

# 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 (MCMC) 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 are compared in an example concerning minke whales in the northeast Atlantic.

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

## 1 Introduction

Line transecting together with point sampling are the most widely used techniques for estimating abundance of wild animal or plant populations (Buckland *et al.*, 2004). The spatial point pattern of the animal or plant positions is often clustered relative to a Poisson process. We consider a Cox point process model (defined and motivated in Section 3.1) for such clustered populations,

and develop likelihood-based methods to infer the model parameters from line transect data on animal positions.

In a line transect survey an observer traverses an area at fixed speed along a predetermined transect line. The transect line is often a zigzag consisting of a number of transect legs which are possibly broken into segments due to changes in sighting condition. The observer records the position of each sighted animal and possibly covariate data on sighting conditions. These data may in the first place be used for estimating the unknown detection probability  $p(x, y)$  of observing an animal located at  $(x, y)$ . Our theme, however, is not the estimation of  $p$ . Viewing the animal positions as a realization of a spatial point process, we instead focus on estimation of the animal density and clustering parameters, assuming the detection probability  $p$  given. Assuming independent detection, the set of detected animal positions is a thinned version of the population point process with thinning probability  $1 - p(x, y)$ . Line transect data were first regarded as thinnings of point processes in Schweder (1974) and Schweder (1977).

With  $n$  animals observed over the transect,

$$\tilde{\lambda} = n / \int p(x, y) dx dy \quad (1)$$

is the moment estimate of mean animal density, and also the maximum likelihood estimate of the intensity parameter of a spatially homogeneous Poisson population process when animals are detected independently of each other. Under clustering, this estimate is inefficient and to evaluate its variance it is necessary to quantify the degree of clustering. We address these issues using a parametric Cox process model for the clustering, and likelihood methods to infer the unknown parameters.

A computationally easy approach to parameter estimation 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. For a stationary point process with intensity  $\lambda$ ,  $\lambda K(t)$  is the expected number of further points within distance  $t$  from a typical point of the point process. A non-parametric estimate of the  $K$ -function may be obtained from line transect data as discussed in Baddeley *et al.* (2000). Animal positions are observed within narrow strips along the transect and an unbiased estimate of  $K(t)$  can only be obtained for  $t$  smaller than the strip width. The estimate is moreover highly variable, see Section 5.4.

Other methods of inference are based on projecting the detected points onto the transect line whereby a one-dimensional point process is obtained.

Cowling (1998) (see also Aldrin *et al.*, 2003) considers the  $K$ -function for the projection of a thinned Neyman-Scott process. The thinning probability is here assumed centred Gaussian with constant scale parameter. In practice, however, the thinning probability is usually varying along the transect according to sighting conditions, see e.g. Skaug *et al.* (2004). In Buckland *et al.* (2004, Chapter 4), the projected process is assumed to be inhomogeneous Poisson with intensity depending on observed covariates. Skaug (2006) considers a one-dimensional Cox point process with random intensity modulated by a latent two-state Markov process.

In this paper we develop likelihood-based inference for a thinned spatial Cox process both in a frequentist and a Bayesian setting. The inference is implemented using simulation methodology. A distinct advantage of our approach (as opposed to the one in Cowling, 1998) is that we do not need simplifying assumptions regarding the functional form of the detection probability. Our approach can moreover easily be adapted to other sampling designs of the distance type (Buckland *et al.*, 2004). It can also be extended to take into account large scale heterogeneity due to spatially varying covariates (see Section 6). This would extend the approach in Hedley and Buckland (2004) who consider spatially inhomogeneous Poisson processes.

Applications of maximum likelihood estimation and Bayesian inference for spatial Cox processes in general are still very rare in the literature. Our paper therefore also serves as a case study of general computational issues concerning the implementation of likelihood-based inference for spatial Cox processes.

Our discussion will be focused on a particular line transect study of minke whale abundance in the northeast Atlantic, see Section 2. Sections 3 and 4 describe our model and the computational approach while Section 5 contains an application to the minke whale data including model assessment using the  $K$ -function. Section 6 contains some final remarks.

## 2 Minke whales

Minke whales (*balaneoptera acutorostrata*) are subject to commercial whaling in the northeast Atlantic. Catch quotas are calculated using, among other sources of information, periodic abundance estimates (International Whaling Commission, 2004). Skaug *et al.* (2004) gave the most recent abundance estimate of about 107 000 (coef. of var. 0.13) summering minke whales in the northeast Atlantic including waters around Jan Mayen. The total area is divided into blocks and the abundance estimate is based on separate line transect surveys within the blocks. Each block was surveyed once in one of the years 1996-2001 except the Lofoten block which was surveyed twice.

Here we focus on the survey block named VSS located west of Spitzbergen. This block was visually surveyed in 1999 with 50 whales observed over a transect with  $m = 7$  transect legs (see Figure 1). A few comments regarding maximum likelihood estimation for a neighbouring survey block VSN (not shown) are given in Section 5 and 6. Following Skaug *et al.* (2004), we regard the whales as immobile since the vessel travels much faster than minke whales usually do.

