molecules can be described with a spatial point pattern (Lagache et al., 2013; Kellner et al., 2007; Jenei et al., 2009). Ripley's K function has been used in the field of neuroscience primarily to describe the spatial pattern of neurons in 3-dimensions (Jafari-Mamaghani et al., 2010; Hansson et al., 2013; Eglen and Wong, 2008; Millet et al., 2011). Recent research explored distance functions to a single point reference, (Joyner et al., 2014) which is similar but different than our method. This paper explores a function based on Ripley's K function to characterize the level of aggregation of a point pattern relative to a *mass*. Analysis of spatial aggregation relative to a mass is a novel application.

This work was motivated by the increasing recognition that communication between neurons depends on the fine architecture of presynaptic (neurotransmitter-releasing) and postsynaptic (neurotransmitter-sensing) elements in the brain. The spatial pattern of release sites about a neuron is of physiological significance to synaptic transmission but has not yet been described mathematically. Towards this end, image stacks of neurotransmitter-releasing objects relative to a neuron were rendered in 3-dimensions and analyzed using a function similar to Ripley's K function. If we consider the objects as events, these images can be reduced to a 3-dimensional spatial point pattern distributed around a neuron. The objective in this application is to ascertain whether the objects are aggregated about the surface of the neuron, and if so to what degree. Applying the usual Ripley's K function to this point pattern, ignoring the neuron body, would measure the spatial aggregation among the objects, but would not tell us whether these objects tend to be aggregated about neuronal surfaces. A different function other than the K function is needed to assess whether the objects are aggregated *about* a neuron.

This paper consists of five additional sections. Since the motivation for this new method is derived from Ripley's K function, section 2 briefly reviews the construction of Ripley's K. Section 3 introduces the Aggregation about a Mass function using study regions with a point pattern and a mass. Construction of this method mimics the construction of Ripley's K, but differences are noted and discussed. Section 4 shows results of a simulation study assessing how well this new method characterizes the spatial aggregation under various known conditions. Then, by applying the Aggregation about a Mass function to neuroscience, analysis of events about a neuron is conducted in section 5. Finally, a discussion of important topics and future studies are discussed.

2. Ripley's K Function

2.1. Definition and Estimation

For a homogenous point pattern, Ripley's K function (Ripley, 1976, 1977) is defined by:

$\lambda K(h) = E(\text{number of additional events within distance } h \text{ of a single arbitrary event})$

(1)

where λ = the intensity or the mean number of events per unit area, assumed constant throughout the study region $\mathcal{R}$, and E is the expectation function. To estimate $K(h)$, let d_{ij} be the distance between the i th and j th observed events in $\mathcal{R}$ and define the indicator function

$$I(d_{ij} \leq h) = \begin{cases} 1 & \text{if } d_{ij} \leq h, \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

To incorporate an edge correction, consider a circle centered on event i that contains j . Let $w(d_{ij})$ be the proportion of the circumference of this circle which lies within $\mathcal{R}$. Since we assume the point pattern is homogeneous and the intensity is constant, then a simple estimate of intensity is $\hat{\lambda} = n/R$, where n is the number of events in $\mathcal{R}$, and n/R is the number of events per unit area in $\mathcal{R}$. This gives the usual estimate of $K(h)$ as:

$$\hat{K}(h) = \frac{R}{n^2} \sum_{i \neq j} \frac{I(d_{ij})}{w(d_{ij})}. \quad (3)$$

Other edge corrections can be found in Cressie (1993). A good graphical explanation of the estimated K function can be found in Bailey and Gatrell (1995, p. 93).

2.2. Hypothesis testing using $\hat{L}(h)$ and envelopes

$\hat{K}(h)$ can be plotted against values of distance h but little inference about the level of spatial dependence can be made from such a figure. $\hat{K}(h)$ must be compared to $K(h)$ under CSR (complete spatial randomness). The theoretical value of $K(h)$ under CSR is $K(h) = \pi h^2$. We would like to compare $\hat{K}(h)$ to πh^2 to assess the type of spatial dependence present. Typically the centered L function is used:

$$L(h) = \sqrt{\frac{\hat{K}(h)}{\pi}} - h. \quad (4)$$

where $K(h)$ is replaced by $\hat{K}(h)$ to give the estimate $\hat{L}(h)$ of $L(h)$. Positive values of $\hat{L}(h)$ indicate clustering since we are observing more event pairs within h compared to a point pattern under CSR. Conversely, negative values of $\hat{L}(h)$ indicate regularity at distance h within the point pattern.

