Basic Analysis of Spatial Patterns

Department of Mining and Mineral Engineering
The University of Leeds
January 2003
1. Introduction
The first step towards spatial pattern modelling is the spatial randomness analysis. A rich source of methods can be found in the literature but in general terms the methods can be classified into three categories:

- Distance analysis
- Quadrat count analysis
- Second moment analysis

Distance analysis uses the distances between events or the distances between events and selected points. Quadrat count analysis uses the number of events falling into the quadrats of certain shape, either located randomly or according to a pre-arranged pattern. Second moment analysis mainly deals with the K-function analysis though the second order intensity and the covariance density also fall into this category. In this report, each category of the analysis is discussed in turn and some examples are given.

2. Distance analysis
As illustrated in Figure 2.1, we use \( t, y \) and \( x \) to represent three different measures of distances. \( t \) is the distance between any pair of events and is referred as interevent distance. \( y \) is the distance between an event and its nearest one and is named the nearest neighbour distance. \( x \) is the distance between a selected point in the region and its nearest event and is referred to as the point to nearest event distance. For the later two types of distances, if the interest is concerned with the \( k \)-th nearest event, the distances are then called the \( k \)-th nearest neighbour distance or the \( k \)-th point to nearest event distance.

2.1 Distance distribution theory
For a point process in \( \mathbb{R}^d \), under the assumption of homogeneous Poisson process, the probability that there is no event within distance \( y \) of an arbitrary event can be obtained by the inter-event property of the Poisson process and can be expressed as:

\[
P[\text{no further event within distance } y \text{ of an event}] = e^{-\lambda V_y},
\]

where \( V_y \) is the \( d \)-dimensional volume of a sphere with radius \( y \) and is given below:

\[
V_y = \frac{\pi^{\frac{d}{2}} \cdot y^d}{\Gamma(1 + \frac{d}{2})}
\]

Therefore the probability distribution function of measure \( y \) and its density function are:

\[
\begin{align*}
G(y) &= 1 - e^{-\lambda V_y} \quad \text{(probability function)} \\
g(y) &= \frac{d \cdot \lambda \cdot \pi^{\frac{d}{2}} \cdot y^{d-1}}{\Gamma(1 + \frac{d}{2})} \cdot e^{-\lambda V_y} \quad \text{(density function)}
\end{align*}
\]

For two dimensional case,

\[
\begin{align*}
G(y) &= 1 - e^{-\lambda \pi y^2} \quad \text{(probability function)} \\
g(y) &= 2\pi\lambda \cdot y \cdot e^{-\lambda \pi y^2} \quad \text{(density function)}
\end{align*}
\]

If we consider up to the \( k \)th nearest event of an arbitrary event, the joint distribution function for \( y_1, y_2, \ldots, y_k \) can be derived as follows:

Based on the inter-event probability property of a Poisson process, the following statements can be written for a selected event:

\[
\begin{align*}
P[\text{no further event within distance } y_1] &= e^{-\lambda V_{y_1}} \\
P[\text{no further event within distance } y_2 \text{ on the condition that there is an event at } y_1] &= e^{-\lambda (V_{y_2} - V_{y_1})} \\
P[\text{no further event within distance } y_3 \text{ on the condition that there are events at } y_1, y_2] &= e^{-\lambda (V_{y_3} - V_{y_1})} \\
\cdots \cdots \\
P[\text{no further event within distance } y_k \text{ on the condition that there are event at } y_1, y_2, \ldots, y_{k-1}] &= e^{-\lambda (V_{y_k} - V_{y_{k-1}})}
\end{align*}
\]

The distribution functions for these events are:
Therefore the joint distribution density function is given by:

\[
g_k(y_1, y_2, \cdots, y_k) = \lambda^k e^{-\lambda V_k}
\]

and the probability function is:

\[
G_k(y_1^*, y_2^*, \cdots, y_k^*) = \int_{V_{y_1}^*}^{V_{y_2}^*} \cdots \int_{V_{y_{k-1}}^*}^{V_{y_k}^*} \lambda^k e^{-\lambda V_1} dV_1 \cdot dV_2 \cdots dV_k
\]

\[
= \prod_{i=1}^{k-1} (V_i^* - V_{i-1}^*) \cdot \lambda^{k-1} \cdot (e^{-\lambda V_{i-1}} - e^{-\lambda V_i})
\]

with \( V_0^* = 0 \)

To account for the edge effect, it is sufficient to replace \( V_i^* \) in the equation by \( V_i^* \cap \mathcal{R} \), where \( V_i^* \) is the d-dimensional sphere volume with radius \( y_i^* \) and \( \mathcal{R} \) is the region under consideration.

The same theory can be applied to the distribution of \( x \), the nearest event distance to a randomly selected point, i.e., the following relations can be obtained:

\[
\begin{align*}
F(x) &= 1 - e^{-\lambda V_1} \quad \text{(probability function)} \\
f(x) &= \frac{d \cdot \lambda \cdot \pi^{\frac{d}{2}} \cdot x^{d-1} \cdot e^{-\lambda V_1}}{\Gamma(1 + \frac{d}{2})} \quad \text{(density function)}
\end{align*}
\]

For two dimensional case,

\[
\begin{align*}
F(x) &= 1 - e^{-\lambda x^2} \quad \text{(probability function)} \\
f(x) &= 2\pi\lambda \cdot x \cdot e^{-\lambda x^2} \quad \text{(density function)}
\end{align*}
\]

and for the nearest k events:

\[
\begin{align*}
F_k(x_1, x_2, \cdots, x_k) &= \lambda^k e^{-\lambda V_k} \\
F_k(x_1^*, x_2^*, \cdots, x_k^*) &= \prod_{i=2}^{k} (V_{i-1}^* - V_{i-2}^*) \cdot \lambda^{k-1} \cdot (e^{-\lambda V_{i-2}} - e^{-\lambda V_{i-1}})
\end{align*}
\]

with \( V_0^* = 0 \)

For the inter-event distance \( t \), Bartlett [ ] derived the distribution density in 1964 for homogeneous Poisson process within a two dimensional square, which is given in the following equation:
\[
\begin{align*}
    h(t) &= \begin{cases} 
    \frac{2\pi t}{L^3} - \frac{8t^2}{3L^3} + \frac{2t^3}{L^3} & (t^2 \leq L^2) \\
    \frac{4t \cdot \arcsin(\frac{2L^2}{t^2} - 1)}{L^3} - \frac{4t}{L^2} + \frac{8t \cdot \sqrt{t^2 - L^2}}{L^3} - \frac{2t^3}{L^4} & (L^2 < t^2 \leq 2L^2)
    \end{cases} 
\end{align*}
\]

where \( L \) is the size of the square. Diggle [] gives the distribution function for a unit square as follows:

\[
    H(t) = \begin{cases} 
    \pi \cdot t^2 - \frac{8 \cdot t^3}{3} + \frac{t^4}{2} & 0 \leq t \leq 1 \\
    \frac{1}{3} - 2 \cdot t^2 - \frac{t^4}{2} + \frac{4 \sqrt{t^2 - 1} \cdot (2 \cdot t^2 + 1)}{3} + 2 \cdot t^2 \cdot \arcsin(\frac{2}{t^2} - 1) & 1 < t \leq \sqrt{2}
    \end{cases}
\]

2.2 Implementation issues

For \( t, y \) and \( x \) distance analysis, the distribution can be calculated in the forms of either histogram or cumulative histogram. For cumulative histogram calculation, they can be expressed as:

\[
    \begin{align*}
    \hat{H}(t) &= \frac{\#(t_{ij} \leq t)}{\frac{1}{2}n(n-1)} & i, j = 1, 2, \cdots, n, i \neq j, \text{number of data points} \\
    \hat{G}(y) &= \frac{\#(y_i \leq y)}{n} & i = 1, 2, \cdots, n, \text{number of data points} \\
    \hat{F}(x) &= \frac{\#(x_i \leq x)}{m} & i = 1, 2, \cdots, m, \text{number of selected points}
    \end{align*}
\]

For histogram calculation, which is analogous to density distribution functions, the ranges of \( t, y \) or \( x \) are divided into \( C \) number of classes first and the relative frequency (distribution density) values are given by:

\[
    \begin{align*}
    \tilde{h}(t^k) &= \frac{\#(t_{ij}^k - 0.5\Delta < t_{ij} \leq (t^k + 0.5\Delta))}{\frac{1}{2}n(n-1)} & i, j = 1, 2, \cdots, n, i \neq j \\
    \tilde{g}(y^k) &= \frac{\#(y_i^k - 0.5\Delta < y_i \leq (y^k + 0.5\Delta))}{n} & i = 1, 2, \cdots, n \\
    \tilde{f}(x^k) &= \frac{\#(x_i^k - 0.5\Delta < x_i \leq (x^k + 0.5\Delta))}{m} & i = 1, 2, \cdots, m
    \end{align*}
\]

where \( i, j \) and \( m \) have the same meanings as the last equation, \( k \) is the distribution class id and \( \Delta \) is the interval dividing the ranges of \( t, y \) and \( x \).

To compare the results with theory, however, edge effect has to be taken into account. For the calculation of \( \hat{G} \) and \( \hat{F} \), the edge effect can be corrected by the following relations:
where \(d_i\) is the distance of the selected event (out of \(n\) events) or the selected point (out of \(m\) points) to the nearest border (edge) of the region \(\mathcal{R}\) being considered. This treatment is equivalent to imposing a safe-guarding area (or volume) around the edge of the region (see Upton []) and discard any event (or point) falling within it. Note the guarding area (volume) changes in size according to the distance \(t\), \(y\) or \(x\) being considered. Another approach to edge correction specifically in two-dimensional case when the region being considered \(\mathcal{R}\) is a rectangle is to virtually join the opposite side of the rectangle and turn the region into a torus. Distances are then calculated on the basis of this virtual region. For example, the nearest event of an event located at the bottom left corner of the region can be an event located at the top right corner of the region.

The random points selected for the calculation of \(\tilde{F}\) or \(\tilde{F}\) values can be chosen randomly or from a fixed grid. Diggle [] suggests using a \(k\times k\) regular grid and increasing \(k\) until \(\tilde{F}\) or \(\tilde{F}\) effectively stabilizes throughout its range.