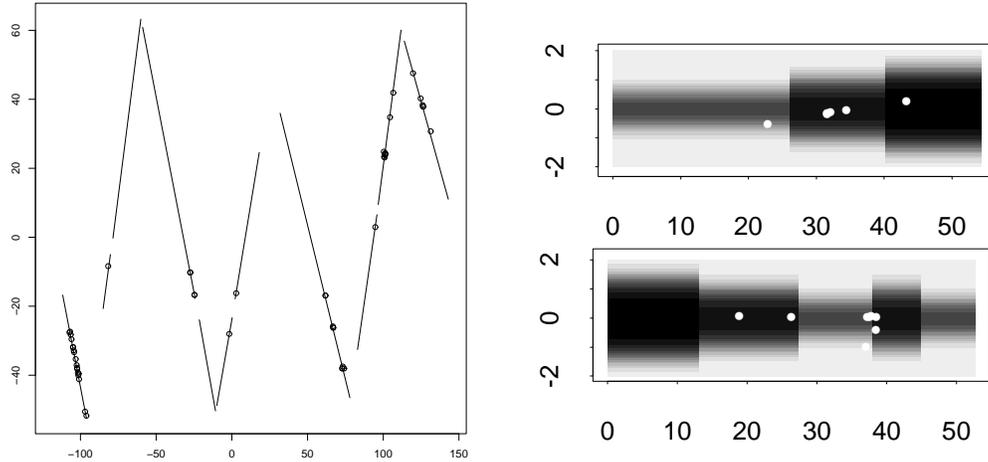


Figure 1: Left: transect and observed whales for the VSS block with distances in km. The transecting was broken when sighting conditions became unsuitable due to sea state, fog, or darkness, and restarted along the transect leg when conditions improved. Right: gray scale plots of the detection probability for the two rightmost transect leg segments in the VSS block. Dark means high detection probability and white dots show positions of observed whales. The numbers on the  $x$  and  $y$  axes refer to distances in km.

The probability of detecting a whale is considerably less than one even when located right on the transect. Let  $Q(x, y, \tilde{x})$  denote the hazard probability of initially detecting a whale surfacing at position  $(x, y)$  when the ship is at position  $(\tilde{x}, 0)$  on a transect leg along the  $x$ -axis. Sightings must be forward of perpendicular to the vessel so  $Q(x, y, \tilde{x})$  is zero for  $x \leq \tilde{x}$ . 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  $p(x, y) = 1 - \exp\left(-\phi \int_{-\infty}^x Q(x, y, \tilde{x}) d\tilde{x}\right)$ . To estimate  $Q(x, y, \tilde{x})$ , 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  $Q(x, y, \tilde{x})$  can then be obtained from trinomial data with outcomes: surfacing whale observed from a) both platforms b) only from first platform c) only from

second platform, see Skaug *et al.* (2004) for further details.

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 leg 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 positions 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 data, and a bias-corrected estimate is obtained, see Skaug *et al.* (2004).

For our present purpose of fitting a spatial cluster model, the detection probability is given by the estimate obtained in Skaug *et al.* (2004) and we ignore the uncertainty due to the estimation of the detection probability. The right plots in Figure 1 show the estimated detection probabilities for two transect leg segments in the VSS block. The detection probability depends on covariates (sea state, glare, observation team etc) recorded every hour, and cannot easily be explicitly given here. An important feature of our likelihood-based approach is that it easily accommodates 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 leg. Within the time-span of traversing a transect leg 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 leg along the  $x$ -axis and origin at the start of the transect leg. The whales in the vicinity of the transect leg are regarded as a subset of a planar stationary point process  $X$  whose intensity  $\lambda$  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  $\xi$ . 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 northeast 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. The distribution of a Cox process in the plane is governed by a non-negative random intensity function  $Z = \{Z(\xi)|\xi \in \mathbb{R}^2\}$ . Given a realization  $z$  of the random intensity function, the Cox process is a Poisson process with intensity function  $z$ . In this paper we consider an example of a so-called shot-noise Cox process (Brix, 1999; Møller, 2003; Møller and Waagepetersen, 2003). The random intensity function is given by

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

where the kernel  $k$  is a probability density and  $\Phi$  is a homogeneous marked Poisson process. That is,  $\Phi = \{(c, \gamma)|c \in C\}$  where  $C$  is a homogeneous Poisson process and given  $C$ , the marks  $\gamma > 0$  are independent and identically distributed. This Cox process can also be viewed as a cluster process, i.e. conditional on  $\Phi$ ,  $X$  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. Conditional on  $\Phi$ ,  $\gamma$  is the expected number of points in the cluster. The process  $Y$  of observed whales is a shot-noise Cox process with random intensity function

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

### 3.2 Parametric model

In the minke whale context we parametrize the model by  $\theta = (\kappa, \alpha, \omega)$  where  $\exp(\kappa)$  is the intensity of cluster centres,  $\alpha = \mathbb{E}\gamma$  is the mean number of whales per cluster, and  $\omega$  is the scale parameter (spread) of the kernel. The whale intensity  $\lambda = \alpha \exp(\kappa)$  is the parameter of main interest.

