Estimation of Clustering Parameters Using Gaussian Process Regression

Paul Rigby1, Oscar Pizarro2, Stefan B. Williams2

1 Australian Institute of Marine Science, Townsville, Queensland, Australia, 2 Australian Centre for Field Robotics, The University of Sydney, Sydney, New South Wales, Australia

Abstract

We propose a method for estimating the clustering parameters in a Neyman-Scott Poisson process using Gaussian process regression. It is assumed that the underlying process has been observed within a number of quadrats, and from this sparse information the distribution is modelled as a Gaussian process. The clustering parameters are then estimated numerically by fitting to the covariance structure of the model. It is shown that the proposed method is resilient to any sampling regime. The method is applied to simulated two-dimensional clustered populations and the results are compared to a related method from the literature.

Introduction

In ecological studies, consideration of spatial structure can lead to useful insights regarding the process of interest (see, for example [1–3]). If the process exhibits clustering, then parameterised models which describe the spatial structure can provide a greater understanding of the behaviour of the species [4]. However, accurately determining parameters that describe a population can be a difficult problem within the natural environment. Measurements are often expensive and difficult to make, and so usually only a sparse sample of a population is available. In this paper we propose a new method for estimating the parameters of a cluster process from a sparse set of quadrat samples with arbitrary design, i.e. any sampling design such as transects, random etc. can be used without damaging the estimator.

The key theoretical framework upon which this work is built is the Gaussian process (GP), which can be defined as a collection of random variables, any finite number of which have a joint Gaussian distribution [5]. The Gaussian process framework can provide a useful tool for modelling stochastic processes and has seen much attention recently within the machine learning community, where it is used to solve regression and classification problems. GP regression has been widely studied and applied to naturally occurring populations (see, for example, [10–13]). The estimation is performed by fitting a theoretical covariance function to the empirical GP counterpart by numerical optimisation.

The standard approach to estimating the parameters of a cluster process is based upon the K-function developed by [14]. This estimation procedure assumes that the spatial process has been mapped over the whole survey area. This is not always practical and so alternative methods have been developed based upon line transect surveys. Most of these have only been developed to estimate the mean intensity of a population based upon a partial mapping, and cannot separate all of the parameters. However, Cowling [4] developed a method for estimating all the parameters in a two-dimensional Neyman-Scott process based upon a one-dimensional K-function along the transect line. An error in Cowling’s derivation of the K-function was corrected in [15], where it was concluded that the corrected method clearly outperformed competing methods.

In this paper an alternative method for Neyman-Scott parameter estimation is developed, based upon GP regression. An experimental arrangement similar to Cowling is used, with the same transect sampling design. Experiments with alternative sampling designs are then tested to demonstrate the resilience of the technique.

Methods

Related Work

Within the ecological literature the standard approach to estimating the parameters of a cluster process based upon Ripley’s
K-function [14]. For a stationary isotropic process with intensity \( \lambda \), it is defined as

\[
K(h) = \tau^{-1} E[\text{number of events within distance } h \text{ of an arbitrary event}]
\]  

The K-function for a Neyman-Scott process of dimension \( d \) is given by Cressie [16]

\[
K(h) = \frac{\pi^{d/2} \lambda^{d/2}}{\Gamma(1 + d/2)} \exp\left[ \frac{E[\mu(\mu-1)] F(h)}{\lambda E[\mu]^2} \right], \quad h > 0
\]

where \( \lambda \) is the intensity of the parent process, \( \mu \) is the number of events per cluster and \( F(h) \) is the distribution function for the distance between two events in the same cluster.

If \( K(h, \mu, \lambda) \) is the K-function evaluated at estimates of \( \mu \) and \( \lambda \), and \( \hat{K}(h) \) is a nonparametric estimator obtained from the data, then a least-squares estimate of \( \mu \) and \( \lambda \) is obtained by minimising the ad-hoc criterion:

\[
D(\hat{\mu}, \hat{\lambda}) = \int_0^{h_0} \{ (\hat{K}(h))^2 - (K(h, \hat{\mu}, \hat{\lambda}))^2 \} \, dh
\]

where \( c \) and \( h_0 \) are tuning constants.

