

Report of the 2nd TOSSM (Testing of Spatial Structure Models) Workshop¹

1. INTRODUCTORY ITEMS

The Workshop was held at the Unit of Evolutionary Biology/Systematic Zoology, Potsdam University, from 17 to 21 March 2006. Participants are listed in Annex A. Donovan was elected Chair. On behalf of the participants, he thanked the Steering Group and particularly Ralph Tiedemann for organising the meeting and providing the venue. Bravington, Tiedemann and Punt acted as rapporteurs with assistance from Donovan as appropriate.

Annex B shows the adopted Agenda.

2. REVIEW OF TOSSM TO DATE

2.1 The original concept and objectives of TOSSM

An adequate understanding of population structure is frequently a crucial element in enabling informed conservation and management decisions. A number of analytical techniques are available that use genetic data to address population structure questions. Although several of these techniques have been published with a few simulation or empirical examples, none have been thoroughly tested to quantify their performance in a conservation and management context. There is also a dearth of comparative studies, leaving scientists and managers uncertain as to what are the most appropriate techniques to use under particular circumstances.

In particular, some whale species are subject to spatially-concentrated mortality (through aboriginal whaling, commercial whaling, and/or bycatch and ship strikes) in situations where population structure is uncertain, and where there is at least some risk of long-term local depletion. In order to provide a systematic tool for deciding what degree of spatial management is necessary, the IWC began the Testing of Spatial Structure Models (TOSSM) project in 2003 (IWC, 2004).

The purpose of TOSSM is to develop and implement a simulation framework to allow evaluation of genetic methods used in inferring population structure in two ways:

- (1) in general terms (i.e. whether the methods achieve what they set out to achieve); and
- (2) from a specifically IWC viewpoint (i.e. the management implications of using them in the context of e.g. the RMP and AWMP).

The framework for TOSSM follows the general simulation-testing pattern pioneered by the IWC Scientific Committee when developing the RMP and AWMP. Specifically, the TOSSM framework:

- (1) simulates the long-term population dynamics of a structured population, to obtain a simulated 'current' population with individual genotypes;

- (2) samples a subset of the population to obtain spatially-structured genetic data;
- (3) feeds these data into an automated version of a population structure method, the results being expressed as management boundaries²;
- (4) simulates future population dynamics, using the chosen boundaries to control the spatial distribution of mortality as managed by e.g. the RMP or AWMP;
- (5) assesses the performance of the method against several (competing) criteria, primarily the avoidance of local over-exploitation and the avoidance of unnecessary management boundaries that result in underestimation of sustainable yield.

The framework is designed not only to allow evaluation of existing methods but also to assist in the development of new methods. The focus of the IWC's Scientific Committee effort will be evaluation of the use of the various methods via conservation and yield performance statistics similar to those used in developing the RMP and the AWMP. It is important to note that in this context, the IWC objective is not primarily concerned with whether the methods reflect some biological 'truth' but rather on whether the resultant boundaries fulfil its management objectives.

The original workshop envisaged a two phase process. The initial phase was to develop the framework and carry out explanatory analyses of relatively simple situations (see Fig. 1).

The purpose of the exploratory phase is to improve understanding of the basic (comparative and absolute) properties of a variety of methods that have been used or suggested for use in determining stock structure and how they are implemented, as well as providing a feel for the behaviour and interaction of the various factors included in the simulation framework. It will also provide a framework for new methods to be developed. It is a fact that many existing methods do not have as their objective the determination of an appropriate set of boundaries for management.

On the basis of the results of Phase I, Phase II will be developed to provide a more thorough and increasingly realistic examination of methods used to infer stock boundaries in management, perhaps ultimately expanding into more than exclusively genetic techniques as well as providing insights into the implications of choosing various 'units-to- conserve'.

It is thus clear that the TOSSM project is complex in terms of execution and interpretation. Moreover, the 'feedback' or iterative approach is fundamental to its success. With respect to interpretation, there are

² Note that few if any standard population genetics methods actually express their results in terms of recommendations about management boundaries, so that a considerable (but unavoidable) amount of extra work may be required to implement the word 'automated'.

¹ Presented to the meeting as SC/58/Rep6.

philosophical differences between the traditional 'evolutionary' paradigm (*sensu* Waples and Gaggiotti, 2006), and the management paradigm (close to the 'ecological' paradigm *sensu* Waples and Gaggiotti *op. cit.*). In the former, it is important to test and understand whether the methods achieve what they set out to achieve e.g. to determine the number of biological populations. In the latter, where management performance is the arbiter, a method can be considered acceptable even if it incorrectly determines the number of biological populations, provided that it prevents local depletion and provides satisfactory yield. Definitions of 'local' and 'satisfactory' will be informed by feedback from simulation results at a variety of geographical scales and biological and genetic assumptions.

2.2 Progress prior to the Workshop

The original TOSSM workshop report (IWC, 2004) identified the following six work modules, each of which has to be completed before the simulation performance testing can actually begin:

- (1) Genetic simulation;
- (2) Biology and population dynamics;
- (3) Sampling;
- (4) Catch strategy;
- (5) Adaptation of boundary-settings methods for testing;
- (6) Integrating all the above to allow a complete test to be run.

Considerable progress has been made in all of these; by the start of the Workshop, all were operational to a greater or lesser extent, although (1) and (5) required further work. The specific progress can be illustrated by the fact that:

- there are 20 simulated datasets for Archetypes I, II and IV (for definition of these archetypes see Item 3.2 and Fig. 1);
- a draft control program has been developed and distributed as a package in the language R, that takes the base simulated data and generates genetic samples, calls a boundary setting algorithm (BSA), simulates the harvest, etc.;
- exploratory results are available from applying a limited number of genetic methods to the simulated data, either as full BSAs inside the control program, or as stand-alone analyses that will suggest how the analysis output might be processed to form an automated decision-making tool for boundary setting.