### 2.3 Monte Carlo reference simulation

An effective approach to test if a point pattern is a particular realisation from a point process is by the help of null hypothesis test based on a certain number of Monte Carlo reference simulation. To proceed with this test, the statistics of the data, \(s_d\), is calculated first. Then \(k\) number of independent simulations based on the model defining the point process are conducted and the corresponding statistics for each simulation, \(s_1, s_2, \ldots, s_k\), are calculated. \(s_d, s_1, s_2, \ldots, s_k\) are then rearranged in ascending order. Then under the null hypothesis with a significance level \(\alpha\), the ranking of \(s_d\) within the sequence must be:

\[
\begin{align*}
    j &\leq (k + 1) \times (1 - \alpha) \quad \text{for upper tail test} \\
    j &> (k + 1) \times \alpha \quad \text{for lower tail test} \\
    (k + 1) \times \alpha / 2 &\leq j \leq (k + 1) \times (1 - \alpha / 2) \quad \text{for two tail test}
\end{align*}
\]

where \(j\) is the ranking of \(s_d\) in the sequence. Any \(j\) value not honouring the above condition will lead to the rejection of the hypothesis.

For example, if the inter-event distance of a point process is being analysed, one of the obvious choices of statistics for null hypothesis testing is the squared differences between the calculated \(\tilde{H}(t)\) and the theoretical \(H(t)\), i.e.,

\[
s_d = \int (\tilde{H}(t) - H(t))^2 dt
\]

where \(H(t)\) is the probability distribution of inter-event distances of the point model to be tested. \(k\) independent realisations are generated and each of the squared
differences between $\hat{H}_1, \hat{H}_2, \cdots, \hat{H}_k$ and $H$ are calculated. The above criteria is then applied to test the hypothesis. In this case, it is an upper tail test. For instance, if 99 simulations are used and the significance level is 5%, then any ranking of $s_d$ above 96 will reject the hypothesis.

Note if the theoretical value of $H(t)$ is unknown, however, it can be approximated by the average value derived from the simulations, i.e.,

$$H(t) = \bar{H}(t) = \frac{1}{k} \sum_k \hat{H}_k(t)$$

The above example is an overall test of $\hat{H}(t)$ for the whole range of $t$ and only gives the picture of average behaviour. The test can also be tested on different $t$ basis. In this case, $s_d = \bar{H}_d(t)$, and $s_1 = \hat{H}_1(t), s_2 = \hat{H}_2(t), \cdots, s_k = \hat{H}(k)$, the same criteria can then be applied. In this case, however, it is a two tail test as too small or too large $s_d$ means the departure of the statistics from the model at that particular distance. As the test depends on $t$ value and therefore it is far more comprehensible to use graph to display the test results. Based on the simulations, if we plot the hypothesis acceptance envelope of $\hat{H}(t)$ against the whole range of $t$, and same as the $H_d(t)$ calculated from the data in the same graph, the rejection of the hypothesis can be concluded if any part of $\hat{H}_d(t)$ goes outside the hypothesis acceptance envelope. This test is more robust and gives more details about the discrepancy between the point model and the data and therefore will be adapted for current research.

The above descriptions can apply similarly to other statistics such as $\hat{G}(t)$ or $\hat{F}(t)$.

Note, however, the test result is correct only if a rejection conclusion is reached. A particular testing not rejecting the hypothesis does not necessarily means the acceptance of the model for the data set. To accept a model with confidence for a point pattern, a few different statistics should normally be rigorously tested.

Note also in the above arguments, the point model can refer to any known models. If the homogeneous Poisson process model is used, the test will be against the discrepancy between the data points and a complete spatial random (CSR) point pattern. In other words, the test is against the departure of the data from CSR.

With respect to the edge effect issue, it can either be taken into account or ignored if statistics from data are only to be compared to those from Monte Carlo simulation using the same treatment. If the results are to be compared with the theoretical values, however, edge corrections must be considered. See examples below.

2.4 Distance analysis on some generated patterns
This section simply displays some distance analysis results for some artificially generated patterns. The underlying models for these pattern are fully defined.

2.4.1 Homogeneous Poisson point pattern
Figure 2.2 is a realisation of a homogeneous Poisson process for a rectangular region $\mathcal{R}=(0,100)\times(0,100)$ with density $\lambda=0.01$. The distance statistics $\hat{H}(t), \hat{h}(t), \hat{G}(y), \hat{g}(y), \hat{F}(x) and \hat{f}(x)$ are shown in the following graph from Figure 2.3 to Figure 2.5. The Monte Carlo simulation results are also plotted based on
100 simulations. Three statistics are displayed from the Monte Carlo, the average simulated value, the minimum and the maximum acceptance envelope values with 95% confidence based on a two tails test described in the above section. As can be seen in the homogeneous case, the calculated statistics agree well with the Monte Carlo test results in all occasions.

Figure 2.2 Homogeneous Poisson points

Figure 2.3 (a) \( H(t) \) vs distance \( t \)

Figure 2.3 (b) \( h(t) \) vs distance \( t \)

Figure 2.4 (a) \( G(y) \) vs distance \( y \)

Figure 2.4 (b) \( g(y) \) vs distance \( y \)
2.4.2 Non-homogeneous Poisson point pattern

Figure 2.6 is a realisation of a non-homogeneous Poisson process for the same region but with the density function defined as:

\[ \lambda(u, v) = 0.1 \cdot e^{-u - 2v} \]

where \( u \) and \( v \) are the horizontal and vertical coordinates. The distance statistics \( \hat{H}(t), \hat{h}(t), \hat{G}(y), \hat{g}(y), \hat{F}(x) \) and \( \hat{f}(x) \) are shown in Figure 2.7 to Figure 2.9 respectively. As can be seen from \( \hat{H}(t) \) and \( \hat{h}(t) \) analysis, the distributions start to depart from the statistics obtained from homogeneous simulation. The degree of departure depends on the data but it cannot easily be quantified by the analysis. The \( \hat{G}(y) \) and \( \hat{g}(y) \) analysis does not even suggest the non-homogeneity for the pattern. \( \hat{F}(x) \) and \( \hat{f}(x) \) only show a very modest degree of departure.

In general for non-homogeneous case, \( \hat{h}(t) \) will tend to be negative skewed as there will always be some point aggregation compared with the average distribution. This feature is clearly visible from the figure. More discussion about this point will follow in the next section.
2.4.3 Poisson cluster point pattern

Figure 2.10 is a realisation of a Poisson cluster process for the same region. The parent process in this example is a homogeneous Poisson process with density \( \lambda = 0.005 \). Each parent produces a fixed number of 20 daughters. Daughter points are
uniformly distributed around their parent within a circle of radius of 5 and centred at their parent location. The realisation consists of daughter points only.

The distance statistics $H(t), h(t), G(y), g(y), F(x)$ and $f(x)$ for the realisation are shown from Figure 2.11 to Figure 2.13. All figures suggest a certain degree of departures of the point pattern from homogeneous case. Only $h(t)$, however, demonstrates some interesting results. $h(t)$ shows multi-modes characteristics. The first mode peaks at around $t = 5$, which suggests from the histogram analysis point of view that there is a point aggregation with the average inter-event distance inside the aggregation at about 5. This is exactly the size of the cluster we specified when generating the pattern. There are also some other modes that suggest aggregation at different level. The overall trend of $h(t)$, however, follows more or less the curve for homogeneous process, which suggests that the underlying parent process is homogeneous. Recall from the last section for non-homogeneous case, the distribution of $h(t)$ will tend to be negative skewed. Figures of other distance analysis always also demonstrate considerable amount of deviation of this pattern from homogeneous case, but the connection between the results and the defined cluster process is not obvious.

![Figure 2.10 Poisson cluster process points](image)

![Figure 2.11 (a) $H(t)$ vs distance $t$](image)

![Figure 2.11 (b) $h(t)$ vs distance $t$](image)
To act as a further example, another realisation of a Poisson cluster process is analysed. In this realisation, the daughter points are distributed according to the following binormal distribution:

\[ h(x, y) = \frac{1}{2\pi \cdot 3.5^2} e^{-\frac{x^2+y^2}{2\cdot 3.5^2}} \]

with variance \( \sigma^2 = 13.5 \). The parent process again is homogeneous. The points image is shown in Figure 2.14.

Distance analysis results are displayed in Figure 2.15 - 2.17. By comparing these figures with those of Figure 2.11 – 2.13, it is not difficult to conclude that great similarities in the results between the two point realisations. The average distance between points from the same parent is \( \sqrt{2} \times 3.5 = 5 \) which is correctly identified in the \( \hat{h}(t) \) analysis result by the peak of the first mode.
Figure 2.14 Poisson cluster process points

Figure 2.15 (a) $\hat{H}(t)$ vs distance $t$

Figure 2.15 (b) $\hat{h}(t)$ vs distance $t$

Figure 2.16 (a) $\hat{G}(y)$ vs distance $y$

Figure 2.16 (b) $\hat{g}(y)$ vs distance $y$
These are not selected examples. In fact clustered points pattern always show the same characteristics, i.e., multi-modes in the $h(t)$ graph and the overall trend implicitly represent the parent process. In other words, if $h(t)$ calculated from a particular data set shows the same or similar characteristics, a cluster process can then be used as the underlying model. If there is no evidence of multi-mode $h(t)$, a non-homogeneous model may be more suitable.

This, however, only serves as a general rule of thumb and should not be treated as a fixed relation. For example, in two extreme cases, if the number of clusters (aggregated points cloud) is small compared to the whole point population, or if the distribution of daughters is more spread out, resulting in greater mixtures between daughters from different parent, the multi-mode judgement based on $h(t)$ may not be suitable, or may even provide false directions. Other limitations include the applications of the rule in the case of anisotropic distribution of daughter process. In this situation, directional distance analysis may help resolve the problem.

Note the rule described here can only help to judge if the point pattern has clusters and help to establish the most likely cluster size if it has. It does not reveal any information about the distribution of daughter process itself.

### 2.4.4 Cox point pattern

Figure 2.18 is a realisation of a Cox process for the same region. The Cox model is defined as a normal distribution with mean and variance defined as follows:

$$
\begin{align*}
\text{mean}(u,v) &= 0.1 \cdot e^{-\frac{-u^2}{100}} \\
\text{variance}(u,v) &= 0.015 \cdot e^{-\frac{-u^2}{100}}
\end{align*}
$$

where $u$ and $v$ are the horizontal and vertical coordinates. The mean basically defines a mean density field which is analogous to that used in the example of Figure 2.6. At each location, the mean together with the variance that is about 15% the mean value, define a normal distribution for the density at that location. A random value is then generated from this distribution to serve as the realisation of the density field at the
location. By comparing the image with that of Figure 2.6, it is obvious the realisation of Cox process expresses more randomness which renders the defined underlying model such as the mean less obvious. This is certainly due to the extra random component which controls the realisation of the random density field. This feature of Cox process obviously imposes extra difficulties to identify an effective parametric model for the modelling exercise.