The above estimation procedure assumes that the spatial process has been mapped over the whole survey area. This is not always practical and so Cowling and Aldrin [4,15] developed a method for estimating all the parameters in a two-dimensional Neyman-Scott process based upon a one-dimensional K-function along a transect line. The key steps in this method are reproduced below.

Cowling introduces a normal detection function \( g(x) \), which is the probability of detecting an offspring at a distance \( x \) from the transect line.

\[
g(x) = g_0 \exp\left(-\frac{x^2}{2\sigma^2}\right)
\]

where \( g_0 \) and \( \sigma \) are typically estimated from external data, and are assumed known. The K-function for the detected points projected onto the transect line is then derived as

\[
K_1(h) = 2h + \frac{2\Phi(h/\rho \sqrt{2}) - 1}{2\sqrt{\pi \lambda \sigma^2 + \rho^2}}
\]

where \( \Phi \) is the distribution function of the standard normal distribution. The empirical K-function is given by

\[
\hat{K}_1(h) = 2n^{-1} L \sum_j \sum_{i<j} I(|y_i - y_j| < h)
\]

where \( L \) is the length of the transect, \( n \) is the number of detected points, and \( y_i \) and \( y_j \) are the positions along the transect line. The parameters \( \lambda \) and \( \rho \) are then estimated by fitting the theoretical K-function to its empirical counterpart. Furthermore, given that \( \mu \) can be estimated by substituting \( E(n) \) by the observed \( n \), and \( \lambda \) and \( \sigma \) by their estimates.

### Gaussian Process Regression

In this section the GP regression methodology is outlined. For a fuller explanation the reader is directed to [5]. Consider a dataset \( D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \) \( x \in \mathbb{R}^d \)

which contains \( n \) observations of some scalar variable \( y \), taken at locations \( x \). The dataset in equation 8 can be more compactly represented as \( D = \{X, y\} \) where \( X \) is an \( n \) by \( d \) array of measurement locations, which will be referred to as training points, and \( y \) is a vector of the observations at those locations. Similarly, if predictions are to be made at more than one location, \( X \), refers to an \( m \) by \( d \) array of test points, and \( y_\ast \) is the predicted output at those locations.

The distribution of the training outputs and the test outputs is jointly Gaussian with dimension \( n + m \), mean \( \mu \) and covariance \( \Sigma \)

\[
\begin{bmatrix} y \\ y_\ast \end{bmatrix} \sim \mathcal{N}(\mu, \Sigma)
\]

\[
\equiv \mathcal{N}\left( \mu, \begin{bmatrix} K(X, X) + \sigma^2_I & K(X, X_\ast) \\ K(X_\ast, X) & K(X_\ast, X_\ast) \end{bmatrix} \right)
\]

where \( \mu \) is a matrix containing the mean function evaluated at each of the \( n \) training points and \( m \) test points.

\[
\mu = [\mu(x_1) \ldots \mu(x_n), \mu(x_{n+1}) \ldots \mu(x_{n+m})]^T
\]

The covariance matrix in equation 9 has been partitioned to give the covariance matrices between the test points \( K(X_\ast, X_\ast) \), training points \( K(X, X) \), the cross covariance between both sets \( K(X, X_\ast) \) and its transpose \( K(X_\ast, X) \). The values of \( y \) in the data set are not the actual function values, only noisy realisations of them. To account for this \( \sigma_n^2 \) is added to the leading diagonal of the training covariance matrix.