2.3 Objectives of the Workshop

The following objectives for the present Workshop were agreed by the Scientific Committee (IWC, 2006, p. 172):

- (1) Consolidate progress on the simulation framework;
- (2) Present results of preliminary runs using existing adapted methods;
- (3) Discuss how to adapt existing boundary-setting methods including:
 - (a) establishing tuneable 'back end' rules for deciding how many boundaries;
 - (b) 'front-end' rules for preliminary sample grouping;
- (4) Discuss what boundary-setting methods should be considered in the first phase;
- (5) Discuss adjustments to the first set of simulated data;

Decide on priorities for further simulated datasets (e.g. more complex population archetypes, more realistic genetics, incorporation of physical tags).

3. GENERAL: GENETIC MODELLING AND RELATED ISSUES

3.1 The status of RMETASIM

A key difference between TOSSM and other simpler *ad hoc* testing frameworks, is that TOSSM tries to mimic realistic allele frequency distributions, in just the same way that they could have evolved over enormous timescales. When originally discussing how to develop population genetic datasets, there had been considerable discussion over the applicability of coalescent models, see Hudson (1990) for a review of coalescent theory. The great advantage of coalescent models is their speed. However, at the time of the first TOSSM Workshop, concerns were expressed about whether coalescent methods would be flexible enough to incorporate important and peculiar facets of whale biology and population dynamics, such as sex-biased dispersal, historical bottlenecks, etc. The alternative approach, IBM (individual based model) has the advantage of being extremely flexible from a biological perspective (e.g. vital rates, location, history and behaviour of each individual) but has the disadvantage, especially for large population sizes of being extremely slow.

At the current Workshop, it was noted that up-to-date coalescent simulators are in fact able to handle a wide variety of scenarios of relevance to TOSSM. Some IBM simulation will always be necessary, e.g. to deal with 'epigenetic inheritance' such as learned feeding ground preferences, and to generate family histories; in particular, the harvesting phase of TOSSM obviously needs an IBM. It is entirely possible, in principle, to meld the two approaches, and indeed the original intention in TOSSM was to use a coalescent simulation (SIMCOAL) to deal quickly with the archaic long-term population history and generate realistic allele frequency distributions that are correlated across populations; next, the slower individual-based simulations in RMETASIM would be run over many fewer years to adjust for details of whale biology that are hard to handle in coalescent models, and also to generate family relatedness data. However, the simulation of datasets intersessionally unearthed some bugs in RMETASIM which made it impossible to 'hot-start' the simulation with a coalescent before the 2nd Workshop. It was therefore necessary to use extremely long runtimes to generate any datasets at all, and this greatly restricted the number of datasets and archetypes that could be generated before the workshop.

Before and during the Workshop, RMETASIM was modified such that it better supports the TOSSM objectives. In particular the major bug that interfered with translation of changes in RMETASIM R objects to their internal C++ representation was identified and repaired (although another has been identified that has not yet been fixed). The most important modification is development of R code to interface with coalescent simulations. It is now possible to generate input parameters for SIMCOAL from within R and automatically call the SIMCOAL executable. The output from SIMCOAL is in ARLEQUIN format and can now be parsed by R routines and integrated into RMETASIM objects. To further reduce 'burn-in' times when transitioning from coalescent to individual-based simulations, several estimators of the central parameter of the neutral theory theta ($4N\mu$) were implemented on RMETASIM objects. These estimates can be used to tune input parameters for SIMCOAL so its output better matches allelic and sequence configurations implied by choices of population sizes, demography, and mutation rates implemented in RMETASIM. Finally, R functions to output time-aggregated

genetic data from `run.tossm()` were developed. Mitochondrial sequences and diploid multilocus microsatellite genotypes are output in ARLEQUIN and GENEPOP formats, respectively³.

The Workshop thanked Strand for his work in fixing bugs in RMETASIM and his co-operation in linking it with a coalescent simulator (SIMCOAL). This should remove the bottleneck on dataset generation. An R interface to SIMCOAL has also been written; this makes it relatively easy for people to quickly generate their own archaic datasets if they want to investigate specific dispersal rates, population sizes, etc, outside of the agreed TOSSM set.

3.2 Archetypes to be considered in TOSSM

The Workshop confirmed the initial archetypes agreed at the initial TOSSM Workshop and these are illustrated in Fig 1.

4. BOUNDARY SETTING METHODS

4.1 Progress on linking boundary setting methods into the framework

The structure of the simulation in the control program is illustrated in Fig. 2. The `run.tossm` module can be used either just to generate genetic samples from the simulated dataset, or to test the full application of a candidate boundary setting method right through the harvest phase. BSAs need to be written as R functions. The BSA function

³ Technical note: Strand also managed to get a full run of `run.tossm()` on a mac unix-based machine, although it has not been confirmed if the output is identical to that produced under Windows. He suggests that `run.tossm()` would easily lend itself to parallelization on Linux clusters, if needed.

might simply be a wrapper that writes out data, calls an external genetic analysis program (e.g. STRUCTURE), reads in the results, and post-processes them to decide on boundaries; or it might involve a statistical analysis written directly in R. Further details are given in Annex C. The Workshop thanked Bravington, Strand and Punt for their work in this regard.

4.1.1 Structure of a single TOSSM run

The first step is to specify basic details of biology and management, such as the archetype, the relation between breeding and harvest grounds, the way in which feeding ground preferences are inherited, the historical exploitation level, which BSA is to be applied and with what parameters, etc. Effectively, this means specifying a number of parameters when invoking an R function.

