Antarctic minke whale abundance estimates from the second and third circumpolar IDCR/SOWER surveys using the SPLINTR model

MARK V. BRAVINGTON[1] AND SHARON L. HEDLEY[2]

Contact e-mail: mark.bravington@csiro.au

ABSTRACT

Abundance estimates of Antarctic minke whales (*Balaenoptera bonaerensis*) from the second and third circumpolar series of International Whaling Commission International Decade of Cetacean Research (IDCR) / Southern Ocean Whale and Ecosystem Research (SOWER) Antarctic surveys are presented. A customized modelling approach for spatial modelling of line transect data is developed (the 'SPLINTR' model) which comprises three main modelling components: (1) estimation of detection probability, assuming Trackline Conditional Independence (TLCI; Laake, 1999; Laake and Borchers, 2004); (2) using smooth polytomous regression, spatial estimation of mean school size; and (3) using a spatial soap-film smoother (Wood *et al*., 2008) with local clustering incorporated using a Markov-modulated Poisson Process (Skaug, 2006), estimation of school density.

KEYWORDS: ANTARCTIC MINKE WHALE; MODELLING; SOWER; SOUTHERN OCEAN; ABUNDANCE ESTIMATE; SCHOOL SIZE; SURVEY – VESSEL.

INTRODUCTION

Antarctic minke whale (*Balaenoptera bonaerensis*) assessment surveys have been conducted under the auspices of the International Whaling Commission since 1978/79. The series of surveys were from 1978/79-1995/96 called the International Decade of Cetacean Research (IDCR) surveys; and then subsequently Southern Ocean Whale and Ecosystem Research (SOWER) surveys. Each year, in the austral summer, the survey covered a longitudinal region of up to about 60° of the Southern Ocean (generally south of 60°S) resulting in three circumpolar sets of surveys (hereafter referred to as CP1, CP2 and CP3). The design of the CP1 survey was significantly different from CP2 and CP3, and so in this paper, we focus on the latter two series, i.e. on data from 1985/86 onwards. An extensive description of the three sets of surveys is given in Matsuoka *et al*. (2003); here we very briefly summarize the basic survey set-up (from CP2 onwards). Note, that in any particular survey, there may have been changes to what is described below – such deviations are detailed in Matsuoka *et al*. (2003).

The surveys generally operated using two very similar vessels, the *Shonan Maru* and the *Shonan Maru No. 2*. Survey strata were defined as 'Southern' and 'Northern', and usually one vessel operated in each strata, although during the course of a survey, each vessel would survey sometimes in the north and sometimes in the south. The southern boundary of the southern survey strata was defined by the 'estimated ice edge' – a logistical boundary beyond which the vessels (which were not ice-strengthened) were not able to navigate safely. The vast majority of the survey effort was conducted in open water, although some effort in low ice concentrations (typically no more than 10%) was occasionally conducted. Ostensibly though, estimates of minke whale abundance from the IDCR/SOWER surveys relate to estimates in open water only. The survey transects were laid out according to a zig-zag survey design, intended to ensure good coverage throughout each survey. The transect locations were not strictly randomized – the start-point for the northern strata effort was almost always from a corner of the stratum – although some degree of 'randomness' was introduced by the start-point of the southern stratum effort being at the ice edge, a constantly moving feature. The inter-stratum boundary (between northern and southern strata) was determined by the latitudinal width of the southern stratum, so was also affected by the ice edge location. Despite intentions to the contrary, the realized survey coverage is sometimes very uneven (see Fig. 1, for example), and thus in such situations, design-based estimates which rely on equal coverage probability may be biased.

There were three sighting platforms located on each vessel (Figure 2): the 'Top' platform located in the barrel (Platform A) from which two observers searched; the 'IOP' platform (Platform B) located on the mast but below the barrel, from which one observer searched, and the Front or Upper Bridge (Platform C), from where between two and six observers searched, and data were recorded. Two modes of survey were in operation: Closing mode, during which time Platforms A and C operated together; and Independent Observer (IO) mode, during which time all three platforms operated but Platforms A and B searched independently of each other, and there was 'one-way' independence between these two and Platform C, i.e. C was informed of sightings made by A or B, but not *vice versa*. When a sighting was made in Closing mode, the vessel immediately changed course and steamed towards the direction of the sighting, in an attempt to 'close' on the school and, amongst other things, to confirm its size and species composition. In IO mode, when a sighting was made (by any platform), the vessel remained on its designated course as in typical Passing Mode survey. If a sighting was made from either Platforms A or B, then Platform C was informed of the sighting, and the role of the personnel on Platform C was directed towards finding and tracking the sighting(s), until either it was judged to have passed abeam, or it was judged to have been seen by both Platforms A and B, with the 'duplicate status' being determined by the Platform C personnel. During such a time, Platform C was still able to make new sightings, but its primary role was the assessment of duplicate status. It was also possible to record a 'triplicate' sighting in a similar manner, provided that a sighting was seen first by platform C (and subsequently by A then B, or B then A). The primary objective of Closing mode was to obtain data for reliable estimation of mean school size (since recorded school sizes in IO mode tend to be underestimates of true school size – see later). The primary objectives of IO mode were to obtain an estimate of $g(0)$, that could be used to adjust for uncertain trackline detection probability, and to provide reliable data on school

[1] CSIRO Mathematical & Information Sciences, Marine Laboratory, Castray Esplanade, Hobart 7001, TAS, Australia.
[2] The School House, Denhead, St Andrews, Fife KY16 8PA, United Kingdom.

encounter rate, to be used in density estimation. Estimation of encounter rate from Closing mode data from these surveys is biased (and depending on the nature of the local clustering the direction of the bias may vary (Haw, 1991)).

In addition to the standard 'survey' data collected on IDCR/SOWER, all cruises have conducted some effort in Experimental mode, with the 1984/85 cruise dedicated to experiments only. Routine experiments include biopsy sampling, photo-identification, collection of resightings data in IO mode and the Estimated Angle and Distance Experiment (Matsuoka *et al*., 2003). Of particular relevance to this paper, however, is the 'SSII experiment,' first conducted on the 1984/85 and 1985/86 cruises, and recently (since 2006/07) either repeated as an experiment or conducted as an alternative mode to standard Closing mode. During SSII, randomly selected sightings in Passing/IO mode (IO mode only began in 1986/87) are closed upon when they are judged to have passed abeam of the vessel, and school sizes are recorded before and after closure. Provided that observers make no attempt to correct for mis-underestimaton during the course of the experiment by modifying their pre-closure estimates, these data provide extremely useful information for assessing the extent of this feature of IDCR/SOWER data, in various conditions.

Using the IDCR/SOWER data and conventional line transect estimation (e.g. Buckland *et al.*, 2001), estimates from Branch (2006) suggest that an appreciable decline in estimates of Antarctic minke whale abundance has been observed, from 786,000 (CV=0.094) from CP2 to 338,000 (CV = 0.079). These estimates, however, rely on the unrealistic assumption that all minke whale schools seen on the trackline were detected with certainty (i.e. $g(0)=1$, where $g(y)$ is the probability of detection at perpendicular distance y). Although methods do exist to estimate $g(0)$ from double platform line transect data - such as the IO mode data on the IDCR/SOWER surveys (e.g. Laake and Borchers, 2004; Skaug and Schweder, 1999), there are some uniquely challenging features of the data collection protocols, and of the Antarctic itself, which render straightforward application of existing approaches unreliable, particularly when examining trend in abundance. Together with estimation of $g(0)$ (and in some cases linked to its estimation), perhaps the main challenges are as follows:

i. the average school size varies in space (even within the designed strata), and has reputedly decreased over the decades during which the surveys have been conducted;

ii. the density of schools varies in space (even within the designed strata), in a way that is correlated with average school size;

iii. the average sighting conditions vary in space (even within the design strata), in a way that is correlated both with average school size and with school density. Near the ice edge, conditions tend to be better, schools tend to be larger, and density is higher. This means that non-covariate-based estimates of, e.g. effective strip width, based on all observations are biased towards conditions close to the ice, where more sightings happen to be made because of better weather;

iv. the coverage within a stratum is often very uneven despite good intentions in the survey design, because weather cannot be controlled for, and the ice edge location is highly variable in some regions;

v. linked to (iv) above, existing line transect spatial model-based approaches (e.g. Hedley and Buckland, 2004) that could be applied when design-based estimates might be considered unreasonable, in practice suffer from the following issues of their own:

   - where survey coverage is poor near the edge of the survey region, particularly towards the corners, typical smoothers tend to extrapolate linearly. This is biologically unreasonable and since smoothers typically describe log-abundance rather than abundance, any increasing linear trend has a disproportionate impact on the abundance estimate, resulting in positively biased , imprecise estimates.

   - where there are complex survey boundaries, e.g. where a narrow peninsula of ice or land sticks out into the middle of a body of water, data from on side of the peninsula will leak across to the other side of the peninsula because the two sections of water are close by a simple distance metric ('as the crow flies') – but not by a more reasonable biological distance metric ('as the whale swims').

vi. $g(0)$ and nominal[1] effective strip widths depend strongly on school size (as well as sighting conditions). In IO mode, $g(0)$ could be estimated from the independent platforms, except that school size is frequently underestimated in IO mode;

vii. in Closing mode, school size is reliable, but there is no independent-platform data to estimate $g(0)$;

viii. a different combination of platforms operate in IO and Closing modes. Platform C is one-way dependent in IO mode (so there are no data to determine whether Platform C would have seen a school initially sighted by A or B), and in Closing mode, the platforms are not independent at all (closure begins immediately upon a school being sighted – there are no data on whether the other platform – A or C – would have eventually seen the school);

ix. the data records do not include independent estimates of school size or identification of species by platform. Rather they are a 'joint effort' between observers on the platforms that sighted the school;

x. there were changes over time in observer habits and experience (Mori *et al.*, 2003), and in platform setup – considerable structural modifications were made to the *Shonan Maru* platforms preceeding the 1998/99 survey, a new IOP platform was installed on the *Shonan Maru No. 2* at the same time, and further modifications were made to her prior to the 1999/2000 survey (Matsuoka *et al.,* 2003);

xi. schools are somewhat clumped spatially, on a small scale.

[1] Nominal effective strip width, i.e. unadjusted for $g(0)<1$.

Further, it is often useful to be able to produce estimates for arbitrary spatial extents, for example to compare only areas covered in both CP2 and CP3, or to investigate the possible effects of different amounts of pack-ice habitat on apparent abundance.

Over the past 12 years or so, addressing these issues adequately in order to obtain reliable estimates of Antarctic minke whale abundance from the IDCR/SOWER surveys has been a primary focus of the In-Depth Assessment sub-committee of the IWC Scientific Committee. Within the Scientific Committee, three modelling approaches, largely customized for analysis of these data but also applicable to other, much less problematic situations, have been developed. The three methods are the Integrated Model (IM; Cooke and Leaper, 1998; Cooke, 2009), the OK method (Okamura *et al*., 2003, 2005; Okamura and Kitakado, 2008) and the SPLINTR (SPatial LINe TRansect) model (Bravington, 2008, 2009). Both the IM and the OK method utilize cue-based two-dimensional hazard probability models to estimate (and correct for uncertain) trackline detection probability, and to estimate effective strip width. Bias-corrections are estimated and applied in order to estimate school size. The IM approach uses Fourier series in a spatial model for school encounter rate. In contrast, the OK method uses stratified estimation, based on a Horvitz-Thompson estimator, for encounter rate. The SPLINTR model assumes Trackline Conditional Independence (or 'point' independence; Laake and Borchers, 2004) to estimate $g(0)$, and estimation of effective strip widths are based on perpendicular distances; the sighting, not the cue, is treated as treated as the unit of observation. Since these parameters depend on school size, and true school sizes are usually unknown for IO mode, school size error models have been developed to handle errors in recorded school sizes. These two components of the model are estimated in tandem – they depend on each other – and then school encounter rate is modelled using a spatial soap-film model designed for regions which may be topographically complex (Wood *et al*., 2008), but which also accommodates local clustering (Skaug, 2006). This paper details the SPLINTR model, and presents estimates of Antarctic minke whale abundance from it, applied to real and simulated IDCR/SOWER data. Note that an overview of the similarities and differences between the three methods is provided in Table 1 of IWC (2008).

The paper is structured as follows. The next section gives some necessary notation for describing the model, and describes its basic structure, including a brief description of estimation of variance through the various model components. The Data section documents the (largely standardized) data used in these analyses. The Results section presents model options used in the analysis, and tabulates the resulting estimates of abundance and their variance. Some diagnostics useful for assessing the fit to the models are presented. We have been careful to document precisely how the diagnostics have been calculated; the detail is important, as discussed at the recent IWC Workshop on Abundance Estimates in St Andrews (IWC, 2009).

