QLK5-CT1999-01609

dst^2

Development of structurally detailed statistically testable models of marine populations

Final Report: 1 January 2000 to 31 August 2004

Volume 1

Commencement date: 01-01-2000 **Completion date**: 31-08-2004

Co-ordinator:

Dr. G. Stefánsson
Marine Research Institute
PO Box 1390
Skúlagata 4 Phone: +354 575 2000
Reykjavík Fax: +354 575 2001
Iceland email: gunnar@hafro.is

Participants:

1	Co-ordinator:	Marine Research Institute, Reykjavík, Iceland	MRI
2	Contractor:	Institute of Marine Research, Bergen, Norway	IMR
3	Contractor:	Danish Institute for Fisheries Research, Charlottenlund, Denmark	DIFRES
4	Contractor:	Science Institute, University of Iceland	SCUI
5	Contractor:	University of Bergen, Norway	UiB
6	Contractor:	Fisheries Research Services, Aberdeen, UK	FRS
7	Contractor:	The Centre for Environment, Fisheries and Aquaculture Science, Lowestoft, UK	CEFAS
8	Contractor:	Institut Francais de Recherche pour l'Exploitation de la Mer, Nantes, France	IFREMER

Contents

Preface

The dst^2 project

This report describes the results of an exercise in modelling marine ecosystems as they relate to fish. This work has been funded as an EU project, dst^2 , co-funded by others (including the institutes involved). The report consists of two main volumes and an accompanying Gadget User Guide. These three volumes constitute the final report from the dst^2 project.

Research on multispecies and technical interactions and how to model these interactions has been ongoing for many decades, culminating in several computer programs which were developed in the late 20th century for the purpose of simulating or estimating the effects of multispecies or technical interactions. Such work has been undertaken under the auspices of several international bodies and grant agencies. As described in this introduction, this particular project follows from many earlier projects, undertaken in several countries.

The project was coordinated by the Marine Research Institute (MRI) of Reykjavik, with other partners being the Institute of Marine Research (IMR) in Bergen, the University of Bergen (UiB), the Science Institute of the University of Iceland (SCUI) in Reykjavik, the Marine Laboratory (FRS) in Aberdeen, The Centre for Environment Fisheries and Aquaculture Science (CEFAS) in Lowestoft, Danish Institute for Fisheries Research (DIFRES) and Institut Français de recherche pour l'exploitation de la mer (IFREMER) in Nantes.

The main goal of the dst^2 project was to develop a platform in the form of a computer program or modules for dealing for multispecies models. The approach follows and builds on from several earlier approaches. The resulting program, Gadget, is now fully functional and is the main output from the project.

This preface describes the background to the project and the report structure along with some principles used for modelling, as detailed in various sections of the report.

Project background

Models of the dynamics of fish populations are usually developed in response to certain questions of interest. The greatest emphasis has traditionally been placed on questions related to the management of fisheries and hence the population models have been developed to answer questions relating to effects of different regulations, such as changes in fishing effort, quotas, mesh sizes, protected areas and so on.

As management has been improved or changed in different areas, the questions also change. Thus, once harvest from a target species comes under control, questions arise concerning bycatch species. Similarly, if the target species biomass gets seriously reduced, questions arise concerning stock and recruit relationships.

The last decades of the twentieth century saw the development and implementation of multispecies methods of assessing fish stocks, statistical methods for assessing population abundance and methods of assessing the uncertainty in predicting the development of individual stocks.

As seen in various parts of this report and earlier publications, each of these three issues, multispecies interactions, statistical models and uncertainty estimation can have a major impact on the perceived or likely range of population abundance and potential or safe harvest from the resource.

Biological model components

Stefansson and Palsson (1998) detail a number of approaches to multispecies models and specified the model types which needed to be used for modelling Arcto-boreal systems such as waters around Iceland or the Barents Sea.

Several multispecies models were suggested in the last few decades of the twentieth century. The most famous one was probably MSVPA, as initially developed for the North Sea. This model yielded important results in terms of stock sizes and species interactions, including predation mortality (Gislason and Helgason 1985). This model, however, does not incorporate the effect of prey abundance on predator growth. This effect seems to be very important in Arcto-boreal systems and needs to be incorporated in any model which is to be used as a description of the ecosystem.

The first true multispecies ecological model of an Arcto-boreal system taking individual growth issues into account was MULTSPEC (Bogstad et al. 1992 and Bogstad et al. 1997). This model incorporated many species and could potentially allow for varying growth due to diet constraints, annual variability in migration and thus spatial overlap. Thus, MULTSPEC included many of the essential building blocks of a model of an Arcto-boreal system, but there was a conspicuous lack of symmetry in the model development and implementation. Thus, not only were data sets for e.g. cod and capelin different, but also the very model formulation.

Naturally, these concerns are not at all unique to Arcto-boreal waters. The question of predator starvation due to the lack of prey is a potential concern in any system and a generic model needs to allow for such a possibility.

As noted by Stefansson and Palsson (1998), data analysis has indicated that a number of processes need to be included in any model which is to be of general use as a fish-oriented ecosystem model: The model needs to be able to incorporate predation mortality, since it appears that predation by one species can significantly affect recruitment and adult survival of another. Other important factors include growth. It is well known that growth in many systems is quite variable and in some cases the abundance of a prey species is known to affect the growth of a predator. Thus the model should incorporate adequate growth equations based on consumption. Further, some prey species have migration routes with fast migration rates leading to variable spatial overlap with predators and a requirement for taking spatial and temporal scales into account in such a manner that the models need to be spatially disaggregated with temporal scales often considerably shorter than a year. A consequence of this is that migration needs to be taken in to account. Since mature fish often have a directed migratory behaviour, a further consequence is that the difference in behaviour of mature and immature fish needs to be incorporated in the models.

Statistical approaches

In accordance with specifications set out by Stefansson and Palsson (1998), Gadget has been developed as a forward simulation model using statistical estimation through weighted combinations of several log-likelihood criteria. Although that particular framework has been used to generate case-specific multispecies models, the statistical and modelling issues that arise are perfectly general to not only statistical multispecies model or, but many other complex nonlinear statistical models which incorporate several data sources. The statistical issues discussed in this report are thus perfectly generic problems but they happen to materialize when developing, implementing and verifying models based on Gadget.

As seen in this report, there are certain issues which need to be addressed in order to ascertain whether the models may potentially be adequate (necessary conditions). If these issues are not taken into account, then the models may be quite misleading. As always, models will be improved and new conditions will be found necessitating further improvements.

The proper statistical estimation methods have been seen to be crucial for providing appropriate uncertainty estimates. Lack of statistical rigour has been seen to result in serious underestimates of true uncertainty and lead to e.g. overly optimistic predictions of the effects of management actions (Patterson et al, 2000, Patterson et al, 2001, Restrepo et al, 2000, Gavaris et al, 2000 and Stefansson et al, 1997).

One important facet of proper statistical treatment is to use data in the proper place. This is done by using models which are completely determined by their parametrization and data are (as far as possible) used only for estimation of these parameters.

Most data are best entered in the form of a likelihood function. This is clear for survey indices of abundance, which contain considerable variation and for stomach content data, which is typically very variable.

It is important that the incorporation of data is flexible. Since the ecosystem does not change simply due to new data being collected, it should be possible to incorporate a new data source without the need for a complete upheaval of the model. It should be possible to incorporate a number of different data sources in the model through the use of log-likelihood components. The weighted sum of these components gives a criterion for comparing different values of parameters such as recruitment, migration rates and parameters in growth functions.

This type of model is statistically based and therefore allows for formal testing of whether or not certain interactions exist. This attribute is sorely lacking in many fisheries models (VPA, MSVPA). Although clearly not a panacea, the potential for formal testing considerably enhances the types of conclusions that can be drawn from data. In particular, if it is found that the inclusion of capelin explains a considerable portion of variation in data on growth, then that clearly provides an indication of the importance of this interaction. This type of result is much stronger than one based on assuming a relationship between stomach contents, feeding and growth and using those assumptions to predict the likely effects on future yields.

In the initial stages of the project it was envisaged that the formal testing mechanisms would mainly be useful for model comparisons. It has been found, however, that model development becomes even more an art in these complex models than in simpler models such as multiple regression. These complex models have also illustrated other issues normally ignored in fisheries models.

For example, model development requires selection of aggregation mechanisms for data not normally an issue in simple statistical models. These further require selection of likelihood functions, where there are often several options, but care must be taken not to inadvertently use the same data several times or to put too much weight on individual data sets. Considerable effort has been expended on these various issues and the results of these analyses are presented in the report.

As seen in section 4.4 in this report, specially developed bootstrapping methods provide a promising approach for estimating uncertainty in the context of these complex multispecies models. This is an important conclusion from the current project and indicates a need for developing these bootstrapping methods further. In particular, it appears likely that bootstrapping entire data sets (age-length tables, length distributions, mark-recapture experiments) from the lightly aggregated datawarehouse and within the spatio-temporal blocks used by Gadget may be the way forward.

Reporting

Earlier dst^2 papers and reports

A large project such as dst^2 inevitably includes a considerable number of reports, project proposals, research publications and technical reports (not to mention a variety of financial reports). Each of the first three years led to an annual dst^2 report, as listed on p. A.1.

This final report is the culmination of the dst^2 project and thus includes summaries or complete results from most tasks undertaken under the auspices of the dst^2 project.

Structure of the report

This report consists of several parts, each one of which consists of several chapters. The chapters in volume I correspond to model development and data bases whereas volum II contains output from the case studies. In addition to these two main report volumes the accompanying Gadget User Guide gives a complete description of the software developed to a large extent within this project.

A chapter in the report refers to a collection of related subprojects. Most sections within a chapter consist of a single paper (or summary of a paper) on a specific topic.

References

Bogstad, B., Tjelmeland, S., Tjelda, T., Ulltang, Ö. (1992). Description of a multispecies model for the Barents Sea (MULTSPEC) and a study of its sensitivity to assumptions on food preferences and stock sizes of minke whales and harp seals. International Whaling Commission, SC/44/O 9.

Bogstad, B., Hiis Hauge, K., and Ulltang, Ø. 1997. MULTSPEC - A Multi-Species Model for Fish and Marine Mammals in the Barents Sea. J. Northw. Atl. Fish. Sci. 22: 317-341.

Gavaris, S., K. R. Patterson, C. D. Darby, P. Lewy, B. Mesnil, A. E. Punt, R. M. Cook, L. T. Kell, C. M. O´Brien, V. R. Restrepo, D. W. Skagen, and G. Stefansson (2000). Comparison of Uncertainty Estimates in the Short Term Using Real Data. CM 2000/V:03

Gislason, H. and Helgason, Þ. 1985. Species interaction in assessment of fish stocks with special application in the North Sea. Dana 5: 1-44.

Patterson, K.R., R.M. Cook, C.D. Darby, S. Gavaris, V.R. Restrepo, A.E. Punt, B. Mesnil, D. W. Skagen, G. Stefansson, M. Smith (2000). Validating three methods for making probability statements in fisheries forecasts. ICES C.M.2000/V:06

Patterson, K., Cook, R., Darby, C, Gavaris, S., Kell, L., Lewy, P, Mesnil, B., Punt, A., Restrepo, V., Skagen, D. W. and Stefansson, G. 2001. Estimating uncertainty in fish stock assessment and forecasting. Fish and fisheries, 2, 125-157.

Restrepo, V.R., K.R. Patterson, C.D. Darby, S. Gavaris, L.T. Kell, P. Lewy, B. Mesnil, A.E. Punt, R.M. Cook, C.M. O´Brien, D.W. Skagen, and G. Stefansson (2000). Do Different Methods Provide Accurate Probability Statements in the Short Term? ICES CM 2000/V:08

Stefansson, G., Sigurjonsson, J. and Vikingsson, G.A. (1997). On Dynamic Interactions Between Some Fish Resources and Cetaceans off Iceland Based on a Simulation Model. Northw. Atl., Fish. Sci. 22 357-370.

Stefansson, G. and Palsson, O. K. (1998). A framework for multispecies modelling of Boreal systems.

Reviews in Fish Biology and Fisheries, 8:1, 101-104.

Part I

Standardised fisheries database

Chapter 1

Organizing the data

Large and complex models are set up in order to incorporate, depict or describe a lot of information in one manner or another. The information is embedded in the models through model equations which have some model parameters. Once the equations have been fixed, the models are still usually too generic to be of any use. Thus, the model parameters must be restricted in order for the models to have practical use.

The incorporation of information through putting restrictions on parameters can be done either by fixing parameters to set values or by estimating parameters through comparisons with data.

The use of assumed values should not be belittled as such methods can be useful for evaluating the influence of different assumptions and no data may be needed for this purpose. This, however, is a biomathematical exercise rather than a biostatistical one, which is the purpose of the present project. Naturally, parameters may be fixed to specific values based on historical knowledge rather than in an ad-hoc manner.

If data are to be used to estimate the various unknown parameters of a multispecies model, then the amount of data needed can become quite enormous and also quite diverse. Routine monitoring of fish stocks typically includes bottow trawl surveys (or acoustic surveys with the auxiliary pelagic trawl), yielding at a minimum data on species composition, length distribution of catches. It is usually also relatively easy to obtain information on sex ratios, individual weight, stomach content (i.e. species and possibly length composition) and in many cases also the age of individual fish. In addition to these survey samples, similar biological samples are routinely obtained from commercial landings (or professional on-board monitoring) as well as records of total landings and effort or catch-per-unit-effort (CPUE). In some cases direct measurements of catches or discards are available.

These data form the fundaments for any stock assessment and for any estimation in multispecies models.

Individual institutes collect these data, each in their own fashion, often predicated by national regislation, tradition or pragmatic considerations due to fleet behaviour, finances or weather. As a result, organisational data bases are highly disparate with regard to format and content.

Any computer program needs a specified format to be used for input data. For a statistical estimation procedure such as Gadget, it is sensible to standardize the input formats. The standardization thus takes place outside the program. However it is normally not sensible to use the same format for data exchange or data storage. The reason for this is that the format used for storage should be more flexible. Notably, one would commonly want to test different levels of aggregation and therefor the data should be stored at a fairly disaggregated level whereas estimation might only use summary data in a given run.

A particular problem with complex and data-hungry models is the need for complex input data, typically in complex input files. Data in such files are highly error-prone and there is much to be said for automating the generation of such files as far as possible.

This has led to the definition of a standardized data base system for fisheries data. Once such a data base has been set up it can be used to generate data files for Gadget.

Within dst^2 case studies, several institutes each had their own data bases and in some cases data were even only available in published papers. Given the differences in data bases, a standardized method for uploading data into the standardized data base was needed. This was done by defining a simple ASCII flat-file format for all data. Upload programs were therefore written to read, validata and upload these data.

The following chapters describe this data base system along with upload and extraction mechanisms.

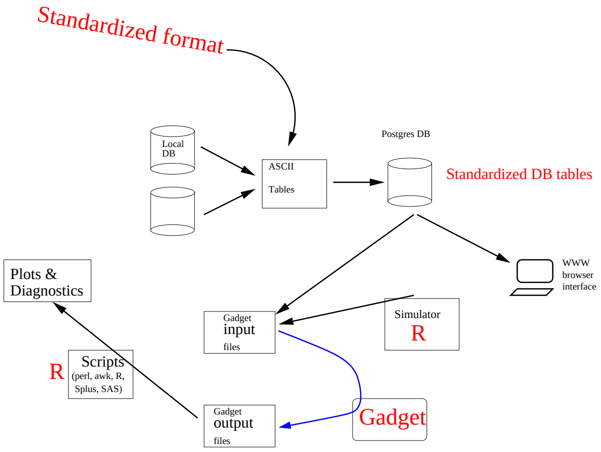


Figure 1.1: Data flow in Gadget models based on real data or simulated input

Following the tradition of international cooperation within fishery science, the data bases are all based on publicly available systems. This ensures that not only can the source code for Gadget be freely used, but also the data base tools.

Chapter 2

Standardised fisheries database

Database structure: The database is hierarchical with a normalized table structure, thus reducing the repetition of information and allowing fast lookup. Rather than use raw data on individual fish or tows, most items correspond to low-level aggregates, such as the number of fish within a length group measured for age. This avoids a complete repetition of the institutional databases, yet the scale and structure should be such that most queries by fishery scientists can be handled by the structure.

Exchange format: An ASCII File Exchange Format (the AFEF) was developed with the intention to create the standard format for all involved parties. Given that this is a flat-file format, no indexing information is included but instead considerable redundancy is used. The tradeoff between speed and complexity is of course quite obvious but it is also obvious that indexing information can not practically be inserted in situations where the ASCII files are generated by hand.

CORBA: Initial project definitions had planned the use of CORBA, XML, JPS etc. As the project developed it was found that better solutions were available for the task at hand. Nevertheless it is of interest to test the use of some of these technologies to link standardised fisheries databases in different countries. The main contribution could be seen in combining several technologies in one application (CORBA / Java / PHP / RDBMS) and using cryptography algorithms (SSL / SSH / HTTPS) to ensure good data security. Such a setup with two hosts was successfully tested at the MRI.

Gadget: From the most important point of view in the dst^2 project (i.e. that of Gadget), the developed database and the import and export tools now allow for the reduction of time of data preparation and extraction (from weeks to hours). Also the data selection is now very flexible depending and is based on setting parameters in a control file.

Improvements: There are some places where improvements can be done. It is the better deletion ability (this may prove to be a complicated task because of the way records are indexed and corresponding information stored in the `KEYS` table) and more query performance tuning. With better understanding of the required functionality recoding some PHP scripts or their parts would be desirable for better readability and structure of the code.

Credits

A list of people who contributed one way or another in the DW part of the dst^2 project follows:

Database structure: The initial database structure was designed by Peter Sandbeck (`pes@dfu.min.dk`) of the Danish Institute for Fisheries Research (DIFRES).

Exchange format: The initial AFEF version was contributed by Alessandro Gimona (`a.gimona@marlab.ac.uk`). An experimental upload program with a web interface was developed by Peter Sandbeck (`pes@dfu.min.dk`) and Andreas Rahelt at DIFRES.

CORBA: The query system was developed at the University of Iceland by Óskar Audunsson (`oskara@hi.is`) and Bergsteinn Einarsson (`bergste@hi.is`), under the guidance of Helgi Thorbergsson (`thorberg@hi.is`). The code clean-up, the first SSL version and class documentation was done by Orri Eiriksson (`orroloth@planewalker.com`).

Data availability: For the Celtic Sea the AFEF data are available from Verena Trenkel (`Verena.Trenkel@ifremer.fr`) and John Pinnegar (`j.k.pinnegar@cefas.co.uk`), for the North Sea from Marco Kienzle (`m.kienzle@marlab.ac.uk`). The standardised fisheries database for Icelandic waters can be freely accessed on-line.

2.1 Datawarehouse Overview

Vojtěch Kupča
Marine Research Institute, Reykjavik, Iceland

Introduction

dst^2 is a 4-year fisheries project aimed at modeling and simulation of marine ecosystem using software called *Gadget*. Gadget needs for its run fair amounts of ecosystem data. These are typically available from the institutional databases. The main disadvantage is that these databases are usually very heterogeneous and often lack robustness and consistency. To enable easier and potentially faster access to these databases, with the stress being put on the creation of Gadget input data files, a solution of DW with aggregated statistical quantities was proposed. The following text will mostly contain Icelandic specific point of view if not stated otherwise.

Data flow

First it makes sense to describe the main stages the data will be passing through. In the dst^2 project, several dependent stages exit. Listed in order, these are:

1. collection of ecosystem data
2. input and storage of data in the institutional database
3. extraction and transformation process
4. ASCII file exchange format (AFEF)
5. upload of AFEF data into DW (`dst2up.php`)
6. extraction of data in Gadget formats (GF) (`dst2ext.php`)
7. Gadget simulation using extracted GF data
8. output of simulation, postprocessing, presentation

This paper contains mainly the information on points 3-6 in the above list, i.e. from the time the data are available in the institutional database until they are used as an input to Gadget.

Data preparation

This stage covers points 2 and 3 in the data flow list. Often the available data are in an unconsolidated state and especially in the dst^2 project an idea of generating GF files directly from the institutional database would shortly prove infeasible due to the above reason. Therefore a better approach is to devise a certain clean-up process that will also transform and prepare the data for an upload into another database, which in our case will be the DW. For the matter of consistency between different involved institutes there was proposed a use of standardized data exchange format in the form of plain text (or ASCII) files (AFEF). AFEF is rather a simple DW input format structured in columns separated by tabulators. Its specification can be found in section 2.2. With a large-scale transformations using hundreds of views and external UNIX tools it was possible to create close to 300 MB (or over 2 million lines) of data in the AFEF for the Icelandic part of the DW.

2.2 Data warehouse architecture

Vojtěch Kupča
Marine Research Institute, Reykjavik, Iceland

Introduction

First let us briefly define our usage of the term *data warehouse*: in general DW is regarded as special type of arrangement of one or more databases that contain data extracted and transformed from heterogeneous data sources. Moreover, the core part ('reconciled' after extraction-transformation process) of the DW is typically a *detailed* and *normalized* database. Denormalization is convenient for a special part of the DW, called a *data mart* (DM). DMs are specially suited for use in analytical decision support and must respond quickly to queries sent to them. Speed is achieved mainly by denormalizing tables of the reconciled data part of the DW and by aggregating queried data (see figures 2.1 & 2.2). In the case of MRI, the extraction and transformation is done by querying institutional database (with the help of views) and storing data in an AFEF. This ensures that other institutes can import data into the same structure of DW. The structure was agreed upon during the first general dst^2 meeting and finetuned in the next year. After inspecting the structure of the final version of DW (see the following section) it can be seen that it is, in fact, a mix of properties of a DM (aggregation) and a reconciled DW part (usually normalized). Since this theoretical description of data warehousing can be considered rather as a rough guideline, it is fully acceptable to call our projected databases DWs, with the possibility to build DMs on top of them, if required later on in the project.

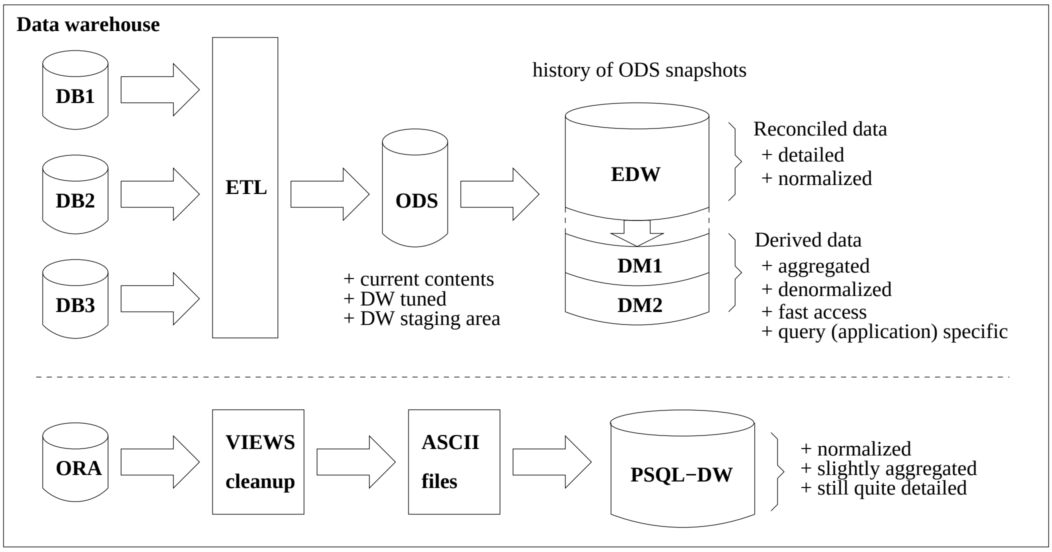


Figure 2.1: General data warehousing architecture and the correspondence to the internal MRI structure. Terms: ETL: extract-transform-load process, ODS: operational datastore, EDW: enterprise data warehouse, DM: data mart, ORA: Oracle database, PSQL-DW: PostgreSQL data warehouse.

Structure

At present the DW is structured around 48 tables (`KEYS` for keeping database consistency, `REP` for database logging, three with min/max length and age information, 23 lookup tables and 20 data tables) that split up as follows:

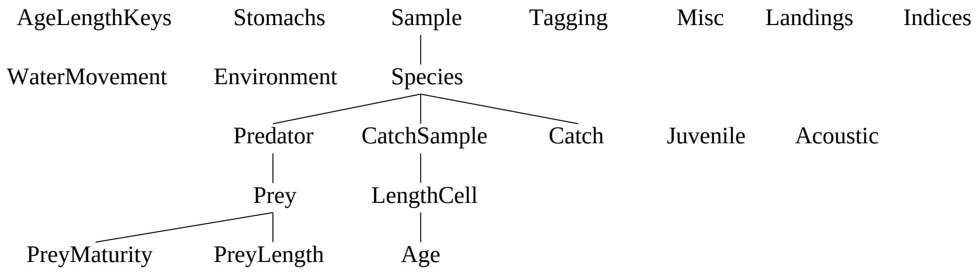


Figure 2.2: **DW tables hierarchy.** Note that table Species is never directly uploaded, but is implicitly populated on the upload of one of the tables one level lower in the hierarchy. This also means that to upload one of the tables Catch, CatchSample, Juvenile, Acoustic or Predator it is enough to have a corresponding row in table Sample.

no.	code	table name	purpose
98	`REP`	report	info on DW operations
99	`KEYS`	keys	table of keys computed from AFEF records
		l_*	23 lookup tables
1	`SAM`	sample	top-level table containing info on time, place, vessel- and gear-types
2	`ENV`	environment	fine-scale (week, depth) data on environmental conditions (temperature, etc.)
3	`WAM`	watermovement	data on water mass movements
4	`CAT`	catch	data on total catches, landings, etc.
5	`CAS`	catchsample	summary data on biological samples (number of fish measured, etc.)
6	`LEC`	lengthcell	biological sample data with lengthcell, sex and mat. stage information
7	`AGE`	age	age-length key data
8	`MSC`	misc	misc data used in stock assessments
9	`ACO`	acoustic	acoustic data
10	`TAG`	tagging	data on fish tagging
11	`JUV`	juvenile	juvenile data
12	`SPE`	species	info on species, stocks and market categories of catch or catch-sample data
13	`PRD`	predator	stomach content data on predator
14	`PRE`	prey	stomach content data on prey
15	`PRL`	preylength	length information about prey
16	`PRM`	preymaturity	maturity and weight information about prey
17	`LND`	landings	official landings data; for the purpose of MRI only
18	`ALK`	agelengthkeys	age-length key data not fitting the main DW structure; used by FRS and IFREMER
19	`STM`	stomachs	stomach content data not fitting the main DW structure; for the purpose of CEFAS
20	`IND`	indices	survey indices; for the purpose of IFREMER
		d_maxleng	maximum length for species
		d_maxages	maximum age for species
		d_minmax	maximum and minimum lengths for species at age

As can be seen, the DW includes biological sample data, catch data, stomach data, tagging data and tables 18 and 19 for the incomplete data that do not fit the common DW structure. The stomach content tables create similar data hierarchy as tables with biological samples. The last three tables in the list (`d_maxages, d_maxleng, d_minmax`) contain periodically updatable data on the minimum and the maximum lengths and ages for a given species used in Gadget output routines.

In more detail, table 18 stores age-length keys which have been created for the Celtic Sea. Timestep and area are not available for these data and to avoid the potential replication of data stored in tables 6 and 7 a separate table was created. Table 18 could also be used to store age-length keys for other areas which could be used whenever aging for a timestep or area is considered insufficient. Primary key data (e.g. timestep, area) are often not available for Celtic Sea stomach data so table 19 was created for the storage of stomach data which cannot be stored in the standard hierarchical structure.

The DW structure resembles natural hierarchy, where biological sampling data are stored in four levels of 1-N relation, meaning that one line from one hierarchy level can have N followers in the table one level lower. This poses certain inevitable restrictions on the speed of data retrieval but on the other hand reduces the amount of stored data in the database (normalization). The tables `CAT, ACO` and `JUV` are in 1-1 relation with `SPE` whose only purpose is to logically divide the data of different origin. For relating individual tables together principles of (compound) primary keys and foreign key references are used. There are also several standalone tables which do not have a reference to any data table. On a table level the overall structure can be well grasped from figure 2.2. For column level description one can refer to web page (2) or read directly the table creation script `dwC.sql` in the DW `Setup` subdirectory.

Software requirements

It was proposed that the DW would be a portable, platform independent database. For this reason there was chosen the Linux(3) operating system and a free RDBMS PostgreSQL (4). Since the PostgreSQL is only an application running on a computer system the portability of DW depends on the particular setup of a system and also the capabilities of individual users. The programming language chosen to develop upload and extraction parts (see sections 2.4 and 2.5) was PHP(5). PHP is an interpreted language designed for generation of dynamic HTML content. Its command line version (sometimes called CGI version) can be used as an interpreter like for instance Perl.

Installation

Before one can install the DW, there have to be met several conditions:

- installed Linux 2.4.x or later (distributed in Red Hat, Slackware or similar)
- installed PostgreSQL 7.3.x or later (and setup one user with database creation permissions)
- installed PHP 4.3.2 or later (compiled with option `-with-pgsql` which includes PostgreSQL database API by using `libpq.so` from PostgreSQL directory tree; it is also necessary to set `register_globals=On` in the PHP initialization file `php.ini`)

Then it is possible to install the database and all its tables. In the DW subdirectory `Setup` there is a script `setup.sh` which can be used to create the database, functions, views and upload data into lookup tables. It is necessary to keep in mind that this script cannot know all the technical details of particular PostgreSQL installation and it may be needed to check the output of setup in the file `setup.log`. For the best control over all taken actions it is advisable to get acquainted with the contents of all files in the `Setup` directory, especially files `dwC.sql` (table definitions), `dwL.sql` (lookup table data) and the `setup.sh`.

Metadata

Metadata contain information about the data stored in the DW. They are split into five topics (biological samples data, stomach data, catch data, acoustic data and tagging data) and include the information on time, position and species in each topic. For biological samples and stomach data the sampling type information is also available. Since these information are mainly aggregated quantities, an on-the-fly querying would take a considerable amount of time. Therefore in the DW subdirectory `Html/Meta` there is an SQL script `dst2mtabs.sql` which creates a set of aggregated meta data tables `meta_samp, meta_stom, meta_catch, meta_tagg` and `meta_acou`. These tables are only several thousand lines in size and their querying takes at most just a couple of seconds. It is convenient to run the table creation script periodically with an arbitrary update period, like e.g. from `cron`. Depending on the particular installation one would use a cron command like:

```
/usr/local/postgresql-7.3.2/bin/psql -f /dw/Html/Meta/dst2mtabs.sql
-d dw0603 -U dst2w >/dev/null 2>&1
```

The whole metadata system is accessible either from the `psql` console or from a PHP web interface created for that purpose. The web interface queries mentioned metadata tables and presents returned results in the table form. It is possible to choose between summary information or manually specify species, year, month and division for a detailed information about individual metadata topics. At the MRI it is currently possible to refer to the metadata website (6). If one wants to change the active database for metadata, it is necessary to adjust connection variables in the `dst2mdefs.php` file and the above mentioned cron entry.

2.3 ASCII file exchange format

Vojtěch Kupča and Lorna Taylor
Marine Research Institute, Reykjavik, Iceland

Introduction

The ASCII file exchange format (AFEF) defines a simple text-based file format for fisheries data suitable for uploading into the standardised fisheries database. The format was designed in such a way that the user does not need to be concerned with the technical details of the database design. This approach has both advantages (limited technical skills of a user needed) and disadvantages (redundancy).

The data are, in most cases, aggregated. This reduces the size of the database and also enables data to be shared between countries as details (particularly from logbook data) are not available.

The structure and content of the database are based on discussions involving all partners at the 2000 project meeting in Nantes. The original definition was then expanded upon by Alessandro Gimona (FRS).

Description

The AFEF format consists of 20 tables and in most cases it follows the hierarchy of database tables described in Section 2.2. Additional tables are also required, known as ‘Lookup tables’, which are static. These lookup tables contain definitions of the codes used in the database tables and data such as substrate type, area (size) of divisions etc. In the following list of tables (starting on page 16), each mandatory column is indicated by NN. Within the hierarchy it is necessary for a corresponding entry to be included in the higher tables for data to be uploaded into the database.

Data can be stored on four levels of spatial aggregation, termed ‘region’, ‘division’, ‘subdivision’ and ‘gridcell’. These must form a hierarchy. In the case of Iceland, region is Iceland, the divisions are large areas which are stratified by depth into subdivisions and gridcells are statistical subrectangles. Of the Icelandic data, most data are stored on the subdivision level and no data are currently stored at the level of gridcell. It is possible to store data at different levels of spatial aggregation. The definition of the spatial hierarchy will depend on the region (information on the definition of areas for Icelandic waters is given in Section 9.1) and the intended use of the database.

All the non-static tables are indexed using the columns `institute`, `year`, `quarter`, `month`, `region`, `division`, `subdivision` and `gridcell`. These columns define when, where and by whom the data were collected.

The main hierarchy of the database contains data from the biological sampling of fish. These tables also include the columns `vesselclass`, `vesselsubclass`, `gearclass` and `gearsubclass`, the combination of which can be used to define a ‘metier’ or fleet. The level of information stored in these columns is at the user’s discretion. An example would be to store vessel length (eg in 3 categories) as `vesselclass` and further detail on the type of vessel as `vesselsubclass`. `gearclass` could be type of gear eg bottom trawl, gill net with `gearsubclass` containing information on mesh size.

The following descriptions are based on the hierarchical table order rather than by table number. Each column will typically only be described the first time it appears in the hierarchy.

Table 1: Sample defines the position and time of the sample along with information on the vessel and gear (as defined above) along with the sampling institute (`institute`). Two additional columns (`areaaggregate` and `timeaggregate`) describe the scale at which data are stored e.g. for

data stored at the level of subdivision “3” would be entered in the `areaggregate` column. Look up tables describe the codes for these columns.

Table 12: Species defines the species, stock and marketcategory. `species` contains the NODC code for the fish species, `stock` can be used if more than stock of a species exists within a region (`stock` codes will be defined for each region where necessary). `marketcategory` can be used to categorise length disaggregated data collected from the commercial fishery. This table is a subset of table 5 and/or 4. See also Table 1.

Table 5: CatchSample defines the source of data and summarises the amount of data available on length and age. These data provide information on the statistical properties of the data. The columns describing the source of data are: `samplingtype` which defines the type of data collected eg harbour, sea and discard for commercial samples and to define surveys eg shrimp survey and spring groundfish survey, and `samplingstrategy` describes the sampling design eg random, stratified, length stratified. Codes for these must be definied in the corresponding lookup tables. The rest of the columns describe the availability of data and number of samples in each sampling unit. `lengthsamplesnum` contains the number of samples for which length data were collected and `fishmeasurednum` the number of fish length measurements. `agesamplenum` contains the number of stations at which fish ages were taken and `fishagednum` the number of fish aged. `weightsamplenum` contains the number of samples where fish were weighed and `fishweightnum` the number of weights available. Intra-haul correlation as described by Pennington and Volstad (1994) means that the number of hauls going into a sampling unit is of interest. See also Tables 1 and 12.

Table 6: LengthCell primarily contains data on length, sex and maturity. The `lengthcell` refers to a length interval which is defined in a lookup table in mm. Species can be stored on different levels of lengthcell intervals, eg shrimp on 1mm and cod on 10mm. `sexcode` is the sex and `maturitystage` the maturity stage, these may not contain the same level of detail as available in the original data source eg maturity stages could be aggregated into immature, prespawning, spawning, spent or simply immature and mature. `fishnum` contains the number of fish within each lengthcell, sex and maturity stage. `weightmean` and `weightmeansd` contain the mean weight (in kg) and standard deviation of weight for fish in that lengthcell. The `surveyindex` and `surveyindexsd` columns enable a survey index to be calculated from the length data. If the length distribution from the survey does not represent the entire catch, the `surveyindex` column can then be used to scale the survey length distribution to the total. `diseaserate` could be used to contain data on eg rates of infection by *Ichthyophonus hoferi* in herring. `gonadweight` is the mean weight of gonads from a sample of `gonadnum` fish. See also Tables 1, 12 and 5.

Table 7: Age contains age data in addition to the length data in Table 6. The lengthcells used in Table 7 must correspond to those in Table 6 for the data to be uploaded into the database. If age data are only available on 5cm intervals but lengths on 1cm, then the length data should be stored on 5cm intervals (or Table 18 used to store age data). All these data must constitute a subset of the length, sex and maturity data in Table 6. `age` is the age, `agenum` the number at age, `weightagemean` the mean weight at age, `weightnum` the number of fish used to calculate the mean weight, `gonadweight` mean weight of gonads and `gonadnum` number of gonads used to calculate the mean weight. See also Tables 1, 12, 5 and 6.

Table 13: Predator defines the source of stomach content data and summarises the number of samples and number of predators from which the data are derived. These data provide information on the statistical properties of the stomach content data. As in Table 5, `samplingtype` and `samplingstrategy` describe the source of data. `samplingtype` may be identical to those in Table 5 with data on the same fish contained in both parts of the hierarchy. For stomach data, `samplingstrategy` may indicate whether stomach data were taken from pooled or individual predators. `stomachsamplesnum` contains the number of samples/hauls the stomach data were taken in as it is informative to know whether one or many hauls are contained in the aggregated data. `predatorlength` is equivalent to `lengthcell` in Table 6 (but not necessarily on the same scale), `predatorsex`, `predatormaturity` and `predatorage` indicate the sex, maturity and age of the predator. Each predator stomach can be categorised as full, empty

or regurgitated (for a predator showing no signs of regurgitation) and the number in each of these categories is stored in `fullnumber`, `emptynumber` and `regurgitatednumber`. The total number of stomachs sampled is then the sum of the `fullnumber`, `emptynumber` and `regurgitatednumber` columns. The `proportionfeeding` can also be recorded. See also Tables 1 and 12.

Table 14: Prey summarises data on the prey. Prey can be categorised into `preygroup` and/or `preyspecies`. `preygroup` is a broad classification eg fish, isopod etc and `preyspecies` a more precise definition. It may only be necessary or useful to store a few prey at the species level. `digestionstage` is code indicating the extent of digestion of the prey. To take account of different sampling methods, it is possible to store data by prey biomass and/or number. `preycount` is the number of a prey and `preyweight` the biomass. `propbynumber` and `propbyweight` are the number (or weight) as proportions of each prey in a sampling unit (position, time, predatortype). See also Tables 1, 12 and 13.

Table 15: PreyLength contains the length distribution of the prey species, potentially by prey sex. `preylength` is the length group of the prey, analogous to `lengthcell` in Table 6. `preysex` is the sex code of the prey and `preynumber` is the number by prey length group and sex. The length distribution can also be stored as proportions in `propldist`. See also Tables 1, 12, 13 and 14.

Table 16: PreyMaturity contains data on the number of prey by length, sex and maturity stage and/or the mean weight of the prey by length, sex and maturity. If this table is used it may replicate some data from Table 15. `preylength` and `preysex` are the same as in Table 15, `preymaturity` is the maturity stage of the prey, `preynumber` is the number of prey by length, sex and maturity and `meanpreyweight` the mean weight of prey by length, sex and maturity. See also Tables 1, 12, 13 and 14.

Table 4: Catch contains catch information based on logbook data and follows Table 12. `vesselnum` contains the number of vessels in the sampling unit and `recordnum` the number of logbook records. These provide some information on the amount of data contributing to the estimate of landings for that spatial and temporal unit. `power` and `grt` cam be used to class the vessels by power and weight and these categories would be described in look up tables. `landings` is the total landings from an spatial/temporal unit according to the logbooks. If available `discards` would also be included. `scaledlandings` are the logbook landings scaled to sum to the official landings statistics for the region. `effort` should be stored in hours and `cpue` in kg/hour. Use of these final 2 columns may, however, depend on gear type. See also Tables 1 and 12.

Table 17: Landings is a record of the official landings statistics for a region. The columns are analogous to those in Tables 1 and 12 with `landings` the landings in kg. This table is independent of the hierarchy.

Table 11: Juvenile contains data from juvenile or egg surveys where data are stored as densities. `m2num` is the number per m^2. See also Tables 1 and 12 for column descriptions.

Table 9: Acoustic contains data from acoustic surveys. Column `sa` contains raw back-scattering coefficient values scaled to the corresponding spatial resolution. `numbers` contains transformed sa-values into numbers of fish. `weight` contains estimated corresponding biomass. This table is independent of the hierarchy. See also Tables 1 and 12 for column descriptions.

Table 10: Tagging contains mark-recapture data and unlike the other tables, data on individual fish are stored. Fish marked but not recaptured should be stored as well those those recaptured. Data are required on the time (`yearreleased`, `quarterreleased`, `monthreleased` and `dayreleased`) and position (`regionreleased`, `divisionreleased`, `subdivisionreleased`, `gridcellreleased`) of tagging, along with data on the fish at time of release (`lengthreleased`, `weightreleased`). For recaptured fish, the time and position of recapture are stored in columns `yearrecaptured`, `quarterrecaptured`, `monthrecaptured`, `dayrecaptured`, `regionrecaptured`, `divisionrecaptured`, `subdivisionrecaptured`, `gridcellrecaptured`. Columns for data on the recaptured

fish are: `lengthrecaptured`, `weightrecaptured`, `agerecaptured`, `sexrecaptured`, `maturityrecaptured` and `gearrecaptured`.

Table 2: Environment will contain data on the environment and zooplankton. This is stand alone table and can contain data on a finer spatial and temporal scale than most other tables. Time can be stored by `week` and it is also possible to store data by depth `depthstratum`. The data are `temperature` in ^{o}C, salinity in PSU and zooplankton in mg dry weight/m^3. As this table is currently not used the units of measurement could be changed and could in any case be different in different regions. These data may be interpolated rather than observed values.

Table 3: WaterMovement This table will contain information on water movement as generated by hydrographical models but is currently not used. As for Table 2, data can be stored by week and depth. The columns allow for the user to define the scale at which data on movement are stored.

Table 8: Miscellaneous This is an independent table. The motivation for it was to provide a place where data used as input to eg annual VPAs could be stored along with the output. The table is therefore able to store eg catch in numbers at age by year along with the year in which the catch in numbers data were used. The columns not previously described are: `sourceyear` the year in which the data were used, `fishlengthmean` mean length at age, `fishweightmean` mean weight at age, `propmature` the proportion mature at age, `catchnumbers` catch in numbers at age, `surveyindex` survey index by age, `cpue` cpue by age, `stocknumbers` the calculated abundance at age and `fishingmort` the calculated fishing mortality. When more than one timeseries of cpue and/or survey indices in used they can be diferentiated through use of columns such as `samplingtype` and `gearclass`.

Table 18: AgeLengthKeys contains age-length key data. Age data may be stored here rather than in Table 7 if the age and length data are available on different scales or when the provenance of age data is not fully known. This table could also be used to store default age-length keys which could be used for spatial/temporal scales with few age data. The columns are analogous to those in Table 7. `calcmethod` can be used to identify the way in which the age-length key was calculated eg precalculated, where only the key is available or from age, length data.

References

Pennington, M. and J. H. Volstad (1994) Assessing the Effect of Intra-Haul Correlation and Variable Density on Estimates of Population Characteristics from Marine Surveys. Biometrics, 50, 725–732

Format

DW tables can be split into several basic categories: **a) 1, 12**: tables holding position, time and species information; **b) 5, 6, 7**: biological sampling data; **c) 2, 3**: environmental information; **d) 4**: catch and logbook data; **e) 13, 14, 15, 16**: stomach content data; **f) 8, 10, 17, 18**: standalone tables; **g) 9, 11**: acoustic, juvenile data.

```
----------------------------

ASCII file exchange format
(AFEF)

                     04/2005

General remarks:

1) columns are separated by
   tabs
2) star (*) denotes key
   attribute set that has to
   be unique
3) NN denotes NOT NULL
   attribute designating
   required information
4) column 0 is a three-letter
   tablecode

-----------------------------


TABLE 1: Sample

  0  SAM
  1* institute         NN
  2* year              NN
  3* quarter
  4* month
  5* region            NN
  6* division          NN
  7* subdivision
  8* gridcell
  9* vesselclass
 10* vesselsubclass
 11* gearclass
 12* gearsubclass
 13  areaaggregate     NN
 14  timeaggregate     NN


TABLE 2: Environment

  0  ENV
  1* institute         NN
  2* year              NN
  3* quarter           NN
  4* month NN
  5* week
  6* region            NN
  7* division          NN
  8* subdivision
  9* gridcell
 10* depthstratum      NN
 11  temperature
 12  salinity
 13  zooplankton


TABLE 3: WaterMovement

  0  WAM
  1* institute         NN
  2* year              NN
  3* quarter           NN
  4* month NN
  5* week              NN
  6* region            NN
  7* divisionfrom      NN
  8* subdivisionfrom
  9* gridcellfrom
 10* depthstratumfrom  NN
 11* divisionto        NN
 12* subdivisionto
 13* gridcellto
 14* depthstratumto    NN
 15  exchangecoef

TABLE 4: Catch

  0  CAT
  1* institute         NN
  2* year              NN
  3* quarter           NN
  4* month
  5* region            NN
  6* division          NN
  7* subdivision
  8* gridcell
  9* vesselclass
 10* vesselsubclass
 11* gearclass
 12* gearsubclass
 13* species           NN
 14* stock
 15* marketcategory
 16  vesselnum
 17  recordnum
 18  power
 19  grt
 20  landings          NN
 21  discards
 22  effort
 23  cpue
 24  scaledlandings


TABLE 5: CatchSample

  0  CAS
  1* institute         NN
  2* year              NN
  3* quarter           NN
  4* month
  5* region            NN
  6* division          NN
  7* subdivision
  8* gridcell
  9* vesselclass
 10* vesselsubclass
 11* gearclass
 12* gearsubclass
 13* species           NN
 14* stock
 15* marketcategory
 16* samplingtype      NN
 17* samplingstrategy
 18  lengthsamplesnum
 19  fishmeasurednum
 20  agesamplesnum
 21  fishagednum
 22  weightsamplesnum
 23  fishweightnum


TABLE 6: LengthCell

  0  LEC
  1* institute         NN
  2* year              NN
  3* quarter           NN
  4* month
  5* region            NN
  6* division          NN
  7* subdivision
  8* gridcell
  9* vesselclass
 10* vesselsubclass
 11* gearclass
 12* gearsubclass
 13* species           NN
 14* stock
 15* marketcategory
 16* samplingtype      NN
 17* samplingstrategy
 18* lengthcell        NN
 19* sexcode
 20* maturitystage
 21  fishnum           NN
 22  weightmean
 23  weightmeansd
 24  surveyindex
 25  surveyindexsd
 26  diseaserate
 27  gonadweight
 28  gonadnum


TABLE 7: Age

  0  AGE
  1* institute         NN
  2* year              NN
  3* quarter           NN
  4* month
  5* region            NN
  6* division          NN
  7* subdivision
  8* gridcell
  9* vesselclass
 10* vesselsubclass
 11* gearclass
 12* gearsubclass
 13* species           NN
 14* stock
 15* marketcategory
 16* samplingtype      NN
 17* samplingstrategy
 18* lengthcell        NN
 19* sexcode
 20* maturitystage
 21* age               NN
 22  agenum            NN
 23  weightagemean
 24  weightnum
 25  gonadweight
 26  gonadnum


TABLE 8: Misc

  0  MSC
  1* institute         NN
  2* year              NN
  3* quarter
  4* month
  5* region            NN
  6* division
  7* subdivision
  8* gridcell
  9* vesselclass
 10* vesselsubclass
 11* gearclass
 12* gearsubclass
 13* samplingtype      NN
 14* species           NN
 15* stock
```

```
 16* age               NN
 17* sexcode
 18* maturitystage
 19* sourceyear
 20  fishlengthmean
 21  fishweightmean
 22  propmature
 23  catchnumbers
 24  surveyindex
 25  cpue
 26  stocknumbers
 27  fishingmort


TABLE 9: Acoustic

  0  ACO
  1* institute         NN
  2* year              NN
  3* quarter           NN
  4* month
  5* region            NN
  6* division          NN
  7* subdivision
  8* gridcell
  9* gearclass
 10* gearsubclass
 13* species           NN
 14* stock
 15  sa                NN
 16  numbers
 17  weight


TABLE 10: Tagging

  0  TAG
  1* taglabel          NN
  2* species           NN
  3* stock
  4  yearreleased      NN
  5  quarterreleased   NN
  6  monthreleased     NN
  7  dayreleased       NN
  8  regionreleased    NN
  9  divisionreleased
 10  subdivisionreleased
 11  gridcellreleased
 12  lengthreleased
 13  weightreleased
 14  yearrecaptured
 15  quarterrecaptured
 16  monthrecaptured
 17  dayrecaptured
 18  regionrecaptured
 19  divisionrecaptured
 20  subdivisionrecaptured
 21  gridcellrecaptured
 22  lengthrecaptured
 23  weightrecaptured
 24  agerecaptured
 25  sexrecaptured
 26  maturityrecaptured
 27  gearrecaptured


TABLE 11: Juvenile

  0  JUV
  1* institute         NN
  2* year              NN
  3* quarter           NN
  4* month
  5* region            NN
  6* division          NN
  7* subdivision
  8* gridcell
  9* gearclass
 10* gearsubclass
 11* species           NN
 12* stock
 13  caughtnum
 14  m2num
```

```
TABLE 12: Species

  0  SPE
  1* institute         NN
  2* year              NN
  3* quarter           NN
  4* month
  5* region            NN
  6* division          NN
  7* subdivision
  8* gridcell
  9* vesselclass
 10* vesselsubclass
 11* gearclass
 12* gearsubclass
 13* species           NN
 14* stock
 15* marketcategory


TABLE 13: Predator

  0  PRD
  1* institute         NN
  2* year              NN
  3* quarter           NN
  4* month
  5* region            NN
  6* division          NN
  7* subdivision
  8* gridcell
  9* vesselclass
 10* vesselsubclass
 11* gearclass
 12* gearsubclass
 13* species           NN
 14* stock
 15* marktetcategory
 16* samplingtype      NN
 17* samplingstrategy
 18  stomachsamplesnum NN
 19* predatorlength    NN
 20* predatorsex
 21* predatormaturity
 22* predatorage
 23  fullnumber        NN
 24  emptynumber       NN
 25  regurgitatednumber


TABLE 14: Prey

  0  PRE
  1* institute         NN
  2* year              NN
  3* quarter           NN
  4* month
  5* region            NN
  6* division          NN
  7* subdivision
  8* gridcell
  9* vesselclass
 10* vesselsubclass
 11* gearclass
 12* gearsubclass
 13* species           NN
 14* stock
 15* marktetcategory
 16* samplingtype      NN
 17* samplingstrategy
 18* predatorlength    NN
 19* predatorsex
 20* predatormaturity
 21* predatorage
 22* preygroup
 23* preyspecies
 24* digestionstage
 25  preycount     NN or 26 NN
 26  preyweight    NN or 25 NN
```

```
TABLE 15: PreyLength

  0  PRL
  1* institute         NN
  2* year              NN
  3* quarter           NN
  4* month
  5* region            NN
  6* division          NN
  7* subdivision
  8* gridcell
  9* vesselclass
 10* vesselsubclass
 11* gearclass
 12* gearsubclass
 13* species           NN
 14* stock
 15* marktetcategory
 16* samplingtype      NN
 17* samplingstrategy
 18* predatorlength    NN
 19* predatorsex
 20* predatormaturity
 21* predatorage
 22* preygroup
 23* preyspecies
 24* digestionstage
 25* preylength        NN
 26* preysex
 27  preynumber        NN


TABLE 16: PreyMaturity

  0  PRM
  1* institute         NN
  2* year              NN
  3* quarter           NN
  4* month
  5* region            NN
  6* division          NN
  7* subdivision
  8* gridcell
  9* vesselclass
 10* vesselsubclass
 11* gearclass
 12* gearsubclass
 13* species           NN
 14* stock
 15* marktetcategory
 16* samplingtype      NN
 17* samplingstrategy
 18* predatorlength    NN
 19* predatorsex
 20* predatormaturity
 21* predatorage
 22* preygroup
 23* preyspecies
 24* digestionstage
 25* preylength        NN
 26* preysex
 27* preymaturity
 28  preynumber      NN
 29  meanpreyweight


TABLE 17: Landings

  0  LND
  1* year              NN
  2* quarter
  3* month
  4* region            NN
  5* vesselclass
  6* vesselsubclass
  7* gearclass
  8* species           NN
  9* stock
 10  landings
```

```
TABLE 18: AgeLengthKeys            7* subdivision                16* samplingtype
                                   8* gridcell                   17* samplingstrategy
  0  ALK                           9* vesselclass                18* lengthcell        NN
  1* institute         NN         10* vesselsubclass             19* sexcode
  2* year              NN         11* gearclass                  20* maturitystage
  3* quarter                      12* gearsubclass               21* age               NN
  4* month                        13* species           NN       22  agenum            NN
  5* region            NN         14* stock                      23  calcmethod
  6* division                     15* marketcategory
```

Lookup table example

Lookup tables control contents of the main DW columns. Note that the following list was edited and should have an informative character only. The full listing is available on the DW homepage (1).

```
l_areaaggregate.txt

1       region
2       division
3       subdivision
4       gridcell
```

```
l_areasofareas.txt

101     52395.71
102     29935.84
103     37101.68
...
2021    264677.73
2031    278527.32
2041    96807.41
```

```
l_digestionstage.txt

101     least digested
102     digested
103     most digested
```

```
l_division.txt

101     DW area 101
102     DW area 102
103     DW area 103
104     DW area 104
105     DW area 105
106     DW area 106
107     DW area 107
108     DW area 108
109     DW area 109
110     DW area 110
111     DW area 111
112     DW area 112
113     DW area 113
114     DW area 114
115     DW area 115
201     DW area 201
202     DW area 202
203     DW area 203
204     DW area 204
301     DW area 301
```

```
l_gearclass.txt

101     long line
102     gillnets
103     hand line
105     danish seine
106     bottom trawl
107     pelagic trawl
109     nephrops trawl
110     purse seine
114     shrimp trawl
115     dredge
117     trap
120     various gears
201     demersal passive gear
202     pelagic passive gear
203     undetermined passive gear
204     demersal active gear
205     pelagic active gear
206     undetermined active gear
301     mobile
302     passive
303     polyvalent bot
304     polyvalent mobile
305     polyvalent passive
```

```
l_gearsubclass.txt

1026    gillnet, 6’’
1027    gillnet, 7’’
1028    gillnet, 8’’
...
305     hooks
306     pts
307     pot
```

```
l_gridcell.txt

-1764   1146
-1763   1146
-1762   1146
...
33D5    3016
33D6    3016
33D7    3016
```

```
l_institute.txt

1       MRI
2       IMR
3       DIFRES
4       SCUI
5       UiB
6       FRS
7       CEFAS
8       IFREMER
9       RIVO
10      ICES
```

```
l_maturitystage.txt

1       immature
2       prespawning
3       spawning
4       spent
6       mature
10      immature shrimp
11      mature shrimp
```

`l_preygroup.txt`

```
101     fish
102     isopods
103     zooplankton
104     amphipods
105     euphausids
106     shellfish
107     echinoderms
108     worms
109     anemones
110     jellyfish
120     other
```

`l_region.txt`

```
1       Icelandic waters
2       North Sea
3       Celtic Sea
4       Irish Sea
5       Atlantic Ocean
```

`l_samplingstrategy.txt`

```
1       simple random
2       stratifed random
3       systematic random
4       cluster
5       individual stomach sampling
6       pooled stomach sampling
```

`l_samplingtype.txt`

```
101     sea sampling
102     harbour sampling
103     sampling by fishermen
110     research, not survey
130     icelandic ground fish survey
131     offshore shrimp survey
133     0-group survey
134     gillnet survey
135     autumn survey
137     inshore shrimp survey
139     capelin survey
201     sea sampling
202     harbour sampling
203     pelagic trawl survey
204     bottom trawl survey
301     survey (sea sampling)
302     market (harbour sampling)
303     mixed (survey + basket)
```

`l_sexcode.txt`

```
M       male
F       female
X       mixed
```

`l_species.txt`

```
1000000000     Fish Larvae
1000000199     Broken Shell
1501000000     Phaeophyceae
...
9218021601     Orcinus Orca
9219020000     Balaenopteridae
9999999999     Tubeworms
```

`l_stock.txt`

```
101     icelandic shrimp inshore
102     icelandic shrimp offshore
110     icelandic herring
111     norwegian herring
201     north sea herring
```

`l_subdivision.txt`

```
1011    DW subarea 101.1
1012    DW subarea 101.2
1013    DW subarea 101.3
...
3014    DW subarea 301.4
3015    DW subarea 301.5
3016    DW subarea 301.6
```

`l_substrate.txt`

```
1       hard rock bottom
2       unconsolidated bottom
3       hard rock shore
...
125     rock and sediment
126     clay and sand
127     undifferentiated solid rock
```

`l_timeaggregate.txt`

```
1       year
2       quarter
3       month
4       week
5       day
```

`l_vesselclass.txt`

```
101     <12m
102     12-24m
103     >24m
201     <12m
202     12-24m
203     >24m
301     <12m
302     12-24m
303     >24m
```

`l_vesselsubclass.txt`

```
101     research
102     commercial
103     commercial, quota
104     commercial, days
105     foreign
106     freezer
201     research
202     commercial
203     foreign
```

2.4 Data upload

Vojtěch Kupča
Marine Research Institute, Reykjavik, Iceland

Introduction

Data upload is done by the PHP program `dst2up.php` found in the DW `Upload` subdirectory. This program recognizes AFEF and uses static column definitions from file `keydef.php` (which, in fact, represent a fixed mapping between AFEF and DW columns) to correctly insert or update rows in the DW. General AFEF file characteristics are (see section 2.3):

- empty lines or lines starting with hash (#) are regarded as comments, and are thus ignored
- each line of the file starts with one of the table codes: `SAM, ENV, WAM, CAT, CAS, LEC, AGE, MSC, ACO, TAG, JUV, PRD, PRE, PRL, PRM, LND, ALK, STM, IND`
- after the table code a tabulator separated list of values follows

The upload of data has to follow the hierarchy of the DW. This, for example, means that data on age (lowest table in the hierarchy) cannot be uploaded until the corresponding information in higher-level tables is available. In such a case upload program will recognize the missing reference from the higher table and report an error. Therefore the tables must be uploaded in the correct order starting from e.g. table `SAM`, followed by `CAS, LEC, AGE`, as appropriate for the hierarchy of tables. Standalone tables `MSC, TAG, ALK, STM, IND` (including `SAM`) can be uploaded without any precautions. Of course, in all tables, values for some fields are searched for in the lookup tables, so a corresponding lookup table record has to exist, otherwise the row fails to upload and the database referential integrity constrain returns an error message. Let us now summarize reasons for which a row may fail to upload:

- missing reference to a higher-level table
- missing lookup table value
- unresolved collision of a key[1] (if switch `-s` is used).
- bug in upload program (use switch `-v`)

As was agreed in dst^2 meetings, the upload program is a *general* upload program meaning that it not only allows adding new rows, but also updating already existing rows. Combining rows from different tables in one file is also possible. To determine if the current row is new or is only rewriting an old one, an initial query has to be sent to the database. This query is somewhat slower and in a case of new row insertion is not necessary, because checking of uniqueness of key attributes[2] is done by database unique constraints. Check is necessary in a case of update where key attributes are the same, only non-key attributes are changed. Therefore, for the insertion of a new record it is sufficient to run upload program with `-s` option, that turns off initial checking. For row update or collision resolution it is necessary to run upload program in its 'normal operation' mode. List of upload program options follows:

---

[1] The upload program tries to speed up management of the database by introducing an idea of hashed key attribute lookup from a special table called `KEYS`. Therefore every uploaded row has a descriptive record in this table that contains full string key (computed as a concatenation of key fields) and a hashed integer key (computed by a `crc32()` hash function). This hashed key is fast to work with, but with a tiny probability has an inconvenience of two or more string keys will collide into the same hash key. The upload program was tested to work fine with one key collision (double collisions should work too, but are very unlikely to occur; estimated at $p = 10^{-8}$).

[2] Key attributes are those that create unique logical groups of data in each table. For example in table Sample columns *year, quarter, month, region, division, subdivision, gridcell, institute, vesselclass, vesselsubclass, gearclass, gearsubclass* create such a group and thus any combination of these values has to be unique.

```
dst2up.php [options] [datafile]

options:
 -h hostname     specify hostname
 -d database     specify database
 -u username     specify username
 -dup [table]    delete database or table and upload into the database
 -del [table]    delete database or table and exit
 -f              force delete (no prompting)
 -s              skip initial key check
 -l              list table codes
 -v              debug output
 -help           print help
```

An example of usage of `dst2up.php` can be:

```
php -q -c ~/bin dst2up.php -d oko2 -s Input/had_tab1.pre |
tee Logs/had_tab1.pre.log
```

In this case flag `-c` says where to look for the `php.ini` file, `-d` specifies the database name `oko2`, `-s` skips the initial check for the existence of the row in the database[3]. Sometimes you may want to keep a log of all operations and at the same time watch the progress on the screen; this is achieved using the UNIX command `tee`. On the screen you can see several symbols that show the operations done on every line of AFEF. Since the upload of one row affects more tables, the whole operation is treated as atomic and implemented as a transaction. It is either committed as a whole, or rollbacked. The symbols you can encounter are: `[S], [!S], [U], [!U], [I], [!I], [C], [!R], [CC]`, that correspond to successful select, update or insert queries (or unsuccessful ones denoted by '!') and either committed or rollbacked transactions, the last symbol informs of a committed double-collision. It is necessary to add that if the complexity of exceptions in a hashing system due to collisions should pass a certain limit, one could think about different hash function (e.g. `rc5`) or discarding the hashing mechanism altogether, possibly at the expense of speed.

During every run a file named as the input with appended `.err` is created. This file contains a list of problematic rows that could not be added for some reason. Typically you want to run the upload program with this file without the `-s` option; collisions or updates should then be resolved, possible missing references will appear in file with extension `.err.err`.

Deletion

Using `dst2up.php` it is also possible to delete data from the whole tables. This process includes deleting data *and* deleting all management information from the `KEYS` table. When you are deleting data manually from `psql` console, always be sure what you are doing. With many records the deletion may take considerable amount of time. For example if you wanted to delete manually all tagging data, you would issue an SQL command:

```
dw=> DELETE FROM tagging; DELETE FROM keys WHERE tablenr = 10;
```

If you wanted to delete more hierarchy levels (say from the lengthcell table), the command would be:

```
dw=> DELETE FROM age; DELETE FROM lengthcell;
DELETE FROM keys WHERE tablenr IN (6,7);
```

Deleting from the age table first would somewhat speed up the deletion. Note that you have to remove all rows in the hierarchy lower then the table lengthcell, i.e. delete the rows from table `KEYS` with

---

[3] It should be noted here that the speed-up is roughly 20% and it is up to the individual if he/she finds this improvement useful.

`tablenr` 6 and 7. For the purpose of deletion and logging, all rows uploaded into the DW have as their last column a unique number `batchid` which is the same for all rows uploaded during the same session (i.e. from the same file). One can use this feature to delete only certain files from the database (using `WHERE batchid = ...`) as they are listed in the table `REP`.

Faster upload

Since the data are more or less static and are not expected to change very much, it makes sense to consider the following simplified upload and delete scenario. After obtaining the data from the institutional database one can upload the part he/she thinks to be fairly static into the database using the upload program. It could be, for example, all species of the years 1950–2000. After the data are uploaded it is possible to backup the database scheme objects using the PostgreSQL utility `pg_dump`. The schema is dumped using the parameter `-s` and it has to be split into two parts, one with tables and the other with constraints (this is necessary for faster upload of data). The data are dumped from individual tables using the parameter `-t`:

```
pg_dump -U dst2w -h haflygna -t sample dw1203 > 01
```

`pg_dump` does not save the sequence numbers when dumping the schema structure. Therefore the sequence numbers have to be obtained by

```
SELECT nextval('sample_sampleid_seq');
```

from the 'static' database and recreated in the new database using

```
SELECT pg_catalog.setval ('sample_sampleid_seq',
  sequence_number, true);
```

for every sequence. This is crucial for resuming the upload into a new database from the place where we left off. Afterwards the `psql` restores the static part of the data from the dumped files (parameter `-f`). The (slow) upload program is then used for the last years only (e.g. 2001 to 2003). The described procedure is a basis of the functionality of several scripts in the DW subdirectory `Static`. Using the shell script `dst2load.sh` it is possible to quickly create the database, its structure and upload most of the needed data. Afterwards, recent ('dynamic') data can be uploaded using the script `Upload/updyna.sh` (may need some adjustments). In fact, if PostgreSQL and PHP were already installed and all required data are extracted from the institutional database, building a DW can require minimum or no user control. In a test environment with 1.6GHz processor the static data (of 1950-2000) were uploaded in about 10 minutes and dynamic (2001-2003) in about 65 minutes resulting in a significant speed improvement. For more installation information refer to the `README` file in the `Static` subdirectory.

2.5 Data export

Vojtěch Kupča
Marine Research Institute, Reykjavik, Iceland

Introduction

This part deals with extraction of data from the DW in the GF. Once there are some data uploaded into the DW one can access them in two main ways:

a) directly using the native command line interface *psql* as a filter. An example would be a query sent directly from the UNIX shell environment:

```
echo "SELECT * FROM sample" | psql -d dw0603
```

This, however, requires some prior knowledge of the structure of the DW tables. It is also necessary to know the format of the Gadget input files, to be able to modify the output of the previous command appropriately. This way of accessing DW can come in handy in certain situations such as checking correctness of the data when other (indirect) access methods are used.

b) Better suited for outputting data in GF is a program called `dst2ext.php` which takes a list of user-specified parameters and outputs the corresponding Gadget file. Using `dst2ext.php` it is also possible to generate matrix formated output, possibly more suitable for further processing. Before running `dst2ext.php` it is necessary to run shell script `dst2dat.sh` from the DW subdirectory `Gadget`. This script creates helper tables `d_maxages, d_maxleng, d_minmax` which are used to obtain limit values for ages and lengths of individual species. The `dst2dat.sh` script should be run whenever a substantial change in the data occurs. For Iceland a typical run would be:

```
dst2dat.sh haflygna dw0603 dst2w IS
```

The parameters supplied by user are hostname, database name, database user and the region code. Afterwards the system is ready to be used by the `dst2ext.php` which has a simple command line syntax:

```
dst2ext.php control_file [-v]
```

Control file

As can be seen from the above, the complexity of parameter setup is moved into a separate *control file*. The control file contains a list of variables (keywords) and values that impose restrictions on the output, and also choose among several output types. The control file format is simply a list of keywords on the left hand side separated by one or more tabulators and a value on the right hand side. If some required keyword is missing or has no value, the program tries to set or obtain a default value from the database. There are two major groups of keywords; one of a general character and another Gadget specific being used in individual GF files. The tables 2.1 & 2.2 contain a complete list of allowed keywords that can occur in the control file.

Table 2.1: Keywords for general parameters

keyword	brief description	example
General parameters		
`hostname`	hostname of the database server	`haffugl`
`database`	database to connect to	`oko1`
`user`	user to log in to the database	`vojtech`
`groupfile`	absolute path to the group file	`/home/dst2/grp.txt`
`filetype`	gadget or stock assessment filetype	`GALD`
`institute`	research institute code	`MRI`
`years`	years to be output	`2000`
`areas`	areas to be worked with	`101`
`timestep`	time aggregation level	`YEA`
`gearclass`	gearclass DW code	`101`
`gearsubclass`	gearsubclass DW code	`101`
`vesselclass`	vesselclass DW code	`101`
`vesselsubclass`	vesselsubclass DW code	`101`
`species`	species ICES code or 3-letter abbrev.	`HER`
`stock`	DW stock number or DW code	`NSH`
`samplingtype`	samplingtype DW code	`101`
`maturitystage`	maturitystage DW code	`1`
`lengthcell`	lengthcell in mm	`20`
`lengthcellmin`	minimum lengthcell in mm	`100`
`lengthcellmax`	maximum lengthcell in mm	`500`
`agemin`	minimum age	`0`
`agemax`	maximum age or plusgroup	`10+`
`lmax`	length upper bound	`1500`
`amax`	age upper bound	`20`
`marketcategory`	marketcategory DW code	`101`
`samplingstrategy`	samplingstrategy DW code	`101`
`sex`	DW sexcode	`F`
`predatormaturity`	predator maturitystage DW code	`1`
`predatorsex`	predator DW sexcode	`M`
`preygroup`	prey group DW code	`101`
`preyspecies`	prey species ICES code or 3-letter abbrev.	`1`
`digestionstage`	digestion stage DW code	`103`
`preylengthcell`	prey lengthcell in mm	`5`
`preylengthcellmin`	prey minimum lengthcell in mm	`80`
`preylengthcellmax`	prey maximum lengthcell in mm	`120`

Table 2.2: Keywords for Gadget data

keyword	brief description	example
Gadget-specific parameters		
`alphabeta`	parameters for weight-length relationship	`8.85e-6 3.0257`
`alkeyscount`	used in initial files; reports alk numbers	`1`
`altthreshold`	limit for querying table `ALK` for age-length keys	`75`
`averageweights`	switch for computing ref. weights from lengths	`1`
`bootstrap`	switch used to enable bootstrapping of areas	`1`
`fleetnames`	fleet labels in fleet file	`net trawl`
`fleetscale`	switch used for rescaling gear-based landings	`1`
`outputdir`	output directory for `*.agg` files	`Input`
`prediction`	last year of prediction	`2006`
`prefix`	used in fleet, prediction, likelihood and stock files	`cod`
`preys`	selection of species for prey file	`CAP HAD *`
`stockname`	label used in the tagging file or as a stock label	`had`
`surveynames`	survey labels in fleet data file and stock-split alks	`IGFS shrimp`
`survey`	survey weights in fleet file	`30000 40000`
`switch`	switch used in prediction file	`1`
`tagarea`	specifies tagging area	`2`
`tagid`	identifier of the tagging experiment	`tagid1`
`tagloss`	tag loss rate used in tag file	`0`
`tagtime`	time of tagging experiment	`1992,4,1992,5`
`temperature`	temperature used in the area file	`5`
`timesteplabel`	label for timestep column used in survey files	`3`
`weightedsums`	use weighted sums for all filetypes with ALK data	`1`
`zerotimestep`	parameter for skipping datasets in output	`1`

A few lines taken from the control file specifying year, species and the output filetype look like:

```
filetype        GALK
years           1998
areas           101,102
species         COD
```

Here `filetype` requests Gadget-formated age-length keys output (for detailed description of all Gadget filetypes it is advisable to refer to section 2.7 or to (1)), areas, years and species are obvious. In a control file keywords can have single values (e.g. `years` 1998) or multiple values (e.g. DW `areas` 101,102). Multiple values are arranged in a *list* separated by comma. However, certain parameters allow multiple *lists* (separated by space) to be assigned to them. For example one may be interested in an age-length keys split by commercial fishing season (spring, autumn). In such a case the output will have two parts, each for one season. It is then possible to define keyword `timestep` as:

```
timestep        1,2,3,4,5 6,7,8,9,10,11,12
```

You can notice a space between months 5 and 6. For a complete example of a control file see (1). One useful feature that has not been fully explained is the keyword `groupfile` which expects a path as a value:

```
groupfile       /home/hafkaldi/vojtech/dw/dst2ext.grp
```

Group file

Group file holds aliases for one or more lists of values of a keyword. Group file has similar syntax to the control file, except it contains three columns – definition type, identifier (alias), group(s). The group file definition for the above example with `timestep` would then be:

```
timestep        SEASON   1,2,3,4,5 6,7,8,9,10,11,12
```

which then makes it possible to use the following line in the control file:

```
timestep        SEASON
```

Allowed values for definition types in the group file are the same as keywords in the control file (except for keywords not directly related to data selection from the DW): `years, areas, timestep, gearclass, gearsubclass, vesselclass, vesselsubclass, samplingtype, maturitystage, stock, sex, predatormaturity, predatorsex, preygroup, digestionstage`. However, for outputting GF data, in one output file you never want to have groups of data split on the following keywords: `vesselclass, vesselsubclass, samplingtype, stock, sex` as these splits require the creation of separate input files for Gadget. Naturally, multiple lists make sense only with `years, areas` and `timestep` keywords and in special cases of `GFLT`, (keywords `gearclass, gearsubclass, samplingtype`[1]) `GLDS` and `GSTO` (`maturitystage, sexcode`) filetypes (for information on filetypes see page 27). If multiple lists appear in Gadget output with the previously mentioned keywords a notice is displayed in the output data:

```
; Notice:  Specified Gadget output contains multi-groups.
```

In another example you may want to see output in a given parameter setup for several years. Therefore it is convenient to define a group file alias `98-02` as:

---

[1] `samplingtype` in `GFLT` output is not directly used for querying but for generating survey information at the end of the file.

```
years    98-02  1998 1999 2000 2001 2002
```

On the other hand line

```
years    98-02  1998,1999,2000,2001,2002
```

would hardly have any use for it aggregates data from years 1998 to 2002 into one combined output. This is, however, convenient with the `areas` definition:

```
areas    ALL    101,102,103,104,105,106,107,108,109,110,111,112
```

If there are keywords composed of multiple lists of values then the number of output datasets[2] will correspond to the Cartesian product of the number of lists in different keywords. For a complete example of a group file, refer to (1).

Filetypes

As mentioned before, the keyword `filetype` specifies the type of output file. Besides the Gadget format filetypes there are also four stock assessment output types. So far the following functions were implemented (the left column contains the value of the keyword `filetype`):

a) for Gadget:

`GALK`	age-length table
`GLD`	length distributions
`GAD`	age distributions
`GAVL`	average lengths at age
`GAVW`	average weights at age
`GSTO`	multi-stock age-length table
`GLDS`	multi-stock length table
`GAVG`	average weights in lengthcell
`GARE`	area file
`GTIM`	time file
`GFLT`	fleet file (total landings by fleets)
`GINI`	initial age-length table and length distributions
`GREC`	average length of recruited fish
`GPRB`	prey biomass
`GPRN`	prey numbers
`GPRL`	prey lengths
`GPRW`	prey biomass (from `GPRL` by $w = \alpha\, l^{\beta}$)
`GPDB`	predators by lengthcell based on prey biomass
`GPDN`	predators by lengthcell based on prey numbers
`GPDL`	predators by lengthcell based on prey lengths
`GPDW`	predators by lengthcell based on prey weights
`GTAG`	tagging file
`GTAD`	tagged fish by lengthcell
`GTAR`	recaptured fish

b) for stock assessment:

`ALT`	age-length table
`ALK`	age-length keys matrix
`ALD`	age-length distributions matrix and length distributions
`CNR`	catches in numbers by age

---

[2] Dataset is in most cases an output for a unique combination of `years,timestep` and `area`.

The difference between **a)** and **b)** is in the formating of output. While GF is defined by Gadget, the stock assessment output data are arranged in a matrix or a table.

General DW keywords

Let us have an example in which we want to specify the range of fish lengths and ages to be used. It is possible to use keywords `lengthcell, lengthcellmin, lengthcellmax, agemin` and `agemax`. For instance in a case of `GALK` filetype, setup

```
filetype        GALK
lengthcell      20
lengthcellmin   450
lengthcellmax   510
agemin          7
agemax          9+
lmax            1500
amax            20
```

would result in an output like:

```
;
; year  step    area    age     length  number
;
1998    1       1       age7    len1    3
1998    1       1       age7    len2    3
1998    1       1       age7    len3    10
1998    1       1       age8    len1    0
1998    1       1       age8    len2    1
1998    1       1       age8    len3    0
1998    1       1       age9    len1    0
1998    1       1       age9    len2    0
1998    1       1       age9    len3    0
```

We see that with `GALK` filetype, age and length columns are assigned labels. These are defined in files `len.agg` and `age.agg` which are created during the run of `dst2ext.php`. In the case of age labels, numbers are corresponding to real ages:

```
age7    7
age8    8
age9    9
```

For length labels the numbering has no relation to real lengths and increases starting from one. The last interval is open on the upper bound so that the length of 510 (or 51cm) is not included in the selection.

```
len1    45      47
len2    47      49
len3    49      51
```

Based on the type of output other files are also output: `area.agg` (labels for individual areas) and `allarea.agg, alllength.agg, allage.agg` (labels for all areas, all lengths and all ages). It is necessary to remark that these files are created 'silently' for each run of `dst2ext.php`. They always appear in the output directory specified by the control file keyword `outputdir` (if left empty, directory

`Input` is taken as a default value and is automatically created if necessary). It can be easily seen that running `dst2ext.php` more times continually overwrites already existing files unless the value of `outputdir` or `filetype` has changed. This is especially true with helper script `dst2input.php` that potentially makes the generation of GF files easier (for more on `dst2input.php` see section 2.5). The usage of `dst2ext.php` is however recommended whenever the direct control over the input parameters and output files is needed.

For filetypes `GALK, GLD, GAD, GSTO, GLDS, GAVL, GAVW` two other files are also created. First it is the `likelihood` file with likelihood component information and the second is a default `penaltyfile`. In `likelihood` file, the sections corresponding to different filetypes are added in each run. If the section for a given filetype already exists then only an update of this section is made. Based on the survey or catch output type, the header of each section contains a string of the form `filetype.suffix`, where `suffix` is obtained from the `surveynames` or `fleetnames` keyword. The header information should not be changed during the extraction of data, since it is used by the extraction program to differentiate different sections of the `likelihood` file. An example of a `GALK` section header follows:

```
;
; DW
; GALK.net
```

For filetype `GFLT` an extra file `fleet` with information on all fleet components is created. The name of the file is prefixed by the value of the keyword `prefix`. For a detailed description of all Gadget-specific keywords see page 30.

If we continue now in the above example we can demonstrate that stock assessment filetype `ALT` would produce an age-length table (matrix) with absolute numbers of fish.

```
; [ALT]
; 450   470     490
3       3       10
0       1       0
0       0       0
```

Notice that in this example `agemax` is specified as a plus group so that the last row contains cumulative sums for all ages equal to or greater then 9. If `lengthcellmax` and/or `agemax` are omitted, their value is then one of the two possibilities: **a)** user specifies fixed maximum values using the control file keywords `lmax` and `amax` (`amax` is also used as an upper bound of age plusgroup). **b)** if no values are specified by `lmax` and `amax`, the program sends an extra query to the database to obtain the maximum age and length.

Also notice that `lengthcell` is 20 meaning that the resulting length groups will be spaced by 2cm intervals. If we define a *step* to be the difference of two consecutive lengths stored in the database, then it holds that `lengthcell` has to be $>=$ *step* for a given species (*step* in mm for different species is: shrimp = 0.5, capelin = 5, herring = 5, other species = 10). This ensures that length groups are never smaller then what the database has to offer. If `lengthcell` is not specified, 10mm is used as the default value.

For `GREC` filetype it is necessary to pay attention to the handling of years specified in the control file. In the case of Iceland, years less then 1985 (where IGFS data are not available) are all assigned an average of the years 1985 to 1989. For later years, either individual values or averages over several years can be taken. The following example demonstrates the idea.

Control file:

```
filetype   GREC
years      1984 1985,1986 1987 1988,1989,1990
```

Simplified `GREC` output:

```
; yr  ...  meanlen
1984       20.2       ; average of 85-89
1985       21.3       ; average of 85-86
1986       21.3       ; average of 85-86
1987       18.7       ; 87 alone
1988       19.9       ; average of 88-90
1989       19.9       ; average of 88-90
1990       19.9       ; average of 88-90
```

Finally let us remark that with some keywords (currently `gearclass`, `gearsubclass`, `vesselclass`, `vesselsubclass` and `timestep`) it is possible to use the reserved word `NULL` as a value. `NULL` allows to select an empty (unknown) value for a given attribute.

Gadget-specific DW keywords

Apart from the general keywords there are about 20 which have an influence on the output in particular Gadget files. First we give a short description of each of them and then in the following subsections show the usage of some of them in the stomach content files and tagging files.

`alphabeta`

> α and β parameters for weight-length relationship $w = \alpha l^\beta$. It can be used in the `GREC` filetype, `GAVG` filetype (see `averageweights`) and `GPRW` filetype. It is assigned a *space* separated values of α and β.

`alkeyscount`

> Used in `GINI, GREC, GAVL, GAVW` to show how many age-length keys were used for computation of each record.

`altthreshold`

> Sometimes it can happen that the age-length key data will be too scarce and will mostly contain just 0's or 1's. Using such data could result in a poor age-length distributions and mean lengths calculations. `altthreshold` controls the level at which auxiliary age-length key table `ALK` starts to be queried for possibly better result. The value of `altthreshold` can either be a proportion of 0's and 1's in the whole table (value 0–1) or the minimum sum of all elements in the matrix (values greater then 1); the value of 1 disables querying of table `ALK` (this is also the default value).

`averageweights`

> Says if the reference weights in `GAVG` filetype should be computed from weight-length relationship (value 1) or not (value 0, default). If set to 1 the `alphabeta` keyword has to be set as well.

`bootstrap`

> If set to 1, enables area bootstrapping (default is 0).

`fleetnames`

Contains space separated list of fleet labels which are used in the `GFLT` filetype, the corresponding `fleet` file and the `likelihood` file.

`fleetscale`

Used in the `GFLT` filetype, it rescales the gear-based catch number to year/species totals.

`outputdir`

Sets the output directory where `dst2ext.php` stores all aggregate files and where the `dst2input.sh` (see page 34) places all data files.

`prediction`

If needed, sets the last year of prediction. Output files `GARE, GTIM, GFLT, GREC` and prediction file are changed accordingly.

`prefix`

A common prefix used by `dst2input.sh` for most of the created GF files. Also used in the `GSPE` file and the `likelihood` file.

`preys`

Specifies what preys should be used for stomach content related files. `preys` contains the species labels separated by space. The value of `*` represents all prey species (see page 32).

`stockname`

The value is used in the `GSTO` and `GLDS` files, the `likelihood` file and the `GTAG` file. More values are separated by space.

`surveynames`

The value is used in the `GFLT` file, the fleet file and the `likelihood` file. More values are separated by space.

`survey`

Used in the fleet file, contains a space separated list of weights assigned to surveys specified by `surveynames`. The numbers and order of values in both keywords have to match.

`switch`

Switch used as the value of the last column (called `q`) in the prediction file.

`tagarea`

The integer value (less then or equal to the total number of areas) specifies the current tagging area.

`tagid`

The label used in all tagging files to identify the tagging experiment (see page 33).

`tagloss`

The label used in the `GTAG` file (see page 33).

`tagtime`

Specifies the time of the tagging experiment. It contains space separated year/month pairs in which the tagging experiment took place (see page 33).

`temperature`

Used in the `GARE` file.

`timesteplabel`

If non-empty its value is used for timestep column in all survey likelihood files, all stomach data files (`gearclass` and `gearsubclass` have to be empty) and `GREC` file.

`weightedsums`

If set to 1 and `lengthcell` is greater then 10 (=1cm), then all age-length keys based files are summed differently (see page 35 for details).

`zerotimestep`

If set to -1 it keeps all datasets; if set to n (for $n >= 0$) it removes a dataset if the sum of its elements is $<= n$.

Stomach content filetypes

As mentioned on page 27 there are four predator `GPDB, GPDN, GPDL, GPDW` and four prey filetypes `GPRB, GPRN, GPRL, GPRW`. The first two of each group are obtained from tables 13 and 14: `PRD` and `PRE`. They require the predator information to be specified by keywords `species, lengthcell, lengthcellmin` and `lengthcellmax` and the prey information by keyword `preys`. `preys` contains space separated list of species the user wants to select. Special value of `*` (asterisk) defines all other species. If `*` is included in the selection with the `GPRB` filetype then the species proportions in each key group have to add up to 1. For `GPRN` the absolute numbers of preys eaten by chosen predator are shown.

For filetypes `GPDL, GPDW, GPRL` and `GPRW` the table `PRL` (with prey lengths) is also queried. A set of extra parameters is used in these filetypes to define prey lengths and species:

```
preyspecies           HAD
preylengthcell        20
preylengthcellmin     80
preylengthcellmax     140
alphabeta             0.0000065 3.07
```

In this case the `preyspecies` is used to specify a *single* prey species in the `GPDL, GPDW, GPRL` or `GPRW` files (notice the difference from `GPDB, GPDN, GPRB` and `GPRN` where *multiple* species can be selected using the keyword `preys`). For prey species there are defined different minimal steps then for predators. The smallest difference in lengthcells are 1mm for nephrops and shrimp and 5mm for other prey species. This directly determines the minimum value of `preylengthcell` keyword. The values of predator and prey lengthcells can be combined to get the desired aggregated output:

```
;
; year  step    area    predlen preylen number
;
1998    1       area1   len1    plen1   1
1998    1       area1   len1    plen2   2
1998    1       area1   len1    plen3   0
1998    1       area1   len2    plen1   1
1998    1       area1   len2    plen2   1
1998    1       area1   len2    plen3   3
```

In the above example of `GPRL` filetype we can define `lengthcell = 30`, `lengthcellmin = 600` and `lengthcellmax = 650` to obtain two predator lengthcells and three prey lengthcells. The definition of labels used in the output is found in aggregation files created in the specified output directory (keyword `outputdir`). Based on the filetype they are called `preyb.agg, preyn.agg, preyl.agg` or `preyw.agg`. The
`alphabeta` keyword was not used here as it is only required by the `GPRW` filetype to derive weight proportions using the weight-length relationship.

Tagging filetypes

The tagging files are generated using filetypes `GTAG, GTAD` and `GTAR`. Tagging consists of one or more tagging experiments. It is possible to extract individual tagging experiments, not all of them. This is considered to be a better way as some extra keywords would have to be otherwise introduced. To easily combine the outputs of several tagging experiments shell command `cat` can be used. As an example of one tagging experiment let us have the following control file setup:

```
areas            101,102 103,104
timestep         1,2,3 4,5,6 7,8,9 10,11,12
stockname        codmat
tagid            tagid1
tagarea          2
tagtime          1992,3,1992,4,1993,3
tagloss          0
lengthcell       20
lengthcellmin    400
lengthcellmax    600
```

In `GTAG` output the user specifies dates of the tagging experiment using keyword `tagtime`. It is a comma separated list of year,month pairs as shown in the example above. The value for `tagarea` is obtained from the corresponding keyword in the control file. The database is then queried for the last recapture associated with all given tagging dates in a given tagging area. The values of minimum tagging date and maximum recapture date are then used for the values of `tagyear, tagstep, endyear` and `endstep`.

In `GTAD` filetype, the tagging dates and areas are again defined by `tagtime` and `tagarea`. In our example the `tagarea` value of 2 selects the area 103,104. `lengthcell, lengthcellmin` and `lengthcellmax` define the length of the tagged fish. All zero values (the lengths at which there were no tags released) in the output are skipped.

Selection of the recaptured fish is based on the same release parameters like in the `GTAD` filetype, i.e. `tagid, tagtime, tagarea`. In addition keyword `gearclass` selects the type of gear used for recapture (if available). Areas are assigned labels according to which area group they fall into as specified by `areas`. Records with areas outside of the working region are removed. The same is also true if the recapture month is not known. Recapture lengths are assigned labels based on the lengthcell information (`len.agg` is created). Unknown recapture lengths are usually assigned -1. If, however, the `lengthcell` covers the whole interval between `lengthcellmin` and `lengthcellmax`, then the recaptured fish (including those of unknown recapture lengths) will all be assigned to the `len1` length group.

For all tagging filetypes the value of timestep is computed from the month of release or recapture and the number of timesteps defined by `timestep` keyword. For example, if the release month in `GTAD` was 4, then based on the above definition of `timestep` the assigned output timestep value will be 2 (uniformly distributed timestep is assumed).

Debugging

For the purpose of debugging and inspecting in detail all actions of `dst2ext.php` there is an optional parameter `-v`. It displays all SQL queries that were sent to the database to obtain required result. The queries are only printed and are not actually executed. This option can be also useful for comparing output GF data with the contents of the database.

Helper script

For easier control over generated files there exists a wrapper shell script `dst2input.sh` for `dst2ext.php`. It takes several command line arguments as follows:

```
./dst2input.sh control_file output_dir filetype(s)
```

`control_file` is the path to the control file, `output_dir` is the directory for placing output files (in a case the directory does not exist, user is interactively prompted for its creation), `filetype(s)` is a space separated list of Gadget files that should be output. These are:

`GSPE, GALKc, GALKs, GLDc, GLDs, GADc, GADs, GAVLc, GAVLs, GAVWc, GAVWs, GSTOc, GSTOs, GLDSc, GLDSs, GAVG, GREC, GINI, GARE, GTIM, GFLT, GPDB, GPDN, GPDL, GPDW, GPRB, GPRN, GPRL, GPRW, GTAG, GTAD, GTAR`.

The idea behind the wrapper script is to be able to generate as many files from one control file as possible. It can be seen that some filetypes have appended `c` (catch) or `s` (survey) to their names. `dst2input.sh` takes the specified control file and in a case of catch files substitutes value of `101`, `102`, `103` for `samplingtype` (101 is sea sampling, 102 is harbour sampling, 103 is sampling by fishermen) and takes values of `gearclass` and `gearsubclass` unchanged. In a case of survey file it takes the specified `samplingtype` and leaves `gearclass` and `gearsubclass` unchanged. With `GAVG` (average weights), `samplingtype 130`, `135` (IGFS, autumn survey) is taken. For other filetypes unchanged parameters are used. These settings are tuned for Iceland, for specifics of other regions refer to page 35.

Names of the files are in most cases created with a prefix, stem and suffix. Prefix is obtained from the control file keyword `prefix` while suffixes of survey and catch files are substituted for the first value of the control file keywords `surveynames` and `fleetnames`, respectively. For tagging files the value of keyword `tagid` is taken as their prefix. To combine more tagging experiments (with different prefixes) together, shell command `cat` can be used as shown below:

```
# first tagging experiment

dst2input.sh Control/tagid1.ctl Input GTAG GTAD GTAR

# second tagging experiment

dst2input.sh Control/tagid2.ctl Input GTAG GTAD GTAR

# to combine both experiments together

cat Input/tagid1.tagfile  Input/tagid2.tagfile  > tagfile
cat Input/tagid1.tag.data Input/tagid2.tag.data > tag.data
cat Input/tagid1.tag.like Input/tagid2.tag.like > tag.like
```

As can be seen files `tagfile, tag.data` and `tag.like` contain both tagging experiments combined.

Names of created files are in most cases prefixed by a string to help distinguish between sets of data files for different models. As a common prefix, the value of the control file keyword `prefix` is used. For a complete list of output files of `dst2input.sh` see page 36.

Regional differences

Several changes and additions that reflect the region differences are incorporated into the extraction program.

For the **Celtic Sea** these are:
a) The year / species total landings are obtained from the table 17: Landings in `GFLT` output. If there is no corresponding information in table 4: Catch on proportions among individual gears and more then 1 timestep is used, then the total landings amount is split uniformly over all specified timesteps.
b) Age-length table is obtained from table 18: AgeLengthKeys.
c) In `GREC` removed 1985 as the start year of surveys (as used in Iceland for IGFS).
d) In `GPRN` the prey numbers (`preycount`) are obtained from the table 19: Stomachs.
e) In `GPRL` the prey lengths are obtained from table 19: Stomachs.
f) In `dst2input.sh` the survey and the catch samplingtypes have been assigned codes 301, 303 and 302, 303.
g) For unavailable weight data it is still possible to create the reference weights filetype `GAVG` by setting the keyword `averageweights` to 1 and specifying alpha and beta using the `alphabeta` keyword. Length are then computed from the weight-length relationship $w = \alpha\, l^{\beta}$.
h) For `GINI` the year 1985 is not used (like in Iceland) as the start year.
i) It is possible to use `weightedsums` for other way of adding age-length keys for `lengthcell` higher then the basic step. The following example explains the computations:

	`lengthcell = 10`		`lengthcell = 20`
age / length	500	510	500
3	4	1	(4/6+1/11)*17/2 = 6.44
4	2	8	(2/6+8/11)*17/2 = 9.01
5	0	2	(0/6+2/11)*17/2 = 1.55

For the **North Sea** the changes are:
a) In `dst2input.sh` the survey and the catch samplingtypes have been assigned codes 202, 203, 204 and 201.
b) Some age-length data can be obtained from table 18: AgeLengthKeys.

The main month of survey and the start year of survey (used for computation of survey timestep in the `GFLT` file) are defined for each samplingtype (specified using the `samplingtype` keyword) by the following table:

samplingtype	main month	first year	description
130	3	1985	IGFS
131	7	1987	offshore shrimp survey
133	8	1989	0-group survey
134	4	1996	gillnet survey
135	10	1995	autumn survey
137	10	1988	inshore shrimp survey
202	3	1990	harbour sampling
203	4	1991	pelagic trawl survey
204	1	1985	bottom trawl survey
301	3	1983	survey (sea sampling)

Web interface access

For the convenience of looking at or analyzing raw data in the DW, a short php script `dst2web.php` was developed, that allows users to access DW content through a web-based interface. The functionality includes sending an SQL command to the database, retrieving its result and presenting it in an HTML table. Users can also browse results by pages or save all returned rows into an ASCII file. For easy use there are predefined table aliases and groups of useful joins that enable the composition complex multiline queries within seconds. Note that this script requires an installed web server(7) with a compiled PHP module; `register_globals=On` have to be set in the `php.ini` file. One can check if the settings are correct by running a one-liner php script `<? phpinfo(); ?>`.

dst2input.sh

```
Extraction of Gadget input data from DW

usage: ./dst2input.sh control_file output_dir filetype(s)

filetypes: GARE : area
           GTIM : time
           GSPE : ^
           GAVG : ^.refweights
           GINI : ^.init
           GREC : ^.rec
           GFLT : ^.fleet.data
           GALKc: ^.alkeys.$
           GALKs: ^.alkeys.$
           GLDc : ^.ldist.$
           GLDs : ^.ldist.$
           GADc : ^.adist.$
           GADs : ^.adist.$
           GAVLc: ^.meanle.$
           GAVLs: ^.meanle.$
           GAVWc: ^.meanwg.$
           GAVWs: ^.meanwg.$
           GSTOc: ^.alkstock.$
           GSTOs: ^.alkstock.$
           GLDSc: ^.ldstock.$
           GLDSs: ^.ldstock.$
```

```
GPDB : ^.predatorb
GPDN : ^.predatorn
GPDL : ^.predatorl
GPDW : ^.predatorw
GPRB : ^.preyb
GPRN : ^.preyn
GPRL : ^.preyl
GPRW : ^.preyw
GTAG : *.tagfile
GTAD : *.tag.data
GTAR : *.tag.like
```

Remarks:
List of GF files which are possible to generate using `dst2input.sh`. Prefixˆis substituted for a value of the control file keyword `prefix`, suffix `$` takes the value of `surveynames` or `fleetnames`. Tagging prefix `*` corresponds to the value of keyword `tagid`. At the end of the command line follows a space separated list of files to be generated; an example run could be `dst2input.sh Control/dst2ext.ctl Input GAVLs GALKs GLDs`.

dst2bs.sh

Using the shell script `dst2bs.sh` one can create many control files at once with areas randomly selected with replacement. `dst2bs.sh` is based on the Linux kernel's random number generator which gathers certain unpredictable environmental noises from hardware I/O devices to output close-to random data through the character device `/dev/urandom`. The first word is taken from the output (variable `rndnr`) modulo current number of areas (variable `mod`) computing the index of an area to be chosen.

2.6 Bootstrapping of areas

Vojtěch Kupča
Marine Research Institute, Reykjavik, Iceland

Abstract: A bootstrapping method is introduced for use with estimation procedures which use data aggregated over spatial units in the DW.

Introduction

The bootstrap (see e.g. Efron and Tibshirani 1993) is a general tool which can be used to estimate uncertainty associated with an estimation procedure such as Gadget. A general method for bootstrapping data sources as complex as those used in Gadget has been proposed by Kupča et al. (2005). This section describes the implementation.

Bootstrapping can be enabled in the extraction program by setting the keyword `bootstrap` to 1. It is effective on the following actions (see section 2.7): `alkeys` and `length` (for filetypes `GALK, GLD, GAD, GAVL, GSTO, GLDS, GINI, GREC`), `lndngs` (`GFLT`, for `fleetscale` set to 0), `preywg, preyln` (`GPDB, GPDN, GPDL, GPDW, GPRB, GPRN, GPRL, GPRW`), `tagrel, tagrec` (`GTAD, GTAR`). To allow for fast creation of many control files needed for bootstrapping of areas, there is a script `dst2bs.sh`. It is used as follows:

```
./dst2bs.sh control_file output_dir number area_label
```

where `control_file` is a location of the source control file, `output_dir` is a directory to put all the created files to, `number` says how many iterations we want to do and `area_label` is the groupfile label to be expanded for areas. For example, if we consider a 2-area case (where groupfile label `TWOAREA` expands to `101,102,103 104,105`), then `dst2bs.sh ctlfiles/celtic.ctl boot 3 TWOAREA` will create three files suffixed by a sequence number in the directory `boot`. For technical details of the implementation of bootstrapping see page 132.

Bootstrapping – technical details

To allow for area bootstrapping one has to rethink the way a query is processed. Normally when selecting areas (divisions or subdivisions) it is specified in the `WHERE` clause which areas should be considered. In the PostgreSQL `WHERE` acts as a filtering condition on returned rows, which means it passes on only those rows for which the condition is evaluated as `TRUE`. The duplicates are removed in the output. With bootstrapping, however, we typically wish to select areas based on the random selection *with replacement* where areas can repeat, as it is shown in the following example:

```
SELECT col FROM tab WHERE areas IN (101,101,102)
```

Possibly the only way to keep the duplicate rows in the output is to rewrite the query using `UNION ALL` as:

```
SELECT col FROM tab WHERE areas IN (101)
UNION ALL
SELECT col FROM tab WHERE areas IN (101)
UNION ALL
SELECT col FROM tab WHERE areas IN (102)
```

If the aggregation is also included in the query, then there have to be two SELECT levels of which the outer SELECT re-aggregates and orders the partial UNION ALL selections. The final query would then be:

```
SELECT col FROM
  (SELECT col FROM tab WHERE areas IN (101)
   UNION ALL
   ...
  ) AS fromlabel
GROUP BY ...
ORDER BY ...
```

Moreover, aggregation type in the outer SELECT changes or stays unchanged as follows: **a)** on `avg` the bootstrapping has no effect; this concerns the action `avgwgt` in filetype `GAVG`, **b)** `sum` stays as it is, **c)** `count` changes to `sum`.

References

Efron, B., R.J. Tibshirani. 1993. An introduction to the bootstrap. Chapman and Hall, New York. 436p.

Kupča, V., Trenkel, V., Taylor, L. and Stefansson, G. 2005. Bootstrapping multiple data sources for estimating uncertainty in parameter estimates in complex models. Section 5.6 of this report.

2.7 Gadget input filetypes

Vojtěch Kupča
Marine Research Institute, Reykjavik, Iceland

```
+--------------------------------------------------------------------------------------------+
| type  | action | description    | parameters      | joining | files                        |
|-------+------------------------------------------------------------------------------------|
| GALK  |                                                                                    |
|-------+------------------------------------------------------------------------------------|
|       |        | selects number |                 |         |                              |
|       |        | of aged fish   |                 |         |                              |
|       |        | to obtain      |                 |         | $prefix.alkeys.$survenames   |
|       |        | age-length     |                 |         | (GALKs)                      |
|       |        | table, for     |                 | SAM >   | $prefix.alkeys.$fleetnames   |
|       |        | stock          |                 | SPE >   | (GALKc)                      |
|       | alkeys | assessment     | all             | CAS >   | len.agg                      |
|       |        | also           |                 | LEC >   | age.agg                      |
|       |        | converting     |                 | AGE     | area.agg                     |
|       |        | numbers to     |                 |         | likelihood                   |
|       |        | proportions to |                 |         | penaltyfile                  |
|       |        | get age-length |                 |         |                              |
|       |        | keys           |                 |         |                              |
|-------+------------------------------------------------------------------------------------|
| GLD   |                                                                                    |
|-------+------------------------------------------------------------------------------------|
|       |        |                |                 |         | $prefix.ldist.$survenames    |
|       |        |                |                 |         | (GLDs)                       |
|       |        |                |                 | SAM >   | $prefix.ldist.$fleetnames    |
|       |        | selects length |                 | SPE >   | (GLDc)                       |
|       | length | distributions  | all             | CAS >   | len.agg                      |
|       |        |                |                 | LEC     | area.agg                     |
|       |        |                |                 |         | allage.agg                   |
|       |        |                |                 |         | likelihood                   |
|       |        |                |                 |         | penaltyfile                  |
|-------+------------------------------------------------------------------------------------|
| GALD* |                                                                                    |
|-------+------------------------------------------------------------------------------------|
|       | length | see GLD        |                 |         |                              |
|       |--------+----------------+-----------------+---------|                              |
|       | alkeys | see GALK       |                 |         |                              |
|       |--------+----------------+-----------------+---------|                              |
|       |        | rescaling      |                 |         |                              |
|       |        | age-length     |                 |         |                              |
|       |        | keys (from     |                 |         | len.agg                      |
|       |        | alkeys output) |                 |         | age.agg                      |
|       |        | using length   |                 |         | area.agg                     |
|       | sld    | distributions  |                 |         |                              |
|       |        | to get         |                 |         |                              |
|       |        | age-length     |                 |         |                              |
|       |        | distributions  |                 |         |                              |
|       |        | (from length   |                 |         |                              |
|       |        | output)        |                 |         |                              |
|-------+------------------------------------------------------------------------------------|
| GAD   |                                                                                    |
|-------+------------------------------------------------------------------------------------|
|       | length | see GLD        |                 |         | $prefix.adist.$survenames    |
|       |--------+----------------+-----------------+---------| (GADs)                       |
|       |        |                |                 |         | $prefix.adist.$fleetnames    |
|       | alkeys | see GALK       |                 |         | (GADc)                       |
|       |        |                |                 |         | age.agg                      |
|       |--------+----------------+-----------------+---------| area.agg                     |
|       |        | computing sums |                 |         | alllength.agg                |
|       | agesum | at ages from   |                 |         | likelihood                   |
|       |        | alkeys output  |                 |         | penaltyfile                  |
|-------+------------------------------------------------------------------------------------|
| GAVL  |                                                                                    |
|-------+------------------------------------------------------------------------------------|
|       | alkeys | see GALK       |                 |         |                              |
|       |--------+----------------+-----------------+---------|                              |
|       | length | see GLD        |                 |         |                              |
|       |--------+----------------+-----------------+---------|                              |
|       | sld    | see GALD       |                 |         |                              |
|       |--------+----------------+-----------------+---------|                              |
|       |        | compute        |                 |         |                              |
|       |        | average        |                 |         |                              |
|       | avgald | lengths at age |                 |         | $prefix.meanle.$survenames   |
|       |        | from sld       |                 |         | (GAVLs)                      |
|       |        | output         |                 |         | $prefix.meanle.$fleetnames   |
```

```
|       |--------+----------------+-----------------+---------| (GAVLc)                       |
|       |        | computes       |                 |         | age.agg                       |
|       |        | standard       |                 |         | area.agg                      |
|       |        | deviation at   |                 |         | likelihood                    |
|       | stddev | age based on   |                 |         | penaltyfile                   |
|       |        | data from sld  |                 |         |                               |
|       |        | and avgald     |                 |         |                               |
|       |        | outputs        |                 |         |                               |
|       |--------+----------------+-----------------+---------|                               |
|       |        | computes       |                 |         |                               |
|       | number | number of fish |                 |         |                               |
|       |        | at ages from   |                 |         |                               |
|       |        | sld            |                 |         |                               |
|-------+-------------------------------------------------------------------------------------|
| GAVW  |                                                                                     |
|-------+-------------------------------------------------------------------------------------|
|       | alkeys | see GALK       |                 |         |                               |
|       |--------+----------------+-----------------+---------|                               |
|       | length | see GLD        |                 |         | $prefix.meanwg.$survenames    |
|       |--------+----------------+-----------------+---------| (GAVWs)                       |
|       | sld    | see GALD       |                 |         | $prefix.meanwg.$fleetnames    |
|       |--------+----------------+-----------------+---------| (GAVWc)                       |
|       |        | selects        |                 | SAM >   | age.agg                       |
|       |        | average        |                 | SPE >   | area.agg                      |
|       | avgwgt | weights and    | all             | CAS >   | likelihood                    |
|       |        | variances at   |                 | LEC >   | penaltyfile                   |
|       |        | age            |                 | AGE     |                               |
|       |--------+----------------+-----------------+---------|                               |
|       | number | see GAVL       |                 |         |                               |
|-------+-------------------------------------------------------------------------------------|
| GSTO  |                                                                                     |
|-------+-------------------------------------------------------------------------------------|
|       |        | see GALK,      |                 |         | $prefix.alkstock.$survenames  |
|       |        | includes       |                 |         | (GSTOs)                       |
|       |        | stockname to   |                 |         | $prefix.alkstock.$fleetnames  |
|       |        | produce alkeys |                 |         | (GSTOc)                       |
|       | alkeys | output split   |                 |         | len.agg                       |
|       |        | by maturity    |                 |         | age.agg                       |
|       |        | stage and/or   |                 |         | area.agg                      |
|       |        | sex groups     |                 |         | likelihood                    |
|       |        |                |                 |         | penaltyfile                   |
|-------+-------------------------------------------------------------------------------------|
| GLDS  |                                                                                     |
|-------+-------------------------------------------------------------------------------------|
|       |        | see GLD,       |                 |         | $prefix.ldstock.$survenames   |
|       |        | includes       |                 |         | (GLDSs)                       |
|       |        | stockname to   |                 |         | $prefix.ldstock.$fleetnames   |
|       |        | produce length |                 |         | (GLDSc)                       |
|       | length | output split   |                 |         | len.agg                       |
|       |        | by maturity    |                 |         | allage.agg                    |
|       |        | stage and/or   |                 |         | area.agg                      |
|       |        | sex groups     |                 |         | likelihood                    |
|       |        |                |                 |         | penaltyfile                   |
|-------+-------------------------------------------------------------------------------------|
| GAVG  |                                                                                     |
|-------+-------------------------------------------------------------------------------------|
|       |        | selects        | areas, species, | SAM >   |                               |
|       |        | average        | samplingtype,   | SPE >   |                               |
|       | weight | weights for a  | lengthcell,     | CAS >   | $prefix.refweights            |
|       |        | given          | lengthcellmin,  | LEC     |                               |
|       |        | lengthcell     | lengthcellmax   |         |                               |
|-------+-------------------------------------------------------------------------------------|
| GSPE  |                                                                                     |
|-------+-------------------------------------------------------------------------------------|
|       |        |                | areas,          |         |                               |
|       |        |                | lengthcell,     |         |                               |
|       | stock  | generates      | lengthcellmin,  |         | $prefix                       |
|       |        | stock file     | lengthcellmax,  |         |                               |
|       |        |                | agemin, agemax, |         |                               |
|       |        |                | stockname       |         |                               |
|-------+-------------------------------------------------------------------------------------|
| GARE  |                                                                                     |
|-------+-------------------------------------------------------------------------------------|
|       |        | generates area | years,          |         |                               |
|       | arsize | file, queries  | timestep,       |         | area                          |
|       |        | l_areasofareas | areas,          |         |                               |
|       |        | for area size  | temperature     |         |                               |
|-------+-------------------------------------------------------------------------------------|
| GTIM  |                                                                                     |
|-------+-------------------------------------------------------------------------------------|
|       |        | generates time | years,          |         |                               |
|       | time   | file           | timestep,       |         | time                          |
|       |        |                | prediction      |         |                               |
|-------+-------------------------------------------------------------------------------------|
```

```
| GFLT  |                                                                                    |
|-------+------------------------------------------------------------------------------------|
|       |        | if fleetscale |                 |         |                               |
|       |        | = 1, selects  |                 |         |                               |
|       |        | total         |                 |         |                               |
|       |        | year/species  |                 |         |                               |
|       |        | landings and  | years, species, |         |                               |
|       |        | rescales      | vesselclass,    | SAM >   |                               |
|       | lndngs | subsequent    | vesselsubclass, | SPE >   |                               |
|       |        | queries split | stock,          | CAT     | $prefix.fleet.data            |
|       |        | on gearclass, | marketcategory, |         | $prefix.predict.data          |
|       |        | gearsubclass; | samplingtype    |         | $prefix.fleet                 |
|       |        | samplingtype  |                 |         |                               |
|       |        | used for      |                 |         |                               |
|       |        | generated     |                 |         |                               |
|       |        | survey output |                 |         |                               |
|       |--------+---------------+-----------------+---------|                               |
|       |        | selects total |                 | SAM >   |                               |
|       | lndngs | landings      | all (for catch) | SPE >   |                               |
|       |        |               |                 | CAT     |                               |
|-------+------------------------------------------------------------------------------------|
| GINI  |                                                                                    |
|-------+------------------------------------------------------------------------------------|
|       |        | see GALK,     |                 |         |                               |
|       |        | selects alk   |                 |         |                               |
|       |        | data for      |                 |         |                               |
|       |        | minimum of    | all except      |         |                               |
|       | alkeys | years         | gearclass and   |         |                               |
|       |        | specified (but| gearsubclass    |         |                               |
|       |        | >= 1985**) in |                 |         |                               |
|       |        | the first     |                 |         |                               |
|       |        | timestep      |                 |         |                               |
|       |--------+---------------+-----------------+---------|                               |
|       |        | see GLD,      |                 |         | $prefix.init                  |
|       |        | selects ld    |                 |         |                               |
|       |        | data for      |                 |         |                               |
|       |        | minimum of    | all except      |         |                               |
|       | length | years         | gearclass and   |         |                               |
|       |        | specified (but| gearsubclass    |         |                               |
|       |        | >= 1985**) in |                 |         |                               |
|       |        | the first     |                 |         |                               |
|       |        | timestep      |                 |         |                               |
|       |--------+---------------+-----------------+---------|                               |
|       | sld    | see GALD      |                 |         |                               |
|-------+------------------------------------------------------------------------------------|
| GREC  |                                                                                    |
|-------+------------------------------------------------------------------------------------|
|       |        | selects       |                 |         |                               |
|       |        | maximum length|                 |         |                               |
|       | recmax | for species at| species         |         |                               |
|       |        | age = agemin  |                 |         |                               |
|       |        | from d_minmax |                 |         |                               |
|       |--------+---------------+-----------------+---------|                               |
|       |        | see GALK,     |                 |         |                               |
|       |        | selects an alk|                 |         |                               |
|       |        | data, all ages|                 |         |                               |
|       |        | are selected  |                 |         |                               |
|       |        | for length    |                 |         |                               |
|       |        | distributions,|                 |         |                               |
|       |        | but only      | all except      |         |                               |
|       |        | agemin is     | gearclass and   |         |                               |
|       | alkeys | shown,        | gearsubclass,   |         |                               |
|       |        | timesteplabel | timestep = YEA  |         | $prefix.rec                   |
|       |        | is used for   |                 |         |                               |
|       |        | timestep; for |                 |         |                               |
|       |        | years < 1985, |                 |         |                               |
|       |        | average of    |                 |         |                               |
|       |        | years         |                 |         |                               |
|       |        | 1985-1989 is  |                 |         |                               |
|       |        | taken**       |                 |         |                               |
|       |--------+---------------+-----------------+---------|                               |
|       |        | see GLD,      |                 |         |                               |
|       |        | selects       | all except      |         |                               |
|       | length | average length| gearclass and   |         |                               |
|       |        | and standard  | gearsubclass,   |         |                               |
|       |        | deviation for | timestep = YEA  |         |                               |
|       |        | recruited fish|                 |         |                               |
|       |--------+---------------+-----------------+---------|                               |
|       | sld    | see GALD      |                 |         |                               |
|-------+------------------------------------------------------------------------------------|
| GPRB  |                                                                                    |
|-------+------------------------------------------------------------------------------------|
|       |        | selects prey  |                 |         |                               |
|       |        | biomass       |                 |         |                               |
```

```
|       |        | specified by   |                 | SAM >   | $prefix.preyb                |
|       | preywg | preys;         | all (for        | SPE >   | len.agg                      |
|       |        | computes       | stomach)        | PRD >   | area.agg                     |
|       |        | proportions    |                 | PRE     | preyb.agg                    |
|       |        | per predator   |                 |         |                              |
|       |        | length         |                 |         |                              |
|-------+------------------------------------------------------------------------------------|
| GPRN  |                                                                                    |
|-------+------------------------------------------------------------------------------------|
|       |        | selects prey   |                 | SAM >   | $prefix.preyn                |
|       | preywg | numbers        | all (for        | SPE >   | len.agg                      |
|       |        | specified by   | stomach)        | PRD >   | area.agg                     |
|       |        | preys          |                 | PRE     | preyn.agg                    |
|-------+------------------------------------------------------------------------------------|
| GPRL  |                                                                                    |
|-------+------------------------------------------------------------------------------------|
|       |        | selects prey   |                 | SAM >   | $prefix.preyl                |
|       |        | lengths        | all (for        | SPE >   | len.agg                      |
|       | preyln | distribution   | stomach)        | PRD >   | area.agg                     |
|       |        | for a given    |                 | PRE >   | preyl.agg                    |
|       |        | preyspecies    |                 | PRL     |                              |
|-------+------------------------------------------------------------------------------------|
| GPRW  |                                                                                    |
|-------+------------------------------------------------------------------------------------|
|       |        | selects prey   |                 |         |                              |
|       |        | lengths        |                 |         |                              |
|       |        | distribution   |                 |         |                              |
|       |        | for a given    |                 | SAM >   | $prefix.preyw                |
|       |        | preyspecies as | all (for        | SPE >   | len.agg                      |
|       | preyln | biomass        | stomach)        | PRD >   | area.agg                     |
|       |        | ratios,        |                 | PRE >   | preyw.agg                    |
|       |        | alphabeta is   |                 | PRL     |                              |
|       |        | used in the    |                 |         |                              |
|       |        | weight-length  |                 |         |                              |
|       |        | relationship   |                 |         |                              |
|-------+------------------------------------------------------------------------------------|
| GPDB  |                                                                                    |
|-------+------------------------------------------------------------------------------------|
|       |        | see GPRB,      |                 |         |                              |
|       |        | outputs        |                 |         | $prefix.predatorb            |
|       | preywg | predator by    |                 |         | len.agg                      |
|       |        | lengthcell     |                 |         | area.agg                     |
|       |        | corresponding  |                 |         |                              |
|       |        | to GPRB        |                 |         |                              |
|-------+------------------------------------------------------------------------------------|
| GPDN  |                                                                                    |
|-------+------------------------------------------------------------------------------------|
|       |        | see GPRN,      |                 |         |                              |
|       |        | outputs        |                 |         | $prefix.predatorn            |
|       | preywg | predator by    |                 |         | len.agg                      |
|       |        | lengthcell     |                 |         | area.agg                     |
|       |        | corresponding  |                 |         |                              |
|       |        | to GPRN        |                 |         |                              |
|-------+------------------------------------------------------------------------------------|
| GPDL  |                                                                                    |
|-------+------------------------------------------------------------------------------------|
|       |        | see GPRL,      |                 |         |                              |
|       |        | outputs        |                 |         | $prefix.predatorl            |
|       | preyln | predator by    |                 |         | len.agg                      |
|       |        | lengthcell     |                 |         | area.agg                     |
|       |        | corresponding  |                 |         |                              |
|       |        | to GPRL        |                 |         |                              |
|-------+------------------------------------------------------------------------------------|
| GPDW  |                                                                                    |
|-------+------------------------------------------------------------------------------------|
|       |        | see GPRW,      |                 |         |                              |
|       |        | outputs        |                 |         | $prefix.predatorw            |
|       | preyln | predator by    |                 |         | len.agg                      |
|       |        | lengthcell     |                 |         | area.agg                     |
|       |        | corresponding  |                 |         |                              |
|       |        | to GPRW        |                 |         |                              |
|-------+------------------------------------------------------------------------------------|
| GTAG  |                                                                                    |
|-------+------------------------------------------------------------------------------------|
|       |        | generates      | species,        |         |                              |
|       |        | tagfile, for   | timestep,       |         |                              |
|       |        | the start      | areas,          |         |                              |
|       |        | dates of the   | lengthcell,     |         |                              |
|       |        | tagging        | lengthcellmin,  |         |                              |
|       | tagend | experiment     | lengthcellmax,  | TAG     | $tagid.tagfile               |
|       |        | finds the      | tagid, tagtime, |         |                              |
|       |        | corresponding  | tagarea,        |         |                              |
|       |        | last year and  | tagloss,        |         |                              |
|       |        | month of       | tagstock        |         |                              |
```

```
|       |        | recapture      |                 |         |                              |
|-------+------------------------------------------------------------------------------------|
| GTAD  |                                                                                    |
|-------+------------------------------------------------------------------------------------|
|       |        |                | species,        |         |                              |
|       |        |                | timestep,       |         |                              |
|       |        | selects        | areas,          |         |                              |
|       | tagrel | numbers of     | lengthcell,     | TAG     | $tagid.tag.data              |
|       |        | tagged fish by | lengthcellmin,  |         |                              |
|       |        | lengthcell     | lengthcellmax,  |         |                              |
|       |        |                | tagid, tagtime, |         |                              |
|       |        |                | tagarea         |         |                              |
|-------+------------------------------------------------------------------------------------|
| GTAR  |                                                                                    |
|-------+------------------------------------------------------------------------------------|
|       |        |                | species,        |         |                              |
|       |        | selects        | timestep,       |         |                              |
|       |        | recaptured     | areas,          |         |                              |
|       | tagrec | fish           | lengthcell,     | TAG     | $tagid.tag.like              |
|       |        | associated     | lengthcellmin,  |         | len.agg                      |
|       |        | with GTAD      | lengthcellmax,  |         |                              |
|       |        |                | tagid, tagtime, |         |                              |
|       |        |                | tagarea         |         |                              |
+--------------------------------------------------------------------------------------------+

 *   optional type not used by Gadget
 **  for Iceland only
 >   natural join
```

2.8 Distributed approach

Vojtěch Kupča
Marine Research Institute, Reykjavik, Iceland

In this part the distribution of the DW using CORBA technology is discussed. A description of the solution and interaction of the query system and the user interface is provided. A detailed diagram of the system architecture is supplied in figure 2.3.