The steps carried out in each simulation run are as follows:

- (1) Archaic population history:
 - (a) ~1,000,000 years of coalescent;
 - (b) X(<<1,000,000) years of IBM to generate individual histories.
- (2) Historical pre-depletion, e.g. 50 years of hard whaling on one population.
- (3) Collection of ‘genetic data’ and ‘abundance data’.
- (4) Calls the BSA to decide on Small Areas.
- (5) Applies RMP:
 - (a) remove TAC for five years, collect new ‘abundance data’, recalculate TAC;
 - (b) (might collect more ‘genetic data’ occasionally and re-run BSA);

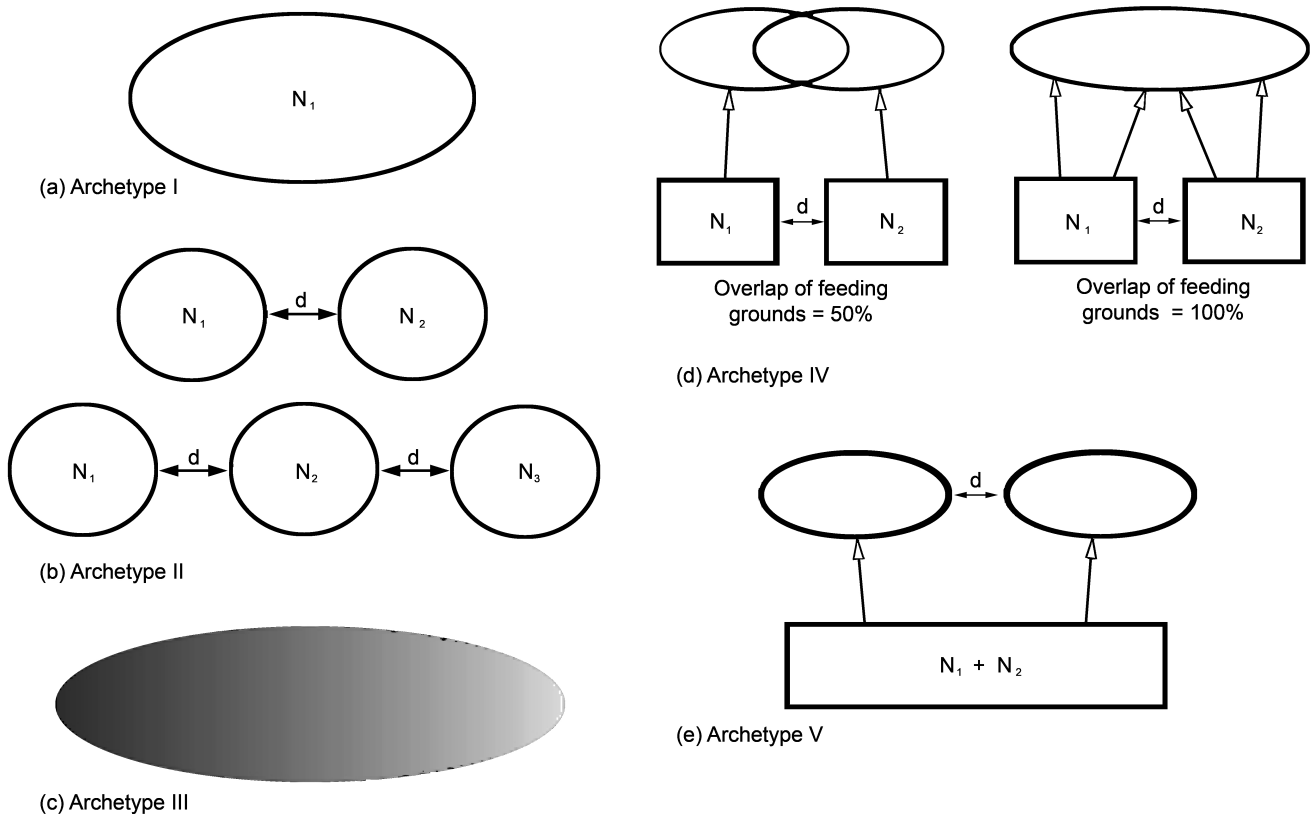


Fig. 1. (a) Archetype I. Panmixia. (b) Archetype II. Stepping-stone. There can be either 2 or 3 populations. Dispersal occurs only between adjacent populations. (c) Archetype III. Diffusion-type isolation-by-distance. (d) Archetype IV. Two discrete breeding grounds with feeding grounds that overlap partially or completely. Ovals indicate feeding grounds while rectangles depict breeding grounds. Open-ended arrows indicate migratory routes while closed arrows indicate dispersal. (e) Archetype V. A single breeding stock with two separate feeding grounds. Animals follow their mothers to the feeding ground and exhibit strong feeding ground fidelity. Ovals indicate feeding grounds while rectangles depict breeding grounds. Open-ended arrows indicate migratory routes while closed arrows indicate dispersal due to females occasionally changing feeding grounds.

