

# Spatial distribution of northeastern Atlantic minke whales 1996-2001

**Magne Aldrin**

Norwegian Computing Center,  
P.O. Box 114 Blindern, N-0314 Oslo, Norway  
email: magne.aldrin@nr.no

**Marit Holden**

Norwegian Computing Center (as above)  
email: marit.holden@nr.no

**Tore Schweder**

Norwegian Computing Center (as above) and  
Department of Economics, University of Oslo,  
P.O. Box 1095 Blindern, N-0317 Oslo, Norway  
email: tore.schweder@econ.uio.no

February 6. 2002

## **Abstract**

Minke whales in the northeastern Atlantic are observed by line transect surveys. Based on these data, we estimate separate Neyman-Scott cluster processes for the spatial distribution of the whales within several smaller regions of the survey area. Parameter estimates are found by fitting a K-function based on simulated data to its empirical counterpart based on observed data. The method aims to account for important details of the observational process, for instance time varying detection probabilities.

## **1 Introduction**

To obtain abundance estimates of minke whales in the northeastern Atlantic, yearly surveys have been conducted in the period 1996-2001 (Skaug et. al. 2002). As a part of the analysis, one needs to model the spatial distribution of the whales.

In the present paper, we assume that the whales are distributed according to a Neyman-Scott process (e.g. Cressie 1993, p. 662), and estimate the parameters from the surveys in the years 1996-2001. The surveyed area of northeastern

Atlantic is divided into several blocks, and estimates are given per surveyed block per year. Furthermore, blocks are aggregated into so called small areas, and estimates are also given for each of these small areas, as well as for the total area made up by all blocks.

The present paper updates the results from the 1995 survey presented in Schweder et. al. (1997) and Koppervik (1996), based on new data and another estimation methodology. In the former analysis of the data from the 1995 survey, the parameters in the Neyman-Scott model were estimated by fitting a theoretical K-function (Ripley, 1977) to its empirical counterpart. However, these methods do not account for: i) that the surveys are carried out along several disjoint transects within a block; ii) that the whales are observed from two platforms and iii) that the probability of detecting whales varies over time according to weather conditions et cetera. In the present paper we use a modified K-function method, based on Monte Carlo simulations, that takes care of these aspects.

## **2 Data**

The survey area is divided into 19 blocks, which again are aggregated into 5 small areas, see Table 1. Most blocks are surveyed only one year. Survey data has been collected in the following way: A vessel is moving along a transect line. The vessel has two platforms, A and B, each manned with two observers, who observe whales and record their positions relative to the transect line. Within a block, there will usually be several transects, and we will assume that these are independent, i.e. whales observed along one transect are not in the same cluster as whales observed along another transect. Table 1 gives, for each block and year, an overview of number of transect pieces, total transect length and how many whales that are observed from each platform. The surveys data are described in more detail in Skaug et. al. (2002).

**Table 1** Summary of survey data. The unit of transect length is kilometer.

small area	block	year	no. of transects	total transect length	no. of whales pl. A	no. of whales pl. B
EB	BAE	1996	3	313	1	2
EB	BAE	2000	36	3293	34	42
EB	GA	2001	8	1022	36	25
EB	KO	2000	11	464	1	2
EB	KO	2001	11	859	15	12
EB	FI	1996	11	1659	63	58
EB	NOS	1996	53	4166	77	48
EB	NOS	2001	18	875	11	12
EC	LOC	1996	11	930	2	0
EC	LOC	2000	8	1081	5	5
ES	VSI	1999	6	428	14	11
ES	VSN	1999	6	482	25	18
ES	VSS	1999	9	688	31	30
ES	SVI	1999	20	1092	4	6
ES	SV	1999	11	996	33	27
ES	NON	1999	9	945	13	12
ES	BJ	1999	7	898	21	21
ES	BAW	1999	15	870	23	7
CM	JMC	1997	14	614	20	18
CM	NVN	1997	36	1935	30	30
CM	NVN	2001	4	66	1	0
CM	NVS	1997	28	1747	42	44
CM	NVS	2001	8	161	1	0
EN	NSC	1998	22	2535	27	19
EN	NS	1998	43	3958	89	63

Except for block NOS 2001 (where there is no need for a model), we will in later sections estimate spatial models for all blocks and years with at least 20 observed whales in total for the two platforms. Furthermore, we will estimate models for the small areas EB, ES, CM and EN, using all available data from all blocks within the area in question. Finally, we will estimate a model for the total area made up by all blocks, using all available data.

The position of each observed whale is given by the projection onto the transect line ( $y$ -coordinate) and by the distance  $x$  from the whale to the transect line. Appendix A gives plots with locations of the detected whales relative to their current transect lines. A few times the position of a whale may be in front of the end of the current transect, i.e. the  $y$ -coordinate is larger than the length of the transect. These observations are not included in the present analysis (or in Table 1 or in the appendix), because the detection probability for these are different (lower) than for the others.

Whales that are in the neighbourhood of the transect lines may be seen from both platforms, from one of the platforms or not at all. Firstly, whales may not be seen because they are too far away or the weather conditions are bad. Secondly, they may not be seen because they are diving, which certainly introduces some degree of dependency in detecting whales from the two platforms.

The detection functions for platform A or B are given by the probability to detect a present whale from platform A or B as a function of  $x$ , the perpendicular distance from the transect line. In addition to depend on platform, the detection functions vary over time, depending on certain covariates related to weather and observers. In the main abundance estimation procedure (Skaug et. al. 2002), the detection functions for the two platforms are assumed to be of a certain parametric form, and are estimated by the survey data. Taking into account the dependency between the two platforms, one can also define a total detection function for detecting a whale from at least one of the platforms. Two important parameters of the detection function are

- $g_0$  = the probability of detecting a whale located at the transect line, and
- $\omega$  = effective strip half-width = 0.5 times the area under the detection function.

In the present analysis,  $g_0$  and  $\omega$  may vary within transects and between transects, depending on 12 different combinations of covariates, see Table 2. These values are estimated from the total abundance estimation procedure described in Skaug et. al. (2002), and in the present analysis we will assume that these are known.

**Table 2** Summary of detection curves. The half strip widths are in meters. The  $P_D$  column is defined in Section 3.

condition number	$g_{0A}$	$\omega_A$	$g_{0B}$	$\omega_B$	$g_{0(A \cup B)}$	$P_D$
1	0.4561208	364.752	0.4169210	300.863	0.6391470	0.8130422
2	0.4561208	364.752	0.3441731	204.057	0.6030068	0.7957160
3	0.3761852	243.491	0.4169210	300.863	0.5937961	0.7869120
4	0.3761852	243.491	0.3441731	204.057	0.5501243	0.7605580
5	0.5626222	591.930	0.5176490	485.109	0.7461506	0.8716638
6	0.5626222	591.930	0.4287049	319.118	0.7097282	0.8565335
7	0.4688125	387.458	0.5176490	485.109	0.6999598	0.8470467
8	0.4688125	387.458	0.4287049	319.118	0.6526158	0.8206652
9	0.3397373	198.953	0.3118700	168.769	0.5072085	0.7337587
10	0.3397373	198.953	0.2638855	123.828	0.4786042	0.7171073
11	0.2844261	142.034	0.3118700	168.769	0.4714673	0.7106050
12	0.2844261	142.034	0.2638855	123.828	0.4393564	0.6888696

### 3 Statistical models for the distribution and detection of whales

#### 3.1 Model for spatial distribution

We will assume that within each block and year, the whales are distributed according to the following special version of a Neyman-Scott process (see for instance Cressie, 1993, p. 662, for a more general definition):

- Cluster centres are Poisson distributed with intensity  $\lambda$  (per 1000 square kilometer).
- The number of whales within each cluster is Poisson distributed with mean  $\mu$ , and is independent between clusters.
- The positions of the whales relative to their cluster centres are independent and have an isotropic bivariate normal distribution with variance  $\rho^2$  in both x- and y-direction ( $\rho$  has unit kilometer).

The overall intensity of whales is  $\mu\lambda$  (per square kilometer).

At the cluster center, the intensity of whales *from that cluster* will be  $\mu/(2\pi\rho^2)$ .

Dividing this intensity by the overall intensity gives the ratio

$$r = [\mu/(2\pi\rho^2)]/(\mu\lambda) = 1/(2\pi\lambda\rho^2) \quad , \quad (1)$$

which can be interpreted as a measure of the degree of clustering.

For a given value of the overall intensity  $\mu\lambda$ , the model will be equivalent to a pure spatial Poisson model with intensity  $\mu\lambda$  when  $\rho \rightarrow \infty$  or when  $\mu \rightarrow 0$ .

### 3.2 Model for the detection process

We will not use the same detection functions as assumed in the main abundance procedure (Skaug et. al. 2002). Instead we will use a simpler function form, but with parameters  $g_{0A}$ ,  $\omega_A$ ,  $g_{0B}$ ,  $\omega_B$  and  $g_{0(A \cup B)}$  as estimated by Skaug et. al. (2002). For platform A, we assume the detection function

$$g_A(x) = g_{0A} \cdot \exp(-x^2/2\sigma_A^2) , \quad (2)$$

where  $g_{0A} = g_A(0)$  is the detection probability at  $x=0$ . The effective strip width is

$$\int_{-\infty}^{\infty} g_A(x) = \sqrt{2\pi}\sigma_A g_{0A} = 2\omega_A , \quad (3)$$

where  $\omega_A$  is the effective strip half-width. Thus, for given values of  $g_{0A}$  and  $\omega_A$ ,  $\sigma_A$  can be calculated by  $\sigma_A = 2\omega_A/\sqrt{2\pi}g_{0A}$ . The detection function for platform B is defined similarly. The detection function (2) was used by Cowling (1998). It fits with the spatial normality of the clusters, and allow closed form integrals.

As mentioned in Section 2, there is a dependency between the two platforms. Independency would require that  $g_{0A} \cdot g_{0B} = g_{0A} + g_{0B} - g_{0(A \cup B)}$ , which is not fulfilled in Table 2. We will model the dependency such that the values of  $g_{0A}$ ,  $g_{0B}$  and  $g_{0(A \cup B)}$  from Table 2 are preserved.

Assume that each whale is *detectable* with a probability  $p_D$ , independent of the distance  $x$  from the transect line. If it is detectable, it can be detected independently from platform A and B, with probabilities depending on  $x$ . On the other hand, if it is *not detectable*, the probability for detecting the whale is 0 for both platforms. One interpretation of this concept is that a whale is detectable if it has at least one surfacing when the vessel passes.