Both the conditional and marginal distributions of a joint Gaussian distribution are Gaussian. It is this property which makes the Gaussian distribution appropriate for stochastic modelling, as closed form expressions for these distributions can be derived. Because \( y \) is known, it is possible to condition the joint Gaussian prior distribution on the observations [17] to give expressions for the mean and variance of the posterior GP:

\[
p(y_\ast | X, y, X_\ast) \sim \mathcal{N}(y_\ast, \Sigma_\ast)
\]

where

\[
y_\ast = K(X_\ast, X) [K(X, X) + \sigma_n^2 I]^{-1} y
\]
As in the case of kriging, the appropriateness of the GP model is entirely dependent on the choice of covariance function which has form arbitrarily selected by the user. Within GP literature [5] a popular form for the covariance function is the squared exponential

\[
K(x_i, x_j) = \sigma_f^2 \exp\left(-\frac{1}{2l^2} |x_i - x_j|^2\right)
\]

(14)

where \(l\) is a length scale that determines the strength of correlation between points. As two points are separated by a large difference, the covariance will tend to zero and the GP variance at test points far from measurements will tend to the underlying variance of the function \(\sigma_f^2\). The parameters \(\sigma_n, \sigma_f\) and \(l\) can all be varied, and doing so will affect the resulting GP model. The free parameters associated with any form of covariance function are referred to as hyperparameters. Also of interest is the marginal likelihood \(P(y|X)\) which can be obtained directly by considering that \(y \sim N(0, K(X, X) + \sigma_n^2 I)\).

\[
\log P(y|X) = -\frac{1}{2} Y^T (K(X, X) + \sigma_n^2 I)^{-1} Y - \frac{1}{2} \log |K(X, X) + \sigma_n^2 I| - \frac{n}{2} \log 2\pi
\]

(15)

The marginal likelihood, or its logarithm, gives a measure of how well the covariance function explains the training data. The absolute value of the log marginal likelihood (LML) is dependent on the dataset but for any given dataset the LML can be used to compare different forms of covariance function and tune the hyperparameters.

Essentially, instead of trying to fit a parametrised model to the underlying function \(f\), the GP (or the geostatistical) approach uses

\[
\Sigma_n = K(X_n, X_n) - K(X_n, X) [K(X, X) + \sigma_n^2 I]^{-1} K(X, X_n)
\]

(13)

Figure 1. Two quadrats (shaded grey boxes) sampling a clustered distribution. Consider a realisation of a Neyman-Scott process where invisible parent events independently produces of children. The survey field is of size \(W \times W\) and is to be sampled using square quadrats of side \(w\), thus the measured quantity is the number of children observed in each quadrat.

\[\text{doi:10.1371/journal.pone.0111522.g001}\]
a parametrised model to describe the covariance. The key assumption is that the covariance of the process can be described using a simple parametric model.

As one would expect, in equation 12 the estimate of $y$ is a weighted average of the observations $y$. The form of equation 13 is also intuitive. The term $K(X,X)$ represents the prior covariance between the test points before any observations are made. When observations are made at locations $X$ then the covariance in the prior will be decreased. The extent of this decrease depends upon the correlation between the observation locations and the test points, this is captured in $K(X,X)$. As a covariance function is by definition positive semi-definite, observations always result in information gain. However, these observations are also correlated with each other, and not as informative as would be expected if they were independent. Hence the inverse term $[K(X,X) + \sigma^2 I]^{-1}$ decreases the information gain accordingly. It is also interesting to note that $y$ does not appear in equation 13, thus the variance only depends on the location of the observations, not on the value of the observations themselves.

The main computational burden in computing the mean function and the variance from equations 12 and 13 comes from the inversion of the training point covariance matrix, $K(X,X)$.