SPLINTR MODEL STRUCTURE

The basic structure of the SPLINTR model is depicted schematically in Figure 3. Fundamentally, it is a three-part model, comprising the following three main components necessary to address the (possibly unique) intricacies of modelling the IDCR/SOWER data:

1. Estimation of detection functions and g(0), conditional on recorded school sizes (which we know to be underestimates of the true school sizes in IO mode), as well as on other covariates/factors. Termed the 'distance sampling' part of SPLINTR in this paper.
2. A spatial model for the frequency distribution of true school sizes.
3. A spatial model for school encounter rate/density.

Because of errors in recorded school sizes, the first two components of SPLINTR must be estimated simultaneously. (To see why this must be the case, consider a region where most minke whales occur in schools of size 1. Given a *recorded* school of size 1 in IO mode, the chances are that in such a region, it really is a school of size 1. In contrast, in a region where whales frequently occur in larger schools, recorded sizes in IO mode would often tend to be underestimates of the true school size, so that there would be a higher probability of a school recorded as of size 1 being in fact a school of size 2 or more.) It is thus also necessary to introduce a sub-component of the school size estimation model to describe school size mis-underestimation[3], and this too, must be fitted simultaneously with the first two components of the model.

Once these components of the model have been estimated, the true school size distribution at every place can be predicted, and the probability of seeing a school of given (true) size under given conditions can be calculated. Hence, for any bit of trackline, the probability of seeing a school if one was present (i.e. the *local* effective strip width (ESW) integrated across possible school sizes) may be estimated. This local ESW enters the third main component of SPLINTR – a spatial school encounter rate/density model – via an offset term. We term this third component the DOSS (Density-Of-Schools-Spatial) model. It is based on the number of schools seen per section of trackline and also incorporates fine-scale clustering (variation that cannot be well represented by broad-scale smoothers) using Skaug's (2006) Markov-modulated Poisson Process. The resulting school density surface may then be multiplied by the estimated true school size surface to give whale density. Summing over the spatial region of interest, an estimate of whale abundance in that region is obtained.

Further details of the model components are given below (and in the Appendices), but first, some notation is introduced. For the distance-sampling parts of the model, the quantities observed are perpendicular distance (y), estimated (i.e. recorded) school size (s_e), and platform who made the sighting, (h), given x, z, o and m where:

- x is spatial location;
- z is environmental conditions (and can be in the form of an 'interaction', e.g. vessel*sightability). Assume for now that z is a

[3] In fact, examination of the SSII data suggest there are a very small number of cases where school sizes are over-estimated, but mis-underestimation describes the usual situation in IDCR/SOWER of IO mode school size being recorded as less than, or equal to, the true value.

discrete-valued factor, although the form of dependence of various parameters on the underlying components of z and s can be arbitrarily complex – see below;

- o is the fact that it was observed at all, i.e. in IO mode, that h is one of *AB, Ab, aB, Cab* (where upper case denotes 'seen' and lower case 'not seen') rather than *abc*; and in Closing mode, that h is $A \cup C$ rather than *ac*;
- m is survey mode (IO or Closing).
- s is true school size, grouped into categories. We assume that $s_e \equiv s$ in Closing mode, but in IO mode, $s_e \leq s$ (for the IDCR/SOWER data at least; perhaps not for the simulated data).

The statistical task is to model $\mathsf{P}[ys_e h \mid xzom]$, which can then be used to compute ESW as a function of *x, z* and *m*. The first two model components (estimation of $\mathsf{P}[h \mid yszm]$ $\forall h$, and $\mathsf{P}[s \mid x]$) plus the school size error model ($\mathsf{P}[s_e \mid yszo]$), would give us what we need. Note that for the school size error model, we assume that the chance of error does not depend on which platform(s) saw the school. In practice, presumably, the more platforms that saw the school, plus the number of times that it was seen, would affect the school size error, but the data are not recorded sufficiently consistently for this to be assessed reliably. Furthermore, the data indicate that school size error is not dependent upon the perpendicular distance of the initial sighting, therefore the school size error model becomes $\mathsf{P}[s_e \mid szo]$. Henceforth in this paper, we refer to these model components collectively as SPAMASSS (Sighting-Probability-And-Misunderestimation-And-Spatial-School-Size).

School size distribution, and school size error, are potentially handled slightly differently in SPLINTR compared to OK or IM. A fully nonparametric distribution can be used for school size, i.e. not restricted to Negative Binomial etc. A fully nonparametric school size error model can also be used. Since the St Andrews workshop, though, we have implemented various simpler bias models, including the OK and IM variants, in order to keep the number of parameters down. The results in this paper are for a Binomial school size error model.

As an aside, we note that IDCR/SOWER data contain a field describing the 'Confirmation' status of each sighting. Sightings in Closing Mode are frequently classified as confirmed, i.e. species identified and school size ascertained with confidence, whereas the majority of IO mode sightings are unconfirmed. Conventional line transect analyses of IDCR/SOWER data (e.g. Branch and Butterworth, 2001; Branch, 2006) use confirmed sightings to estimate mean school size. However, because the probability of confirmation depends on school size (so that successful confirmation is less likely for smaller schools, particularly singletons), this can lead to positively biased estimates of mean school size. Because of the complicated nature of confirmation, and because the process by which a school is classified as confirmed is fundamentally different between the two modes, in this paper we do not use confirmation status at all, either in IO mode or in Closing mode. Instead, we assume all school sizes in Closing mode are correct (whether confirmed or not), and that all school sizes in IO mode are possibly in error (whether confirmed or not).

Some motivation for and details of SPAMASSS

The SPAMASS components of the model were developed primarily considering the following attributes of, or assumptions about, IDCR/SOWER data: (1) detection of a minke whale school is assumed independent on the trackline, but not necessarily elsewhere; (2) even if there were no errors in recorded school sizes, it is still necessary to estimate a local school size distribution because pooling across group sizes (to get stratified estimates) does not account for different *mixtures* of detection functions in different strata; and (3) there are errors in recorded school sizes, at least in IO mode.

The first of these, namely 'Trackline Conditional Independence' (TCLI; Laake, 1999; Laake and Borchers, 2004) is an established method for the two-platform case common to many line transect marine mammal surveys, to handle the phenomenon whereby the seen-by-both detection function of two platforms is very different to the product of the individual detection functions. (Typically, the seen-by-both distribution is much higher in the tail of the distribution – at higher perpendicular distances – than it should be, if full independence (FI) applied.) Indeed, this phenomenon was noted for the IDCR/SOWER data by Ashbridge *et al*. (1998). TLCI allows the seen-by-both distribution to follow the form it wants to, and replaces the FI assumption by the assumption that detection is independent if the animal is directly on the trackline, but not otherwise. It can be shown to be equivalent to a covariate-stratified version of the Direct Duplicate method of Palka (1995) under some circumstances; it is essentially a likelihood-based formalization of the latter, and allows standard likelihood-based inference on the results.

During IDCR/SOWER surveys, however, there were two modes of survey (IO and Closing) plus a rather unusual platform semi-independent arrangement, which differed by survey mode. In particular, in IO mode, the only useful data we have from Platform C is for schools that were not seen by either of the other two platforms. Accommodation of this platform arrangement, whilst retaining as much of the data as possible, is not straightforward, but it is algebraically possible to reformulate the two-platform TCLI equations (Bravington *et al.*, 2006). See Appendix 1 for full details.

Spatial group size distribution can be modelled by a smooth polytomous regression (see, e.g. Harrell, 2001). A dummy variable μ is estimated as a function of location *x*, and $\mathsf{P}[S = s \mid x] = \Phi(\delta_s = \mu(s))$ where Φ is the cumulative Normal distribution. δ_s is a fixed cutpoint, estimated for each *s* but fixed in space. The distribution is partly nonparametric, in that if μ were constant, the δ_s would provide the flexibility to give any distribution to *s* (for example, this is not restricted to commonly-used assumed school size distributions such as the negative binomial). However, the polytomous form does constrain how the distribution can vary; given the limited size of the dataset for each year, this is probably a good thing.

Although it is theoretically possible to estimate IO mode mis-underestimation probabilities by direct comparison between "nearby" school size distributions from Closing and IO mode, and to separate mis-underestimation parameters from other distance-sampling

parameters, this requires very high variability in true school size distribution from place to place, and places a heavy burden on the IDCR/SOWER data. Therefore, in contrast to the OK and IM methods, we directly incorporate data from the 2007, 2008 and 2009 SSII and SSIII experiments into our likelihood, in essence to generate prior distributions on the probability of observed size given true school size.

Unusual feature: Bivariate-increasing dependency

Sometimes common sense suggests that certain parameters should increase with two covariates; for example, many scale parameters should probably increase both with increasing true school size (*Struth*) and with increased Sightability conditions (*Sig*). Specifying an additive dependence via a '*Struth* + *Sig*' model formula might achieve this, but may not be flexible enough; on the other hand, because of limited data, a full-interaction model via '*Struth* * *Sig*' may give nonsensical estimates where, for example, scale decreases with school size at some Sightability values. Bivariate-increasing independence was thus developed to allow situations such as the following to be included in the model formulae: full interaction, but non-decreasing with school size for given Sightability, and non-decreasing with Sightability for given school size.

Unusual feature: the Z-change model

In theory, it might be optimal to fit the spatial smooths for school density simultaneously with the spatial smooths for school size, rather than fitting sequentially with one model depending on the other's outputs. In practice, however, this is completely infeasible, so as a compromise, we incorporate a Z-change component into the SPAMASSS likelihood. The idea is to use some of the information on relative encounter rates in different conditions (e.g. far more sightings made in Sightability 4 than in Sightability 2 at nearby locations) in order to inform, and aid the fitting of, the SPAMASSS likelihoods. On a large spatial scale, there is substantial confounding between weather conditions and location, so it is logically incorrect to include all the encounter rate data directly in the SPAMASSS model. However, the DOSS model only deals with large spatial scales. When sighting conditions change during a stretch of effort, the Z-change model allows SPAMASSS to use pairwise comparisons of numbers of sightings in a limited time period before and after the stretch, in a region where local density and local mean school size is effectively constant. As long as the distance covered before and after the change is considerably smaller than the knot spacing[4] in the DOSS model, then there is no confounding (and no "double use" of data), because the spatial scales are completely different. The comparison is basically Binomial, allowing for the differences in time interval and the intrinsic difference in ESW that occurs in the different conditions before and after the change. However, because of imprecision in recording the time of the change in conditions, and because of fine-scale clustering of schools, each Binomial probability might vary. The comparison we use is therefore a Beta-Binomial model for '#schools before, given #schools before and #schools after' (with an additional parameter for overdispersion).

Preliminary results from using the Z-change model on the IWC simulated datasets were encouraging (Palka, 2009). From experience with the SOWER, the Z-change model can largely eliminated unwanted "residual patterns", in terms of say a trend with weather conditions in the proportion of schools seen versus predicted.

Fitting SPAMASSS

The SPAMASSS model is at heart just a very complicated mixed-effect model (Wood, 2006), in which the coefficients describing each year's spatial surface are the random effects, and all the distance-sampling parameters are the fixed effects. The covariance of the random effects are governed by smoothing parameters, which are estimated in the same 'layer' as the fixed effects. Evaluation of the log-likelihood for the combined fixed effects and smoothing parameters requires integration over possible values of the spatial coefficients, which is done via Laplace approximation; this is the same strategy used by the ADMB-RE software, although we have used our own code. All necessary derivatives, of which there are a great many, are computed by Automatic Differentiation,based on Tapenade.

Some details of the DOSS component of SPLINTR

Once we have estimated $\mathsf{P}[S \mid x]$ and $\mathsf{P}[o \mid sz]$, we can find the probability of sighting any group that happens to be present at a point x_i when $z = z_i$; it is $\sum_s \mathsf{P}[s \mid x_i]\mathsf{P}[o \mid sz_i]$. In particular, we can do this over the whole trackline, in effect estimating a continuously-varying ESW. School density can then be estimated by assuming it follows a Poisson process (see below) with true local rate following a smooth curve modelled by another soap-film smoother, but modified by the ESW. In other words, over a short time interval δt of survey effort near x_i :

$$\mathsf{P}[\texttt{Number seen} = k] = (\rho_i ESW_i)^k \exp(-\rho_i ESW_i)/k!$$
$$\rho_i = \exp\left(X_i^D \beta^D\right)$$

[4] In the present analyses, knot spacing is set at 90 nautical miles. We have not had time to examine in any detail the sensitivity of our estimates to this choice but experience suggests that it is not critical as long as the knot spacing is sufficiently small to have enough degrees of freedom to represent the underlying 'truth' reasonably well, whilst large enough to maintain reasonable computational efficiency.

where X_i^D is the ith row of the smoother design matrix, and β^D is the parameters of the smoother. In practice, the effort data is broken into short snippets (of 15 minutes in this paper) with constant z in each, and the number of sightings within the snippet is counted.