In order to assess how positive (or how negative) $\hat{L}(h)$ must be in order to conclude significant clustering (or regularity) simulation envelopes are used (Ripley, 1977). Since

envelopes are used as a test of significance, special consideration of null models and p-values is needed. Baddeley et al. (2014) clarify when it is appropriate to use certain types of simulation envelopes. The use of pointwise and global envelopes are presented in later sections. Bailey and Gatrell (1995) provide a useful explanation of how one would simulate point patterns in order to create envelopes. Illian et al. (2008, p. 215) neatly lay out the formation of Ripley's K in dimensions greater than 2. An application of Ripley's K function can be found in Cressie (1993, p. 580).

3. The Aggregation about a Mass Function

3.1. Definition and Estimation

Data that consist only of event locations whose spatial pattern can be analyzed with Ripley's K function. Now consider a large object or mass in the study region along with events distributed around the mass. An example might be a nuclear power facility within a neighborhood with events defined as locations of deceased birds. In this example, one might hypothesize that event locations are spatially dependent on the location of the mass. If the mass was small relative to the study region, a single point could represent the mass. But, if the mass is large relative to the study region and there is an interest in clustering at small scales about the mass, a single point representation of the mass might be misleading. If the mass is not uniform in shape, the distance from the single point mass representation to the mass's boundary would vary. Thus, distances found would not represent events about the mass's boundary appropriately. Instead of considering a single point mass representation, we consider the mass's boundary which could be represented by a closed surface or many points estimating the closed surface, typically in 2 or 3-dimensions. In our case, we want the mass to be "solid" and not part of the study region. This allows events to only fall outside the mass within some study region.

When it is appropriate to consider a mass within the study region, and spatial aggregation about that mass is of concern, methods similar to Ripley's K function can be used. We consider searching for events within a distance h of the mass (Figure 1). When we do this we create a "band" (or "buffer" in 3D) about the mass that has width h . These bands vary as a function of h . Figure 1 illustrates one such band about the mass.

In order to test if event locations are dependent on the mass surface, we assume that in the absence of the mass the events exhibit CSR. Then for a homogenous point pattern, we define the Aggregation about a Mass function as:

$$\lambda A_M(h) = E(\text{\#of events within distance } h \text{ of a mass}) \quad (5)$$

where λ is the intensity. This definition differs from Ripley's K only in the expectation since we are now considering all distances to be "mass to event" rather than "event to event". This difference highlights that the A_M function investigates first order properties of the point pattern around the mass. The K function is designed to study the interaction between events.

To estimate A_M we take a similar approach as to how the K function is estimated. Let d_i be the minimum distance from the i th event to the mass, and let

$$I_h(d_i \leq h) = \begin{cases} 1 & \text{if } d_i \leq h, \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

If we sum equation $I_h(d_i \leq h)$ over all $i = 1, \dots, n$ for a given h where n is the number of events in the study region $\mathcal{R}$, then we are counting the number of events within each band of width h . This summation is an estimate of the expectation in equation 5. Note that as h varies the bands are nested, which makes $\sum_{i=1}^n I_h(d_i \leq h)$ a cumulative count of events. Note also that d_i is defined as the *nearest* distance to the mass. A single event has many distances to the boundary of the mass; d_i is the smallest of those distances. Since we assume the point pattern to be homogeneous with constant intensity, we estimate λ by n/R . It should be noted that R is the area (or volume in 3D) of only the study region which does not include the mass. Thus, our estimate of A_M is

$$\hat{A}_M(h) = \frac{R}{n} \sum_{i=1}^n I_h(d_i \leq h). \quad (7)$$

Now consider an edge correction for A_M . Graphically the A_M function has bands about the mass. Let $w(d_i)$ be the proportion of the band's area inside $\mathcal{R}$. One might also consider using the proportion of the band's outer circumference (or surface area in 3D) inside $\mathcal{R}$. Then for each i we scale the value of $I(d_i \leq h)$ by its corresponding $w(d_i)$. The estimate of A_M becomes

$$\hat{A}_M(h) = \frac{R}{n} \sum_{i=1}^n \frac{I_h(d_i \leq h)}{w(d_i)}. \quad (8)$$

This edge correction effectively extrapolates the count of events outside the study region,

making $\sum_{i=1}^n \frac{I_h(d_i \leq h)}{w(d_i)}$ invariant to the location of the mass about the study region. Moving the mass and its relative point pattern about the study region is essentially a translation of the mass and events. When this translation is performed different values of n are found, since some events will fall outside the study region depending on the translation. As a result, n also needs to be adjusted for the edge. Let n_e = the number of events within the largest distance h from the mass, adjusted by the edge correction. Notationally,

$$n_e = \sum_{i=1}^n \frac{I_h(d_i \leq \max h)}{w(d_i)}. \quad (9)$$

This correction of n will allow the values of $\hat{A}_M(h)$ to be roughly invariant to the location of the mass within $\mathcal{R}$. We should note that n_e is solely an adjustment to the scale of our estimate. Later, when we simulate point patterns under CSR, we use n events and not n_e events. It should also be noted that the values of h are chosen by the researcher. With these edge adjustments, our final estimate of A_M is

$$\hat{A}_M(h) = \frac{R}{n_e} \sum_{i=1}^n \frac{I_h(d_i \leq h)}{w(d_i)}. \quad (10)$$

