# Annex I

# Report of the Working Group on Stock Definition

**Members:** Bravington (Chair), Baker, Brandao, Butterworth, Cipriano, Donovan, Goto, Hammond, Hatch, Iñíguez, Kitakado, LeDuc, Lyrholm, Martien, Nakatsuka, Øien, Olafsdottir, Palsbøll, Park, Pastene, Perrin, Polacheck, Punt, Rosenbaum, Schweder, Skaug, Taylor, B., Tiedemann, Walløe, Wang.

## 1. ELECTION OF CHAIR AND APPOINTMENT OF RAPPORTEURS

Bravington was elected Chair. Martien and Skaug acted as rapporteurs.

## 2. ADOPTION OF AGENDA

The adopted agenda is given in Appendix 1.

## 3. REVIEW OF DOCUMENTS

The documents discussed were SC/56/SD1-9.

## 4. STATISTICAL AND GENETIC ISSUES RELATING TO STOCK DEFINITION

SC/56/SD2 presented a new method called SiteSnipper for improving the statistical power of detecting population structure in cases where the haplotypic diversity is high and/or the sample size is limited. The improvement of genetic sequencing technology has lead researchers to increase the length of mitochondrial DNA sequences with the expectation of improved power. For population structure studies, however, high diversity coupled with limited sample size can result in the opposite effect: lowered statistical power because many individuals have unique haplotypes that do not contribute to tests using frequency comparisons. Unique haplotypes are defined as haplotypes represented by a single sampled individual. SiteSnipper sequentially snips out basepairs that result in the greatest reduction in unique haplotypes. Simulations showed that the Type I error rate of SiteSnipper was unbiased when the number of sites to snip was specified in advance.

SC/56/SD2 included two case studies, demonstrating that power was indeed improved but that how to choose a stopping criterion was not obvious. With Steller sea lions, the first few snipping steps dramatically reduced the number of unique haplotypes and substantially improved the power. With common dolphins, only a few haplotypes were reduced per step, but nevertheless statistical power was again improved, in a more gradual way than with Steller sea lions. In neither case did p-values decrease monotonically as the number of sites snipped increased. The author invited suggestions about appropriate unbiased stopping criteria.

The Working Group noted that this method could be useful when sample sizes are low relative to genetic diversity. It may also help for highly abundant species such as common dolphins, for which there is virtually no prospect of obtaining a large enough sample to adequately characterize haplotype frequencies. Further simulation testing was encouraged, in particular to assess statistical power.

Alternatives to collapsing rare haplotypes were discussed, including combining all haplotypes represented by a single individual into a single 'other' category. However, this is likely to be less powerful than the genetically-based algorithm in SC/56/SD2, since this takes account of the relationships between the haplotypes and therefore retains more information. Another suggestion was to generate a haplotype network and then compare clade frequencies rather than haplotype frequencies. However, this is very similar to the method of Topiary Pruning, which was designed for evolutionary phylogenetic analyses and does not perform well for questions of population structure. The author commented that she has tried several evolutionary methods for addressing this problem, including Topiary Pruning and Tree Puzzle; none of these methods seemed to perform well because they removed too many haplotypes or basepairs from the dataset in the first step. A further alternative could be the SNN method (Hudson, 2000), which evaluates the proportion of 'nearest-neighbour' (i.e. genetically close) haplotypes that are in the same stratum as a given sample. SNN is specifically designed for use when rates of exchange are high enough to leave very little phylogeographic signal, and has been shown to be more powerful than $\chi^2$ in many situations.

There was some discussion of possible stopping rules for deciding when to stop snipping basepairs. If snipping was halted at the most significant p-value, then that p-value would be a biased estimate of overall significance. This could be corrected by developing a permutation test that incorporates the adaptive stopping rule, similar to that used with Boundary Rank in Martien and Taylor (2002).

SC/56/SD9 reported on the performance of Boundary Rank (BR; Martien and Taylor, 2001) in a simple simulation of two populations distributed over a one-dimensional region, with a (linear) cline in their genetic traits. The underlying situation was symmetric about the midpoint of the region considered, so that the algorithm should (on average) select this point as demarcating the inter-population boundary. While the algorithm provided unbiased results for uniform sampling of the region, non-trivial bias becomes evident when the sampling is preferentially to one side of the optimal dividing line. Furthermore, the distribution of boundary placement about this optimum is quite wide, and this width appears to decrease only rather slowly as sample size increases. The

results suggest the need for close inspection of BR results if a cline structure is suspected, rather than automatically accepting its output.