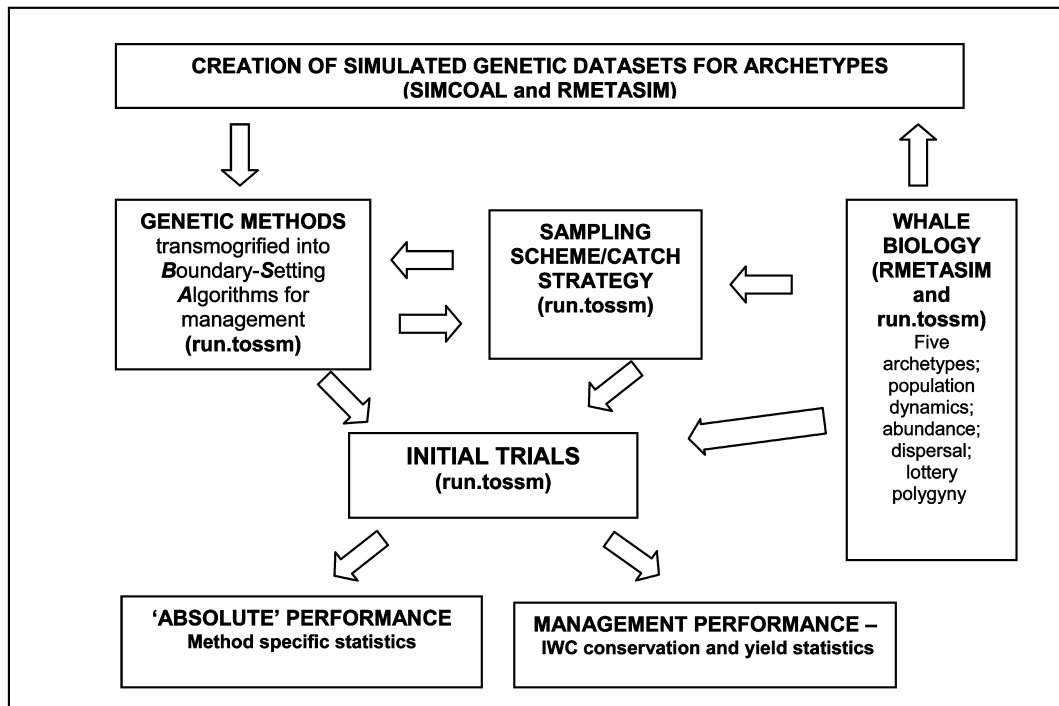


Fig. 2. How the initial trials in TOSSM are set up and run. Items in brackets are pieces of software where various options need to be set in order to make things happen.

(6) (Optional post-RMP relaxation phase: no harvest for Y years).

During all of steps (3)-(6), the program keeps track of 'true' numbers of animals and catches and effort.

In practice, the archaic simulations in step (1) are carried out separately, and each simulated dataset is archived; steps (2)-(6) can then be run many times on the same underlying population, for different BSAs and exploitation rates etc.

4.2 Genetic methods that might be turned into Boundary-Setting Algorithms (BSAs)

This section briefly introduces the ideas behind several methods that might be useful to examine soon in TOSSM and which some participants in the process have expressed their willingness to test. This list is not a comprehensive list of available genetic methods, but the Workshop believes it represents an achievable initial set. It will be clear from the descriptions below that different archetypes are likely to require different types of method; in fact, it may well be that there is no single method that could be expected to work completely 'blind', i.e. with no idea about the underlying reproductive biology and life history. A combination of different types of method may often be needed.

4.2.1 Models not requiring breeding-ground samples

All five methods below assume an underlying scenario of normally non-interbreeding populations subject to some degree of mixing (possibly zero) in the sampling/harvesting sites. They do not attempt to estimate low-level gene exchange through occasional interbreeding, and do not require samples from animals with known population-of-origin (e.g. samples on breeding grounds, as well as on feeding grounds). These methods are potentially able to estimate the number of populations present, as well identifying spatial structure. Note, though, that even if a method is able to clearly identify distinct breeding populations, it does not necessarily follow that separate management is required – for example, if the breeding

populations are evenly mixed throughout the potential harvest area in proportion to their abundance. Hence, to implement a BSA, several stages of different genetic analyses may be needed, e.g. first to identify populations and then to estimate migration rates.

(1) STRUCTURE (Pritchard *et al.*, 2000) is a program and algorithm which groups individuals into assumed breeding populations, based on multilocus genotypes (microsatellites); spatial location is not used directly. It forms groups within which concordance to Hardy-Weinberg equilibrium and linkage equilibrium is maximized. STRUCTURE can also identify hybrid individuals, by assigning percentages of the genotype to the most likely population of origin, and can also provide probabilistic assignment of individuals to their populations of origin. The number of groups, K , is normally an input parameter, but can optionally be estimated according to posterior probability. This may be particularly useful in the context of TOSSM. Although STRUCTURE itself does not use spatial information, given information on sample location, a potential boundary between two adjacent locations could, for example, be tested based on the extent to which specimens from the two locations appear to belong to the same group. The ability to identify potential migrants could also be of value.

(2) GENELAND (Guillot *et al.*, 2005) is an R package that implements a geographically constrained Bayesian clustering algorithm. The algorithm is very similar to STRUCTURE with the main exception that GENELAND can make explicit use of spatial location data, constraining the clustering so that the resultant clusters are geographically contiguous. The geographic constraint can be turned off, in which case GENELAND implements STRUCTURE. GENELAND itself is not set up to directly decide whether or not to place boundaries, so some extra decision rule has to be built in; it is possible, for example, to run the program twice, once to estimate number of populations, and then fixing that value in a second run to propose boundaries.

(3) MIXPROP is a new method developed by Kitakado for estimating area-wise mixing proportions of multiple stocks using individual multi-locus genotypes. The method does not need any baseline populations, and it can incorporate various structures such that the mixing proportions gradually change across sampling areas. The number of populations is selected based on the maximum values of the integrated likelihood functions under several numbers of populations. The method originally aims at not setting any boundaries but estimating mixing proportions; hence, an additional step to actually choose the boundaries based on the boundary setting strategy tested to date is just tentative. It is planned to extend MIXPROP to incorporate gene flow rate under island models.

(4) SEQUENTIAL HYPOTHESIS TESTS can be applied to samples from different areas, aggregating potential groups that are not sufficiently different.