**Estimation of Clustering Parameters**

Consider a realisation of a Neyman-Scott process where invisible parent events are Poisson distributed with intensity $\lambda$ per unit area; each parent independently produces a Poisson
Table 1. Population parameters.

| Population | \( \lambda \) | \( \rho \) | \( \mu \) |
|------------|--------------|------------|---------|
| s1         | 0.02         | 2.0        | 12.5    |
| s2         | 0.005        | 2.0        | 50.0    |
| s3         | 0.00125      | 2.0        | 200.0   |
| s4         | 0.02         | 4.0        | 12.5    |
| s5         | 0.005        | 4.0        | 50.0    |
| s6         | 0.00125      | 4.0        | 200.0   |
| s7         | 0.02         | 8.0        | 12.5    |
| s8         | 0.005        | 8.0        | 50.0    |
| s9         | 0.00125      | 8.0        | 200.0   |

Parameters of simulated clustered populations.
doi:10.1371/journal.pone.0111522.t001

Figure 3. Intensity per unit quadrat for one realisation of populations s1…s9. The point process is converted into the intensity map by dividing the survey into unit quadrats and counting the number of events lying within each quadrat.
doi:10.1371/journal.pone.0111522.g003
distributed number of children with intensity \( \mu \); the positions of the children relative to their parents are independent and have an isotropic bivariate normal distribution with variance \( \rho^2 \) in the \( x \) and \( y \) directions.

The survey field is of size \( W \times W \) and is to be sampled using square quadrats of side \( w \), thus the measured quantity is the number of children observed in each quadrat. \( N \) is the unknown number of clusters within the region \( W^2 \), centred on parents

\[
\{ (px_1, py_1), (px_2, py_2), \ldots, (px_N, py_N) \}.
\]

The covariance of the counts \( c_1 \) and \( c_2 \) measured in two quadrats covering regions \( Q_1 \) and \( Q_2 \), separated by distance \( h \) (see Figure 1) is defined as:

\[
\text{cov}(c_1, c_2) = E[(c_1 - E[c_1])(c_2 - E[c_2])].
\]

Let \( \Phi_i \) be a bivariate Gaussian distribution centred at a parent location \( (px_i, py_i) \)

\[
\Phi_i = \frac{1}{2\pi\rho^2} \exp \left[ -\frac{1}{2\rho^2} ((x-px_i)^2 + (y-py_i)^2) \right]
\]

then the expected number of children from cluster \( i \) that will fall within \( Q_1 \) is given by

\[
E[c_{1,i}] = \mu \int_{Q_1} \Phi_i \, dx \, dy
\]

and the covariance between two quadrats separated by distance \( h \) can be obtained by substituting into equation 16:

\[
\text{cov}(c_{1,i}, c_{2,j}) = \mu \int_{Q_1} \Phi_i \, dx \, dy \int_{Q_2} \Phi_j \, dx \, dy
\]
Expanding this expression gives

$$cov(c_1, c_2) = E \left[ \mu \sum_{i=1}^{N} \int q_1 \left( \Phi_i - \frac{1}{W^2} \right) \, dx \, dy \right]$$

$$\left\{ \mu \sum_{i=1}^{N} \int q_2 \left( \Phi_i - \frac{1}{W^2} \right) \, dx \, dy \right\}$$

which because of the independence of the cluster locations can be simplified to

$$cov(c_1, c_2) = E \left[ \mu^2 \sum_{i=1}^{N} \int q_1 \left( \Phi_i - \frac{1}{W^2} \right) \, dx \, dy \right]$$

$$\left\{ \int q_2 \left( \Phi_j - \frac{1}{W^2} \right) \, dx \, dy \right\}$$
The placement of cluster $i$ is arbitrary, and could be anywhere within the region $W^2$ with equal probability. Evaluating the expectation operator, and noting that $N = \lambda W^2$, gives

$$
\int \int q_i \left( \Phi_i - \frac{1}{W^2} \right) dx dy
$$

The placement of cluster $i$ is arbitrary, and could be anywhere within the region $W^2$ with equal probability. Evaluating the expectation operator, and noting that $N = \lambda W^2$, gives

$$
cov(c_1, c_2) = \mu^2 \lambda \int \int W^2 \left( \Phi_i - \frac{1}{W^2} \right) dx dy
$$
Table 2. Neyman-Scott Parameter Estimates.