3.2. Aggregation about a Mass Function under CSR

Plots of $\hat{A}_M(h)$ are not very meaningful unless we compare the values to what we expect A_M to look like under CSR. This is a difficult task since the mass could take on shapes that are irregular. Two-dimensional shapes such as circles or rectangles would make this task easier. Consider the mass to be circular with radius r in a 2-dimensional study region. Then the “bands” about the mass will be donut shaped with a width of h . The area of this band is computed as $\pi(h+r)^2 - \pi r^2$, so that using equation 5:

$$\lambda A_M(h) = \lambda [\pi(h+r)^2 - \pi r^2]. \quad (11)$$

Thus, under CSR, $A_M(h) = \pi(h+r)^2 - \pi r^2$ when the mass is a circle of radius r . In this case, we can scale $\hat{A}_M$ to the value of A_M under CSR. Similarly if the mass is a rectangle, the area of the bands can be computed analytically and $\hat{A}_M$ can be scaled. If the true area of bands can be found, one should utilize this information. However, if the mass is irregular and band areas are difficult or impossible to find, the value of A_M under CSR must be estimated.

To estimate A_M under CSR, we simulate point patterns under CSR. This is conceptually equivalent to randomly distributing events in the study region, but outside the mass. Each time a random point pattern is created $\hat{A}_M(h)$ is calculated. Let $\hat{A}_M^s(h)$, $s = 1, 2, \dots, m$ be the m simulated functions under CSR. We can then take the median (or mean) value of all the $\hat{A}_M^s(h)$ for each h to be the estimated value of $A_M(h)$ under CSR. The number of events we randomly distribute is dependent on n , the number of events observed in $\mathcal{R}$.

We can also use the $\hat{A}_M^s(h)$ values, $s = 1, 2, \dots, m$, to create pointwise envelopes according to some α level. Let the $\hat{A}_M$ lower envelope be the $\alpha/2$ percentile of the $\hat{A}_M^s(h)$ values at each h . Similarly, let the $\hat{A}_M$ upper envelope be the $1 - \alpha/2$ percentile of the $\hat{A}_M^s(h)$ values. The $\hat{A}_M(h)$ function and two $\hat{A}_M$ envelopes can be scaled by the simulated estimate of $A_M(h)$ under CSR to create B_M functions. B_M is similar to the centered L function. Notationally,

$$\hat{B}_M(h) = \hat{A}_M(h) - \text{median}[\hat{A}_M^s(h)] \quad (12)$$

$$B_M \text{ Upper Envelope} = A_M \text{ Upper Envelope} - \text{median}[\hat{A}_M^s(h)] \quad (13)$$

$$B_M \text{ Lower Envelope} = A_M \text{ Lower Envelope} - \text{median}[\hat{A}_M^s(h)]. \quad (14)$$

Positive values of $\hat{B}_M(h)$ indicate aggregation about the mass since we are observing more points within the bands than would be expected under CSR. Negative values of $B_M(h)$ indicate a repulsion from the mass since we are observing fewer events within the bands relative to CSR. The notion of repulsion here is different from its interpretation with Ripley's K . In our case, we view repulsion as the events avoiding the mass.

By viewing the plots of $\hat{B}_M(h)$ with envelopes, the significance of the negative or positive values of $\hat{B}_M(h)$ can be determined. It is appropriate to use pointwise envelopes to determine significance only if we decide to inspect the B_M function at one particular value of h . (Baddeley et al., 2014; Loosmore and Ford, 2006). This is typically not performed by researchers and is certainly not our goal in the application of A_M . Global envelopes (Ripley, 1977; Baddeley et al., 2014) do not suffer from this special consideration and will correctly determine significance levels. When testing for significant aggregation about a mass, we consider our null model to be that the events exhibit CSR when the mass is present in the study region. If the $\hat{B}_M$ function strays outside the global envelopes we conclude a departure from CSR and significant event to mass dependence.

