Inferring spatial cluster structure from line transect data

Hans J. Skaug

May 11, 2009

Department of Mathematics, University of Bergen, Norway.

Abstract

We propose a new method, based on the Markov modulated Poisson process, for fitting a spatial Neyman-Scott process to line transect data. The method is applied to sighting survey data for Northeastern Atlantic minke whales for the period 2002-2007. Variation in effective strip half-width along the transect line is accounted for. We obtain parameter estimates for the Neyman-Scott process by survey block.

Keywords: Clustering, Line transect surveys, Markov modulated Poisson process, Minke whales.

1 Introduction

The estimation approach used for Northeastern Atlantic minke whales uses a Neyman-Scott process to represent the spatial distribution of whales (Hagen & Schweder 1994, Aldrin et al. 2002, Skaug et al. 2004). Here we present a new method for estimating the parameters of the Neyman-Scott process, and apply it to sighting survey data from the period 2002-2007 (Bøthun et al. 2009).

The spatial positions of individuals in a population may be viewed as the 'points' in stochastic point process. Assume that a line transect survey (Buckland et al. 2001) has been conducted, possibly for the purpose of estimating animal abundance. We can 'project' each detected animal onto the transect line, by marking the position of the observer at each time an animal is detected (Schweder 1977). This generates a one-dimensional point process on the transect line (the encounter process). Figure 1 shows the projections for a multi year sighting survey for minke whales conducted in the Northeastern Atlantic during the period 2002-2007. The aim of the present paper is to estimate the parameters in the underlying two-dimensional point process from the line transect data.

An important characterization of the spatial point process is the degree of clustering measured relative to a Poisson process. Large scale variation in animal density along the transect line may be explained in terms of geographical covariates as in Hedley & Buckland (2004), but clustering on a smaller scale is most conveniently treated as a stochastic phenomenon. Small-scale clustering may be of biological interest, and affect the precision of line transect estimators

of abundance. The presence of small-scale clustering in the minke whale surveys is apparent in Figure 1. Consider as an example the survey block GA in which we can say that there are between 5 and 10 clusters, depending on how we count. Under the model to be fit later the expected number of clusters is eight.

A two-dimensional Neyman-Scott (NS) process with Poisson cluster size has been used to model whale populations (Hagen & Schweder 1994, Schweder et al. 1999, Cowling 1998, Skaug et al. 2004, Waagepetersen & Schweder 2006). Most of these authors have estimated the parameters of the NS process by minimizing the distance between theoretical and empirical versions of Ripley's K function for the one dimensional point process. This is in effect a moment estimator, and the approach does not have the benefits of likelihood methods, such as model selection via the AIC criterion. Waagepetersen & Schweder (2006) used a MCMC algorithm to evaluate the likelihood function, but the method was computationally intensive.

We present a two-dimensional Cox process with two levels of animal density: ρ_1 (background) and ρ_2 (high density regions). The high density regions are assumed to be disk-shaped with a stochastic radius R, and disks-centers located according to a homogeneous Poisson process. We choose the distribution of R such that the 'projection process' (onto the transect line) has properties similar to a Markov modulated Poisson process (MMPP). The likelihood function for the MMPP is readily available (Skaug 2006), and hence maximum likelihood estimation for the proposed Cox process is computationally feasible.

The goal of the paper is to present the cluster process, to investigate its properties and make comparison to the NS process. The rest of the paper is organized as follows: Section 2 reviews the MMPP and introduces the moment matching criteria. Section 3 applies the methodology to real data.

2 Theory

We start out by reviewing the MMPP in the context of line transect surveys.

2.1 The Markov modulated Poisson processes

The MMPP is a generalization of the ordinary homogeneous one-dimensional Poisson process, in that it allows the encounter rate (number of animals detected per unit distance) to attain two levels: λ_1 (low) and λ_2 (high). Which of these levels is attained at a given point is determined by a latent continuous time Markov process with switching intensities μ_1 ($\lambda_1 \to \lambda_2$ transition) and μ_2 ($\lambda_2 \to \lambda_1$ transition). The λ's are related to the underlying (spatial) animal density ρ through $\lambda_i = 2w\rho_i$ ($i = 1, 2$), where w is the effective strip half-width. The framework allows w to vary along the trackline as a function of covariates related to sighting conditions etc. The parameter vector to be estimated is $\theta = (\rho_1, \rho_2, \mu_1, \mu_2)$, while w is treated as an exogenous variable, i.e. it is assumed to have been estimated using ordinary line transect methods.