To deal with clustering at a finer scale than the density surface can model, we use a reparametrization of Skaug's (2006) Markov-modulated Poisson process (MMPP), which flips at random according to a Markov process between two states. In one state, it intensifies the local density ρ by a constant multiplier, and in the other it attenutates ρ by a constant divisor. The divisor is chosen so that the mean of the modified process remains 1. The two parameters are the rate of state-flipping, and the multiplier, which control the width and the intensity of the clumps. Both parameters can depend on a (presumably spatial) covariate, such as distance from ice edge, e.g. to allow more clumpiness near the ice than further away.

As noted above, Closing mode presents a conundrum for density estimation. The act of closure could disrupt nearby aggregations, and it is hard to know what to do about secondary[5] sightings that are inevitably made during a closure attempt. Logically, though, it seems reasonable to include Closing mode effort up to and including the start of the first closure attempt, because until that point the protocols are the same as IO mode (except that the B platform is not operating). After a closure begins, the rest of the transect is deemed unusable for density estimation (though any sightings may be used in school size estimation). This approach, of using time to first sighting, only gives unbiased estimates of density if encounter rates are truly Poisson, but it is known to be biased if there is clumpiness. However, if the MMPP model deals correctly with clustering, then estimation should be unbiased. Closing mode effort and sightings used this way add quite a lot of information, particularly in regions when encounter rates are low and sighting conditions are poor. Although this seems to be helpful for at least some of the simulated datasets, limited trials on the SOWER data suggest that our current clustering model is too simple and that incorporation of CL-mode still leads to negative bias, so we have not used Closing mode for density at all in this paper, or in our analyses of simulated data.

The density surface model is parametrized by the smoothing parameters λ^D, and by the parameters of the Markov-modulation. As with SPAMASSS, a REML-style approach is used to maximise over the approximate marginal likelihood with respect to those parameters, and the spatial density parameters β^D are estimated conditional on the outer parameters.

Variance estimation

Whilst the basic structure of SPLINTR is quite simple (Figure 3), the associated complexities of optimizing functions based on a large number of parameters could not be achieved without Automatic Differentiation (e.g. Skaug and Fournier, 2006). Nonetheless, the approach we have taken has the considerable appeal that all the model components are likelihood-based. Furthermore, the time series of abundance estimates abundance is just a multivariate function of the estimated parameters, so an asymptotically-valid estimate of variance can be derived from the Delta-method and the inverse Hessian (at least, conditional on the estimated smoothing parameters for spatial group size and spatial density). Since the SPAMASSS estimates directly affect the fitting of the DOSS model, it is necessary to propagate uncertainties across the models, which can again be done using techniques from Automatic Differentiation. A large number of parameters are required for the sighting-probability models, in particular, and so the appeal to asymptotics may be questionable; however, many of the parameters are very restricted in the extent to which they can affect either the fit or the estimates of abundance.

DATA USED IN THESE ANALYSES

The data used in our analyses were those extracted from the IWC Database Estimation and Software System (DESS) to form a 'standard' database (Burt, 2004). In addition, data from the SSII experiments in 1984/85 and 1985/86, and from 2006/07 – 2008/09 have been used to help estimate the extent and nature of school size error (in IO mode).

For various components of the model, data from CP2 and CP3 were analysed using all primary on-effort activity codes (i.e. BC, BA, BL, BR, BK in Closing mode and BO, BH, BI, BU, BQ in IO mode). However, for school density estimation, only IO mode effort and effort up to the first sighting in Closing mode was used.

Perpendicular distance truncation of bias-corrected sighting distances was at 1.5n.miles. Sightings of 'like-minkes' and 'undetermined minkes' were included, together with confirmed Antarctic minke whales (i.e. species codes 04, 91, 92 and 39).

We have used both Beaufort and Sightability in analysis. Generally, Sightability variations have more of an effect on sighting probabilities than Beaufort does; however, the effects of Sightability in CP2 are less marked and there may be some differences between vessel responses to Sightability in CP2. The results shown in this paper in fact use Beaufort in CP2 and Sightability in CP3, giving slightly better fits, although overall abundance estimates are not much affected by changing this.

So far, we have only used 'Definite' duplicates in SPLINTR, because of lack of time to investigate the effect of including Possibles as well.

[5] Additional sightings that are made during closure on a sighting made from the trackline.

IMPLEMENTING SPLINTR

SPLINTR is mainly written in the Delphi programming language, but as can be seen below, the models may all be run by a simple (if a little lengthy) call in R (R Development Core Team, 2009). Multivariate optimization has been achieved by using Automatic Differentiation (AD). The models were too complex and non-standard for implementation in standard AD software, so it was necessary to code the AD routines directly. We used the AD tool *Tapenade* (Inria, 2002) for this.

For illustrative purposes, we could fit the full SPLINTR model (i.e. SPAMASSS and DOSS components) using a call in R similar to the following (some data-subsetting and recoding arguments omitted for brevity):

```
spam1 <- fit.sower4( data= sower.dat,  just.SPAMASSS= F,    SS.cut = c(1:10,11,15,20), general.start.form = ~Sig * Struth, g0.form = ~(Sig + Struth)*Platform,   scale.form.general = ~Vessel,   scale.y_AUBUC.form = ~Vessel + Sig + Struth,   shape.form.general = ~1,   fit.Ccond = FALSE, ss.yr.scale.formula = ~factor(ygroup.code),  ssbias.form=~1, use.CL.dens=T)
```

although in practice, we would usually fit the SPAMASSS components first, and then the DOSS model. The call demonstrates some of the model options that need to be selected (currently, and this is not anticipated to change in the near future, there is no formal model – i.e. covariate – selection in the code). Specifically we have:

- SS.cut = specification of school size categories
- general.start.form = starting value formula, used to chop the data into subsets – here the full interaction between Sightability ('Sig') and true school size ('Struth').
- g0.form = formula for dependence of g(0) on covariates
- scale.formula.general = general formula for dependence of detection function scale parameter(s) on covariates
- scale.y_ $A \cup B \cup C$ .form = specific formula for dependence of scale parameter of the detection function of all platforms combined on covariates
- shape.formula.general = general formula for dependence of detection function shape parameter(s) on covariates
- fit.Ccond = a logical value giving choice of parametrization of the TLCI equations for the IDCR/SOWER platform setup. See Appendix 1 for further details.
- ss.yr.scale.formula = formula for specifying how the spatial school size distribution varies in time – here the spatial distribution is freely allowed to vary by year. In practice, there have been no problems implementing this, and it requires fewer assumptions than other possibilities, such as a steady trend over time (~year) or no dependence on year (~1).
- ssbias.form = formula for specifying the linear predictor of the logit of the *b* parameter in the school size error model (see Appendix 1), such that $\mathrm{E}[S_e \mid s, o] = b * (s\text{-}1)$. This can depend on covariates and true school size, but in the example above is set to ~1 to estimate a fixed bias probability for all sighting conditions and true school sizes.

A typical fit to IDCR/SOWER data from one CP series takes about an hour to run, about half for each component . Variances take a little less time to compute, as a post-hoc step

RESULTS

It is very easy to construct models with several hundred SPAMASSS parameters (excluding the spatial surfaces), just by allowing large numbers of interactions. That is not a good idea, for many reasons. We have undertaken a moderate amount of informal model selection in order to avoid this and other pitfalls, but no formal selection procedures have been used. Our guiding principles have been:

- to get vaguely plausible-looking patterns of parameter estimates;
- to get satisfactory diagnostics, both graphically and in terms of expected/observed encounter rate;
- to exclude a priori unlikely interactions in order to keep the number of parameters manageable.

Generally speaking, abundance estimates have not been that sensitive to model choice, although goodness of fit has been more so. The models we have arrived at in this paper are (where Z is Beaufort or Sightability depending on CP):

g0 ~ Vessel*Platform + Biv(Beauf,SS) [cp2] or Vessel*Platform + Vessel:Biv(Beauf, SS) [cp3]

scale for y_AUBUC, A_By, B_Ay ~ Vessel + Biv(Z,SS)

scale for AUB_AUBUCy ~ Vessel

all ginfinity and shape params: ~ Vessel

SSbias ~ Z

MMPP: fixed transition rate and high-low ratio

Results are given in the Figures and Tables; note that we have not yet run the Additional Variance code, which will inflate CVs substantially. As the Figures at the end of the paper show, the graphical diagnostics seem good. They do not, of course, test the fundamental assumption of Trackline Conditional Independence, since this is intrinsically untestable within a one-dimensional (perpendicular-distance-based) model. We have also compared observed and expected numbers of schools encountered across different covariates (e.g. Vessel, Sightability, proximity to ice, ...). This has been a very useful tool in excluding some models, although with an MMPP it is not necessary for observed and expected numbers of sightings to tally exactly, unlike the case with Poisson-distributed sightings.

Spatial smooths appear well-behaved; there is no suggestion of "flapping around in the corners" even in years with very poor coverage in the regions that we have fitted across (which are not necessarily the regions intended for the survey to cover). However, the fit to CP3 2004 has a problem; there is a considerable excess of predicted over observed sightings that year, and the abundance estimate for the year is substantially higher than under a Poisson-variability-only model. The Ross Sea certainly did have a great many minkes in it that year, and there was a very high concentration in the centre; what seems to be happening is that the bump in the centre is spreading out to an unjustifiable extent. This may be linked to the knot spacing, which for computational reasons is 120nm in these particular fits. The consequence of the spacing is that no feature smaller than about 120nm (a day's travel) can be well-reflected in the smoother, and the MMPP has to soak up any phenomena of this order, in the process sometimes seeming to introduce "extra" sightings. It will be worth checking whether finer knot spacing (without changing the effective smoothing) just in this one year can eliminate this unwelcome fit.

Overall, though, there is one aspect of the results that is very surprising: the abundance estimates are quite low, being not much above the "standard model'' despite including an estimate of g0. While there are some sources of positive bias in the standard estimate arising from non-random spatial distribution of boat, whales, and weather, it is not clear whether these are big enough to overcome the substantial bias that comes from assuming g0=1. Further investigations are required..

DISCUSSION

In this paper, we have introduced the 'SPLINTR' model – currently, a bespoke model for analysing the IDCR/SOWER data from the CP2 and CP3 series – but in the near future, with a far wider range of possible applications for other line transect survey data. It is our belief that by addressing fully the complex issues that are apparent in the IDCR/SOWER data (and to a lesser extent in the IWC simulated datasets) that other survey data will generally be simpler cases of what is possible in SPLINTR, and thus will be fairly readily analysed with simple modifications to the existing framework.

In the Introduction, we listed eleven items, in addition or related to $g(0)$ estimation that without appropriate consideration, may lead to unreliable estimation of abundance from the IDCR/SOWER data. It seems appropriate to revisit these issues in this section, and state how, or to what extent, these issues have been addressed by the SPLINTR model.

i-iii. the average school size, density of schools and average sighting conditions vary in space (even within the designed strata)…

- SPLINTR includes a sightability covariate in estimation of $g(0)$, as well as school size. Since detectability also depends on school size, and true school sizes are generally not available in IO mode, the distance sampling components of SPLINTR are estimated simultaneously with a spatial model for school size error. The spatial element of the school size error model is important, since true school size varies locally, generally in smaller regions than in the defined survey strata. The 'DOSS' component of SPLINTR incorporates the ESW from the distance sampling and school size estimation parts ('SPAMASSS'), and its associated uncertainty; density of schools is then modelled using a spatial smoother developed for complex surveys regions (see below).

iv-v. the coverage within a stratum is often very uneven…and existing spatial models have undesirable properties at the edge of a survey region, and in areas of complex topography…

- SPLINTR is a model-based estimation framework, that does not rely on randomized transects. For density estimation (including prediction beyond the surveyed transect strips), coverage is not required to be even – but it helps. By developing and implementing an appropriate smoother (the soap-film smoother; Wood *et al*., 2008), with 'smoother taming' properties designed to prevent undesirable extrapolations in regions of poor coverage, the effects on the estimates of uneven realized survey coverage have been reduced. Furthermore, the soap-film smoother was developed in order to address issues associated with complex survey boundaries. Indeed, Wood *et al*. (2008) demonstrate its ability to perform well in such regions (see their Aral Sea chlorophyll example), even when there are few data.

vi. g(0) and nominal effective strip widths depend strongly on school size (as well as sighting conditions). In IO mode, g(0) could be estimated from the independent platforms, except that school size is frequently underestimated in IO mode

- See (*i-iii*) above.

vii. in Closing mode, school size is reliable, but there is no independent-platform data to estimate g(0)

- SPLINTR assumes that all school sizes in Closing mode are correct, both confirmed and unconfirmed, and uses these data to fit a spatial smooth polytomous regression model of true school sizes (separately for each year). The school size experiment data (SSII) are used to fit a model for school size error in IO mode, where it is assumed that all school sizes *may* have been recorded in error. Given estimates of true school sizes in IO mode, then $g(0)$ can be modelled appropriately (see (*i-iii*) above).

