Report of the SOWER Abundance Workshop*

Members: Borchers, Bravington, Buckland, Burt, Cooke, Hedley, Kitakado, Palka (Chair).

1. OPENING REMARKS AND WELCOMING ADDRESS

The meeting was held in St. Andrews, Scotland from 7-10 April 2009 at the Centre for Research into Ecological and Environmental Modelling, University of St Andrews. The convenor, Palka, welcomed all the participants and thanked Burt and Hedley for hosting this workshop. She also thanked Borchers and Buckland for returning to the IWC SC and helping to review and improve the three new analytical methods being developed. The Terms of Reference were reviewed, which were to facilitate the completion of the abundance estimates of Antarctic minke whales using the IDCR/SOWER data collected during the three circumpolar series, CPI, CPII and CPIII, where the emphasis is on the two most recent CP series.

The Workshop noted that Okamura and Skaug contributed workshop papers but were unable to attend. Skaug joined the Workshop via conference phone call to discuss one of the papers.

Palka was elected Chair.

2. ADOPTION OF AGENDA

The agenda is given as Annex A.

3. REVIEW OF AVAILABLE DOCUMENTS

A list of available documents appears as Annex B.

4. METHODS USED TO ESTIMATE ABUNDANCE

The IDCR/SOWER survey data are being analysed by four analytical methods:

(1) the 'standard' method (Branch, 2006);
(2) a hazard probability method developed by Okamura and Kitakado (2008; SC/A09/AE1; AE3; AE4), referred to as the OK method;
(3) an integrated hazard probability method developed by Cooke (2008; SC/A09/AE5; AE6), referred to as the integrated model (IM) method; and
(4) a spatial point independence method developed by Bravington, Hedley, Wood and Peel (2008; SC/A09/AE2), referred to as the SPLINTR method.

Before the methods were discussed, Bravington provided an overview of the characteristics and challenges of the IDCR/SOWER Antarctic minke whale abundance data. Detailed specifications of each analytical method are summarised below.

4.1 Standard method

Okamura and Kitakado (2008) presented estimates of Antarctic minke whale abundance from the IDCR/SOWER surveys conducted between 1978/79 and 2003/04, grouped into three circumpolar sets of surveys, CPI, CPII and CPIII. Abundance was estimated using the IWC 'standard' line transect methodology. Some of the estimation options in this paper were different from those previously adopted in Branch and Butterworth (2001). Using these modifications, the circumpolar abundance estimates were: 645,000 (CV=0.143) for CPI; 786,000 (CV=0.094) for CPII and 338,000 (CV=0.079) for CPIII. When adjusted for comparable areas and when 'like-minkes' were included, the ratio of CPI to CPII to CPIII was 0.97:1.00:0.39. The CPIII:CPII ratio for individual IWC *Management Areas* was also low, ranging between 0.18 and 0.52 except for Area VI where the ratio was 1.59.

The standard method has been discussed in detail during previous Scientific Committee meetings, so this Workshop did not discuss this method, except as compared with the other methods.

4.2 OK method

The Okamura and Kitakado (OK) method is a type of hazard probability model developed for the North Atlantic minke whales and it was extended to deal with the school size error problems, semi-independent platform, and the measurement errors of timings in recording. Cue production is defined as a Poisson process linked to the logarithm of school size. The detection function is estimated separately for circumpolar series and Management Area. The detection function is modeled with the radial distance and angle, and it is dependent on true school size, weather condition on the Beaufort scale, vessel and platform. The school size is estimated separately for circumpolar series and survey strata (West and East). The distribution of school size is assumed to be negative binomial, in which the mean parameter is linked to the distance from the ice edge. School size is also categorised as 'confirmed' and 'unconfirmed' status. 'Confirmed' school sizes are assumed to be the true school sizes. 'Unconfirmed' school sizes are dealt with as biased estimates of the true school sizes. The confirmation probability of school size is related to true school size, sighting distance, and weather condition, where the probabilities in Passing and Closing modes have separate parameters. The likelihood function of the OK method is based on the joint probability for duplicate status, group size confirmation status, and observations of school size as well as distances to detected animals. Abundance estimation is based on a Horvitz-Thompson-like estimator. The variance estimation in encounter rate is based on an empirical estimator using replicate lines, and that for other parameters are derived from the Fisher information matrix. Spatial issues are dealt with by stratification (in CP series and Management Areas (MA)) and distance from ice-edge for the school size and by stratification (in CP and strata) for the detection function.

*Presented to the meeting as SC/61/Rep9.

The method described above is a modification of the former OK model (Okamura and Kitakado, 2008), where the modifications are based on suggestions made in SC/60. During SC/60, the points raised were: (i) examine the use of the hazard function of radial distance and angle (say $Q(r,\theta)$) instead of that of forward and perpendicular distances (say $Q(x,y)$) with the goal to get a better fit to the radial distance and angle data; (ii) estimate the densities in CPII and CPIII separately; and (iii) investigate a suitable way to describe the spatial variation of mean school size across the areas. Regarding the last two items, the paper SC/A09/AE3 reported the progress, and these have been incorporated into the current method summarised in the last paragraph.

Regarding the first item (i), SC/A09/AE1 and SC/A09/AE4 investigated the sensitivity of density estimation to the use of several forms of the hazard function. SC/A09/AE1 explored the robustness of different functions of the hazard probability model to estimate the effective strip half-width and perpendicular distance density. Four alternative parametric forms of the hazard probability function were explored. Model form 1 was that used in SC/60/IA9, which is based on distance ahead and perpendicular distance, and model forms 2-4 were based on radial distance and angle, of which model form 4 was that used in Skaug *et al.* (2004) which estimated abundance of North Atlantic minke whales. The exploration involved performing a pairwise comparison of the four alternative forms by minimising the Kullback-Leibler distance between the models. It was found that with one exception, the approximating effective strip half-width laid within 10% of the true value for all 12 comparisons considered. Though, two of the forms performed slightly better than the others (models 3 and 4). The overall conclusion was the hazard probability model appears to be robust.