Consider a transect leg of length L, and let $y_1, \ldots, y_n$ be the points in the interval $[0, L]$ at which animals are detected. Skaug (2006) gives an expression for the MMPP likelihood based on these data. The log-likelihood, summarized over contributions from different transect legs, is maximized w.r.t. θ using AD Model Builder (Fournier 2001). We allow all four parameters to vary across survey block, using a parametrization which in case of ρ_1 reads $\rho_{1,b} = \rho_1 \exp(\delta_{\rho_1,b})$,

where $\sum_b \delta_{\rho_1,b} = 0$. In effect, we fit a separate MMPP to each survey block. For identificability reasons discussed below, we impose a weak prior (mean-zero and variance 10) on all the δ's, also those for ρ_2, μ_1 and μ_1.

The stationary distribution, which applies to the beginning of each new survey leg, is given by

$$\Pr(\lambda_1) = \frac{\mu_2}{\mu_2 + \mu_1} \qquad \text{and} \qquad \Pr(\lambda_2) = \frac{\mu_1}{\mu_2 + \mu_1}. \tag{1}$$

The overall overall animal density is

$$D = \frac{\mu_2\rho_1 + \mu_1\rho_2}{\mu_2 + \mu_1}. \tag{2}$$

We note that the Poisson process can be obtained as a limit for the MMPP in different ways: 1) $\lambda_1 \to \lambda_2$, 2) $\{\mu_1 \to 0\}$ or $\{\mu_2 \to \infty\}$ which means that the encounter rate is λ_1 all of the time, or 3) $\{\mu_2 \to 0\}$ or $\{\mu_1 \to \infty\}$ fixing the encounter rate at λ_2. To make the model identifiable in the case that data truly comes from a Poisson process we impose a weak prior on all the δ's, as mentioned above.

2.2 The MMPP-2

Imagine a two dimensional cluster process with disk-shaped high density areas (see Figure 2). Given the location and size of disks, points are distributed according to an inhomogeneous Poisson process with density ρ_2 inside disks, and density ρ_1 outside. The radius R of a given disk is assumed to be random with density $h(r)$. Assuming that the transect line is placed randomly in the plane, the conditional distribution of R for a disk that happens to be intersected by the transect line is $h^*(r) \propto r \cdot h(r)$. Cluster centers are randomly distributed with low density so that overlapping disks do not become a problem. We shall refer to this process as the MMPP-2, although 'MMPP' strictly speaking is a misnomer.

Consider a disk that is intersected by the transect line, and denote by V the length of the cord inside the disk (Figure 2). It may be shown that $V = 2R\sqrt{1-U^2}$, where U is uniformly distributed on $(0,1)$. Because we want to match the MMPP-2 with the one-dimensional MMPP we make the assumption that V is exponentially distributed with rate μ_2. Through the relationship $R = V\left(2\sqrt{1-U^2}\right)^{-1}$ this assumption induces a distribution on R. However, it may shown that the resulting density satisfies $\lim_{r\to 0} h^*(r) > 0$, and hence $h(r) \propto r^{-1} \cdot h^*(r)$ is not a proper density. The solution chosen here is to instead assume that V has a gamma distribution with rate μ_2 and shape $1+\varepsilon$. By choosing $\varepsilon > 0$ small we get a good approximation of the exponential distribution, while ensuring that $h(r)$ is proper.

We define a cluster as a high density region containing at least one animal. This requirement alters the probability density of R relative to the population of all disks: $h^\dagger(r) \propto h(r)(1 - \exp(-2\pi\rho_2 r^2)$. For small r we have $h^\dagger(r) \propto r^2 h(r)$ which makes it clear that we are selecting against small disks. Further, since $h^\dagger(r) \propto r^2 h(r) \propto r h^*(r)$ we see that $h^\dagger(r)$ is a proper density even in the situation where V is exponential. Figure 3a) shows $h^\dagger(r)$ when $\mu_2 = 0.079$ km^{-1}, which is representative for the minke whale data.

2.3 Neyman-Scott process

The two-dimensional NS cluster process is defined as follows:

- Cluster centers constitute a spatial Poisson process with intensity with γ^{NS}.
- The cluster size M, i.e. the number animals within a cluster, is Poisson distributed with mean μ^{NS}.
- The position of points relative to the cluster center are drawn independently from an isotropic binormal distribution with standard deviation ρ^{NS}.