The distance statistics $H(t)$, $h(t)$, $G(y)$, $g(y)$, $F(x)$ and $f(x)$ are shown in Figure 2.19 to Figure 2.21 respectively. Though the results show very similar features as the non-homogeneous point pattern, as compared with Figure 2.7 to Figure 2.9, they tend to get closer to CSR pattern because of the reason stated above.
2.5 Distance analysis of two example point dataset

In this section, two practical examples are analysed. The points data patterns are derived from two fracture trace images which are also displayed. For each of the two datasets, there are clearly two distinct sets of fractures which are almost perpendicular to each other. This is clearly the demonstration of two different geological formations and it may be beneficial to analyse the two fracture sets in each example separately. The separation of fracture sets and their distinct modelling are the fundamental steps toward the hierarchy modelling described by Lee & Einstein 

Figure 2.22 (a) Example 1 (fracture traces)
Figure 2.22 (b) Example 1 (Points data set)

Figure 2.23 (a) Example 1 – fracture trace 1
Figure 2.23 (b) Example 1 - fracture set 1

Figure 2.24 (a) Example 1 – fracture trace 2
Figure 2.24 (b) Example 1 - fracture set 2
Since cumulative histograms for nearest event distance y and point to nearest event distance x do not actually give clear pictures about the distance characteristics and therefore in the following analysis they will not be presented to save some space of the report. Figure 2.28 shows the three distance analysis histogram for the example 1 for fracture trace set 1 and 2 combined. Analysis for trace set 1 only is displayed in Figure 2.29, while for trace set 2 in Figure 2.30. For the second practical example, the corresponding analysis results are given in Figure 2.30 to Figure 2.32.
Figure 2.28 (a) $h(t)$ for E.1 – F.1 + F.2

Figure 2.28 (b) $h(t)$ for E.1 – F.1 + F.2

Figure 2.28 (c) $g(y)$ for E.1 – F.1 + F.2

Figure 2.28 (d) $f(x)$ for E.1 – F.1 + F.2

Figure 2.29 (a) $h(t)$ for E.1 – F.1

Figure 2.29 (b) $h(t)$ for E.1 – F.1
Figure 2.29 (c) \( g(y) \) for E.1 – F.1

Figure 2.29 (d) \( f(x) \) for E.1 – F.1

Figure 2.30 (a) \( h(t) \) for E.1 – F.2

Figure 2.30 (b) \( h(t) \) for E.1 – F.2

Figure 2.30 (c) \( g(y) \) for E.1 – F.2

Figure 2.30 (d) \( f(x) \) for E.1 – F.2
Figure 2.31 (a) $h(t)$ for E.2 – F.1 + F.2

Figure 2.31 (b) $h(t)$ for E.2 – F.1 + F.2

Figure 2.31 (c) $g(y)$ for E.2 – F.1 + F.2

Figure 2.31 (d) $f(x)$ for E.2 – F.1 + F.2

Figure 2.32 (a) $h(t)$ for E.2 – F.1

Figure 2.32 (b) $h(t)$ for E.2 – F.1
For example 1, analysis for all three cases (fracture trace set 1 only, fracture trace set 2 only or fracture trace set 1 + set 2) show very similar results. A brief list of some of the common features are given below:

1. Serious departure from homogeneous Poisson process.
2. Negative skewed histograms for the inter-event distance analysis suggest in general a non-homogeneous type of Poisson process should be used for the modelling.
3. Only single mode can be observed in $h(t)$ for all three cases which suggest there is no cluster features in the point patterns.

4. Compared with points from fracture trace set 2, points from fracture trace set 1 show less deviation from homogeneous case which reveals that for this set of fractures, there is less aggregation, i.e., fractures tend to be more evenly distributed. The degree of deviation can be obtained by inspecting the histogram. Greater skew of the set 2 suggest more serious points aggregation.

5. A logical geological extension to the above point is that fracture trace set 1 was created before set 2. When set 2 fractures were being formed at later geological activity, certain part of the rock was weakened more seriously than other parts by the existing set 1 fractures and therefore attract more new fractures being created. More fracture aggregation would be the results of this action. This argument, of cause, is only a speculation based on the distance analysis results. More verifications need to be done from different angles, especially from the geological history of the site.

6. A common feature for non-homogeneous point patterns is the significantly high proportion of small nearest event distances compared with homogeneous case. This is so as more points aggregation means more small inter-event distance point pairs and more tendency toward non-homogeneity. This feature is clearly visible from $g(y)$ analysis of the three cases.

7. Since a regular grid covering the whole area of the region being studied is used to calculate the point to nearest event distance $x$, the distribution $f(x)$ can then be viewed as certain measure for points distribution across the whole area. For the case when the whole area is covered by points realised from a homogeneous point process, $f(x)$ will show identical characteristics as $g(y)$. For cases when points only occupy certain area of the whole region, the distribution of $f(x)$ will tend to uniform. An extreme case is shown in Figure 2.34 below.

Figure 2.34 (a) An extreme case
Analysis results show typical features of non-homogeneous point processes. The most interesting result is the distribution of $f(x)$ shown in Figure 2.34 (e). As most of the area in the region being studied is not covered by data (or to be precisely, covered by data with zero point density value), $f(x)$ tends to be uniformly distributed.

As for the example 1 dataset, considerable amount of space in the region is not covered by data points and therefore $f(x)$ tends to skew to the right compared with the simulated results derived by the Monte Carlo simulations. As a result, the degree of uniformality of $f(x)$ may be used as an index to describe the proportion of the space being occupied by the point cloud. More consideration may be given to this speculation at later stage.

For example 2 data set, all analysis basically suggest similar features to those of example 1 data set. Non-homogeneous process is again suggested for the fracture trace set 1 only, fracture trace set 2 only or the combined fracture traces. Whether or not the two fracture traces form a hierarchy depends on some rigorous test of dependency between the sets. Relations between different point sets will be covered in later research.
3. Quadrat count analysis

Quadrat count analysis is a kind of variance analysis. It uses the measures of number of points inside quadrats located inside the region. There is no special restrictions for the shape and size of a quadrat provided that the size is reasonable compared to the volume (area) of the region. For a particular analysis, however, the shape of size of quadrats are fixed.

3.1 Theoretical background

From its origination, quadrat count analysis is to test samples from Poisson distribution. For a Poisson distribution with mean \( m \), the probability of obtaining sample value \( N \) is given as:

\[
P(N) = \frac{m^N}{N!} e^{-m}
\]

For any volume \( V \) located at \( X \) in a Poisson point field, the mean of the Poisson distribution at the location will be:

\[
m_*(X) = \int_V \lambda(X) \cdot dX
\]

For homogeneous Poisson process, \( m_*(X) = \lambda \cdot V \) is not dependent on location \( X \), but on the size of the volume \( V \) only. Therefore, if we define a quadrat with size \( V \) and use the quadrat to sample the region, independently, we should obtain a serial of samples from a single Poisson distribution with mean \( \lambda \cdot V \).

Suppose that \( n \) number of samples marked as \( N_1, N_2, \ldots, N_n \), are obtained by the sampling. The most obvious test to see if they are from a Poisson distribution is to derive a histogram, hence a distribution, from the samples and use the goodness-of-fit testing technique. There is another quick and effective way to reach the same or similar conclusion which uses the equality property between mean and variance of a Poisson distribution. We do not, however, know the mean and variance of the Poisson distribution we are testing and therefore they can only be estimated from the samples. If \( \bar{N} \) is the mean value of the sample, the ratio between the variance and the mean of the samples, without any distribution assumption, is:

\[
r = \frac{1}{(n-1) \cdot \bar{N}} \sum_{i=1}^{n} (N_i - \bar{N})^2
\]

If a Poisson distribution is assumed, \( r \) should be close to 1 for the hypothesis to hold. Though this only tests the samples obtained, if the hypothesis holds, the logical extension will be that samples are from a homogeneous Poisson points field, provided sampling is representative and is done independently.

If the hypothesis being tested does not hold and we are still assuming Poisson points field, the only implication is that the Poisson process is not homogeneous. In this case, the ratio \( r \) helps to describe the degree of departure of the point pattern from homogeneity. It is for this reason, \( r \) in quadrat count analysis is given a special name, the index of dispersion, i.e.,

\[
\text{index of dispersion (IOD)} = \frac{1}{(n-1) \cdot \bar{N}} \sum_{i=1}^{n} (N_i - \bar{N})^2
\]
There are about half a dozen of different indices, such as index of cluster size, index of patchiness, index of mean crowding. For a more detail listing, see Upton []. These indices are derived more or less based on IOD and do not actually reveal more information about the point pattern and therefore they will not be discussed here.

Based on the definition, significantly large IOD indicates large variations for the number of points inside the quadrats which implies point aggregation. Significantly small IOD means the variations between quadrat count is small and the point pattern is more regular. Diggle [] suggest that IOD test is powerful against point aggregation, but weak against point regularity.

3.1.1 Goodness-of-fit test

There is also an alternative way to look at IOD. For a homogeneous Poisson process, the expected number of points within the quadrats located anywhere in the region is a constant $\lambda \cdot V$, where $V$ is the volume of the quadrat. Therefore, conditioned on the total point number, the Pearson’s goodness-of-fit criterion can be used to test the departure of the samples from this constant distribution, i.e.,

$$\chi^2 = \sum_{i=1}^{n} \left( \frac{N_i - \bar{N}}{\bar{N}} \right)^2$$

which can be approximated by $\chi^2_{n-1}$ on the condition that $n > 6$ and $\bar{N} > 1$. When this criterion is used, lower tail and upper tail tests are possible. Lower tail (significantly small $\chi^2$ value) is used to test against regularity and upper tail (significantly large $\chi^2$ value) is used to test against point aggregation.

Note $\chi^2 = (n-1) \cdot IOD$. In other words, these two analyses are equivalent.