The Workshop thanked Skaug for joining the Workshop via conference call to discuss this paper. This paper helped address one of the issues that arose in SC/60. It was particularly good to confirm that the hazard probability model is robust and it was enlightening to see that the models based on radial distance and angle performed better than models based on distance ahead and perpendicular distance. However, to fully interpret these results the Workshop suggested a revised paper be submitted to SC/61, where the analysis is rerun using data that are truncated at 1.0 or 1.5km perpendicular distance and also provide an additional figure that compares the probability of sightings in addition to the probability density function. It was also suggested that the parameter values of the forms be checked to make sure that the simulated distributions are similar to the actual distributions seen in the IDCR/SOWER minke whale data. After this is complete, it was suggested that the most favourable model be used in the final OK method to estimate minke whale abundance in CPII and CPIII.

Complementing SC/A09/AE1, SC/A09/AE4 examined the sensitivity of the density estimation to hazard probability model using a subset of the IWC simulation data (Palka and Smith 2004; 2005). Three kinds of hazard probability models were employed: (x,y)-logit, (r,θ)-logit, and (r,θ)-separate models. The goodness-of-fits in (r,θ)-models were generally better than that in (x,y)-model. This was also supported by the diagnosis plots and AIC-based model selection. In this sense, use of $(r\ \theta)$-model improves fitting the data. However among the three models investigated, the (x,y)-logit model resulted in the least biased estimates of whale density. These results suggest that the density estimation was robust to the form of the hazard probability function, and is consistent with the general conclusions in SC/A09/AE1.

The Workshop welcomed both of these investigations. It was noted that the various forms of the hazard probability model appeared to be robust; however, some forms were less biased than others. The Workshop encouraged the developers of the OK method to use the (r,θ)-separate model or which ever model performs best when SC/A09/AE1 is updated. In addition, it was suggested that when analysing the simulated scenarios, a measure of overall performance that incorporates both bias and precision, such as the root mean square, is more appropriate than the mean.

4.3 SPLINTR method

Bravington *et al.* (2008) and SC/A09/AE2 presented the SPLINTR method developed by Bravington, Hedley, Wood and Peel. There are four components to the model:

(1) the probability that a given combination of platforms will detect a school that is at a given perpendicular distance from the trackline, for given school size and sighting conditions [assuming independence of platforms for sightings on the trackline, given conditional independence];
(2) the spatial pattern of mean true school size (MSS), and the frequency distribution of school size, for each year [polytomous regression on school size categories, overlaid on spatial smoother];
(3) the probability distribution of *recorded* school size in IO mode, for given *true* school size, perpendicular distance, and sighting conditions [various models, including binomial and Poisson error models as in OK and IM; Binomial used in final versions of fits]; and
(4) the spatial pattern of density-of-schools (regardless of group size) within each year, including fine-scale local clustering [Markov-modulated Poisson process], as well as large-scale spatial variability [spatial smoother].

Sub-models 1-3 (collectively called SPAMASSS – 'Sighting Probability And Misunderestimation And Spatial School Size') are fitted simultaneously by approximate maximum likelihood (Laplace approximation), and are used to estimate a continuously-varying ESW along the survey tracks, and continuously-varying MSS across the whole survey area. This step also uses a likelihood based on relative encounter rates in short intervals of time across changes in sighting conditions; this allows the use of relative density information, while avoiding confounding with large-scale spatial correlations between sighting conditions and density.

Sub-model 4 (DOSS – 'Density of Schools Spatial') is then fitted conditionally on the ESW estimates, to estimate continuously-varying school density. The statistical model for encounter rate uses the MMPP (Skaug, 2006) to deal with fine-scale clustering.

Abundance within an arbitrary region is estimated by multiplying local school density by local MSS and integrating across that region. Variances are propagated via the delta-method. The two spatial smoothers used are soap-film smoothers which are specially designed to cope with irregular boundaries and to 'tame' under-sampled corners - two issues which have plagued spatial modelling of whales in the past. The analyst has considerable flexibility in choosing how the various parameters may depend on true school size and various aspects of 'sighting conditions': for example, sea state, decade, vessel, and any combination/interaction.

The Workshop was encouraged with the progress on the development of this method and looked forward to seeing abundance estimates at the SC/61 annual meeting. Several suggestions were made to explore the influence of some of the unique characteristics of this method. For example, it was suggested that a sensitivity analysis be conducted to explore the effect on using only encounter rate data up to and including the first sighting from the Closing mode in the density model for the IDCR/SOWER data.

4.4 Integrated model method

Cooke (2008), SC/A09/AE5 and SC/A09/AE6 presented the Integrated Method developed by Cooke. The Integrated Model (IM) method for analysing whale sightings data addresses the following issues: incomplete detection on the trackline; heterogeneity in the detection function due to school size and environmental conditions; estimation of density from incomplete spatial coverage, and from non-synoptic surveys of the region across years. The approach is to estimate the effective strip width of the sightings surveys as a function of school size, environmental conditions and other factors, and to use this as a measure of the effective effort (search area) in the fitting of a spatial model to the distribution of sightings. The parameters of the model for strip width are estimated together with the spatial model parameters in a single maximum likelihood computation, so that the uncertainty in strip width is fully accounted for in the estimation of the spatial model parameters. Resulting estimates of abundance and their variance/covariances take account of the uncertainty in the strip widths and in the fitted spatial model. Questions of model selection are solved by modelling all but the essential parameters as random effects, and fitting their variances with a model of much lower dimension than the model for the parameters themselves.

The Workshop welcomed the insights made on the two approaches to smoothing spatial models of density using covariance models (SC/A09/AE5) and looked forward to seeing these approaches applied to the IDCR/SOWER Antarctic minke whale data to obtain abundance estimates. The Workshop also restated some of the suggestions made at the previous workshop. It was suggested that the school size bias model could be enhanced if the relationships documented in the school size experiments conducted in 2006/07 were used. In the large scale whale density function a relationship other than the log(distance from ice) will probably be needed when analysing the IDCR/SOWER data, in contrast to the simulated data.