For a given parameter setting $(\rho_1, \rho_2, \mu_1, \mu_2)$ of the MMPP we choose the parameters of the NS process so that the processes matches each other as closely as possible. Three moment conditions are needed in order to determine γ^{NS}, μ^{NS} and ρ^{NS}. There are many ways of choosing these, and we here select three conditions with a simple interpretation:

- **Expected distance to cluster center.** In the MMPP-2 consider a (non-empty) cluster with disk radius R drawn from $h^{\dagger}(r)$. Select one cluster member at random and denote by S the distance to the center of the disk. The expectation of S is given as

$$E_{MMPP}(S) = \int_0^{\infty} f(s|R=r)h^{\dagger}(r)dr, \tag{3}$$

  where $f(s|R=r) = 2sr^{-2}$ is the conditional density of S. The corresponding expectation under the NS process is obtained by noting that $(S/\rho^{\text{NS}})^2$ has a chi-square distribution with two degrees of freedom, implying that $E(S) = \rho^{\text{NS}}\sqrt{\pi/2}$. By matching $E(S)$ for the two processes we

$$\rho^{\text{NS}} = \frac{E_{MMPP}(S)}{\sqrt{\pi/2}}. \tag{4}$$

- **Expected cluster size.** Conditionally on $R = r$ we have that M has a Poisson distribution with mean $2\pi r^2\rho_2$ under the MMPP-2, and hence

$$E_{MMPP}(M) = \int_0^{\infty} 2\pi r^2 \rho_2 h^{\dagger}(r)dr. \tag{5}$$

  Under the NS process M by definition, but because we are only considering non-empty clusters we must use a zero-truncated Poisson distribution with expectation $\mu^{\text{NS}}/[1 - \exp(-\mu^{\text{NS}})$. Matching this expression with $E_{MMPP}(M)$ and solving (numerically) yields a value for μ^{NS}.

- **Overall density.** The overall animal density should match, implying

$$\gamma^{\text{NS}} = \frac{D}{\mu^{\text{NS}}}, \tag{6}$$

  where D is given by (2). This follows from the fact that the overall density under the NS process is $\gamma^{\text{NS}}\mu^{\text{NS}}$.

It is not easy to obtain closed form expressions for $E_{MMPP}(S)$ and $E_{MMPP}(M)$. Appendix A describes a simulation procedure for evaluating these quantities numerically.

Figure 3a) compares the distribution of R under the two processes for ρ^{NS} determined using the above criterion. The distribution of S is much more heavy tailed under the MMPP. Figure 3b) shows a similar comparisons for M, and it is seen that M is very overdispersed under MMPP-2.

3 Application to Northeastern minke whale surveys

Shipboard sighting surveys have been conducted in the eastern North Atlantic since 1987 with the purpose of estimating minke whale abundance. In order to obtain regions with relative homogeneous whale density the total survey area has been divided into distinct survey block (Figure 1). For the present analysis blocks with low survey effort was been merged (see caption of Figure 1), yielding a total of 18 survey blocks. Whale abundance has traditionally been calculated by survey block, and a total estimate is obtained by adding together block specific estimates (Skaug et al. 2004). The total survey area is too large to be covered within a single year. Instead a sequence of yearly surveys with partial coverage (approximately 1/6) has been used. In the present analysis we use data from the period 2002-2007. The multiyear aspect is not very important for the present analysis, but explains why some survey block (NC1 and NOS) have been covered twice.

Survey blocks, tracklines and spatial location of sightings are shown in Figure 1. Two survey vessels, both with two independent observer platforms (A and B), participated in the survey each year. Both platforms had two individual observers, but we shall treat these as a single entity (the 'platform'). There were totally 354 periods of continuous survey effort (transect legs). Breaks in survey effort of duration less than 30 minutes were ignored, i.e. the two adjacent legs were joined under the assumption that the survey vessel had not moved during the break. Table 1 summarizes transect lengths and the number of observations by survey block. Details about the data collection protocol may be found in Skaug et al. (2004).

Estimates of the effective strip half-width (w) as a function of covariates are taken from Bøthun et al. (2009). Because the covariates vary along the transect line, with values being recorded every hour, w is a piecewise constant function of time. We recall that encounter rate is related to whale density through $\lambda_i = 2w\rho_i$ $(i = 1, 2)$.

Data are analysed separately by platform, and jointly (A+B) by multiplying the likelihoods of the two datasets. The latter may be viewed as pseudo-likelihood method, because the two data sets are not independent, since the platforms have been exposed to the same animals. A full likelihood approach would require duplicate identification, i.e. to determine which animals were detected by both platforms, and then fitting the MMPP to the merged dataset. However, duplicate identification is not an error free process (Bøthun et al. 2009), and by using the pseudo-likelihood approach we avoid problems arising from erroneous duplicate identification. A disadvantage of the pseudo-likelihood is that empirical Fisher information may not be used to calculate parameter uncertainty.

