Mathematical Statistics
Stockholm University

Spatial Point Pattern Analysis of Neurons Using Ripley's K-function and Voronoi Tessellation

Mehrdad Jafari Mamaghani

Examensarbete 2008:16

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se

Mathematical Statistics
Stockholm University
Examensarbete **2008:16**,
http://www.math.su.se

Spatial Point Pattern Analysis of Neurons Using Ripley's K-function and Voronoi Tessellation

Mehrdad Jafari Mamaghani*

November 2008

Abstract

The aim of this project is to apply and develop methods for statistical analysis of spatial point patterns. Spatial point pattern analysis is widely used within biological fields of inferential statistics. This text is constructed upon applications and developement of such analysis on distribution of neurons. Unknown distributions in statistics are in principle investigated using non-parametric tools. Two such tools within the spatial point patterns' field are Ripley's K-function and Voronoi Tessellation. These methods have widely been used to study the 2-dimensional distribution of biological phenomena in the past decades. Confocal microscopy has now given the possibility of acquiring data for expanding these studies to 3-dimensional domains and thus attaining more information. An authentic study in this case requires development of consistent tools. The tool chosen to develop here is Ripley's K-function and its edge correction term for operations in 3-dimensional domains. The operability of this function along with its corresponding function in 2D, and Voronoi tessellation is confirmed by different types of simulations. These methods are later used to investigate the distribution of neurons in samples obtained from a mouse brain.

KEY WORDS: *Ripley's K-function, Edge Correction in 3D, Voronoi Tessellation, Homogenous Poisson Process, Poisson Cluster Process, Simple Poisson Inhibition Process.*

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91 Stockholm, Sweden. E-mail: mjm@6kye.com. Supervisor: Mikael Andersson.

Foreword and Acknowledgements

This paper constitutes a Master's thesis worth of 30 ECTS credit points in mathematical statistics at the Mathematical Institution of Stockholm University. The project has been ordered by Patrik Krieger at the Department of Neuroscience, Karolinska Institute.

I would like to thank my supervisors Patrik Krieger at Karolinska Institute for investing his trust in me and for introducing me to this fascinating field of biostatistics, and indeed Mikael Andersson at Stockholm University for his constant, cordial and enlightening assistance throughout the entire project. Also thanks to all my friends and other people who've made this process more exciting and colorful.

Finally, I dedicate an immense amount of gratitude to my parents for the lifelong support and the infinite inspiration.

Mehrdad Jafari Mamaghani

Contents

1 Introduction

In the present paper we have used statistical methods to analyze the 3-dimensional distribution of neurons. The distribution of neurons can be described as a spatial point pattern, a multi-dimensional stochastic process. In some brain diseases the normal organization of neurons appears to be changed. This reorganization of neurons can many times be subtle, and thus improved statistical methods are necessary to adequately analyze the $3D$ distribution of neurons.

To analyze the distribution of neurons we have used non-parametric methods such as Ripley's K-function and analysis of Voronoi tessellation. Much of our focus has been to develop a K-function for 3-dimensional observations. With the technological advancement of confocal microscopy, large amounts of $3D$ data sets can now be obtained. This demands a revision of earlier methods used to analyze cell distributions. The particular challenge in this case has been the correction of 3-dimensional edge effects in Ripley's K-function, which is later described in details.

The samples in this case are taken from the layer 5 in neocortex of a mouse brain. The neocortex is layered from 1 to 6 and is the largest portion of the cerebral cortex, the thin layer of gray matter on the surface of the cerebral hemisphere that develops from the telencephalon and folds into gyri. The cerebral cortex reaches its highest development in man and is responsible for intellectual faculties and higher mental functions.
The samples are 50-100 μm thick. Since the analyzed neurons have a cell diameter of 15-20 μm the process of edge correction is sensitive to false geometrical assumptions. One other challenge here is the non-stationary nature of a few of our samples.

Our goal here is the following:

- To perform a qualitative investigation of our obtained samples using the mathematical tools mentioned above,
- Introducing Ripley's K-function to analysis of point patterns in 3-dimensional domains,
- Comparing Ripley's K-function and Voronoi tessellation's operability in $2D$,
- Using simulations of different types of Poisson process to monitor the differences between different point patterns.

The entire computational process has been implemented in MatLab from the very first step to the last one.

The following section illustrates the theoretical background of the methods used to investigate the spatial point patterns present among the neurons. The results obtained by simulation are clarified in Section 3. Section 4 represents the results acquired using the mentioned methods and finally in the Discussion section we put our statistical inferences into words and discuss possibilities of further improvements.

2 Theoretical Design

A spatial point pattern can be easily seen as the expansion of a one-dimensional stochastic point process to higher dimensions. A very intuitive example of such expansion is the following of a Poisson process; a one-dimensional Poisson process with intensity λ demonstrates the frequency of points occurring along a line or an axis. Likewise would a spatial Poisson process in $2D$ demonstrate the frequency of occurring points on a plane where the points are spread across the area of the plane domain. Note that we hereafter use the term event when we refer to the point positions in the point process.

The distribution of different sorts of spatial point processes is usually classified in three main classes (Diggle 2003):

- Aggregation (clustering)
- Regularity
- Complete spatial randomness (CSR)

Note that this classification is more of a way of illustrating the major differences between different distributions and not a definitive one.

Aggregation is a state of randomness where there is a significant clustering trend. As a biological phenomenon, this usually occurs where parent events produce offsprings in a neighbouring domain of their own. Whether it is the spread of a certain plant or the reproduction of a bacteria.

Regularity corresponds to a state of randomness where the events are located in a more or less uniform manner. In such state there are observable distances between most of the events. This behaviour could best be described in terms of biological phenomena (trees) competing for nutrition.

CSR corresponds to a state of randomness that could be described by a homogenous Poisson process where the number of events in any domain D

with area $|A|$ or volume $|V|$ follows a Poisson distribution with mean $\lambda|A|$, $\lambda|V|$ and when given a number of events they're independent and identically distributed according to a uniform distribution (Diggle 2003).

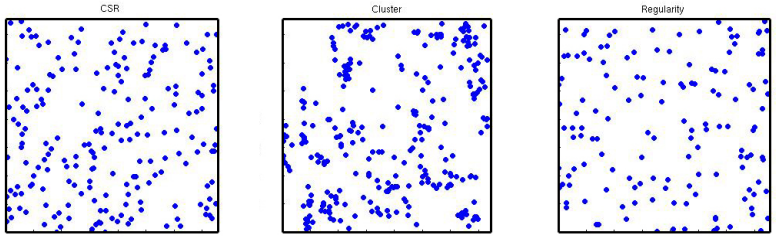

Figure 1: *The three main types of spatial point patterns*

2.1 Poisson Process

The Homogenous Poisson Process

The homogenous Poisson process is at the very structural core of spatial point patterns' theory. The properties of the homogenous Poisson process are identical with those of CSR and it is therefore a tool of great strength in spatial point pattern analysis. Diggle (2003) states the following postulates for the homogenous Poisson process (in $2D$):

- For some $\lambda > 0$, and any finite planar region A, $N(A)$ follows a Poisson distribution with mean $\lambda|A|$.
- Given $N(A) = n$, the n events in A form an independent random sample from the uniform distribution on A.
- For any two disjoint regions A and B, the random variables $N(A)$ and $N(B)$ are independent.

Poisson Cluster Process

Introduced by Neyman and Scott (1958), the Poisson cluster process is used for modelling spatially *clustered* point processes. The first step in the construction of a Poisson cluster process is to generate a number of so called parent events using a homogenous Poisson process with an intensity parameter ρ. There are then a random number of so called offspring put into the neighbourhood of each parent. The number of offspring for every parent is an independent and identically distributed

outcome of a random variable S. The distances at which offspring lie from their corresponding parent, are then independent and identically distributed with a random variable D. The distance directions, too, are indepedent and identically distributed.
In order to conduct a simulation, we therefor use a random Poisson process to generate the number of offspring per parent (that is $S \sim Po(\cdot)$), and determine their Euclidean distance according to the random variable $D \sim \exp(\cdot)$. The directions are based on outcomes of random variables $\varphi \sim U(-\pi, \pi)$ and $\phi \sim U(-\frac{\pi}{2}, \frac{\pi}{2})$ so that the $3D$ parent-to-offspring Euclidean distances follow the set:

$$\begin{aligned} x &= D \sin\varphi \cos\phi \\ y &= D \sin\varphi \sin\phi \\ z &= D \cos\varphi. \end{aligned}$$

Naturally, the clustering pattern would depend on the number of parents and their offsprings in general and the distance varible D in particular. Also the total intensity λ_c would be the result of the mean number of parent events and the mean number of offsprings. That is:

$$\lambda_c = \rho \cdot E[S] \tag{1}$$

Simple Poisson Inhibition Process

The simple Poisson inhibition process (hereafter referred to as Poisson Inhibition process) is a process corresponding to the properties of a regularly distributed spatial point process. This process can be first structured as a simple Poisson process with intensity τ which is then made 'regular' by eliminating all pairs of events lying within a distance δ or less from each other. The value of δ is essential to the outcome of the inhibition process and the magnitude of regularity. A low value of δ would generate a process with only slight differences from a homogenous Poisson process and a high value of δ would demonstrate high regularity and hence less events (see **7.2**). The intesity of events λ_i in Poisson inhibition process is (Diggle, 2003):

$$\lambda_i = \tau \exp(-\pi\tau\delta^2) \tag{2}$$

that is the initial intensity τ times the probability that an arbitrary event survives, $\exp(-\pi\tau\delta^2)$.

