Likelihood-based inference for clustered line transect data

Waagepetersen, Rasmus Plenge; Schweder, Tore

Publication date:
2004

Document Version
Publisher's PDF, also known as Version of record

AALBORG UNIVERSITY

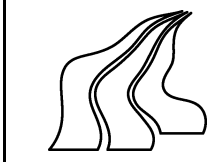
Likelihood-based inference for clustered line transect data

by

Rasmus Waagepetersen and Tore Schweder

December 2004 R-2004-29

Department of Mathematical Sciences
Aalborg University
Fredrik Bajers Vej 7 G ▪ DK - 9220 Aalborg Øst ▪ Denmark
Phone: +45 96 35 80 80 ▪ Telefax: +45 98 15 81 29
URL: www.math.aau.dk/research/reports/reports.htm



ISSN 1399–2503 ▪ On-line version ISSN 1601–7811

Likelihood-based inference for clustered line transect data

Rasmus Waagepetersen
Institute of Mathematical Sciences, Aalborg University
Fredrik Bajersvej 7G, DK-9220 Aalborg
rw@math.aau.dk

Tore Schweder
Department of Economics, University of Oslo
P.O. Box 1095 Blindern, N-0317 Oslo

Abstract

The uncertainty in estimation of spatial animal density from line transect surveys depends on the degree of spatial clustering in the animal population. To quantify the clustering we model line transect data as independent thinnings of spatial shot-noise Cox processes. Likelihood-based inference is implemented using Markov Chain Monte Carlo methods to obtain efficient estimates of spatial clustering parameters. Uncertainty is addressed using parametric bootstrap or by consideration of posterior distributions in a Bayesian setting. Maximum likelihood estimation and Bayesian inference is compared in an example concerning minke whales in the North Atlantic. Our modelling and computational approach is flexible but demanding in terms of computing time.

Keywords: shot-noise Cox process, simulation-based inference, spatial point process, thinning.

1 Introduction

Line transecting is a survey technique aimed at estimating the abundance of animals in an area (Buckland *et al.*, 2004). An observer traverses at fixed speed through the area along a predetermined transect line, and records data on sighting conditions and on the sighted animals. The transect is often a zigzag and may be broken due to changes in sighting condition. For simplicity, consider one straight leg of the line, and term this the transect. Let $p(x, y)$ denote the probability of observing an animal at a location (x, y) assuming that

the transect of length L runs along the y-axis with start at the origin. If m animals are observed, the standard moment estimate of animal density λ is

$$\tilde{\lambda} = m/\int_0^L \int_{-\infty}^{\infty} p(x,y)\mathrm{d}x\mathrm{d}y. \tag{1}$$

This is also the maximum likelihood estimate when the animal population is distributed spatially as a homogeneous Poisson process with intensity λ and the animals are detected or not detected independently of each other with detection probability $p(x,y)$. In spatial point process terminology, the detected animals are given by an independent thinning of the animal population with an inhomogeneous thinning probability $1-p(x,y)$. Line transect data were first regarded as thinnings of point processes in Schweder (1974) and Schweder (1977). Many animal populations are more spatially variable and clustered than a homogeneous or inhomogeneous Poisson process. Line transect data for such populations may naturally be modelled as independent thinnings of spatial Cox point processes.

A computationally cheap way to obtain parameter estimates for a Cox process is to match a non-parametric estimate of a second order summary statistic with its theoretical expression depending on the unknown parameters. This approach was first taken for line transect data by Hagen and Schweder (1995) who used the so-called K-function. A non-parametric estimate of the K-function may be obtained from line transect data as discussed in Baddeley *et al.* (2000). Due to strong edge effects for line transect data, it is, however, hard to obtain an efficient estimate. Cowling (1998) (see also Aldrin *et al.*, 2003) considers the K-function for the one-dimensional point process obtained from projecting a thinned Neyman-Scott process onto the transect line. The thinning probability is here assumed centred Gaussian with constant scale parameter. In practice, however, the thinning probability is usually varying along the transect line according to sighting conditions, see e.g. Skaug *et al.* (2004).

In this paper we develop efficient likelihood-based inference for a thinned Cox process both in a frequentist and a Bayesian setting. The inference is implemented using simulation methodology like parametric bootstrap and Markov chain Monte Carlo. A distinct advantage of our approach is that we do not need any simplifying assumptions regarding the nature of the detection probability. Our modelling and computational approach can moreover easily be adapted to other sampling designs of the distance type (Buckland *et al.*, 2004) and to nonstationary clustered animal populations.

Our discussion will be focused on a particular line transect study of minke whale abundance in the North Atlantic, see Section 2. Section 3 and Section 4 describes our model and the computational approach while Section 5 contains an application to the minke whale data. Section 6 contains some final remarks.