(5) BOUNDARY RANK (Martien and Taylor, 2002) is a related procedure that uses χ^2 -per-degree-of-freedom to decide which boundaries to merge. However, it does not include a rule for deciding when to stop aggregating, so this needs to be imposed externally.

4.2.2 Models requiring breeding-ground samples

These breeding-sample models rely on data from animals whose breeding population is *known* (so that, in particular, the total number of populations must be known); the quantity to be estimated is the long-term rate of gene exchange between populations. Thus they are quite different in intent from the no-breeding-sample models, where the population of origin need not be known, but for which the rate of migration is not be estimated. Assuming the underlying population archetype is appropriate, breeding-sample models are fairly easy to adapt for purposes of making ‘yes/no’ decisions about pre-specified boundaries. However, in a context where there are no pre-specified potential boundaries, additional steps would be required prior to applying a breeding-sample model.

(1) GENERIC F_{ST} quantifies genetic divergence among predefined groups, and can be translated into a matrix of migration measures (N_m). Gene flow can be estimated over an evolutionary timescale. F_{ST} can be calculated in a hierarchical grouping framework. In TOSSM, F_{ST} could be useful in validating predefined potential boundaries. N_m (and its confidence interval) is considered a valuable measure to capture the degree of connectivity among populations, and could serve as a criterion for deciding whether to manage separately or jointly, simply by choosing a threshold. Note, however, that the reliability of N_m estimates based on GENERIC F_{ST} relies on assumptions which are often not met by real populations.

(2) BAYESASS is a Bayesian approach to estimating recent migration between predefined populations, by identifying likely migrants. BAYESASS differs from F_{ST} in two important ways. First, it operates on an ecological rather than evolutionary timescale; and second, it can detect and estimate *asymmetric* migration rates. It provides a valuable measure to describe demographic independence and, like F_{ST} , could fairly easily be adapted in principle to provide a management criterion.

4.2.3 Future work

The Workshop agreed that the above methods were useful to investigate within the TOSSM framework in the near future. It assigned a ‘champion’ to each who will take the lead role in exploring how it might be turned into an automated boundary setting method and incorporate this into the

framework. Automation in terms of both establishing tuneable ‘back end’ rules for deciding how many boundaries and appropriate ‘front end’ rules for preliminary sample grouping is essential if completion of sufficient simulation trials is to be covered in a feasible time-frame. The Workshop stressed that this is **not** a competitive exercise and it is envisaged that many problems (both conceptually and in writing the BSA code) will be common to several methods and may be resolved by an email discussion group. Issues which cannot be resolved that way will need further discussion in the Scientific Committee.

Many of the genetic methods involve MCMC (Markov chain Monte Carlo) methods, and it is not a trivial task to automate this; in practice, when faced with a single dataset, most practitioners spend considerable time visually examining details of the convergence statistics to decide whether the MCMC has worked properly or whether further runs are required. Considerable experience has been gained in other fora (e.g. fisheries management) forming with automation of MCMC diagnostics; Punt offered to assist with this.

The methods and their champions are as follows:

- (a) MIXPROP (Kitakado)
- (b) BAYESASS (Gaggiotti)
- (c) GENELAND/STRUCTURE (Martien/Tiedemann)
- (d) Sequential hypothesis testing (Punt)
- (e) Boundary Rank (Martien)
- (f) F_{ST} -based estimate of dispersal (Waples)

In addition, it will be very useful to examine the two rather simple BSAs: ‘always manage separately’ and ‘never manage separately’, regardless of the data. By testing these two for a range of different dispersal and mixing rates, we will learn about where suitable threshold dispersal and mixing rates might lie for deciding whether or not to manage two areas separately *in a perfect-information scenario*. This may be very helpful in designing BSAs that use a criterion based on, say, an upper confidence limit of estimated dispersal rate. Bravington offered to champion these BSAs.

5. PROGRESS ON TRIALS

5.1 Scenario considerations

It is vital to have an agreed set of shared simulated datasets which allow a comparison of different methods on the same data. The Workshop reconsidered the original TOSSM choices and agreed some minor changes. Annex D gives the detailed specifications of the scenarios underlying the datasets, covering:

- (a) Demographic structure;
- (b) Dispersal, mutation, and breeding;
- (c) Parameterising the coalescent and hot-starting RMETASIM;
- (d) Harvesting and application of the catch control rule;
- (e) Sampling.

In all, the number of scenarios for the initial stage is 80 [5 (archetypes) $\times 8$ (dispersal rates) $\times 2$ (sample sizes for the genetics data)]. 500 replicates of each scenario are ultimately intended, but it may take some time before all 8,000 datasets are available.

5.2 Integration of modules

Considerable progress was achieved at the Workshop, in particular by Strand and Bravington, as discussed under Item 4. The Workshop was pleased to see that a number of successful preliminary runs had been undertaken by

Martien, Kitakado, and Tiedemann. The importance of these runs was not in the results themselves but rather in the fact that the process was able to be taken through the full framework.

5.3 Data archiving and use of the control program

The operational structure will be as follows: 'archaic' populations simulated by coalescence plus a relatively short individual-based simulation will be generated (by Martien) and available via a TOSSM website (http://www.iwcoffice.org/commission/sci_com/TOSSMworkshop.htm), along with the R package 'run.tossm' containing the control program. The control program governs both the simulation of genetic sample data, and the application of each BSA (see Item 4). Specific instructions will be provided on how to run the control program to ensure identical genetic datasets are provided to each BSA. This is slightly unorthodox compared with the usual IWC procedure of archiving the entire dataset, including genetic samples; the reason for this design choice is to allow flexibility in future with multi-year genetic samples and multiple applications of BSAs. Agreement and discussion of coalescent parameters will be dealt with by the intersessional steering group.