viii. *a different combination of platforms operate in IO and Closing modes,,,*

- A complication here is that Platform C (the Front/Upper Bridge) makes a high proportion of sightings in Closing mode, but we have little information to estimate its sighting efficiency in IO mode, because sightings by C are only informative if neither A nor B saw them. In SPLINTR, we have a component of the model that implements standard two-platform TLCI assumptions, but in a customized way for the platform setup in IDCR/SOWER. We currently have two possible parametrizations for this component of the model. The first (requiring estimation of 17 parameters even in the simplest case), rather undesirably contains estimation of $\mathrm{P}[C \mid aBy]$ - C saw it at *y*, given B saw it and A missed it – a function that we have no *a priori* notion of how it should behave, and hence of how to diagnose potential lack-of-fit. With an additional assumption (see Appendix 1), the number of parameters reduces to 13, and model-fitting becomes somewhat easier.

ix. *the data records do not include independent estimates of school size or identification of species by platform…*

- this does not directly affect SPLINTR as it stands, but it does restrict the potential to include components of the model that might incorporate these measurement error issues.

x. *there were changes over time in observer habits and experience (Mori et al., 2003), and in platform setup…*

- The expectation that a decline in overall observer experience would provide some explanation for the appreciable decline in conventional line transect estimates (Branch, 2006) was not able to be demonstrated by Mori *et al.* (2003). Whilst anecdotally this may well be the case, in practice, the effects are less clear, perhaps because some inexperienced observers are indeed equivalent to experienced ones, or perhaps because of inexperienced and experienced observers were frequently paired on the same platform. Therefore SPLINTR does not directly accommodate 'observer experience', and, as agreed in the SC, the estimates of abundance presented in this paper include 'like-minkes' (the notion being that with less experience, observers were less likely to identify a sighting as an Antarctic minke school, so their inclusion would ameliorate this problem). 'Survey platform', however, i.e. *h* for 'who saw it' in this paper, is included in various components of the model. The vessel modifications were substantial; we have not attempted to account for these directly in the modelling framework, but as a sensitivity test, have produced some estimates for all years of CP3, and some for CP3 split into the years preceding, and following completion of, the modifications.

xi. schools are somewhat clumped spatially, on a small scale.

- SPLINTR currently addresses this issue by incorporating local clustering in the spatial model for density of schools using Skaug's (2006) MMPP. The main advantage of doing so is increased confidence in our variance estimates; failure of the smooth for school density to model local clustering ('over-smoothing') would result in our estimates being too precise. Allowing more flexibility in the smooth functions themselves ('over-fitting') would also lead to unreliable variance estimates.

It should be noted that SPLINTR currently has no component which incorporates measurement error. Simulation results presented to this meeting (SC/60/IA??) demonstrate the extent to which estimates may be biased due to such error. Likewise, no attempt at all has been made to accommodate incorrect assignment of duplicate status, although the results presented in this paper do examine the sensitivity to including only 'Definite' duplicates compared with 'Definite+Probable' duplicates.

Finally, and we almost do not which to contemplate this, SPLINTR is effectively two modular components: SPAMASSS and DOSS. In theory at least, it is possibly to replace parts of the SPAMASSS models with cue-based hazard proability models akin to those used by the IM (e.g. Cooke, 2009) and OK (e.g. Okamura and Kitakado, 2008) methods, and to bolt on the spatial DOSS model to the respective model outputs. Likewise, it is feasible (although in view of the spatial issues we have outlined above, we do not think it would be very sensible) to use SPAMASSS to model the distance sampling and school size components of an IDCR/SOWER analysis, and then to use stratified (e.g. Horvitz-Thompson like) estimation for school density.

ACKNOWLEDGEMENTS

Firstly, thanks are expressed to all researchers, Captains and crew for their efforts during the course of the IDCR/SOWER surveys. Particular appreciation is due to Paul Ensor, Cruise Leader on many of these surveys, for his assistance in interpreting the data and steering the direction of some of our approaches to analysing them. Thanks to Louise Burt and Marion Hughes for their assistance, often at short notice, in providing DESS data, including the SSII data. David Peel programmed earlier versions of the SPAMASSS model, and the spatial models used in SPLINTR would not have been possible without Simon Wood's work on soap-film smoothers. The model development has been greatly helped by the work led by Debi Palka in producing simulated IDCR/SOWER datasets, and we have benefited from many useful discussions with other SC members, particularly Justin Cooke, Toshihide Kitakado, Hiroshi Okamura, Tom Polacheck, and Hans Julius Skaug. Toshihide Kitakado also kindly provided the programs used in this paper to calculate additional variance.

REFERENCES

Ashbridge, J., Borchers, D.L. and Burt, M.L. Estimates of Antarctic minke density from the first six IWC/IDCR independent observer surveys. Paper SC/50/CAWS34 presented to the IWC Scientific Committee, April 1998 (unpublished).

Branch, T.A. 2006. Abundance estimates for Antarctic minke whales from three completed circumpolar sets of surveys, 1978/79 to 2003/04. Paper SC/58/IA18 presented to the IWC Scientific Committee, May 2006 (unpublished) 28pp.

Branch, T.A. and Butterworth, D.S. 2001. Southern Hemisphere minke whales: standardised abundance estimates from the 1978/79 to 1997/98 IDCR-SOWER surveys. *J. Cetacean Res. Manage.* 3:143-174.

Bravington, M. 2009. Preliminary results from the SPLINTR model for SOWER data. Paper SC/A09/AE2 presented to the IWC Workshop on SOWER Abundance Estimates, St Andrews, April 2009 (unpublished) 15pp.

Bravington, M.V. 2008. Abundance estimates for Antarctic minke whales from SOWER/IDCR data, using trackline-conditional independence and spatial models. Paper SC/60/IA17 presented to the IWC Scientific Committee, June 2008 (unpublished). 18pp.

Bravington, M.V.., Peel, D. and Hedley, S.L. More abundance estimates for Antarctic minke whales, using different methods. Paper SC/58/IA15 presented to the IWC Scientific Committee, June 2006 (unpublished). 10pp.

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. and Thomas, L. 2001. *Introduction to Distance Sampling*. Oxford University Press, Oxford. 448pp.

Cooke, J.G. 2009. A integrated method for analysis of IDCR/SOWER data and TRANSIM simulated data sets. Paper SC/A09/AE6 presented to the IWC Workshop on SOWER Abundance Estimates, St Andrews, April 2009 (unpublished) 11pp.

Cooke J.G. and Leaper R. 1998. A general modelling framework for the estimation of whale abundance from line transect surveys. Paper SC/50/RMP21 presented to the IWC Scientific Committee, June 1998 (unpublished) 16pp.

Harrell, F.E. 2001. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis.* 2nd Edn. Springer, New York. 568pp.

Haw, M.D. 1991. An investigation into the differences in minke whale school density estimates from passing mode and closing mode survey in IDCR Antarctic assessment cruises. *Rep. int. Whal. Commn* 41:313-330.

Hedley, S.L. and Buckland, S.T. 2004. Spatial models for line transect sampling. *J Agr Biol Environ Stat* ,9: 181–199.

Inria. 2002. TAPENADE 2.1 user's guide. Online reference manual (unpublished). Available from http://www.inria.fr/rrrt/rt-0300.html.

International Whaling Commission. 2008. Report of the SOWER Abundance Workshop. SC/60Rep4 presented to the IWC Scientific Committee, June 2008 (unpublished). 17pp.

International Whaling Commission. 2009. Report of the SOWER Abundance Workshop, St Andrews, April 2009. SC/60Rep? presented to the IWC Scientific Committee, June 2009 (unpublished). ??pp.

Laake, J.L. 1999. Distance sampling with independent observers: reducing bias from heterogeneity by weakening the conditional independence assumption. Pp. 137-148. *In:* G.W. Garner, S.C. Amstrup, J.L. Laake, B.F.J. Manly, L.L. McDonald and D.G. Robertson (eds.) *Marine Mammal Survey and Assessment methods.* Balkema, Rotterdam. 287pp.

Laake, J.L. and Borchers, D.L. 2004. Methods for incomplete detection at distance zero. In Buckland, S.T., Anderson D.R., Burnham, K.P, Laake, J.L., Borchers, D.L. and Thomas, L. (eds) Advanced Distance Sampling. Oxford University Press, Oxford, pp.108-189.

Matsuoka, K., Ensor, P., Hakamada, T., Shimada, H., Nishiwaki, S., Kasamatsu, F. and Kato, H. 2003. Overview of minke whale sightings surveys conducted on IWC/IDCR and SOWER Antarctic cruises from 1978/79 to 2000/01. *J. Cetacean Res. Manage.* 5:173-201.

Mori, M., Butterworth, D.S., Brandao, A., Rademeyer, R.A., Okamura, H. and Matsuda, H. 2003. Observer experience and Antarctic minke whale sighting ability in IWC/IDCR-SOWER surveys. *J. Cetacean Res. Manage*. 5:1-11.

Okamura, H. and Kitakado, T. 2008. Abundance estimates of Antarctic minke whales from the historical IDCR/SOWER survey data using the OK Method. Paper SC/60/IA8 presented to the IWC Scientific Committee, June 2008 (unpublished). 19pp.

Okamura, H., Kitakado, T., Hiramatsu, K., and Mori, M. 2003. Abundance estimation of diving animals by the double-platform line transect method. *Biometrics* 59: 512-520.

Okamura, H., Kitakado, T., and Mori, M. 2005 An improved method for line transect sampling in Antarctic minke whale surveys. *J. Cetacean Res. Manage.* 7: 97-106.

Palka, D. 1995. Abundance estimate of the Gulf of Maine harbor porpoise. *Rep. Int. Whal. Commn.*, (special issue) 16:27-50.

Palka, D. 2009. Preliminary results of OK, BHWP, integrated and standard analytical methods when applied to simulated data, 2004 – 2008. Paper SC/A09/AE9 presented to the IWC Workshop on SOWER Abundance Estimates, St Andrews, April 2009 (unpublished) 11pp.

Palka, D.L., and Smith, D.W. 2004. Update of specifications of data simulating the IDCR/SOWER surveys – 2004. Paper SC/56/IA6 presented to the IWC Scientific Committee, July 2004 (unpublished). 15pp.

Palka, D.L., and Smith, D.W. 2005. Description of 2005 simulations of the IWC/SOWER Southern Hemisphere minke whale abundance surveys. Paper SC/57/IA2 presented to the IWC Scientific Committee, June 2005 (unpublished). 8pp.

R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Skaug, H.J. 2006. Markov modulated Poisson processes for clustered line transect data. *Environmental and Ecological Statistics*. 13: 199-211.

Skaug, H.J., Fournier D. 2006. Automatic Approximation of the Marginal Likelihood in non-Gaussian Hierarchical Models; *Computational Statistics and Data Analysis*. 51: 699-709

Skaug, H.J. and Schweder, T. 1999. Hazard models for line transect surveys with independent observers. *Biometrics* 55:29-36.

Wood, S.N. 2006. *Generalized Additive Models: An Introduction in R.* Chapman and Hall/CRC, Boca Raton. 416pp.

Wood, S.N., Bravington, M.V. and Hedley, S.L. 2008. Soap film smoothing. *J. R. Statist. Soc. B*, 70: 931-955.

Appendix 1

ALGEBRAIC AND ESTIMATION DETAILS FOR THE SPAMASSS MODEL COMPONENTS

There are three platforms in IDCR/SOWER: A (top/barrel); B (IOP) and C (front/upper bridge). In what follows, we denote 'seen' by Platform X in upper case, and 'not seen' in lower case, e.g. Ab denotes 'seen by A and not seen by B.'