2 Minke whales data example

Skaug *et al.* (2004) is concerned with statistical analysis of minke whales line transect

data collected from 1996 to 2001 in various survey blocks covering the northeast Atlantic. Here, we focus on the survey block named VSS located west of Spitzbergen. This block was visually surveyed in 1999 with 50 whales observed (see Figure 1). A few comments regarding a neighbouring survey block VSN (not shown) are given in Section 5 and 6.

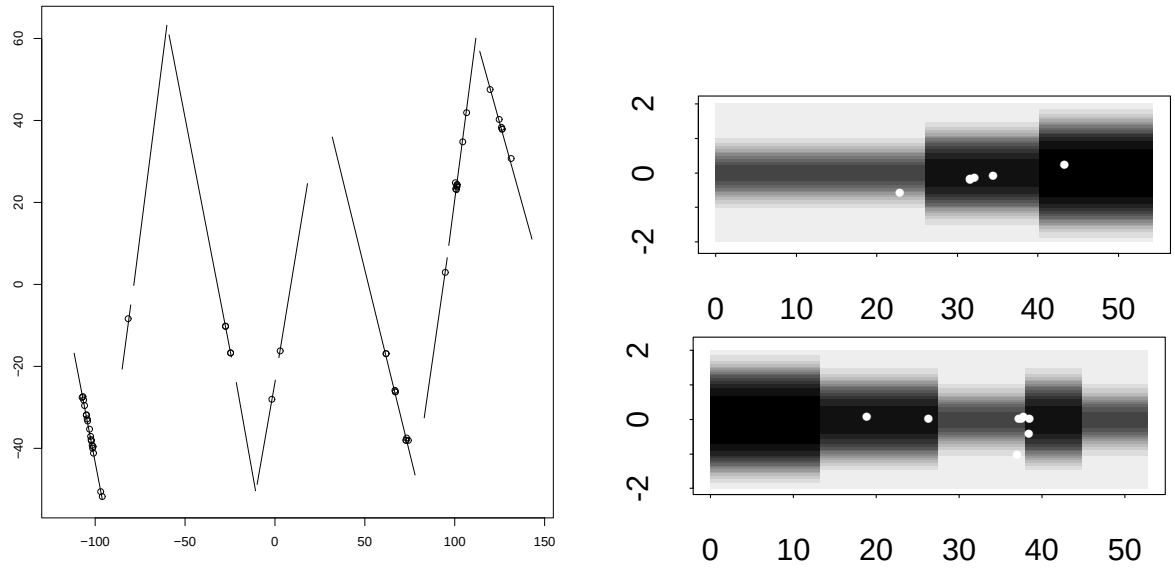


Figure 1: Left: transects and observed whales for the VSS block with distances in km. The transecting was broken when sighting conditions became unsuitable, and restarted according to a protocol. Right: gray scale plots of the detection probability for the two leftmost transects in the VSS block. Dark means high detection probability and white dots show positions of observed whales. Distances are in km both on the x and y axes.

The probability of detecting a whale is considerably less than one even when located right on the transect. Let $Q(x, y, \tilde{y})$ denote the hazard probability of initially detecting a whale surfacing at position (x, y) when the ship is at position $(0, \tilde{y})$, $\tilde{y} < y$, along the transect. Assuming that the whales surface according to a Poisson process in time with intensity $\phi > 0$ and that the ship moves at unit speed, the detection probability is $p(x, y) = 1 - \exp\big(-\phi \int_{-\infty}^{y} Q(x, y, \tilde{y}) \mathrm{d}\tilde{y}\big)$. To estimate $Q(x, y, \tilde{y})$, a double platform design is used in the minke whale surveys. There is no communication between the platforms and from each platform, tracks of successive surfacings of detected whales are recorded. An estimate of $Q(x, y, \tilde{y})$ is then obtained from matched pairs of tracks observed from both platforms and tracks observed from just one platform.

In the Norwegian minke whale surveys, radial distance from the ship to the surfacing whale is estimated by eye, and the angle between the transect and the sighting line is estimated by way of an angle board fixed to the rim of the barrel or platform fence. Time and ship position are accurately measured, but the angle and particularly radial distance measurements are rather imprecise. Due to measurement error, tracks and surfacings from the two platforms might be wrongly matched. This induces bias in the estimation of $p(x, y)$.

The bias is estimated by regression analysis on simulated estimates, and a bias-corrected estimate is obtained.

For our present purpose of fitting a spatial cluster model, the detection probability is regarded as known and given by the estimate obtained in Skaug *et al.* (2004). The right plots in Figure 1 show the estimated detection probabilities for two transects in the VSS block. Note that the detection probabilities vary due to dependence on sea state, visibility, observation team, and other covariates. An important feature of our likelihood-based approach is that it easily accomodates the spatially varying detection probability.