3.1.2 Considerations for choice of quadrats and scheme of sampling

There is no restriction on the shape of quadrats though in general circles and rectangles are used for 2D region and sphere and cuboid are used for 3D region. As for the size of quadrats, different research are trying to derive the optimal quadrat size (see Upton []) but no agreement can be reached. This obviously is an application dependent issue but as a rule of thumb, the size of quadrats should be chosen such that the mean point count should be at least 1.

There are two issues for the sampling scheme: the number of quadrats to use and the locations of quadrats. Obviously there are a great range of choices here but the main concern for the selection is to satisfy the conditions of independent and representative samplings. For independent sampling, regularly spaced mutually exclusive quadrats can serve the purpose and for representative sampling, randomly located quadrats may be a better choice. It is also possible to impose the mutually exclusive condition to the random quadrats though the performance may become an issue when large number of quadrats are used.

Figure 3.1 shows the difference between the random quadrat sampling and contiguous sampling scheme. As can be seen, random sampling may not be able to produce totally independent samples because of the overlapping of some quadrats. If, however, the total number of quadrats used is not too high, the effect of sample dependency on the analysis results may not be severe. There is research in distance sampling analysis about the upper limit of the number of samples for independent
sampling, which is \( n = 0.1 \times N \), where \( N \) is the total number of Poisson points (Diggle []). For quadrat count analysis, no similar value is reported to my readings. Some research effort can be directed to the investigation into the limitation if it became an issue. For the contiguous quadrat sampling, the grid can float in the region to get the best representative samples. If performance is not an issue, covering the whole region with the sampling grid may be a safer choice. Random selection of a certain number of quadrats from a contiguous grid can also be used in some cases.

![Random quadrats and Contiguous quadrats](image)

Figure 3.1 Quadrats used for quadrat count analysis

### 3.1.3 Agglomerative quadrat count analysis

This is another kind of variance analysis other than IOD, which was introduced by Greig-Smith based on the contiguous quadrat data. The method starts by partitioning the whole region into \( m \times m \) regular quadrats. Neighbourhood quadrats are agglomerated into blocks. At the first step, blocks and quadrats are identical, i.e., number of quadrats in each block \( q = 1 \). At the second step, each block contains two quadrats, \( q = 2 \), which can be done by horizontal or vertical agglomeration. At the third step, each block contains four quadrats, \( q = 4 \), and so on. At each stage, the squares of number of points inside the blocks are summed to form quantity \( T_q \), expressed as follows:

\[
T_q = \sum_i (N_i^q)^2
\]

and the Greig-Smith variance is then calculated as:

\[
G_q = 2 \cdot T_q - T_{2q} \quad q = 1, 2, 4, 8, ...
\]

For homogeneous Poisson point pattern, \( G_q \) will be more or less constant. If there is clusters, however, \( G_q \) is claimed to be able to reach a peak at a value of \( q \) which indicates the cluster size. In my opinion, this claim may not necessarily be true as peaking in this case also depends on spatial arrangement of point pattern. Examples will be given in later section for this argument.

In our analysis, in order to eliminate the directional effect imposed by the original proposal, intermediate blocks showing differences when horizontal or vertical
agglomeration is used are not to be used. This leaves the Greig-Smith variance being calculated by:

\[ G_q = 4 \cdot T_q - T_{4q} \quad q = 1, 4, 16, 64, \ldots \]

### 3.2 Quadrat count analysis of simulated point patterns

In this section, we are going to present some quadrat count analysis for four different types of simulated point patterns, homogeneous, non-homogeneous, cluster and Cox Poisson processes. The intention here is trying to make the connection between the point patterns and the expected quadrat count analysis results.

#### 3.2.1 Homogeneous Poisson points

Figure 3.2 (a) shows a realisation of a homogeneous Poisson process. Five different types of quadrat count analysis are conducted. Figure 3.2 (b) is the analysis by random quadrat method where 50 random quadrats are generated for each quadrat size. Quadrat sizes in this case are measured in relative scale, i.e., percentage of the ranges of the region in horizontal or vertical directions. Quadrats with too small size will result in too low value of mean counts in the quadrats. Quadrats with too large size will increase the interaction between quadrats and hence the correlation between the samples in the random quadrats case, or result in too few total number of quadrats usable for sampling in the case of regular grid quadrats. Both of the cases should be avoided in the analysis. In the results presented below, the sizes of quadrats range from 0.01 (1%) to 0.5 (50%) of the size of the region.

![Figure 3.2 (a) Homogeneous points](image1.png)

![Figure 3.2 (b) By random quadrats](image2.png)
As can be seen from Figure 3.2 (b), the index of dispersion (IOD) remains more or less constant at the value of around 1.0 for different quadrat sizes. It also follows the curve of the average IOD based on the Monte Carlo simulation. The $\chi^2$ values for different quadrat sizes are also calculated and shown as the green curve in the Figure. The number of degrees of freedom in this case is 49 and the 95% confidence critical value is 67. As can be seen from the figure, most of the $\chi^2$ values are not significant. Only the $\chi^2$ value for the quadrat size of about 0.2 exceeds the critical value. This phenomenon, however, is not consistent, as it is not the case during other running sessions (i.e., with different random quadrats) and therefore the $\chi^2$ values for different quadrat sizes should be considered not significant. In the figure, mean count and quadrat count variance are also calculated and displayed. For homogeneous case, these two statistics will have quadratic increase with the size of quadrats, as can be seen from the figure.

For quadrat count analysis using regular quadrats, three different options are used. In the examples shown here, the size of the grid covers 80% of the region. The first option is for the grid to be fixed at certain location, for example, starting at 0.1 and ending at 0.9 (relative scales), and the result is shown in Figure 3.2 (c). The second option is for the whole grid to be located randomly in the region and the result is shown in Figure 3.2 (d). The last option is to fix the grid at certain location but only a certain proportion of the quadrats inside the grid are selected randomly for the
analysis and the result for this option is shown in Figure 3.2 (e). All results in these three figure show very similar features. Compared with random quadrat analysis, two interesting difference are obvious. Firstly, the \( \chi^2 \) values for small quadrat size are extremely significant implying that in small scales the point process is not homogeneous. This is true as any point process can be viewed as non-homogeneous if the scale used is small enough. The reason this feature does not show up in random quadrat analysis is considered to be due to the fixed number of random quadrats used (50 compared with about 6000 number of quadrats used in regular grid quadrat analysis) and in the case of small quadrat size, 50 samples may not be representative. Secondly, note the difference between the shapes of the 95% confidence envelopes for the homogeneous Monte Carlo simulation. The differences between the upper and lower 95% envelope values are vanishing (and it should be) as the quadrat size decreases in regular grid quadrat case but this is not so in random quadrat case. Non-representative samples or sample correlations may be the reason behind. These features suggest that in general, regular grid quadrat analysis should give a more reliable analysis result.

For the Greig-Smith analysis, the result is presented in Figure 3.2 (f). There is no significant variations in the variance except for the large quadrat sizes. For large quadrat sizes, however, the number of quadrats available for calculating the variances is normally small and the values are considered less reliable than the smaller quadrats cases. In this example, for quadrat sizes less than 25% of the size of the region, the variances are more or less constant, which implies no point aggregations detected according to the intentions of the analysis. Note also the empirical variances follow quite well the variances from Monte Carlo simulations.

3.2.2 Non-homogeneous Poisson points

For non-homogeneous Poisson process, we use the same example used in the distance analysis. Figure 3.3 (a) shows a realisation of a non-homogeneous Poisson process with the density defined as:

\[
\lambda(u,v) = 0.1 \cdot e^{-100(u-2v)^2}
\]

The corresponding quadrat count analysis results are shown from Figure 3.3 (b) to (f).
From these figures, it is not difficult to conclude from IOD values that serious departure from homogeneous point process has been detected by the analysis. The $\chi^2$ values are all extremely significant which also implies the departure of the point pattern from CSR. Greig-Smith variances also show the discrepancy from CSR but no sign of point aggregations.

One of the interesting points shown by these figures is for the cases of analysis when the quadrat sizes are small. For these cases, the results do not actually suggest departure of the point pattern from CSR. This can be considered as one of the weak points by quadrat count testing. In other words, the quadrat count analysis is not sensitive for small quadrat sizes.

### 3.2.3 Poisson cluster points

We will again use the same cluster process used for distance analysis. Figure 3.4 (a) is a realisation of a Poisson cluster process where the parent process is a homogeneous Poisson process with density $\lambda=0.005$, each parent produces a fixed number of 20 daughters and daughter points are uniformly distributed around their parent within a circle of radius of 5 and centred at their parent location. The realisation consists of daughter points only.

Figure 3.4 (b) – (f) display the results of the quadrat count analysis. Serious departure from CSR is again evident in all results.
Compared with the analysis results for non-homogeneous points, Figure 3.3, the considerable difference is the IOD values or G-S variances for small quadrat sizes. The greater values in these figures suggest there is more point aggregations in smaller scale in this example compared with the non-homogeneous case presented above. This is true as the cluster process creates point aggregations in the scale of 10 (or 0.1 in relative scale). From these figures, it is only possible to conclude that point aggregations happen in the scale roughly less than 20 (or 0.2 relative), but nothing more details can be obtained.
Quadrat count analysis is considered to be good at detecting cluster sizes but based on our examples this is not generally the case. Several factors contribute to the effectiveness of this detection. The most important factor is that the quadrats used are arranged in such a way that coincides with the locations of clusters. An idealised example is given in Figure 3.5 (a). In this example, the quadrats used happen to be in such a way that the main parts of most of point clusters are contained within the quadrats. In this case, there will be a peak value for IOD for this quadrat size (which is 10), as can be seen from Figure 3.5 (b). The cluster size in this case can easily identified from such a quadrat count analysis. However, in practice, this kind of quadrat arrangement is unlikely to be always the case in cluster point pattern analysis and therefore detecting cluster sizes by quadrat count analysis is not always reliable. For example, the same point pattern as Figure 3.5 (a) is analysed again using the same regular grid, but repositioned in a slightly different location as shown in Figure 3.5 (c), the analysis result is displayed in Figure 3.5 (d). As can be seen, the peak value of IOD implying the cluster size disappears all together. The similar result is obtained by random quadrat count analysis as displayed in Figure 3.5 (e) and (f). The cluster size in the later few cases fails to be detected.