4.5 Comparison of methods

The structures of the four methods have not varied substantially from that described last year (IWC, 2008). Refer to table 1 of that report for a detailed comparison of the four methods. This table will be updated at SC/61 to include the characteristics of the final methods that will be presented at SC/61.

5. RECENT SOWER EXPERIMENTAL DATA

The 2004/05 field season in the Antarctic started a series of IWC/SOWER cruises that included experiments which have been very useful in providing information both to develop optimal survey designs and methodology for future SOWER or other similar surveys, and to address issues pertaining to the analyses of the existing IDCR/SOWER survey data. The experiments included using BT mode to estimate $g(0)$ for minke whales, trialling new equipment to collect the sightings data, collecting extensive dive time data from different minke whale school sizes, and exploring new methods to estimate abundance of species other than minke whales. The relevant experiments have already improved the analytical methods now being used to estimate the minke whale abundance using the CPII and CPIII data, as well as assisting in conditioning the IWC simulated data scenarios. They also will be able to help interpret any abundance estimates and provide confidence that components of the estimates, such as $g(0)$, are in the appropriate range.

SC/A09/AE9 described the analysis and results of the Buckland-Turnock (BT) search mode experiments that were conducted on the IWC/SOWER cruises in 2005/06 to 2007/08. Of particular interest were the experiments conducted on the 2005/06 and 2006/07 cruises, as these allowed the estimation of the probability of detection for the topman in the barrel. In BT mode, observers are divided into primary and tracker observers with the tracker searching far ahead of the vessel to detect animals before the primary. Thus the tracker sets up trials for the primary observers. A successful trial is one in which the primary detects the same animal (duplicates). In the implementation on these surveys, the topman in the barrel acted as the primary observer and searched as usual in normal standard passing (NSP) mode. The tracker was located on the bridge and searched with higher powered binoculars. School size was an important explanatory variable and the expected probability of detection on the trackline for different school sizes indicated that detection was substantially lower for singletons compared to schools of two or more animals. Beaufort sea state also had more influence on the detection of singletons than larger schools. An important consideration in this type of analysis was whether there had been responsive movement of the animals to the vessel. Plots of perpendicular distances of duplicate sightings at the time of detection by the tracker and subsequently by the primary did not indicate that there was responsive movement before the animals were detected. However, it was suggested that the data were insufficient to tease out any indications of responsive movement unless the reactions were severe. Errors in angle and radial distance measurements may also mask any patterns.

The Workshop believed these experiments would provide an appropriate range of estimates of $g(0)$ that could be compared to estimates made using the three new analytical methods. But first the Workshop suggested a couple of things to more fully explore these data and then to submit an updated paper at SC/61. It was suggested that the probability of detection on the trackline for different school sizes could be modelled using covariates or strata for school size and Beaufort sea state. This could be used to describe how small group sizes (in particular groups of size one) differ from larger group sizes, if they do differ.

Because the Workshop considered all of the experiments as an important tool to evaluate and interpret any abundance estimates that will be presented in SC/61 it suggested that the results of all of the experiments be reviewed, and during the SC/61 meeting, utilise these results to evaluate or interpret the abundance estimates that will be presented at SC/61. The Workshop also noted that the dive time data collected during some of the experimental cruises might be another source of data to investigate if the minke whales are responding to the survey vessel by either avoiding or being attracted to the survey vessel.

6. DIAGNOSTICS

The Workshop examined diagnostics plots and tables to evaluate the goodness of fit of the SPLINTR analytical method (SC/A09/AE2 electronic plots). Though the results are considered preliminary, the diagnostics plots indicate that the models in SPLINTR fitted the data well. It was noted that the observed perpendicular distance data were more spiked at the original than the predicted perpendicular distance, particularly when stratified by school size.

Examining these diagnostics and those presented at SC/60 for the OK method lead to a discussion on the most appropriate diagnostics that should be presented at SC/61 for any analytical method that will be presenting results. To investigate the fit to a variety of aspects of the data, the list of diagnostics was updated and is presented as Annex C. The Workshop **recommended** that quantitative measures of goodness-of-fit were necessary to determine if apparent lacks of fit were significant. The Workshop also **recommended** that a template be developed to assist in standardising the presentation of the diagnostics, which will facilitate easy comparison of the methods and ensure the diagnostics from the various analysis methods were calculated in a similar manner. Bravington said he would distribute such a template to the other developers.

7. ABUNDANCE ESTIMATES

The Workshop reviewed the specifications of the sets of abundance estimate contained in IWC (2008) that are requested from each developer, and agreed to revise them slightly, mainly to make them more explicit and to reduce duplication.

To facilitate the comparison of estimates from the different analytical methods it was **recommended** that the following abundance estimates from CPII and CPIII be presented at SC/61:

(1) Stratum estimates for each individual survey using the strata boundaries in the ‘standard’ database. CVs for each estimate should be calculated without additional variance.
(2) Estimates for the ‘additional variance blocks’ as defined in Annex D for the area south of a ‘common northern boundary’. CVs for each block estimate should be calculated without additional variance. The variance-covariance matrix of the full set of block estimates should be computed for input into the additional variance calculations (see Item 8 below).
(3) Estimates for the longitudinal blocks defined in Annex D, but with the actual northern boundaries of the surveyed strata instead of the common northern boundary. CVs for each block estimate should be calculated without additional variance. The variance-covariance matrix of the full set of block estimates should be computed for use in (4).
(4) Using the results of (3), circumpolar abundance estimates should be calculated for each Management Area for each of CPII and CPIII using the ‘survey once’ approach. The blocks to be included in the ‘survey once’ totals are identified in Annex D. Circumpolar totals for each of CPII and CPIII should also be computed. CVs for each Management Area abundance estimate and for each circumpolar abundance estimate should be computed using the variance-covariance matrix of the abundance estimates by block, without additional variance.
(5) ‘Best’ estimates for each Management Area in each of CPII and CPIII, and also circumpolar estimates for CPII and CPIII, for the area from the ice edge north to 60°S (or alternative northern boundary as determined by the analyst), where the definition of ‘best’ estimate and its CV is determined by the analyst. The abundance estimates presented here should not include extrapolations into the pack ice or polynyas, but the analysts are free to present such estimates as separate additional material in relation to the IA agenda item on whale abundance within the pack ice.