In order to evaluate the SPAMASSS likelihood, a probability model for $\mathsf{P}[ys_e h \mid xzo]$ is needed. One way to break down the joint probability is as follows. As usual, when a variable appears inside braces on the RHS of a conditional probability, e.g. the z in $\mathsf{P}[s \mid \{z\}x]$, then that variable is irrelevant to the conditioning. Note also that, if o doesn't appear on the RHS of the conditional probability, then the implication is that the corresponding distribution is the underlying true distribution, not what was observed. Note further that $\mathsf{P}[o \mid h \cdots] \equiv 1$. Finally note that we assume, without loss of generality, that the truncation distance is 1, and thus neglect the factor dy which would otherwise appear in many of the equations below. Thus it is reasonable to write $\mathsf{P}[y \mid \cdots] = 1 \;\; \forall \; y$ provided that the … do not include h or o.

$$
\begin{aligned}
\mathsf{P}[ys_e h \mid xzo] &= \sum\nolimits_s \mathsf{P}[ys_e h \mid \{x\} zos]\mathsf{P}[s \mid xzo] \\
\mathsf{P}[s \mid xzo] &= \frac{\mathsf{P}[o \mid s\{x\}z]\mathsf{P}[s \mid \{z\}x]}{\mathsf{P}[o \mid xz]} \\
&= \frac{\mathsf{P}[o \mid sz]\mathsf{P}[s \mid x]}{\sum_{s'} \mathsf{P}[o \mid sz]\mathsf{P}[s \mid x]} \\
\mathsf{P}[ys_e h \mid zos] &= \mathsf{P}[s_e \mid yzh\{o\}s]\mathsf{P}[hy \mid ozs] \\
\mathsf{P}[hy \mid ozs] &= \frac{\mathsf{P}[o \mid hyzs]\mathsf{P}[hy \mid sz]}{\mathsf{P}[o \mid sz]} \\
&= \frac{1 \times \mathsf{P}[h \mid ysz]\mathsf{P}[y \mid sz]}{\mathsf{P}[o \mid sz]} \\
&= \frac{\mathsf{P}[h \mid ysz]}{\mathsf{P}[o \mid sz]}
\end{aligned}
$$

Combining everything:

$$
\begin{aligned}
\mathsf{P}[ys_e h \mid xzo] &= \sum\nolimits_s \mathsf{P}[s_e \mid yzhs]\frac{\mathsf{P}[h \mid ysz]}{\mathsf{P}[o \mid sz]}\frac{\mathsf{P}[o \mid sz]\mathsf{P}[s \mid x]}{\sum_{s'} \mathsf{P}[o \mid s'z]\mathsf{P}[s' \mid x]} \\
&= \frac{\sum_s \mathsf{P}[s_e \mid yzhs]\mathsf{P}[h \mid ysz]\mathsf{P}[s \mid x]}{\sum_s \mathsf{P}[o \mid sz]\mathsf{P}[s \mid x]}
\end{aligned}
$$

For the denominator, note that $\mathsf{P}[o \mid sz] = \sum_h \mathsf{P}[h \mid sz] = \sum_h \int \mathsf{P}[h \mid ysz]dy$, i.e. it uses the same things already occurring in the numerator.

The modification to Closing mode is pretty simple. We don't need to worry about mis-underestimation, and there is only one possible platform combination $(A \cup C)$, so

$$
\mathsf{P}[ys \mid A \cup C] = \frac{\mathsf{P}[A \cup C \mid ysz]\mathsf{P}[s \mid x]}{\sum_s \mathsf{P}[A \cup C \mid sz]\mathsf{P}[s \mid x]}
$$

Therefore, the submodels required are:

- a school size error model for mis-underestimation $\mathsf{P}[s_e \mid yzhs]$
- a spatial school size model $\mathsf{P}[s \mid x]$
- a probability of detection model $\mathsf{P}[h \mid ysz]$, i.e. $\mathsf{P}[AB \mid ysz]$, $\mathsf{P}[Ab \mid ysz]$, $\mathsf{P}[aB \mid ysz]$, $\mathsf{P}[abC \mid ysz]$, $\mathsf{P}[A \cup C \mid ysz]$.

TWO PARAMETRIZATIONS OF THE PROBABILITY OF DETECTION MODEL $\mathsf{P}[h \mid ysz]$ FOR IDCR/SOWER PLATFORMS

The key here is to find some basic set of functions that allow us to reconstruct all the $\mathsf{P}[h \mid ysz]$, plus the C-mode case $\mathsf{P}[A \cup C \mid ysz]$, in a way that is guaranteed to be consistent with the laws of probability, and can be expected to have sensible underlying shapes. The following set suffice (all potentially depending on s and z):

- $g_0(A)$, $g_0(B)$, $g_0(C)$ as single parameters
- $\mathsf{P}[y \mid A \cup B \cup C]$, $\mathsf{P}[A \mid By]$, $\mathsf{P}[B \mid Ay]$ – all standard distance-sampling hazard-rate functions

- $\mathsf{P}[A \cup B \mid A \cup B \cup C, y]$, logit $\mathsf{P}[C \mid aBy]$ – logit $\mathsf{P}[C \mid aby]$ (shape only – value at $y = 0$ is set by the g_0 s and the TLCI assumption)

The last four are all conditional-probability functions. It is not obvious *a priori* whether they should increase or decrease with y. However, a rough look at the IDCR/SOWER data (this has to be done for all school sizes together, because true s is unknown) does give moderately clear suggestions for which way each one goes, except for the last which can only be inferred indirectly from the difference between the shapes of the distance functions for IO and Closing mode. But, whether the curve rises or falls, the functional form must ensure that the conditional probabilities stay between 0 and 1, without necessarily tending to either value, and should have a flat section near $y = 0$ like a hazard probability function does. In fact, we have represented them as scaled and shifted hazard functions. Their value at $y = 0$ is set by the g_0 s and the assumption of independence on the trackline. As y increases from 0, they are assumed to start flat and then to rise (or fall) to another asymptote somewhere between 0 and 1. It therefore takes three further parameters apart from the g_0 s to describe each term: the asymptote as $y \to \infty$, the turning point when the curve changes from leaving one asymptote to approaching another, and the abruptness of the transition. Specifically, each curve has a shape parameter b, a scale parameter s, and a far-distance asymptote g_∞ (a descriptive but somewhat misleading name). In total, 17 parameters are required to describe the distance-sampling functions for given $\{s, z\}$, although some of the 17 might well be shared across $\{s, z\}$ strata. For example, there is actually very little information to estimate the g_∞ parameters of $\mathsf{P}[C \mid aBy]$, i.e. they have minimal effect on fit; so we might as well keep them fixed across z and/or s.

It is clear *a priori* that the g_0 s, $\mathsf{P}[y \mid A \cup B \cup C]$, $\mathsf{P}[A \mid By]$ and $\mathsf{P}[B \mid Ay]$ should all increase with both s and z for fixed y, and this helps in deciding on good models to use. It is not clear how either $\mathsf{P}[A \cup B \mid A \cup B \cup C, y]$ or $\mathsf{P}[C \mid aBy] / \mathsf{P}[C \mid aby]$ should behave with increasing s or z. However, these functions have much less effect on the estimated ESWs and are much harder to estimate because they are based on indirect comparisons, so it is reasonable to use much simpler formulae to describe them.

An alternative parametrization, which avoids estimation of the awkward conditional probabilities associated with Platform C (i.e. particularly the $\mathsf{P}[C \mid aBy]$ part), may be derived by instead making the assumption that

$$\text{logit}(\mathsf{P}[C \mid aBy])\text{-logit}(\mathsf{P}[C \mid aby])=\text{logit}(\mathsf{P}[A \mid By])\text{-logit}(\mathsf{P}[A \mid by]).$$

The implication is that B has the same effect (on the logit scale) on $\mathsf{P}[C \mid ay]$, for which we have little data, as on $\mathsf{P}[A \mid y]$, for which we have good data.

A MODEL FOR MIS-UNDERESTIMATION OF SCHOOL SIZES

Previous incarnations of the SPLINTR model (e.g. the BHWP model; Bravington, 2008) used a parameter-hungry, constrained smooth polytomous regression to model errors in school size. Here we describe a substantially simpler form of model, which requires estimation of a school size bias parameter, b, which can depend on true school size, s, and sighting conditions, z.

As pointed out by Cooke (2009), a simple multiplicative bias model for school size estimation would be inappropriate because an observed school size cannot be less than 1. We compare the following three potential distributional assumptions:

i. $(S_e - 1) \mid s, o \sim$ Bin (s-1,b), i.e. the negative Binomial (as used in the IM; Cooke, 2009);

ii. $(S_e - 1) \mid s, o \sim$ Poi (s-1,b), i.e. the Poisson (as used in the OK model; Okamura and Kitakado, 2008);

iii. $\mathsf{P}[S_e \mid s, o] =$ cumWeib(S_e,bs,γ)- cumWeib(S_e-1,bs,γ) where cumWeib is the CDF of a Weibull distribution with scale parameter bs and, controlling the degree of skewness in the tail, the shape parameter γ.

The Poisson assumption allows over-estimation for s>1, but neither the Poisson nor the negative Binomial model permit over-estimation when the true school size is 1. The Weibull form given above, however, does allow this possibility.

Initial inspection of a portion of the SSII experimental data, reveals that of 24 sightings of true school size 1, three were initially over-estimated (as of size 2). Given the IDCR/SOWER protocols of recording the number of whales in a school that were seen, rather than the number that might have been there, the most plausible explanation for over-estimation is that the sighted school split during the closure attempt (P.H. Ensor, pers. comm.) Nonetheless, it seems reasonable to conclude that such over-estimation is a rather rare event in IDCR/SOWER, and thus all three of the distributional assumptions above should therefore be reasonable. However, for purely pragmatic purposes of using the same school size error model for both the real data and the IWC simulated data (Palka and Smith, 2004; 2005), the Weibull form is used in this paper since over-estimation is a common feature in the simulation scenarios to date.

As implemented, the parameter b is allowed to vary by school size and conditions, the parameter γ is not; the latter is simply fitted separately for each data set (year). The model operates on finely-categorized school sizes, i.e. separately for schools of size 1, 2, 3,…, 10, and then for each of the groupings 11-12, 15-19 and 20+.

SPATIAL SCHOOL SIZE MODEL $\mathsf{P}[S = s \mid x]$

The general idea is a smooth polytomous regression, whereby

$$\mathrm{P}[S = s \mid x] = \Phi\left(\zeta_s^{SS} - X\beta^{SS}\right) - \Phi\left(\zeta_{s-1}^{SS} - X\beta^{SS}\right)$$

where $\zeta_0^{SS} \equiv -\infty$, $\zeta_{s_{\max}}^{SS} = \infty$, the remaining ζ^{SS} s are increasing parameters to be estimated, X is a design matrix of a soap-film smoother (Wood *et al.*, 2008), and β^{SS} is a parameter vector to be estimated. There is also a penalty term of the form $-\lambda^{SS}\beta^{SS\mathsf{T}}S\beta^{SS}$ which needs to be added on to the log-likelihood; the matrix S is pre-specified, and the two elements of λ^{SS} are estimated later by a Residual Maximum Likelihood (REML)-like Laplace approximation.

SPAMASSS ESTIMATION

The coefficients β^{SS} are random effects, and need to be integrated out in order to compute a proper likelihood for all the other parameters ω (g0-related, relationships between shape and scale parameters and covariates, mis-underestimation, smoothing parameters). This is done by approximate REML, first using an inner optimization to find the best-fitting spatial parameters $\hat{\beta}_\omega^{SS}$ for a given ω, and then using a Laplace approximation based on the joint log-likelihood $\Lambda\left(\omega, \beta^{SS}\right)$:

$$\tilde{\Lambda}(\omega) \approx \Lambda\left(\omega, \hat{\beta}_\omega^{SS}\right) - \frac{1}{2}\left\|\frac{d^2\Lambda}{d\beta^2}\right\| + \frac{1}{2}\left\|S_\omega\right\| + c$$

where S_ω is the quadratic smoothness penalty on β^{SS}. Various Automatic Differentiation tricks are used to compute $\tilde{\Lambda}(\omega)$ and $d\tilde{\Lambda} / d\omega$, along the lines of Skaug and Fournier (2006).