4. Aggregation about a Mass Function Simulation Study

4.1. Simulations and Figures

To provide insight into how the Aggregation about a Mass function performs, a simulation study was performed. Simulated or contrived point patterns were generated. Simulations were performed using a 2-dimensional point pattern (instead of 3-dimensions) to increase clarity of discussion. The patterns were produced on a 1 unit by 1 unit study region. A mass within the study region is represented by a 2-dimensional circle with radius 0.1 and centered at (0.5, 0.5). We will refer to the three patterns as patterns 1, 2, and 3. Pattern 1 is the obvious first choice; a random point process. Computationally this implies sampling the x and y coordinates from two uniform distributions. Pattern 2 is an aggregated point pattern. This pattern is produced by sampling distances from the mass using an exponential distribution with mean distance θ^{-1} . If not otherwise stated, distances were sampled from an exponential distribution with mean 2^{-1} ($\theta = 2$). Direction from the mass was random. This creates a pattern where events are more likely to occur near the mass. As θ increases the point pattern exhibits a higher degree of aggregation about the mass. Pattern 3 is also an aggregated pattern but in a donut-like fashion. Events were sampled randomly but forced to occur between .1 and .2 units away from the mass, thus resulting in a “donut of events” around the mass. Figure 2 shows the three patterns.

The use of 95% pointwise envelopes are included in the simulation figures in order to better visually address possible departures from completely random patterns at multiple scales. If significance testing is of interest global envelopes are more appropriate. (Baddeley et al., 2014)

The first simulation was performed with different numbers of events. n was set to be 10, 100, and 500. The resulting average $\hat{B}_M$ functions and envelopes based on 1000 simulations are shown in Figure 3. In all cases of n , the $\hat{B}_M$ functions describe the patterns appropriately. Pattern 1, the random point process, remains within the envelopes for all values of h . $\hat{B}_M$ for Pattern 2, an aggregated process, immediately exceeds the envelopes to indicate aggregation about the mass. And pattern 3, an aggregated process where all events occur between 0.1 and 0.2 units from the mass, increases to a spike between .1 and .2. These characteristics of the $\hat{B}_M$ functions occur with all simulations hereafter. As the number of events, n , increases the $\hat{B}_M$ functions and pointwise envelopes become less noisy as expected. Also, as n increases, the envelope values narrow toward 0.

Figure 4 was created using two maximum distances from the mass, 0.3 and 0.6. In other words, the $\hat{B}_M$ function only performed calculations up to the respective maximum distance. Again, the $\hat{B}_M$ functions describe patterns 2 and 3 appropriately. In all cases, the $\hat{B}_M$ functions and pointwise envelopes terminate at 0 for the maximum value of h chosen. This is an artifact of the $\hat{B}_M$ function scaling by n_e and is similar to that of Ripley's K function. In both aggregated patterns in Figure 4, the scale of aggregation is inflated when the max distance is increased, especially with pattern 3.

The second simulation is shown in Figure 5. 1000 simulations were used to show the effect of moving the mass from the center of the 1 unit by 1 unit study region, (0.5, 0.5), toward a corner of the region at, (0.3, 0.3). Doing this heightens the effect of the edge correction built into the $\hat{A}_M$ and $\hat{B}_M$ functions. Patterns 1 and 3 exhibit no bias and are identical when the mass is moved. Pattern 2 exhibits little to no bias by showing more aggregation when the center is at (0.3, 0.3). Also the envelopes become slightly more narrow and approach zero more abruptly when the maximum distance is 0.6.

Figure 6 only uses point patterns similar to that of pattern 2. Like in pattern 2 explained above we use the exponential distribution to generate distances from the mass for each event. Different parameter (θ) values are used to vary the degree of aggregation about the mass. Specifically θ was set to 1, 5, and 10, where smaller choices of θ correspond to less aggregation about the mass. The simulation shows that the $\hat{B}_M$ functions illustrate greater aggregation when θ is increased as expected. $\hat{B}_M$ functions and envelopes are means of 1000 simulations.

The results from the simulations illustrate that the $\hat{B}_M$ function is stable and measures aggregation well when a mass is present in the study region. The correction to the $\hat{B}_M$ function, n_e , almost completely removes any bias when an edge effect is present. This is very useful since the mass might be located near the edge. The $\hat{B}_M$ function works very similar to Ripley's K and L functions, but measures a different kind of aggregation. Notice that when using pattern 3, the "aggregated donut", there are no events located past a distance

of 0.2 from the mass. Yet, the function still indicates significant clustering beyond 0.2. The function does however immediately begin to taper back toward zero.

It is possible to compare the quantity of aggregation by using methods similar to the index of association found in Fajardo et al. (2006). Ratios of the areas (or distances) between the 95% pointwise envelopes and the $\hat{B}_M$ function can be found. However, as with Figure 4, one should consider using the same maximum distance, h , before comparing. Comparing two different aggregated patterns like pattern 2 and pattern 3 could be misleading as well. These two patterns represent different kinds of aggregation and it is difficult to determine which is *more* aggregated about the mass.