| Population | Parameter | Value | GP Estimator | Cowling Estimator |
|------------|-----------|-------|--------------|-------------------|
| s1         | \( \lambda \times 10^2 \) | 2.00  | 2.20 (1.21)  | 3.40 (7.62)       |
|            | \( \rho \)           | 2.00  | 2.14 (0.58)  | 1.87 (0.90)       |
| s2         | \( \lambda \times 10^1 \) | 1.25  | 1.36 (0.60)  | 1.37 (0.84)       |
|            | \( \rho \)           | 5.00  | 4.94 (1.81)  | 4.23 (1.98)       |
| s3         | \( \lambda \times 10^3 \) | 5.00  | 2.07 (0.32)  | 1.88 (0.36)       |
|            | \( \rho \)           | 5.00  | 5.56 (1.92)  | 7.02 (3.40)       |
| s4         | \( \lambda \times 10^2 \) | 1.25  | 1.26 (0.52)  | 0.86 (0.40)       |
|            | \( \rho \)           | 2.00  | 2.07 (0.24)  | 1.90 (0.31)       |
| s5         | \( \lambda \times 10^1 \) | 2.00  | 2.16 (0.71)  | 3.31 (1.53)       |
|            | \( \rho \)           | 5.00  | 4.92 (10.31) | -                 |
| s6         | \( \lambda \times 10^2 \) | 1.25  | 1.52 (1.21)  | -                 |
| s7         | \( \lambda \times 10^3 \) | 5.00  | 5.03 (2.52)  | 7.28 (8.13)       |
|            | \( \rho \)           | 4.00  | 4.21 (0.81)  | 3.62 (1.31)       |
| s8         | \( \lambda \times 10^1 \) | 5.00  | 5.94 (2.78)  | 6.67 (6.16)       |
| s9         | \( \lambda \times 10^4 \) | 1.25  | 1.20 (0.49)  | 1.17 (0.76)       |
|            | \( \rho \)           | 8.00  | 8.83 (3.84)  | -                 |
| s10        | \( \lambda \times 10^2 \) | 2.00  | 1.40 (2.11)  | -                 |
|            | \( \rho \)           | 8.00  | 8.83 (3.84)  | -                 |
|            | \( \mu \times 10^{-1} \) | 1.25  | 3.72 (1.78)  | -                 |
| s11        | \( \lambda \times 10^1 \) | 5.00  | 6.64 (8.57)  | -                 |
|            | \( \rho \)           | 8.00  | 8.87 (3.70)  | -                 |
| s12        | \( \lambda \times 10^3 \) | 5.00  | 7.87 (6.28)  | -                 |
|            | \( \rho \)           | 8.00  | 8.56 (1.80)  | 5.51 (2.22)       |
| s13        | \( \lambda \times 10^0 \) | 1.25  | 1.45 (0.97)  | 3.92 (6.13)       |
|            | \( \rho \)           | 8.00  | 8.56 (1.80)  | 5.51 (2.22)       |

Mean and standard deviation (in parentheses) of GP parameter estimates from 100 simulated runs.
doi:10.1371/journal.pone.0111522.t002

\[
\int \int q^2 \left( \Phi - \frac{1}{W^2} \right) \text{d}x \text{d}y \int \text{d}p_x \text{d}p_y
\]  \( (22) \)

In this implementation Equation 22 was evaluated by solving the inner integrals \( \text{d}x \) and \( \text{d}y \) to give a closed form solution in terms of the error function, and then the outer integrals \( \text{d}p_x \) and \( \text{d}p_y \) were evaluated numerically.

**Results and Discussion**

Estimation of simulated process parameters

Events were simulated from nine clustered populations using a two-dimensional Neyman-Scott model within a 100 \( \times \) 100 square survey area. In each case the product \( \mu \lambda \) was equal to 0.25, hence the expected number of events is the same for each population, however the extent of clustering varies. One realisation of each population is shown in Figure 2 and the simulation parameters are given in Table 1. Note that s3 exhibits the most clustering, and s7 the least. Edge effects were minimised by simulating the process for a 200 \( \times \) 200 region, and then only considering events lying within a central 100 \( \times \) 100 square.