5.4 Performance statistics

The Workshop noted that there are a number of stages in the evaluation process of a method, particularly in terms of assessing whether it achieves what it says it can achieve (remembering that most of the methods are not designed to set management boundaries). This is discussed fully in the original TOSSM report (IWC, 2004). Clearly performance statistics will be case specific in such circumstances. The general questions that the methods address usually fall into one or more of the following categories:

- (1) If the number of stocks⁴ is known, does a method accurately stratify the data?
- (2) Given a pre-stratified dataset, does the method define the correct number of stocks?
- (3) Given a dataset with a large number of potential stock boundaries and no information on the number of stocks, how well does the method define stocks?

The third category is clearly the most difficult.

With respect to the IWC's management objectives, the Workshop agreed that the performance statistics available for evaluating the RMP and AWMP should prove an adequate basis.

In both categories, the Workshop agreed that the actual statistics and tabular/graphic formats should be developed and refined in the light of examination of the preliminary

⁴ The appropriate meaning of 'stock' depends on the demographics of each scenario.

results. 'Champions' were encouraged to suggest appropriate statistics for the methods they are investigating.

6. FUTURE WORK AND TIMETABLE

The Workshop identified a number of tasks throughout its report for future work. Clearly the most important concern is the generation and archiving of datasets, addressing computational issues with respect to dataset generation and the automation of BSAs. Responsible persons are outlined under the relevant Items and the whole process will be guided by a Steering Group (to be decided at the 2006 Annual Meeting).

In addition, Donovan and Bravington agreed to produce a short worked example, that would demonstrate the steps that might be followed in developing a BSA that translates genetic data into management decisions for situations similar to those considered by TOSSM. Such an example would be very helpful in presenting TOSSM to geneticists who do not normally work in a fisheries or whaling-management context.

The Workshop stressed that one of the most successful aspects of the Workshop had been to allow time for development work on RMETASIM and run.tossm. It recognised the importance of Workshops in galvanising ideas and progress. It suggests that a further Workshop should be held in two years to assess progress with Phase I and to begin to design Phase II.

REFERENCES

- Guillot, G., Mortier, F. and Estoup, A. 2005. GENELAND: a computer package for landscape genetics. *Mol. Ecol. Notes* 5:945-59.
- Hudson, R.R. 1990. Gene genealogies and the coalescent process. pp1-44. In: Futuym, D.J. and Antonovics, J.D. (eds), *Oxf. Surv. Evol. Biol.* Oxford University Press, New York.
- 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.
- International Whaling Commission. 2006. Report of the Scientific Committee. Annex I. Report of the Working Group on Stock Definition. *J. Cetacean Res. Manage. (Suppl.)* 8:171-176.
- 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].
- Pritchard, J.K., Stephens, M. and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945-59.
- Waples, R.S. and Gaggiotti, O. 2006. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Mol. Ecol.* 15(6):1419-39.

Annex A

List of Participants

Greg Donovan
International Whaling Commission
The Red House, 135 Station Road
Impington, Cambridge
CB24 9NP, UK
Greg.donovan@iwcoffice.org

Per Palsbøll
Department of Genetics,
Microbiology and Toxicology
Stockholm University
106 91 Stockholm
Sweden
Per.palsboll@gmt.su.se

Oscar Gaggiotti
Laboratoire d'Ecologie Alpine
UMR CNRS 5553
BP53
38041 Grenoble Cedex 9
France
oscar.gaggiotti@ujf-grenoble.fr

Robin Waples
Northwest Fisheries Science Center
2725 Montlake Boulevard East
Seattle
WA 98112
USA
Robin.waples@noaa.gov

Allan Strand
College of Charleston
66 George Street
Charleston
South Carolina 29424
USA
stranda@cofc.edu

Andre Punt
University of Washington
School of Aquatic and Fishery
Science
Box 355020
Seattle
WA 98195-5020
USA
aepunt@u.washington.edu

Valerio Ketmaier
University of Potsdam
Karl-Liebknecht-Strasse 24-25
Haus 25, Golm
D-14476
Germany
ketmaier@rz.uni-potsdam.de

Mark Bravington
CSIRO
PO Box 1538
Castray Espi
Hobart
TAS 7001
Australia
Mark.bravington@csiro.au

Hans Skaug
Dept. of Mathematics
Johannes Brunsgt 12
Bergen 5008
Norway
skaug@mi.uib.no

Toshihide Kitakado
Tokyo University of Marine Science
and Technology
5-7. Konan 4
Minato-ku
Tokyo 108-8477
Japan
kitakado@s.kaiyodai.ac.jp

Naohisa Kanda
Institute of Cetacean Research
Genetic Ecology Section, Research
Division
Toyomi Shinko Bldg
4-5 Toyomi-Cho
Chuo-ku
Tokyo 104-0055
Japan
kanda@cetacean.jp

Ralph Tiedemann
University of Potsdam
Karl-Liebknecht-Strasse 24-25
Haus 25, Golm
D-14476
Germany
tiedeman@rz.uni-potsdam.de