Fig. 3.5 (a) Special case of clusters  
Fig. 3.5 (b) IOD results

Fig. 3.5 (c) Different grid location  
Fig. 3.5 (d) IOD results
If we just rearrange a few of the point clusters at different locations, as shown in Figure 3.5 (g), the peak value of IOD will also disappear even we use the same grid as used in Figure 3.5 (a), as shown in Figure 3.5 (h). This reveals a very serious disadvantage of using quadrat count analysis: it is not a reliable analysis in the sense that features can always be detected. It is far too sensitive to both the locations of clusters and the quadrats used for the analysis. Both factors must coincide with each other for the analysis to really reveal the cluster features.
It is the same story when Greigh-Smith variance analysis is used. For the point pattern displayed in Figure 3.5 (a), the Greigh-Smith variance analysis results for the grid size of 70x70 is given in Figure 3.5 (i). The peak value of the variance at the grid cell size of about 10 is obvious. However, this result again is sensitive to the changes in the grid used. Figure 3.5 (j) gives the analysis result for the same point pattern but using the 75x75 grid. As can be seen, the peak value present in Figure 3.5 (i) disappears. If we use the 70x70 grid to analysis the point pattern in Figure 3.5 (g), the peak value is not present either. This demonstrates the similar conclusion reached above that detecting cluster size by quadrat count analysis is not a reliable tool.

Quadrat count analysis can provide a reliable results for detecting the departure of the point pattern from CSR, but not for detecting the size of clusters.

### 3.2.4 Cox point pattern

For this analysis, we will use the same Cox process used for the distance analysis. The Cox model is defined as a normal distribution with mean and variance defined as follows:

\[
\begin{align*}
\text{mean}(u,v) &= 0.1 \cdot e^{-u-2v} \\
\text{variance}(u,v) &= 0.015 \cdot e^{-u-2v}
\end{align*}
\]

where \(u\) and \(v\) are the horizontal and vertical coordinates. At each location, the mean together with the variance that is about 15% the mean value, define a normal distribution for the density at that location. A random value is then generated from this distribution to serve as the realisation of the density field at the location. A realisation of this Cox process is given in Figure 3.6 (a).

---

**Fig. 3.6 (a) Cox points**

**Figure 3.6 (b) By random quadrats**
Results from quadrat count analysis are given from Figure 3.6 (b) – (f). These figures show similar features as the analysis for non-homogeneous case, Figure 3.3. Apart from the conclusion that there is serious departure of the point pattern from CSR, no other specific features are apparent.

3.3 Quadrat count analysis of two example point dataset
The two example data sets and their sub-sets presented in Figure 2.22 – 2.27 are analysed again using the quadrat count analysis discussed in the previous sections. The results are presented below:
Fig. 3.8 (a) Example 1 – set 1

Figure 3.8 (b) By random quadrats

Figure 3.8 (c) By regular quadrats - 1

Figure 3.8 (d) By regular quadrats - 2

Figure 3.8 (e) By regular quadrat - 3

Figure 3.8 (f) By Greig-Smith grid
Fig. 3.9 (a) Example 1 – set 2

Figure 3.9 (b) By random quadrats

Figure 3.9 (c) By regular quadrats - 1

Figure 3.9 (d) By regular quadrats -2

Figure 3.9 (e) By regular quadrat - 3

Figure 3.9 (f) By Greig-Smith grid
Fig. 3.10 (a) Example 2

Figure 3.10 (b) By random quadrats

Figure 3.10 (c) By regular quadrats - 1

Figure 3.10 (d) By regular quadrats - 2

Figure 3.10 (e) By regular quadrat - 3

Figure 3.10 (f) By Greig-Smith grid
Fig. 3.11 (a) Example 2 – set 1

Figure 3.11 (b) By random quadrats

Figure 3.11 (c) By regular quadrats - 1

Figure 3.11 (d) By regular quadrats - 2

Figure 3.11 (e) By regular quadrat - 3

Figure 3.11 (f) By Greig-Smith grid
These figures reveal nothing more than the conclusion that the point pattern being analysed have serious departure from CSR. Any point aggregation features can not be concluded from these results. Many of the figures presented here are solely for the purpose of giving a complete set of the analysis.

3.5 General conclusions
Quadrat count analysis is an effective tool for detecting the departure of the point pattern from CSR. It is weak for quantifying any underlying features of the point
pattern. Quite often, characteristics such as cluster size may not be able to be detected by this method.

4. K-function analysis
The last pattern analysis tool to be discussed in this report is the K-function analysis, which belongs to the category of second moment analysis of point density. The analysis is equivalent to the variogram analysis in geostatistical modelling and it reveals some valuable spatial correlations for the point density measurement.

4.1 Theoretical background
As stated, K-function is about point density, naturally a starting point of the analysis will be the definition of the point density, denoted as $\lambda(X)$:

$$\lambda(X) = \lim_{|dX| \to 0} \frac{E[N(dX)]}{|dX|}$$

where $X$ is the location variable, $dX$ is an infinitesimal volume containing location $X$, $N(V)$ is the number of points within volume $V$ and $E[..]$ is the expected value.

Similarly the second-order point density, denoted as $\lambda_2(X)$ is defined as follows:

$$\lambda_2(X,Y) = \lim_{|dX|,|dY| \to 0} \frac{E[N(dX) \cdot N(dY)]}{|dX| \cdot |dY|}$$

where $Y$ is also a location variable. Further, we define a covariance density $\gamma(X,Y)$ that is directly related to $\lambda(X)$ and $\lambda_2(X,Y)$, analogous to covariance definition of two random variables,

$$\gamma(X,Y) = \lambda_2(X,Y) - \lambda(X) \cdot \lambda(Y)$$

For stationary point processes, or homogeneous point processes, $\lambda(X)$ will be a constant value $\lambda$, independent of locations. For these cases, $\lambda_2(X,Y) \equiv \lambda_2(X - Y)$, $\gamma(X,Y) = \gamma(X - Y)$, i.e., $\lambda_2(X,Y)$ and $\gamma(X,Y)$ are also location independent. If the process is also isotropic, $\lambda_2(X,Y) \equiv \lambda_2(t)$ and $\gamma(X,Y) = \gamma(t)$, where $t$ is the distance between locations $X$ and $Y$.

Using these definitions, supposed $\lambda(X)$ can be evaluated accurately at all locations within the region $\mathcal{R}$ being considered, the random points are then transformed into a random field quantified by the density values. Tools such as geostatistics can then be used to analyse and model the variable. There are, however, two main points which obstruct us to go directly in this route for the modelling. Firstly, the estimation of point density $\lambda(X)$ is difficult to be conducted in an objective and accurate manner. Secondly, geostatistics only describes the model which is correct in a global scale and is lack of descriptions for local details. This characteristic will find the technique difficult to model sensibly some special point processes which require both global and local models. A handy example for this case is the cluster process which needs the global model to describe the distributions of clusters within the region, and the local model to describe the point distributions within clusters. This argument does not imply that geostatistics is not applicable in the case of point process modelling. It may be worthwhile to do some comparison analysis at later stage.
Meanwhile something suitable for modelling spatial correlations of point density for point process is needed. K-function is such a tool. A simple and practical definition for K function is as follows:

\[
K(t) = \frac{E[\text{number of further events within distance } t \text{ of an arbitrary event}]}{\lambda}
\]

For formal definition of K function based on reduced Palm distribution, please see Cressie \[\ldots\]. The reason that the definition includes the point density value as its denominator is for normalisation of the expression of \(K(t)\) by point density values, hence eliminating the scaling effect of the density value. In other words, \(K(t)\) corresponds to the expected number of further events within distance \(t\) of an arbitrary event when the point density is a unity. Take the stationary point process as an example. According to the definition of point density \(\lambda(X)\), the expected number of events within a volume of \(V\) can be expressed as:

\[
N(V) = \int_V \lambda(X) \cdot dX
\]

For stationary cases, \(N(V) = \lambda \cdot V\), and therefore the K function in these cases will be:

\[
K(t) = \frac{\lambda \cdot V(t)}{\lambda} = V(t)
\]

which is independent of density value \(\lambda\), i.e., same point patterns with difference only in density values can be described with the same model. For two dimensional case, \(K(t) = \pi t^2\).

Since K function is also a second order measure directly related to point density, their relation can be formerly established. Conditioned on a known arbitrary event located at \(X\), we can find the conditional probability of another event at location \(Y\) as:

\[
P\{N(dY) > 0 / N(dX) = 1\} = \frac{P\{N(dY) > 0, N(dX) = 1\}}{P\{N(dX) = 1\}}
\]

where \(dX\) and \(dY\) are infinitesimal volume centered at location \(X\) and \(Y\). If \(dY\) is done in such a way that only one event is possible inside the volume, we can then have the following relations:

\[
E\{N(dY) = 1 / N(dX) = 1\} = 1 \cdot P\{N(dY) = 1 / N(dX) = 1\} = P\{N(dY) = 1 / N(dX) = 1\}
\]

\[
E\{N(dY) = 1, N(dX) = 1\} = 1 \cdot P\{N(dY) = 1, N(dX) = 1\} = P\{N(dY) = 1, N(dX) = 1\}
\]

\[
E\{N(dX) = 1\} = 1 \cdot P\{N(dX) = 1\} = P\{N(dX) = 1\}
\]

The conditional expectation of the number of events within \(dY\) is then:

\[
E\{N(dY) / N(dX) = 1\} = 1 \cdot P\{N(dY) = 1 / N(dX) = 1\} + 0 \cdot P\{N(dY) = 0 / N(dX) = 1\}
\]

\[
= P\{N(dY) = 1 / N(dX) = 1\} = \frac{P\{N(dY) = 1, N(dX) = 1\}}{P\{N(dX) = 1\}} = \frac{E\{N(dY) = 1, N(dX) = 1\}}{E\{N(dX) = 1\}}
\]

\[
= \frac{dX \cdot dY}{E\{N(dX) = 1\}} \cdot dY \quad \rightarrow \frac{\lambda_2(X,Y)}{\lambda(X)} \cdot dY
\]

Note \(dY\) is only an infinitesimal volume centred at location \(Y\) and to get the total expected number of events within distance \(t\) of the location \(X\), which is \(\lambda \cdot K(t)\) according to definition, \(dY\) must be integrated over the ball centred at \(X\). For example, in two dimensional case, the integration must be done over the area of the circle centred at \(X\), i.e.,
\[ \lambda(X) \cdot K(X, t) = \int_0^t \int_0^{2\pi} \frac{\lambda_2(X, y)}{\lambda(X)} \cdot y \cdot dy \cdot d\theta = \frac{2\pi}{\lambda(X)} \int_0^t \lambda_2(X, y) \cdot y \cdot dy \]

Note \(dy\) here means the integration over \(y\) and is different to the meaning of \(dY\) discussed above. The above relation can be re-arranged as:

\[ \lambda_2(X, t) = \frac{[\lambda(X)]^2}{2\pi t} \cdot K'(X, t) \]

For stationary case, \(X\) can be dropped from the relation:

\[ \lambda_2(t) = \frac{\lambda^2}{2\pi t} \cdot K'(t) \]

For point process in \(d\)-dimensions, similar integration can be done and the result is given below:

\[ \lambda_2(t) = \frac{\lambda^2 \Gamma(1 + \frac{d}{2})}{d\sqrt{\pi d} \cdot t^{d-1}} \cdot K'(t) \]

This establishes the formal relation between \(\lambda_2(t)\) and \(K(t)\).