In discussion, it was noted that when genetic diversity is so low that the number of haplotypes is independent of the sample size, then there is a correlation between $\chi^2$ per degree of freedom (the distance measure used by the BR algorithm) and sample size when sample size is low. This correlation could explain the bias seen in SC/56/SD9 which assumed only two haplotypes, because simulated sample size in some of the tests was lowish. For real populations, the number of observed haplotypes usually increases with sample size, nearly eliminating the correlation between sample size and $\chi^2$ per degree of freedom. The results in SC/56/SD9 are interesting, though it was noted that the management implications of any bias or variance in boundary placement in the isolation-by-distance situation simulated in the paper are not obvious; the number of areas selected may be more important than their position. It was suggested that the simulations in SC/56/SD9 could be improved by using a more realistic genetic model, for instance with more haplotypes, and the Working Group noted that the TOSSM project should soon be able to produce such datasets.

SC/56/SD8 proposed a new Bayesian method to estimate mixing rates and stock structure of whales using multilocus genotype data. The method does not rely on the existence of an area containing only a single stock. In addition, the model can incorporate functional or smoothness structure as well as no spatial structure into mixing rates. The incorporation of such structures will help to improve the estimation performance, especially if the sample sizes in areas are small. If the abundance estimates and the DNA data for multiple years are available, the method can infer temporal changes in the spatial distributions of stocks. The method has been applied to a small amount of simulated data and the tests showed the method performed well. It can therefore be a candidate model to be tested on TOSSM datasets when these are available.

The method in SC/56/SD8 assumes that allele frequencies are uncorrelated between subpopulations, as in Cui *et al*. (2002), and this assumption may not be appropriate when divergence is relatively recent. Pritchard *et al*. (2000) describe a method for constructing prior allele frequency distributions with correlations between populations, and it was suggested that this approach might be adapted for use with SC/56/SD8. The Working Group encouraged testing of the SC/56/SD8 method via TOSSM.

SC/56/SD3 described a pilot study where 288 mother-foetus pairs in the Norwegian minke whale DNA register were used to obtain partial DNA profiles for the fathers of the foetuses. The father profiles were subsequently matched against the male part of the register. Three close-to-definite paternities were found. Such data has the potential to yield information about stock structure, migration routes, and other biological parameters. Based on the 'recaptures' of fathers in the database, the number of reproductively active males in the population could also be estimated, although the estimate was very imprecise.

A somewhat similar study was recently published in Krützen *et al*. (2004), though with a much smaller population and different objectives. Two methods were used in that paper, one similar to SC/56/SD3 and the other using the program CERVUS. The latter is based on classifying candidate fathers into either 'possible' or 'impossible', rather than assigning probabilities in the possible cases. Full exclusion is needed if the goal is to estimate the reproductive success of individual males. However, fractional paternities as in SC/56/SD3 make maximal use of data and avoid assigning too many paternities to males with common genotypes.

SC/56/SD1 described the use of pairs of individuals that can be genetically identified as parent and offspring. This approach is particularly useful in cases when gene flow is high or recent exchange is of interest, for which traditional population genetic approaches are inappropriate or have low power. Twenty microsatellite loci were typed in 306 humpback whales from the Gulf of Maine, and pairs of individuals that shared at least one allele per locus were used to estimate heterogeneity in the spatial distribution of close relatives. The analyses revealed some significant incidences of spatial heterogeneity in the northern part of the Gulf of Maine.

In discussion, the Working Group noted the potential power of the approach in SC/56/SD1; it was pointed out, though, that if animals travel widely but stay in closely-related groups which are likely to be sampled together, then there is potential for misinterpreting social structure as population structure. While this could be a real concern for some species, photo-identification of Gulf of Maine humpbacks has not found any evidence of an association between mothers and their offspring, suggesting that the signal detected in this area is unlikely to be due to social structure. Bravington (2002) had suggested an approach to distinguishing between family and population structure, by incorporating differences between times as well as locations of samples.

The methods in SC/56/SD1 are fairly simple, and improvements are planned in several areas: to use more sophisticated measures of possible relatedness, and to take account of scoring errors in microsatellite loci. The Working Group encouraged further work, including estimation of exchange rates, and noted that the approach might be developed for testing on TOSSM simulated data.