Annex B

Agenda

1. Introductory items
 - 1.1 Convenor's welcoming remarks and arrangements for the meeting
 - 1.2 Election of Chair
 - 1.3 Appointment of rapporteurs
 - 1.4 Adoption of Agenda
2. Review of where we are with TOSSM
 - 2.1 The objectives of TOSSM
 - 2.2 The original concept
 - 2.3 Progress up to the 2005 Annual Meeting
 - 2.4 Work prior to the Workshop
3. General: genetic modelling and related issues
 - 3.1 The status of RMETASIM
 - 3.2 Coalescence and other approaches
 - 3.3 Archetypes – now and future
 - 3.4 Generated datasets
 - 3.4.1 What is now available
 - 3.4.2 Priorities for future work
4. Boundary setting methods: 'front ends', 'back ends' and other unsavoury matters
 - 4.1 Progress on linking boundary setting methods with the testing framework
- 4.2 Existing methods
- 4.3 Proposed new methods for consideration
5. Progress on 'trials'
 - 5.1 Scenario considerations
 - 5.1.1 Biological issues including number of stocks, *MSYR*, depletion
 - 5.1.2 Harvesting strategies (applying the RMP)
 - 5.1.3 Sampling schemes
 - 5.1.4 Revised specification of scenarios
 - 5.2 Modelling issues including integration of modules (this will form a major part of the workshop)
 - 5.2.1 Integration of genetic, population dynamics and RMP modules
 - 5.2.2 Linking boundary setting methods
 - 5.2.3 Validation, documentation and archiving of models
 - 5.3 Performance statistics and presentation of results
 - 5.4 Review results of preliminary runs
6. Future work and timetable

Annex C

Technical Specifications of the Initial TOSSM Trials

The following specifications provide the basis for an initial set of (relatively exploratory) trials. Trials based on these specifications will be conducted for each of the five stock structure archetypes.

1. FUNDAMENTAL STRUCTURE OF THE SIMULATIONS

There are up to 4 phases in a simulated population's existence: 'ancient' (wherein happens all the long-term genetic simulations, without any whaling), 'historical' (wherein the population may get depleted but not under RMP), 'managed' (wherein the RMP is implemented using whatever management boundaries are chosen), and entirely optional 'recovery' (no harvesting).

A general framework that can accommodate all archetypes is as follows. The 'species' is divided into breeding populations (BPs), with some migration. It's assumed that all the human-related action happens in Fully

Internally Mixed Areas (FIMAs) – genetic sampling, abundance estimation, and harvest. FIMAs might correspond to feeding grounds, breeding grounds, migration routes, or wherever, and can be defined on a really small spatial scale. They are only relevant during the 'historical' and 'management' phases. As is often the case for whales, there is no 'data' collected from the BPs themselves. However, pure BP samples can be obtained by creating one FIMA for each BP and not having any mixing on the FIMAs.

The boundary-setting options are to group some or none of the FIMAs together; FIMAs can't be split. Hence, if there are only two FIMAs, the only boundary-setting decision is whether to join them or keep them separate. With more FIMAs, there are more options, and it is possible to examine BSAs which decide where to place at most one boundary, or even BSAs which try to decide whether and where to place several boundaries simultaneously. Not all options are relevant for all initial trials.

2. DEMOGRAPHIC STRUCTURE

The same life history parameters are assumed for all archetypes. Tables 1 and 2 list the values for the parameters of the stage-structured population dynamics model. Density-dependence is modeled by linearly interpolating the values in the state-transition/survival and reproduction matrices based on the depletion of the total (i.e. all stage) population size, i.e.

$$X_t = (P_t^B / K^B) X_K + (1 - P_t^B / K^B) X_0$$

where

- X is the matrix used during year t ,
- X_K is the matrix at carrying capacity,
- X_0 is the matrix in the limit of zero population size,
- P_t^B is the total (breeding) population size at the start of year t ,
- K^B is the pre-exploitation number of animals on the breeding ground, and
- P_t^B / K^B is the depletion¹ of the total population size at the start of year t .

This assumes that density-dependence is a function of the size of the population on each breeding ground (rather than feeding ground). Animals are assigned randomly to the feeding grounds in which they are found (option 'Randomsim' in run.tossm). Future trials may involve some site-specificity. Figure 3 shows the yield curve. *MSYL* for the parameters in Tables 1 and 2 occurs at 55% of carrying capacity with an *MSYR* of 3.7%.

¹ Depletion is the conventionally defined at the IWC as the numbers remaining divided by the initial numbers, a high value for the depletion therefore corresponds to a less impacted population.

Table 1

Estimates of survival rate (σ_i) and duration (T_i) for each stage and the resulting parameter estimates for the fixed stage duration model. P_i is the probability of an individual surviving and remaining in stage i , while G_i is the probability of an individual surviving and moving into the next stage. F_i is the fertility rate for stage i . Because the probability that a lactating female gave birth is 1.0, F_i is simply equal to the juvenile survival rate.

Stage	σ_i	T_i	P_i	G_i	F_i
Near zero population density					
Juve1	0.94	3	0.73	0.210	0
Juve2	0.94	1	0	0.94	0
Fertile female	0.946	1	0	0.946	0
Lactating female	0.946	1	0	0.946	0.94
Adult male	0.954	-	0.954	-	0
At carrying capacity					
Juve1	0.925	5	0.768	0.157	0
Juve2	0.925	4	0.720	0.205	0
Fertile female	0.946	3	0.648	0.300	0
Lactating female	0.946	1	0	0.946	0.925
Adult male	0.954	-	0.954	-	0

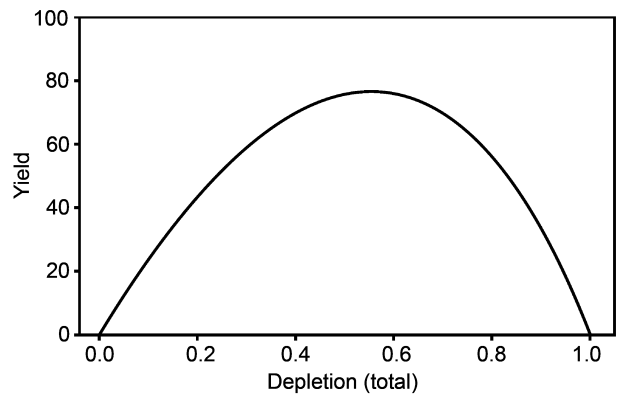


Fig. 1. Yield versus depletion of the total population size for the parameters in Table 1.