3 Spatial point process modelling of whale positions

In this section we discuss the modelling of whale positions observed along one transect. Within the time-span of traversing a transect the whales are regarded as immobile and occur at spatial locations $\xi = (x, y)$ where these locations are relative to a coordinate system with the transect along the y-axis and origin at the start of the transect. The whales in the vicinity of the transect are regarded as a subset of a planar stationary point process X whose intensity λ is the parameter of main interest. The process Y of positions of observed whales is regarded as an independent thinning of X with thinning probabilities $1 - p(\cdot)$ where $p(\xi)$ is the probability of detecting a whale positioned at ξ. In practice $p(\cdot)$ has bounded support so that Y is a finite point process.

3.1 Shot-noise Cox processes

Minke whales in the north Atlantic tend to form loose and variable clusters, partially due to stochastic clustering in the prey distribution (Skaug *et al.*, 2004). Therefore a Cox process seems an appropriate model for the whale positions. In this paper we consider so-called shot-noise Cox processes (Brix, 1999; Møller, 2003; Møller and Waagepetersen, 2003). A shot-noise Cox process X in the plane is a Cox process with random intensity function

$$Z(\xi|\Phi) = \sum_{(c,\gamma)\in\Phi} \gamma k(\xi - c) \tag{2}$$

where Φ is a Poisson process on $\mathbb{R}^2 \times]0, \infty[$ and k is a probability density. A shot-noise Cox process can also be viewed as a cluster process, i.e. given Φ, X is distributed as a superposition of Poisson processes $X_{(c,\gamma)}$, $(c, \gamma) \in \Phi$, each with intensity function $\gamma k(\cdot - c)$ where c is the cluster centre and γ is the expected number of points in the cluster. A parametric model for Φ is considered in Section 3.2 below. The process Y of observed whales is a shot-noise Cox process with random intensity function

$$Z_Y(\cdot|\Phi) = p(\cdot)Z(\cdot|\Phi).$$

Our kernel k is a truncated bivariate Gaussian density with scale parameter $\omega > 0$, i.e.

$$k((x,y);\omega) = 1[\max(|x|,|y|) < T\omega] \exp\big(-(x^2+y^2)/(2\omega^2)\big)/(2\pi\omega^2 c(T)) \tag{3}$$

where $T > 0$ and $c(T)$ is a normalizing constant ensuring that k integrates to 1. In our application, $T = 3$ so that $c(3) = 0.9973$. Working with a k of bounded support is very advantageous for computational reasons, see Appendix A.

3.2 Shot-noise G Cox processes

A shot noise G Cox process (Brix, 1999) is obtained when Φ in (2) has intensity function of the form

$$\zeta(c,\gamma) = \chi(\gamma) = \exp(\kappa)\beta^\alpha \gamma^{\alpha-1} \exp(-\beta\gamma)/\Gamma(\alpha) \tag{4}$$

where $\kappa \in \mathbb{R}$, $\alpha > -1$, and $\beta > 0$. With $\alpha = 0$ a so-called Poisson-gamma process (Wolpert and Ickstadt, 1998) is obtained.

The process of cluster centres $C = \{c \in \mathbb{R}^2 : (c,\gamma) \in \Phi \text{ for some } \gamma\}$ is not locally finite when $\alpha \leq 0$. We henceforth assume $\alpha > 0$ in which case Φ can be viewed as a locally finite marked point process where C is a homogeneous Poisson process of intensity $\exp(\kappa)$ and the mark γ associated with a cluster centre c is gamma distributed with mean α/β and variance α/β^2. Then $\lambda = \exp(\kappa)\alpha/\beta$ and for a region A the overdispersion index (i.e. the ratio between the variance and mean of the number of points in $X \cap A$) is approximately $1 + (1+\alpha)/\beta$.

A shot-noise G Cox process with $\alpha > 0$ is an example of a Neyman-Scott process with negative binomial numbers of points in each cluster. The use of gamma distributed marks γ in (2) adds additional flexibility compared to more common examples of Neyman-Scott processes like the Thomas process (first used in the line transect context by Hagen and Schweder, 1995) for which the random intensity function is obtained by a superposition of Gaussian kernels all multiplied with the *same* positive parameter.

Our rather small data set provides only indirect observation of the γ's, so only the ratio α/β is well determined by the data. For numerical reasons (see Section 5) one may therefore either fix $\alpha = 1$ or $\beta = 1$ whereby respectively exponential or standard gamma distributed marks are obtained. In the sequel we restrict attention to the standard gamma case $\beta = 1$.