The point process was then converted into an intensity map by dividing the survey area into unit quadrats and counting the number of events lying within each quadrant. One realisation of the intensity maps for each population is shown in Figure 3. These maps served as a ‘ground truth’, and represent the function that the GP attempted to model using sparse data.

In order to be consistent with the experiment carried out by Cowling, nine vertical, equally spaced transects of length 100 were used as the training data for the GP. All quadrats along each transect were used as training data, the effective area surveyed was thus 9% of the total study area.

The GP models were generated using a squared exponential covariance function with additive Gaussian noise. The hyperparameters for the covariance function were selected by maximising the LML.

Figures 4 and 5 show the resulting GP mean and variance for the realisation of the populations depicted in Figures 2 and 3. The variance plots show how the uncertainty increases with distance from the vertical transects. For populations s1, s2 and s3 where \( \rho = 2 \) the variance increases relatively rapidly from the sample...
locations, as the offspring are clustered closely to the parents and inference is weak beyond the extents of a cluster. For populations \(s_7, s_8\) and \(s_9\) where \(\rho = 8\), the pattern is much smoother. The variance is almost constant across the survey field, only rising at the edges where extrapolation is occurring rather than interpolation.

For each population, Figure 6 shows the average covariance between samples plotted against their separation. The GP squared
The parameters of the Neyman-Scott process were then estimated by fitting the theoretical covariance function given in Equation 22 to the optimised GP covariance function. In practice this was performed by minimising

\[ \sum_{h=1}^{H} [k(h,\sigma_f,\lambda) - \text{cov}(h,\lambda,\rho,\mu)]^2 \]

subject to \( \lambda H = \hat{\lambda} H \) (23)

where \( k(h,\sigma_f,\lambda) \) is the squared exponential covariance function given in Equation 14 evaluated for two points distance \( h \) apart with hyperparameters \( \sigma_f \) and \( \lambda \); \( \text{cov}(h,\lambda,\rho,\mu) \) is the theoretical covariance function; \( \hat{\lambda} H \) is an estimate of the mean intensity obtained by taking the mean of the training points; \( H \) is the maximum range for the minimisation such that \( k(h \geq H,\sigma_f,\lambda) \approx 0 \).

The simulation was repeated 100 times and the means and standard deviations of the parameter estimates are given in Table 3.

| Sampling Strategy | Parameter Value | GP Estimator |
|-------------------|-----------------|--------------|
| Transects         | \( \lambda \times 10^3 \) 1.25 | 1.26 (0.52) |
|                   | \( \rho \) 2.00 | 2.07 (0.24) |
|                   | \( \mu \times 10^{-1} \) 2.00 | 2.16 (0.71) |
| Random Points     | \( \lambda \times 10^3 \) 1.25 | 1.61 (1.03) |
|                   | \( \rho \) 2.00 | 1.87 (0.39) |
|                   | \( \mu \times 10^{-1} \) 2.00 | 1.97 (0.82) |
| Blocks            | \( \lambda \times 10^3 \) 1.25 | 1.35 (1.51) |
|                   | \( \rho \) 2.00 | 2.08 (0.34) |
|                   | \( \mu \times 10^{-1} \) 2.00 | 2.86 (1.98) |
| Random Walk       | \( \lambda \times 10^3 \) 1.25 | 1.29 (0.88) |
|                   | \( \rho \) 2.00 | 2.05 (0.29) |
|                   | \( \mu \times 10^{-1} \) 2.00 | 2.46 (1.55) |

Mean and standard deviation (in parentheses) of GP parameter estimates from 100 simulated runs, for population s3 with different sampling strategies. doi:10.1371/journal.pone.0111522.t003