In 1999, the Scientific Committee had commented that thus far little use has been made of whale vocalisations to define stocks, and recommended research examining the relationship between acoustic variation and genetic variation. Blue and fin whales were noted as particularly relevant for such studies, as preliminary data suggest that their songs show less inter-annual, inter-individual and intra-individual variability than has been documented in other baleen whale species. SC/56/SD6 addressed this point. Geographic variation in songs produced by male fin whales (*Balaenoptera physalus*) in the North Pacific and North Atlantic Oceans were quantified for comparison with estimates of genetic distance based on Y chromosome haplotypes, mitochondrial haplotypes and autosomal microsatellite genotypes. The goal of the research presented in SC/56/SD6 was to identify particular acoustic features concordant with patterns of gene flow mediated by male fin whales, female fin whales, or both. Four hundred and eight singers from eleven geographic regions in two ocean basins were analysed for a total of 127,094 song notes. Three note types were documented in fin whale song: the 20Hz pulse, or 'regular' note, the lower frequency 'backbeat', and the higher frequency 'upsweep.' Upsweeps were recorded only from two regions in the North Atlantic, and thus can be used to distinguish between singers from the two ocean basins. Three statistical measures (median frequency, bandwidth, and duration) were used to estimate variation in note features. In addition, the timing between consecutive notes of the same type was calculated. Fin whale song composition varied significantly among sampled geographic

regions in number of note types, proportion of note types, note features and inter-note interval distributions. Based on recursive partitioning analysis, 82% of sampled fin whale song bouts were accurately categorised to the regions in which they were recorded. Generalised addition models (GAMs) found differences in the distributions of inter-note intervals (timing of notes) to be the strongest determinants of differences among regional singing behaviour. Regional differences among singers remained apparent despite seasonal and annual variation in song parameters.

Five geographic regions were sampled both acoustically and genetically. Genetic distance based on maternally, paternally and bi-parentally inherited DNA showed trends toward negative correlations with acoustic distance, although only significantly when the influence of geographic distance was removed. Such patterns provide preliminary evidence of character displacement between regions in which singers are sympatric, either physically or acoustically. More generally, results of integrating data types suggest that, for the most part, significant acoustic differentiation among regions is not reflected in estimates of genetic divergence. Patterns of male acoustic dissimilarity may therefore represent discontinuities in fin whale movement and/or social behaviour that either are too recent to be reflected in the evolutionary history of genomes, or vary over relatively short periods of time (decades). Thus, in some cases, variation in acoustic behaviour may be relevant to understanding current fin whale demographics despite lack of signal in slower-resolving genetic markers.

The Working Group discussed the role of learning in fin whale song and how that might affect the usefulness of acoustic data for investigating population structure. If call types are stable over time, they would likely make good markers for population structure. For example, preliminary analyses suggest that blue whale songs may be stable over at least 40 years (McDonald *et al*., 2003). However, if immigrant whales quickly learn to imitate the whales already there, then acoustic data may suggest far more structure than is actually relevant for management. Nevertheless, some degree of acoustic behavioural flexibility could potentially be valuable, as calls may well be plastic enough to allow a clear acoustic boundary to establish much more quickly than a clear genetic boundary. At a minimum, acoustic differentiation may prove useful in designing sampling or tagging schemes to further elucidate population structure.

The ability to discriminate between the songs of different fin whale populations is very good in many instances, and the potential for acoustic assignment of individuals or groups to overlapping subpopulations was discussed. In some regions, the Gulf of California for instance, songs are so distinct that assignment would probably be easy. However, in other areas, particularly the mid-latitudes, songs are less distinct and distinguishing between overlapping subpopulations would be far more difficult.

There are some other limitations to the utility of acoustic data. While some species, such as blue whales, are acoustically active year round, other species, such as minke whales, appear to only sing/call during the breeding season. Thus, acoustic data would be useless in determining their population structure during migration or on the feeding grounds. Similarly, in many species it is only the males that sing. Thus, for these species, acoustic data can only provide insight into the male component of the population. In this sense, acoustic data are the opposite of genetic data, which traditionally have been focused on investigating the female portion of the population through the examination of mtDNA. In the case of the fin whales in SC/56/SD8, though, call type correlates with divergence at both the mtDNA and Y-chromosome markers, suggesting that the structure uncovered by acoustic data is reflective of both sexes.