Introduction

As a demonstration of linking several databases containing the marine ecosystem data it was decided that CORBA technology will be used. Its main advantages were first seen in the platform and language independence. However, with better understanding of the project these started to be not so significant contributions since only one type of language and operating system were used. Moreover, real problems of the distributed database solutions were encountered that were not thought of at the beginning. This report describes the results of an attempt to implement the system despite the mentioned obstacles. As it was stated before (8), more interesting from the point of view of truly usable and useful system is the usage of asynchronous replication implemented e.g. by the PostgreSQL Replicator (9) or later by native distributed PostgreSQL facilities currently developed within the PostgreSQL Replication Project (10).

Solution description

CORBA is an object oriented framework (specification) for interconnecting distributed computer systems regardless of underlying software and hardware differences. Implementation of CORBA called JacORB with Java binding[1] was used. The system has two main parts, one that processes all requests and does all internal data processing (called the *query system*) and the other which allows the end-user to send requests to the query system and receive corresponding results (called the *interface*). Query system is programmed in Java and is composed of the following parts:

in one site only:

- *PHP server* for communication with the user interface part
- *CORBA client* which controls distribution of queries to all involved query systems and combination of partial results in temporary tables of an auxiliary database
- *CORBA name service* for looking up references to objects by names
- *auxiliary PostgreSQL database* for combination of partial results

in all involved sites:

- PostgreSQL database system with setup dst^2 DW tables
- ORB (Object Request Broker) with defined functions for accessing dst^2 DW

Interface part is web-based and is programmed in the PHP and HTML.

The idea behind distributed querying can be described in several steps:

1. user fills in the request using the interface part

---

[1] The Java binding in JacORB is a mechanism that maps CORBA specification details (e.g. data types) to the corresponding entities in Java programming language.

2. request is sent to the CORBA client part for processing
3. CORBA client contacts all involved ORBs to process the request
4. ORBs send queries to (usually local) dst^2 DWs
5. ORBs report back the results to the CORBA client
6. CORBA client stores partial results inside of the temporary table of an auxiliary database and tries to combine all queries as if only one system was queried
7. CORBA client hands over the result to the PHP server
8. PHP server forwards 'raw' result to the user interface for processing and presentation

Communication between the interface and the PHP server (points 2 and 8) are realized using a simple communication protocol described in (8). In the following sections, some points from the above list are described in more detail.

ORB functions

What an ORB (the core CORBA component which actually performs requests sent to it) can do is defined in the *ORB interface definition* using language called IDL (Interface Definition Language). The ORB interface defined in `QR.idl` contains these functions:

- `long doQuery (in string theQuery);`
  execute query on a connected database
- `long getNumberOfRows();`
  return number of rows
- `long getNumberOfColumns();`
  return number of columns
- `stringArray getRow(in long rowNumber);`
  return array with one row specified by its number
- `stringArray getColumnNames();`
  return array of column names
- `long getTypeOfColumn(in long columnNumber);`
  return column type code
- `boolean login (in string username, in string database, in string password);`
  open database connection and login to the database
- `void logout();`
  log off the database and close connection

where `stringArray` is a CORBA container type defined as: `typedef sequence<string> stringArray;`. For a detailed description it is best to refer to the implementation file `QueryRetrieverImpl.java`.

Database setup

On all involved sites there has to be installed the PostgreSQL database (traditionally PostgreSQL runs on port 5432). dst^2 DW table definitions are stored in `dwC.sql` which is a part of the database setup scripts. In addition, on CORBA client site there has to be installed an auxiliary database system with the write permission for creation of temporary tables. These are used for combination of results from different sites.

User interface

User can interact with the whole system using the web interface. For demonstration purposes and simplicity it was sufficient to provide only a subset of functionalities and features of dst^2 DW database. User can therefore choose from the following options:

a) output types: age-length keys (ALK), age-length distributions with average lengths at age (ALD), catches in numbers by age and total landings (CNR),
b) available databases: Iceland or the Celtic Sea,
c) areas: dst^2 divisions covering Iceland and the Celtic Sea,
d) year, month, species (cod or haddock), gearclass and vesselclass selection,
e) lengthcell range, age range and lengthcell interval selection,
f) it is possible to generate output in two forms: one is to display result formated as HTML tables, the other is to save output in an ASCII text file.
On the web page a user can find information on how to use the interface and also the map of dst^2 DW divisions.

Limitations and problems

As it was mentioned in (8) it was not a goal to build a versatile and robust distributed database system; instead a demonstration of simple linking of several databases using CORBA was intended. Therefore there are some limitations to the design as well as several other ad hoc solutions.

- The database definitions have to be identical, the same query is sent to all databases.
- The CORBA client part is responsible for sending queries to the involved ORBs and retrieving the results. Later it has to combine partial results from each database into the correct unified result. It tries to do so by storing the results in a temporary PostgreSQL table and rerunning the same query on the whole table. This approach is simple and works for some queries but for many other it fails and certain modifications to query morphology are required. Let us show a simple example where this technique fails. If we consider only one database with table `A` as

a	b
1	5
2	10

  and we do `SELECT a FROM A WHERE b = 5;`, we obtain the first row and the first column of the table only. Running the query from the temporary table as `SELECT a FROM temp1 WHERE b = 5;` is not correct since the column `b` is not selected in the first query. We can say that problem occurs if `SELECT` and `WHERE` clauses do not coincide column-wise. A simple workaround for this is to split the above query into three: one that creates a new table from the result (using `SELECT INTO`) and returns only boolean value, the second that selects everything from this table, and the third that drops this table. More problems would arise if we needed to use aggregate functions with a `GROUP BY` clause. An example is an average function `AVG()`: if we consider tables `A` and `B` having both one column `a`, where table `A` contains values 1, 1, 4 (whose average is 2) and table `B` contains values 3, 5 (average 4). It is then obvious that the true average of 2.8 is not equal to the average of partial results. For more information see Mariposa project (11).

- Performance is dependent on how much data is being moved around the network (typically tens of kilobytes) and how complicated the queries are. In extreme cases queries can take a couple of minutes to process. This is due to the fact that many tens of thousands of rows are being selected from the databases and large matrices manipulated by the web interface program.

- The application uses *threading* for non-blocking execution of queries. From the point of view of a program threads are separate parallel execution paths that share the common memory. For the operating system, creation of a thread in Java means a creation of a new process. During run it may happen that the number of threads (processes) increases. This behaviour was seen in our application and it is was not tested what happens if the application is run for a long time.

Security

SSH

With distributed applications it is reasonable to address the issue of security of data being sent among involved sites. Normally data are passing the network in a plain unencrypted way so that anyone can potentially eavesdrop the communication. On an insecure network (like e.g. Internet) it is often needed to protect the content being sent. In our case, JacORB has a built-in set of proxy classes (called *Appligator*) that provide a client with a 'gateway' for calling server functions. For our purpose, it is possible to use the Appligator to direct all client-server communication to a specified port and tunnel this port using `ssh` (secure shell) in a secure way over the network. Section 10.4 of the JacORB Programming Guide (12) provides further details on how to setup this configuration.

SSL

Another possibility to implement a secure data connection is to use the SSL (Secure Socket Layer). SSL encapsulates higher level network protocols in an encrypted data stream using symmetric cryptography protocols such as DES or RC4. In JacORB the SSL is implemented through the Java Secure Socket Extension (JSSE). This solution is not suitable for passing data through firewalls since it still opens the same high ports for communication as in the normal setup (though this time in a secure way).

Either of the above described ways makes sure that client-to-server communication is secure. Eventually, however, the data are sent all the way back from the PHP server and interface part to the user's browser in an open form (see figure 2.3). Therefore, for the part between PHP server and interface the JSSE (for server) and PHP's SSL (for client) is convenient to use. For details on how to configure JacORB to use the SSL see section 11 of the JacORB Programming Guide (12) and JSSE Reference Guide (13). For the part between the web interface and user's browser HTTPS[2] is needed. These measures ensure that no third party is able to see the passing data until they are presented to the (authorized) end user. Connections between servers (or client) and databases are local and are considered as trusted, since they are interceptable through a local loopback device only. The same is true for the connection between client and `ssh`'s tunneling port.

Installation and running

In this section we describe a way to install and run the application from the general point of view. For a working example with two hosts please refer to page 50. Before it is possible to run the whole system, there has to be made a number of installation steps first:

---

[2] HTTPS is a secure HTTP protocol running typically on port 443. This protocol uses mechanisms of public key cryptography to protect the data content. It is based on the open source implementation of SSL called OpenSSL (14). To enable the use of SSL by the web server it is necessary to configure `mod_ssl` module accordingly.

- installed latest OpenSSL (14)
- installed latest Apache (7)
  - for the encryption between the web server and the end user it is necessary to install Apache module `mod_ssl` (15) and create public/private key pair as described in (16). Public key is usually distributed in the form of the so called *certificate file* which contains some extra authentication information
  - **remark**: when signing the *Certificate Signing Request* (CSR) by the *Certifying Authority* (CA) using `sign.sh` script, one has to use different "Organizational Unit Name" or "Common Name"
- installed Java environment SDK-1.4.1 (17), it is necessary to set variable `JAVA_HOME` accordingly (e.g. `export JAVA_HOME=/usr/lib/j2sdk1.4.1_02`)
- installed latest Apache Ant (Java based build-tool) (18)
- installed latest PostgreSQL database (compiled with `-with-java` option) (4), notice that a path to the `ant` binary may need to be set (e.g.
  `export ANT=/usr/local/ant-1.5.4/bin/ant`)
- installed latest JacORB (19)
- for database access, the `jdbc:postgresql` driver is used, it is therefore necessary to inflate PostgreSQL subdirectory jar-file `.../share/java/postgresql.jar` into the JacORB subdirectory `classes` and add this directory to `CLASSPATH` variable which tells JacORB where to look for java classes (e.g.
  `export CLASSPATH= $CLASSPATH:/usr/local/JacORB_1_4_1/classes:.`
- add JSSE package path to the `CLASSPATH`
  (e.g. `export CLASSPATH= $CLASSPATH:$JAVA_HOME/jre/lib/jsse.jar`)
- copy the template `jacorb.properties` file to your `$HOME` directory and do the following changes:
  - unset the usage of CORBA name service, i.e. remove property
    `ORBInitRef.NameService`
  - set `jacorb.ProxyServerID` to some name, which is the same for all involved sites
    (e.g. `jacorb.ProxyServer.ID=Appligator`)
  - set `jacorb.ProxyServerURL` to point to the URL written by Appligator (e.g.
    `jacorb.ProxyServer.URL=http://130.208.66.191/~vojtech/AP_Ref.ior`)
    for more sites, CORBA client will access several Appligators in more networks, for that purpose the above property name has to be extended by the network and netmask information:
    ```
    jacorb.ProxyServer.URL-130.208.66.0-255.255.255.0=
    http://130.208.66.191/~vojtech/APIs_Ref.ior
    jacorb.ProxyServer.URL-130.208.64.0-255.255.255.0=
    http://130.208.64.141/~vojtech/APFr_Ref.ior
    ```
  - set `jacorb.giop_minor_version=0`
- setup the PostgreSQL database(s) with dst^2 DW tables (and upload data using `dst2up.php`, see section 2.4); empty auxiliary database `temp` has to be created at CORBA client site (use `createdb`)
- installed PHP (5) with the shared Apache module (compiled with `-with-apxs -with-openssl` options) and the PHP interface program
- extended variable `$PATH` to include directories with executables of Java, JacORB, Ant and PostgreSQL

- install OpenSSH (20) with `ssh` client and `sshd` server
- once the application is extracted in some directory it can be optionally recompiled (run command `ant` in the directory with `build.xml` makefile) and should be ready for launching. In the `Implementation` subdirectory there are several small scripts which should make the start-up easier:
    - First it is necessary to create the server keyentry (public/private key pair) stored in the server keystore (this keystore is in the application's `KeyStore` subdirectory) using `keytool` utility as e.g.:

      ```
      keytool -keyalg "SunX509" -genkey -alias vojtech -keyalg RSA
      -validity 7 -keystore keystore
      ```

      The keyentry is then used in the communication between the PHP server and PHP interface. The generated `keystore` file should be placed into the CORBA subdirectory `Implementation/KeyStore` under the name `server_keystore`.
    - adjust and run `appligator.sh`. This file removes old Appligator reference and creates a new one in the specified directory
    - adjust and run `server.sh`. It is necessary to set one additional JacORB property, which indicates that Appligator is used: on the server side the property

      ```
      org.omg.PortableInterceptor.ORBInitializerClass.org.jacorb.proxy.
      ProxyServerInitializer
      ```

      has to be set. This is done by `-D` switch on the Java command line.
    - before starting the PHP server, utility `GetFile` copies the Appligator IOR (inter object reference) to the local machine. To do this you can adjust and run `getfile.sh`. This somewhat clumsy handling of IORs is needed due to the buggy Appligator when used in conjunction with the CORBA name service (applies to JacORB 1.4.1 or earlier). CORBA name service (or nameserver) handles IORs of application objects and simplifies reference lookups.
    - in an `ssh` setup it is necessary to patch the Appligator references in such a way that they point to the local host port where `ssh` will listen and tunnel all traffic to the `sshd` on the other communication end and pass it on to the remote Appligator object. To do this one has to run `fixior.sh` to change the reference to 127.0.0.1 (localhost), port 111111.
    - adjust and run `phpserver.sh`. Similarly as in the case of server, the property

      ```
      org.omg.PortableInterceptor.ORBInitializerClass.org.jacorb.proxy.
      ProxyClientInitializer
      ```

      has to be set.
    - next `ssh` has to be invoked in a port forwarding mode (see `man ssh` like e.g.:

      ```
      ssh -l dst2user -T -L 11111:dst2.ifremer.fr:22222
      dst2.ifremer.fr
      ```
    - connect through a browser to the web interface over the HTTPS protocol (for the Icelandic installation see (21), it is possible to change connection settings in the DW subdirectory file `Dist/dpsql.php`, function `send_query`)

It it necessary to remember that the numbers of ports and the names of databases and reference files are variable and has to be setup and checked for correctness in appropriate files or command lines.

Example

In this section we follow up on the previous information and describe a simple example with two hosts. Let us assume we have two sites, one running CORBA client / PHP server (called *client*, IP 130.208.66.191) and both running one CORBA server (*server1* and *server2*, IPs 130.208.66.191 and 130.208.64.141, resp.). Notice hosts being on different networks. To setup a working environment one has to adjust `jacorb.properties` files, correctly manage IOR files and copy them to the right locations. First we have to remove references to CORBA name service from all files, since the name service does not work properly with the Appligator. Then we set properties as shown in the following table:

client	`jacorb.ProxyServer.ID=Appligator`
	`jacorb.ProxyServer.URL-130.208.66.0-255.255.255.0=http://130.208.66.191/~vojtech/APIs_Ref.ior`
	`jacorb.ProxyServer.URL-130.208.64.0-255.255.255.0=http://130.208.66.191/~vojtech/APFr_Ref.ior`
	`jacorb.giop_minor_version=0`
	`org.omg.PortableInterceptor.ORBInitializerClass.org.jacorb.proxy.ProxyServerInitializer`
	
	Remark: the last property can be set externally using java `-D` option

server1	`jacorb.ProxyServer.ID=Appligator`
	`jacorb.giop_minor_version=0`
	`org.omg.PortableInterceptor.ORBInitializerClass.org.jacorb.proxy.ProxyServerInitializer`
	
	Remark: it is possible to use the same properties file like for the client, because the last property can be set externally by java `-D` option

server2	`jacorb.ProxyServer.ID=Appligator`
	`jacorb.ProxyServer.URL=http://130.208.64.141/~vojtech/APFr_Ref.ior`
	`jacorb.giop_minor_version=0`
	`org.omg.PortableInterceptor.ORBInitializerClass.org.jacorb.proxy.ProxyClientInitializer`

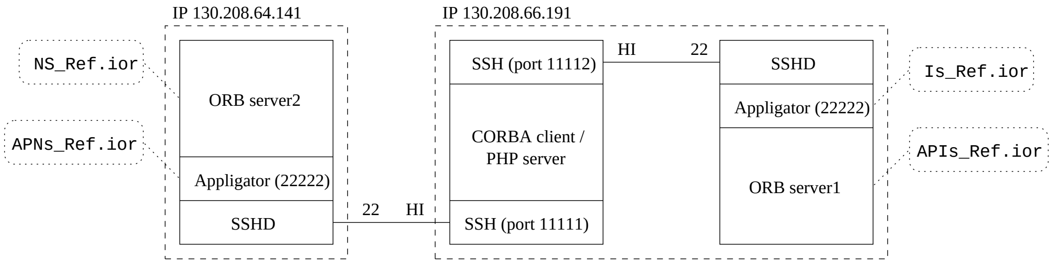
After starting Appligators in both locations on port 22222 client has to copy both Appligator references to the local host, move them to public locations specified by the above two client properties `jacorb.ProxyServer.URL` (this is done by `GetFile` utility which downloads an IOR file over the HTTP), fix their references using `fixior` utility (we choose to run ssh tunneling on ports 11112 (for server1) and 11111 (server2). The following sequence of steps has to be taken.

server1: `appligator 22222 /home/vojtech/public_html/APIs_Ref.ior`
server2: `appligator 22222 /home/vojtech/public_html/APFr_Ref.ior`
client: `java QR.GetFile http://130.208.66.191/~vojtech/APIs_Ref.ior /home/vojtech/APIs_Ref.ior`
client: `cp /home/vojtech/APIs_Ref.ior /home/vojtech/public_html/APIs_Ref.ior`
client: `java QR.GetFile http://130.208.64.141/~vojtech/APFr_Ref.ior /home/vojtech/APFr_Ref.ior`
client: `cp /home/vojtech/APFr_Ref.ior /home/vojtech/public_html/APFr_Ref.ior`
client: `fixior 127.0.0.1 11112 /home/vojtech/public_html/APIs_Ref.ior`
client: `fixior 127.0.0.1 11111 /home/vojtech/public_html/APFr_Ref.ior`

Afterwards the servers can be started up on both hosts (writing IOR references `Is_Ref.ior` and `Fr_Ref.ior` to public locations) and CORBA client on a client host. Finally it is necessary to forward all network traffic through two client-to-server ssh connections as:

client: `ssh -l vojtech -T -L 11111:130.208.64.141:22222 130.208.64.141`
client: `ssh -l vojtech -T -L 11112:130.208.66.191:22222 130.208.66.191`

This way we forward local client calls to port 11112 to server1, port 22222 and calls to port 11111 to server2, port 22222 where Appligator proxies are listening for GIOP requests. Upon inspecting the contents of generated network traffic on client it should be noticeable whether the application behaves as intended or not. The following figure helps visualize the whole situation. Notice that ssh traffic between the client and server1 can pass a firewall only if port 22 is left open. Also notice that public HTTP areas have to be writable by server1 and server2 users and web servers have to be running on both sites.



To automate the ssh start-up one can get rid of the password prompting by generating a public / private keypair, copying the public key to the remote host and adding it to the authorized public keys file. In our particular example one has to run:

client: `ssh-keygen -t rsa` (empty password)
client: `scp /home/vojtech/.ssh/id_rsa.pub vojtech@130.208.64.141:/home/vojtech/id_rsa_client.pub`
server2: `cat /home/vojtech/.ssh/id_rsa_client.pub » /home/vojtech/.ssh/authorized_keys2`
server1: `cat /home/vojtech/.ssh/id_rsa.pub » /home/vojtech/.ssh/authorized_keys2`

The whole procedure of running server1 and client with forwarding to server2 can be initiated by running

```
control.sh start
```

in the DW subdirectory `Corba-2host/Implementation`. Local forwarding to server1 has to be started in another terminal window on client. Similarly, one can start the other ORB server and Appligator at server2 site.

Changes

There had to be made certain minor corrections or changes to the original source code: **a)** instead of the original `jdbc:odbc` driver, the `jdbc:postgresql` driver taken from PostgreSQL installation tree was used, **b)** problem with rerunning the query described in the section on limitations was fixed, **c)** creation of auxiliary tables changed to the creation of temporary tables that are dropped automatically on connection close, **d)** queries containing aggregate functions return one row of result even when being empty; in `QueryRetrieverImpl.java` it is first necessary to check the content of the first returned row using `if(rs.getString(1) != null)` before increasing the number of rows by `rowCount++;`.

Results

It was possible to run the application with full encryption on one computer as well as between two hosts. Using the network monitoring tool `ethereal` one could analyze the internal workings of the application. Its communication had three basic parts:

- the communication between CORBA client and CORBA servers (ORBs) was properly tunneled through `ssh`, i.e. high port-to-22 or 22-to-high port communication occurred (22 is a `sshd` port),

- the communication between the PHP server and PHP interface was encrypted, first the open text certificate information (with public key) was distributed, followed by the SSL encrypted data stream,
- finally the data were passed to the end user over the SSL encrypted HTTP channel (HTTPS), port 443.

This kind of setup where all vulnerable communication is secured by strong cryptography protocols can be regarded as the software virtual private network (VPN).

Appligator problems

The main problem was the buggy JacORB component Appligator. It was not able to run together with the CORBA name service. Therefore an auxiliary solution was to hand over object references ’by hand’. For that purpose the `GetFile` utility was made. Consequently the communication protocol between the PHP server and PHP interface had to be slightly changed. The value of the `HOSTLIST` command is now of the form *ior_url_reference:dbname:database:password* where the *ior_url_reference* is the URL of an ORB’s IOR without the `http://` prefix. For a complete description of the protocol see (8).

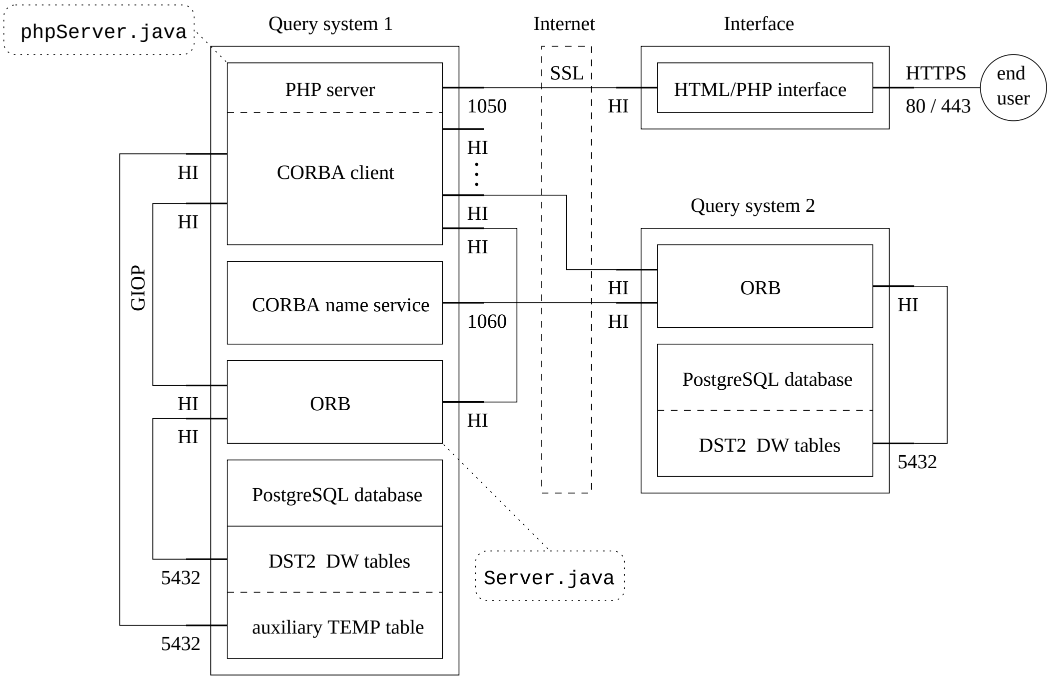


Figure 2.3: **System architecture:** In the example above CORBA name service is started on port 1060; since then CORBA objects can be referenced by names bound to the references. Afterwards the server (ORB) is started on one of the JacORB's high ports (marked HI); first it needs to register with the naming service to be usable by other components. Later its IDL definition says what functions the server can do (implemented in `QueryRetrieverImpl.java`): e.g. query local dst^2 DW database, count returned rows and columns, etc. Then the *PHP server / CORBA client* program is run. It has two parts, the PHP server listens on some agreed port (e.g. 1050) for incoming web-client requests over a defined protocol and submits these requests to the second part, CORBA client. It's purpose in brief is to connect to user-specified servers (ORBs), send them appropriate SQL requests, wait for results, combine them using the local auxiliary database table, and through the PHP server hand over the result to the end-user interface. It can be seen that more ORBs similar to *Query system 2* can be connected in the place of three dots in the above picture. Since all shown components communicate over the network, their physical location is almost arbitrary; they can all reside on one host or can be scattered in many places. The above diagram is just one of many possible configurations. GIOP is a General Inter-ORB Protocol.

Remark: The name service is not used due to the non-functional Appligator component, instead the server IORs are obtained from URL locations.

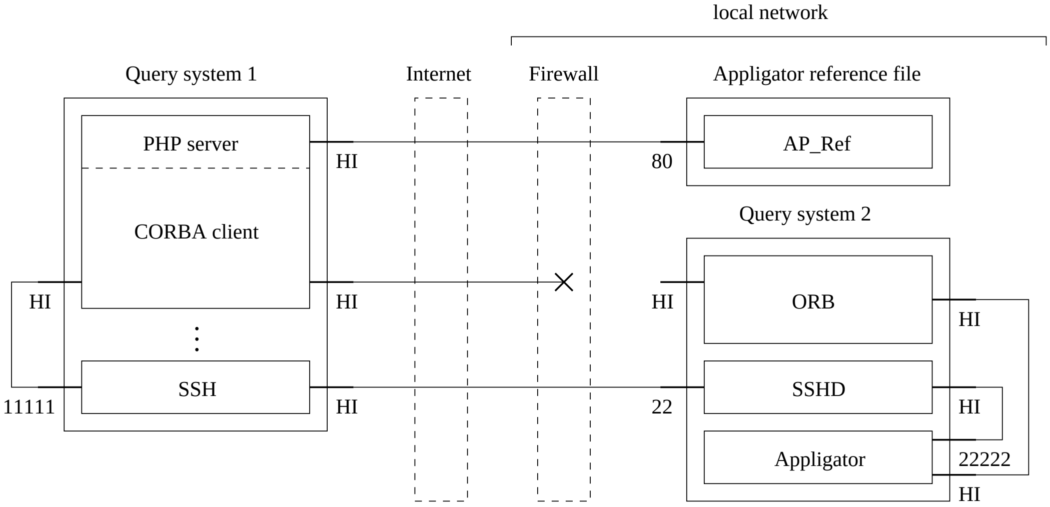


Figure 2.4: **Secure system architecture using Appligator and `ssh` Client:** (*Query system 1*) first reads the Appligator object reference from the file `AP_Ref`. Because `AP_Ref` is patched to refer to 127.0.0.1, the client thinks that the Appligator is found on the local host on port 11111. On this port, however, listens the `ssh` which tunnels connection in a secure way through the port 22 and hands it over to the real Appligator listening on the server side (*Query system 2*) on port 22222. The Appligator serves as a proxy for communication between client and server. This way it is possible to overcome difficulties with restrictions of high-port communication over firewalls.

References

1 DW home page:
`http://www.hafro.is/~vojtech/dw`

2 DW column structure:
`http://www.hafro.is/~vojtech/dw/dw-struct.html`

3 Linux:
`http://www.kernel.org`

4 PostgreSQL:
`http://www.postgresql.org`

5 PHP:
`http://www.php.net`

6 Metadata web interface (local access):
`http://haflygna.hafro.is/Html/Meta`

7 Apache web server:
`http://www.apache.org`

8 Demonstration of CORBA technology for distributed database:
`http://www.hafro.is/~vojtech/dw/dw-corba02.pdf`

9 PgReplicator:
`http://pgreplicator.sourceforge.net`

10 PostgreSQL replication:
`http://gborg.postgresql.org/project/pgreplication`

11 Mariposa project:
`http://s2k-ftp.cs.berkeley.edu:8000/mariposa/src/alpha-1/mariposa-manual.pdf`

12 JacORB documentation:
`http://www.jacorb.org/documentation.html`

13 JSSE documentation:
`http://java.sun.com/j2se/1.4.1/docs/guide/security/jsse/JSSERefGuide.html`

14 OpenSSL:
`http://www.openssl.org`

15 `mod_ssl`:
`http://www.modssl.org`

16 `mod_ssl` FAQ:
`http://httpd.apache.org/docs-2.0/ssl/ssl_faq.html`

17 Java:
`http://java.sun.com`

18 Ant:
`http://ant.apache.org`

19 JacORB:
`http://www.jacorb.org`

20 OpenSSH:
`http://www.openssh.org`

21 Distributed database web interface (local access):
`https://haflygna.hafro.is/Dist`

Part II

Models and model components for multispecies, multifleet models

Chapter 3

Introduction

Models and Model Components

When developing a model of an ecosystem the developers need to draw a line indicating the level of detail to be included in the model. In the models presented in this report, the approach is a detailed one, at least in potential. Thus a requirement from the outset has been to include the ability to model a predator where considerable data and information are availeble. However, another requirement is the ability to include in the model a prey where only poor data may be available.

Thus one is led to a scenario where an implementation may be either a simple, aggregate model with poor data or a highly disaggregated model with data of high quality. These models can all be implemented using a computer program, Gadget, which is described in detail in the following chapter.

When complex models are developed, these consist of a myriad of small components. In particular a spatially disaggregated multispecies model must inevitably include model components which describe predation, migration and growth to name some of the more important components of Gadget. Each of these components can be studied to some extent in isolation before incorporating it into a large all-encompassing model. The following sections therefore describe the development and tests of some of these model components.

In addition to defining a model, it must be fitted to data. Naturally, Gadget has methods for doing this, but some care needs to be taken when parameter estimation is undertaken. If will be seen in subsequent chapters that this is notrivial as several issues arise when fitting complex models to multiple data sources. Methods have therefore been developed for estimating the weight to be given to each data source and subsequently for verifying whether the model can simultaneously explain all data sources.

Simple model verification is implemented using an ecosystem simulator, which can generate data for an ecosystem as complex as the one modelled by Gadget.

Needless to say, estimation of uncertainty is important and there is no applicable standard method for doing this in general when the models and data are as complex as in the present setting. For this reason special bootstrapping methods have been developed, taking into account the correlation structure between and within the multiple data sources used in fishery science. Initial tests indicate this to be a promising approach.

Chapter 4

Gadget

Introduction

Fishery science has developed many ways of looking at fish populations. Thus one has many conceptual models and correspondingly many mathematical models. The difference between pure imagination and models of the real world is the need for the models to be able to fit to observations. However the data sets are as different as the research projects since not only fish species are different but also scientists.

Gadget has been developed to take these different population models and data sets into account. Thus, Gadget is intended to combine the benefits of multispecies models which take prey mortality into account with models which describe the effect of consumption on predator growth. Formal statistical methods are used for estimation of parameters, as has been done in a few single-species models of fish populations. Implementations of the individual components of these models have been available in a variety of computer programs, but none have allowed the use of multiple data sources to fit statistically multispecies models in a parametric manner as described in the following pages.

4.1 An Overview of Gadget, the Globally applicable Area Disaggregated General Ecosystem Toolbox

James Begley[1] **and Daniel Howell**[2]

[1] **Marine Research Institute, Reykjavik, Iceland**
[2] **Institute of Marine Research, Bergen, Norway**

Abstract: Gadget is the Globally applicable Area Disaggregated General Ecosystem Toolbox. Gadget is a powerful and flexible framework that has been developed to model complicated statistical marine ecosystems within a fisheries management and biological context, and can take many features of the ecosystem into account. Gadget allows the user to include a number of features of the ecosystem into the model: One or more species, each of which may be split into multiple components; multiple areas with migration between areas; predation between and within species; growth; maturation; reproduction and recruitment; multiple commercial and survey fleets taking catches from the populations. Gadget works by running an internal forward projection model based on many parameters describing the ecosystem, and then comparing the output from this model to observed measurements to get a likelihood score. The model ecosystem parameters can then be adjusted, and the model re-run, until an optimum is found, which corresponds to the model with the lowest likelihood score. This iterative, computationally intensive process is handled within Gadget, using a robust minimisation algorithm. Gadget has successfully been used to investigate the population dynamics of stock complexes in Icelandic waters, the Barents Sea, the North Sea, the Irish and Celtic Seas and the Sofala Bank fishery of Mozambique. This paper describes the structure and main components of an ecosystem model developed using the Gadget framework.

Introduction

Gadget is a software tool that has been developed to model marine ecosystems, including both the impact of the interactions between species and the impact of fisheries harvesting the species. Gadget simulates these processes in a biologically realistic manner, and uses a framework to test the development of the modelled ecosystem in a statistically rigorous manner (Anon 2002; Anon 2003).

Gadget can run complicated statistical models which take many features of the ecosystem into account. Gadget works by running an internal model based on many parameters, and then comparing the data from the output of this model to observed data to get a goodness-of-fit likelihood score. The parameters can then be adjusted, and the model re-run, until an optimum is found, which corresponds to the model with the lowest likelihood score. The Gadget framework consists of three parts:

- a parametric model to **simulate** the ecosystem
- statistical functions to **compare** the model output to data
- search algorithms to **optimise** the model parameters

The files required for a Gadget model are all in plain text ASCII format, enabling them to be easily generated, read and edited. A data warehouse program exists that can output selected data in the correct format for Gadget (Anon 2003; Kupča 2004). A graphical front end is in development to produce the files required for the simulation model (Anon 2004).

Gadget is a freely available, open source program. The software and associated documentation can be downloaded and used free of charge, and is available from the Gadget website at `www.hafro.is/gadget`. Since the source code is freely available it is possible to examine the program to ensure that the program is correctly implementing the model selected by the user. Furthermore users can add features to the code as required, providing the flexibility to adapt the core program to a wide range of fisheries situations. These improvements can then be submitted to the maintainers of the program for inclusion in future releases of the code.

Gadget has been been developed from Bormicon (Stefánsson and Pálsson 1997) and Fleksibest (Frøysa et al. 2002), which in turn were influenced by the Multspec model (Bogstad et al. 1997). The recent development of Gadget has been partially funded through EU grant QLK5-CT1999-01609. Gadget has successfully been used to investigate the population dynamics of various stock complexes in Icelandic waters, the Barents Sea, the North Sea, the Irish and Celtic Seas, and the Sofala Bank fishery of Mozambique.

Simulation Using Gadget

A Gadget model is a parametric forward-simulation model of an ecosystem, typically consisting of various fish populations and their interactions. Such a model simulation will track the changes to the fish populations due to both biological actions (e.g. growth) and also the interaction between populations (e.g. predation).

Gadget works by keeping track of the number N_{smalrt}, and mean weight W_{smalrt}, of fish in a ”population cell”. The population cell consists of all the fish of a given species (s), and maturation stage (m), with a specified age (a) and length (l), living in a specified region (r), on a specified timestep (t). Mathematically, this is simply a pair of numbers (over each of the six indices) that can be adjusted by various functions to simulate the effect of biological processes on the population cell.

Clearly, for any moderately complex model, this will involve the storage of a large amount of data, and a number of steps have been taken to reduce this. Gadget runs a forward simulation, Markovian, model so that the population at time $t + 1$ is only dependant on the population at time t, and not any earlier timestep, which means that Gadget only has to store the population from the previous timestep. Also, Gadget combines the species and maturation stage of the fish into a single index (termed ”population group” or, perhaps erroneously, ”stock”) since it is assumed that differing maturation stages of a species are biologically different (e.g. there can be a different growth rate for an immature fish compared to a mature fish) so that they can effectively be considered as different species by the simulation.

Stocks

A Gadget ”stock” or ”population group” is a group of fish which are all considered to share similar biological characteristics (e.g. growth, mortality, maturation ogive). A typical such stock could be an entire species, all mature fish in a species, or even all mature females from one region in a species.

Each stock is defined by specifying the length groups, age groups, and length-weight relationship to be used, along with the functions that are to be used to simulate the biological processes that affect the stock. The oldest age group and the longest length group of the stock are both treated as plus groups. A full description of each of the available functions that are available to simulate biological processes in Gadget is outside the scope of this paper, and interested readers are referred to the Gadget User Guide (Begley 2004) for a full list.

Growth

Modelling the change in length, and corresponding mean weight, due to growth is an important process in a marine ecosystem model, and there are a number of growth functions that are included within the Gadget framework. All the growth functions work by calculating a mean increase (in either length or weight) and then implementing this increase by moving a number of fish from one length group up to longer length groups.

For fish in a given length cell, the mean increase in either weight, or length, is calculated according to a growth equation (for example by a function based on a Von Bertalanffy growth equation). The corresponding increase in length, or weight, is then calculated, based on the length weight relationship defined for the stock.

The mean increase in length and weight is then translated into a statistical distribution of actual growths around that mean increase. This distribution has the effect of splitting the calculated increase up into a number of discrete length groups, so that a proportion of the fish increase in length by zero length groups, some by one length group, some by two and so on, such that the mean of all the increases is equal to the calculated mean growth. The statistical distribution that is used by Gadget is the beta-binomial distribution, which gives a high degree of flexibility from a single parameter.

Predation

For any model of an ecosystem that has more than one stock, it is usually of interest to model the interaction between these stocks, typically by considering the consumption of a prey by a predator. When this is extended to consider an ecosystem with many stocks, each predator can consume more than one prey and each prey can be consumed by more then one predator.

For each predator-prey relationship, the preference of a predator from length group L for a prey from length group l is modelled using a suitability function, $S(L, l)$. Multiplying this suitability value by the prey biomass will then give the prey abundance available for each predator to consume. There is a maximum amount $H(L)$ that each predator can consume (this is typically obtained from laboratory experiments), which is multiplied by the fraction of the total food abundance that the prey accounts for, to give the biomass of prey that each predator would want to consume, assuming that enough food is available to meet the needs of all predators. This is then scaled by multiplying by the number of predators in length group L.

The effect of the predation is calculated for each predator-prey relationship in turn, and stored without actually applying this effect to the ecosystem until all the predation for that timestep has been calculated. Then the overall mortality induced in the prey species by all the predators is calculated, and prey biomass is removed from the model. If the total consumption of any prey by all the predators amounts to more than the biomass of prey available, then the model runs into ”understocking”. In this case, the consumption by the predators is adjusted so that no more than 95% of the available prey biomass is consumed, and a penalty is applied to the likelihood score obtained from the simulation.

Maturation

For any model that has many stocks modelling different maturation stages of the same species it is important to be able to simulate the maturation process by moving fish from one stock to another. This is done by calculating the proportion of each age-length cell of the immature stock that will mature on a given timestep, and then moving this proportion into the corresponding age-length cell for the mature stock.

The proportion of the immature stock that will mature is calculated from a maturity ogive that is calculated from parameters specified by the user. The maturation probabilities may be age and/or length dependant, and, by using multiple stocks, maturation may also be made sex-dependant. Note that when moving the fish from one stock to another, the total number of fish in the ecosystem is kept constant, and that there is no mortality associated with the maturation process.

Recruitment and Reproduction

In order to track the development of an ecosystem over time, it is necessary to add new fish into the model to replace the fish that are removed (e.g. by commercial fishing). Gadget has two mechanisms to simulate this - either by adding recruits into the simulation or by modelling a closed life-cycle system.

Adding recruits to the stock takes a simplistic approach. The total number of recruits on any given timestep is specified (either as an absolute number or a parameter), along with the mean length and standard deviation of the length. A Normal distribution is then used to calculate a length distribution for the recruits, and a length - weight relationship is used to calculate a mean weight for each of the length cells of the recruits. These new recruits are then simply added to the youngest age group of the stock. The model may then be optimised to select the values for each recruitment parameter that produces the population giving the best fit to the available data.

An alternative approach that can be taken is to model a closed life-cycle system, where the number of recruits to a stock is governed by the spawning population. The total number of recruits is calculated (either based on a function of the fecundity, or the spawning stock biomass, of the 'parent' stock). The mean length and standard deviation of the length of the recruits is specified by the user, enabling a Normal distribution for the lengths of the recruits to be calculated, and these new recruits can then be added to the simulation in a similar manner to that described above. This spawning process can also affect the parent stock, and the mortality and weight loss attributable to the spawning process is also modelled by Gadget.

Migration

For any model that has a stock that is found on more than one region within the simulation, it is important to be able to simulate the migration of the stock as the fish move from one region to another. This is done in Gadget by calculating the proportion of each age-length cell of the stock that will move from one region to another on a given timestep, and then removing this proportion from the original region and adding it to the corresponding age-length cell for the destination region.

The proportion of the stock that will migrate from one region to another is calculated from migration matrices specified by the user. These migration matrices M_{ij} specify the proportion of the stock that will migrate from region i to region j on each timestep.

Other Processes

At the end of each year in the simulation, all the fish that are currently in the simulation will increase in age. This is done by moving all fish up one age group. Note that the oldest age group is modelled in Gadget as a plus group, and so this age group doesn't increase in age.

In addition to the mortality induced by some of the other processes (in particular predation), Gadget allows for a proportion of the stock to be removed due to other natural causes. This is simply modelled by specifying the proportion of each age group that will be removed from the simulation on each timestep.

Fleets

A Gadget "fleet" is treated as a simplified predator of the various prey stocks. Thus a fleet can be thought of a "stock" that has a single age group, a single length group, which doesn't grow, mature, migrate, recruit or reproduce. The only process that is of interest is the predation, where the fleet will 'consume' an amount of the prey stocks.

Gadget has two methods that can simulate effect on the ecosystem of the removal of fish due to fishing effort - either by specifying the biomass of the fish that is caught by the fleet (as recorded by landings data), or by specifying the fishing effort of the fleet. Both methods have similarities with the predator-prey interaction described above, simplified slightly since the predator has only one length group.

The effect of specifying the biomass of the fish that is caught by the fleet is to fix the 'consumption' by the predator of the prey to the specified biomass. Since there is only one length group for the predator, this consumption can be shared amongst the length groups of the prey, using the suitability function as a multiplier. This approach to the simulation of the predation is termed "TotalFleet", and requires the landings data to be known.

Alternatively, by specifying a fishing effort parameter, the ’consumption’ by the predator of the prey is fixed to a given proportion of the available prey biomass, after taking the suitability function into account. This approach to predation is termed ”LinearFleet”, and doesn’t require any landings data, so it can be used for a prediction model running the simulation into the future.

Comparisons Using Gadget

For a Gadget simulation, the data requirements are minimal - either none or just the overall catch for the fleets for each timestep, depending on the approach taken to model the fishing effort. However, for the model to be statistically testable, likelihood data is required. The data that is used depends on what aspects of the simulation are of particular interest, and could be length distributions, age length keys, survey abundance indices, mean length or weight at age, stomach content data or returns from tagging experiments. Unlike most fisheries models, catch at age data is not required, though it can be used if available.

Each data set that is to be used is assigned to a likelihood component. This specifies the statistical relationship to be used when comparing model results to the data. There are a number of likelihood components that are available to compare a Gadget simulation to likelihood data. A number of factors will influence the choice of the likelihood components, but the most important two factors to consider are availability and relevance of the data that is used to make the comparisons.

To make best use of the data that is available, it is sometimes necessary to aggregate sparse data, or even to exclude data from some timesteps completely. Gadget will only make a data comparison for timesteps that have data available, but a zero entry (from a sparse data source) is taken to mean zero, and not data missing due to poor sampling, which can lead to the comparison being unrepresentative of the data. The likelihood data does not need to be continuous, and the level of aggregation used doesn’t need to be the same for all likelihood components used in the model.

Various likelihood types are used to define the various likelihood components that can be used to calculate the ”goodness of fit” of the Gadget model to the available data. Each likelihood component will calculate a likelihood score for that individual component. A weighted sum of all the likelihood scores is then used to calculate an overall likelihood score. A full description of each of the available likelihood components that can be used in Gadget is outside the scope of this paper - interested readers are referred to the Gadget User Guide for a full list (Begley 2004).

Distribution Data

A common source of data for fisheries studies is age- or length- distribution data, either from sampling the commercial catch or from government surveys. This data can be used in Gadget by specifying a ”CatchDistribution” likelihood component, and the distribution data can either be aggregated into age groups (giving a distribution of length groups for each age), length groups (giving a distribution of age groups for each length) or into age-length groups. The likelihood score that is calculated gives some measure as to how well the distribution from the model fits to the distribution observed in the sampling process.

Various different functions to compare the distribution data are available in Gadget, ranging from a simple sum-of-squares comparison (as shown in equation 4.1 below) through to more complicated multivariate Normal distributions. The choice of equation used to compare the distribution data is left to the user.

$$\ell = \sum_{time} \sum_{regions} \sum_{ages} \sum_{lengths} \Big(P_{tral} - \pi_{tral}\Big)^2 \tag{4.1}$$

where P is the proportion of the distribution from the data sample, and π is the corresponding proportion from the model sample, for that time/region/age/length combination

Abundance Data

Another common source of data that can be used in a Gadget model is information from abundance indices that are calculated from standardised government surveys. This data can be used by specifying a ”SurveyIndices” likelihood component, and the abundance data can either be aggregated into age groups or length groups. This likelihood component is used to compare the relative abundance of a stock in the Gadget simulation to the indices calculated from a standardised survey for that stock.

The likelihood score is calculated by fitting a linear regression line to the difference between abundance indices calculated within the model and abundance indices calculated directly from standardised survey data, as given by equation 4.2. The model will calculate an index for the population, and then fit a linear regression line between these calculated indices and those input from data files.

$$\ell = \sum_{time} \Big(I_t - (\alpha + \beta \widehat{I_t}) \Big)^2 \tag{4.2}$$

where I is the abundance index from the standardised survey, and $\widehat{I}$ is the corresponding index calculated in the Gadget model.

The exact format of this linear regression equation can vary, depending on survey index data available. It is possible to take the log of the indices and the modelled data before fitting the linear regression line. The slope and intercept of the linear regression line are controlled by the parameters alpha and beta, and it is possible to fix these to specified numbers, or let Gadget calculate these to get the best fit to the modelled data.

Statistical Data

An alternative use for data obtained by sampling the commercial catch or government surveys is to calculate the mean of some biological measurement and then compare this mean value to the corresponding value obtained from the simulated ecosystem data. This data is used in Gadget by specifying a ”CatchStatistics” likelihood component, and is typically used to compare mean length at age, or mean weight at age.

The model will calculate the mean length (or weight) of the stock that is caught according to the model parameters, and aggregate this into specified age groups. The likelihood score is calculated as a function of the difference between the mean data calculated by the model and that caught according to the landings data and specified in the data file. Various functions are available for the user to make this comparison, of the form shown in equation 4.3 below.

$$\ell = \sum_{time} \sum_{regions} \sum_{ages} \Big(\frac{(x_{tra} - \mu_{tra})^2}{\sigma_{tra}^2} N_{tra} \Big) \tag{4.3}$$

where x is the sample mean length from the data, and μ is the corresponding mean length calculated from the model, σ is the standard deviation of the length, calculated from the model and N is the sample size for each time/region/age combination

Stomach Data

Data obtained from analysing the stomach contents of various predators can be used to give an indication of the diet composition of the stock. This data is used in Gadget by specifying a ”StomachContent” likelihood component, and is typically used to compare the ratios of different prey stocks found in the diet of the predator. Care is needed when making this comparison, since the data will give information on the stomach content at the time of capture, where as the simulation can currently only give information about the modelled consumption of the prey.

The likelihood score is calculated by comparing the ratio of the consumption of different preys by a predator in the model to the ratio of the preys found in the stomach contents data specified in the input file. Various functions are available for the user to make this comparison, of the form shown in equation 4.4 below.

$$\ell = \sum_{time} \sum_{regions} \sum_{prey} \Big(P_{trp} - \pi_{trp} \Big)^2 \tag{4.4}$$

where P is the proportion of the stomach content from the data sample, and π is the corresponding proportion from the model sample, for that time/region/prey combination

Tagging Data

Data obtained from analysing the returns information from tagging experiments can be used to give an indication of the migration patterns of the stock. This data is used in Gadget by specifying a ”Recaptures” likelihood component, and also specifying the tagging experiments that have taken place (Hannesson et al. 2004).

The likelihood score is calculated by comparing the recaptures data obtained from the tagging experiment with an expected recapture value. The model will calculate an expected recaptures value for the tagging experiment, and then use a Poisson function (as shown in equation 4.5 below) to compare this to the recaptures obtained from the tagging experiment, as given in the data files.

$$\ell = \sum_{time} \sum_{regions} \sum_{length} \Big(N_{trl} + \log \nu_{trl}! - N_{trl} \log \nu_{trl} \Big) \tag{4.5}$$

where N is the recapture sample size from the tagging experiment, and ν is the corresponding sample size from the model, for that time/area/age/length combination

Optimisation Using Gadget

The overall likelihood score gives a single measure as to how well observations from the modelled ecosystem fit to the data that has been provided as likelihood components. Varying the value of the parameters that have been used to simulate the ecosystem will cause the fit to the data to change, and the overall likelihood score will also change to reflect this. By using a search algorithm to iteratively adjust the parameters, it is possible to find a set of parameter values that give the lowest likelihood score, and thus correspond to the best fit of the model to the data. The functional form of the model (number of population groups, choice of growth equation etc.) remain fixed, while the parameters are optimised. This computationally intensive search process is handled by Gadget, using several robust algorithms, which can be used either on their own or by combining them into a single hybrid algorithm.

Hooke & Jeeves

Hooke & Jeeves (Hooke and Jeeves 1961) is a simple and fast local minimisation algorithm. From an initial starting point the algorithm takes a step in various ’directions’, by adjusting one parameter at a time, and conducts a new model run at each new point. If the new likelihood score is lower than the old one then the algorithm uses the new point as it's best guess, and re-starts from this new point. However, if the new likelihood score is higher, then the algorithm returns to the old point, and tries again by adjusting a different parameter. The search proceeds in series of these steps, each step slightly smaller than the previous one. When the algorithm finds a point which it cannot improve on with a small step in any direction, it accepts this point as being the 'minimum', and records this point to a data file before exiting.

Simulated Annealing

Simulated Annealing (Corona et al. 1987) is a global minimisation algorithm. From an initial starting point the algorithm takes a random step in various 'directions', by adjusting one parameter at a time, and conducts a new model run at each new point. If the new likelihood score is lower than the old one then the algorithm stores the new point as it's best guess, and re-starts from this new point. However, if the likelihood score is higher then the algorithm may still accept this point, based on the probabilistic "Metropolis Criteria", and thus the algorithm can escape from a local minimum. The algorithm exits when a stable point is found which cannot be improved on with a small step in any direction, and the Metropolis Criteria rejects all the steps away from the current best point. The best point is then accepted as being the 'minimum', and is recorded to a data file, and the algorithm exits.

The Metropolis Criteria will accept a move to a point with a higher likelihood score based on a function of both the size of the move and a parameter termed "temperature", and is given in equation 4.6 below:

$$\begin{aligned} M &= e^{\frac{-\Delta F}{T}} \\ P &= \begin{cases} 1 & \text{if } M > r \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{4.6}$$

where ΔF is the change in likelihood score, T is the "temperature" of the algorithm, and r is a random number between 0 and 1.

Note that when the "temperature" is very high ($T \rightarrow \infty$), the Metropolis Criteria will always accept any move, and the Simulated Annealing algorithm will simplify to a form of a random search algorithm. Conversely, when the temperature is very low ($T \rightarrow 0$), the Metropolis Criteria will always reject any move, and the Simulated Annealing algorithm will simplify to a form of a local search, similar to Hooke & Jeeves. By slowly reducing the temperature of the algorithm, the number of moves that are accepted by the Metropolis Criteria are reduced and the algorithm will find a minimum.

Summary

The Gadget software provides a "tool box" for constructing, running and optimising statistical marine ecosystem models. It is open source, and freely downloadable, with documentation and examples available online. The software consists of three parts; a biologically realistic parametric forward-simulation model; a suite of statistical functions to compare the model output to data; an optimisation suite to adjust the parameters of the simulation model to give the best possible fit to the data.

Gadget provides the user with a choice of functional forms for the different biological processes within the model. The suite of statistical functions allows for a wide variety of different data sources to be incorporated into the modelling process in a statistically rigorous manner. There is a high degree of flexibility in the utilisation of different data sources, allowing for different time coverage and aggregation levels between the data sources. A hybrid optimisation algorithm is incorporated within Gadget, combining the wide-area searching capabilities of Simulated Annealing with the fast local convergence of Hooke & Jeeves. By linking the Gadget software with the custom designed data warehouse it is possible to produce new or modified Gadget models with minimum effort.

Gadget models have been, and are being, constructed for a variety of ecosystems, ranging from arctic cod to sub-tropical shrimp. The models have been used to investigate the biology of fish population and ecosystems as well as in an assessment context. The flexible framework provided by Gadget provides an important tool for generating statistically rigorous, biologically realistic, age and length structured models.

References

Anon. 2002. Development of Structurally Detailed Statistically Testable Models of Marine Populations (dst^2). QLK5-CT1999-01609. Progress Report for 1 January to 31 December 2001. Marine Research Institute Report No. 87, Marine Research Institute, Reykjavik, Iceland.

Anon. 2003. Development of Structurally Detailed Statistically Testable Models of Marine Populations (dst^2). QLK5-CT1999-01609. Progress Report for 1 January to 31 December 2002. Marine Research Institute Report No. 98, Marine Research Institute, Reykjavik, Iceland.

Anon. 2004. ACFM Report of the Working Group on Methods of Fish Stock Assessments, ICES CM 2004/D:03 Ref. G. 11-18 February 2004 Lisbon, Portugal.

Begley, J. 2004. Gadget User Guide, available online from the Gadget web page.
`http://www.hafro.is/gadget`

Bogstad, B., Hiis Hauge, K., and Ulltang, Ø. 1997. MULTSPEC - A Multi-Species Model for Fish and Marine Mammals in the Barents Sea. Journal of Northwest Atlantic Fisheries Science. vol 22: pp317-341.

Bogstad, B., Howell, D., and Åsnes, M.N. 2004. A closed life-cycle model for Northeast Arctic cod. ICES C.M.2004/K:26.

Corana, A., Marchesi, M., Martini, C., and Ridella, S. 1987, Minimising Multimodal Functions of Continuous Variables with the 'Simulated Annealing' Algorithm. ACM Transactions on Mathematical Software vol 13: pp262-280.

Frøysa, K.G., Bogstad, B., and Skagen, D.W. 2002. Fleksibest - an age-length structured fish stock assessment tool with application to Northeast Arctic cod (Gadus morhua L.). Fisheries Research vol 55: pp87-101.

Hannesson, S., Jakobsdóttir, A., Begley, J., Taylor, L.A., and Stefánsson, G. 2004. On the use of tagging data in statistical multi-species multi-area models of marine populations. ICES CM 2004/FF:30.

Hooke, R., and Jeeves, T.A. 1961. Direct Search Solution of Numerical and Statistical Problems. ACM Journal. vol 8: pp212-229.

Kupča, V. 2004. A standardised database for fisheries data. ICES CM 2004/FF:15.

Stefánsson, G., and Pálsson, Ó.K. 1997. BORMICON. A Boreal Migration and Consumption Model. Marine Research Institute Report 58, Marine Research Institute, Reykjavik, Iceland.

Taylor, L.A., Begley, J., Kupča, V., and Stefánsson, G. 2004. A Simple implementation of Gadget for cod in Icelandic waters. ICES CM 2004/FF:23.

4.2 Growth models in population simulations

Gunnar Stefansson
Marine Research Institute, Reykjavik, Iceland

Abstract: This paper describes methods to test and implement growth submodels within general population dynamics models.

Introduction

Early models of fish population dynamics tended to focus on individual aspects of the dynamics, one at a time. Thus, the virtual population analysis (VPA) of Gulland (19659 contains only mortality components, the von Bertalanffy growth curve includes only growth, single-species yield per recruit analysis normally only considers a fixed growth and mortality pattern, although some simple density dependence can also be included (Beverton and Holt, 1957). Similarly, early multispecies models only considered the effects of predation on prey mortality (MSVPA) as in Helgason and Gislason (1979) or other single effects such as of consumption on predator growth (Magnusson and Palsson, 1991).

More comprehensive single species models such as versions of Stock Synthesis (Methot, 1989) include many of these components at the same time, e.g. a mortality model, and a growth model and a thus selection pattern based on length rather than age can be used.

Recent multispecies models of fish population dynamics such as MULTSPEC (Tjelmeland and Bogstad, 1998) and Bormicon (Stefansson and Palsson, 1997) include not only mortality and growth functions, but the growth functions can be dynamic and will typically depend on the amount consumed. The varous processes which need to be accounted for in these models are described e.g. in Stefansson and Palsson (1998). At a given time step and area, these models typically have as their main internal component the number of fish in an age and length group. Growth models become an issue of how to transfer fish between length groups. Both MULTSPEC and Bormicon include very simple models of these updates.

This paper describes a growth modelling technique which is considerably more flexible and lends itself to statistical estimation and evaluation methods. This particular approach is developed as a part of a new area-disaggregated multispecies model, Gadget.

Models of the general population dynamics

This paper will focus on and compare different models of fish growth. In order to do this some assumptions need to be made about the underlying population dynamics, since data on length or weight of fish are collected within variable environmental and harvest regimes.

The basic population dynamics model is the formulation of Bormicon (Stefansson and Palsson, 1997), implemented using the Gadget program code (Begley, 2004).

A model of average population growth

The models considered in this section are assumed to be based on consumption. Thus, initial growth in weight could be computed using energy conversion techniques or other methods using stomach content data. Growth in weight subsequently defines growth in length. Thus, it is assumed that a growth in length is predicated.

Suppose a group of fish of a specific age, a and in a length group with mean length l are destined to grow by ΔL. If the growth of these fish has the Markovian property with respect to length, then it only depends on the initial length and predicated growth during a time step.

A simple model of length growth is the von Bertalanffy model,

$$L_t = L_\infty \left(1 - e^{-Kt}\right),\ t \in \mathbf{R}_+, \tag{4.7}$$

describing the length of a group of fish (or an individual) at each time step, t. If this model is assumed to be applicable for a group of fish, then the growth during a time period is

$$\begin{aligned}
\Delta L &= L_{t+\Delta t} - L_t && (4.8)\\
&= L_\infty \left(1 - e^{-K(t-\Delta t)}\right) L_\infty \left(1 - e^{-Kt}\right) && (4.9)\\
&= L_\infty \left(e^{-Kt} - e^{-K(t-\Delta t)}\right) && (4.10)\\
&= L_\infty e^{-Kt} \left(1 - e^{K\Delta t}\right) && (4.11)
\end{aligned}$$

Suppose fish of a given age, a, happen to be of length l. In terms of their size, they can be considered of biological age t which corresponds to this size, i.e. t is defined by $L_t = l$:

$$t = -\frac{1}{K} ln \left(1 - \frac{l}{L_\infty}\right) \tag{4.12}$$

Equivalently, the growth is given by

$$\begin{aligned}
\Delta L &= L_\infty \left(1 - \frac{L}{L_\infty}\right) \left(1 - e^{K\Delta t}\right) && (4.13)\\
&\simeq KL_\infty \left(1 - \frac{L}{L_\infty}\right) \Delta t, && (4.14)
\end{aligned}$$

where the approximation is based on the Taylor approximation to the exponential function.

Either equation 4.13 or 4.14 can be used to define growth during a time step in simulations and 4.14 is probably the more common of the two. Naturally, 4.14 is also the basis for the von Bertalanffy model itself, as 4.14 provides the differential equation

$$\frac{dL}{dt} = K \left(L_\infty - L\right)$$

whose solution is 9.1. Early Gadget models used the approximation 4.14, but it was found that the exact equation was a considerable improvement, particularly for the purpose of debugging models where it is highly desirable to obtain exact numerical results.

For a prespecified growth, ΔL, e.g. from consumption estimates, equations 4.8-4.12 are not needed. However, when consumption models are not used, equation 4.12 can be used in 4.8 in order to specify the average growth for the group of fish which start out at length L. In the present paper the growth in weight and growth in length may potentially be linked so that a length-weight relationship of the form

$$W = cL^b \tag{4.15}$$

is maintained. In this relationship, c becomes the condition factor, W/L^3 when $b = 3$ but c has little direct meaning otherwise.

It should be noted that when l is greater than L_∞, the biological age is undefined as the fish has exceeded it maximum length. In this case it will be assumed that the fish do not grow further.

Also, the average growth of a group of fish next needs to be translated into a new distribution of these fish in a subsequent time step. This will be handled in subsequent sections.

Growth in length from growth in weight

One set of growth functions which has been proposed is to base length growth on weight growth and compute weight growth from consumption. In order to verify different methods of implementing growth, it is useful to be able to link these approaches together. A generic method for implementing this will be described, but in the context of von Bertalanffy growth in length.

If the von Bertalanffy function is assumed to hold on length scale and a length-weight relationship holds, $W = aL^b$, then the following relationships are obtained for the length and weight updates:

$$\frac{\Delta W}{\Delta L} \simeq \frac{dW}{dL} = cbL^{b-1} \tag{4.16}$$

and from this it follows that

$$\begin{aligned}
\Delta W &\simeq \frac{dW}{dL}\Delta L && (4.17)\\
&= cbL^{b-1}\Delta L && (4.18)\\
&\simeq KL_\infty cbL^{b-1}\left(1-\frac{L}{L_\infty}\right)\Delta t && (4.19)\\
&= KL_\infty cb\left(L^{b-1}-\frac{L^b}{L_\infty}\right)\Delta t && (4.20)\\
&= KL_\infty^{b-1}cb\left(\left(\frac{L}{L_\infty}\right)^b-\left(\frac{L}{L_\infty}\right)^b\right)\Delta t && (4.21)
\end{aligned}$$

Adding temperature as a variable, T, it is seen that qrowth in weight can be updated using general formulae of the form

$$\Delta W = q_0 e^{q_1 T}\left(\left(\frac{W}{q_2}\right)^{q_4}-\left(\frac{W}{q_3}\right)^{q_5}\right)\Delta t \tag{4.22}$$

where the q's are constants. This is the growth update equation used in BorMiCon, the Gadget predecessor described in Stefansson and Palsson (1997). The von Bertalanffy growth is obtained through the choice

$$\begin{aligned}
q_0 &= cbKL_\infty^b && (4.23)\\
q_1 &= 0 && (4.24)\\
q_2 = q_3 &= W_\infty = cL_\infty^b && (4.25)\\
q_4 &= (b-1)/b && (4.26)\\
q_5 &= 1 && (4.27)
\end{aligned}$$

It follows that for given (mean) growth in weight, mean length growth can subsequently be implemented using

$$\Delta L := \frac{\Delta W}{cbL^{b-1}}, \tag{4.28}$$

which uses the derivative of the length-weight relationship (4.15) to deduce length growth from growth in weight.

The above description of growth in weigth uses a Taylor approximation to the length-weight relationship. Although used extensively (see e.g. Stefansson, 1997), this has the same problems as using the Taylor approximation to length growth. Notably model verification can become quite difficult as seen when using simulated data as in Sigurdardottir et al (2005).

For this reason it would be desirable to use exact formulae such as

$$\begin{aligned}
\Delta W &= c\left\{(L+\Delta L)^b - L^b\right\} && (4.29)\\
& && (4.30)
\end{aligned}$$

in place of 4.17. This equation can of course easily be inverted to provide ΔL as a function of other values in this equation (i.e. to specify length growth from consumption), but this does not resolve all issues, as will be clear below. The Gadget program still uses 4.17, as inherited from BorMiCon.

The equations in this section describe possible methods for implementing and verifying average population growth, either by specifying ΔW based on consumption or by specifying length growth ΔL based on predicating von Bertalanffy length growth. Either approach can be used as a basis for the following sections.

Updating length distributions

Fish population dynamics are modelled in MULTSPEC and Bormicon through forward simulations of fish populations, allowing fish to migrate between areas, die, grow, mature and spawn. The basic unit in these models is the number of fish in a certain model "cell". The fish in a "cell" are in the same age and size group, in the same region and time step. When this basic model formulation is used, the numbers in a "cell" need to be updated during a given time step, so as to reflect all processes being modelled. In addition to growth, these processes include migration, spawning, natural and fishing mortality. Only growth will be considered in this document.

A given amount consumed predicates an average growth for the fish in a given "cell". This growth can be either in weight or length or both, but only length growth will be considered here.

A common approach is to start with a predicted average length increment (ΔL) based on consumption and to try to distribute fish in the length class into upper length classes in a reasonable manner. Simple techniques may use only few upper length intervals and use a simple ad-hoc update scheme. The update scheme needs to be evaluated in terms of its ability to provide adequate eventual length distributions. This sets some immediate bounds on the dispersion at each time step, since an overly high or low variance in the length update will quickly result in inadequate final length distributions at age for the oldest ages.

The length update scheme can most easily be implemented through a look-up table, where a discrete set of ΔL-values is provided along with the distribution to be used for reallocating the length group when the chosen growth is ΔL. This is undesirable for many reasons. Firstly, the setup is completely rigid as there is no built-in parameter to describe possible possible deviations of growth from the specified distribution and hence data on growth may adversely affect parameters in other parts of a complex model only because of incorrect specification of the rigid relationship. Secondly, a simple discrete (rounded) lookup provides a nondifferentiable likelihood function which will result in estimation problems later on.

What is needed is a way to specify a flexible parametric distribution with enough parameters to allow minimal flexibility to track length distributions of an age group, yet with enough parsimony in parameters to allow for the estimability of the parameters involved.

A formal model for the update

Although a first step might be to attempt to estimate individual probabilities, these would result in far too many parameters. Another approach would be to estimate variance, skewness and kurtosis and go from these to transition probabilities, but there is no trivial transformation between the two.

A flexible distribution such as the 4-parameter inverse lambda distribution could probably be used (Ramberg et al, 1979), but parameter estimation tends to be difficult. Similarly, a binomial distribution (or even a a (truncated) Poisson) can be used, but both are completely rigid, since the value of ΔL completely specifies the single free parameter in each of these distributions (assuming the number of permissible length group increases to be fixed).

The beta-binomial distribution can be used as a simple alternative. This approach can be formulated so as to provide a single estimable parameter in addition to the mean, which is specified by ΔL.

First consider the binomial distribution which is defined for integers, $x = 0, \ldots, n$ by

$$\binom{n}{x} p^x (1-p)^{n-x} = \frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x+1)} p^x (1-p)^{n-x}$$

Using this distribution for the issue at hand, for a given n, the other parameter, p, of this distribution, is fully defined since $\mu = np$ and the mean growth is given as the specified ΔL, which fixes $p = \Delta L/n$. Although this distribution can certainly be used, it is clear that no flexibility is allowed at all and in fact it would be quite unlikely for such a rigid distribution to satisfy the specified requirements of attaining the correct final distribution of length at age.

A common approach to more flexibility is to allow the parameter p itself to come from another distribution, often the beta distribution.

The beta distribution is defined for arbitrary values of $\alpha > 0$ and $\beta > 0$ by

$$f(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}, \quad 0 \le p \le 1 \tag{4.31}$$

and it is well-known that the mean of this distribution is given by

$$E(p) = \frac{\alpha}{\alpha+\beta} \tag{4.32}$$

Thus, rather than using a fully specified binomial distribution, more flexibility is obtained by using this combined beta-binomial distribution. This approach results in the following marginal distribution of the length increments:

$$\begin{aligned}
P[X=x] &= \int_{p=0}^{1} P[X=x|p] f(p) dp \\
&= \int_{p=0}^{1} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \\
&= \frac{\Gamma(n+1)\Gamma(\alpha+\beta)}{\Gamma(x+1)\Gamma(n-x+1)\Gamma(\alpha)\Gamma(\beta)} \int_{t=0}^{1} p^{x+\alpha-1}(1-p)^{n-x+\beta-1} \\
&= \frac{\Gamma(n+1)\Gamma(\alpha+\beta)}{\Gamma(x+1)\Gamma(n-x+1)\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(n+\alpha+\beta)} \\
&= \frac{\Gamma(n+1)}{\Gamma(n-x+1)\Gamma(x+1)} \frac{\Gamma(\alpha+\beta)}{\Gamma(n+\alpha+\beta)} \frac{\Gamma(n-x+\beta)}{\Gamma(\beta)} \frac{\Gamma(x+\alpha)}{\Gamma(\alpha)}
\end{aligned}$$

Since for any positive number, y, the relationship $\Gamma(y+1) = y\Gamma(y)$ holds, it also follows that for any integer $x \ge 1$ and $\alpha, \beta > 0$,

$$\frac{\Gamma(x+\alpha)}{\Gamma(\alpha)} = \begin{cases} 1 \text{ for } x = 0 \\ \alpha \text{ for } x = 1 \\ (x-1+\alpha)(x-2+\alpha)(x-3+\alpha) \cdot \ldots \cdot (\alpha+1)\alpha \text{ for } x > 1 \end{cases}$$

and

$$\frac{\Gamma(n-x+\beta)}{\Gamma(\beta)} = \begin{cases} 1 \text{ for } n-x=0 \\ \beta \text{ for } n-x=1 \\ (n-x-1+\beta)(n-x-2+\beta)(n-x-3+\beta)\cdot\ldots\cdot(\beta+1)\beta \text{ for } n-x>1 \end{cases}$$

and finally, for $n \geq 1$:

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(n+\alpha+\beta)} = \frac{1}{(n-1+\alpha+\beta)(n-2+\alpha+\beta)(n-3+\alpha+\beta)\cdot\ldots\cdot(1+\alpha+\beta)(\alpha+\beta)}$$

It follows that for $n \geq 1$ the above probabilities can be rewritten as

$$P[X=x] = \begin{cases} \frac{\Gamma(n+1)}{\Gamma(n-x+1)\Gamma(x+1)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(n+\alpha+\beta)} \cdot \frac{\Gamma(n-x+\beta)}{\Gamma(\beta)} \cdot \frac{\Gamma(x+\alpha)}{\Gamma(\alpha)} & \text{for } x=0 \\ \frac{\Gamma(n+1)}{\Gamma(n-x+1)\Gamma(x+1)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(n+\alpha+\beta)} \cdot \frac{\Gamma(n-x+\beta)}{\Gamma(\beta)} \cdot \frac{\Gamma(x+\alpha)}{\Gamma(\alpha)} & \text{for } x=1 \\ \frac{\{n\cdot(n-1)\ldots\cdot(n-x+1)\}}{x!} & \\ \cdot \quad \frac{(n-x-1+\beta)(n-x-2+\beta)(n-x-\ 3+\beta)\cdot\ldots\cdot(\beta+1)\beta}{(n-1+\alpha+\beta)(n-2+\alpha+\beta)(n-3+\alpha+\beta)\cdot\ldots\cdot(1+\alpha+\beta)(\alpha+\beta)} & \\ \cdot \quad (x-1+\alpha)(x-2+\alpha)(x-\ 3+\alpha)\cdot\ldots\cdot(\alpha+1)\alpha & \text{for } 1 \leq x \leq n-2 \\ \frac{\{n\cdot(n-1)\ldots\cdot(n-x+1)\}}{x!} & \\ \frac{\Gamma(n+1)}{\Gamma(n-x+1)\Gamma(x+1)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(n+\alpha+\beta)} \cdot \frac{\Gamma(n-x+\beta)}{\Gamma(\beta)} \cdot \frac{\Gamma(x+\alpha)}{\Gamma(\alpha)} & \text{for } x=n-1 \\ \frac{\Gamma(n+1)}{\Gamma(n-x+1)\Gamma(x+1)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(n+\alpha+\beta)} \cdot \frac{\Gamma(n-x+\beta)}{\Gamma(\beta)} \cdot \frac{\Gamma(x+\alpha)}{\Gamma(\alpha)} & \text{for } x=n \end{cases}$$

It should be noted that in the case of growth by length groups within a short time interval, only low values of n and x are needed.

Thus, probabilities can readily be generated from this beta-binomial distribution, given specified values of α, β and n. The last parameter of these, n, will usually be assumed, outside an estimation procedure.

It is also reasonably easy to see that the mean of the beta-binomial distribution is given by $\mu = nE[p] = \frac{n\alpha}{\alpha+\beta}$. If β is taken as a parameter to be estimated, the requirement $\Delta L = \mu$ therefore implies $\alpha = \frac{\beta\Delta L}{n-\Delta L}$.

This approach will therefore be implemented in Gadget by defining a new growth function with a single estimable parameter β, to be set (along with n) to an initial value in a specification file.

Growth after starvation

The general case will involve consumption and therefore growth in length and growth in weight will not always conform strictly to a length-weight relationship. Notably, if fish do not get enough food, they will typically lose weight but not get shorter. This effect needs to be accounted for in the models.

Define the function g with

$$g(x) := l_0 + x(l_1 + l_2 x), \tag{4.33}$$

the piecewise linear function f with

$$f(x) := \begin{cases} 0 & \text{if } l_3 + l_4 x \leq 0, \\ l_5 & \text{if } l_3 + l_4 x \geq l_5, \\ l_3 + l_4 x & \text{otherwise,} \end{cases}$$

and let

$$r(L) := \frac{W - W_{ref}(L) g(\psi(L))}{W}, \tag{4.34}$$

where $l_0, \ldots, l_5$ are constants and $\psi(L)$ is the feeding level. In the case of von Bertalanffy's growth function, when there is no consumption, the feeding level is an additional parameter, and is generally put equal to 0.5.

The formula used for the length increase can then be written as

$$\Delta L := \begin{cases} \frac{\Delta W}{cbL^{b-1}} f(r(L)) & \text{if } \Delta W > 0, \\ 0 & \text{if } \Delta W < 0. \end{cases} \tag{4.35}$$

This is a feasible approach and is an option in Gadget, but is normally low on the list of priorities since obtaining basic growth is a difficult task even in the absense of starvation issues.

Linking growth and length update distributions

In general, growth in weight, ΔW and growth in length, ΔL, may be determined using different formulae. The use of an update mechanism such as the beta-binomial distribution subsequently converts (mean) length growth into update probabilities for length distributions.

Some models, such as Gadget, maintain an average weight at length. In order to maintain the total growth in weight, a deviation from the length-weight relationship may need to be implemented. When going from length growth to weight growth through length update mechanisms and a length-weight relationship, one needs to consider the weight update for each lenght group.

Suppose fish of length L are destined to grow by ΔL and this results in a series of updates where n_i fish grow by δl_i. The weight increase of each of these will be $cL^b - c\left(L + \delta l_i\right)^b$. The overall weight increase will therefore become

$$\Delta W = \sum_i \left(cL^b - c\left(L + \delta l_i\right)^b \right)$$

Model flexibility

Future work includes a close scrutiny of how variations in the β parameter affect the growth update and how this affects the final length distributions, conditional on all other model components.

Simulated data is also needed to illustrate the interactive effects of growth parameters and parameters in the selection pattern.

Process error

It is clear that there will always be considerable unexplained variation in growth and it is also clear that this variation is persistent, i.e. it can not be treated as independent random errors (e.g. Millar, 1991).

Future work therefore must also consider the possibility of including process error, e.g. in the form of annual variation in β. A simple form for this is to assume $\beta_y \sim n(\beta, \sigma^2)$ where β is an unknown parameter but β_y is the annual effect in the growth.

Evaluation of the growth functions

Future testing of these models must include likelihood functions linking growth or modelled length distributions to data from surveys or commercial catches. In either case, however, the comparison will be confounded due to the well-known intra-haul correlation and overdispersion commonly found in fisheries data (Pennington and Völstad, 1994).

In addition, problems arise from the compound nature of the data sets used in these models: Since the models not only include growth but also other processes which can only be estimated by including other data sources such as abundance indices from surveys, there are many data sets in the full likelihood function to be used in the models. The question therefore arises on how to weight together these components. If probability distributions of the data could be fully specified, maximum likelihood could be used, but the question is particularly pertinent since the individual components are difficult to fully specify in light of the overdispersion mentioned above. In fact it turns out that model results are quite sensitive to the weighting use for the different data sets and therefore it is quite important to develop methods for the estimation of the weighting factors (Stefansson, 1998).

References

Beverton, R.J.H. and Holt, S.J. (1957). On the Dynamics of Exploited Fish Populations. Fish and Fisheries Series - 11. Chapman and Hall.

Begley, J. 2004. Gadget User Guide. Available from the Gagdet web page:
`http://www.hafro.is/gadget`.

Gulland, J. (1965). Estimation of mortality rates. Annex to Arctic Fisheries Working Group Report. ICES, C.M. Gadoid Fish Comm., (3). 9pp (mimeo)

Helgason, T. and Gíslason, H. (1979). VP-analysis with species interaction due to predation ICES C.M. 1979/G:52

Magnússon, K. G. og Palsson, O. K. (1991). Predator-prey interactions of cod and capelin in Icelandic waters ICES mar. Sci. Symp., 193 : 153-170 964

Methot, R.D. (1989). Synthetic estimates of historical abundance and mortality for northern anchovy. Amer. Fish. Soc. Symp. 6: 66-82.

Millar, R. B. (1991). Modelling the effect of environment on growth of cod: fitting to growth increment data versus fitting to size-at-age data ICES, C.M. G:27

Pennington, M. and Völstad, J.H. (1994). Assessing the Effect of Intra-Haul Correlation and Variable Density on Estimates of Population Characteristics from Marine Surveys. Biometrics 50.

Ramberg, J. S., Dudewicz, E. J., Tadikamalla, P. R. and Mykytka, E. F. (1979). A probability distribution and its uses in fitting. Technometrics 21, No. 2 201-214

Sigurdardottir, A. J., Olafsdottir, E. I. and Taylor, L. (2005). An R program to simulate Gadget population dynamics. This report.

Stefansson, G. and Palsson, O. K. (1997). BORMICON A Boreal Migration and Consumption model Marine Research Institute Tech. Rep. 58.

Stefansson, G., Palsson, O. K. (1998:). A framework for multispecies modelling of Boreal systems. Reviews in Fish Biology and Fisheries, 8: 101-104.

Stefansson, G. (1998). Comparing different information sources in a multispecies context. In: F. Funk [et al.] (eds): Fishery Stock Assessment Models : Proceedings of the international symposium; Anchorage 1997. 15th Lowell Wakefield Fisheries Sympo

Stefansson, H. N. (1997).Mathematical model. In Stefansson, G. and Palsson, O. K. (1997). BORMICON A Boreal Migration and Consumption model Marine Research Institute Tech. Rep. 58.

Tjelmeland, S., and Bogstad, B. 1998. MULTSPEC - A review of a multispecies modelling project for the Barents Sea. Fisheries Research 37: 127-142.

4.3 Spawning, Recruitment and Reproduction

James Begley
Marine Research Institute, Reykjavik, Iceland

Abstract: Gadget is a software tool that has been developed to model marine ecosystems, and to follow the changes to the populations in these ecosystems over time. A Gadget model is a parametric forward-simulation model of an ecosystem, typically consisting of various fish populations and their interactions. Such a model simulation will track the changes to the fish populations using as series of biologically realistic functions. In order to track the development of an ecosystem over time, it is necessary to add new fish into the model to replace the fish that are removed (e.g. by commercial fishing). Gadget has two mechanisms to simulate this - either by adding recruits into the simulation directly or by modelling a closed life-cycle system, as part of the spawning process of the mature stocks.

Spawning

Gadget simulates how the spawning process can affect the mature stocks, with the mortality and weight loss attributable to the spawning process modelled by Gadget. During the simulation of the spawning process, some proportion of the mature stock will spawn, and of those that spawn, some of the stock will be removed due to the spawning mortality whilst the rest will have their weight reduced.

In Gadget models, spawning is considered to be an annual event, that takes place on the same timestep and the same area in each year. When the adult stock spawns, the population of the age-length cells of the adult stock are reduced due to the mortality induced by the spawning process in the fish that spawn (as shown in equation 4.36). In addition to the reduction in numbers, the mean weight of the surviving fish in the age-length cells are adjusted to take the reduction in weight of the fish that spawn, and the change in population, into account (as shown in equation 4.37).

$$N = N\big(1 + P(e^{-m} - 1)\big) \tag{4.36}$$

$$W = W\frac{(1 + P((2 - w)e^{-m} - 1))}{(1 + P(2e^{-m} - 1))} \tag{4.37}$$

where:
N is the population of the age-length group
W is the mean weight of the fish in the age-length group
P is the proportion of the fish in the length group that will spawn
m is the spawning mortality of the fish in that length group
w is the spawning weight loss of the fish in that length group

The values for P, m, and w are given by the following length selection functions, based on the mean length of the length group.

Length Selection

The length selection functions determine the proportion of a length group that will be selected, based on a function of the mean length of that length group. Currently there are three selection functions defined in Gadget, although this can easily be extended in the future, and the valid length selection functions are:

- a **constant** selection function, that has no dependence on the length of the stock, as shown in the equation below, and is specified by the user defining the value of α in the input file.

$$s(l) = \alpha$$

- a **straight line** selection function, that has a linear dependence on the length of the stock, as shown in the equation below, and is specified by the user defining the values of α and β in the input file.

$$s(l) = \alpha l + \beta$$

- an **exponential** selection function, that has a logarithmic dependence on the length of the stock, as shown in the equation below, and is specified by the user defining the values of α and l_{50} in the input file - note that the length dependence is actually dependant on the difference between the length and l_{50}, which is the length of the stock with a 50% probability of selection.

$$s(l) = \frac{1}{1 + e^{\alpha(l - l_{50})}}$$

Recruitment

The process of adding new recruits directly in to a stock in a Gadget simulation is a straight forward one, with the new recruits simply added to the youngest age group of the stock, along with a length distribution that is generated by Gadget from parameters that are specified by the user in the input files.

The total number of recruits on any given timestep is specified (in units of 10,000 fish) in the input file (either as an absolute number or as a parameter that Gadget can estimate), along with the mean length and standard deviation of the length of the recruits. A Normal distribution is then used to calculate a length distribution for the recruits, and a length-weight relationship (as shown in equation 4.38) is used to calculate a mean weight for each of the length cells of the recruits, with the parameters α and β also specified in the input file. These new recruits are then added to the youngest age group of the stock.

$$W = \alpha L^{\beta} \tag{4.38}$$

Reproduction

An alternative approach that can be taken is to model a closed life-cycle system, where the number of recruits to a stock is governed by the spawning population. The total number of recruits is calculated (either based on a function of the fecundity, or the spawning stock biomass, of the 'parent' stock). The mean length and standard deviation of the length of the recruits is specified by the user, enabling a Normal distribution for the lengths of the recruits to be calculated, and these new recruits can then be added to the simulation in a similar manner to that described above.

The total number of recruits is given by a recruitment function that is calculated as part of the spawning process, and these recruits are then added to the youngest age group of the spawned stock at the start of the following timestep. Currently there are four recruitment functions defined in Gadget, although this can easily be extended in the future, and the valid recruitment functions are:

- a **fecundity** recruitment function, that calculates the number of recruits to be added to the spawned stock as a function of the length, age, number and weight of the spawning stock, as shown in equation 4.39 below.

$$R = p_0 \sum_{ages} \sum_{lengths} l^{p_1} a^{p_2} N_{al}^{p_3} W_{al}^{p_4} \tag{4.39}$$

- a simple **spawning stock biomass** recruitment function, that calculates the number of recruits to be added to the spawned stock as a simple proportion of the spawning stock biomass of the spawning stock, as shown in equation 4.40 and equation 4.41 below.

$$S = \sum_{ages} \sum_{lengths} N_{al} W_{al} \tag{4.40}$$

$$R = \mu S \tag{4.41}$$

- a **Ricker** recruitment function, that calculates the number of recruits to be added to the spawned stock as a function of the spawning stock biomass (see equation 4.40), based on the Ricker recruitment relationship shown in equation 4.42 below.

$$R = \mu S e^{-\lambda S} \tag{4.42}$$

- a **Beverton Holt** recruitment function, that calculates the number of recruits to be added to the spawned stock as a function of the spawning stock biomass (see equation 4.40), based on the Beverton Holt recruitment relationship shown in equation 4.43 below.

$$R = \frac{\mu S}{\lambda + S} \tag{4.43}$$

The total number of recruits calculated from the recruitment function is added to the youngest age group of the spawned stock at the start of the following timestep. The lengths of the newly spawned stock are calculated from a Normal distribution about a mean length, with a standard deviation, specified in the input files by the user. The mean weight of the fish in these age length cells is calculated from the length-weight relationship (as shown in equation 4.38), with the parameters α and β specified in the input file.

4.4 On the use of tagging data in statistical multispecies multi-area models of marine populations

Sigurdur Hannesson[1], Audbjorg Jakobsdottir[1], James Begley[1],
Lorna Taylor[1], and Gunnar Stefansson[1,2]

[1]**Marine Research Institute, Reykjavik, Iceland**
[2]**University of Iceland, Reykjavik, Iceland**

Abstract: With the use of multispecies models of marine stocks and recognition of spatial differences in spatial overlap, migration rates play an increasingly important role. Further, traditional estimates of growth based on samples from fishing gear are confounded with the selection pattern in an obvious way, exacerbated when multispecies issues are considered. For these and other reasons there is a need to explicitly include tagging data as a component of fisheries models. A statistical multispecies multi-area framework has been extended to predict tag returns and subsequently incorporate tagging data in likelihood components, to be used when estimating migration rates. The information content of such data is not clear *a priori* but simulation experiments indicate that bootstrapped uncertainty estimates may be reasonable indicators.

4.4.1 Introduction

Recent multispecies models incorporate multiple areas in order to account for species overlap, an important concern when considering species interactions (Stefansson and Palsson 1998; Begley 2004). This implies an immediate need to account for migration in such models and these migrations are typically largely unknown or at least poorly known. The rates therefore need to be estimated and the errors associated with these estimates need to be ascertained.

This is all the more important in the light of recent developments in the management of fisheries, where marine protected areas have become steadily more acknowledged as a potential management tool. Such management procedures have been criticised for lack of rigour since they may not have been subject to the same level of testing as traditional management systems (Hilborn et al. 2004).

Although in some cases it is possible to use simulation of different migration scenarios to evaluate the effects of marine protected areas (e.g. Stefansson and Rosenberg 2004), a more common requirement may be to obtain adequate estimates for a given environment, since this is likely to allow some leeway in how e.g. MPAs are set up. This requires estimates of migration rates and the evaluation of corresponding harvest control rules will require knowledge of the corresponding error rates.

This paper describes several simulation experiments designed to investigate the effects of using tagging data to estimate migration rates. The experiments are intended to shed some light on the estimability of migration rates in the context of statistical multispecies multi-area models.

Most of this paper uses simulated quarterly data on two areas with a stock migrating between the areas while subject to harvest mortality, natural mortality and other processes.

Bootstrap methods are designed and used to ascertain the estimation errors associated with maximum likelihood estimation of migration rates in the nonlinear population dynamics models. Simulated populations are used to verify the estimation of error rates.

It is of considerable importance to know how tagging experiments should be conducted in order to minimise the error of estimated migration rates. A simulation experiment is conducted in order to shed some light on this issue in the section on experimental design on page 100.

Finally, the paper considers a case study with a real data set.

4.4.2 General framework

General multispecies population dynamics

The following describes in fairly generic terms the population dynamics which can be incorporated in current spatially explicit multispecies models of marine animal populations. All of these processes are implemented in a computer program, Gadget, designed for generic multispecies modelling. For details of Gadget, see Begley and Howell (2004) or the Gadget User Guide (Begley 2004). If considered important in the parent population, these processes also need to be taken into account when implementing tagged sub-populations.

Denote by N_{almsrt} the number (N) of animals of species s, age a, length l and maturity stage m, alive in region r and timestep t. This group of individuals is a subset of the entire collection of animals of a given species. Such a collection will be referred to as a substock. Mathematically, this is just a collection of numbers, but they are usually unknown and will need to be estimated using statistical techniques.

Normally there are only two maturity stages within the model, immature and mature animals ($m = 0, 1$), though in the case of closed life-cycles, there is a need to include the egg and larval stage as a separate population. In many cases only a fixed area, time step and maturity stage are considered and in this case such indices are omitted and the notation is simplified to N_{als}.

Several processes can affect such a group of animals and cause a change in numbers in the group.

- The **migration** a population undertakes on a given time step is described by matrices, $\mathbf{P} = (p_{r_1 r_2})_{r_2=1,\ldots,R}^{r_1=1,\ldots,R}$, containing the proportion $p_{r_1 r_2}$ of the population that moves from area r_2 to area r_1 and R is the number of areas. Hence, if $\mathbf{u} = (u_1, ..., u_R)$ are abundance numbers by area for an age-length-maturity-species group in a population, and $\mathbf{P}$ is such a matrix, then the area distribution after the migration has taken place, is $\mathbf{Pu}$. When the migration matrices vary by age, time, maturity stage etc, the notation $\mathbf{P}_{almst}$ may be needed.
- **Maturation** involves shifting fish from a population of immature fish to another population of mature fish. This is done using proportions (as in the migration process), which are designed to mimic the resulting proportion of mature fish in each age-length cell.
- **Consumption** of a prey by a predator is defined through a set of equations which describe the desired food supply of a predator and the suitability of each prey class. This results in the amount consumed by the predator as well as the number of prey which die due to predation. Consumption will not be considered in this paper.
- **Growth** in length within multispecies population dynamics models is most naturally implemented through growth in weight since the weight increase is more naturally linked to consumption. Alternatively growth may be according to a fixed growth schedule in either length or weight. In any case, since the model is based on the frequency in length groups, growth in length needs to correspond to transfers of fish between length groups in such a manner that the predicted mean growth is maintained for the length group. This is implemented through the use of a length-update probability distribution, where the updating distribution is a beta-binomial which has the mean length increment as one parameter and the other is a free parameter (β) and can be either estimated or set.
- **Natural mortality** in addition to predation/fishing is usually taken as a fixed number for each species, but it may of course vary with time, age or size.
- **Ageing**, i.e. increasing the age by one year occurs at the last time step of a year (except for the oldest age group which is usually a plus group).
- A (mature) population may **spawn** to generate offspring and lose biomass. This possibly results in spawning mortality.

Tagged sub-populations

A tagged sub-population can be envisaged as a subgroup, $N'_{almsrte}$ of the numbers in a population, N_{almsrt}, where e denotes a given tagging experiment. There may of course be many tagging experiments. At the time of tagging everything is known but the age of the fish and the maturity stage. Therefore, the fish is distributed to age groups by lengths depending on the age distribution of the population itself at the time T of tagging: $N'_{almsrTe} = N'_{.lmsrTe} \cdot N_{almsrT}/N_{.lmsrT}$. If needed, the maturity stage can be determined by length and length distribution of the fish in the tagging experiment but it could depend on the stock.

In addition to all the processes described above, tag loss is taken into account. There is a parameter γ referred to as tag loss. At each time step $1 - e^{-\gamma}$, a proportion of the tagged subpopulation, loses its tags. This action acts on the tagged substock at the same time as natural mortality acts on the population.

Ignoring for the moment this additional process, the new subgroups have exactly the same dynamics as the original sub-population. Assuming that the number of tagged fish is low compared to the population totals, the tagged individuals do not affect the dynamics of the population as a whole. Therefore, the tagged dynamics can be viewed conditionally on the dynamics of the whole population, though from an implementational point of view, these components will normally be simulated simultaneously. It is clear that computational effort will increase as the number of tagging experiments increases.

For technical reasons it is useful to think not only of the number of tagged fish, but also of the proportion of tags, $\tilde{p}_e = N'_e/N$, in a given group (with a,l,m,s,r and t fixed) .

The tagged sub-populations need to be subjected to the exact same processes as described for the original populations. Mortality processes affect the numbers in a given cell through reduction alone, resulting in no modification of the proportion of tagged fish, so the values of $\tilde{p}_e$ remain fixed throughout these processes and can be used subsequent to the process to recompute the tagged numbers. On the other hand the numbers need to be used when migration is implemented as in this case the proportions can change drastically and are not easy to track until after the migration process has been completed. Thus judicious use alternately of numbers and proportions can simplify implementation considerably, particularly with regard to computational time.

Observation model

Since the same population dynamics process applies to the tagged sub-population, the predicted number of tag returns is just the predicted tagged catch, $\hat{C}_{almsrfte}$. For further information about the catch see Begley's paper (2004). The corresponding data or observed count, $C_{almsrfte}$, should have this expected value if the model is correct. The new subscript, f, refers to fleet, as there may be many fleets in operation.

Likelihood components

The models used in this paper and implemented in Gadget have deterministic forward projection models as a basic internal simulation mechanism. Any such forward projection is based on assumptions, several of which may be formulated in terms of parameters which are potentially unknown but can be estimated.

A forward projection of the system results in a set of predicted values of all model components, including the catch in numbers of each length and age group at each time step. This also provides predicted mean lengths at age along with predictions of arbitrary aggregations of catches in weight or numbers. Any of these can be used to obtain likelihood components if corresponding data are available.

The estimation itself proceeds by minimising the sum of all negative log likelihood components, possibly after weighting each appropriately.

In addition to proper likelihood components it is common to add penalties when bounds are exceeded and a special penalty component which enters as the squared difference between predicted and observed landings, when the model does not provide enough catch in an area.

The mean length at age can be entered as sums of squares and data on length distributions can be entered as Gaussian log-likelihoods (sums of squares) or using either multinomial or multivariate Gaussian distributions.

All of Gadget's likelihood components are described in Begley's paper (2004), but for the purposes of the project described in this paper a component for using tag-recapture data has been added as follows.

The treatment of these recapture data is initially taken as Poisson counts as is proposed in the paper by Hilborn (1990). Since the Poisson parameter is now $\hat{C}_{almsrfte}$, the corresponding likelihood contribution becomes

$$\Pi_e e^{-\hat{C}_{almsrfte}} \frac{\hat{C}_{almsrfte}^{C_{almsrfte}}}{C_{almsrfte}!}.$$

This is the ideal situation, applicable when information about length, maturity, age and location of recaptures is available for each individual recaptured fish. In cases when one or more of these pieces of information are unavailable, aggregation will be needed in one form or another. Normally, the area, time period and tagging experiment are available, but one or more of the others may be missing. It is not clear how to combine data with varying levels of information, but it is certainly possible to only use the information which tends to always be available:

$$\Pi_e e^{-\hat{C}_{....rfte}} \frac{\hat{C}_{....rfte}^{C_{....rfte}}}{C_{....rfte}!},$$

where "." denotes aggregation over the corresponding index.

The emphasis in this approach is on predicting the observed counts. It follows that this is mainly designed to resolve issues of numbers, i.e. provide some added resolution to mortality rates and population numbers.

4.4.3 Simulation, data and estimation

The level of information available in tagging data is not clear, nor is the required number of tagging experiments or number of fish in each tagging experiment in order to obtain a given reduction in variance of migration parameters. Similarly it is not at all clear how the nonlinearity of the population models affects the estimates. A simulation approach has been developed for the purpose of verifying the assumption and estimation accuracy in multispecies, multi-area models. Then an example using real data was considered.

Biological parameters in simulated stock

The model is run for twenty years with four (quarterly) time steps each year. A single stock is simulated, but the stock is distributed across two areas, with an age range of 1-10 years and length range of 4.5–90.5 cm, with 1 cm intervals, giving 86 length groups.

The simulated stock grows according to a von Bertalanffy growth function, Gadget's `lengthvbsimple`, with parameter values $L_\infty = 95$, $\kappa = 0.08$. The weight at length relationship is a simple power relationship, $w = cl^b$ with $c = 1 \cdot 10^{-5}$ and $b = 3$. The maximum length group growth was 6 and in the beta-binomial, β was set at $\beta = 12$.

Recruitment enters the stock once a year and for simplicity the simulated recruitment enters as one million recruits in each area, each year.

The stock will be assumed to migrate every time step. In order to simulate variable spawning, returning and feeding behaviour, there will be three different migration matrices. These will assume the same migration pattern for the second and third time step each year.

These matrices are:

$$\mathbf{P_i} = \begin{pmatrix} 1 & p_i \\ 0 & 1-p_i \end{pmatrix}$$

with $i = 1, 4$, $p_1 = 0.4$ and $p_4 \equiv 0$; and

$$\mathbf{P_j} = \begin{pmatrix} p_j & 0 \\ 1-p_j & 1 \end{pmatrix}$$

where $j = 2, 3$ and $p_2 = p_3 = 0.4$.

Notably, two of the migration rates are always fixed to be the same in the simulations. This is of interest since it will be used to test how consistent the estimates are.

In the baseline cases the fishing fleet catch a fixed proportion of the stock each year, implemented as 6.25% of the stock each quarter, that is the fishing mortality was set to be constant, $F = 0.25$. Natural mortality was set to 0.2 for all ages.

In some real situations the catch information is disaggregated by season and area and this will provide some bounds on the migration parameters. In the present models, the catch from the stock was aggregated across areas so those likelihood components did not affect the estimation of migration rates. This is done in order to avoid confusion with which part of the model provides information on migration.

Simulated stocks and components

A simulation model has been developed in the R statistical package, simulating the population dynamics described in the previous section. This simulation model is an extension of Olafsdottir and Sigurdardottir (2004), with the addition of simulating recaptures from tagging experiments. Notably, simulated data is generated completely separately from the estimation procedure which is to be verified.

The simulation predicts the expected number of tag returns as the catches from the tagged stock, $\hat{C}_{almsrfte}$. When generating stochastic data, these are generated from a Poisson distribution with this parameter, and the generated number is rounded to the nearest integer.

These simulated data are then fed into the estimation procedure which give estimated migration rates (or other parameters if needed).

Repeating these simulations provides repeated sets of estimated migration rates. These can then be used to obtain the variability in the estimated migration rates (as well as bias, which appears negligible). Now, these values are obtained by simulating different “realities” and can be obtained to an arbitrary accuracy simply by repeating the simulations often enough. It should be noted that these then become “true” standard errors of migration rate estimates, i.e. these give $\sigma_{\hat{p}_i}$.

In real situations one attempts to estimate the errors using some numerical procedure (such as Hessian matrices or bootstrapping) which gives $\hat{\sigma}_{\hat{p}_i}$. The simulation approach can be used to verify whether the estimates are realistic at least within the bounds of the simulated world.

Real Data

A two area, two stock component model was considered using the biological assumptions detailed in Taylor et al. (2004) and Taylor and Stefansson (2004). There are immature and mature cod stock components and only the mature fish can migrate between the two areas. There are twelve time steps in each year. Many parameters were kept fixed, such as growth and maturity parameters. Migration rates and two sets of initial population parameters and recruitment parameters were estimated using tagging data, survey indices and age distribution data. This is because migration is linked to initial population and recruitment. This will change the biomass which affects the number of recaptures.

Immature cod is one to ten years old and 4–130 cm long with 1 cm intervals. It grows according to a von Bertalanffy growth function, Gadget's `lengthvbsimple` (Begley 2004). It can grow a maximum of ten centimetres per time step. There is recruitment in the first time step every year and it matures.

Mature cod is three to twelve years old and 20–140 cm long with 1 cm intervals. It grows according to a von Bertalanffy growth function, Gadget's `lengthvbsimple`, and it can grow a maximum of ten centimetres per time step. It migrates and the migration pattern is similar to that in the simulated model and will be described in details later.

There are two fleets, commercial and the ground fish survey.

There is a lot of likelihood components but the most important here are the recapture likelihood components. The other ones are:

- Bound likelihood component. It gives penalty if the bounds given for the estimated parameters are exceeded.
- Migration penalty likelihood component. Gives a penalty if there is a negative entry in migration matrix.
- Catch distribution likelihood components. Sums of squares of the difference of the proportion of the data and the model sample of caught fish by time, area and age to estimate the initial population.
- Understocking likelihood component. Sums of squares of overconsumption in the model are included to ensure that no more fish are removed from the stock than a certain maximum proportion (Begley 2004, Taylor and Stefansson 2004).
- Survey indices likelihood components (Taylor et al. 2004).

All those likelihood components are described by Begley (2004).

The two areas in the model are on the one hand Area 1, the south of Iceland, including the West-Fjords, and on the other hand Area 2, the north. See picture 4.1 where Area 1 consists of areas 101, 102, 107 and 108 and Area 2 consists of areas 103, 104 and 105.

The migration rates are expected to be very low because of the area definition and since there are monthly time steps. Therefore they have been scaled so that they are between 0 and 100, instead of being between 0 to 1 before.