### 4.2 Estimation of \(K(t)\)

Analogous to the variogram in geostatistics, \(K(t)\) can be estimated empirically from point pattern realisations. As \(\lambda \cdot K(t)\) is defined as the expected number of further events within distance of \(t\) of an arbitrary event, a direct estimate can be obtained for a points realisation with \(N\) number of events as:

\[ \hat{K}(t) = \frac{1}{N} \sum_{i=1}^{N} M_i(t) = \frac{1}{N} \sum_{i=1}^{N} \text{(number of further events within distance } t \text{ of event } i) \]

To estimate \(M_i(t)\), the number of further events within distance \(t\) of event \(i\), we assign indicator values for all the events except event \(i\) as follows:

\[ I_{ij}(t) = \begin{cases} 1 & \text{if } d_{ij} \leq t \\ 0 & \text{otherwise} \end{cases} \quad \text{where } d_{ij} \text{ is the distance between event } i \text{ and event } j \]

Then,

\[ M_i(t) = \sum_{j=1}^{N} I_{ij}(t) \quad j \neq i \]

This leads to a simple estimate of \(K(t)\) as:

\[ \hat{K}(t) = \frac{1}{\lambda \cdot N} \sum_{i=1}^{N} \sum_{j=1}^{N} I_{ij}(t) \]

where \(\lambda\) can be replaced with the empirical intensity \(N/V(\mathcal{R})\). This estimate does not include pair of events for which event \(j\) is outside the region \(\mathcal{R}\) and is not observable. In other words, the edge effect is not taken into account and the estimate is biased. To obtain an unbiased estimate for \(K(t)\), we can use the guard volume techniques discussed in Section 2 of this report. This approach, however, effectively throw away a considerable amount of valuable points. Another approach is to take into account the conditional probability \(p_{ij}\) that event \(j\) is observed given that the distance between the event and the event \(i\) is \(d_{ij}\).
For two dimensional case, $p_{ij}$ can be calculated as the proportion of the circumference inside the region $\mathcal{R}$ of the circle centred at $i$ and with radius $d_{ij}$. As shown in Figure 4.1, $p_{ij}$ is the proportion of the solid circumference line over the whole circumference of the circle.

When the circle is fully enclosed by the region $\mathcal{R}$, $p_{ij}=1$. In other words, the edge does not affect the pair. When the circle is partially enclosed, $p_{ij}<1$, which means there are possibilities some events with the same distance as event $j$ to event $i$ are outside the region $\mathcal{R}$ and therefore the calculated indicator value $I_{ij}(t)$ must be compensated. To compensate $I_{ij}(t)$, the weight $w_{ij}=1/p_{ij}$ is used to increase the indicator value going into the calculation of $K(t)$.

For analysis purpose, the graph of $\{K(t)-\pi t^2\}$ vs $t$ is normally used instead of $K(t)$ vs $t$. Because $K(t)=\pi t^2$ for homogeneous Poisson point process, the plot of $\{K(t)-\pi t^2\}$ vs $t$ will be a horizontal line with the value of 0 for homogeneous cases. For cases other
than homogeneous one, the plot will demonstrate directly the degree of departure of the point pattern from CSR.

4.3 Implementation issues
The difficult part for unbiased estimate of $K(t)$ lies in the estimate of the edge correction weights, $w_{ij}$. For two dimensional case with a rectangular region, Diggle [] give the solution of $w_{ij}$ as follows:

$$
p_{ij} = \begin{cases} 
\cos^{-1}\left(\frac{\min(d_1, d_y)}{d_y}\right) + \cos^{-1}\left(\frac{\min(d_2, d_y)}{d_y}\right) & \text{if } d_{ij}^2 \leq d_1^2 + d_2^2 \\
\frac{\pi}{2} - \frac{\cos^{-1}\left(\frac{d_1}{d_y}\right) + \cos^{-1}\left(\frac{d_2}{d_y}\right)}{2\pi} & \text{if } d_{ij}^2 > d_1^2 + d_2^2
\end{cases}
$$

where $d_1 = \min(x_i, a-x_i)$, $d_2 = \min(y_i, b-y_i)$ and $a$ and $b$ are the sizes of the rectangles in $x$ and $y$ directions. $d_1$ and $d_2$ as defined are the shorter distance of $i$ to the two vertical edges and the two horizontal edge of the rectangular region.

The above equations are only applicable when $d_{ij}$ is in the range of $[0, \frac{1}{2}\min(a, b)]$, which quite often may not be adequate in practical analysis, especially when the region is an extremely slim (flat) rectangle (i.e., considerable difference in $a$ and $b$). This restriction provides only a partial edge correction for $K(t)$ evaluation in the region for small $t$ values, not the whole range of interest. We approach the calculation of $p_{ij}$ purely by numerical approximation described below.

As shown in Figure 4.1 (b), we divide the circumference of the circle into $L$ equal length segments. The probability $p_{ij}$ is then calculated as:

$$
p_{ij} = \frac{\sum_{L} (s_i)_{|x_i \text{ is inside } \mathcal{R}}}{\sum_{L} [(s_i)_{|x_i \text{ is inside } \mathcal{R}} + (s_i)_{|x_i \text{ is outside } \mathcal{R}}]}
$$

If the value of $L$ is large enough, the numerical method should provide a very good approximation. Acceptable $L$ value obviously depends on the size of the region. In most of the cases we use the value of 50 for $L$ and the differences between the numerical results and those calculated by the above equations for small $d_{ij}$ values are negligible.

The advantage of using numerical approximation is that the $p_{ij}$ can be calculated for large $d_{ij}$ value up till the maximum possible distance in the region. This is required for a complete edge correction for the evaluation of $K(t)$ in the region. As discussed above, there is possibility that $w_{ij}$ becomes unbounded for too large $d_{ij}$ values, which will push the corrected $K(t)$ values for large $t$ towards infinity. Another advantage of this approach is that the technique can be readily adapted for higher dimensional cases where no analytical solutions for $p_{ij}$ are not available. More on this point when we come to the stage to deal with three dimensional problems.
The second issue needs discussing is the evaluation of second order point density \( \lambda_2(t) \) and the covariance density \( \gamma(t) \). From Section 4.1, we understand the evaluations of these two densities involve the calculation of the derivatives of \( K(t) \), i.e., \( K'(t) \). As only discrete values of \( K(t) \) are available, \( K'(t) \) can only be evaluated numerically at the \( t \) where \( K(t) \) value is available. The average of the forward and backward derivatives at \( t \) is used as \( K'(t) \) at that distance.

As illustrated in Figure 4.2, the forward and back derivatives at distance \( t \) can be written as:

\[
K'(t)\bigg|_{\text{forward}} = \frac{K(t + \Delta t) - K(t)}{\Delta t}
\]

\[
K'(t)\bigg|_{\text{backward}} = \frac{K(t) - K(t - \Delta t)}{\Delta t}
\]

\[
K'(t) = \frac{1}{2}(K'(t)\bigg|_{\text{forward}} + K'(t)\bigg|_{\text{backward}})
\]

where \( \Delta t \) is the division increment used for the evaluation of \( K(t) \).

### 4.4 K-function analysis of generated point patterns

We will start the K function analysis on some point datasets generated from known point process. This should help us to build the correspondence between the characteristics of \( K(t) \), \( \gamma(t) \) and the underlying point process. At the current stage of the progress, only two dimensional analysis is available and will be presented.

#### 4.4.1 Homogeneous Poisson point process

For homogeneous Poisson process, the analytical solutions for \( K(t) \), \( \lambda_2(t) \) and \( \gamma(t) \) are simple. In this case, \( \lambda(X) = \lambda \), i.e., the point density is independent of point locations. From the definition and relations, the followings can be obtained:

\[
\begin{align*}
\lambda_2(t) &= \lambda^2 \\
K(t) &= \pi t^2 \quad \text{or} \quad K(t) - \pi t^2 = 0 \\
\gamma(t) &= 0
\end{align*}
\]

As can be seen, \( \gamma(t) = 0 \), which implies there is no correlation between point densities in different locations for this process.
The first data set generated is from a homogeneous Poisson process with $\lambda=0.02$ for the same region $(0,100) \times (0,100)$ used in previous demonstrations. Figure 4.3 (a) is one realisation of the process.

![Figure 4.3 (a) Homogeneous points](image1)

![Figure 4.3 (b) $[K(t)-\pi t^2]$ vs $t$](image2)

The functions of $K(t)$, $\lambda_2(t)$ and $\gamma(t)$ for this point pattern are given in Figure 4.3 (b). The $K$ function values from 100 Monte Carlo simulations are also plotted. The green and pink lines are the 95% confidence envelope of $[K(t)-\pi t^2]$ when the underlying process is homogeneous. As can be seen from the figure, $[K(t)-\pi t^2]$ from the data follows more or less the average value from the Monte Carlo simulations, and is oscillating around 0, implying homogeneous point pattern. $\gamma(t)$ is also oscillating around 0 which is also the behaviour of homogeneous Poisson point pattern as discussed above. Note this analysis is conducted up till the distance of 90.

To demonstrate the effect of edge correction on the evaluation of $K(t)$, $\lambda_2(t)$ and $\gamma(t)$, we have the following comparison. Figure 4.4(a) and (b) shows the empirical values of $K(t)$, $\lambda_2(t)$ and $\gamma(t)$ with edge correction imposed for a homogeneous Poisson process. Figure (4.4 (c) and (d) gives the evaluated values without edge correction. As can be seen, $K(t)$, $\lambda_2(t)$ and $\gamma(t)$ values without edge correction could be very misleading. Significant departure from values they should have is obvious after a very short distance. All functions suggest wrongly a non-CSR point pattern. The function values with edge correction, on the other hand, agree well with analytical results and all suggest the correct point process.