3.1 Results

Table 2 shows parameter estimates when a separate MMPP is fit to each survey block. There is a large variation in parameter values across survey blocks, as measured by the ratio of the largest to the smallest value. This variation is partly due to real biological variation, but is also caused by statistical uncertainty, as many survey blocks have few observations (Table 1). Generally, the μ parameters vary more than the ρ parameters, and variation is less for the joint dataset (A+B), where statistical uncertainty is reduced.

Table 4 performs model selection via the AIC criterion to see if both $\mu = (\mu_1, \mu_2)$ and $\rho = (\rho_1, \rho_2)$ need to be block specific. The AIC criterion weights the likelihood value against

the number of parameters in the model. Since there are 18 survey blocks, there are 17 free parameters associated with each of μ_1 and μ_2, and hence 34 free parameters in total for a block effect in μ. Similarly, there are 34 parameters associated with a block effect in ρ. The table also compares AIC values with and without variation in w along the transect line. The number of parameters used in AIC for the two options regarding w is the same, because w has the status as an 'offset' in the model. It is seen that the lowest AIC values (and hence the best fit) is obtained for the model where μ is constant across survey block. This holds for both platform A and B.

For platform A the model with a varying w has a better fit than the one with a constant w. This is to be expected, as variation in strip width must affect the animal encounter rate. Somewhat surprisingly, the same is not the case for platform B, where the model with constant w has the best fit, except for the model where both of μ and ρ are constant across survey block.

Table 3 shows parameter estimates for the NS process obtained from inserting values from Table 2 into (4)–(6).

4 Discussion

A method for fitting a two dimensional Neyman-Scott (NS) process to line transect data has been suggested. The approach involves two intermediate models (one- and two dimentional MMPPs), and may thus seem ad-hoc compared to the full likelihood approach of Waagepetersen & Schweder (2006). The advantage of our method is computational simplicity. It takes less one than 2 minutes on a standard work station to fit all the models in Table 4. The availability of a likelihood function provides a recipe for how to estimate covariate effects and account for variation in effective strip half-width under the MMPP. In addition, end-effect's, i.e. how to handle right-truncation resulting when the effort is shut down due to bad weather conditions, are easily handled under the MMPP (Skaug 2006).

A way of inferring two-dimensional cluster structure from line transect data have been introduced. The main mathematical contribution is the formulae (4)-(6). It may seem that the road taken is unnecessarily long, going via the MMPP and the 'two dimensional MMPP'. However, the fact that a likelihood function is available under the MMPP makes the fitting of the model is much simpler than the L-function based approach in Aldrin et al. (2002). Accounting for covariates, such as a survey block effect, follows a standard recipe under a likelihood framework.

5 Acknowledgments

I am grateful to Gjermund Bøthun, Siri Hartvedt and Nils Øien for help and useful discussions discussions.

References

Aldrin, M., Schweder, T. & Holden, M. (2002), Spatial distribution of whales based on the Norwegian minke whale survey 1996-2001, Paper SC/54/RMP4 presented to the Scientific

Committee of the International Whaling Commission.

Bøthun, G., Øien, N. & Skaug, H. (2009), Abundance of NA minke whales, Paper SC/57/SD2.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L. & Thomas, L. (2001), *Introduction to Distance Sampling; Estimating abundance of biological populations*, Oxford: Oxford University Press.

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

Fournier, D. (2001), An introduction to AD Model Builder Version 6.0.2 for use in nonlinear modeling and statistics, Available from http://otter-rsch.com/admodel.htm.

Hagen, G. & Schweder, T. (1994), Point Clustering of Minke Whales in the Northeastern Atlantic, *in* A. S. Blix, L. Walløe & Ø. Ulltang, eds, 'Whales, Seals, Fish and Man', Vol. 4 of *Developments in Marine Biology*, Elsevier, pp. 27–34.

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

Schweder, T. (1977), Point process models for line transect experiments, *in* J. Barra, ed., 'Recent developments in statistics', North Holland, pp. 221–242.

Schweder, T., Skaug, H. J., Langaas, M. & Dimakos, X. (1999), 'Simulated likelihood methods for complex double-platform line transect surveys', *Biometrics* **55**, 678–687.

Skaug, H. J. (2006), 'Markov modulated Poisson processes for clustered line transect data', *Environmental and Ecological Statistics* **13**, 199–211.

Skaug, H. J., Øien, N., Bøthun, G. & Schweder, T. (2004), 'Current abundance of minke whales (Balaenoptera acutorostrata) in the Northeast Atlantic; variability in time and space', *Canadian Journal of Fisheries and Aquatic Sciences* **61**, 870–886.