Let for the moment  $A$  and  $B$  denote the events that a whale is detected by platform A or B, respectively, and let  $D$  denote the event that a whale is detectable. The probability that the whale is detected from both platforms is

$$\begin{aligned} P(A \cap B) &= P(A \cap B|D) \cdot P(D) + P(A \cap B|D^C) \cdot P(D^C) \\ &= (P(A) \cdot P(B))/P(D) + 0 . \end{aligned} \quad (4)$$

From standard probability theory we have that

$P(A \cap B) = P(A) + P(B) - P(A \cup B)$ , which combined with (4) gives

$$P(D) = \frac{P(A) \cdot P(B)}{P(A) + P(B) - P(A \cup B)} . \quad (5)$$

This holds for all  $x$ , including  $x = 0$ , which gives

$$p_d = P(D) = ((g_{0A} \cdot g_{0B}) / (g_{0A} + g_{0B} - g_{0(A \cup B)})) . \quad (6)$$

The final model for the simultaneous detecting process is then the following: Each whale is detectable with a probability  $p_D$ . If it is detectable, it is seen by platform A and platform B independently of each other, with probabilities  $g_A(x)/p_D$  and  $g_B(x)/p_D$ , respectively.

### 3.3 Expected number of observed whales

Assume that we consider a block with  $m$  disjoint pieces of transects surveyed in a specific year. Each transect  $i$  may be divided into  $m_j$  segments, with constant values of  $g_0$  and  $\omega$  within each segment. Let  $L_{ij}$  denote the length of the  $j$ th segment within the  $i$ th transect, and let  $g_{0Aij}$ ,  $\omega_{Aij}$ ,  $g_{0Bij}$  and  $\omega_{Bij}$  denote the corresponding ‘‘known’’ parameters of the detection function. Let  $n_{ij}^A$  and  $n_{ij}^B$  be the number of whales detected within that segment, from platform A and B, respectively. The expected number of detected whales from for instance platform A will be

$$E(n_{ij}^A) = \sqrt{2\pi}\mu\lambda\sigma_{Aij}g_{0Aij}L_{ij} . \quad (7)$$

This is obtained by simply integrating the detection function, see for instance Aldrin, Holden and Schweder (2001). Furthermore, let  $n$  be the sum of the total number of detected whales for each platform (i.e. duplicates are counted twice). The expectation of  $n$  is then given by

$$E(n) = \sum_i^m \sum_j^{m_j} \sqrt{2\pi}\mu\lambda L_{ij} (\sigma_{Aij}g_{0Aij} + \sigma_{Bij}g_{0Bij}) . \quad (8)$$

## 4 Estimation method

We will base our estimation method only on the  $y$ -coordinates of the detected whales, ignoring the  $x$ -coordinates. The rationale for this is that the distance to the transect line is usually considerably shorter than the transect length, and that

most of the information about the spatial pattern is contained in the coordinate along the transect.

#### 4.1 The K-function

The so called K-function (Ripley, 1977) will be a central concept in our estimation procedure. The K-function of a stationary spatial point process with intensity  $\tau$  is defined as

$$K(h) = \tau^{-1} E \left( \begin{array}{l} \text{number of extra points within distance } h \\ \text{of a randomly chosen point} \end{array} \right). \quad (9)$$

Assume for the moment that we have observed whales from *one* platform along *one* transect with *constant*  $g_0$  and  $\sigma$ . The  $y$ -coordinates of the detected whales will constitute a one-dimensional stationary Neyman-Scott process. The theoretical K-function of this process is given in Aldrin, Holden and Schweder (2001), as a function of  $\lambda$  and  $\rho$  (and  $\sigma$ , which is assumed known). Thus, given  $\lambda$  and  $\rho$ , the K-function does not depend on  $\mu$ . By using (8), one may equivalently express the K-function as a function of  $\mu$  and  $\rho$  (and  $g_0$  and  $\sigma$ ).

The K-function of such a one-dimensional point process on a transect can be estimated by

$$\hat{K}(h) = 2n^{-2}L \sum_j \sum_{i < j} I(|y_i - y_j| < h) , \quad (10)$$

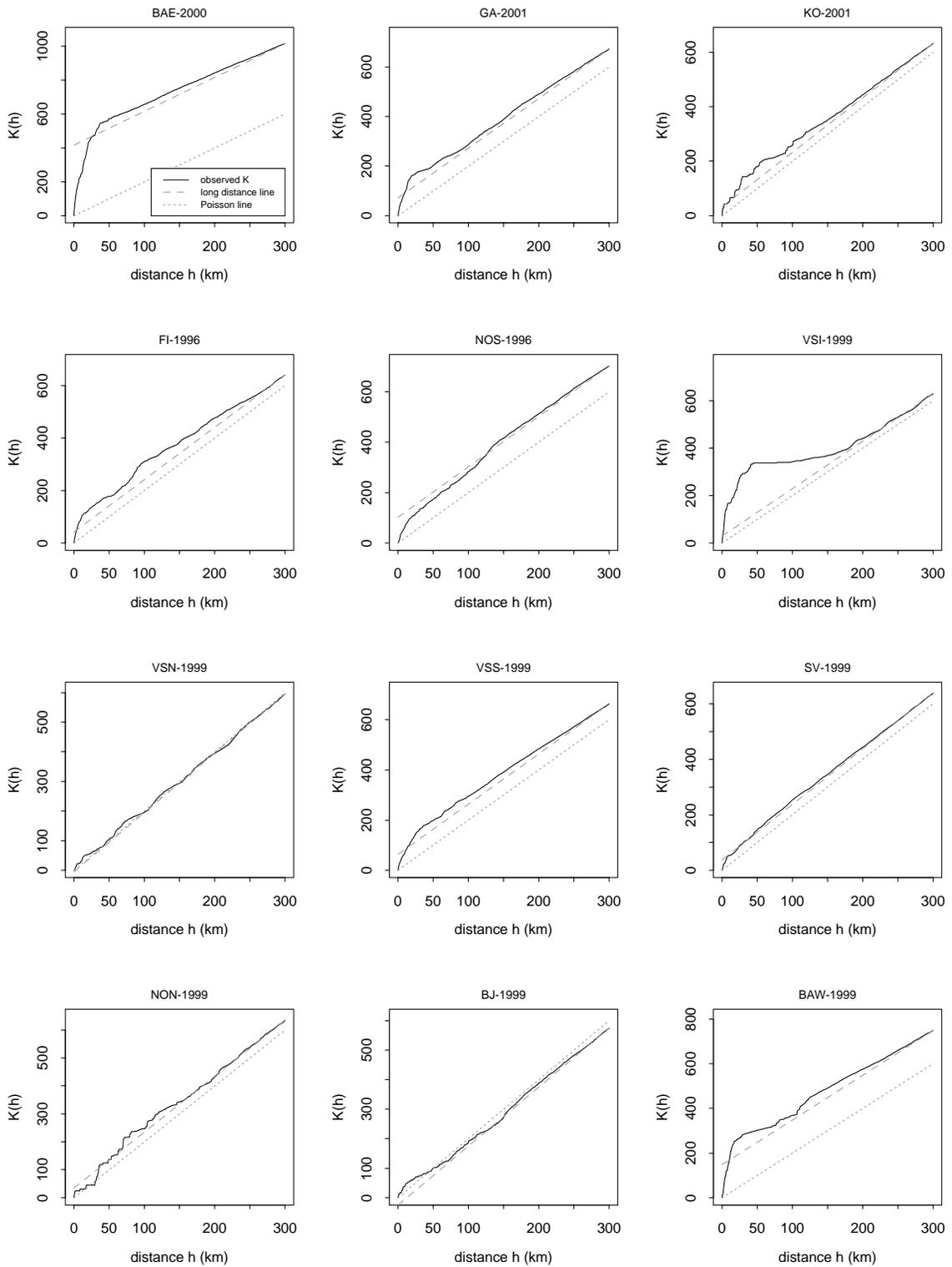
where  $L$  is the length of the transect,  $n$  is the number of detected whales, and  $y_i$  and  $y_j$  are the positions along the transect line. In principle, the parameters of the Neyman-Scott process can be estimated by minimising a measure of the distance between the theoretical and empirical K-function, as in Aldrin, Holden and Schweder (2001), Cowling (1998) and Schweder et. al. (1997). However, the theoretical K-function is defined for an infinitely long transect, whereas the estimate (10) will be influenced by end effects. Therefore, fitting the theoretical K-function to the empirical K-function may be problematic even in the situation with observations from one platform only, only one transect and with a homogeneous detection function. In our real situation we even have two platforms in parallel, several transects (many of these are short) and inhomogeneity within and between transects.