Additional variance should not be included in the CVs presented for sets (1) through (4), but will be calculated as discussed in Item 8. In the case of the analysts’ own ‘best estimates’ (set 5), the CVs may include a component of additional variance which is at the discretion of the analyst in a manner consistent with their approach to multi-year abundance estimation.

The blocks required to implement the ‘common northern boundary’ for estimate set (2) are listed in Annex D, Table 1 and displayed in map form in Annex D, Fig. 1. Data files of the boundary points were determine by Branch and corrected by Okamura, and circulated to all developers. Each IWC Management Area (I through VI) is divided into two or three sectors (East and West, or East, Middle and West) based on which longitudinal sector was covered in each survey.

The Workshop confirmed that the set of estimates presented in Okamura and Kitakado (2008) for the OK method is already complete with respect to the above specifications (except for the variance-covariance matrix required for (2)) and thanked the authors for preparing these. The Workshop appreciated the large amount of work involved in generating these estimates.

8. ESTIMATION OF ADDITIONAL VARIANCE AND TRENDS

8.1 Additional variance

SC/A09/AE7 presented models and a method for estimating additional variance. The issue of additional variance arises from the fact that the estimated sampling variances for the abundance estimates do not account for variability of abundance levels due to especially inter-annual changes in distribution of whale population in the surveyed areas. The additional variance should be taken into account when abundance estimates for Management Areas as well as for the total are used to compare between the two CP series. To estimate the additional variance, the abundance estimates and their associated sampling errors according to the so-called additional variance blocks (see IWC, 2008 and Annex D) are required. Then, the additional variance is estimated with an REML method under log-normal random effect models for abundance estimates. Population trends can be incorporated into the model in several ways such as an exponential yearly growth (Model 1), fixed-CP effect (Model 2), and interaction of CP and Management Area-effects (Model 3). The paper also described an extended model for simultaneous estimation of the difference in abundance level between CPII and CPIII and the number of whales in sea-ice (see also Kitakado and Okamura, 2008).

During discussion of this paper, the Workshop suggested that it could be better to modify Model 3 so that it has an interaction between the exponential yearly trend and the Management Area.

The Workshop discussed the circumstances in which additional variance should be taken into account. On the assumption that the abundance estimation methods are capable of determining the sampling variance of the abundance estimates for each stratum at least approximately correctly, the additional variance should reflect primarily the real inter-annual variation from year to year (additional to any deterministic trend that may be fitted) in the true abundance in each block.

The Workshop agreed that additional variance should not be included in CVs presented for abundance estimates for individual survey strata in each year, but that it should be included in the variance of circumpolar or Management Area estimates that have been obtained by combining data across years.

The Workshop agreed to use abundance estimates for the blocks specified in Annex D as the basis for additional variance calculations. Each developer is requested to provide variance-covariance matrices for their abundance estimates for these blocks (see estimate set (2) in section 7). A 'block' means a sector (E, M, or W) of a Management Area (I through VI) surveyed in a given year, south of the 'common northern boundary'.

SC/A09/AE7 presented some alternative models for the estimation of additional variance. The Workshop agreed with the general approach of this paper, and selected the following specific model:

$$\log \hat{N}_{b,y} = \log \mu_b + (y-2000)\varphi_{A(b)} + \gamma_{b,y} + \varepsilon_{b,y} \quad (1)$$

where:

$\hat{N}_{b,y}$: input estimate of abundance for block b in year y (blocks and years as listed in Annex D)

μ_b: fixed effect to be estimated for block b

φ_A : trend parameter for Management Area A

$A(b)$: Management Area to which block b belongs

$\gamma_{b,y}$: independent random effects to be estimate

$\varepsilon_{b,y}$: input sampling errors of log-abundance estimates by year and block.

This model allows for different trends in each Management Area. However, because there are insufficient data points for reliable unconstrained estimation of area-specific trends, the φ_A should be estimated as random effects around a common mean $\varphi_A = \varphi+\sigma\nu_A$ where ν_A are distributed as N(0,1), φ is a mean trend parameter and σ is a variance parameter to be estimated. The year 2000 was selected as a convenient origin year, with no particular significance.

The Workshop recognised that the estimate of τ, the additional CV, obtained by this method may turn out not be very precise, but agrees that this approach is better than to ignore the additional variance.

To preserve the usual correspondence between variances on the log scale and CVs on the normal scale, the following assumptions are made. The $\gamma_{b,y}$ effects are assumed to be distributed as N (-0.5 log(1+τ^2), log(1+τ^2)) where τ is the 'additional CV' to be estimated. The $\varepsilon_{b,y}$ are assumed to have a multivariate normal distribution with means -0.5 log(1+$cv^2_{b,y}$) where $cv_{b,y}$ is the input CV of the abundance estimate for block b in year y, and the variance-covariance matrix $\tilde{V}$ is defined by:

$$\tilde{V}_{b_1,y_1;b_2,y_2} = \log\left(1 + \frac{V_{b_1,y_1;b_2,y_2}}{\hat{N}_{b_1,y_1}\hat{N}_{b_2,y_2}}\right) \quad (2)$$

where:

V is the input variance-covariance matrix of the abundance estimates (on the normal scale).

After fitting the model, circumpolar abundance estimates and CVs are to be computed for each circumpolar series for the following reference years:

- CPII: 1989 (1988/89 season)
- CPIII: 1999 (1998/99 season)

The reference years were chosen as the approximate mid-points of each series. The circumpolar abundance estimates are computed as follows. Model (1) is fitted and the parameter estimates obtained. The model-predicted value of $N_{b,y}$ is computed for each reference year (1989 and 1999), for each block b, along with the variance-covariance matrices of these predicted values. The model-predicted values of $N_{b,y}$ are:

$$\tilde{N}_{b,y} = \hat{\mu}_b \exp\left(y - 2000\hat{\varphi}_{A(b)}\right) \quad (3)$$

where y = 1989 or 1999. These predicted values are totalled for each of CPII and CPIII to produce two circumpolar estimates, and their (2×2) variance-covariance matrix.