4 Likelihood-based inference

The set of spatial locations with positive probability of detecting a whale is essentially a union of narrow bands around the line transects. The geometry of this set is rather complicated from a computational point of view. We therefore use a composite likelihood approach: log likelihood functions are computed for each transect separately and then added to obtain a composite log likelihood function based on all of the transects. Since

there is very little overlap between data from the narrow strips around different transects, we believe that the loss of information is small compared to using the likelihood function.

For a data set with n transects, and letting $\theta = (\kappa, \alpha, \omega)$, the composite log likelihood is $l(\theta) = \sum_i \log L_i(\theta)$ where $L_i(\theta)$ is the likelihood function based on the ith transect. In the following, we refer for sake of brevity to the composite likelihood function and maximum composite likelihood estimates as the likelihood function and maximum likelihod estimates, respectively.

In Section 5.1 we use a profile likelihood approach where $l(\theta)$ is maximized with respect to (κ, α) for a finite set of ω values using Newton-Raphson. The reason for using the profile likelihood approach is that it is computationally very involved to compute the first and second derivatives with respect to the kernel scale parameter ω. The likelihood may further be highly multimodal as a function of ω, see Figure 2, in which case gradient based maximization is not reliable. We are not aware of theoretical results concerning the properties of maximum likelihood estimates for spatial Cox processes so we use a parametric bootstrap to investigate the repeated sampling properties of our estimates.

The score function, information matrix, and log likelihood ratios $l(\theta_2) - l(\theta_1)$ are obtained by summing the corresponding quantities obtained from the log likelihood functions $\log L_i(\theta)$ for each transect. In the following we therefore restrict attention to the likelihood function $L_i(\theta)$ based on data for just one transect.

4.1 Likelihood function for one transect

Let S_i denote the bounded support of the detection probability for the ith transect, see the right plots in Figure 1. The conditional density of $Y_i = Y \cap S_i$ given Φ is given by

$$f_i\big(y_i | Z_Y(\cdot|\Phi;\omega)\big) = \exp\big(-\int_{S_i} Z_Y(\xi|\Phi;\omega))\mathrm{d}\xi\big) \prod_{\eta \in y_i} Z_Y(\eta|\Phi;\omega) \tag{5}$$

and the likelihood function for the ith transect is

$$L_i(\theta) = \mathbb{E}_{(\kappa,\alpha)} f\big(y_i | Z_Y(\cdot|\Phi;\omega)\big) = \mathbb{E}_{(\kappa,\alpha)} f\big(y_i | Z_Y(\cdot|\Phi_i;\omega)\big)$$

(see Chapter 10 in Møller and Waagepetersen, 2003, for details) where Φ_i is the finite point process of marked cluster centres which may contribute with offspring in S_i. That is, for a given ω, $\Phi_i = \{(c,\gamma) \in \Phi | c \in E_i\}$ where E_i is the smallest rectangle containing S_i with the property that a cluster centre c outside E_i has zero probability of contributing with an offspring inside S_i.

Approximations of likelihood ratios $L_i(\theta_2)/L_i(\theta_1)$ are obtained using bridge sampling (Gelman and Meng, 1998; Møller and Waagepetersen, 2003). To compute approximate derivatives of $\log L_i(\theta)$ (Section 4.2) or bridge sampling likelihood ratios $L_i(\theta_2)/L_i(\theta_1)$ we need conditional simulations of the "missing data" Φ_i given Y_i. An algorithm for this is discussed in Appendix A. This algorithm also forms the backbone in an algorithm for posterior simulation in a Bayesian setting, see Section 4.3.

4.2 Computation of log likelihood derivatives

Consider a fixed ω and let $\psi = (\kappa, \alpha)$. Denote by n_i the number of points in Φ_i, let $l_i = \sum_{(c,\gamma)\in\Phi_i} \log\gamma$ and $V_\theta(Y_i, \Phi_i) = \mathrm{d}\log f_i(y_i, \Phi_i; \theta)/\mathrm{d}\psi$ where $f_i(y_i, \phi_i; \theta)$ is the joint density of (Y_i, Φ_i), see (7) in Appendix A. Following Section 8.6.2 in Møller and Waagepetersen (2003), the score function for the ith transect is given by

$$u_i(\psi) = \mathbb{E}_{\theta,y_i} V_\theta(X_i, \Phi_i) = (\mathbb{E}_{\theta,y_i} n_i - \exp(\kappa)|E_i|, \mathbb{E}_{\theta,y_i} l_i - \Psi(\alpha)\mathbb{E}_{\theta,y_i} n_i)$$

where $\mathbb{E}_{\theta,y_i}$ denotes conditional expectation with respect to Φ_i given $Y_i = y_i$ and Ψ is the digamma function. Similarly, the observed information matrix is