## 4.2 Estimation based on the K-function of an artificial transect

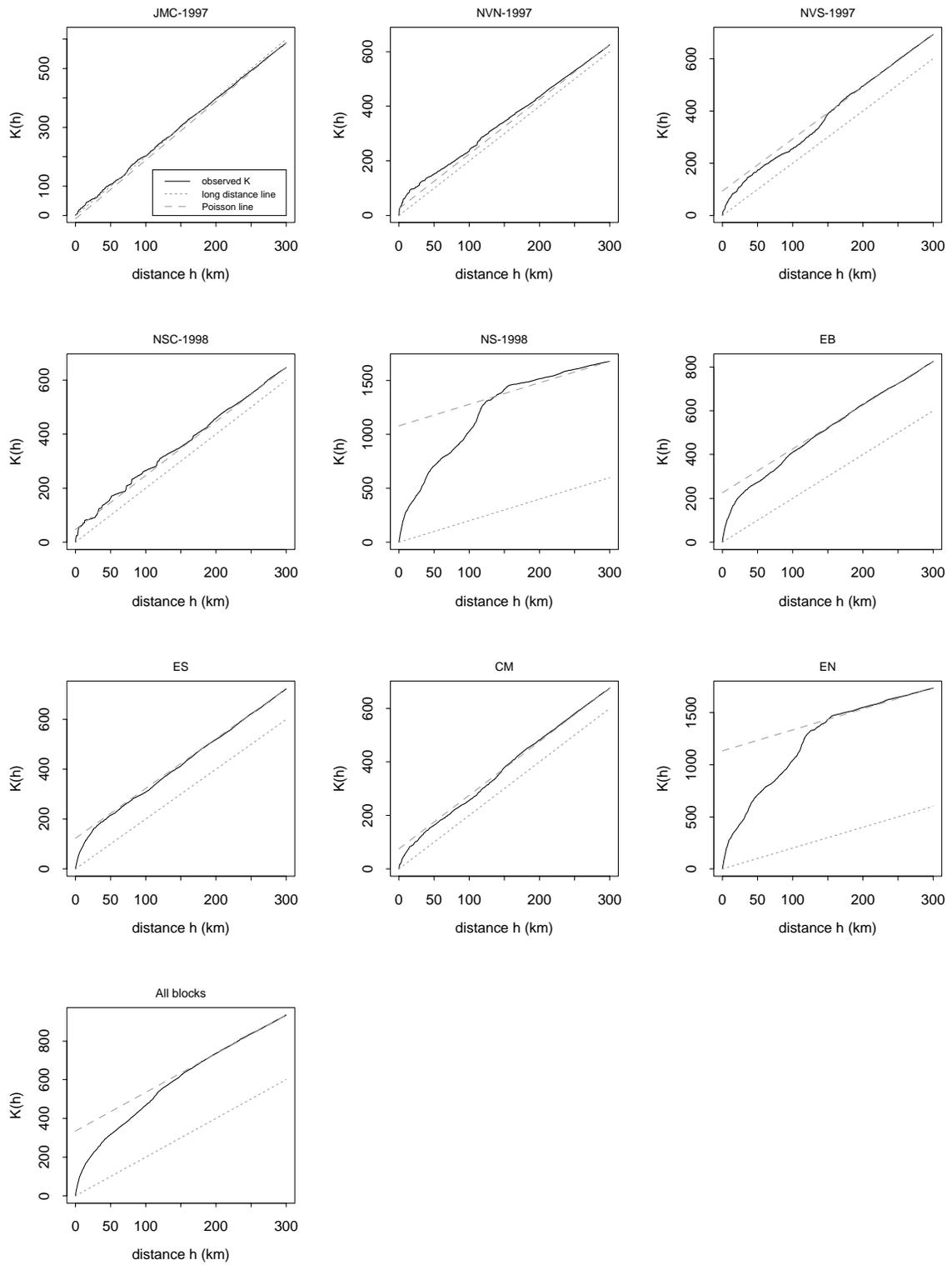
To handle this complicated situation we have constructed another estimation method based on a K-function. The main idea is to first link the observed data from all the transects together into one long artificial transect. In this way we join transects that in reality are unrelated, which at first glance may seem not to be a good idea. However, in the next step we simulate data for given parameter values, and treat the simulated data in the same way as the observed data. When comparing K-functions from observed and simulated data, we have then done the same “mistakes” on both types of data. This simulation-based estimation method is similar in spirit to the method of simulated likelihood in Schweder et. al. (1999). We will now describe the method in more detail.

Let again  $m$  denote the number of independent transects within a block within a year. We have thus  $2m$  data series for the two platforms. These data series are linked together into one long artificial transect, denoted by  $T_1^{data}$ . The order of the linking is totally random, as is the direction of each series. Certainly, this artificial transect is an arbitrary choice out of  $(4m)!$  possibilities. Therefore, we permute the data series again, giving  $T_2^{data}$ , and join the new artificial transect to the previous. We continue this procedure until the artificial super long transect includes at least 12 500 detections. Then we calculate the empirical K-function  $\hat{K}^{data}(h)$  of the final artificial transect  $T_{long}^{data}$ .

Figure 1 shows these empirical K-functions for the various blocks and years, and the four small areas (that we will estimate models for). In addition, two straight lines are shown in the plots. The dotted line is the theoretical K-function for a pure Poisson process. It goes through origo, and has a slope of 2. An empirical K-function close to this Poisson line indicates that the data belonging to it follow a pure Poisson process. The dashed line has also a slope of 2, but goes through the point where the empirical K-function ends. All theoretical (one dimensional) K-functions will reach a slope of 2 when  $h$  becomes large enough. Therefore, the range of the clusters is indicated by that  $h$  where the empirical K-function becomes a tangent to this “long distance line”. We observe that the K-functions for blocks/years VSN-1999, BJ-1999, JMC-1997 and perhaps NON-1999, indicates pure Poisson behaviour.



**Figure 1** Observed K-functions, with long distance lines and Poisson lines.



**Figure 1 cont.**

The next step is to find a theoretical counterpart to  $\hat{K}^{data}$ , that depends on  $\lambda$  and  $\rho$ . It is probably impossible to find an analytical expression, but we may use Monte Carlo simulation.

Firstly, for given values of  $\lambda$  and  $\rho$ , a corresponding value of  $\mu$  is calculated from (8), by substituting  $E(n)$  by the observed number  $n$ . Then, for the given values of  $\lambda$ ,  $\rho$  and  $\mu$ , whales detected around the  $i$ th transect line is simulated by the following procedure: First, cluster centres are simulated in a rectangle

$(-\delta, L_i + \delta) \times (-b, b)$  centred along the transect line, where the width  $2b$  is large compared to  $\rho$  and  $\sigma$ , and the extra length  $\delta$  in the ends is large compared to  $\rho$ . Then the whales within each cluster are simulated, but whales with  $y$ -coordinates less than 0 or greater than  $L_i$  are rejected. Finally, the whales are randomly detected from platform A and B according to the detection model described in Section 3.2, with values of  $g_0$  and  $\sigma$  that vary along the transect. This is repeated for  $i=1, \dots, m$ . This gives us  $2m$  simulated data series that each corresponds to an empirical counterpart in the real data. The simulated data series are linked together into one artificial transect,  $T_1^{sim}$ , in exactly the same order and direction as the first artificial transect based on the data,  $T_1^{data}$ . The whole procedure is repeated, but by independent simulations, giving  $T_2^{sim}$ , which is joined to  $T_1^{sim}$ . This is repeated until we have simulated a super long artificial transect  $T_{long}^{sim}$  of the same length as the empirical counterpart  $T_{long}^{data}$ . Finally, we calculate the empirical K-function  $\hat{K}^{sim}(h; \lambda, \rho)$  of  $T_{long}^{sim}$ .

Now, by varying  $\lambda$  and  $\rho$ , these parameters can be estimated by minimising

$$\int_0^{h_0} [\sqrt{\hat{K}^{data}(h)} - \sqrt{\hat{K}^{sim}(h; \lambda, \rho)}]^2 dh, \quad (11)$$

for a suitable value of  $h_0$ , with a corresponding estimate of  $\mu$ . By studying the various curves in Figure 1, we have chosen to use  $h_0 = 300$  kilometres, which is about 15 times as long as in Schweder et. al. (1997). A few models is re-fitted with  $h_0 = 150$  kilometres.

The integral in (11) is calculated by simply summing the integrand values for each 3/5 kilometer from 0 to 300 kilometres. Due to simulation uncertainty, the value of  $\hat{K}^{sim}(h; \lambda, \rho)$  evaluated at two pairs of  $\lambda, \rho$  very close to each other may differ. This makes it difficult to use a optimization procedure based on the derivatives, such as Newton-Raphson. Instead, we have minimised (11) by a grid search on various values of  $\lambda$  and  $\rho$ .

Earlier work by Hagen and Schweder (1995), Koppervik (1996) and Schweder et al. (1997) may look similar to our procedure described above, since they linked the real transects into one artificial transect as we do. However, their procedures differ from ours in several important aspects: Firstly, they did not use permuted repetitions. Secondly, they compared the empirical K-function to the theoretical K-function of a homogenous process with constant detection function, without taking into account the effect of joining independent transects or variability in sighting efficiency.

### 4.3 Uncertainty evaluation

We evaluate the uncertainty of the parameters by the following parametric bootstrap experiment: First, we get estimates  $\hat{\lambda}$ ,  $\hat{\mu}$  and  $\hat{\rho}$  from the observed data from a block within a year, or from all available data from a small area. Then we simulate  $S$  new data sets from the estimated model. For each simulated data set, we simulate exactly the same number of transects with the same lengths and the same varying  $g_0$  and  $\omega$  as in the original data, but in other aspects, the simulations are independent. For each simulated data set  $s$ , we get estimates  $\hat{\lambda}^s$ ,  $\hat{\mu}^s$  and  $\hat{\rho}^s$ , for  $s=1, \dots, S$ .

The distribution of  $\hat{\lambda} - \hat{\lambda}^s$  is highly variable and very asymmetric, whereas the distribution of  $\log(\hat{\lambda}) - \log(\hat{\lambda}^s)$  is more symmetric (but may have mean different from 0). We calculate the mean  $b$  and the standard deviation  $sd$  of  $\log(\hat{\lambda}) - \log(\hat{\lambda}^s)$ , and construct a 95% confidence interval for  $\log(\lambda)$  as  $(\log(\hat{\lambda}) + b - 2sd, \log(\hat{\lambda}) + b + 2sd)$ . This gives a 95% confidence interval for  $\lambda$  as  $(\hat{\lambda} \exp(b - 2sd), \hat{\lambda} \exp(b + 2sd))$ . Confidence intervals for  $\mu$ ,  $\rho$ , the overall intensity  $\mu\lambda$  and the ratio  $r = 1/(2\pi\lambda\rho^2)$  are constructed in a similar way.

All confidence intervals will be presented as relative intervals  $(l, u)$ , such that the relative limits  $l$  and  $u$  have to be multiplied by the estimate in question to give the interval in original scale. For instance, if the relative confidence interval for  $\lambda$  is  $(l, u)$ , the absolute interval will be  $(\hat{\lambda} \cdot l, \hat{\lambda} \cdot u)$ .

Due to computational limitations, we have let  $S$  be as low as 20. Therefore, our uncertainty estimates will be very rough, but they will at least be useful as an indication of the uncertainty.

## 5 Results

The parameters in the Neyman-Scott model are now estimated by the procedure described in Section 4.2, with  $h_0 = 300$  kilometres. Figure 2 shows the empirical and the fitted K-functions.