5. Application of Ripley's K Function to Cholinergic Neurotransmission

5.1. Rationale and implications for investigating the spatial distribution of cholinergic terminals in the hippocampus

The release of the excitatory neurotransmitter acetylcholine (ACh) into the hippocampus critically modulates learning and memory through the activation of postsynaptic receptors (Hasselmo and Giocomo, 2006; Ruivo and Mellor, 2013). Cholinergic dysfunction in the hippocampus has been implicated in diverse disease states, including Alzheimer's disease, epilepsy, and autism spectrum disorders. The source of ACh arises from fibers of the medial septum-diagonal band of Broca (MS-DBB), which diffusely innervate all areas of the hippocampus. At a cellular and synaptic level, there is a growing debate as to whether some cells are specifically targeted for cholinergic neuromodulation. Therefore, understanding the spatial relationships between cholinergic fibers and their targets could yield important insight into disease states. Release of acetylcholine from synaptic terminals, or boutons, are thought to activate neurons through two different modes that depend on the spatial relationship to their neuronal targets. At one extreme, cholinergic transmission employs a "classical" mode in which a synapse is targeted to the immediate vicinity (within 50 nm) of membrane-bound receptors that are associated with the synapse. This mode, called point-to-point transmission, entails the release of ACh at a high concentration to activate synaptic receptors within 1 ms. At the opposite extreme, ACh is released at some distance, diffuses in 3-dimensions, and ultimately decays to a low concentration that activates high affinity receptors distributed along the surface of the neuron. This mode is called volumetric transmission (Vizi et al., 2004). These two modes are not mutually exclusive. However, the concentration of ACh that reaches a neuron depends upon the precise spatial arrangement and density of ACh release sites relative to the neuronal surface. Importantly, the enzyme acetylcholinesterase (AChE) rapidly degrades ACh in the extracellular space, which underscores the complex temporal and spatial relationships of ACh release sites relative to neurons. AChE inhibitors, the only current therapy for treatment of Alzheimer's Disease, block the degradation of ACh, thereby making volumetric transmission more efficacious by altering spatiotemporal aspects of diffusion between ACh release sites and postsynaptic neurons.

The spatial organization of ACh release sites relative to neuronal types is poorly understood. Using an antibody to the vesicular acetylcholine transporter (vAChT), there is some evidence that ACh release sites are preferentially clustered at distinct subtypes of neurons

(Dougherty and Milner, 1999; Henny and Jones, 2008; Ludkiewicz et al., 2002, 2000; Leranath and Frotscher, 1987), suggesting that the spatial arrangement of ACh release sites varies across cell types and between regions. In contrast, volumetric transmission could be defined as a random pattern of release sites that release ACh onto nonspecific postsynaptic targets. Therefore, by examining the spatial distribution of cholinergic boutons, principles could emerge that confer cell-type specificity to cholinergic transmission. Towards this end, we applied the Aggregation about a Mass function (A_M) to examine the spatial distribution of cholinergic boutons in the vicinity of hippocampal neurons. We find that this statistic can reveal targeting of boutons to the immediate vicinity of a neuron, distinguishing the empirical spatial distribution of these boutons from a random distribution.

5.2. Confocal imaging and image processing in Image J

Inhibitory interneurons in the hippocampus were visualized with GAD65-GFP mice (López-Bendito et al., 2004), a transgenic mouse line that labels distinct subpopulations of inhibitory interneurons with green fluorescent protein (GFP) (Wierenga et al., 2010). Cholinergic terminals were visualized with an antibody to the vesicular acetylcholine transporter (vAChT), similarly to previously described (Cea-del Rio et al., 2010). GAD65 GFP-positive neurons and vAChT-positive terminals were sequentially imaged using an Olympus Fluoview 1000 confocal microscope. All analyzed images were gathered using a 60× objective at 2× magnification. For the image stack in Figure 7A, ninety-three slices of 1024×1024 images were gathered at 0.5 micron z-step size. Images were exported to TIFF image formats, deconvolved with Huygens Essential Software (Scientific Volume Imaging Co., The Netherlands) and further processed using ImageJ software. To eliminate subjectivity and to maintain a uniformity of methodology across images, images first were thresholded using an automated Otsu threshold algorithm in both red and green channels, reducing the image data to binary. This image segmentation process limited the inclusion of background fluorescence in subsequent processing. Thresholded images were then counted at a minimum voxel size using the ImageJ plugin ObjectCounter3D. The minimum voxel size was calculated from a vAChT bouton size of 0.5 microns in diameter, which was based on empirical measures and available literature (Wouterlood et al., 2007, 2008). For the confocal stack in Figure 7A, 9737 objects (or “events”) were detected based on a minimum object size of 128 voxels. The x, y, and z coordinates of each of the 9737 objects were exported from Object-Counter3D as a text file. The somatodendritic domain of the GAD65-GFP+ neuron was isolated by setting the object size to a minimum of 2,500 voxels, which excluded smaller objects in the green channel. The isolated neuronal body (green channel) was then exported to MATLAB as a TIFF file.