### 4.1 Genetic data quality

During this year's meeting, the Scientific Committee identified a number of issues related to quality of microsatellite data and caveats about their interpretation. Genetic data are discrete, with well-understood mechanisms of variation. However, there is some degree of noise and inaccuracy which can compromise and even bias results. The sources of error include (but are not limited to) database errors in linking samples to specific whales, compromised DNA from inadequate sample handling, inappropriate markers, null alleles, allelic dropout, stutter bands and anomalous/intermediate allele sizes. There was not enough time to discuss this in detail this year, but the Working Group agreed to put this on next year's agenda, and **encouraged** the presentation of collaborative papers on this topic next year.

## 5. TESTING OF SPATIAL STRUCTURE METHODS (TOSSM)

### 5.1 Update on intersessional progress

At the 2003 IWC meeting, the Scientific Committee and the Stock Definition Working Group instigated the TOSSM project, following a workshop in La Jolla, USA, in January 2003. The main aim of TOSSM is to develop simulation tools that can be used to examine the performance of current and future genetic techniques in determining population structure. The focus is on a management context, where the genetic techniques are used to suggest management boundaries, which in turn are used to set or subdivide quotas according to some rule; the performance of different genetic methods is ultimately to be assessed in terms of how well a simulated management regime performs if the suggested boundaries are used. The Scientific Committee's experience of studying population structure, e.g. in developing *Implementation Simulation Trials* for common minke whales in the North Pacific, has shown that genetic data does not usually provide unequivocal evidence for specific boundaries for use in management. Furthermore, few boundary-placement techniques have been subject to any form of simulation testing. Even those that have, cannot be considered to have undergone the level of extensive simulation testing to incorporate uncertainty that has been a feature of, for example, the IWC's work on the RMP and AWMP. This is perhaps not surprising, given the scope and complexity of developing suitable genetically-specified simulation datasets. This is the most difficult part of TOSSM, and has been the focus of work during the past year.

Such a complex project inevitably has to proceed in an iterative fashion, as with the development of the RMP and AWMP. As proposed last year, the first phase of TOSSM is to develop datasets for a simplified set of scenarios, in order to keep the task manageable and to avoid generating datasets that are so difficult and complicated that it is impossible to work out why they cause problems for particular methods.

The Report of the TOSSM Workshop (IWC, 2004) identified six work modules, each of which has to be completed before the simulation performance testing can actually begin. Intersessional work on the modules was done by the TOSSM developers: Tallmon, Martien and Tiedemann. Progress on each module is detailed below.

*Module 1: Genetic simulation model*

The most important and time-consuming aspect of the TOSSM project is the development and validation of a simulation model that can be used to generate simulated genetic data against which the performance of different methods can be tested. In addition to generating genetic data, the model will also be coupled with a harvest model (module 4) to simulate the impact of catches on the population(s) using the boundaries identified by whatever method is being tested.

Work on this module is nearly complete. SC/56/SD7 described progress on validating and modifying an existing program, RMETASIM, for use as the TOSSM simulation model. RMETASIM comprises a simple and flexible front-end in the language R which interfaces to the genetic simulation program METASIM (Strand, 2002), and is fairly easy to adapt to the requirements of TOSSM. RMETASIM has now been tested across a range of conditions and been found to be a generally well-functioning model that will be adequate to meet the needs of the TOSSM project. The population genetic functions of the model accurately matched theoretical expectations for haploid and diploid loci across an array of scenarios that included isolation, migration, no mutation, and mutation. The structure of the model proved flexible enough to allow the incorporation of a complex whale demography that included a realistic age at first reproduction and inter-birth interval. One error has been identified in the RMETASIM demographic model, and it is being fixed. Making allowance for that error, the output of the model met theoretical expectations from both demographic and genetic theory across a range of conditions.