Tagging Data

Tag-recapture data for the years 1991–1997 was extracted from a data warehouse (Kupca 2004; Kupca and Sandbeck 2003; Taylor 2003a; Taylor 2003b). All tagged fish were aggregated by month, year and area giving 13 tagging experiments. Only mature fish were considered but since there is no information on maturity stage at the time of tagging, fish of minimum length of 40 cm were taken to be mature. Those tagging experiments are summarised in table 4.1.

Tag loss was fixed to 0.02 for all the thirteen tagging experiments, while the parameters were estimated. There are more tagged fish that lose its tags than die at each time step since natural mortality is 0.2 over the year but tag loss is 0.02 for every month which equals 0.24 over the year. The recapture rate was about 21% so tag loss seems to be a bit high. Tagging mortality and non-reporting rate are not taken into

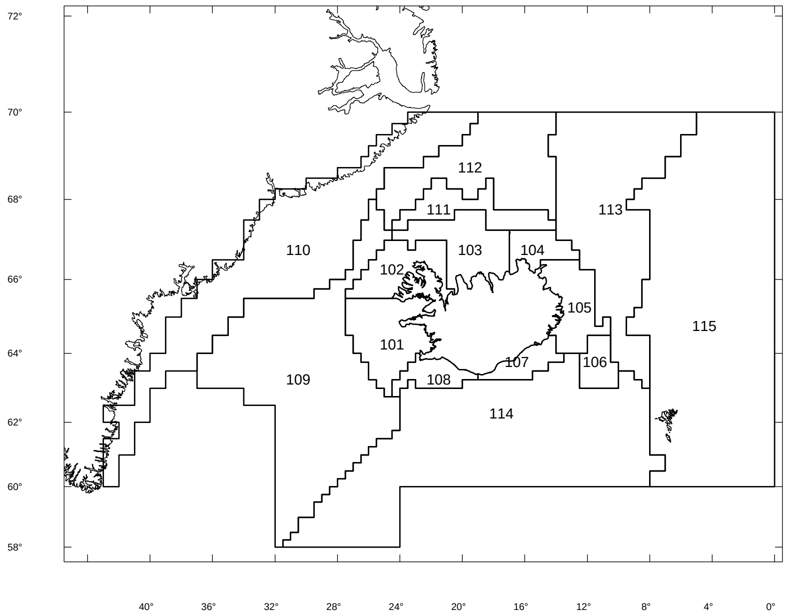


Figure 4.1: This is the data warehouse area definition. Area 1, south, consists of areas 101, 102, 107 and 108. Area 2, north, consists of areas 103, 104 and 105.

Tag id	Area	Tag year	Tag step	End year	N_t	N_r	r
tag91.4n	Area 2	1991	4	1997	521	181	0.347
tag91.5n	Area 2	1991	5	1994	97	22	0.227
tag92.4s	Area 1	1992	4	1998	218	23	0.106
tag92.4n	Area 2	1992	4	1995	1014	290	0.286
tag93.3s	Area 1	1993	3	1998	702	192	0.274
tag93.4s	Area 1	1993	4	2002	468	53	0.113
tag93.4n	Area 2	1993	4	2002	2818	1073	0.381
tag93.5s	Area 1	1993	5	1996	55	8	0.145
tag94.4s	Area 1	1994	4	2000	3486	426	0.122
tag95.4s	Area 1	1995	4	2001	4028	668	0.166
tag96.4s	Area 1	1996	4	2001	905	167	0.185
tag96.5s	Area 1	1996	5	1999	218	91	0.417
tag97.4s	Area 1	1997	4	2002	2263	461	0.204

Table 4.1: N_t is the number of tagged fish, N_r the number of recaptures and r is the proportion recaptured of the total number of tagged fish

account in Gadget so tag loss parameter represents all of those three. Therefore, its value was chosen to be 0.02, a little bit higher, on a yearly basis, than natural mortality.

The weights on the recapture data in the likelihood function were all equal to one.

Migration

The stock will be assumed to migrate every time step. In order to simulate variable spawning, returning and feeding behaviour, there will be three different migration matrices. These matrices are:

$$\mathbf{P_i} = \begin{pmatrix} 1 & p_i/100 \\ 0 & 1 - p_i/100 \end{pmatrix}$$

with $i = 1, 4$, p_1 to be estimated and $p_4 \equiv 0$; and

$$\mathbf{P_j} = \begin{pmatrix} p_j/100 & 0 \\ 1 - p_j/100 & 1 \end{pmatrix}$$

with $j = 2, 3$, $p_2 \equiv p_3$ to be estimated.

Now, in January to April (inclusive), migration matrix $\mathbf{P_1}$ is used, in May to August (inclusive) and December, migration matrix $\mathbf{P_2}$ is used and in September to November (inclusive) there are no migrations ($\mathbf{P_4}$ is used).

One can see from the matrices above that it is assumed that no fish will move to areas other than the two areas involved.

Model implementation

The above population dynamics models and estimation procedures have been implemented within Gadget. Estimation is in all cases implemented with an initial global search, the Simulated Annealing algorithm, followed by a local search, the Hooke and Jeeves algorithm, both implemented within Gadget and described by Begley and Howell (2004) and Begley (2004).

4.4.4 Bootstrap estimation of uncertainty

Bootstrap methods have been developed to shed some light on the issue of estimability and accuracy of parameter estimates.

Some care needs to be taken when implementing the bootstrap for tagging data. In particular it is very clear that the fish within a single tagged sample are not randomly sampled from the entire population of wild fish and hence the traditional bootstrap (Efron and Tibshirani 1993) needs to be modified. This is done by taking the sampling units for bootstrapping purposes to be individual tagging experiments.

For a scenario with a collection of N tagging experiments, bootstrapping is implemented by sampling N experiments (with replacement) from this collection, giving a bootstrapped sample. Given the bootstrapped sample, the parameters are estimated giving a bootstrap parameter estimate. This procedure is then repeated to obtain repeated bootstrap estimates of parameters, subsequently providing estimated standard errors.

It is not known *a priori* whether the bootstrap procedure will provide reasonable estimates of parameter values. As with any proposed method, the bootstrap needs to be evaluated and this will be done within the context of a simulated population where the variability of migration estimates is known.

4.4.5 Results

Model verification

Consider first a simple estimation exercise. The tagged fish are sampled by length from a normal distribution with a mean length of 50 cm and a standard deviation of 7 cm.

The estimated parameters are only the three migration rates, p_1, p_2 and p_3, where it is also of interest to use the knowledge $p_2 = p_3$ and verify whether this decreases the estimation variability.

Initial tests using deterministic recaptures with either large (50,000) or small (500) numbers of tags indicates complete replicability of results since in all cases this provides estimates of p_i to two correct decimal places. It was also verified that the choice of weights on likelihood components does not affect the estimation of migration rates (within this accuracy range), ascertaining that the area-aggregated likelihood data do not affect these rates.

50,000 fish tagged

When there are as many as 50,000 fish tagged, the estimated values of the parameters are quite close to the "true" values. There will always be some error in the recapture likelihood because the real recaptures, which are integers, are compared to the predicted values from Gadget, which are real (positive) numbers.

	p_1	p_2	p_3
True values	0.4	0.4	0.4
Initial values	0.50	0.40	-
Estimated values			
Assume $p_2 = p_3$, tagging and catch data in likelihood, low wt on tags	0.403	0.402	-
Assume $p_2 = p_3$, only tagging data in likelihood	0.403	0.402	-
Initial values	0.41	0.42	0.38
Estimated values			
All rates estimated freely	0.401	0.409	0.388

Table 4.2: Point estimates of migration rates based on 50 000 simulated tags and stochastic (Poisson) recaptures.

In no case is there an obvious problem or bias, though there is some deviation when estimating all three values as opposed to restricting the rates to be equal.

500 fish tagged

The point estimates naturally deviate more from the true values. The difference in the estimates of p_2 and p_3 when all three are estimated freely is now in the opposite direction from the one before. Again there is no obvious indication of bias. On the other hand, the runs with stochastic recaptures poorly estimated the migration rates in comparison with the "true" values from Splus (the simulated values).

Potentially the optimisation might be sensitive to the initial values used, but repeated optimisation again with other initial values of the migration parameters gave the same estimates as before.

	p_1	p_2	p_3
True values	0.4	0.4	0.4
Initial values	0.50	0.40	-
Estimated values			
Assume $p_2 = p_3$, tagging and catch data in likelihood, low wt on tags	0.525	0.428	-
Assume $p_2 = p_3$, only tagging data in likelihood	0.521	0.427	-
Initial values	0.41	0.42	0.38
Estimated values			
All rates estimated freely	0.530	0.349	0.546

Table 4.3: Point estimates of migration rates based on 500 simulated tags and stochastic (Poisson) recaptures.

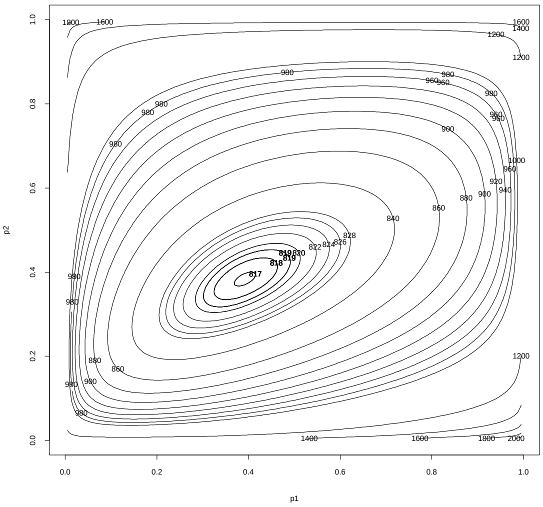


Figure 4.2: Contour lines of the likelihood score as a function of migration rates p_1 and p_2

The likelihood score for the migration rates in the domain $]0,1[\times]0,1[$ was computed using Gadget. A plot of the likelihood surface for the migration rates p_1 and p_2 can be seen in figure 4.2. It is seen in this figure that, as a function of these two parameters, the negative log-likelihood surface is quite well behaved with a single minimum.

First simulations

In all of what follows, recaptures are stochastic.

Consider two models for estimating migration rates: in the first model each tagging experiment is considered an independent experiment and the number tagged are treated as a single sub-population throughout each simulation. With this approach the number of sub-populations and computational requirements increase drastically as the number of experiments increases.

It is therefore of considerable interest to test a second model, where all tagged animals are considered a part of the same tagged population and during a simulation period, tagged animals are simply added to the existing tagged population.

The following subsections consider two tagging scenarios. Firstly a scenario with forty tagging experiments, where tagging occurs in all time steps on both areas for five years. The second scenario considers eight tagging experiments, ie. tagged in every time step on both areas but the year of tagging is randomly chosen of six years for each of the eight tagging experiments.

All steps – all areas – all years

A tagging experiment is simulated in each of 2 areas every quarter in each of 5 years, yielding 40 tagging experiments.

Each tagging as a separate experiment

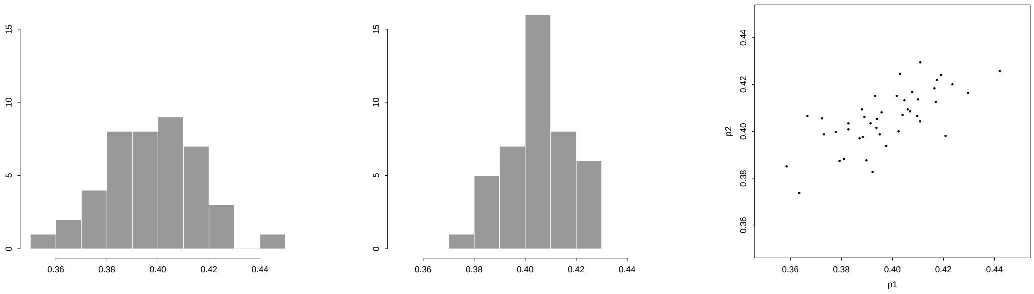
The number of tagged fish in each experiment is normally distributed (and rounded to the nearest integer) with a mean of 280 fish and standard deviation of 30. For such a generated data set, the migration rates are estimated. The procedure is repeated 43 times in order to verify the variability in the estimated migration rates.

The results are below:

	true value	mean	std. dev.
p_1	0.4	0.398	0.018
p_2	0.4	0.406	0.012

and the correlation between p_1 and p_2 was 0.698.

Figure 4.3 shows the distribution of migration rates p_1 and p_2, and a scatter plot of migration rate p_1 vs migration rate p_2 showing the correlation between those two parameters.



(a) Distribution of p_1 (b) Distribution of p_2 (c) Scatter plot p_1 vs p_2

Figure 4.3: Estimation of migration rates when each tagging experiment is considered separately

An aggregated experiment

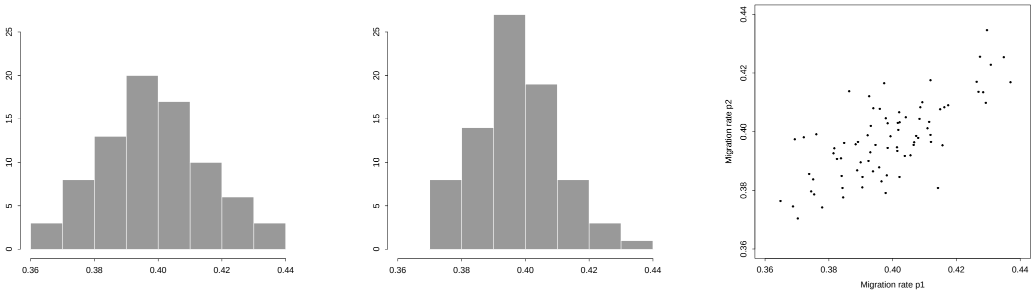
Consider again the forty experiments conducted above. This time consider them as a single aggregated experiment, where subsequently tagged fish are added to the existing tagged stock during the simulation.

The advantage of this approach is that the simulation time is reduced by an order of magnitude.

The results are below:

	true value	mean	std. dev.
p_1	0.4	0.399	0.017
p_2	0.4	0.398	0.013

and the correlation between p_1 and p_2 was 0.719, whereas it was 0.698 earlier.



(a) Distribution of p_1 (b) Distribution of p_2 (c) Scatter plot p_1 vs p_2

Figure 4.4: Estimation of migration rates when the tagging experiments are aggregated

Figure 4.4 shows the distribution of migration rates p_1 and p_2, and a scatter plot of migration rate p_1 vs migration rate p_2 showing the correlation between those two parameters.

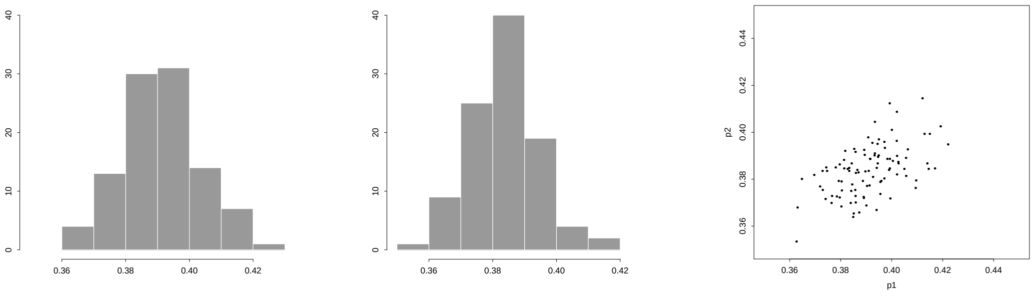
Bootstrap estimation of uncertainty

The same scenario as before is considered here. Bootstrapping was done 100 times to obtain 100 estimates of migration rates, providing estimated standard errors.

The results are below:

	base value	mean	std. dev.
p_1	0.392	0.391	0.012
p_2	0.383	0.384	0.011

and the correlation between p_1 and p_2 was 0.518. The "base value" in this case is the migration rate as estimated from the initial simulated data from which the bootstrap samples are simulated. It should be noted that the bootstrap mean is similar to the base value, indicating that there is not a great bias. Also notable is the fact that the standard errors here are about 0.011, which is slightly lower than the simulated values of 0.012–0.018, but certainly of the same order.



(a) Distribution of p_1 (b) Distribution of p_2 (c) Scatter plot p_1 vs p_2

Figure 4.5: Estimation of migration rates when bootstrapping the tagging experiments

Figure 4.5 shows the distribution of migration rates p_1 and p_2, and a scatter plot of migration rate p_1 vs migration rate p_2 showing the correlation between those two parameters.

All steps – all areas – years chosen randomly

Now, there are eight tagging experiments, two at each time step from the year, one from each area. The year at tagging doesn't matter and is randomly chosen of six years for each of the eight tagging experiments. Tagging data was simulated, 100 data sets, using Splus, as before. Then Gadget was run with optimisation to estimate the two migration rates.

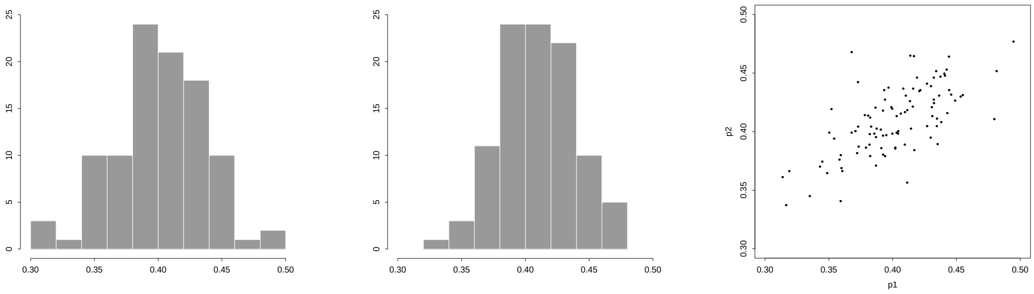
Each tagging as a separate experiment

The results are below:

	true value	mean	std. dev.
p_1	0.4	0.402	0.0353
p_2	0.4	0.410	0.0294

and the correlation between p_1 and p_2 was 0.671.

Figure 4.6 shows the distribution of migration rates p_1 and p_2, and a scatter plot of migration rate p_1 vs migration rate p_2 showing the correlation between those two parameters.



(a) Distribution of p_1 (b) Distribution of p_2 (c) Scatter plot p_1 vs p_2

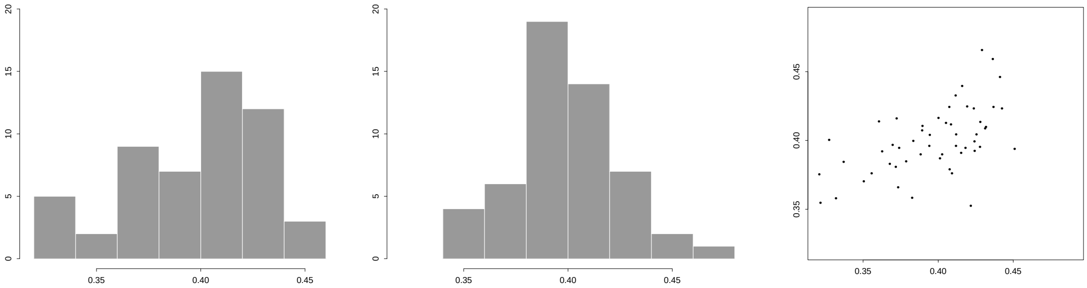
Figure 4.6: Estimation of migration rates when each tagging experiment is considered separately, using tagging data from a single year selected at random

An aggregated experiment

The results are below:

	true value	mean	std. dev.
p_1	0.4	0.397	0.0332
p_2	0.4	0.400	0.0244

and the correlation between p_1 and p_2 was 0.554. Note that standard deviation of the migration rates is lower in the multiple experiment than it was in the single experiment



(a) Distribution of p_1 (b) Distribution of p_2 (c) Scatter plot p_1 vs p_2

Figure 4.7: Estimation of migration rates when the tagging experiments are aggregated, using tagging data from a single year selected at random

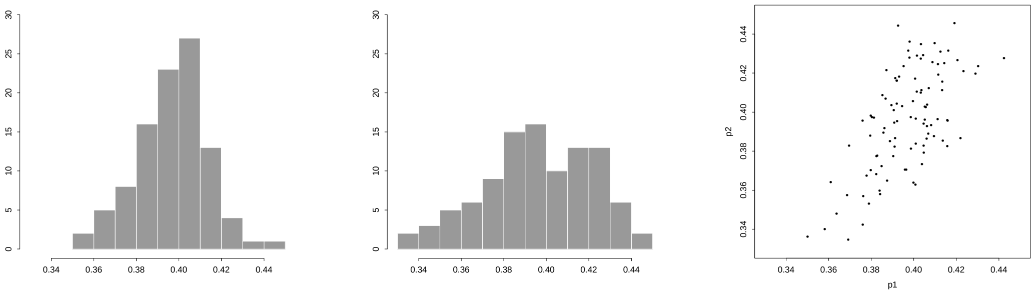
Figure 4.7 shows the distribution of migration rates p_1 and p_2, and a scatter plot of migration rate p_1 vs migration rate p_2 showing the correlation between those two parameters.

Bootstrap estimation of uncertainty

The same scenario with eight tagging experiments is considered again. Bootstrapping was done 100 times. The results are below:

	base value	mean	std. dev.
p_1	0.400	0.397	0.0162
p_2	0.399	0.396	0.0257

and the correlation between p_1 and p_2 was 0.626, slightly lower than before.



(a) Distribution of p_1 (b) Distribution of p_2 (c) Scatter plot p_1 vs p_2

Figure 4.8: Estimation of migration rates when bootstrapping the tagging experiments, using tagging data from a single year selected at random

Figure 4.8 shows the distribution of migration rates p_1 and p_2, and a scatter plot of migration rate p_1 vs migration rate p_2 showing the correlation between those two parameters.

Changing Fleet and Fishing Mortality

In all the above work, Gadget's linearfleet (catching some proportion of the population) has been used with fishing mortality, F, constant for all (both) areas, all timesteps and all years. The question now arises whether the above results will hold with variable fishing mortality and what happens when other fishery models are used (Gadget's totalfleet which uses biomass caught). In the following model one pseudo stock and one fleet are used, as before. There are 40 tagging experiments, as before. The number of tagged fish in each experiment is normally distributed with a mean of 280 fish and standard deviation of 30 fish. Mean length in each experiment is 50 cm with standard deviation of 7 cm.

Linearfleet – variable F

Before, $F =$ 0.25 for all time steps. Now, F was set to 0.25, 0.375, 0.125, 0.25 for those four time steps respectively. The results are similar to the results when using Gadget's linearfleet and fixed F.

Totalfleet

This model is almost the same as before, only using the biomass caught each time step, in both areas, instead of just F before. A simple von Bertalanffy growth function is used with weights dependent of length.

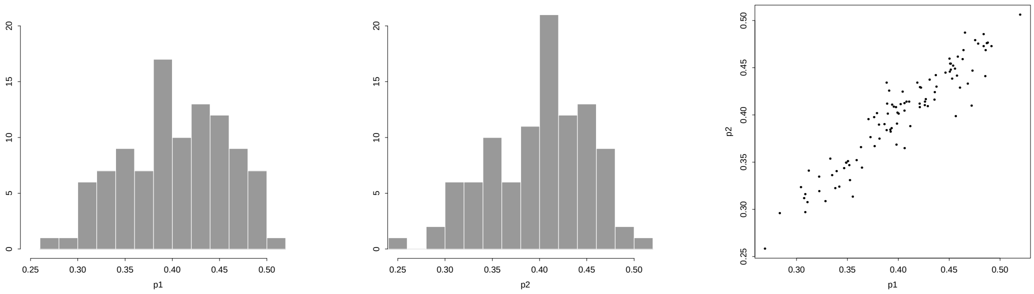
When simulating the pseudo stock and the recaptures, a certain proportion of the stock, in terms of number of fish, was caught and there were no information about how much the fish caught weighed. However, Gadget's totalfleet uses biomass caught instead of proportion as used before and this conflict causes some inaccuracy when estimating the migration rates.

Constant F

Using totalfleet and stochastic recaptures, simulating recapture data 100 times and estimating migration rates for every data set gives the following results:

	true value	mean	std. dev.
p_1	0.4	0.404	0.055
p_2	0.4	0.400	0.052

and the correlation between p_1 and p_2 was 0.945 which is very high. Note that the standard deviation is about three times higher than before, when using linearfleet



(a) Distribution of p_1

(b) Distribution of p_2

(c) Scatter plot p_1 vs p_2

Figure 4.9: Estimation of migration rates when bootstrapping the tagging experiments with constant F

Figure 4.9 shows the distribution of migration rates p_1 and p_2, and a scatter plot of migration rate p_1 vs migration rate p_2 showing the correlation between those two parameters.

Variable F

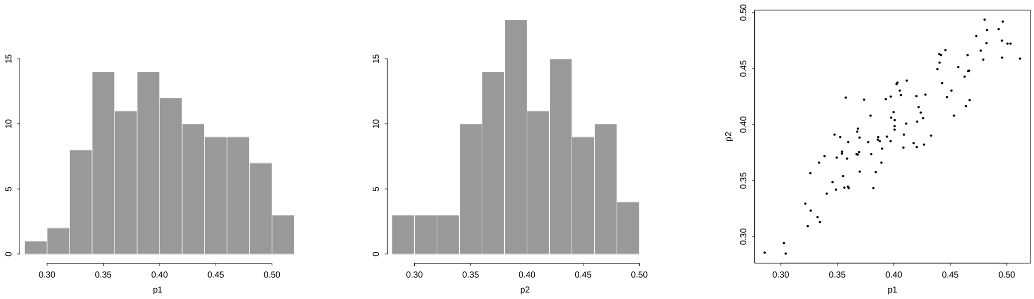
For variable F, one can use Gadget's concept of "totalfleet" along with stochastic recaptures. Simulate recapture data 100 times and estimate migration rates for every data set.

Before, $F = 0.25$ for all time steps. Now, F was set to 0.25, 0.375, 0.125, 0.25 for those four time steps resp. The results are below:

	true value	mean	std. dev.
p_1	0.4	0.403	0.0527
p_2	0.4	0.402	0.0484

and the correlation between p_1 and p_2 was 0.894 which is very high but a bit lower than before, when F was constant.

Figure 4.10 shows the distribution of migration rates p_1 and p_2, and a scatter plot of migration rate p_1 vs migration rate p_2 showing the correlation between those two parameters.



(a) Distribution of p_1 (b) Distribution of p_2 (c) Scatter plot p_1 vs p_2

Figure 4.10: Estimation of migration rates when bootstrapping the tagging experiments with variable F

Experimental Design

Scenarios have been set up to verify the effect of number of fish in each tagging experiment and number of tagging experiments to the estimated migration rates and especially the standard deviation of this estimation.

Scenarios

Each scenario (A – G) is done 100 times for each number of tagged fish in each tagging experiment. There are two sets of numbers of fish tagged in each experiment, $N_1 : 50, 100, 150, \ldots, 1000$ and $N_2 : 100, 200, 300, \ldots, 2000$.

In all scenarios, we have the same underlying model in Gadget, the only difference is in the tagging file and the recapture file. All scenarios take place over five years except scenarios C and F.

Scenario A: Five years, four steps each year, tagging experiment done in one area. Use N_1.

Scenario B: They are all done on the same year, at every step (four steps) and in both areas on every step. Use N_2.

Scenario C: They are done every other year at every time step and on both areas. Use N_1.

Scenario D: They are done in five years, at the third time step each year and on both areas every time. Use N_2.

Scenario E: They are done over five years, at every time step and on both areas. Use N_1.

Scenario F: Every time step on both areas. The year of tagging is randomly chosen of six years in the model. Use N_2.

Scenario G: They are done on five years, every other time step (first and third) on both areas. Use N_1.

These scenarios are summarised in table 4.4.

Scenario	Number of years	Tagging in quarters	Tagging in areas	Number of tagging experiments
A	5	1,2,3,4	1	20
B	1	1,2,3,4	1,2	8
C	5*	1,2,3,4	1,2	40
D	5	3	1,2	10
E	5	1,2,3,4	1,2	40
F	6**	1,2,3,4	1,2	8
G	5	1,3	1,2	20

Table 4.4: Scenario design summary. Notes *every other year, and **the year is randomly chosen from six years for each of the eight tagging experiments.

Statistical Analysis

In order to compare the entire sets of standard deviations obtained a simplifying model is needed. Since the standard deviation is expected to vary in inverse proportion to the square root of total number tagged, the following two linear models are defined:

$$\sigma_{i,Jk} = a + \frac{b}{\sqrt{n_{Jk}}} \tag{4.44}$$

$$\sigma_{i,Jk} = a_J + \frac{b}{\sqrt{n_{Jk}}} \tag{4.45}$$

where $i = 1, 2$ and J denotes the scenario, A – G (use some subset of the set of all scenarios), and k denotes measurement in scenario J. Note that the intercept, a, tells us something about standard deviation of the migration rates. So if (for fixed value of b) $a_L < a_J$ for some scenarios L, J then scenario L is “better” than scenario J. An F-test for linear models is used to see if there is any difference between those two models and if the second model is “better” then there is some scenario with the lowest intercept which is then assumed better than other scenarios. The standard deviations in the scenario are compared by total number of tagged fish but not by number in each experiment. Note that the means of the standard deviation from each of the numbers in N_1 or N_2 are used but not individual measurements. Also, the data used are the means of the standard deviation for each number of fish in each tagging experiment (the N_i’s), but not all of the data itself.

All Scenarios: When comparing the two models (with and without the factor of scenario) we find there is a significant difference between the σ’s from all those seven scenarios. Therefore it does matter how we set up the scenario.

Note that all of the above scenarios (A – G) are done on two areas except scenario A. We should also note some connection between scenarios. In the figures below, the σ’s from scenario A are labelled with ’A’ etc.

A and E: Scenario E is an extension of scenario A as they are both done on the same time steps but A is done in one area while E is done on both areas.

As we see in figure 4.11, σ_1 is higher in scenario A than in scenario E. However, σ_2 seems to be lower in scenario A than in scenario E. This was confirmed when comparing the two models, mentioned above, using data from scenarios A and E. This can be explained by scenario A being done in just one area but E in both areas.

B and E: Scenario E is an extension of scenario B as B is done on one year but E is done on five years.

As can be seen in figure 4.12, right two pictures, σ seems to be lower in scenario B than in scenario E.

When comparing the two models (with and without factor of scenario) we find there is a significant difference between σ from scenario B and scenario E with lower values in both σ_1 and σ_2 in scenario

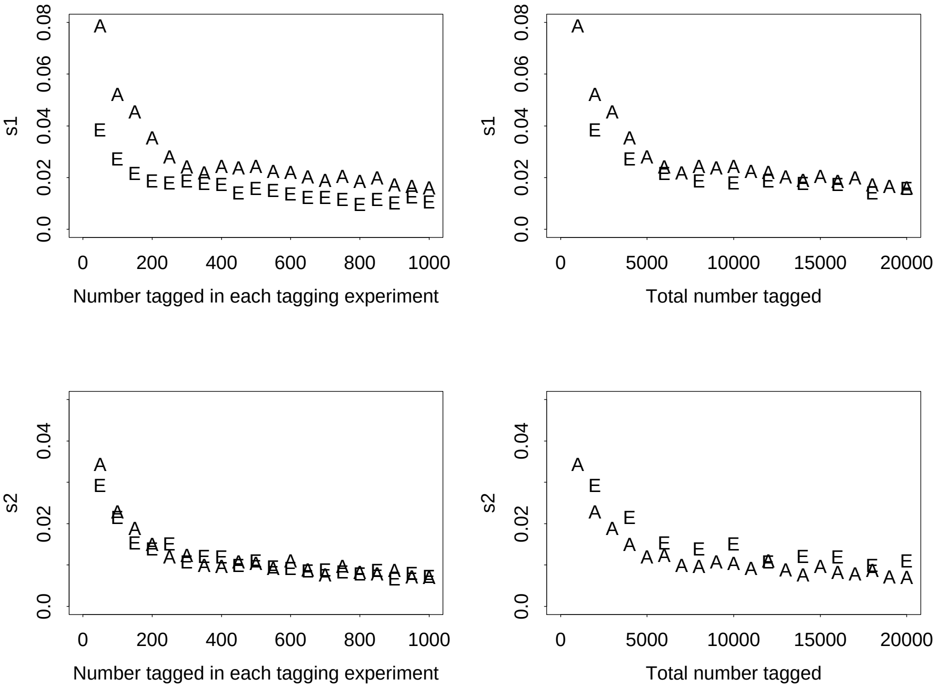


Figure 4.11: Two upper figures show σ_1 as the two lower figures show σ_2. The two figures to the left show σ_i as a function of number tagged in each tagging experiment as the right two figures show σ_i as a function of total number tagged.

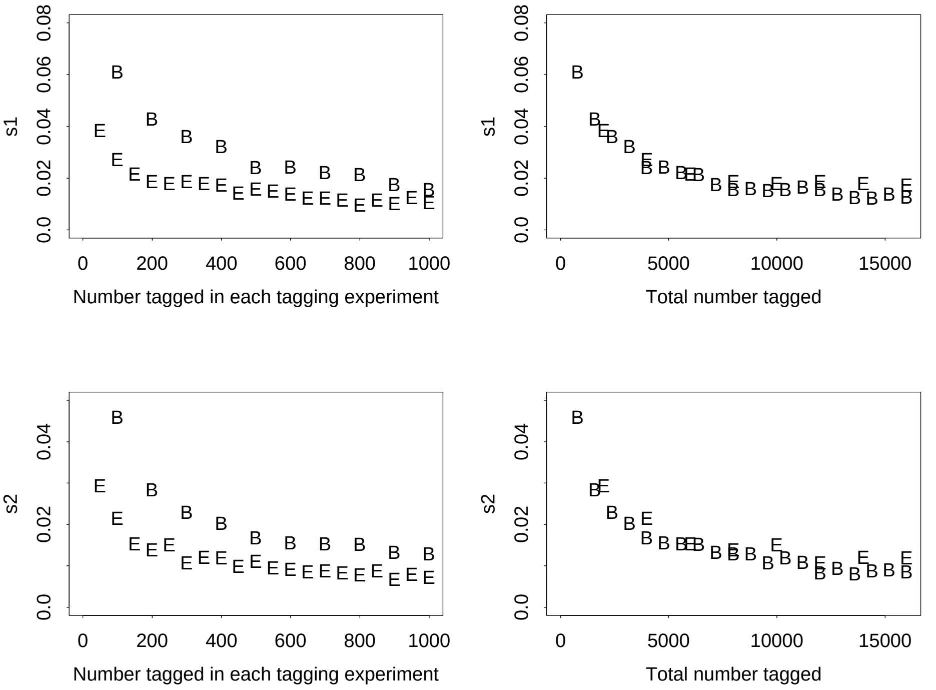


Figure 4.12: Two upper figures show σ_1 as the two lower figures show σ_2. The two figures to the left show σ_i as a function of number tagged in each tagging experiment as the right two figures show σ_i as a function of total number tagged.

B. Therefore one can conclude that scenario B is better than scenario E.

However when only looking at number of fish tagged in each tagging experiment in the two pictures on the left, scenario E is better than scenario B.

B and F: Scenarios B and F are similar as they are both done in both areas in every time step (1–4) but in B all the tagging experiments are done on the same (fixed) year but the year at tagging in F is chosen randomly (for every one of the eight tagging experiments).

When comparing the two models (with and without factor of scenario) there is no significant difference between the σ_i's from scenarios B and F.

D, E and G: Scenario G is an extension of scenario D as D is done in one time step each year as G is done twice (every other time step). Scenario E is an extension of scenario G as G is done on every other time step in the year as E is done in every time step of the year.

As can be seen in figure 4.13, right two pictures, σ seems to be lower in scenario D than in both scenarios E and G.

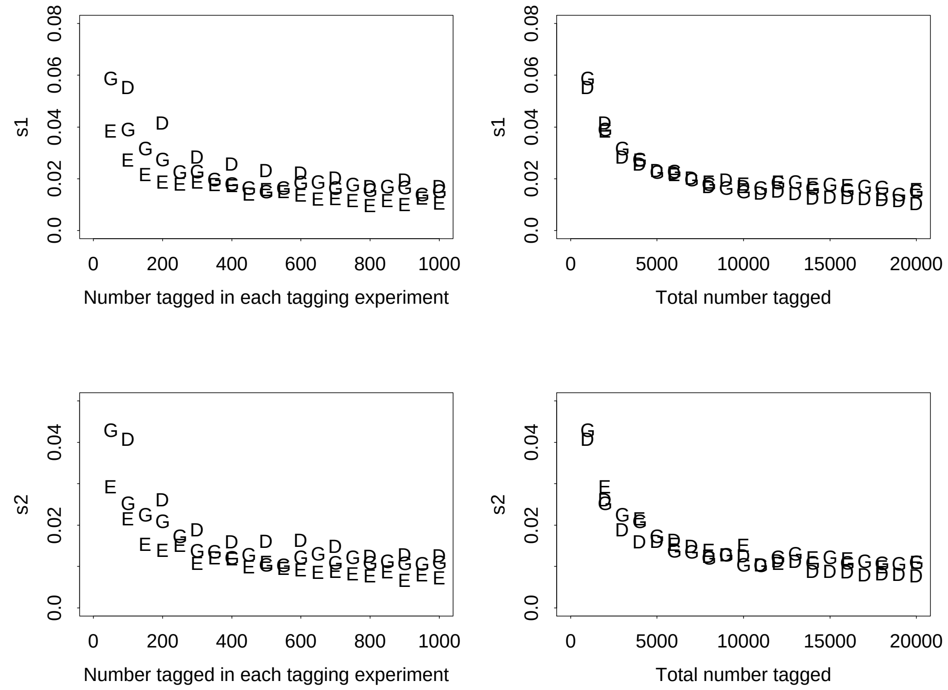


Figure 4.13: Two upper figures show σ_1 as the two lower figures show σ_2. The two figures to the left show σ_i as a function of number tagged in each tagging experiment as the right two figures show σ_i as a function of total number tagged.

When comparing the two models (with and without factor of scenario) we get that there is a significant difference between σ from those three scenarios with lowest values in both σ_1 and σ_2 in scenario D. When comparing D and E and then D and G, scenario D showed the lowest intercept and one is led to conclude that scenario D is better than scenarios E and G.

C and E: Scenarios E and C are the same except C takes place every other year (total of five years as mentioned above) and E in every year.

When comparing the two models (with and without factor of scenario) there is no significant difference between the σ_i's from scenarios C and E.

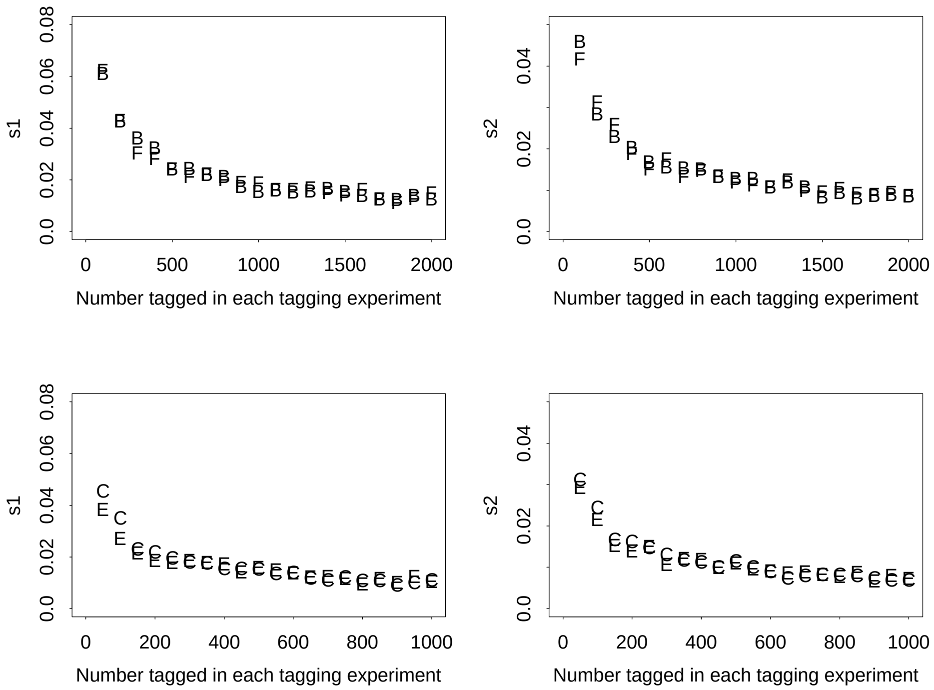


Figure 4.14: Upper two pictures show comparison between B and F and the lower two show comparison between C and E. Left two pictures show σ_1 and the right two show σ_2

Further results: Now the scenarios were compared two by two along with some other combinations. The results are shown in table 4.5.

The first line, with all scenarios, tells us that there is a difference (w.r.t. standard deviation of estimation of migration rates) between the scenarios. Scenarios B, D and F are similar and so are scenarios C, E and G. However there is some difference if we add scenario A to the model.

Conclusions

If the idea is to tag a fixed number of fish in each tagging experiment, then there should be many tagging experiments to minimise the variance of the migration rates. However if the idea is to tag a fixed number of fish in total, then there should be fewer tagging experiments (but in both areas – see scenarios A and E) to minimise the variance of the migration rates.

Exp.	Pr(F) - σ_1	Pr(F) - σ_2	lowest intercept		Significance	
			σ_1	σ_2	σ_1	σ_2
all	0	0	F	A	1	1
DEG	9.9e-05	5.4e-05	D	D	1	1
BDF	0.32	0.38	B	B	0	0
CEG	0.084	0.010	G	G	0	1
ACEG	6.4e-08	2.4e-14	G	A	1	1
A, B	1.3e-10	0.0017	B	A	1	1
A, C	0.0051	1.2e-10	C	A	1	1
A, D	4.4e-10	0.00013	D	A	1	1
A, E	0.0014	4.1e-11	E	A	1	1
A, F	3.1e-10	0.0012	F	A	1	1
A, G	4.5e-06	4.024e-07	G	A	1	1
B, C	4.5e-10	2.7e-11	B	B	1	1
B, D	0.32	0.22	B	B	0	0
B, E	3.4e-06	3.1e-09	B	B	1	1
B, F	0.52	0.93	F	F	0	0
B, G	0.00048	0.00070	B	B	1	1
C, D	6.3e-10	2.8e-08	D	D	1	1
C, E	0.20	0.28	E	E	0	0
C, F	5.6e-09	3.4e-09	F	F	1	1
C, G	0.017	0.0076	G	G	1	1
D, E	8.2e-06	1.1e-06	D	D	1	1
D, F	0.15	0.2635775	F	F	0	0
D, G	0.0020	0.012	D	D	1	1
E, F	4.3e-06	7.7e-08	F	F	1	1
E, G	0.58	0.090	G	G	0	0
F, G	0.00023	0.0011	F	F	1	1

Table 4.5: The scenarios compared are listed in the first column. Next two columns show probability of F when comparing models with and without factor (different intercepts or not), the former giving this probability for σ_1 and the latter one shows the same for σ_2. The fourth and fifth column show which scenario had the lowest intercept for σ_1 and σ_2 respectively and the last two columns tell whether there is a significant difference between scenarios for σ_1 and σ_2 respectively (if Pr(F) is less than 0.05 then there is considered to be a significant difference and then “1” is written in the appropriate column).

Real Data

The model described in section 4.4.3 was run in Gadget and parameters were estimated. Then bootstrap methods were conducted in order to get confidence interval for the migration rates.

Results

The model predicted fewer recaptured fish than were actually recaptured. However, the proportion of recaptured fish by area were similar.

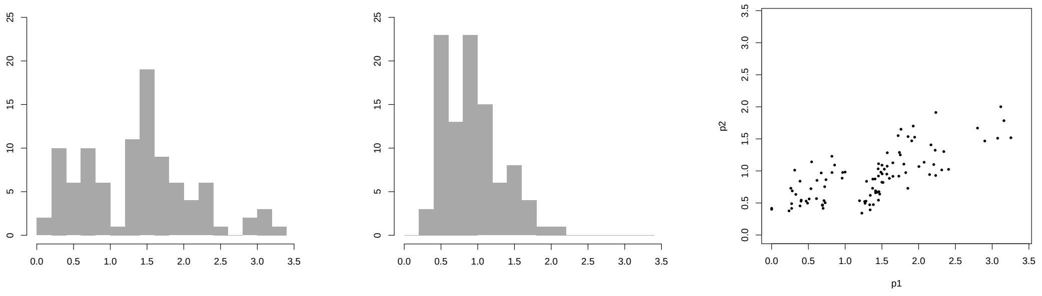
The migration rates were estimated to be $p_1 = 1.559$ and $p_2 = 0.997$.

Bootstrapping

Bootstrapping was done and parameters were estimated for 97 samples. Those samples were randomly chosen with the restriction that there had to be at least one tagging experiment from each area. The results are following (divide by 100 to get the actual migration rates and standard deviation):

	base value	mean	std. dev.	bias	rel. bias
p_1	1.559	1.356	0.748	-0.202	-0.130
p_2	0.997	0.916	0.387	-0.080	0.081

and the correlation between p_1 and p_2 was 0.726. The estimates are biased as can be seen in the above table. Confidence intervals, 95%, based on the bootstrap data are (0.246,3.100) for migration rate p_1 and (0.396,1.750) for migration rate p_2. Boxplot of the migration rates from the bootstrap runs can be seen in figure 4.16. The means are close to the base value but the standard deviation is relatively high.



(a) Distribution of p_1 (b) Distribution of p_2 (c) Scatter plot p_1 vs p_2

Figure 4.15: Estimation of migration rates when bootstrapping tagging experiments

Figure 4.15 shows the distribution of migration rates p_1 and p_2, and a scatter plot of migration rate p_1 vs migration rate p_2 showing the correlation between those two parameters.

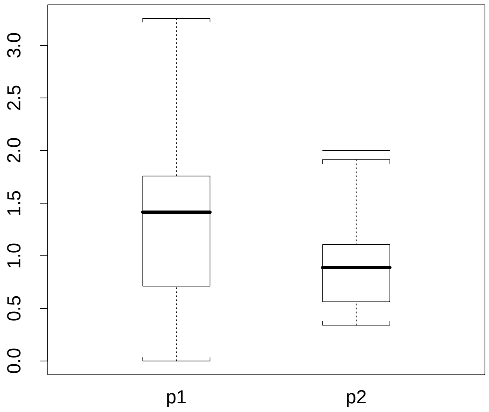


Figure 4.16: Boxplot of the estimated migration rates from bootstrapping

Note from the scatter plot and the histogram of migration rate p_1 that the points lie in three intervals. The first is where p_1 is less than 1.1, the second one is where p_1 is between 1.1 and 2.6 and the third one is where p_1 is greater than 2.6. The interesting thing is that the tagging experiment tag93.4n never occurs in the models corresponding to the points in the first interval, one to three times in the models corresponding to the points in the second interval and two to five times in the models corresponding to the points in the third interval. Note that this tagging experiment is very big, with 2818 fish tagged, and the return rate is high, 38.1% with over 1000 recaptured fish it represents almost a third of the total number of recaptured fish.

4.4.6 Discussion

Adding tagging data to existing models of (multi-species) multi-area models allows the estimation of migration rates, at least in principle. Results in this paper indicate that these migration rates are indeed estimable, at least for few areas and when decent estimates of some other model parameters have been obtained from other sources. The estimators seem to be biased when using real data set.

These initial simulations indicate that repeated tagging experiments may, at least in some cases, be treated as a single experiment, where new tags are added to tagged populations already in the model. This greatly reduces the computations required. It is not known how variable survival rates or other complications might affect these results. It matters much how the likelihood components are weighted. It was shown that a tagging experiment with a high number of tagged fish and recaptured changes the estimate of the migration rates. The weights should depend on number of comparisons or number of fish recaptured.

In addition to estimating migration rates, a no less important issue is obtaining the errors of these estimates. This is nontrivial since the models are highly nonlinear and several earlier simulations indicate that simple-minded approaches to variance estimation in fisheries models provide highly misleading results, even in single-stock, single-area models (Gavaris et al. 2000; Restrepo et al. 2000; Patterson et al. 2001).

Bootstrap-based methods based on resampling entire tagging experiments appear to adequately capture the uncertainty in estimation of migration parameters. Although this result may be conditional on the models used, the results obtained appear quite promising.

Gadget is simple when it comes to reducing the number of tagged fish. Only tag loss in time is considered here but there is also tagging mortality which should reduce the initial number of tagged fish. Then there is non-reporting rate which appears in the end where the predicted recaptures are reduced by some ratio. Those two cannot be distinguished from each other and should appear as one parameter to be estimated (Hilborn 1990). The tag loss rate can be estimated by double tagging some sample of the fish and see how many of them are returned with only one tag.

In addition to this layout, it must be noted that there is information in the recapture data about the growth of each recaptured fish. This information is only used implicitly in the above and not at all when the likelihood components are aggregated over length- and age-groups. Gadget is a Markovian model, neglecting length of fish before the previous time step. It is therefore not trivial how information on the history of the individual fish should be brought forward into a new likelihood component, but this is certainly not insoluble and needs to be explored. The easiest way would be to put each length group as a specific tagging experiment but that would increase the computational time dramatically. It is, however, possible to compare the mean length of recaptured fish from the model and observed recaptures and this has in fact been implemented within Gadget.

Only one likelihood function was considered but there are more possibilities like multinomial likelihood (Hilborn 1990).

Only example where there are two areas have been considered in this paper. Adding more areas to a model increases the computing time much and there are 30 parameters for every area to be estimated, excluding the migration rates whose number would increase as well. Data is also not available on all areas as needed.

References

Begley, J. 2004. Gadget User Guide. Available from the Gagdet web page: `http://www.hafro.is/gadget`.

Begley, J., D. Howell. 2004. An overview of Gadget, the Globally applicable Area–Disaggregated General Ecosystem Toolbox. ICES CM 2004/FF:13.

Efron, B., R.J. Tibshirani. 1993. An introduction to the bootstrap. Chapman and Hall, New York. 436p.

Gavaris, S., K.R. Patterson, C.D. Darby, P. Lewy, B. Mesnil, A.E. Punt, R.M. Cook, L.T. Kell, C.M. O'Brien, V.R. Restrepo, D.W. Skagen, and G. Stefansson 2000. Comparison of Uncertainty Estimates in the Short Term Using Real Data. CM 2000/V:03

Hilborn, R. 1990. Determination of Fish Movement Patterns from Tag Recoveries using Maximum Likelihood Estimators. Can. J. Fish. Aquat. Sci. Volume 47, 635–643.

Hilborn R., K. Stokes, J.-J. Maguire, T. Smith, L.W. Botsford, M. Mangel, J. Orensanz, A. Parma., J. Rice, J. Bell, K.L. Cochrane, S. Garcia, S.J. Hall, G.P. Kirkwood, K. Sainsbury, G. Stefansson, and C. Walters. 2004. When Can Marine Protected Areas Improve Fisheries Management? Ocean & Coastal Management. Accepted.

Kupca, V. 2004. A standardized database for fisheries data. ICES CM 2004/FF:15.

Kupca, V., P. Sandbeck. 2003. dst^2 datawarehouse structure and data import. In *dst^2 : Development of structurally detailed statistically testable models of marine populations*, Volume 98 of *Technical Report*, pp. 37–46. Marine Research Institute, Reykjavik.

Olafsdottir, E.I. and A.J. Sigurdardottir. 2004. A multispecies population simulation model in R. Unpublished MS.

Patterson, K.R., R. Cook, C. Darby, S. Gavaris, L. Kell, P. Lewy, B. Mesnil, A.E. Punt, V.R. Restrepo, D.W. Skagen and G. Stefansson. 2001. Estimating uncertainty in fish stock assessment and forecasting. Fish and fisheries, 2, 125–157.

Restrepo, V.R., K.R. Patterson, C.D. Darby, S. Gavaris, L.T. Kell, P. Lewy, B. Mesnil, A.E. Punt, R.M. Cook, C.M. O'Brien, D.W. Skagen, and G. Stefansson. 2000. Do Different Methods Provide Accurate Probability Statements in the Short Term? ICES CM 2000/V:08

Stefansson, G. and O.K. Palsson. 1998. A framework for multispecies modelling of Boreal systems. Reviews in Fish Biology and Fisheries, 8:1, 101–104.

Stefansson, G. and A.A. Rosenberg. 2004. Combining control measures for managing fisheries under uncertainty: quotas, effort limitation and protected areas. J. Royal Society. Accepted.

Taylor, L.A. 2003a. Datawarehouse for Icelandic waters. In *dst^2 : Development of structurally detailed statistically testable models of marine populations*, Volume 98 of *Technical Report*, pp. 64–75. Marine Research Institute, Reykjavik.

Taylor, L.A. 2003b. Definition of areas in Icelandic waters. In *dst^2 : Development of structurally detailed statistically testable models of marine populations*, Volume 98 of *Technical Report*, pp. 222–230. Marine Research Institute, Reykjavik.

Taylor, L.A., J. Begley, V. Kupca, and G. Stefansson. 2004. A simple implementation of Gadget for cod in Icelandic waters. ICES CM 2004/FF:23.

Taylor, L.A., and G. Stefansson. 2004. Gadget models of cod-capelin-shrimp interactions in Icelandic waters. ICES CM 2004/FF:29.

4.5 An R program to simulate Gadget population dynamics

Ásta Jenný Sigurdardottir, Elínborg Ingunn Olafsdottir and Lorna Taylor
Marine Research Institute, Iceland

Abstract: In order to test Gadget, a simulation tool was programmed in R to replicate the main Gadget population processes. This tool can be used as a source of independent simulated data with which the more complex Gadget model can be tested. The program is called 'Example' and was developed to replicate Gadget2.0.03. In this paper 'Example' is described along with a brief user guide.

Introduction

Example is a simple R program which imitates a pseudo stock by using a few of the most used properties of Gadget. The program was designed, were possible, to make use of processes optional (as they are in Gadget). The program replicates Gadget2.0.03 simulating stocks called *example.imm* and *example.mat*.

The program

Each year is divided into timesteps of equal length. Two stocks are simulated, representing immature and mature stock components, which live on 2 areas. Each stock grows and is subject to natural mortality and there is recruitment into the immature stock. Optional processes are:

- 0 to 2 fleets, representing the commercial fleet and a survey.
- Migration.
- Movement between the stocks, representing maturation.
- Consumption such that the mature stock is a predator of the immature stock.

Fleets are modelled as predators, as in Gadget. The survey takes place in the first timestep of each each year and the commercial catch takes place in each timestep for every year. Maturation from stock A to stock B is modelled by moving the oldest agegroup of A into B with the age increasing if done on the last timestep of the year. This replicates the Gadget process *doesmove*.

The order of calculations is the same as in Gadget and is as follows:

1. Migration between areas
2. Consumption, including catch by the fleets
3. Natural mortality
4. Growth
5. Recruitment
6. Remove the stock, here immature, that will move
7. Increase the age
8. Replace the stock, here to the mature, that has moved and increase the age.

The Stock

The First Timestep

The simulation is started with an initial population, the age of which ranges from the minimum age of the immature stock to the maximum age of the mature stock. This population is calculated as follows:

Let n_a be the number of age a individuals in the first timestep, μ_a the mean length at age a and σ_a the standard deviation of length at age a. For the minimum age (representing recruitment) see page 113. For a given constant mortality Z_0 we get

$$n_a = n_{a-1}e^{-Z_0} \tag{4.46}$$

to calculate the number at age for all a.

The number in lengthgroup i at age a in timestep 1 can then be calculated from:

$$N_{i,a,1} = n_a \left(\Phi \left(\frac{l_{i+1} - \mu_a}{\sigma_a} \right) - \Phi \left(\frac{l_i - \mu_a}{\sigma_a} \right) \right) \tag{4.47}$$

where l_i and l_{i+1} are the endpoints of lengthgroup i, $N_{l,a,t}$ is the number at age a in lengthgroup l at timestep t and Φ is the probability function for Normal distribution.

NB: in Gadget (which is programmed in C++) the value of Φ is approximated whereas R uses integration. To ensure compatibility between the models, the initial population for Gadget is entered directly from the initial population file rather than calculated from a Normal distribution. While this is an option within Gadget, it is not the standard method.

Calculations

Migration

Migration takes place in every timestep and migration is assumed to be constant for all years.

Migration at timestep t is defined as a 2×2 transition matrix $P_t := [p_{ij}]$ where p_{ij} is the proportion moving from area j to area i. **NB:** for P to be a transition matrix $\sum_i p_{ij} = 1$, for a fixed j.

If $N1_t$ is a matrix containing the abundance in numbers in area 1 at timestep t before migration and $N2_t$ is the same number for area 2, the numbers after migration will be

$$\begin{aligned} N1_t &= p_{11} \cdot N1_t + p_{12} \cdot N2_t \\ N2_t &= p_{21} \cdot N1_t + p_{22} \cdot N2_t \end{aligned} \tag{4.48}$$

As there are only 2 areas and the sum of the transition matrix columns must be 1 it is only necessary to enter the values of the first line into the migration file in the R program. In the Gadget2.0.03 migration file, the proportion moving must be removed from the old area and added to the new area.

Consumption

The following variables are used in the consumption calculations:

l	Lengthgroup in prey
L	Lengthgroup in predator
A	Areasize
H	The density (biomass per area unit) of available food at which the predator can consume half maximum consumption
Δt	Length of timestep
$M_{pred}(L)$	Maximum consumption
$\psi_{pred}(L)$	Fraction of M_{pred} consumed
$N_{pred}(L)$	Number of predator $pred$ in lengthgroup L
$N_{prey}(l)$	Number of prey $prey$ in lengthgroup l
$W_{prey}(l)$	The mean weight of prey of length l
$S_{pred,prey}(L,l)$	Suitability of $prey$ at length l for $pred$ at length L
$C_{prey,pred}(L,L)$	Total weight predator of length L consumes of prey of length l

Predation

The formula for C is as follows:

$$\begin{aligned} C_{pred,prey}(L,l) &= N_{pred}(L)M_{pred}(L)\Psi_{pred}(L)\frac{F_{pred,prey}(L,l)}{\sum_{l,prey} F_{pred,prey}(L,l)} \\ &= N_{pred}(L)M_{pred}(L)\frac{\sum_{l,prey} F_{pred,prey}(L,l)}{\sum_{l,prey} F_{pred,prey}(L,l)+HA}\frac{F_{pred,prey}(L,l)}{\sum_{l,prey} F_{pred,prey}(L,l)} \\ &= N_{pred}(L)M_{pred}(L)\frac{F_{pred,prey}(L,l)}{\sum_{l,prey} F_{pred,prey}(L,l)+HA} \end{aligned} \tag{4.49}$$

where

$$\begin{aligned} F_{pred,prey}(L,l) &= S_{pred,prey}(L,l)N_{prey}(l)W_{prey}(l) \\ M_{pred}(L) &= m_0 e^{(m_1 T - m_2 T^3)} L_{pred}^{m_4} \Delta t \end{aligned}$$

The suitability function for predation used in the R model is:

$$S_{pred,prey}(L,l) = \frac{\delta}{1+e^{-\alpha-\beta l-\gamma L}} \tag{4.50}$$

With one predator, one prey and otherfood the equation becomes:

$$\begin{aligned} C_{L,l} &= N_L M_L \Psi_L \frac{F_{L,l}}{\sum_l F_{L,l}+OA} \\ &= N_L M_L \frac{F_{L,l}}{\sum_l F_{L,l}+OA+HA} \end{aligned} \tag{4.51}$$

where O is the density of otherfood.

Catch

Catch is implemented in R using the *Linearfleet* option in Gadget.

Let $C_{fleet,prey}(l, a, t)$ be the number of age a prey, in lengthgroup l caught at timestep t, then

$$C_{fleet,prey}(l, a, t) = F_{l,t} N_{prey}(l, a, t) \Delta t \tag{4.52}$$

with $F_{l,t} = S_l F_y$ where F_y is constant for each year,

$$S_l = \frac{1}{1 + e^{-\alpha - \beta l}} \tag{4.53}$$

is the suitability function and α and β are constants.

OverConsumption

For each prey an upper limit needs to be set on the total amount consumed by all predators so as not to obtain more consumption than available biomass. Consumption is limited to 95% (R_M) of the available biomass. This is implemented by scaling target consumption by all predators.

Let

$R_{prey}(l)$	Ratio consumed
R_M	Maximum Ratio Consumed

Then

$$R_{prey}(l) = \frac{\sum_{pred} \sum_L C_{pred,prey}(L, l)}{N_{prey}(l) W_{prey}(l)}. \tag{4.54}$$

If $R_{prey}(l) > R_M$ consumption is adjusted as follows

$$C_{pred,prey}(L, l) = R_M N_{prey}(l) W_{prey}(l) \frac{C_{pred,prey}(L, l)}{\sum_{pred} C_{pred,prey}(L, l)} \tag{4.55}$$

Natural Mortality

Natural mortality at age a is implemented in the usual manner:

$$N_{l,a+1,t+\Delta t} = N_{l,a,t} e^{-M_a}. \tag{4.56}$$

Growth

Growth is according to a von Bertalanffy equation (equation 4.57) with the $lengthvbsimple$ growth function from Gadget implemented.

$$\mu_a = L_\infty (1 - e^{-\kappa a}). \tag{4.57}$$

For a fish of age a and length l, mean length growth $DeltaL$ is then calculated as:

$$\Delta L = L_\infty(1 - \frac{l}{L_\infty})(1 - e^{-\kappa \Delta t}). \tag{4.58}$$

The length distribution is updated using the beta-binomial distribution, ie the probability of growing x lengthgroups, given maximum lengthgroupgrowth n, is

$$P[X = x] = \frac{\Gamma(n+1)}{\Gamma(n-x+1)\Gamma(x+1)} \frac{\Gamma(\alpha+\beta)}{\Gamma(n+\alpha+\beta)} \frac{\Gamma(n-x+\beta)}{\Gamma(\beta)} \frac{\Gamma(x+a)}{\Gamma(\alpha)} \tag{4.59}$$

with $\alpha = \frac{\beta \Delta L}{n - \Delta L}$ to preserve the mean lengthgrowth according to equation (4.58). **NB:** the expected value of ΔL should be taken into consideration when fixing n.

Let $G = [g_{ij}]$ be the length updating matrix where g_{ij} is the probability of growing from lengthgroup i to lengthgroup j obtained from equation (4.59).

$$N_{l,a+1,t+\Delta t} = \sum_{l' \leq l} g_{l'l} N_{l,a,t} \tag{4.60}$$

with $N_{l,a,t}$ as described for the initial population for $a >$ min a (cf 110).

Recruitment

The timestep (or timesteps) on which recruitment takes place is defined by the user.

Given n_t recruits, at timestep t, with mean length μ and standard deviation of length σ, the number of recruits in lengthgroup i is calculated by:

$$N_{i,1,\mathrm{t}} = n_t(\Phi(\frac{l_{i+1} - \mu}{\sigma}) - \Phi(\frac{l_i - \mu}{\sigma})) \tag{4.61}$$

As for the initial population, the number of recruits in each length groups is given in the recruit input file (cf page 110).

User Guide

All the scripts to run the simulation program, output files for Gadget, run Gadget and plot results are contained a simple directory structure:

$$Example \left\{ \begin{array}{l} Splus \left\{ \begin{array}{ll} GadgetInputfiles & \\ Plots & \left\{ \begin{array}{ll} Num & \{ \quad src \\ Length & \{ \quad src \\ Mean & \{ \quad src \end{array} \right. \end{array} \right. \\ \\ Gadget \left\{ \begin{array}{l} out \\ Read \end{array} \right. \end{array} \right.$$

The `R` Program

The Gadget simulation scripts are stored in the Splus directory. In addition, the program can write input files and likelihood files for Gadget (written to Gadgetinputfiles) and plot the results from the stock calculations (Plots).

R vs Splus

The default settings are for `R` but the simulation scripts can run in `Splus` with minor modifications to one file (files.splus). To edit files.splus to run under `Splus` the write.table commands for `R` need to be replaced with those for `Splus`. For example, if you want the program to run in `Splus` instead of `R`:

```
# Splus
#write.table(veidisurv,"./GadgetIn ... dimnames.write=F)
#write.table(veidicomm,"./GadgetIn ... dimnames.write=F,append=T)
# R
write.table(veidisurv,"./GadgetIn ... row.names=F,col.names=F,quote=F)
write.table(veidicomm,"./GadgetIn ... col.names=F,quote=F,append=T)
```

should be changed to

```
# Splus
write.table(veidisurv,"./GadgetIn ... dimnames.write=F)
write.table(veidicomm,"./GadgetIn ... dimnames.write=F,append=T)
# R
#write.table(veidisurv,"./GadgetIn ... row.names=F,col.names=F,quote=F)
#write.table(veidicomm,"./GadgetInp ... ,col.names=F,quote=F,append=T)
```

Stock Calculations and Gadget Inputfiles

To simulate the stock it is sufficient to source("stock.splus"). To output Gadget files, it is either necessary to source("makefiles.splus") or to set makeinputfiles=1 in var.splus, in which case makefiles.splus is run automatically from stock.splus.

The files in the Splus directory are:

stock.splus: The main program

var.splus: Contains most parameters needed for the program. The parameters describing the stocks and catch (eg number of recruits, maximum age, minimum length) are set in this file ad can be edited by users.

initial.splus: Initializes the rest of variables.

function.splus: Contains most functions used in the program. The function for migration is in migration.splus and the functions which output Gadget files are in files.splus.

migration.splus: Calculates migration. The parameters for the migration matrix are defined here.

files.splus: Contains functions which write Gadget input files and likelihood files. If Gadget is run with these input files (along with the appropriate stockfiles etc) the results should be the same as those from the simulation program.

makefiles.splus: Makes all the files that are available from files.splus. To define which files should be output, those not required should be commented out using a # at the start of the line.

Result Plots

To plot the results from Gadget against the results from the simulation, the perl program *runandmove.p* (in Example/Splus/Plots) should be run and then plot.splus sourced. The plots are then stored in the Num, Mean and Length directories.

The files required for plotting are:

plot.splus: Makes the plots.

statistics.splus: Calculates mean length and total number which are required in the plotting program.

Gadget

Gadget Inputfiles

The R program creates the following inputfiles:

$area$

$example.refw$ reference weight file

$$\left.\begin{array}{l} example.imm.init \\ example.mat.init \end{array}\right\} \text{initial population files for immature and mature stock}$$

$fleet.data$ total catch per year (in kilos)

$example.rec$ renewal (recruitment) data file

aggregation files:

$$\left.\begin{array}{l} len.agg \\ age.agg \end{array}\right\} \text{aggregation files}$$

and likelihood data files:

$example.surveyindices$ survey index likelihood data

$example.ldist.catch$ catch length distribution likelihood data

$example.adist.catch$ catch age distribution likelihood data

$example.alkeys.catch$ catch age length distribution likelihood data

The following files need to be written/edited by the user:

$main$

$printfile$

$time$

$$\left.\begin{array}{l} allage.agg \\ allarea.agg \\ alllen.agg \end{array}\right\} \text{aggregation files}$$

$$\left.\begin{array}{l} example.imm \\ example.mat \end{array}\right\} \text{stock files for immature and mature stocks}$$

if the stock migrates:

$example.migration$ migration data file

if the stocks are caught:

$fleet$ fleet file

if the stocks eat:

$otherfood$ otherfood file

$otherfood.data$ contains available otherfood (biomass per area unit)

These files are contained in the directory Example/Gadget.

Estimating Parameters

To test the optimisation routines in Gadget, it is possible to use the simulation program to simulate a stock and then extract data, adding error, to use as likelihood data. The results from the estimation can then be compared to the known values in var.splus. Section 6.1 contains some examples of this.

Two classes of likelihood data can be output, survey indices and catch distribution (length distributions, age distributions and age-length keys).

Likelihood data can be output as follows:

example.surveyindex is calculated by multiplying the number N in each area per timestep by an error term, i.e. if $N^e_{A,t}$ is the number in area A at timestep t then

$$N^e_{A,t} = e^{(\epsilon - \frac{\sigma^2}{2})} \sum_l \sum_a N_{l,a,t,A}$$

where $\epsilon \sim N(0, \sigma^2)$.

σ is defined in var.splus

example.ldist.catch is calculated by multiplying the catch in numbers in each lengthgroup by a proportion ρ and by an error term, i.e. if $N^e_{l,A,t,f}$ is the length distribution from the catch at length l, area A, timestep t and fleet f with error then,

$$N^{e,p}_{l,A,t,f} = \rho e^{\left(\varepsilon - \frac{\sigma^2}{2}\right)} \sum_a N_{l,a,t,A,f}$$

where $\varepsilon \sim N(0, \sigma^2), 0 \leq \rho \leq 1$ and $N^{e,p}_{l,A,t,f}$ is the number of length l and age a in area A caught by fleet f in timestep t.

Examples indicating the use of these likelihood data for parameter estimation are available in 2 directories:

estimate.recruits: An example estimating the number of recruits in each timestep by survey indices.

estimate.alpha.beta: An example estimating the catch suitability parameters α and β using length distribution likelihood data. (α and β are called salphacomm and sbetacomm in var.splus).

estimate.alpha.beta.recruits: An example including estimate.recruits and estimate.alpha.beta.

README files with instructions are available in all directories.

Chapter 5

Model verification

This chapter deals with modelling issues which are of general concern in fisheries but which become crucial when modelling multispecies interactions.

It will be seen in section 5.2 that a common distributional assumption used in fisheries, namely a multinomial distribution for length data does not hold. This is in fact seen without any complicated modelling and the basic tests used are therefore typical examples of goodness-of-fit tests that can be used to evaluate suggested likelihood functions based on the data alone. Section 5.3 takes this one step further by using simple models of a single data source, to set up distributional assumptions which can then be used as a likelihood component in more complex models. The point to note with these approaches is that the measuring stick, i.e. the likelihood function, can in many cases be defined and evaluated before complex models are tested. Once the data are being used within a complicated model it becomes difficult to disentangle distributional properties of the data and model verification.

When using multiple data sources two major issues come up. The first of these is how weights should be chosen for each data source or log-likelihood component and the second is how one can detect and resolve apparent conflicts between data sources. Section 5.4 demonstrates how weights can be attached to different data sources in a fairly objective manner. It turns out that the weights can be determined more-or-less independently of whether data inconsistencies appear. In order to resolve inconsistencies, the first step is to detect them. This is done in section 5.7 where statistical approaches to this problem are developed and explored. Once such inconsistencies have been detected, the next step is to attempt to resolve them and though this may be a nontrivial task, it is sometimes possible to add parameters to the model in such a manner as to make the model more flexible and thus better explain the various data sets simultaneously.

Given the large number of difficult issues which arise when using standard likelihood functions, it is of some importance to test a variety of such functions, potentially developing new ones as the need arises. Section 5.5 illustrates how a multivariate normal distribution can be used to alleviate a number of the distributional problems encountered in some fisheries data sets. It is seen that the multivariate normal distribution may provide a feasible alternative to the multinomial distribution.

5.1 Goodness of fit tests for Gadget likelihood functions

Gunnar Stefansson
Marine Research Institute, Skulagata 4, P. O. Box 1390
121 Reykjavik, Iceland,
gunnar@hafro.is

Abstract: Gadget is a program which can simulate the development of multiple fish stocks in several areas, while being harvested by an arbitrary number of fleets. The program can further estimate parameters by comparing diverse model outputs to several data sources which can be modelled using arbitrary likelihood functions.

The adequacy of the model as an explanation of the data can be evaluated using any number of goodness-of-fit tests. This paper describes some of these tests, new ones are developed and preliminary results are reported on the applications of these tests to real data.

Introduction

When several data sources are combined in an analysis one issue which emerges is the specification and verification of likelihood functions to be used for comparing the data to the model.

It must be noted that, when a specific test indicates that a specified distribution should be rejected, this may imply either a failure of the distributional assumption or a failure of the model providing parameters for the distribution. The rejection must therefore not be taken unilaterally to indicate a problem with the distribution, but rather an indication of a general problem. Subsequent analysis of the raw data and of the model behaviour is needed to specify exactly where the problem lies.

Multinomial distributions

The simplest method to evaluate whether data really come from a multinomial distribution is to compare the observed count in each cell to the predicted count based on the modelled proportions, computing

$$X^2 = \sum_{i=1}^{p} \frac{(O_i - E_i)^2}{E_i}$$

using obvious notation for observed counts, expected counts and the number of categories. If the model estimates of expected counts are correct and the data truly satisfy the multinomial assumption, then $X^2 \sim \chi^2_{p-1}$ and the hypothesis is usually rejected if X^2 is too large, i.e. if $X^2 > \chi^2_{p-1,1-\alpha/2}$.

In this connection it must be remembered that when using Gadget, usually a huge number of observations are available so that the expected counts (which are of course estimated) are normally considered fixed.

Combining several tests for multinomial distribution

Several different multinomial χ^2-tests can be combined fairly easily if the same number of categories is used for each test. Thus, if results from independent comparisons should all satisfy $X_i^2 \sim \chi^2_p,\ i = 1, \ldots, n$ when the distributional assumptions are satisfied, then in addition to comparing each one to $\chi^2_{p,1-\alpha/2}$, the collection of test statistics can be verified to come from a χ^2_p-distribution.

The combination test can be done either using a new χ^2-test or be based on a Kolmogorov test.

In practise this has been implemented using those length distributions which cut across several length classes. Here the modelled and observed length distributions have been aggregated into a fixed number (e.g. 5) length groups, thus fixing the number of degrees of freedom.

Alternative tests for multinomial distribution

In addition to the usual χ^2-tests, several alternative tests exist to evaluate the goodness-of-fit of the multinomial assumption. Some of these have been used for typical fishery data sets.

First, when generalized linear models are used, the resulting fitted deviance for the multinomial distribution follows a χ^2-distribution. This has been evaluated using fairly extended models of the mean response and found to be an inadequate assumption for some tested cases (Stefansson and Palsson, 1997). This test could in principle be used in place of the test above based on X^2.

Second, it is possible to design special tests for whether individual pairs of cells follow the binomial distribution. For tested cases it is found that the observed variances from data sets are much greated than predicted from a binomial variance (Hrafnkelsson and Stefansson, 2002).

Thirdly, it can be evaluated whether the correlations between counts in pairs of length groups are slightly negative ($-np_ip_j$) as predicted from multinomial theory, and this assumption is similarly found to be totally invalidated for data sets tested (Hrafnkelsson and Stefansson, 2002).

These last two tests are of a different nature from the first and the tests based on X^2.

Gaussian distributions

If a Gaussian (and independence) assumption is used, the corresponding negative likelihood components are proportional to terms of the form

$$\sum_{i=1}^{n} (y_i - \mu_i)^2 ,$$

each of which has a distribution proportional to a χ^2_n distribution, if the means μ_i are known. In practice, the number of observations will be large and the subtraction of the degrees of freedom due to parameter estimation will not usually be an issue.

The nuisance parameter, σ^2 needs to be estimated, usually from the same data with

$$\hat{\sigma^2} = \frac{\sum_{i=1}^{n} (y_i - \hat{\mu}_i)^2}{n-q},$$

where q denotes the number of estimated parameters.

In the current setting, the number of estimated parameters in the Gadget setting is a highly dubious concept, since the parameters are estimated based on minimizing many more sums simultaneously. It is probably the most reasonable approach to assume $q = 0$, but some simulation testing is needed to evaluate this.

Normality of the standardised residuals (before squaring) is usually tested using Kolmogorov's D-statistic, but sometimes using χ^2, multinomial fashion, after appropriate grouping. Neither test is truly nonparametric (i.e. distribution-free) when the mean is estimated as here (cf Moore and Spruill, 1975) and in particular, χ^2 would be more applicable if a minimum χ^2-criterion is used for estimation, rather than maximum likelihood.

Combining several tests for the Gaussian distribution

Within Gadget a large number of likelihood components will typically be of the Gaussian type.

Tests of independence

Given that many data sets are indexed by time, it is natural to consider the autocorrelation in the various likelihood residual. Specific tests for Gadget output or generic likelihood components are not currently available but need to be developed as a part of future research.

These tests can not be easily integrated into the other χ^2-tests above and are therefore not considered further in this paper.

Evaluating combined likelihoods

One remaining question pertains to how it is possible to combine the various likelihood components into a single test statistic, ideally giving more power to identify conflicts between model and data. Thus, although it is certainly possible to apply a battery of tests to all components, it would be quite useful to have a single statistic which could be applied to the likelihood components as a whole, resulting in a single test for detecting unlikely values.

Naturally, this can be done if the likelihood functions are "structured" so that all components have the same degrees of freedom and are independent.

Results

In order to evaluate the adequacy of the usual statistical assumptions of (log-)normality and a multinomial distribution for various Gadget input data sets in a typical scenario, a sample scenario was taken and a battery of tests conducted.

The majority of tests for the multinomial distribution have rejected this as an assumption. Thus it is seen that there is not an immediate need to further develop such tests. Rather, the immediate need is to modify these assumptions, i.e. find distributional assumptions which apply in the fisheries setting. Although some work has been undertaken in this fashion (Hrafnkelsson and Stefansson, 2002), practical likelihood functions do not yet exist for fisheries length distributions.

Similarly, a very large number (about 45%) of tests for normality were rejected. It follows that the emphasis needs to be on modifying the assumption rather than further developing these tests.

Discussion

It is well-known that under fairly general assumptions, $-2lnL$, for each component group, does have an asymptotic χ^2_{n-q} distribution for large sample size n in each group, if the number q reflects the number of parameters estimated by this component group.

In the Gaussian case, this holds also for individual components with $n = 1$ and $q = 0$, which might be expected to apply approximately when there are many data points in total (across all components), relative to the number of parameters (including nuisance parametrs).

In general, for enough data, it is seen that the likelihood components can each be expected to follow a χ^2_{n-q}-type distribution, where n is the number of data items in each block and q is approximately 0. It is therefore an interesting future project to evaluate under what general conditions this can be extended down to the extreme case of $n = 1$.

It would be a particularly useful exercise to develop a single test statistic which could be used as a generic tool for testing the adequacy of the Gadget composite likelihood function. Such an indicator would immediately point a finger to further required analysis of individual deviations.

At present, however, the goodness-of-fit tests are seen to be sufficiently developed that they can reject most currently used likelihood functions used in fisheries. Further development therefore must await more appropriate distributional assumptions.

Acknowledgements

Much of the statistical testing reported on in this paper was undertaken by Ingileif Hallgrimsdottir.

References

Hrafnkelsson, B. and Stefansson, G. 2004. Analysis of categorical length data from groundfish surveys. Can. J. Fish. Aquat. Sci. 61: 1135-1142.

Moore, D. S. and Spruill, M. C. (1975). Unified large-sample theory of general chi-squared statistics for tests of fit. Annals of Statistics 3, No. 3. 599-616

Stefansson, G. and Palsson, O. K. (1997). Statistical evaluation and modelling of the stomach content of Icelandic cod (Gadus morhua). Can. J. Fish. Aquat. Sci. 54:1, 169-181.

5.2 Analysis of categorical length data from groundfish surveys

Gunnar Stefansson
Marine Research Institute, Reykjavik, Iceland

Several data sources in fisheries consist of counts in different cells. This is true for length distributions, catches in numbers at age and age-length frequency tables. Traditionally, several approaches exist for handling such data.

Different approaches have been used for modelling such data, typified by length distributions obtained from biological sampling of commercial or survey catches of fish. A simple approach would simply assume that the count or proportion in each length group comes from a Gaussian or binomial distribution. Both are clearly incorrect since the data correspond to counts into several length groups. Further, it is known that there are several problems with the binomial distribution, including the treatment of the weighting factor. From a theoretical viewpoint, the weighting factor should be based on $\pi_l(1 - \pi_l)$ where π_l is the true proportion in each length group. Since this proportion is unknown, one issue is whether to use the observed or modelled proportions in its place and either approach causes problems.

It has been claimed in the literature that the use of either the Gaussian or binomial should be replaced by a multinomial assumption but this is clearly not enough since the same overdispersion issue will arise when using the multinomial distribution. One approach to alleviate the overdispersion issue is to reduce the effective sample size, i.e. allow for overdispersion in the multinomial distribution. It is seen, however, that the deviations from distributional assumptions are much more serious than merely due to overdispersion. This is intuitively obvious since fish of similar size tend to behave similarly and not to be found with fish of different sizes. This implies a correlation structure which depends on the distance between length groups and contradicts the multinomial assumption, which implies a correlation structure between length cells independent of distance between length groups.

In Hrafnkelsson and Stefansson (2004) it is demonstrated how none of these methods are appropriate to length distributions in fisheries. A different method is tested and the corresponding model of length distributions is found to give a considerably better description of the data.

A serious problem with the model in Hrafnkelsson and Stefansson (2004) is that it does not provide a simple estimation procedure, nor a probability distribution easily incorporated into a likelihood estimation procedure such as Gadget.

References

Hrafnkelsson, B. and Stefansson, G. 2004. Analysis of categorical length data from groundfish surveys. Can. J. Fish. Aquat. Sci. 61: 1135-1142.

5.3 Distributions and models for survey abundance data

J. Brynjarsdóttir and G. Stefansson
Marine Research Institute, Reykjavik, Iceland

The paper by Brynjarsdottir and Stefansson (2004) develops models which describe the abundance of demersal species caught in a trawl survey. It is found in this paper that although there are several important predictor variables, there are also some features of the data not caught by the model. In particular, the distibution of abundance in non-empty hauls can neither be described by a gamma nor a lognormal distribution (though the lognormal gives a slightly better fit).

One curious aspect of the data illustrated in this analysis is a change in the survey CV in certain years. Such behavior will invalidate the assumption of heterogeneous variances in the lognormal distribution and of heterogeneous CV in the gamma density. It is not clear why such changes occur, but they will clearly affect (even invalidate) inferences made.

Brynjarsdottir, J. and Stefansson, G. 2004. Statistical Analysis of Cod Catch Data from Icelandic Groundfish Surveys. Fisheries Research. In press.

5.4 The weighting issue

Gunnar Stefansson
Dept. of Mathematics
University of Iceland and Marine Research Institute
Reykjavik, Iceland

As a part of dst^2, considerable work has been put into developing methods for evaluating the quality of the models and deciding what model may be appropriate.

It is generally acknowledged that models of population dynamics need to include realistic biological and statistical assumptions if they are to be of any use. The “realistic” complexity of these models is defined not only by the underlying biology but also by what the available data can reasonably determine. As complexity increases through inclusion of interactions new issues arise, which need to be resolved before these complex models can be reliably applied. Specific issues in new models for management include proper statistical fitting to the data and incorporation of time-dependent and spatially explicit biological interactions. These current problem issues are presented in Stefansson (2004) and some potential solutions are discussed.

When fitting a complex model several data sources are usually needed. Each data source (i) is compared to the model predictions through a negative log likelihood which is considered a function of the parameter vector ($l_i(\theta)$). Given relative weightings to each data set, the task becomes the estimation of the parameter vector by minimizing the combined log-likelihoods vix

$$min_\theta \sum_i w_i i_i(\theta)$$

A method is proposed in Stefansson for estimating the weighting factors, w_i. It is seen that the choice of these weights can be very important.

A related concern is whether one can detect the failure of the model to fit all data sets well simultaneously. This matter is addressed in section 5.7.

Stefansson, G. 2004. Issues in multispecies models. Natural Resource Modeling, 16(4): 415–437.

5.5 An Application of Multivariate Normal in a Statistical Fisheries Model

Bjarki Þór Elvarsson
Marine Research Institute, Reykjavik, Iceland

Abstract: This report looks into an implementation of a multivariate normal likelihood function in GADGET's catch distribution class. It was implemented in hope that it would surpass the multinomial likelihood function used for the length distribution as the multinomial one had shown it was a very poor fit to real data.

Implementation

GADGET provides an estimate of the catch using any of GADGET's suitability functions. This estimated proportion of fish caught per length group, π_{tral}, is dependent on time t, position r and age a as stock catch varies accordingly. The working assumption was that $(\mathbf{P_{tra}} - \mathbf{\Pi_{tra}})$, where P_{tral} denotes real catch proportions, came from a multivariate normal distribution with mean 0 and variance-covariance matrix Σ. Furthermore some correlation structure was assumed so that the entire variance-covariance matrix, Σ, could be denoted with as few as two parameters.

The implementation consists of two modules, one that defines the correlation pattern for the variance-covariance matrix at the beginning of each simulation run and the other which is updated at each time step and performs the likelihood calculation.

The matrix module could be programmed to create any positive definite symmetric matrix as the matrix code is entirely independent from the likelihood calculations but currently an autocorrelation pattern is implemented to GADGET. With an autocorrelation assumption it implied that we can write a measurement x_i as in (5.1).

$$x_i = \sum_{k=1}^{n} \alpha_k x_{i-k} + \epsilon_i \tag{5.1}$$

that is as a linear combination of the foregoing n variables with added error $\epsilon_i \sim N(0, \sigma^2)$. From (5.1) we derive a correlation pattern for the variance-covariance matrix $\Sigma = (\sigma_{ij})_{i=1,j=1}^{p,p}$

$$\sigma_{ij} = \begin{cases} \sum_{k=1}^{min(n,i)} \alpha_k \sigma_{i,j-k} + \delta_i^j \sigma^2 & \text{if } i \geq j \\ \sum_{k=1}^{min(n,j)} \alpha_k \sigma_{i-k,j} & \text{otherwise} \end{cases} \tag{5.2}$$

where δ_i^j is the Kronecker symbol and $(\alpha_k)_{k=1}^n$ are the autocorrelation coefficients which need to be estimated. This pattern is quite flexible as it can assume the role of univariate normal by setting the lag, n, to zero as well as any arbitrary lag (but a large lag could not be feasible due to excessive number of parameters that need to be optimised).

The likelihood calculation module calculates the likelihood score of the multivariate normal using the matrix pattern defined earlier as its variance-covariance matrix. The likelihood score is calculated from the usual log-transform of the likelihood function in (5.3)

$$\ell = -AT\frac{p}{2}log(2\pi) - \frac{AT}{2}log|\Sigma| - \sum_{r=1}^{A}\sum_{t=1}^{T} (\mathbf{P_{tra}} - \mathbf{\Pi_{tra}})\Sigma^{-1}(\mathbf{P_{tra}} - \mathbf{\Pi_{tra}}) \tag{5.3}$$

where $\Sigma^{-1}(\mathbf{P_{tra}} - \mathbf{\Pi_{tra}})$ is calculated using a LU decomposition on Σ, A is the number of areas and T the number of years.

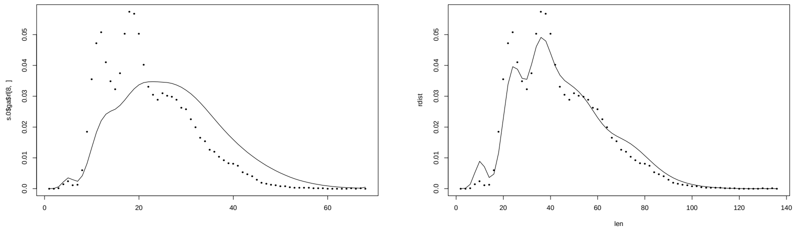
It can be seen that equation (5.3) is defined only when $det(\Sigma) \neq 0$. It is therefor important that the parameters are not estimated so the determinant is optimised to zero. This is done in GADGET by awarding a penalty whenever the determinant is optimised to zero.

A few things should be noted about this new length distribution model. It can be set up as it were a previously implemented sum of squares model by setting the lag as zero and fixing the variance to 1. When multivariate normal likelihood function is set up like this the resulting likelihood score should be the same from both sum of squares and multivariate normal. Differences in the likelihood score could be due to the calculation of $\Sigma^{-1}(\mathbf{P_{tra}} - \mathbf{\Pi_{tra}})$ and $|\Sigma|$ in (5.3). Another thing that is worth considering is that there are no lower bounds (or even upper bounds) for the multivariate normal likelihood function as opposed to previous likelihood functions in GADGET who have a lower bound of zero. This could cause problems because a optimiser will be more prone to optimise the length distribution and other likelihood components will become negligible. from $\hat{S}$).

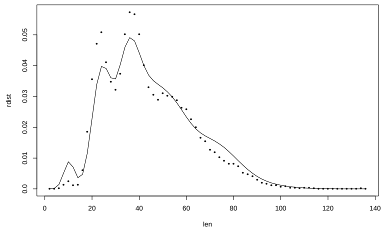
Testing for normality

A variable X is said to be $N(\theta, \Sigma)$ if it is possible to write X as $\theta + A\mathbf{u}$ where A is a nonsingular factorisation of Σ such that $\Sigma = AA'$ and $\mathbf{u} \sim N(\mathbf{0}, \mathbf{I})$, that is $\mathbf{u} = (u_i)_{i=1}^n$ where $u_i \sim N(0,1)$ and $(u_i)_{i=1}^n$ are i.i.d. So to test our assumption that $(\mathbf{P_{tra}} - \pi_{\mathbf{tra}})$ came from a multivariate normal distribution with mean $\mathbf{0}$ and variance Σ, where Σ is our autocorrelation matrix, $(\mathbf{P_{tra}} - \pi_{\mathbf{tra}})$ need to be scaled with $\Sigma^{-1/2}$ ($\Sigma^{-1/2}$ is the inverse of A the factorisation mentioned above).

The assumption was tested with cod length distribution data from ground fish survey from the years 1985 to 2001 in a single area model with autocorrelation lags 0 (with σ fixed),1,2,3,4 and 10 and 2cm lengthgroups. To illustrate the difference between the lags length distribution plots which can be seen from figure (5.1) were created using data from the year 1993.



(a) lag 0 (b) lag 3



(c) lag 10

Figure 5.1: Model (lines) and survey (points) distribution data from the year 1998 compared. Data was aggregated into 2cm lengthgroups.

It should also be noted that the lag was set to a higher value than all the autocorrelation coefficients after the fourth were optimised to an absolute value smaller then 0.01. Also there is no visible difference between GADGET's approximations from lag 1 and up. There should then be no surprise that a results from a normality tests were similar for lags 1 to 10. The residuals were tested for normality used a Wilk-Shapiro test and a sample was rejected if p-value was < 0.05.

lag	likelihood score	% accepted
0	8.4226024	89.7
1	-261865.58	91.2
2	-261829.39	75
3	-263609.07	79.4
4	-267211.94	86.8
10	-267408.69	86.8

As mentioned above the scaled residuals should be univariate normal so $\Sigma^{-1/2}(\mathbf{P_{tra}} - \pi_{\mathbf{tra}})$ should be a vector of i.i.d normal variables. This was tested again with the Wilk-Shapiro normality test with the following results:

lag	% accepted
0	0
1	94.1
2	100
3	100
4	100
10	100

This of course only tells us that each of the vectors $\Sigma^{-1/2}(\mathbf{P_{tra}} - \pi_{\mathbf{tra}})$ is normal but the correlation between the lengthgroups is another thing.

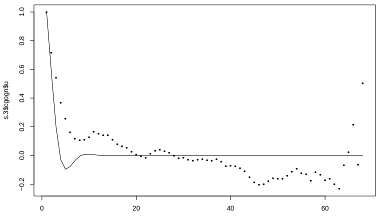


Figure 5.2: Autocorrelation from real data (points) compared with estimated autocorrelation (lines) with lag 3

As can be seen from figure (5.2) the GADGET estimate does not fit very well to real correlation data. A better estimate of the autocorrelation could be found by fitting a model with a higher lag but at the expense of a greater computation time.

The search for a suitable matrix

The autocorrelation approach described above could have a drawback, a full blown variance-covariance matrix

$$\hat{S} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})'$$

is not always ideally described with few parameters, which is to be expected given that you are trying to estimate n^2 variables with as few as two or three parameters. So how far off could our estimate be? A suitable matrix A would effectively minimises the difference $||\Sigma - A||$ for all x. So a series of bootstrapping was conducted to get a estimate of the distribution of $||\Sigma - \hat{\Sigma}||$ to see what is a necessary condition for a suitable matrix.

The data in the bootstrapping were the same as before or the Icelandic groundfish survey data from the years between 1985 and 2001 but now aggregated to 1cm lengthgroups. The distribution data in figure (5.3) was then compared with GADGET's estimate of Σ, Σ_g, from an optimum point and compared it the bootstrap estimate, $\hat{\Sigma}$, that is $||\Sigma_g - \hat{\Sigma}||$ was calculated to see whether or not the estimate fits into the distribution.

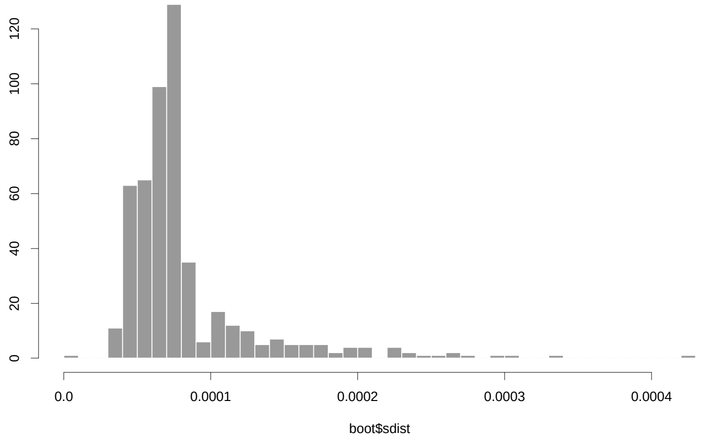


Figure 5.3: Distribution of $|\Sigma - \hat{\Sigma}|$.

The GADGET estimate fitted quite well to the distribution of $||\Sigma_g - \hat{\Sigma}||$, optimising with lag 6 resulted in a difference of $8.9 * 10^{-5}$ and lag 12 $8.6 * 10^{-5}$. It should be noted that $||\hat{\Sigma}|| \approx 9 * 10^{-5}$ and $||\Sigma_g|| \approx 6 * 10^{-5}$

Multivariate Logistic

The multivariate logistic likelihood function is yet another model built into Gadget's catch-distribution class. The multivariate logistic model it is assumed that the observed proportions P_{lt} in each length group can be represented as

$$P_{tral} = \frac{e^{x_{tral}}}{\sum_{l'=1}^{L} e^{x_{tral'}}} \tag{5.4}$$

where

$$x_{tral} = log(\pi_{tral}) + \epsilon_{tral} - \frac{1}{L}\sum_{l'=1}^{L}(log(\pi_{tral'}) + \epsilon_{tral'})$$

π_{tral} is our GADGET estimation of the catch-distribution, $\epsilon_{tral} \sim N(0, \sigma^2)$ and L is the number of length groups. Then the likelihood function for the length distribution is (Schnute et al)

$$L = (\sqrt{2\pi}\sigma)^{T(L-1)} L^{T/2} exp(-\frac{1}{2\sigma^2} \sum_{l=1}^{L} \sum_{t=1}^{T} \nu_{tral}^2) \tag{5.5}$$

where

$$\nu_{lt} = log(P_{tral}) - log(\pi_{tral}) - \frac{1}{L} \sum_{l=1}^{L} (log(P_{tral}) - log(\pi_{tral}))$$

and as usual the $-log()$ transform of (5.5) was used in the calculation of the likelihood score.

The results, when optimising to the Icelandic groundfish survey data, were less than stellar as can be seen from figure (5.4)

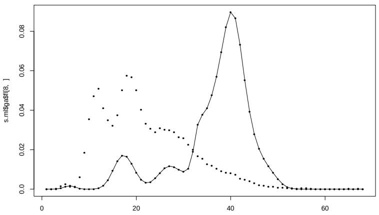


Figure 5.4: Survey data (points) from the year 1993 compared with the gadget estimate (lines) from optimisation with multivariate logistic function

5.6 Bootstrapping multiple data sources for estimating uncertainty in parameter estimates in complex models

Vojtěch Kupča
Marine Research Institute

Verena Trenkel
Laboratoire Maerha
Ifremer

Lorna Taylor
MRI

Gunnar Stefansson
University of Iceland
and MRI

Abstract: A bootstrap method is introduced for estimating the uncertainty in parameter estimates when multi-parameter complex models are fitted to multiple data sources.

Introduction

When a single data source is being used and a sum of squares is the appropriate fitting criterion, it is well known that the Hessian or Jacobian matrices can be used to obtain estimates of uncertainty, i.e. variance estimates for the parameters obtained from the minimization.

This only applies, however, when the data really do come from a Gaussian distribution. If this assumption is inappropriate some other method must be used. Naturally, alternative likelihoods can in principle be used and this approach has been extensively developed in the theory of generalized linear models (McCullagh and Nelder 1989). However, when the distributional properties of the data are not well understood or the models are now correct, these approaches will fail as seen in several examples in fishery science (Patterson et al., 2001).

When multiple data sources are considered, a different issue arises, which is how the data sets should be weighted. Given the problems involved in using a single data sources, it would seem rather futile to try to use analytical or parametric approaches such as the ones above.

Bootstrap methods (Efron and Tibshirani, 1993) provide a general nonparametric mechanism for estimating uncertainty in any estimation method. In introductory textbooks it is assumed that the data are simple measurements without correlation. However semi-parametric approaches have also been developed to sample residuals from a model, possibly from a distribution (parametric bootstrap) and even with a correlation structure.

In the current setting the correlation structure is non-trivial, needing to take into account the fact that fish within a sample are more similar than between samples. Even resampling entire samples of fish will not quite suffice since this implies that samples of ages and lengths at the same station will be correlated through the intra-haul correlation (Pennington and Volstad, 1994). There is a need to undertake resampling in such a manner as to resample simultaneously all samples taken in the same unit of time and area.

A proposal

A typical fisheries data base consists of data from a variety of sources. Every sample from each data source can be sorted by sampling location and time. A model such as Gadget operates on certain time steps and uses area definitions which are always fairly large, though there may be several. Within any such spatio-temporal unit or “Gadget cell” there will normally be several samples.

Naturally, one will not get the correct results if bootstrapping is implemented by resampling individual measurements within each Gadget cell.

In principle one could resample the raw biological samples. However this has two basic problems. The first is a pragmatic one in that normally Gadget input data is stored in a data base system which consists of some aggregates and the samples are lost. The second problem is a statistical one and has to do with the correlation across samples as described above. This even applies to samples from small commercial vessels which may well be fishing on the same school of fish during the same short time interval and similar landing location.

It would therefore seem appropriate to base the resampling directly on the data which has already been blocked into “data cells” in the standardized data base. It is therefore proposed that resampling be implemented by allowing sampling with replacement from the data warehouse.

The data within the standardized data base is aggregated within some small time step (typically a month) and some small unit area (typically a quarter of a statistical square, though full statistical squares may also be used).

The data within a Gadget run are usually aggregated across many such squares. Thus a Gadget area may consist of e.g. 100 such rectangles. Naturally, if quarters or years are used as a Gadget time step, then the aggregations would be corresponding in time as well.

The aggregation method from the standardized data base to Gadget input files varies somewhat depending on the data source. Some data are simply added up whereas others consist of e.g. mean length at age and go through a computational mechanism.

Any one of these aggregation methods can be based on resampling entire data cells from the data warehouse. Each such resampling will lead to a new Gadget data set. A Gadget estimation run based on such a data set will result in a resampled parameter estimate. The collection of all such estimates form a bootstrap sample.

Implementation

An implementation of the bootstrap as described here is given by Kupča, V. (2005), indicating how bootstrapping can be done directly from a data base. This is only the simplest case, where a single data source is considered, albeit a data source with internally correlated data sets. A case study of the method is given by Hannesson et al (2005).

Discussion

A simple example of this approach is given in 4.4, where resampling of entire tagging experiments is considered.

References

Efron, B., R.J. Tibshirani. 1993. An introduction to the bootstrap. Chapman and Hall, New York. 436p.

Hannesson, S, Jakobsdottir, A., Begley, J., Taylor, L. and Stefansson, G. 2005. On the use of tagging data in statistical multispecies multi-area models of marine populations. Section 4.4 of this report.

Kupča, V. 2005. Bootstrapping of areas. Section 2.6 of this report.

McCullagh, P. and Nelder, J. A. 1989. Generalised Linear Models, 2^{nd} ed. Chapman and Hall, London.

Patterson, K., Cook, R., Darby, C, Gavaris, S., Kell, L., Lewy, P, Mesnil, B., Punt, A., Restrepo, V., Skagen, D. W. and Stefansson, G. 2001. Estimating uncertainty in fish stock assessment and forecasting. Fish and fisheries, 2, 125-157.

Pennington, M. and J. H. Volstad. 1994. Assessing the effect of intra-haul correlation and variable density on estimates of population characteristics from marine surveys. Biometrics, 50: 725-732.

5.7 Verifying consistency when fitting complex statistical models to multiple data sets with applications to fish population dynamics and fisheries.

Gunnar Stefansson
Dept. of Mathematics, University of Iceland
and Marine Research Institute

Lorna Taylor
Marine Research Institute

Abstract: Modern models of fish population dynamics have grown in complexity and size, being routinely based on formal statistical methods, sometimes taking into account multispecies concerns as well as the effects of fishing and, more rarely, climate. These models have drawn forward the needs for new statistical, mathematical and informatics techniques in the field of fishery science. The models tend to be highly non-linear, use considerable amounts of data and have large numbers of unknown parameters.

Introduction

Fishery science deals with the analysis of the interactions between fisheries and the ecosystem, possibly taking into account economic and social aspects. In order to analyse such interactions data are collected from the ecosystem, mainly by sampling and measuring fish in various types of surveys and from the fishery in the form of recording the catches, effort and biological sampling of the catches. These two types of biological samples routinely include measurements of the total amount caught, length measurements of individual fish, expensive age measurements from a subset of these and stomach content analysis from an even smaller subset.

This leads to enormous data sets typically involving several commercial fleets, hundreds of thousands of length measurements, tens of thousands of age readings and thousands to tens of thousands of analyses of stomach contents by prey species.

Different models of the population dynamics of fish stocks have been able to utilise varying amounts of these data. In recent decades models have been developed to use all of the data through likelihood components when describing species and fleet interactions in spatially disaggregated models. This paper thus considers the scenario when large and disparate data sets are used and weightings to be determined. In this scenario questions arise concerning whether the model can adequately explain all the data sets simultaneously.

A technique for investigating the weighting problem has been proposed earlier (Stefansson, 1998). The method is based on iteratively putting heavy weights on individual data components in order to estimate the minimum value of each likelihood component, as described in detail by Stefansson (2004). Having solved the weighting issue to some extent, the next question is how well the model can fit all the data sets. Using only one of the data sets (or putting a hight weight on it) will give a set of parameter estimates which may vary considerably from that obtained from another data source. Since a large model must have the ability to accommodate all the data sets, there is a need to verify how different these various parameter estimates can be allowed to become without being significantly different in a formal sense.

Sensitivity analyses: Testing for model stiffness

Problem statement

In ordinary linear regression it is customary to estimate the influence of each observation on the corresponding predicted value, or on each parameter estimate, using hat matrices and similar techniques.

In constrained nonlinear optimisation it is similarly customary to evaluate the effect of each constraint on the objective function through the marginal effects of alleviating the constraints, resulting in shadow values.

When highly complex and compound models are analyzed constraints are implicit. For example, when growth and mortality submodels are analyzed within a framework of a single model, it is implicitly assumed that the two submodels can be unified through some restrictions on the parameter space. It may not be clear, however, how these constraints work or can be alleviated through more flexible parametrization. Further, the complex models tend to be associated with large composite data sets, making it infeasible to evaluate the effect of each observation on every prediction or estimate.

A simple simulated example

To mimic this issue, take the simple true state of nature (the data generator) to consist of independent measurements around two straight lines

$$Y_{ij} \sim n(\alpha_i + \beta_i x_{ij}, \sigma_i^2), \quad j = 1, \ldots, n_i \;\; i = 1, 2,$$

and write $n = n_1 + n_2$.

Suppose the experimenter does not know that the slopes are actually different, but merely collects y-data corresponding to different x-values and for each data set (i). This simple example, taken from Stefansson (2004), will be used to develop and verify test statistics, to be used in more generic settings.

Given the general infeasibility of working with the individual residuals, it is more natural to consider how the individual data groups imply different directions for each parameter. Equivalently, the partial derivative of each likelihood component with respect to each parameter can be computed in order to evaluate these directions.

Derivatives and elasticities

Write S_i for the negative log likelihood for the i'th data source and θ_j for the j'th parameter, i.e. component j of the entire parameter vector $\boldsymbol{\theta}$. The weighting methodology of Stefansson (2004) attaches a high weight to data set i to obtain an estimate of the whole parameter vector, $\boldsymbol{\theta}^{(i)}$, corresponding to indications in data set i.

In the general case, these derivatives need to be evaluated numerically, but for the above theoretical example, one can use $S_i = \hat{\sigma}_i^2$, to obtain

$$\frac{\partial S_i}{\partial \alpha_j} = 0 \text{ for } i, j = 1, 2 \tag{5.6}$$

$$\frac{\partial S_i}{\partial \beta} = -\frac{2}{n_i} \sum_{j=1}^{n_i} \left(y_{ij} - \hat{\alpha}_i - \hat{\beta} x_{ij} \right) x_{ij} \tag{5.7}$$

The fact that $\frac{\partial S_i}{\partial \alpha_j} = 0$ holds exactly in the simple example is a consequence of each intercept being completely determined by minimizing the corresponding sum of squares (maximizing the likelihood components). On the other hand, β is determined through the total likelihood function and depends on both components. It follows that although the expected value of each residual, $\left(Y_{ij} - \hat{\alpha}_i - \hat{\beta} x_{ij}\right) x_{ij}$, viewed as a random variable, is zero when $\beta_1 = \beta_2$, partial sums of the observed residuals will not in general be zero even under this hypotheses.

Partial sums of the residuals will, however, in the simple example be observations from Gaussian distributions when the model is correct. Hence, the partial residual sums can be used to form test statistics for the null hypothesis of the model being correct, as will be shown on page 138.

Such derivatives (or related elasticities) have been used in earlier analyses to investigate the local behaviour of each likelihood component with respect to changes in each parameter (e.g. Stefansson and Palsson, 1997) and found to give useful indications in some cases regarding the strength of the “pull” of each data set.

The testing framework

Naturally, the hypothesis $H_0 : \beta_1 = \beta_2$ in the simple example is a linear hypothesis in a linear Gaussian model. Although this simplified framework is typical of the general case, it might appear to simple at first sight.

Since this is a linear hypothesis in a in a linear Gaussian model, the hypothesis can in principle be tested in a standard (and trivial) manner. The point is, however, that the test statistic to be developed should be a generic test for model stiffness. A generic version of this test can be developed by considering the models for each data set separately, $\mathbf{Y}^{(1)} = \mathbf{X}^{(1)}\boldsymbol{\theta}^{(1)} + \boldsymbol{\epsilon}^{(1)}$ and $\mathbf{Y}^{(2)} = \mathbf{X}^{(2)}\boldsymbol{\theta}^{(2)} + \boldsymbol{\epsilon}^{(2)}$. This can be phrased as a single linear model in matrix form and the hypothesis of the two data sets being consistent can now be reduced to $H_0 : \boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}^{(2)}$.

In the current setting this is a linear hypothesis in a linear model, which would usually (with equal variances) be tested with an F-statistic of the form:

$$\begin{aligned} & \frac{(SSE(R) - SSE(F))/(p-q)}{SSE(F)/(n-p)} \\ = \ldots = & \frac{(||\mathbf{X}^{(1)}\hat{\boldsymbol{\theta}}^{(1)} - \mathbf{X}^{(1)}\hat{\boldsymbol{\theta}}||^2 + ||\mathbf{X}^{(2)}\hat{\boldsymbol{\theta}}^{(2)} - \mathbf{X}^{(2)}\hat{\boldsymbol{\theta}}||^2)/p}{(||\mathbf{y}^{(1)} - \mathbf{X}^{(1)}\hat{\boldsymbol{\theta}}^{(1)}||^2 + ||\mathbf{y}^{(2)} - \mathbf{X}^{(2)}\hat{\boldsymbol{\theta}}^{(2)}||^2)/(n-p)} \end{aligned}$$

where $\hat{\boldsymbol{\theta}}$ refers to estimation in the reduced model (i.e. under H_0).

Unfortunately this resulting F-test for consistency of the parameters contains the same sums of squares as go into the estimation of weighting factors in the case of unknown and different variances. For this reason it is not feasible to combine the case of uncertain weights with the otherwise obvious F-test for model consistency, and a different test statistic will be developed in what follows.

An F statistic for consistency

A different F-test may be derived by comparing model predictions for data set 1 as obtained from parameter estimates from data set 1 or the rest of the data viz $\mathbf{X}^{(1)}\hat{\boldsymbol{\theta}}^{(1)} - \mathbf{X}^{(1)}\hat{\boldsymbol{\theta}}^{(2)} = \mathbf{X}^{(1)}\left(\hat{\boldsymbol{\theta}}^{(1)} - \hat{\boldsymbol{\theta}}^{(2)}\right)$.