Waagepetersen, R. & Schweder, T. (2006), 'Likelihood-based analysis of clustered line transect data', *Journal of Agricultural, Biological, and Environmental Statistics* **11**(3), 264–279.

A Sampling algorithm

The following algorithm is used to sample evaluate the expectations (3) and (3). Each step below is repeated $N = 10,000$ times. This is not made explicit in our notation except when necessary.

1. Draw $V \sim \text{gamma}(\text{shape} = 1.01, \text{rate} = \mu_2)$.
2. Draw $U \sim \text{uniform}(0,1)$ and calculate $R^* = V/\left(2\sqrt{1-U^2}\right)$, so that $R^* \sim h^*(r)$.
3. Sample R from $R_1^*, \ldots, R_N^*$ with probability proportional to $1/R_i^*$, so that $R \sim h(r)$.

4. Given R, draw M from a Poisson distribution with mean $2\pi R^2 \rho_2$.

5. If $M \doteq 0$ remove the corresponding R from the sample. The number of 'survivors' $R^\dagger$ satisfies $R^\dagger \sim h^\dagger(r)$. This step reduces the sample size.

6. Given $R^\dagger$, define $S = R^\dagger \sqrt{U}$, where $U \sim \text{uniform}(0, 1)$ (not the same variable as in step 2).

When evaluating (3) we take $E_{MMPP}(S)$ to be a trimmed mean (5% on both sides) of the S values sampled in step 6 above. Similarly, when evaluating (4) we take $E_{MMPP}(M)$ to be a trimmed mean of the non-zero M's in step 4.

			A			B			A+B		
		L	n	D	τ	n	D	τ	n	D	τ
BAW	03	847	14	0.037	3.4	11	0.035	3.4	25	0.036	3.4
BAE	07	2055	31	0.032	5.1	25	0.035	12.3	56	0.03	4.4
BJ	03	1235	45	0.061	5.4	34	0.067	11.8	79	0.065	7.6
FI	02	1992	26	0.026	4.8	29	0.036	5.8	55	0.031	5.8
NOS	06	2045	29	0.029	3.7	24	0.027	8.6	53	0.027	5.2
KO	07	926	21	0.051	2.3	18	0.055	2.1	39	0.053	2.1
GA	07	1331	36	0.062	8.7	28	0.057	7.7	64	0.06	8.2
JMC	05	621	35	0.129	6.5	32	0.137	1.5	67	0.134	3.5
LOC	06	1378	26	0.027	5.7	21	0.027	8.5	47	0.027	7.6
NC1	04	1290	8	0.016	2	5	0.013	1.3	13	0.014	1.8
NC1	06	842	13	0.033	1.5	6	0.018	1.4	19	0.026	1.5
NC2	04	357	10	0.032	5.2	6	0.028	4.6	16	0.031	4.8
NON	03	756	12	0.038	1.5	10	0.038	2.1	22	0.038	1.7
NOS	02	4273	71	0.04	4.2	60	0.038	5.4	131	0.039	4.4
NS	04	2115	9	0.007	1.8	7	0.007	1.3	16	0.007	1.6
NVN	05	1705	26	0.025	11.9	21	0.031	5.6	47	0.027	7.2
NVS	05	1788	6	0.006	1.9	7	0.01	2	13	0.008	2.2
SV	03	1614	31	0.045	10.9	32	0.048	7.3	63	0.046	9.1
Total		27170	449			376			825		

Table 1: Estimated whale density D (km^{-2}) under the MMPP by survey block and platform. Also shown is transect length L (km), and the number of observations n. Platform A+B means pooling data from A and B as explained in the main text.