In addition, abundance estimates are computed using formula (2) for each Management Area for each of the reference years, 1989 and 1999.

Kitakado kindly volunteered to supply computer code for this model, so that each analyst can run the model on their estimates, and can compute the above circumpolar and Management Area estimates, for each abundance estimation method.

8.2 Trends

Among the outputs of the additional variance model of the preceding section will be Area-specific estimates of trend in abundance, φ_A for each of the six Management Areas, along with their standard errors. The circumpolar trend is not a simple arithmetic mean of the Area-specific trends, because they should be weighted by abundance. The circumpolar annual 'trend' estimate is essentially the difference between the 1989 and 1999 circumpolar estimates, divided by the time interval,$(N^{CP}_{1999}/N^{CP}_{1989})^{1/10}$-1. The quantity should be computed along with a standard error that takes account of the covariance between the two circumpolar estimates.

The workshop noted that the assumption of an exponential trend in each Management Area is required for the purpose of standardising the estimates to a common reference year for the calculation of notionally synoptic circumpolar abundance estimates. When expressing the final estimate of difference between CPII and CPIII as an annual trend, it should not be assumed that this represents an ongoing trend that would be expected to continue.

9. SIMULATED DATASETS

The SPLINTR method was used to analyse the simulated datasets from scenarios 39 through 54 (SC/A09/AE9). These results were also compared to the results from the OK model presented in SC/60 and the standard method. These scenarios incorporate the following factors:

(1) recorded data includes measurement errors in the time, radial distance, angle and school size;
(2) whale density follows a non-linear gradient;
(3) some groups are incorrectly assigned as a duplicate sighting;
(4) individual whales within a group surface in a non-synchronised fashion;
(5) location correlation. That is, simultaneously, there is a school size gradient (larger schools near the ice edge), vertical density gradient (higher density near the ice edge and a non-linear gradient) and weather gradient (better weather near the ice edge); thus, the detection function is dependent on school size and weather.

To investigate the effects of the factors and to attempt to de-alias the 2-way and higher interactions from the main effects of these factors, a Mirror-Image Fold-over partial factorial design was developed. To quantify and determine the significance of the effect of these factors, the mean percent bias, (observed-actual))/actual, of the 100 replicates within each scenario was regressed against the factors (where -1 indicates absence of a factor in a scenario and +1 indicates the presence of a factor in a scenario). As expected the mis-identification of duplicate sightings lead to a very large significant biased density of whales. Unfortunately, even though the scenarios were designed to de-alias the 2-way and higher interactions, in these scenarios (39-54), there were significant 2-way interactions, in contract to the previous sets of scenarios (1-16 and 17-32). Thus, it is not possible to confidently attribute a single factor to the lack of fit for scenarios 39-54.

To investigate the possible mis-identification of duplicates in the CPII and CPIII data, the numbers of non-duplicates, definite, possible and remote duplicates were tabulated by CP series and vessel (Annex E). This indicated that the number of possible duplicates changed over time and appeared differ by ship. At this time, it was not possible to investigate why this has occurred, but it seemed possibly due to heterogenities in weather or group sizes, the underlying density of whales or the location of the detected groups relative to the ship. Related to this, it was noted that the number of groups detected only by the bridge team in the simulated datasets was much lower than that in the CPII and CPIII data. The Workshop suggested the reason why the simulated datasets had such low numbers be further investigated, and if possible create an additional scenario that has more bridge sightings. The Workshop also suggested a similar table be constructed from the simulated data to determine if the simulated data resemble the actual data.

Despite the difficulty in interpreting the results from scenarios 39-54, it was obvious that mis-identified duplicates could cause biased abundance estimates when using any analytical method. It is not possible to determine the level of mis-identification of duplicates in the CPII and CPIII data, because experiments to determine this were not conducted. The Workshop does not (and cannot) know for sure, but it was considered likely that the levels of duplicate mis-identification in the simulations were probably higher than in the real data.

To determine bounds on the effect of mis-identified duplicates when analysing the CPII and CPIII data, the Workshop **recommended** that for each analytical method presented at SC/61, two sets of abundance estimates be presented, where one set assumes only the definite duplicates are the true duplicate sightings and the other set assumes both the definite and possible duplicates are the true duplicate sightings. In addition, the workshop suggested that a few additional scenarios could be developed that would allow the effects of solely mis-identified duplicates to be investigated further, after it was demonstrated that the level of mis-identification in the simulations is plausible.

The Workshop **recommended** that for all the analysis methods, the results from all the simulated datasets using their most up-to-date analysis methods be presented to SC/61. To investigate the robustness of components of the abundance estimate, it was suggested to compare the true values to the estimated values of, not only the density of whales (as done during the last couple of years), but also the average group size and density of schools. In addition the robustness of estimated CVs could be investigated by reporting the frequency distribution of the percentiles that the true density is of the distribution of the estimated density (assuming the distribution of the estimated density is log normally distributed given the point estimate and estimated CV). It was also suggested that the root mean square is a more appropriate measure of overall level of performance than the mean when reporting the degree of bias of an analysis method over more than one scenario.

It was also noted that the simulations do not investigate all issues involved in the abundance estimation process and do not capture all of the characteristics of the IDCR/SOWER surveys. For example, the abundance estimates of the IDCR/SOWER data are based on data from multiple years, while the simulated datasets are essentially from a single year; the boundaries of the simulated data are straight, in contrast to the wiggly boundaries of the real data; and the simulated data could incorporate the dive time patterns documented during one of the IWC/SOWER experimental cruises (Hedley and Ensor, 2006).