$$\begin{aligned} j_i(\psi) &= -\mathbb{E}_{\theta,y_i}\mathrm{d}V_\theta(X_i, \Phi_i)/\mathrm{d}\psi^\mathsf{T} - \mathbb{V}\mathrm{ar}_{\theta,y_i} V_\theta(X_i, \Phi_i) = \\ &\begin{bmatrix} \exp(\kappa)|E_i| & 0 \\ & \Psi'(\alpha)\mathbb{E}_{\theta,y_i} n_i \end{bmatrix} - \begin{bmatrix} \mathbb{V}\mathrm{ar}_{\theta,y_i} n_i & \mathbb{C}\mathrm{ov}_{\theta,y_i}[n_i, l_i] - \Psi(\alpha)\mathbb{V}\mathrm{ar}_{\theta,y_i} n_i \\ & \mathbb{V}\mathrm{ar}_{\theta,y_i}[l_i - \Psi(\alpha) n_i] \end{bmatrix} \end{aligned} \tag{6}$$

where $\mathbb{V}\mathrm{ar}_{\theta,y_i}$ and $\mathbb{C}\mathrm{ov}_{\theta,y_i}$ denotes conditional variance and covariance, and where the matrices are symmetric with only the upper triangle shown. The first term in (6) is the conditional expectation of the observed information in the case where n_i and l_i is observed.

We could reparametrize letting $\kappa := \log(\kappa - \Gamma(\alpha))$ in which case an exponential family density with sufficient statistic $t_i = (n_i, l_i)$ would be obtained for Φ_i. Then we obtain particularly neat expressions for the score function and observed information: $u_i(\psi) = \mathbb{E}_{\theta,y_i} t_i - \mathbb{E}_\theta t_i$ and $j_i(\psi) = \mathbb{V}\mathrm{ar}_\theta t_i - \mathbb{V}\mathrm{ar}_{\theta,y_i} t_i$. However, with the original parametrization a more well-conditioned observed information matrix is obtained.

The expectations appearing in the score function and the information matrix can not be evaluated analytically. In order to estimate the expectations using importance sampling methods (see Section 8.6.2 in Møller and Waagepetersen, 2003) we use conditional simulations of Φ_i given $Y_i = y_i$, see Appendix A.

4.3 Bayesian approach

A Bayesian approach is possible if prior information is available concerning θ. Assuming independence between transects, this leads to consideration of the posterior distribution of $(\theta, (\Phi_i)_{i=1}^n)$ with density

$$p(\theta, (\phi_i)_{i=1}^n | (y_i)_{i=1}^n) \propto \pi(\theta) \prod_{i=1}^n f_i(y_i, \phi_i; \theta)$$

where $\pi(\theta)$ is the prior density. An MCMC algorithm for posterior simulation can be obtained by combining the MCMC algorithm from Appendix A with Metropolis-Hastings updates for θ (see e.g. Robert and Casella, 2004, for background on MCMC).

5 Application to whale data

Precise Monte Carlo estimation of the score function and in particular the observed information and log likelihood ratios requires large MCMC samples. Hence our approach to maximum likelihood estimation is demanding in terms of computing time. The Bayesian approach on the other hand is computationally much less demanding, see Section 5.3. To give an idea of the computational complexity we report below computing times on a 2.4 GHz/256 MB Intel 4 processor.

5.1 Maximum likelihood estimation

Estimates $(\kappa_l, \alpha_l) = \arg\max_{(\kappa,\alpha)} l(\kappa, \alpha, \omega_l)$ and $\lambda_l = \exp(\kappa_l)\alpha_l$ are obtained for different values $\omega_l = l/10\text{km}$, $l = 2, \ldots, 30$, using Newton-Raphson. Occasionally, Monte Carlo error results in negative definite Monte Carlo estimates of the observed information so we use a Marquardt-Levenberg variant of the Newton-Raphson algorithm where positive terms are added to the diagonal of the estimated observed information when it is negative definite.

The left plot in Figure 2 shows the profile log likelihood function for ω obtained by cumulating log likelihood ratios $l(\theta_{l+1}) - l(\theta_l)$ (with $\theta_l = (\kappa_l, \alpha_l, \omega_l)$) obtained using bridge sampling. The profile likelihood function for VSS has a well-defined maximum for $\omega = \omega_6 = 0.6$ with