This difference is based on $\mathbf{y}^{(1)}$ only through linear combinations of columns of the $\mathbf{X}^{(1)}$-matrix (as in $\mathbf{H}^{(1)}\mathbf{y}^{(1)}$ where $\mathbf{H}^{(1)}$ is the hat-matrix). These columns are orthogonal to the columns of the matrix forming the residuals from the first model fit, $\mathbf{y}^{(1)} - \mathbf{X}^{(1)}\hat{\boldsymbol{\theta}}^{(1)}$ ($= \left(\mathbf{I}^{(1)} - \mathbf{H}^{(1)}\right)\mathbf{y}^{(1)}$). It follows that the ratio

$$F_c := \frac{||\mathbf{X}^{(1)}\hat{\boldsymbol{\theta}}^{(1)} - \mathbf{X}^{(1)}\hat{\boldsymbol{\theta}}^{(2)}||^2/p}{||\mathbf{y}^{(1)} - \mathbf{X}^{(1)}\hat{\boldsymbol{\theta}}^{(1)}||^2/(n_1 - p)} \tag{5.8}$$

is a ratio of independent sums of squares.

Apart from being one of many obvious choices, it is not too hard to ascertain that this statistic is asymptotically (as $n_2 \to \infty$) equivalent to the likelihood ration statistic for testing the above null hypothesis.

The likelihood ratio is proportional to

$$\frac{||\mathbf{y}^{(1)} - \mathbf{X}^{(1)}\hat{\boldsymbol{\theta}}||^{n_1}||\mathbf{y}^{(2)} - \mathbf{X}^{(2)}\hat{\boldsymbol{\theta}}||^{n_2}}{||\mathbf{y}^{(1)} - \mathbf{X}^{(1)}\hat{\boldsymbol{\theta}}^1||^{n_1}||\mathbf{y}^{(2)} - \mathbf{X}^{(2)}\hat{\boldsymbol{\theta}}^2||^{n_2}} \tag{5.9}$$

where $\hat{\boldsymbol{\theta}}$ is the combined estimate. This is asymptotically (as $n_2 \to \infty$) equivalent to

$$\frac{||\mathbf{y}^{(1)} - \mathbf{X}^{(1)}\hat{\boldsymbol{\theta}}^1||^{n_1}}{||\mathbf{y}^{(1)} - \mathbf{X}^{(1)}\hat{\boldsymbol{\theta}}^1||^{n_1}} \tag{5.10}$$

(up to a multiplicative constant) and this is equivalent to the above F-test (up to an additive constant and change of power).

If the parameter vector is completely determined by each data set so that e.g. $\hat{\boldsymbol{\theta}}^{(2)}$ is a consistent estimator of $\boldsymbol{\theta}^{(2)}$, then the asymptotic distribution (as data set 2 gets larger) of the ratio (5.8) will be the same as that of

$$\frac{||\mathbf{X}^{(1)}\hat{\boldsymbol{\theta}}^{(1)} - \mathbf{X}^{(1)}\boldsymbol{\theta}^{(2)}||^2/p}{||\mathbf{y}^{(1)} - \mathbf{X}^{(1)}\hat{\boldsymbol{\theta}}^{(1)}||^2/(n_1 - p)}. \tag{5.11}$$

When the parameters are equal, under H_0, and the rank of $\mathbf{X}^{(1)}$ if p, it follows that, F_c has an F-distribution with p and $n_1 - p$ degrees of freedom, asymptotically as n_2 gets large.

The asymptotic distribution of the test statistic, F_c (i.e. the distribution of (5.11), does not require assumptions on the equality of variances, since it is only based on data set 1 and predictions of data set 1. Given the asymptotic properties of predictions based on the other data sets and independence of all measurements, the null distribution of F_c is formally only based on normality and constancy of variance of measurements within data set 1. It will be seen that for asymptotics to kick in, however, some finite sample properties will be required on data set 2.

The simplified example illustrates this scenario but it also illustrates some of the problems involved in the approach. For example the slope can indeed be estimated with each data set separately, but the two intercepts α_1 and α_2 can not. It follows that in order to estimate the entire parameter vector both data sets need to be used, but in order to obtain approximate independence and an F-distribution, the weight of the primary data set is made arbitrarily large when estimating the respective parameter, as suggested by Stefansson (2004). In such model fits the degrees of freedom only make approximate guides.

The simple example suggests how the degrees of freedom may possibly be corrected in general: Suppose data set 1 can estimate $r_1 \leq p$ of the parameters, and data set 2 estimates $r_2 \leq p$ of them (in the linear model this can be formalized in terms of estimable functions and ranks of the X-matrices). Of the p parameters which enter the computations of the numerator, not all will be independently estimated by the two data sets if $r_2 < p$. In this case $p - r_2$ parameters in the $\hat{\boldsymbol{\theta}}^{(2)}$-vector are actually re-estimated through data set 1, basically repeating the estimation in the $\hat{\boldsymbol{\theta}}^{(1)}$-vector. Further, data set 1 could really only estimate r_1 parameters to begin with and therefore the degrees of freedom in the numerator should be approximated by $r_1 - (p - r_2) = r_1 + r_2 - p$. In the denominator the degrees of freedom should obviously be replaced by $n_1 - r_1$. These guidelines should only be considered indications, as it is likely that in a given situation some further analysis would be appropriate.

It should be noted that although the technique is similar to cross-validation (Neter et al, 1996), the formal approach and applicability to nonlinear models is somewhat different.

The use of asymptotics to get the approximate F distribution implies that this method should only be used when n_2 is considerably greater than n_1. The complex models and data which initiated this research typify this situation: There is a large number of data sources and the interest is in sequentially verifying each one against the entire remaining set. In this case n_1, though usually large compared to the number of parameters, is always much smaller than the number of data points in the remaining data sets.

Under the null hypothesis that the two parameter vectors are equal, with common values $\boldsymbol{\theta}$, the expected value of the numerator SS can be written as the sum of the terms

$$E\left[||\mathbf{X}^{(1)}\hat{\boldsymbol{\theta}}^{(1)} - \mathbf{X}^{(1)}\boldsymbol{\theta}||^2\right] = p\sigma^2 \tag{5.12}$$

(seen using the geometric basis methods of Scheffe, 1959) and

$$E\left[||\mathbf{X}^{(1)}\hat{\boldsymbol{\theta}}^{(2)} - \mathbf{X}^{(1)}\boldsymbol{\theta}||^2\right]. \tag{5.13}$$

Since (5.12) is the correct expected value and (5.13) is a positive number (which gets smaller as n_2 increases), it is seen that the test tends to be too liberal and it is only level-α as $n_2 \to \infty$.

Expression (5.13) provides guidance as to when the F_c statistic may be appropriate. If n_2 is large compared to n_1, the difference in (5.13) will become negligible. However, if n_1 is large compared to n_2, problems will occur regardless of how large n_2 is. This is easily observed in the case when $\mathbf{X}^{(1)}$ has only two different rows, each repeated $n_1/2$ times. In this case the sum in (5.13) consists of only two different values, each repeated $n_1/2$ times. As the ratio n_1/n_2 increases, the sum can be made arbitrarily large. It follows that F_c should only be used when n_2 is considerably larger than n_1.

Although there has been no mention of or implicit assumption on σ_2^2 when developing the test statistic, it is clear from (5.13) that an increase in σ_2^2 will lead to a poorer control of the α-level through inappropriateness of the asymptotic properties. Basically, increasing σ_2^2 will have a similar effect as increasing the ratio n_1/n_2 Unfortunately, this statement is not generalized into the nonlinear framework in an obvious manner: The variability of interest is really the variability expressed in $V[\hat{\boldsymbol{\theta}}^{(2)}]$ and this can become arbitrarily complex as the model gets nonlinear.

As a possible rule-of-thumb, to use F_c in the general case one should try to ensure that n_2 is considerably larger than n_1 and the variances of the parameter estimates using data set 2 be no greater than than those from data set 1.

An alternative: A t-test for consistency

The following asymptotics can be used to derive an alternative general consistency test. Consider one of the data sources, say i. In this case one can define the statistic

$$U_{ij} = \frac{\partial S_i}{\partial \theta_j} \tag{5.14}$$

in the generic case, which is

$$U_i = \frac{\partial S_i}{\partial \beta} \quad = \quad -\frac{2}{n}\sum_{j=1}^{n_i}\left(y_{ij} - \hat{\alpha}_i - \hat{\beta}x_{ij}\right)x_{ij} \tag{5.15}$$

for the simple example. In the case of the linear Gaussian model the variance of the statistic can be computed. Write $\mathbf{e} := \mathbf{Y} - \hat{\mathbf{Y}}$ where

$$\mathbf{Y}' = (y_{11}, \ldots y_{1n_1}, y_{21}, \ldots, y_{2,n_2})$$

is the vector of all measurements and $\hat{\mathbf{Y}}$ is the the vector of all fitted values.

Define the vector of n_1 measurements and n_2 zero values,

$$c_1 = \frac{-2}{n}(x_{11}, x_{12}, \ldots, x_{1n_1}, 0, 0, \ldots, 0)'$$

and correspondingly the vector c_2.

Denote by $\mathbf{W}$ the diagonal matrix of variances of measurements. Thus, $w_{kk} = \sigma_i^2$ where $i = 1$ for $k = 1, \ldots, n_1$ and $i = 2$ for $k = n_1 + 1, \ldots, n_1 + n_2$ but $w_{kj} = 0$ if $k \neq j$. In the general case, $\mathbf{W}$ would correspondingly contain a line for each datum and the diagonal should be the variance of the datum (constant within each data source). Each of these diagonal elements can be estimated using separate minimisations for each data set, placing a high weight on each in turn.

Further, define the vectors

$$\mathbf{b_i} = \mathbf{c}_i' \left(\mathbf{I} - \mathbf{X} \left(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{W}^{-1} \right).$$

Given these definitions it follows that the residuals $\mathbf{e}$ and residual statistics $U_i,\ i = 1, 2$ take the form

$$\begin{aligned} U_i = \mathbf{c}_i'\mathbf{e} &= \mathbf{c}_i' \left(\mathbf{Y} - \hat{\mathbf{Y}} \right) \\ &= \mathbf{c}_i' \left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}} \right) \\ &= \mathbf{c}_i' \left(\mathbf{Y} - \mathbf{X} \left(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{W}^{-1}\mathbf{Y} \right) \\ &= \mathbf{c}_i' \left(\mathbf{I} - \mathbf{X} \left(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{W}^{-1} \right) \mathbf{Y} \\ &= \mathbf{b}_i'\mathbf{Y}. \end{aligned}$$

From this it is seen that the variance of U_i is given by

$$\sigma_{U_i}^2 := V[U_i] = V[\mathbf{b'}_i\mathbf{Y}] = \mathbf{b'}_i V[\mathbf{Y}]\mathbf{b}_i = \mathbf{b'}_i \mathbf{W} \mathbf{b}_i. \tag{5.16}$$

It is well known that

$$U_i/\sigma_{U_i} \sim n(0, 1),$$

but of course the above σ_{U_i} involves unknown parameters through $\mathbf{W}$. Asymptotic theory (Slutsky's theorem, e.g. Randles&Wolfe (1979)) implies the equality of some asymptotic distributions, where notably

$$\frac{U_i}{\hat{\sigma}_{U_i}} \tag{5.17}$$

will have the same asymptotic distribution as

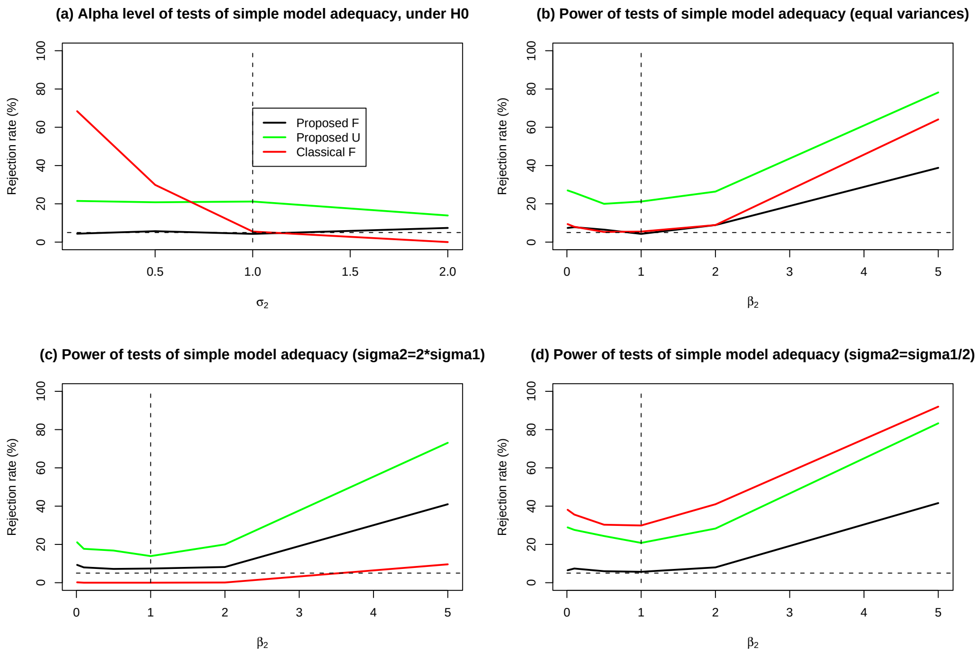
$$\frac{U_i}{\sigma_{U_i}} \sim n(0, 1) \tag{5.18}$$

where $\hat{\sigma}_{U_i}$ is any consistent estimator of σ_{U_i}, as obtained through any consistent estimator of $\mathbf{W}$.

It follows that a very simple test of H_0 is obtained using the U-statistic where the variances in $\mathbf{W}$ are estimated through separate weighted minimization of the negative log-likelihoods under the assumptions $\sigma_i \rightarrow 0$.

Results

Simulation results are given in Fig. 5.5. It is seen that the proposed F_c statistic appears to hold its α-level within the range of values of σ_2 considered. On the other hand, the proposed U-statistic does not hold its level. This can of course happen when sample sizes are too small for asymptotic properties to kick in.


(a) Alpha level of tests of simple model adequacy, under H0

(b) Power of tests of simple model adequacy (equal variances)

(c) Power of tests of simple model adequacy (sigma2=2*sigma1)

(d) Power of tests of simple model adequacy (sigma2=sigma1/2)

Figure 5.5: Power of several consistency tests: (a) Under the null hypothesis, as a function of σ_2, assuming $\sigma_1 = 1$ (b) As a function of the slope, assuming equal variances (c) As a function of the slope, assuming $\sigma_2 = 2\sigma_1$, (d) As a function of the slope, assuming $\sigma_2 = \sigma_1/2$

It is seen that the traditional F-test performs miserably in terms of its level and can of course not be used for testing consistency in even the simplest of models when the variances are unequal, as will normally be the case. Most notably, the traditional F-test fails when σ_2 is low (or n_2 is high) and this could easily happen in real situations. The reason for this failure is probably the fact that when the model itself is incorrect and σ_2 is low, the estimates of β are based on data set 2 and the common estimated variance is low. Therefore the large discrepancy observed due to a higher σ_1 is attributed by the statistic to a discrepancy in β since there is no allowance for heterogeneous variances.

The proposed F_c statistic is therefore the only one which can potentially be used, among the three considered here. The other panels of Fig. 5.5 indicate the power behaviour of this statistic.

This form of analysis can be used to e.g. estimate only the T years of recruitment by upweighting a set of survey indices (e.g. ages 1, 2 and 3+ giving $3T$ data points) and comparing this to what one obtains by estimating the entire parameter vector from the rest of the data. Here, $r_1 = T$, $n_1 = 3T$ and $r_2 = p$. In this case the approach will give the relative weights to the survey data as compared to the remaining data sets. Repeated minimisations such as those obtained in Taylor *et at*, 2004 give the kinds of outputs required for the analyses presented in this paper.

Discussion

The work presented in this paper has been undertaken in order to shed some light on how to proceed with complex models which are fit to several data sets. It is certainly feasible in many cases to estimate weights to be given to each data set but formal tests have been developed in the present paper to verify internal model consistency with regard to the various data sets. The F_c statistic can in principle be used to test whether a specific part of the data provides the same estimates as “the remainder” of the data.

Simulations conducted indicate that the proposed test performs well in some cases and appears to hold the nominal α-level, under the assumption that “the remainder” of the data provides more information than the part being tested.

References

Neter, J., Kutner, M. H., Nachtsheim C. J. and Wasserman, W. (1996). *Applied linear statistical models.* McGraw-Hill. Boston. 1408pp.

Randles, R. H. and Wolfe, D. A. (1979). *Introduction to the theory of nonparametric statistics.* John Wiley and Sons, New York. 450pp.

Scheffe, H. (1959) *The analysis of variance.* John Wiley&Sons. New York. 477 pp.

Stefansson, G. and Palsson, O. K. (eds.) 1997. *Bormicon: A Boreal Migration and Consumption Model.* Marine Research Institute technical report. 253 pp.

Stefansson, G. (1998). *Comparing different information sources in a multispecies context. In: F. Funk et al. (eds.). Fishery Stock Assessment Models: Proceedings of the 15th Lowell Wakefield Fisheries Symposium.* Anchorage 1997.

Stefansson, G. (2004) *Issues in multispecies models. Natural Resource Modelling.* 16(4): 415–437.

Taylor, L. A., Begley, J., Kupca, V and Stefansson, G. (2004) *A simple implementation of Gadget for cod in Icelandic waters.* ICES C.M. 2004/FF:23.

5.8 Further published/in press papers

Fleksibest – an age-length structured fish stock assessment tool with application to Northeast Arctic cod (*Gadus morhua* L.).

Frøysa, K. G., Bogstad, B., and Skagen, D. W

Abstract: In this paper we describe a state-space matrix formulation for an age-length-structured single species fish population model. The aim is to focus on how such a model can be described by formulating the relevant biological processes as matrix operators. First, a very simple model is described, including only mortality (survival), growth (in length) and recruitment. The model is then extended to include maturation and transition between different biological stages. The model is also extended to include multi-area and migration between areas. It is shown that much of the basic structure is kept when new features are added. In doing this we make explicit the large-scale structure of such a fish population model, independent of the precise details of the formulation of the equations governing biological processes such as maturation and growth.

All the biological processes are assumed to be length dependent in this paper. It should be noted that although we here present the model in terms of length-based biological processes the structure can readily accommodate the addition of age-based processes by incorporating pre-multiplied matrices into the formulation, thus allowing for full generality in an age- and length- structured model.

The matrices are assumed to be linear operators in this paper, but extension to non-linearity is possible. This can be done in a similar way as the non-linear extensions to the Leslie matrix framework.

Frøysa, K. G., Bogstad, B., and Skagen, D. W. 2002. Fleksibest – an age-length structured fish stock assessment tool with application to Northeast Arctic cod (*Gadus morhua* L.). Fisheries Research 55: 87-101.

Chapter 6

Optimisation

Optimising Likelihood Functions in Nonlinear Statistical Models

Modern models in fishery science tend to be of a statistical nature, typically involving estimation of parameters through the minimisation of negative log likelihood functions where the “mean function” is a highly nonlinear function of the vector of parameters. The term “mean function” is basically the same here as used in the theory of generalised linear models (e.g. McCullagh and Nelder, 1983) and the likelihood functions are used to describe the probability distribution of the data around the specified mean function. Given that these models do not fit into the relatively simple framework of (generalised) linear (or additive) models, the various methods developed for estimation commonly used in statistical packages will rarely apply.

When using the Gadget framework for statistical estimation of parameters in a multispecies, multiarea framework the user may develop models of arbitrary complexity. As the complexity increases so will the number of parameters to be estimated and also the number of data sets used. The issues which arise are not at all unique to Gadget, however, though here they are only considered in that context.

Numerical difficulties abound when estimating parameters of such models. Examples of difficulties due to nonlinearities can be found in section 6.1 of this report and difficulties due to multimodality in section 5.1. Naturally, analytical differentiation is not feasible and hence gradient-based nonlinear minimisation methods such as BFGS or DFP (Bertsekas, 1999) are not directly applicable.

However, when the response surface is unimodal numerical differentiation can be used in principle. It is well known, however, that scaling of parameters is important even in methods which should in principle be self-scaling (see section 6.1). It is therefore often important to estimate parameters in the correct order or to re-scale the parameters in order to have them all of the same magnitude. In addition, several methods have been developed for initial scaling of the gradient, e.g. using initial numerical estimates of the Hessian matrix, as described in Begley (2004).

Early model versions, such as several implemented using BorMiCon, resulted in non-differentiable likelihood functions. For this reason algorithms such as the Hooke and Jeeves method (Hooke and Jeeves, 1961) have been implemented. In order to handle potentially multimodal situations, but also to insist on more robust searches of the parameter space, a global search algorithm Simulated Annealing (Corana et al, 1987) has been implemented.

As they increase in complexity, such models become quite computationally intensive, and this is exacerbated when repeatedly evaluating the models during minimisation, particularly when using global searches. Methods have therefore been developed for estimating parameters using parallel versions of these various minimisation algorithms, as described in section 6.2, using PVM (Geit et al, 1994). Needless to say, these make an enormous difference, particularly when run on large computer networks.

References

Bazaraa, M. S., Sherali, H. D. and Shetty, C. M. 1993. Nonlinear Programming. Theory and Applications. John Wiley and Sons. 638pp.

Begley, J. 2004 Gadget User Guide. In prep.

Bertsekas, D.P. 1999. Nonlinear Programming. Athena Scientific 2^{nd} edition, 777pp.

Corana, A., Marchesi, M., Martini, C., and Ridella, S. 1987, Minimising Multimodal Functions of Continuous Variables with the 'Simulated Annealing' Algorithm. ACM Transactions on Mathematical Software vol 13: pp262-280.

Geist, A., Beguelin, A., and Dongarra, J. 1994. PVM: Parallel Virtual Machine: A Users' Guide and Tutorial for Network Parallel Computing (Scientific and Engineering Computation). MIT Press.

Goffe, W.L., Ferrier, G.D., and Rogers, J. 1994. Global Optimisation of Statistical Functions with Simulated Annealing Journal of Econometrics. vol. 60: pp65-100.

Hooke, R., and Jeeves, T.A. 1961. Direct Search Solution of Numerical and Statistical Problems. ACM Journal. vol 8: pp212-229.

McCullagh, P. and Nelder, J. A. 1989. Generalised Linear Models, 2^{nd} ed. Chapman and Hall, London.

6.1 Testing the optimisation routines in Gadget using simulated data.

Lorna Taylor[1] **and James Begley**
Marine Research Institute, Iceland

Abstract: This paper describes testing of optimisation in Gadget using data output from the R simulation program described in section 4.5. Likelihood surfaces for selection and growth are also presented.

Introduction

An important feature of Gadget is the link between the modelled population and samples from the observed population. These observation models allow for statistical testing of the parameters or for the parameters to be estimated within the Gadget framework.

To test the ability of the Gadget optimisation routines to converge to the correct solution, data are required from a simulated population where the parameter values are known. These ‘true’ data can then be subject to varying degrees of error to test the sensitivity of the model to data quality.

In order to generate these data, an independent simulation version of Gadget is required. The simulation model used was written in R and is described in Sigurdardottir and Olafsdottir (2005) (section 4.5).

The Gadget simulator

Overview/program

The Gadget simulator replicates some of the more commonly used features of Gadget, with use of some of these features optional. In addition to replicating the Gadget stock calculations, the program generates input files and likelihood data files for Gadget. The most complex model possible consists of two substocks, with the younger substock maturing into the older substock at a given age and timestep, ie the maturation process is not modelled. Both substocks are subject to growth and natural mortality and live on two areas. Two fleets are possible, eg one representing the commercial fleet and the other an annual survey, with the timesteps on which these fleets operate optional. There can be migration between the areas. The mature substock can also predate upon the immature stock. As with Gadget, the fleets are modelled as predators. The order in which the population processes are modelled on each timestep is identical to that in Gadget.

As Gadget is a forward simulation model, it is started with an initial population. The length distribution of each age group of the population is assumed to be Normally distributed with a given mean and standard deviation (parameters associated with the initial population can be optimised). The youngest age group in each year, in implementations without the closed lifecycle, is entered into the population in a similar manner. Differences between R and C++ (C++ approximates the probability function of the Normal distribution, whereas R uses integration to obtain the most accurate representation) require these initial populations to be entered from a file, rather than calculated, to ensure compatibility between the R model and Gadget. While this is an option within Gadget, it is not the standard method.

---

[1] An earlier version of the report was partly funded by the EU project EFEP (Q5RS–2001–01685)

Model specification

The model runs for 10 years, with 4 timesteps, of equal length, in each year. There are two areas, on which two fleets operate: one which is considered to be commercial and operates on each timestep and a survey which only fishes on the first timestep of each year. The stock consists of two substocks, one exists from age 1 to age 3 and moves into the other stock in the final timestep of its third year. The older substock exists from age 4 to age 10 and predates on the younger substock. All processes are modelled on a length interval of 1, with the length of the stock ranging from 5 to 90. The growth parameters are the same for both substocks and both are caught by both the commercial fleet and survey.

The growth function

The simplest Gadget growth function, `lengthvbsimple`, is used in the simulation model, where the mean length increase in each timestep is given by ΔL_i.

$$\Delta L_i = (q_0 - L_i)\left(1 - e^{-q_1 \Delta t}\right)$$

and q_0 and q_1 are the parameters which are equivalent to L_∞ and κ in the von Bertalanffy growth function.

The other growth functions are more complex and as they use Taylor approximations there are compatibility problems between the C++ and `R` methods of calculation. As with Gadget, mean length growth is calculated and the length distribution updated using the beta-binomial distribution (section 4.2). The beta-binomial distribution is controlled by three parameters: the mean length growth (ΔL_i) from the growth function, a fixed limit on the number of growth length groups (n) and a parameter (β) which controls the spread of the distribution. The ability to generate a wide range of distributions from estimating β alone makes this a very flexible distribution.

Parameter estimation

To test the robustness and efficiency of parameter estimation in Gadget, the simulation program can be used to simulate the population dynamics. The modelled population can then be sampled to generate likelihood data with which the optimization routines in Gadget can be tested by comparing the parameters calculated from the optimization with the known parameters. Error can also be added to the sampled data. Fleet suitability and recruitment parameters were estimated, both separately and simultaneously.

Recruitment

The number of the youngest age group entering the population in each year (which will be referred to as recruitment) can be estimated. The most influential likelihood components in estimating recruitment are survey indices. To test estimation of recruitment, two sets of survey indices were output from the `R` simulator: abundance at age 1 and abundance of the rest of the population. The age 1 index is the key source of information on the relative abundance of the recruits, whereas the index of older fish (given information on growth and mortality) helps determine the overall abundance of recruitment. These indices represent the first timestep of each year and were generated with varying degrees of error. Index data are output separately for both areas creating 2 survey index likelihood components - one for each area with the 2 age groups combined in one likelihood component.

If $N^e_{A,t}$ is the number in area A at timestep t, with error applied, then

$$N^e_{A,t} = e^{\left(\varepsilon - \frac{\sigma^2}{2}\right)} \sum_l \sum_a N_{l,a,t,A},$$

where $\varepsilon \sim N(0, \sigma^2)$ and $N_{l,a,t,A,}$ is the number of length l and age a in area A in timestep t.

In this example, values of σ were 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1. For each value of σ there are 20 numbers (10 years and 2 areas) for each index and these are illustrated in figure 6.1. For each value of σ an optimizing run was conducted, in which the total number of age 1 fish at the start of each year in each area was estimated.

The value of recruitment used to generate the data was 500 000 in each area and year. Runs were conducted with two sets of initial parameters, the first within 100% of the true value and the second within 1000% of the true value.

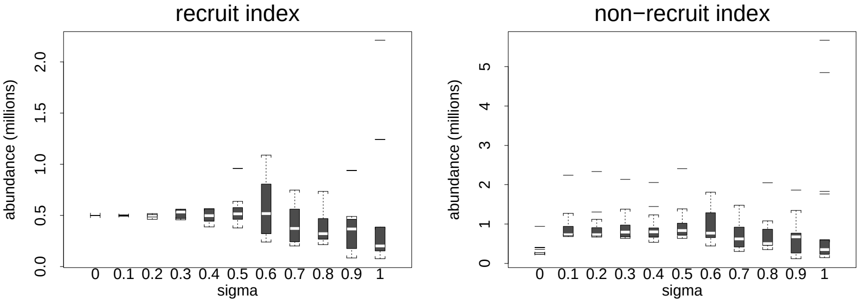


Figure 6.1: Boxplots of the likelihood data representing survey indices calculated from the R simulation model.

Catch selection pattern

The fleet suitability function has logarithmic dependence on the length of the prey i and the predator L :

$$S(l, L) = \frac{\delta}{1 + e^{-\alpha - \beta l - \gamma L}}$$

In the R simulator γ = 0 removing the dependence on predator length and δ = 1 fixing the maximum suitability to 1 with α and β defined by the user.

Length distributions are the key source of information to estimate these parameters. A proportion of the length distribution caught by the commercial fleet for each timestep and year can be output from the R simulator with the aggregation by length defined. The likelihood data used are equivalent to sampling 0.1% of the commercial catch and error was added as for the survey indices. The data were aggregated either in 1, 5, or 10cm length groups.

If $N^{e,p}_{l,A,t,f}$ is a proportion of the length distribution from the catch at length l, area A, timestep t and fleet f with error then,

$$N^{e,p}_{l,A,t,f} = \rho e^{\left(\varepsilon - \frac{\sigma^2}{2}\right)} \sum_a N_{l,a,t,A,f}$$

where $\varepsilon \sim N(0, \sigma^2), 0 \leq \rho \leq 1$ and $N^{e,p}_{l,A,t,f}$ is the number of length l and age a in area A caught by fleet f in timestep t.

Four sets of runs were conducted: 1) estimating both α and β of the selection pattern with likelihood length distributions aggregated on 1cm intervals 2) α fixed and estimating β with data on 1cm intervals 3) α fixed and estimating β with data on 5cm intervals and 4) α fixed and estimating β with data on 10cm intervals. A sum of squares in 1cm intervals was calculated as a measure of the difference between the original suitability curve and those estimated.

Objective functions

Two different objective functions were used, a sum of squares for length distributions and a log linear regression for survey indices. The sum of squares compares the length distribution as proportions from the model with the length distribution as proportions from the sampled data:

$$\ell = \sum_{time} \sum_{areas} \sum_{ages} \sum_{lengths} \left(p_{tral} - \pi_{tral}\right)^2$$

where p is the proportion in the data in that time/area/age/length group and π is the proportion in the model in that time/area/age/length group. The survey index objective function compares the survey index with the model index:

$$\ell = \sum_{time} \left(\ln\left(I_t\right) - \left(\alpha + \beta \ln\left(N_t\right)\right)\right)^2$$

where $I_t$is the observed index and N_t the index of the modelled population. In the following analyses the intercept (α) was estimated and the slope (β) fixed equal to 1.

Recruitment and selection

Recruitment and the fleet selection pattern were also estimated simultaneously. This was done in two ways 1) with equal weights on all likelihood components and 2) with the appropriate weights calculated using the weighting protocol described in Sections 5.4 & 9.3. For the first method the likelihood data were identical to those used when estimating one type of parameter. For the second method, the indices were split by age with the recruitment index a separate likelihood component from that representing the rest of the data and another set of index data were generated. Two sets of indices are required to calculate the weights as explained in Section 9.3. The second method was only used for one level of variance with $\sigma = 1$. In both cases all parameters were estimated simultaneously.

Optimisation specifications

All Gadget optimisation runs were run until the optimising criteria were met, ie the runs did not stop because the maximum number of iterations/function evaluations had been completed. Both Simulated Annealing and Hooke and Jeeves were used in all cases.

Simulated annealing does not scale the parameters which means that the step length adjustment factor is not equally suitable for parameters of different orders of magnitude. For this reason, when more then one parameter is estimated the parameters are scaled to a similar order of magnitude.

With more than one likelihood component ie when estimating recruitment and recruitment and selection simultaneously, each likelihood component was given an identical weight. This does not mean that each component has equal influence in the estimation (see section 5.4).

The same optimisation settings were used for all runs estimating the same parameters and were as follows:

Estimating fleet selection:

Simulated annealing

max iterations	halt criterion (ϵ)	starting temperature
30 000	0.00001	20

Hooke and Jeeves

max iterations	min step length (ϵ)
20 000	1e-6

Estimating recruitment:

Simulated annealing

max iterations	halt criterion (ϵ)	starting temperature
30 000	0.0001	100

Hooke and Jeeves

max iterations	min step length (ϵ)
20 000	1e-5

Estimating selection and recruitment (method 1):

Simulated annealing

max iterations	halt criterion (ϵ)	starting temperature
30 000	0.0001	100

Hooke and Jeeves

max iterations	min step length (ϵ)
20 000	1e-4

Estimating selection and recruitment (method 2):

Simulated annealing

max iterations	halt criterion (ϵ)	starting temperature
30 000	0.1	7500

Hooke and Jeeves

max iterations	min step length (ϵ)
20 000	1e-4

Estimation results and discussion

Recruitment

In figure 6.2 it can be seen that the initial values have little influence on the final parameter estimation, especially at lower levels of sampling error. The likelihood score is also similar for both runs at all values of σ except the largest (figure 6.3). As the error (σ) in the index increases, the total likelihood score increases as shown in figure 6.3 and the variance of the estimated recruitment increases but is always less than the input variance (figure 6.4).

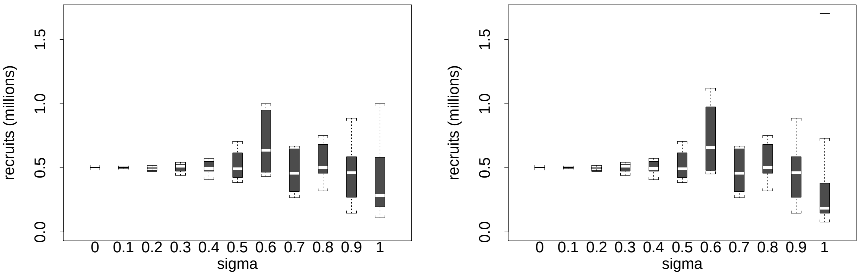


Figure 6.2: Boxplots of parameter estimates with different values of σ from two different sets of initial values.

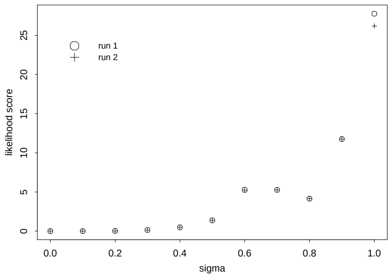


Figure 6.3: Total likelihood scores for each value of σ

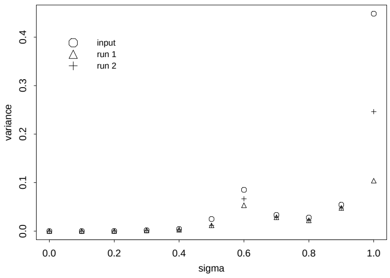


Figure 6.4: Variance of the input and estimated recruitment.

Fleet selection

The values of α and β used to generate the data were -0.45 and 0.2 respectively. The estimated values of α and β where both parameters were estimated are shown in figure 6.5 and the estimates of β from the runs with α fixed in figure 6.6.

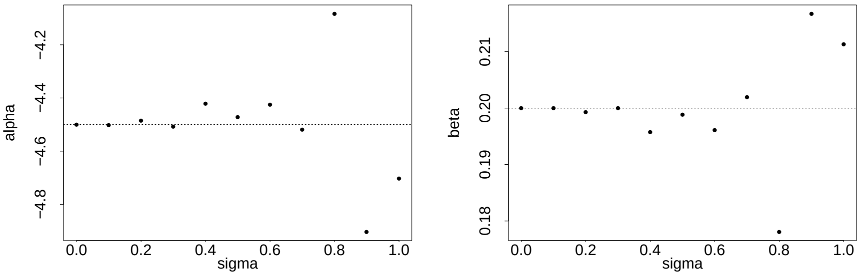


Figure 6.5: Estimated values of the α and β parameters of the fleet suitability function with both α and β estimated.

The greatest differences between the true and estimated curves are when both α and β are estimated or when the likelihood data are more aggregated (figure 6.7). Comparing the true curve with those estimated in these cases shows that the estimated curve is very close to the true curve (figures 6.8 & 6.9). As the amount of error in the likelihood data increases the total likelihood score increases (figures 6.10 & 6.11) and aggregating the length distribution results in poorer estimates of β (figures 6.6 & 6.7).

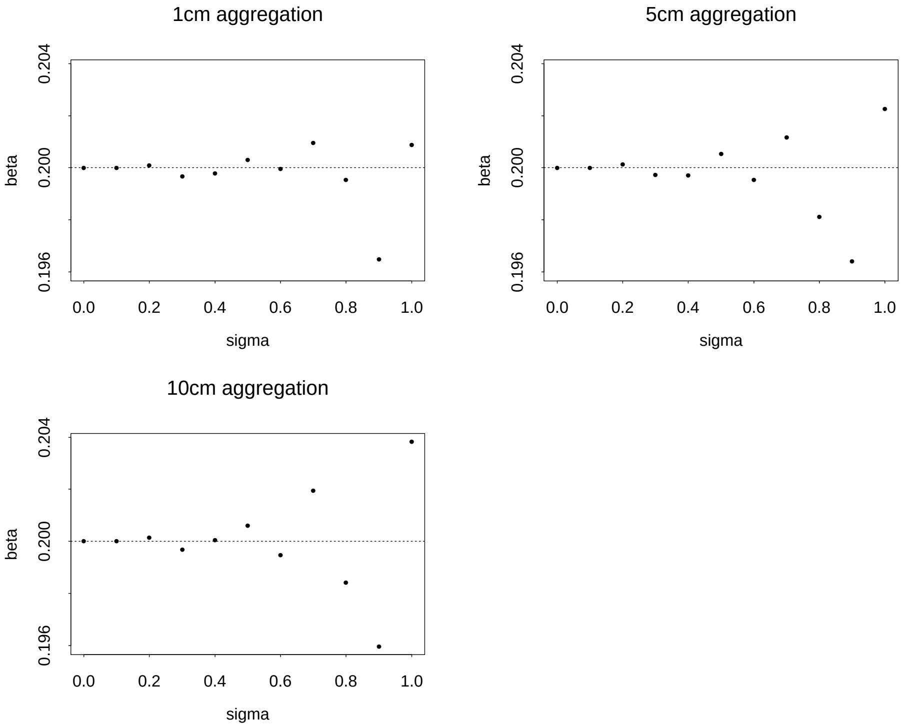


Figure 6.6: Estimated values the β parameter of the fleet suitability function with the length distribution aggregated on 1cm, 5cm and 10cm intervals.

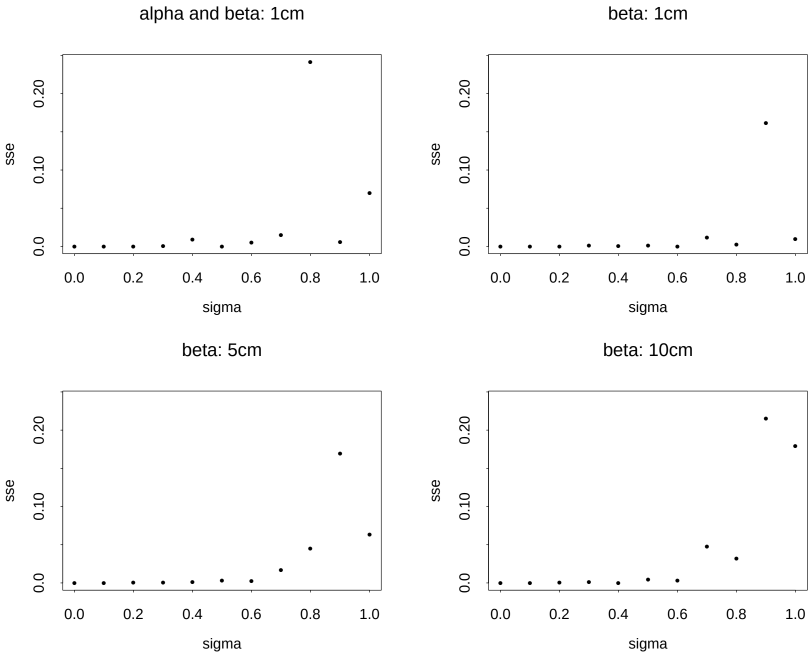


Figure 6.7: Sum of squares by σ for the four set of model runs.

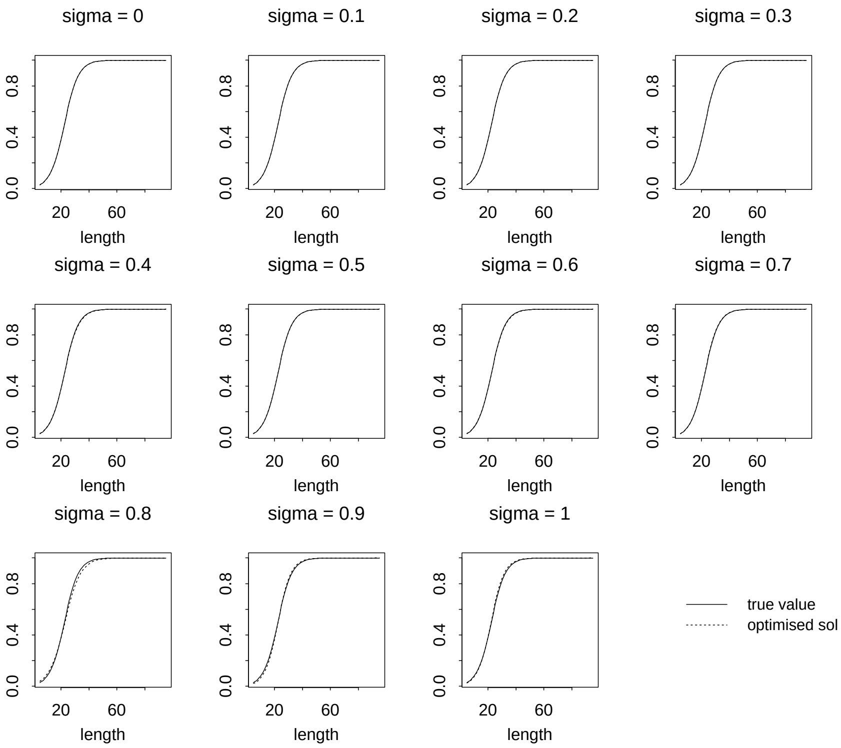


Figure 6.8: Estimated fleet suitability curves: α and β estimated, with likelihood data aggregated on 1cm intervals. The true line is solid and the optimized dashed.

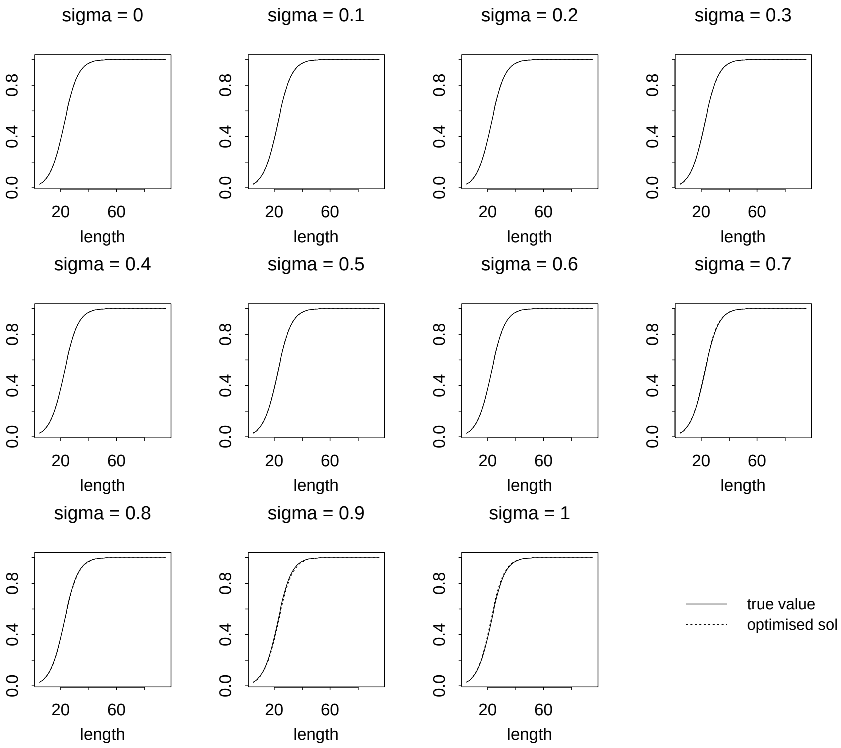


Figure 6.9: Estimated fleet suitability curves: α fixed and β estimated, with likelihood data aggregated on 10cm intervals. The true line is solid and the optimized dashed.

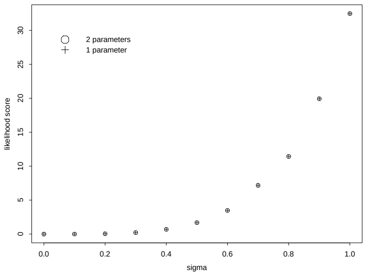


Figure 6.10: Total likelihood score plotted against σ for the 2 cases where the data were aggregated on 1cm intervals.

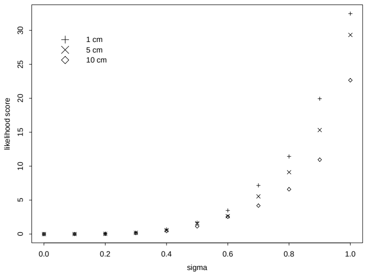


Figure 6.11: Total likelihood score plotted against σ for the 3 cases with α fixed and data aggregated on 1cm, 5cm and 10cm intervals.

Recruitment and selection (method 1)

With fewer fixed parameters (and ad hoc weighting of likelihood components) estimation is less successful. The estimated β and recruitment values are plotted in figures 6.1 & 6.1 respectively.

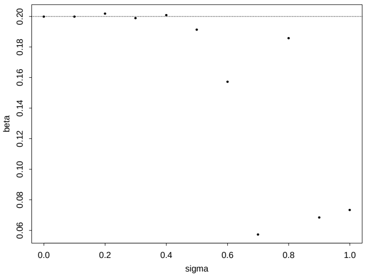


Figure 6.12: β estimated at different levels of σ with both recruitment and selection estimated.

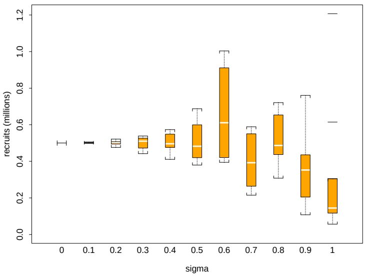


Figure 6.13: Recruitment estimated at different levels of σ with both recruitment and selection estimated.

The estimated selection patterns are, however, in most cases still very close to the true selection curves (figures 6.14 & 6.15). The selection and recruitment parameters are confounded and as the length at 50% selected increases the number of recruits entering the population decreases. As before, the amount of sampling error in the likelihood data increases, the total likelihood score increases (figure 6.16).

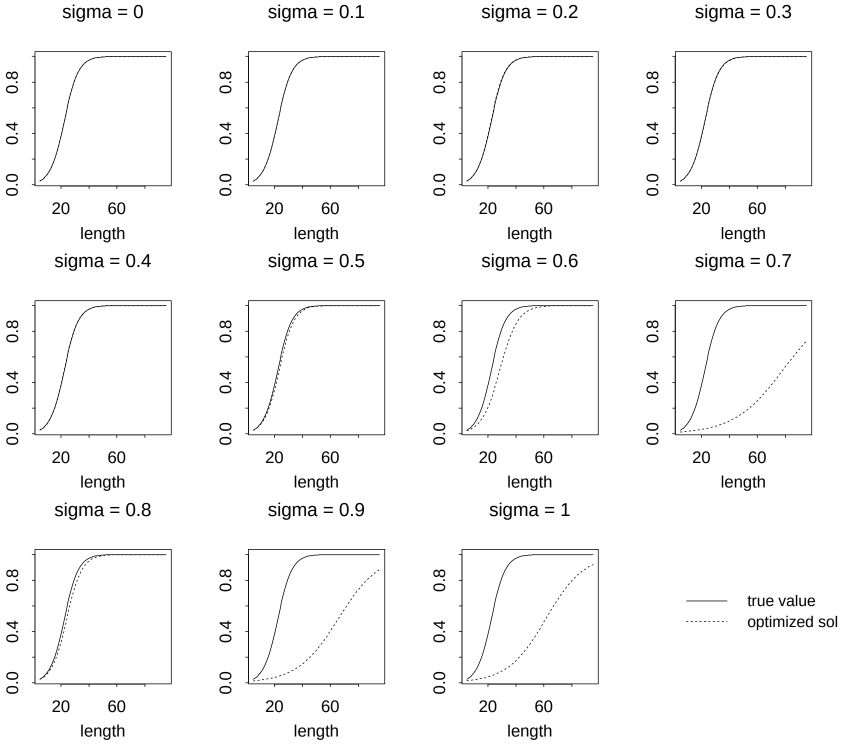


Figure 6.14: Estimated fleet suitability curves, estimated from likelihood datasets with sigma varied. The true line is a solid and the optimized dashed.

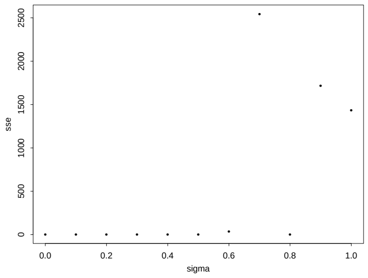


Figure 6.15: Sum of squares by σ for the fit between the true and estimated selection patterns.

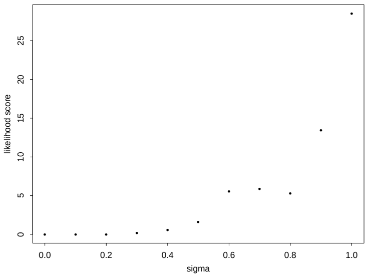


Figure 6.16: The total likelihood score plotted against σ.

Recruitment and selection (method 2)

Comparing the results from method 1 with $\sigma = 1$ with those from method 2, it can be seen in figures 6.1 & 6.1 that appropriate weights are essential when more than one likelihood dataset is used. For both selection and recruitment the values estimated with the calculated weights are considerably closer to the true values than those with the weights set to 1. The values of recruitment estimated from both methods are shown in figure 6.1 (with the recruitment indices in figure 6.1.

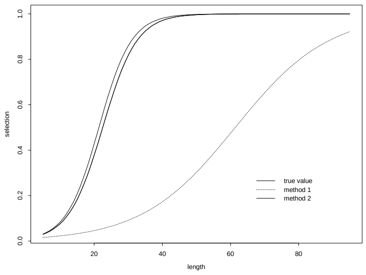


Figure 6.17: β estimated at different levels of σ with both recruitment and selection estimated.

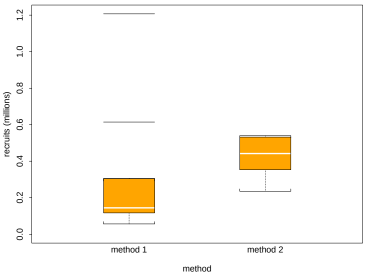


Figure 6.18: Recruitment indices generated with $\sigma = 1$. The true level of recruitment is indicated by the solid line.

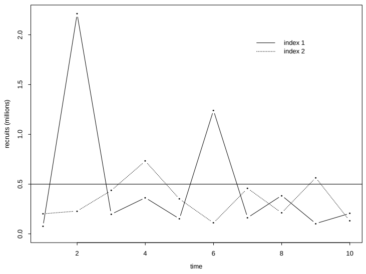


Figure 6.19: Recruitment estimated by the 2 methods. 1) with all weights = 1 and 2) with appropriate weights calculated. The true level of recruitment is indicated by the solid line. Recruitment was estimated simultaneously with selection.

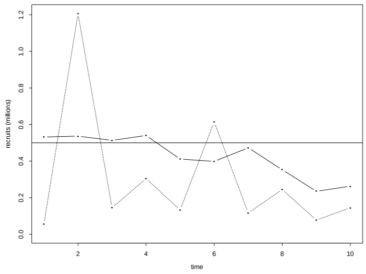


Figure 6.20: Recruitment estimated with $\sigma = 1$ with weighting according to methods 1 and 2. with both recruitment and selection estimated.

Likelihood surfaces

Parameters are often related and in some cases closely related, eg the suitability function and growth function parameters, which means that altering one results in a different optimum for the other. It is of interest to consider the likelihood surface associated with such parameters. It may also be desirable to reformulate the parameters to generate a surface more amenable to optimisation, or to use a different objective function. Alternatively, the likelihood surface may indicate that it is sufficient to optimise only one of a pair of parameters.

Method

To plot the likelihood surface for the selection parameters the parameters of the suitability function are varied simultaneously with all other parameters constant. The only likelihood data component is the catch length distribution with no error. The range over which α and β are varied is from -10% to +10% of the true value in steps of 0.25%.

Equivalent surfaces for the growth parameters were generated using length distribution and age-length distribution likelihood data.

Results and discussion

Fleet selection

When α and β are both varied, the surface (figures 6.21 & 6.22) has a trench with steep sides which indicates the parameters are correlated. As the gradient is very shallow along the bottom of the trench, any point on the bottom of the trench is therefore close to the minimum. Moving along the trench, however, requires a very short step length. There is therefore little to be gained in estimating both parameters and estimating one is sufficient (and faster) as long as the other is fixed to a suitable value.

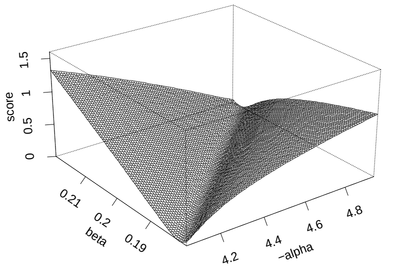


Figure 6.21: Likelihood surface for the selection pattern parameters as a perspective plot.

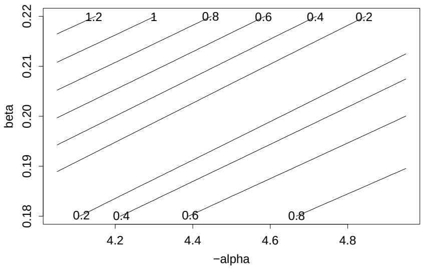


Figure 6.22: Likelihood surface for the selection pattern parameters as a contour plot.

Growth

When both growth function parameters (L_∞ and k) are varied, the surface (figures 6.23 & 6.24) also has a trench which is more steep sided towards lower values of L_∞ then towards higher values. As with selection, the parameters are correlated and there is little to be gained from estimating both parameters.

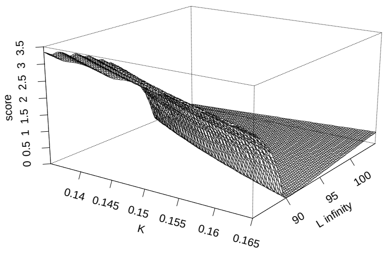


Figure 6.23: Likelihood surface for the growth parameters as a perspective plot generated using length distribution data.

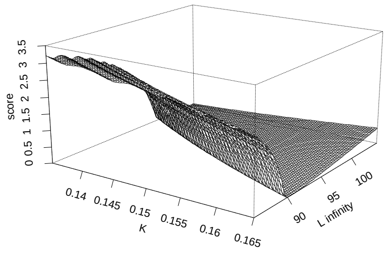


Figure 6.24: Likelihood surface for the growth parameters as a perspective plot generated using age-length distribution data.

6.2 Paramin, a Composite Parallel Algorithm for Minimising Computationally Expensive Functions

Gunnar Stefánsson[1,2], Auðbjörg Jakobsdóttir[1], Bjarki Þór Elvarsson[1], and James Begley[1]

[1]**Marine Research Institute, Reykjavik, Iceland**
[2]**University of Iceland, Reykjavik, Iceland**

Abstract: Minimisation of computationally expensive (i.e. time-consuming) functions of many parameters tends to be hampered by available computing power. Examples of such functions in biometrics include outputs from algorithms which conduct deterministic simulations for a given set of input parameters, and each simulation gives rise to a value of a likelihood function. This paper describes a parallel implementation of a composite algorithm derived from the Simulated Annealing algorithm, the Hooke & Jeeves algorithm and the BFGS algorithm. The composite algorithm is based on computing numerical derivatives, line searches and coordinate searches in parallel, through the use of multiple processors or computers, by computing each function value on a single processor.

Introduction

The problem of minimising complicated nonlinear functions is quite common. The present paper describes an algorithm for undertaking such minimisations by using a number of Unix or Linux workstations on a local-area network.

Although the problem description is generic in nature, the development of the algorithm was motivated by complicated likelihood functions of many variables, typically 50–500, as used in the class of ecosystem models described in Stefánsson and Pálsson (1998). Such a likelihood function is a sum of likelihood components corresponding to output from a multi-year simulation of a marine ecosystem. The arguments to the function are unknown parameters, to be estimated statistically, though some parameters may be kept fixed. It is not uncommon that computation of a single function value may take up to a full minute on a fast workstation and no analytical derivatives are available. Given this setting, a single non-symmetric numerical gradient evaluation can commonly take a full hour on a single CPU.

Thus there is a need to somehow distribute the single task of obtaining

$$min\{f(x) : x\epsilon R^n\} \tag{6.1}$$

where $f(x)$ may not be unimodal or continuously differentiable everywhere, but will be assumed to be continuous everywhere and twice continuously differentiable near a unique global minimum.

Several programs have been developed to implement parallel or distributed computing. Some of these are designed to allow a user to send a single difficult task to a powerful machine. In the present setting this is clearly not the appropriate solution, since it is easy to get a large number of processors which have total computing power far exceeding that of a single processor for the same price. For example, when evaluating the gradient numerically, for large n, it is trivial to distribute the computation of $f(x + he_i)$ and of course one will easily obtain an almost linear speedup since there is very little cost involved in data transfer between processors, even on a fairly slow network.

Considerable development and testing of parallel versions of gradient-based or quasi-Newton algorithms has taken place and it has been found that it is possible to obtain close to linear increases in performance by implementing different variants of these methods on different processors, possibly combining estimates every few iterations.

A fundamental problem with this approach in the current problem setting is that a single gradient evaluation will take a long time on each computer. As an example, if each function evaluation takes about

1 minute and numerical gradient computation for 60 variables can be spread over 60 processors, then a total of 60 gradient evaluation and in fact 60 quasi-Newton steps could be taken in one hour. If different gradient-based methods are used on each machine, they will only have completed one gradient evaluation and one quasi-Newton step each during the same amount of time.

Alternatively, threads or other methods can be used in order to distribute a single function evaluation across CPUs. This approach requires the application itself to be modified, highly specialised for parallelisation and non portable and will therefore not be considered further.

Other parallel methods for minimisation have emphasised using parallel techniques internally in the minimisation algorithm, e.g. as a part of obtaining matrix decomposition. This is clearly not the bottleneck in the present situation where computations within the minimisation algorithm are dwarfed by to the time taken for function evaluations.

The approach followed for this paper, therefore, will be to implement a minimisation algorithm where the function evaluations are distributed across a computer network, evaluating each single function call on a single processor. The minimisation algorithm can place requests for several function values at once and the time required will not depend on the number of requests as long as they are no more than the number of available processors. It is therefore an important design issue never to request only a single function value, but always to request enough function evaluations to keep all processors busy, as long as any function value is needed.

The remainder of the paper describes the implementation of an algorithm designed to make use of the possibility of obtaining simultaneous function evaluations on several computers.

Algorithm

Given that there is no guarantee ahead of time that the function is unimodal, there is a need to involve and test a global optimiser, at least to ensure that the optimiser will get to the vicinity of the minimum. The Simulated Annealing algorithm has been chosen here for this purpose.

The functions in the problem of primary interest may have some problem areas such as non-differentiability far away from the optimum. Since the primary problem situation corresponds to simulations of an ecosystem, the likelihood function may have special problems for ”strange” parameter values, due to certain modelled species becoming extinct etc. This is one reason for including a minimiser which does not require differentiability. Several such methods are available but the Hooke & Jeeves algorithm has been used in the combined algorithm.

However, close enough to the optimum the usual properties of likelihood functions should apply so it is expected to exhibit behaviour similar to a quadratic function. In addition, in statistical estimation there is general interest in obtaining the inverse Hessian matrix at the optimum. For this reason, as well as to obtain local convergence properties, the BFGS algorithm has been incorporated into the combined algorithm.

The performance of the combined algorithm is determined by the relative number of function calls allowed, or the accuracy required, for each method. These pairs of numbers are therefore defined as the primary parameters in the combined algorithm.

The parallel processing mechanism for all the methods used in this combined algorithm was implemented using the freely available Parallel Virtual Machine (PVM) (Geist et al., 1994) libraries and functions, using a low-level C++ interface that was developed as part of the work for this study.

Simulated Annealing Overview

Simulated Annealing (SA) (Corona et al., 1987) is a global minimisation algorithm. From an initial starting point the algorithm takes a random step in various directions, by adjusting one (random) parameter at a time, and evaluates the function at each new point. If the new function value is lower than the old one then the algorithm stores the new point as it's best guess, and re-starts from this new point. However, if the function value is higher then the algorithm may still accept this point, based on the probabilistic ”Metropolis Criteria”, and thus the algorithm can escape from a local minimum. The algorithm exits when a stable point is found which cannot be improved on with a small step in any direction, and the Metropolis Criteria rejects all the steps away from the current best point.

The Metropolis Criteria will accept a move to a point with a higher function value based on a function of both the size of the move and a parameter termed ”temperature”, as is given in the equation below:

$$\begin{aligned} M &= e^{\dfrac{-\Delta F}{T}} \\ P &= \begin{cases} 1 & \text{if } M > r \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{6.2}$$

where ΔF is the change in function value, T is the ”temperature” of the algorithm, and r is a random number between 0 and 1.

Note that when the ”temperature” is very high ($T \to \infty$), the Metropolis Criteria will always accept any move, and the algorithm will simplify to a form of a random search algorithm. Conversely, when the temperature is very low ($T \to 0$), the Metropolis Criteria will always reject an uphill move, and the algorithm will simplify to a form of a local search, similar to Hooke & Jeeves. By slowly reducing the temperature of the algorithm, the number of moves that are accepted by the Metropolis Criteria are reduced and the algorithm will find a minimum.

Simulated Annealing in Parallel

It is reasonably straightforward to parallelise the SA algorithm, since the only change required is to request all parameter updates simultaneously. Once a new, improved (or rather accepted) point is found, the algorithm sets this to being the initial point and then subsequent requests are then made from this new point.

Apart from the minor changes to accommodate parallel computation by requesting all function values simultaneously, the SA algorithm used is derived from that presented by Corana et al. (1987), with the various improvements and modifications presented by Goffe et al. (1994).

It should be noted that when the algorithm accepts a new point, there may be many requests still in the SA queue, which are then outdated (i.e. not in accordance with the initial definition of the algorithm) and are likely to be rejected. This implies that the following improvement could potentially improve the SA algorithm when it is being run in parallel:

- the queue for the SA algorithm should be a LIFO queue, so that the outdated points are lost (or put on the back burner) as soon as a new point is accepted and a new suite of points requested
- the ”temperature” parameter that controls the convergence of the algorithm plays a crucial rôle in determining the efficiency of the parallel algorithm, since if a large number of new points are being accepted then most of the remaining points in the SA queue will be outdated and thus rejected, which will result in the loss of most of the increase in performance gained by parallelising the algorithm

Hooke & Jeeves Overview

Hooke & Jeeves (HJ) (Hooke and Jeeves, 1961) is a simple and fast local minimisation algorithm. From an initial starting point the algorithm takes a step in various directions, by adjusting one parameter at a time, and evaluates the function at each new point. If the new function value is lower than the old one then the algorithm uses the new point as it's best guess, and re-starts from this new point. However, if the new function value is higher, then the algorithm returns to the old point, and tries again by adjusting a different parameter. The search proceeds in series of these steps, with each step slightly smaller than the previous one, until the algorithm exits when a point is found which cannot be improved on with a small step in any direction.

Hooke & Jeeves in Parallel

The parallel version presented here is derived from the sequential implementation, with the main difference being to request as many coordinate searches as there are processors available, beginning in a single direction for each coordinate. Upon getting negative results (no function improvement), the remaining coordinates are searched and subsequently the other direction of each such negative coordinate is requested. In the case of a sequence of negative results there is no numerical difference between the parallel and sequential versions of the algorithm, except for the sequence of function evaluations.

When a positive result is obtained, the new requests must be issued from this revised intermediate point, disregarding any results that may be in the HJ queue based on a search from the previous point. This means that the stepsize parameter is important when determining the efficiency of the parallel algorithm, particularly when running on a network with a large number of processors available, since if a large number of new points are being accepted then most of the points in the HJ queue will be rejected, which will result in the loss of most of the increase in performance gained by parallelising the algorithm. It follows that the most efficient approach to take is to re-issue coordinate searches from each new intermediate best point, with a maximum number of calls for each coordinate to ensure that the current estimate of the optimum point is regularly updated.

Apart from these changes to accommodate parallel computation, the HJ algorithm used is derived from that presented by Hooke and Jeeves (1961), with the various improvements and modifications presented by Kaupe (1963).

BFGS Overview

Broyden–Fletcher–Goldfarb–Shanno (BFGS) is a quasi-Newton minimisation algorithm that uses information about the gradient of the function at the current point to calculate the best direction to look in to find a better point. Using this information, the BFGS algorithm can iteratively calculate a better approximation to the inverse Hessian matrix, which will lead to a better approximation of the minimum value. From an initial starting point, the gradient of the function is calculated and then the algorithm uses this information to calculate the best direction to perform a linesearch for a point that is ”sufficiently better”. The linesearch that has been used to look for a better point in this direction is the ”Armijo” linesearch. The algorithm will then adjust the current estimate of the inverse Hessian matrix, and restart from this new point. If a better point cannot be found, then the inverse Hessian matrix is reset and the algorithm restarts from the last accepted point.

The Armijo linesearch calculates the stepsize that is to be used to move to a point that is ”sufficiently better”, in the direction given by the gradient calculation, to be β^n, where n is the first integer that satisfies the Armijo criterion given by the following inequality:

$$f(x) - f(x + \beta^n) \geq -\sigma\beta^n \nabla f(x)^T \tag{6.3}$$

where $\nabla f(x)$ is the gradient of the function at the current point.

BFGS in Parallel

There are two main areas where the parallel version of the BFGS algorithm presented here differs from the sequential version – performing the linesearch and calculating the gradient of the function to be minimised.

When performing the linesearch in parallel, it is possible to evaluate many Armijo criteria simultaneously, for values of $n = 1, 2, \ldots, p$. Since these are tested in the sequence they are returned, the simplest approach would be to accept the first returned value to satisfy the criterion. In the case of many processors and requests this has the obvious disadvantage that there is a high probability that the first value returned will correspond to a value of n which is too high, resulting in a much smaller stepsize than could be used. Alternatively, one could wait for all the results from the first round and find the exact solution to the Armijo criterion, i.e. request a series of values of the form β^n and accept the one which satisfies the criterion and corresponds to the smallest such integer n. Naturally, waiting for all the requested values may inflict considerable delays since this may result in waiting for results from a processor which happens to be much slower than the average for some reason. A much better approach, therefore, is not to wait for all the results, but only ”most” of them. The algorithm implemented therefore waits until the first (somewhat arbitrary) 80% of all those requested are returned and selects the smallest acceptable n from those returned.

Apart from this change to accommodate parallel computation of the Armijo linesearch criterion, and the use of numerical derivatives to calculate the gradient detailed below, the BFGS algorithm used is derived from that presented by Bertsekas (1999).

Gradient Calculation

For the functions to be minimised, analytical derivatives are not always available and hence numerical differentiation is used when calculating the gradient of the function. Some care needs to be taken in the numerical evaluation of derivatives. Given the expensive nature of the function to be evaluated, the approach used here is to initially assume that it is sufficient to estimate the gradient through the components:

$$\frac{f(x) - f(x - \delta e_i)}{\delta} \tag{6.4}$$

The choice of δ is nontrivial and the approach taken is to initially set δ to a small value but at the same time to evaluate the loss of digits and to convert to a larger δ when the need is indicated. If this does not suffice, symmetric gradient evaluation is reverted to, using:

$$\frac{f(x + \delta e_i) - f(x - \delta e_i)}{2\delta} \tag{6.5}$$

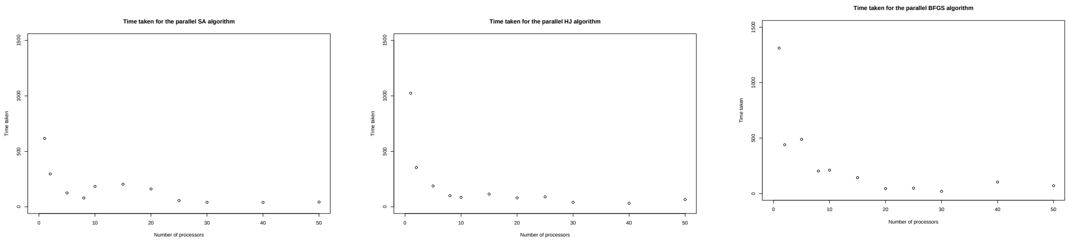
This overall approach seems to work reasonably well for the present problem, i.e. asymmetric gradient evaluation can be used while getting into the vicinity of the minimum, after which there is sometimes a need to evaluate the gradient with even more accuracy, using:

$$\frac{8(f(x + \delta e_i) - f(x - \delta e_i)) + f(x - 2\delta e_i) - f(x + 2\delta e_i)}{12\delta} \tag{6.6}$$

The requests for function evaluations are all made simultaneously. The gradient computation needs to wait for the last evaluation and this is the only such case in the entire combined algorithm. Hence bookkeeping of the speed of the different processors is important since a fast processor which is currently busy may be able to return a requested function sooner than an slower processor, even if the slower processor is currently idle and available.

Results and Discussion

The parallel versions of the algorithms have been tested using the network version of Gadget, running the well-defined haddock example model presented in Anon (2003) that is available for download from the Gadget website (`www.hafro.is/gadget`). The tests were performed on a Class I Beowolf Cluster, comprising of 130 2.8 GHz PC's, each running Fedora Linux Core 1. Each algorithm run started from the same point, chosen such that each of the three algorithms could converge. The time taken for each of the three algorithms to converge, when running on different numbers of processors, is shown in figure 6.2. All three algorithms appear to scale reasonably well, with the time taken to find a point that meets the convergence criteria of the algorithm inversely proportional to the number of processors used.



(a) SA (b) HJ (c) BFGS

Figure 6.25: Results from a speed test on 1 to 50 processors for the parallel versions of the 3 algorithms

A comparison between the relative performance of the three algorithms is shown in figure 6.26, scaled so that the time taken to run the algorithm on a single processor is 1000 seconds. It should be noted that this comparison can give a slightly mis-leading picture of the relative performance of the algorithms, since the convergence criterion differs between algorithms. SA is a global search algorithm, so the convergence criterion is looser than that for the other two algorithms.

An obvious hindrance in the scaling is when the number of processors available exceeds the number of parameters to be optimised - the example used here has 37 parameters, and it can be seen that there is little improvement once this number of processors has been reached. All the algorithms used here could be adapted to make better use of the extra processors if they are available, but this would require better book-keeping to keep track of the requests that have been sent. However, in practical terms, it is unlikely that there will be many optimisations where the number of processors available is more than the number of parameters to be optimised, so this has not been investigated further.

The SA algorithm selected here requires an order of magnitude more function calls for the convergence criterion to be satisfied than either of the others, which is not surprising since it is a global search algorithm. However, for the composite algorithm, all that is required from the SA component is that the search area for the local searches that follow will be in the region of the optimum, without actually finding the optimum. By adjusting the parameters downwards from those recommended by Corana et al. (1987), it is possible to get the SA algorithm to move the search area to the general region of the optimum reasonably quickly. Once the algorithm is close to the optimum, SA is inefficient, so choosing the point to stop using SA and start using one of the other algorithms requires some care to ensure that the composite algorithm makes good use of the processing time available.

The HJ algorithm is the simplest algorithm used here, and, partly due to its simplicity, it scales fairly well. This algorithm would be the simplest algorithm to scale to use more processors than parameters, by taking a step in both the positive and negative directions at the same time, but this would only take the limiting point up to twice the number of parameters. Since the HJ algorithm appears to perform as well as BFGS in the test cases, there is a need to evaluate whether the use of the inverse Hessian for variance estimation can be replaced by bootstrap methods in given applications.

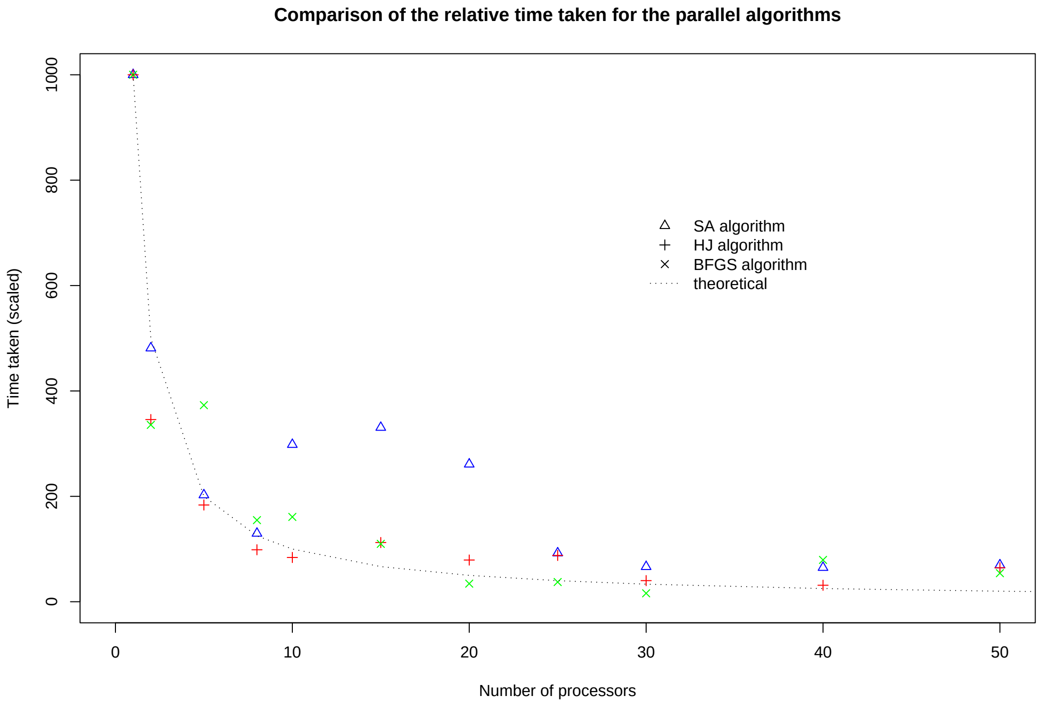


Figure 6.26: Comparison of the relative performance of the 3 parallel algorithms

Naturally, for a well-behaved problem, the BFGS algorithm is expected to behave best among the methods considered, but this needs to be tested on the particular functions of interest in a given problem. A potentially serious problem with BFGS is the requirement of a positive definite Hessian matrix as it is not at all uncommon for a Hessian matrix to be very poorly numerically determined and in many cases the effects of some parameters are negligible leading to a Hessian matrix with zero or negative eigenvalues. Possibly the use of correction mechanisms such as those used in the Levenberg–Marquardt version of the Gauss–Newton algorithm (replacing H by $H + \Delta I$ where Δ is a large enough number) might be used to alleviate this problem, but this has not been investigated.

If SA and/or HJ has been run before BFGS there are a large number of evaluations $(x, f(x))$ available before BFGS starts. These could in principle be used to get an estimate of the Hessian at the starting point for BFGS, using a judicial weighting mechanism. Alternatively, H could be initialised to $diag(h_{ii})$, where the elements h_{ii} are obtained by fitting a parabola through $(x - \delta e_i, f(x - \delta e_i))$, $(x, f(x))$ and $(x+\delta e_i, f(x+\delta e_i))$ (if the HJ algorithm was used earlier). These could have considerable potential in getting a better starting direction than the current implementation, which uses $H = I$ for the initial estimate.

In the case where the number of processors is large the Armijo linesearch algorithm for BFGS described above is not ideal, due to the discarding of 20% of the results. This might result in the choosing of a point which would firstly not have been chosen by the sequential Armijo (which is something that can happen in any case) and secondly, and more importantly, the point chosen might not be close to the one picked by the sequential one as they could also be discarded. Therefore the resulting stepsize could be too small and the function evaluations needed to converge will somewhat increase. There are a few obvious remedies to the Armijo linesearch, which need to be evaluated in more detail. Firstly the algorithm could be adjusted to simply decrease the discarded percentage (which will decrease the probability of picking a poor point, at a cost of increased processing time), secondly the value of β could be increased in the hope that that cures the problem and thirdly a limit on the maximum number of processors used in the Armijo linesearch could be implemented. An alternative approach that needs further consideration would be to find the minimum along the search direction, since this would cost (in terms of computing time) the same as the Armijo linesearch, but should result in better convergence.

Acknowledgements

The parallel versions of the various minimisation algorithms have been programmed mainly by Þórdís Linda Þórarinsdóttir and Kristjana Ýr Jónsdóttir, who also rewrote the SA program from the FORTRAN version presented by Goffe et al. (1994) to C++. Testing and debugging work was carried out by Daniel Howell.

References

Anon. 2003. Development of Structurally Detailed Statistically Testable Models of Marine Populations (dst^2). QLK5–CT1999–01609. Progress Report for 1 January to 31 December 2002. Marine Research Institute Report No. 98, Marine Research Institute, Reykjavik, Iceland.

Bertsekas, D.P. 1999. Nonlinear Programming. Athena Scientific 2^{nd} edition, pp22-61.

Corana, A., Marchesi, M., Martini, C., and Ridella, S. 1987, Minimising Multimodal Functions of Continuous Variables with the ’Simulated Annealing’ Algorithm. ACM Transactions on Mathematical Software vol 13: pp262-280.

Geist, A., Beguelin, A., and Dongarra, J. 1994. PVM: Parallel Virtual Machine: A Users’ Guide and Tutorial for Network Parallel Computing (Scientific and Engineering Computation). MIT Press.

Goffe, W.L., Ferrier, G.D., and Rogers, J. 1994. Global Optimisation of Statistical Functions with Simulated Annealing Journal of Econometrics. vol. 60: pp65-100.

Hooke, R., and Jeeves, T.A. 1961. Direct Search Solution of Numerical and Statistical Problems. ACM Journal. vol 8: pp212-229.

Kaupe, A.F. 1963. Algorithm 178: Direct Search. ACM Communications. vol 6: pp313-314.

Stefánsson, G., and Pálsson, Ó.K. 1998. A Framework for Multispecies Modelling of Boreal Systems. Reviews in Fish Biology and Fisheries. vol. 8.1: pp101-104.

Chapter 7

External Models

External estimation of process parameters

When developing large models in any field, choices need to be made as to what should be estimated in the models and what should remain as assumptions or externally estimated. The dst^2 project and Gadget framework are no exception to this.

A particular problem issue in multi-area models is the collection of migration parameters. These are unlikely to be well-estimated in any scenario and if these were to be estimated as free parameters, the sheer number of parameters quickly exceeds any reasonable limits. Naturally it is important to use separately migration rates for feeding and spawning behavior and therefore a temporal component typically needs to be added on top of having migration rates between all (adjacent) areas (Stefansson and Palsson, 1998). The approach taken in BorMiCon (Stefansson and Palsson, 1997) was to use ad-hoc methods to reduce the number of parameters, basically by assuming that many of them had to be equal. This approach is clearly not adequate.

The migration rates are a particularly important issue in the Gadget setting, since although it is seen in 4.4 that migration rates can in principle be estimated, it is also abundantly clear that the more areas are used, the more difficult the estimation will be and the results in 4.4 clearly only apply to the case of few areas.

At the other extreme one could reduce the parameter set immensely by assuming some underlying continuous movement model describing the movement of fish in continuous time. Given that most assessment methods (including Gadget) work on discrete scales, the use of such an internal model would require considerable internal computational power in order to undertake the integration required to determine the movement from block to block in a given time step.

For these reasons several processes were addressed outside the Gadget framework. In particular, migration issues were handled by migration models developed as stand-alone models intended to appropriately describe such processes alone. Similarly, other subprojects included the direct estimation of consumption and the effects of different control measures.

Such results provide important information on how these processes can be incorporated into the Gadget framework. For example, knowledge on spatial dynamics can be brought into Gadget in the form of migration matrices, based on results from the external models, considerably reducing the difficulty of estimation within Gadget.

Stefansson, G. and Palsson, O. K. (1997). BORMICON A Boreal Migration and Consumption model Marine Research Institute Tech. Rep. 58.

Stefansson, G., Palsson, O. K. (1998:). A framework for multispecies modelling of Boreal systems.

Reviews in Fish Biology and Fisheries, 8: 101-104.

7.1 Individual-based multi-species models of cod, capelin and herring

Geir Huse
Institute of Marine Research, Box 1870 Nordnes, N–5817 Bergen, Norway.
Email: geir.huse@imr.no

Introduction

This report describes modelling work related to "WP 4.1 Feeding and consumption" of the EU funded project "Development of structurally detailed statistically testable models of marine populations" (QL K5 – CT 1999 – 01609). The work related to Deliverables 4.1.1 and WP 4.1.2 of that project was reported on in Huse (2003) We have interpreted "Deliverable 4.1.5 Final vital statistics to GADGET" to mainly regard the implementation of spatial detail in GADGET and this is reported in another enclosed report (Huse *et al.*, 2004).

Predator-prey interactions are of great importance in fisheries science, particularly as we attempt to move away from management of marine resources based on single species assessments, towards an ecosystem-based approach. Spatial overlap between predator and prey is a prerequisite for predation to occur. Consequently, it is important to address spatial processes in order to understand and enable quantification of predator-prey interactions (Magnússon, 1995). In general, fisheries assessment models are highly aggregated with regard to space, time and biological detail in order to allow parameter estimation. Such models are therefore not designed to study species interactions. A notable exception is multispecies models, which have been developed to address the impact of species interactions on the population dynamics of fish (Helgason and Gislason, 1979; Pope, 1979; Bogstad *et al.*, 1997; Stefánsson and Palsson, 1998). However, such models generally use large spatial and temporal scales, and thus the modelling of processes such as predation, growth and migration is rather coarse. In WP 4.1 we approach species interactions from a different model philosophy by using individual-based models with a fine spatial resolution and short time steps to simulate the movement, growth and survival of cod, capelin and herring in the Barents Sea. The consumption of herring and capelin by cod emerges from simulating encounters between the species, and growth is calculated using bioenergetics models. Below some of the work performed in WP 4.1 during the project is shown, related to modelling movement of fish and feeding and consumption.

Modelling movement of fish

An important element in the predator-prey modelling using bottom-up approaches as here, is the implementation of movement behaviour. Different methods for modelling fish movement have been attempted during the project including artificial neural networks (ANN, Huse and Giske, 1998; Huse *et al.*, 1999), adapted random walk (ARW, Huse, 2001), and rule-based approaches (Huse *et al.*, 2004; Vabø *et al.*, 2004). The ARW concept for adaptive modelling of movement behaviour, developed during the project, is shown in Figure 1. The concept is based on evolving threshold values for when to depart a habitat and where to go, using a genetic algorithm. A habitat is departed if the fitness value associated with it is below the evolved threshold value. Movements are simulated using a modification of the random walk technique. A random walk is defined as random movement where displacement in each direction has the same probability. The new technique is inspired by adaptive walks in fitness landscapes, as originally introduced by Wright (1968), but is here used for movements in geographic landscapes. Under the new concept, random numbers and threshold values are used to determine movement. Rather than having equal probabilities for movement in each direction as in random walk, the probabilities are biased as the result of the adaptive process. This allows some directions to be preferred over others. A comparison with ANNs reveal that ARW is an efficient modelling technique (Huse, 2001). Particular advantages are the readily interpretation of the movement threshold values (Fig. 1) and the straight forward link to observations.

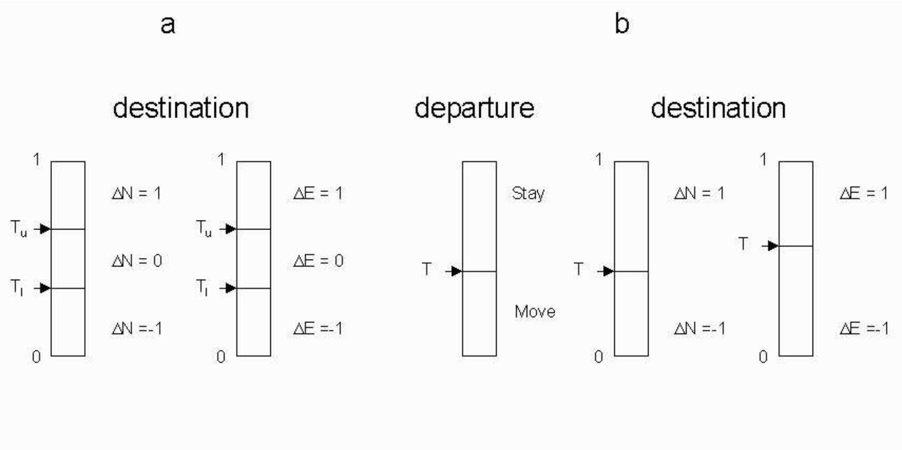


Figure 1: The concept for adapted random walk based on destination rules only (a) and in combination with departure rules (b). The threshold values are adapted using a genetic algorithm. During adaptation the threshold values (T_l, T_u and T) can move up and/or down to achieve the best value. Destinations are determined by drawing random numbers and comparing these with the threshold values. Movement is then performed according to which section of the bar the random number belongs. For the departure (b) the fitness value (or an environmental cue) in the square currently occupied by the individual in question, is compared with the threshold value and if the current fitness is greater than the treshold the fish stays, else it moves according to the destination rule.

Work on general predator-prey interactions has resulted in a manuscript that is currently being revised (Huse, submitted). This manuscript presents a model of cod-capelin interactions where behaviour is simulated using artificial neural networks (ANN), adapted over hundreds of generations by a genetic algorithm. Specifically the manuscript focuses on the causes for the northwards summer migration that the Barents Sea capelin performs. This migration has generally been labelled a feeding migration, but may also be motivated from predator avoidance of cod in cold Arctic waters. The main objective of the study is to attempt an evaluation of the relative importance of growth and survival in bringing about the northwards migration of capelin during summer. Simulations with different levels of capelin forage, varying light conditions, with and without a predator refuge, and different levels of alternative prey for cod are performed. The results suggest that the northwards migration is much more likely to evolve when there is a refuge against cod predation in cold water (below 1.5°C). Even a 100% increase in forage density in the northern areas does not promote a northward migration in itself. Some distribution maps of capelin during the spawning season are shown in Figure 2. Here the model is run with different levels of alternative prey for cod. At low levels of alternative prey, the population size of cod remains low and the capelin is abundant at the spawning ground and has a fairly spread out distribution in the southern part of the Barents Sea (Fig. 2a). When the population size of cod is increased (Fig. 2b), the distribution of capelin is more distinct, with a spawning aggregation at the spawning area and the rest of the stock distributed further north along the polar front. At the highest level of alternative prey when cod is most abundant, the entire stock except for a few spawning individuals stays in a narrow refuge area along the polar front (Fig. 2c). This illustrates the importance of predation in shaping distribution patterns of fish stocks.

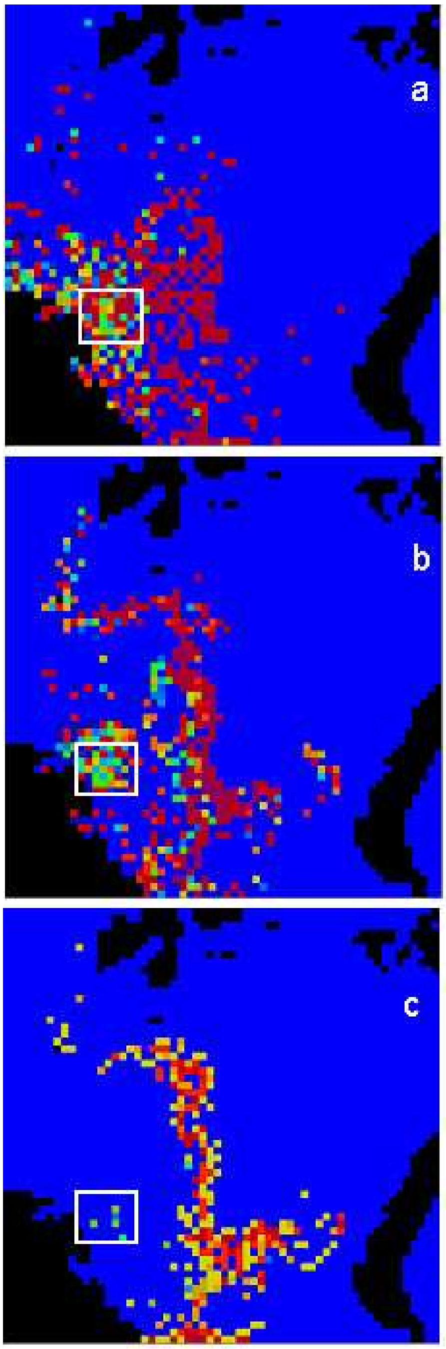


Figure 2: The distribution of prey at the start of the spawning in simulations with different level of alternative prey. The capelin spawning area used in the model is indicated by the white rectangle.

In addition to the modelling of adaptive movement, rule-based approaches were investigated where the aim was to mimic patterns observed in the field. It is important to address movement patterns at different spatial scales in order to understand spatial dynamics. Examples of simulated search behaviour of individual cod using different movement strategies are shown in Figure 3. The diversity in search pattern between the different strategies is pronounced. In these simulations we investigated the profitability of different search strategies in finding bait (or prey), and the "counter current" rule (Fig. 3e and f) consistently performed better than the other strategies in locating odour plumes from the bait. These search patterns resemble observed search patterns of real cod from Ramfjord (Vabø *et al.*, 2004).

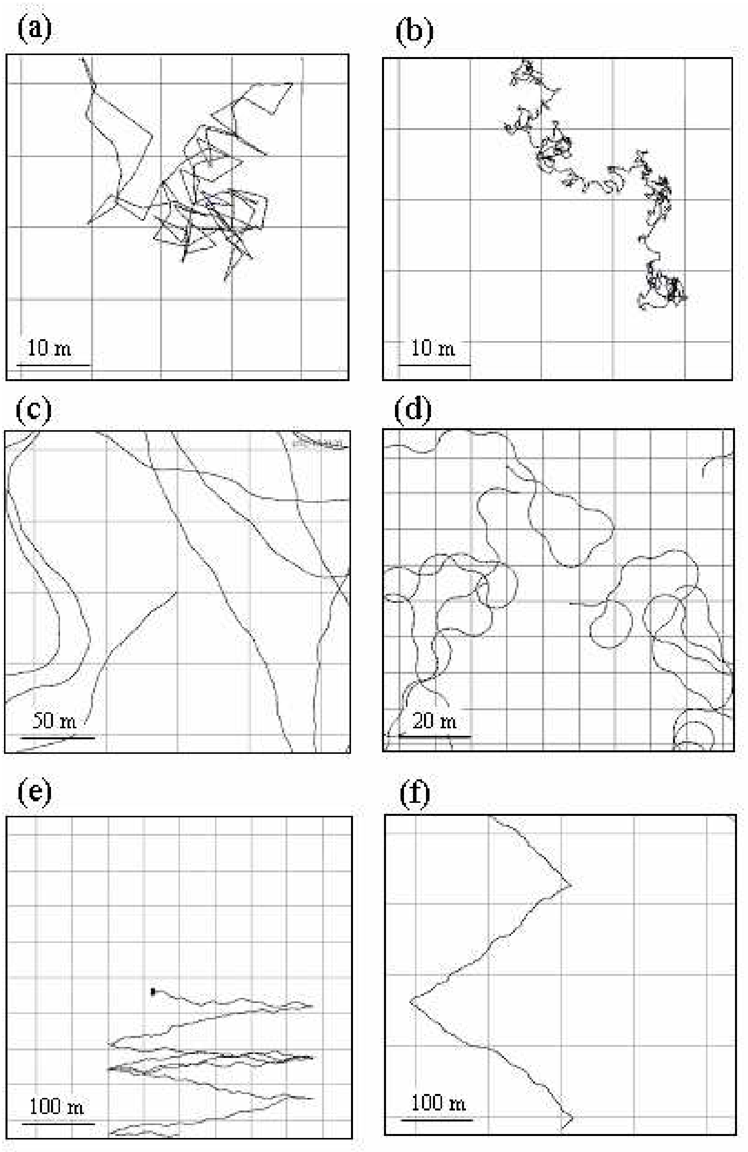


Figure 3: Trajectories of different search rules. The general current direction is towards the bottom of the figure. The scale varies among the figures and is indicated in the bottom left corner of each subplot. a) “random walk” dt=10, b) “random walk” dt=2, c) “random turn”, dt=10s, dá =1°, d) “random turn”, dt=10s, dá =5°, e) “counter current” dt=600s, dá =80°, f) “counter current” dt=600s, dá =45°. Here dt and dá refer to length between turns and turn angles respectively.

An example of rule-based movement in a large scale simulation is shown in Figures 4 and 5. The figures show the predicted monthly distribution of capelin and cod stocks respectively (Huse *et al.*, 2004). Here, the directed movement is based on movement vectors and temperature boundaries, with stochastic and advective components added. The simulations are initiated from survey data on cod and capelin. The observed abundances of cod and capelin are divided among altogether 40 000 super-individuals, each representing millions of individuals. These super-individuals are then moved about according to the rules. Simulations are validated against observed distributions of the two species after 6 and 12 months of simulation, and the results of the best simulation are shown here. The predicted distributions give a fair fit to the observations, which are shown to the right. The separation of the mature and immature capelin associated with spawning, is seen during February to April (Fig. 4).

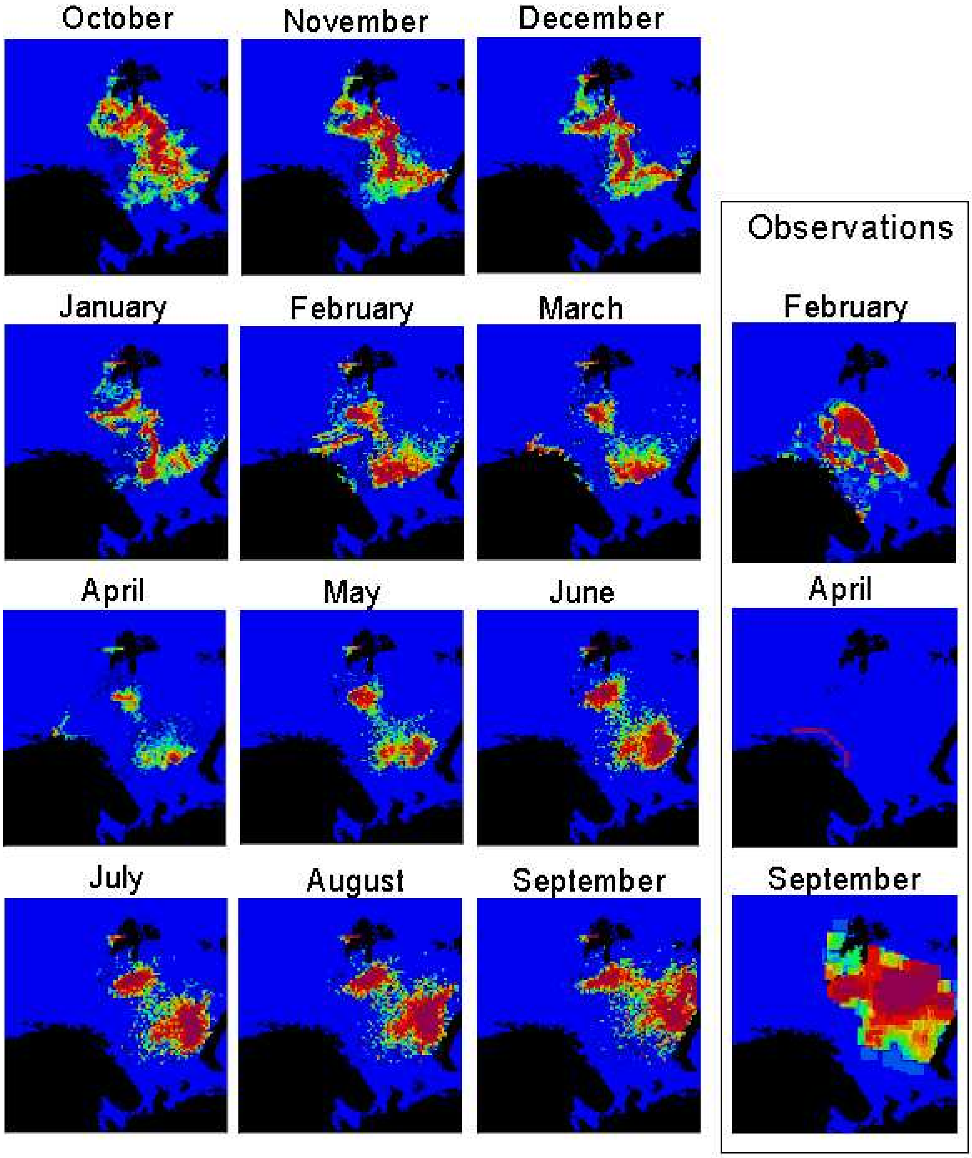


Figure 4: Predicted spatial distribution of capelin from October 1990 to September 1991. Observed spatial distributions of capelin for February and September are shown to the right. The observed spawning area in 1991 is inside the purple fields for the April panel, but no acoustic observations for the remaining part of the stock are made at that time.

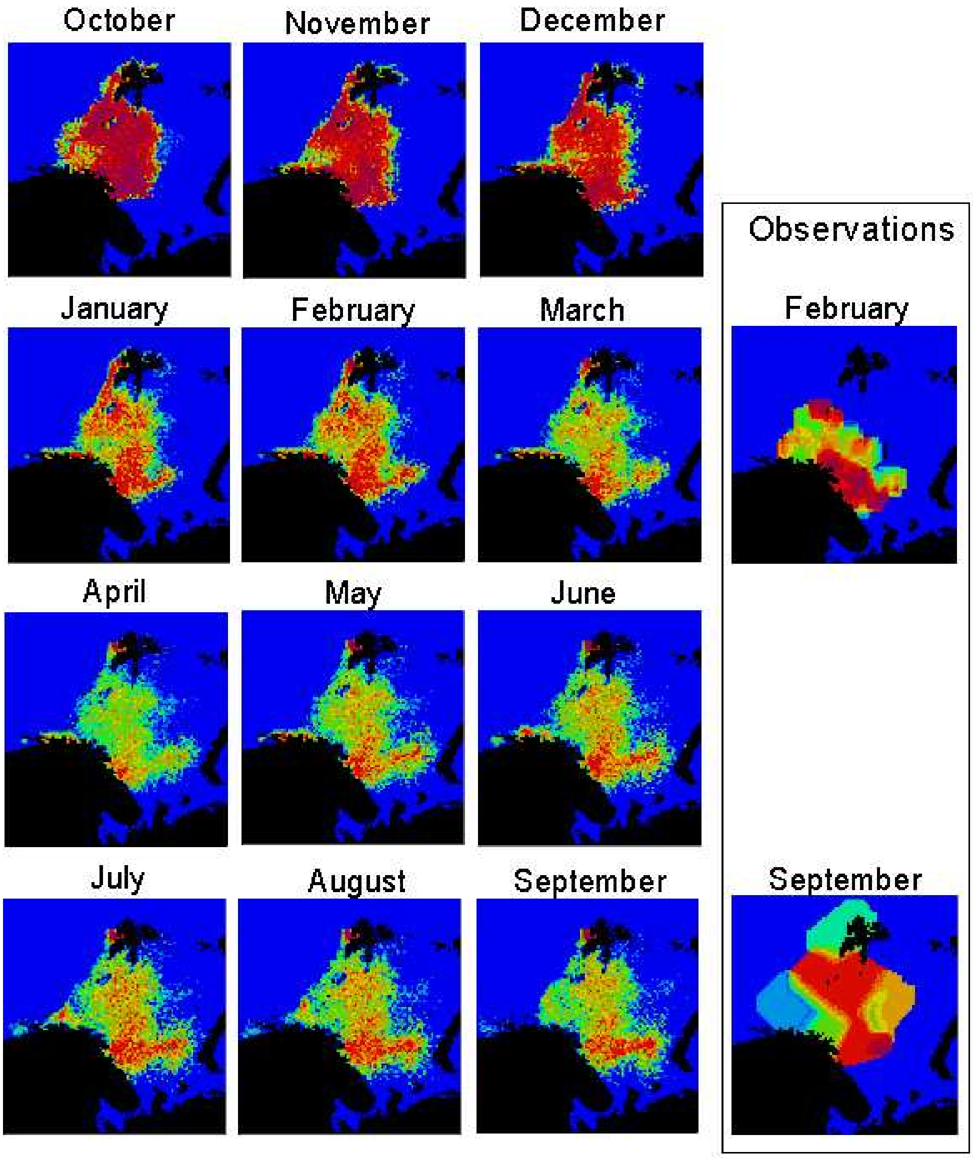


Figure 5: Predicted spatial distribution of cod from October 1990 to September 1991. Observed spatial distributions of cod for February and September are shown to the right.

The relatively high effort on modelling fish movements related to WP 4.1 is motivated from the particular approach taken based on individual-based modelling. Indeed many different simulations confirm that spatial dynamics can have a pronounced effect on feeding and consumption. For example the study by Huse *et al.* (2004) showed that consumption estimates varied depending on the movement models, and the best movement model also produced the consumption estimate closest to that obtained in other studies. Introducing a simple rule stating that cod should move in a randomly selected direction when the local capelin density is zero increased the consumption estimate by 30%. This suggests that emphasis needs to be put on exploring how behavioural rules in predators and prey affects their interactions.

Feeding and consumption

In addition to the deliverables report (Huse, 2003), feeding and consumption within WP 4.1 has mainly been reported in Huse *et al.* (2004). Here growth of cod and capelin was modelled using bioenergetics models (Hewett and Johnson, 1992), and the interaction was simulated using the movement model as described above. Predicted average capelin weight at age was in fair agreement with observations (Fig. 6), but were not significantly correlated (O2 vs. P2: $r^2 = 0.34$, $p > 0.05$, O3 vs. P3: $r^2 = 0.22$, $p > 0.05$). The average observed weight for two year old capelin during the period was 11.7 g. The results indicate that the simple zooplankton model used without any component taking into account capelin density or differences in zooplankton production is insufficient to explain the observed inter annual variation capelin growth.

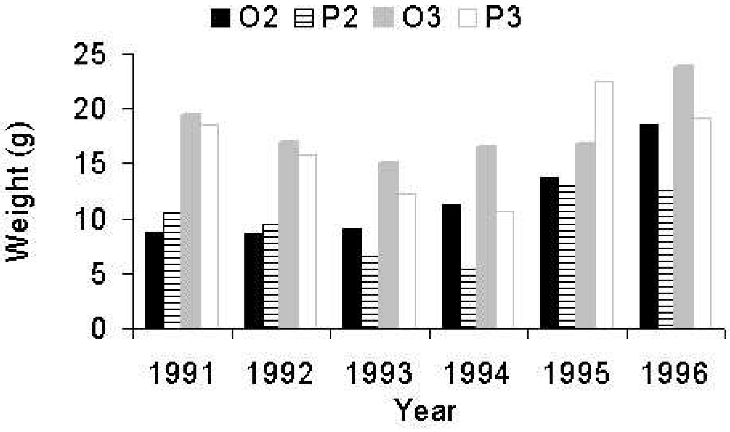


Figure 6: Comparison between observed and predicted weight for two (O_2 and P_2) and three year old (O_3 and P_3) capelin at the end of the simulation in September.

Cod was assumed to have a constant feeding level of other food (OF) in addition to the foraging on capelin, which depended on the local capelin density. Cod foraging on capelin was calculated square by square in two steps. First an initial estimate of the consumption of each cod super-individual was calculated by letting each super-individual feed sequentially according to the capelin density in the square, taking into account the reduction in capelin biomass caused by cod feeding. Then the biomass of capelin eaten by cod super-individuals in each square was summed. Next the total capelin consumption in the square was divided by the cod biomass in the square to yield an estimate of biomass of capelin consumed per biomass of cod, which then was used to calculate the total consumption by each cod super-individual. Cod weight at age was predicted with a greater accuracy than for capelin (Fig. 7). The observed average weight of six year old cod during the study period was 2968.7 g. Both for three and six year olds, significant correlations between observed and predicted values were found (O3 vs. P3: $r^2 = 0.91$, $p < 0.01$, O6 vs. P6: $r^2 = 0.69$, $p < 0.05$). Still there were some problems, for example in 1992, when weight at age was predicted to be much higher than observed. The poor growth in the latter part of the study period following the collapse of the capelin stock was predicted. Cod of about 75-85 cm had the highest growth relative to other size groups (Fig. 8).

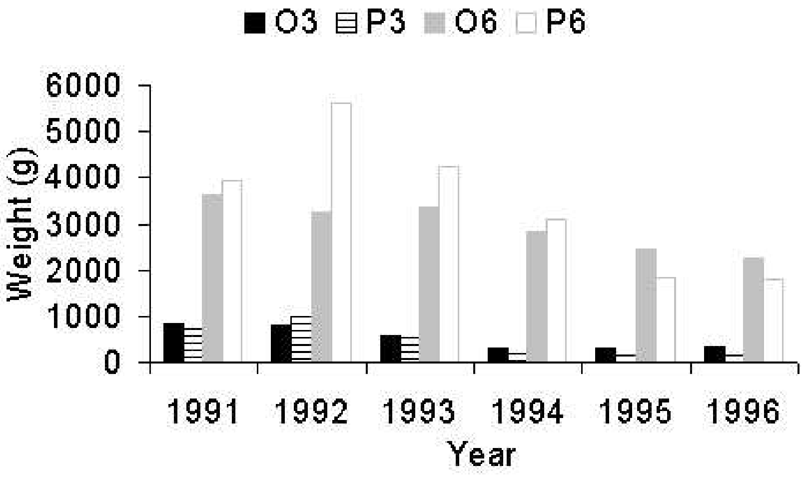


Figure 7: Comparison between observed and predicted weight for three (O_3 and P_3) and six (O_6 and P_6) year old cod at the end of the simulation in September using simulation 10.

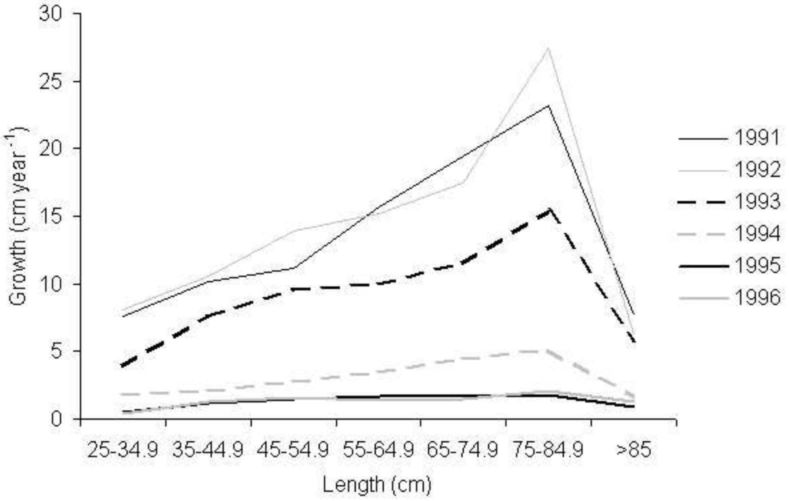


Figure 8: Size dependent growth of cod during 1991–1996, grouped by initial length.

Capelin biomass was predicted to vary in a similar pattern as the observed biomass data although there were some discrepancies, for example in 1992 when the predicted biomass was quite a bit higher than observed (Fig. 9). The maximum predicted capelin consumption by cod was about 1.9 mill. tonnes, in 1991–1992. There was considerable inter-annual variation in the biomass of capelin consumed by cod (Fig. 9), reflecting the variation in capelin biomass. Also there was a pronounced seasonal variation in capelin consumption by cod, with a peak in February-April coinciding with capelin overwintering and spawning. Capelin mortality due to cod predation was generally higher in the two year olds than in the three year olds (Table 1). The mortality attributed to cod predation in the two year old capelin varied from 15 to 30% per year.

	Year					
Age group	1991	1992	1993	1994	1995	1996
2	0.154	0.163	0.241	0.291	0.223	0.288
3	0.065	0.111	0.21	0.35	0.21	0.191
4	0.044	0.144	0.111	0.218	0.147	0.099

Table 1: Age dependent mortality of capelin during 1991–1996. Age group refers to age at the end of the simulation.