Tables and figures

		A				B				A+B			
		ρ_1	ρ_2	μ_1	μ_2	ρ_1	ρ_2	μ_1	μ_2	ρ_1	ρ_2	μ_1	μ_2
BAW	03	1.11	43	1.1	15.92	1.37	53.3	0.7	16.11	1.24	47.2	0.9	16.35
BAE	07	0.12	12.1	1.1	3.12	1.99	158	0.1	12.11	1.69	66.9	0.4	18.78
BJ	03	1.21	24.5	1.8	6.67	4.05	666	0.2	57.95	3.84	160	0.4	25.52
FI	02	1.77	76.2	0.2	20.77	2.11	55.8	0.3	9.81	1.97	67.8	0.2	13.25
NOS	06	0.21	28	0.9	8.55	1.24	51	0.1	4.68	0.94	36.7	0.3	6.56
KO	07	3.6	105	0.9	62.75	0.96	42.4	3.4	27.28	2.93	76.2	1.6	48.54
GA	07	1.92	41.4	0.7	6	2.25	35.8	0.6	4.73	2.03	38	0.7	5.31
JMC	05	5.26	150	1.5	27.08	3.12	368	9.8	326.52	4.38	142	3.5	50.13
LOC	06	1.73	85	0.3	21.41	1.66	180	0.1	22.61	1.81	131	0.1	21.17
NC1	04	0.66	40.7	0.7	30.19	1.12	21.1	0.3	46.8	0.87	39.4	0.5	32.87
NC1	06	1.9	101	2.3	159.73	1.64	23.1	0.3	44.63	1.82	102	1.3	173.35
NC2	04	0.37	28.4	1.1	9.79	0.16	22.1	0.9	6.74	0.2	25.3	1.1	8.49
NON	03	3.64	16.9	0.4	30.39	0.58	20	2.5	12.48	3.46	22.4	0.5	24.93
NOS	02	1.02	32.7	0.9	8.51	1.91	42.4	0.3	6.74	1.38	35.9	0.6	8.09
NS	04	0.45	32	0.4	45.1	0.62	17.7	0.3	57.42	0.52	28.8	0.4	50.28
NVN	05	0.58	15.9	0.2	1.29	1.19	32	0.3	5.18	0.7	18.9	0.3	2.46
NVS	05	0.12	9.33	0.6	11.12	0.46	48.3	0.5	44.3	0.41	49.5	0.3	40.87
SV	03	1.8	43.4	0.3	4.39	2.26	49.3	0.4	6.48	2	45.7	0.3	5.17
max/min		43.8	16.1	11.5	123.8	25.3	37.6	98.0	69.8	21.9	8.5	35.0	70.5

Table 2: Parameter estimates (times a factor 100 for convenience) for the MMPP by survey block and observer platform. The unit of ρ is km^{-2} and that of μ is km^{-1}.

		A			B			A+B		
		γ^{NS}	μ^{NS}	ρ^{NS}	γ^{NS}	μ^{NS}	ρ^{NS}	γ^{NS}	μ^{NS}	ρ^{NS}
BAW	03	0.0028	13.3	0.9	0.0023	15.2	0.9	0.0026	13.9	0.9
BAE	07	0.00063	50.7	3	0.00076	46	0.8	0.0021	14.5	0.7
BJ	03	0.0021	29.2	1.7	0.0045	14.9	0.2	0.0038	16.9	0.5
FI	02	0.0019	13.5	0.7	0.0012	30	1.1	0.0014	22.4	0.9
NOS	06	0.0013	22.8	1.4	0.00034	79.7	1.8	0.00069	39.1	1.6
KO	07	0.011	4.7	0.4	0.008	6.9	0.7	0.01	5.2	0.4
GA	07	0.0013	48.9	1.6	0.00092	62.2	1.9	0.0011	54.2	1.7
JMC	05	0.0086	15	0.5	0.086	1.6	0.1	0.019	7	0.4
LOC	06	0.0019	14.4	0.7	0.0013	21.4	0.5	0.0014	19	0.6
NC1	04	0.0026	6.1	0.6	0.0048	2.7	0.6	0.0025	5.5	0.6
NC1	06	0.019	1.7	0.2	0.0062	2.9	0.6	0.017	1.5	0.2
NC2	04	0.0016	19.4	1.3	0.001	26.9	1.7	0.0015	21.2	1.4
NON	03	0.01	3.8	0.8	0.0034	11.2	1.2	0.007	5.4	0.8
NOS	02	0.0016	25.4	1.4	0.00091	41.7	1.5	0.0014	28.8	1.4
NS	04	0.002	3.5	0.6	0.0035	2	0.6	0.0024	2.9	0.5
NVN	05	0.00011	228.9	5.1	0.00063	48.9	1.8	0.00027	99.4	3.2
NVS	05	0.00073	8.2	1.6	0.0022	4.5	0.5	0.0016	4.9	0.5
SV	03	0.00058	77.4	1.9	0.00097	49.7	1.5	0.00072	63.9	1.7

Table 3: Parameter estimates for the NS process by survey block and observer platform.

Block effect	None		ρ		μ		μ and ρ	
Platform	A	B	A	B	A	B	A	B
w const	26.1	19	2.9	0	25	25.7	40.6	36.4
w varying	17	15.3	0	6.2	12.2	24.4	34.1	45.2

Table 4: AIC values (normalized within plattform) for models with and without a survey block effect. The top row shows which parameter of the MMPP has a survey block effect. Data for plattform A and B are treated separately, and results are given both with and without variation in effective strip width (w).