2.2 Stationarity

One significant assumption before entering the field of non-parametric methods (Ripley's K-function and Voronoi Tessellation) is the property of stationarity, which is simply invariance to translation and rotation. Non-stationarity *in our case* is a state of spatial distribution where the events are scattered in a certain stream throughout the sample. Rather than covering the entire sample randomly, the distribution of events leaves 'empty spaces' in the sample.
The greater deviation from this assumptions, the more biased the conclusions will be.

In practice, it is usually impossible to obtain samples that display total stationarity. In order to reach maximum stationarity without eliminating any observations and due to absence of non-stationarity in one dimension of our samples, we choose to apply the following rotation matrices in $2D$ (see **7.3** for rotation matrices in $3D$):

$$R_{cw}(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} \text{ and } R_{ccw}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

where R_{cw} is the clockwise and R_{ccw} the counter-clockwise rotation. θ is the angle obtained by length of the diagonal stream constituted by the events, and the longest edge of the initial sample domain[1].

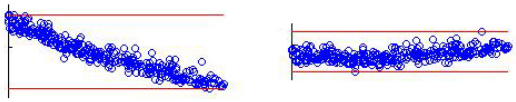

Figure 2: *How the matrices operate in* $2D$

[1]This interpretation is not a conclusive one as there are other versions of these rotation matrices. The ones mentioned above are, however, the ones used in the computational procedure of this project and have been proven to be correct. See the corresponding post in References for further reading.

2.3 Ripley's K-function

Ripley's K-function is an intuitive approach for detecting deviations from general assumptions inside or within samples. It has a non-parametric nature and is therefore a natural first step in investigating spatial point patterns due to its independence of initial assumptions regarding the properties of the point pattern. K-function, regardless of dimensional parameters, can simply be stated as

$$K(t) = \frac{E[\#\textit{Cells within a distance } t \textit{ of an arbitrary cell }]}{\textit{The total cell intensity}}$$

The intensity of events in a domain is usually denoted as λ and is estimated by the ratio $\frac{n}{|A|}$ in $2D$ and $\frac{n}{|V|}$ in $3D$ where n is the number of events in a sample domain of area $|A|$ or volume $|V|$.
Throughout the text, we choose to plug in the estimations instead of using the notation λ as a distinguishing characteristic between the fomulae in $2D$ and $3D$.

The way in which Ripley's K-function operates can be described as screening the number of events in a surrounding circle/sphere of radius t where an arbitrary event i resembles its center. Now if the intensity of observed events in this neighbourhood is more than the unit overall intensity (clustering), $K(t)$ will increase *faster* than the expected rate under CSR, $\lambda|A|$ in $2D$ or $\lambda|V|$ in $3D$. Vice versa, a weaker presence of events would cause $K(t)$ to increase *slower* than the expected rate under CSR (regularity). In a hypothetical state of complete spatial randomness (CSR), $K(t)$ would increase with the exact same rate as $\lambda|A|$ or $\lambda|V|$.
How this property is used is described in details below.

2.3.1 Ripley's K-function in $2D$

The following is Ripley's K-function for an arbitrary event i (Ripley 1976):

$$\widehat{K}_i(t) = |A| \cdot \frac{\sum_{j \neq i}^{n} e_i(t)^{-1} I[D(i,j) \leq t]}{n} \tag{3}$$

where A is the area of the domain of interest, n is the number of cells in the sample, $e_i(t)$ is the edge correction term for a circle of radius t with cell (event) i as its center. $I[\cdot]$ is the indicator function and $D(i,j)$ represents the Euclidean distance from cell i to cell j and hence a matrix of n rows and n columns where the diagonal is neglected simply because it represents the

distance from cell i to itself, that is 0.
The general $\widehat{K}(t)$ for an entire sample is:

$$\widehat{K}(t) = \frac{\sum_i^n \widehat{K}_i(t)}{n}. \tag{4}$$

Consistently, an unbiased expected value for the K-function under CSR is:

$$E[\widehat{K}(t)] = \pi t^2. \tag{5}$$

Furthermore, a natural way of detecting K-function's deviations from its expected value under CSR would be to create the function $\widehat{K}(t) - \pi t^2$.
We also choose to use the expression $\widehat{K}(t)/\pi t^2$ to monitor the magnitude of deviation in a more apparent way.
The stronger the hypothesis of complete randomness is in the sample studied by the K-function, the less departure we shall expect from its expected value. Finally to create a single K-function for different samples in a group, assuming homogenous samples, we assemble $\widehat{K}_{group}(t) = \frac{\sum \widehat{K}(t)}{\#\text{ samples in a group}}$.

2.3.2 *Edge Correction in 2D*

The edge correction term $e_i(t)$ is used to adjust the value of $K(t)$ when a circle of radius t in $2D$ expands beyond the boundaries of the sample domain. Not taking this into account would result in an underestimation of event intensity for large values of t ($t_{max} = \frac{\text{shortest edge}}{2}$).
The edge correction approach we've preferred to use here is Ward & Ferrandino's global correction from 1999 (Marcon and Puech 2003). This formula was chosen after a close investigation of results obtained by Ripley's and Besag's element-wise correction algorithms. The aim is to achieve the most apparant convergence towards the corresponding expected values.

$$e(t) = 1 - \frac{4}{3\pi}\left(\frac{t}{L} + \frac{t}{W}\right) + \left(\frac{11}{3\pi} - 1\right)\left(\frac{t^2}{LW}\right)$$

where L and W are the rectangular dimensions of the sampling domain.
Another strength in Ward & Ferrandino's correction is its global nature and hence swift and computationally cheap evaluation. Using the global term $e(t)$ instead of the element-wise $e_i(t)$ would simplify $K_i(t)$ into:

$$\widehat{K}_i(t) = |A| \cdot \frac{\sum_{j \neq i}^n I[D(i,j) \leq t]}{n \cdot e(t)}.$$

2.3.3 Ripley's K-function in $3D$

An extension of K-function from $2D$ to $3D$ is easily obtained by assuming 3-dimensional measures instead of those in $2D$. That is replacing the area parameter A by volume V and redefining the distance matrix D for 3-dimensional inter-event distances.

$$\widehat{K}(t) = |V| \cdot \frac{\sum_{i=1}^{n} \sum_{j \neq i} e_i(t)^{-1} I[D(i,j) \leq t]}{n^2}. \tag{6}$$

Likewise the expected value under CSR is

$$E[\widehat{K}(t)] = \frac{4\pi t^3}{3}. \tag{7}$$

The function $\widehat{K}_{group}(t)$ follows the same pattern as above.

2.3.4 *Edge Correction in 3D*

The edge correction term in $3D$ monitors the proportion of spherical volumes outside the sample boundaries. The challenge upon us here is of a complete different nature than the corresponding one in $2D$. Considering the shape of our samples in $3D$ and their nearly negligible thickness, it is obvious that a function operating as Ripley's K, demands paramount carefulness regarding its correction component. That is, the demand to evaluate these corrections at already small values of t and taking into account the geometrical characteristics in $3D$.

In order to calculate the K-function in $3D$ we need to observe the number of cells within a sphere of radius t with an arbitrary cell i as its center. It is easy to imagine that these expanding spheres will start to grow beyond the boundaries of the sample domain at a very early stage due to the slim thickness of our samples. The correction term in such case represents the proportion of volume inside the sample domain boundaries.
Our first step is to study the volume proportions of spheres exceeding merely one boundary and thus cut by one single plane.