*Module 2: Biology and population dynamics*

Many aspects of the biological and population dynamics to be used during Phase I of TOSSM, including the population structures, abundances, dispersal rates, mating system and MSYR, were specified during the TOSSM workshop. However, the details of the life history and mutation models were not. SC/56/SD5 proposed detailed life history matrices for Phase I of TOSSM. These matrices are based on life history data for eastern Pacific gray whales and satisfy the specification for MSYR agreed during the TOSSM workshop. This species was chosen because there are many data available on its life history, and because the life history seems fairly typical of baleen whales. Published data were used to parameterise stage-based life history matrices, using a fixed stage duration model (Caswell, 2001) for the five stages: juvenile 1, juvenile 2, fertile female, lactating female and adult male. Two juvenile stages were used to allow for better control of age at first reproduction. The use of separate fertile and lactating classes for females allowed enforcement of a minimum two year inter-birth interval. Stage-based matrices were developed using life history parameter estimates near zero population density and near carrying capacity. The resulting matrices produced rates of increase, generation times and birth and lactation rates very similar to those obtained from a full age-based model. The life history matrices near zero population density and carrying capacity differ with respect to juvenile survival rates, mean age at first reproduction and mean inter-birth interval. Density dependence can be implemented by varying the life history matrix used in the model linearly between these two matrices as a function of abundance relative to carrying capacity.

SC/56/SD7 presented ranges of published mutation rates in humans for both mitochondrial and microsatellite loci, and showed that these result in distributions of genetic diversity comparable to what is found in real baleen whale datasets. In discussion, it was noted that an equal mutation rate had been assumed for all sites/loci; this is not typical of real populations, and may have implications for identifying population structure. It is possible to use locus specific mutation rates for microsatellite markers, and a range of mutation rates can be explored.

Discussions ranged over several detailed issues in the demography and genetics. The Working Group agreed that TOSSM should go ahead with the broad genetic and demographic parametrisations identified, with fine details being agreed intersessionally in the Steering Group (e.g. concerning variable mutation rates across loci). The basic model structure is quite satisfactory, and the flexible nature of RMETASIM should make it easy to change any numerical specifications in future.

*Module 3: Sampling scheme*

During last year's meeting, it had been agreed that for Phase I of TOSSM two sampling schemes would be simulated: evenly distributed sampling and 'gappy' sampling. There are several routines built in to RMETASIM that can be used to sample from the simulations. Thus, all that needs to be done in order to complete this work module is to write a simple R function to assign appropriate geographic coordinates to the samples obtained from RMETASIM.

*Module 4: Catch strategy*

A basic outline for the catch strategy was agreed during the 2003 TOSSM Workshop, and many of the details were specified as given in Punt (2003). Considerable progress has been made toward implementing this strategy; most of the code is written but not yet debugged. The harvest model inputs the locations of the actual population boundaries, the catch histories for the populations, the locations of the stock boundaries suggested by a particular analytical method, and the other harvest-related parameters specified during the TOSSM workshop (e.g., CV for abundance estimates, location of the harvest, etc.). RMETASIM will be used to simulate the population dynamics. In each year, the harvest model receives from RMETASIM a list of individuals currently alive in the simulation. It uses this information to simulate abundance estimates and CVs for each stock and will use the CLA to calculate the quota for each stock. The Fortran code for the CLA has been provided by Cherry Allison. The harvest model then 'kills' an appropriate number of recruited individuals within the harvested area, as specified in Punt (2003). An updated list of surviving animals is returned to RMETASIM, which runs for one more time step before the process is repeated.

*Module 5: Adapting genetic boundary methods for automated use*

No progress was made on this, since the rest of the modules need to be completed before any genetic boundary method can be tested.

*Module 6: Integrating modules*

The Working Group saw a PowerPoint presentation of RMETASIM, using program code from the appendix of SC/56/SD7. RMETASIM can be interfaced with a 'boundary finding' method as follows. First, RMETASIM is run for long enough to equilibrate the populations. Then it simulates a set of genetic samples before saving its state and

stopping. The boundary-finding algorithm can then be applied to the simulated samples, producing a set of boundaries. Finally, RMETASIM is restarted; it reloads its state, and runs in combination with the catch strategy module for the period over which management performance is measured.

SC/56/SD7 included some data on the time required for simulating a dataset. For large populations especially, runtime can be very high. The main bottleneck is the time required to run the population to genetic equilibrium. If coalescent code can be successfully incorporated, runtime should be greatly reduced. For Phase I, though, it seems sensible to focus on the smallest population size considered by the TOSSM Workshop, namely 7,500 animals.