Figure 1: Survey block definitions (blue lines), transect lines (black lines) and minke whale sightings (black diamonds) made by platform A in the survey period 2002-2007. The following blocks have been merged, i.e. treated as a single survey block: {SV,VSN,VSS}→SV, {FI,FI1,FI2}→FI, {BAW,BA1}→BAW.

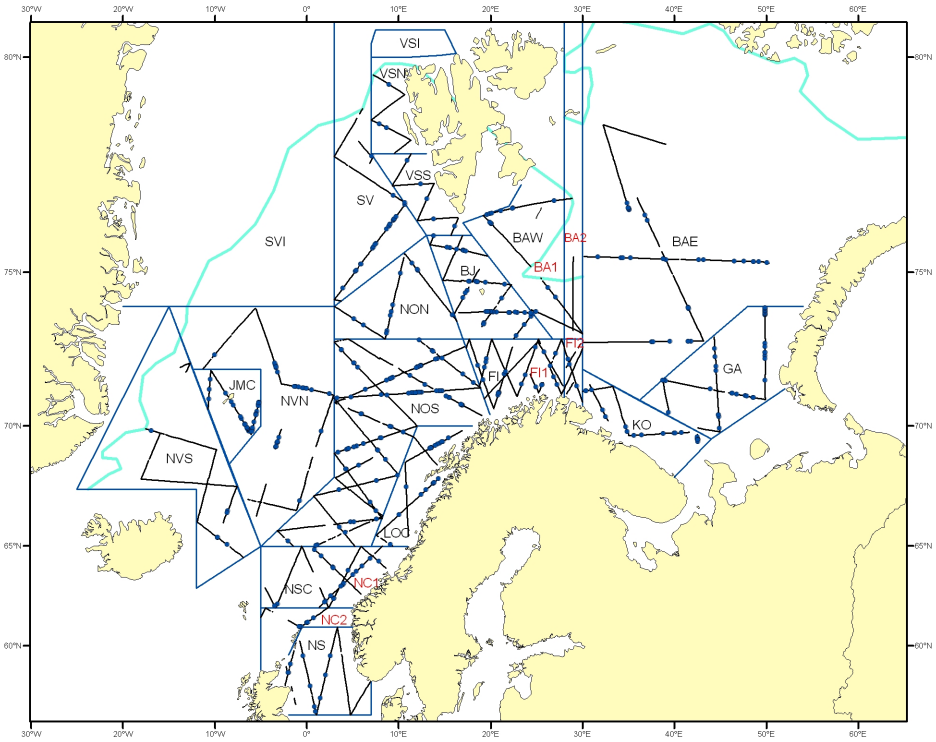
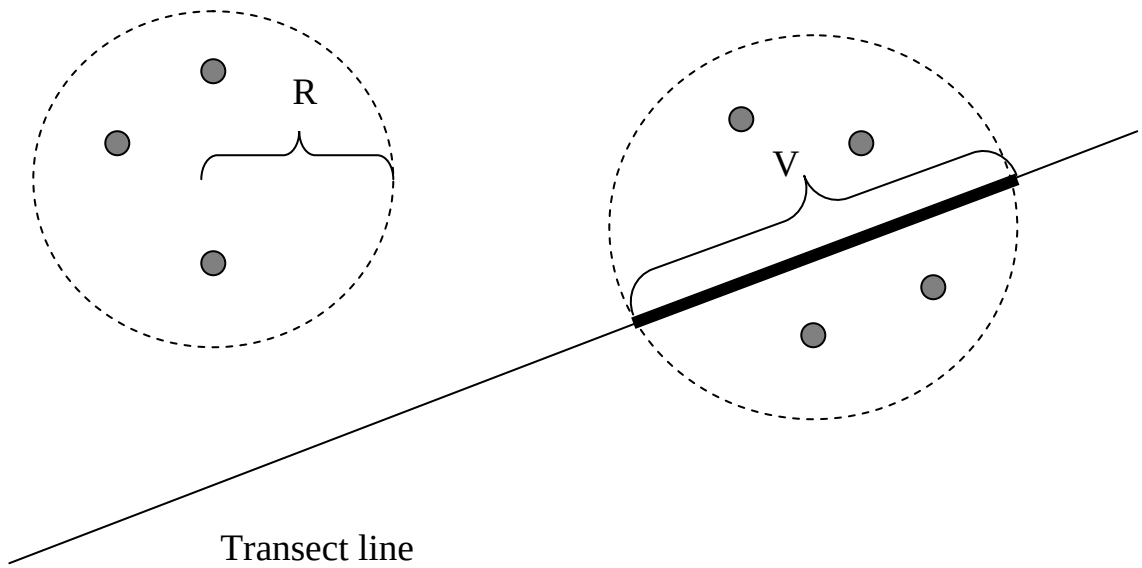


Figure 2: Two dimentional MMPP with disk-shaped high density areas (dashed) and points uniformly distributed inside these. The transect line (solid line) intersects one of the clusters, and the part of the transect line inside the disk (thick solid line) becomes a period with high encounter rate for the one dimentional MMPP. The cluster radius R and length of high density period V are also indicated.