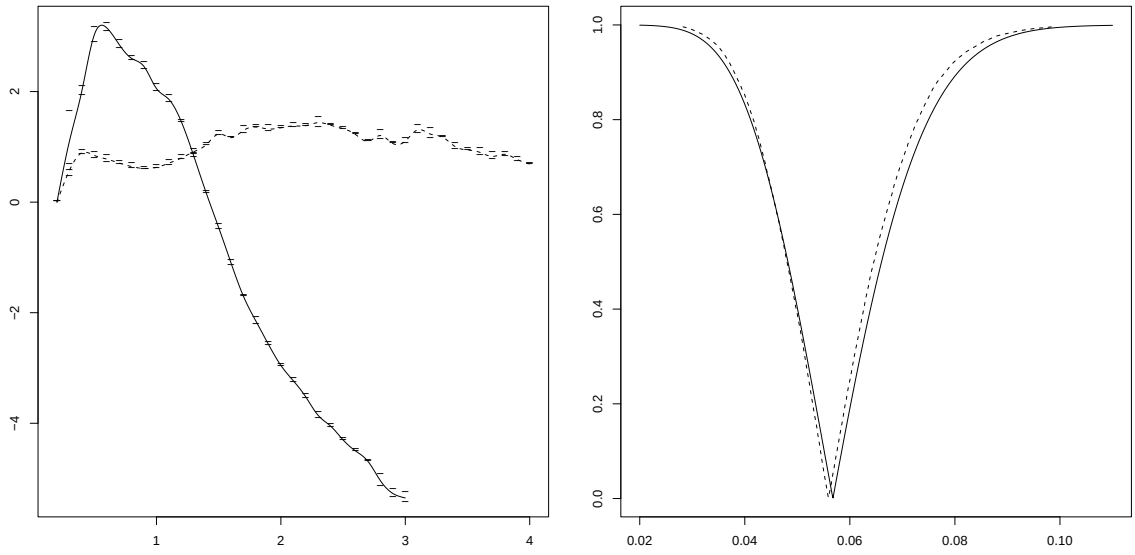


Figure 2: Left: Profile log likelihood functions $l_p(\omega) = \max_{(\kappa,\alpha)} l(\theta)$ obtained for the VSS (solid line) and VSN (dotted line) blocks by cumulating log likelihood ratios $l(\theta_{l+1}) - l(\theta_l)$. The small vertical bars indicate Monte Carlo confidence intervals for the differences $l(\theta_{l+1}) - l(\theta_l)$. Right: confidence (solid line) and posterior credibility nets (dotted line) for λ, see Section 5.2 and Section 5.3.

corresponding estimates $\kappa_6 = 13.8$, $\alpha_6 = 2.4$ and $\lambda_6 = 0.06$. In Section 6 we comment on

the second more flat and multimodal profile likelihood function for the VSN block. The small vertical bars in Figure 2 indicate Monte Carlo confidence intervals for the log likelihood ratios $l(\theta_{l+1}) - l(\theta_l)$. We consider the Monte Carlo error for the estimates $(\kappa_l, \alpha_l, \lambda_l)$ in the simulation study in Section 5.2. The computation of an estimate (κ_l, α_l) and a log likelihood ratio $l(\theta_{l+1}) - l(\theta_l)$ took around 70 minutes.

To illustrate how $(\kappa_l, \alpha_l, \lambda_l)$ depends on ω_l, a collection of estimates are given in Table 1. The estimates κ_l and α_l vary considerably whereas λ_l is rather insensitive to the value of ω.

ω_l	0.2	0.4	0.6	0.8	1
κ_l	15.3	14.5	13.8	13.4	13.1
α_l	0.5	1.2	2.4	3.7	4.9
λ_l	0.06	0.06	0.06	0.06	0.06

Table 1: A collection of estimates κ_l, α_l and λ_l obtained for different values of ω.

5.2 Parametric bootstrap

The repeated sampling properties of the parameter estimates are studied using a parametric bootstrap based on 100 independent synthetic data sets obtained by simulation under the fitted model with parameters equal to the maximum likelihood estimates obtained in Section 5.1. It is very time consuming to repeat the whole profile likelihood procedure for each simulated data set. We therefore use an adaption of the parametric bootstrap where ω is assumed known and equal to the maximum likelihood estimate. However, Table 1 suggests that this only implies a slight underestimation of the sampling variability for the maximum likelihood estimate $\hat{\lambda}$ of main interest.

Bootstrap estimates of the means for the sampling distributions of the estimates κ_6, α_6 and λ_6 for fixed $\omega = \omega_6 = 0.6$ and the standard moment estimate (1) are 13.9 (13.8), 2.3 (2.4), 0.06 (0.06), and 0.06 (0.06) with the parameter values used for the bootstrap simulation given in parantheses. The 2.5% and 97.5% quantiles are (13.3;14.6), (0.7;4.5), (0.03;0.08), and (0.03;0.08). The estimates of κ, α, and λ for fixed ω seem close to unbiased but displays considerable variation. The right plot in Figure 2 shows a so-called confidence net for λ obtained from its confidence distribution (Schweder and Hjort, 2002) estimated from the bootstrap simulations. For each level of confidence on the vertical axis, the horizontal interval from the left to the right branch of the net provides a tail-symmetric confidence interval.