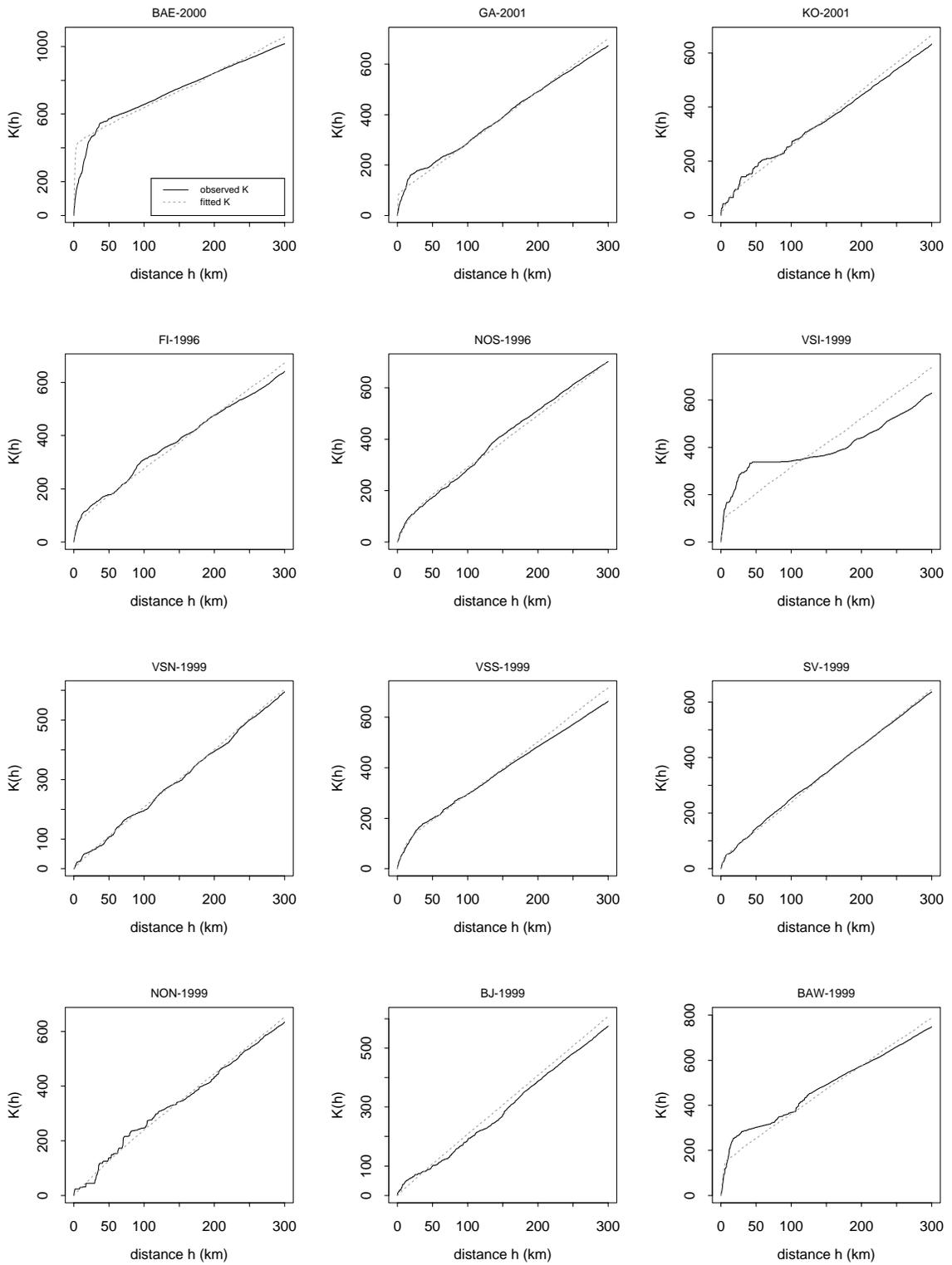
We see that the fitted K-functions for VSN-1999, NON-1999, BJ-1999 and JMC-1997 are very close to a straight line, which means that the fitted models are close to pure Poisson models. We therefore re-estimate pure Poisson models for these blocks/years.

For most of the blocks/years or small areas (e.g. KO-2001 and FI-1996), the estimated K-functions seem to fit very well. However, there are some exceptions.

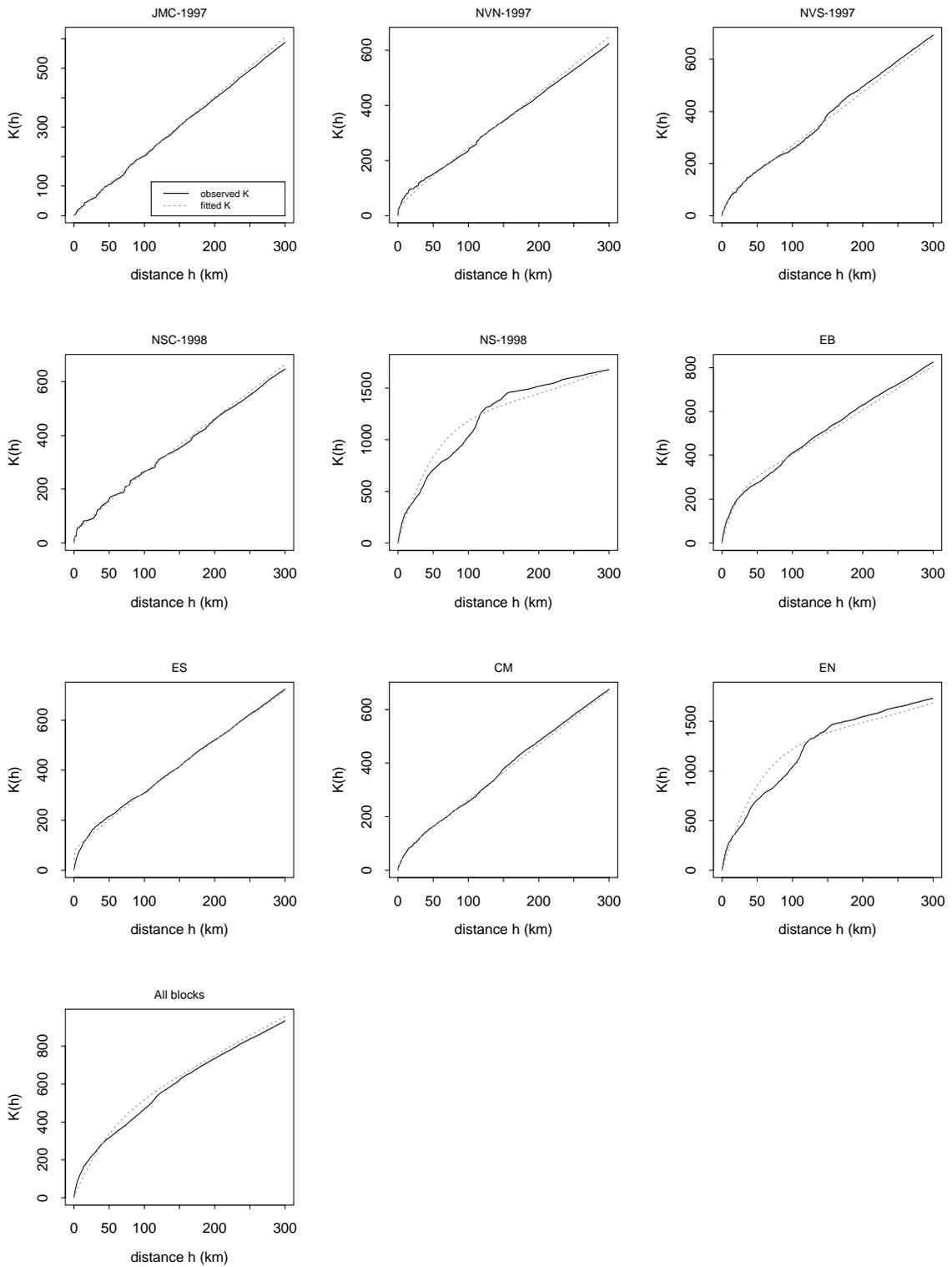
For BAE-2000, GA-2001, ES and the total area made up by all blocks, the estimated K-functions fit less well for small distances. The models for these blocks/areas are therefore re-fitted with  $h_0 = 150$  kilometres. The resulting K-functions are shown in Figure 3. ES now have a satisfactory fit for all distances, whereas GA-2001 and the total area made up by all blocks have improved fits for small distances, at the cost of slightly worse fits for large distances. Thus for these three blocks/areas, our final estimates are based on the fits with  $h_0 = 150$  kilometres. Concerning BAE-2000, the fit has improved slightly for small distances, but it has become clearly worse for large distances. For BAE-2000, we have therefore kept the original estimates based on the fit with  $h_0 = 300$  kilometres.

VSI-1999 is another block where the fit is bad. Figure A6 in Appendix A shows that the data consist of one cluster with many observed whales, and one single whale far away from the cluster. This gives the peculiar form of the empirical K-function. Even if the estimated K-function fits badly, we regard it as a reasonable compromise between a good fit for small distances versus large distances.

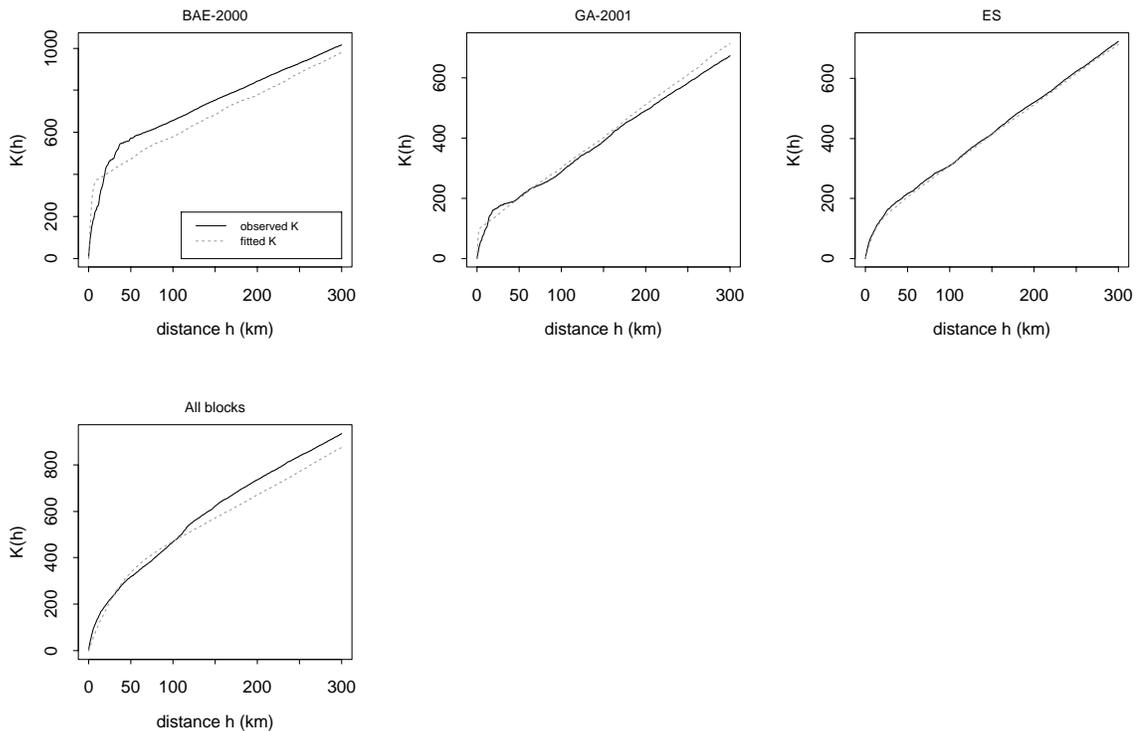
The fits for NS-1998 and EN, which includes NS-1998, seem satisfactory. However, Figure A3 in Appendix A shows that most of the observed whales in NS-1998 lie within one transect, and that these whales may further be divided into 3-4 clusters. Thus these clusters seem to be further clustered into a meta cluster. Such inhomogeneity is not covered by the Neyman-Scott process. Therefore, even if the fitted K-functions seem to be reasonably good, the Neyman-Scott model is obviously not a good one for these blocks/areas. The same argument applies for the total area made up by all blocks.