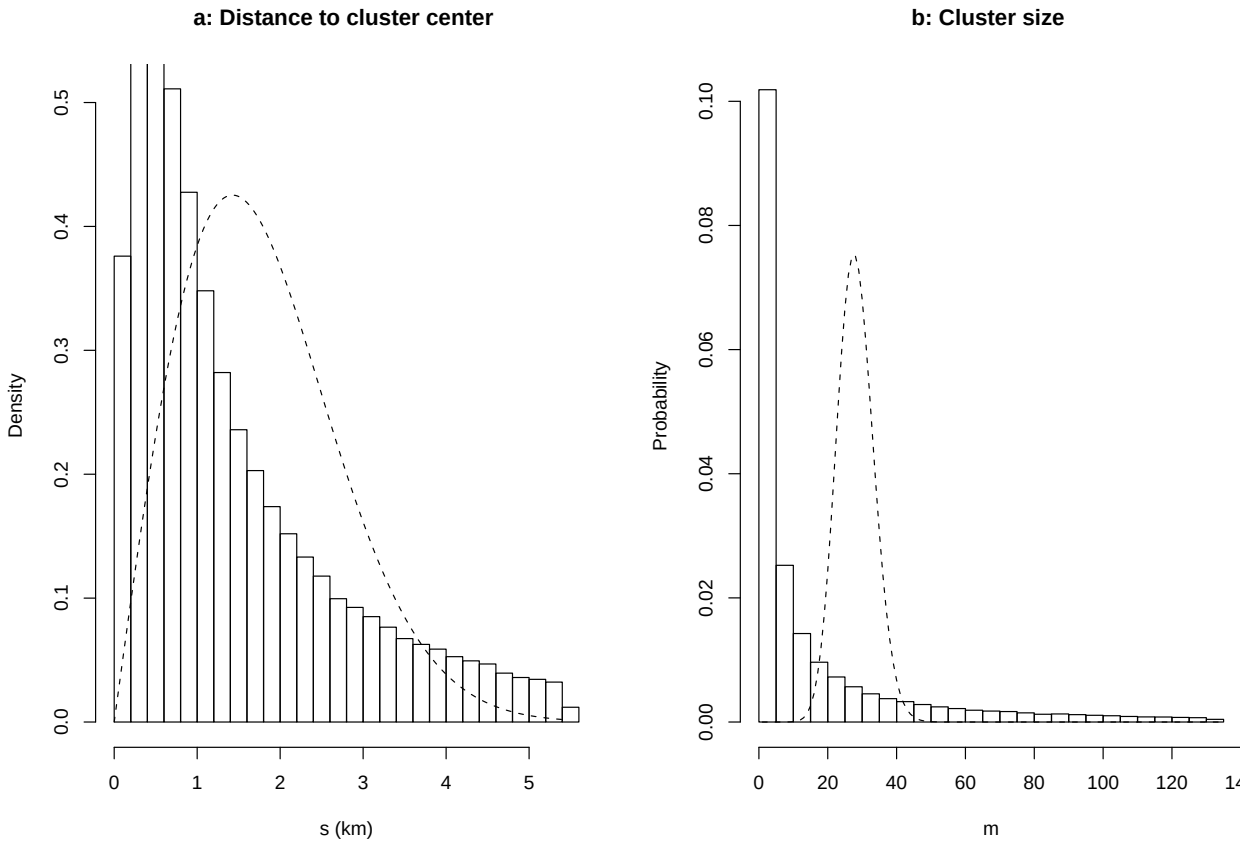


Figure 3: Comparison of MMPP-2 and NS process for minke whale data, with the NS process being matched to the MMPP-2 using the moment criteria from section 2.3. a) Density of S (distance to cluster center for randomly selected cluster point), b) Probability distribution of M (number of points in non-empty clusters). The parameters values used in the figure is $\mu_2 = 0.079$ and $\rho_2 = 0.033$ which corresponds to the MMPP fit to platform A data without any survey block effect.