Figure 8. Sensitivity of method to sample size. Summarised results of 100 runs of GP modelling of population s3, for 6 transect sample designs with varying numbers of evenly spaced transects. The error bars show one standard deviation from the mean estimate. doi:10.1371/journal.pone.0111522.g008
Table 2. Corresponding results from Cowling are also reproduced as a benchmark, however a direct comparison is not entirely appropriate for two reasons. Firstly, because to our knowledge these results were not recomputed and republished once an error in Cowling’s method was corrected in [15]. Secondly, Cowling’s method assumes that an observer is travelling along the one-dimensional transect line, and events become exponentially more difficult to detect as perpendicular distance from the transect increases. Hence although the effective area surveyed was 9% of the total study area (as in this experiment) the detection function ensures that only some fraction of events are detected. This contrasts with the GP estimator, where it is assumed that all events within a quadrat are detected.

Table 2 shows that in the strongly clustered populations s2, s3 and s6 the clustering parameters can be estimated much more reliably than in the weakly clustered populations s4, s7 and s8. In the weakly clustered populations Cowling’s method fails to detect any noticeable clustering, and so parameter estimation was not attempted. The GP estimator typically detects the weak clustering and tends to fit a smooth function with a very large length scale (as shown in Figures 4(d),(g) and (h)). However, occasionally (i.e. 9 runs out of 100 for population s7) the GP fails to detect the clustering, and instead overfits a spiky function with a very short length scale. Despite these outliers, on average the GP gives reasonable, if highly variable estimates for the weakly clustered populations.

In all cases the GP estimator tends to overestimate $\rho$ and $\mu$. All the parameter estimates compare favourably with the K-function method of Cowling, however it must be remembered that the two experiments involve a different form of information loss: Cowling’s method assumes that not all events are detected, and the GP method aggregates all event locations into a single quadrat count and so some high resolution position data is disregarded.

Alternative Sampling Designs

The previous section demonstrated that a GP can be used to estimate clustering parameters with an accuracy that compares favourably with existing methods. However the real advantage of this method is its resilience to arbitrary sampling designs. The experiment was repeated for the most clustered population s3 using three different sampling strategies: samples at uniformly distributed random points; ‘block’ sampling (nine 10x10 grids of samples with equal spacing between each grid, see figure 7); and a random walk starting at a random location. Figure 7 shows the GP mean and variance, along with the sample locations for the first run of this experiment. The transect and block patterns were identical for each run, but the random point and walk patterns were different for each run.

All of the new surveys were designed to give the same 9% coverage of the survey field. The results are shown in Table 3, with the corresponding results from Table 2 reproduced for comparison.

For this population, on balance the transect sample seems to be the most consistent estimator of the parameters. We postulate that this is because transect sampling gives reasonably even coverage of the survey field, plus contiguous samples. The former is required for the model to detect the intensity of the clusters ($\lambda$), while the latter helps the model fit to the distribution of the clusters themselves ($\mu$ and $\rho$). Random point sampling provides the best coverage of the survey field and so the GP mean is likely to be the most representative of the true pattern. This is evident from Figure 7(d) where almost all of the clusters are detected and modelled. However, the random point sample provides the worst estimate of $\rho$. The other sampling designs all contain contiguous samples, which would be preferred in this case because the covariance is only detectable with separation between quadrats of up to approximately 5 units. Figure 7(f) provides an indication of why the block sample estimates have a very high variance; the GP mean will be unrepresentative of the underlying pattern on many of the runs if little or no clusters fall within the blocks. However when a cluster does fall within one of the blocks it will be modelled precisely, hence the good estimate of $\rho$. These results show that while some survey designs are better than others, the GP estimator is applicable to any survey design and provides a similar estimate of the parameters in all cases.

To investigate how sensitive the method is to sample coverage, the transect sample experiment was repeated with a varying number of transect lines. As before, the experiment was run with 100 realisations of population s3 with equally spaced transect lines covering 5% to 30% of the survey area in steps of 5%. The resulting parameter estimates are shown in Figure 8. From this figure, the diminishing returns on increasing sample size are apparent. The variability in the estimate decreases significantly from 5% to 20% coverage, but increasing coverage further has little noticeable effect.