**Figure 2** Observed and fitted K-functions, with  $h_0 = 300$  kilometres.



**Figure 2** cont.



**Figure 3** Observed and re-fitted K-functions, with  $h_0 = 150$  kilometres.

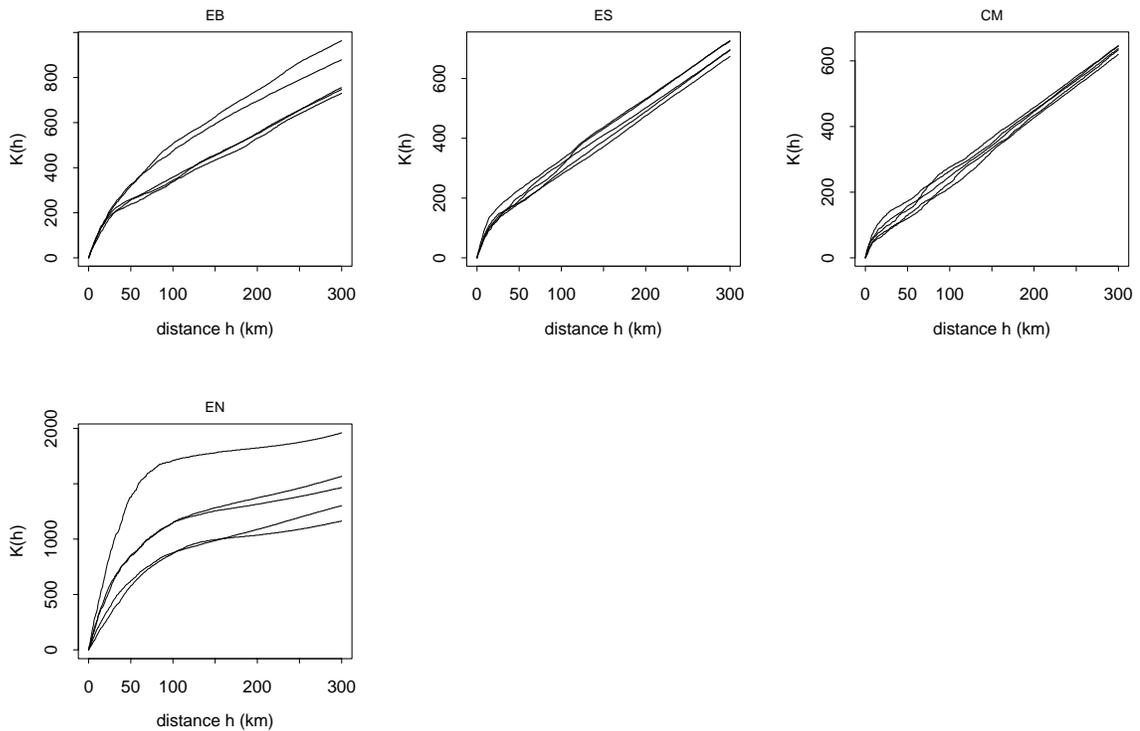
The parameter estimates, with relative 95% confidence intervals, are given in Table 3. We observe that the uncertainty is huge. Since the overall intensity  $\mu\lambda$  is estimated by substituting the expected number of whales by the observed number in (8), the estimate of  $\mu\lambda$  is unbiased, but the lower and upper confidence limits may differ by a factor between 2 and 5. For the other parameters the lower and upper limits often differ by a factor of 100. However, even if the uncertainty is high, we are clearly able to estimate models that are significantly different from pure Poisson models.

Note that confidence intervals are not given for NS-1998 and EN. The reason is that this would take too much computer time, because the computer time increases rapidly when  $\rho$  becomes as large as for these two areas.

**Table 3** Parameter estimates and relative 95% confidence intervals.

small area	block	year	$\lambda$ (per 1000 km <sup>2</sup> )	$\mu$	$\rho$ (km)	$\mu\lambda$ (per km <sup>2</sup> )	$r$
EB	BAE	2000	<b>0.615</b> (0.14,9.8)	<b>41.4</b> (0.09,7.8)	<b>1.21</b> (0.13,4.7)	<b>0.0255</b> (0.45,2.2)	<b>175.6</b> (0.20,9.5)
EB	GA	2001	<b>3.387</b> (0.14,4.2)	<b>22.9</b> (0.33,6.2)	<b>0.85</b> (0.31,4.9)	<b>0.0776</b> (0.45,2.7)	<b>65.3</b> (0.19,3.7)
EB	KO	2001	<b>1.086</b> (0.01,8.1)	<b>26.5</b> (0.16,80.2)	<b>7.03</b> (0.06,8.6)	<b>0.0288</b> (0.61,2.0)	<b>3.0</b> (0.05,818)
EB	FI	1996	<b>2.938</b> (0.06,5.9)	<b>23.3</b> (0.14,22.4)	<b>1.68</b> (0.23,9.1)	<b>0.0685</b> (0.66,1.7)	<b>19.2</b> (0.14,5.2)
EB	NOS	1996	<b>0.348</b> (0.07,6.2)	<b>97.0</b> (0.15,18.0)	<b>13.22</b> (0.29,6.3)	<b>0.0338</b> (0.72,1.6)	<b>2.6</b> (0.09,7.9)
ES	VSI	1999	<b>1.252</b> (0.02,1.6)	<b>25.9</b> (0.43,55.8)	<b>2.48</b> (0.15,20.0)	<b>0.0324</b> (0.33,2.6)	<b>20.6</b> (0.01,391)
ES	VSN	1999	x	x	x	<b>Po 0.0781</b>	x
ES	VSS	1999	<b>0.463</b> (0.05,0.8)	<b>168.9</b> (1.21,21.8)	<b>8.20</b> (0.39,9.0)	<b>0.0782</b> (0.51,2.3)	<b>5.1</b> (0.04,44.3)
ES	SV	1999	<b>4.501</b> (0.02,10.9)	<b>12.0</b> (0.09,58.7)	<b>2.13</b> (0.1,17.3)	<b>0.0540</b> (0.73,1.8)	<b>7.8</b> (0.08,7.5)
ES	NON	1999	x	x	x	<b>Po 0.0328</b>	x
ES	BJ	1999	x	x	x	<b>Po 0.0262</b>	x
ES	BAW	1999	<b>0.943</b> (0.07,4.2)	<b>26.8</b> (0.26,13.5)	<b>2.48</b> (0.31,5.0)	<b>0.0253</b> (0.44,2.2)	<b>27.4</b> (0.22,7.2)
CM	JMC	1997	x	x	x	<b>Po 0.0672</b>	x
CM	NVN	1997	<b>21.494</b> (0.07,5.2)	<b>1.6</b> (0.20,17.1)	<b>0.15</b> (0.11,8.7)	<b>0.0337</b> (0.81,1.5)	<b>312.0</b> (0.01,226)
CM	NVS	1997	<b>0.709</b> (0.05,1.7)	<b>73.4</b> (0.52,28.3)	<b>7.28</b> (0.48,8.0)	<b>0.0521</b> (0.67,2.0)	<b>4.2</b> (0.09,8.2)
EN	NSC	1998	<b>5.981</b> (0.04,8.3)	<b>3.5</b> (0.11,31.0)	<b>1.55</b> (0.09,3.5)	<b>0.0208</b> (0.65,1.7)	<b>11.1</b> (0.10,353)
EN	NS	1998	<b>0.00425</b>	<b>11004.9</b>	<b>51.13</b>	<b>0.0468</b>	<b>14.3</b>
EB	-	96-01	<b>0.112</b> (0.12,4.6)	<b>333.9</b> (0.24,7.8)	<b>14.90</b> (0.40,3.4)	<b>0.0373</b> (0.79,1.3)	<b>6.4</b> (0.59,1.6)
ES	-	1999	<b>0.463</b> (0.27,1.5)	<b>82.3</b> (0.78,3.5)	<b>8.20</b> (0.90,2.0)	<b>0.0381</b> (0.77,1.5)	<b>5.1</b> (0.44,1.7)
CM	-	97-01	<b>1.444</b> (0.09,4.3)	<b>30.6</b> (0.23,11.7)	<b>5.09</b> (0.35,3.5)	<b>0.0442</b> (0.76,1.4)	<b>4.3</b> (0.25,6.8)
EN	-	1998	<b>0.00425</b>	<b>8529.8</b>	<b>51.13</b>	<b>0.0363</b>	<b>14.3</b>
All blocks	-	96-01	<b>0.041</b> (0.27,3.4)	<b>882.9</b> (0.32,3.8)	<b>27.06</b> (0.49,2.1)	<b>0.0365</b> (0.89,1.2)	<b>5.3</b> (0.62,1.6)

The uncertainty calculations are based on simulating new data sets from the estimated models, and we will show the effect of this in more detail. Figure 4 shows empirical K-functions from 5 simulated data sets for each of the small areas. For each small area, all simulations have the same expected number of observed whales as has been observed in the real data sets, and the parameters are constant. The variability in these curves explain why it is relatively difficult to estimate the parameters in the Neyman-Scott model precisely, particularly since we have observed only 20-200 whales.