There are several analytical formulas for calculating the semi-spheres or so called 'caps' illustrated in Figure 3. The equation used here is

$$V = \frac{(\pi h)(3r^2 + h^2)}{6} \tag{8}$$

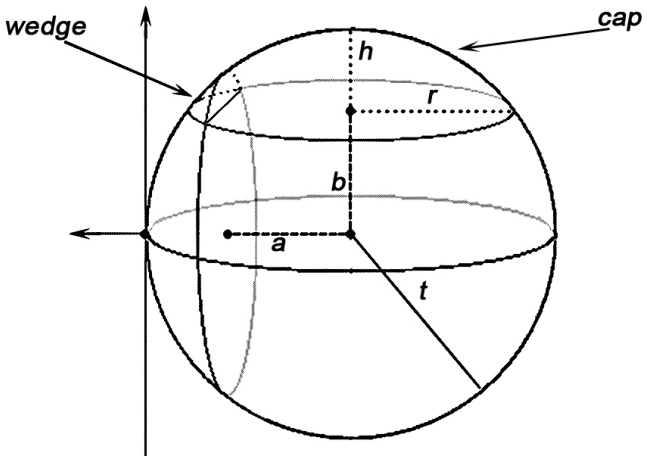

Figure 3: *Caps and wedges*

where h is the height of the spherical cap and r is its base.
We should also keep in mind the possibility of spheres intercepting two parallel planes for large values of t.

To confine the correction component solely to spheres cut by a single plane is however far from sufficient. This limitation would demand an extensive truncation of observations in order to evaluate Ripley's K-function in a mathematically consistent way. Thus would a second step be to evaluate the volume proportions of spheres intercepting two perpendicular planes.

Here we are once again interested in calculating the volume of semi-spheres lying outside the sample domain. This is as mentioned above calculated in a simple manner. As it appears in Figure 3 above there are intersections between every pair of semi-spheres. This means that without taking the intersection into account, we'll have an overestimation of the proportion lying outside the domain if we simply add the volumes of the semi-spheres. We are thus interested in calculating the volume of this intersection (wedge). These wedges lack a general symmetry[2] and we therefore need to employ methods of integration to evaluate their volumes.

There are then events in the corners of the sample domain where their spherical neighbourhood is cut by three planes. To evaluate their corresponding volume proportions seems to be the natural step in the construction of our edge correction term. This is, however, computationally expensive and time demanding. As a measure of precaution we choose to truncate the events in the corners of the domain to firstly observe their population in regard

[2]The 'base' and the 'height' of the wedge are dependent on the position of the event.

to the total number of events, and secondly their effect on $K(t)$ based on simulations. We find out that these corner events compose, in average, 5% of the total event population (where $t_{max} = \frac{\text{shortest edge}}{2}$) and their absence has no visible effect on the outcome of $K(t)$.
As a result, we introduce the following as our edge correction term in $3D$:

$$\begin{aligned} e_i(t) &= 1 - \frac{c_i(t) - w_i(t)}{4\pi t^3/3} \\ &= 1 - \frac{\{{c_i}^x(t) + {c_i}^y(t) + {c_i}^z(t)\} - \{{w_i}^{xy}(t) + {w_i}^{xz}(t) + {w_i}^{yz}(t)\}}{4\pi t^3/3}. \end{aligned}$$

Where ${c_i}^d(t)$ is the outside-of-sample volume of the sphere with radius t intercepted by plane d and ${w_i}^{d_1 d_2}$ is the volume of the wedge caused by the intersection of two semi-spheres defined on planes d_1 and d_2.

To simplify this we introduce the following notations for an event i with coordinates (x_i, y_i, z_i):

$$\begin{aligned} dx_l &= |x_i - x_l| \\ dx_u &= |x_i - x_u| \\ dy_l &= |y_i - y_l| \\ dy_u &= |y_i - y_u| \\ dz_l &= |z_i - z_l| \\ dz_u &= |z_i - z_u| \end{aligned}$$

where x_l, for instance, represents the lower domain boundary along the x-axis and x_u its corresponding upper boundary. Then

$$\begin{aligned} {c_i}^x(t) &= {c_i}^{x_l}(t) + {c_i}^{x_u}(t) \\ &= \frac{\pi(t - dx_l) \cdot (3(t^2 - {dx_l}^2) + (t - dx_l)^2)}{6} \\ &+ \frac{\pi(t - dx_u) \cdot (3(t^2 - {dx_u}^2) + (t - dx_u)^2)}{6} \end{aligned}$$

Each ${w_i}^{\cdot\cdot}(t)$ is then constituted as it follows:

$$
\begin{aligned}
{w_i}^{xy}(t) &= {w_i}^{x_l y_l}(t) + {w_i}^{x_l y_u}(t) + {w_i}^{x_u y_l}(t) + {w_i}^{x_u y_u}(t) \\
{w_i}^{xz}(t) &= {w_i}^{x_l z_l}(t) + {w_i}^{x_l z_u}(t) + {w_i}^{x_u z_l}(t) + {w_i}^{x_u z_u}(t) \\
{w_i}^{yz}(t) &= {w_i}^{y_l z_l}(t) + {w_i}^{y_l z_u}(t) + {w_i}^{y_u z_l}(t) + {w_i}^{y_u z_u}(t).
\end{aligned}
$$

The term ${w_i}^{xy}(t)$ can further be dissected into:

$$
\begin{aligned}
{w_i}^{xy}(t) &= {w_i}^{x_l y_l}(t) + {w_i}^{x_l y_u}(t) + {w_i}^{x_u y_l}(t) + {w_i}^{x_u y_u}(t) \\
&= \int_{dx_l}^{\sqrt{t^2 - {dy_l}^2}} \int_{-\sqrt{t^2-({dx_l}^2+{dy_l}^2)}}^{+\sqrt{t^2-({dx_l}^2+{dy_l}^2)}} (\sqrt{t^2-(u^2+v^2)} - dy_l) du dv \\
&+ \int_{dx_l}^{\sqrt{t^2 - {dy_u}^2}} \int_{-\sqrt{t^2-({dx_l}^2+{dy_u}^2)}}^{+\sqrt{t^2-({dx_l}^2+{dy_u}^2)}} (\sqrt{t^2-(u^2+v^2)} - dy_u) du dv \\
&+ \int_{dx_u}^{\sqrt{t^2 - {dy_l}^2}} \int_{-\sqrt{t^2-({dx_u}^2+{dy_l}^2)}}^{+\sqrt{t^2-({dx_u}^2+{dy_l}^2)}} (\sqrt{t^2-(u^2+v^2)} - dy_l) du dv \\
&+ \int_{dx_u}^{\sqrt{t^2 - {dy_u}^2}} \int_{-\sqrt{t^2-({dx_u}^2+{dy_u}^2)}}^{+\sqrt{t^2-({dx_u}^2+{dy_u}^2)}} (\sqrt{t^2-(u^2+v^2)} - dy_u) du dv
\end{aligned}
$$

For example, the volume of the wedge illustrated in Figure 3 is

$$
V_{wedge} = \int_{a}^{\sqrt{t^2-b^2}} \int_{-\sqrt{t^2-(a^2+b^2)}}^{+\sqrt{t^2-(a^2+b^2)}} (\sqrt{t^2-(u^2+v^2)} - b) du dv.
$$

Further dissection of ${w_i}^{xz}(t)$ and ${w_i}^{yz}(t)$ follows the same pattern as above.

Thus each $e_i(t)$ could contain up to 12 integral expressions (wedge volumes) and 6 analytical expressions (semi-sphere volumes), for each value of t.
Due to the immense amount of calculations and its local (element-wise) nature, the edge correction procedure in $3D$ is much more time demanding than the corresponding (global) one in $2D$. More of this is discussed in **5.2**.