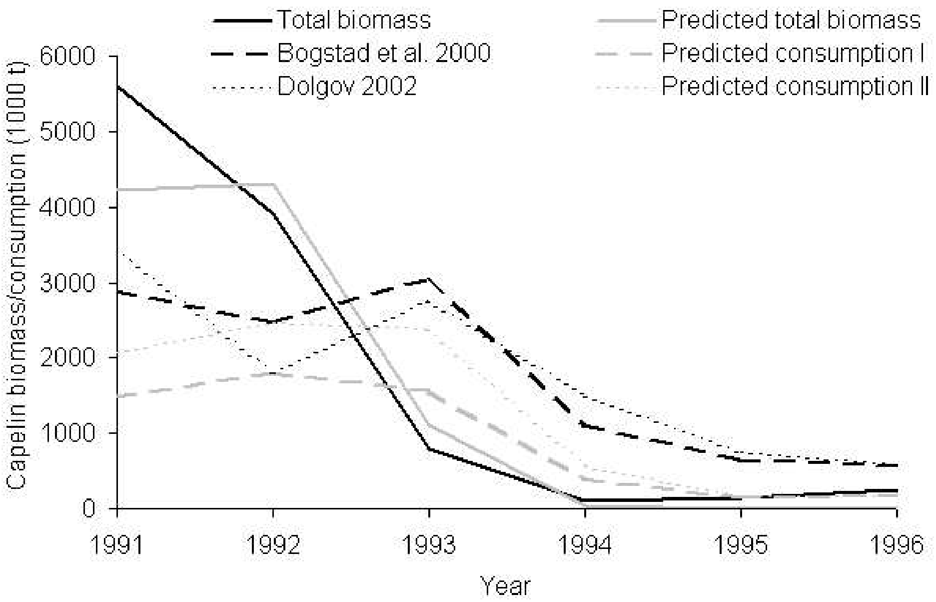


Figure 9: Observed and predicted total capelin biomass (2 years and older), and capelin consumption by cod during 1991–1996. Bogstad *et al.* 2000 and Dolgov 2002 refer to previously reported consumption estimates. Predicted consumption I refers to output from the best movement simulation, while Predicted consumption II refers to a simulation where cod move in a randomly chosen direction if no capelin is present.

Discussion

The modelling work performed in WP 4.1 has integrated fine scale movement processes with predator-prey interactions between fish stocks. We believe that it is important to make this link and the results shows that spatial processes can have a pronounced effect on the predator-prey interaction. Consequently it is important to address such issues within fisheries science. The modelling approach developed in the project is very flexible as illustrated above, and allows many aspects of real fish populations to be implemented. The ambition behind this line of work is to develop a modelling approach that can be used to study interactions between fish populations as well as contribute to fisheries management. The results obtained in this project with regard to simulated spatial distribution, growth and consumption are promising.

The results from the consumption model illustrate the close link between the cod, capelin and herring, which has been pointed out by many previous authors (Mehl, 1989; Hamre, 1994; Gjøsæter, 1998). The model captured the inter-annual dynamics of cod growth fairly well, with relatively high growth during the first part of the period, and lower growth towards the end of the period following the collapse of the capelin stock. Capelin growth is known to vary inter-annually as a function of climatic conditions (Gjøsæter and Loeng, 1987; Skjoldal *et al.*, 1992) and density dependency (Hopkins and Nilssen, 1991). In 1991 and 1992 the predicted capelin weight at age was similar to the observed values. During 1993, 1994, and 1996 growth was underestimated, while in 1995 it was overestimated. Clearly there is some inter-annual variation in capelin growth that is not captured by the model. It is difficult to point to the specific causes of these discrepancies, but most likely they are linked to prey dynamics and/or bioenergetics. These issues need to be addressed in future studies.

Herring in the Barents Sea is not investigated with the same regularity as cod and capelin. This made it difficult to develop a movement model for herring migration (see Huse, 2003). Consequently we have put a lot more effort on studying cod-capelin interactions than cod-herring interaction during the project.

The consumption model presented here (Huse *et al.*, 2004) describes the basic rates of growth, mortality and consumption of the target stocks, which are essential components of population models including fisheries assessment models. This makes it straight forward to investigate the effects that various spatial processes have on the model results. At present, spatial dynamics are rarely considered in fisheries assessment models. In ‘traditional’ multispecies models where migration is included (e.g. MULTSPEC, Bogstad *et al.*, 1997), it is not process-based. Inter annual variation in overlap between cod and capelin during the capelin spawning migration could for example be included in the assessment methodology presently used for capelin (Gjøsæter *et al.*, 2002). To our knowledge this model (Huse *et al.*, conditionally accepted) is the first attempt to simulate the interaction between two fish populations in a model, which is initiated by observed data on spatial distribution, size, and age, and includes daily movement and a fine scale environment description. Models such as the one described in WP 4.1 can be valuable tools for studying how spatial processes affect the growth, survival and interaction of fish populations, and may for example be used to address climate change effects on fish stocks and issues related to ecosystem based management. There is need for further investigations into how different movement rules affects the interactions between the target species, and the validity of other assumptions, in particular regarding capelin growth. Rather than seeing this contribution as an end result, it should be viewed as a starting point for future explorations of spatial IBM applied to fish populations where migration is an important part of the stock dynamics.

Acknowledgements

I am indebted to Øyvind Ulltang and Geir Odd Johansen for their enthusiastic cooperation throughout the dst^2 project. This work was funded by the European Commission through the project "Development of structurally detailed statistically testable models of marine populations" (QLK5–CT1999–01609).

References

Bogstad, B., Hauge, K. H., and Ulltang, Ø. 1997. A multispecies model for fish and marine mammals in the Barents Sea. Journal of Northwest Atlantic Fisheries Science, 22: 317–342.

Gjøsæter, H. 1998. The population biology and exploitation of capelin (*Mallotus villosus*) in the Barents Sea. Sarsia, 83: 453–496.

Gjøsæter, H., Bogstad, B., and Tjelmeland, S. 2002. Assessment methodology for Barents Sea capelin, *Mallotus villosus* (Müller). ICES Journal of Marine Science, 59: 1086–1095.

Gjøsæter, H., and Loeng, H. 1987. Growth of the Barents Sea capelin *Mallotus villosus* in relation to climate. Environmental biology of fishes, 20: 293–300.

Hamre, J. 1994. Biodiversity and the exploitation of the main fish stocks in the Norwegian-Barents Sea ecosystem. Biodiversity and Conservation, 3: 473–492.

Helgason, T., and Gislason, H. 1979. VPA-analysis with species interaction due to predation. ICES CM 1979/G: 52, 10 pp.

Hewett, S. W., and Johnson, B. L. 1992. Fish Bioenergetics Model 2. University of Wisconsin Sea Grant Institute, Madison, 80 pp.

Hopkins, C. C. E., and Nilssen, E. M. 1991. The rise and fall of the Barents Sea capelin (*Mallotus villosus*) – a multivariate scenario. Polar Research, 10: 535–546.

Huse, G. 2001. Modelling habitat choice in fish using adapted random walk. Sarsia, 86: 477–483.

Huse, G. 2003. Deliverables 4.1.1 and 4.1.2. Development of structurally detailed statistically testable models of marine populations (QL K5 – CT 1999 – 01609), Progress report for 1 January to 31 December 2002.

Huse, G. submitted. Adaptive predator and prey migration in a large scale environment. The American Naturalist.
Huse, G., and Giske, J. 1998. Ecology in Mare Pentium: an individual-based spatio-temporal model for fish with adapted behaviour. Fisheries Research, 37: 163–178.

Huse, G., Johansen, G. O., Gjøsæter, H., and Bogstad, B. 2004. Studying spatial and trophic interactions between capelin and cod using individual-based modelling. ICES Journal of Marine Science, 61: 1201-1213.

Huse, G., Sigurdsson, S., Dereksdottir, E. H., and Tjelmeland, S. 2004. Implementation of spatial detail in GADGET: a comparison of different approaches. Development of structurally detailed statistically testable models of marine populations (QL K5 – CT 1999 – 01609), Final report.

Huse, G., Strand, E., and Giske, J. 1999. Implementing behaviour in individual-based models using neural networks and genetic algorithms. Evolutionary Ecology, 13: 469–483.
Magnússon, K. G. 1995. An overview of the multispecies VPA - theory and applications. Reviews in Fish Biology and Fisheries, 5: 195–212.

Mehl, S. 1989. The Northeast Arctic cod stock's consumption of commercially exploited prey species in 1984–1986. Rapports et Procès-Verbaux des Réunions du Conseil International pour l'Exploration de la Mer, 188: 185–205.

Pope, J. G. 1979. A modified cohort analysis in which constant natural mortality is replaced by estimates of predation levels. ICES CM 1979/H: 16, 7 pp.

Skjoldal, H. R., Gjøsæter, H., and Loeng, H. 1992. The Barents Sea ecosystem in the 1980s: ocean climate, plankton and capelin growth. ICES Marine Science Symposia, 195: 278–290.

Stefánsson, G., and Palsson, O. K. 1998. A framework for multispecies modelling of arcto-boreal systems. Reviews in Fish Biology and Fisheries, 8: 101–104.

Vabø, R., Huse, G., Fernö, A., Jørgensen, T., Løkkeborg, S., and Skaret, G. 2004. Search behaviour of fish towards baited fishing gear: confronting theory and simulations with field observations. ICES Journal of Marine Science, 61: 1224–1232.

Wright, S. 1968. Evolution and the genetics of populations. Volume 1: Genetic and Biometric Foundations. University of Chicago Press, Chicago, 470 pp.

7.2 Mathematical models for migration of fish: An overview of work carried out by SCUI under the dst^2 project

Kjartan G. Magnússon and Sven Th. Sigurdsson
Science Institute, University of Iceland, Dunhaga 3, 107 Reykjavík, Iceland

The migration models that we have developed since the autumn of 2001 under the dst^2 project have mainly been motivated by the migration patterns of pelagic fish, in particular feeding and spawning migrations of capelin in the Barents Sea and the seas around Iceland. The models fall into two main categories, with each category in turn being divided into two sub-categories:

1. **Spatially discrete models**. Individual fish or small schools represented as particles.
    (a) **Simple particle models**. Each particle represents a school of fixed size.
    (b) **Superparticle models.** Particles may represent schools of differing sizes that may fuse together into a larger particle or split into smaller ones.
2. **Spatially continuous models.** The fish is represented in terms of a continuous density function.
    (a) **Dynamic models.** The density is the actual density of the fish.
    (b) **Kolmogorov type models.** The density function is a probability density function showing the probability that the fish is at a given location at a given time.

A common feature of all our models, as they have been developed so far, are:

1. They are **two-dimensional** in space and thus in particular do not include any effects of depth
2. The **direction of motion,** taking into account such effects as temperature, food supply and attraction to spawning grounds is **governed by** the gradient of a so-called **comfort function**.
3. The motion includes a **stochastic component**.
4. The effect **of sea currents** is included.
5. **Maturity stage** has in general **not** been included explicitly.
6. **Mortality** has **not** been included, although such an extension is fairly straight-forward in the case of the spatially continuous models.

The basis of the simple particle model and some of its most important characteristics are described in a published paper [1]. Report [1] is an earlier version of this work. In a separate paper in this report it is described how the model can be applied to the task of simulating migration of capelin in the seas around Iceland. A slightly more detailed account of this is presented in report [8]. A noteworthy aspect of this implementation is that it makes use of exactly the same triangularization of the region under consideration as is done in the finite element implementation of the continuous models. Therefore the same external files on e.g. sea temperature, sea currents and land boundaries can be used in all the models.

A preliminary program has been written for the superparticle model and implemented for capelin migrations in the Barents sea. There is, however, as yet no publication that describes this model.

The basis of the Kolmogorov type model with an application to capelin migrations in the Barents Sea is presented in a published paper [2], and an application of such a model to capelin migrations around Iceland is presented in a second published paper [4]. An account of this application is also presented in a separate paper in this report. Reports [2], [6], and [7] are earlier versions of similar work. Report [7] also contains a description of the finite element implementation of the model. In terms of the stochastic component this model has the important distinction that this component is incorporated into the probability density function computed by the model, but the computation itself is deterministic, i.e. it does not include any random element. The remaining models all contain a random element so that repeated computations have to be carried out in order to obtain results on the most likely movement.

The basis of the dynamic model with an application to capelin migrations in the Barents Sea is presented in a published report [3], which also contains a description of the finite element implementation of the model. An important distinction of this model is that it allows us to incorporate the effect of some preferred density, so that the fish is drawn together if its density is below the preferred level, but pushed apart if the density is above this level. This leads to some interesting results in terms of school formation. A mathematical analysis of these effects in terms of attractors of such models is presented in a published paper [3], but restricted mainly to the 1-dimensional case. Report [4] is an earlier version of that work and report [5] deals with further mathematical analysis of the model related to regularity and well-posedness.

As well as analysing these models in their own right and and designing efficient numerical implementations we have considered how they could be incorporated into a fishing assessment tool such as GADGET, by using the models to construct the migration matrices that GADGET makes use of in order to take into account migration. The most suitable model for this purpose is the Kolmogorov type model, because it is the only model that computes the probability density of the position of fish directly, as mentioned above. We describe a possible procedure for the construction of migration matrices in two separate papers in this report. Similar ideas are also presented in a published paper [4], as well as in reports [6] and [7].

While we have demonstrated qualitatively how mathematical migration models can reflect important aspects of fish migration patterns, at least in the case of capelin, it remains to carry out a more quantitaive comparison, using e.g. sounding measurements and/or recaptured tags, and to estimate the values of the free parameters of the models from such a comparison. The central idea is that by constructing an appropriate comfort function and then estimating parameters of this function, that are independent of environmental factors such as sea temperature or location of spawning grounds, the model can subsequently be used to assess the effects of changes in environment on the migration pattern. The GADGET packet in its present form does include the possibility for such an assessment. In two separate papers in this report we present some preliminary ideas on how one may carry out such a parameter estimation. We see this as the most important next stage in further development of the mathematical migration models.

Published papers

1. S. Hubbard, P. Babak, S. Th. Sigurdsson, K. G. Magnússon. 2004. A model of the formation of fish schools and migrations of fish. Ecological Modelling, 174, 359-374

2. K. G. Magnússon, S. Th. Sigurdsson, P. Babak, S. F. Gudmundsson, E. H. Dereksdóttir. 2004. A continuous density Kolmogorov type model for a migrating fish stock, Discrete and Continuous Dynamical Systems-Series B, Volume 4 (3), 695-704

3. P. Babak, K. G. Magnússon, S. Th. Sigurdsson. 2004. Dynamics of group formation in collective motion of organisms, Mathematical Medicine and Biology, Volume 21, (4), 269-292

4. K. G. Magnússon, S. Th. Sigurdsson, E. H. Dereksdóttir. (In press). A simulation model for capelin migrations in the North-Atlantic. Nonlinear Analysis: Real World Problems.

Science Institute Reports

1. S. Hubbard, P. Babak, S. Th. Sigurdsson, K. G. Magnússon. 2001. A model of the formation of fish schools and migrations of fish. Science Institute, University of Iceland. Report RH-28-2001.

2. S. Th. Sigurdsson, K. G. Magnússon, P. Babak, S. F. Gudmundsson, E. H. Dereksdóttir. 2002. A continuous density Kolmogorov type model for a migrating fish stock. Science Institute, University of Iceland. Report RH-24-2002.

3. K. G. Magnússon, S. Th. Sigurdsson, K. G. Magnússon, P. Babak, S. F. Gudmundsson, E. H. Dereksdóttir. 2002. Dynamic continuous model of fish migration. Science Institute, University of Iceland. Report RH-25-2002.

4. P. Babak, S. Hubbard, K. G. Magnússon, S. Th. Sigurdsson. 2002. Dynamics of group formation in collective motion of organisms. Science Institute, University of Iceland, Report RH-28-2002.

5. P. Babak, K. G. Magnússon, S. Th. Sigurdsson. 2003. Well-posedness and regularity of solutions for dynamical system arising in collective motion of organisms in 1D bounded domains. Science Institute, University of Iceland, Report RH-12-2003.

6. E. H. Dereksdóttir, K. G. Magnússon and S. Th. Sigurdsson. 2003. Simulations of spawning migrations of capelin in Icelandic waters based on temperature-current- and potential fields. Science Institute, University of Iceland, Report RH-16-2003.

7. E. H. Dereksdóttir, K. G. Magnússon and Sven Th. Sigurdsson. 2003. Simulations of spawning migrations of capelin in the Barents sea based on temperature- and potential fields. Science Institute, University of Iceland, Report RH-17-2003.

8. K. G. Magnússon, Sven Th. Sigurdsson and B. Einarsson. 2004. A discrete and stochastic simulation model for migration of fish with application to capelin in the seas around Iceland. Science Institute, University of Iceland, Report RH-20-2004

Conference presentations

1. K.G. Magnússon. Models of spawning and feeding migrations of pelagic fish species in the North Atlantic. International Conference on Mathematical and Theoretical Biology. Hilo. Hawaii. July 16-19 2001.

2. S. Sigurdsson. Discrete particle models and continous density models of fish migration. Workshop on Computational Biology, The Fields Institute. Toronto. Canada. November 29-December 2, 2001.

3. S. Sigurdsson. Continuous density-velocity model of fish migration. Conference on Mathematical Modelling of Population Dynamics. Bedlewo. Poland. June 24-28, 2002

4. K.G. Magnússon. Density-velocity models of fish migrations. ICM2002 Satellite Conference on Mathematical Biology. Guilin. China. August 15-18, 2002.

5. K.G. Magnússon. Models of fish migration of pelagic fish in the North Atlantic. Alcala 2nd International Conference on Mathematical Ecology. Alcala de Henares. Spain. September 5-9, 2003.

6. Estimation of fish movement parameters from tagging data with application to Atlantic cod (Gadus Morhua). ICES. CM 2004/FF:25. Vigo. Spain. September 22-25, 2004.

7. A simulation model for capelin migrations in the North-Atlantic. ICES. CM 2004/FF:34. Vigo. Spain. September 22-25, 2004.

Conference poster presentations

1. Petro Babak. Dynamics of Group Formation in Collective Motion of Organisms. EuroMech 422. Pattern Formation by Swimming Micro-Organisms and Cells. Leeds, England, December 2001.

2. Simon Hubbard. Models of spawning and feeding migrations of pelagic fish species in the North-Atlantic, EuroMech 422. Pattern Formation by Swimming Micro-Organisms and Cells. Leeds, England, December 2001.

3. Petro Babak and Simon Hubbard. Discrete models for moving social organisms. Statistical analysis and computer modelling, From Worker to Colony: Understanding the Organisation of Insect Societies, Cambridge, December 2001.

4. K. G. Magnússon, Sven Th. Sigurdsson and E. H. Dereksdóttir. Spawning migrations of capelin. “Computational and Mathematical Population Dynamics", Trento, Italy, 21.-26.6.2004

In addition, results from the project have been presented at a number of venues in Iceland, including the Marine Research Institute, and also at the Institute of Marine Research, Bergen, Norway.

7.3 A simulation model for capelin migrations in the North-Atlantic

Kjartan G. Magnússon[1]**, Sven Th. Sigurdsson and Eva Hlín Dereksdóttir**
Science Institute, University of Iceland, Dunhaga 3, 107 Reykjavík, Iceland

Abstract: Some pelagic fish species undertake extensive feeding and spawning migrations covering distances of several hundred miles. We present a model of such migrations where the dynamics are governed by an advection-diffusion type equation with the advection being determined by environmental fields for temperature, food density and oceanic currents, in addition to a force field generated by an attracting spawning region. The model is motivated by the migrations of the capelin stock in the Central North- Atlantic. A simulation of the spawning migration of this stock is presented. We also discuss how the continuous migration model can be reduced to a simpler model based on transition or migration matrices.

Key words: fish migration, advection-diffusion equation, capelin, environmental fields

Introduction

The seasonal spawning and feeding migrations of fish – or birds for that matter – are as yet not fully understood. Knowledge about what controls these regular movements of fish or bird populations is limited but some insight has been achieved, particularly as regards birds, which are easier to track and experiment with. Fish are more difficult, but the path of some pelagic species, such as capelin, herring, sardines and anchovies, which migrate in huge shoals may be monitored by visual and acoustic means. These species undertake extensive feeding and spawning migrations over several hundred miles (e.g. Vilhjálmsson, 1994, 2002; Misund et al., 1998; Gjøsæter, 1998).

The spatio-temporal distributions of fish and the timing and path of migrations are influenced by many factors. It is likely that there is a genetic component determining the general direction and time of the migrations, which are then controlled by a variety of environmental factors such as the temperature distribution, salinity, bottom topography, oceanic currents, density of food, and internal variables such as the physiological state and the state of maturity of each fish. Furthermore, the population density may influence the extent of feeding migrations. Due to lack of information and the high degree of uncertainty about possible causal relationships it is not feasible to attempt to take all of these factors into account in a simulation model. We present in this paper a model where some of the factors are included, i.e. a model driven by environmental conditions consisting of a temperature field, a food density field and a vector field of oceanic currents. These environmental variables determine the local forces affecting the movement of the fish schools. A key component in the model is an attraction towards the spawning grounds, which is modelled as a long-distance force analogous to the attraction of charged particles towards an electrically charged body or of particles of mass towards a large mass concentration. This force is used in the absence of any clear knowledge about why fish spawn where they do or how they find their way and may be viewed as a proxy for all the unknown factors, for example magnetic fields, chemical signals or even learnt behaviour, which drive the spawning migrations.

The migrations of capelin in the Barents Sea and in the waters around Iceland provide the main motivation for this work. Both stocks undertake a northward feeding migration in summer and a return spawning migration in winter, after which nearly all the spawners die. The migrations of the Iceland stock are of more interest from a modelling point of view since the stock has to go round the island to reach the spawning grounds on the south coast. We will therefore restrict our attention to this stock, although the model presented here could equally be applied to the Barents Sea stock (for an application of an earlier version of the model to this stock see Magnússon et al. (2004)). The broad migration

---

[1] Fax: +354-552 8911 e-mail: kgm@hi.is web-address: http://www.raunvis.hi.is/∼kgm/

pattern for the Icelandic capelin stock is the following (for details on the migrations of capelin in Icelandic waters see Vilhjálmsson (1994, 2002) which are the sources of all the information used in this work about capelin behaviour in Icelandic waters): the maturing stock usually migrates north to the Jan Mayen region in the Central North Atlantic in early summer, returns south arriving at the northwest coast of Iceland in early winter and subsequently migrates around the island in a clockwise direction to spawn on the south coast in March – April (see Fig. 1). The prevailing direction of the coastal currents around Iceland is clockwise (Fig. 2), which would seem to indicate that the current direction is the determining factor for the migration route around the island. However, a component of the spawning stock occasionally migrates in a counter-clockwise direction to the spawning grounds, the size of this component being very variable. The reason for this deviation is not known, but may be related to the temperature distribution and/or the state of maturity of the migrating stock. The spawning migrations of this stock have been monitored by acoustic surveys since the 1970's and a number of features have been observed: schools tend not to cross fronts between cold and warm water masses but migrate along such boundaries; cruising speeds are highly variable; the geographical path is fairly constant, but may vary, particularly due to changing locations of cold water fronts and certain isotherms and the spawning stock may enter the spawning grounds in a number of different runs, possibly along different routes. Our aim is to construct a simulation model, which emulates these movements and can be used to investigate and possibly test hypotheses about the effect of environmental conditions on migrations. It is quite likely that the state of maturation (roe content) of the individual fish is crucial in determining the timing and perhaps also the migration route. This will not be explored here but will be addressed in future work. Although the capelin migrations are the primary motivation behind this work, the model is nevertheless generic since it describes general movements between feeding and spawning grounds, which are influenced by the distributions of temperature, food density and oceanic currents and by the location of spawning grounds and any obstacles, such as islands, which the migrating fish must go round. In Hubbard et al. (2004) a prototype for a discrete and individual based migration model for capelin in Central North Atlantic is presented. The migrating fish are regarded as a self-propelled interacting particle system and the model describes migrations between feeding and spawning areas with an "obstacle" in between these areas. The fish are drawn to the spawning grounds by asymmetric noise in velocity and to the feeding grounds by a food density field. A completely different approach is adopted in the model presented here.

Most migration models belong to either of two types (see for example Beverton and Holt, 1957; Quinn and Deriso, 1999; Okubo and Levin, 2001): compartment models with the movements between compartments being described by transition matrices whose elements $m(i,j,t)$ are the probabilities of a fish moving from compartment i to compartment j between times t and t+1, and advection- diffusion models where the velocity may be (partly) determined by an environmental gradient. Compartment models are more empirical; the overall area where the movements and migrations take place is divided into a number of sub-areas and the values of the matrix elements obtained by a combination of guesswork and statistical estimation from tagging and/or survey data (Vilhjálmsson et al., 1997; Tjelmeland and Bogstad, 1998). The number of parameters to be estimated is generally large relative to the amount of data available and consequently some parameterization of the transition probabilities is required. These models are by their very nature discrete in the space variables and predict the changes in population density in the individual sub-areas and may be continuous or discrete with respect to time. Continuous advection-diffusion models predict changes in local population densities and contain essentially only two parameters, the advection velocity and the diffusion coefficients, although the values of these parameters are in general dependent on time and space. Furthermore diffusion may be anisotropic leading to yet more parameters. In general, the information requirements for such models are the values of the advection velocity and diffusion in each of a set of specified sub-areas. Continuous advection-diffusion models and discrete compartment models can be linked; Deriso et al. (1991) estimated diffusion and advection values (which can be input into a continuous model) by using a Markovian matrix model and we will describe below how results from numerical simulations with a continuous model may be aggregated into a compartment model with transition (or migration) matrices.

Advection-diffusion models were used by McCall (1990) to describe fish movements (and other dynamic population variables) and by Sibert et al. (1999), which used tagging data to estimate movements of skipjack tuna. We consider here a similar model. A stochastic differential equation for the position of a fish leads to Kolmogorov's forward equation for the time evolution of the probability density function for the position. The probability density may be viewed as the equivalent to the distribution of mass density, which is described by an advection-diffusion type equation in two space variables. The advection velocity will be modelled as a function of environmental conditions (temperature, food density and oceanic currents) and a force field directed towards the spawning grounds. This field will depend on the geometry of the domain i.e. the location and shape of the spawning grounds and of the obstacles, which the migrating fish must go round to reach them. An earlier version of this model was presented in Magnússon et al. (2004) with an application to capelin migrations in the Barents Sea. These migrations are simpler since there are no island obstacles between the feeding and spawning grounds as is the case with capelin migrations in the Iceland – Jan Mayen area. The long distance attracting force in the Barents Sea application was taken to be in the direction of the shortest distance to the spawning grounds. This formulation does not work in the present case since the modelled fish generally fail to go round the obstacle. This present model is an extension of the earlier one, applicable to domains with a more complex geometry and also takes oceanic currents into account, which were not included in the early version. Numerical simulations of the spawning migrations of Icelandic capelin are presented and loosely compared to actual distributions obtained from acoustic surveys. Finally, a method whereby the results from the continuous model can be aggregated to a compartment model with transition matrices is described.

The primary aim of this paper is to demonstrate that an advection-diffusion model with the advection direction determined by environmental variables (temperature and currents) and a spawning attraction field can produce distributions very similar to those observed in acoustic surveys. At this stage no attempt is made to estimate model parameters statistically by comparing calculated and observed distributions.

Model description

Consider the motion of a single fish. We can assume that the fish is undergoing a biased random walk in the plane with discrete time and spatial steps and by taking the continuum limit obtain a Kolmogorov type equation for the probability density of the location of the fish (see Magnússon et al, 2004). Alternatively, we can assume that the position in a domain Ω is governed by a two dimensional stochastic differential equation

$$d\boldsymbol{X}_t = (\boldsymbol{V}_t, t)\, dt + B\left(\boldsymbol{X}_t, t\right) d\boldsymbol{W}(t) \tag{1}$$

where $\boldsymbol{X}_t = (X, Y)$ is the random variable for the position of the fish, $\boldsymbol{V} = (v_1, v_2)$ is a given velocity, *B* is a given 2x2 matrix and *W(t)* is a continuous 2 dimensional Brownian motion. The conditional probability density at time *t* for the position of the fish, given that the fish is at (x_0, y_0) at time t_0,

$p = p\left(x, y, t | x_0, y_0, t_0\right)$

satisfies Kolmogorov's forward equation (Øksendal, 2000)

$$\frac{\partial p}{\partial t} = -\nabla \cdot (\boldsymbol{V} p) + \frac{1}{2}\left(\frac{\partial^2 \left(D_1 p\right)}{\partial x^2} + 2\frac{\partial^2 \left(D_{12} p\right)}{\partial x \partial y} + \frac{\partial^2 \left(D_2 p\right)}{\partial y^2}\right), \quad (x, y) \in \Omega \tag{2}$$

where

$$BB^T = \begin{bmatrix} D_1 & D_{12} \\ D_{12} & D_2 \end{bmatrix}$$

If ***V*** and *B* are constant, then it is easy to see that

$E\left[\boldsymbol{X}_t - \boldsymbol{X}_0\right] = \boldsymbol{V} \cdot t$

and

$Cov\left[\boldsymbol{X}_t\right] = \left(BB^T\right) \cdot t$

Equation (2) resembles an advection – diffusion equation but differs from it in that it contains a mixed second derivative term and the second derivative is taken of the multiple of the diffusion coefficients *D* and the density *p*. If the diffusion is isotropic then $D_{12} = 0$ and $D_1 = D_2 = D$, say, and equation (2) reduces to the equation:

$$\frac{\partial p}{\partial t} = -\nabla \cdot (\boldsymbol{V} p) + \frac{1}{2}\left(\frac{\partial^2 (Dp)}{\partial x^2} + \frac{\partial^2 (Dp)}{\partial y^2}\right) \qquad (3)$$

We will employ this simpler model in the simulations described here, since it is not possible to distinguish between models (2) and (3) on the basis of the available measurements, but our numerical implementation of the model, using finite elements can readily be extended to the more general case. We will also assume for simplicity that the diffusion coefficient is constant – in which case equation (3) reduces to an advection-diffusion equation.

If we assume that there is zero probability of a fish escaping through the boundary of Ω, the natural boundary condition is

$$\left(-v_1 p + \frac{\partial}{\partial x}(D_1 p) + \frac{\partial}{\partial y}(D_{12} p)\right) n_1 + \left(-v_2 p + \frac{\partial}{\partial y}(D_2 p) + \frac{\partial}{\partial x}(D_{12} p)\right) n_2 = 0, (x, y) \in \partial\Omega \quad (4)$$

where (n_1, n_2) denotes the outward unit normal vector to the boundary. Under the simplifying assumptions about the diffusion coefficients, this becomes simply

$$-(\boldsymbol{V} \cdot n)\, p + \nabla (Dp) \cdot \mathrm{n} = 0, \qquad (x, y) \in \partial\Omega \qquad (4)$$

The advection velocity ***V*** is determined by the environmental fields and the location of the fish relative to the spawning grounds. It is the vector sum of the velocity of the fish relative to the surrounding fluid and the velocity of the current, denoted by V_p and V_c respectively.

The direction of the velocity vector V_p is determined by the gradient of a function $U(x, y; t)$, which we will refer to as a comfort function (Reed and Balchen, 1982). This function will incorporate the factors, which are believed to affect the “comfort” or well-being of the fish, such as temperature, food density, location relative to the spawning grounds, etc. The fish move on a time-varying “comfort surface” defined by $U = U(x, y; t)$ constantly attempting to maximize their comfort, i.e. moving in the direction of the spatial gradient of *U*. The comfort function used here is a linear combination of functions of temperature and food density and a potential function

$$U(x, y; t) = \alpha_1(t)\, r(T(x, y; t)) + \alpha_2(t)\, s(f(x, y; t)) + \alpha_3(t)\, \phi(x, y) \qquad (5)$$

where *T(x,y;t)* and *f(x,y;t)* are the temperature and food density respectively at location *(x,y)* at time *t* and *r* and *s* are given functions defined as follows:

$$r(T) = \begin{cases} -(T - T_1)^4 & if \quad T \le T_1 \\ 0 & if \quad T_1 \le T \le T_2 \\ -(T - T_2)^2 & if \quad T_2 \le T \end{cases} \qquad (6)$$

and

$$s(f) = \frac{f}{h+f} \qquad (7)$$

where T_1, T_2 and *h* are constants. The functional form of *r* implies that the preferred temperature range is between T_1 and T_2 degrees. The fish tend to move towards areas where the temperature is within the preferred range and this tendency is stronger in cold waters. The preferred temperature range may vary within the year. The function *s* is an asymptotic and increasing function of food density, *f*. This functional form implies that the tendency to move towards areas of higher food density decreases with increasing density and is virtually nil at high food densities. The α – coefficients will be specified later.

The function ϕ is a potential function for the attraction towards the spawning grounds and is defined a s follows:

Let the region in which the fish can move be denoted by Ω. We define the boundary of Ω to be any lines, which the fish are not able to cross, such as the coastlines and possibly certain isotherms. The region is in general not simply connected since it may have one or more “holes” in it. Assume that the spawning grounds cover a sub-area S of Ω and that migrating fish are attracted towards S, in the sense that they experience a “force field” pulling them towards it. The spawning region may be regarded as a continuous sink spread over S, to use an analogy from fluid dynamics, and we define the sink density (analogous to mass density or charge density) in the spawning region S by $\rho\,(x, y)$. The total attraction strength (sink strength) of the spawning area is therefore given by

$$m = \iint_S \rho dA$$

The potential ϕ is given as the solution of Poisson’s equation

$$\Delta\phi = -\rho \qquad \text{in} \quad \Omega \qquad (8)$$

subject to no-flux condition at the boundaries of Ω, i.e.

$$\frac{\partial\phi}{\partial n} = 0 \qquad \text{on} \quad \partial\Omega \qquad (9)$$

The force field attracting the fish towards the spawning grounds is given by the gradient of ϕ $\nabla\phi$. The gradient is orthogonal to the equi-potential lines and is therefore tangent to the set of lines known as streamlines, which the fish would travel along if there were no other factors influencing the motion, than possible obstacles in the form of holes within Ω, that the fish cannot cross.

To take a very simple example, we can consider an infinite region Ω with no boundaries and an attracting region consisting of a single point of sink strength *m*, located at the origin. The attraction density is $\rho\,() = m\delta\,()$, where $\delta\,()$ is the delta function, the potential function is

$$\phi\,(x, y) = -\frac{m}{2\pi} \ln\sqrt{x^2 + y^2} \qquad (10)$$

and the streamlines are simply straight lines through the origin. The second example is of an attracting spawning region south of an island – Iceland in this case - with the fish approaching from the north and encountering the island obstacle. The attraction density ρ, is taken to be uniform over the attracting region, the spawning region in this case. Fig. 3 shows the force field attracting the fish towards a spawning area to the south of the island. The fish move clockwise or counter-clockwise depending on where they arrive from and on the location of the attracting spawning region. Note that in order to obtain this field we have to specify boundary conditions at the outer boundaries of the region. Rather than viewing them as non-obstacle boundaries we specify the ϕ – values on the outer boundary according to (10), approximating the spawning region by a single point.

The gradient of the comfort function gives the direction of the velocity vector $\boldsymbol{V}_p$, but the speed must be specified external to the model. The value of this parameter – which can be regarded as the average cruising speed- may be deduced from observations in experimental tanks and in the field, and is typically about one body-length per second or about 15 km/day for 17-18 cm capelin. However, cruising speeds may be as high as 20-30 nautical miles per day (Vilhjálmsson, 1994, p. 53). Alternatively, the speed may be estimated by comparing simulated to observed distributions. The advection velocity in equation (2) or (3) is the vector sum of $\boldsymbol{V}_p$ and $\boldsymbol{V}_c$ where the latter is a vector field, which is specified external to the model:

$$\boldsymbol{V} = \begin{cases} v\frac{\nabla U}{|\nabla U|} + \boldsymbol{V}_c & \nabla U \neq 0 \\ \boldsymbol{V}_c & \nabla U = 0 \end{cases} \tag{11}$$

It might be argued that the reaction of the fish to the current field should be included in the comfort function, but it is probably the velocity relative to the surrounding fluid that depends more directly on other environmental fields than the overall velocity.

Numerical simulations

We present here a number of simulations of the spawning migration of capelin in Icelandic waters designed to examine the influence of environmental factors on the spatio-temporal distribution. The biological and oceanographic background and the simulation framework is as follows:

The spawning stock spends the summer on the feeding grounds north towards the island of Jan Mayen (see Fig. 1), migrates south in autumn – probably following the eastern boundary of the cold East Greenland current – and arrives at the northwest coast of Iceland in early winter (Oct.-Nov.). Ideally, the simulated migrations should include the feeding migration. However, since there are no observations of the capelin distribution on the feeding grounds and very limited temperature or food density measurements, we will restrict the simulations to a six-month period from October 1 to April 1 when the spawning is taking place. The spatial domain is therefore limited to the waters around Iceland and only the spawning migrations around the island are simulated. The spawning stock is surveyed acoustically in October and again in January/February every year - Fig. 4 shows typical distributions in these months. These observations provide data, which can be used to calibrate the simulation model and to estimate parameters.

Temperature measurements are taken at a number of transects in October/November and in February/March. These measurements are interpolated (and extrapolated to some extent) spatially and temporally to give a daily temperature field for the period October-April. We have selected two years, which are sufficiently different to highlight the effect of the temperature distribution, 1994-1995, which was a cold year, and 2000-2001, a warm year. Fig. 5 shows the temperature fields in November 1994 and 2000 and in March 1995 and 2001. The fields are probably fairly accurate close to the shore, but some extrapolation combined with guesswork was required to obtain the fields further offshore.

At present, there are no data for oceanic currents in a format suitable for these simulations. However, a general picture is available of the currents in Icelandic waters, which has been drawn up based on some measurements and on general knowledge (see Fig. 2). This general picture was used to construct a hypothetical current field, which reflects the effects of the real field based on the information currently available (Fig. 6). Note that the coastal current circulates Iceland in a clockwise direction. The speed of the hypothetical current is in the interval [0.75, 15] km/day and only varies with approximate distance from the shore. Note that the upper value is similar to a typical cruising speed of one body length per second or 15 km/day.

Since the capelin feed very little on spawning migrations, the coefficient of the food density term in the comfort function is set to zero. The gradient of the comfort function is therefore a weighted sum of the temperature gradient and the attraction force field shown in Fig. 3, i.e.

$$\nabla U\left(x,y\right)=\alpha_1 r'\left(T\left(x,y\right)\right)\nabla T\left(x,y\right)+\alpha_3\nabla\phi\left(x,y\right)$$

The preferred temperature range varies from [0.5,4.0] degrees on October 1 to [3.0,7.5] on April 1, based on various field observations showing that capelin spend most of the year in cool waters (0-3°C) but spawn in somewhat warmer waters. Nevertheless, the empirical foundation for the preferred ranges is rather weak and further investigations (field observations, tank experiments and modelling work) are required. Vilhjálmsson (1994, p. 60) gives spawning temperatures of 5-7°C. The relative weighting of the two factors in the comfort function is as follows:

$\alpha_3=0$ for the first 80 days and thereafter

$$\lambda=\frac{\alpha_1}{\alpha_3}=\begin{cases}\frac{1}{24} & \text{if } 80\leq t<130\\ \frac{1}{1920} & \text{if } 130\leq t<180\end{cases}$$

The spawning migration is simulated over a six-month period from October 1 to April 1. The stock is initially confined to a feeding area to the northwest of Iceland, this starting distribution is chosen to resemble a typical observed October distribution off the northwest coast. The temperature effect is active throughout the period and the same applies to the current. The field generated by the spawning attraction is constant in time, but is only active from mid December, since the spawning migration usually starts in December (Vilhjálmsson, 1994, p.85). Similarly, the current field is also independent of time. Thus, once the attraction becomes active, only the temperature field is changing. However, because of the way the overall advection velocity is calculated (taking the vector sum of $\boldsymbol{V}_p$ and $\boldsymbol{V}_c$), the speed relative to the surrounding fluid (i.e. $||\boldsymbol{V}_p||$) - which changes with time - will also influence the shape of the overall velocity field.

The preferred speed $v=||\boldsymbol{V}_p||$ changes with time as follows: it is constant for the first 80 days 5 km/day, then increasing rapidly to 25 km/day according to the expression

$$v\left(t\right)=\min\left(5+25\cdot\frac{\left(t-80\right)^3}{50^3},25\right)$$

The hypothesis underlying this assumed increase is that the speed is related to the stage of maturity (i.e. roe content), which is increasing while the capelin are migrating. The fish are usually not ready to spawn until March-April and there is therefore no urgency to reach the spawning grounds for the first half or so of the period. Field observations confirm that the maturing capelin migrate slowly eastwards in October – January, while roe content is low (Vilhjálmsson, 1994, p. 129). When the roe content approaches its maximum of approximately 25% the urgency to reach the spawning grounds increases and consequently the speed as well.

The diffusion is assumed to be isotropic and constant and the value of the diffusion coefficient is set at *D*=20.00 km^2 per day. This value is more or less arbitrary but was chosen to give a reasonably realistic spread in the distribution and gives a root mean square displacement due to diffusion of vD=4.5 km per day.

The purpose with these simulations is twofold. Firstly, to demonstrate that the advection-diffusion model can give a reasonably accurate picture of the spawning migrations of capelin around Iceland using the “tools” available: a temperature field, a current field and force field generated by an attracting spawning region. Given that this can be achieved in a satisfactory way, the second objective is to investigate the effects of the various factors on the migration route. In particular, we wish to see how the fraction taking the western-or anti-clockwise - route around Iceland varies. It has been postulated is that this fraction depends on the temperature distribution north and northwest of Iceland for the following reason. The southward migration from the feeding grounds is believed to follow the eastern edge of the cold East Greenland Current. The location of this border is shifted westwards in warm years and the southward migrating fish will therefore arrive in Icelandic coastal waters further west in warm years. In addition, the preferred temperature range may also drive the capelin east in cold years. We will investigate this temperature effect in two ways: by comparing the fractions taking the western route in two different years, a cold year and a warm year and by using a more westerly initial distribution in October.

Equation (2) is solved by a numerical scheme based on a linear staggered Galerkin finite element approximation with upwinding in space and a second order Runge-Kutta approximation in time. The size of the triangle elements varies from 90.1 km^2 in the south to 68.8 km^2 furthest north and the time-step is 0.1 day. One of the benefits of the finite element approach is that we can locally take care of the fact that the motion takes place on a sphere without having to introduce spherical co-ordinates or some mapping projection. For more details see Dereksdóttir et al (2003).

Figures 7 and 8 show a few “snapshots” of the simulated spawning migrations over the six-month period from the beginning of October to the end of March in the two years selected – the cold year 1994-1995 and warm year 2000-2001. We refer to these two simulations as “base case simulations”. The maturing capelin is located northwest off Iceland on October 15. The difference between the two distributions on October 15 is due to different temperature distributions; areas with temperatures within the preferred range being more extensive in 1994 (see Fig. 5). The capelin then migrate slowly eastwards for the next two months governed only by the temperature and current fields. On February 1 the majority of the stock is at the southeast corner, but a component is taking the westerly counter-clockwise route, this component being much larger in the warm year (2001). The western component is found off the west coast at the height of the spawning season in March when the main component is spawning at the south and southwest coast. A small part of the spawning stock stays in the northern fjords to spawn as does in fact happen occasionally (see Fig. 1). In order to condense the overall spatio-temporal pictures in Figs 7-8, the proportion of the spawning stock taking the western route (P_{west}), the eastern route (P_{east}) and the proportion remaining off the north coast (P_{north}) was calculated for all the simulations described here, base case simulations as well as most of the sensitivity tests. These statistics are shown in Table 1. Note that 19% take the westerly route in the warm year, but only 2% in the cold year and that a higher number spawns on the north coast in 1995.

A few more simulations were carried out in order to test the sensitivity to some factors. Spatio-temporal pictures like Figs 7-8 are not shown here for these simulations, but such pictures are given in Dereksdóttir et al (2003). In order to investigate the importance of currents, spawning attraction and temperature in determining the motion, we removed each of the three factors –one at a time- in the 1994-1995 simulation. When the current is removed all the stock takes the western route a small part (3%) being trapped in northern fjords by cold offshore temperatures in 1995. When the temperature effect is removed (in fact making 2000-2001 identical to 1994-1995), the fraction taking the western route is between that in those years, no fish spawn on the north coast and the distribution is more spread out. On the other hand, when the spawning attraction is removed so that only temperature and currents govern the movements, the stock fails to reach the spawning grounds and is located far off the eastern coast when it should be spawning on the south coast. This might however be an appropriate scenario for immature capelin which “sometimes follow the spawning migrations to the east Icelandic area in large numbers” (Vilhjálmson, 1994, p.79).

Some further sensitivity analyses where carried out by changing the assumptions in the base case simulation. Increasing diffusion by a factor of 10 in 1994-1995 i.e. setting D=200.0 resulted in a greater spread as expected and a slightly higher fraction going west. When a more westerly initial distribution was assumed, more fish take the western route as expected, but the change in 2000-2001 is dramatic, 80% go that way.

Finally, increasing the weight of the temperature effect in the comfort function (relative to the spawning attraction) resulted in minor changes. The most notable effect is that if temperature weighs more heavily in the comfort function, then the stock will temporarily concentrate in the cooler waters off the southeast coast in February instead of moving further west towards the warmer spawning areas. This is in fact an interesting result since “real life” capelin do in fact sometimes halt on migration off the southeast coast, presumably to wait until the roe content is sufficiently high and the fish ready to spawn (Vilhjálmsson, 1994, p. 50). The fractions taking the eastern and western routes are hardly affected.

What can be deduced from these numerical experiments is the following:

- The stock fails to reach the spawning grounds without the spawning attraction
- No fish take the eastern route around Iceland if there is no current field
- The temperature field is influential in determining the fractions taking the different routes to the spawning grounds; lower temperatures send a higher fraction east
- The distribution in October – when the spawning fish arrive from the northern feeding grounds- is important, a more westerly distribution means that more fish take the western route.
- The value of the diffusion coefficient is not critical, but affects the extent of the spatial distribution and has a small influence on the size of the fraction spawning on the north coast.
- The stock halts temporarily of the southeast coast if the temperature weighs more heavily in the comfort function.

Thus, both the spawning attraction and the oceanic currents are necessary. The temperature field and the location of the spawning stock in October are crucial in determining the fractions taking the eastern and western routes to the spawning grounds.

The simulations shown here are primarily for illustrative purposes; i.e. to demonstrate that the advection-diffusion model can give a reasonably accurate picture of the spawning migrations of capelin around Iceland and to make preliminary investigations of the effect of environmental variables. Indeed, the pictures, which have been drawn up here of the overall spatio-temporal migration pattern of the simulated capelin are quite similar to what has been observed with real life capelin in the past (Vilhjálmsson, 1994, 2002), provided both the current field and the attraction field are used. Temperature does also play an important role in the fine tuning of the spatial distribution as can been seen by noting that a much larger part of the spawning stock takes the westerly route in the warm year 2000-2001 and by noting the effect of a more westerly October distribution, which is in all likelihood caused by a more westerly boundary of the East Greenland Current, which is due to warmer temperatures north of Iceland.

Compression of results

The spatio-temporal pictures (Figs 7-8) provide good qualitative descriptions of the simulated migration patterns. It may however be useful or even necessary to compress this pictorial information to a set of numbers, which can be used to make direct quantitative comparisons between different simulations as well as between simulated and observed distributions. One rather crude but informative way whereby the simulation results can be condensed is to calculate the fractions taking the different routes to the spawning grounds as described in the previous section. Another possibility is to divide the relevant area into N sub-areas based on the oceanographical, hydrographical and biological characteristics of the region. The standard division of the waters around Iceland (see Vilhjálmsson et al., 1997) is shown in Fig. 9. The simulated fraction in each sub-area can then be calculated at different points in time.

Table 2 shows the relative distribution in each of the sub-areas at four different times for the two base case simulations. The influence of temperature is apparent one month after the start of the simulation: in the warm year 2000 nearly all the stock is concentrated in sub-area 11, whereas in the cold year 1994, 32% are found in sub-area 3, nearer the coast. On the other hand, on February 15 the stock is slightly more concentrated in the cold year, with sub-area 2 having a reasonable fraction of the stock during the warm year 2001 whereas only a small fraction of the stock is in compartment 2 in the colder year 1995. These fractions resemble the proportions of the stock that takes the westerly route each year.

Another noticeable difference in the distributions can be seen on January 1. A considerably larger proportion of the stock has reached sub-area 6, which is due east of the island, during the cold year 1995 than during the warm year 2001 when most of the stock is still in sub-area 3 north of the island. This behaviour can be directly related to temperature. Because of the cold front coming in from the north as winter progresses the stock is pushed further south in early 1995 than in 2001. Temperatures north of Iceland in January 1995 were around 1-2°C, but around 5°C in 2001. On April 1, there is a sizeable component off the west coast in the warm year but only a negligible fraction in 1995.

This migration model may ultimately constitute a part of a larger multi-species model incorporating migrations, recruitment, predation, catches, growth, etc. However, the overall computing requirements are too great at present. It therefore becomes essential to try to reduce the continuous model to a simpler one, which is easier to use and less computer intensive.

One way whereby this can be achieved is converting the continuous model into an equivalent transition matrix model (also referred to as migration matrix model). A transition matrix at time *t* is an *NxN* matrix *M(t)* where the element in row *i* and column *j*, *m(i,j,t)*, is the fraction of fish in sub-area *i* which moves to sub-area *j* between times *t* and *t*+1 (this time-step is typically one or two months). If $\boldsymbol{n}(t)$ is the distribution vector at time *t* (i.e. the density in each of the *N* sub-areas) then $\boldsymbol{n}(t+1)$ is obtained by

$$\boldsymbol{n}^T\,(t+1) = \boldsymbol{n}^T\,(t)\,M(t)$$

While the partial differential equation governing the evolution of the density *p* (equation (2)) remains linear we can apply the principle of superposition, i.e. if p_i is the solution corresponding to initial condition p_{0i}, then $\sum_i p_i$ is the solution corresponding to initial condition $\sum_i p_{0i}$. We can therefore calculate each of the rows in the transition matrix separately as follows. Assuming that all the mass is initially confined to sub-area nr. 1 (see Fig. 9), we solve equation (2) and after one time-step, compute the fraction of the initial mass in each of the *N* sub-areas. The fraction in sub-area *j*, *j*=1,2,…,*N*, is element (1,j) in the transition matrix. This exercise therefore gives the first row in the transition matrix at time *t*. Repeating for sub-areas 2,3,…,*N*, the transition matrix is obtained. The initial distribution within each sub-area is set on the basis of the temperature distribution, i.e. the fish are concentrated (10 times higher density) in the areas where the temperature is within the preferred range.

Using the continuous simulations with the settings described above, the six migration matrices for October through to March with a time step of one month were computed and these in turn used to compute the relative density in each of the sub-areas using the same initial October distribution as for the continuous model. As a test of the sensitivity of the results to the size of the time-step two transition matrices with a time step of three months were also computed and the relative densities calculated using these matrices.

Furthermore, in order to get some indication of the sensitivity to the spatial resolution, the initial distribution was taken to be uniform over the whole sub-area when calculating the migration matrices. Setting the initial distribution in this way can cause problems since the sub-area division may not be the most suitable for calculating migration matrices, in particular the very large sub-areas outside the coastal areas. Table 3 shows the relative densities on January 1 and April 1 in 1995 and in 2001 as computed by the continuous model and by using one-month matrices and three-month matrices. These matrices are computed using either a preferred or a uniform distribution within each sub-area.

The matrix model reproduces the simulated distribution on January 1 1995 fairly well, less so in 2001 the main difference being that more of the stock has advanced to sub-area 6 off the east coast by the matrix model while 43% is still in the northern sub-area 3 according to the continuous model. The agreement is also fairly good on April 1 1995 and 2001 for the matrix model with a time step of three months, but for a one-month time step a higher fraction ends up in the southeast sub-area 9 and a correspondingly lower fraction in the western sub-areas 1 and 10 compared to the continuous model. The difference between the models with matrices based on preferred and uniform distributions is small.

Note that while the main fraction of the initial distribution in each sub-area coincides with the areas of preferred temperature when calculating the matrices as described above the distribution is taken to be uniform within these preferred areas and the fish are therefore effectively re-distributed within each sub-area according to the preferred temperature at the beginning of each transition step. Furthermore, the temperature in some sub-areas lies entirely outside the preferred range and the distribution is taken to be uniform over the whole sub-area in that case. We also force the fish into sub-areas where it is never found, e.g. there is hardly any fish in sub-areas 14 and 15 in the continuous simulations, but when calculating the migration matrices it is put into these areas at the beginning of the transition step.

In general, if the size of the compartments is large compared to the distance travelled per unit of time, then transition matrix models will tend to exaggerate the spread compared to an advection-diffusion model. Furthermore, within large sub-areas environmental heterogeneity can have a substantial influence on the motion in a continuous model, which cannot be addressed in a matrix model. These considerations emphasize the importance of the choice of sub-areas, especially their size in relation to the speed of the fish. It seems tempting to try to include more a-priori information about the whereabouts of the fish within each sub-area at the start of a transition step. However, it must be kept in mind that the aim is to use the transition model for predictive purposes and care must be taken not to introduce bias into these predictions. Clearly some balance must be struck between the sizes of the sub-areas (keeping their total number small) and the time-steps of the transition process that matches the speed of the fish. The outer sub-areas shown in Fig. 9 are in all likelihood too large for the purpose of a transition matrix model of capelin migrations.

The essential purpose of the advection-diffusion model is to predict migrations conditional on environmental conditions. When compressing the continuous model into a migration matrix model, the matrices must therefore depend on environmental conditions. Rather than estimating the matrices directly from observations, as is commonly done in such models, we propose to use the observations to estimate values of coefficients in the advection-diffusion model, which are independent of these conditions, e.g. the functions $\alpha_i\left(t\right)$, *i*=1,2,3 in equation (5), the constants T_1 and T_2 in equation (6) and the speed $v\left(t\right)$. This reduces the number of values to be estimated and increases the number of observations behind each parameter estimate. The transition matrices *M(t,q)* for a specified environmental category *q*, e.g. a warm, average or cold year, can then be obtained from the advection-diffusion model as described above.

Discussion

The migration pattern of capelin in Icelandic waters is quite complex and variable, depending on environmental conditions. It is therefore of some interest to see if these migrations can be satisfactorily modelled and the model used to predict the route and timing of the migration and to test hypotheses about the effect of the environmental variables. The model described in this paper is an advection-diffusion type model, where the advection is determined by environmental fields (temperature and currents) and a force field drawing the spawners towards the spawning grounds. The main objective is to investigate whether such a model can give a realistic picture of the spawning migration of capelin around Iceland. From the preliminary results it would appear that this is indeed the case. What can be deduced from the numerical experiments described in the previous section is that both the spawning attraction and the oceanic currents are required.

Furthermore, it seems that what determines the variations in the fraction taking the western route to the spawning grounds - apart from the currents which are the same in all simulations - are the temperature fields and the east-west location of the pre-spawning stock off the northwest peninsula in October-November. In reality, the return migration from the northern feeding grounds to the waters off northwest Iceland is usually “along the eastern border of the East Greenland – Iceland current, in mild years even reaching westwards over the outer part of the East Greenland shelf” (Vilhjálmsson, 2002). Furthermore, Vilhjálmsson (1994, p. 58) suggests that the “migration directly south and southeast towards the west and/or southwest coast spawning grounds is associated with a western distribution of the spawning stock in late autumn and early winter”. This hypothesis seems to be borne out by the simulations presented in this paper. It may therefore be tentatively concluded that the location of the eastern border of the cold East Greenland-Iceland current and consequently the temperature field in the waters northwest of Iceland – i.e. on the Iceland- Greenland ridge and in the Iceland-Greenland Channel – is highly important in determining the size of the fraction taking the western route.

If the true distribution of capelin in October/November is determined by the temperature fields north and northwest of Iceland, as it would appear, then it is essential to have good temperature measurements from this region. However, actual measurements from the offshore areas are lacking; the temperature fields for these areas used in the simulations, were obtained by extrapolations and some guesswork. Since temperature and currents are the only factors determining the spatial distribution before the spawning attraction becomes active in December, it is clear that reasonably accurate temperature data from the regions where the capelin returning from the feeding grounds are found in October – November are critical. In fact, more detailed environmental fields are desirable. Ideally, measurements of both temperature and currents should be coupled to an appropriate hydrographical model in order to interpolate the measurements both in time and space. Such models should become available in the not too distant future and will then also provide more detailed current values to use as input into this migration model.

A few limitations of the present model are apparent, apart from the shortcomings of the oceanographic data available at present. Firstly, only the last six months in the life cycle of the capelin have been modelled, that is the spawning migration from the waters between Greenland and Iceland to the spawning ground on the south coast. The feeding migrations of the maturing stock north towards Jan Mayen have not been considered. At present, it does not appear feasible to extend the simulations much further into this region since virtually no data exist with which to calibrate the model, with the exception of a small amount of capelin data from the 1970's. This part of the migration cycle is therefore likely to remain a closed book for the time being at least. The distribution of the immature capelin can however be modelled, since it does not migrate north to the same extent as the maturing stock and is surveyed regularly on feeding grounds off the north coast. This part of the stock does not experience any attraction to spawning grounds and its distribution is probably only governed by the density of food in addition to currents and temperature.

A physiological variable – the stage of maturity is the primary candidate - is an obvious omission in the present model. However, the migration speed may be linked to the stage of maturity (roe content), which increases slowly to begin with (the speed also being slow at that stage of the migration) and faster as the time of spawning in March-April approaches. Since we assume a similar pattern of change in migration speed, it may be argued that a maturity variable has been implicitly included in the model. Nevertheless, it is probably necessary to model the stage of maturity explicitly and attempt to link this variable to the speed in a more rigorous manner.

The simulations in this paper are based on a crude tuning of the model, simply choosing parameter values so as to give a spatio-temporal picture corresponding roughly to what is known about past spawning migrations. However, this is of little use in predicting future migrations. What is required is a proper statistical estimation of key model parameters. The distribution of the spawning stock in October and in January/February has been charted in acoustic surveys since the 1970's. Furthermore, the distributions of the catches are also available. This data is now being prepared in a suitable format at the appropriate spatial resolution. The advection – diffusion model is solved by finite element methods on a triangular grid, which gives the computed density in each of the triangles. The survey data will also be given on the same triangular grid, allowing a comparison between the computed and observed densities. The observed distribution in October is taken as the initial distribution and the distribution simulated until January/February, the time of the next survey. A comparison between the computed and observed distributions can then be made. A possible objective function is

$$\Phi\left(\theta\right) = \sum_i \left(p_i^{comp} - p_i^{obs}\right)^2$$

where p_i^{comp} and p_i^{obs} are the computed and observed relative densities in triangle #i in January/February and θ is a set of parameters. The summation is over all triangles where there are observations, zero or otherwise. The aim is to minimize this function with respect to key model parameters, such as λ, the weighting of the temperature effect relative to the spawning attraction; some parameters specifying the time dependence of the speed; the preferred temperature range etc. Note that the computed distribution in January/February is *conditional* on the given distribution in October.

Since the computed densities are probability densities, another possibility is to us a likelihood function. Assume that there are *N* fish in October. Then the likelihood of the observed distribution in January/February is (using a multinomial distribution)

$$L(\theta) = \frac{N!}{\prod\limits_i n_i!} \prod_i p_i^{comp}(\theta)^{n_i}$$

where n_i is the number of fish in found in triangle #i. The log-likelihood function is

$$\ln L(\theta) = \sum_i n_i \ln p_i^{comp}(\theta) - \sum_i \ln(n_i!) + \ln(N!)$$

Since the last two terms do not depend on θ, the function to be maximized is

$$l(\theta) = N \sum_i \tfrac{n_i}{N} \ln p_i^{comp}(\theta) = N \sum_i p_i^{obs} \ln p_i^{comp}(\theta)$$

Note that this function has a maximum if $p_i^{comp}(\theta) = p_i^{obs}$ for all *i*. We have ignored the mortality in the period between the surveys, but assuming that the numbers in each triangle are affected proportionately in the same way, this should not matter.

Once estimates of the parameters are obtained, the tuned model can be used to test hypotheses and predict migrations conditional on environmental conditions. There are strong indications that the migration pattern of mature capelin has been changing in the past few years and that this may be linked to a warming of the waters north off Iceland. This highlights the importance of a simulation model linking migration patterns to oceanographic conditions, which can be used to test such hypotheses and to predict the spatial distribution and migration route.

References

P. Babak, K.G. Magnússon and S. Th. Sigurdsson. Dynamics of group formation in collective motion of organisms. Mathematical Medicine and Biology, in press.

A. Czirok and T. Vicsek. 2000. Collective behaviour of interacting self-propelled particles. Physica A 281: 17-29.

R.J.H. Beverton and S.J. Holt. 1957. On the dynamics of exploited fish populations. Fish. Invest. Ser. II. Mar.Fish. G.B. Minist. Agric. Fish. Food, **19** (1957**)**.

E.H. Dereksdóttir, K.G. Magnússon and S. Sigurdsson. 2003. Simulations of migrations of capelin in the Icelandic waters based on temperature-, current- and potential fields. Report RH-17-2003, Science Institute, University of Iceland. (Available at http://www.raunvis.hi.is/∼kgm/).

R.B. Deriso, R.G. Punsly, and W.H: Bayliff. 1991. A Markov movement model of yellowfin tuna in the Eastern Pacific Ocean and some analysis for international management. Fisheries Res. 11: 375-395

H. Gjøsæter. 1998. The population biology and exploitation of capelin (Mallotus villosus) in the Barents Sea. Sarsia 83: 453-496.

S. Hubbard, Petro Babak, Sven Sigurdsson and K.G. Magnússon. A model of the formation of fish schools and migrations of fish. Ecological Modelling 174(2004): 359-374

A.D. MacCall, 1990. *Dynamic Geography of Marine Fish Populations.* University of Washington Press, Seattle

K.G. Magnússon, S. Sigurdsson, P. Babak, S.F. Gudmundsson and E.H. Derekesdóttir. A continuous density Kolmogorov type model for a migrating fish stock. Discrete and continuous dynamical systems

– Series B 4(3): 695-704

O.A.Misund, H. Vilhjálmsson, S.H. i Jákupsstovu, I. Røttingen, S. Belikov, Ó. Ástthorsson, J. Blindheim, J. Jónsson, A. Krysov, S.A. Malmberg, S.A., and S. Sveinbjørnsson. Distribution, migration and abundance of Norwegian spring spawning herring in relation to the temperature and zooplankton biomass in the Norwegian Sea as recorded by coordinated surveys in spring and summer 1996. Sarsia 83 (1998), 117-127.

H.-S. Niwa. 1996. Newtonian dynamical approach to fish schooling. J. theor. Biol. 181: 47-63.

A. Okubo. 1986. Dynamical aspects of animal grouping: swarms, schools, flocks and herds. Adv. Biophys. 22:1-94.

A. Okubo and S.L. Levin. 2001. *Diffusion and Ecological Problems: Modern Perspectives*. Springer

B. Øksendal. 2000. *Stochastic Differential Equations*. Springer.

B.L. Partridge. 1982. The structure and function of fish schools. Scientific American 246: 90-99. June 1982

T.J. Quinn and R.B. Deriso. 1999. *Quantitative fish dynamics*. Oxford University Press.

M.J. Reed and J.G. Balchen. 1982. A multidimensional continuum model of the fish dynamics and behaviour: Application to the Barents Sea capelin (*Mallotus villosus*) Modelling, Identification and Control 3: 65-109.

J.R. Sibert, J. Hampton, D.A. Fournier, and P.J. Bills. 1999. An advection-diffusion-reaction model for the estimation of fish movement parameters from tagging data with application to skipjack tuna (Katsuwonus pelamis). Can. J. Fish Aquat. Sci. 56: 925-938.

S. Sigurdsson, K.G. Magnússon, P. Babak, S.F. Gudmundsson and E. H. Dereksdóttir. 2002. Dynamic continuous model of fish migrations. Report RH-25-2002, Science Institute, University of Iceland, http://www.raunvis.hi.is/∼kgm/

S. Tjelmeland and B. Bogstad. 1998. *Biological Modelling*, in “Models for Multispecies Management”, (ed. T. Rødsethe). Physica-Verlag..

H. Vilhjálmsson. 1994. The Icelandic capelin stock. Journal of the Marine Research Institute Reykjavik, Vol. XIII, No. 1. 281 pp.

H. Vilhjálmsson. 2002. Capelin (Mallotus villosus) in the Iceland-East Greenland-Jan Mayen ecosystem. ICES J. Mar. Sci. 59: 870-883.

H. Vilhjálmsson, H.N. Stefánsson, A.B. Lund, H. Sigurgeirsson and H. Björnsson. 1997. Capelin migrations, in BORMICON A Boreal Migration and Consumption Model, G. Stefánsson and Ó.K. Pálsson (eds). Marine Research Institute Report 58.

Table 1. Fractions taking the western and the eastern route to the spawning grounds and the fraction spawning in the north for the different simulation cases.

Simulation	P(west)	P(east)	P(north)
1994-1995: base case	0.02	0.93	0.05
2000-2001: base case	0.19	0.80	0.01
1994-1995: no current	0.97	0.00	0.03
No temperature effect	0.09	0.91	0.00
1994-1995: 10-fold diffusion	0.05	0.94	0.01
1994-1995: westerly Oct. distrb.	0.16	0.77	0.07
2000-2001: westerly Oct. distrib.	0.80	0.19	0.01

Table 2. The relative distribution (%) in each of the sub-areas at four selected times for the two base case simulations. The initial condition is the same in the two years.

Sub-area	Initial distrib.	November 1		January 1		February 15		April 1	
		1994	2000	1995	2001	1995	2001	1995	2001
1	0	0	0	0	0	0	2.2	0.4	7.8
2	8.1	0.4	0	0.1	0	1.7	13.5	0.1	2.3
3	22.6	32.3	1.8	12.4	43.5	0	0	0	0
4	0	0.1	0	4.0	1.2	2.7	3.2	1.9	0.6
5	0	11.7	1.0	10.6	27.0	6.2	0.6	1.6	0
6	0	1.1	0.1	49.5	15.1	5.2	2.1	1.1	0.1
7	0	0	0	9.4	0.1	0	0	0	0
8	0	0	0	5.4	0.1	1.3	1.0	0.3	0
9	0	0	0	1.0	0	81.5	74.4	10.9	8.7
10	0	0	0	0	0	0.5	0.1	79.9	77.5
11	69.0	52.7	96.6	0.2	10.3	0	0	0	0
12	0	1.5	0.4	6.8	2.6	0	0	0	0
13	0	0	0	0.6	0.0	0.9	2.9	3.7	2.9
14	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0
16	0.3	0.1	0.1	0	0	0	0	0	0

Table 3. The relative density in each of the sub-areas from the advection-diffusion model and migration matrix models with time steps one month and three months. The column heading 3x1 refers to results using three one-month matrices and so on.

1994-1995		January 1					April 1				
Sub area	Initial distrib	Cont. model	Preferred		Uniform		Cont. model	Preferred		Uniform	
			3x1	1x3	3x1	1x3		6x1	2x3	6x1	2x3
1	0	0	0	0	0	0	0.4	0.5	0.7	0.8	0.9
2	8.1	0.1	1.2	0.2	1.7	0.3	0.1	0.1	0.1	0.1	0.2
3	22.6	12.4	8.6	13.5	11.0	16.0	0	0	0	0	0
4	0	4.0	2.4	4.5	2.5	5.3	1.9	0.6	2.1	0.8	2.5
5	0	10.6	9.7	9.7	9.6	10.4	1.6	0.1	1.2	0.1	1.4
6	0	49.5	44.8	46.9	42.5	43.4	1.1	0.1	0.9	0.1	1.0
7	0	9.4	13.7	8.8	10.6	8.1	0	0	0	0	0
8	0	5.4	4.5	4.5	4.1	4.0	0.3	0	0.2	0	0.3
9	0	1.0	1.7	0.8	2.4	0.7	10.9	20.0	10.4	19.2	10.8
10	0	0	0	0	0.1	0	79.9	73.2	80.7	73.4	79.2
11	69.0	0.2	1.4	0.2	1.8	0.2	0	0	0	0	0
12	0	6.8	10.6	9.9	11.7	10.2	0	0	0	0	0
13	0	0.6	0.6	0.5	0.5	0.4	3.7	5.4	3.6	5.5	3.7
14	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0.7	0.5	1.5	0.9	0	0	0	0	0
16	0.3	0	0	0	0	0	0	0	0	0	0

2000-2001		January 1					April 1				
Sub area	Initial distrib	Cont. model	Preferred		Uniform		Cont. model	Preferred		Uniform	
			3x1	1x3	3x1	1x3		6x1	2x3	6x1	2x3
1	0	0	0	0	0	0	7.8	2.9	4.6	4.9	7.6
2	8.1	0	0.5	0.8	1.0	0.9	2.3	0.5	1.4	1.0	2.5
3	22.6	43.5	21.0	27.5	25.0	30.7	0	0	0	0	0
4	0	1.2	0.3	0.7	0.3	0.6	0.6	0	0.4	0	0.7
5	0	27.0	9.3	15.2	9.4	13.6	0	0	0	0	0
6	0	15.1	32.8	28.0	24.7	24.8	0.1	0	0	0	0.1
7	0	0.1	4.1	0.5	3.4	0.5	0	0	0	0	0
8	0	0.1	0.6	0.3	0.6	0.2	0	0	0	0	0
9	0	0	0.2	0	0.7	0	8.7	19.1	9.0	17.1	8.5
10	0	0	0	0	0	0	77.5	72.6	81.5	72.5	77.6
11	69.0	10.3	9.0	8.5	11.3	10.8	0	0	0	0	0
12	0	2.6	18.1	17.5	15.3	16.4	0	0	0	0	0
13	0	0	0	0	0.1	0	2.9	4.8	3.0	4.3	2.8
14	0	0	0	0	0	0	0	0	0	0	0
15	0	0	4.0	1.1	8.1	1.3	0	0	0	0	0
16	0.3	0	0	0	0	0	0	0	0	0	0

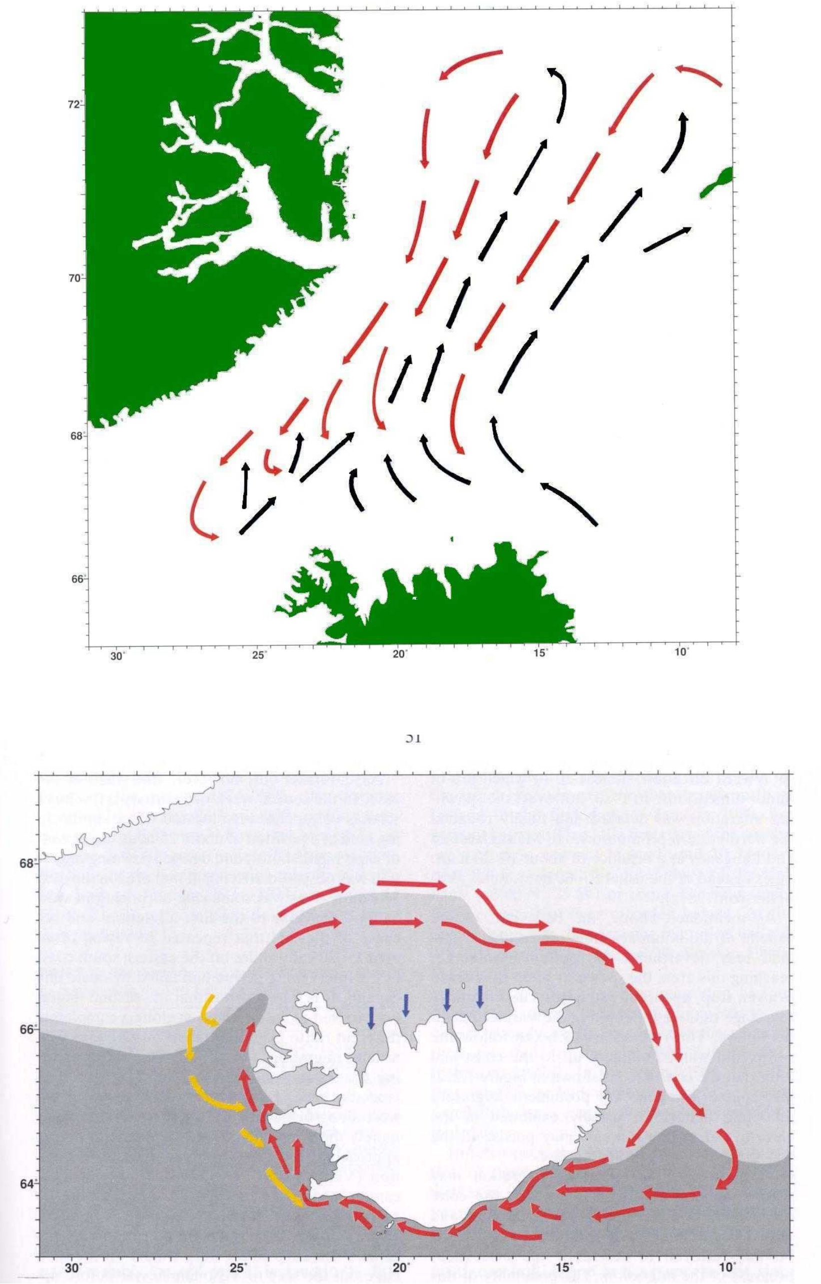


Figure 1. Feeding and spawning migrations of capelin in the Central North-Atlantic (from Vilhjálmsson, 1994).

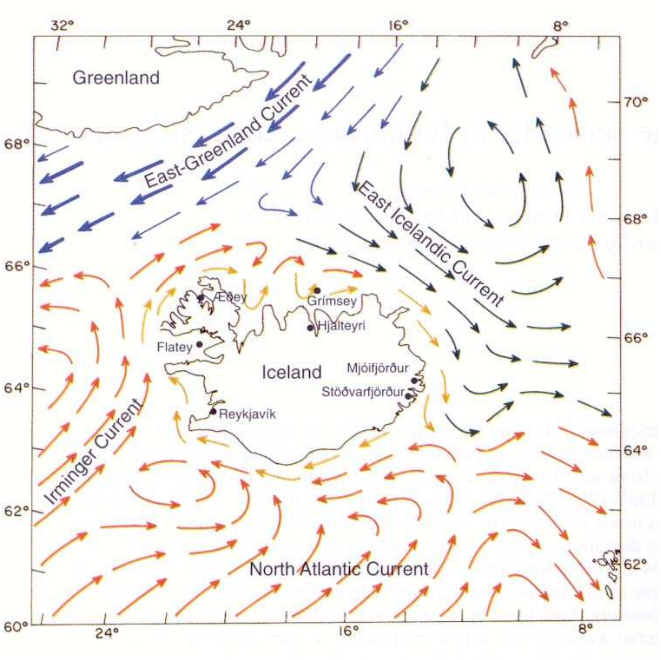


Figure 2. Oceanic currents around Iceland (from Vilhjálmsson (1994)).

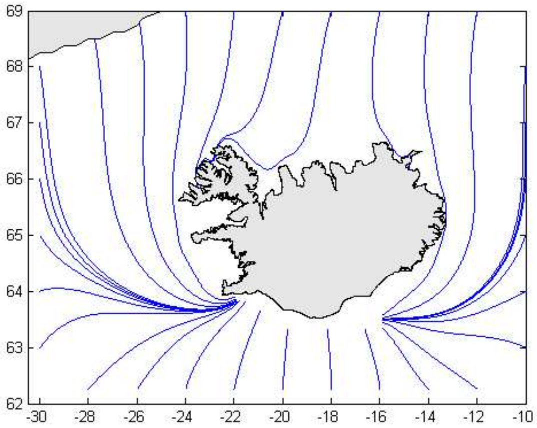


Figure 3. Iceland and the field lines for the attraction force generated by a spawning area on the south coast.

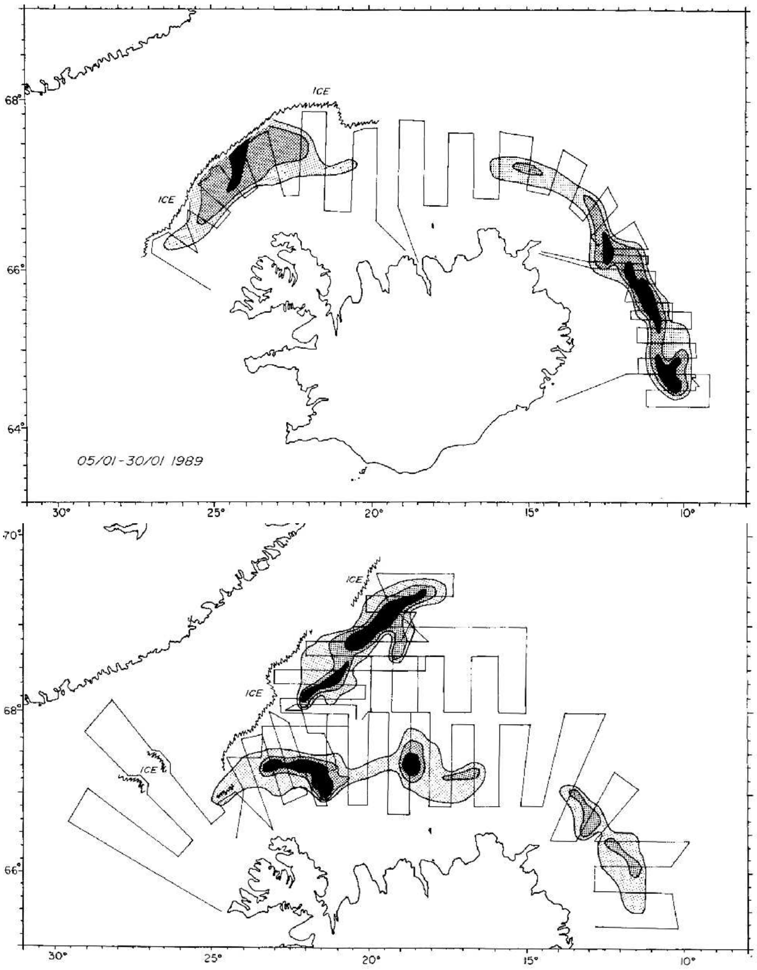
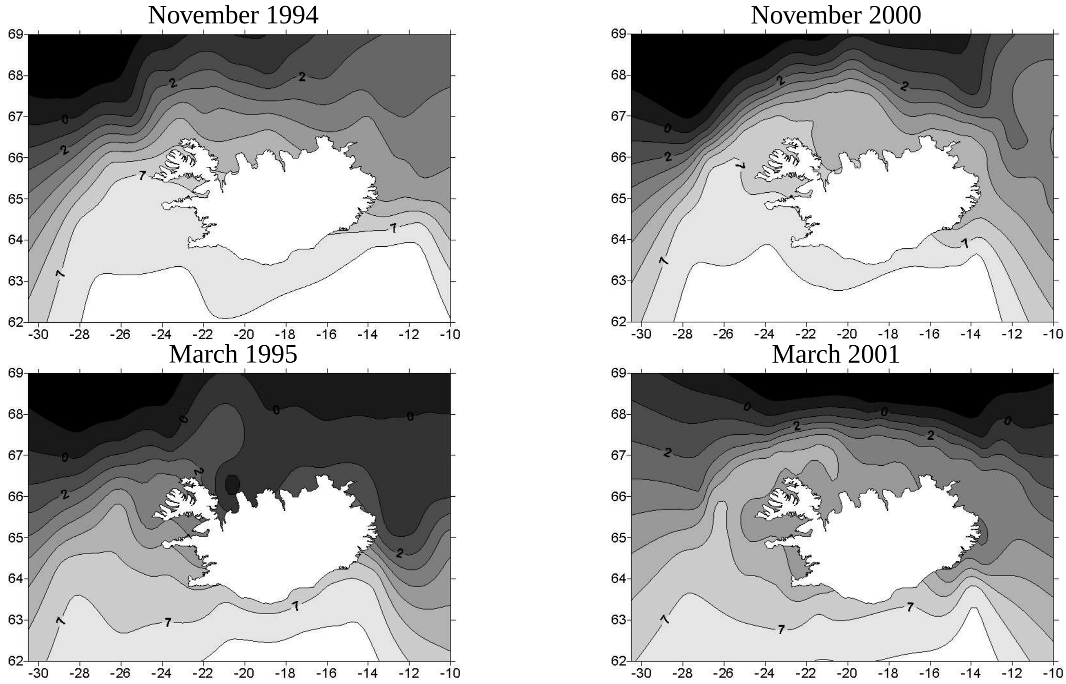


Figure 4. Mesaured capelin distribution and survey tracks in October 1988 (upper) and January 1989 (lower) (from Vilhjálmsson, 1994)


November 1994

November 2000

March 1995

March 2001

Figure 5. Temperature fields in November 1994 and 2000 and March 1995 and 2001.

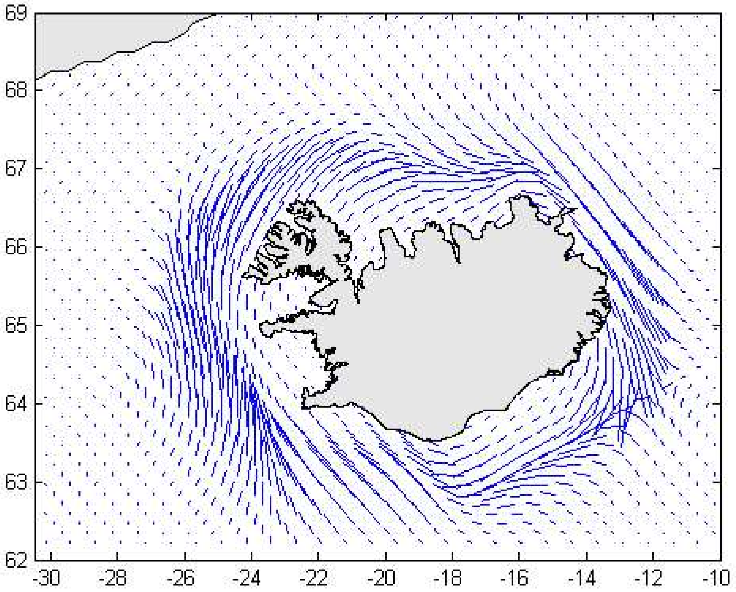
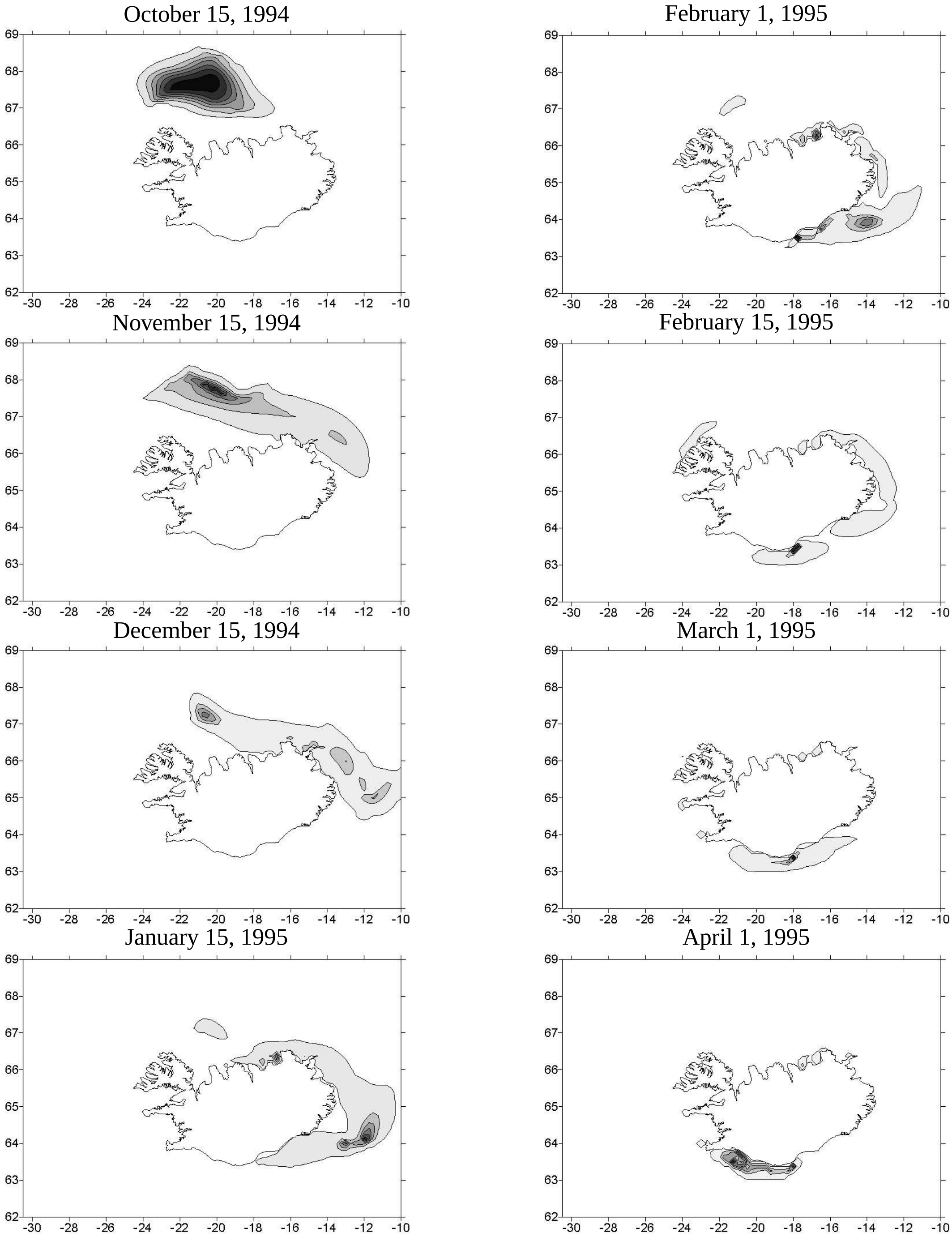


Figure 6. Hypothetical field of oceanic currents around Iceland. The speed of the current is given by the length of the arrows, i.e. darker colours indicate greater speed.



October 15, 1994

February 1, 1995

November 15, 1994

February 15, 1995

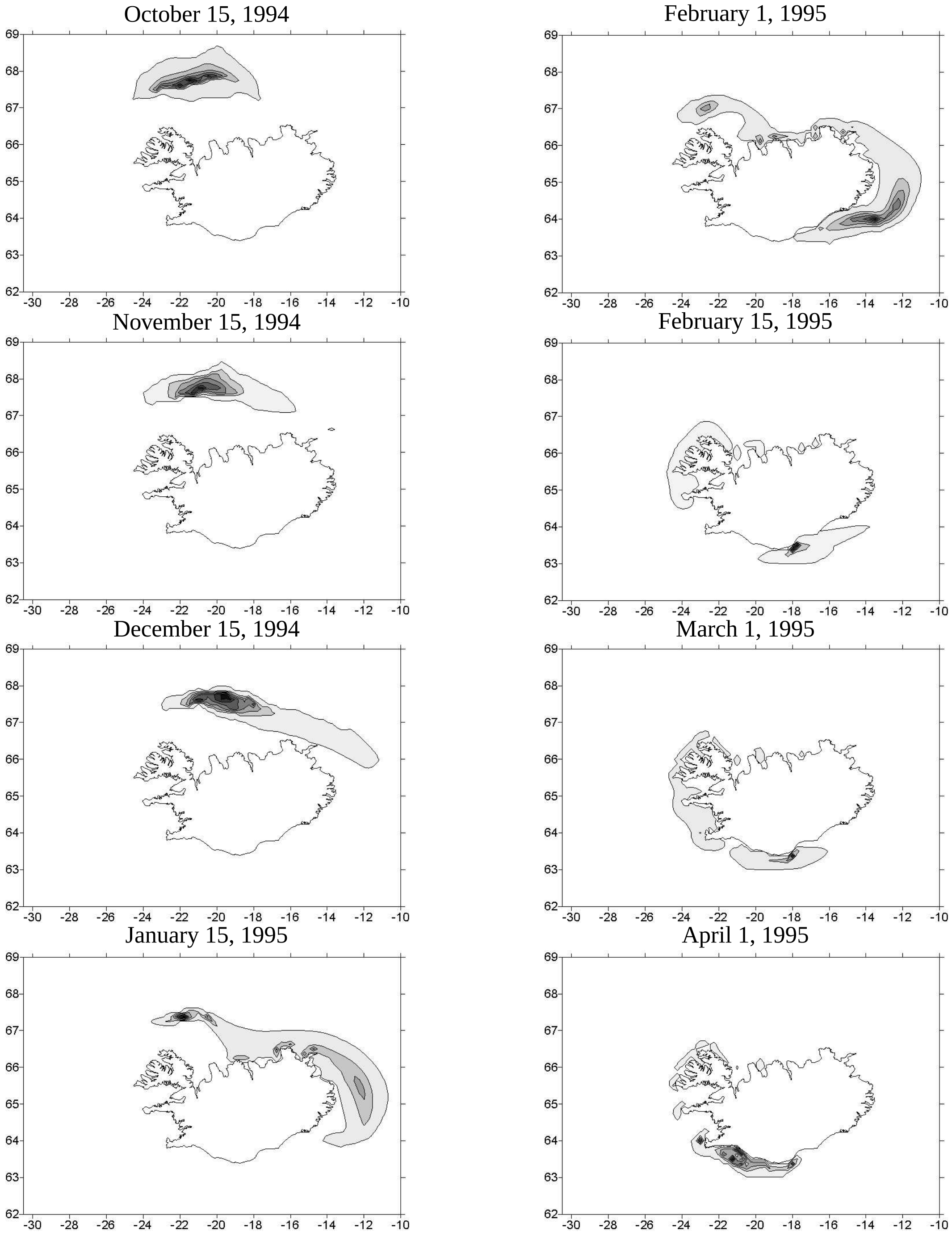
December 15, 1994

March 1, 1995

January 15, 1995

April 1, 1995

Figure 7. Simulated spawning migration of capelin around Iceland from October 1 1994 to April 1 1995 using fields of temperature, spawning attraction and oceanic currents (base case simulation). P$_{west}$=2%, P$_{east}$= 93%, and P$_{north}$= 5% (see main text).


October 15, 1994

February 1, 1995

November 15, 1994

February 15, 1995

December 15, 1994

March 1, 1995

January 15, 1995

April 1, 1995

Figure 8. Simulated spawning migration of capelin around Iceland from October 1 2000 to April 1 2001 using fields of temperature, spawning attraction and oceanic currents (base case simulation). P$_{west}$= 19%, P$_{east}$= 80% and P$_{north}$= 1%

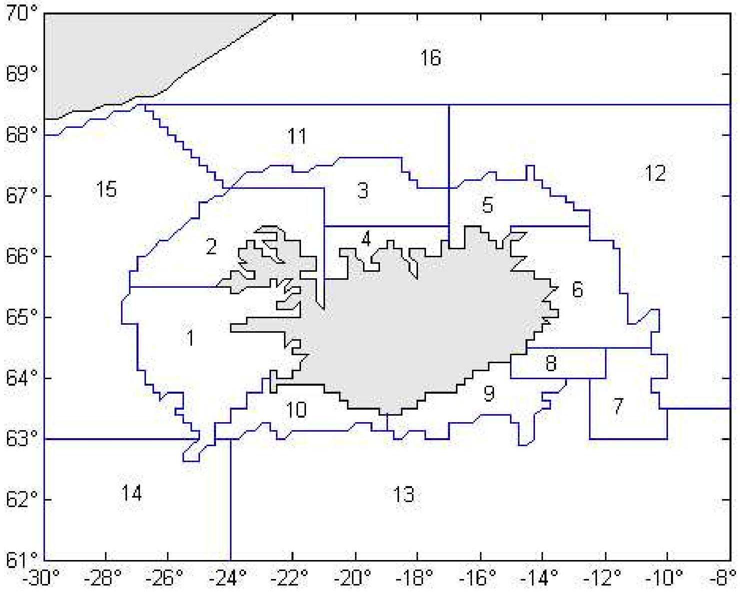


Figure 9. The sub-areas used in the transition matrix migration model for the waters around Iceland

7.4 Estimation of fish movement parameters from tagging data, with application to Atlantic cod (*Gadus morhua*)

Petro Babak, Kjartan G. Magnússon and Sven Th. Sigurdsson
Science Institute, University of Iceland, Dunhaga 3, 107 Reykjavík, Iceland

Abstract: The conditional probability density for the location of fish can be modelled by Kolmogorov's forward equation. The parameters of such a model are estimated from Cod tagging data around Iceland during the period from 1993 to 1997. The movement parameters in the Kolmogorov model are advection - velocity and diffusion parameters. These parameters are estimated based on compartmental divisions around Iceland and on the given time divisions. Three time intervals are considered within each year: spawning migration from January 1 until April 30, spawning from May 1 until May 31 and feeding migration from June 1 until December 31.
The maximum likelihood function is the main tool in parameter estimation. The likelihood function is calculated from the probability density function, which is obtained from the Kolmogorov's model.

Keywords: fish migration, advection-diffusion equation, cod, parameter estimation.

Model description

The movement of a single particle-fish can be described by a two dimensional stochastic differential equation

$$d\boldsymbol{X}_t = \boldsymbol{V}\left(\boldsymbol{X}_t, t\right) dt + B\left(X_t, t\right) d\boldsymbol{W}(t) \qquad (1)$$

where $\boldsymbol{X}_t = (X, Y)$ is the random variable for the position of the fish, $\boldsymbol{V} = (v_1, v_2)$ is a given velocity, *B* is a given 2x2 matrix and *W(t)* is a continuous 2 dimensional Brownian motion. The conditional probability density at time *t* for the position of the fish, given that the fish is at (x_0, y_0) at time t_0, $p = p\left(x, y, t | x_0, y_0, t_0\right)$, satisfies Kolmogorov's forward equation (Øksendal, 2000)

$$\frac{\partial p}{\partial t} = -\nabla \cdot (\boldsymbol{V} p) + \frac{1}{2}\left(\frac{\partial^2 (D_1 p)}{\partial x^2} + 2\frac{\partial^2 (D_{12} p)}{\partial x \partial y} + \frac{\partial^2 (D_2 p)}{\partial y^2}\right), \quad (x, y) \in \Omega \quad (2)$$

where

$$BB^T = \begin{bmatrix} D_1 & D_{12} \\ D_{12} & D_2 \end{bmatrix}$$

If ***V*** and *B* are constant, then it is easy to see that

$$E\left[\boldsymbol{X}_t - \boldsymbol{X}_0\right] = \boldsymbol{V} \cdot t$$

and

$$Cov\left[\boldsymbol{X}_t\right] = \left(BB^T\right) \cdot t$$

Equation (2) resembles an advection – diffusion equation but differs from it in that it contains a mixed second derivative term and the second derivative is taken of the multiple of the diffusion coefficients *D* and the density *p*. If the diffusion is isotropic then $D_{12} = 0$ and $D_1 = D_2 = D$, say, and equation (2) reduces to the equation:

$$\frac{\partial p}{\partial t} = -\nabla \cdot (\boldsymbol{V} p) + \frac{1}{2}\left(\frac{\partial^2 (Dp)}{\partial x^2} + \frac{\partial^2 (Dp)}{\partial y^2}\right) \qquad (3)$$

We will employ this simpler model in the work described here, since it is not possible to distinguish between models (2) and (3) on the basis of the available measurements, but our numerical implementation of the model, using finite elements can readily be extended to the more general case. We will also assume for simplicity that the diffusion coefficient is constant – in which case equation (3) reduces to an advection-diffusion equation.

If we assume that there is zero probability of a fish escaping through the boundary of Ω, the natural boundary condition becomes

$$\left(-v_1 p + \frac{\partial}{\partial x}(D_1 p) + \frac{\partial}{\partial y}(D_{12} p)\right) n_1 + \left(-v_2 p + \frac{\partial}{\partial y}(D_2 p) + \frac{\partial}{\partial x}(D_{12} p)\right) n_2 = 0, (x,y) \in \partial\Omega$$

where (n_1, n_2) denotes the outward unit normal vector to the boundary. Under the simplifying assumptions about the diffusion coefficients, this becomes simply

$$-(\mathrm{V} \cdot \mathrm{n})\, p + \frac{1}{2}(\nabla (Dp) \cdot n) = 0, \qquad (x,y) \in \partial\Omega \qquad (4)$$

Finally, an initial conditon at time $t = t_0$,

$$p(t,x,y)|_{t=t_0} = p_0(x,y) \qquad (5)$$

has to be specified.

When using this model for simulations the parameters ***V*** and *D*, that may in general depend both on time and location, have to be specified. Our approach has been to view ***V*** as the vector sum of the velocity of the fish relative to the surrounding waters and the velocity of the water current, denoted by $\boldsymbol{V}_p$ and $_c$ respectively. The direction of the velocity vector $\boldsymbol{V}_p$ has, in turn, been determined by the gradient of a function $U(x,y;t)$, which we have referred to as a comfort function (Reed and Balchen, 1982). This function incorporates the factors, which are believed to affect the "comfort" or well-being of the fish, such as temperature, food density, location relative to the spawning grounds, etc. The fish move on a time-varying "comfort surface" defined by $U = U(x,y;t)$ constantly attempting to maximize their comfort, i.e. moving in the direction of the spatial gradient of *U*. Typically, the comfort function is a linear combination of functions of temperature and food density and a potential function

$$U(x,y;t) = \alpha_1(t)\, r(T(x,y;t)) + \alpha_2(t)\, s(f(x,y;t)) + \alpha_3(t)\, \phi(x,y) \qquad (5)$$

where *T(x,y;t)* and *f(x,y;t)* are the temperature and food density respectively at location *(x,y)* at time *t* , *r* and *s* are given functions, and the function ϕ is constructed in such a way that it draws the fish towards specified spawning grounds, around any obstacles that may be in its way (Dereksdottir et al., 2003). The timedependent α - coefficients are specified so that their relative weights vary, depending on whether the time corresponds to eg. a spawning period or a feeding period, and the same holds true for the specified speed $v = \|\boldsymbol{V}_p\|$

The specification of the diffusion parameter, *D*, has been somewhat more speculative, and we have often opted for a single constant value.

The model described above does not take into account mortality. This can, however, readily be done by adding the term $-Zp$to the right hand side of (2) or (3), where the parameter *Z* is the sum of fishing mortality and natural mortality, and will in genera like ***V*** and *D* depend both on location and time. In this case the interpretation of the conditional probability, $p = p\left(x, y, t | x_0, y_0, t_0\right)$, changes. It is now the probability density at time *t* for the fish having survived AND being at position (x, y), given that the fish is at (x_0, y_0) at time t_0.

Parameter estimation

We have applied the model described above, without a mortality term, to capelin migrations both in the Barent's sea and the waters around Iceland, where the parameters in the comfort function (5) have been chosen from computational experimentation aimed at getting results that agree qualitatively with observed migration patterns (Magnusson et al., 2004, Dereksdottir et al., 2003). Note that by estimating the parameters of a comfort function, rather then those of the differential equation, we are eliminating the changes caused by changes in environmental conditions. The implication of this is, firstly, that we can include data sampled under different environmental conditions in a single parameter estimate, and, secondly, that the model becomes more robust for predictions of migration patterns under new environmental conditions.

We are now interested in extending this work in two ways. Firstly, to establish a more quantitative appraoch to the parameter estimation, and secondly to try to apply it to other fish species, in particular to cod (Gadus morhua). One of the potential advantages of estimating parameters quantitatively for cod, compared to capelin, is that we have much more extensive tagging data in the latter case.

Our general approach to the parameter estimation is as follows: We divide the year into a fixed number, *T*, say, of time intervals and assume that the parameters, ***V*** and *D*, remain constant within these intervals. These parameter values are furthermore considered to be periodical with a period of one year. We divide he spatial domain Ω into a fixed number of compartments, *M*, say. Thus, excluding mortality, the total number of parameters to be estimated is $3MT$. In general, we assume, however, that the diffusion parameter remains constant over the whole spatial domain in which case this number reduces to $(2M + 1)T$. For the numerical simulation we divide the domain into small triangular elements, in the order of a few thousand, and apply a Galerkin finite element method for discretization in space and a second order Runge-Kutta method for integration in time. We group the tagging data set both in terms of time and location by including in the same cohort items in adjacent elements that have been released within a span of of a few weeks.

For one such cohort, number *r*, say, we can now run a simulation with the initial value, p_0, being restricted to those few elements where the tagged fish of that cohort was realeased and normalized in such a way that the integral of p_0 over these elements equals 1. After *n* timesteps the integral of the calculated *p*-value for a given element number *i* (the avearage *p*-value at its cornerpoints times the area of the element), which we denote as $p_{r,i}^n$ can the be interpreted as the probability that a tagged fish from this cohort will be located within that element at that timepoint. We note that if mortality is taken into account the effect of tag shedding should be added to fishing mortality and natural mortality in this context.

If we have in addition some measure of the probaility of a fish being caught that happens to be within element *i* between timesteps *n*and *n*+1, denoted by F_i^n, the probability of catching *C* fish within this timeinterval will be $\left(F_i^n p_{r,i}^n\right)^C$, assuming that the interval is short enough for *p* not to change significantly within it. The probaility of not catching any tagged fish will be

$$1 - \sum_{C=1}^{\infty} \left(F_i^n p_{r,i}^n\right)^C = 1 - \frac{F_i^n p_{r,i}^n}{1 - F_i^n p_{r,i}^n}.$$

Thus define:

$$P_{r,i}^{n,C} = \begin{cases} \left(F_i^n p_{r,i}^n\right)^C \text{ if } C \geq 1 \\ 1 - \frac{F_i^n p_{r,i}^n}{1 - F_i^n p_{r,i}^n} \text{if } C = 0 \end{cases} \qquad (6)$$

Now if the actual number of caught tagged fish within element *i* between these timesteps *n* and *n*+1 and from cohort *r* is in fact $C_{r,i}^n$

then by a maximum likelihood argument we would like to choose the parameters of our model, ***V*** and *D*, in such a way that the product:

$$\prod_{r=1}^{R} \prod_{n=1}^{N_r} \prod_{i=1}^{I} P_{r,i}^{n,C_{r,i}^n} \qquad (7)$$

is maximized, where *R* denotes the total number of cohorts, N_r

the total number of simulated timesteps for cohort number *r*, and *I* the totla number of finite elements (not to be confused with *M*, the total number of compartments). In equation we are making the assumption that the recapturing events are independent. Also note that if we were to calculate the actual probability of recapturing tagged fish in accordance with this assumption and the given recapture data the product (7) would have to be multiplied with an appropriate multinomial coefficient.

Instead of maximizing (7) we minimize the negative logarithm of this expression, using a numerical minimization procedure that makes use of a numerical estimate of the gradient with respect to the parameters at each iteration step. Since the approximation of each partial derivative involves re-running the simulation with a shift in the given parameter value, the total number of simulation runs to be carried out at each iteration becomes $R\,(3MT + 1)$. While this will in general be a large number, this is balanced by the fact that all the runs can be carried out in parallell.

The approach described above may be contrasted with that used by Adam and Sibert (2002), where they used tagging data from skipjack tuna off the Maldives in order to estimate similar parameters. While the diffusion-reaction model is exactly the same as shown in equation (3) in the case of constant diffusion, except for the fact that it includes mortality, the dependent variable, *p*, is interpreted as the total number of tagged fish at a given time and location, rather than as a probability density of its location at the given time. Accordingly, for a given cohort the initial condition is the actual number of tagged fish being released at a given location. The total number of tagged fish subsequently being caught within location *i* between timesteps *n* and *n*+1 is then assumed to follow a Poission distribution, the probability of catching exactly *C* fish being $e^{-\mu}\frac{\mu^C}{C!}$, μ being the expected number. This expected number is in turn taken to be $F_i^n p_{r,i}^n$ where F_i^n has a similar interpretation as in equation (6) above but $p_{r,i}^n$ is now the expected number of tagged fish being situated within element *i* between timesteps *n* and *n*+1. Denoting similarly, as in equation (6), $F_i^n p_{r,i}^n$ with $P_{r,i}^n$ the parameter estimate in this case amounts to maximizing the probability

$$\prod_{r=1}^{R} \prod_{n=1}^{N_r} \prod_{i=1}^{I} \frac{e^{-P_{r,i}^n} \left(P_{r,i}^n\right)^{C_{r,i}^n}}{C_{r,i}^n!} \qquad (7)$$

An analogous approach is advocated by Sibert et al., (2000), in the case of estimating parameters in a migration model for yellowfin and bigeye tuna in the area south of Honolulu, and in Stefansson et al.(2003) for estimation of parameters in a migration matrix model from tagging data.

Numerical experiment

We have made use of tagging data for cod made available to us by the Marine Research Institute of Iceland in order to establish the feasability of using it for quantiative parameter estimation.The area being considered ranges from 11°W to 27°30'W and from 63°N to 67°45'N (see Fig.1).

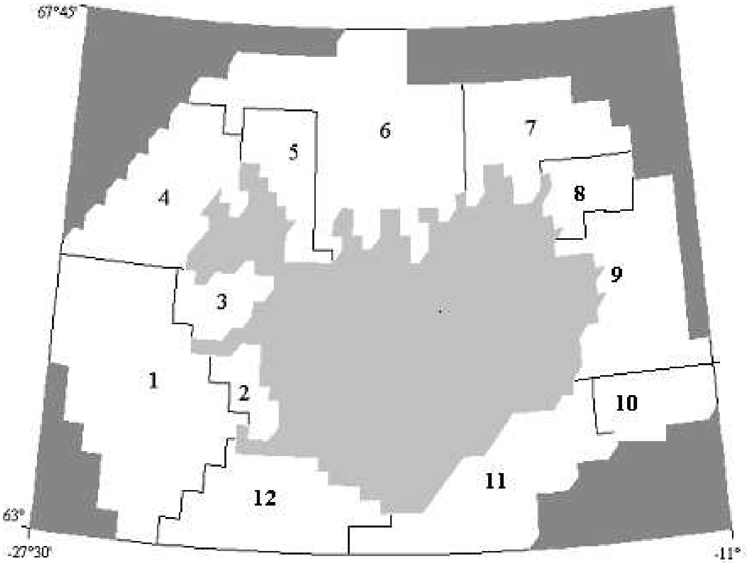


Fig.1. The map of Icelandic ocean waters divided into 12 compartments. The areas in the corners of the map are excluded from consideration, since there were no observations there.

In our finite element discretization of this region we use a 67x39 regular spatial grid in spherical coordinates, with increments of 15' in longitude and 7.5' in latitude. Each grid square gets divided into two triangular elements so this amounts to just over 5000 elements, but since the land area of Iceland is excluded, as well as the corners of the region, where there are no observations of cod, this element number is reduced by more than a half. For this experiment we applied an orthogonal projection of the Earth surface onto a tangential plane at, the point 19°15'W, 65°22'30"N. This allowed us to apply the model in (x, y) - coordinates as presented above. The fact that the computational grid becomes irregular after the projection is of no great consequence in the finite element setting.

In this experiments we divided the year into three timeintervals and assumed that the parameters were constant within these intervals. These are from January 1^{st} to April 30^{th} which corresponds to the spawning migration period, from May 1^{st} to May 31^{st} which corresponds to the spawning period when the active movement of fish can be neglected and they spread mainly by random forces, and from June 1^{st} to December 31^{st} which corresponds to the feeding migration period.

The sea around Iceland was divided into 12 compartments as shown in Fig.1.These are similar to so-called Bormicon areas, used by The Marine Research Instiute of Iceland in their Gadget-model (Anonymous, 2003), but have been slightly modified. The parameters are assumed to be constant within each compartment. Assuming that during the second time interval in May the velocity component can be neglected and the diffusion coefficient is independent of location, the number of parameters is reduced to 12*2*2+1*3 = 51.

The data set made available to us contains time and locations of tagged and recaptured cod. Tagging dates vary between 1993 and 1997, and recapture time is up to 5 years after tagging. In this experiment we restricted ourselves to the years 1994 and 1995 both for tagging and recapture times. Within this time interval the data set consists of 1098 items shown in figure 2. Grouping the data set both in terms of time and location by including in the same group items in adjacent intervals that have been tagged within a span of of a few days the number of distinct tag cohorts was reduced to 86.

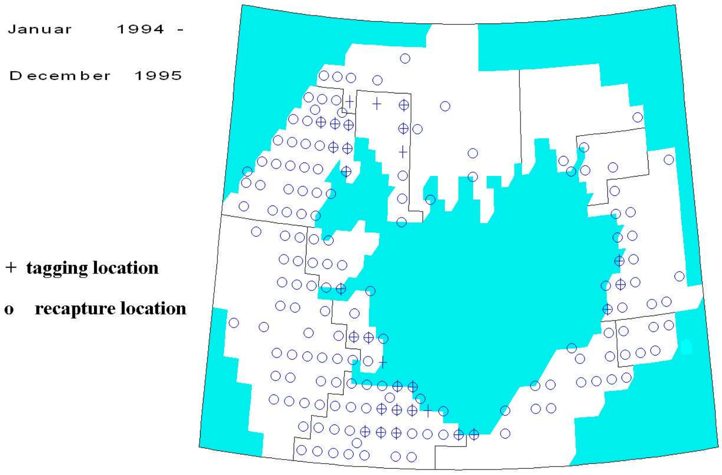


Fig. 2. Tagging and recapture locations of cod, marked with + and o resp., in the years 1994-95, that are used for parameter estimation.

In the absence of any reliable information on fishing effort, we simply set $F_i^n = 1$ if any tagged fish was caught within element *i* between timesteps *n* and *n*+1 and set $F_i^n = 0$ otherwise.

In this experiment we used the FORTRAN subroutine MINF1 in the Fujitsu Scientific Subroutine Library in order to minimize the negative logarithm of expression (6). This routine is bsed on a quasi-Newton method developed by Fletcher(1972), and uses only function values. The results obtained are shown in fig. 3 and in table 1. Since time was measured in days an distance in kilometers the the unit for ***V*** is km/day and for *D* km^2/day.