![Figure 4.4 (a) $[K(t)-\pi t^2]$ vs $t$ (with edge correction)](image3)

![Figure 4.4 (b) $K(t)$ vs $t$ (with edge correction)](image4)
4.4.2 Non-homogeneous Poisson point process

The same non-homogeneous Poisson process used in the previous analysis for the same region is also used here for the $K$ function analysis. The density function is defined as:

$$\lambda(u, v) = 0.1 \cdot e^{-\frac{u-2v}{100}}$$

Figure 4.5 (a) is one of the realisations and Figure 4.5 (b) is the evaluated values for $K(t)$, $\lambda_2(t)$ and $\gamma(t)$ with edge correction.

As can be seen, all functions $[K(t) - \pi t^2]$, $\lambda_2(t)$ and $\gamma(t)$ display significant departures from the values for CSR case. A few interesting points are worth listed:

- $[K(t) - \pi t^2]$ curve is of parabolic shape concaving upward. It increases as $t$ increases when $t$ is small, peaks at a certain distance (to be discussed) and then decreases as $t$ increases for large $t$ values. From definition, $K(t)$ is directly proportional to the expected number of points within an area of $\pi t^2$. Therefore a positive value of $[K(t) - \pi t^2]$ implies that the actual number within the area is greater than that to be expected if the points are evenly distributed in the region (i.e., homogeneous distribution of points). In other words, point aggregation occurs within the area $\pi t^2$ defined by the distance. Negative $[K(t) - \pi t^2]$, on the other hand, signifies the actual number is lower than that expected for even distribution case. This will always happen in the cases of large
distance $t$, hence large areas covered. These behaviours described are exactly the basic characteristics of non-homogeneous point distribution across the region as a whole.

- The distance $t$ when $[K(t) - \pi t^2]$ peaks represents a balancing point when the degree of point aggregations start decreasing as the area $\pi t^2$ increases. This point corresponds to the point where covariance $\gamma(t)$ changes signs from positive to negative, implying negative correlations between point density separated beyond this distance. After the balancing point, $[K(t) - \pi t^2]$ value continues decreasing as $t$ increases and eventually becomes negative, implying lower number of points than expected. For very large distances, $[K(t) - \pi t^2]$ becomes unbounded. This may be caused by the unbound property of the edge correction weight $w_{ij}$ and therefore should be discarded. Note in this example, the distance of this balancing point is about 50, half of the size of the edge of the rectangular region.

- As discussed, covariance density $\gamma(t)$ will change from positive to negative as $t$ increases. The distance where $\gamma(t) = 0$ represents the boundary within which correlation between the point density of two locations is positive. Density values of two locations separated more than this distance will have negative correlation between them. As mentioned above, this distance corresponds to the peak value of $[K(t) - \pi t^2]$. Recall from geostatistics, the structural analysis always suggests the range of influence beyond which the two random variables will no longer correlate, i.e., correlation coefficient = 0. This seems not to be the case in point density analysis. The influence of point density in one location on another location can either be positive or negative (except for homogeneous cases). The reason for this is due to the construction of the reference homogeneous (average) point distribution used for density analysis, which takes all points and the whole region. The $K$ function analysis is actually about the difference between the actual and the reference point patterns and therefore density variables in the whole region are correlated.

- The shape of the curve for $\gamma(t)$ may be important. It may be used to reveal the characteristics of the underlying density model of the point process. We will come back to this point in later discussions.

### 4.4.3 Poisson cluster process

Again we will use the same cluster process used in the previous analysis. The parent process in this is a homogeneous Poisson process with density $\lambda = 0.005$. Each parent produces a fixed number of 20 daughters. Daughter points are uniformly distributed around their parent within a circle of radius of 5 and centred at their parent location. The realisation consists of daughter points only.

(In the following analysis, I use the word “bump”. Please If you can think of a better suggestion).

Figure 4.6 (a) and (b) are the generated pattern and the results for empirical $K(t)$, $\lambda_2(t)$ and $\gamma(t)$. The most interesting feature of the graph is the bumpy characteristics. This is actually the feature unique to cluster point processes. To demonstrate this, we start with just one cluster, such as the one shown in Figure 4.7 (a). The $K(t)$, $\lambda_2(t)$ and $\gamma(t)$ for this point set are given in Figure 4.7 (b) and (c). As can be seen, one bump is present for $[K(t) - \pi t^2]$, but none for others. For two clusters, the results are shown in Figure 4.8 and two bumps are observed for $[K(t) - \pi t^2]$ and one for $\lambda_2(t)$ and $\gamma(t)$. For
three clusters, Figure 4.9 shows four bumps for \([K(t) - \pi t^2]\) and three for \(\lambda_2(t)\) and \(\gamma(t)\). In fact, for \(n\) clusters, the number of bumps present in the curve of \([K(t) - \pi t^2]\) vs \(t\) will be \(\frac{1}{2} n(n - 1) + 1\) and the \(\frac{1}{2} n(n - 1)\) for the curves of \(\lambda_2(t)\) and \(\gamma(t)\). This kind of behaviour can be explained below:

![Figure 4.6 (a) Cluster points](image1)

![Figure 4.6 (b) \([K(t) - \pi t^2]\) vs \(t\)](image2)

![Figure 4.7 (a) One cluster](image3)

![Figure 4.7 (b) \([K(t) - \pi t^2]\) vs \(t\)](image4)

![Figure 4.7 (c) \([K(t) - \pi t^2]\) vs \(t\)](image5)

![Figure 4.8 (a) Two clusters](image6)
For one cluster case, Figure 4.7 (a), when \( t \) is within the scale of the cluster size, the \( K \) function analysis is equivalent to the analysis of a small region \( A \) containing all the points. The results will be similar to that shown in Figure 4.4 (d) as edge correction is not an issue here. As \( t \) increases and gets over the scale of the cluster size, \( K(t) \) will stay unchanged as all points are already included. \( [K(t) - pt^2] \) will decrease but \( \lambda_2(t) \) will remain as zero because for any two locations separated by this distance, point density for one of the locations will always be zero. \( \gamma(t) \) will also remain constant when \( K(t) \) remain the same.

For two cluster case, Figure 4.8 (a), when distance \( t \) is within the scale of the cluster size, the behaviour of \( K(t) \), \( \lambda_2(t) \) and \( \gamma(t) \) are the same as in the single cluster case except the absolute function values. As \( t \) increases and before the area defined by \( t \) spans both cluster \( A \) and \( B \), \( K(t) \) and \( \gamma(t) \) will also stay unchanged and \( \lambda_2(t) \) will remain as zero because for any two locations separated by this distance, point density for one of the locations will always be zero.
remain as zero. As the area starts include points from both clusters at the same time, values for these functions will start increasing until the maximum separate distance (in the example 58) is reached. After that, $K(t)$ and $\gamma(t)$ will again stay unchanged and $\lambda_2(t)$ will again become zero.

As for three cluster case, Figure 4.9 (a), $K(t)$, $\lambda_2(t)$ and $\gamma(t)$ have the same behaviour as those discussed above. Only the absolute values of the functions are different. As distance $t$ increase, $K(t)$, $\lambda_2(t)$ and $\gamma(t)$ will remain unchanged before the area defined by $t$ can possibly cover points from at least two out of the three clusters. The function values will then start increasing until they stabilise at another level. Note in this case, there are three different distances separating the clusters and therefore three further stabilising stages, or three further bumps can be observed.

For the case of $n$ clusters, similar characteristics as those discussed above can be expected for $K(t)$, $\lambda_2(t)$ and $\gamma(t)$. As mentioned above, the total number of bumps is equal to $\frac{1}{2}n(n-1)+1$ for $[K(t)-\pi^2]$ and $\frac{1}{2}n(n-1)$ for $\lambda_2(t)$ and $\gamma(t)$. The actual numbers, however, may be different depending on the cluster distributions across the region. For example, if all clusters are separated by the same distance only one bump will be observed as points from all clusters will come into effect at the same distance. Another possible case will be when the differences in distances between clusters are small or there are too many clusters in the region, it will not possible to distinguish two or more very close bumps. In extremely case when $n\rightarrow\infty$, i.e., there are infinite number of clusters separated by all possible distances, an infinite number of bumps will make up the curves which will actually come out as smooth curves, i.e., no bumps at all.

The number of bumps for $[K(t)-\pi^2]$ is always one more than $\frac{1}{2}n(n-1)$. The first bump corresponds to the behaviour of points within clusters and the rest of $\frac{1}{2}n(n-1)$ bumps are the behaviours between points from different clusters. Therefore, the first bump provides a handy tool for the estimation of average size of the clusters. For example, all the figures given above show correctly the cluster size of about 10 for the point process.