*Overall*

It is clear that excellent progress has been made in specifying and implementing this very complex task. The Working Group expressed its thanks to Tallmon, Martien, and Tiedemann for their efforts over the past year. The project is a world first in its focus on appropriate boundaries for resource management, and should become of great value to the IWC and to cetacean conservation.

### 5.2 Directions for further work

Several technical issues remain to be sorted out with RMETASIM. First, the coding error mentioned under Module 1 needs to be fixed. The original author of RMETASIM is correcting it and expects to have a corrected version available within the next two weeks. Second, density dependence needs to be implemented in the model. This will be achieved by interpolating between two sets of demographic parameters, those expected at zero population density and those expected at carrying capacity, as described in SC/56/SD5. With this approach, density dependence can be adjusted to act on any life history stage. The code for density dependence has been written and will be incorporated once the demographic error has been fixed. The third remaining task is the incorporation of a coalescent model for initialising haplotypes and allele frequencies. The necessary scripts for integrating the coalescent program SIMCOAL (Excoffier *et al*., 2000) into RMETASIM are currently being written. This task should be completed by the end of this month.

The Working Group **agreed** that, although there are various issues of detail in the model specifications which might warrant further attention, priority should be given to assembling a working version of all the modules. Experience in developing other complex testbeds, such as operating models for RMP *Implementation Simulation Trials*, has shown that it pays to get the whole process working before spending too much time on the details. In order to keep the task manageable, the Working Group **agreed** that initial testing efforts should go towards Archetypes I and II (no population structure, and a simple two-stock structure). However, specification and coding of the more interesting archetypes can be continued at the same time

It was noted that the CLA may not be the best tool for learning about the intrinsic performance of different methods, because of its complexity and low catch rates. Ultimately, it will of course be necessary to test candidate methods on management procedures relevant to the IWC, but at the moment the goal is to develop some basic insights into the performance and limitations of different boundary setting methods. This may be accomplished more easily if a simpler catch strategy is implemented, for instance a constant harvest rate. RMETASIM is flexible enough to make this easy.

It is expected that simulated datasets will become available for testing during the coming year. This means that some attention needs to be given to preparing existing genetic methods for linking to RMETASIM, so that some preliminary results are available for next year's meeting. Some existing methods can likely be adapted easily by the TOSSM developers; these include BR and SAMOVA (Dupanloup *et al*., 2002). For both methods, some automatic rule for initial grouping will need to be specified, as will an automated criterion for deciding whether to use a boundary or not. The new method in SC/56/SD8 is also suitable for testing, and should require less in the way of additional rules. As noted last year, tests will include simple summary statistics such as whether the biologically 'right' number of boundaries is selected, as well as performance criteria more directly linked to management, such as depletion, average catch, and average catch location.

The TOSSM Steering Committee, consisting of Bravington, Martien, Kitakado, Kanda, Skaug, Tallmon, Taylor and Tiedemann, will provide any necessary guidance intersessionally to the three developers.

## 6. USING NON-GENETIC TAGGING DATA TO IDENTIFY POPULATION STRUCTURE

Information on population structure is necessary to develop *Implementation Simulation Trials (ISTs)* under the RMP. Currently, the Committee relies heavily on genetic data; other forms of data have not been used to the same extent to develop plausible stock structure hypotheses to be tested in the *ISTs*, primarily because these other data are much sparser than genetic data. However, information on animal movements has the potential to be extremely valuable by providing an additional window on stock structure, on a very different time scale to genetic data.

Developments in telemetry technology have been significant and fairly rapid in recent years. Three primary telemetry devices are currently available: conventional VHF radio telemetry tags, satellite-linked telemetry tags, and pop-up archival tags (PATs). Conventional VHF radio telemtery is not particularly conducive to studies of stock structure because animals need to be tracked at sea, in the air or on land. Satellite-link telemetry, however, works remotely throughout the world. The location information is of variable accuracy, depending on the position and behaviour of the animal. However, even the lowest quality locations are adequate for addressing stock structure hypotheses. PATs have been developed and successfully deployed on a number of large pelagic fish species. Positional information from PATs is not as accurate as satellite-link telemetry, but would be of sufficient accuracy for refining stock structure hypotheses and defining sub-areas and *Small Areas* in RMP *ISTs*. In the context of refining population structure hypotheses, an advantage of PATs over direct satellite-link telemetry may be greater flexibility in attachment.