The Workshop suggested that when the simulated datasets and results of analysing these datasets are documented the characteristics of the simulation be compared to that in the actual IDCR/SOWER data collected during CPII and CPIII and the recent experiments. For example, compare the following characteristics: twodimensional (radial distance and angle) detection functions by platform, group sizes, and weather; surfacing patterns for synchronised and non-synchronised diving groups, spatial distribution of groups, percent of duplicate to non-duplicate sightings (to demonstrate effects of mis-identified duplicates), error in school size, errors in radial distance and angles, and percent of sightings made by the three platforms.

10. WORK PLAN

The workshop **recommended** that the three method developers send to the IWC Secretariat papers providing abundance estimates of the CPII and CPIII data and complete descriptions of the methods two weeks before SC/61. To facilitate the discussions during the IA sub-committee, the members of this workshop, and others whom want to join, may want to meet during the first couple days of SC/61 to finalise discussions of the methods used to obtain the abundance estimates (not the interpretation of the estimates) and then report a summary of these discussions to the IA sub-committee.

A timeline of due dates to encourage the completion of the abundance estimates that was developed by the Workshop are:

April 17 - Bravington send diagnostic templates to all of the developers;

May 1 - Kitakado send code for additional variance for each analysis method via email;

May 23 - send papers to members of this Workshop that describe the details of analysis methods and abundance estimates, preferable also the diagnostics;

May 25 - supply Palka results from simulated datasets.

The Workshop noted the difficulty of reviewing these complex analysis methods so suggested that any paper that is not sent to the Secretariat and members of this Workshop by May 23 at their 6pm, may not be able to be fully evaluated by the Scientific Committee during the SC/61 meeting.

The Workshop concluded at 15:35 on 10 April 2009.

REFERENCES

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, June 2006, St Kitts and Nevis (unpublished). 28pp. [Paper available from the Office of this Journal].

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(2):143-74.

Bravington, M., Hedley, S.L., Wood, S.N., Peel, D. 2008. Estimating Antarctic minke whale abundance from SOWER data, using spatial modelling and point independence. Paper SC/F08/A9 presented to the SOWER Abundance Workshop, 7-10 February 2008, Seattle, USA (unpublished). 19pp. [Paper available from the Office of this Journal]

Cooke, J.G. 2008. An integrated method for analysis of IDCR/SOWER data and TRANSIM simulated data sets. Paper SC/F08/A8 presented to the SOWER Abundance Workshop, 7-10 February 2008, Seattle, USA (unpublished). 10pp. [Paper available from the Office of this Journal].

Hedley, S.L. and Ensor, P. 2006. A preliminary look at minke whale dive time data from the 2004/2005 IWC-SOWER survey. Paper SC/58/IA21 presented to the IWC Scientific Committee, May 2006, St, Kitts and Nevis, West Indies (unpublished). 7pp. [Paper available from the Office of this Journal].

IWC. 2008. Report of the SOWER Abundance Workshop, 7-10 February 2008, National Marine Fisheries Service, National Marine Mammal Laboratory and the Watertown Hotel, Seattle, USA. *J. Cetacean Res. Manage. (Suppl.)* 11: 455-463.

Kitakado, T and Okamura, H. 2008. A method for simultaneous estimation of difference in abundance level between two survey periods and the numbers of whales in sea-ice. Paper SC/F08/A3 presented to the SOWER Abundance Workshop, 7-10 February 2008, Seattle, USA (unpublished). 13pp. [Paper available from the Office of this Journal]

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, Santiago, Chile, June 2008 (unpublished). 19pp. [Paper available from the Office of this Journal].

Palka, D.L. and Smith, D.W. 2004. Simulating the IDCR/SOWER surveys – 2004. Paper SC/56/IA6 presented to the IWC Scientific Committee, July 2004, Sorrento, Italy (unpublished). 15pp. [Paper available from the Office of this Journal].

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, Ulsan, Korea (unpublished). 8pp. [Paper available from the Office of this Journal].

Skaug, H.J., Øien, N., Bøthun, G. and Schweder, T. 2004. Current abundance of minke whales (*Balaenoptera acutorostrata*) in the Northeast Atlantic; variability in time and space. *Canadian Journal of Fisheries and Aquatic Sciences*, 61: 870-886.

Skaug, H.J. 2006. Markov modulated Poisson processes for clustered line transect data. *Env. and Eco. Stat.* 13: 199-211.

Annex A

Agenda

1. Opening remarks and welcoming address
2. Adoption of agenda
3. Review of available documents
4. Methods used to estimate abundance
    4.1 Standard method
    4.2 OK method
    4.3 SPLINTR method
    4.4 Integrated model method
    4.5 Comparison of methods
5. Recent SOWER experimental data
6. Diagnostics
7. Abundance estimates
8. Estimation of additional variance and trends
    8.1 Additional variance
    8.2 Trends
9. Simulated datasets
10. Work plan

Annex B

List of Available Papers

SC/A09/AE
1. Kleppe, T.S., Skaug, H.J. and Okamura, H. Robustness of the hazard probability model.
2. Bravington, M. Preliminary results from the SPLINTR model for SOWER data.
3. Okamura, H. and Kitakado, T. Progress report on the OK method.
4. Okamura, H., Kitakado, T. And Skaug, H.J. Sensitivity analysis of different hazard functions using the simulated data.
5. Cooke, J. A note on smoothing options for spatial density models.
6. Cooke, J. A integrated method for analysis of IDCR/SOWER data and TRANSIM simulated data sets.
7. Kitakado, T. and Okamura, H. Estimation of the additional variance.
8. Burt, M.L., Ensor, P. and Borchers, D.L. Detection probability of Antarctic minke whales: analyses of the BT mode experiments conducted on the IWC-SOWER cruises 2005/06-2007/08.
9. Preliminary results of OK, BHWP, integrated and standard analytical methods when applied to simulated data, 2004-2008.

SC/60/IA

8. Okamura, H. and Kitakado, T. Abundance estimates of Antarctic minke whales from the historical IDCR/ SOWER survey data using the OK method.