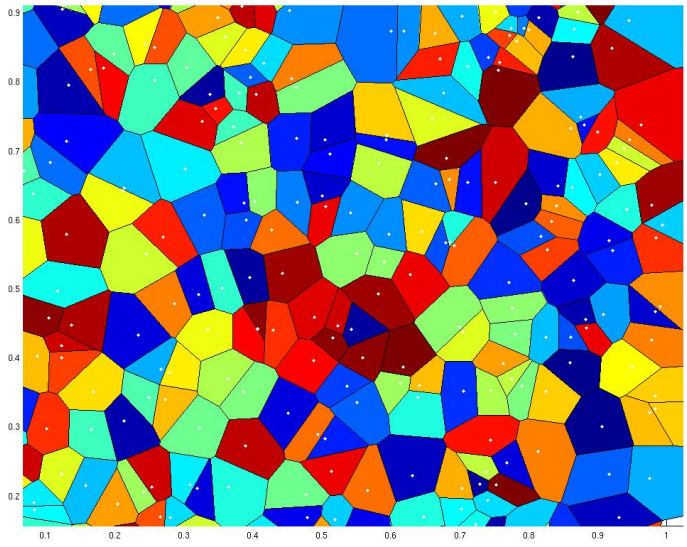

Figure 4: *Realization of Voronoi tessellation on CSR*

2.4 Voronoi Tessellation

Named after Georgy Voronoi, a Voronoi tessellation (also called a Voronoi decomposition or a Dirichlet tessellation) is a decomposition of a region prescribed by distances to a set of events in the region (see Duyckaerts et al. 2001). The region in this case is the 2-dimensional version of our sampling domain, which is obtained by eliminating the thickness dimension. Such decompositions result in sets of polygons associated with the events so that every arbitrary point inside any polygon lies closest to the associated event than any other neighbouring event.

A property that the Voronoi tessellation shares with Ripley's K-function is its non-parametric nature and as mentioned above it is a natural choice when we are dealing with samples without any knowledge of their parametric family.

A big difference, however, is the way in which we draw inferences from the tessellation. The resulting polygons are as expected variable in the size of their areas and this depends naturally on the intensity of other events in their neighbourhood. Thus calculating the polygon areas and the deviation among them would be an intuitive way to characterise the existing point pattern. The expression below has been used to evaluate the polygon areas:

$$A_{polygon} = \frac{1}{2}\sum_{i=0}^{n-1}(r_i s_{i+1} - r_{i+1}s_i)$$

Where (r_i, s_i) represent the coordinates for the ith vertex in an arbitrary polygon with n vertices numbered *clockwise* and where $(r_0, s_0) = (r_n, s_n)$.

Now to draw inferences from these areas we'll deploy the cofficient of variation. An approach that has been applied frequently in the majority of articles dealing with Voronoi tessellaton as of the writing moment.

The coefficient of variation (CV) is a normalised measure of dispersion that is obtained by creating a ratio with the standard deviaton as the numerator and the mean as the denominator (hence the normalisation).

With a clustering trend as the most plausible one, the CV tends to be high in comparison to the CV in a state of CSR. Correspondingly the CV tends to be in comparison to CSR lower with a regular pattern as the underlying factor (see Results 3.2 for more details).

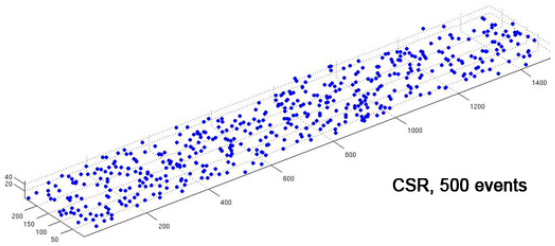

Figure 5: *Simulation of a homogenous Poisson process (CSR). Dimensions* $x = 1500\mu m$, $y = 250\mu m$, $z = 50\mu m$

3 Simulations

We have executed three major types of simulations. The simulations of homogenous Poisson process, Poisson cluster process and Poisson inhibition process. The practical and computational approach in creating these simulations is consistent with the corresponding descriptions given earlier.
Common parameters for all simulations are the dimensional lengths of the domain, which on average follows the same size as the real samples. The Poisson process is later on simulated by generating 500 events in the 3-dimensional domain. The Poisson cluster process is designed to take in more variability in form of parent and offspring intensity, and the clustering intensity determined by the random parent-to-offspring Euclidean distances. The Poisson inhibition process is variable in number of events to be initially generated, τ, and the least permissible distance variable, δ. The latter one has a great effect on the outcome of the simulations and needs extra carefulness in choice of a suitable value. This is discussed in **5.4** and illustrated in **7.2** and **7.4**.
A few samples of these simulations in $3D$ with different input parameters are given in Figures 5-7. An illustration of simulations in $2D$ is represented in Figure 1.

The parameters associated with the Poisson cluster process and the Poisson inhibition process are chosen to reach an intensity equivalent to the intensity in CSR according to (1) and (2).
In the following figures ρ, γ and ϵ are associated with the number of parents, offspring and the distance parameter respectively (where $S \sim Po(\gamma)$ and $D \sim \exp(\epsilon)$).

In order to conduct Ripley's K-function and Voronoi tessellation in $2D$, the 'thickness dimension' represented by z has been eliminated. The follow-

ing simulation results are used as a background for our interpretation of the mouse brain samples.

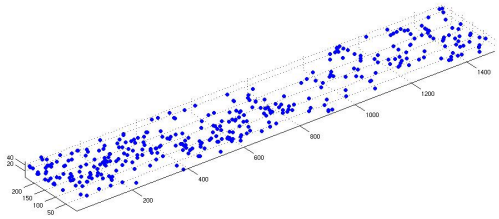

Figure 6: *Simulation of a Poisson cluster process with* $\rho = 50$, $\gamma = 10$, $\epsilon = 100$. *Dimensions* $x = 1500\mu m$, $y = 250\mu m$, $z = 50\mu m$

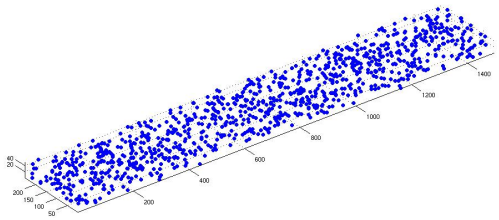

Figure 7: *Simulations a Poisson inhibition process with 1000 initial events and* $\delta = 0.0148$. *Dimensions* $x = 1500\mu m$, $y = 250\mu m$, $z = 50\mu m$

3.1 Ripley's K-function

3.1.1 CSR

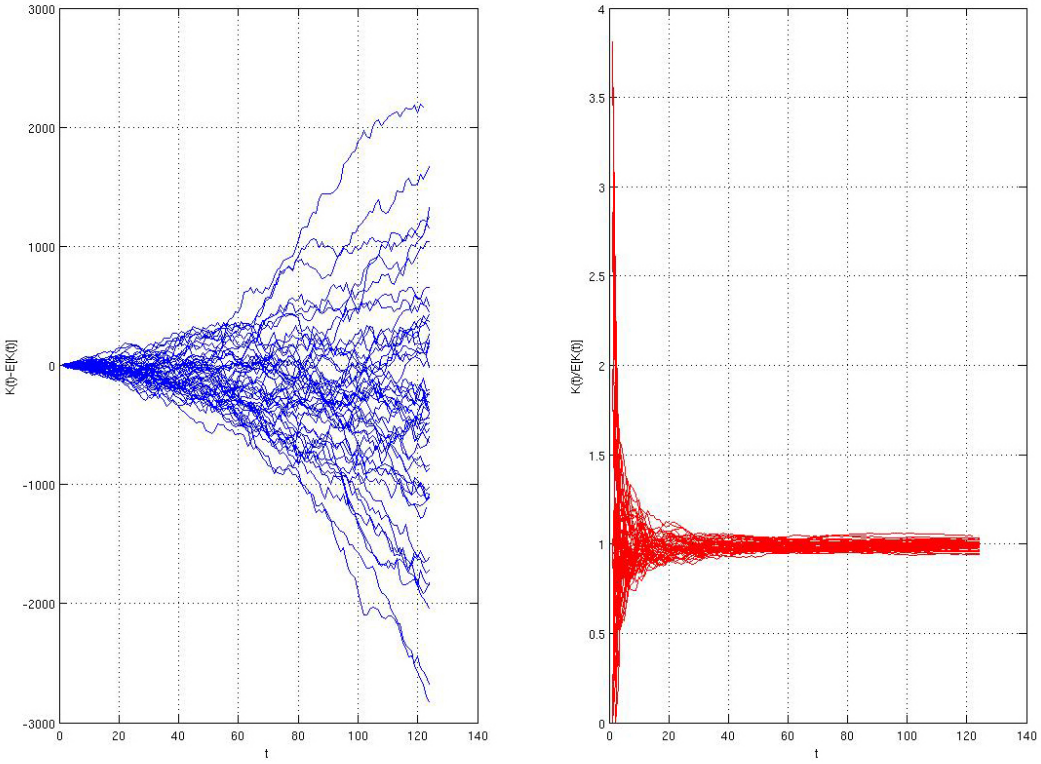

Figure 8: *$K(t) - \pi t^2$ and $K(t)/\pi t^2$ in 2D for 50 simulations of CSR with 500 events in each simulation*