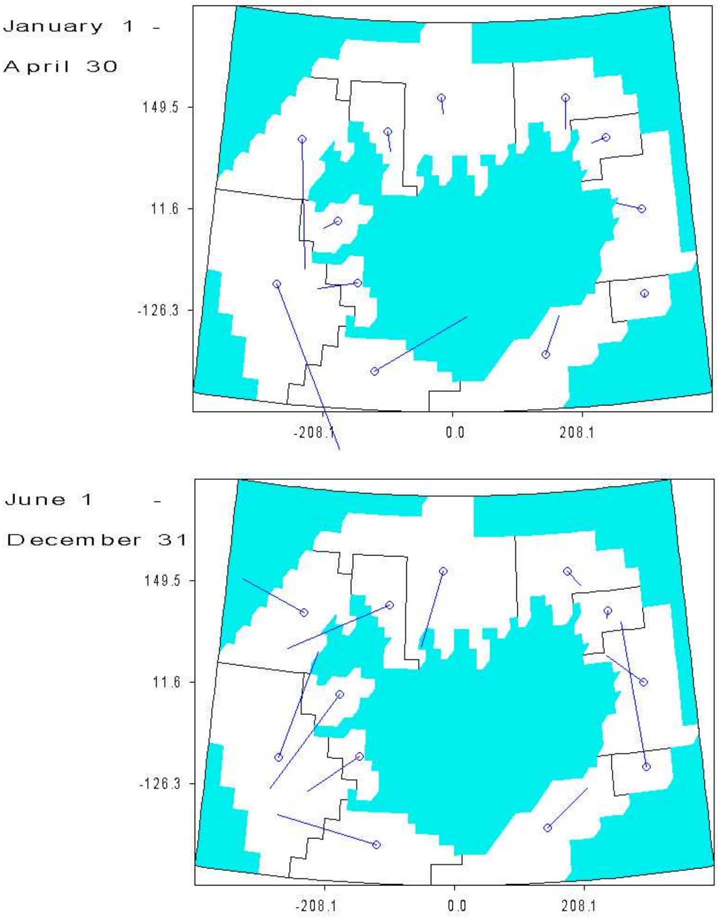


Fig.3. Estimated velocity-parameter values for cod in the 12 compartments in the seas around Iceland. The direction is away from the small circle.

Compart-ment	**Jan 1 – Apr 30**			**May 1-May 31**			**June 1 – Dec 31**		
	V_x	V_y	D	V_x	V_y	D	V_x	V_y	D
1	1.99	-4.49	6.49	0.0	0.0	7.14	0.62	1.43	4.71
2	-0.64	-0.07	-	-	-	-	-0.84	-0.47	-
3	-0.23	-0.10	-	-	-	-	-1.13	-1.27	-
4	0.04	-1.76	-	-	-	-	-0.98	0.46	-
5	0.05	-0.27	-	-	-	-	-1.63	-0.59	-
6	0.02	-0.22	-	-	-	-	-0.35	-1.03	-
7	-0.01	-0.43	-	-	-	-	0.21	-0.20	-
8	-0.23	-0.07	-	-	-	-	-0.02	-0.10	-
9	-0.41	0.07	-	-	-	-	-0.59	0.36	-
10	-0.00	-0.04	-	-	-	-	-0.40	1.96	-
11	0.21	0.52	-	-	-	-	0.64	0.53	-
12	1.48	0.73	-	-	-	-	-1.58	0.40	-

Table1. Estimated velocity- and diffusion-parameter values for cod in the 12 compartments in the seas around Iceland. The unit for ***V*** is km/day and for *D* km^2/day.

Discussion

The results presented above may be compared to those presented in Sigurgeirsson et al. (1997), showing migration of mature cod between compartments in the BORMICON-model, for the same timeperiods. The migration pattern in this model between compartments is described in terms of migration matrices. In the period January-April the agreement is fair in most compartments, the most striking descrepancy between the results being in compartments 11 and 12, where the directions of the velocity are almost diametriclly opposite each other. In the period June-December the agreement is also fair in most compartments, except that the cod tends to move away from region 6 according to the results above, but into the region according to the results presented in Sigurgeirsson et al. (1997). The absolute values are also within reasonable limits. A diffusion value of 6 km^2/day corresponds to a root mean square displacement due to diffusion of v6=2.5 km per day. In our simulation experiments for capelin we have found that a diffusion value of 20 km^2/day gives reasonable results. Parameter estimation of diffusion in the experiments of Adam and Sibert (2002) for skipjack tuna resulted in values of $\sim$1000 square nautical miles per month $\approx$ 100km^2/day. In our opinion the results do, however, not merit much analysis at this stage since the main purpose of our experiment was to establish the feasibility of carrying out such a quantitative parameter estimation for our migration model. But while the results in themselves are encouraging, much work remains to be done.

Perhaps most importantly in the case of cod, an appropriate expression for the formulation of the comfort function has to be established. While it seem feasible, according to this experiment, to estimate the $\boldsymbol{V}$ and D parameters directly, we have also pointed out that by estimating the parameters of a comfort function, rather then those of the differential equation, we are eliminating the changes caused by changes in environmental conditions. The important implication of this is, firstly, that we can include data sampled under different environmental conditions in a single parameter estimate, and, secondly, that the model becomes more robust for predictions of migration patterns under new environmental conditions.

We have not considered the effect of mortality, which may affect the results of the parameter estimation more in the case of cod than eg. capelin. The inclusion of a mortality term into the model does effectively not change the result if the mortality parameter is constant both with respect to location and time but it will do so if this is not the case. In this context we also point out that in the parameter estimation work of Adam and Sibert (2002) referred to above, the main onus is in fact on the estimation of these mortality parameters.

Finally, the tagging data has to be analyzed more closely, both with respect to age distribution and the effects of varying fishing effort.

References

M. S. Adam and J. R. Sibert. 2002. Population dynamics and movements of skipjack tuna (*Katsuwonus pelamis*) in the Maldivian fishery: analysis of tagging data from an advection-diffusion-reaction model. Aquat. Living Resour. 15: 13–23.

Anonymous. 2003. Definitions of areas in Icelandic waters. dst^2. Development of Structurally Detailed Statisticlly Testable Models of Marine Populations. Progress Report for 1 January to 31 December 2002. Marine Research Institute, Iceland. Rep. no. 98. 222-230.

E.H. Dereksdóttir, K.G. Magnússon and S. Sigurdsson. 2003. Simulations of migrations of capelin in the Icelandic waters based on temperature-, current- and potential fields. Report RH-17-2003, Science Institute, University of Iceland. (Available at http://www.raunvis.hi.is/~kgm/).

R. Fletcher. 1972. Fortran Subroutines for minimization by quasi-Newton methods, Report 7125, AERE, Harwell, England.

K.G. Magnússon, S. Sigurdsson, P. Babak, S.F. Gudmundsson and E.H. Derekesdóttir. 2004. A continuous density Kolmogorov type model for a migrating fish stock. Discrete and continuous dynamical systems – Series B 4(3): 695-704

B. Øksendal. 2000. *Stochastic Differential Equations*. Springer.

M.J. Reed and J.G. Balchen. 1982. A multidimensional continuum model of the fish dynamics and behaviour: Application to the Barents Sea capelin (*Mallotus villosus*) Modelling, Identification and Control 3: 65-109.

J.R. Sibert, K. Holland and D. Itano. 2000. Exchange rates of yellowfin and bigeye tunas and fishery interaction between Cross seamount and near-shore FADs in Hawaii. Aquat. Living Resour. 13. 225-232.

H. Sigurgeirsson, H. Björnsson, G. Stefánsson, H.N. Stefánsson. 1997. Cod Migrations. BORMICON. A Boreal Migration and Consumption model. Marine Research Institute. Rep. no. 58. 93-99.

G. Stefansson, A. Jakobsdottir and S. Hannesson. 2003. Including tagging experiments in simulations of fish populations. dst^2. Development of Structurally Detailed Statisticlly Testable Models of Marine Populations. Progress Report for 1 January to 31 December 2002. Marine Research Institute, Iceland. Rep. no. 98. 119-121

7.5 A discrete and stochastic simulation model for migration of fish with application to capelin in the seas around Iceland

Kjartan G. Magnússon, Sven Th. Sigurdsson and Baldvin Einarsson
Science Institute, University of Iceland, Dunhaga 3, 107 Reykjavík, Iceland

Abstract: An individual based, discrete and stochastic model for the collective motion of fish presented in Hubbard *et al.* (2004) is applied to the task of simulating migrations of capelin in the seas around Iceland. In this application the individual particles may be viewed as small schools of fish. that are self-propelled and interacting in such a way that the motion is governed on one hand by a tendency to imitate the motion of other particles in a local neighbourhood and on the other hand by external environmental vector fields for temperature, food density and oceanic currents, as well as a force field generated by an attracting spawning region. In addition there is a stochastic component. The implementation of this model is based on a triangularization of the region under consideration which is shown to be advantageous both in terms of computational efficiency and access to external data. A typical simulation is presented showing the migration pattern of capelin that is situated North-West off Iceland at the beginning October and migrates to spawning grounds south off Iceland by the end of March the following year.

Key words: Fish migration; Interacting particle model; Stochastic model; Capelin; Environmental fields.

Introduction

Many species of fish undertake extensive movements between feeding and spawning grounds. Such directed movements between specific areas — as opposed to random dispersal — are referred to as migrations. Fish migrations are as yet a poorly understood phenomenon; the general route the fish take – which may be genetically "programmed" in some way – are usually fairly well known as well as the approximate timing, but there is considerable year to year variation, presumably due to varying oceanographical conditions. The migration is also influenced by the physiological condition of the fish which is govened by the food density and distribution. How the spatio-temporal characteristics of migrations are affected by the environment is not well known, apart from some general observations. Oceanographic variables which clearly have an effect on migrations are temperature, oceanic currents and food density amongst others.

The capelin (*Mallotus villosus*) in the central North–Atlantic is a good example of a migrating stock illustrating the above. The migrations of the Icelandic capelin are discussed in detail in Vilhjálmsson (1994) and the main features summarized in Hubbard *et al* (2004). We will therefore only give a very brief account here. The spawning stock appears in the waters off the north-east coast of Iceland in late autum- early winter from feeding grounds in the north around the island of Jan Mayen. The stock moves slowly clockwise around Iceland and ends up on the spawning grounds on the south and south-west coast in March-April. A component of the stock takes a more direct route, counter-clockwise around the island. The size of this component is very variable and is possibly connected to the temperature distribution in the ocean. In this paper we will attempt to model the effects of oceanographical conditions on the spawning migrations of capelin in Icelandic waters.

The collective movements of groups of animals can be modelled as a continuous system in time and density (Okubo and Levin, 2001; Toner and Tu, 1998; Babak *et al*., 2004), or as individual based system, which may be discrete (e.g. Vicsek *et al*., 1995) or continuous (Niwa, 1998) in time. The migrations of capelin around Iceland and in the Barents Sea have been modelled by a continuous density Kolmogorov type model (Magnússon *et al*, 2004; Magnússon *et al*., in press) and by a discrete individual based stochastic model (Hubbard *et al*, 2004), where individual fish or schools are regarded as self-propelled interacting particles. Both these models are generic in the sense that they can be applied to any stock and area.

The model presented here extends the one presented in Hubbard *et al.* (2004) firstly in that it includes a field of currents that may carry the fish along. Since typical sea current speeds can be of the same order of magnitude as the speed of fish relative the surrounding sea this may have a significant effect on the migration pattern. Secondly we present a more general approach to deal with attraction to spawning grounds. This allows us, in particular, to include complex obstacles (e.g. islands) in the path of the fish, without the problem of fish getting "stuck" by being pulled into these obstacles by the attraction. This new approach further allows us to include the effect of attraction to spawning grounds in the same comfort function that includes the effects of sea temperature and food density on the motion on the fish. Our approach here is identical to the one adopted in Dereksdóttir *et al.* (2003), in the context of a continuous Kolmogorov-type model of fish migration. In the context of the particle model presented here, the result of this new approach is that we now first group the effects of sea temperature on the direction of motion together with the effects of spawning attraction, rather than first grouping them with the effects of alignment as is done in Hubbard *et al.* (2004).

We also present here a new implementation of the model by introducing a triangularization, that may be fully unstructured, of the region under consideration and subsequently applying the model locally within these triangular elements. Our main reasons for adopting such an implementation are:

1. It allows us to deal effectively with data described in terms of longitude and latitude. Much of the available data on oceanographical conditions as well as fishery is described in this way.
2. It allows us to introduce good approximations to obstacles of complex shape by describing them as a union of unstructured triangular elements.
3. It provides us with a framework of keeping track of neighbour particles, within the same element or within neighbour elements, which in turn makes the calculation of alignment computationally more efficient.
4. It allows us to share data files and make effective comparisons with continuous migration models that we have also developed, since a finite element implementation of those models is based on exactly the same triangular subdivision.

We finally demonstrate the feasibility of this approach by presenting a numerical simulation of capelin migration around Iceland towards its spawning grounds from October 1 2000 till April 1 2001, using the same data as in Dereksdóttir *et al.* (2003).

Mathematical model

The model consists of a collection of particles (fish or small fish schools) moving in a plane, within a prescribed domain. The particles are self-propelling, but are also being carried by a time dependent vector field of currents, $\boldsymbol{V}_c(\boldsymbol{x};t)$, specified external to the model. Particle *i* has speed $v_i(t)$ relative to the field of currents, and a direction of motion given by the angle θ_i. The general dynamical equation for the position of the *i*-th particle is

$$\boldsymbol{x}_i(t+\Delta t) = \boldsymbol{x}_i(t) + \left(\boldsymbol{V}_i(t) + \boldsymbol{V}_c(\boldsymbol{x}_i(t);t)\right)\Delta t \qquad (1)$$

where

$$\boldsymbol{V}_i(t) = \frac{v_i(t)}{|(1-\beta_i(t))\,\boldsymbol{p}_i(t) + \beta_i(t)\boldsymbol{q}_i(t)|}\left((1-\beta_i(t))\,\boldsymbol{p}_i(t) + \beta_i(t)\boldsymbol{q}_i(t)\right) \qquad (2)$$

and the unit vector $\boldsymbol{p}_i(t)$ represents the effect of neighbouring particles on the direction of motion, whereas the unit vector $\boldsymbol{q}_i(t)$ is determined by the gradient of a function $U\left(\boldsymbol{x};t\right)$, which we refer to as a comfort function (Reed and Balchen, 1982). This function will incorporate the factors, which are believed to affect the "comfort" or well-being of the fish, such as temperature, food density, location relative to the spawning grounds, etc.

$\theta_{p,i}(t)$, the direction angle of $\boldsymbol{p}_i(t)$ is calculated similarly as in Vicsek et al. (1995) i.e. by

$$\theta_{p,i}(t) = \langle \theta_j(t-\Delta t)\rangle_{i,\rho} + \xi_i^t \qquad (3)$$

where $\langle \theta_j \rangle_{i,\rho}$ denotes the average angle of motion of all particles inside a circle of radius ρ centered on the *i*-th particle. The average angle is calculated by taking the direction angle of the average of individual direction vectors of these particles at the previous time $t-\Delta t$. The term ξ_i^t is a random perturbation of the direction angle which we choose either as a uniform noise within the interval $[-\eta/2, \eta/2]$ and add in fact to the angle of $(1-\beta_i(t))\,\boldsymbol{p}_i(t) + \beta_i(t)\boldsymbol{q}_i(t)$ rather than $\boldsymbol{p}_i(t)$, or as a "directional noise" described in more detail below.

The comfort function used here is the same as presented in Magnússon *et al.* (2004). It is a linear combination of functions of temperature and food density and a potential function

$$U(\boldsymbol{x};t) = \alpha_1(t)\,r(T(\boldsymbol{x};t)) + \alpha_2(t)\,s(f(\boldsymbol{x};t)) + \alpha_3(t)\,\phi(\boldsymbol{x}) \qquad (4)$$

where *T*(***x***;*t*) and *f*(***x***;*t*) are the temperature and food density respectively at location ***x*** at time *t* and *r* and *s* are given functions defined as follows:

$$r(T) = \begin{cases} -(T-T_1)^4 & \text{if} \quad T \le T_1 \\ 0 & \text{if} \quad T_1 \le T \le T_2 \\ -(T-T_2)^2 & \text{if} \quad T_2 \le T \end{cases} \qquad (5)$$

and

$$s(f) = \frac{f}{h+f} \qquad (6)$$

where T_1, T_2 and *h* are constants. The functional form of *r* implies that the preferred temperature range is between T_1 and T_2 degrees. The fish tend to move towards areas where the temperature is within the preferred range and this tendency is stronger in cold waters. The preferred temperature range may vary within the year. The function *s* is an asymptotic and increasing function of food density, *f*. This functional form implies that the tendency to move towards areas of higher food density decreases with increasing density and is virtually nil at high food densities. The α-coefficients in (4) will be specified later.

The function ϕ is a potential function for the attraction towards the spawning grounds and is defined as follows:

Let the region in which the fish can move be denoted by Ω. We define the boundary of Ω to be any lines, which the fish are not able to cross, such as the coastlines and possibly certain isotherms. The region is in general not simply connected since it may have one or more "holes" in it. Assume that the spawning grounds cover a sub-area *S* of Ω and that migrating fish are attracted towards *S*, in the sense that they experience a "force field" pulling them towards it. The spawning region may be regarded as a continuous sink spread over *S*, to use an analogy from fluid dynamics, and we define the sink density (analogous to mass density or charge density) in the spawning region *S* by $\rho\left(\boldsymbol{x}\right)$. The total attraction strength (sink strength) of the spawning area is therefore given by

$$m = \iint_S \rho dA \qquad (7)$$

The potential ϕ is given as the solution of Poisson's equation

$$\Delta\phi = -\rho \qquad \text{in } \Omega \qquad (8)$$

subject to no-flux condition at the obstacle boundaries of Ω, i.e.

$$\frac{\partial\phi}{\partial n} = 0 \qquad \text{on } \partial\Omega \qquad (9)$$

The force field attracting the fish towards the spawning grounds is given by the gradient of ϕ, $\nabla\phi$. The gradient is orthogonal to the equi-potential lines and is therefore tangent to the set of lines known as streamlines, which the fish would travel along if there were no other factors influencing the motion, than possible obstacles in the form of holes within Ω, that the fish cannot cross.

In the case of an infinite region Ω with no boundaries or obstacles and an attracting region consisting of a single point of sink strength *m*, located at the origin, the attraction density is $\rho\left(\boldsymbol{x}\right) = m\delta\left(\boldsymbol{x}\right)$, where $\delta\left(\boldsymbol{x}\right)$ is the delta function. The corresponding potential function is

$$\phi\left(\boldsymbol{x}\right) = -\frac{m}{2\pi}\ln\left|\boldsymbol{x}\right| \qquad (10)$$

and the streamlines are simply straight lines through the origin. In the case of a finite region we have to specify boundary conditions at the outer boundaries of the region. Rather than viewing them as non-obstacle boundaries we specify the ϕ-values on the outer boundary according to (10), approximating the spawning region *S* by a single point.

One way of quantifying the weight β in equation (2) is to let it depend on the time constants associated with how quickly the particle adjusts its direction to that governed by alignment on one hand and the gradient of the comfort function on the other. Assuming that the former time constant is Δt and the latter is $\sigma\Delta t$ we have that

$$\beta = \frac{1/(\sigma\Delta t)}{1/\Delta t + 1/(\sigma\Delta t)} = \frac{1}{1+\sigma} \qquad (11)$$

Also note in this context that equations (1) and (2) may be viewed as an explicit time discretization of the following dynamical system:

$$\frac{d}{dt}\begin{bmatrix} \boldsymbol{x}_i(t) \\ \hat{\boldsymbol{V}}_i(t) \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{V}}_i(t) + \boldsymbol{V}_c(\boldsymbol{x}_i, t) \\ \frac{1}{\tau}\left(\boldsymbol{V}_i(t) - \hat{\boldsymbol{V}}_i(t)\right) \end{bmatrix} \qquad (12)$$

with $\boldsymbol{V}_i(t)$ being defined according to equation (2) and the time constant of how quickly the velocity $\boldsymbol{V}_i(t)$ adjusts to the set velocity $\boldsymbol{V}_i(t)$ being $\tau = \Delta t$.

Returning to the question of noise, an alternative to the formulation presented above, with uniform noise around the angle of the vector $(1-\beta_i(t))\,\boldsymbol{p}_i(t) + \beta_i(t)\boldsymbol{q}_i(t)$, is to adopt the approach given in Hubbard *et al.* (2004) and introduce a “directional noise”, $\hat{\xi}_i^t$, around $\theta_{p,i}(t)$, the angle of the vector $\boldsymbol{p}_i(t)$ in the direction of $\theta_{q,i}(t)$, the angle of the vector $\boldsymbol{q}_i(t)$x. We can for example define the probability density function of $\hat{\xi}$ as a linear function on a finite interval around 0

$$p_{\hat{\xi}}(x) = \frac{2\gamma}{\eta^2}x + \frac{1}{\eta} \qquad -\frac{\eta}{2} \leq x \leq \frac{\eta}{2} \tag{13}$$

where

$$\gamma = \kappa_\theta \frac{\theta_{q,i}(t) - \theta_{p,i}(t)}{\pi} \qquad 0 \leq \kappa_\theta \leq 1 \tag{14}$$

(see Fig. 1). In general, κ_θ will be taken to be either zero or one, i.e. the directional noise is switched on or switched off. Note that $-\pi < \theta_{q,i}(t) - \theta_{p,i}(t) \leq \pi$ and thus $|\gamma| \leq \kappa_\theta \leq 1$ ensuring that $p_{\hat{\xi}}(x) \geq 0$. The expected value of $\hat{\xi}$ is

$$E\left[\hat{\xi}\right] = \frac{\gamma\eta}{6} \tag{15}$$

and hence

$$E\left[\theta_{p,i}(t) + \hat{\xi}_i^t\right] = (1-\beta)\theta_{p,i}(t) + \beta\theta_{q,i}(t) \qquad \text{with } \beta = \frac{\kappa_\theta \eta}{6\pi} \tag{16}$$

Thus if $\kappa_\theta = 1$, $\eta = 2\pi$ and the time constant associated with alignment is Δt then the expected time constant associated with adjustment to the direction of the gradient of the comfort function will be $\sigma\Delta t$ where $\sigma = \frac{1}{\beta} - 1 = 2$ (cf. (11)) and this time constant increases as η decreases.

The variance is

$$V\left[\hat{\xi}\right] = \frac{\eta^2}{36}\left(3 - \gamma^2\right) \tag{17}$$

Thus the variance with the directional nose “off”, i.e. $\kappa_\theta = 0$, is $\frac{\eta^2}{12}$, but decreases as the directional noise is “switched on”, i.e. $\frac{\eta^2}{18} \leq V\left[\hat{\xi}_{\kappa_\theta=1}\right] \leq \frac{\eta^2}{12}$, depending on $\theta_{q,i}(t) - \theta_{p,i}(t)$.

We can also introduce in a similar manner a noise component, ζ_i^t, to the speed with probability density given by

$$p_\varsigma(x) = \frac{2\lambda}{\mu^2}x + \frac{1}{\mu} \qquad -\frac{\mu}{2} \leq x \leq \frac{\mu}{2} \tag{18}$$

where

$$\lambda = \kappa_v \max\left\{\min\left\{v_0 - v_i(t), 1\right\}, -1\right\} \qquad 0 \leq \kappa_v \leq 1 \tag{19}$$

Here v_0 is a reference speed or an average cruising speed. Thus particles which are moving slower than the reference speed are more likely to speed up than slow down and particles moving faster are more likely to slow down. The above formulation of λ ensures that $|\lambda| \leq \kappa_v \leq 1$ and hence that $p_\zeta(x) \geq 0$.

Numerical implementation

In our numerical implementation of the mathematical model we introduce a triangularization of the region under consideration, that may be fully unstructured. Obstacles (islands) within the region are simply specified by identifying some of the triangular elements as “land” elements. For each triangular element ΔABC we introduce a local (*x*,*y*)-coordinate system as shown in Fig. 2 where the local coordinates of the corner points

$$A = \langle 0, 0\rangle \quad B = \langle a, 0\rangle \quad C = \langle b, c\rangle \qquad (20)$$

are readily evaluated from the distances between the corner points:

$$a = |AB|, \quad b = \frac{a^2 + |AC|^2 - |BC|^2}{2a}, \quad c = \sqrt{|AC|^2 - b^2} \qquad (21)$$

When these corner points are specified globally in terms of longitude, Θ, and latitude, Φ, as is usually the case in our applications, these distances are in turn readily calculated from the corresponding cartesian coordinates:

$$X = R\cos\left(\frac{\Phi\pi}{180}\right)\cos\left(\frac{\Theta\pi}{180}\right), \quad Y = R\cos\left(\frac{\Phi\pi}{180}\right)\sin\left(\frac{\Theta\pi}{180}\right), \quad Z = R\sin\left(\frac{\Phi\pi}{180}\right) \qquad (22)$$

R = 6350km being the radius of the earth. We may calculate these distances either as direct cartesian distances or as great circle distances, the approximation in either case being that we are approximating the triangular element on the sphere with a flat element. This approximation has a negligible effect in our application. We avoid on the other hand having to introduce any global projection on to a plane which may result in more serious distortions. In addition to these local coordinates we calculate the angle from the global west-east direction to the edge *AB* as:

$$u = \tan^{-1}\left(\frac{\Phi_B - \Phi_A}{\cos\Phi_A\left(\Theta_B - \Theta_A\right)}\right) \qquad (23)$$

where the value of the angle is changed by π when $\Theta_A > \Theta_B$.

For each particle, number *i*, under consideration we keep the following information:

i	the identification number of the particle.
E ($\hat{E}$)	the number of the element that the particle is within now (after motion).
x, y ($\hat{x}, \hat{y}$)	the local coordinates of the particle within that element now (after motion).
θ ($\hat{\theta}$)	the local angle of the previous (next) direction of motion of the particle.
v	the present speed of the particle.

For each nodal point, number *j*, in the net of elements we keep the following information:

j	the identification number of the nodal point.
È, Ö	the global coordinates of the point (longitude and latitude).
U	the present value of the comfort function at the point.

For each triangular element, number *k*, in the net we keep the following information:

k	the identification number of the element (-1 if the element is a “land” element).
N_A, N_B, N_C	the identification numbers of its three corner points
a,b,c	the local coordinate values of its corner points
u	the angle that the *AB*-edge makes with the global west-east direction
E_A, E_B, E_C	the identification numbers of its three neighbour elements, adjacent to the element edges opposite the points *A,B,C* (0 if there is no neighbour element).
$\boldsymbol{V}_c$	the present velocity current vector within the element

In addition, for the sake of computational efficiency, we may maintain a list of the numbers of those elements that presently contain any particles, as well as maintaining for such elements a list of the numbers of those particles that are presently within the element. On the basis of this information the positions of the particles can now be updated in turn as follows:

We calculate the angle of the alignment unit vector$p_i(t)$ locally within the element, by letting the alignment only depend on those particles that are within a specified distance, ρ, as well as being within the same element or one of its three neighbour elements, cf. Fig. 3. The second additional restriction is included for the sake of computational efficiency. In order to check whether a particle at point *P'* , within a neighbour element, should affect the alignment of the particle at point *P*, we firstly check whether $|PQ| > \rho$, where *Q* is the projection of *P* onto the common edge. If that is so we subsequently check whether $(|PQ| + |P'Q'|)^2 + |QQ'|^2 > \rho^2$ where *Q'* is the projection of *P'* onto the common edge. The lengths of these line segments can all be easily calculated locally. If such a particle in a neighbour element is to be included the corresponding local angle θ' can be changed into the local coordinates of the original element by the transformation

$$\theta = \theta' + u' - u \qquad (24)$$

cf. Fig. 4. The direction angle of $p_i(t)$ can subsequently be calculated as:

$$\tan^{-1}\left(\frac{\sum\limits_j \sin\theta_j}{\sum\limits_j \cos\theta_j}\right) \qquad (25)$$

where we sum over all particles satisfying the distance conditions, and the value of the angle is changed by π when the denominator of the argument is negative.

We calculate the angle of the comfort gradient unit vector $\boldsymbol{q}_i(t)$ locally within the element, from the specified comfort values $U_A(t),\ U_B(t),\ U_C(t)$, at its three corner points, assuming that the comfort function is linear within the element. Then we have that within the element

$$\nabla U(t) = \frac{-1}{ac}\begin{bmatrix} c\&-c\&0 \\ a-b\&b\&-a \end{bmatrix}\begin{bmatrix} U_A(t) \\ U_B(t) \\ U_C(t) \end{bmatrix} \quad \text{and} \quad \boldsymbol{q}(t) = \frac{1}{|\nabla U(t)|}\nabla U(t) \qquad (26)$$

noting that the columns of the matrix are in fact outward normal vectors on the edges opposite *A*, *B*, *C* respectively whose lengths are that of the corresponding edge.

Finally, since the particle speed, *v*(*t*), is specified with the particle and the current velocity $\boldsymbol{V}_c(t)$ with the element we can now calculate the new position of the particle according to equation (1) as well as the corresponding angle of the direction of motion in the local coordinates, and subsequently establish whether the new position falls within the same element or not. If it does we simply update these values.

If it does not, we first have to calculate the point of intersection, *I*, with the outgoing edge (see Fig. 4), the distance that the particle has travelled when hitting the edge ($|PI|$ in Fig. 4), and the distance of the point of intersection from the preceeding corner point in an anticlockwise direction ($|BI|$ in Fig. 4). From this we can readily establish the local coordinates of this point within the neighbour element, as well as the local angle of direction, θ', from equation (24), and hence continue the motion for the remaining distance $|IP'|$. This may lead us to a new edge in which case the above procedure is repeated. It is also possible that there is no neighbour element or that it turns out to be a land element. In that case the particle is reflected back into the same element for the remaining distance, the local angle of motion changing from θ to $\theta + 2(w - \theta) = 2w - \theta$, where *w* is the angle between the outgoing edge and the edge *AB* of the element (see Fig. 5). If the particle changes elements we update the new element number as well as the coordinates and angle with respect to that element. Note that, while dealing with the particles in turn, we have to keep separate records of positions and angles before and after motion since the alignment depends on the motion of the neighbour elements during the previous time step.

Note that with this implementation we are applying a reflective boundary condition, according to the terminology introduced in Hubbard *et al.* (2004), both at outer boundaries as well as obstacle boundaries. We may effectively replace these conditions by repulsion boundary conditions, according to the same terminology, by introducing at such boundary points temperature values that are well outside the interval $[T_1, T_2]$ in equation (5) (eg. 1000°C), thus forcing $\boldsymbol{q}_i(t)$ to point away from the boundary.

Equations (8) and (9), determining the potential part of the comfort function (4), are solved by a Galerkin finite element method with piecewise linear functions on the same triangular net thus giving directly the values of this function at the nodal points. The temperature and food density parts of the comfort function are specified externally and have to be interpolated onto these nodal points. The same holds true for the sea current values except that they have to be interpolated onto the centroids of the elements.

When displaying the positions of the particles we need to know their global coordinates. These can approximately be calculated from the local coordinates as follows:

$$\Theta = \Theta_A + \frac{x\cos u - y\sin u}{R\cos\Phi_A} \qquad \Phi = \Phi_A + \frac{x\sin u + y\cos u}{R} \tag{27}$$

Simulation example

We demonstrate the feasibility of the approach described above by presenting a numerical simulation of capelin migration around Iceland towards its spawning grounds from October 1 2000 till April 1 2001, using the same data and the same choice of parameters as in Dereksdóttir *et al.* (2003).

The area being considered is an area around Iceland extending from longitude 10°W to 30.5°W and from latitude 62°N to 69°N shown in Fig. 6. The triangular elements are aligned in such a way that one side has a fixed longitude and one side a fixed latitude. Each element spans 0.25° in longitude and 0.125° in latitude. Thus we have in total 9184 elements and 4731 nodal points. The size of the triangular elements varies from 90.1 km^2 in the south to 68.8 km^2 furthest north. The land elements describing Iceland are shown in Fig. 6. Greenland also extends into the north-west corner of this area, but has not been marked since the temperature of the sea in fact prevents the capelin from entering that part of the area.

In the comfort function we ignore the effect of food density by setting $\alpha_2 = 0$ in equation (4). Thus the vector $\boldsymbol{q}_i(t)$ in equation (2) only depends on the ratio $\alpha_1(t)/\alpha_3(t)$. $\alpha_3 = 0$ for the first 80 days and thereafter

$$\alpha_1/\alpha_3 = \begin{cases} \frac{1}{24} & \text{if} 80 \leq t \leq 130 \\ \frac{1}{1920} & \text{if} 130 \leq t \leq 182 \end{cases} \tag{28}$$

The preferred temperature range $[T_1, T_2]$ in equation varies from [0.5,4.0] degrees on October 1 to [3.0,7.5] on April 1, the end values increasing linearly in time throughout this period. This is based on various field observations as described more fully in Magnússon *et al.* (in press).

The radius of alignment, ρ, is set equal to 10 km. Since typical lengths of sides in the triangular elements are 12-16 km this implies that the circle of alignment will always extend into neighbour elements as shown in Fig. 3, and that some particles within that circle will in general not be counted for.

The weight-factor $\beta_i(t)$ in equation (2) is set equal to 0.9 for all particles at all times.

The noise in the direction is uniform around the angle $(1 - \beta_i(t))\, p_i(t) + \beta_i(t) q_i(t)$ with a total range of $\eta = 80°$ (i.e. we do not implement the “directional noise”).

The speed $v_i(t)$ changes with time as follows: it is constant for the first 80 days, 5 km/day, then increasing rapidly to 25 km/day during the following 50 days, and remaining at that value after that according to the expression

$$v\,(t) = \min\left(5 + 20 \cdot \frac{(t-80)^3}{50^3}, 25\right) \tag{29}$$

The hypothesis underlying this assumed increase is that the speed is related to the stage of maturity (i.e. roe content), which is increasing while the capelin are migrating as described more fully in Dereksdóttir *et al.* (2003). There is no noise included with the speed.

In order to include the effect of carrying sea currents we use the same hypothetical current field that is constructed in Dereksdóttir *et al.* (2003), which is meant to reflect the effects of the real currents around Iceland based on the information currently available (see Fig. 7). Note that the coastal current circulates Iceland in a clockwise direction. The speed of the hypothetical current is in the interval [0.75, 15] km/day and only varies with approximate distance from the shore.

The time step Δt is set at 0.1 days which implies that at the maximum speed of 25 km/day it takes the particle approximately 5 time steps to traverse each element. The numbers of particles is 1096 and at the start of the simulation at October 1 they are distributed over an area north-west of Iceland, as shown in Fig. 5, similar to the one used in the simulations in Dereksdóttir *et al.* (2003). The spawning area is specified south of Iceland as shown in Fig. 8. That figure also shows the gradient directions of the corresponding potential field obtained by solving equations (8) and (9).

Some snapshots of the result of the simulation are shown in Fig. 9. In this particular simulation 61% of all particles (fish) end up reaching the spawning ground by going round the west cost of Iceland, whereas 37% reach it by going round the east coast, 2% remaining at the north coast. With the present values of parameters these percentages vary considerably (remembering that there is a random angle parameter), and in three other runs the percentage going round the west coast was only 12%, 28%, and 34% respectively. It has further been noted that when the noise is not included and also when there is no alignment virtually no particles travel along the west coast. This is more in agreement with the “base” simulation presented in Dereksdóttir *et al.* (2003) for the same period, using a continuous Kolmogorov-type model, where 19% reach the spawning ground by going round the west coast, 80% reach it by going round the east coast, while 1% remain at the north coast. Note, however, that this result is sensitive to the position of the initial distribution. With a more westerly distribution Dereksdóttir *et al.* (2003) present a simulation where the situation is reversed, i.e. 80% reach the spawning ground by going round the west coast and 19% reach it by going round the east coast.

Discussion

While this simulation example demonstrates the feasibility of adopting the modelling approach presented above, it remains to carry out much more extensive experimentation in order to establish the sensitivity of the results to changes in parameter values, including the effect of particles grouping themselves into schools, and possibly splitting up again as observed e.g. in the simulations presented in Hubbard *et al.* (2004). Following that it is obviously of interest to see how the parameters can be adjusted to obtain as good fit as possible with real acoustic measurements of capelin.

As noted in the introduction, the approach of the triangularization of the simulation region introduced in this work has a number of advantages. In particular, it allows us to share data files and make effective comparisons with continuous migration models. Thus it is of interest to investigate the agreement of the results from the particle model with results obtained by the Kolmogorov continuous migration model described in Magnússon *et al.* (2004), and implemented on exactly the same tringularized domain with the same comfort function and the same field of sea currents in Dereksdóttir *et al.* (2003). In this respect it should be noted, however, that the output of the Kolmogorov model is the probability density of a fish being located at a given position at a given time subject to some initial probability distribution, whereas the output of particle model presented above is the actual position at a given time. In order to obtain comparable results one would thus have to make repeated runs of the stochastic particle model.

It is also of interest to compare the results of the particle model with the results of a continuous dynamic models like the one presented in Sigurdsson *et al.* (2002) where the output is the actual density of fish. It is e.g. of interest to establish how the particle model and the continuous model compare in terms of computational efficiency. However, here it should be noted that while we can incorporate the effect of a comfort function in a similar way with these two approaches, and apply them on the same tringularized region, it is easier to deal with effects of alignment in the particle setting, while it is easier to incorporate the effect of some preset preferred density in the continuous density setting. The role of such preset density is to draw the fish towards each other if the actual density falls below the preferred density and push them apart if the actual density exceeds the preferred density. We have already noted that the interplay between a stochastic component and the effect of alignment may play a role in the formation of schools. In Babak *et al.* (2004) it is shown that the existence of some preferred density may also play a role in the formation of school patterns.

References

P. Babak, K. G. Magnússon, S. Th. Sigurdsson. 2004. Dynamics of group formation in collective motion of organisms, Mathematical Medicine and Biology, Volume 21, (4), 269-292

E. H. Dereksdóttir, K. G. Magnússon and S. Th. Sigurdsson. 2003. Simulations of spawning migrations of capelin in Icelandic waters based on temperature-current- and potential fields. Science Institute, University of Iceland, Report RH-16-2003

S. Hubbard, P. Babak, S. Th. Sigurdsson, K. G. Magnússon. 2004. A model of the formation of fish schools and migrations of fish. Ecological Modelling, 174, 359-374

K. G. Magnússon, S. Th. Sigurdsson, P. Babak, S. F. Gudmundsson, E. H. Dereksdóttir. 2004. A continuous density Kolmogorov type model for a migrating fish stock, Discrete and Continuous Dynamical Systems-Series B, Volume 4 (3), 695-704

K. G. Magnússon, S. Th. Sigurdsson, E. H. Dereksdóttir. (In press). A simulation model for capelin migrations in the North-Atlantic. Nonlinear Analysis: Real World Problems.

H.-S. Niwa. 1998. Migration dynamics of fish schools in heterothermal environments. J. theor. Biol., 193, 215-231.

A. Okubo and S.L. Levin. 2001. Diffusion and Ecological Problems: Modern Perspectives. Springer.

M.J.Reed and J.G.Balchen. 1982. A multidimensional continuum model of the fish dynamics and behaviour: Application to the Barents Sea capelin (*Mallotus villosus).* Modelling, Identification and Control, 3, 65-109.

S. Th. Sigurdsson, K. G. Magnússon, P. Babak, S. F. Gudmundsson, E. H. Dereksdóttir. 2002. Dynamic continuous model of fish migration, Science Institute, University of Iceland. Report RH-25-2002

J. Toner, J. and Y. Tu.1998. Flocks, herds and schools: A quantitative theory of flocking. Physical Review E, 58 (4), 4828-4858.

T. Vicsek, A. Czirok, E. Ben-Jacob, I. Cohen, O.Shochet. 1995. Novel type of phase transition in a system of self-driven particles. Phys. Rev. Lett. 75 (6), 1226-1229.

H. Vilhjálmsson.1994. The Icelandic capelin stock. Journal of the Marine Research Institute Reykjavik, Vol. XIII, No. 1. 281 pp.

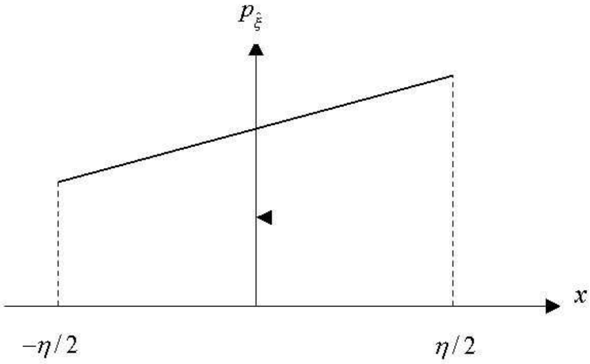


Figure 1. Probability density function of the random perturbation in direction angle.

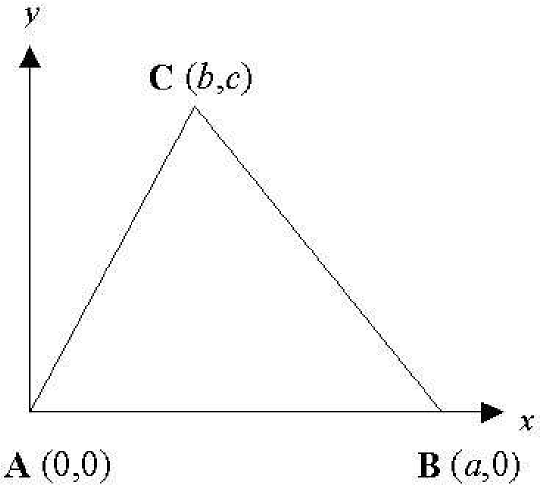


Figure 2. Local coordinates of triangular element.

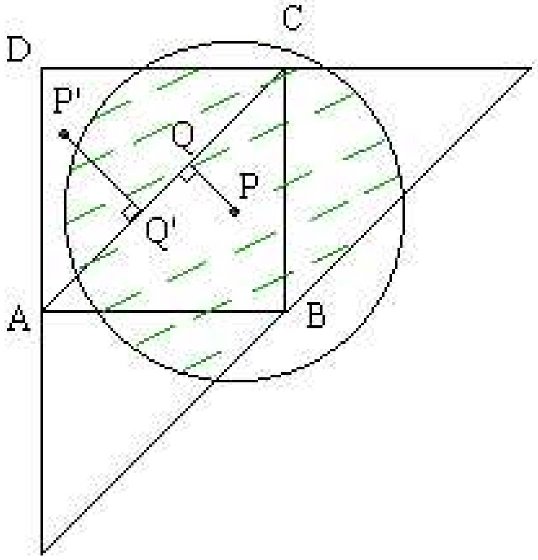


Figure 3. Circle of alignment around point *P*. Particles within the shaded area will affect the direction of a particle at *P*.

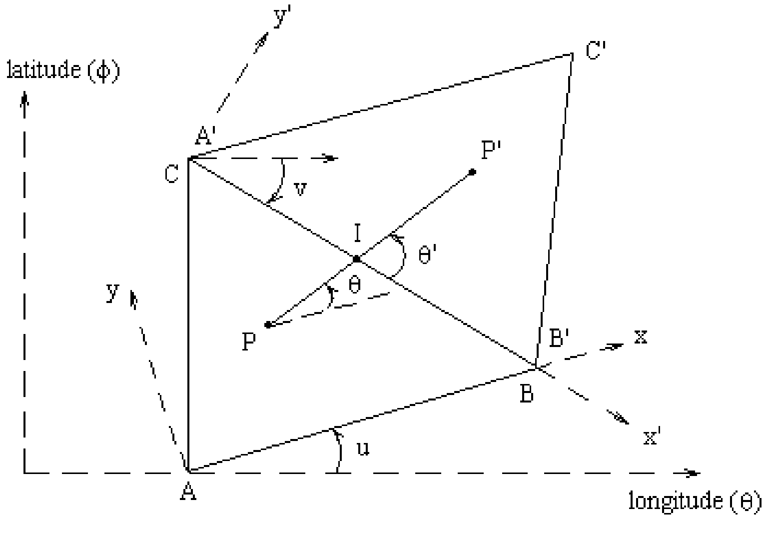


Figure 4. Relationship between local and global directions of motion as particle moves from one element into another.

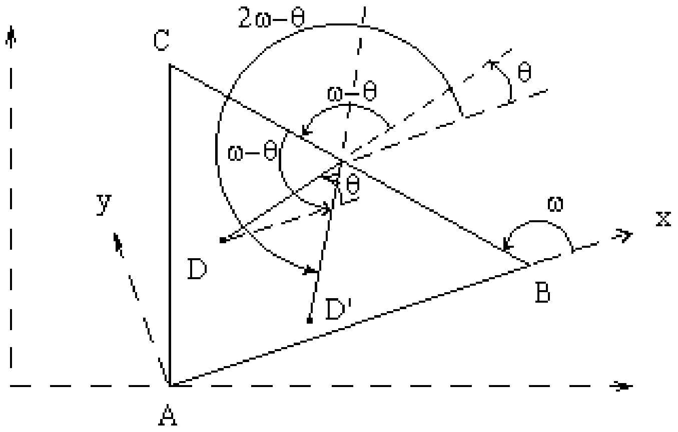


Figure 5. Relationship between local and global directions of motion as particle is reflected from land boundary.

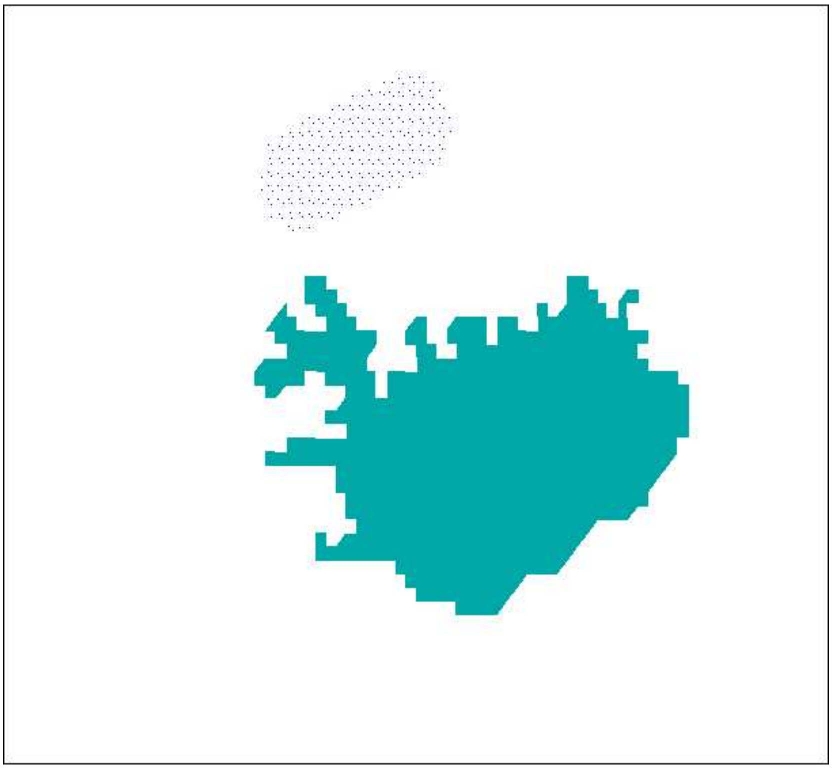


Figure 6. Area of simulation. The dots North-West of Iceland show the initial distribution of particles on October 1

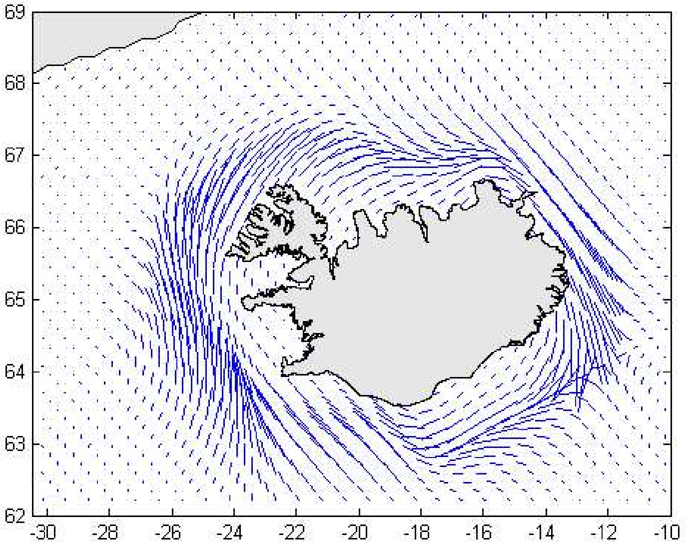


Figure 7. Hypothetical field of oceanic currents around Iceland. The speed of the current is given by the length of the arrows, i.e. darker colours indicate greater speed

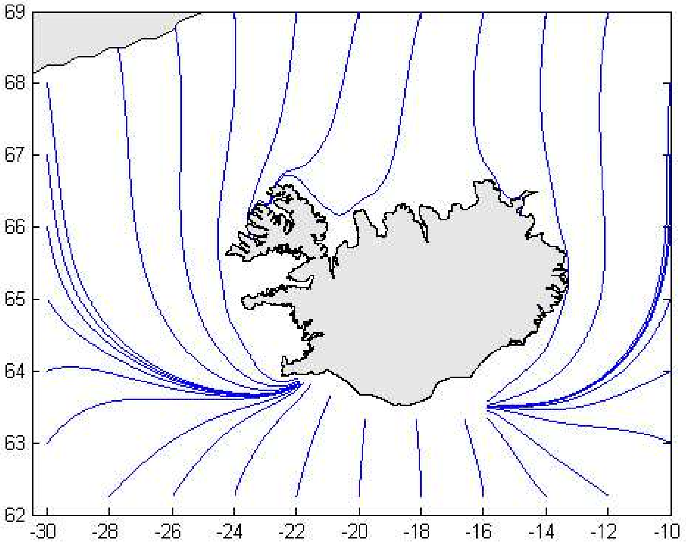


Figure 8. Iceland and the field lines for the attraction force generated by a spawning area on the south coast.

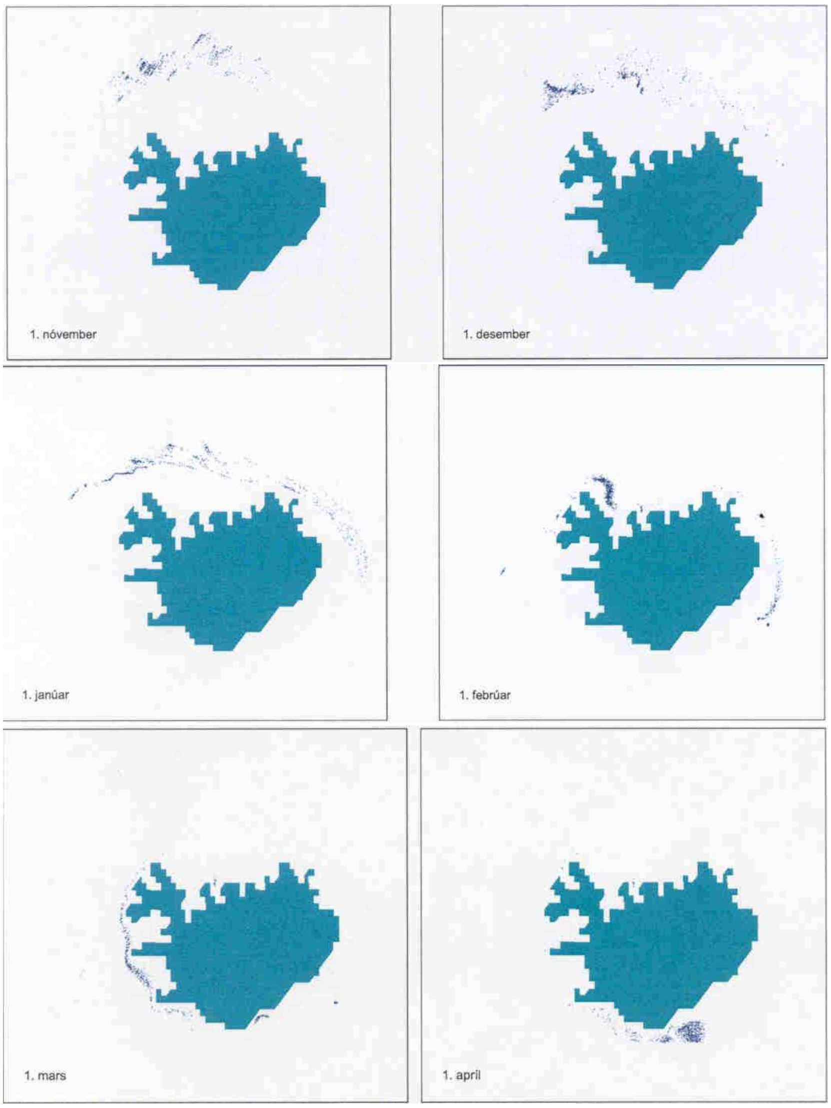


Figure 9. Snapshots of results from a simulation run showing the migration of capelin from October 1 2000 until April 1 2001. The initial distribution on October 1 is shown in Figure 6.

7.6 Implementation of spatial detail in GADGET: a comparison of different modelling approaches

Geir Huse, Eva Hlin Dereksdottir, Kjartan G. Magnusson, Sven Sigurdsson, Sigurd Tjelmeland

G. Huse, S. Tjelmeland: Institute of Marine Research, Box 1870 Nordnes, N-5817 Bergen, Norway; Phone: +47 55238500; fax: +47 55238584; E.H. Dereksdottir, K.G. Magnusson, S. Sigurdsson, Science Institute, University of Iceland, Dunhagi 3, 107 Reykjavik, Iceland

Introduction

This report presents three different approaches to modelling fish migrations, using the Barents Sea capelin as a case study. The report also describes how of migration matrices can be output for GADGET. The background for the report is that several research groups associated with the dst^2 project have been working with modelling migration. Two meetings on this topic were arranged in Bergen to discuss modelling of fish migrations in general and implementation in GADGET specifically.

The three models are presented below are: The super-individual model (SI) model, the Kolmogorov model, and the Borealis model. All three models are individual-based models, although, from a computational point of view the Kolmo-gorov model is more like a continuous model. Individual-based modelling (IBM) has become an important technique in ecology, and has been used extensively in studying advection in larval fish (see review by Grimm, 1999; Huse *et al.*, 2002). The advantage of this approach is that the implementation of spatial detail is very flexible for such models. Many different approaches to moving individuals are available and can be used interchangeably in model simulations. Although abundant fish populations cannot be represented individual by individual, the super-individual approach allows scaling of individuals so that an “individual” actually represents millions of identical siblings (Scheffer *et al.*, 1995). This makes IBM flexible with regard to population dynamics as well. In addition to describing the models and some results, we provide a short discussion of implementation of spatial detail in GADGET. Capelin was chosen as a case study for presenting the various migration modelling approaches since this is a migratory species and there is good availability of spatial distribution data, though at a rather course temporal resolution. In addition temperature fields for the Barents Sea are readily available for inclusion in migration models.

The Barents Sea capelin

The Barents Sea is dominated by few but abundant species of fish, which makes the area prone to fluctuations in fish biomass caused by species interactions and fisheries. The most abundant fish species are cod *(Gadus morhua)* and capelin *(Mallotus villosus)* (Fig. 1). The interaction between these species is important for understanding the fluctuations of the Barents Sea ecosystem (Hamre, 1994). The geographical distribution of the capelin is closely associated with the ice edge in the Barents Sea throughout large parts of the year. During summer the capelin follow the retracting ice edge and prey on the secondary production in the rich marginal ice zone (Hassel *et al.*, 1991). As fall approaches, the capelin moves southwards and overwinter along the ice border in central parts of the Barents Sea (e.g. Hamre and Monstad, 1980; Gjøsæter, 1998). In spring the adults migrate to the coast of Northern Norway where they spawn. Cod prey extensively on capelin during the overwintering and spawning periods in winter, but also during other times of the year (Mehl, 1989; Mehl and Sunnanå, 1991; Dolgov, 2002).

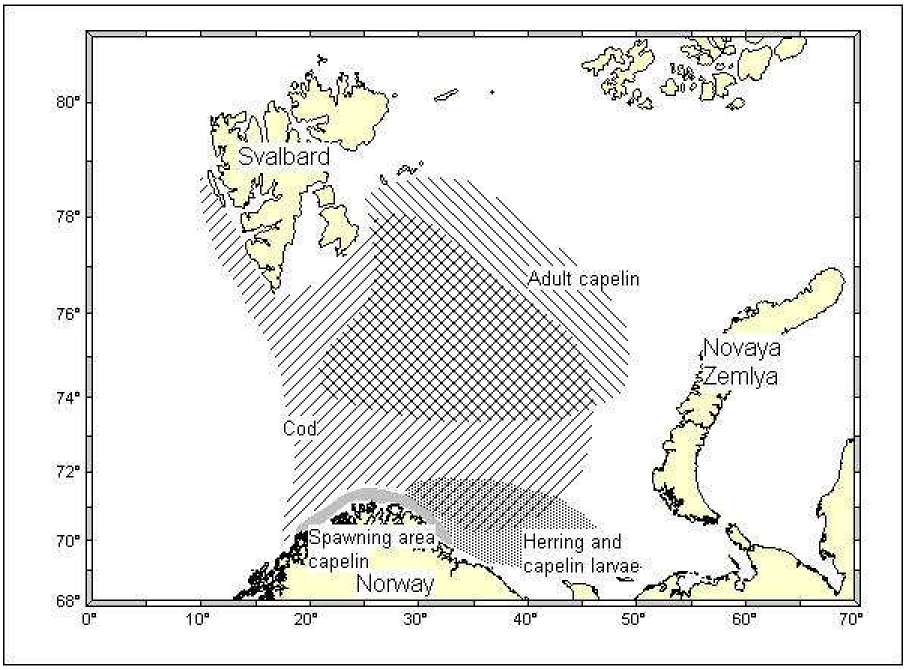


Figure 1. The Barents Sea with the main features of the distributions of cod and capelin. Modified from Huse and Gjøsæter (1999).

7.6.1 The super-individual model: description and results

The Individual-Based Model

For a general introduction to IBM see DeAngelis and Gross (1992), Grimm (1999) or Huse *et al.* (2002). Fish populations can be represented in IBM by using super-individuals (Scheffer *et al.*, 1995). A super-individual represents many identical individuals, here referred to as siblings. This is an efficient way to maintain the individual-based structure, and still be able to simulate the large population sizes that occur in fish populations. An attribute vector *AV* (Chambers, 1993) is applied to keep track of the states used to specify super-individuals. Mortality operates on the super-individual and number of siblings of super-individuals is thus decreased in proportion to the mortality rate. The following attribute vector was here used*:*

$$\mathrm{AV} = (\mathrm{a}, \mathrm{l}, \mathrm{w}, \mathrm{n}, \mathrm{x}, \mathrm{y}) \quad (1)$$

where a is age, l is length, w is weight, n is the number of identical siblings, x and y are spatial positions of a super-individual. The model domain consists of a 90 (x-axis) by 100 (y-axis) grid of 20x20 km squares. The observations were initiated from survey data (see below) as average length and weight, and number of each age group in the squares. Each age group in each of the squares with observations were initially allocated a super-individual, with length, weight and internal number corresponding to the observations for that square. After all the observations had been allocated to super-individuals, new super-individuals were generated from the existing ones until the population consisted of 20 000 super-individuals. This was performed by picking the most abundant super-individual (donor), and generating a new one by copying the attributes of the donor. The abundance was then divided by two, both for the donor and the new super-individual to maintain mass balance (Rose *et al.*, 1993). The number of super-individuals is a trade off between having a high resolution of the population, but at the same time keeping computer running time within reasonable limits. Daily time steps were used.

Temperature

Temperature data for 1990-1996 were taken from a digital temperature atlas for the Barents Sea based on IMR's hydrographic measurements at different, spatially referenced locations. The average temperature in the upper 200 m was used as an estimate of capelin ambient temperature, since this is the regular range of capelin vertical distribution (Gjøsæter, 1998). The temperature data were converted to the regular polar stereographic grid constituting the model domain. The data were spatially interpolated in each depth strata using a partial spline interpolation method based on nearest neighbours in the G3GRID procedure in the SAS® software package. Before interpolation, temperature fields generated by an ocean circulation model (Ådlandsvik and Sundby, 1994) were used to provide temperature data in grid cells at the edge of the temperature atlas, to control for boundary effects cause by the spatial interpolation technique. Simple temporal interpolation between measurements in the third and first quarters of the year, were performed to obtain daily temperature fields. This was done due to the greater sampling frequency in these quarters compared with the other two quarters. The Kola section in the Barents Sea (70°30' N to 72°30' N along 33°30' E, Tereshchenko, 1996) is measured monthly. There are fairly linear relationships between the monthly temperatures at the Kola section during the period September to March (linear regression, $p < 0.001$, $r^2 = 0.99$) and from March to September (linear regression, $p < 0.001$, $r^2 = 0.92$) for the monthly average temperatures in the six year study period. This supports the usage of linear interpolation between the third and first quarters, which correspond approximately to September and March. Current fields were used in calculating fish advection, applying the same depth ranges as used for temperature (see above). The current fields were generated by the ocean circulation model.

Survey data

Data from a survey in September-October (Gjøsæter *et al.*, 1998) were used to initiate the capelin super-individuals. The survey is an annual acoustic survey involving altogether four research vessels from Norway and Russia. During this time of the year the capelin stock is distributed over wide areas of the northern and central Barents Sea. The capelin are undertaking vertical migrations, but are normally not found near the surface or the sea bed, and are, therefore, ideally distributed for acoustic stock estimation (Toresen *et al.*, 1998). The methods used and the results from the whole time series going back to 1972 are summarized in (Gjøsæter *et al.*, 1998). Acoustic data on capelin were directly converted to the temperature grid. Biological information from hauls during the survey is compiled in a geographical block system (Gjøsæter *et al.*, 1998), and the average values from these blocks were used in initiating the capelin super-individuals.

In addition to the surveys in fall used to initiate the distribution, a winter survey in January-March (Jakobsen *et al.*, 1997) was used to validate the model predictions. This survey targets demersal fish, in particular cod and haddock. However, acoustic abundance of capelin is also recorded. It should be noted that the spatial coverage of this survey is limited by ice. Furthermore the simulated distributions were compared with capelin distributions in the following fall. Information on the spawning area of capelin in different years was taken from Gjøsæter (1998), and these data were used to evaluate the simulated spawning migration. Age groups 1-4 were studied. As customary in fisheries science, age is updated January 1. Intervals of 0.5 cm body length were used.

Migration

Movement of the super-individuals was performed using a simple model based on direction, velocity and temperature boundaries (Table 1). Swimming was assumed to have directed, d_X and d_Y, and stochastic components s_x and s_y along the x and y axes. Current vectors were added to the swimming to generate the net displacement of super-individuals. The directed movement was generated from the directions and relative velocities listed in Table 1. The subscripts w and m refer to immature and mature individuals in winter, and s refers to summer. Temperature has previously been shown to be important in determining the distribution of capelin (Ozhigin and Luka, 1985). Some previous observations (e.g.

Anon., 1974; Hamre and Monstad, 1980) suggest that the capelin align their horizontal distribution with temperature fronts in the upper waters. Summer and winter temperatures were therefore implemented in the migration model. TL_w and TL_m gives the temperature below which the directed winter movement (DL_w and VL_w or DL_m and VL_m) is induced for juvenile and mature capelin respectively. If directed movement is not induced (temperature > TL_w) during winter, only the stochastic movement and advection is used to calculate the new position of a super-individual. TL_s gives the temperature below which the capelin is not allowed to move during summer. Also, capelin was not allowed to enter the shallow banks in the Barents Sea, such as the Skolpen bank in the southern part of the Sea, based on observations that the capelin tends to avoid such areas (e.g. Dommasnes *et al.*, 1974). Since the horizontal position of temperature fronts may vary with depth, the temperatures at 50 rather than average temperatures in the water column, were used in the movement model.

Both d_X, d_Y, s_x and s_y have a range of [–0.5,0.5], and relative displacement thus is in the range [–1,1]. During the spawning migration when movement is highly directional, the stochastic component was left out, and instead d_x and d_y had the range [–1,1]. The relative displacement was multiplied by the maximum daily swimming distance D_{max}, to yield the displacement in each direction (M_X and M_Y):

$\mathrm{M}_X = (\mathrm{d}_X + \mathrm{s}_x)\cdot\mathrm{D}_{max}$ (2a)
$\mathrm{M}_Y = (\mathrm{d}_Y + \mathrm{s}_y)\cdot\mathrm{D}_{max}$ (2b)

D_{max} was calculated as the distance swum at maximum swimming speed over a day. Capelin was generally assumed to have a maximum swimming speed of one bdl s^{-1}, but during the spawning season the maximum swimming speed was set to three bdl s^{-1} (Hafsteinsson and Misund, 1995). The movement distance was then calculated as:

$$\mathrm{D_{tot}} = \sqrt{\mathrm{M_X^2}+\mathrm{M_Y^2}} \qquad (3)$$

If D_{tot}>D_{max}, the movement vector was corrected accordingly by $\mathrm{D}_{max}/\mathrm{D}_{tot}$. After the distance swum by the fish had been calculated, the current vectors (C_X and C_Y) generated by the ocean circulation model were added:

$$\Delta\mathrm{X} = \mathrm{M_X}+\mathrm{C_X} \quad (4a)$$

$$\Delta\mathrm{Y} = \mathrm{M_Y}+\mathrm{C_Y} \quad (4b)$$

where ΔX and ΔY is the net displacement in each direction used to calculate the movement vector. Since the polar-stereographic grid is tilted, a correction was made to get true heavenly directions, and allow implementation of strategies such as “move south”.

Table 1. The specification of the movement criteria used in different simulations. The subscripts w, m refer to immature and mature individuals in winter, and s refers to summer. DL, VL and TL refer to directional movement, swimming velocity and temperature thresholds of capelin. S is south, SW is southwest, SSW is south southwest, WSW is west southwest, N is north, NW is northwest, NNW is north northwest, NNE is north northeast.

Sim	DL_w	DL_s	VL_w	VL_s	TL_w	TL_s	DL_m	TL_m
1	S	NNE	0.8	0.4	3.0	0.0	SW	3.5
2	SSW	NNW	1.0	0.2	2.0	-0.5	SSW	4.0
3	WSW	N	0.7	0.1	2.5	-1.0	SW	3.0
4	SW	N	1.0	0.2	2.5	-1.5	SSW	4.0
5	SW	NNE	0.9	0.1	2.5	-0.5	S	4.5
6	S	N	1.0	0.3	3.0	-1.5	S	5.0
7	SW	NNW	1.0	0.2	3.5	0.0	S	4.0
8	SSW	N	0.8	0.2	2.5	0.5	SSW	4.5
9	SW	NE	0.9	0.1	3.0	-0.5	S	4.0
10	SW	N	0.6	0.1	2.5	-1.5	SW	4.0
11	S	NNW	1.0	0.2	2.5	-2.0	SSW	3.5

12	SSW	N	1.0	0.2	3.0	-1.0	SSW	4.0
13	SW	N	1.0	0.3	3.0	-1.5	SW	3.5
14	SW	NNE	1.0	0.3	2.5	-1.0	SSW	4.0
15	SSW	N	1.0	0.2	2.5	-1.5	S	4.0

Comparing simulations and observations

The study was performed by testing various parameter values in the movement models (Table 1) for capelin, and validate them against observed distributions. The range of movement parameter values tested was selected based on previous experience and test runs of the model, and many unreasonable parameter values, that for example yield northwards movement in winter, were left untested.

The year was divided into a winter and a summer period (Fig. 2). The simulations were initiated on October 1. Comparisons of predicted and observed spatial distribution are made in February and September, while the proportion of mature capelin reaching the observed spawning ground is determined in April. The combined success of the winter and spawning migrations of mature capelin was calculated as the proportion of mature individuals entering the observed spawning areas during February-April. Predicted and observed abundances were compared using root mean square (RMS) calculated as:

$$\mathrm{RMS} = \sqrt{\frac{1}{\mathrm{S}}\sum_{\mathrm{j=1}}^{\mathrm{jmax}} (\mathrm{P_j} - \mathrm{O_j})^2} \qquad (5)$$

where P_j and O_j are the predicted and observed relative fish abundances (proportion of total abundance found in square j), jmax is the total number of squares, and S is the number of squares with observations and/or predictions. Finally the RMS was summed over the entire study period. The different simulations were evaluated using RMS, and low RMS relative to other simulations indicates better performance. All correlations referred to below are Pearson correlations.

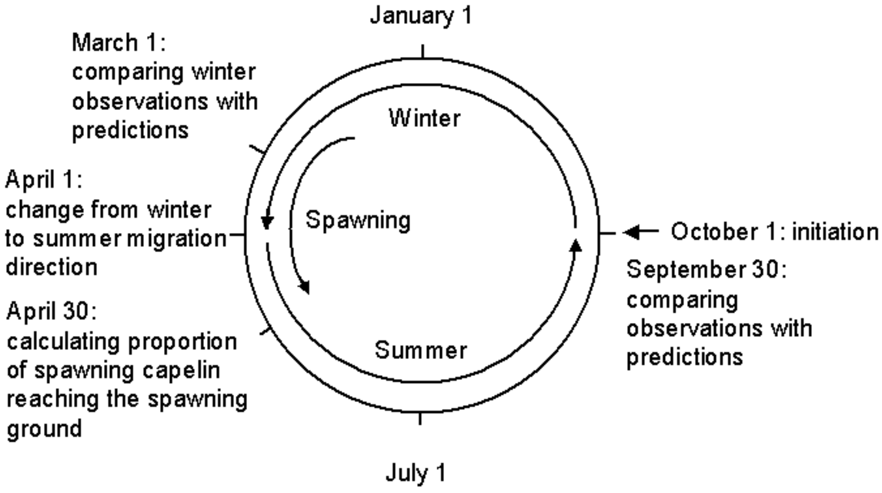


Figure 2. The temporal structure of the study with model initiation in fall and comparisons with observed distributions in March, spawning in April, and distributions in September. The arrows indicate the time span for using the winter, summer and spawning migration directions.

Results

The results of the different simulations are shown in Table 2. Simulation 10 had the overall best fit to observations as it produced the second lowest sum of RMS for the four criteria and also had the second highest proportion of mature individuals reaching the spawning area. This simulation used the 2.5 and 4.0°C thermoclines for immature and mature capelin, and relied on southwest- and northwards migrations during winter and summer respectively.

Table 2. Comparison of the model predictions against observations. February and September (RMS) refer to the root mean square of predicted-observed distributions, Spawning (Proportion) is the proportion of mature capelin reaching the observed spawning area, WL_2 is the predicted average weight (g) for two year old capelin.

	February	Spawning	September	
Simulation	RMS	Proportion	RMS	WL_2
1	0.0043	0.62	0.0060	7.0
2	0.0043	0.53	0.0030	6.7
3	0.0033	0.77	0.0017	8.1
4	0.0044	0.70	0.0020	10.7
5	0.0042	0.63	0.0018	11.4
6	0.0054	0.25	0.0027	6.5
7	0.0054	0.54	0.0018	34.2
8	0.0039	0.54	0.0037	7.2
9	0.0042	0.54	0.0016	17.1
10	0.0033	0.76	0.0017	9.5
11	0.0052	0.38	0.0019	5.8
12	0.0051	0.53	0.0021	10.8
13	0.0040	0.75	0.0016	12.6
14	0.0043	0.70	0.0029	8.6
15	0.0051	0.40	0.0027	9.0

Examples of the predicted and observed spatial distribution of capelin are shown in Figure 3. A notable feature is the separation of the immature and mature components of the stock, which takes place in February-April. The predicted spatial distribution of capelin resembles the observed capelin distribution quite well. The model results give a much wider distribution of the capelin in February than the survey indicates. However, the survey coverage is limited by ice, and parts of the stock are probably distributed under the ice (north and east of the survey area) at this time of the year. Also, the survey in 1991 did not cover the entire ice-free part of the Barents Sea. Thus, the difference between observed and modelled distributions in February is probably smaller than Figure 4.

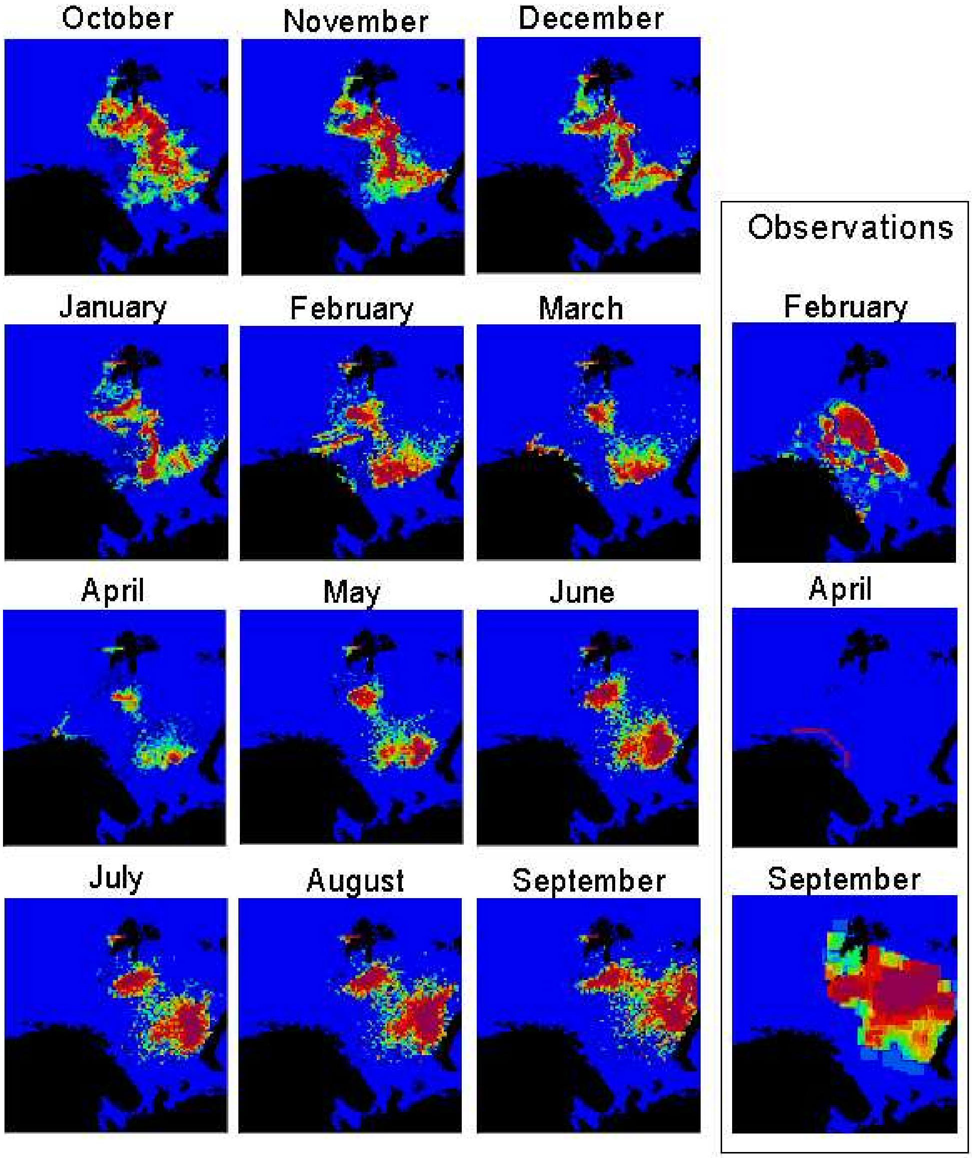


Figure 3. Predicted spatial distribution of capelin from October 1990 to September 1991 using simulation 10. Observed spatial distributions of capelin for February and September are shown to the right. The observed spawning area in 1991 is inside the purple fields for the April panel, but no acoustic observations for the remaining part of the stock are made at that time.

Table 3 shows an example of a migration matrix generated by the SI model for October 1981. It gives the proportion fish that move from one MULTSPEC area to another area, and stay. These proportions have been generated from the individual-based model.

Table 3. Example of migration matrix for immature capelin in GADGET generated by the SI model for October 1990. The second row and first column gives the MULTSPEC areas.

	Immature						
	1	**2**	**3**	**4**	**5**	**6**	**7**
1	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	0.968	0.039	0.004	0	0
4	0	0	0	0.749	0	0	0
5	0	0	0.026	0	0.983	0	0.011
6	0	0	0	0.212	0	0.994	0.005
7	0	0	0.006	0	0.013	0.006	0.984

Predicted distributions of immature and mature capelin using the SI model are given in Table 4 for the different MULTSPEC areas. This is an alternative way of implementing spatial detail in GADGET. More examples of distribution matrices and migration matrices are given in Appendix 2.

Table 4. Example of distribution of Barents Sea capelin generated by the SI model to fit GADGET. The distributions are for October 1990.

Area	Immature	Mature
1	0.000	0.000
2	0.000	0.000
3	0.034	0.000
4	0.002	0.000
5	0.158	0.012
6	0.189	0.669
7	0.618	0.320

7.6.2 The Kolmogorov model: description and results

Model description

Consider the motion of a single fish. We can assume that the fish is undergoing a biased random walk in the plane with discrete time and spatial steps and by taking the continuum limit obtain a Kolmogorov type equation for the probability density of the location of the fish (see Magnússon *et al.,* in press). Alternatively, we can assume that the position in a domain Ω is governed by a two dimensional stochastic differential equation

$$d\boldsymbol{X}_t = \boldsymbol{V}\left(\boldsymbol{X}_t, t\right) dt + B\left(X_t, t\right) d\boldsymbol{W}(t) \qquad (6)$$

where $\boldsymbol{X}_t = (X, Y)$ is the random variable for the position of the fish, $\boldsymbol{V} = (v_1, v_2)$ is a given velocity, *B* is a given 2x2 matrix and *W(t)* is a continuous 2 dimensional Brownian motion. The conditional probability density at time *t* for the position of the fish, given that the fish is at (x_0, y_0) at time t_0, $p = p\left(x, y, t | x_0, y_0, t_0\right)$, satisfies Kolmogorov's forward equation (Øksendal, 2000)

$$\frac{\partial p}{\partial t} = -\nabla \cdot (\boldsymbol{V} p) + \frac{1}{2}\left(\frac{\partial^2\left(D_1 p\right)}{\partial x^2} + 2\frac{\partial^2\left(D_{12} p\right)}{\partial x \partial y} + \frac{\partial^2\left(D_2 p\right)}{\partial y^2}\right), \quad (x, y) \in \Omega \quad (7)$$

where

$$BB^T = \begin{bmatrix} D_1 & D_{12} \\ D_{12} & D_2 \end{bmatrix}$$

If ***V*** and *B* are constant, then it is easy to see that

$$E\left[\boldsymbol{X}_t - \boldsymbol{X}_0\right] = \boldsymbol{V} \cdot t$$

and

$$Cov\left[\boldsymbol{X}_t\right] = \left(BB^T\right) \cdot t$$

Equation (7) resembles an advection-diffusion equation but differs from it in that it contains a mixed second derivative term and the second derivative is taken of the multiple of the diffusion coefficients *D* and the density *p*. If $D_{12} = 0$ and diffusion is isotropic, then equation (7) reduces to an advection-diffusion type equation:

$$\frac{\partial p}{\partial t} = -\nabla \cdot (\boldsymbol{V} p) + \frac{1}{2}\left(\frac{\partial^2 (Dp)}{\partial x^2} + \frac{\partial^2 (Dp)}{\partial y^2}\right) \qquad (8)$$

It is not possible to distinguish between models (7) and (8) on the basis of the available measurements and we will therefore employ the simpler model in the simulations described here. However, anisotropic diffusion is not impossible, for example diffusion may be greater along the direction of motion than perpendicular to it. We will also assume for simplicity that the diffusion coefficient is constant although diffusion might be higher on feeding grounds or spawning grounds when fish are searching for food or spawning sites (cf. Magnússon *et al.,* in press).

If we assume that there is zero probability of a fish escaping through the boundary of Ω, the natural boundary condition is

$$\left(-v_1 p + \frac{\partial}{\partial x}\left(D_1 p\right) + \frac{\partial}{\partial y}\left(D_{12} p\right)\right) n_1 + \left(-v_2 p + \frac{\partial}{\partial y}\left(D_2 p\right) + \frac{\partial}{\partial x}\left(D_{12} p\right)\right) n_2 = 0, (x, y) \in \partial\Omega \quad (9)$$

where (n_1, n_2) denotes the unit outward normal vector to the boundary. The advection velocity ***V*** is determined by the environmental fields and the location of the fish relative to the spawning grounds. It is the vector sum of the velocity of the fish relative to the surrounding fluid and the velocity of the current, denoted by V_p and V_c respectively.

The direction of the velocity vector V_p is determined by the gradient of a function $U\left(x, y; t\right)$, which we will refer to as a comfort function (Reed and Balchen, 1982). This function will incorporate the factors, which are believed to affect the "comfort" or well-being of the fish, such as temperature, food density, location relative to the spawning grounds, etc. The fish move on a time-varying "comfort surface" defined by $U = U\left(x, y; t\right)$ *c*onstantly attempting to maximize their comfort, i.e. moving in the direction of the spatial gradient of *U*. The comfort function is of the form

$$U\left(x, y; t\right) = \alpha_1\left(t\right) r\left(T\left(x, y; t\right)\right) + \alpha_2\left(t\right) s\left(f\left(x, y; t\right)\right) + \alpha_3\left(t\right) \phi\left(x, y\right) \qquad (10)$$

where *T(x,y;t)* and *f(x,y;t)* are the temperature and food density respectively at location *(x,y)* at time *t* and *r* and *s* are given functions defined as follows:

$$r(T) = \begin{cases} -\left(T - T_1\right)^4 & if \quad T \leq T_1 \\ 0 & if \quad T_1 \leq T \leq T_2 \\ -\left(T - T_2\right)^2 & if \quad T_2 \leq T \end{cases} \qquad (11)$$

and

$$s(f) = \frac{f}{h + f} \qquad (12)$$

where T_1, T_2 and *h* are constants. The functional form of *r* implies that the preferred temperature range is between T_1 and T_2 degrees. The fish tend to move towards areas where the temperature is within the preferred range and this tendency is stronger in cold waters. The preferred temperature range may vary within the year. The function *s* is an asymptotic and increasing function of food density *f*. This functional form implies that the tendency to move towards areas of higher food density decreases with increasing density and is virtually nil at high food densities when the density is sufficiently high to satisfy the requirements of the fish.

The function ϕ is a potential function for the attraction towards the spawning grounds and is defined as follows: Let the region in which the fish can move be denoted by Ω. We define the boundary of Ω to be any lines, which the fish are not able to cross, such as the coastlines and possibly certain isotherms. The region is in general not simply connected since it may have one or more "holes" (islands) in it. Assume that the spawning grounds cover a sub-area S of Ω and that migrating fish are attracted towards S, in the sense that they experience a "force field" pulling them towards it. The spawning region may be regarded as a continuous sink spread over S, to use an analogy from fluid dynamics, and we define the sink density (analogous to mass density or charge density) in the spawning region S by $\rho\left(x, y\right)$. The total attraction strength (sink strength) of the spawning area is therefore given by

$$m = \int\int_S \rho dA$$

The potential ϕ is given as the solution of Poisson's equation

$$\Delta\phi = -\rho \qquad \text{in} \quad \Omega \qquad (13)$$

subject to no-flux condition at the boundaries of Ω, i.e.

$$\frac{\partial \phi}{\partial n} = 0 \qquad \text{on} \quad \partial\Omega \qquad\qquad (14)$$

The force field attracting the fish towards the spawning grounds is given by the gradient of ϕ $\nabla\phi$. The gradient is orthogonal to the equi-potential lines and is therefore tangent to the set of lines known as streamlines, which the fish would travel along if there were no other factors influencing the motion, apart from possible obstacles in the form of holes within Ω, that the fish cannot cross.

To take a very simple example, we can consider an infinite region Ω with no boundaries and an attracting region consisting of a single point of sink strength *m*, located at the origin. The attraction density is $\rho\left(\right) = m\delta\left(\right)$, where $\delta\left(\right)$ is the delta function, the potential function is $\phi\left(x, y\right) = -\frac{m}{2\pi} \ln \sqrt{x^2 + y^2}$ and the streamlines are simply straight lines through the origin. We describe below the field generated by an attracting spawning region off the coast of North Norway and Russia.

The gradient of the comfort function gives the direction of the velocity vector V_p, but the speed must be specified external to the model. The value of this parameter – which can be regarded as the average cruising speed- may be deduced from observations in experimental tanks and in the field, and is typically about one body-length per second or about 15 km/day for 17-18 cm capelin. However, cruising speeds may be as high as 20-30 nautical miles per day (Vilhjálmsson, 1994, p. 53). Alternatively, the speed may be estimated by comparing simulated to observed distributions. The velocity in the Kolmogorov equation is the vector sum of V_p and V_c where the latter is a vector field, which is specified external to the model. It might be argued that the reaction of the fish to the current field should be included in the comfort function, but the velocity relative to the surrounding fluid probably depends more directly on other environmental fields than the overall velocity.

Numerical simulations

We present here simulations of the spawning migrations of capelin in the Barents Sea and of the spatio-temporal distribution of immature capelin in 1990-1991. The migrations of the mature capelin are simulated over the period October 1 to May 31.
This is a summary of more detailed results presented in Dereksdóttir *et al.* (2003b), including comparisons with simulations from years 1980-81 and 1993-94, and sensitivity analysis of some of the free parameters.

The temperature field used in the simulations is based on twice yearly measurements made available by the Institute of Marine Research in Bergen, Norway, with spatial and temporal interpolation and some extrapolation. The assumed preferred temperature range (Fig. 4) is taken to be time-varying and different for immature and mature capelin. The field lines generated by an attracting spawning region on the coast of North Norway and Russia are shown in Fig. 5a. The attraction density ρ is taken to be near uniform over the spawning region. Since there are at present no data on food density, in a suitable format, we construct an artificial food density field using an attracting region in the northeast corner of the area. The field lines are shown in Fig. 5b. The temperature in the “attracting feeding region” and the whole northeast corner is too low for the simulated (and real-life capelin as well) to enter this region. The effect is therefore to draw the fish towards the true feeding grounds in the northeast between Svalbard and Novaya Zemlya, rather than into this hypothetical attracting region. The simulation model includes the option of using fields of oceanic currents as described in the previous section. However, at present we do not have access to such data from the Barents Sea and are therefore cannot include this factor at present. The present model has been adapted to describe the spawning migrations of capelin in Icelandic waters where current fields are included, see Dereksdóttir *et al.* (2003a) and turn out to play a significant role. Fig. 5 shows the level curves of the comfort function, with and without spawning attraction.

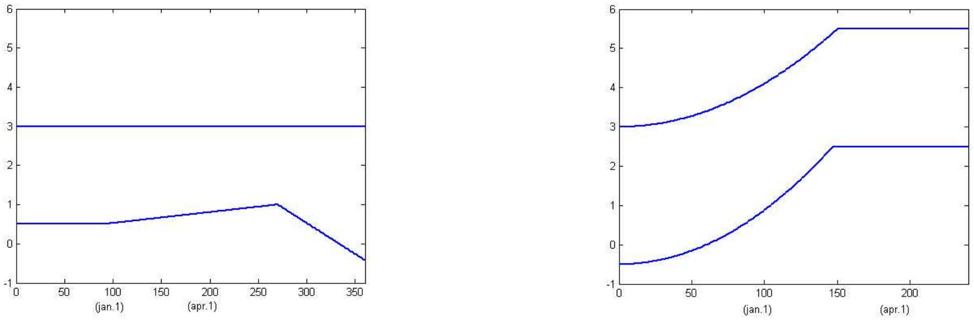

a) b)

Figure 4. The preferred temperature range in °C as a function of time for immature capelin (a) and mature capelin (b).


a) b)

Figure 5. The Barents Sea and the field lines for the attraction force generated by a spawning area on the coast of northern Norway (a) and a feeding area in the northeast (a).

Since we use an attracting feeding area as a proxy for a food density field, the comfort function is given by

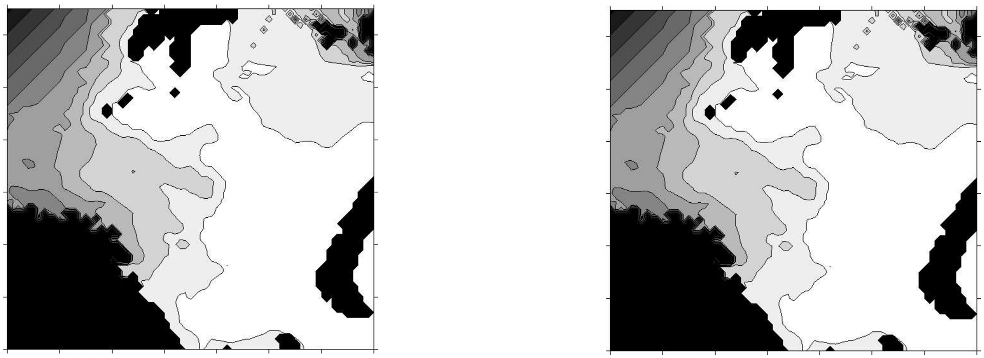
$$U(x, y; t) = \alpha_1(t)\, r(T(x, y; t)) + \alpha_2(t)\, \psi(x, y) + \alpha_3(t)\, \phi(x, y)$$

where ϕ and ψ are the potential fields generated by the attracting spawning and feeding regions respectively whose field lines are shown in Fig. 6. The simulation of the spawning migration from October 1 to May 31 spans a total of 243 days. The values of the α – coefficients in the comfort function are as follows (note that only the relative values matter):

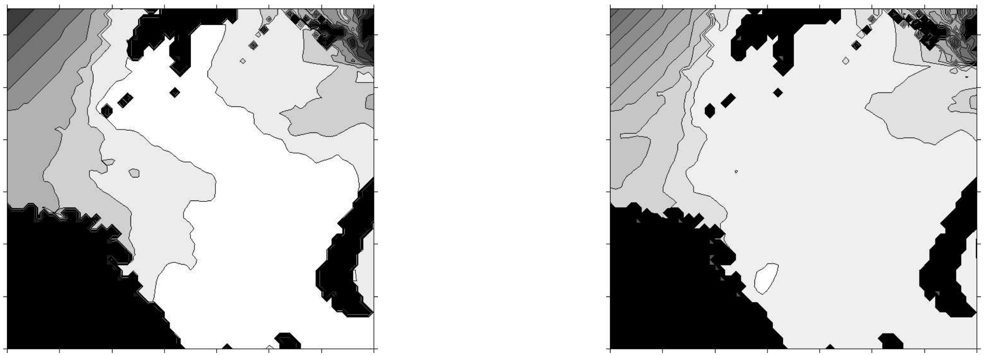
	Oct. 1 – Nov. 20	Nov. 20 – Jan. 1	Jan. 1 – March 1	March 1 – June 1
α_1:	1	1	1	1
α_2:	0.0005	0	0	0
α_3:	0	0	0.65	10

The temperature effect dominates until January 1, with a slight feeding attraction for the first 50 days. The spawning attraction becomes active on January 1, the temperature effect is stronger until March 1 and the spawning attraction dominates thereafter. The speed $s = ||\boldsymbol{V}_p||$ is set at 18 km/day and the diffusion coefficient *D* at 800 km^2/day. The values of the α– coefficients, *s* and *D* are selected such that the simulated distribution resembles those observed in surveys. The initial distribution on October 1 is obtained from acoustic surveys in September each year.

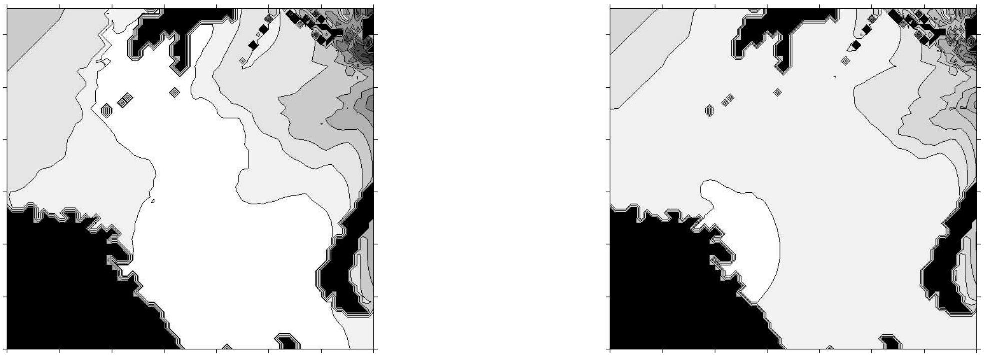
December 1 1990 December 1, 1990



January 1, 1991 January 1, 1991



February 1, 1991 February 1, 1991



April 1, 1991 April 1, 1991

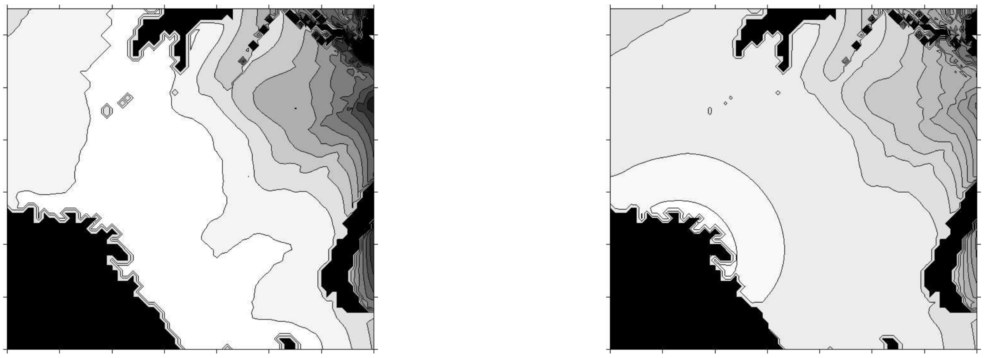


Figure 6. The level curves for the comfort function in 1990-1991 with (right) and without spawning attraction (left), lighter shades represent higher levels of comfort.

Equation (8) is solved by a numerical scheme based on a linear staggered Galerkin finite element approximation with upwinding in space and a second order Runge-Kutta approximation in time. The area that is simulated is divided into 2400 triangular elements of size 200 km^2 each and the time-step is 0.1 day. For more details on the numerical methods see Dereksdóttir *et al*. (1993a). Figs 7 and 8 show the simulated spawning migration with "snapshots" of the distribution of the mature capelin at eight selected points in time.

In order to facilitate a compression of the simulation results from the model, which is continuous in space and time, we use the seven MULTISPEC areas, shown in Fig. 9. These are taken from Tjelmeland and Bogstad (1998), and Table A3 in the appendix shows results from the same simulation, aggregated to show the percentage in each sub-area.

Reduction into transition matrix models

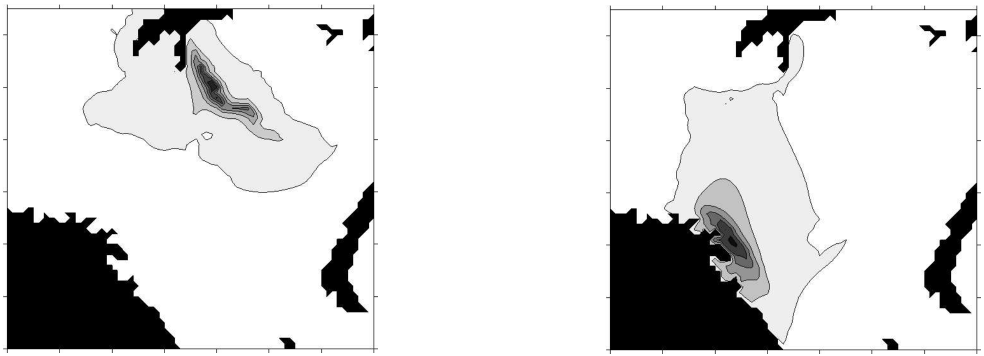
The longterm aim with the migration model is to constitute a part of a larger multi-species model like GADGET incorporating migrations, recruitment, predation, catches, growth, etc. However, the overall computing requirements are too great at present. It is therefore necessary to try to reduce the continuous model to a simpler one, which is easier to use and less computer intensive.

One way whereby this can be achieved is converting the continuous model into an equivalent transition matrix model (also referred to as migration matrix model). A transition matrix at time *t* is an *NxN* matrix *M(t)* where the element in row *i* and column *j*, *m(i,j,t)*, is the fraction of fish in sub-area *i* which moves to sub-area *j* between times *t* and *t*+1 (this time-step is typically one or two months). If ***n***(*t*) is the distribution vector at time *t* (i.e. the density in each of the N sub-areas) then ***n***(*t*+1) is obtained by

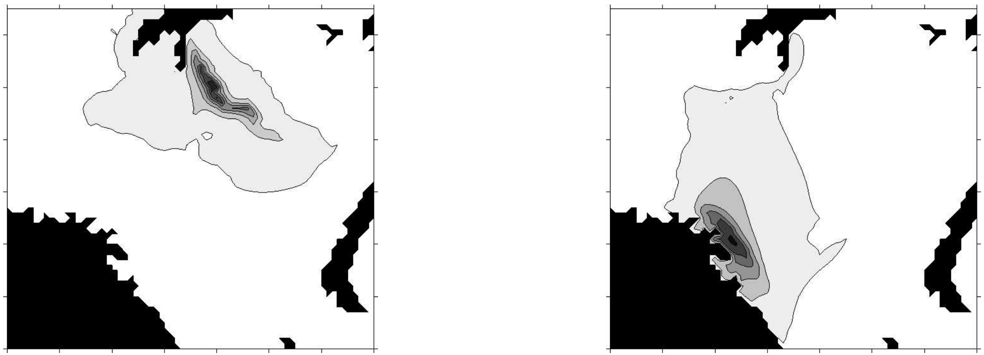
$$\boldsymbol{n}^T\,(t+1) = \boldsymbol{n}^T\,(t)\,M(t)$$

Since the partial differential equation governing the evolution of the density *p* (equation (8)) is linear we can apply the principle of superposition, i.e. if p_i is the solution corresponding to initial condition p_{0i}, then $\sum_i p_i$ is the solution correspond-ing to initial condition $\sum_i p_{0i}$. We can therefore calculate each of the rows in the transition matrix separately. Assuming that all the mass is at time *t* confined to sub-area nr. 1, we solve equation (8) with this as the initial condition and at the end of one time-step, compute the fraction of the initial mass in each of the *N* sub-areas which we also refer to as compartments in this context. The fraction in compartment *j*, *j*=1,2,…,*N*, is element (1,j) in the transition matrix. This exercise therefore gives the first row in the transition matrix at time *t.* Repeating for sub-areas 2,3,…,*N*, the transition matrix is obtained. There are, however, a number of issues that have to be considered.

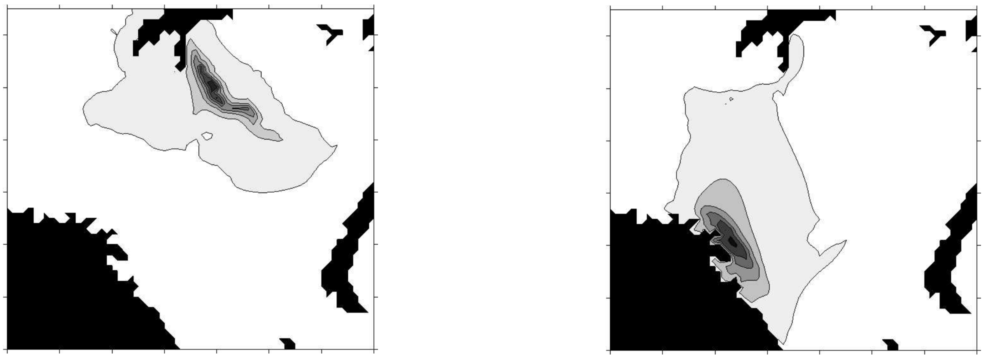
October 15, 1990 February 15, 1991



November 15, 1990 March 15, 1991



December 15, 1990 April 15, 1991



January 15, 1991 May 15, 1991

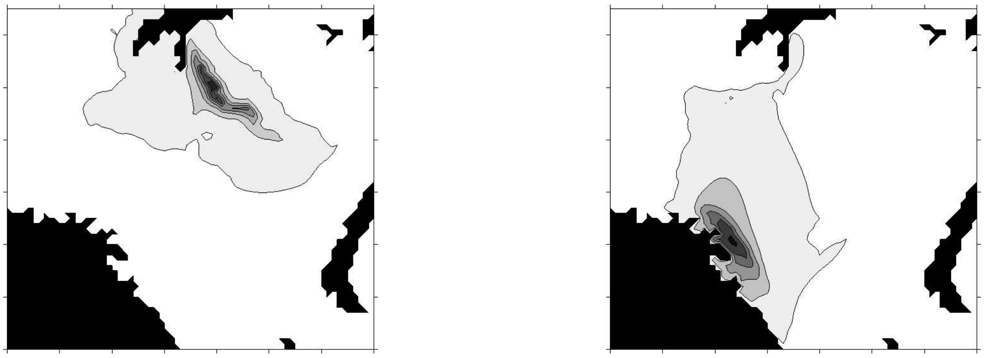
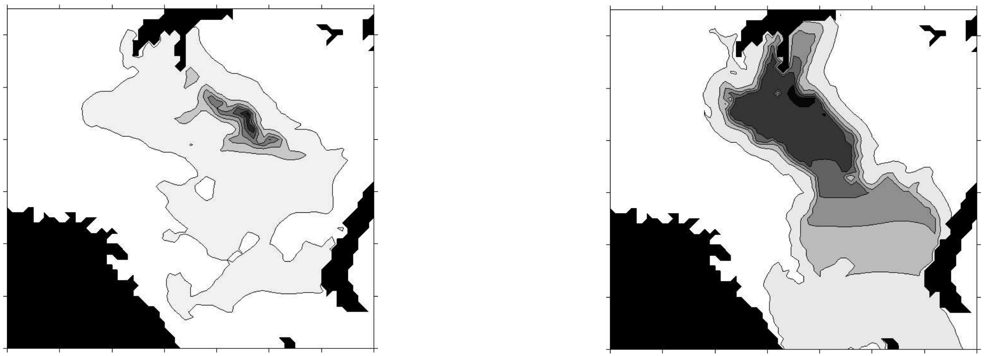


Figure 7. Simulated spawning migration of mature capelin in the Barents Sea from October 1990 to May 1991.

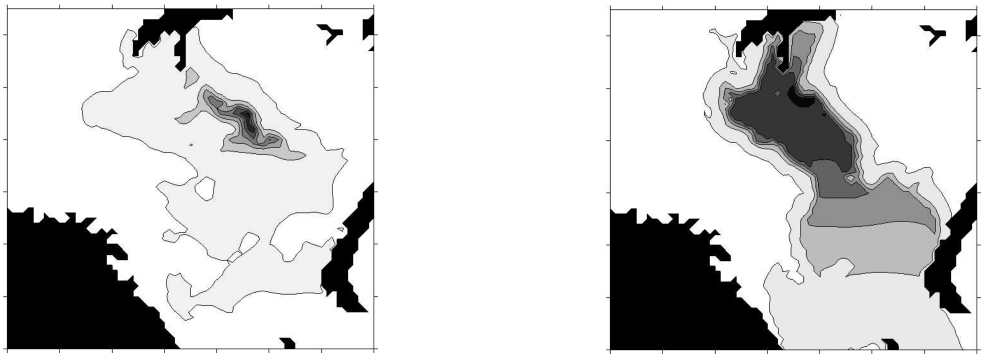
October 15, 1990

April 15, 1991



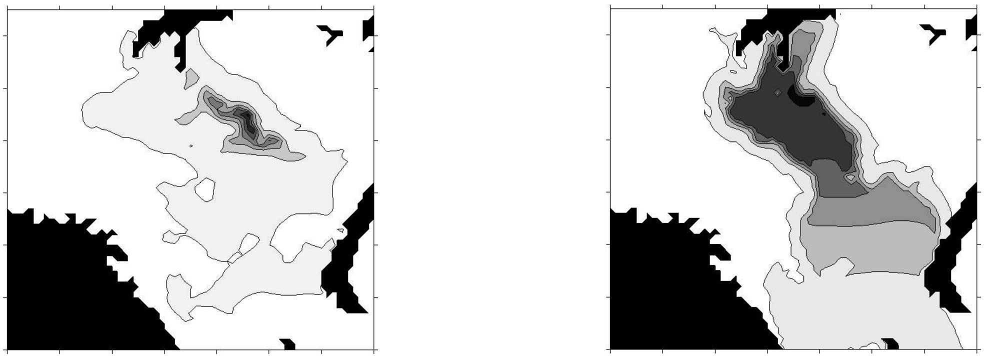
December 15, 1990

May 15, 1991



January 15, 1991

June 15, 1991



February 15, 1991

August 15, 1991

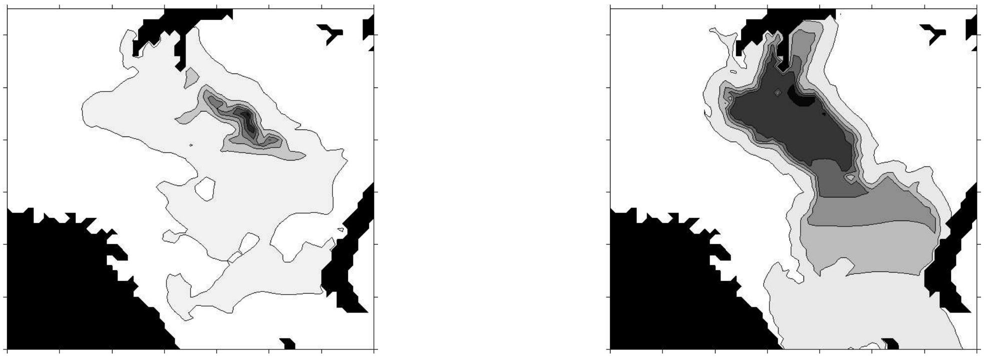


Figure 8. Simulated spawning migration of immature capelin in the Barents Sea from October 1990 to August 1991

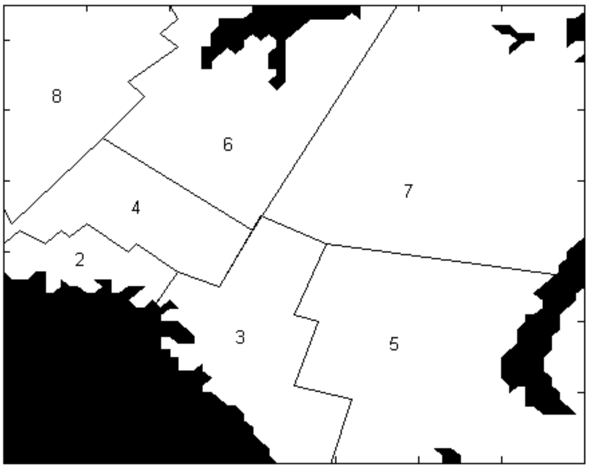


Figure 9. The division of the Barents Sea into sub-areas used in the transition matrix migration model.

Firstly, we must consider how to spread the initial mass over the given compartment. Keeping in mind that the aim is to use the transition model for predictive purposes we have to be careful not to introduce bias into these predictions. With no a-priori information on the whereabouts of the fish the consistent choice is to spread the fish initially uniformly over the whole compartment. Since the comfort function in the model tends to attract the fish towards higher function values an alternative choice is to spread the fish so that its density reflects the values of this function at the initial time, *t*. We have adopted a simple implementation of the latter approach where we let the initial density within a compartment be ten times higher in areas of preferred temperature than in the remaining area of the compartment. In Table 5 we show a migration matrix for February 1991 based on the area subdivision in Figure 9.

Table 5. Calculated migration matrix for mature capelin in the Barents Sea for February 1991. The entry in row *i* and column *j* shows the percentage moving from compartment *i* to compartment *j*.

	2	**3**	**4**	**5**	**6**	**7**	**8**
2	72.0	20.3	7.6	0	0	0	0
3	4.1	93.7	1.8	0.4	0	0	0
4	49.4	34.1	15.1	0	1.3	0	0.2
5	0.4	82.6	0.4	16.5	0.0	0.1	0
6	21.7	24.8	32.9	0.1	19.4	0.8	0.3
7	6.1	47.2	9.7	11.6	9.8	15.5	0
8	14.7	1.7	35.1	0	36.4	0.1	12.1

A second issue is the length of the time step. From the viewpoint of the larger multispecies model a smaller time step allows more resolution but a larger time step reduces the amount of data that has to be transferred between the migration model and the larger model. A possible compromise is to use a time step of 1 month say, and to define a number of distinct environmental categories or scenarios, e.g. warm, average or cold years, and assign each year for which environmental data are available to one of these categories. In that case 36 transition matrices have to be transferred. In order to estimate migrations over longer time intervals we multiply together such monthly matrices for consecutive time steps as is done in a transition matrix model. Note, however, that with this approach the fish are effectively re-distributed within each compartment, according to the preferred temperature, at the beginning of each transition step. Thus the transition matrix models will tend to exaggerate the spread compared to an advection-diffusion model. Furthermore, within large sub-areas environmental heterogeneity can have a substantial influence on the motion in a continuous model, which cannot be addressed in a matrix model.