We assume that  $\gamma$ , the number of whales per cluster, is standard gamma distributed with shape and scale parameter  $\alpha$  and 1, respectively, and that the kernel is a truncated bivariate Gaussian density with scale parameter  $\omega > 0$ , i.e.

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

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 advantageous for computational reasons, see Appendix B. 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  $2 + \alpha$ .

In addition to being an example of a shot noise G Cox process (see Appendix A), our shot-noise Cox process is moreover an example of a Neyman-Scott process with negative binomial numbers of points in each cluster. From

the point of view of constructing a random intensity function, the use of gamma distributed marks  $\gamma$  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 the Thomas process, the random intensity function is obtained by a superposition of Gaussian kernels all multiplied with the *same* positive parameter. The Thomas model does not seem to allow for the amount of dispersion seen in our example, see Figure 1 where there are many isolated points and at least one cluster with many points.

## 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 transect legs. 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 leg separately and then added to obtain a composite log likelihood function based on all of the transect legs.

For a survey with  $m$  transect legs, we use the composite log likelihood  $l(\theta) = \sum_{i=1}^m \log L_i(\theta)$  where  $L_i(\theta)$  is the likelihood of the data from the  $i$ th leg. Dependence between the likelihood components  $L_i(\theta)$  is due to clusters of whales which can be observed from more than one transect leg. For our whale data the spatial extent (determined by  $\omega$ ) of the clusters is small relative to the separation between the transect legs. It thus seems reasonable to consider the likelihood components independent. We therefore in the following refer to the composite likelihood function and maximum composite likelihood estimates as the likelihood function and maximum likelihood estimates, respectively.

In Section 5.1 we use a profile likelihood approach where  $l(\theta)$  is maximized with respect to  $(\kappa, \alpha)$  for a finite set of  $\omega$  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  $\omega$ . The likelihood may further be highly multimodal as a function of  $\omega$ , 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 and information matrix are obtained by summing the corresponding quantities obtained from the log likelihood functions  $\log L_i(\theta)$  for each transect leg. Similarly, for two values  $\theta_1 = (\kappa_1, \alpha_1, \omega_1)$  and  $\theta_2 =$

$(\kappa_2, \alpha_2, \omega_2)$  of the parameter vector, the log likelihood ratio  $l(\theta_1) - l(\theta_2)$  is given by the sum  $\sum_{i=1}^m \log(L_i(\theta_1) - L_i(\theta_2))$ . It therefore suffices to work out the likelihood function for one generic transect leg.

## 4.1 Likelihood function for one transect leg

To simplify notation we drop the transect leg index  $i$  in this and the following section. Let  $S$  denote the bounded support of the detection probability for a transect leg - in our application  $S$  is a narrow rectangular strip around the transect leg, see the right plots in Figure 1. The conditional density of  $Y_S = Y \cap S$  given  $\Phi$  in (2) is the Poisson process density

$$f(y|Z_Y(\cdot|\Phi;\omega)) = \exp(|S| - \int_S Z_Y(\xi|\Phi;\omega)d\xi) \prod_{\eta \in y} Z_Y(\eta|\Phi;\omega). \quad (4)$$

Note that the conditional density only depends on  $\Phi$  through the finite point process of cluster centres with positive probability of contributing with offspring inside  $S$ . More specifically, this finite point process is  $\Phi_E = \{(c, \gamma) \in \Phi | c \in E\}$  where  $E$  is the rectangle  $\{\xi \in \mathbb{R}^2 | \exists \eta \in S : k(\xi - \eta; \omega) > 0\}$ . The likelihood function for the transect leg is thus

$$L(\theta) = \mathbb{E}_{(\kappa, \alpha)} f(y|Z_Y(\cdot|\Phi;\omega)) = \mathbb{E}_{(\kappa, \alpha)} f(y|Z_Y(\cdot|\Phi_E;\omega))$$

where  $\mathbb{E}_{(\kappa, \alpha)}$  denotes expectation with respect to  $\Phi$  or  $\Phi_E$  whose distributions depend on  $\kappa$  and  $\alpha$ . Denote by  $v$  the number of points in  $\Phi_E$  and let  $w = \sum_{(c, \gamma) \in \Phi_E} \log \gamma$ . The marginal density of  $\Phi_E$  is given by the Poisson process density of the cluster centres times the densities of the standard gamma marks:

$$f(\phi; \kappa, \alpha) = \exp(|E|(1 - \exp(\kappa)) \exp(\kappa)^v \prod_{(c, \gamma) \in \phi} \gamma^{\alpha-1} \exp(-\gamma)/\Gamma(\alpha) = \exp(|E|(1 - \exp(\kappa)) + (\kappa - \log \Gamma(\alpha))v + (\alpha - 1)w - \sum_{(c, \gamma) \in \phi} \gamma). \quad (5)$$