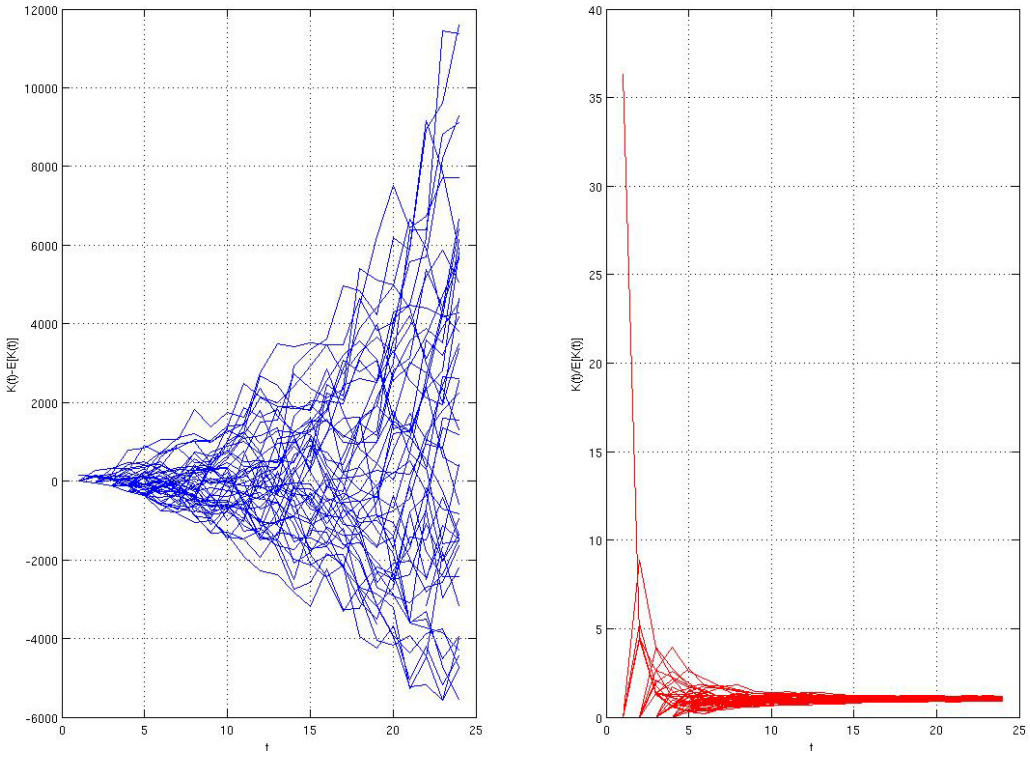

Figure 9: $K(t) - \frac{4\pi t^3}{3}$ *and* $K(t)/\frac{4\pi t^3}{3}$ *in* $3D$ *for 50 simulations of CSR with 500 events in each simulation*

3.1.2 Poisson Cluster Process

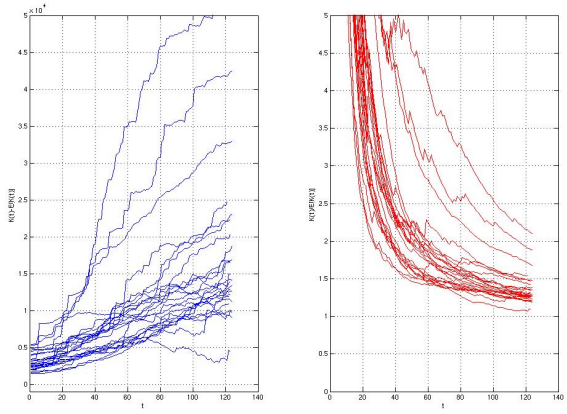

Figure 10: $K(t) - \pi t^2$ *and* $K(t)/\pi t^2$ *in* $2D$ *for 50 simulations of Poisson cluster process with* $\rho = 50$, $\gamma = 10$, $\epsilon = 100$

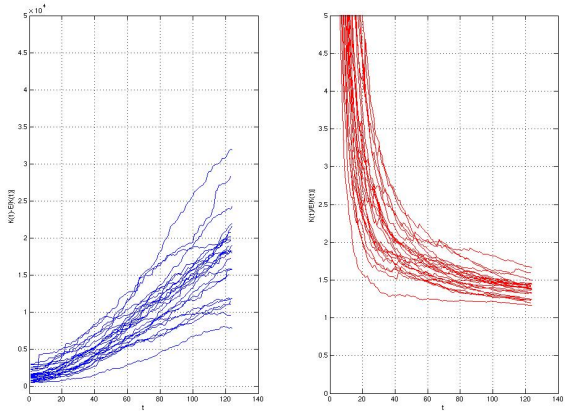

Figure 11: $K(t) - \pi t^2$ *and* $K(t)/\pi t^2$ *in* $2D$ *for 50 simulations of Poisson cluster process with* $\rho = 20$, $\gamma = 25$, $\epsilon = 175$

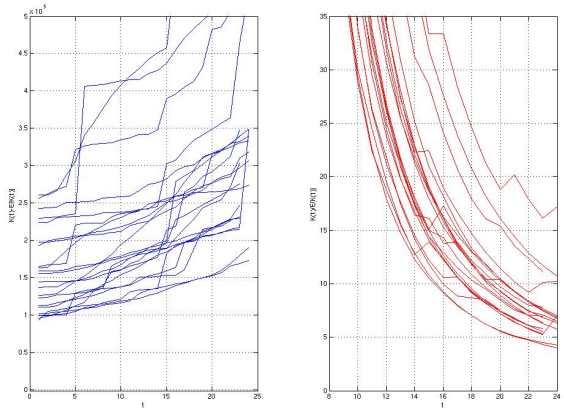

Figure 12: $K(t) - \frac{4\pi t^3}{3}$ *and* $K(t)/\frac{4\pi t^3}{3}$ *in* $3D$ *for 25 simulations of Poisson cluster process with* $\rho = 50$, $\gamma = 10$, $\epsilon = 100$

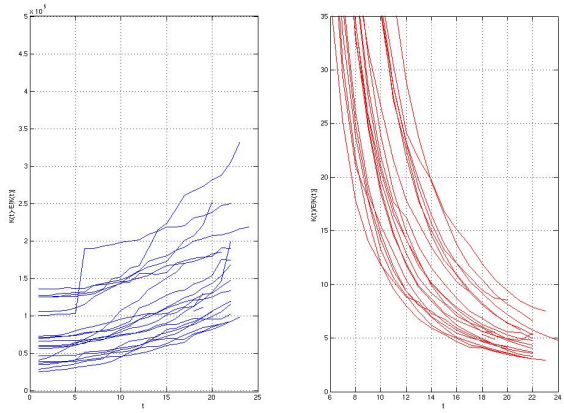

Figure 13: $K(t) - \frac{4\pi t^3}{3}$ *and* $K(t)/\frac{4\pi t^3}{3}$ *in* $3D$ *for 25 simulations of Poisson cluster process with* $\rho = 20$, $\gamma = 25$, $\epsilon = 175$

3.1.3 Poisson Inhibition Process

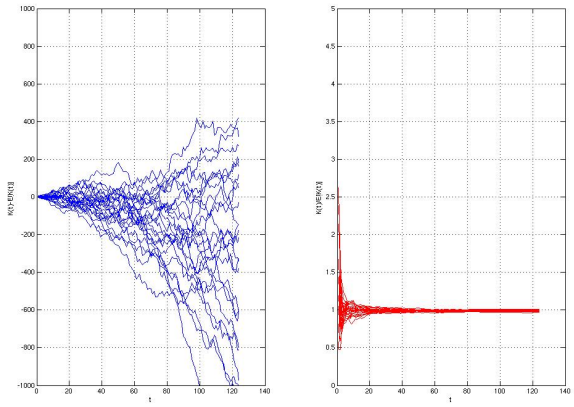

Figure 14: $K(t) - \pi t^2$ *and* $K(t)/\pi t^2$ *for 50 simulations of Poisson inhibition process with 1000 initial events and* $\delta = 0.0148$