For each simulated data set we considered two independent optimizations in order to assess the Monte Carlo error on the parameter estimates. The estimated Monte Carlo standard deviations for κ_6, α_6, and λ_6 are 0.07, 0.2, and 0.002. The Monte Carlo standard deviations seem reasonably small compared with the variability of the bootstrap distributions.

The average time used for a bootstrap simulation and subsequent optimization is around 30 minutes.

5.3 Bayesian inference

From a numerical point of view the Bayesian approach is very advantageous since Monte Carlo estimation of posterior expectations is rather simple compared with maximization of the likelihood function.

Hedley and Buckland (2004) mentions that minke whales in the Antarctic comes in pods of 1-3 animals. This may not be relevant for the north Atlantic minke whales but we use this information to illustrate a Bayesian approach. About 90% of the probability mass of a negative binomial distribution with mean $\alpha = 2$ and variance $2\alpha = 4$ falls on $\{0, 1 \ldots, 5\}$. It therefore seems reasonable to use an informative $N(2, 1)$ prior (truncated at zero) for α. We further impose uniform priors on $\exp(\kappa)$ and ω.

The marginal posterior means and 2.5% and 97.5% quantiles for κ, α, ω, and λ are 13.9 (13.2;14.7), 2.2 (1.0;3.5), 0.7 (0.4;1.0), and 0.06 (0.04;0.08). A posterior credibility net for λ is shown in the right plot in Figure 2, i.e., for a probability q on the vertical axis, the horizontal interval from the left to the right branch of the credibility net provides a tail-symmetric q posterior credibility interval. The posterior means are very similar to the maximum likelihood estimates. The credibility net for λ agrees well with the confidence net, but is slightly narrower.

The MCMC computations for the Bayesian analysis took about 20 minutes.

6 Discussion

The stationary shot-noise Cox process with $\beta = 1$ does of course not fit all line transect data. In addition to the VSS block, we considered another block, VSN, for which the profile likelihood function (see Figure 2) was nearly flat with multiple local maxima. This indicates a lack of fit of the model. An obvious extension of our approach would allow β to be estimated from the data. However, it seems more important to allow for nonstationarity. Hedley and Buckland (2004) models line transect data using nonstationary Poisson processes depending on spatial covariates. Such covariates could easily be incorporated in our model either through a linear model for κ or by multiplying a log-linear term to the random intensity function (2). The model would then account for both small scale clustering and large scale heterogeneity given by the covariates.

Another very flexible class of Cox processes is given by log Gaussian Cox processes (Møller *et al.*, 1998) where the log random intensity function is a Gaussian process. For these processes we would not need to consider extended regions E_i since the marginal distribution of $X \cap B$ is known for a log Gaussian Cox process X on $\mathbb{R}^2$ and a bounded region B. Certain discretization problems must on the other hand be addressed, see Waagepetersen (2004).

Our approach to maximum likelihood estimation is computationally demanding, but the computing times do not seem prohibitive - especially when compared to the time spent on collecting the line transect data. The computing time can moreover easily be reduced by running optimization and bridge sampling computations in parallel on several computers. When priors can be elicited in a sensible way the Bayesian approach is a computationally much easier alternative.

The spatial distribution for territorial animals may be underdispersed relative to a Poisson process. For this case we refer to Baddeley *et al.* (2000) where inference for a thinned repulsive point process is considered.

References

Aldrin, M., Holden, M. & Schweder, T. (2003). Comment on Cowling's "Spatial Methods for Line Transect Surveys". *Biometrics* **59**, 186–188.

Baddeley, A. J., Møller, J. & Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* **54**, 329–350.

Brix, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Advances in Applied Probability (SGSA)* **31**, 929–953.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L. & Thomas, L. (2004). *Advanced distance sampling*. Oxford University Press.

Cowling, A. (1998). Spatial methods for line transect surveys. *Biometrics* **54**, 828–839.

Gelman, A. & Meng, X.-L. (1998). Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statistical Science* **13**, 163–185.

Hagen, G. S. & Schweder, T. (1995). Point clustering of minke whales in the northeastern Atlantic. In: *Whales, Seals, Fish and Man* (eds. A. Schytte Blix, L. Walløe and Ø. Ulltang), Elsevier, Amsterdam, 27–33.

Hedley, S. L. & Buckland, S. T. (2004). Spatial models for line transect sampling. *Journal of Agricultural, Biological, and Environmental Statistics* **9**, 181–199.

Møller, J. (2003). Shot noise Cox processes. *Advances in Applied Probability* **35**, 614–640.

Møller, J. & Waagepetersen, R. P. (2003). *Statistical inference and simulation for spatial point processes*. Chapman and Hall/CRC, Boca Raton.

Møller, J., Syversveen, A. R. & Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics* **25**, 451–482.

Robert, C. P. & Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer-Verlag, 2nd edition.