Table 1. Estimates of G0 and ESW by conditions, platform, and CP series

(a) CP2, Beaufort (Good=0-2, Bad=3+)

		1	2	3-4	5-9	10+
Good	g0A	0.27	0.56	0.69	0.72	0.81
	g0B	0.13	0.34	0.48	0.50	0.64
	g0C	0.32	0.60	0.72	0.74	0.83
	ESW	0.33	0.69	0.89	1.03	1.30

		1	2	3-4	5-9	10+
Bad	g0A	0.26	0.39	0.62	0.65	0.74
	g0B	0.12	0.20	0.39	0.43	0.53
	g0C	0.31	0.44	0.65	0.68	0.76
	ESW	0.29	0.46	0.65	0.69	1.04

(b) CP3, Sightability

		1	2	3-4	5-9	10+
Sig4+	g0A	0.38	0.54	0.69	0.72	0.80
	g0B	0.21	0.32	0.48	0.51	0.64
	g0C	0.44	0.59	0.73	0.76	0.85
	ESW	0.52	0.78	0.99	1.02	1.38

		1	2	3-4	5-9	10+
Sig3	g0A	0.31	0.39	0.67	0.71	0.79
	g0B	0.16	0.21	0.46	0.50	0.62
	g0C	0.37	0.45	0.71	0.75	0.84
	ESW	0.41	0.64	0.85	0.91	1.33

		1	2	3-4	5-9	10+
Sig2	g0A	0.12	0.37	0.51	0.57	0.62
	g0B	0.06	0.20	0.30	0.36	0.41
	g0C	0.15	0.43	0.57	0.63	0.68
	ESW	0.16	0.62	0.76	0.81	0.98

Table 2. Estimates by AVB block, without Common Northern Boundary. No process error

	1986	1987	1988	1989	1990	1991	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
1W					24,200										18,300			
					0.41										0.24			
1M					63,600			30,500										
					0.23			0.19										
1E					29,100			35,000						6,350				
					0.26			0.22						0.32				
2W		26,400										19,100						
		0.23										0.21						
2E		115,000									32,100							
		0.26									0.21							
3W			80,600				30,600											
			0.54				0.15											
3E			6,560				115		25,000									
			0.50				0.69		0.18									
4W				23,400					10,700									
				0.18					0.20									
4E				37,200									25,300					
				0.24									0.30					
5W	41,000															13,600		
	0.38															0.21		
5M	80,200															31,800		
	0.15															0.22		
5E	161,000															164,000		
	0.21															0.10		
6W						25,500				35,700								
						0.24				0.24								
6E						33,900									18,300			
						0.36									0.24			

Table 3. Estimates by AVB block, with Common Northern Boundary (no process error version)

	1986	1987	1988	1989	1990	1991	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
1W					17,000										17,600			
					0.27										0.25			
1M					48,000			26,600										
					0.16			0.19										
1E					20,000			32,200						4,120				
					0.19			0.21						0.33				
2W		25,500										18,900						
		0.22										0.21						
2E		99,200									26,900							
		0.17									0.20							
3W			58,900				18,300											
			0.37				0.14											
3E			5,890				18		24,100									
			0.52				0.74		0.18									
4W				23,600					10,200									
				0.18					0.20									
4E				37,000									25,000					
				0.23									0.29					
5W	38,700																13,900	
	0.39																0.21	
5M	77,600																31,900	
	0.15																0.22	
5E	161,000																164,000	
	0.21																0.10	
6W						26,100				36,200								
						0.24				0.23								
6E						27,100									17,600			
						0.31									0.25			

Table 4. Estimates by original SOWER stratum. Area in n.mi2; P=abund est; E[S]=mean true school size; esw_w = P[seeing a randomly-chosen whale within 1.5nmi of transect, times 1.5nmi]

			Area	P	E[S]	esw_w				Area	P	E[S]	esw_w
A1 CP2	1990	EN	155,000	22,300	1.67	0.41	A4 CP2	1989	BN	17,600	9,530	2.56	0.39
		ESBA	62,600	14,200	1.89	0.39			BS	6,930	5,760	2.68	0.41
		WN	166,000	29,600	1.73	0.36			EN	181,000	13,300	1.52	0.39
		WS	43,900	17,300	2.03	0.42			ES	52,300	10,500	2.01	0.44
CP3	1994	EN	297,000	17,600	1.27	0.73			WN	157,000	7,380	2.05	0.38
		ES	71,100	23,000	1.74	0.54			WS	58,500	10,800	2.01	0.36
		WN	249,000	7,800	1.12	0.39	CP3	1995	EN	149,000	5,940	1.21	0.66
		WS	52,200	14,200	1.58	0.59			PRYD	20,000	3,930	1.29	0.57
	2000	EN	59,000	2,540	1.63	0.66		1999	EN	171,000	6,480	1.10	0.59
		ES	24,600	1,120	1.67	0.56			ES	69,800	5,660	1.25	0.51
		WN	110,000	2,170	1.53	0.57			WN	107,000	6,780	1.29	0.44
		WS	20,500	509	1.41	0.73			WS	42,100	6,580	1.79	0.51
	2001	EN	127,000	2,580	1.79	0.60							
		ES	29,400	3,620	2.98	0.51	A5 CP2	1986	EM	162,000	70,000	1.78	0.38
									EN	277,000	54,900	1.59	0.39
A2 CP2	1987	EBAY	13,600	8,350	2.30	0.69			ES	107,000	51,400	2.37	0.39
		EM	69,400	24,200	3.24	0.46			WM	168,000	46,400	1.84	0.37
		EN	124,000	38,200	3.25	0.48			WN	140,000	22,900	1.49	0.40
		ES1	22,800	5,650	2.11	0.47			WS	98,500	30,700	1.86	0.42
		ES2	42,600	16,100	2.48	0.57	CP3	2002	EN	82,700	2,750	1.33	0.57
		WBAY	9,550	2,700	2.32	0.42			ES	25,700	4,120	1.36	0.69
		WN	94,000	9,920	1.62	0.49			ESA	10,500	1,250	1.64	-
		WS1	8,450	2,560	2.35	0.46			WN	47,600	1,520	1.45	0.57
		WS2	21,300	1,850	1.78	0.52			WS	35,300	5,250	1.62	0.71
		WS3	76,400	13,800	1.73	0.56		2003	EN	137,000	6,090	1.22	0.57
CP3	1997	EN	243,000	17,200	1.35	0.49			ES	127,000	7,640	1.34	0.69
		ES	51,700	3,390	1.63	0.57			W1N	75,900	9,960	1.42	0.38
		WN	113,000	9,180	1.42	0.51			W1S	22,200	3,910	2.22	0.62
		WS	23,800	3,960	1.64	0.59			W2N	100,000	11,300	1.41	0.74
	1998	EN1	85,300	4,550	1.25	0.51			W2S	21,700	4,410	1.98	0.66
		EN2	80,200	3,860	1.34	0.57		2004	MID	132,000	122,000	2.32	0.67
		ES1	46,700	6,870	1.70	0.54			N1	123,000	9,040	1.29	0.41
		ES2	10,200	1,620	1.77	0.64			N2	96,300	11,600	1.34	0.56
		WN	53,100	1,160	1.39	0.52			N3	14,400	2,950	1.87	0.66
		WS	33,300	816	1.50	0.60			ROSS	55,500	20,600	1.53	0.54
	2000	ENA	7,000	218	1.48	0.54							
		ESA	6,220	213	1.59	0.57	A6 CP2	1991	EN	192,000	22,700	1.75	0.32
									ES	108,000	9,130	1.47	0.37
A3 CP2	1988	EN	168,000	4,720	1.32	0.41			WN	214,000	15,200	1.73	0.37
		ES	88,300	3,420	1.37	-			WS	44,400	5,400	2.07	0.35
		WN	150,000	29,600	2.36	-	CP3	1996	EN	242,000	19,400	1.60	0.63
		WS	71,200	26,000	2.21	0.43			ES	72,600	5,990	1.29	0.44
CP3	1993	EN	149,000	3,540	1.69	0.51			WN	97,900	9,050	1.44	0.43
		ES	22,700	1,510	2.20	0.64			WS	34,400	1,830	1.56	0.62
		WN	207,000	13,500	1.78	0.71		2001	WN	252,000	13,200	1.49	0.57
		WS	61,100	11,600	1.77	0.60			WS	43,900	5,720	1.74	0.63
	1995	ES	59,400	13,600	1.58	0.54							
		WN	149,000	7,040	1.28	0.55							
		WS	52,700	5,240	1.65	0.67							

Table 5. ``Survey-once'' estimates, without CNB

		MA1	MA2	MA3	MA4	MA5	MA6	Total
CP2	abund	117,000	141,000	87,200	60,700	282,000	59,400	747,000
	CV	0.22	0.23	0.51	0.18	0.15	0.25	0.13
CP3	abund	55,200	51,200	55,800	36,000	209,000	54,000	461,000
	CV	0.15	0.17	0.13	0.23	0.10	0.18	0.09

Table 6. Comparisons of observed & predicted pre-closure school size estimates in SSE2&3, 2007 & 2008

(a) CP2 by True SS

		1	2	3-4	5-9	10+
1	expo	63.00	0.00	0.00	0.00	0.00
	obso	60	3	0	0	0
2	expo	23.81	18.19	0.00	0.00	0.00
	obso	23	18	1	0	0
3-4	expo	7.65	13.98	9.37	0.00	0.00
	obso	7	10	12	2	0
5-9	expo	1.30	4.58	11.45	2.67	0.00
	obso	1	6	9	4	0
10+	expo	0.01	0.10	1.20	5.42	2.27
	obso	0	0	2	3	4

(b) CP2 by Beaufort

		1	2	3-4	5-9	10+
Good	expo	25.43	13.20	9.51	3.30	0.56
	obso	24	9	12	6	1
Bad	expo	70.32	23.66	12.51	4.79	1.72
	obso	67	28	12	3	3

(c) CP3 by True SS

		1	2	3-4	5-9	10+
1	expo	63.00	0.00	0.00	0.00	0.00
	obso	60	3	0	0	0
2	expo	17.92	24.08	0.00	0.00	0.00
	obso	23	18	1	0	0
3-4	expo	4.14	12.14	14.71	0.00	0.00
	obso	7	10	12	2	0
5-9	expo	0.48	2.33	10.60	6.59	0.00
	obso	1	6	9	4	0
10+	expo	0.02	0.15	1.44	4.36	3.03
	obso	0	0	2	3	4

(d) CP3 by Sightability

		1	2	3-4	5-9	10+
S4+	expo	29.13	13.15	12.18	5.52	3.02
	obso	33	9	11	6	4
S2-3	expo	56.44	25.55	14.56	5.43	0.01
	obso	58	28	13	3	0

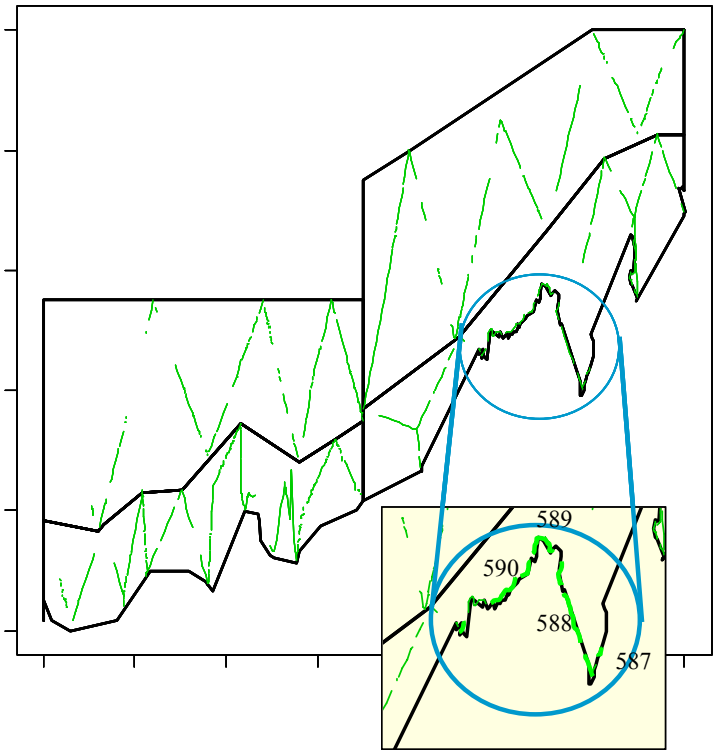


Figure 1 Transects covered on effort during 1989/90 survey of Area I. Highlighted section shows transects which tracked the ice edge, and could lead to bias from a design-based estimate. Numbers are transect labels.

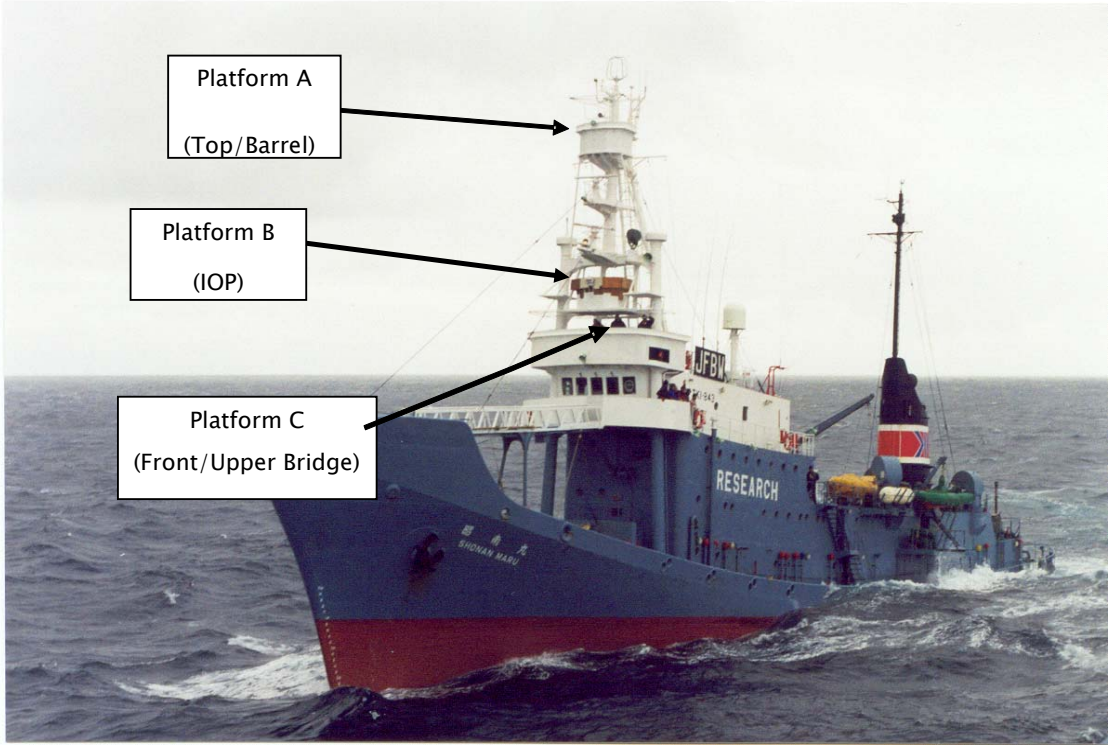


Figure 2 Platform locations on the Shonan Maru.