5.3. Image processing within MATLAB and R

From this revised TIFF file, MATLAB was used to find the surface points of the neuron. Three text files containing image data were created, two of which were the neuron surface points and the object locations both in ordered triplets. The third is a 3-dimensional binary array indicating the voxels that intersect the neuron volume. (This text file is a vector where location within the vector preserves the 3-dimensional array indices.) The text files were then uploaded into the statistical program R and the Aggregation about a Mass function was implemented. Text files are provided with the supplementary materials.

In this application, the neuron is the mass and the 9737 objects are the events. The analysis shown in Figure 8 is from a single neuron. Inference is taken from the $\hat{B}_M$ function plotted with global envelopes in Figure 8. Global envelopes were created from 999 simulated study regions with point patterns under CSR. The $\hat{B}_M$ function wanders outside the shaded gray global envelope illustrating that the objects are significantly aggregated (p-value < .001) at distances roughly between 2 and 5 units from the neuron surface.

6. Discussion

The Aggregation about a Mass function and its application to neuroscience is a novel and useful method for determining aggregation about a mass. The simulation study included demonstrates that the Aggregation about a Mass function operates well when a mass is present in the study region. The simulation study also increases knowledge of how the A_M reacts to different kinds of spatial dependence upon a mass. If the Aggregation about a Mass function is used, the simulation study could be utilized to guide interpretation of results.

The $\hat{B}_M$ function and pointwise envelopes terminate at the zero value as seen in the many examples. As mentioned in section 4 and Figure 4 this is due to the scaling of $\hat{B}_M$ by n_e . Researchers should not “search” for aggregation about the mass near the maximum h distance. The maximum h distance should be chosen further than where aggregation is of interest.

The application to the Aggregation about a Mass function has demonstrated feasibility of the approach, that there is aggregation of ACh release points about a mass. Therefore, consistent with previous anatomical studies (Dougherty and Milner, 1999; Henny and Jones, 2008; Ludkiewicz et al., 2002, 2000; Leranthe and Frotscher, 1987), we have shown that the connectivity between cholinergic neurons and their cellular targets is non-random, that the cholinergic fibers preferentially make contact with the neuron. The natural progression from this study would be to look at multiple neurons and compare neurons that exhibit aggregation and randomness. This raises several points. First, although possible to detect in our simulations, we have not demonstrated whether neurons exist in which events (ACh release sites) are repulsed from the neuronal mass. We are planning to explore this question in a future study. Secondly, as the volume of study area is scaled up, there are inevitable complications that occur with multiple neurons in the study area. This is not studied in this paper, but the Aggregation about a Mass function could be adjusted for such study regions. Events would need to be assigned to a single mass or multiple masses depending on the strength of event to mass spatial dependence. This consideration of multiple masses is of importance, as existing images within our data set have multiple neurons within the study region. Original data were gathered by locating a single neuron and then imaging it and the surrounding area. Future work and imaging could further explore practical limits to the study area and find an ideal study region size. As of now the ideal study region size is one that contains a single neuron.

In the future, we would like to incorporate this method into existing computational frameworks, such as the RipleyGUI environment (Hansson et al., 2013). In addition, there

are opportunities to increase computational efficiency, either by streamlining or compiling the code or by making the code compatible with supercomputing resources.

There are assumptions in our approach that are clearly oversimplifications. For example, we assume that the release sites are point sources. Future work could incorporate complex structures to examine relative to a neuron. Finally, future work could incorporate diffusion of neurotransmitter from the release sites under different conditions (i.e. under circumstances in which ACh degradation has been blocked).

In conclusion, as spatial data within neuroscience are becoming more prevalent, our new methodology will provide an adequate tool for addressing spatial relationships and will uncover more in depth information when point processes become more complex.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was partially supported by from The University of Montana Department of Mathematical and Statistical Sciences (MOS) and National Institutes of Health grant R01 NS069689 (JIL). JB was supported by the Montana Integrative Learning Experience for Students (MILES) program, through a grant to the University of Montana from the Howard Hughes Medical Institute-Undergraduate Science Education Program (Grant No. 52005905), and a Center for Structural and Functional Neuroscience Summer Undergraduate Research Fellowship (NCRR grant P20RR015583). P20RR015583, P20RR017670 and P20GM10356 grants supported core imaging facilities.