Methods to analyse the tracks that can be generated by these devices are not yet particularly well developed. However, even at the simple level, relatively small amounts of data have the potential to show, for example, that what have previously been considered for management purposes as two separate stocks could or should be managed as a single stock (e.g. SC/56/BRG5).

Telemetry data provide information complementary to that from genetic data for defining stock structure hypotheses. Lack of significant genetic differences can make it difficult to assess the plausibility of multi-stock hypotheses and to decide on the minimum sizes of sub-areas and *Small Areas* for RMP implementation. A major problem with genetic data is the statistical power when considering the 'null' hypothesis of a single stock. Since it is generally not safe to interpret 'no significant difference' as meaning 'no population structure' based on genetic data alone, it is not obvious where it is safe to stop the process of progressively finer management subdivisions. If it is apparent from telemetry data that animals are mixing on the harvest grounds over a spatial scale of hundreds of kilometres, then there is no reason to expect that differential depletion is possible on smaller scales, and no reason to subdivide the management area more finely than that.

Relatively small amounts of information from telemetry data can provide information on within-year movement ranges. Multi-year telemetry data from the same individual could provide information on longer-term mixing, although current tag attachments limit the ability to collect such data. It is important to note that lack of movement by individuals between areas cannot by itself be taken to indicate the existence of separate stocks. Thus, while lack of difference in genetic data does not necessarily mean a single stock, seasonally confined spatial movements from telemetry data do not necessarily mean multiple stocks. Telemetry data also have the potential to determine the location and number of breeding areas where this information is lacking and thus provide a firmer basis for developing stock structure hypotheses for animals during migration or on their feeding grounds.

The potential uses of telemetry data in evaluating stock structure are numerous. Telemetry studies can be used to identify breeding grounds and clarify their relationship to feeding grounds. They can also help to distinguish between a constant, low rate of dispersal and rare, large (i.e. multi-individual) dispersal events. Data from telemetry studies can be of help in directing sampling design for genetic studies, for instance, by identifying which age/sex classes should be sampled in order to reduce the amount of noise in genetically-based population structure analyses.

Several disadvantages still remain for telemetry data in evaluating stock structure. It was suggested that sub-populations might sometimes be linked by dispersal events which are common enough over management timescales to affect demographics, but rare enough to make detection via telemetry unlikely unless many years are sampled. The logistics of tag attachment are also difficult for some species, and further work on attachment will be needed before focussed telemetry research becomes feasible for some species of interest to the Committee.

There was considerable discussion of how to interpret telemetry data when formulating stock structure hypotheses. The complementary questions of 'How much movement do we have to see before considering two areas to be the same stock?' and 'How many tracks are needed showing no interchange before we would consider two areas to be separate stocks?' need to be addressed. Comparing the results of telemetry studies to those of genetic studies can help with this. It may also be possible to develop statistical models for the likelihood of observed telemetry data given different population structure hypotheses.

This year, the Working Group had discussed three types of information that can potentially yield estimates of dispersal rates and movement rates: telemetry data, close-kin analyses (e.g. SC/56/SD1, SC/56/SD3), and nuclear or mtDNA frequency data. The Chair drew attention to some of the advantages and disadvantages of the three types. Telemetry data can only give information about movements over a few years, and within a single generation; however, this is highly relevant to spatial scale of harvesting. A major plus is that there is no need to actively collect samples from all parts of a population's range, because the animals collect the data by themselves. Close-kin studies describe movements over a generation or two, which is particularly relevant to studies of potential local over-exploitation; however, close-kin studies do require relatively large numbers of loci by current standards, and may require large sample sizes when abundance is high. Gene frequency data usually describe dispersals over evolutionary timescales of hundreds of generations, although changes can occur quite rapidly under some conditions; the problems lie chiefly in estimating rates of dispersal, and in interpreting very long-term phenomena in terms of relatively short-term management issues.

The Working Group encouraged further work on using tagging data to study population structure, and looked forward to the submission of more papers next year.

## 7. UNIT-TO-CONSERVE

The point of this Item is to allow consideration of various possible definitions of unit-to-conserve, and their corresponding implications for management; see IWC (2003, p.49). No papers were received this year. As last year, the Working Group noted that the TOSSM project is deliberately structured to allow investigation of how different units-to-conserve would respond to management, and that results from the first phase of TOSSM should help to inform discussions of unit-to-conserve in future.