Approximations of likelihood ratios  $L(\theta_2)/L(\theta_1)$  are obtained using bridge sampling, see Appendix C. To compute approximate derivatives of  $\log L(\theta)$  (see Section 4.2) or bridge sampling likelihood ratios  $L(\theta_2)/L(\theta_1)$  we need conditional simulations of the “missing data”  $\Phi_E$  given  $Y_S$ . An algorithm for this is discussed in Appendix B. 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  $\omega$  and let  $V_\theta(Y_S, \Phi_E) = d \log f(Y_S, \Phi_E; \theta) / d(\kappa, \alpha)$  where

$$f(y, \phi; \theta) \propto f(y|Z_Y(\cdot|\phi; \omega)) f(\phi; \kappa, \alpha) \quad (6)$$

is the joint density of  $(Y_S, \Phi_E)$ , see (4) and (5). Following Section 8.6.2 in Møller and Waagepetersen (2003), the score function is given by

$$u(\kappa, \alpha) = \mathbb{E}_{\theta, y} V_\theta(Y_S, \Phi_E) = (\mathbb{E}_{\theta, y} v - \exp(\kappa)|E|, \mathbb{E}_{\theta, y} w - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \mathbb{E}_{\theta, y} v)$$

where  $\mathbb{E}_{\theta, y}$  denotes conditional expectation with respect to  $\Phi_E$  given  $Y_S = y$  and  $\Gamma(\alpha)$  is the gamma function. Similarly, the observed information matrix is

$$j(\kappa, \alpha) = -\mathbb{E}_{\theta, y} dV_\theta(Y_S, \Phi_E) / d(\kappa, \alpha)^\top - \text{Var}_{\theta, y} V_\theta(Y_S, \Phi_E) =$$

$$\begin{bmatrix} \exp(\kappa)|E| & 0 \\ \frac{d^2 \log \Gamma(\alpha)}{d\alpha^2} \mathbb{E}_{\theta, y} v & \end{bmatrix} - \begin{bmatrix} \text{Var}_{\theta, y} v & \text{Cov}_{\theta, y}[v, w] - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \text{Var}_{\theta, y} v \\ \text{Cov}_{\theta, y}[w, v] - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \text{Cov}_{\theta, y} v & \text{Var}_{\theta, y}[w - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} v] \end{bmatrix} \quad (7)$$

where  $\text{Var}_{\theta, y}$  and  $\text{Cov}_{\theta, y}$  denotes conditional variance and covariance, and where the matrices are symmetric with only the upper triangle shown. The first term in (7) is the conditional expectation of the observed information in the case where  $v$  and  $w$  are observed. The first and second derivatives of the gamma function are known as the digamma and trigamma functions and are available in many statistical or mathematical software packages.

We could reparametrize letting  $\kappa := \log(\kappa - \Gamma(\alpha))$  in which case an exponential family density with sufficient statistic  $t = (v, w)$  would be obtained for  $\Phi_E$ , cf. (5). Then we obtain particularly neat expressions for the score function and observed information:  $u(\kappa, \alpha) = \mathbb{E}_{\theta, y} t - \mathbb{E}_\theta t$  and  $j(\kappa, \alpha) = \text{Var}_\theta t - \text{Var}_{\theta, y} t$ . However, with the original parametrization a more well-conditioned observed information matrix is obtained. A third option is to parametrize in terms of  $(\kappa, \log \lambda) = (\kappa, \kappa + \log \alpha)$ . This gives a somewhat better conditioned observed information than with the  $(\kappa, \alpha)$  parametrization but the expression is rather messy and omitted here.

The expectations appearing in the score function and the information matrix cannot 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  $\Phi_E$  given  $Y_S = y$ , see Appendix B.

## 4.3 Bayesian approach

In the Bayesian framework we introduce a prior density  $p(\theta)$  and consider the joint posterior distribution of  $(\theta, (\Phi_{E_i})_{i=1}^m)$  where we reintroduce the index  $i =$

$1, \dots, m$ , for the  $m$  transect legs. Assuming independence between transect legs, the posterior density is given by

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

A Markov chain Monte Carlo (MCMC) algorithm for posterior simulation can be obtained by combining the MCMC algorithm from Appendix B with Metropolis-Hastings updates for  $\theta$  (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 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  $\omega_l = l/10\text{km}$ ,  $l = 2, \dots, 30$ , 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  $\omega$  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 corresponding estimates  $\kappa_6 = -3.7$ ,  $\alpha_6 = 2.4$  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  $(\kappa_l, \alpha_l)$  and a log likelihood ratio  $l(\theta_{l+1}) - l(\theta_l)$  took around 70 minutes.