Annex C

Diagnostics for IDCR/SOWER Estimates (2009)

For comparisons between methods, and for general sanity checks:

Parameter estimate	Disaggregation[1]
1. ESW for whales in IO mode[2]	Stratum * Year (on map)
2. Mean true school size	Stratum * Year
3. ESW for schools in IO mode[3]	CP series * True school size class * Conditions[4]
4. $g(0)$ for schools in IO mode[3]	CP series * True school size class[5] * Conditions[4] * Platform-combination[6]
5. Contour or colour maps of whale density (S of 60°S, longitudinal range of that year's survey)[7]	Year

For assessing goodness-of-fit via comparison of observed and predicted quantities:

Diagnostic	Disaggregation
6. Histograms[8] of Obs & Pred numbers of IO-mode sightings by perp dist[9] and platform-combination[10]	CP series * Vessel[11]
7. Histos of Obs & Pred numbers of IO-mode sightings by perp dist[9], for each observed school-size class[5]	CP series
8. Histos of Obs & Pred numbers of CL-mode sightings by perp dist[9], for each observed school-size class[5]	CP series
9. Histograms of observed and predicted corrected radial distance[12]	CP series * Vessel
10. Histograms of observed and predicted corrected angle[13]	CP series * Vessel
11. Histograms of observed & predicted school size estimates in IO mode[14]	CP series * Vessel
12. Histograms of observed & predicted school size estimates in CL mode	CP series * Vessel
13. SSE-type data: observed and predicted numbers in recent SSE-type data[15]; true SS category * recorded SS category	(i) Conditions (ii) PerpDist (<=0.3nm or >0.3nm)

Footnotes to tables:
1. Disaggregation specifies the level of disaggregation for the display of results. The level of pooling for estimation purposes is at the discretion of the analyst.
2. IE the probability of seeing a randomly-chosen whale within the perpendicular truncation distance of the tracklines within the stratum (including allowance for possible school size underestimation, i.e. not seeing all whales in a school), multiplied

by twice the truncation distance. Spatial variation in mean school size should be allowed for. In full generality (except for neglecting variations in density across the surveyed strip) this amounts to:

$$2T\frac{\int_{x\in\text{tracklines}}\sum_{s=1}^{\infty}P[SS=s/x]\times E[S_e/x,s,\text{seen}]\times\left(\text{school density at }x\right)\times dx}{\int_{x\in\text{tracklines}}\sum_{s=1}^{\infty}P[SS=s/x]\times s\times\left(\text{school density at }x\right)\times dx}$$

where T is the truncation distance, SS is true school size, and summation is across all tracklines are within the stratum. The middle term in the numerator is defined as:

$$E[s_e/x,s]=def\int_{y=\text{truncdist}}^{\text{truncdist}}P[\text{school of size }s\text{ at }\left(x,y\right)\text{will be seen}]\times\sum_{s_e=1}^{s}P[\text{recorded as }s_e/s,x,y,\text{seen}]\times dy$$

The dependence on x in $E[s_e/x,s]$ arises only through effort-related covariates (such as Beaufort, Vessel, etc.) which vary across the tracklines.

The method of calculating the above quantities will depend on the model being used. Considerable simplifications can be made for methods that lack within-stratum spatial variation in school density or mean school size.

3. Unlike ESW for whales, ESW for schools does not use spatial distribution, nor school size estimation error; it is more like an 'internal' parameter estimate for each model of a theoretical quantity. This also goes for $g(0)$ for schools.
4. Conditions: depends on what's used in the model. Preferably either 3 levels based on Sightability field (Poor: 1-2; Medium: 3; Good: 4-5) or 2 levels based on Beaufort.
5. School size classes: 1, 2, 3-4, 5-9, and 10+.
6. Platform combinations: AB, Ab, aB, Cab where A = Topman, B = IO, C = Upper bridge, capital letters mean 'did see', lower case letters mean 'did not see'. Duplicates include delayed as well as simultaneous duplicates. For example, aB means 'Topman never saw it, IO did, Upper bridge may or may not have seen it'.
7. Only for methods with spatial modelling.
8. Expected values to be plotted as histogram-bars rather than smooth curves, to facilitate comparison with observed values. If the X-axis is distance or angle, then smooth curves should also be added.
9. Perpendicular distance grouped into 12 equally-spaced categories spanning 0-1.5nm. Note that even the *observed* numbers by perpendicular distance category may differ slightly between methods, because of different definitions of distance when a school is seen by multiple platforms.
10. This is a bivariate diagnostic (as per note 2) based on computing: $P[\text{PerpD}\in\text{DCateg}_j,\ \text{PlatComb}=h/x_i,\ z_i,\ \text{IO}]$ over all sightings i, for all PerpDist categories j (see note 8) and Platform-Combination h (see note 6); x_i is the location and year of the sighting, and z_i is the sighting conditions, vessel, etc., that made the sighting. It should be presented as one histogram per Platform-Combination, with the X-axis being perpendicular distance category; note that the sum of observed and expected need not be equal within any single histogram.
11. Vessel K27 (which operated only in a few years of CP2) can be omitted from the graphs.
12. Only for methods which use radial distances to sightings. Radial distance by 0.1n.mile intervals from 0 to 2.5n.miles.
13. Only for methods which use angles to sightings. Angular intervals: intervals: 0-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90+°. The 90+ interval can be omitted.
14. 'Predicted' means 'what the model expects would have been recorded'; i.e. it should allow (among other things) for any mis-estimation of school size.
15. Including at least the SS3 experiment, plus data from SS-type closing-mode operations in more recent years if available.