## 8. WORK PLAN

The main intersessional task will be the furtherance of TOSSM, as described in Item 5. By next year's meeting, it is expected that there will be preliminary results from several boundary-setting methods, at least for the two simplest stock archetypes. Preliminary agenda items for next year are: TOSSM; statistical and genetic issues relating to stock definition, including a discussion of DNA data quality issues; possible definitions of unit-to-conserve and their implications for management; and progress on use of non-genetic data (tagging) in studying population structure.

## 9. ADOPTION OF REPORT

The report was adopted at 15:02 on 6 July 2004.

### REFERENCES

Bravington, M.V. 2002. Estimating population rates using genetic kinship data. Paper SC/54/SD4 presented to the IWC Scientific Committee, April 2002, Shimonoseki, Japan (unpublished) 10pp. [Paper available from the Office of this Journal].

Caswell, H. 2001. *Matrix Population Models*. 2nd Edn. Sinauer Associates, Inc, Sunderland, Massachusetts, USA. i-xxii+722pp.

Cui, G., Punt, A.E., Pastene, L.A. and Goto, M. 2002. Bayes and Empirical Bayes approaches to addressing stock structure questions using mtDNA, with an illustrative application to North Pacific minke whales. *J. Cetacean Res. Manage.* 4(2):123-34.

Dupanloup, I., Schneider, S. and Excoffier, L. 2002. A simulated annealing approach to define the genetic structure of populations. *Mol. Ecol.* 11:2571-81.

Excoffier, L., Novembre, J. and Schneider, S. 2000. SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *J. Hered.* 91:506-9.

Hudson, R.R. 2000. A new statistic for detecting genetic differentiation. *Genetics* 2000(155):2011-4.

International Whaling Commission. 2003. Report of the Scientific Committee. *J. Cetacean Res. Manage. (Suppl.)* 5:1-92.

International Whaling Commission. 2004. Report of the Workshop to Design Simulation-based Performance Tests for Evaluating Methods Used to Infer Population Structure from Genetic Data, 21-24 January 2003, La Jolla, USA. *J. Cetacean Res. Manage. (Suppl.)* 6:469-85.

Krutzen, M., Barre, L.M., Connor, R.C., Mann, J. and Sherwin, W.B. 2004. 'O father: where art thou?' —Paternity assessment in an open fission-fusion society of wild bottlenose dolphins (*Tursiops* sp.) in Shark Bay, Western Australia. *Mol. Ecol.* 13:1975-90.

Martien, K.K. and Taylor, B.L. 2001. A new method of generating hypothesized population structures for continuously distributed species using genetic data. Paper SC/53/SD7 presented to the IWC Scientific Committee, July 2001, London (unpublished). 16pp. [Paper available from the Office of this Journal].

Martien, K. and Taylor, B.L. 2002. Permutation testing of population structure hypotheses generated by Boundary rank. Paper SC/54/SD5 presented to the IWC Scientific Committee, April 2002, Shimonoseki, Japan (unpublished). 6pp. [Paper available from the Office of this Journal].

McDonald, M.A., Hildebrand, J.A. and Mesnick, S.L. 2003. Biogeographic characterization of blue whale song worldwide: Using song to identify populations. Paper SC/55/SH7 presented to the IWC Scientific Committee, May 2003, Berlin (unpublished). 18pp. [Paper available from the Office of this Journal].

Pritchard, J.K., Stephens, M. and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945-59.

Punt, A.E. 2003. Specifications related to evaluating alternative genetic methods in a RMP management context. Paper SC/55/SD3 presented to the IWC Scientific Committee, May 2003, Berlin (unpublished). 5pp. [Paper available from the Office of this Journal].

Strand, A.E. 2002. METASIM 1.0: an individual-based environment for simulating population genetics of complex population dynamics. *Mol. Ecol. Notes* 3:373-6.

## Appendix 1

## AGENDA

1. Election of Chair and appointment of rapporteurs
2. Adoption of agenda
3. Review of documents
4. Statistical and genetic issues relating to stock definition
5. TOSSM (Testing of Spatial Structure Models)
   - 5.1 Update on intersessional progress
   - 5.2 Directions for further work
6. Using non-genetic tagging data to identify population structure
7. Unit-to-conserve
8. Work plan
9. Adoption of report