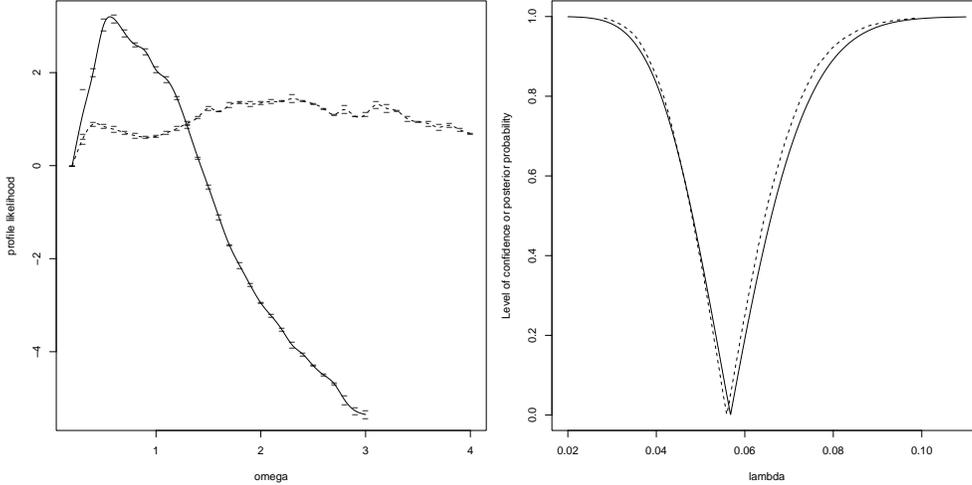


Figure 2: Left: Profile log likelihood functions  $l_p(\omega) = \max_{(\kappa, \alpha)} l(\theta)$  obtained for the VSS (solid line) and VSN (dotted line) blocks by cumulating log likelihood ratios  $l(\theta_{l+1}) - l(\theta_l)$ . The small vertical bars indicate Monte Carlo confidence intervals for the differences  $l(\theta_{l+1}) - l(\theta_l)$ . Right: confidence (solid line) and posterior credibility nets (dotted line) for  $\lambda$ , see Section 5.2 and Section 5.3. For each confidence level/posterior probability on the vertical axis, the horizontal intervals between the left and right branches of the curves provide a confidence and posterior credibility interval, respectively.

$\omega_l$	0.2	0.4	0.6	0.8	1
$\kappa_l$	-2.2	-3.0	-3.7	-4.1	-4.4
$\alpha_l$	0.5	1.2	2.4	3.7	4.9
$\lambda_l$	0.06	0.06	0.06	0.06	0.06

Table 1: A collection of estimates  $\kappa_l$ ,  $\alpha_l$  and  $\lambda_l$  obtained by maximizing the log likelihood  $l(\kappa, \alpha, \omega_l)$  with respect to  $\kappa$  and  $\alpha$ .

To illustrate how  $(\kappa_l, \alpha_l, \lambda_l)$  depends on  $\omega_l$ , a collection of estimates are given in Table 1. The estimates  $\kappa_l$  and  $\alpha_l$  are vary considerably as a function of  $\omega_l$  whereas  $\lambda_l$  is essentially constant.

## 5.2 Parametric bootstrap

The repeated sampling properties of the parameter estimates are studied using a parametric bootstrap based on 100 independent simulated data sets. The data sets are simulated 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  $\omega$

is assumed known and equal to the maximum likelihood estimate. In a full parametric bootstrap we should also maximize with respect to  $\omega$ . However, Table 1 suggests that regarding the estimate of  $\lambda$ , it does not matter much whether we maximize the likelihood function over all three parameters  $\kappa$ ,  $\alpha$ , or  $\omega$  or maximize only over  $\kappa$  and  $\alpha$  for fixed  $\omega$ .

Bootstrap estimates of the means for the sampling distributions of the estimates  $\kappa_6$ ,  $\alpha_6$  and  $\lambda_6$  for fixed  $\omega = \omega_6 = 0.6$  and the moment estimate (1) are -3.6 (-3.7), 2.3 (2.4), 0.06 (0.06), and 0.06 (0.06), respectively, with the parameter values used for the bootstrap simulation given in parantheses. The 2.5% and 97.5% quantiles are  $\kappa_6$ : (-4.2;-2.9),  $\alpha_6$ : (0.7;4.5),  $\lambda_6$ : (0.03;0.08), and moment estimate: (0.03;0.08). The estimates of  $\kappa$ ,  $\alpha$ , and  $\lambda$  for fixed  $\omega$  seem close to unbiased but displays considerable variation. The right plot in Figure 2 shows a so-called confidence net for  $\lambda$  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.

Our estimates are affected by sampling variation, but also by Monte Carlo error due to the likelihood derivatives being evaluated using MCMC. We estimate the Monte Carlo error by pairwise comparison of two independent optimizations for each simulated data set. The estimated Monte Carlo standard deviations for  $\kappa_6$ ,  $\alpha_6$ , and  $\lambda_6$  are 0.1, 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) mention that minke whales in the Antarctic comes in pods of 1-3 animals. 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  $\{0, 1, \dots, 5\}$ . It therefore seems reasonable to use an informative  $N(2, 1)$  prior (truncated at zero) for  $\alpha$ . We further impose uniform priors on  $\exp(\kappa)$  and  $\omega$  on the bounded intervals  $]0.01, 0.2[$  and  $]0.1, 1.5[$ , respectively.