In most cases this estimation is correct. It still can be used when the distance between clusters is less than the cluster size as in this case clusters join together to form large clusters and the first bump can still be used to estimate averaged “joined” cluster size. The only exception is when all clusters mix together to form a “smeared” picture of points in the region so that point clusters visually disappears. In this case, there may be still the first lump in the $[K(t)-\pi^2]$ curve, which may or may not correctly identify the size of the underlying point clusters, or the first bump may not show up at all. To demonstrate this point, look at the following examples. Figure 4.10 (a) is realised from a cluster process with cluster radius of 20 (cluster size = 40) and as can be seen from Figure 4.10 (b), the cluster size is correctly identified from the first bump in the $[K(t)-\pi^2]$ function. When the number of cluster is increased, however, the cluster patterns become mixed up and a smeared point pattern is obtained, as shown in Figure 4.10 (c). The $[K(t)-\pi^2]$ function shown in Figure 4.10 (d) fails to identify any cluster effect at all. It only reveals a non-homogeneous point process.
The feature of the first bump in the $[K(t) - \pi t^2]$ curve discussed above is only broadly correct if the cluster pattern is the dominant characteristics within the region. In point processes, point clusters are present but the dominant features of the whole point pattern is something else such as a non-homogeneous point process. In this case, $[K(t) - \pi t^2]$ curve will still give the bump features discussed above but the first bump disappears as for the distance in the scale of the cluster size, non-homogeneous point process is dominant. Two examples are given in Figure 4.11 where the parent process is a non-homogeneous process with density $\lambda = f(x)$. 10 or 20 daughter points are generated for each parent and they are distributed uniformly within the circle of radius $r = 5$ and centred at parent point. As can be seen from 4.11 (b) and (d), the first bump of the $[K(t) - \pi t^2]$ curve corresponding to the cluster size fails to show up clearly. The curves do still preserve the bump features. Care should be exercised in estimating the cluster size when features other than clustering (such non-homogeneity) are dominant in the point pattern.
From the theoretical side, the $K(t)$ function for a cluster process can be expressed as follows (see Diggle [{\textsuperscript{[1]}}]):

$$K(t) = \pi t^2 + \frac{E[S(S-1)] \cdot H_2(t)}{\rho \cdot E[S]}$$

where $\rho$ is the parent point density, $S$ is the number of daughters per parent, $E[\cdot]$ is the expectation and $H_2$ is the distribution function of the PDF $h_2$ defined as:

$$h_2(Y) = \int h(X) \cdot h(X - Y) \cdot dX$$

and $h_2$ is the PDF of daughter points relative to their parents. For example, if each parent produces a Poisson number of daughter points and daughter points are distributed around their parents according to the bi-variate model:

$$h(u, v) = \frac{1}{2\pi \sigma^2} e^{-\frac{u^2 + v^2}{2\sigma^2}}$$

$\sigma$ is the dispersion variance of daughter points. The $K(t)$ function for this process can be deduced as (see Cressie [{\textsuperscript{[2]}}]):

$$K(t) = \pi t^2 + \frac{1}{\rho} \left[1 - e^{-\frac{t^2}{4\sigma^2}}\right]$$

I think this theoretical solution, however, only takes into account the clustering effect in the scale of the cluster size, i.e., clustering of points from the same parent. The clustering or non-homogeneous effects from points in different clusters are only approximated with a homogeneous term $\pi t^2$ in the equation. In other words, the
The equation is only correct up till the distance $t$ equal to the average size of the clusters. Take the example shown in Figure 4.6. The theoretical curve based on the above equation is only roughly correct up to the distance 10, which is the average size of clusters within the region, as shown in Figure 4.12 below.

4.4.4 Cox point pattern

The same Cox process for the same region used in previous analysis is also used here for $K$ function analysis. The Cox model is defined as a normal distribution with mean and variance defined as follows:

$$
\begin{align*}
\text{mean}(u,v) &= 0.1 \cdot e^{-\frac{u-2v}{100}} \\
\text{variance}(u,v) &= 0.015 \cdot e^{-\frac{u-2v}{100}}
\end{align*}
$$

Figure 4.13 (a) is a realisation of the process and Figure 4.13 (b) shows the corresponding results for $K(t)$, $\lambda_2(t)$ and $\gamma(t)$. The first impression the figure gives is that these curves look very similar to those derived for non-homogeneous process. In fact, the core element of this Cox process is the same as the non-homogeneous density model used for the example of Figure 4.5. The Cox model, in addition, adds another random component to the process and therefore push the point pattern toward more homogeneous. This can be proved by the visually inspecting the point pattern of Figure 4.13 (a) and Figure 4.5 (a), or it can also be proved by the absolute $[K(t)-\pi t^2]$ values of the two point pattern. The discrepancy between $[K(t)-\pi t^2]$ value of the Cox point pattern and the homogeneous case shown in Figure 4.13 (b) is smaller compared to the value in Figure 4.5 (b), and thus reveals a more homogeneous pattern for Figure 4.13 (a). The same conclusion can also be reached by comparing the covariance value in Figure 4.13 (b) and Figure 4.5 (b). The difference between the absolute $\gamma(t)$ value with 0 (for homogeneous case) of the Cox process is smaller compared to that of the non-homogeneous process. Note the $\gamma(t)$ values displayed in Figure 4.13 (b) and Figure 4.5 (b) include the scaling effect of $\lambda^2$. To make the two figures comparable (for $\gamma(t)$), the $\gamma(t)$ curve must be divided by the scaling factor $\lambda^2$ which will be different in these two cases. For the Cox process (Figure 4.13), $\lambda = 0.045$ and $\lambda^2 = 0.002025$. For the non-homogeneous process (Figure 4.5), $\lambda = 0.023$ and $\lambda^2 = 0.000529$. The $\gamma(t)$ curve shown in Figure 4.13 (b) should therefore be scaled down by a factor of 3.8 to be comparable to the $\gamma(t)$ curve in Figure 4.5 (b).
4.5 K-function analysis of two example point datasets
We now turn to the $K$ function analysis of the two actual datasets used in the previous analysis. See the reference for the sources of the datasets.

4.5.1 Data set 1
The $K(t)$, $\lambda_2(t)$ and $\gamma(t)$ function results for the whole of dataset 1 and the two subsets are given in Figure 4.14 below.
These figures support the obvious suggestion that the process is non-homogeneous. They also reveal that the sub-set 1 points are more homogeneous than sub-set 2 points, i.e., there is more point aggregation in sub-set 2, which can be concluded from the $K(t)$ values of the two sub-sets. This is also obvious from the direct visual inspection of the point pattern.

4.5.2 Data set 2
The $K(t)$, $\lambda_2(t)$ and $\gamma(t)$ function results for the whole of dataset 2 and the two subsets are given in Figure 4.15 below. $[K(t)-\pi^2]$

Figure 4.15 (a) Whole dataset -2

Figure 4.15 (b) $[K(t)-\pi^2] $ vs $t$

Figure 4.15 (c) Sub-set 1

Figure 4.15 (d) $[K(t)-\pi^2] $ vs $t$

Figure 4.15 (e) Sub-set 2

Figure 4.15 (f) $[K(t)-\pi^2] $ vs $t$

Again, these figures show features of non-homogeneous process. This example, however, demonstrate the effect of boundary surrounding the points. By inspecting visually Figure 4.15 (a) and (e), it is not difficult to suggest that by some polygonal boundary the point process may be homogeneous, such as one shown in Figure 4.5
(e), but the overall behaviour of the point pattern is non-homogeneous. This may be a totally misperception but nevertheless it is worth some more investigation.

4.6 General discussions
The $K$ function analysis presented in this section is aiming for two purposes:
- Building the correspondence between the function characteristics and the known point pattern
- Proposing the most suitable point process model based on empirical functions.

So far we only demonstrate clear correspondence for two point processes: homogeneous and cluster processes, which in general show clear characteristics in the second order functions. For homogeneous process, \[ K(t) - \pi t^2 \] and \( g(t) \) are varying around zero. For cluster process, the “bump” features can be expected for \[ K(t) - \pi t^2 \], \( \lambda_2(t) \) and \( \gamma(t) \) and the first bump is of particular importance as it can be used to estimate the average cluster size in the process.

For non-homogeneous and Cox processes, second order functions show very similar features which imply that both processes are all non-homogeneous in nature. In other words, an inhomogeneous point pattern can be either modelled by non-homogeneous Poisson model or by the Cox model. Both, if modelled accurately, should statistically give the same answer (on average). Apart from the extra modelling component (freedom) provided by Cox process, Cox modelling is no difference to the non-homogeneous modelling. To illustrate this point, we use a one dimensional example.

In Figure 4.16 (a), a non-constant point density is to be modelled by a non-homogeneous process. In this case, the model $f(X)$ selected should coincide with the actual $\lambda(X)$ for accurate modelling. The specification of $f(X)$ may prove difficult in practical situations and therefore this leads to another choice (out of large number of possible approaches). We can model the “trend” of $\lambda(X)$ by a simple model $g(X)$ and then model the “residuals” by a simple noise model $g'(X)$. $g(X)$ and $g'(X)$ are selected in such a way that their accumulative effect statistically reproduce correctly the original $\lambda(X)$, the process is illustrated in Figure 4.16 (b). The first modelling technique is the direct non-homogeneous modelling and the second one is the Cox modelling.
Nevertheless non-homogeneous model is the foundation of both the modelling processes. It will be necessary at some stage to build the possible correspondence between the characteristics of second order function and some know $\lambda(X)$ (such as linear or quadratic intensity relations). In other words, if similar features are present in the $[K(t)-\pi t^2]$, $\lambda_3(t)$ and $\gamma(t)$ functions, the corresponding intensity function can then be used for the modelling.

The second possible improvement for modelling process can be achieved by coordinate transformation. Take the sub-set 1 of the first example dataset for instance. If the coordinate system is rotated 40° clockwise and shifted as shown in Figure 4.17 (a) below. The $[K(t)-\pi t^2]$, $\lambda_3(t)$ and $\gamma(t)$ functions calculated for the rotated case are shown in Figure 4.17 (b). By comparing this figure with Figure 4.14 (d), it is interesting to notice that the point process tends more homogeneous. In some cases, it is possible to transform a non-homogeneous point pattern to a homogeneous case simply by coordinate transformation and then the process can be modelled more easily and more accurately by the homogeneous process. This seems to contradict the fundamental assumption we have for the point process: stationarity, which states that after simple coordinate transformation characteristics of the process should stay the same. As a matter of fact, there is no contradiction here. The coordinate transformation we are using here is simply to get rid of the white spaces where points do not occupy. With those blank spaces inside the region, the point pattern is inhomogeneous in the scale of the whole region. With white spaces taken out, their effect of introducing non-homogeneity disappears. Simple coordinate transformations do not change the internal structures of the point pattern. For example, the transformation does not change the point aggregation shown in Figure 4.14 (e). This is demonstrated in Figure 4.17 (c) and (d) below.
Simple coordinate transformation does not cut out all the white spaces occupied within the region. This suggests another possible approach to improve the modelling. If a polygonal boundary is defined around the region of interest, it is possible to use a simpler point process for easier and more accurate modelling of the point pattern within the polygonal region. Two examples are given in Figure 4.18 below, where (a) is for sub-set 1 of example data set 1 and (b) is for whole dataset 2. The program we have has not yet implemented this technique yet. However, only by visual inspection, it is not difficult to suggest points within the defined polygon(s) can be modelled by homogeneous point process.

5. References
In the following figures, we try to build the correspondence between the characteristics of second order functions and some known $\lambda(X)$. In other words, in similar features are present in the $[K(t)-\tau_2^2]$, $\lambda_2(t)$ and $\gamma(t)$ functions, the corresponding density function can then be used for modelling.
Figure 4.15 (e) Sub-set 2

Figure 4.15 (f) $[K(t) - \pi t^2]$ vs $t$