As an indicator of these effects we compare in Table 6 the product of 5 monthly transition matrices for mature capelin from October 1990 till February 1991 with a transition matrix over the same period using a single transition step of 5 months, using initial distribution based on preferred temperature. There is a fair agreement that has to be balanced against the amount of transition data that we are prepared to transfer between the models.

Table 6. Calculated migration matrices for mature capelin in the Barents Sea for October 1 1990 till March 1 1991, calculated as a product of five matrices, each with a transition step of one month (top), and as a single matrix with a transition step of five months (bottom). The entry in row *i* and column *j* shows percentage moving from compartment *i* to compartment *j*.

product of five monthly matrices

	2	**3**	**4**	**5**	**6**	**7**	**8**
2	15.5	59.4	14.9	1.6	7.7	0.9	0.2
3	6.6	82.9	4.8	2.9	2.0	0.7	0
4	14.4	61.5	13.9	1.9	7.2	1.0	0.1
5	7.4	80.3	5.8	3.1	2.6	0.9	0
6	14.3	60.4	14.2	2.2	7.5	1.2	0.1
7	10.5	70.9	9.5	3.0	4.8	1.3	0.1
8	16.8	54.1	17.1	1.6	9.1	1.1	0.2

single five month matrix

	2	**3**	**4**	**5**	**6**	**7**	**8**
2	19.6	66.8	11.7	0.2	1.6	0.1	0.1
3	4.3	92.4	2.1	1.0	0.1	0	0
4	18.5	68.2	11.2	0.3	1.7	0.1	0
5	6.0	89.2	3.1	1.5	0.2	0.1	0
6	20.6	61.9	13.6	0.2	3.3	0.2	0.1
7	13.9	75.9	8.0	1.0	1.1	0.1	0
8	24.1	54.8	16.5	0.1	4.2	0.1	0.1

For reference purposes we also include in table A3 in the appendix calculated monthly migration matrices for mature capelin in the Barents Sea for the whole period October 1 1990 till June 1 1991.

The final issue to be considered is the size of compartments where we again must strike a balance between the resolution and the amount of data transferred between models. By reducing the size of the compartments we clearly avoid the problems described above in the context of Table 6 and reduce the additional spread caused by the initial redistribution of mass at each transition step. A finite element realization of the continuous model can indeed be viewed as a compartmental model with thousands of compartments. On the basis of the simulation results it would, however, appear that the division of the Barents Sea into the compartments shown in figure 9 is not the most suitable one for the purpose of calculating migration matrices, since the environmental heterogeneity is too great within some compartments.

In order to summarize the plausibility of replacing a continuous model with a transition matrix model we compare in Table 7 the results of using a continuous model directly to predict the whereabouts of mature capelin over the period October 1 1991 till June 1 1991 with the use of compartmental models with the same initial data. In the latter case we distinguish between using a product of monthly transition matrices and a single transition matrix.

Table 7. Comparison of calculated distribution (in percentages) between compartments of mature capelin in the Barents Sea November 1 1990 till June 1 based on a given initial distribution October 1 1990. Calculated values obtained using a continuous model directly, and using matrix transition models, distinguishing between using a product of monthly transition matrices and a single transition matrix.

		January 1, 1991			March 1, 1991			May 1, 1991		
Compart ment	Initial distrib	Cont model	Trans matrix		Cont. model	Trans matrix		Cont. model	Trans matrix	
			prod.	single		prod.	single		prod.	single
2	0.0	0.0	0.0	0.0	17.9	12.8	17.9	20.8	19.7	21.1
3	0.0	0.3	0.8	0.4	68.7	64.6	67.6	74.6	75.0	74.2
4	0.0	0.0	0.1	0.1	10.9	12.3	11.4	4.5	4.8	4.5
5	1.2	4.7	16.1	5.6	0.5	2.5	0.5	0.1	0.3	0.1
6	60.9	47.0	36.0	48.5	1.9	6.4	2.4	0.0	0.2	0.0
7	37.9	48.0	46.8	45.3	0.2	1.2	0.2	0.0	0.1	0.0
8	0.0	0.0	0.1	0.1	0.0	0.1	0.1	0.0	0.0	0.0

There is consistently a better agreement between the matrix transition model and the continuous model when a single transition matrix is used rather than a product of monthly matrices, as is to be expected. It is, however, notable that the loss in agreement by using monthly matrices does not seem to increase significantly as the total simulation period increases. Thus the results indicate that monthly transition matrices, with the matrices being conditional on environmental conditions, may carry with them sufficiently detailed information on migration to act as a transfer mechanism between the migration model and a larger multispecies model, provided the comparmental division is modified.

7.6.3 The Borealis model: description and results

BorealisMigration is an individ-based migration model system intended for use with Mathematica-based models and programs for use with assessment of fish in the Norwegian Sea – Barents Sea ecosystem. The Borealis system is at present used with management of Barents Sea capelin (Gjøsæter *et al*, 2002) and Norwegian spring spawning herring and work is underway to estimate multispecies harvesting control rules for the Barents Sea.

The fish distribution is represented with a number of assemblages of fish with identical position, sex, age and maturation stage. Each assemblage is further characterized by a number of fish and a geographical distribution of the fish around the assemblage position.

At present only a strictly temperature driven migration model is implemented, where each assemblage moves along the maximum temperature gradient to a preferred temperature in one time step, constrained by a maximum swimming speed. In later versions current, depth and possibly magnetic fields will be added as possible explanatory fields in the migration. The capelin is forced to move due south north of 74 degrees N during October-December, or else the temperature driven model will lead to a substantial large part of the population migrating due west towards Svalbard during this period.

The temperature and current fields are simulated from an oceanographic model. The grid size is 20 km. At present fields from restricted periods of time only are available. An oceanographic model extending to the Florida region is expected to give fields for as long a period as there are acoustic capelin data by the end of August 2003 (Paul Budgell, pers. comm.). The fields are smoothed by a 2-dim 5-point moving average before they are used in BorealisMigration.

The parameters in the migration model are estimated in a classical sense where a likelihood function is constructed by evaluating the probability of observations conditional on the simulation results. As observations are implemented the distribution of acoustic integrator values and the distribution of number of fish on age, sex and Multspec areas. In later versions of the model the geographical distribution of fish in cod stomachs may be added.

BorealisMigration is developed with modelling the migration of mature capelin from the end of September until spawning about April 1 as first priority. The reason for this is that the overlap between capelin and cod (the geographic distribution of cid is measured in February) is important for the model of predation of mature capelin by cod that is used in the capelin assessment model Bifrost.

The contribution to the overall likelihood function from comparing simulated geographical distribution and the geographical distribution of integrator values from acoustic surveys is evaluated by constructing the distribution of integrator values in a square given a true mean integrator value by bootstrapping from the model developed by Tjelmeland (2002). This model is based on the September survey which is directed towards capelin and may therefore underestimate the uncertainty connected to other surveys where the capelin is not a target species.

The contribution to the overall likelihood function from comparing the simulated distribution of fish on age, sex and Multspec areas is evaluated assuming a multinomial distribution. Even if in the present implementation neither sex nor age influences the migration, the distribution on sex and age will depend on the initial distribution on sex and age and thus be helpful in determining plausible migration parameters.

In connection with the dst^2 meeting in Bergen in August 2003 the BorealisMigration was used to model the migration of mature capelin from the end of September 1990 until spawning April 1 1991. The assemblages are based on the WMO squares underlying the capelin assessment. Figure 10 shows the initial distribution. Figure 11 shows gridded likelihood surfaces for the acoustic and biological terms in the likelihood function separately.

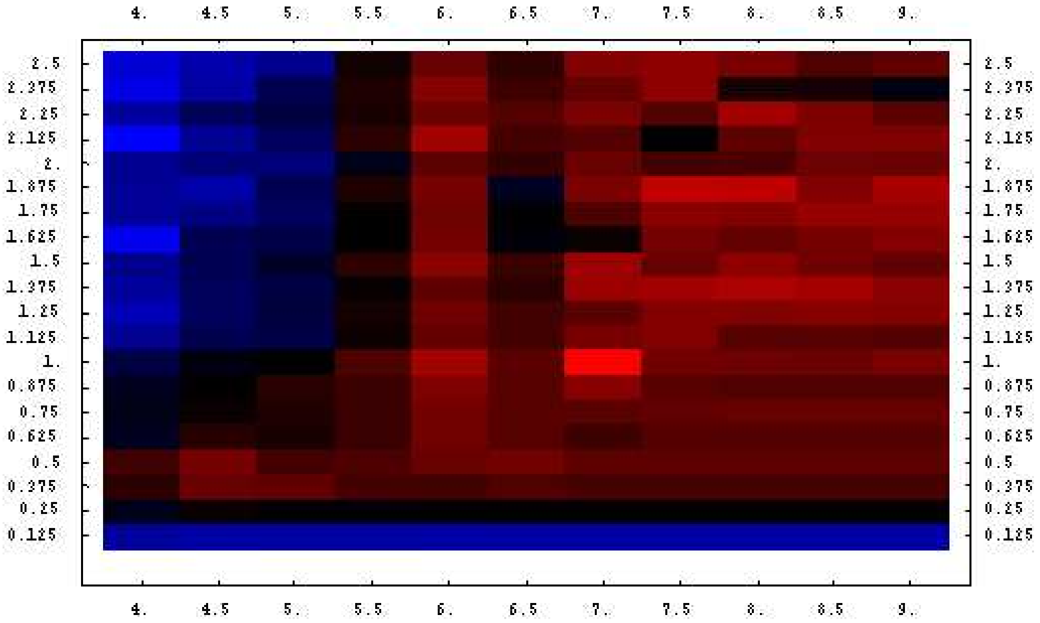


Figure 10. Initial distribution

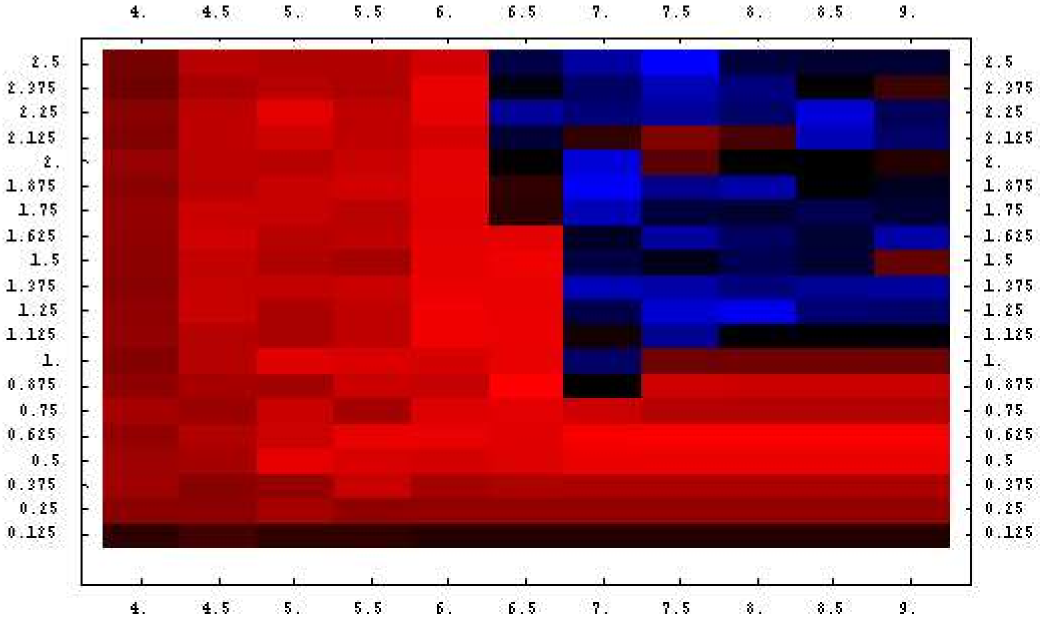
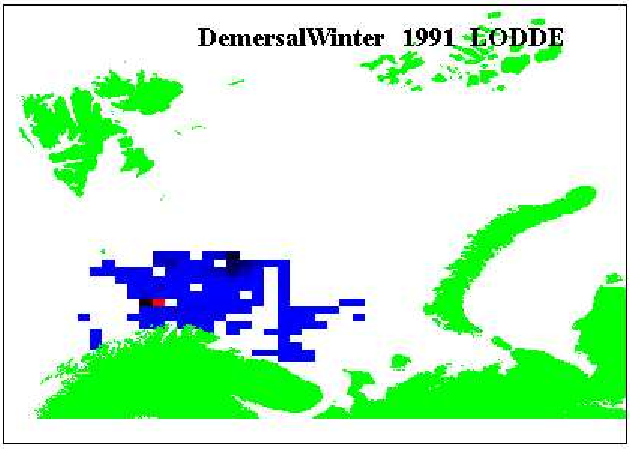


Figure 11. Likelihood surfaces, acoustic upper, biological lower. X-axis is preferred temperature and Y-axis is the maximum swimming speed expressed in body lengths per second.

Both likelihood surfaces show well-defined areas of high likelihood and well-defined areas of low likelihood. However, these areas do not overlap very well for the two different observation sets, which therefore give somewhat conflicting information with respect to the migration model being tested. Figure 12 shows the simulated distribution and the distribution of acoustic integrator values from the February survey.

Figure 13 shows the distribution by the end of March for a preferred temperature of 6.0 degrees and a swimming speed scaling of 1.25, which is well inside the region of high probability of the biological observations. The geographical distribution is here split into a western and an eastern part. Figure 14 shows a comparison of simulated and measured age distribution on Multspec areas for this case. Simulated and observed data agree fairly well, with the exception that there are no observations in area 5.



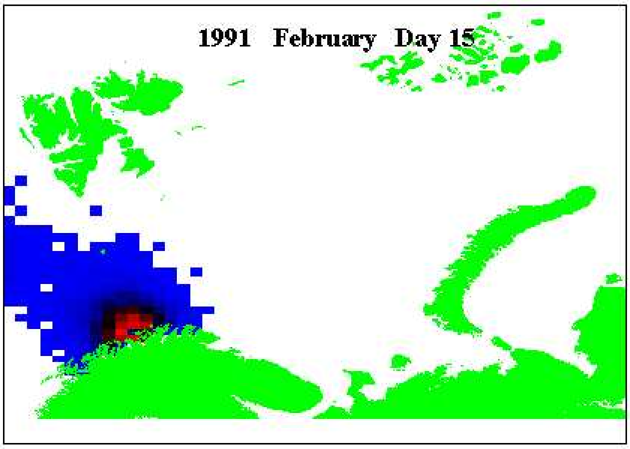


Figure 12. Distribution of acoustic integrator values (upper) and simulated distribution in mid-February (lower).

The simple temperature driven model tested here works surprisingly well, although giving a too large geographical spread by the end of the simulation period. Later additions of current and depth as driving fields may yield more realistic pre-spawning distributions.

A routine output of the simulations are migration matrices – the proportion of fish migrating from each area to other areas during a time step – for the Multspec areas. The migration matrices summarize the model and may be used as input to other models (Gadget, Bifrost).

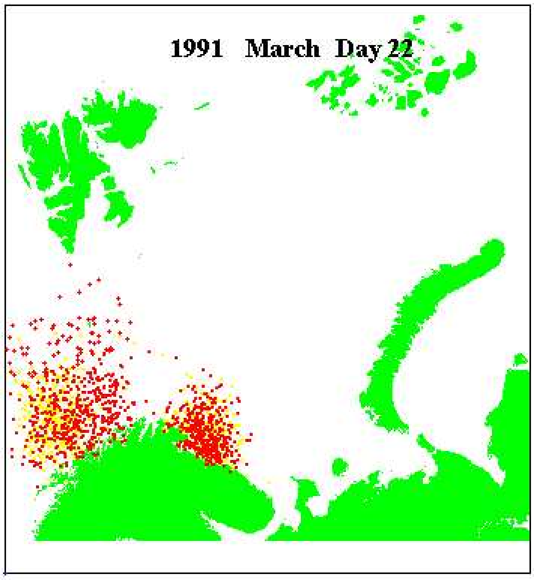


Figure 13. Simulated geographical distribution by the end of March for parameter values giving a high probability of the biological observations.

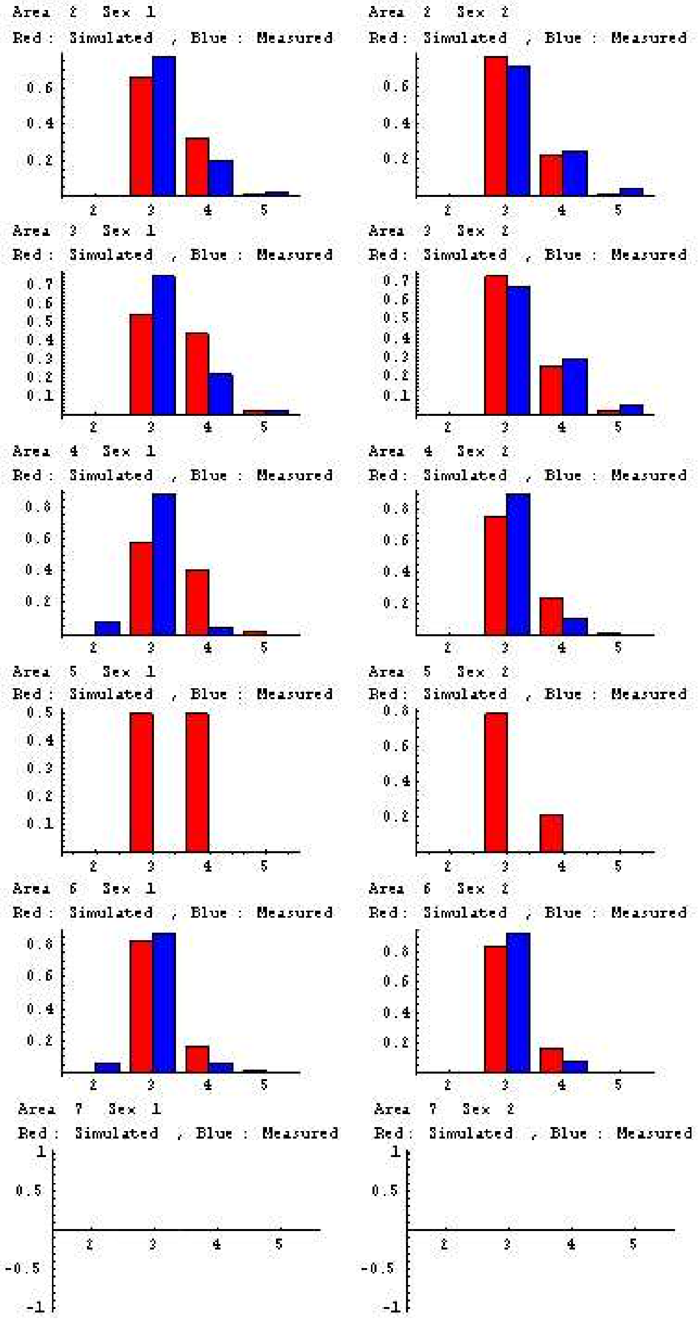


Figure 14. Comparison of simulated and measured age distributions for parameter values giving a high probability of the biological observations.

7.6.4 Discussion

A variety of techniques for simulating fish migration exist. The models applied here are composed independently and vary substantially in their structure. A common feature is that they can all be characterised as individual-based models. IBM is a very flexible technique and is likely to remain an important technique for simulating spatial movement of fish in years to come. GADGET should therefore be fitted to exchange information with such models.

The approach taken in GAGET, at present, is to describe migration in terms of transition matrices between compartments, and to estimate the coefficients of these matrices from e.g. tagging or acoustic data. Thus, perhaps the most natural role of more detailed migration models, like those described above, in the context of GADGET, is to provide a quantitative description of how the coefficients of such transition matrices may depend on environmental factors such as sea temperature. By using tagging or acoustic data to estimate some free parameters in these underlying models, rather than estimating the transition coefficients explicitly, we incorporate into the GADGET model the effects of environmental variability on migration. In the description of the Kolmogorov model, above, we have indicated how results from simulations with that model may be reduced to a matrix transition model. An important feature of that reduction is that it does not depend on any direct a priori assumptions on the whereabouts of the fish at the start of a given simulation period. If the GADGET model is to be used for prediction purposes care has to be taken not to introduce unnecessary bias. However, with the other two models it is less clear how to reduce results from simulations to a matrix model in a similar fashion. The considerable discrepancy between the transition matrices obtained from the super-individual model on one hand and the Kolmogorov model on the other, presented in the appendix below, arises from the fact that they are not derived in the same manner. Thus it would appear that new methods for incorporating movement from IBM into GADGET are needed. A standard format for reading distributions generated from IBM rather than going via the migration matrices should therefore be considered. This should be relatively easy to do, especially if the same spatial resolution is maintained. Table 4 shows an example of such a table, which gives the distribution of the stock for a particular month in a particular year rather than average flows as in the case of the migration matrices.

In conclusion we suggest that the migration matrix implementation that is used in GADGET at the moment is maintained. In addition to this, GADGET should be modified to allow the implementation of distribution matrices. This routine should be accompanied by code for exporting relevant data, such as number, weight and length at age in different areas, to individual-based models. This will make GADGET very flexible with regard to implementation of spatial detail at a low cost in terms of GADGET development or simulation time.

Acknowledgements

We thank Øyvind Ulltang, Daniel Howell and Geir Odd Johansen for participating in the discussions and meetings resulting in this report. This work was funded by the European Commission through the project "Development of structurally detailed statistically testable models of marine populations" (QLK5–CT1999–01609).

References

Anon. 1974. Gyteinnsiget av lodde vinteren 1974. Fisken og havet, 3-1974 (in Norwegian).

Chambers, C. R. 1993. Phenotypic variability in fish populations and its representation in individual-based models. Transactions of the American Fisheries Society, 122: 404-414.

DeAngelis, D. L., and Gross, L. J. (eds) (1992) Individual-based Models and Approaches in Ecology. Chapman & Hall, London. 525 pp.

Dereksdóttir, E. H., Magnússon, K. G. and Sigurdsson, S. 2003. Simulations of migrations of capelin in in Icelandic waters based on temperature-, current and potential fields. Report RH-16-2003, Science Institute, University of Iceland.

Dereksdóttir, E. H., Magnússon, K. G. and Sigurdsson, S. 2003. Simulations of spawning migrations of capelin in the Barents Sea based on temperature- and potential fields. Report RH-17-2003, Science Institute, University of Iceland.

Dolgov, A. V. 2002. The role of capelin (*Mallotus villosus*) in the foodweb of the Barents Sea. ICES Journal of Marine Science, 59: 1034-1045.

Dommasnes, A., Olsen, K., and Shleinik, V. N. 1974. Loddeundersøkelser i Barentshavet i november-desember 1973. Fisken og havet, 2-1974 (in Norwegian).

Gjøsæter, H. 1998. The population biology and exploitation of capelin (*Mallotus villosus*) in the Barents Sea. Sarsia, 83: 453-496.

Gjøsæter, H., Dommasnes, A., and Røttingen, B. 1998. The Barents Sea capelin stock 1972-1997. A synthesis of results from acoustic surveys. Sarsia, 83: 497-510.

Gjøsæter, H., Bogstad, B. And S. Tjelmeland. 2002. Assessment methodology for Barents Sea capelin, Mallotus villosus (Müller). ICES Journal of Marine Science, 59: 1086-1095.

Grimm, V. 1999. Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future? Ecological Modelling, 115: 129-148.

Hafsteinsson, M. T., and Misund, O. A. 1995. Recording the migration behavior of fish schools by multibeam sonar during conventional acoustic surveys. ICES Journal of Marine Science, 52: 915-924.

Hamre, J. 1994. Biodiversity and the exploitation of the main fish stocks in the Norwegian-Barents Sea ecosystem. Biodiversity and Conservation, 3: 473-492.

Hamre, J., and Monstad, T. 1980. Loddeundersøkelser i Barentshavet vinteren 1980. Fisken og havet, 5-1980 (in Norwegian).

Hassel, A., Skjoldal, H. R., Gjøsæter, H., Loeng, H., and Omli, L. 1991. Impact of grazing from capelin (*Mallotus villosus*) on zooplankton: a case study from the northern Barents Sea. Polar Research, 10: 371-388.

Huse, G., Giske, J., and Salvanes, A. G. V. 2002. Individual-based models. *In* Handbook of Fish and Fisheries, pp. 228-248. Ed. by P. J. B. Hart and J. Reynolds. Blackwell Science, Oxford.

Huse, G., and Gjøsæter, H. 1999. A neural network approach for predicting stock abundance of the Barents Sea capelin. Sarsia, 84: 457-464.

Magnússon, K. G., Sigurdsson, S., Babak, P., Gudmundsson, S. F. and Derekesdóttir, E. H. A continuous density Kolmogorov type model for a migrating fish stock. Discrete and continuous dynamical systems – Series B (in press).

Mehl, S. 1989. The Northeast Arctic cod stock's consumption of commercially exploited prey species in 1984-1986. Rapports et Procès-Verbaux des Réunions du Conseil International pour l'Exploration de la Mer, 188: 185-205.

Mehl, S., and Sunnanå, K. 1991. Changes in growth of Northeast Arctic cod in relation to food consumption in 1984-1988. ICES Marine Science Symposia, 193: 109-112.

Ozhigin, V. K., and Luka, G. I. 1985. Some peculiarities of capelin migrations depending on the thermal conditions in the Barents Sea. *In* The Soviet-Norwegian symposium on the Barents Sea capelin, pp. 135-148. Ed. by H. Gjøsæter. Institute of Marine Research, Bergen, Norway.

Reed, M. J. and Balchen, J. G. 1982. A multidimensional continuum model of the fish dynamics and behaviour: Application to the Barents Sea capelin (*Mallotus villosus*) Modelling, Identification and Control 3: 65-109

Rose, K. A., Christensen, S. W., and DeAngelis, D. L. 1993. Individual-based modeling of populations with high mortality: A new method based on following a fixed number of individuals. Ecological Modelling, 68: 273-292.

Scheffer, M., Baveco, J. M., DeAngelis, D. L., Rose, K. A., and van Nes, E. H. 1995. Super-individuals a simple solution for modelling large populations on an individual basis. Ecological Modelling, 80: 161-170.

Tereshchenko, V. V. 1996. Seasonal and year-to-year variations of temperature and salinity along the Kola meridian transect. ICES CM 1996/C: 11, 24 pp.

Toresen, R., Gjøsæter, H., and de Barros, P. 1998. The acoustic method as used in the abundance estimation of capelin (*Mallotus villosus* Müller) and herring (*Clupea harengus* Linné) in the Barents Sea. Fisheries Research, 34: 27-37.

Vilhjálmsson, H. 1994. The Icelandic capelin stock. Journal of the Marine Research Institute Reykjavik, Vol. XIII, No. 1. 281 pp

Ådlandsvik, B., and Sundby, S. 1994. Modelling the transport of cod larvae from the Lofoten area. ICES Marine Science Symposia, 198: 379-392.

Tjelmeland, S. and Bogstad, B. 1998. *Biological Modelling*, in “Models for Multispecies Management”, (ed. T. Rødsethe). Physica-Verlag.

Tjelmeland, S. 2002. A model for the uncertainty around the yearly trawl-acoustic estimate of Barents Sea capelin, Mallotus villosus (Müller). ICES Journal of Marine Science, 59: 1072-1080

Øksendal, B. 2000. *Stochastic Differential Equations*. Springer.

Appendix 1. Participants at the migration meetings

November 2002

Kjartan Magnusson, University of Iceland
Thomas Philip Runarsson, University of Iceland
Sven Sigurdsson, University of Iceland
Eva Hlin Dereksdottir, University of Iceland
Stefan Fr. Gudmunddson, University of Iceland
Sigurd Tjelmeland, Institute of Marine Research
Bjarte Bogstad, Institute of Marine Research
Edda Johannesen, Institute of Marine Research
Daniel Howell, Institute of Marine Research
Øyvind Ulltang, University of Bergen
Geir Huse, University of Bergen
Geir Odd Johansen, University of Bergen

August 2003

Sven Sigurdsson, University of Iceland
Eva Hlin Dereksdottir, University of Iceland
Sigurd Tjelmeland, Institute of Marine Research
Daniel Howell, Institute of Marine Research
Øyvind Ulltang, University of Bergen
Geir Huse, Instiutute of Marine Research

Appendix 2. Distribution tables and migration matrices for GADGET.

Table A1. Distribution tables of capelin at the start of the month during 1990-1991 taken from the super-individual model.

Month	Area	Immature	Mature
10	1	0.000	0.000
	2	0.000	0.000
	3	0.034	0.000
	4	0.002	0.000
	5	0.158	0.012
	6	0.189	0.669
	7	0.618	0.320
11	1	0.000	0.000
	2	0.000	0.000
	3	0.034	0.000
	4	0.000	0.000
	5	0.253	0.067
	6	0.231	0.792
	7	0.482	0.141
12	1	0.000	0.000
	2	0.000	0.000
	3	0.030	0.032
	4	0.000	0.000
	5	0.309	0.054
	6	0.275	0.856
	7	0.386	0.058
1	1	0.000	0.000
	2	0.000	0.000
	3	0.026	0.033
	4	0.000	0.000
	5	0.359	0.074
	6	0.299	0.872
	7	0.316	0.021
2	1	0.000	0.000
	2	0.000	0.303
	3	0.059	0.232
	4	0.000	0.289
	5	0.469	0.046
	6	0.297	0.130
	7	0.174	0.000
3	1	0.000	0.000
	2	0.000	0.558
	3	0.086	0.340
	4	0.000	0.012
	5	0.534	0.012
	6	0.280	0.078
	7	0.100	0.000
4	1	0.000	0.000
	2	0.000	0.569
	3	0.079	0.331
	4	0.000	0.004
	5	0.592	0.022
	6	0.280	0.074

	7	0.049	0.000
5	1	0.000	0.000
	2	0.000	0.574
	3	0.073	0.337
	4	0.000	0.001
	5	0.612	0.017
	6	0.265	0.072
	7	0.050	0.000
6	1	0.000	0.000
	2	0.000	0.000
	3	0.050	0.000
	4	0.000	0.000
	5	0.627	0.000
	6	0.257	0.000
	7	0.066	0.000
7	1	0.000	0.000
	2	0.000	0.000
	3	0.034	0.000
	4	0.000	0.000
	5	0.583	0.000
	6	0.247	0.000
	7	0.135	0.000
8	1	0.000	0.000
	2	0.000	0.000
	3	0.021	0.000
	4	0.000	0.000
	5	0.462	0.000
	6	0.238	0.000
	7	0.279	0.000
9	1	0.000	0.000
	2	0.000	0.000
	3	0.012	0.000
	4	0.000	0.000
	5	0.326	0.000
	6	0.219	0.000
	7	0.443	0.000

Table A2. Migration matrices for 1990-1991 generated by the super-individual model.

1990	10 Immature						
	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	0.968	0.039	0.004	0	0
4	0	0	0	0.749	0	0	0
5	0	0	0.026	0	0.983	0	0.011
6	0	0	0	0.212	0	0.994	0.005
7	0	0	0.006	0	0.013	0.006	0.984
	10 Mature						
	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0

3	0	0	0.797	0	0	0	0
4	0	0	0	0.552	0	0	0
5	0	0	0.063	0	0.985	0	0.01
6	0	0	0.013	0.448	0	0.997	0.028
7	0	0	0.126	0	0.015	0.003	0.962
	11 Immature						
	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	0.952	0.046	0.004	0	0.001
4	0	0	0	0.862	0	0	0
5	0	0	0.038	0	0.983	0	0.012
6	0	0	0	0.092	0	0.996	0.006
7	0	0	0.01	0	0.013	0.004	0.981
	11 Mature						
	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	0.834	0	0.031	0	0.001
4	0	0	0	0.9	0	0	0
5	0	0	0.129	0	0.96	0	0.015
6	0	0	0.005	0.1	0	0.997	0.046
7	0	0	0.032	0	0.009	0.003	0.938
	12 Immature						
	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0
2	0	0.935	0	0.003	0	0	0
3	0	0	0.922	0.02	0.004	0	0.001
4	0	0.065	0	0.886	0	0	0
5	0	0	0.066	0	0.98	0	0.02
6	0	0	0	0.092	0	0.995	0.006
7	0	0	0.012	0	0.016	0.005	0.972
	12 Mature						
	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	0.874	0	0.064	0	0.012
4	0	0	0	0.81	0	0	0
5	0	0	0.116	0	0.929	0	0.025
6	0	0	0	0.19	0	0.999	0.039
7	0	0	0.01	0	0.007	0.001	0.924
1991	1 Immature						
	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0
2	0	0.935	0	0.004	0	0	0
3	0	0.032	0.848	0.028	0.023	0	0.003
4	0	0.032	0	0.792	0	0	0
5	0	0	0.146	0	0.973	0	0.024
6	0	0	0	0.176	0	0.997	0.004
7	0	0	0.006	0	0.004	0.003	0.969
	1 Mature						
	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0

2	0	0.981	0.02	0.02	0	0	0
3	0	0.01	0.901	0.02	0.055	0.012	0.053
4	0	0.009	0.023	0.88	0	0.063	0
5	0	0	0.023	0	0.945	0	0.005
6	0	0	0.034	0.08	0	0.924	0.507
7	0	0	0	0	0	0.001	0.435
	2 Immature						
	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	0.893	0.018	0.017	0	0.015
4	0	0	0	0.824	0	0	0
5	0	0	0.095	0	0.981	0	0.023
6	0	0	0	0.158	0	0.991	0.017
7	0	0	0.011	0	0.001	0.009	0.946
	2 Mature						
	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0
2	0	0.995	0.018	0.075	0	0	0
3	0	0.005	0.978	0.059	0.057	0.004	0
4	0	0	0.002	0.862	0	0.018	0
5	0	0	0.001	0	0.943	0	0
6	0	0	0.001	0.004	0	0.978	0
7	0	0	0	0	0	0	1
	3 Immature						
	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	0.902	0.113	0.011	0	0.021
4	0	0	0	0.856	0	0	0
5	0	0	0.085	0	0.988	0	0.02
6	0	0	0	0.031	0	0.984	0.064
7	0	0	0.012	0	0	0.016	0.894
	3 Mature						
	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0
2	0	1	0	0.085	0	0	0
3	0	0	0.997	0	0.043	0	0
4	0	0	0	0.898	0	0.003	0
5	0	0	0.003	0	0.957	0	0
6	0	0	0	0.018	0	0.997	0
7	0	0	0	0	0	0	1
	4 Immature						
	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	0.922	0.1	0.009	0	0.014
4	0	0	0	0.9	0	0	0
5	0	0	0.07	0	0.991	0	0.012
6	0	0	0	0	0	0.986	0.064
7	0	0	0.008	0	0	0.014	0.91
	4 Mature						
	1	2	3	4	5	6	7

1	1	0	0	0	0	0	0
2	0	1	0	0.077	0	0	0
3	0	0	0.997	0	0.046	0	0
4	0	0	0	0.913	0	0.001	0
5	0	0	0.003	0	0.954	0	0
6	0	0	0	0.011	0	0.999	0
7	0	0	0	0	0	0	1
	5 Immature						
	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	0.925	0	0.006	0	0.005
4	0	0	0	1	0	0	0
5	0	0	0.07	0	0.993	0	0.004
6	0	0	0	0	0	0.991	0.039
7	0	0	0.005	0	0.001	0.009	0.953
	5 Mature						
	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0
2	0	1	0	0.001	0	0	0
3	0	0	1	0	0	0	0
4	0	0	0	0.999	0	0	0
5	0	0	0	0	1	0	0
6	0	0	0	0	0	1	0
7	0	0	0	0	0	0	1
	6 Immature						
	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	0.93	0.086	0.003	0	0.002
4	0	0	0	0.847	0	0	0
5	0	0	0.065	0	0.991	0	0.013
6	0	0	0	0.067	0	0.989	0.027
7	0	0	0.005	0	0.005	0.011	0.957
	6 Mature						
	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	1	0	0	0	0
4	0	0	0	1	0	0	0
5	0	0	0	0	1	0	0
6	0	0	0	0	0	1	0
7	0	0	0	0	0	0	1
	7 Immature						
	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	0.937	0.039	0.002	0	0.001
4	0	0	0	0.896	0	0	0
5	0	0	0.058	0	0.983	0	0.017
6	0	0	0	0.065	0	0.986	0.015
7	0	0	0.006	0	0.015	0.014	0.968
	7 Mature						

	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	1	0	0	0	0
4	0	0	0	1	0	0	0
5	0	0	0	0	1	0	0
6	0	0	0	0	0	1	0
7	0	0	0	0	0	0	1
	8 Immature						
	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	0.935	0.032	0.002	0	0
4	0	0	0	0.968	0	0	0
5	0	0	0.059	0	0.976	0	0.011
6	0	0	0	0	0	0.981	0.011
7	0	0	0.006	0	0.022	0.019	0.978
	8 Mature						
	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	1	0	0	0	0
4	0	0	0	1	0	0	0
5	0	0	0	0	1	0	0
6	0	0	0	0	0	1	0
7	0	0	0	0	0	0	1

Table A3. Distribution tables (in %) of capelin at the start of the month during 1990-1991 according to simulations with the Kolmogorov model. Sub-area 1 does not enter into any of the simulations.

Sub	October 1		November 1		December 1		January 1		February 1	
area	imm.	mat.	imm.	mat.	imm.	mat.	imm.	mat.	imm.	mat.
2	0	0	0	0	0	0	0	0	0	0
3	3.5	0	0.8	0	0.2	0	0.3	0.3	0.6	45.1
4	0.2	0	0.1	0	0	0	0	0	0	11.3
5	16.0	1.2	7.4	0.1	6.0	0.2	11.9	4.7	18.3	4.6
6	18.9	60.9	21.7	39.1	25.8	41.8	32.1	47.0	33.3	34.9
7	61.5	37.9	70.0	60.8	67.9	58.0	55.6	48.0	47.7	4.0
8	0	0	0	0	0	0	0	0	0	0.1

Sub	March 1		April 1		May 1		June 1		July 1	
	imm.	mat.	imm.	mat.	imm.	mat.	imm.	imm	imm	imm
2	0	17.9	0	23.9	0	20.8	0	0	0	0
3	1.3	68.7	1.8	71.3	2.2	74.6	2.5	1.1	0.6	0.3
4	0	10.9	0	4.5	0	4.5	0	0	0	0
5	24.2	0.5	28.3	0.1	31.2	0.1	33.3	10.5	4.2	2.7
6	34.0	1.9	34.0	0.2	33.2	0	32.4	19.3	15.1	13.9
7	40.4	0.2	35.8	0	33.3	0	31.7	69.1	80.0	83.1
8	0.1	0	0.1	0	0	0	0	0	0	0

Table A4. Calculated monthly migration matrices (in %) for mature capelin in the Barents sea for the period October 1 1990 till June 1 1991, according to the Kolmogorov model, using initial distribution based on preferred temperature. The entry in row *i* and column *j* shows percentage moving from compartment *i* to compartment *j*. NOTE that these tables atre the transpose of corresponding tables in Table A2.

October							
	2	**3**	**4**	**5**	**6**	**7**	**8**
2	9.6	9.3	51.5	1.2	17.9	1.1	9.4
3	0.1	27.3	1.3	62.7	1.1	7.7	0.0
4	2.4	10.8	32.4	3.3	39.3	5.8	6.1
5	0.0	2.8	0.0	68.7	0.1	28.4	0.0
6	0.0	0.4	1.5	0.2	73.2	23.5	1.3
7	0.0	0.5	0.1	1.0	14.2	84.2	0.0
8	0.6	0.1	14.0	0.0	61.6	0.7	23.1

November							
	2	**3**	**4**	**5**	**6**	**7**	**8**
2	10.6	10.7	51.7	1.4	17.7	1.1	6.8
3	0.2	40.9	3.2	45.0	2.6	8.1	0.0
4	2.6	11.6	31.4	3.9	40.7	5.0	4.9
5	0.0	2.2	0.0	67.3	0.2	30.4	0.0
6	0.0	0.3	0.5	0.2	72.0	26.6	0.5
7	0.0	0.1	0.0	3.7	6.6	89.6	0.0
8	0.6	0.1	13.5	0.0	62.0	0.3	23.6

December							
	2	**3**	**4**	**5**	**6**	**7**	**8**
2	15.4	14.2	48.4	1.5	16.4	1.6	2.6
3	0.1	50.4	1.8	41.4	1.8	4.6	0.0
4	3.8	14.1	30.0	3.7	40.0	5.6	3.0
5	0.0	7.8	0.0	80.6	0.1	11.5	0.0
6	0.0	0.5	0.8	0.3	81.1	16.8	0.7
7	0.0	0.9	0.0	20.4	9.3	69.4	0.0
8	1.0	0.1	13.3	0.0	62.5	0.2	22.9

January							
	2	**3**	**4**	**5**	**6**	**7**	**8**
2	33.5	0.1	10.9	0.0	0.1	0.0	3.0
3	19.6	92.8	23.9	57.0	9.4	31.1	0.4
4	40.9	2.4	39.4	1.8	14.9	8.0	17.3
5	0.1	4.0	0.2	39.3	0.2	19.3	0.0
6	5.4	0.5	23.4	0.3	69.6	16.6	55.8
7	0.1	0.2	0.8	1.7	5.1	24.9	0.1
8	0.4	0.0	1.4	0.0	0.7	0.0	23.3

February							
	2	**3**	**4**	**5**	**6**	**7**	**8**
2	62.1	13.0	24.6	0.0	0.1	0.0	0.2
3	4.2	83.1	9.9	2.3	0.3	0.2	0.0
4	37.6	19.9	38.4	0.0	3.3	0.1	0.7

5	0.5	56.2	2.5	40.0	0.1	0.7	0.0
6	8.4	10.4	39.7	0.1	38.1	2.1	1.2
7	3.1	27.2	18.1	10.7	13.3	27.5	0.0
8	8.2	0.4	26.8	0.0	39.4	0.0	25.3

March							
	2	**3**	**4**	**5**	**6**	**7**	**8**
2	62.6	13.1	24.2	0.0	0.1	0.0	0.1
3	7.1	78.6	12.8	1.2	0.2	0.1	0.0
4	39.4	27.5	32.4	0.0	0.5	0.0	0.1
5	0.9	66.2	3.5	28.8	0.1	0.6	0.0
6	14.9	10.4	40.0	0.1	31.8	1.3	1.6
7	3.1	22.5	12.6	14.3	14.1	33.5	0.0
8	23.3	0.7	36.3	0.0	24.2	0.1	15.5

April							
	2	**3**	**4**	**5**	**6**	**7**	**8**
2	43.9	20.9	35.1	0.0	0.1	0.0	0.1
3	4.1	80.1	14.1	1.4	0.2	0.1	0.0
4	27.5	29.7	42.1	0.0	0.6	0.0	0.1
5	0.6	68.0	4.2	26.5	0.1	0.6	0.0
6	10.2	10.6	43.5	0.1	32.6	1.4	1.7
7	2.0	23.3	14.5	13.9	12.6	33.7	0.0
8	20.0	0.7	37.8	0.0	24.7	0.0	16.7

May							
	2	**3**	**4**	**5**	**6**	**7**	**8**
2	48.0	11.8	39.9	0.0	0.1	0.0	0.1
3	2.3	81.1	14.7	1.6	0.2	0.1	0.0
4	20.3	29.0	50.0	0.1	0.6	0.0	0.1
5	0.4	69.9	4.9	24.1	0.1	0.6	0.0
6	8.7	10.3	46.9	0.1	31.0	1.4	1.6
7	1.7	24.0	17.8	12.8	11.6	32.1	0.0
8	18.7	0.7	40.1	0.0	24.1	0.1	16.4

7.7 Estimates of cod's consumption of capelin and herring related to prey density, stock overlap and physical factors.

Geir Odd Johansen and Øyvind Ulltang
University of Bergen, Norway.

Abstract: Consumption of different age groups of juvenile Norwegian spring–spawning herring (*Clupea harengus* L.) and different size groups of capelin capelin (*Mallotus villosus* Müller 1776) by northeast Arctic cod (*Gadus morhua* L.) in the Barents Sea in 1992–1997 is estimated using cod stomach content data.

Estimation of digestion time

A new approach to the problem of estimating consumption by fish has been developed and tested in this project. In this approach, consumption is estimated for individual fish using stomach content data and the local environmental temperature as input to a stomach evacuation model. The new method is based on estimation of digestion time for single prey items in the early stage of the digestion process using the difference between fresh weight at ingestion and weight in the stomach at time of sampling. Estimation of the digestion times is done within a gastric evacuation model for cod using area specific sea temperatures. This estimation routine returns the time (t_{max}) it takes for a prey to become digested to a stage where length is no longer measurable. Predation rate is then estimated for all prey with digestion time = t_{max} as number of prey eaten within the time range defined by t_{max}. Details about the new method and the data used are published as scientific paper (Johansen et al. 2004). Parameter estimates for digestion of juvenile herring and capelin are given in Johansen et al. (2004) and Addendum I, respectively.

Estimation of total consumption

The predation rate, which is a *per capita* consumption rate for cod with the prey of current interest in their stomach, is combined with estimates of the proportion of the cod stock consuming that prey and area specific abundance of cod. The proportion is estimated from the stomach data. Cod abundance is estimated by combining swept area estimates from trawl surveys with VPA output. This gives estimates of total consumption of juvenile herring and capelin on a seasonal and yearly basis. Some extrapolation of data in time and space must be done to cover the whole year and area. Details about the extrapolation routines for juvenile herring and capelin are given in Johansen et al. (2004) and Addendum I, respectively. Consumption is given as number of prey of a given age (juvenile herring) or size (capelin) consumed over the period. These estimates are converted to biomass by using age–weight relationships (herring) and length–weight relationships (capelin).

Results and discussion

The consumption estimates and related results are presented in Johansen et al. (2004) and Addendum I. The following will summarise those results which are directly connected to deliverables in the dst^2 project, and present some new results and perspectives on the relationship between consumption and prey density, stock overlap and physical factors.

Total consumption of herring and capelin

Consumption of juvenile herring and capelin by cod in the Barents Sea is presented in tables 1 and 2. The consumption estimates are given as number of prey and biomass consumed. Herring consumption is given for different age groups, while capelin consumption is given by size group. Herring consumption estimated by the new method is compared to herring consumption estimated by other methods, showing difference between estimates from different methods (Johansen et al. 2004). The method developed in the current project is flexible with respect to temporal and spatial scale of a study. Assumptions about initial meal size and undetermined stomach content are avoided, allowing description of large variation in feeding characteristics of a predator. The method also seems to be robust with regard to random variability in observations of prey in single stomachs. The new method is promising as a tool for studying ecological interactions in piscivorous fish.

Table 1. Consumption of juvenile herring in the Barents Sea by northeast Arctic cod in the period 1992–1997. Interval 1–5 refers to five sample intervals of the year: 1 January–31 March, 1 April–31 May, 1 June–9 July, 10 July–15 October, and 16 October–31 December, respectively. The intervals are defined based on the temporal distribution of hauls with stomach samples. The estimates are in billion individuals and 1000 metric tons.

		0 group		1 group		2 group		3+ group			
Year	Interval	Number	Biomass	Number	Biomass	Number	Biomass	Number	Biomass	Total N	Total B
1992	1	-	-	0.539	3.304	1.087	41.816	0.245	26.609	1.871	71.729
	2	-	-	0.332	3.177	0.680	42.699	0.149	15.458	1.161	61.334
	3	-	-	0.113	1.130	0.433	20.316	0.077	8.164	0.623	29.610
	4	2.449	16.045	0.561	11.349	0.139	10.888	0.013	1.735	3.162	40.017
	5	5.876	40.432	0.426	14.919	0.159	12.387	0.003	0.593	6.464	68.331
	Total	8.325	56.477	1.971	33.879	2.498	128.106	0.487	52.559	13.281	271.021
1993	1	-	-	1.756	11.159	0.434	9.727	0.320	28.449	2.510	49.335
	2	-	-	1.119	8.683	0.432	11.532	0.202	24.127	1.753	44.342
	3	-	-	0.632	5.491	0.132	3.255	0.117	7.599	0.881	16.345
	4	1.003	4.878	0.733	15.188	0.000	0.000	0.000	0.000	1.736	20.066
	5	3.568	16.870	0.209	5.152	0.000	0.000	0.000	0.000	3.777	22.022
	Total	4.571	21.748	4.449	45.673	0.998	24.514	0.639	60.175	10.657	152.110
1994	1	-	-	12.379	66.694	4.734	87.963	0.863	40.783	17.976	195.440
	2	-	-	2.031	16.917	2.630	66.801	0.557	51.456	5.218	135.174
	3	-	-	1.743	14.517	0.135	3.642	0.005	0.289	1.883	18.448
	4	17.594	93.398	2.073	40.686	0.225	10.150	0.000	0.000	19.892	144.234
	5	0.027	0.212	0.000	0.000	0.013	0.665	0.000	0.000	0.040	0.877
	Total	17.621	93.610	18.226	138.814	7.737	169.221	1.425	92.528	45.009	494.173
1995	1	-	-	0.067	0.558	0.377	5.717	0.148	7.598	0.592	13.873
	2	-	-	0.184	1.760	1.162	25.955	0.526	28.387	1.872	56.102
	3	-	-	0.012	0.117	0.142	3.607	0.046	2.376	0.200	6.100
	4	0.086	0.457	0.131	2.575	0.226	9.524	0.008	0.937	0.451	13.493
	5	3.149	16.716	0.388	7.768	0.440	21.835	0.023	2.499	4.000	48.818
	Total	3.235	17.173	0.782	12.778	2.347	66.638	0.751	41.797	7.115	138.386
1996	1	-	-	0.002	0.017	0.107	2.843	0.086	4.242	0.195	7.102
	2	-	-	0.000	0.000	0.083	2.317	0.064	5.003	0.147	7.320
	3	-	-	0.000	0.000	0.031	0.987	0.010	0.716	0.041	1.703
	4	0.000	0.000	0.000	0.000	0.061	3.025	0.000	0.000	0.061	3.025
	5	1.425	5.391	0.000	0.000	0.000	0.000	0.000	0.000	1.425	5.391
	Total	1.425	5.391	0.002	0.017	0.282	9.172	0.160	9.961	1.869	24.541
1997	1	-	-	0.047	0.392	0.007	0.181	0.052	3.873	0.106	4.446
	2	-	-	0.011	0.091	0.000	0.000	0.004	0.285	0.015	0.376
	3	-	-	0.007	0.064	0.000	0.000	0.004	0.322	0.011	0.386
	4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	5	0.631	3.083	0.180	4.470	0.000	0.000	0.000	0.000	0.811	7.553
	Total	0.631	3.083	0.245	5.017	0.007	0.181	0.060	4.480	0.943	12.761

Table 2. Consumption of capelin by Northeast Arctic cod in the Barents Sea in the period 1992–1997. Sample intervals are equal to the intervals in table 1. The estimates are in billion individuals and 1000 metric tons.

Capelin sizegroups		< 10 cm		10 - 15 cm		>= 15 cm			
Year	Interval	Number	Biomass	Number	Biomass	Number	Biomass	Total N	Total B
1992	1	1.116	6.699	44.045	490.226	6.111	133.463	51.272	630.388
	2	0.512	3.071	16.220	188.316	3.344	65.280	20.076	256.667
	3	0.377	1.926	1.168	14.676	0.016	0.338	1.561	16.940
	4	48.474	247.703	65.314	651.185	4.208	90.170	117.996	989.058
	5	17.919	91.564	36.686	270.376	1.440	24.674	56.045	386.614
	Total	68.398	350.963	163.433	1614.779	15.119	313.925	246.950	2279.667
1993	1	6.834	41.003	79.321	882.843	31.125	679.773	117.280	1603.619
	2	4.269	25.615	47.992	557.191	16.516	322.384	68.777	905.190
	3	0.010	0.049	1.146	14.406	0.008	0.177	1.164	14.632
	4	4.225	21.592	35.677	355.697	0.986	21.128	40.888	398.417
	5	0.061	0.310	5.504	40.564	0.049	0.841	5.614	41.715
	Total	15.399	88.569	169.640	1850.701	48.684	1024.303	233.723	2963.573
1994	1	2.530	15.182	34.864	388.032	4.750	103.739	42.144	506.953
	2	1.302	7.811	28.541	331.364	2.464	48.098	32.307	387.273
	3	0.229	1.171	1.001	12.577	0.027	0.555	1.257	14.303
	4	2.868	14.657	8.341	83.157	0.196	4.192	11.405	102.006
	5	21.451	109.616	4.107	30.266	0.070	1.192	25.628	141.074
	Total	28.380	148.437	76.854	845.396	7.507	157.776	112.741	1151.609
1995	1	8.154	48.925	35.434	394.384	2.207	48.203	45.795	491.512
	2	0.110	0.661	24.510	284.562	0.765	14.931	25.385	300.154
	3	0.015	0.078	0.437	5.493	0.029	0.616	0.481	6.187
	4	0.265	1.352	5.294	52.780	0.049	1.040	5.608	55.172
	5	0.000	0.000	0.977	7.197	0.050	0.853	1.027	8.050
	Total	8.544	51.016	66.652	744.416	3.100	65.643	78.296	861.075
1996	1	1.690	10.139	8.390	93.380	2.482	54.212	12.562	157.731
	2	0.050	0.297	3.357	38.973	1.325	25.872	4.732	65.142
	3	0.000	0.000	0.173	2.180	0.118	2.468	0.291	4.648
	4	6.653	33.997	10.761	107.288	0.928	19.890	18.342	161.175
	5	2.224	11.364	5.563	41.001	0.948	16.244	8.735	68.609
	Total	10.617	55.797	28.244	282.822	5.801	118.686	44.662	457.305
1997	1	1.056	6.338	6.365	70.845	3.496	76.348	10.917	153.531
	2	0.379	2.274	1.605	18.639	1.367	26.693	3.351	47.606
	3	0.009	0.044	0.037	0.469	0.000	0.000	0.046	0.513
	4	6.963	35.583	10.364	103.327	0.853	18.278	18.180	157.188
	5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Total	8.407	44.239	18.371	193.280	5.716	121.319	32.494	358.838

Ecological perspectives

As consumption rate estimated by the new method is given as number of prey consumed per time unit, independent estimates of natural mortality can be calculated directly, in order to assess the effect of a predator if prey stock size is known. In this project, predation mortality of 1 and 2 years old juvenile herring due to predation by cod is estimated directly from the consumption estimates by combining them with acoustic herring abundance data (Johansen et al. 2004). In summary, the natural annual instantaneous mortality rate of these age groups due to cod predation was 0.05–0.18 in the period 1991–1994, and varied between both years and year–classes. It was found to be about 1/3 of the total mortality

of the 1991 year–class and 1/5 of that of the 1992 year–class. Juvenile herring in the Barents Sea seems to be vulnerable to significant natural mortality due to predation from the cod stock in the area. The variability of this mortality may explain some of the variation in recruitment to the adult stock. The ICES Northern Pelagic and Blue Whiting Fisheries Working Group uses a natural mortality of 0.9 for age 0–2 herring in the Barents Sea. A revision of the value of natural mortality of age 0–2 herring is due, both because of variable abundance of predators and of alternative prey (capelin) and because the variability in year–class strength is so large that a constant natural mortality seems unlikely.
The effect of prey density on consumption by cod is studied by analysing the mean daily predation rate (number of prey eaten per cod per day) for cod that has consumed the prey of interest, as well as the proportion these cod constitute of the sample. Predation rate and proportion are averaged over each half of the year, areas, size groups of cod and size/age groups of the prey. The level of consumption averaged over the whole cod stock in the strata studied above can be found by taking the product of the predation rate and the proportion. The areas used are sampling strata for standard bottom trawl surveys targeting demersal fish in the Barents Sea conducted by the Institute of Marine Research, Bergen, Norway (Fig 1). The analyses are done for selected strata constituting the core area for the prey species and where the coverage of stomach samples is extensive throughout the time series. For capelin, strata 2, 6, 7 and 12 are included in the analyses for first half of the year, while strata 1–12 are included for the second half. For herring, strata 6–8 and 12 are used in the first half, while strata 15–17 and 160 are used for the second half. The strata in first and second half of the year refer to the winter (intervals 1–2) and summer (intervals 3–5) strata systems, respectively (Fig. 1). Cod size groups with <20 stomach samples within a stratum are not included in the analysis, to increase the precision of the proportion estimates.

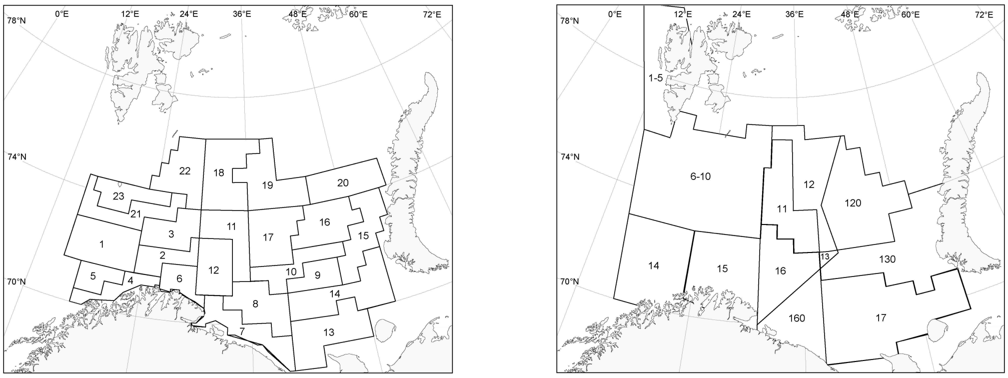


Figure 1. Maps of the sampling strata used in standard bottom trawl surveys targeting demersal fish in the Barents Sea conducted by the Institute of Marine Research, Bergen, Norway. Stratum numbers are given. Left and right panels show the winter strata system and summer strata system, respectively.

Stock size of capelin and juvenile herring are acoustic estimates from standard surveys. The capelin estimates are from surveys in September–October. In the first half of the year, estimates from previous year are used, while estimates from the current year are used in the second half. This is done to correct for spawning mortality in spring. The juvenile herring estimates are from surveys in May–June. In the first half of the year, estimates from the current year are used, while estimates from the next year are used in the second half. This is done to correct for the recruitment of 0–group in late summer.

For capelin, the variation in both the predation rate and the proportion reflects the variation in stock size (Fig. 2 and 3). However, the mean predation rate seems to be less related to the prey stock sizes than proportion. This may reflect variations in turnover rate with temperature, and the stomach capacity of cod. Environmental temperature will affect the rate of metabolism in the cod, and the mean predation rate is likely to vary with this factor. The stomach capacity of cod will put an upper limit to the number of prey the predators are able to ingest. Therefore, the relationship between predation rate and prey stock size should stabilize at an asymptotic level at high prey stock levels. The relationship between stock size and consumption is weak in juvenile herring (Fig. 2 and 3).

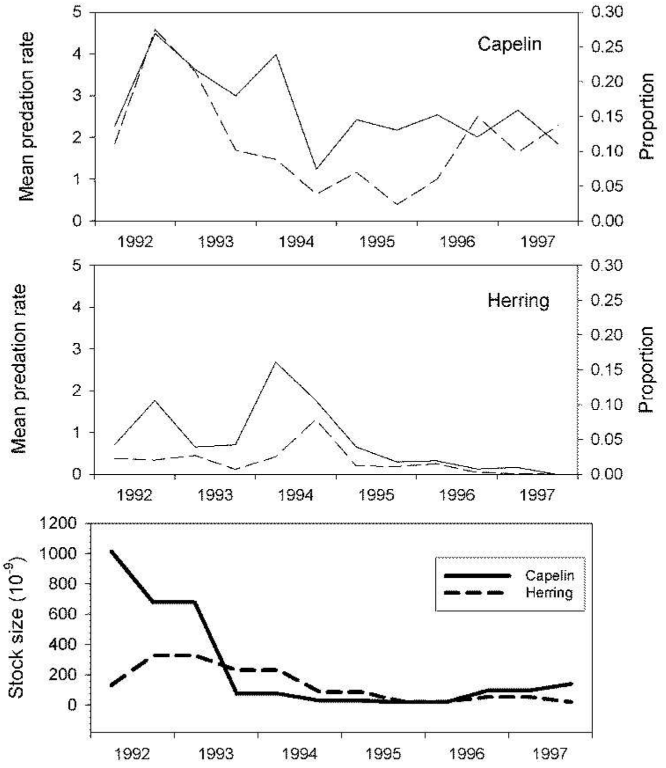


Figure 2. Mean daily predation rate (solid line) and proportion of cod stock (dashed line) eating capelin (a) and juvenile herring (b) in the Barents Sea in 1992–1997. Stock sizes of capelin and juvenile herring in the same period (c). See text for details about the calculations.

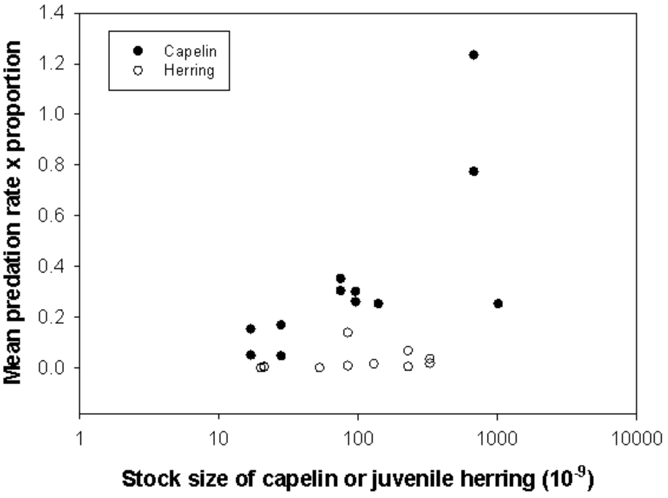


Figure 3. Relationship between average consumption of capelin and juvenile herring by cod in selected areas of the Barents Sea, and abundance of the two prey species in 1992–1997. Average consumption is presented as the product of mean daily predation rate and proportion of the cod stock eating the prey species. Note logarithmic x–axis. See text for details.

The level of consumption of capelin averaged over the whole cod stock in the strata studied above is generally higher than consumption of juvenile herring, particularly in years of high capelin stock (Fig. 4). When these results are interpreted, two important aspects must be considered. First, juvenile herring in the Barents Sea attains larger sizes than capelin. The cod will therefore be able to ingest fewer individuals of juvenile herring than capelin on average, influencing the consumption estimated as number of prey consumed. This effect will vary with the abundance of different year classes of the two prey species. Second, there are indications that cod must reach a certain size before it can handle herring as effectively as capelin (Johansen 2003). The decline of the mean cod size in the Barents Sea the last decades (Nilssen et al. 1994) may have shifted the predator pressure away from juvenile herring. As the consumption estimates presented here are averaged over all cod sizes, high consumption of capelin by abundant small cod sizes may dominate the results. A size–structured analysis should be done to evaluate such an effect. The increased consumption of herring in 1994, when the capelin stock was very low, indicates that a predator–prey interaction cannot be viewed in isolation. Such an interaction is likely to be influenced by other species than the two directly involved. The results are interesting with respect to the coupled dynamics of cod, capelin and juvenile herring in the Barents Sea.

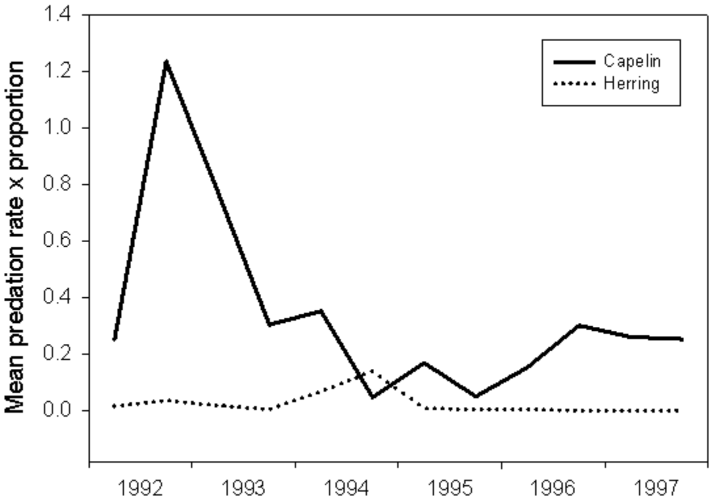


Figure 4. Average consumption of capelin and juvenile herring by cod in selected areas of the Barents Sea in 1992–1997, presented as the product of mean daily predation rate and proportion eating the prey species. See text for details.

The relationship between consumption of different prey by cod and the stock overlap between predator and prey is difficult to assess with the data used in this study. A fruitful study of this should be done with abundance data of predator and prey on a finer temporal and spatial scale than available here. In such a study both the overlap between predator and prey, and the overlap between prey species should be incorporated. The first will determine the availability of prey to the predator, while the overlap between prey species could influence the relative consumption of different prey when they simultaneously overlap with the predator. Situations with simultaneously overlap between the predator and several prey species can shed light on selection between prey species by the predator. The method for estimating consumption developed in this project takes account of spatial structuring by offering a flexible partitioning of the area where predation takes place. The method therefore accounts for variation in spatial overlap between predator prey on large scale, and is suitable for studying this on a smaller scale.

The relationship between consumption of different prey by cod and climatic variation requires longer time series than available in this study. The method for estimating consumption incorporates temperature data on small scale, and effects of climatic variation altering the spatial distribution of temperature fields could be studied with this method. As the environmental temperature affects the rate of metabolism in fish, it is likely that climatic variation can result in variation in total consumption by a predator stock. This effect will depend on the availability of suitable prey.

Conclusion

The estimation method uses only prey in the early stage of the digestion process. The main advantages of this is to circumvent the problem of estimating initial meal size, avoid the use of unidentified stomach content when estimating stomach evacuation and increase the precision when estimating digestion time. Consumption rate is given as number of prey consumed per time unit. This can be directly converted to predation mortality, and is useful for studying behavioural mechanisms governing the predation process.

Cod in the Barents Sea consume on average more individuals of capelin than juvenile herring in the Barents Sea. The generally smaller size of capelin compared to herring and the size–dependent predation success of cod towards herring must be considered when judging the relative importance of these two prey species as food for cod in the Barents Sea.
The study of cod's consumption of capelin and juvenile herring in the Barents Sea presented here emphasises the importance of considering the availability of alternative prey when studying the predator–prey interaction between a predator and one of its prey species.

References

Johansen, G.O. 2003. Size–dependent predation on juvenile herring (*Clupea harengus* L.) by Northeast Arctic cod (*Gadus morhua* L.) in the Barents Sea. Sarsia **88**: 136–153.

Johansen, G. O., B. Bogstad, S. Mehl, and Ø. Ulltang. 2004. Consumption of juvenile herring (*Clupea harengus*) by cod (*Gadus morhua*) in the Barents Sea: a new approach to estimating consumption in piscivorous fish. Canadian Journal of Fisheries and Aquatic Sciences **61**:343-359.

Nilssen, E.M., Pedersen, T., Hopkins, C.C.E., and Pope, J.G. 1994. Recruitment variability and growth of Northeast Arctic cod: influence of physical environment, demography, and predator–prey energetics. ICES Marine Science Symposia **198**: 449–470.

Addendum I. Consumption of capelin (*Mallotus villosus* Müller 1776) by Northeast Arctic cod (*Gadus morhua* L.) in the Barents Sea

Geir Odd Johansen, Bjarte Bogstad, Sigbjørn Mehl, Øyvind Ulltang, Harald Gjøsæter

Johansen, G.O.; Bogstad, B.; Mehl, S.; Ulltang, Ø.; Gjøsæter H. 2003. Consumption of capelin (*Mallotus villosus* Müller 1776) by Northeast Arctic cod (*Gadus morhua* L.) in the Barents Sea.

> **Abstract:** Consumption of different size groups of capelin by Northeast Arctic cod in the Barents Sea in 1992–1997 was estimated using cod stomach data. Number of capelin consumed per cod per time unit was estimated by combining area specific sea temperature and a model for stomach evacuation in cod. Total consumption by the cod stock was also estimated on a seasonal and yearly basis, by combining the consumption estimates with area specific estimates of abundance of cod.

Key words: cod, capelin, predation, consumption, stomach samples, Barents Sea

Geir Odd Johansen / Øyvind Ulltang: University of Bergen. Department of Fisheries and Marine Biology, PO Box 7800, N–5020 Bergen, Norway. Bjarte Bogstad / Sigbjørn Meh / Harald Gjøsæter: Institute of Marine Research, PO Box 1870 Nordnes, N–5817 Bergen, Norway.
Correspondence to Geir Odd Johansen: e–mail: geir.johansen@ifm.uib.no

Introduction

In this study, consumption of capelin by North–east Arctic cod in the Barents Sea is estimated from cod stomach data. Consumption is estimated for individual fish using the local environmental temperature as input for a stomach evacuation model describing the digestion of individual prey items. These estimates are then combined with area specific estimates of cod abundance to calculate total consumption for the sampled area

Materials and methods

Stomach content data from cod were taken from the joint IMR–PINRO stomach database at the Institute of Marine Research (IMR) in Bergen, Norway. This database includes stomachs sampled at both Norwegian and Russian regular demersal fish surveys in 1984–1997. Most of the cod were caught by bottom trawl, mainly in the first quarter of the year and in September/October. A detailed description of the general survey methodology can be found in Jakobsen et al. (1997) and Lepesevich & Shevelev (1997). Some of the data are from surveys of pelagic fish and shrimp. The sampling design has undergone modifications during the period of interest to this study. The maximum number of stomach samples per 5 cm group of cod per haul has changed from two in 1992–95 to one after 1995 (Bogstad *et al.*, 1995; Jakobsen *et al.*, 1997). Details about the sampling procedures are given in Mehl (1989) and Mehl & Yaragina (1992).

In this study, the analyses were restricted to the period 1992–97. In the period 1992–93 there was a gradual change in the methods for recording the prey size, from 5 cm to 1 cm groups. The material sampled before 1992 was therefore considered too imprecise for this study.

The data were divided into five sample intervals of the year, motivated by the temporal distribution of hauls. The sample intervals were defined as: 1 January–31 March, 1 April–31 May, 1 Juny–9 July, 10 July–15 October, and 16 October–31 December, referred to as interval 1–5, respectively. Most of the surveys providing the data were not targeted for stomach sampling, so the temporal and spatial distribution of hauls varied between years and intervals. A detailed description of the temporal structuring, and the temporal and spatial distribution of the hauls representing the raw data is given by Johansen (2002). Cod body length was measured to the nearest cm below, and divided into 10 cm groups in the analyses.

Cod = 90 cm were aggregated in the largest group. Cod smaller than 20 cm were excluded from the data, because of their low propensity to eat fish (Johansen, 2002). Table 1 summarises the data.

Consumption of capelin by individual cod was estimated as number of capelin consumed per time unit, referred to as predation rate. This limited the analysis to prey that was counted. The estimates were based on estimating the digestion times for the consumed capelin, i.e. the time the prey has been digested in the stomach of the predator. Digestion times were estimated by using a gastric evacuation model (GEM), which describes the reduction in weight of a prey in a predator stomach due to digestion as a function of time (Bromley, 1994). A similar approach is described in Mergardt & Temming (1997) for estimating the diel pattern of food intake in whiting (*Merlangius merlangus* L.). Digestion times for capelin were estimated by relating the weight of the partly digested prey (W_p) in the cod stomach to the weight of that prey when it was ingested (W_f). The weight of ingested prey was estimated from a year specific length–weight relationship for capelin estimated from survey data at IMR. This limited the analysis to prey with measured lengths.

This resulted in 5496 capelin observations from 2657 stomachs. A capelin observation is a record of capelin of the same length group and same digestion stage within a cod stomach, and may include several individual capelin.

Gastric evacuation model (GEM)

The GEM used in this study is based on the general GEM:

$$\frac{dS}{dt} = -R \times^{B} \qquad (1)$$

(Jones, 1974), with variables *S* = the stomach content weight (g), R = standard evacuation rate, B = constant defining degree of curvilinearity. Integrating (2) gives:

$$S_t = [S_{0^{(1-B)}} - R(1-B)t]^{\frac{1}{1-B}} \qquad (2)$$

for $B \neq 1$. In this model S_t = stomach content at time *t* (g), *t* = time after ingestion and S_0 = initial stomach content (g). *R* incorporates the effects of temperature, food type, predator size and other factors (Jones, 1974).

Following Temming & Andersen (1994), (2) was expanded to a multivariable model describing the effects of temperature, predator size and meal size:

$$S_t = \left[M^{1-B} - R'e^{AT}W^{C}M^{D}(1-B)t\right]^{\frac{1}{1-B}}, \qquad (3)$$

with additional variables *M* = meal size (g) (substitute for S_0), *T* = temperature (°C), *W* = predator weight (g), and parameters *A* = temperature coefficient, *C*= predator weight coefficient, *D* = meal size coefficient and *R*’ = food type constant. According to Temming & Andersen (1994), the effect of including the meal size correction on *R*’ is negligible when fitting the GEM to experimental data. They recommend the use of a simpler model without this term:

$$S_t = \left[M^{1-B} - R'e^{AT}W^{C}(1-B)t\right]^{\frac{1}{1-B}} \qquad (4)$$

Back-calculation of digestion times for individual cod.

To calculate digestion times of capelin, equation (4) was rearranged to:

$$t = \frac{(S_t^{(1-B)} - M^{(1-B)})}{-R' e^{AT} W^C (1-B)}, \quad (5)$$

where t is an estimate of digestion time. Substituting W_p (g) for S_t and W_f (g) for M enables the estimation of the digestion time of a capelin of weight W_p and length L (cm) in the stomach. Parameters B = 0.50, R' = 0.00749, A = 0.11, and C = 0.305 are estimates from gastric evacuation experiments reported in Temming & Herrmann (2003). Temperature was taken from a digital temperature atlas for the Barents Sea based on IMR's hydrographic measurements. The temperature was averaged for 50 m and below at each station, and then averaged horizontally within each area.

Consumption estimates

Predation rate for cod containing measurable capelin of a given 1 cm size group was estimated as the number of prey in the stomach with an estimated digestion time $\leq t_{max}$, divided by t_{max}. The analyses were restricted to capelin observations less than 6 individual prey. The size groups < 7 cm and >18 cm were aggregated to get sufficient sample sizes. t_{max} was defined as the upper limit of the time range within which all capelin were measurable, and was estimated as follows. For each size group of capelin, digestion time was obtained from (5) using the environmental temperature and setting cod weight equal to the median weight of cod that had measurable capelin of the given size group in the stomach for data from 1992–97 (Table 2). t_{max} was found by plotting the cumulative frequency of measurable capelin as a function of digestion time grouped within 1–hour intervals. Assuming a uniform distribution of cod stomachs at different levels of digestion for the whole material, there is an expected linear relationship between these variables, until the digestion time reaches a level where capelin starts to become immeasurable due to digestion. This breakpoint in the relationship defines t_{max}. The breakpoint was found by visual inspection of the plot and the residuals of a linear regression of the relationship. Assuming that the capelin becomes immeasurable when a certain proportion of its weight is digested, and setting predator weight constant, rearranging (5) gives a log–linear relationship between capelin body length ($TL_{capelin}$) and t_{max}. Linear regression of this relationship at the reference temperature 4°C ($T4$) ($\ln(t_{\max}) = -2.0204 + 1.7657 \ln(TL_{capelin})$, df = 17, r^2 = 0.92, $p < 0.0001$), was used to estimate t_{max} for different 1 cm size groups of capelin (Table 2). t_{max} at other temperatures for the same size groups were estimated by the relationship $t_{\max,T} = \frac{\exp(A \cdot T4)}{\exp(A \cdot T)} \cdot t_{\max,T4}$.

Mean predation rate for cod containing measurable capelin was estimated for each size group capelin a and each 10 cm length group of cod, L as:

$$\mu_{a,L} = \frac{1}{n_L} \sum_{j=1}^{n_L} \sum_{i=1}^{m_{a,j}} \frac{y_{a,j,i}}{t_{\max,T,i}}, \quad (6)$$

where $y_{a,j,i}$ = number of capelin of size a in capelin observation i in cod stomach j with digestion time = t_{max}, $m_{a,j}$ = number of length measurable capelin observations of size a in cod stomach j, n_L = number of cod stomachs in length group L containing length measurable capelin. The mean predation rate is expressed as number of capelin consumed per cod per hour.

Consumption was then calculated for each year, time interval, area, size group of capelin and length group of cod as:

$$C_{y,q,s,a,L} = N_{y,q,s,L} \times P_{y,q,s,L} \times \mu_{y,q,s,a,L} \times h_q, \quad (7)$$

where $N_{y,q,s,L}$ = number of cod of length group *L* in year *y*, interval *q* and area *s*, $P_{y,q,s,L}$ = proportion of all cod stomachs from length group *L* sampled in year *y*, interval *q* and area *s*, containing measurable capelin, $\grave{\imath}_{y,q,s,a,L}$ = the year, interval and area specific mean predation rate as defined in (6), and h_q = duration in hours of interval *q*. The areas used correspond to the strata systems used by IMR during standard bottom trawl surveys on demersal fish in the Barents Sea in winter and autumn (Figure 1). The winter system was used in intervals 1–2, and the autumn system in intervals 3–5. In the following the areas will be denoted *strata*.

Abundance of cod ($N_{y,q,s,L}$) was calculated by scaling the bottom trawl survey index to the VPA estimate of total abundance in the following way: The number at age *A* in the beginning of year *y*, $NV_{y,A}$ is available from the VPA (ICES, 2001). The stock abundance in time period *q* can then be calculated as $NV_{y,q,A} = NV_{y,A}e^{-(F_{y,A}+M_{y,A})t_q/12}$, where t_q is the number of months from the start of the year to the mid–point of time period *q*. *F* and *M* is the fishing mortality and natural mortality, respectively. Adjustments also need to be made for the proportion of the total stock that is outside the survey area: One should adjust for the proportion by age found in the Lofoten (ICES, 2001) and Svalbard areas for the winter survey (intervals 1 and 2). For the summer survey, one may assume that the entire stock is covered.

Assume that a survey estimate of abundance by stratum *s* and length *L* is available at time *q* in year *y*, $n_{,y,q,s,L}$. The total survey estimate (entire area) of age *A* fish is given by $n_{\ y,q,A}$. These abundance indices are calculated in the same way as described in Jakobsen et al. (1997). Age–length keys are calculated on main areas *o* consisting of several strata. Let $R_{y,q,o,A,L}$ be the proportion of fish in year *y*, interval *q* and main area *o*, which is of age *A* and length *L*.
The survey–to–VPA scaling factor by age is given by:

$$\alpha_{y,q,A} = \frac{NV_{y,q,A}}{n_{y,q,A}}$$

One can then calculate abundance by length in each main area using the age–length key, and use this to calculate the abundance by length in each stratum:

The abundance by length in each main area is given by:

$$N_{y,q,o,L} = \sum_{A} n_{y,q,o,A}\alpha_{y,q,A}R_{y,q,o,A,L},$$

and the abundance by length in each strata by

$$N_{y,q,s,L} = \frac{n_{y,q,s,L}}{n_{y,q,o,L}}N_{y,q,o,L}.$$

Note that swept–area estimates from autumn 1995 were used in autumn 1994, because of missing data in 1994.
The consumption estimates were summed over length groups of cod and strata to get total consumption of different size groups of capelin in each interval. Due to variable temporal and spatial coverage in the temperature and stomach data, some extrapolation was needed to estimate the yearly consumption (Appendix 1).

Biomass of capelin consumed was calculated for each year y, time interval q and size group of capelin a as:

$$B_{y,q,a} = C_{y,q,a} \times GMw_{y,q,a},$$

where $\mathrm{C}_{y,q,a}$ is consumption in number of individuals and $\mathrm{GMw}_{y,q,a}$ is the mean of the weight of capelin in the Barents Sea. $\mathrm{GMw}_{y,q,a}$ was estimated from survey data on weight at size for capelin in the Barents Sea from IMR.

Database operations, calculations and statistical analysis were done with SAS® 8.1 for Windows (SAS Institute Inc., Cary, NC, USA).

Results

Table 3 shows cod's consumption in billion individuals and biomass of different size groups of capelin in the Barents Sea in the five sample intervals in 1992–1997. Consumption of capelin was highest in the period 1992–1994. 1992 and 1993 was the years of highest consumption, with total consumption being about 2–3 times higher compared to 1994. The consumption was gradually declining in the period 1994–1997.

References

Bogstad, B., Pennington, M., and Vølstad, J. H. 1995. Cost–efficient survey designs for estimating food–consumption by fish. Fisheries Research (Amsterdam), 23: 37–46.

Bromley, P. J. 1994. The role of gastric evacuation experiments in quantifying the feeding rates of predatory fish. Reviews in Fish Biology and Fisheries, 4: 36–66.

ICES 2001. Report of the Arctic Fisheries Working Group. ICES CM 2001/ACFM: 19.

Jakobsen, T., Korsbrekke, K., Mehl, S., and Nakken, O. 1997. Norwegian combined acoustic and bottom trawl surveys for demersal fish in the Barents Sea during winter. ICES CM 1997/Y: 17.

Johansen, G. O. 2002. Temporal and spatial variation in predation on juvenile herring (*Clupea harengus* L.) by Northeast Arctic cod (*Gadus morhua* L.) in the Barents Sea in 1984–1997. ICES Journal of Marine Science, 59: 270–292.

Jones, R. 1974. The rate of elimination of food from the stomachs of haddock *Melanogrammus aeglefinus*, cod *Gadus morhua* and whiting *Merlangius merlangus*. Journal du Conseil. Conseil International pour l'Exploration de la Mer, 35: 225–243.

Lepesevich, Y. M. and Shevelev, M. S. 1997. Evolution of the Russian survey for demersal fish: from ideal to reality. ICES CM 1997/Y: 9.

Mehl, S. 1989. The Northeast Arctic cod stock's consumption of commercially exploited prey species in 1984–1986. Rapports et Procés–Verbaux des Réunions. Conseil International pour l'Exploration de la Mer, 188: 185–205.

Mehl, S. and Yaragina, N. A. 1992. Methods and results in the joint PINRO–IMR stomach sampling program. *In* Interrelations between fish populations in the Barents Sea. Proceedings of the fifth PINRO–IMR symposium, 12–16 August 1991 Murmansk, pp. 5–16. Ed. by B. Bogstad and S. Tjelmeland. Institute of Marine Research, Bergen.

Mergardt, N., and Temming, A. 1997. Diel pattern of food intake in whiting (*Merlangius merlangus*) investigated from the weight of partly digested food particles in the stomach and laboratory determined particle decay functions. ICES Journal of Marine Science, 54: 226–242.

Temming, A., and Andersen, N. G. 1994. Modeling gastric evacuation without meal size as a variable. A model applicable for the estimation of daily ration of cod (*Gadus morhua* L.) in the field. ICES Journal of Marine Science, 51: 429–438.

Temming, A., and Herrmann, J. P. 2003. Gastric evacuation in cod: Prey–specific evacuation rates for use in North Sea, Baltic Sea and Barents Sea multi–species models. Fisheries Research (Amsterdam), 63: 21–41.

Tereshchenko, V. V. 1996. Seasonal and year–to–year variations of temperature and salinity along the Kola meridian transect. ICES CM 1996/C: 11.

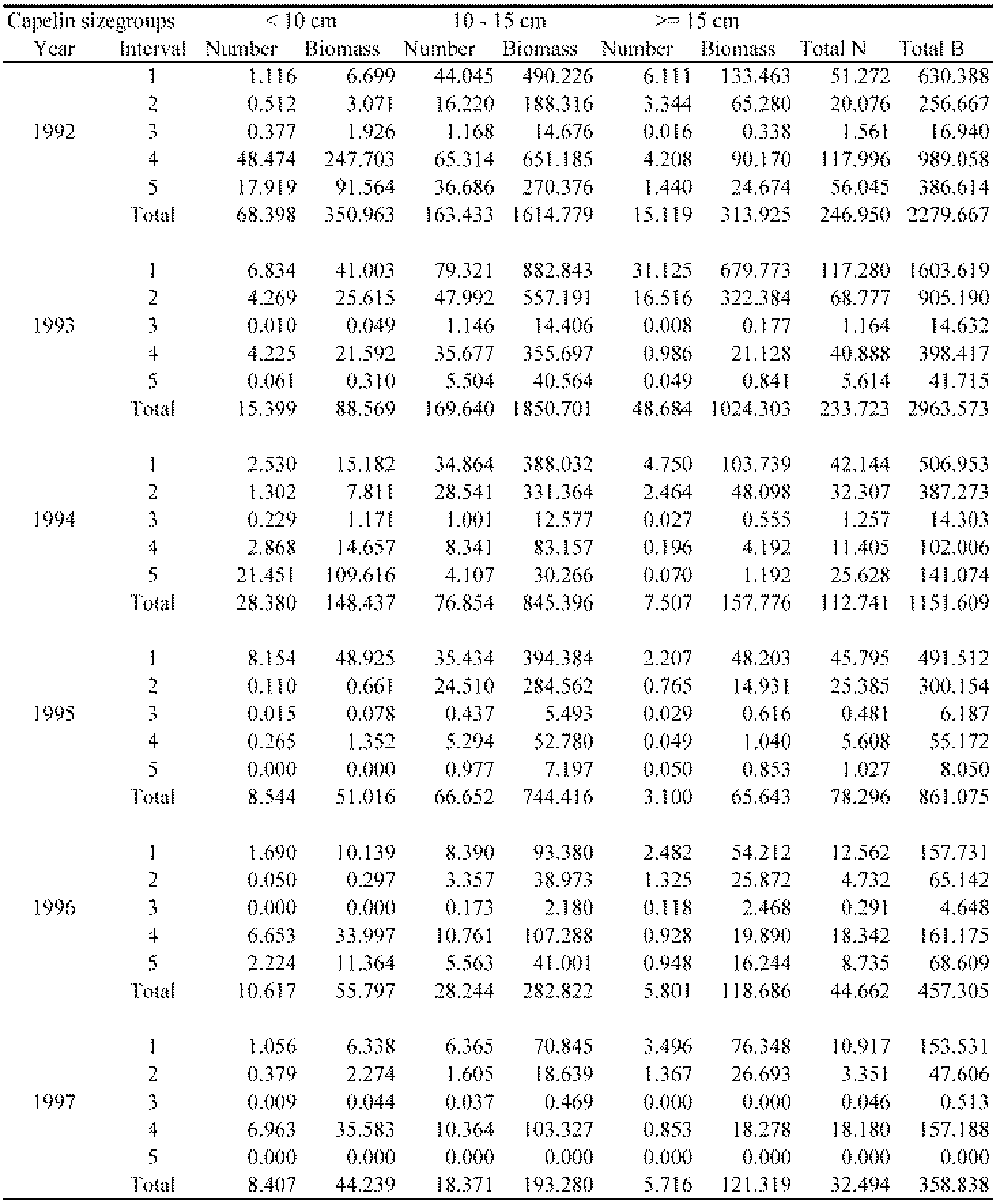
Table 1. Number of trawl hauls with stomach samples from cod = 20 cm each year and interval, and number of cod stomachs by 10 cm length groups included in the analyses.

Cod size										
Year	Interval	Hauls	20–29 cm	30–39 cm	40–49 cm	50–59 cm	60–69 cm	70–79 cm	80–89 cm	= 90
	1	102	261	244	349	316	238	187	129	62
	2	4	0	1	7	3	8	15	4	0
1992	3	8	0	1	15	26	34	56	39	28
	4	97	219	329	258	374	239	215	197	149
	5	5	0	0	0	3	10	12	13	10
	1	150	236	357	433	474	410	221	142	91
	2	10	7	20	65	121	66	17	9	11
1993	3	0	0	0	0	0	0	0	0	0
	4	100	494	432	591	565	643	297	160	126
	5	41	21	33	39	90	260	114	52	75
	1	166	325	376	457	529	587	347	125	79
	2	20	19	33	61	56	50	55	27	24
1994	3	8	31	56	75	68	64	54	20	19
	4	97	186	243	230	445	413	317	111	124
	5	51	3	22	121	433	387	236	45	16
	1	188	351	398	554	763	651	439	232	78
	2	62	1	27	210	575	472	238	89	32
1995	3	28	0	16	89	295	230	96	27	10
	4	143	203	268	408	648	682	335	140	74
	5	81	25	27	85	231	335	139	78	82
	1	268	370	372	482	617	657	356	204	131
	2	45	6	48	199	395	315	183	100	32
1996	3	49	23	49	174	375	387	180	95	121
	4	138	292	282	335	482	567	393	159	74
	5	93	88	89	135	351	404	391	132	63
	1	201	223	288	411	778	631	480	186	109
	2	87	23	86	292	776	590	369	206	77
1997	3	52	1	66	184	427	322	233	144	90
	4	108	209	263	379	573	506	267	158	135
	5	75	31	120	190	350	399	164	91	28

Table 2. Length of capelin in 1 cm groups (TL capelin) and median weight of the cod (W$_{med}$ cod) used to find the expected breakpoint in the relationship between the cumulative frequency of measurable capelin as a function of digestion time grouped within 1–hour intervals. Data are from 1992–97. Maximum time range (t_{max}) in hours within which all the capelin within the given 1 cm group were measurable at 4°C is estimated by: $\ln(t_{\max}) = - - 2.0204 + 1.7657\ln(TL_{capelin})$.

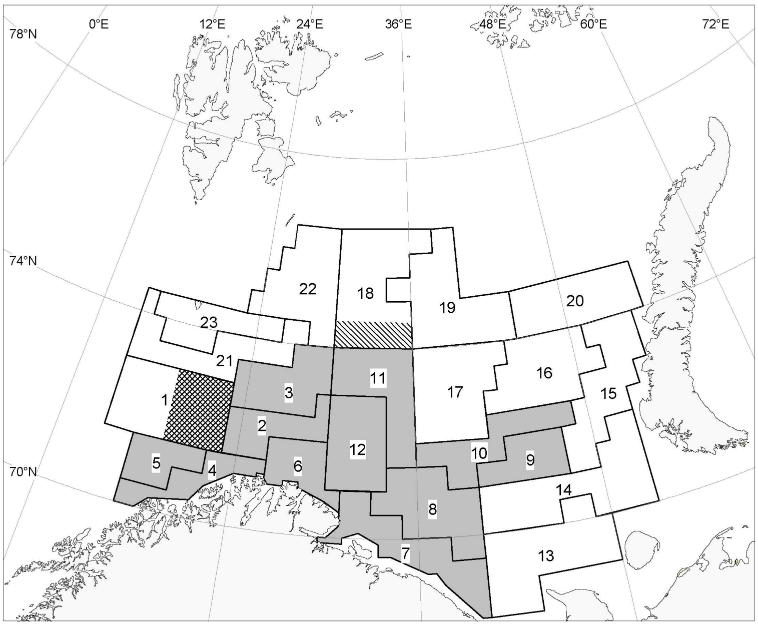
TL capelin (cm)		n		W$_{med}$ cod (kg)		t$_{max}$ (h)	
< 7		53		1.51		2.690	
7 – 8		53		1.86		4.652	
8 – 9		107		1.84		5.803	
9 – 10		224		2.16		7.062	
10 – 11		414		2.37		8.427	
11 – 12		602		2.55		9.895	
12 – 13		1094		2.24		11.465	
13 – 14		927		2.43		13.134	
14 – 15		857		2.40		14.900	
15 – 16		613		2.47		16.762	
16 – 17		314		2.68		18.718	
17 – 18		198		2.38		20.767	
> 18		40		3.98		25.140	

Table 3. Consumption of capelin by Northeast Arctic cod in the Barents Sea in the period 1992–1997. The estimates are in billion individuals and 1000 metric tons.



Capelin sizegroups		< 10 cm		10 - 15 cm		>= 15 cm			
Year	Interval	Number	Biomass	Number	Biomass	Number	Biomass	Total N	Total B
1992	1	1.116	6.699	44.045	490.226	6.111	133.463	51.272	630.388
	2	0.512	3.071	16.220	188.316	3.344	65.280	20.076	256.667
	3	0.377	1.926	1.168	14.676	0.016	0.338	1.561	16.940
	4	48.474	247.703	65.314	651.185	4.208	90.170	117.996	989.058
	5	17.919	91.564	36.686	270.376	1.440	24.674	56.045	386.614
	Total	68.398	350.963	163.433	1614.779	15.119	313.925	246.950	2279.667
1993	1	6.834	41.003	79.321	882.843	31.125	679.773	117.280	1603.619
	2	4.269	25.615	47.992	557.191	16.516	322.384	68.777	905.190
	3	0.010	0.049	1.146	14.406	0.008	0.177	1.164	14.632
	4	4.225	21.592	35.677	355.697	0.986	21.128	40.888	398.417
	5	0.061	0.310	5.504	40.564	0.049	0.841	5.614	41.715
	Total	15.399	88.569	169.640	1850.701	48.684	1024.303	233.723	2963.573
1994	1	2.530	15.182	34.864	388.032	4.750	103.739	42.144	506.953
	2	1.302	7.811	28.541	331.364	2.464	48.098	32.307	387.273
	3	0.229	1.171	1.001	12.577	0.027	0.555	1.257	14.303
	4	2.868	14.657	8.341	83.157	0.196	4.192	11.405	102.006
	5	21.451	109.616	4.107	30.266	0.070	1.192	25.628	141.074
	Total	28.380	148.437	76.854	845.396	7.507	157.776	112.741	1151.609
1995	1	8.154	48.925	35.434	394.384	2.207	48.203	45.795	491.512
	2	0.110	0.661	24.510	284.562	0.765	14.931	25.385	300.154
	3	0.015	0.078	0.437	5.493	0.029	0.616	0.481	6.187
	4	0.265	1.352	5.294	52.780	0.049	1.040	5.608	55.172
	5	0.000	0.000	0.977	7.197	0.050	0.853	1.027	8.050
	Total	8.544	51.016	66.652	744.416	3.100	65.643	78.296	861.075
1996	1	1.690	10.139	8.390	93.380	2.482	54.212	12.562	157.731
	2	0.050	0.297	3.357	38.973	1.325	25.872	4.732	65.142
	3	0.000	0.000	0.173	2.180	0.118	2.468	0.291	4.648
	4	6.653	33.997	10.761	107.288	0.928	19.890	18.342	161.175
	5	2.224	11.364	5.563	41.001	0.948	16.244	8.735	68.609
	Total	10.617	55.797	28.244	282.822	5.801	118.686	44.662	457.305
1997	1	1.056	6.338	6.365	70.845	3.496	76.348	10.917	153.531
	2	0.379	2.274	1.605	18.639	1.367	26.693	3.351	47.606
	3	0.009	0.044	0.037	0.469	0.000	0.000	0.046	0.513
	4	6.963	35.583	10.364	103.327	0.853	18.278	18.180	157.188
	5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Total	8.407	44.239	18.371	193.280	5.716	121.319	32.494	358.838

A



B

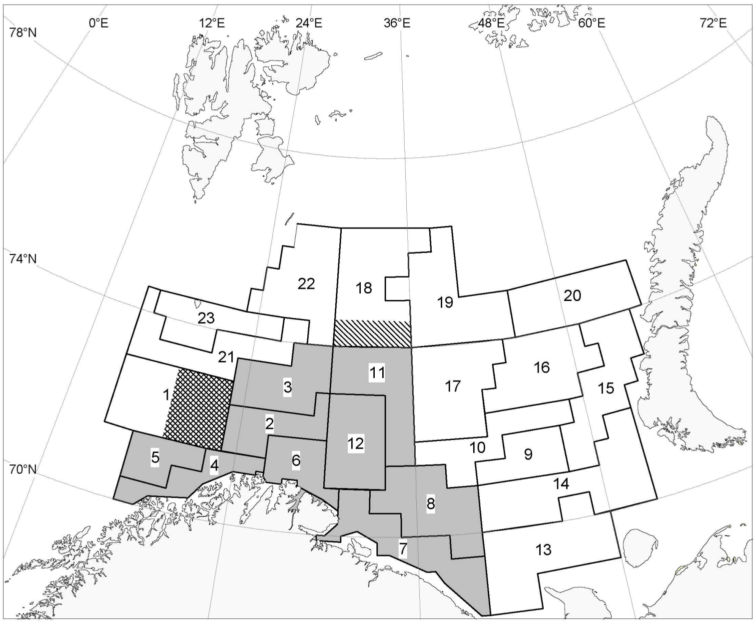
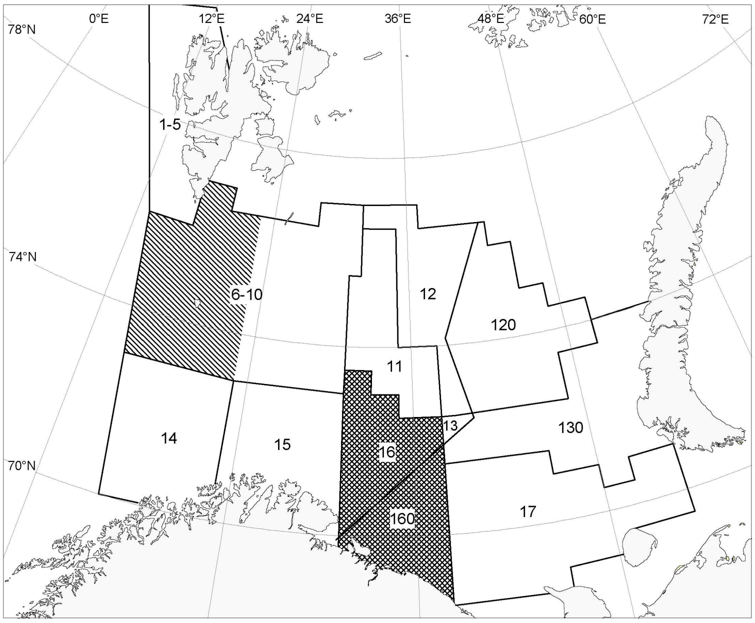


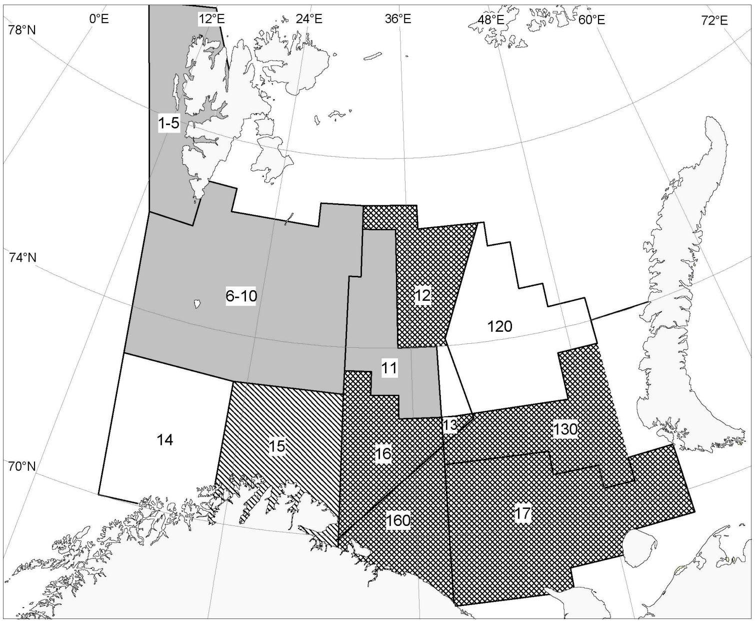
Figure 1. Stratification and extrapolation scheme for estimating consumption of capelin in the Barents Sea in 1992–97. Grey shading, crosshatching and left slanting indicates extrapolation to 100%, 50 % or 25% of the cod population in that stratum, respectively (see Appendix 1). In cases where only a part of a stratum is marked, the marking indicates the spatial distribution of capelin in cod stomachs in that stratum. A, B, C, D and E indicates sampling intervals 1–5, respectively.

Figure 1 continued.

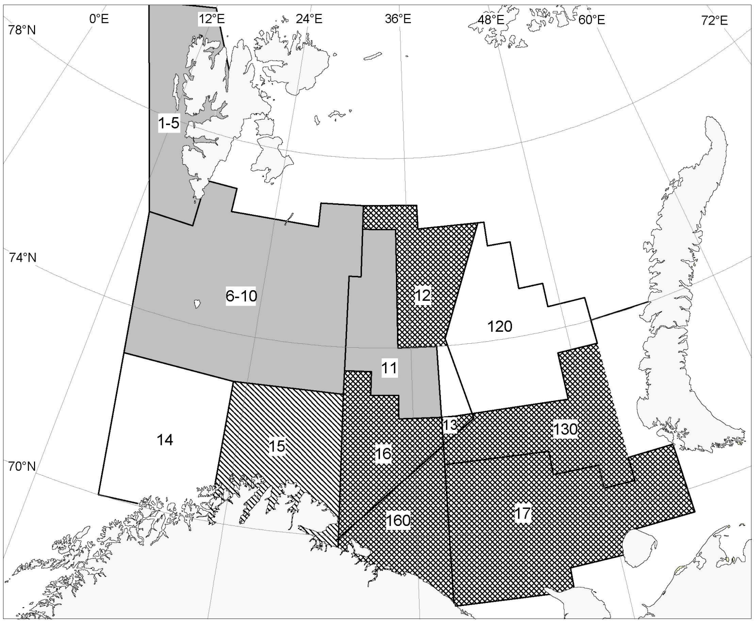
C



D



E


78°N
0°E
12°E
24°E
36°E
48°E
60°E
72°E
74°N
70°N
1-5
6-10
12
120
11
13
130
14
15
16
160
17

Appendix 1

Extrapolation of temperature data

The temperature data were given on a quarterly basis, and data from quarters 1, 2, 3 and 4 were used to estimate temperature in intervals 1, 2, 4 and 5, respectively. For interval 3 temperature data for quarter 2 and 3 were combined. In quarter 4 in 1994–1997 some stratum means were missing, and were estimated as follows: The difference between average temperature in stratum s in year y and quarter Q ($\overline{T_{y,Q,s}}$) and the year and quarter specific mean temperature for the 0–200 m depth range from the Russian hydrographic section off the Kola peninsula ($\overline{K_{y,Q}}$) (Tereshchenko, 1996) were calculated for the period 1992–1997 as: $dif_{y,Q,s} = \overline{T_{y,Q,s}} - \overline{K_{y,Q}}$. A linear regression of these differences in quarters 3 and 4 ($\mathrm{dif}_{y,4,s} = -0.0092 + 0.8566\ (\mathrm{dif}_{y,3,s})$, df = 37, $r^2 = 0.81$, $p < 0.0001$), combined with the mean Kola section temperature in quarter 4 was used to estimate the missing stratum means.

Spatial extrapolation of stomach data

The stomach data were segregated on size groups of cod and size groups of herring. Analyses of the spatial distribution of capelin in cod stomachs were done to set the rules used when extrapolating. If a stratum contained < 5 trawl hauls with stomach samples, predation rates, averaged over the neighboring strata with = 5 hauls with stomach samples in the same interval, was extrapolated in space. The proportion of cod with measurable capelin in their stomachs was calculated for the neighboring strata combined. If neighboring strata also lacked sufficient data, temporal extrapolation from other intervals was carried out, preferably from the same stratum. In intervals 1–2 this was done if the target stratum had < 2 neighboring strata with = 5 hauls with stomach samples, while in intervals 3–5 only 1 neighboring stratum with sufficient data was required. If the stratum contained some hauls with stomach samples, these data were included. In some strata the predation on capelin by cod only occurs in parts of the stratum, or the intensity of predation is lower compared to the main areas. When extrapolating to such strata the proportion of cod with measurable capelin in their stomachs was corrected according to the approximate proportion of the cod population assumed to consume capelin in that stratum (Figure 1).

Temporal extrapolation of stomach data

Temporal extrapolation to an interval was done by transferring the predation rate and proportion from a stratum in the interval of origin to the same stratum in the interval subject to extrapolation. Data from the stratum to which data was extrapolated were included when present. The extrapolation was done to the shaded and hatched strata in the interval, according to the patterns in figure 1. Data from interval 1 was extrapolated to interval 2. Data from interval 4 was extrapolated to intervals 3 and 5. This extrapolation scheme was based on the similar spatial distribution of capelin in cod stomachs in intervals 1–2 and 3–5.

7.8 Multispecies stock dynamics in the North Sea including spatial inhomogeneity

Morten Vinther and Peter Lewy
January 2004
Danish Institute for Fisheries Research

Introduction

Most stock assessment models assume that the spatial distribution of the stock, recruitment, growth, natural mortality and fishing activity remains unchanged over time within the area considered and the presence of spatial inhomogeneity is thus disregarded. If spatial inhomogeneity should be accounted for the area in question has to be divided into two or more subdivisions and stock dynamics within the sub-areas and the migration between sub-areas should be considered.

The lack of spatial homogeneity is especially important for the evaluation of the potential effect of closed areas or local fishing effort reduction, which are a widely used management measure for protection of marine resources (Pastoors et al. 2000; Horwood et al. 1998; Coleman et al. 2004).

In the present paper the stock dynamics have been analysed assuming spatial inhomogeneity in the North Sea. This has been exemplified for the simple case where the North Sea is divided into two sub-areas, a northern and a southern area (Figure 1). Eight North Sea stocks have been considered. Multispecies stock assessments and predictions (Helgason and Gislason 1979; Pope 1979; Sparre 1980; Gislason & Sparre, 1987) including species interaction and migration between the two sub-areas have been carried out and compared to results of analyses of the entire North Sea. Four predators, cod, haddock, whiting and saithe and four prey species, herring, sandeel, Norway pout and sprat have been included. This division of the North Sea seems to be meaningful as the Roundfish stocks mainly occur in the Northern North Sea and sprat mainly in the southern area while sandeel and herring occur in the entire area.

Area-specific single species management measures for the North Sea have previously been investigated (Anon. 1992). In general, historical spatial disaggregated analyses (VPA) are seldom carried out because they are very data demanding and require area divided data for a range of years on catch, effort information and biological data. Such data do not exist for the North Sea. For this reason it is not possible to perform spatial disaggregated VPA's. Even if spatial disaggregated catch and effort were available it is not possible to perform a VPA including migration (Gislason and Sparre, 1994) However, for a single year, 1991, such data were collected in the North Sea including both catch data by fleet (Lewy 1992), stock distribution data (ICES 1998) and stomach contents data (ICES 1997). These data sources make it possible for 1991 to carry out regional species analyses of stock distribution and fishing and predation mortalities.

The multispecies model implemented for the North Sea is probably one of the most investigated multispecies models (see i.e. ICES, 1997) and has a very comprehensive database of stomach contents data (Daan, 1989; ICES, 1997). As mentioned above even for the North Sea it is not possible to perform area divided VPA's and therefore a more simple approach must be applied. In this paper, the outcome of the MSVPA applied to the entire North Sea is combined with stock distribution data, spatial disaggregated catch and stomach contents data to estimate local food suitability coefficients and fishing mortalities. These values are then used as basis for multispecies catch projections for evaluation of local area management measures. In order to evaluate the importance of spatial heterogeneity catch and stock projections based on the North sea divided into the northern and southern areas were compared to projections based on the entire North Sea.

Methods and material

Spatially disaggregated assessment model

The method for assessment of sub-areas is straightforward and based on the MSVPA algorithm applied to the entire North Sea. Gislason and Sparre (1987) give a formal description of MSVPA and only a few details are given here. MSVPA operates by quarter and year d includes the following input and variables.

Input to MSVPA includes the variables (not the complete list):

- C$_{s,a}$ Catch numbers by species (s) and age (a):
- F$_{s,a}$ Terminal fishing mortality
- M1$_{s,a}$ Residual natural mortality
- R$_{s,a}$ Food ration per predator
- W$_{j,a,i,b}$ Mean weight of prey (i), age (b) in the stomach of predator (j), age (a)
- S$_{j,a,i,b}$ Relative stomach contents weight

Output from the MSVPA includes:

- M2$_{s,a}$ Predation mortality
- F$_{s,a}$ Fishing mortality
- Z$_{s,a}$ Total Mortality
- N$_{s,a}$ Stock numbers
- U$_{j,a,i,b}$ Food suitability coefficient (year independent)

Stock assessment by sub-area or region (r) requires in addition, the following data for at least one year:

- DIST$_{s,a,r}$ Stock distribution (relative)
- S$_{j,a,i,b,r}$ Relative stomach contents weight
- C$_{s,a,r}$ Catch numbers

The method for performing the proposed local assessment is first to perform a MSVPA for the total North Sea. The result includes stock numbers and mortality rates for the whole areas. To calculate the area specific parameters (local fishing and predation mortalities and suitability coefficients) stock numbers for the entire North Sea are distributed on sub-areas in accordance with the assumed known stock distribution ($\overline{N}_{s,a,r} = \overline{N}_{s,a} * DIST_{s,a,r}$). When the local stock numbers are estimated, local food suitability coefficients can be calculated from the local stomach contents in accordance to the definition of suitability (Gislason and Sparre, 1987) :

$$U_{i,a,j,b,r} = \frac{\frac{S_{i,a,j,b,r}}{\overline{N}_{i,a,r} * W_{i,a,j,b}}}{\sum\limits_{k,d} \frac{S_{k,d,j,b,r}}{\overline{N}_{k,d,r} * W_{k,d,j,b}}}$$

where the indices *k* and *d* indicate all preys species and all prey ages respectively.