The marginal posterior means and 2.5% and 97.5% quantiles for  $\kappa$ ,  $\alpha$ ,  $\omega$ , and  $\lambda$  are -3.6 (-4.3;-2.8), 2.2 (1.0;3.5), 0.7 (0.4;1.0), and 0.06 (0.04;0.08). A

posterior credibility net for  $\lambda$  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  $\lambda$  agrees well with the confidence net, but is slightly narrower due to the use of prior information.

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

## 5.4 Spatial $K$ -function for whales

We conclude our study of the VSS whale data by considering a non-parametric estimate of the  $K$ -function for the whale process  $X$ . Based on the data  $y_i$  for the  $i$ th transect leg, a non-parametric estimate of the  $K$ -function is given by

$$\hat{K}_i(t) = \sum_{\xi, \eta \in y_i} \frac{1[0 < \|\xi - \eta\| < t]}{\tilde{\lambda}p(\xi)\tilde{\lambda}p(\eta)} w_{\xi, \eta}$$

where  $\tilde{\lambda}$  is the moment estimate (1) of the intensity and  $w_{\xi, \eta}$  is an edge correction factor, see Baddeley *et al.* (2000) or Section 4.3.2 in Møller and Waagepetersen (2003). The edge correction requires that  $t$  is less than the width of the strip  $S_i$  (4 km in our application). Our estimate based on all the transect legs is simply the average  $\hat{K}(t) = \sum_{i=1}^m \hat{K}_i(t)/m$ . Figure 3 shows  $\hat{L}(t) - t$  where  $\hat{L}(t) = \sqrt{\hat{K}(t)/\pi}$ . For a Poisson process,  $L(t) - t = \sqrt{K(t)/\pi} - t$  is zero for all  $t$  so our estimate  $\hat{L}(t) - t$  which takes values larger than zero indicates clustering (the theoretical value of  $L(t) - t$  is  $\sqrt{t^2 + [1 - \exp(-t/(2\omega)^2)]/(\pi\kappa)} - t > 0$  under the shot noise Cox process). The dotted curves in Figure 3 are 95% pointwise confidence bands: for each  $t > 0$  the corresponding values of the dotted curves provide a 95% confidence interval for  $\hat{L}(t) - t$  under the shot noise Cox process. The confidence bands illustrates the large variability of  $\hat{K}(t)$ . Except for very small  $t$ ,  $\hat{L}(t) - t$  falls within the confidence bands. Thus the plot does not provide strong evidence against our model (the  $p$ -value obtained from a Monte Carlo test based on the integrated squared distance between  $\hat{L}(t) - t$  and the theoretical value of  $L(t) - t$  under the fitted shot noise Cox process is 21%).

The  $K$ -function uniquely characterizes the second order properties of a stationary point process. We also investigated the first order properties of our model by comparing the observed counts of whales within each transect leg with their sampling distributions under the fitted shot noise Cox process. Only the high count (18) of whales in the lower left transect leg in Figure 1 falls outside the 0.025% and 0.975% quantiles (0 and 11, respectively).

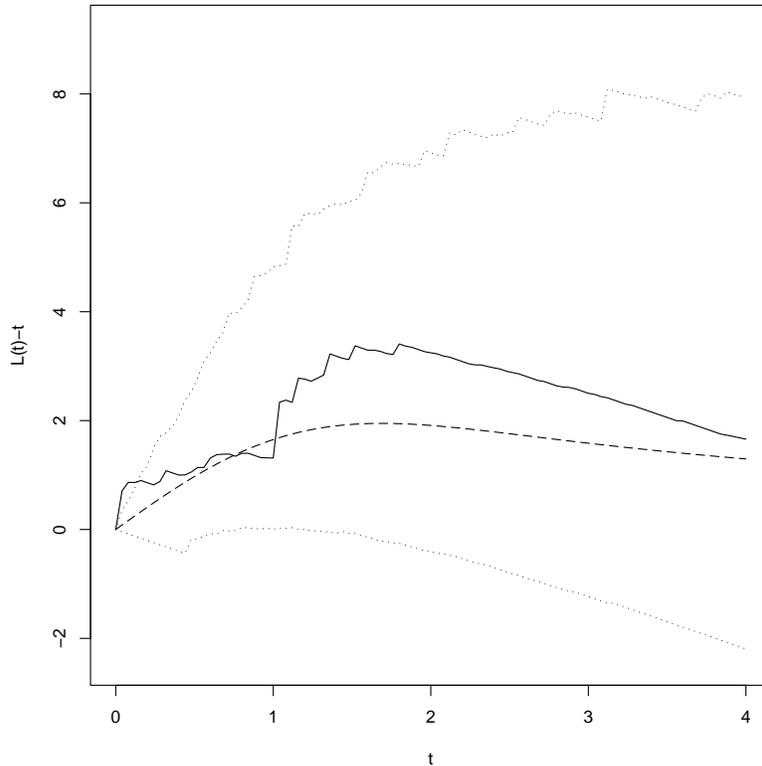


Figure 3: Solid line,  $\hat{L}(t) - t$  (distance  $t$  in km); dotted lines, 95% confidence band based on simulations of the fitted shot noise Cox process; dashed line,  $L(t) - t > 0$  for the fitted shot noise Cox process. A Poisson process has  $L(t) - t = 0$ .