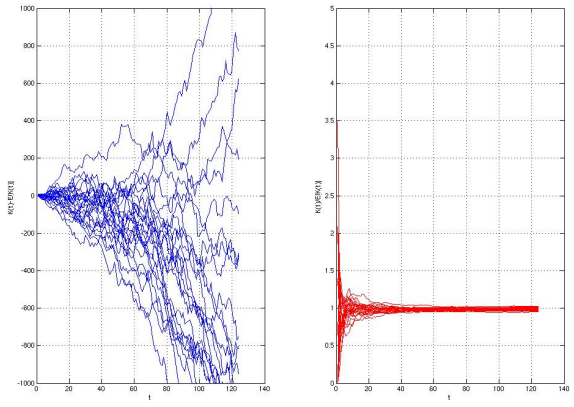

Figure 15: $K(t) - \pi t^2$ *and* $K(t)/\pi t^2$ *for 50 simulations of Poisson inhibition process with 585 initial events and* $\delta = 0.0093$

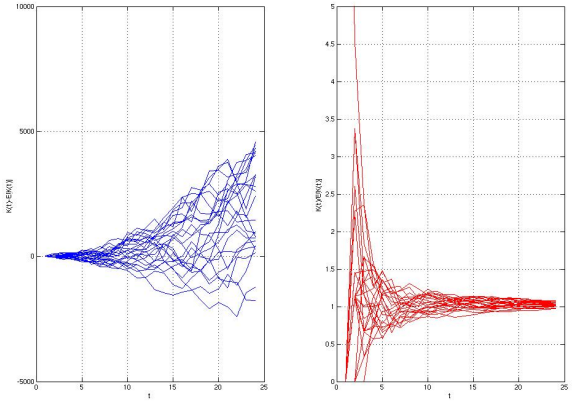

Figure 16: $K(t) - \frac{4\pi t^3}{3}$ *and* $K(t)/\frac{4\pi t^3}{3}$ *in* $3D$ *for 25 simulations of Poisson inhibition process with 1000 initial events and* $\delta = 0.0148$

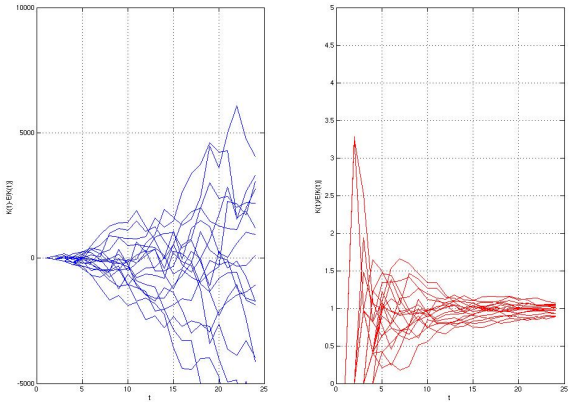

Figure 17: $K(t) - \frac{4\pi t^3}{3}$ *and* $K(t)/\frac{4\pi t^3}{3}$ *in* $3D$ *for 25 simulations of Poisson inhibition process with 585 initial events and* $\delta = 0.0093$

3.1.4 Comments

Figure 8 representing $K(t) - \pi t^2$ and $K(t)/\pi t^2$ under CSR in $2D$ and Figure 9 representing $K(t) - \frac{4\pi t^3}{3}$ and $K(t)/\frac{4\pi t^3}{3}$ under CSR in $3D$, share the same appearance as the plots to the left have a uniform spread towards positive and negative values as t grows, as expected under CSR. The plots to the right demonstrate solid convergence towards 1, also as expected under CSR. Figures 10-13 representing $K(t) - \pi t^2$, $K(t)/\pi t^2$, $K(t) - \frac{4\pi t^3}{3}$ and $K(t)/\frac{4\pi t^3}{3}$ for two different sets of Poisson cluster process, illustrate as expected increasing functions of $K(t) - \pi t^2$ and $K(t) - \frac{4\pi t^3}{3}$. The plots to the right illustrate convergence towards 1 as t grows. There is, however, more uniformity in the cluster process with parameters $\rho = 20$, $\gamma = 25$, $\epsilon = 175$.
Finally Figures 14-17, representing simulations of Poisson inhibition process, are very much like their corresponding figures under CSR. This is due to the small values of δ which make the regularity almost non-existent. This occurence is discussed in **5.4**. To see how larger values of δ would monitor regularity, see **7.4**.

3.2 Voronoi Tessellation

An apparent result here is the high value of CV for the inhibition processes. According to the majority of works done earlier to investigate the relationship between CV and the underlying point pattern, the expected CV value when there is significant regularity is less than 0.33. That is, under the hypothesis of CSR, the CV value is expected to lie in the interval (0.33 , 0.64) with a probability of 0.99 (see Duyckaerts et al.). Why the CV value is higher in this case is believed to be related to the value of δ and hence λ_i and the geometrical characteristics of our simulations i $3D$. More of this is brought up in **5.4** and **7.2**.

Type	Parameters	CV	Ratio
CSR	*500 events*	0.6336	0.7732
Cluster	$\rho = 50$, $\gamma = 10$, $\epsilon = 100$	0.7310	0.5222
Cluster	$\rho = 20$, $\gamma = 25$, $\epsilon = 175$	0.8292	0.6728
Inhibition	*1000 initial events and* $\delta = 0.0148$	0.6109	0.8370
Inhibition	*585 initial events and* $\delta = 0.0093$	0.6213	0.7744

CV-values and ratio of remaining events after elimination from Voronoi Tessellation

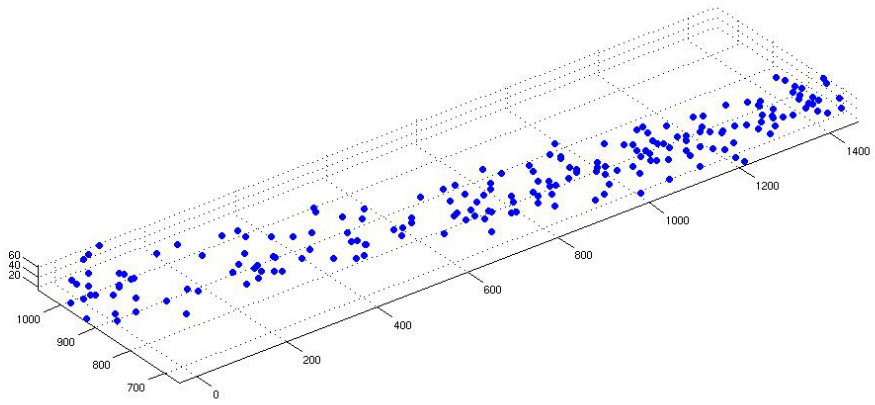

Figure 18: *One of our samples in 'glt' before test of stationarity*

4 Results

The data in this investigation consists of 11 samples of a mouse brain. The samples are sets of 3-dimensional position coordinates that are imported directly into the computer program for analysis. Here we employ the non-parametric methods mentioned earlier in the text. The procedure of analysis went as follows:

There were 5 samples of $'etv'$ and 6 samples of $'glt'$ consisting of three columns representing the positions of the neural cells in the brain tissues. The samples were imported to a program in MatLab and tested for stationarity by simply scattering each cell's coordinates in a plot.

Those samples that showed somewhat great deviation from assumptions of stationarity were then rotated in order to reach minimum area/volume value. This was done by using clockwise and anticlockwise rotation matrices in $2D$ due to absence of non-stationarity in one of the dimensions. All samples were then observed again to avoid loss of data.
The programs written to evaluate Ripley's K-function in $2D$ and $3D$ follow the same outline described earlier in the theoretical section. The graphs obtained by these programs are shown below.