Appendix A. Supplemental Materials

R Code

Two R files are included to accompany sections 4 and 5. These files will reproduce most figures found in these sections. Section 4 code will not reproduce simulations without additional coding, but should be used to explore how B_M reacts toward point patterns. (supplement chapter4.R, supplement chapter5.R)

R Functions

Three R functions needed to run R files. (ModKFun.R, plotK.R, ModKFunGAD.R)

Data

The three data files described in section 5.3 are included in the zipped file DataZip.zip. (Cell 1_1-5 event.txt, Cell 1_1-5 mass.txt, Cell 1_1-5 cell.txt)

Bibliography

- Baddeley A, Diggle PJ, Hardegen A, Lawrence T, Milne RK, Nair G. On tests of spatial pattern based on simulation envelopes. *Ecological Monographs*. 2014; 84(3):477–489.
- Bailey T, Gatrell A. Interactive spatial data analysis. Longman Scientific and Technical. 1995
- Cea-del Rio CA, Lawrence JJ, Tricoire L, Erdelyi F, Szabo G, McBain CJ. M3 muscarinic acetylcholine receptor expression confers differential cholinergic modulation to neurochemically

- distinct hippocampal basket cell subtypes. *The Journal of Neuroscience*. 2010; 30(17):6011–6024. [PubMed: 20427660]
- Cressie, N. Wiley series in probability and mathematical statistics: Applied probability and statistics. J. Wiley; 1993. Statistics for spatial data.
- Dougherty KD, Milner TA. Cholinergic septal afferent terminals preferentially contact neuropeptide y-containing interneurons compared to parvalbumin-containing interneurons in the rat dentate gyrus. *The Journal of neuroscience*. 1999; 19(22):10140–10152. [PubMed: 10559422]
- Eglen SJ, Wong JCT. Spatial constraints underlying the retinal mosaics of two types of horizontal cells in cat and macaque. *Visual neuroscience*. Mar; 2008 25(2):209–214. [PubMed: 18334045]
- Fajardo A, Goodburn JM, Graham J. Spatial patterns of regeneration in managed uneven-aged ponderosa pine/douglas-fir forests of western montana, {USA}. *Forest Ecology and Management*. 2006; 223(13):255–266.
- Hansson K, Jafari-Mamaghani M, Krieger P. Ripleygui: software for analyzing spatial patterns in 3d cell distributions. *Frontiers in Neuroinformatics*. 2013; 7(5)
- Hasselmo M, Giocomo L. Cholinergic modulation of cortical function. *Journal of Molecular Neuroscience*. 2006; 30(1):133–135. [PubMed: 17192659]
- Henny P, Jones BE. Projections from basal forebrain to prefrontal cortex comprise cholinergic, gabaergic and glutamatergic inputs to pyramidal cells or interneurons. *European Journal of Neuroscience*. 2008; 27(3):654–670. [PubMed: 18279318]
- Illian, J., Penttinen, A., Stoyan, H., Stoyan, D. Statistical analysis and modelling of spatial point patterns. Vol. 70. John Wiley & Sons; 2008.
- Jafari-Mamaghani M, Andersson M, Krieger P. Spatial point pattern analysis of neurons using ripley's k-function in 3d. *Frontiers in neuroinformatics*. 2010; 4:9. [PubMed: 20577588]
- Jenei A, Kormos J, Szentesi G, Veres AJ, Varga S, Bodnár A, Damjanovich S, Mátyus L. Non-random distribution of interleukin receptors on the cell surface. *Chemphyschem : a European journal of chemical physics and physical chemistry*. Jul; 2009 10(9–10):1577–1585. [PubMed: 19514033]
- Joyner ML, Seier E, Jones TC. Distances to a point of reference in spatial point patterns. *Spatial Statistics*. 2014; 10(0):63–75.
- Kellner RR, Baier CJ, Willig KI, Hell SW, Barrantes FJ. Nanoscale organization of nicotinic acetylcholine receptors revealed by stimulated emission depletion microscopy. *Neuroscience*. Jan; 2007 144(1):135–143. [PubMed: 17049171]
- Lagache T, Lang G, Sauvonnet N, Olivo-Marin J-C. Analysis of the spatial organization of molecules with robust statistics. *PLoS ONE*. 2013; 8(12):e80914. [PubMed: 24349021]
- Leranth C, Frotscher M. Cholinergic innervation of hippocampal GAD- and somatostatin-immunoreactive commissural neurons. *The Journal of comparative neurology*. Jul; 1987 261(1):33–47. [PubMed: 2887594]
- Loosmore NB, Ford ED. Statistical inference using the g or k point pattern spatial statistics. *Ecology*. 2006; 87(8):1925–1931. [PubMed: 16937629]
- López-Bendito G, Sturgess K, Erdélyi F, Szabó G, Molnár Z, Paulsen O. Preferential origin and layer destination of gad65-gfp cortical interneurons. *Cerebral Cortex*. 2004; 14(10):1122–1133. [PubMed: 15115742]
- Ludkiewicz B, Wójcik S, Spodnik E, Domaradzka-Pytel B, Mory J. Cholinergic endings on various neurones containing calcium binding proteins and glutamic acid decarboxylase in the hippocampus of the rat. *Folia morphologica*. 2000; 59(4):263–270. [PubMed: 11107697]
- Ludkiewicz B, Wójcik S, Spodnik E, Domaradzka-Pytel B, Mory J. Cholinergic innervation and calretinin-immunoreactive neurones in the hippocampus during postnatal development of the rat brain. *Folia morphologica*. 2002; 61(1):7–14. [PubMed: 11905254]
- Millet LJ, Collens MB, Perry GLW, Bashir R. Pattern analysis and spatial distribution of neurons in culture. *Integrative biology : quantitative biosciences from nano to macro*. Dec; 2011 3(12):1167–1178. [PubMed: 22057472]
- Ripley BD. The second-order analysis of stationary point processes. *Journal of applied probability*. 1976:255–266.
- Ripley BD. Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1977:172–212.