Annex D

Specification of blocks for the common northern boundary and additional variance calculations

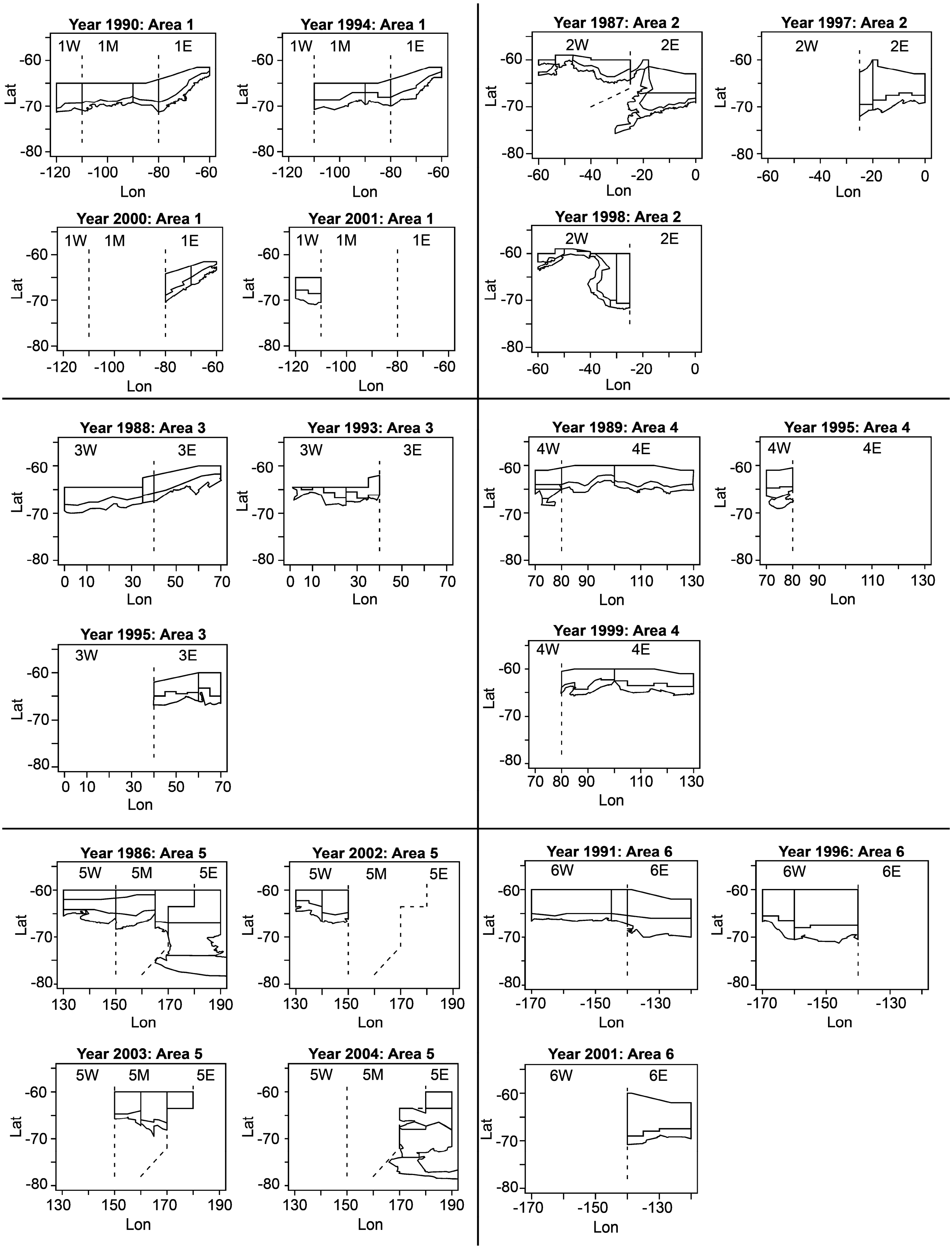

Table 1

List of blocks for use in additional variance calculations.

				CP II						CP III											
Area	Sector	longit. span		1986	1987	1988	1989	1990	1991	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
I	W	120W	110W					II										III			
	M	110W	80W					II			III										
	E	80W	60W					II			X						III				
II	W	60W	25W		II										III						
	E	25W	0EW		II									III							
III	W	0EW	40E			II				III											
	E	40E	70E			II						III									
IV	W	70E	80E				II					III									
	E	80E	130E				II									III					
V	W	130E	150E	II															III		
	M	150E	175E	II																III	
	E	170E	170W	II																	III
VI	W	170W	140W						II				III								
	E	140W	120W						II									III			

Key:

II CPII survey-once block III CPIII survey-once block X Other blocks

NB. Years defined at January 1st (i.e. 1986 refers to the 1985/86 survey season). The boundary between VM and VE does not run along a single meridian (see map). The 1991/92 survey in Area V is not used.

Annex E

Number of duplicate classifications in the IDCR/SOWER data

In IO mode the same group may be seen by two (or three) platforms. Each sighting is listed with a duplicate classification of Definite, Possible or Remote. The following tables show the number of times each classification code is used.

The following options were used to select the data:

- IO mode activity codes: BO, BI
- Species codes: 04, 39, 91, 92
- CP: II=1985/86-1990/91; III=1991/92-2003/04

Table 1 lists the frequencies with which each classification code is recorded. Note a duplicate sighting has two records and a triplicate has three records. The number in parentheses is the number of triplicate sightings.

Table 2 lists the number of unique groups, which was calculated as:

$$Groups = \frac{Duplicates - Triplicates}{2} + \frac{Triplicates}{3} = \frac{93-18}{2} + \frac{18}{3} = 43$$

Note, this calculation does not always work out to whole numbers!

Table 1

List of the frequencies for each classification code.

CP	Vessel	N	D (T)	P (T)	R (T)
II	K27	377	93 (18)	4	2
	SM1	603	253 (35)	47 (2)	14
	SM2	717	405 (78)	17 (3)	14
	All	1,697	751 (131)	68 (5)	30
III	SM1	970	662 (105)	31 (6)	2
	SM2	950	396 (65)	38	9 (1)
	All	1,920	1,058 (170)	69 (6)	11 (1)

Table 2

Approximate number of unique groups by classification code.

CP	Vessel	N	D	P	R
II	K27	377	43	2	1
	SM1	603	120	23	7
	SM2	717	189	5	7
	All	1,697	353	33	15
III	SM1	970	313	14	1
	SM2	950	187	19	4
	All	1,920	500	33	5