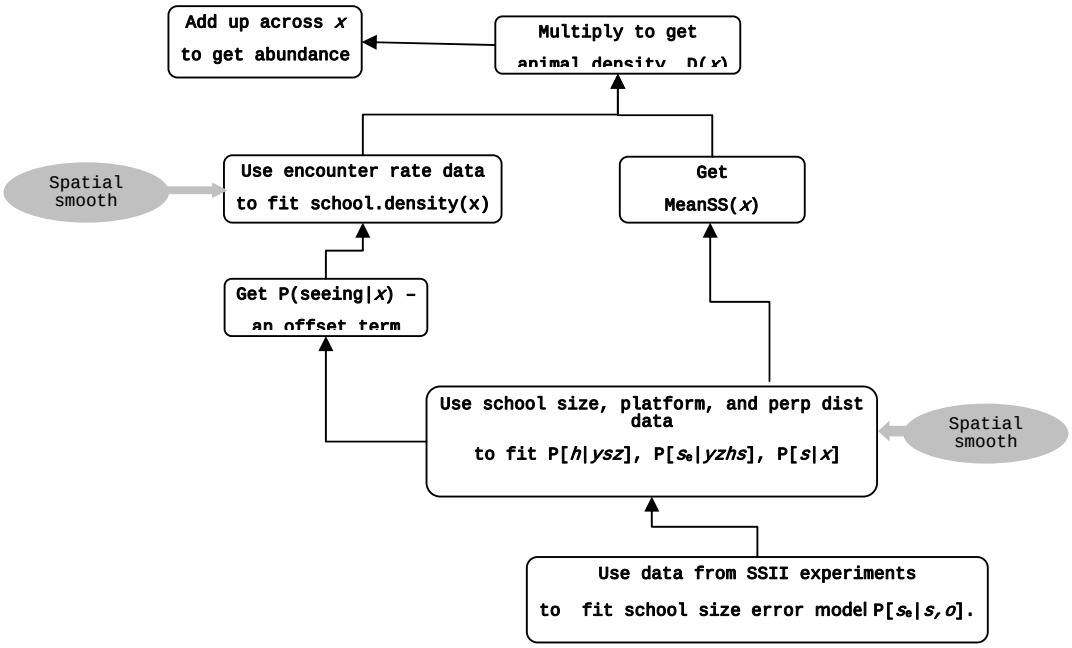


Figure 3 Schematic structure of SPLINTR model.

Figure 4(a). Spatial density estimates for CP2. 1 dot = 10 whales.

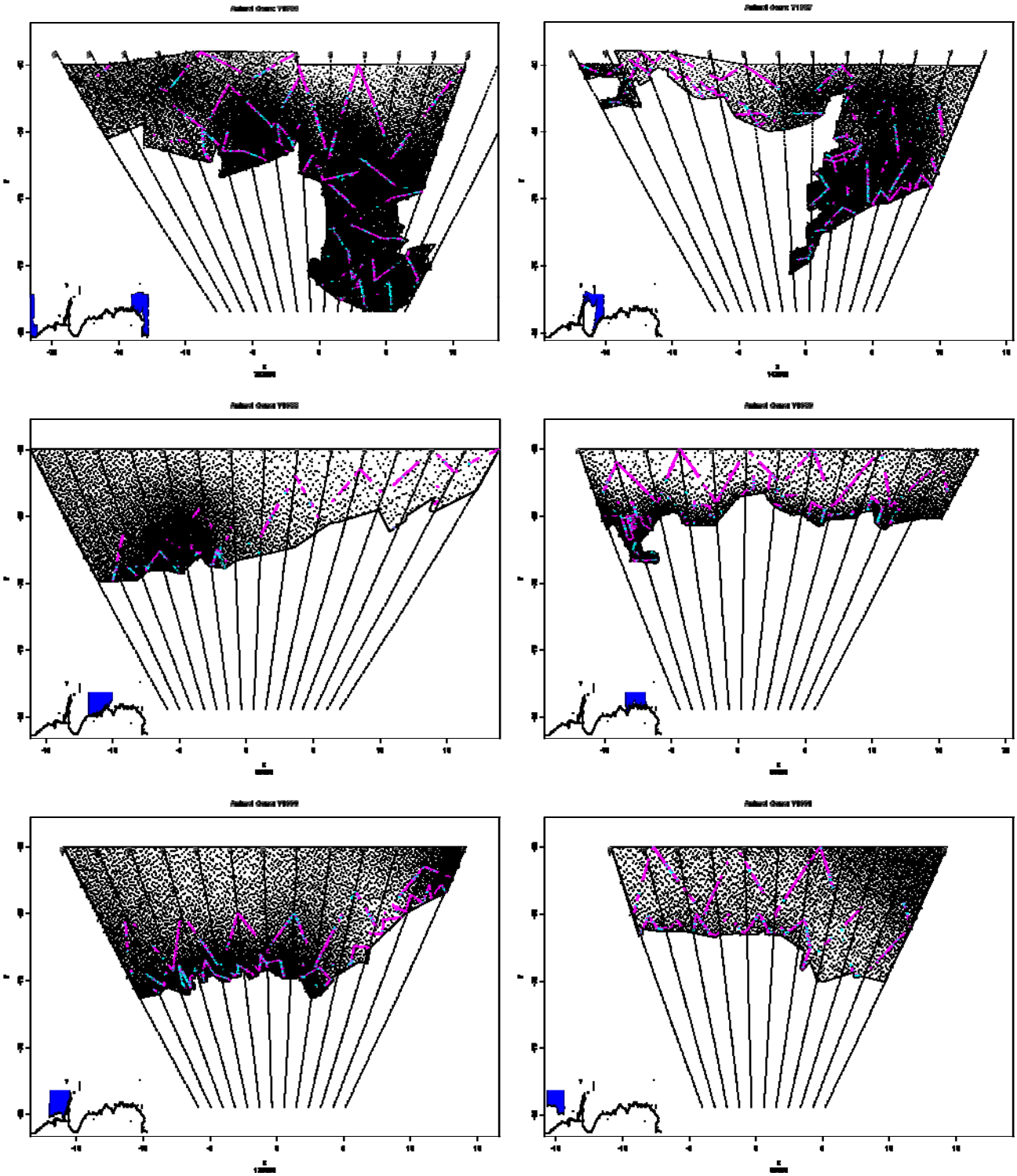
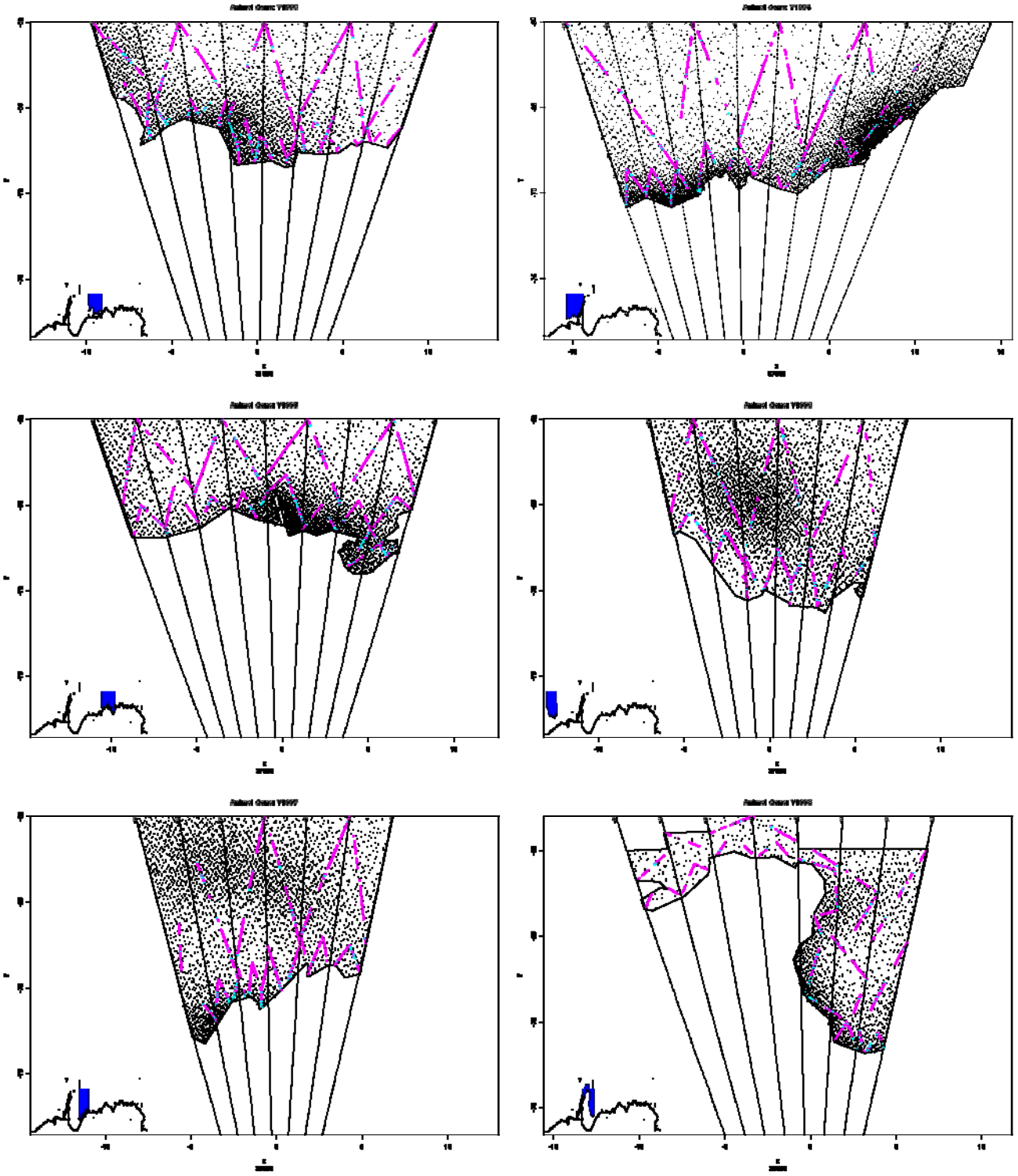


Figure 4(b). Spatial density estimates for CP3. 1 dot = 10 whales.



[Figure 4(b) cont.]