4.1 Ripley's K-function

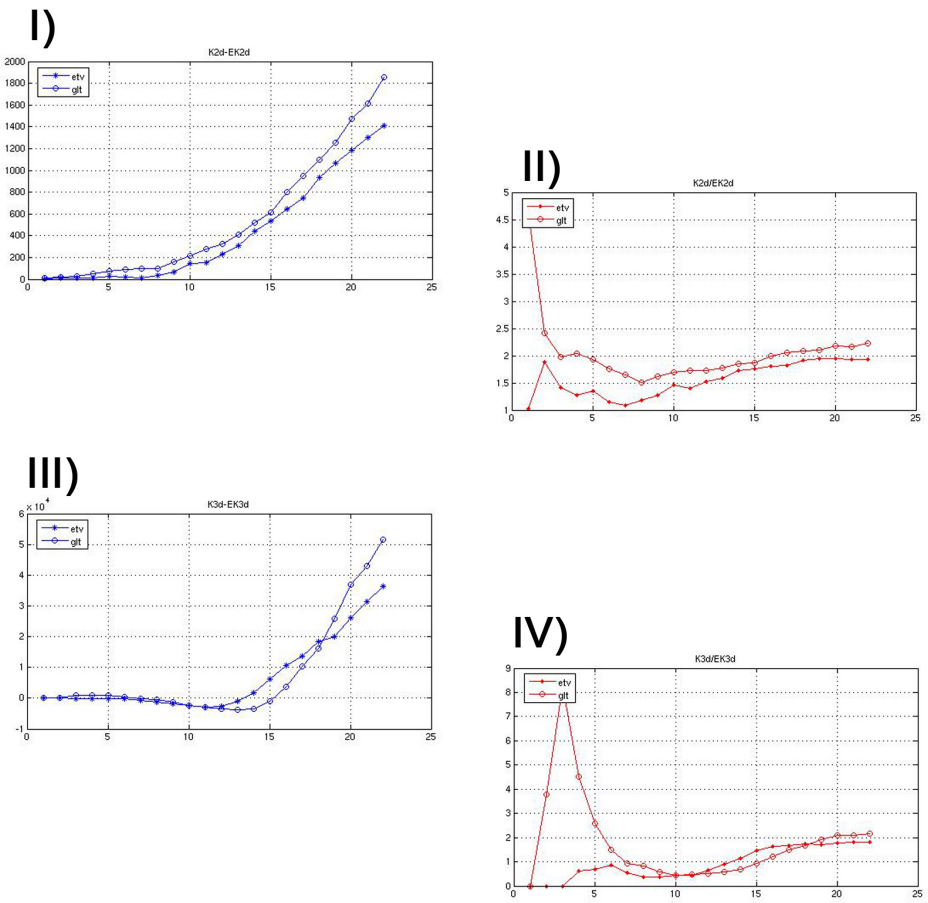

Figure 19:
I) Groupwise $K(t) - \pi t^2$ *in* $2D$
II) Groupwise $K(t)/\pi t^2$ *in* $2D$
III) Groupwise $K(t) - \frac{4\pi t^3}{3}$ *in* $3D$
IV) Groupwise $K(t)/\frac{4\pi t^3}{3}$ *in* $3D$

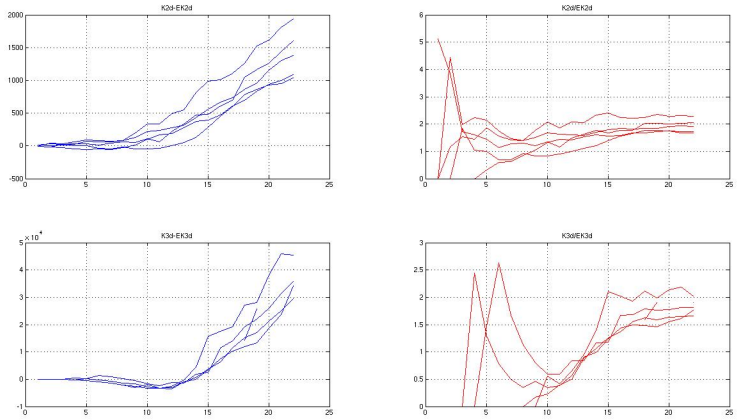

Figure 20: *Ripley's K-function in* $2D$ *and* $3D$ *for 'etv'*

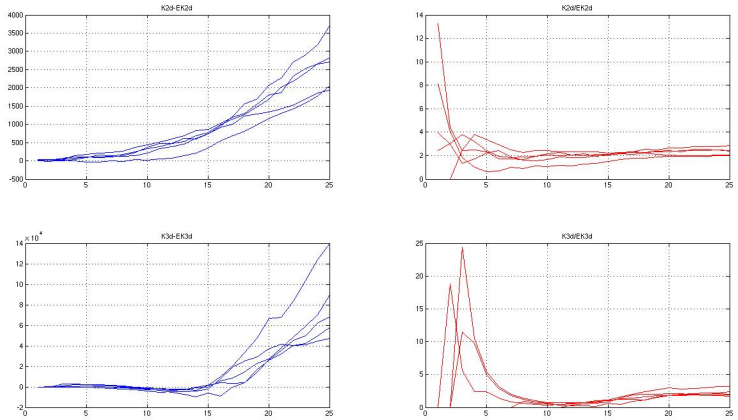

Figure 21: *Ripley's K-function in* $2D$ *and* $3D$ *for 'glt'*

4.2 Voronoi Tessellation

Using the `voronoin` command in MatLab we were able to obtain the locations of vertices for each Voronoi cell with an event as its center. Since an average 25% of these Voronoi cells are bulit on vertices located beyond the boundaries of the sample domains, they create irrationally big areas. These Voronoi cells are neglected in order to minimize the biased variation in the CV.

	CV	Ratio
'etv'	0.8556	0.7345
'glt'	0.9239	0.7674

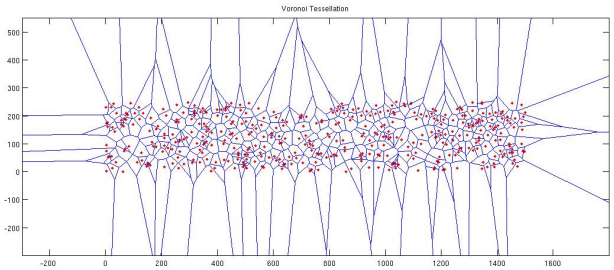

Figure 22: *Voronoi tessellation on a sample in group 'etv'*

5 Discussion

5.1 Aim

Our aim has partially been to perform a computational investigation of spatial point patterns in 3-dimensional domains and hence developing a K-function operating in $3D$. Based on our simulations in MatLab, this function and its edge correction term appear to work without any flaws despite the slight truncation of corner events. Also the results from programs written to evaluate its corresponding $2D$ function, Voronoi tessellation in $2D$ and simulations of manipulated Poisson processes match our expectations.

5.2 Ripley's K and Edge Correcton in $3D$

A lot of resources throughout the project has been devoted to creation of a compatible edge correction term for analysis in $3D$. As mentioned in **2.3.4** the evaluation of this term is of a time-demanding nature due to its local property and the involvement of integrals that require numerical methods in order to be solved. This is, however, despite our slight truncation of corner events. One might suspect that taking these observations into account and thus introducing a corresponding edge correction term for spheres cut by three planes, would reduce the time efficiecy of Ripley's K in $3D$ substantially. A suitable proposal would therefore be to create a global edge correction term for observations in $3D$ due to the extreme efficiency of global terms as we've observed in Ward & Ferrandino's edge correction term in $2D$.

5.3 Extension of Voronoi Tessellation to $3D$

Another possibility of development is extending Voronoi tessellation to $3D$ domains. Locating the vertices of the resulting polyhedrons is somewhat more demanding than the corresponding procedure in $2D$ but nontheless possible. The real problem, however, is the evaluation of polyhedron volumes due to the great variability in the number of vertices and thus their shapes.

5.4 Simulations

In order to maintain consistency when conducting simulations of the Poisson cluster process and Poisson inhibition process, we chose to use parameters that would result in an intensity, λ_c or λ_i, equivalent to the one in CSR with 500 events. This choice in the case of the Poisson inhibition process leads to very small values of δ, which result in an almost non-existent regularity.

This weak presence of regularity is also reflected in the CV-values obtained by Voronoi tessellation in **3.2**. Choosing larger values of δ would cause λ_i to swiftly converge towards zero (see (2) and **7.2**) and lead to simulations with considerably less events. Although such simulations would illustrate more regularity (**7.4**), they would result in inconsistent values of λ_i.

5.5 Sensitivity to Stationarity

We recommend the following as plausible possibilities of improvement regarding non-stationary samples:

- Usage of methods invariant to stationarity. The theoretical structures of Ripley's K-function and Voronoi tessellation are both sensitive to non-stationarity. As it has been mentioned earlier 'empty spaces' in the sample lead these methods to false conclusions regarding the existing point pattern.

- Resampling. Despite the mentioned flaw in Ripley's K-function and Voronoi tessellation, their non-parametric property is a desirable quality. Resampling or simply cutting samples to make them stationary is another possible approach. Yet there is the disadvantage of information loss when samples are reduced.

- Evolution of methods. A reconstruction or insertion of new parameters in the present methods to create immunity against stationarity.

5.6 The Mouse Brain Samples

Regarding our mouse brain samples, we choose to state our inferences with some doubt. This is due to the non-stationary nature of the samples. A property that nevertheless seems to remain after our adjustments of the position coordinates using rotation matrices.
The two groups, $'etv'$ and $'glt'$, of mouse brain samples seem to demonstrate a clear clustering trend. Analysis of Ripley's K-function in $2D$ shows a stronger presence of this trend in $'glt'$ for all values of t. This constant presence is violated in $3D$ for a small interval between $t = 12$ and $t = 19$. This deviation can very possibly be a result of the expansion of analysis to one more dimension and hence taking more information into account. The results from Voronoi tessellation, too, point towards a stronger presence of clustering in $'glt'$.