In these experiments unit quadrat size is arbitrarily used, but in practice choice of quadrat size relative to the survey area would be another variable to consider when designing the sample. Too small a quadrat size will result in a very sparse intensity map, which may result in a GP model that is ‘overfitted’ to a small number of detected events. If the quadrat size is too large, multiple clusters of events will be aggregated within a single quadrat, and this loss of resolution may result in a GP model that is ‘underfitted’, with individual clusters smoothed out. In both these extremes one would expect the estimator to perform poorly. In these experiments, the unit quadrat size is less than the standard deviation of the clusters ($\rho = 2, 4, 8$) but of a similar order of magnitude; in practice we would recommend a similar approach if some prior knowledge of the underlying process is available.

Conclusions

We have shown that a GP can provide a good model for a stationary, isotropic cluster process such as the Neyman-Scott model. As a GP is completely defined by its mean and covariance function, these provide a good proxy for the parameters of a cluster process. The GP model can be used to estimate these parameters, with a precision and accuracy which compares well with other methods within the literature. The main advantage to the GP method is its resilience to arbitrary sampling design.

Author Contributions

Conceived and designed the experiments: PR OP SBW. Performed the experiments: PR. Analyzed the data: PR OP SBW. Contributed reagents/materials/analysis tools: PR. Wrote the paper: PR OP SBW.

References

1. Dunning J, Danielson B, Pulliam H (1992) Ecological processes that affect populations in complex landscapes. Oikos 65: 169–175.

2. Hanski IA, Simberloff D (1997) Metapopulation biology. Ecology, genetics and evolution, Academic Press, San Diego, chapter The metapopulation approach, its history, conceptual domain, and application to conservation. pp. 5–26.
3. Holmes E, Lewis M, Banks J, Veit R (1994) Partial differential equations in ecology: spatial interactions and population dynamics. Ecology 75: 17–29.
4. Cowling A (1998) Spatial methods for line transect surveys. Biometrics 54: 828–839.
5. Rasmussen CE, Williams CKI (2006) Gaussian Processes for Machine Learning. MIT Press.
6. Matheron G (1963) Principles of Geostatistics. Economic Geology 58: 1246–1266.
7. Holmes KW, Nielsen KV, Kendrick G, Baxter K (2006) Designs for remote sampling: review, discussion, examples of sampling methods and layout of scaling issues. Technical report, Cooperative Research Centre for Coastal Zone, Estuary and Waterway Management, Australia.
8. Okabe A, Levin SA (1980) Diffusion and Ecological Problems. Springer.
9. Dunning J, Stewart D, Danielson B, Noon B, Root T, et al. (1995) Spatially explicit population models: current forms and future uses. Ecol Appl 5: 3–11.
10. Matern B (1971) Statistical Ecology Volume 1, Penn State University Press, University Park and London, chapter Doubly stochastic Poisson process in the plane. pp. 195–213.
11. Hagen G, Schweder T (1995) Whales, Seals, Fish and Man, Elsevier Science B. V., chapter Point clustering of minke whales in the north-eastern Atlantic. pp. 27–33.
12. Aldrin M, Holden M, Schweder T (2002) Spatial distribution of northeastern Atlantic minke whales 1996–2001. Technical report, Paper SC/54/RMP3 presented to the Scientific Committe of the International Whaling Commission 2002.
13. Wu BM, Subbarao KV, Ferrandino FJ, Hao JJ (2006) Spatial analysis based on variance of moving window averages. Phytopathology 154 (6): 349–360.
14. Ripley B (1977) Modelling spatial patterns (with discussion). Journal of the Royal Statistical Society B39: 172–212.
15. Aldrin M, Holden M, Schweder T, Cowling A (2003) Comment on Cowling’s ‘Spatial methods for line transect surveys’. Biometrics 59: 186–188.
16. Cressie NAC (1993) Statistics for Spatial Data. Wiley.
17. von Mises R (1964) Mathematical Theory of Probability and Statistics. Academic Press, 200 pp.