The carrying capacity of each breeding stock is assumed to be equal, and the total carrying capacity is 7,500. (Simulations with carrying capacities of 2,500 and 15,000 will also be available later.)

2.1 Genetic structure

The trials are based on annual dispersal rates of 5×10^{-6} , 5×10^{-5} , 5×10^{-4} , 5×10^{-3} , 0.01, 0.03, and 0.05. Most studies of mutation rate report rates of between 10^{-3} and 10^{-4} per locus per generation. However, simulation results with a rate of 5×10^{-4} generate unrealistically low allelic diversities compared to most whale populations, so rates of $1,2,3 \times 10^{-3}$ have been used.

A *sequence* mutation rate of 5×10^{-3} per generation is assumed for the 500bp mtDNA control, which gives realistic haplotype frequencies.

Reproduction is based on lottery polygyny.

2.2 Initializing the population matrix

The demographic and genetic characteristics of each individual in the population to be included in run.tossm are based on the result of coalescence simulations based on using pre-specified values for θ (possibly depending on archetype and dispersal rate) followed by a projection for a number of years.

The initial depletions for the two-stock scenarios are: (a) 0.3 for the stock closest to the coast (where the whaling operations are assumed to occur if the entire feeding ground is treated as a single *Small Area*) and (b) 0.99 for the other stock. All of the catch is taken randomly from the population in the year before the genetics samples are taken.

2.3 Harvesting and application of the catch control rule

Catches are set by *Small Area*, and the boundaries defining the number and size of each *Small Area* are selected by analysing the genetics data. The catch limit for a *Small Area*

Table 2

Updated stage-based matrices for use at (a) zero population density and (b) carrying capacity. Stage class abbreviations are Juve1=juvenile1, Juve2=juvenile2, Fert=fertile female, Lact=lactating female and Male=adult male.

	(a)	Juve1	Juve2	Fert	Lact	Male	(b)	Juve1	Juve2	Fert	Lact	Male
Juve1	0.730	0	0	1.0	0	0	0.768	0	0	1.0	0	0
Juve2	0.210	0	0	0	0	0	0.720	0.720	0	0	0	0
Fert	0	0.47	0	0.946	0	0	0.102	0.102	0.648	0.946	0	0
Lact	0	0	0.946	0	0	0	0	0	0.300	0	0	0
Male	0	0.47	0	0	0	0.954	0.102	0.102	0	0	0	0.954

is based on applying a catch control rule² using the abundance and catch data for that *Small Area*. Harvests are removed randomly, i.e. it is possible to harvest calves. Catches from a *Small Area* are taken in the FIMA closest to the coast. If the catch limit for a *Small Area* exceeds the size of the population in the FIMA closest to the coast, it is removed from the FIMA that is next closest.

² The catch control rule for these initial calculations is based on the *CLA* of the RMP, except that the catch limit from the *CLA* is multiplied by 5.

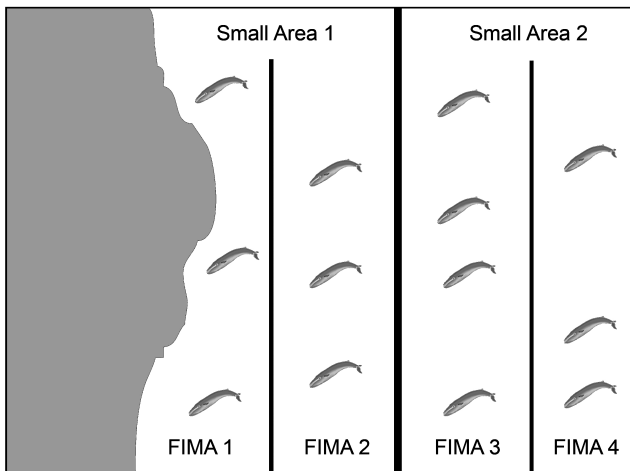


Fig. 2. Example of FIMAs and *Small Areas*. If the catch limit for *Small Area 1* is 4 whales, then 3 will be taken from FIMA 1 because it is closest to the coast, and the remaining whale will be taken from FIMA 2 because there are none left in FIMA 1.

2.4 Sampling

For trials in which there are two FIMAs, the data available to the genetics method are 50 (or 400) individuals taken from each of the FIMAs. The data generated for each individual are 30 microsatellite loci and a microchondrial locus. For trials in which there are more than two FIMAs, the total number of individuals (100 or 800) is taken uniformly from each FIMA. The genetic samples are based on selecting animals at random within each FIMA (without replacement).

The estimates of abundance available to the catch control rule for a given *Small Area* are determined by summing the estimates of abundance for the FIMAs that make up the *Small Area*. The estimate of abundance for year y for a FIMA is assumed to be gamma distributed about the true abundance of all animals (including calves) in the FIMA, with a CV equal to $0.2 / \sqrt{P_y^F / K^F}$ where P_y^F is the total population size in the FIMA (i.e. on the feeding ground), and K^F is the pre-exploitation number of animals in the FIMA.

2.5 Total number of trials

The number of trials for the initial stage is 70 [5 (archetypes) \times 7 (dispersal rates) \times 2 (sample sizes for the genetics data)]. Each trial involves 100 replicates, each of which is based on a different archaic long-term population history. For the purposes of the initial simulations, only 10 replicates for each trial need be analyzed by the 2006 meeting of the Scientific Committee. The number of microsatellite loci for these initial trials will be 10 even though 30 will be generated.