## 6 Discussion

The profile likelihood function (Figure 2) for the VSS data has a well-defined maximum at  $\omega = 0.6$ . In addition to the VSS block, we considered another block, VSN, for which the profile likelihood function was nearly flat with multiple local maxima, see Figure 2. One may speculate whether this indicates aggregation at different spatial scales where large scale aggregation might be explained by spatial covariates concerning, say, sea depth or water temperature. It is also plausible that such covariates might explain the high count of whales in the lower left transect leg in Figure 1 which is also causes the slight lack of fit seen in Figure 3 for small  $t$ . Failing to take possible extra sources of variation into account may lead to an underestimation of the uncertainty of the whale intensity estimate. Including too many sources of variation might on the other hand lead to lack of identification given the

relatively sparse data.

Covariates can easily be incorporated in our model by multiplying a positive function  $e(\cdot)$  depending on the covariates to the random intensity function (2). The random intensity function for the observed whales is then

$$Z_Y(\xi|\Phi) = e(\xi)p(\xi)Z(\xi|\Phi)$$

which accounts for both small scale clustering and large scale aggregation due to the covariates. Note that  $Z_Y(\xi|\Phi) = \tilde{p}(\xi)MZ(\xi|\Phi)$ ,  $\xi \in S$ , where  $M = \max_{\xi \in S} e(\xi)p(\xi)$  and  $\tilde{p}(\xi) = e(\xi)p(\xi)/M$ . Hence, our simulation algorithm from Appendix B is still applicable if we replace  $p(\cdot)$  with  $\tilde{p}(\cdot)$  and the standard gamma marks with  $\Gamma(\alpha, M)$  marks.

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  $E$  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  $B$ . On the other hand, for computational reasons it is required to discretize the Gaussian process and this introduces the problem of choosing a suitable discretization, see Waagepetersen (2004).

Skaug *et al.* (2004) estimated the detection probability  $p$  through a parametric model for the hazard probability  $Q$ . A simulation approach accounting for uncertainty in this estimate and that of the surfacing rate  $\phi$  was employed to obtain an approximate confidence distribution for the abundance of minke whales in the survey area. A similar approach might be taken in our shot-noise Cox process setup to account also for uncertainty in the detection probability.

Our approach to maximum likelihood estimation is computationally demanding, but the computing times do not seem prohibitive. The computing time can moreover easily be reduced by running optimization and bridge sampling computations in parallel on several computers. The Bayesian approach is a computationally much easier alternative. However, the specification of priors may be a controversial issue. For instance, not everyone might agree on the informative prior for  $\alpha$  used in Section 5.3 which in turn induces an informative prior on  $\lambda$ .

**Acknowledgement** We are grateful to the editor and the two referees for detailed and valuable comments and suggestions.

## 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.
- International Whaling Commission (2004). Report of the scientific committee. *Journal of Cetacean Research and Management* **6**, 171–179.
- 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, New York, 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. (2006). Markov modulated Poisson processes for clustered line transect data. *Environmental and Ecological Statistics* **13**, to appear.
- Skaug, H. J., Øien, N., Schweder, T. & Bøthun, G. (2004). Abundance of minke whales (balaneoptera acutorostrata) in the northeast Atlantic: variability in time and space. *Canadian Journal of Fisheries and Aquatic Sciences* **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-C 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. Møller and Waagepetersen (2003) and Robert and Casella (2004).

## A Relation to shot noise G Cox processes

Our shot-noise Cox process is a special case of a shot noise G Cox process (Brix, 1999) which is obtained when  $\Phi$  is a Poisson process with intensity function of the form

$$\zeta(c, \gamma) = \exp(\kappa)\beta^{-\alpha}\gamma^{\alpha-1} \exp(-\gamma/\beta)/\Gamma(\alpha) \quad (8)$$

where  $\kappa \in \mathbb{R}$ ,  $\alpha > -1$ , and  $\beta > 0$ . With our rather small data set we can not determine well both  $\alpha$  and  $\beta$  and hence choose to fix  $\beta = 1$ . Alternatively one might fix  $\alpha = 0$  whereby a so-called Poisson-gamma process (Wolpert and Ickstadt, 1998) is obtained. With  $\alpha \leq 0$  the process of cluster centres  $C = \{c \in \mathbb{R}^2 : (c, \gamma) \in \Phi \text{ for some } \gamma\}$  is not locally finite and this is a

nuisance for computational reasons.

The notation in the following appendices B and C is as in Sections 4.1 and 4.2 where we omit the index  $i$  for the different transect legs.