**Figure 4** Empirical K-functions from 5 different simulated data sets for each of the small areas.

## References

- Aldrin, M., Holden, M. and Schweder, T. (2001). *Estimation of a Neyman-Scott process from line transect data by a one-dimensional K-function*. Norwegian Computing Center, NR-note SAMBA/18/01.
- Cowling, A. (1998). Spatial Methods for Line Transect Surveys. *Biometrics* **54**, 828-839.
- Cressie, N. A. C. (1993). *Statistics for Spatial Data (Revised Edition)*. New York: Wiley.
- Hagen, G. and Schweder, T. (1995). Point clustering of minke whales in the north-eastern Atlantic. In *Whales, Seals, Fish and Man*, A. S. Blix, L. Walløe and Ø. Ulltang (eds), 27-33. Amsterdam: Elsevier Science B.V.
- Koppervik, I. (1996). A note on Neyman-Scott point process parameter for north-eastern Atlantic minke whales. Paper SC/A96/AE26 presented to the Abundance Estimation Working Group of the Scientific Committee. 13 pp.
- Ripley, B. D. (1977). Modelling spatial patterns (with discussion). *Journal of the Royal Statistical Society, Series B* **39**, 172-212.
- Schweder, T., Skaug, H. J., Dimakos, X. K. and Langaas, M. (1997). Abundance of Northeastern Atlantic Minke Whales, Estimates for 1989 and 1995. *Rep. int. Whal. Commn.* **47**, 453-483.
- Schweder, T., Skaug, H. J., Langaas, M. and Dimakos, X. K. (1999). Simulated Likelihood Methods for Complex Double-Platform Line Transect Surveys. *Biometrics* **55**, 678-687
- Skaug, Øien, Bøthun, Schweder (2002). Abundance of northeastern Atlantic minke whales for the survey period 1996-2001. To be published.

## **Appendix A Plots of location of detected whales**

The figures in this section show the positions of the detected whales relative to their pieces of transect lines, separately for each year, block and platform. Each transect is plotted as a solid line, starting at  $y=0$ , and with length equal to the length of the transect. Detected whales are plotted as triangles. The dashed lines divides the various transects. The distances along the y-axis and the x-axis are given in kilometres, but note that the scale is different between the two axes.

The various blocks and years are plotted in the following order:

Figure A1: NOS 1996, NVN 1997

Figure A2: NVS 1997, NSC 1998

Figure A3: NS 1998, BAE 2000

Figure A4: NOS 2001

Figure A5: FI 1996, LOC 1996, BAE1996, JMC 1997

Figure A6: BJ 1999, VSN 1999, VSI 1999, SV 1999

Figure A7: NON 1999, BAW 1999, VSS 1999, SVI 1999

Figure A8: KO 2000, LOC 2000, NVN 2001, NVS 2001

Figure A9: KO 2001, GA 2001

Figure A1

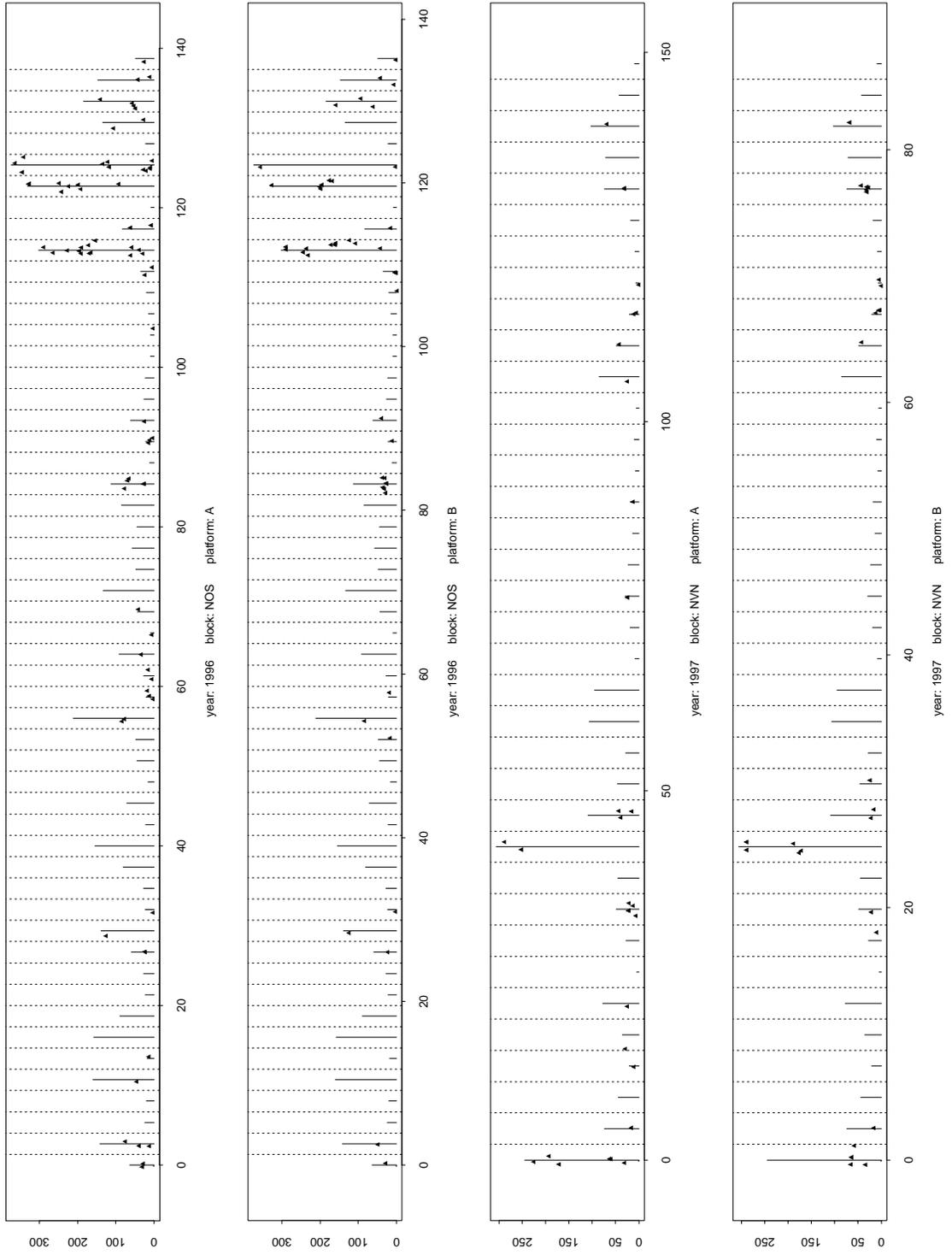