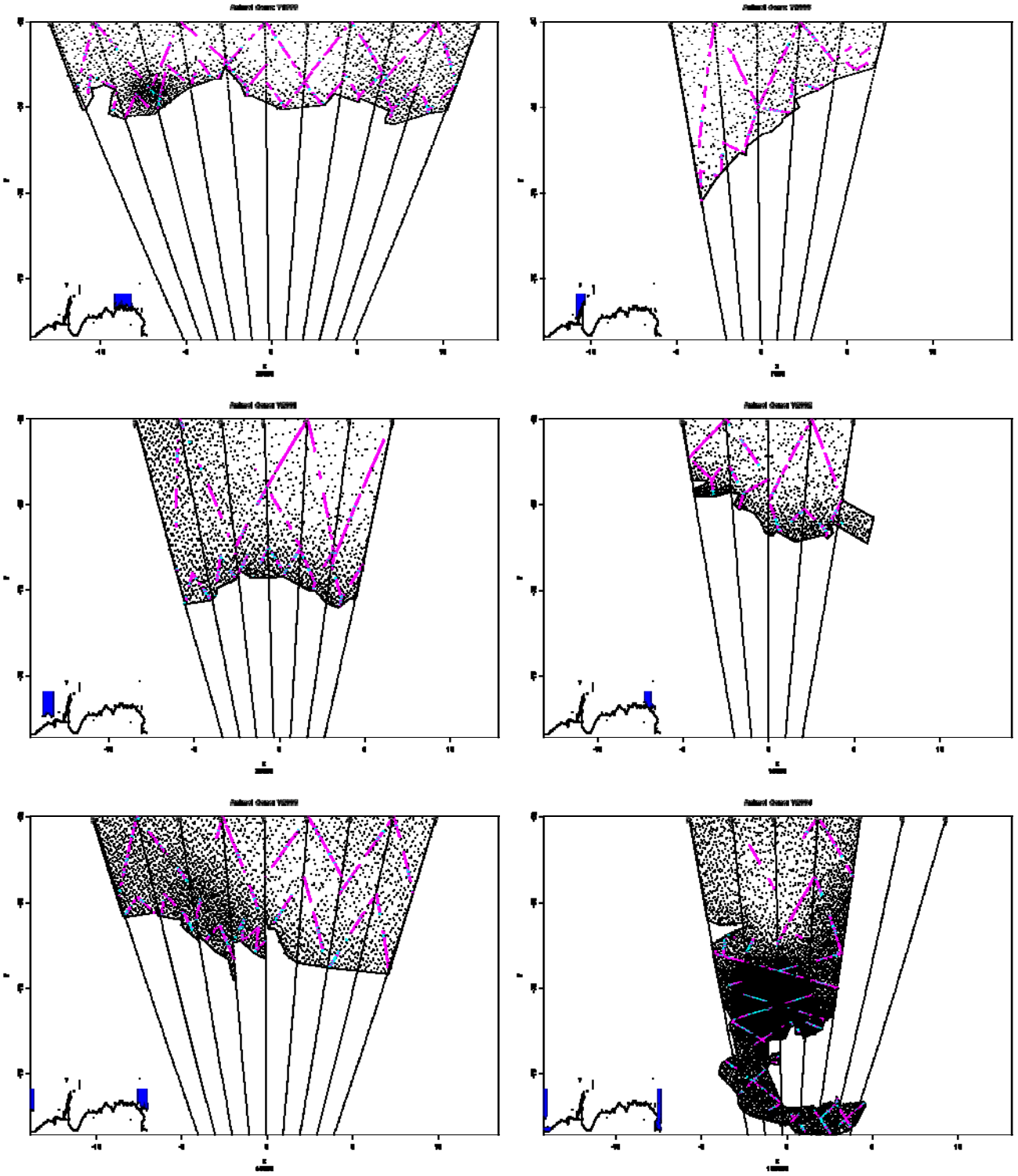


Figure 5. (D#8) SS class(1, 2, 3-4, 5+) and perp dist in CL mode, CP2 & CP3

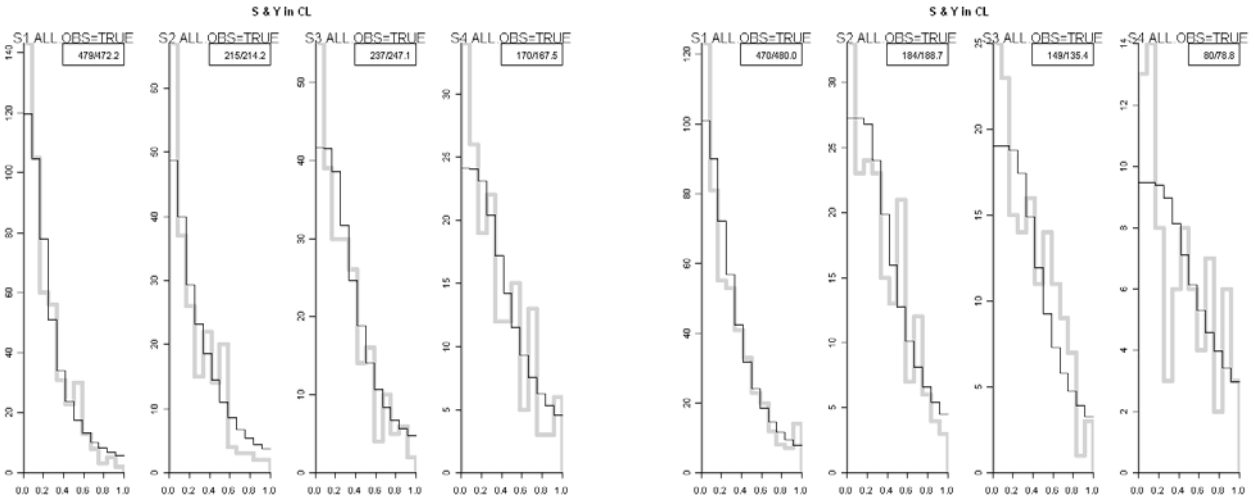


Figure 6. (D#6) Perp dist and platform by Vessel, CP2 & CP3

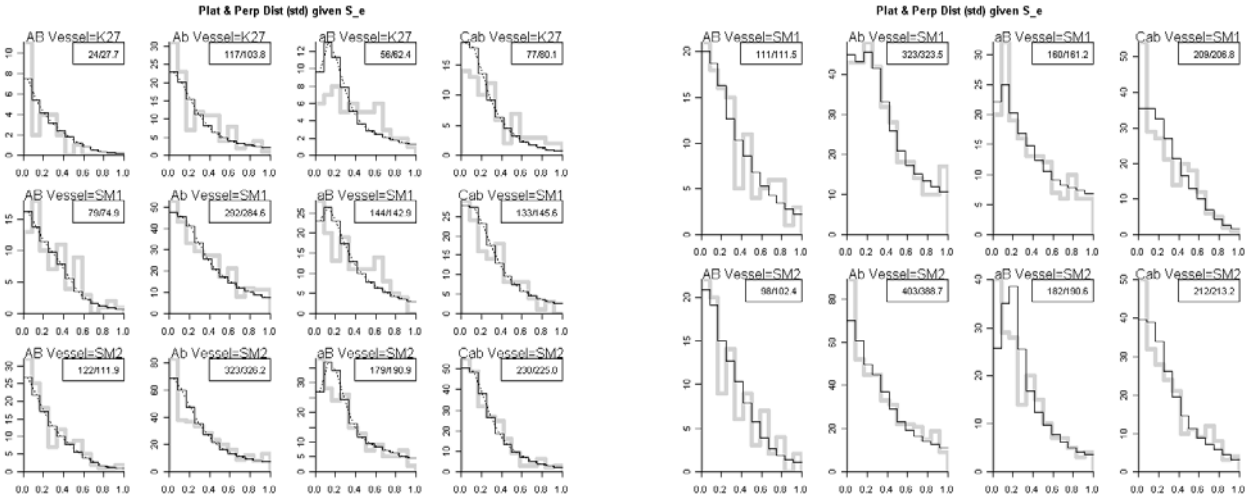


Figure 7. (D#7) Perp dist in IO mode, given observed SS, CP2 and CP3

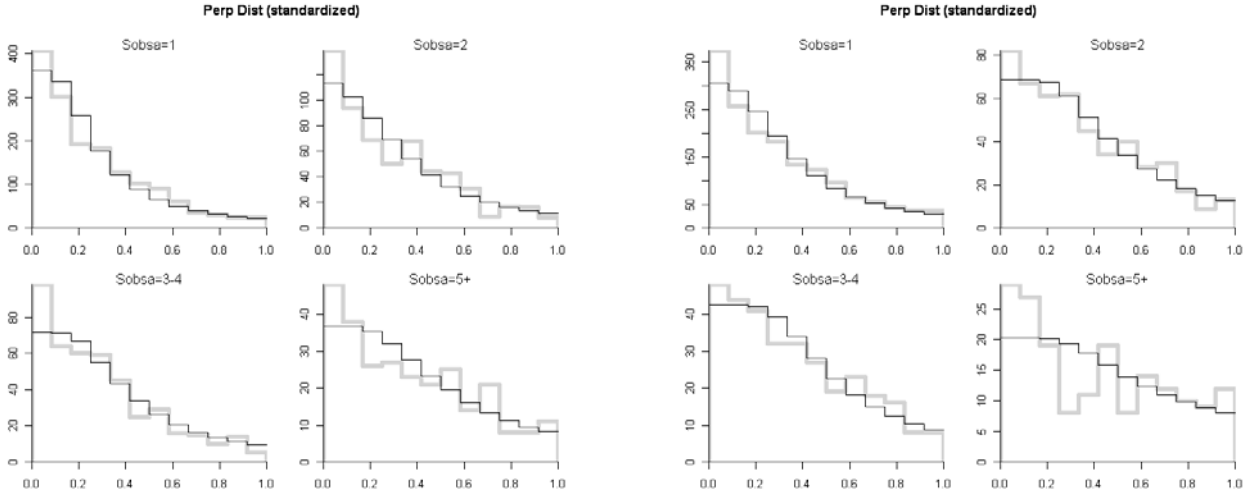


Figure 8. (D#11) Proportions of IO-mode school sizes by Vessel: CP2 and CP3

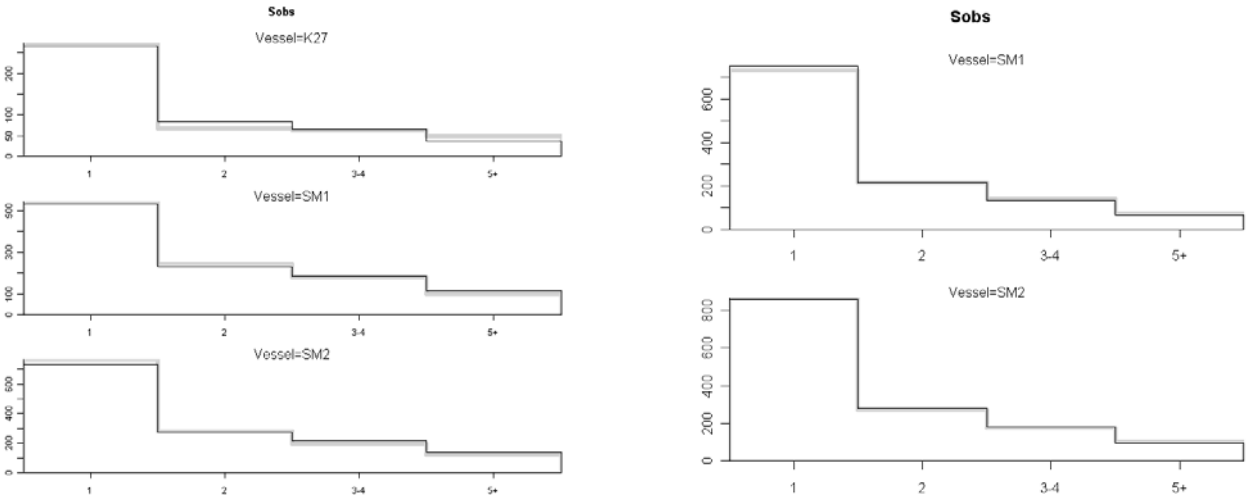
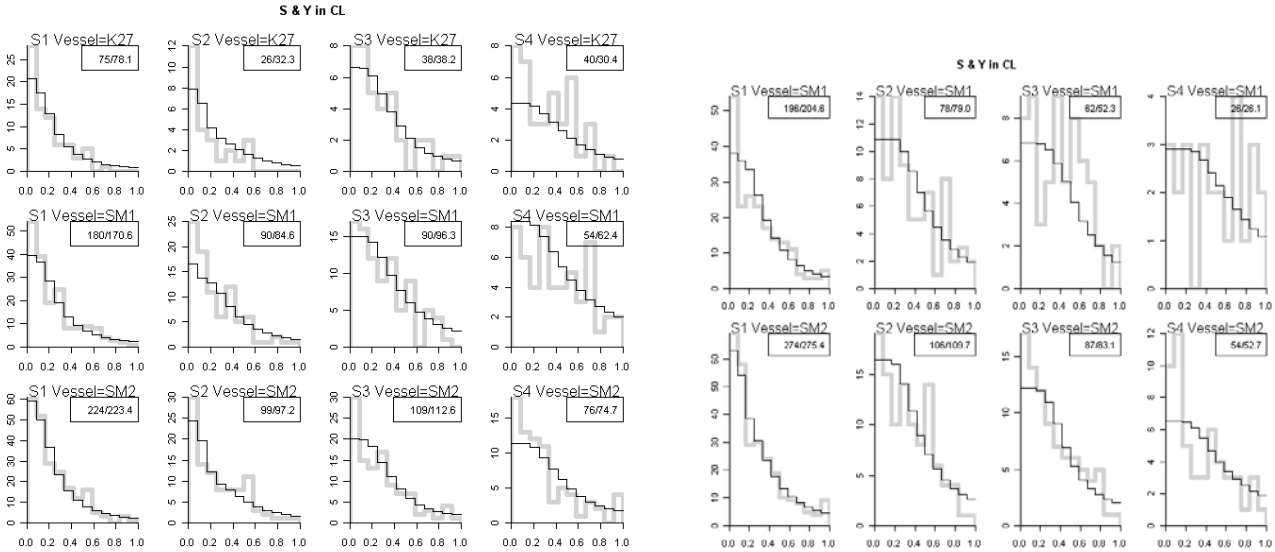



Figure 9. (D#12) SS (and perp dist) by Vessel in CL mode, CP2 and CP3