Schweder, T. (1974). *Transformation of point processes: Application to Animal Sighting and Catch Problems, with special Emphasis on Whales*. Ph.D. thesis, University of California, Berkeley, 183 pp.

Schweder, T. (1977). Point process models for line transect experiments. In: *Recent developments in statistics* (eds. J. R. Barra, B. Van Cutsem, F. Brodeau and G. Romier), North Holland, 221– 242.

Schweder, T. & Hjort, N. L. (2002). Confidence and likelihood. *Scandinavian Journal of Statistics* **29**, 309–332.

Skaug, H. J., Øien, N., Schweder, T. & Bøthun, G. (2004). Abundance of minke whales (balaneoptera acutostrata) in the northeast Atlantic: variability in time and space. *Can. J. Fish. Aquat. Sci.* **61**, 870–886.

Waagepetersen, R. (2004). Convergence of posteriors for discretized log Gaussian Cox processes. *Statistics and Probability Letters* **66**, 229–235.

Wolpert, R. L. & Ickstadt, K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika* **85**, 251–267.

In Appendix A a working knowledge of spatial point process densities and Markov chain Monte Carlo methods is assumed. Background material on these subjects can be found in e.g. Robert and Casella (2004) and Møller and Waagepetersen (2003).

A Conditional simulation using MCMC

The joint density of Y_i and Φ_i with respect to the product of a unit rate Poisson process on S_i and a marked unit rate Poisson process on $E_i \times]0,\infty[$ is given by

$$f_i(y_i,\phi_i;\theta) \propto f_i\big(y_i|Z_Y(\cdot|\phi_i;\omega)\big)\exp\big(|E_i|(1-\exp(\kappa))\big)\prod_{\phi_i}\chi(\gamma;\kappa,\alpha)/\tilde{\chi}(\gamma) \tag{7}$$

where χ is given by (4), $\tilde{\chi}(\cdot)$ is a positive reference mark density on $]0,\infty[$ and $\prod_{\phi_i}$ is short for $\prod_{(c,\gamma)\in\phi_i}\cdot$

Simulations from the conditional distribution of Φ_i given $Y_i = y_i$ with density $f_{i|Y}(\phi_i|y_i;\theta) \propto f_i(y_i,\phi_i;\theta)$ can be obtained using a birth/death MCMC algorithm as described in Chapter 7 in Møller and Waagepetersen (2003). In each MCMC iteration it is then required to compute the integral

$$\int_{S_i} Z_Y(\xi|\phi_i';\omega)\mathrm{d}\xi = \int_{S_i} p(\xi)Z(\xi|\phi_i';\omega)\mathrm{d}\xi$$

appearing in the conditional density $f_{i|Y}(\phi_i'|y_i,\theta)$ (cf. (5) and (7)) when ϕ_i' is the proposal for a new state of the MCMC chain. Numerical quadrature is required to compute the integral due to the rather irregular form of the detection probability $p(\cdot)$.

Instead we may consider a data augmentation approach where we simulate the joint distribution of Φ_i and the unobserved whales $X_i = (X \cap S_i) \setminus Y_i$ within S_i given $Y = y_i$. Given Φ_i, X_i and Y_i are independent Poisson processes and the joint density of (Y_i, X_i, Φ_i) is given by

$$f_i^A(y_i, x_i, \phi_i; \theta) \propto$$
$$p(y_i|y_i \cup x_i) \exp(-\int_{S_i} Z(\xi|\phi_i;\omega)\mathrm{d}\xi) \prod_{\eta \in y_i \cup x_i} Z(\eta|\phi_i;\omega) \prod_{\phi_i} \chi(\gamma;\kappa,\alpha)/\tilde{\chi}(\gamma) \quad (8)$$

where $p(y_i|y_i \cup x_i) = \prod_{\eta \in y_i} p(\eta) \prod_{\eta \in x_i}(1-p(\eta))$. The evaluation of the conditional density $f^A_{i|Y,X}(\phi_i|y_i,x_i;\theta) \propto f_i^A(y_i,x_i,\phi_i;\theta)$ of Φ_i given $Y_i = y_i$ and $X_i = x_i$ is easy since $Z(\eta|\phi_i;\omega)$ is just a sum of scaled truncated Gaussian densities, cf. Section 3.1. The full conditional of X_i given Y_i and Φ_i is simply a Poisson process with intensity function $(1-p(\cdot))Z(\cdot|\phi_i;\omega)$ and is easily simulated using a thinning procedure. We can thus simulate (X_i, Φ_i) given $Y_i = y_i$ by alternating between Gibbs updates for X_i and single point birth/death updates for Φ_i.

For computational speed it is very convenient to work with a kernel of bounded range since a birth or death of a marked cluster centre then only influences the intensity function for the whales in a neighbourhood of the added or removed marked cluster centre.