## B Conditional simulation using MCMC

The conditional density of  $\Phi_E$  given  $Y_S = y$  is proportional to the joint density (6). Simulations from the conditional distribution of  $\Phi_E$  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

$$\int_S Z_Y(\xi|\phi';\omega)d\xi = \int_S p(\xi)Z(\xi|\phi';\omega)d\xi$$

appearing in  $f(y, \phi'; \theta)$  (cf. (4) and (6)) when  $\phi'$  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 consider a data augmentation approach where we simulate the joint distribution of  $\Phi$  and the unobserved whales  $X_S = (X \cap S) \setminus Y_S$  within  $S$  given  $Y_S = y$ . Given  $\Phi_E$ ,  $X_S$  and  $Y_S$  are independent Poisson processes and the joint density of  $(Y_S, X_S, \Phi_E)$  is given by

$$f^A(y, x, \phi; \theta) \propto p(y|y \cup x)f(y \cup x|Z(\cdot|\phi; \omega))f(\phi) \quad (9)$$

where  $p(y|y \cup x) = \prod_{\eta \in y} p(\eta) \prod_{\eta \in x} (1 - p(\eta))$  is the probability of observing  $y$  given all whales  $Y_S \cup X_S = y \cup x$  and

$$f(y \cup x|Z(\cdot|\phi; \omega)) = \exp(|S| - \int_S Z(\xi|\phi; \omega)d\xi) \prod_{\eta \in y \cup x} Z(\eta|\phi; \omega)$$

is the Poisson process density of  $Y_S \cup X_S$  given  $\Phi_S = \phi$ .

The conditional density of  $\Phi_E$  given  $Y_S = y$  and  $X_S = x$  is proportional to  $f^A(y, x, \phi; \theta)$  and easy to evaluate since  $Z(\eta|\phi; \omega)$  is just a sum of scaled truncated Gaussian densities, cf. Section 3.1. The full conditional of  $X_S$  given  $Y_S$  and  $\Phi_E$  is simply a Poisson process with intensity function  $(1 - p(\cdot))Z(\cdot|\phi; \omega)$ . We then simulate  $(X_S, \Phi_E)$  given  $Y_S = y$  using a Gibbs/Metropolis-within-Gibbs algorithm where we alternate between the following two steps:

1. Gibbs update for  $X_S$  given  $\Phi_E$  and  $Y_S$  where the current state of  $X_S$  is replaced by a simulation of a Poisson process with intensity function  $(1 - p(\cdot))Z(\cdot|\phi; \omega)$ .

2. Single point Metropolis birth/death updates for  $\Phi_E$  given  $X_S$  and  $Y_S$ .

For computational speed it is 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.

## C Computation of log likelihood ratios

A likelihood ratio  $L(\theta_2)/L(\theta_1)$  can be calculated using bridge sampling (Gelman and Meng, 1998; Møller and Waagepetersen, 2003):

$$\frac{L(\theta_2)}{L(\theta_1)} = \prod_{l=0}^{k-1} \frac{\mathbb{E}_{\theta_{2l}^b, y}[f^A(y, X_S, \Phi_E; \theta_{2l+1}^b)/f^A(y, X_S, \Phi_E; \theta_{2l}^b)]}{\mathbb{E}_{\theta_{2l+2}^b, y}[f^A(y, X_S, \Phi_E; \theta_{2l+1}^b)/f^A(y, X_S, \Phi_E; \theta_{2l+2}^b)]} \quad (10)$$

where  $f^A$  is given by (9),  $\theta_0^b = \theta_1$ ,  $\theta_{2k}^b = \theta_2$  and  $\theta_j^b$ ,  $j = 1, \dots, 2k - 1$  are “intermediate” parameter values between  $\theta_1$  and  $\theta_2$ , e.g. obtained by linear interpolation. In each factor in (10),  $f^A(y, X_S, \Phi_E; \theta_{2l+1}^b)$  is a “bridge” between  $f^A(y, X_S, \Phi_E; \theta_{2l}^b)$  and  $f^A(y, X_S, \Phi_E; \theta_{2l+2}^b)$ . An approximation of the likelihood ratio is obtained by replacing the conditional expectations with Monte Carlo estimates.

If  $\theta_{j+1}^b$  and  $\theta_j^b$  are not sufficiently “close”, the conditional variance of a ratio  $f^A(y, X_S, \Phi_E; \theta_{j+1}^b)/f^A(y, X_S, \Phi_E; \theta_j^b)$  may be huge so that very large Monte Carlo samples are needed to compute the conditional expectation of the ratio. Some pilot experiments are typically needed in order to determine a suitable number  $k$ .

Note that for each expectation  $\mathbb{E}_{\theta_{2l}^b, y}[f^A(y, X_S, \Phi_E; \theta_{2l+1}^b)/f^A(y, X_S, \Phi_E; \theta_{2l}^b)]$ , the extended window  $E$  must be chosen according to the maximal  $\omega$  value in  $\theta_{2l}^b$  and  $\theta_{2l+1}^b$ , see Section 4.1.