FIGURE A2

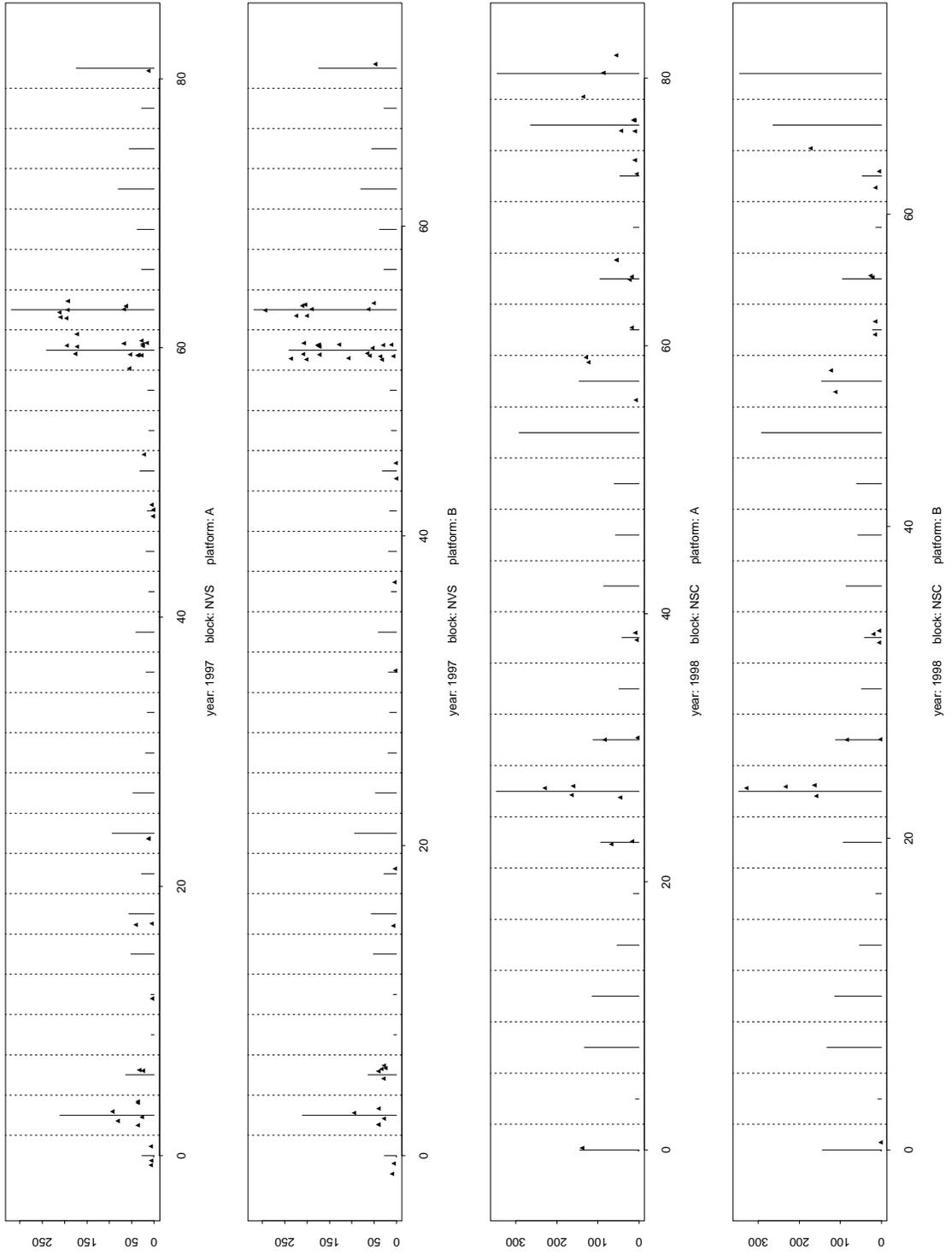


FIGURE A3

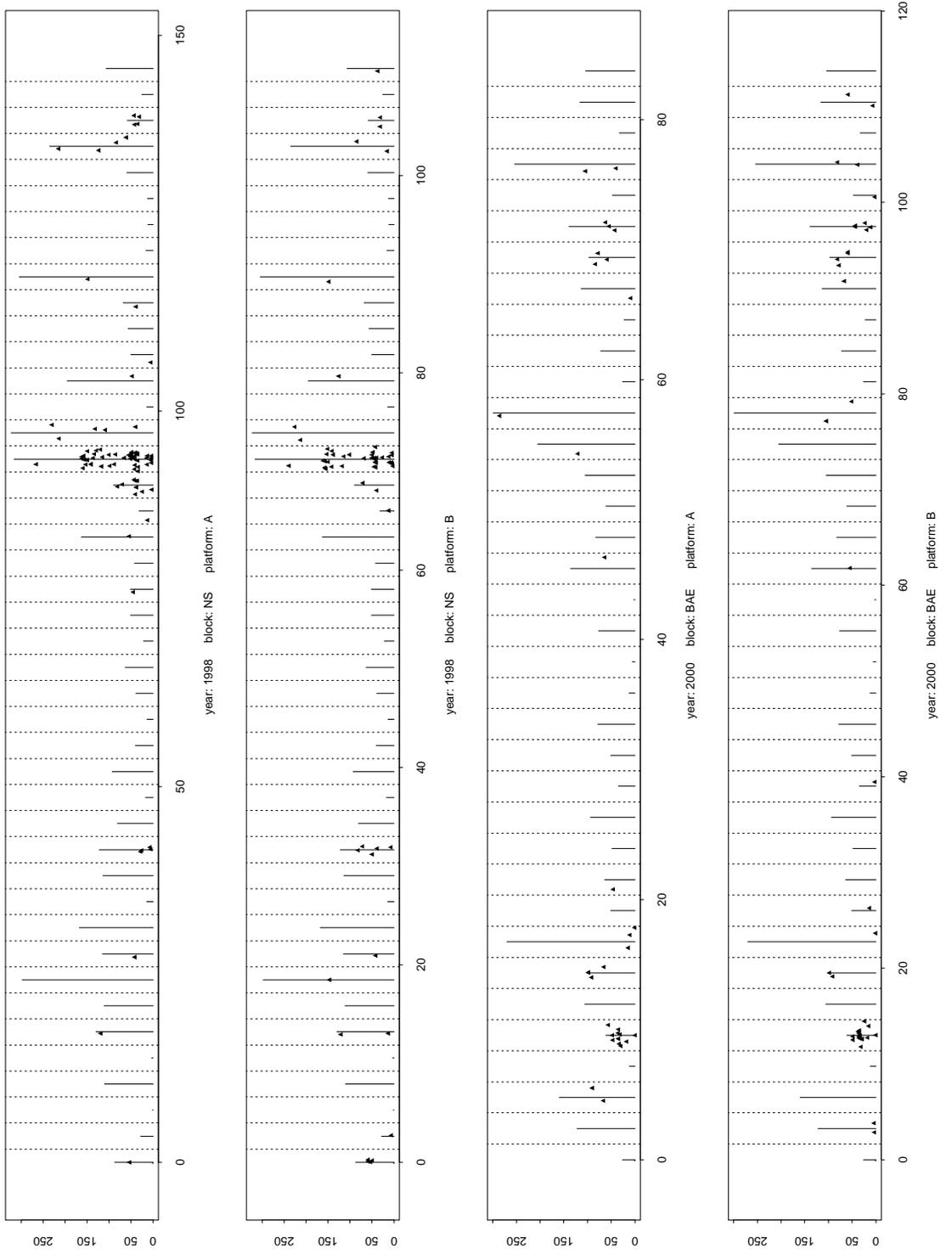


FIGURE A4

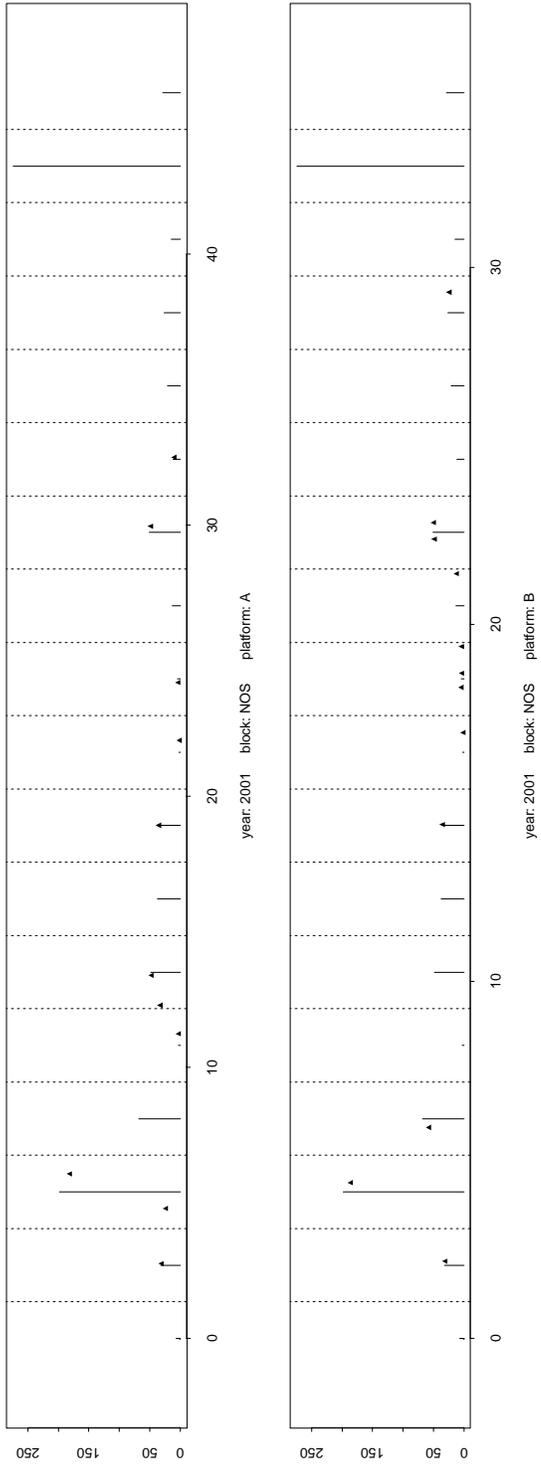


FIGURE A5

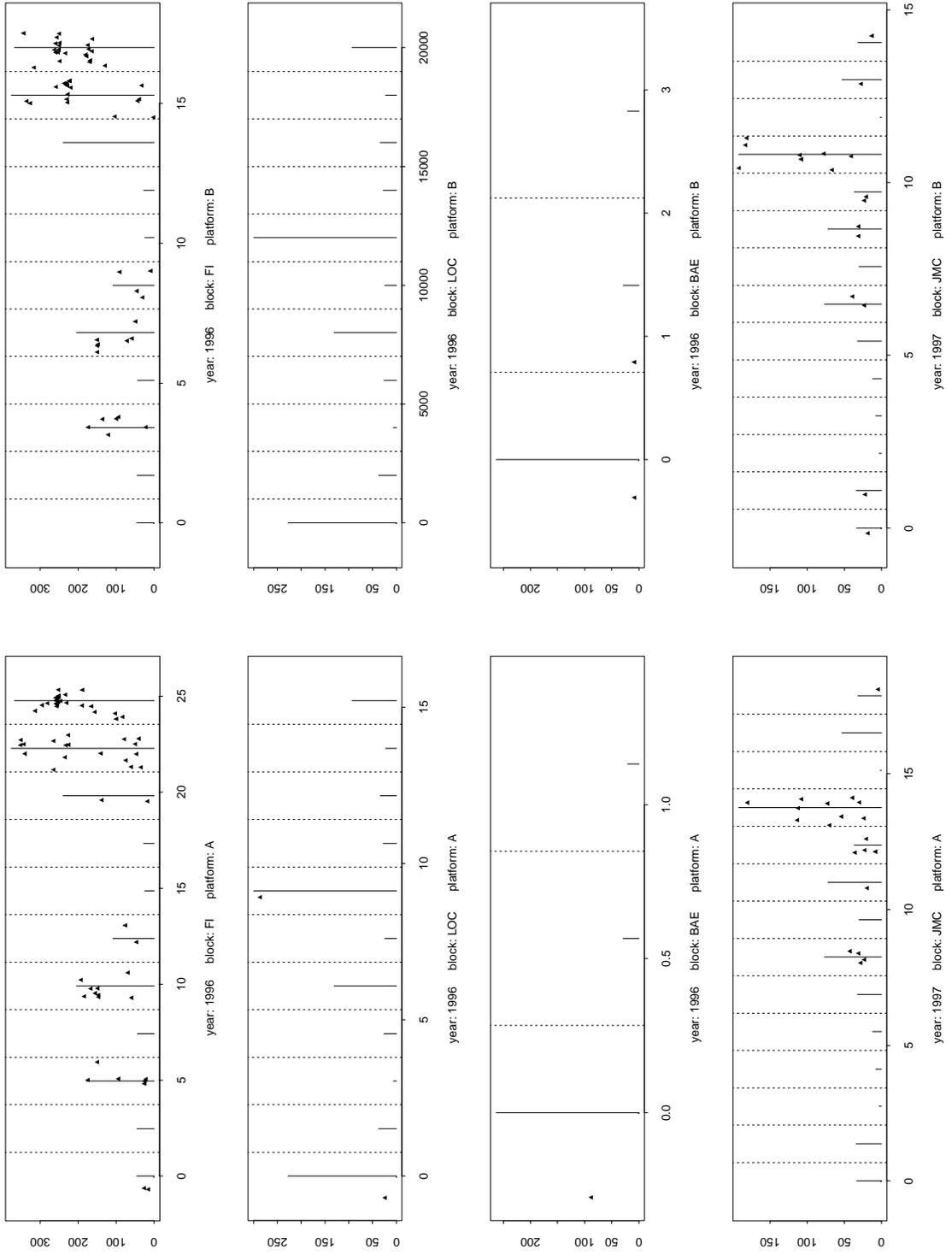


FIGURE A6

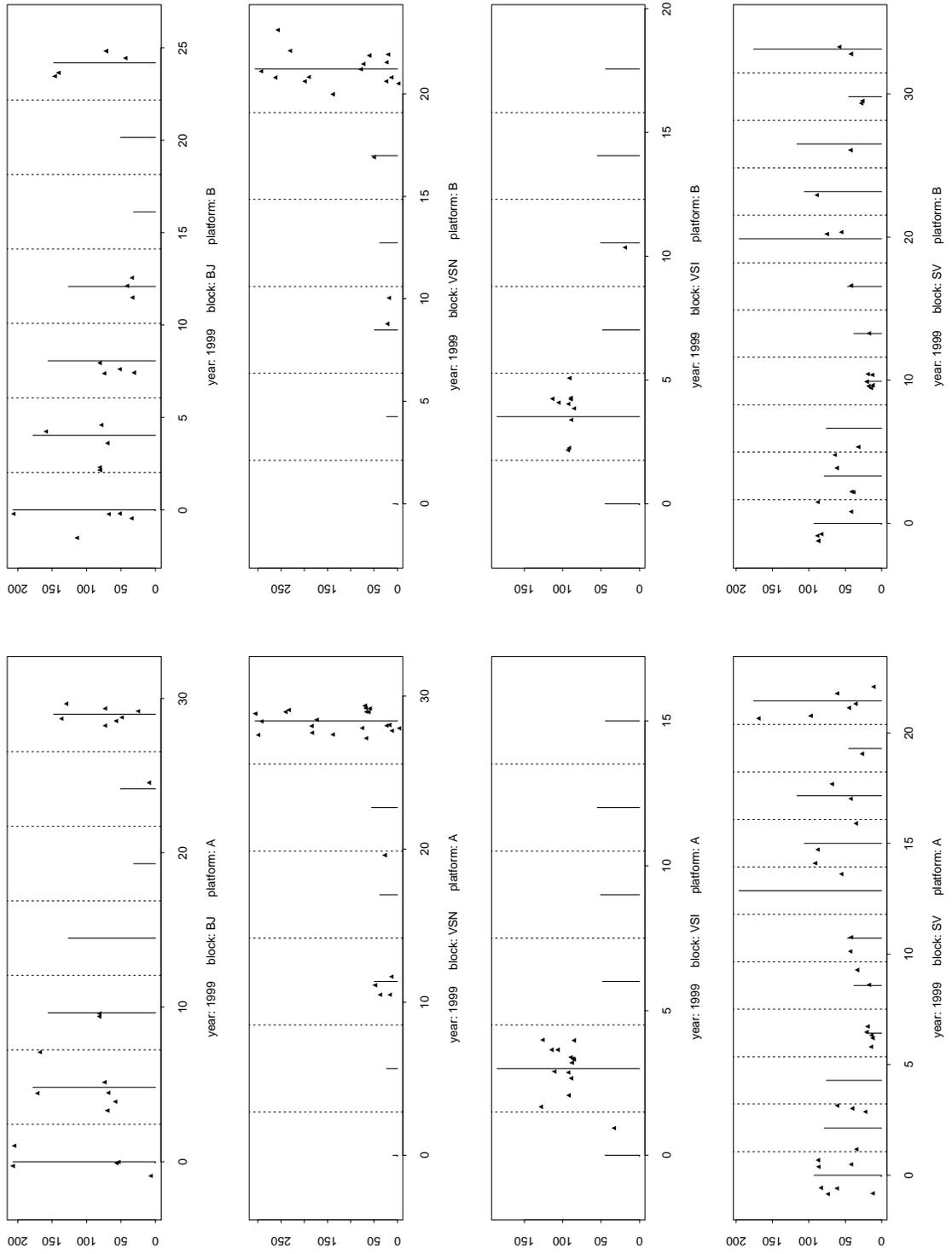
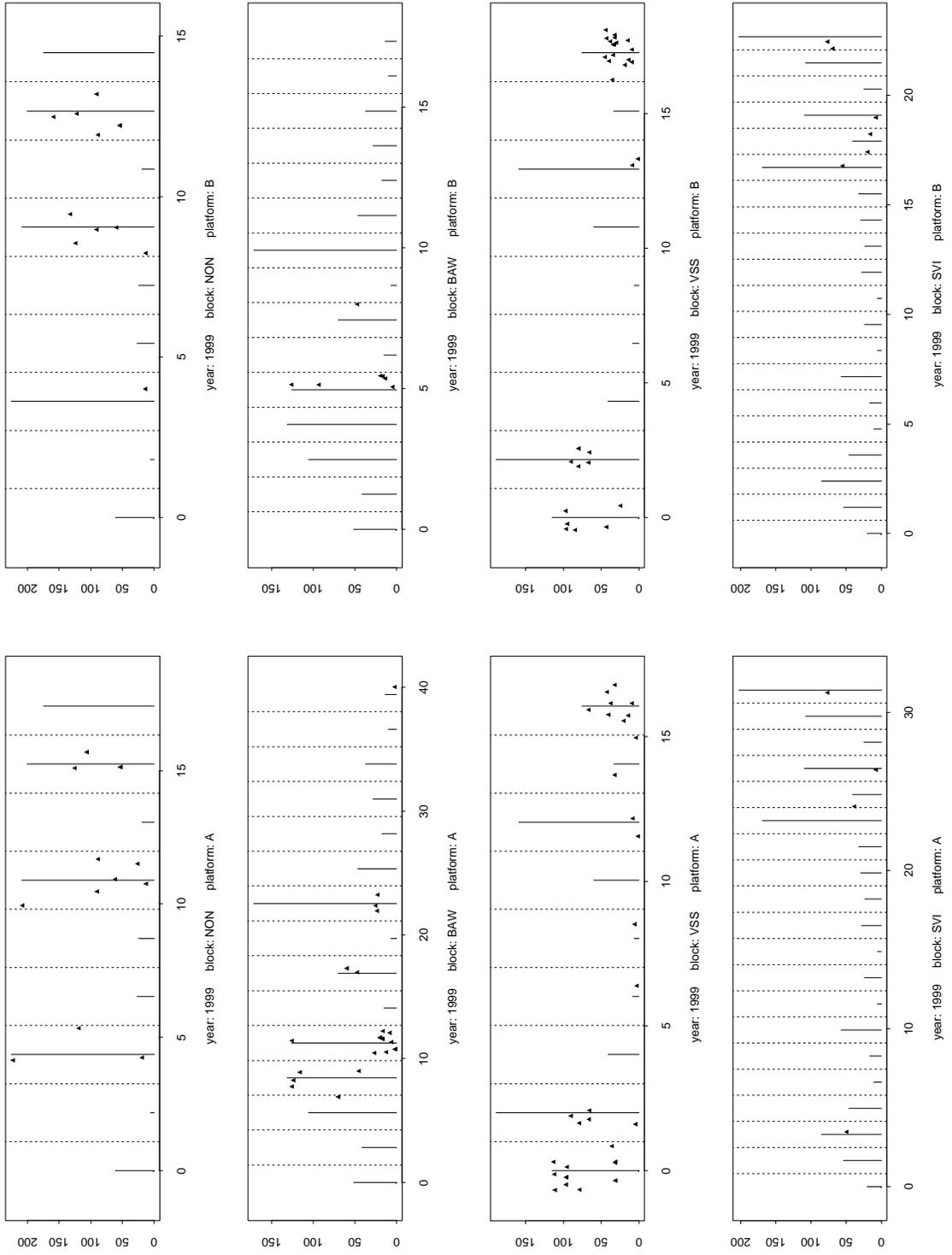


FIGURE A7



**FIGURE A8**

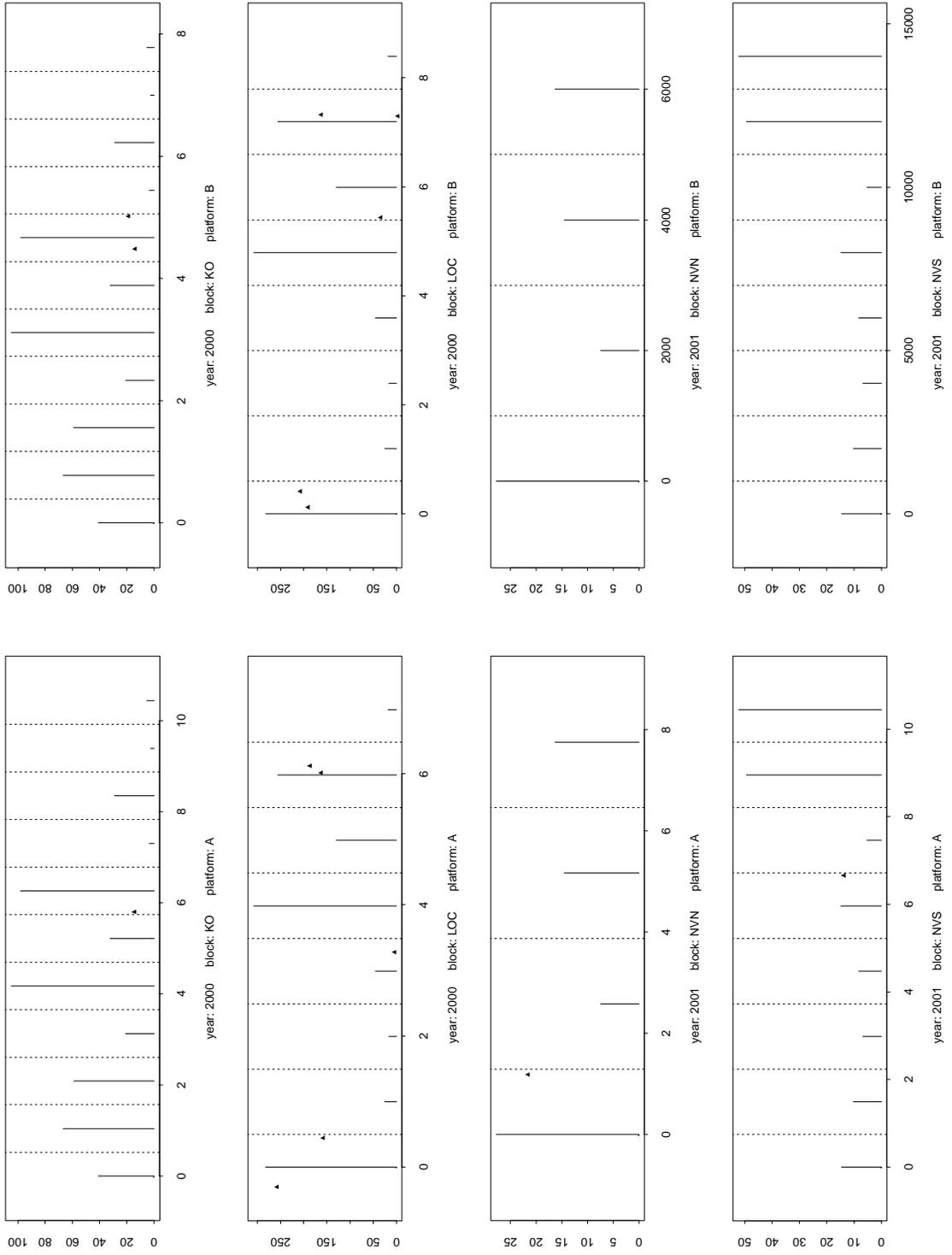


FIGURE A9