- Ruivo LMT-G, Mellor JR. Cholinergic modulation of hippocampal network function. *Frontiers in synaptic neuroscience*. 2013; 5
- Vizi ES, Kiss JP, Lendvai B. Nonsynaptic communication in the central nervous system. *Neurochemistry international*. Sep; 2004 45(4):443–451. [PubMed: 15186910]
- Wierenga CJ, Millner FE, Rinke I, Keck T, Stein V, Bonhoeffer T. Molecular and electrophysiological characterization of gfp-expressing cal interneurons in gad65-gfp mice. *PLoS ONE*. 2010; 5(12):e15915–e15915. [PubMed: 21209836]
- Wouterlood FG, Boekel AJ, Kajiwarra R, Beliën JA. Counting contacts between neurons in 3d in confocal laser scanning images. *Journal of neuroscience methods*. 2008; 171(2):296–308. [PubMed: 18471891]
- Wouterlood FG, Boekel AJ, Meijer GA, Beliën JA. Computer-assisted estimation in the cns of 3d multimarker overlap at the level of individual nerve endings: A confocal laser scanning microscope application. *Journal of neuroscience research*. 2007; 85(6):1215–1228. [PubMed: 17387746]

Figure 1.

Example of a point pattern with a mass included. The band (or buffer in 3D) is created by “searching” at distance h away from the mass surface. Only one band is shown in the figure, but in practice many are created as h varies. The Aggregation about a Mass function counts events within each band, and then compares that count with the expected number of events based on the area of each band.

Figure 2.

Point patterns used in simulations. Study region is 1 unit by 1 unit with a circular mass (radius of 0.1, centered at (0.5, 0.5).) Pattern 1 is a random point pattern, while Pattern 2 and 3 are aggregated patterns. It should be noted that in the point patterns seen here, the number of events (n) is 100. Some simulated point patterns used different values of n , but were created with the same type of random or aggregated fashion. These 3 point patterns are examples of point patterns used in the center graph ($n = 100$ case) of Figure 3.

Figure 3.
 B_M with different number of events used. As the number of events increases, the noise among the $\hat{B}_M$ functions and envelopes are dampened.

Figure 4. Comparing the $\hat{B}_M$ function with different choices of the maximum distance. Increasing the maximum distance with Pattern 3 inflates the peak magnitude of the $\hat{B}_M$ values.

Figure 5. Effect of invoking an edge effect while using the $\hat{B}_M$ function. Little to no bias is seen when an edge effect is invoked, indicating the edge correction of the Aggregation about a Mass function is stable.

Figure 6.

Effect of increased clustered patterns. As the point pattern becomes more clustered (more events near mass), the magnitude of the $\hat{B}_M$ increases.

Figure 7.

This neuron in particular was imaged with 93 slices separated by 0.5 μm . Each slice is then altered for clarity and then split in order to distinguish the neuron from objects (events). Locations of the neuron surface and objects are then taken from the respective image.

Figure 8.

Left: A 3-dimensional representation of a neuron and surrounding objects (events). The figure displays a sample of events (20%) in order to illustrate aggregation about the neuron. Right: The $\hat{B}_M$ function plotted with global envelopes (zone of constant width). Global envelopes were created from 999 simulated study regions. The solid black $\hat{B}_M$ line wanders outside the gray envelope indicating significant aggregation (p-value < .001) at distances roughly between 2 and 5 units from the neuron surface.