6 References

Cotter D., Landau S., Beasley C., Stevenson R., Chana G., MacMillan L., Everall I. (2001). *The Density and Spatial Distribution of GABAergic Neurons, Labelled Using Calcium Binding Proteins, in the Anterior Cingulate Cortex in Major Depressive Disorder, Bipolar Disorder, and Schizophrenia.* Society of Biological Psychiatry, United Kingdom.

Diggle P.J. (2003). *Statistical Analysis Of Spatial Point Patterns.* Oxford University Press Inc., New York.

Diggle P.J., Lange N., Benes M. (1991). *Analysis of Variance for Replicated Spatial Point Patterns in Clinical Neuroanatomy.* Journal of the American Statistical Association, Vol. 86, No.415. (Sep., 1991), pp. 618-625.

Duyckaerts C., Godefroy G. (2001). *Voronoi tessellation to study the numerical density and the spatial distribution of neurones.* Journal of Chemical Neuroanatomy 20 (2003) 83-92.

Duyckaerts C., Godefroy G., Hauw J. (1994). *Evaluation of neuronal numerical density by Dirichlet tessellation.* Journal of Neurosceince Methods 51 (1994) 47-69.

Marcon E., Puech F. (2003). *Generalizing Ripley's K-function to Inhomogenous Populations.*
http://e.marcon.free.fr/publications/index.htm.

Neyman J., Scott E.L. (1958). *Statistical approach to problems of cosmology (with discussion).* Journal of the Royal Statistical Society, B 20, 1-43.

Weisstein, Eric W. *Rotation Matrix. From MathWorld–A Wolfram Web Resource.*
http://mathworld.wolfram.com/RotationMatrix.html.

7 Appendix

7.1 Implementational Structure in MatLab

The programmes written to execute the work in this paper are ordered by Karolinska Institute and shall be made public on behalf of the associated parts.

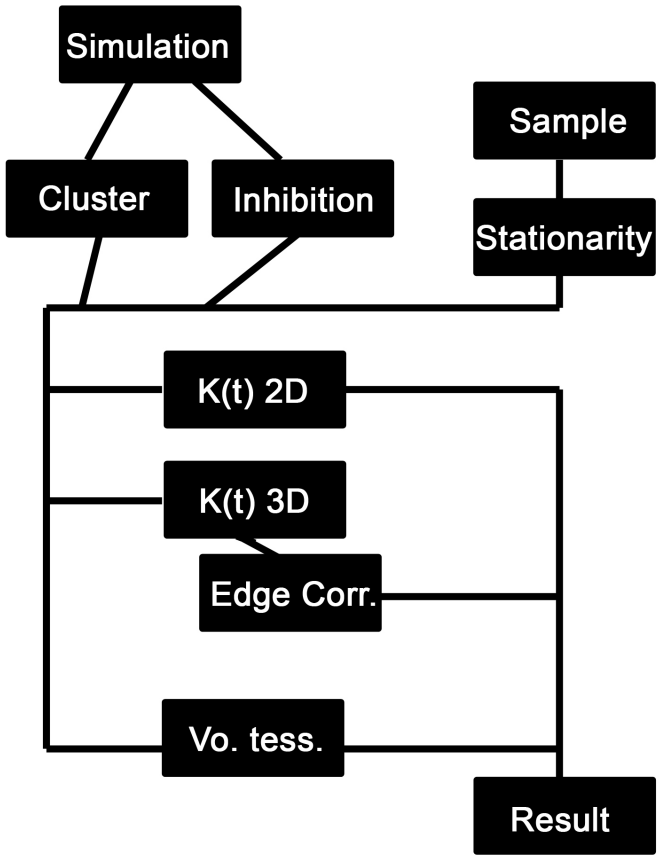

Figure 23: *The computational structure*

7.2 Poisson Inhibition Process Intensity, λ_i

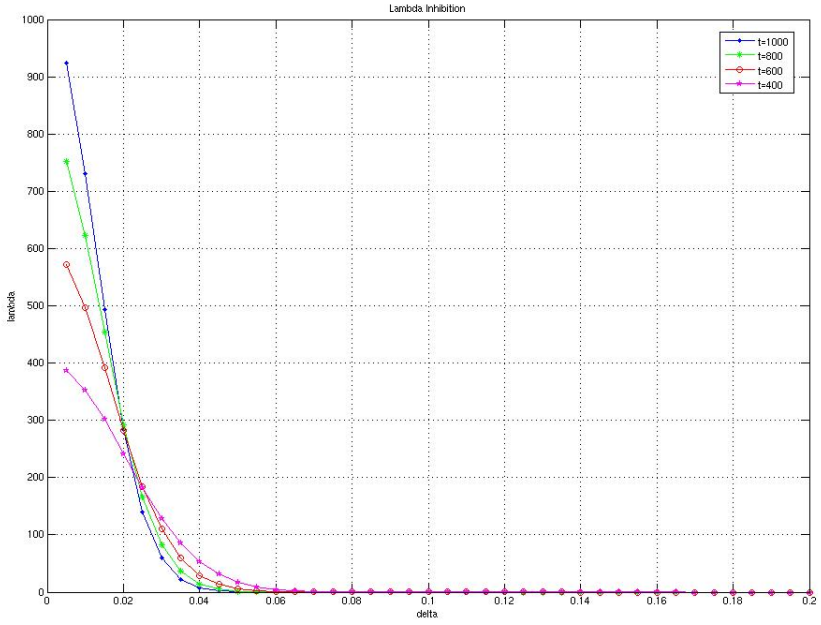

Figure 24: λ_i *for* $\tau = 1000$, $\tau = 800$, $\tau = 600$, $\tau = 400$ *and* $\delta = 0.005, \ldots, 0.02$ *according to* $\lambda_i = \tau \exp(-\pi\tau\delta^2)$.

7.3 Rotation Matrices in $3D$

The following are the clockwise rotation matrices in $3D$ about the x, y and z axes.

$$R_x(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}$$

$$R_y(\theta) = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix}$$

$$R_z(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

7.4 Alternative Simulations of Poisson Inhibition Process

These simulations illustrate the decreasing trend of $K(t) - \pi t^2$ in the interval $(0, \delta)$.

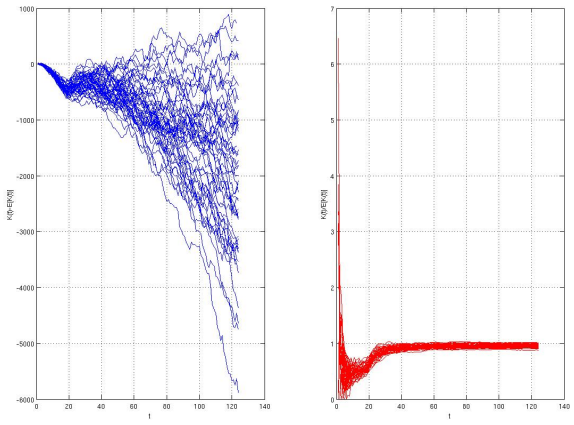

Figure 25: *$K(t) - \pi t^2$ and $K(t)/\pi t^2$ in 2D for 50 simulations of Poisson inhibition process with 700 initial events and $\delta = 20$*

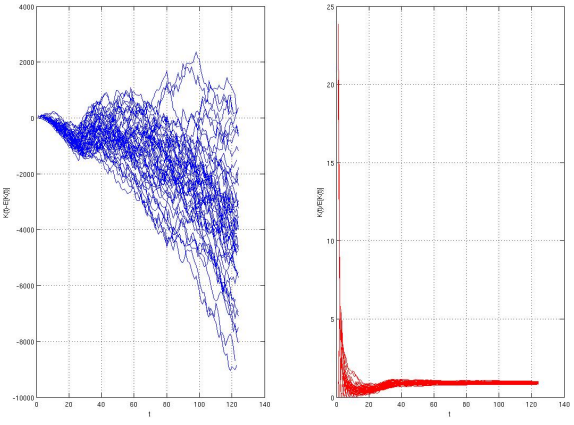

Figure 26: *$K(t) - \pi t^2$ and $K(t)/\pi t^2$ in 2D for 50 simulations of Poisson inhibition process with 500 initial events and $\delta = 27$*