The local predation mortality can then be calculated as defined by Andersen and Ursin (1977) as:

$$M2_{i,a,r} = \sum_{j,b,r} \overline{N}_{j,b,r} * R_{j,b} * \frac{U_{i,a,j,b,r}}{\sum\limits_{k,d,r} \overline{N}_{k,d,r} * W_{k,d,j,b} * U_{k,d,j,b,r}} \qquad (1)$$

Finally, local fishing mortalities can be calculated from local catches and mean local stock numbers: $F_{s,a,r} = \frac{C_{s,a,r}}{\overline{N}_{s,a,r}}$.

Predictions

Comparisons of spatially disaggregated analyses and analyses including the entire North Sea could be carried out by comparing local and entire North Sea F's, suitabilities and predation mortalities. As this includes many parameters and corresponding comparisons we have instead chosen to summarise the many parameters. This was done by considering multispecies catch and stock predictions and by comparing predicted local and entire North Sea biomasses. Moreover, this also enables evaluation of the effect of protected areas.

The forecast multispecies prediction model, MSFOR (Gislason and Sparre,1987) predicts future catches and stock sizes from input fishing mortalities and initial stock sizes. The predation mortality is estimated as given in (1) using the MSVPA estimated food suitability values.

Like MSVPA, MSFOR is originally designed for one area only. To handle more areas migration or redistribution of the stocks has to be accounted for. For the North Sea case the predictions were carried out for the period 1991-2000 based on the situation calculated for 1991, i.e. using local stock size, F's and suitability coefficients. Recruitment at age zero was assumed constant over years and equal to the average over the period 1991-2000. The "migration" was accounted for by redistribution of stock numbers at the start of each quarter using stock proportion by sub-area calculated from the quarterly IBTS survey data. The average spatial stock distribution for the period 1991-1995 was used for all the prediction years, 1991-2000. The choice of constant redistribution over years based on the period 1991-1996 should be considered as a preliminary test example. A better choice will depend on analyses of the spatial distribution of IBTS data in the period.

The MSVPA and MSFOR algorithm, including local area analyses have been implemented using of the so-called 4M package (Vinther et al, 2000), which is an updated and extended version of the MSVPA and MSFOR programs.

Four scenarios were considered:

- **Baseline scenario**, where the North Sea was considered as one stock unit area for all stocks considered.

- **Scenario 1**: The North Sea is divided into the northern and southern sub-areas. For each quarter and year the stocks are spatially redistributed to the two sub-areas according the IBTS survey average distribution in 1991-1995.

- **Scenario 2**: As scenario 1, but no redistribution or migration was assumed

- **Scenario 3**: AS scenario 1, but cod recruitment in the southern North Sea was reduced to a half.

Materials

The required data for assessment by sub-areas comprise spatially disaggregated data on stock, catch and stomach contents. MSVPA operates by quarter and the spatial disaggregated data must be given by quarter as well. The year 1991 was the exception where all the types of data were available. The ICES, International Bottom Trawl Survey (IBTS) supplied stock distribution data; the STCF data base (Lewy et al, 1992) gave catch catches and the ICES stomach sampling program (ICES, 1997) supplied stomach contents data. For all data set, information is given by quarter and ICES rectangles (1^0 x 0.5^0). However , the stratification for sampling of age-length keys has in most cases been the so-called ICES roundfish areas, such that the actual minimum sub-area becomes one of the totally seven roundfish area in the North Sea.

Stock data

For the period 1991-1995 the IBTS was extended to a quarterly trawl survey for the North Sea. Description of the survey method can be found in ICES (1996) and a detailed description of the result for 1991 is given in ICES (1998). IBTS data for stock distribution consist of mean CPUE per ICES rectangle by species and age. The average proportion for 1991-1995 of the stock in a sub-area was calculated as the arithmetic mean over years of the average rectangle CPUE. No attempt was made to correct for varying rectangle area or rectangles without hauls. Catches of older fish were sporadic, and for each species a plus group was defined, which was assumed to represent the distribution of older fish.

As sandeels were very seldom caught in the IBTS survey the survey could not be used to estimate the spatial distribution. Instead the spatially distribution of commercial catches was assumed to represent the stock distribution. Fishery for sandeel takes mainly place in the second quarter. Third quarter has a rather limited effort while the fishery is almost non-existing for the rest of the year. The assessment (ICES, 2001) give catches by half-year for the Northern and southern North Sea. Due to the limited information the catch distribution for the first half-year of 1991 was used as stock distribution key for all quarters and prediction years.

Catch data

The STCF database contains quarterly catch at age data for each of the ICES rectangles in the North Sea in 1991. These data were used as a spatially distribution key to calculate the MSVPA catches by sub-area.

Stomach contents data

Input to the MSVPA is average weight proportion of a prey of the stomach content and the mean (fresh) weight of the prey. Data are given by year, quarter, predator, predator age, prey and prey age. MSVPA operates for the entire North Sea and stomach data reflect average stomach contents for the total population in the whole model area. The procedure used for compilation of data used in the assessment for the northern and southern North Sea is described in the appendix.

MSVPA and MSFOR data and setup.

Table 1 gives the setup of MSVPA with respect to species and stomach contents data. Species where stomachs have been sampled are in the model considered as predators. Only stomachs collected in 1991 were included in the model. This in contrast to the so-called key run, defined at the WG (ICES 1997), which also included stomach data from other years, mainly 1981. The key run includes a group of other predators like “sea birds”, grey gurnards and *raja radiata* with an assumed known biomass, and an observed or derived stomach contents. Predation from this group induce a mortality on the VPA species, but due to problems with the division of stomach contents data on sub-areas, this group was left out of the present spatially disaggregated MSVPA. The catch numbers were extended to year 1998, compared to 1995 used in the key-run.

Table 1. Input data for the North Sea MSVPA

VPA Species	Age groups	No. of stomachs sampled in 1991
Cod	0-11	9700
Haddock	0-10	12883
Saithe	0-10	3020
Mackerel (“Otherpredator”)	0-15	5456
Whiting	0-8	38413
Herring	0-9	
Norway pout	0-3	
Sandeel	0-4	
Plaice	0-15	
Sole	0-15	

In single species VPA and MSVPA fishing mortality values for the oldest age and for all ages the last year must be given as input. These “terminal F” values were estimated in a so-called multispecies tuning (Vinther 2001) which is a technique where the tuning algorithms applied in the ICES stock assessment uses the multispecies natural mortality rates. The resulting terminal Fs from tuning are then used in MSVPA for production of a new set of natural mortality values. This exchange of terminal F and natural mortalities is continued until equilibrium is reached.

Results

It is generally known that haddock, saithe and Norway pout only occurs in the northern North Sea. Since the beginning of the nineties the main part of the cod and whiting stocks also are found in the northern North Sea, 80% and 70% respectively. Moreover, sprat (and the flatfish) is mainly found in the southern North Sea, herring and sandeel are found in both areas. This means that the predators mainly occur in the Northern North Sea. In general, on would therefore expect that that predation mortality is greater in the northern than in the southern North Sea. Table 1 showing the average predation mortality for age groups zero to 2 confirms this expectation.

Table 1. Average predation mortality for age groups 0 to 2 in 1991.

$M2_{0-2}$	Northern North Sea	Southern North Sea
Cod	0.02	1.50
Haddock	0.09	0.02
Whiting	0.47	0.51
Saithe	0.00	0.00
Herring	0.05	0.01
Sandeel	0.93	0.05
Norway pout	0.51	0.15
Sprat	0.10	0.46

Table 1 shows that – except for sprat, cod and whiting – predation mortality in the northern North Sea is 3-19 times that of the southern area. The sprat exception should be considered as an artefact as the sprat stock nearly always occurs in the southern area.

Plots of predicted stock descriptors and yield for the scenarios for 1992 – 2002 are given in the Figures 2-5. Figures 2 and 3 show the development of the Southern and the total stock descriptors compared to the baseline scenario while baseline and scenario 1-3 predictions of SSB for the entire North Sea is compared in Figure 5.

Figures 2-5 further show that all stocks considered in nearly all cases have reached equilibriums after 10 years.

Comparisons of baseline and scenario 1 predictions (Figure 2) of SSB in year 2002 shows that the relative difference between baseline and scenario 1 was relative small for most Roundfish (cod, haddock and saithe (0% - 17%)) while the difference was larger for the prey species, sandeel, Norway pout and sprat (33% - 64%).

For scenario 1 the spatial stock distribution remains unchanged due to the assumption of constant quarterly stock redistribution.

For scenario 2, where no migration between northern and southern North Sea is assumed the spatial distribution of SSB on the two areas are shown in table 2.

Table 2. The proportion of SSB in the southern North Sea in 1992 and 2002 in scenario 2.

Proportion (%)	1992	2002
Cod	0.21	0.38
Haddock	0.04	0.02
Whiting	0.32	0.58
Saithe	0.01	0.11
Herring	0.55	0.74
Sandeel	0.67	0.97
Norway pout	0.03	0.00
Sprat	0.85	0.87

Table 2 shows that for most stock the biomass proportion in the southern area increases from 1992 to 2002 if no migration is assumed.

With respect to the predicted stock biomasses for the total North Sea the difference between scenario 1 and 2 is illustrated in Table 3 showing the relative difference, $(SSB_1/SSB_2 - 1)$, in 2002:

Table 3. Relative difference between scenario 1 and 2 of predicted total North Sea SSB in 2002

	Relative difference
Cod	0.17
Haddock	-0.41
Whiting	0.17
Saithe	-0.08
Herring	-0.33
Sandeel	-0.41
Norway pout	-0.32
Sprat	0.21

Table 3 illustrates the significance of the migration assumptions on the long-term total stock biomass. Considerable differences are found especially for the haddock, sandeel and herring and stocks.

Regarding scenario 3 the relative difference compared to scenario 1 for cod biomass in 2002 was about 80% while only minor changes were found for the other stocks.

Predictions based on spatial divided areas generally differ from those based on the entire area. This is illustrated in Table 4:

Table 4. The relative difference between predicted SSB in 2002 based on a division of North Sea into two areas and the entire North Sea

Relative difference	Scenario 1	Scenario 2
Cod	0.18	0.01
Haddock	- 0.13	0.48
Whiting	0.13	- 0.03
Saithe	0.00	0.09
Herring	0.14	0.69
Sandeel	0.63	1.76
Norway pout	0.34	0.98
Sprat	0.47	0.22

Especially for scenario 2 large differences between area-divided and total area predictions are found for sandeel, Norway pout, herring and haddock. Even for scenario 1, where constant quarterly redistribution is assumed, significant differences are found for sandeel and sprat.

Discussion

Stock assessment is usually conducted on management unit areas implicitly assuming that the spatial interrelationship between stocks and fishing effort remains unchanged over time. If these conditions are not fulfilled or if the effect of marine protected areas should be evaluated it may be necessary to conduct spatially disaggregated stock assessments. Generally, such analyses will differ from analyses based on the entire area because the species composition and the spatial stock distribution and fishing effort differ. Furthermore, for the spatially disaggregated analyses it is important to include possible migration between sub-areas, which is seldom historically known and difficult to predict. The importance of spatial inhomogeneity and migration was analysed for a division of the North Sea into a northern and a southern area in a multispecies context including predation of four predators on four prey species. The effect of migration was considered comparing of a case with constant quarterly redistribution of stocks and the case with no migration. The results showed that significant differences between predictions based on the division of the North Sea into two sub-areas and on the entire North Sea were found: For total North Sea SSB relative stock differences up to 1.76 occurred. Significant changes of the spatial distribution were found as well as for instance the proportion of the whiting and sandeel stocks in the southern North Sea increased from 32% and 67% in 1992 to respectively 58% and 97% in 2002 for the case with no migration. The two migration scenarios considered lead to substantial different long-term biomasses and spatial distribution on the two sub-areas as well.

Migration between the two sub-areas considered dividing the North Sea into a northern and a southern area was accounted for applying quarterly IBTS data, which were available for 1991-1995. These data provides the distribution of stock numbers on the two sub-areas by species and age. Tentatively, it was assumed that the quarterly spatial distribution was constant for the prediction period 1992-2002 and equal to the average distribution for the known period. In principle it would be better to apply the observed spatial distribution for each of the years available, 1992-1995. However, we felt that analyses of these distributions with respect to annually/monthly variations are needed before annual data can be used. Generally, if scientific survey data providing spatial stock distributions for a historical period should be used to predict future distributions this ideally requires analyses of the driving forces affecting stock distributions or migration patterns.

Acknowledgement

The work was partly funded by the EU project QLK5-CT1999-01609: dst^2 Development of structurally detailed statistically testable models of marine populations.

References

Andersen, K. P., and Ursin, E. 1977. A Multispecies Extension to the Beverton and Holt Theory of Fishing, with account of Phosphorus Circulation and Primary Production. Meddr. Danm. Fisk.- og Havunders. **7** 319-435.

Anon. 1992. Report of the meeting of the STCF working group on improvements of the exploitation pattern of the North Sea fish stocks, Charlottenlund 4-14 June 1991

Coleman F.C., Baker P.B., and Koenig C.C. 2004. A review of Gulf of Mexico marine protected areas: Successes, failures, and lessons learned. FISHERIES 29:10-21.

Daan N. (ed), 1989. Database report of the Stomach sampling Project 1981. ICES Cooperative Research Report No. 164

Gislason, H. & Sparre P. 1987. Some theoretical aspects of the implementation of multispecies virtual population anaylisis in ICES. ICES CM 1987/G:51.

Gislason, H. & Sparre P. 1994. Some thoughts on the incorporation of areas and migrations in MSVPA.

ICES C: 1994/mini:15.

Helgason T, Gislason H. 1979. VPA analysis with species interaction due to predation. ICES CM 1979/G:52

Horwood, J. W., Nichols, J. H. and Milligan, S. 1998. Evaluation of closed areas for fish stock conservation. Journal of Applied Ecology 35:893-903.

ICES (1997) Report of the multispecies assessment working group, ICES Headquarter, 11-19 August 1997. ICES CM 1997/Assess:16

ICES 1996. Report of the International Bottom Trawl Survey Working Group. ICES CM 1996/H:1

ICES 1997. Database report of the stomach sampling project 1991. ICES Coop. Res. Rep. No. 219.

ICES 1998. Report of the International Bottom Trawl Survey in the North Sea, Skagerrak and Kattegat in 1991: Quarter 2,3, and 4. ICES 1998/D:8

ICES 2001. Report of the working group on the assessment of demersal stocks in the North Sea and Skagerrak. ICES headquarter 2-12 Oct. 2000. ICES CM 2001/ACFM:07

Lewy, P., Vinther, M., and Thomsen L. 1992. Description of the STCF North Sea database system and the prediction model ABC, Assessment of Bio-economic Consequences of technical measures. ICES CM 1992/d:17.

Pastoors, M. A., Rijnsdorf, A. D. and Van Beeck, F. 2000. Effects of a partially closed area in the North Sea ("plaice box") on stock development of plaice. ICES Journal of Marine Science 57:1014-1022.

Pope JG (1979) A modified cohort analysis in which constant natural mortality is replaced by estimates of predation levels. ICES CM 1979/H16.

Sparre P (1980) A goal function of fisheries (legion analysis). ICES CM 1980/G:40

Vinther, M., Lewy, P and Thomsen L. 2000. Specification and documentation of the 4M Package containing Multi-species, Multi-fleet and Multi-area Models.

Appendix

This section describes the compilation of data to transform stomach data from the individual sample to the average stomach data at population level for one ore more sub-areas. The methods follows the technique given by ICES (1997), but the present description is supplemented by formulas for the actually data transformations done. Moreover, descriptions for handling incomplete data are given as well.

Individual sample information:

Observed stomach contents data include the information shown in the following data hierarchy:

Haul information (*h*) - Quarter of the year (*q*)
- ICES Roundfish area (*r*)
- ICES rectangle (*sq*)
  - **Predator** (*c*)
    - **Predator length class** (*cl*)
    - CPUE , (*CPUE*)
    - **Sample no** (*s*)
    - number of feeding (valid) stomachs, (*NFEED*)
    - number of feeding, but regurgitated stomachs, (*NRGUR*)
    - number of stomachs with skeletal remains only, (*NSKEL*)
    - number of empty stomach, (*NEMPT*)
    - total number of stomachs, (*NTOT*)
    - Total stomach contents of valid stomachs (*WTOT*)
      - **Prey species** (*p*)
        - **Prey length class** (*pl*)
        - Stomach contents weight (*W*)
        - Number of preys

Only stomach contents from the feeding, non-regurgitated, stomachs were sampled. It is assumed that the regurgitated stomachs had a similar stomach content as the (valid) feeding fish and the average stomach contents for a predator length group in haul (potentially including one or more samples) were calculated applying the following relationship:

$$\overline{W}_{h,c,cl,p,pl} = \frac{\sum\limits_{s} W_{s,h,c,cl,p,pl}}{\sum\limits_{s} NFEED_{h,s}} \times \frac{(\sum\limits_{s} NFEED_{h,s} + \sum\limits_{s} NRGUR_{h,s})}{\sum\limits_{s} NTOT_{h,s}}$$

Average stomach contents per roundfish area

ICES rectangles are used as strata in the calculation of average stomach content per ICES roundfish area. If more than one sample are taken from a rectangle, the average stomach content for a predator length class is calculated as a weighted mean, using the number of stomachs sampled as weighting factor.

$$\overline{W}_{sq,c,cl,p,pl} = \frac{\sum\limits_{s} \overline{W}_{sq,s,c,cl,p,pl} NTOT_{sq,s,c,cl}}{\sum\limits_{s} NTOT_{sq,s,c,cl}}$$

The average stomach content of a given predator and length class in a roundfish area are calculated as a weighted mean of the average stomach content per square weighted by the average (square root of) CPUE for the square.

$$\overline{W}_{r,c,cl,p,pl} = \frac{\sum\limits_{sq} \overline{W}_{sq,c,cl,p,pl} \sqrt{\overline{CPUE}_{r,sq,c,cl}}}{\sum\limits_{sq} \sqrt{\overline{CPUE}_{r,sq,c,cl}}}$$

Average square CPUE is the arithmetic mean of the observed CPUEs within a square.

Species and size redistribution of preys

In a few cases a prey item has been identified to species level but the length is not recorded. These items are redistributed on length groups proportionally to the observed length distribution within the species; first within the square and in cases of no “match” within the roundfish area.

A prey item can be partly digested such that species identification is impracticable. However, these items are often identified to a higher taxon (e.i. Gadidae). Such partly identified items belonging to the families (f) Gadidae, Clupeidae, Pleuronectidae and Soleidae were allocated proportionally to the species identified within the family. If a prey item had been assigned to a particular length class the redistribution was made over all family members in this length class. If the prey item was well digested and the length class is unknown, the redistribution was made over all identified family members and size classes. The redistribution was made within a predator size class and was first done using redistribution “keys” from the same ICES rectangle. If there was no key to a particular prey item, the redistribution was made using a redistribution key calculated for the entire roundfish area. In cases with no matches on roundfish level either, the prey item was classified as unknown. The manipulations can be summarised as follows:

1) Redistribute identified species without length using a species-length key by rectangle
2) Redistribute “unmatched” preys from 1) using a species-length key by roundfish area.
3) Classify unmatched remains from 2 as unknown species
4) Redistribute “family members” with length information using a family$_{lenght}$ -species key by rectangle
5) Redistribute unmatched preys from 4) using a key family$_{length}$-species by roundfish area
6) Redistribute family members without length information using a family-species key by rectangle
7) Redistribute unmatched preys from 6) using a family-species key by roundfish area
8) Classify unmatched remains from 5) and 7) as “Other food”

Prey items not identified to the mentioned families were not redistributed. That means that prey in the category “Unidentified fish” are not allocated to species and classified as unknown (Other food).

Age length transformations

All stomach contents observations are done by length classes of predator and preys and must be transformed to age before used in MSVPA. For each roundfish area, there exist a age-length-key (ALK) which gives the proportion of an age class within the length class.

For most species the ALK is constructed from the IBTS CPUE data. First, the average CPUE of a length class is calculated for each ICES rectangle as a simple mean of the available haul’s CPUE. The average CPUE per length class within a roundfish area is then calculated as the arithmetic mean of average CPUE per rectangle. The age distribution within combinations of roundfish area and length class is found from otoliths randomly sampled in strata.

To calculate the mean stomach content of a given predator age group (ca) within a roundfish area the following was done:

$$\overline{W}_{c,ca,p,pl} = \sum_{ca} \frac{ALK_{c,ca,cl}\overline{W}_{c,cl,p,pl}}{\sum\limits_{cl} ALK_{c,ca,cl}}$$

The age distribution of preys was calculated in a similar way as for the predator ages.

$$\overline{W}_{c,ca,p,pa} = \sum_{pl} \frac{ALK_{p,pa,pl}\overline{W}_{c,ca,p,pl}}{\sum\limits_{pa} ALK_{p,pa,pl}}$$

First the ALK by roundfish area was applied if such exist. If there were no ALK for a particularly prey in the roundfish area, an ALK calculated for the total North Sea was used.

For partly digested prey remains (Wrem) without length information, it was assumed that they have had an age and age distribution similar to just estimated age-distribution.

$$\overline{Wnew}_{c,ca,p,pa} = \frac{\overline{Wrem}_{c,ca,p}\overline{W}_{c,ca,p,pa}}{\sum\limits_{pa} \overline{W}_{c,ca,p,pa}}$$

Calculation of average North Sea stomach contents

Average stomach contents of a predator for a combination of roundfish areas are calculated as the mean of the roundfish average value, weighted by the predator abundance and spatial extent of the roundfish area.

Figures

Figure 1. Division of the North Sea into two sub-areas.

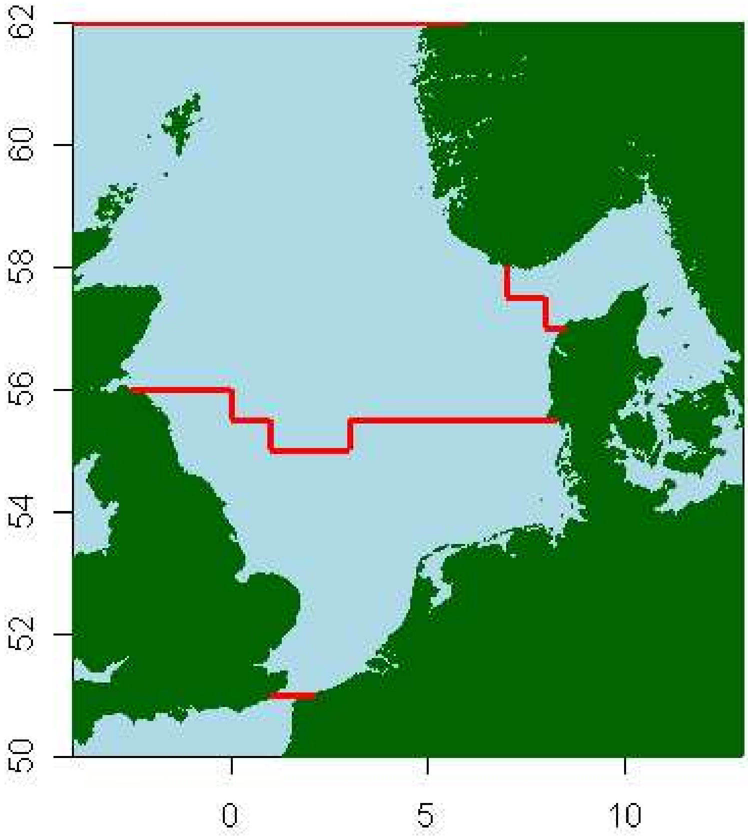
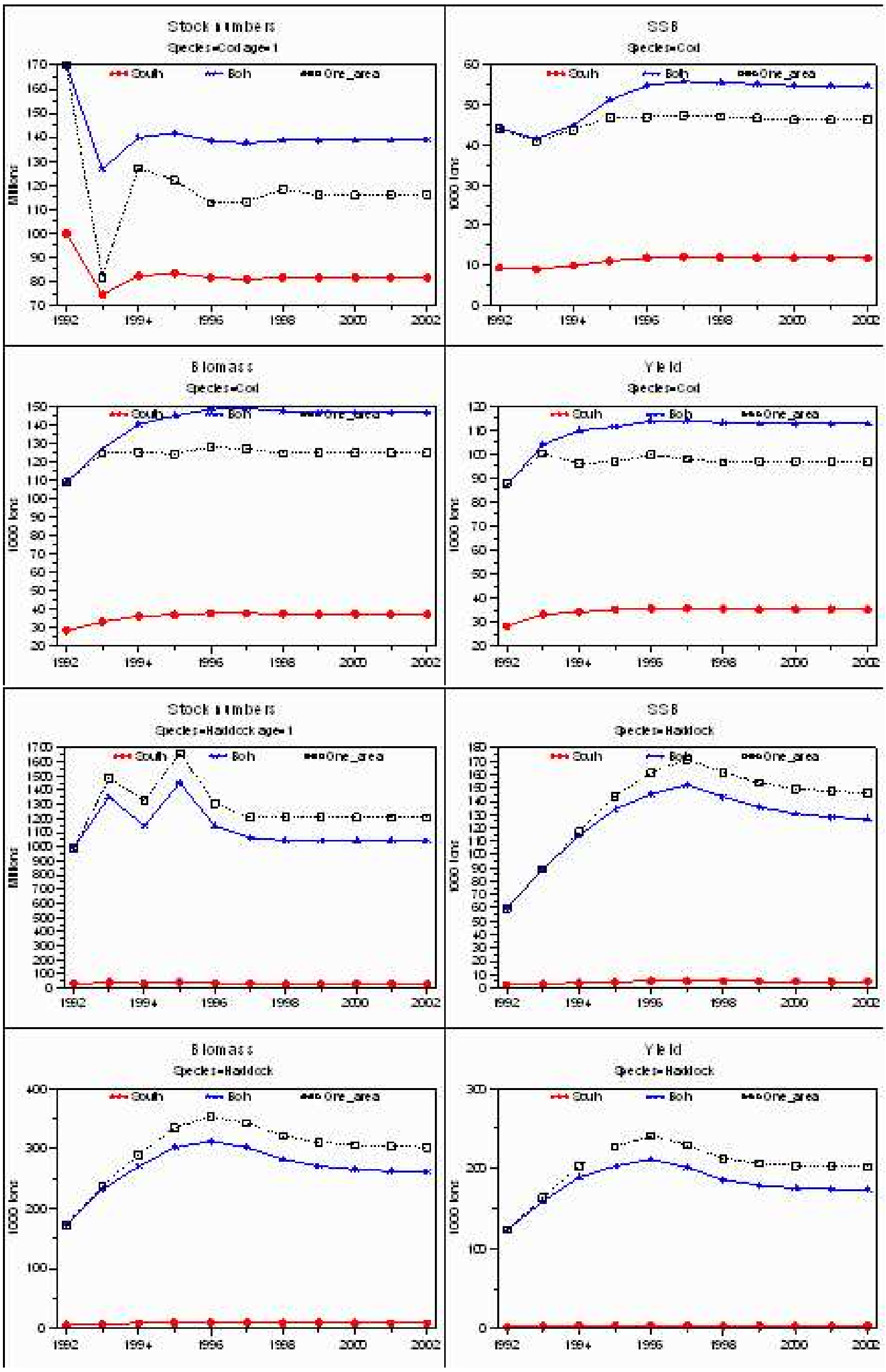
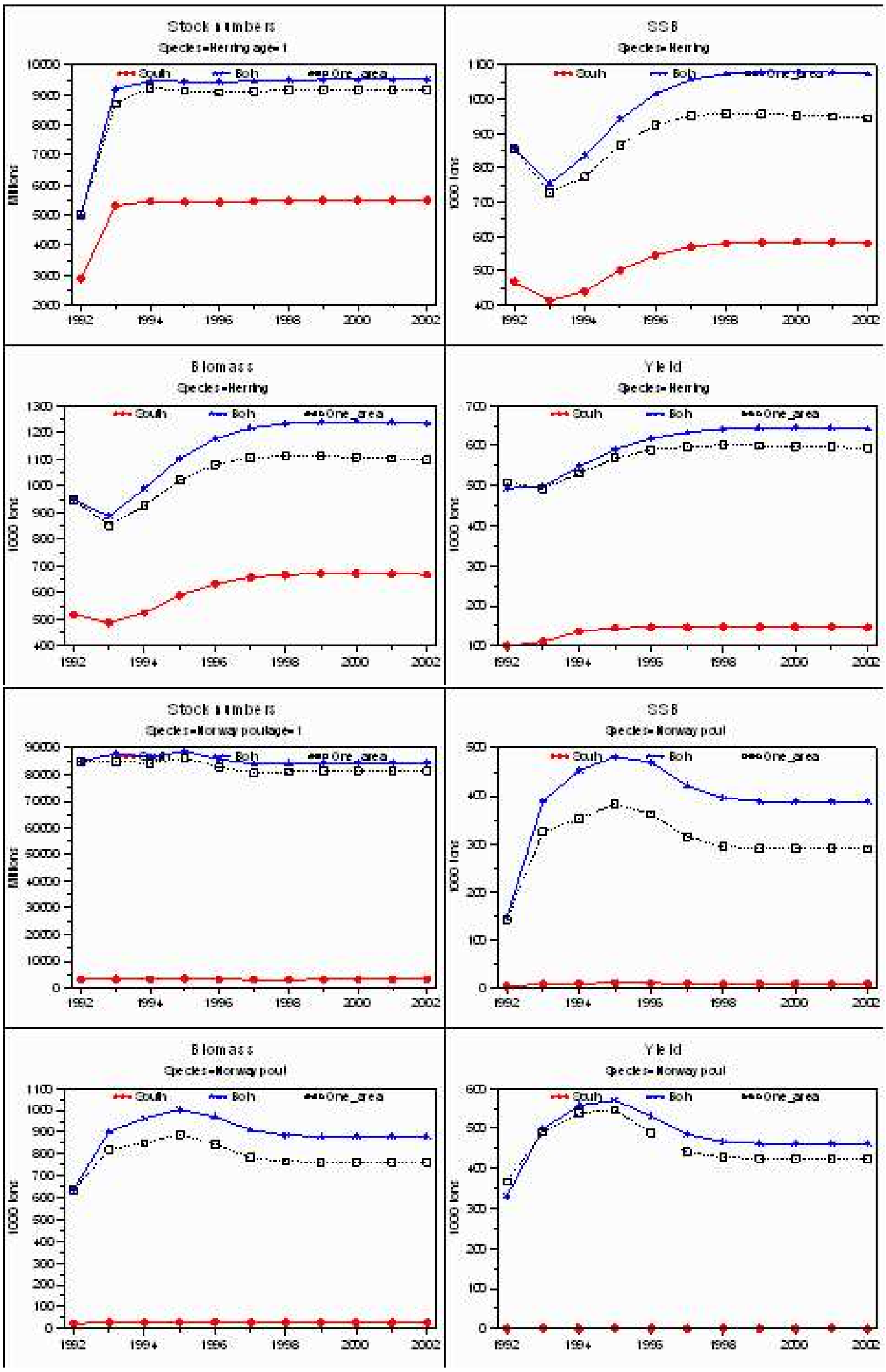
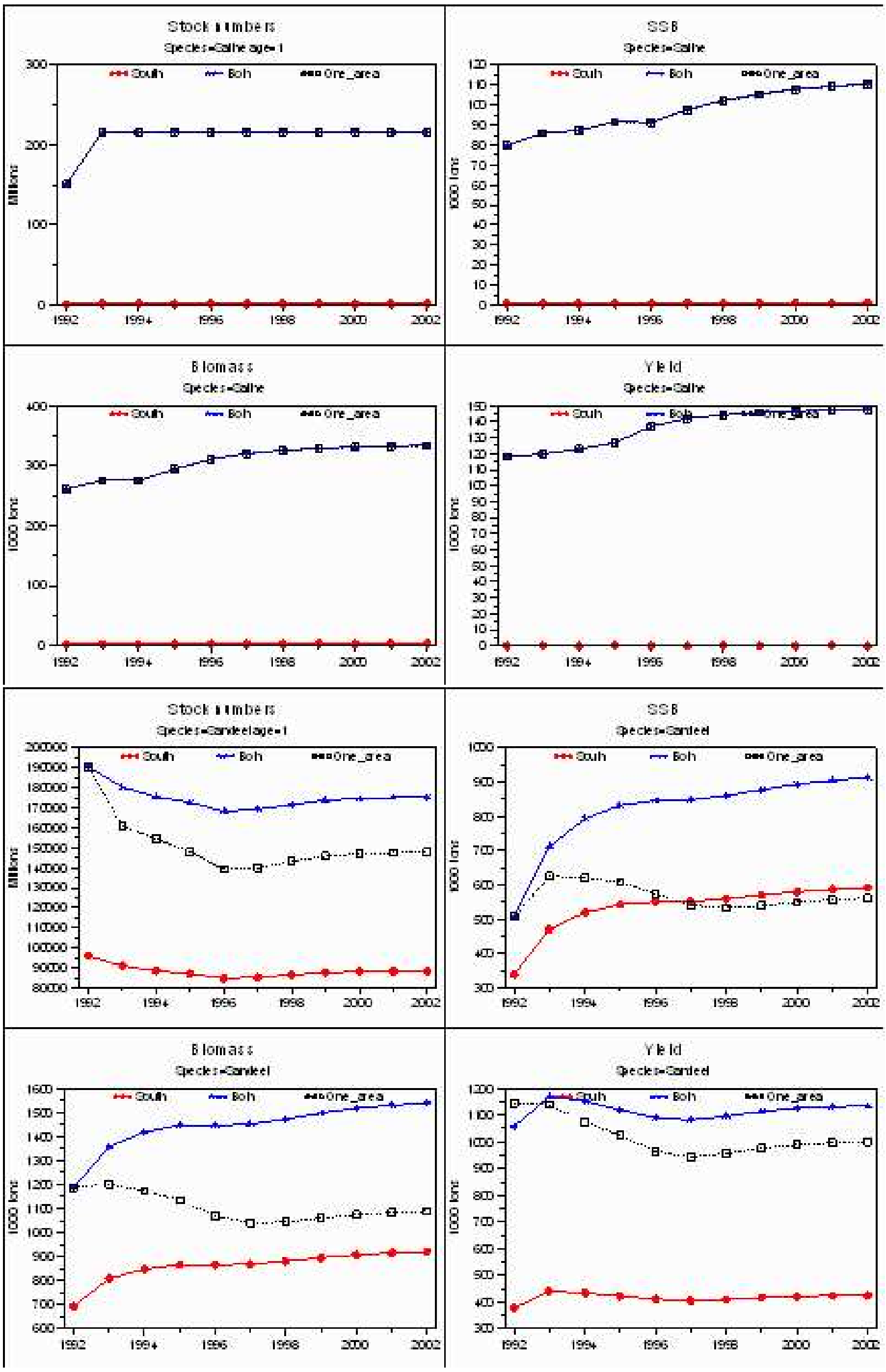


Figure 2. Scenario 1. Predicted stock numbers,SSB, TSB and yield for the southern North Sea (South), the southern+northern North Sea (Both) and the baseline (One area).




Stock numbers
Species=Herring age=1
Millions
South
Both
One_area
SSB
Species=Herring
1000 tons
Biomass
Species=Herring
Yield
Species=Herring
Stock numbers
Species=Norway poutage=1
SSB
Species=Norway pout
Biomass
Species=Norway pout
Yield
Species=Norway pout
1992
1994
1996
1998
2000
2002


Stock numbers
Species=Saithe age=1
South
Both
One_area
Millions
SSB
Species=Saithe
1000 tons
Biomass
Species=Saithe
Yield
Species=Saithe
Stock numbers
Species=Sandeel age=1
SSB
Species=Sandeel
Biomass
Species=Sandeel
Yield
Species=Sandeel
1992
1994
1996
1998
2000
2002

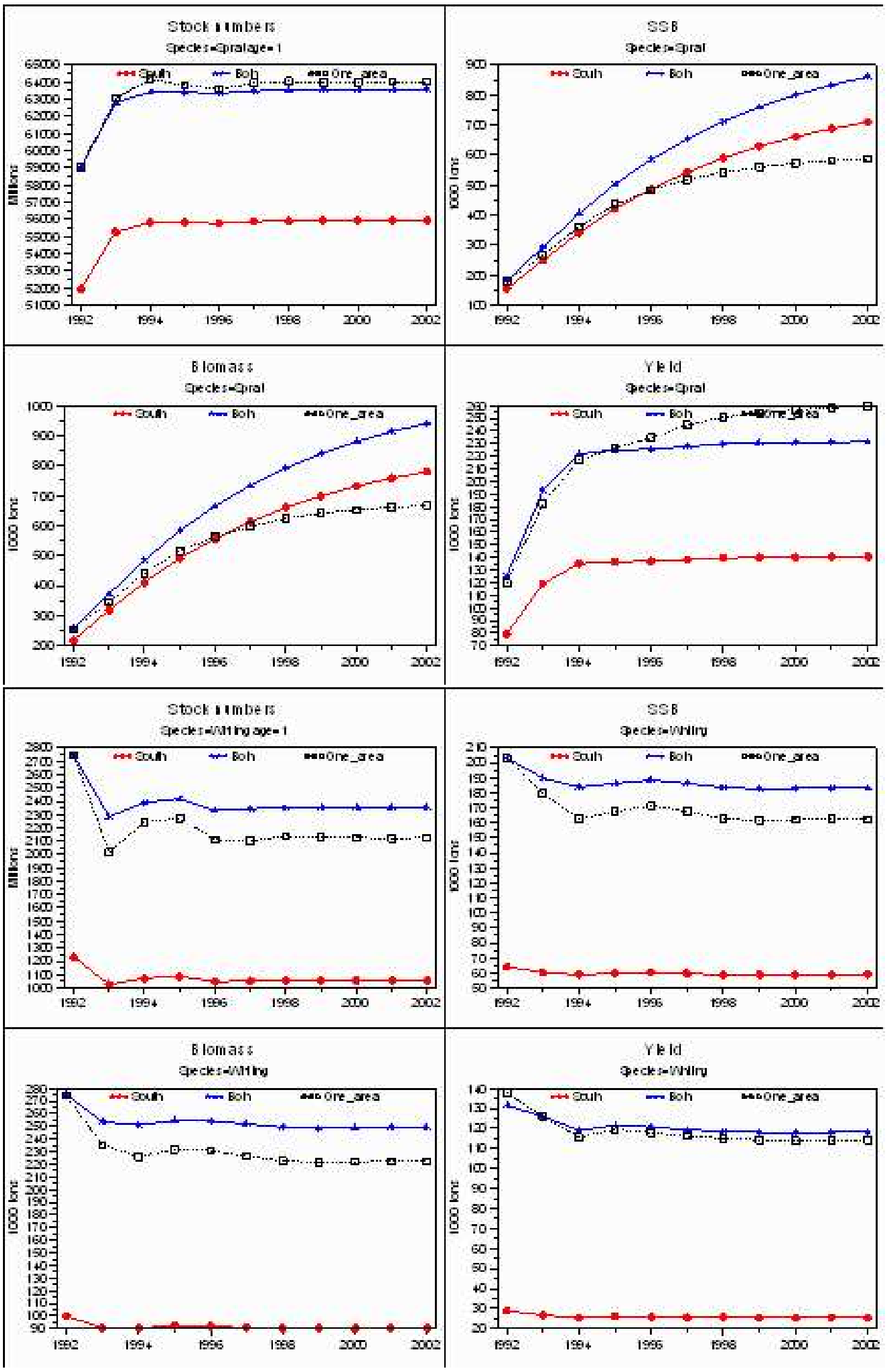
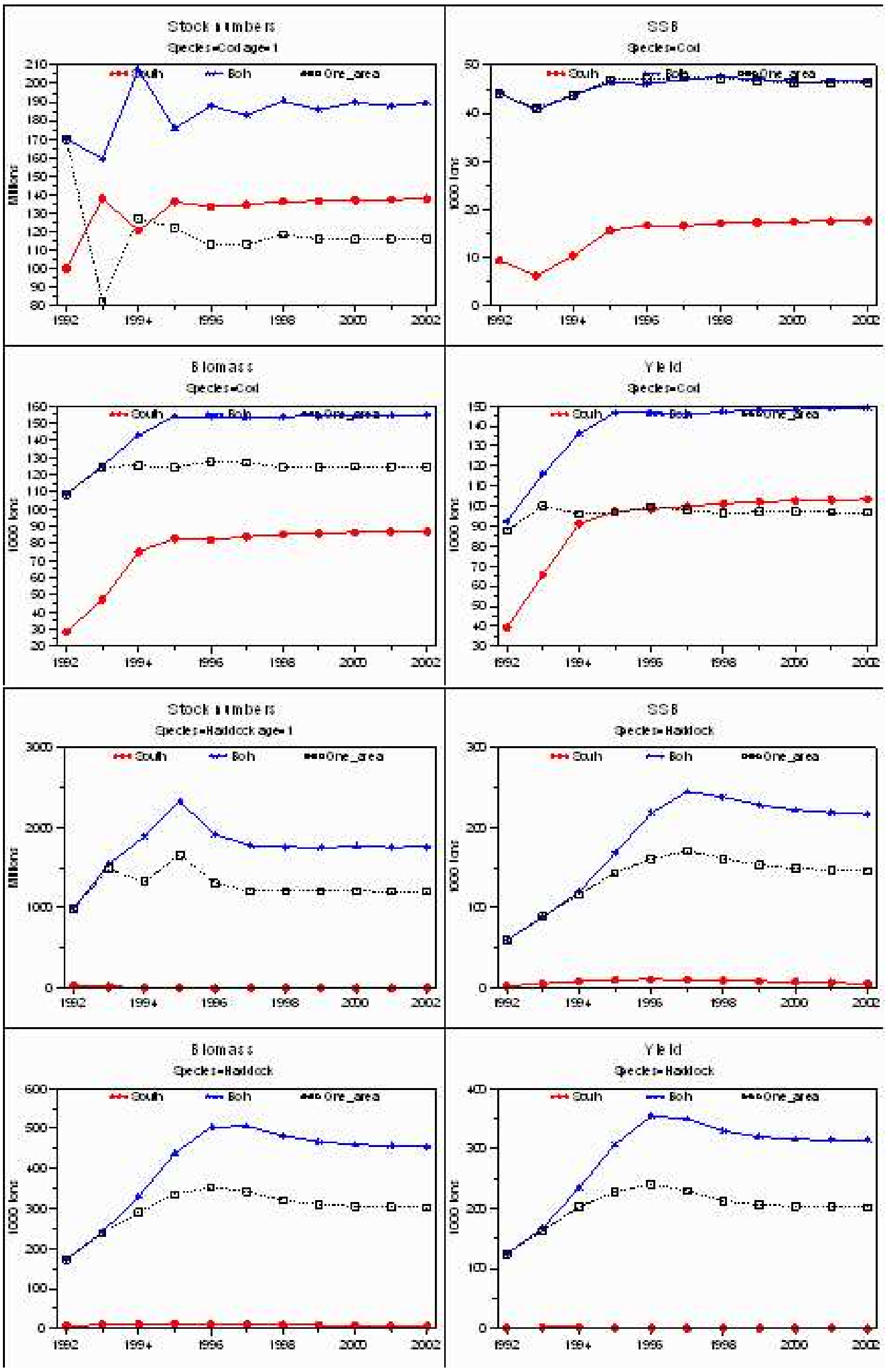
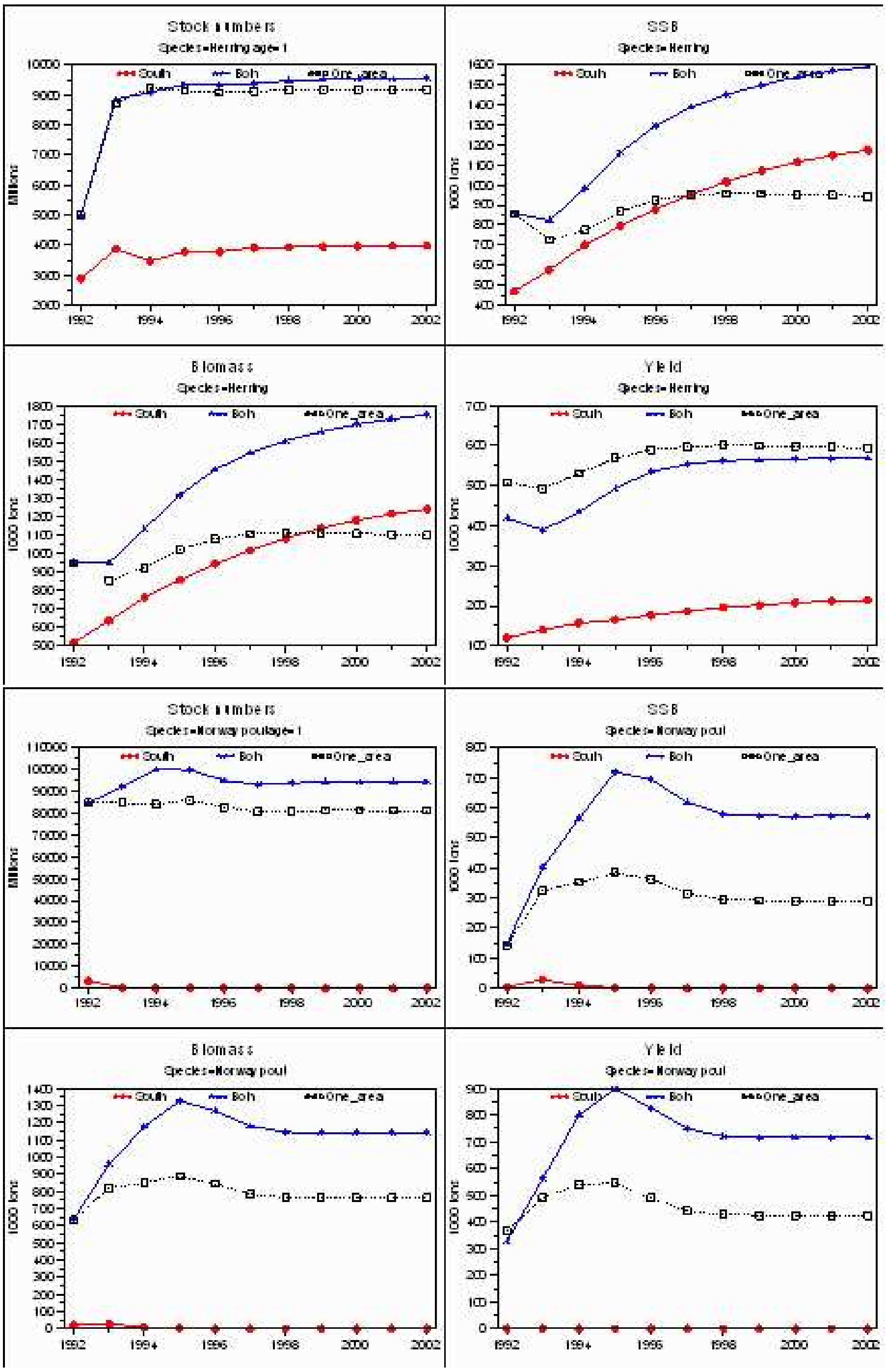
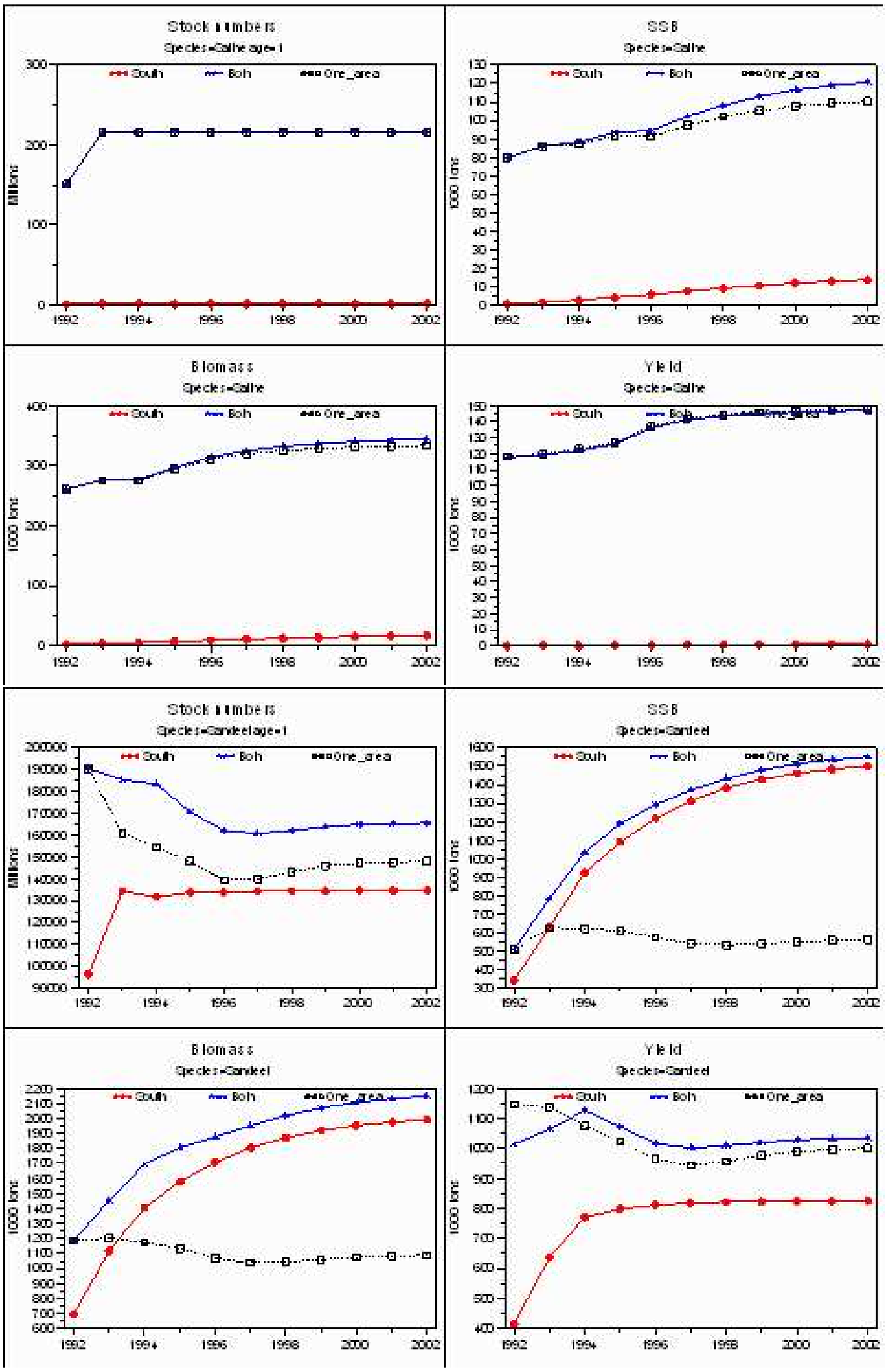

Stock numbers
Species=Sprat age=1
South
Both
One_area
Millions
SSB
Species=Sprat
1000 tons
Biomass
Species=Sprat
Yield
Species=Sprat
Stock numbers
Species=Whiting age=1
SSB
Species=Whiting
Biomass
Species=Whiting
Yield
Species=Whiting
1992
1994
1996
1998
2000
2002

Figure 3. Scenario 2. Predicted stock numbers,SSB, TSB and yield for the southern North Sea (South), the southern+northern North Sea (Both) and the baseline (One area).




Stock numbers
Species=Herring age=1
South
Both
One_area
Millions
10000
9000
8000
7000
6000
5000
4000
3000
2000
1992
1994
1996
1998
2000
2002
SSB
Species=Herring
1000 tons
1600
1500
1400
1300
1200
1100
1000
900
800
700
600
500
400
Biomass
Species=Herring
1800
1700
1600
1500
1400
1300
1200
1100
1000
900
800
700
600
500
Yield
Species=Herring
700
600
500
400
300
200
100
Stock numbers
Species=Norway poutage=1
110000
100000
90000
80000
70000
60000
50000
40000
30000
20000
10000
0
SSB
Species=Norway pout
800
700
600
500
400
300
200
100
0
Biomass
Species=Norway pout
1400
1300
1200
1100
1000
900
800
700
600
500
400
300
200
100
0
Yield
Species=Norway pout
900
800
700
600
500
400
300
200
100
0


Stock numbers
Species=Saithe age=1
South
Both
One_area
Millions
0
100
200
300
1992
1994
1996
1998
2000
2002
SSB
Species=Saithe
1000 tons
0
10
20
30
40
50
60
70
80
90
100
110
120
130
Biomass
Species=Saithe
1000 tons
0
100
200
300
400
Yield
Species=Saithe
1000 tons
0
150
Stock numbers
Species=Sandeel age=1
Millions
90000
100000
110000
120000
130000
140000
150000
160000
170000
180000
190000
200000
SSB
Species=Sandeel
1000 tons
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
Biomass
Species=Sandeel
1000 tons
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
2100
2200
Yield
Species=Sandeel
1000 tons
400
500
600
700
800
900
1000
1100
1200

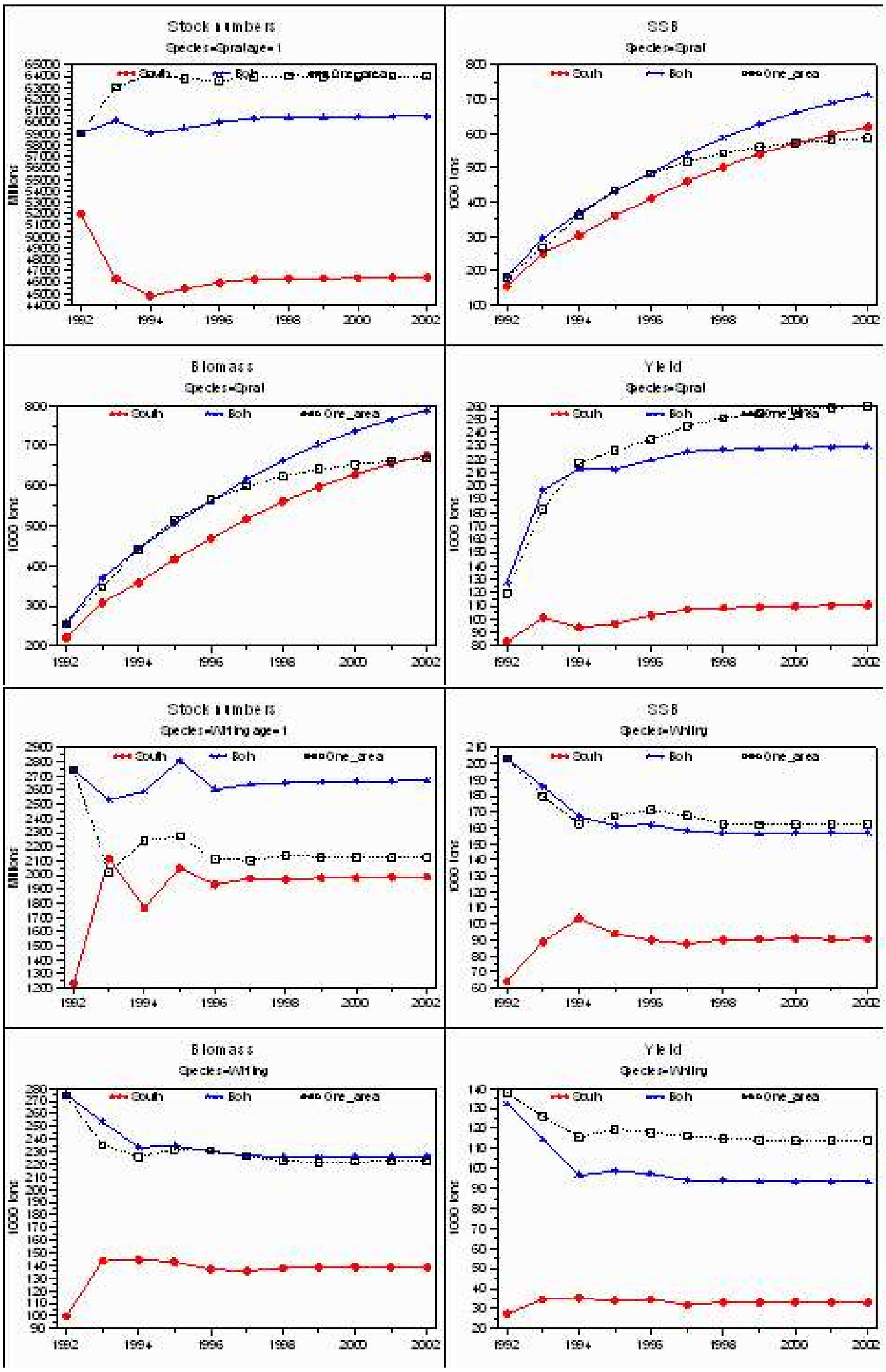
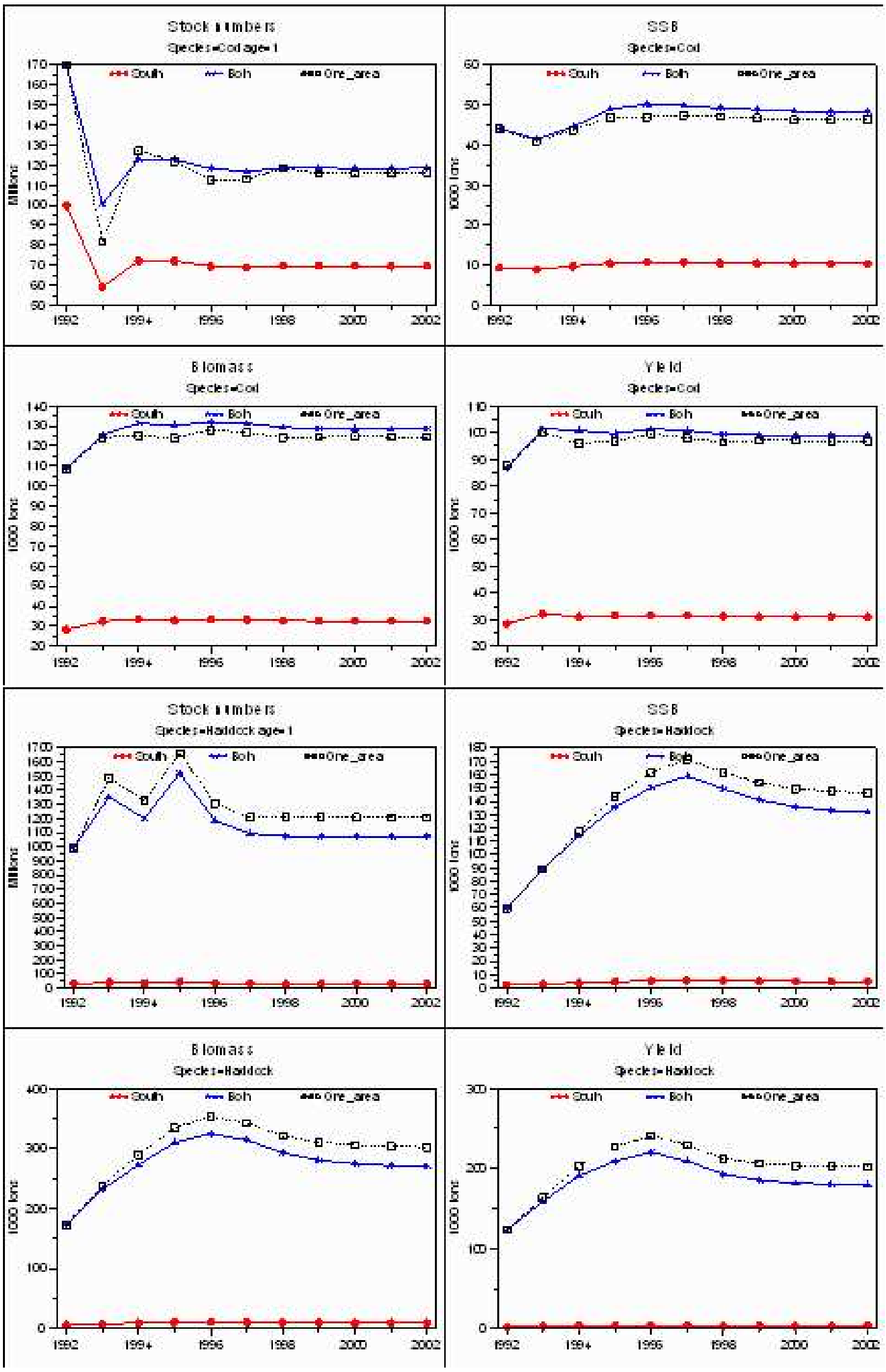
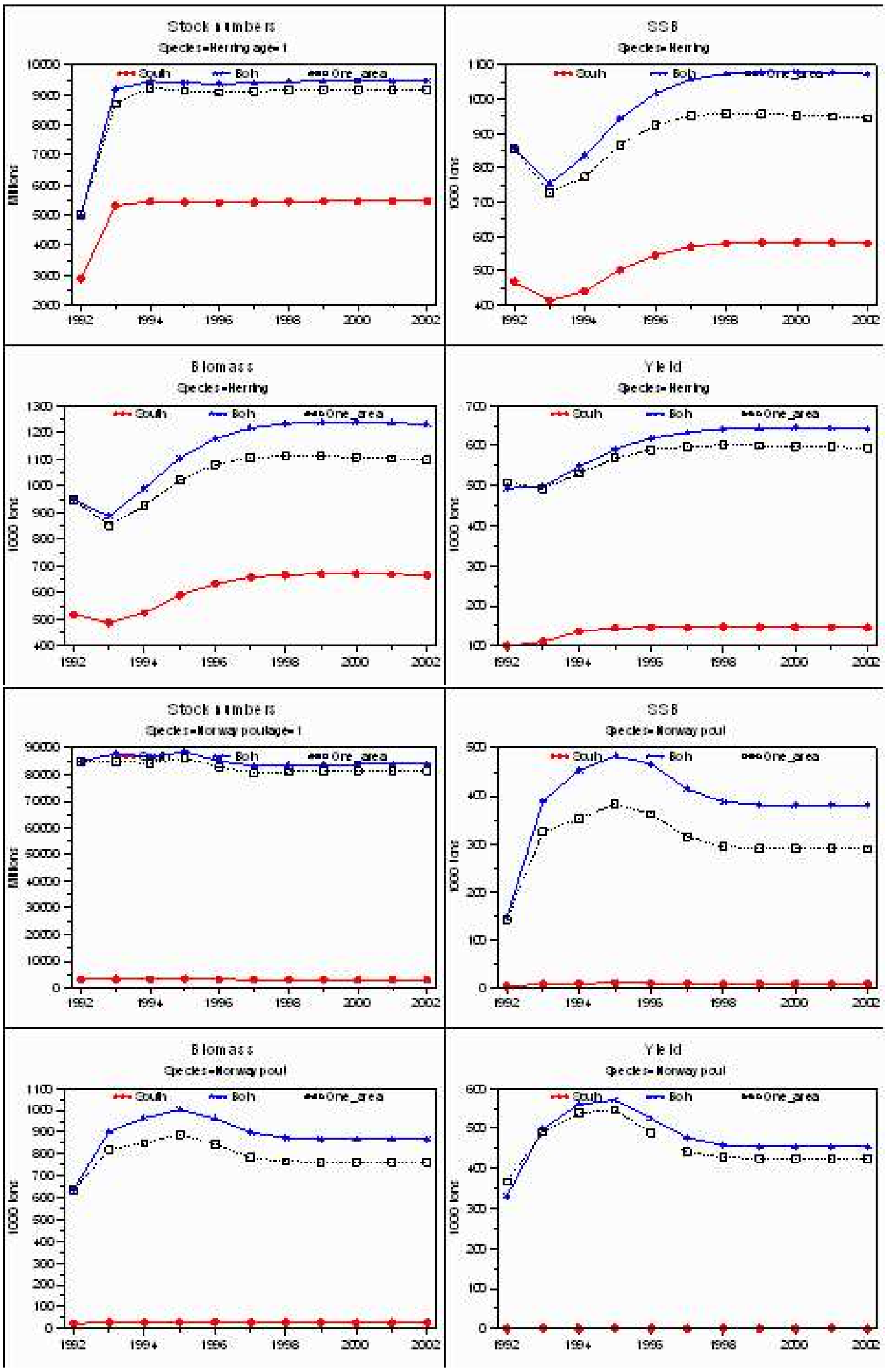
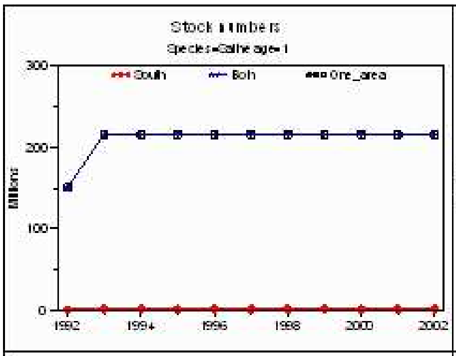
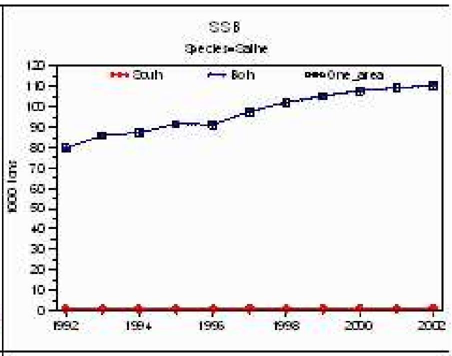

Stock numbers
Species=Sprat age=1
Millions
South
Both
One_area
SSB
Species=Sprat
1000 tons
Biomass
Species=Sprat
Yield
Species=Sprat
Stock numbers
Species=Whiting age=1
SSB
Species=Whiting
Biomass
Species=Whiting
Yield
Species=Whiting
1992
1994
1996
1998
2000
2002

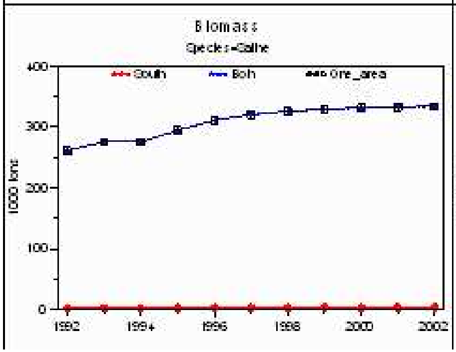
Figure 4. Scenario 3. Predicted stock numbers,SSB, TSB and yield for the southern North Sea (South), the southern+northern North Sea (Both) and the baseline (One area).

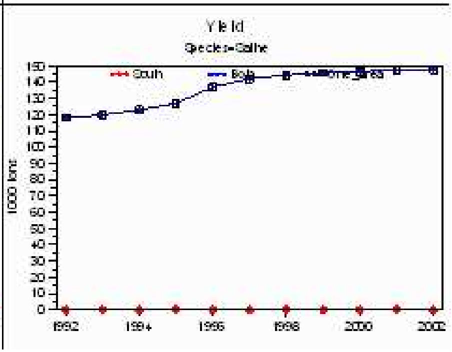


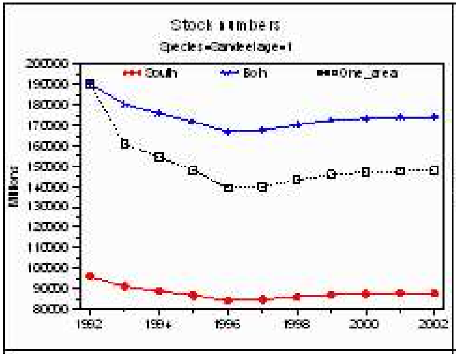

Stock numbers
Species=Herring age=1
Millions
SSB
Species=Herring
1000 tons
Biomass
Species=Herring
Yield
Species=Herring
Stock numbers
Species=Norway poutage=1
SSB
Species=Norway pout
Biomass
Species=Norway pout
Yield
Species=Norway pout
South
Both
One_area
1992
1994
1996
1998
2000
2002

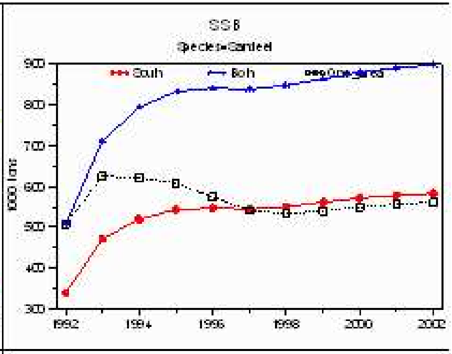

Stock numbers
Species=Saithe age=1
South
Both
One_area
Millions
0
100
200
300
1992
1994
1996
1998
2000
2002

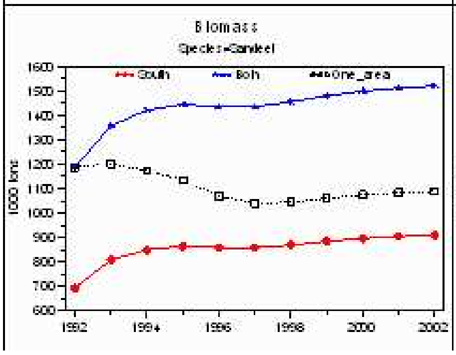

SSB
Species=Saithe
South
Both
One_area
1000 tons
0
10
20
30
40
50
60
70
80
90
100
110
120
1992
1994
1996
1998
2000
2002

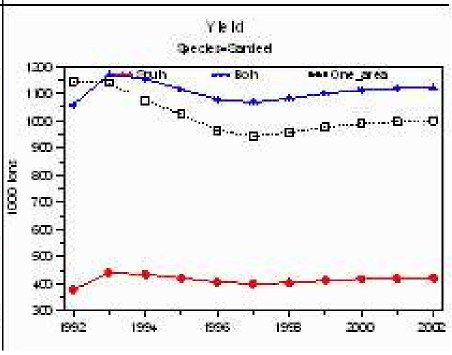

Biomass
Species=Saithe
South
Both
One_area
1000 tons
0
100
200
300
400
1992
1994
1996
1998
2000
2002


Yield
Species=Saithe
South
Both
One_area
1000 tons
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
1992
1994
1996
1998
2000
2002


Stock numbers
Species=Sandeel age=1
South
Both
One_area
Millions
80000
90000
100000
110000
120000
130000
140000
150000
160000
170000
180000
190000
200000
1992
1994
1996
1998
2000
2002


SSB
Species=Sandeel
South
Both
One_area
1000 tons
300
400
500
600
700
800
900
1992
1994
1996
1998
2000
2002


Biomass
Species=Sandeel
South
Both
One_area
1000 tons
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1992
1994
1996
1998
2000
2002


Yield
Species=Sandeel
South
Both
One_area
1000 tons
300
400
500
600
700
800
900
1000
1100
1200
1992
1994
1996
1998
2000
2002

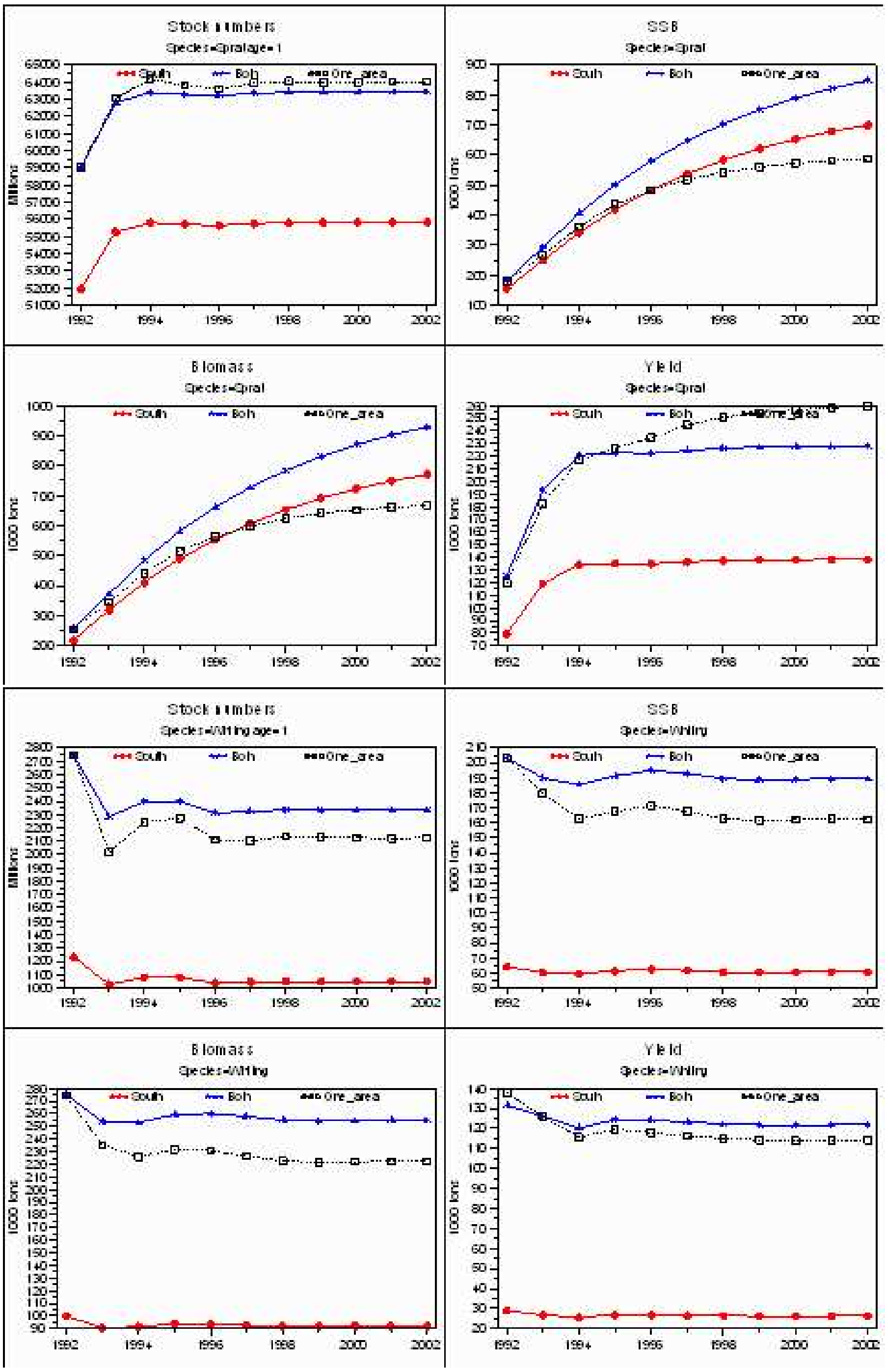
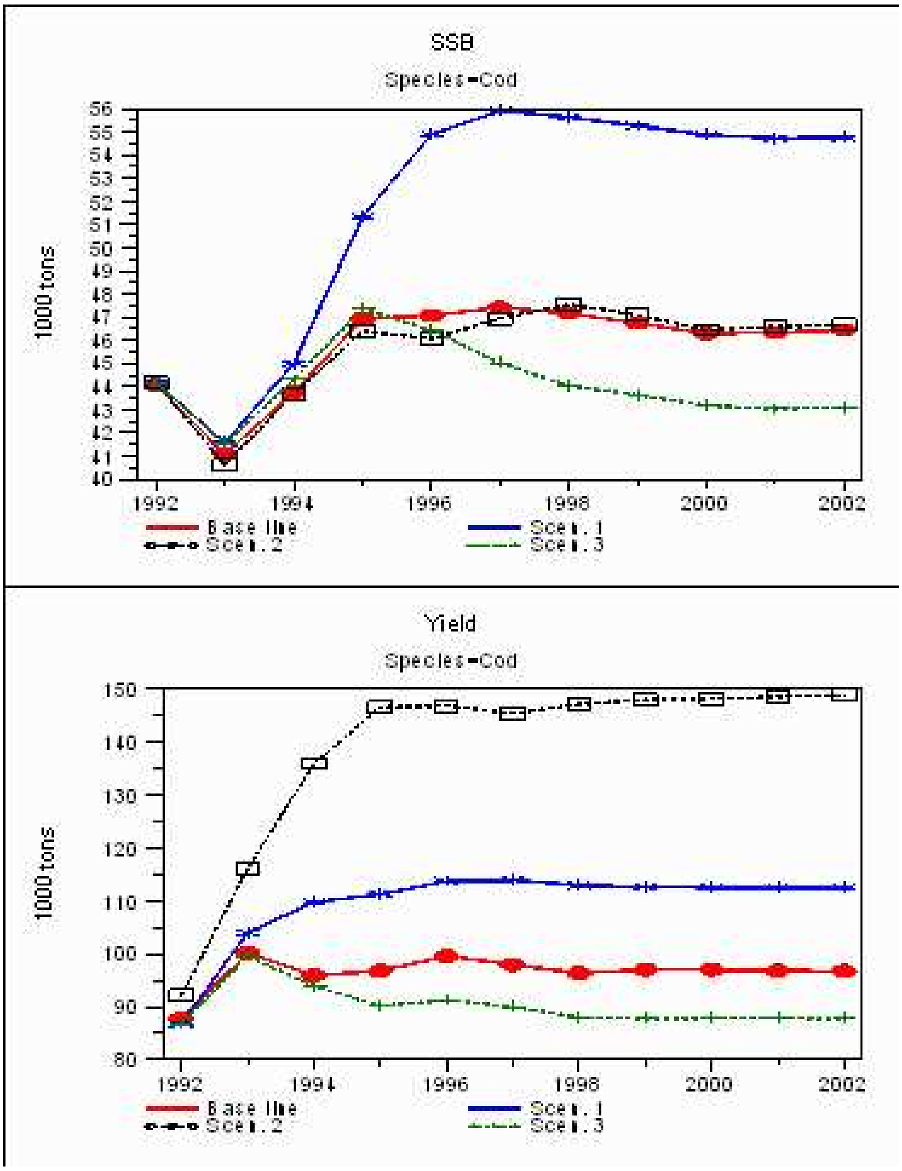
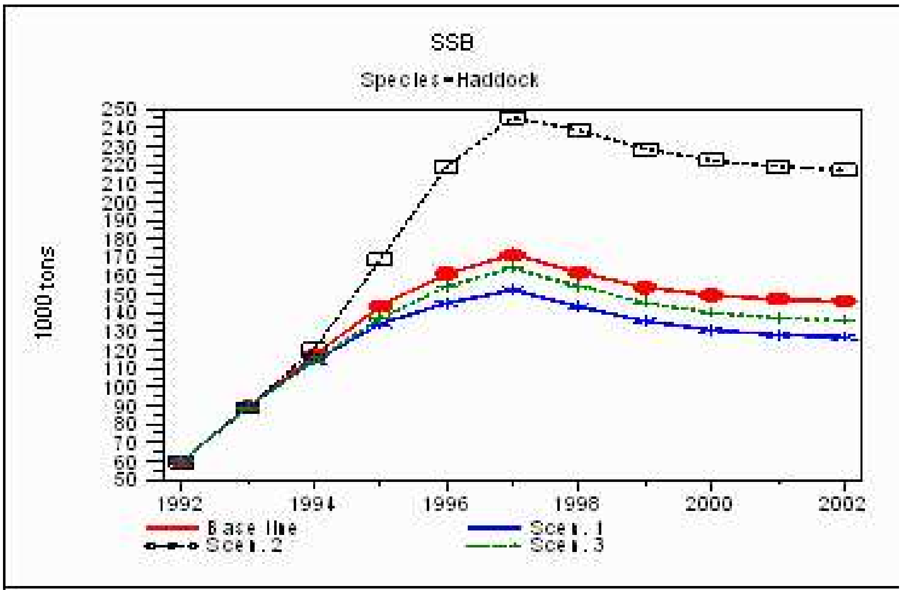
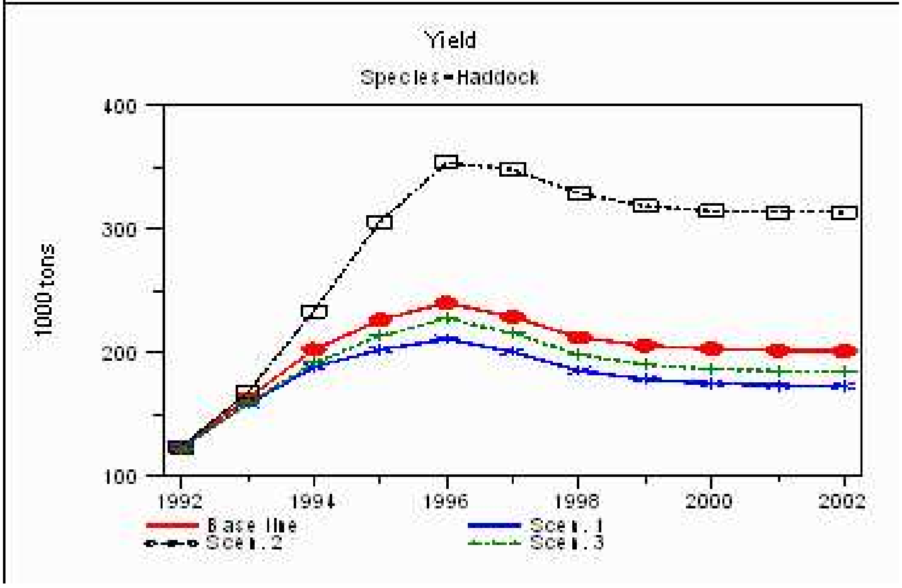
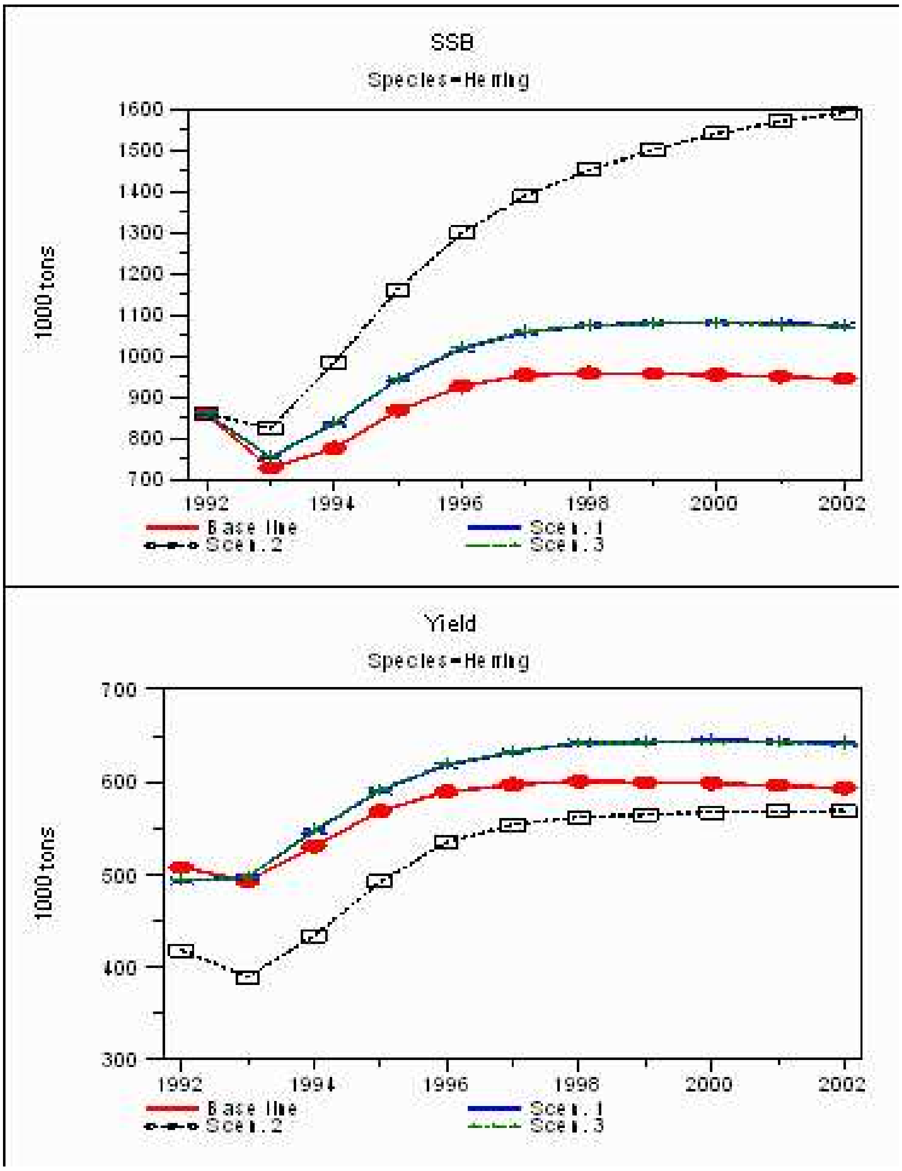

Stock numbers
Species=Sprat age=1
Millions
SSB
Species=Sprat
Biomass
Species=Sprat
Yield
Species=Sprat
Stock numbers
Species=Whiting age=1
SSB
Species=Whiting
Biomass
Species=Whiting
Yield
Species=Whiting
1000 tons
South
Both
One_area
1992
1994
1996
1998
2000
2002

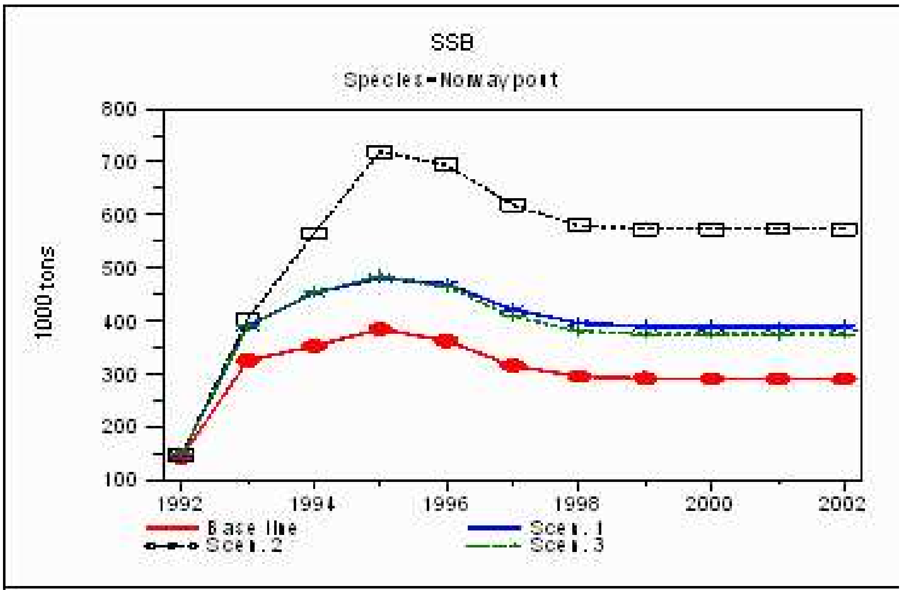
Figure 5. Comparison of predictions of spatial homogeneous (baseline)/ inhomogeneous (scenarios 1-3) stocks in the North Sea.

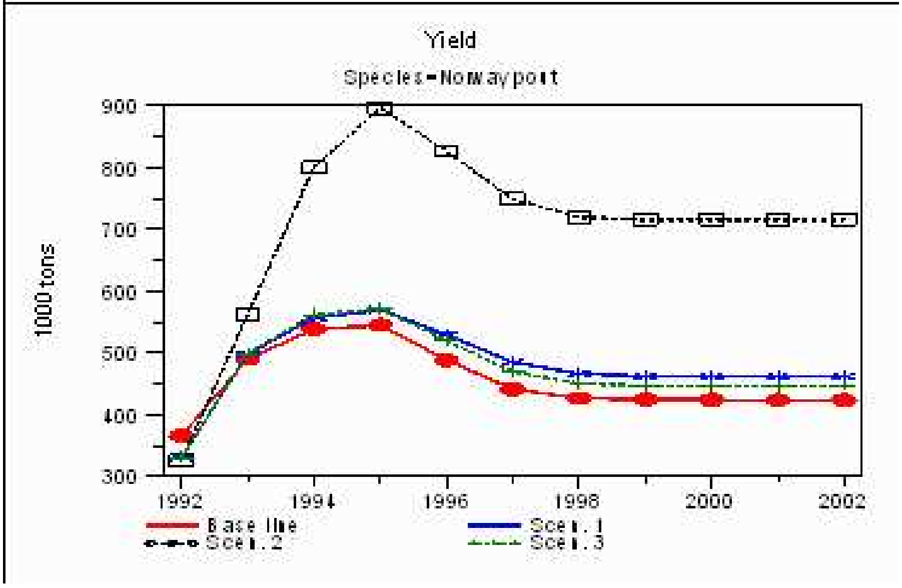


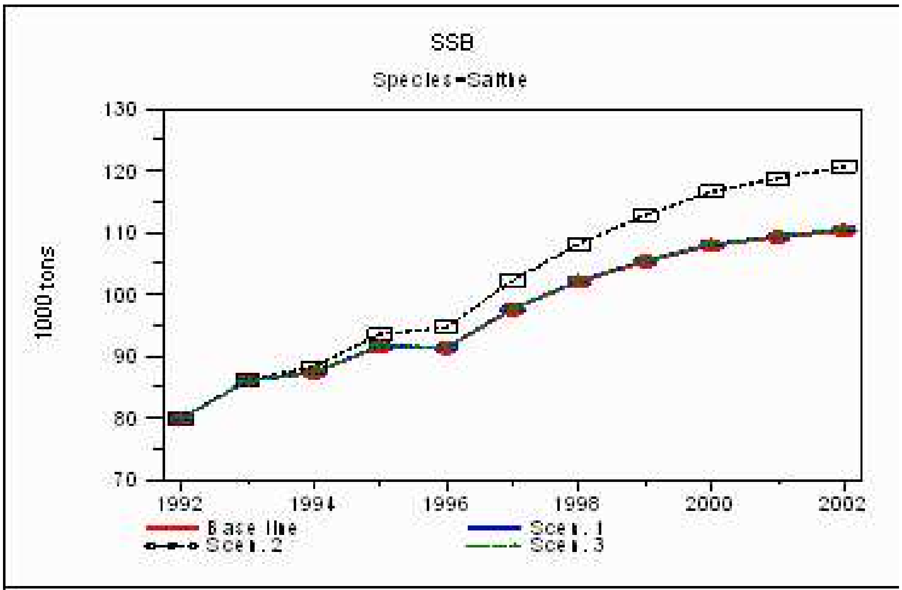

SSB
Species=Haddock
1000 tons
250
240
230
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
1992
1994
1996
1998
2000
2002
Base line
Scen.1
Scen.2
Scen.3

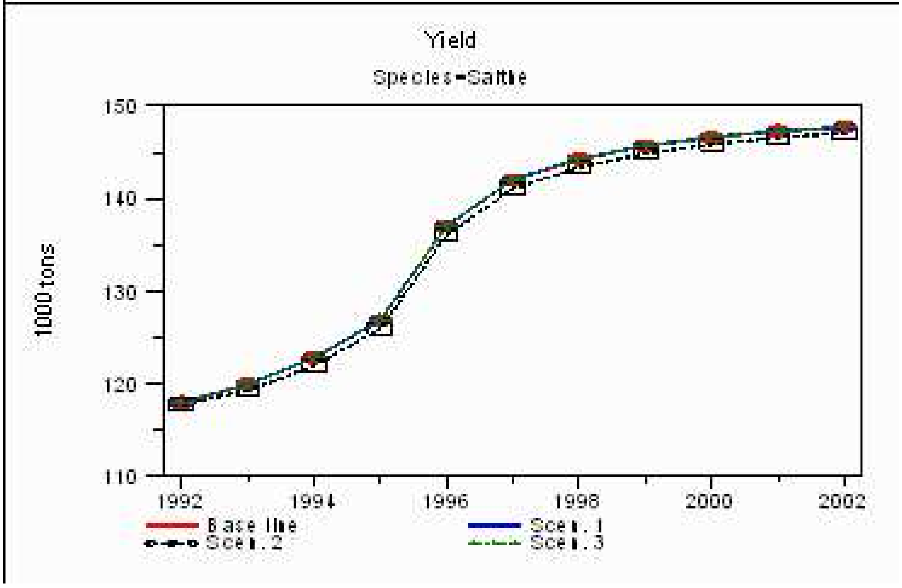

Yield
Species=Haddock
1000 tons
400
300
200
100
1992
1994
1996
1998
2000
2002
Base line
Scen.1
Scen.2
Scen.3

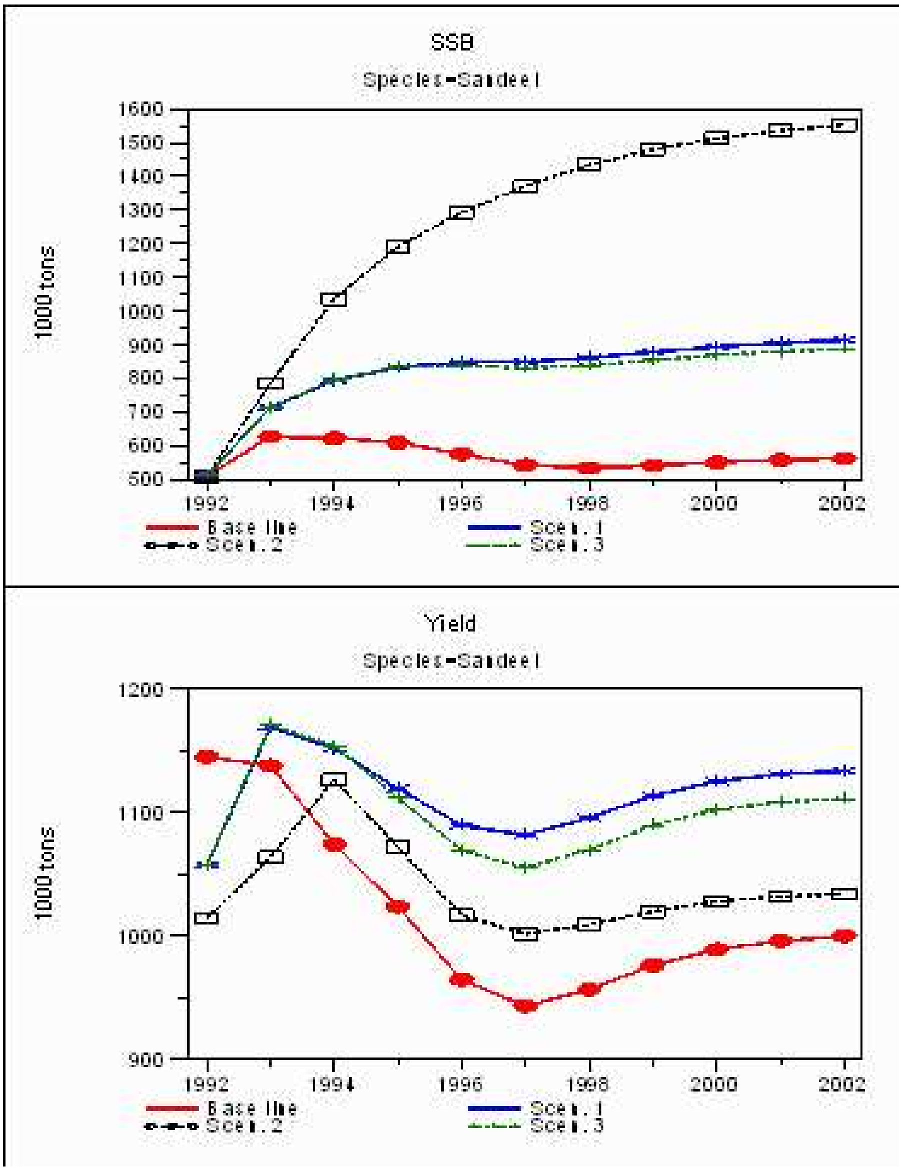

SSB
Species=Herring
1000 tons
1600
1500
1400
1300
1200
1100
1000
900
800
700
1992
1994
1996
1998
2000
2002
Base line
Scen.1
Scen.2
Scen.3
Yield
Species=Herring
1000 tons
700
600
500
400
300
1992
1994
1996
1998
2000
2002
Base line
Scen.1
Scen.2
Scen.3

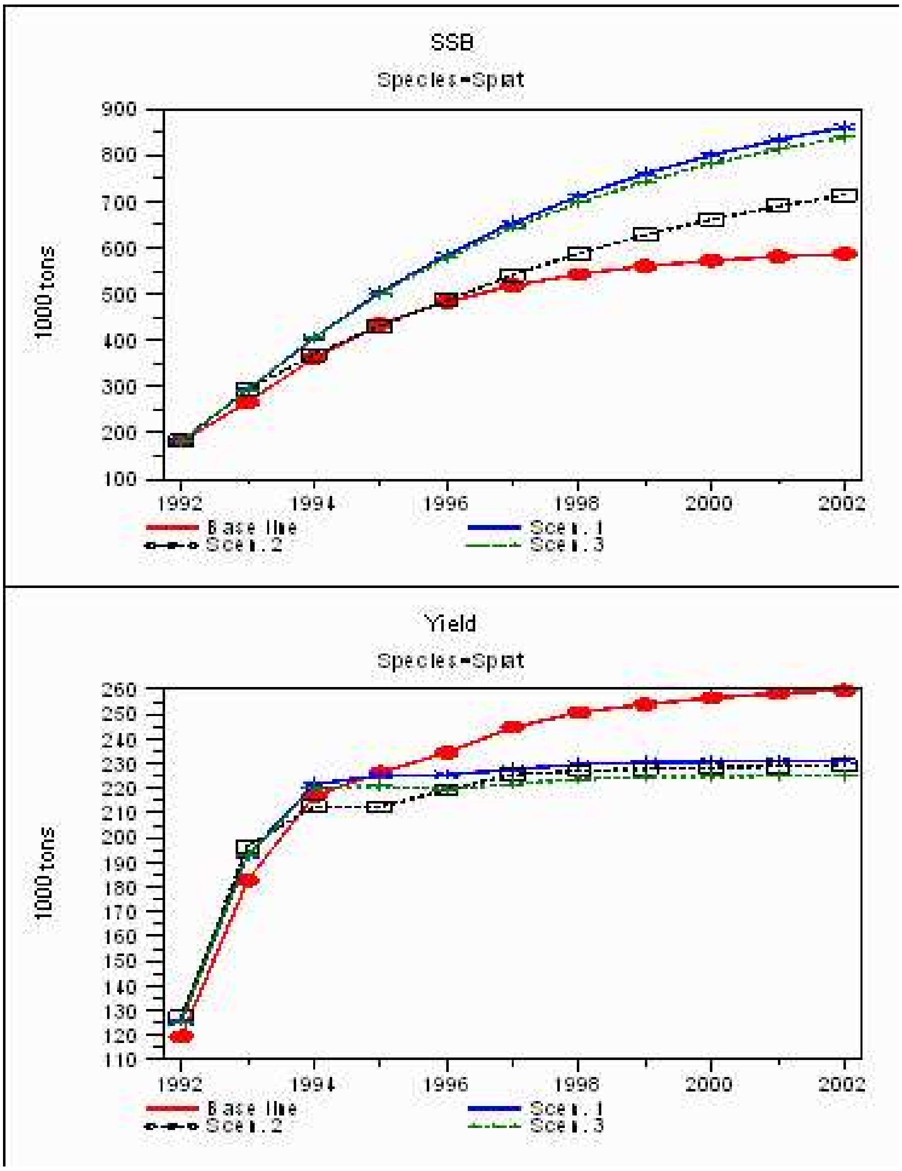

SSB
Species=Norway pout
1000 tons
800
700
600
500
400
300
200
100
1992
1994
1996
1998
2000
2002
Base line
Scen. 1
Scen. 2
Scen. 3

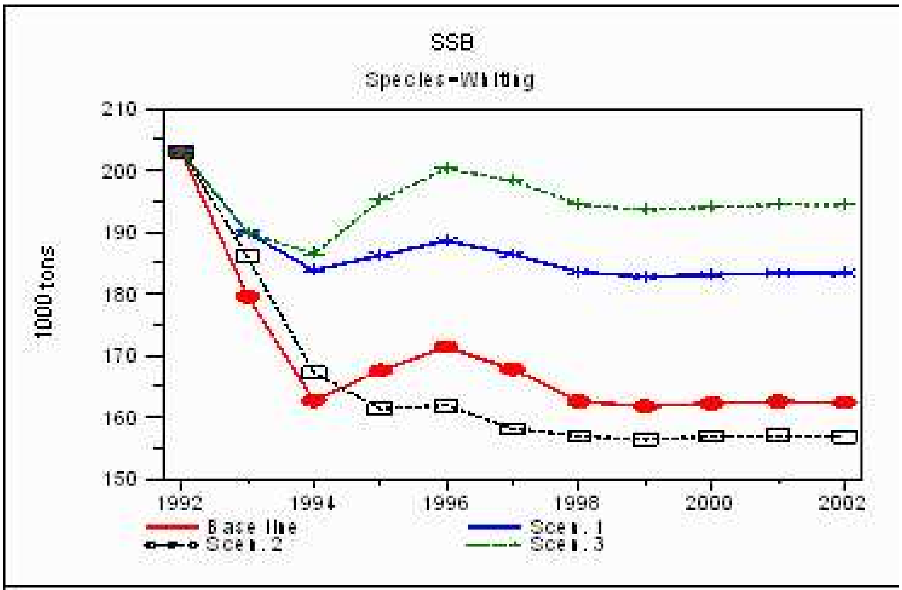

Yield
Species=Norway pout
1000 tons
900
800
700
600
500
400
300
1992
1994
1996
1998
2000
2002
Base line
Scen. 1
Scen. 2
Scen. 3

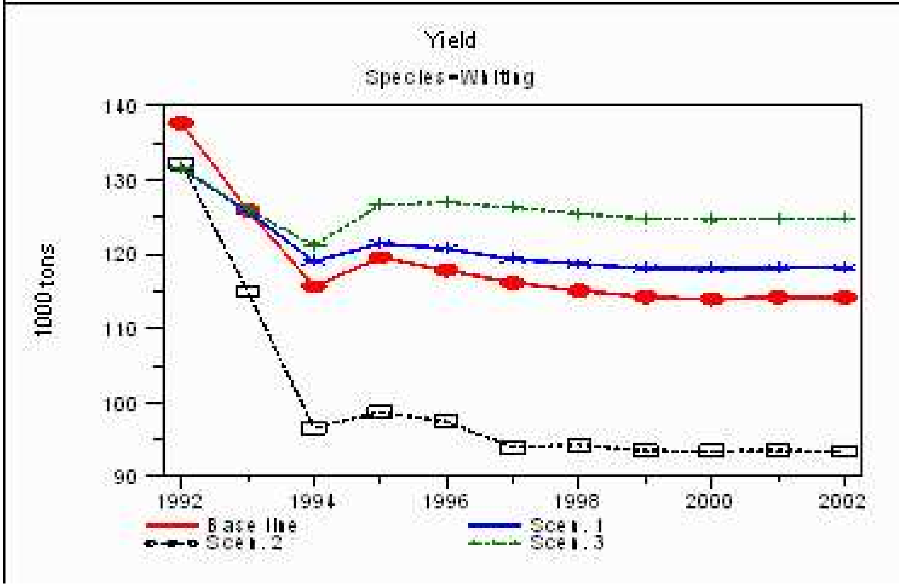

SSB
Species=Saithe
1000 tons
130
120
110
100
90
80
70
1992
1994
1996
1998
2000
2002
Base line
Scen.1
Scen.2
Scen.3


Yield
Species=Saithe
1000 tons
150
140
130
120
110
1992
1994
1996
1998
2000
2002
Base line
Scen.1
Scen.2
Scen.3


SSB
Species=Sandeel
1000 tons
1600
1500
1400
1300
1200
1100
1000
900
800
700
600
500
1992
1994
1996
1998
2000
2002
Base line
Scen.1
Scen.2
Scen.3
Yield
Species=Sandeel
1000 tons
1200
1100
1000
900
1992
1994
1996
1998
2000
2002
Base line
Scen.1
Scen.2
Scen.3


SSB
Species=Sprat
1000 tons
900
800
700
600
500
400
300
200
100
1992
1994
1996
1998
2000
2002
Base line
Scen.1
Scen.2
Scen.3
Yield
Species=Sprat
1000 tons
260
250
240
230
220
210
200
190
180
170
160
150
140
130
120
110
1992
1994
1996
1998
2000
2002
Base line
Scen.1
Scen.2
Scen.3


SSB
Species=Whiting
1000 tons
210
200
190
180
170
160
150
1992
1994
1996
1998
2000
2002
Base line
Scen. 1
Scen. 2
Scen. 3


Yield
Species=Whiting
1000 tons
140
130
120
110
100
90
1992
1994
1996
1998
2000
2002
Base line
Scen. 1
Scen. 2
Scen. 3

7.9 Control measures in relation to ecosystem considerations

Gunnar Stefansson

Under the auspices of the dst^2 project subtasks have dealt with the implications of multispecies models for fishery management. A popular approach to "ecosystem management" is to advocat the use of closed areas as a management tool. However, as noted e.g. in Stefansson (2003), and references therein, it is not at all clear what the effect of using closed areas will be on the fishing mortality for a typical finfish stock. This is further emphasized in Hilborn et al (2004) where it is pointed out that the use of closed areas (Marine Protected Areas, MPAs) needs to be tested just as rigorously as any other management procedures.

As a followup to these concerns, the work presented in Stefansson and Rosenberg (2004) compares several approaches to controlling the take from a resource, notably systems for limiting access through effort controls, quotas and MPAs. It is found that the effect of MPAs may be limited in terms of fishing mortality reductions (or long-term profits) unless very large areas are closed (e.g. more than 50% of the initial biomass is protected). It is also found that using closed areas may become a costly exercise since it will lead to higher expenses given the same income. However, the use of large MPAs may be beneficial in conjunction with or instead of a quota system, if the quota system does not give adequate protection.

Hilborn R., Stokes K., Maguire, J.-J. Smith, T., Botsford, L. W., Mangel M., Orensanz J., Parma A., Rice J., Bell J., Cochrane, K. L., Garcia S., Hall, S. J., Kirkwood, G. P., Sainsbury K., Stefansson G., and Walters C. 2004. When Can Marine Protected Areas Improve Fisheries Management? Ocean & Coastal Management. 47 (3-4): 197-205

Stefansson, G. 2003. Multi-species and ecosystem models in a management context. In Sinclair and Valdimarsson. FAO Conference Proceedings on Fisheries in the Ecosystem.

Stefansson, G. and Rosenberg, A. A. 2005. Combining control measures for managing fisheries under uncertainty: quotas, effort limitation and protected areas. Royal Soc Phil Trans B. Accepted.

7.10 Further published/in press papers

Studying spatial and trophic interactions between capelin and cod using individual-based modelling

Geir Huse, Geir Odd Johansen, Bjarte Bogstad and Harald Gjøsæter

Abstract: The objective of this study was to use spatially explicit individual-based models for simulating the movement, foraging, growth and mortality of cod and capelin in the Barents Sea in order to identify general features in their migration patterns and cod's consumption of capelin. The individual-based models are initiated from survey data, run over one year, and validated against survey information. Directed movement is based on a combination of movement vectors and temperature boundaries, and bioenergetics models are used to calculate growth. Capelin consumption by cod is calculated from local encounters between the species. For capelin, the best movement model can be summarised as: stay southwest of the 2.5^o and 4^o C temperature front at 50 m depth for juvenile and mature individuals respectively in winter, and migrate northwards during summer, but do not pass the -1.5^o C temperature front. The best cod model was to migrate south-southwest during winter and north-northeast during summer, within the temperature range 1 to 8^o C. The annual consumption estimates found here reflected the inter-annual and seasonal pattern from previous studies based on stomach samples, but were generally lower. Consumption estimates varied depending on the movement models, and the best movement model also produced the consumption estimate closest to that obtained in other studies. Introducing a simple rule stating that cod should move in a randomly selected direction when the local capelin density is zero increased the consumption estimate by 30%. This suggests that more emphasis needs to be put on exploring how behavioural rules in predators and prey affects their interactions. Even though there are some discrepancies between predictions and observations, the results achieved by the model with regard to spatial distribution, growth and consumption are promising.

Huse, G., Johansen, G.O., Bogstad, B., and Gjøsæter, H. 2004. Studying spatial and trophic interactions between capelin and cod using individual-based modelling. ICES Journal of Marine Science 61: 1201-1213.

Consumption of juvenile herring (Clupea harengus) by cod (Gadus morhua) in the Barents Sea: a new approach to estimating consumption in piscivorous fish.

Johansen, G. O., B. Bogstad, S. Mehl, and Ø. Ulltang.

Abstract: Consumption of different age groups of juvenile, Norwegian, spring-spawning herring (Clupea harengus) by northeast Arctic cod (Gadus morhua) in the Barents Sea in 1992-1997 is estimated using cod stomach content data. We present a new approach to the problem of estimating consumption by fish. The new method is based on the estimation of digestion time for single prey items based on the difference between fresh weight at ingestion and weight in the stomach at time of sampling. Estimation is based on a gastric evacuation model for cod and area-specific sea temperatures. This is used to estimate the time (tmax) it takes for a prey to become digested to a stage where length is no longer measurable. Predation rate is then estimated for all prey with digestion time <tmax as number of prey eaten in the time range defined by tmax. This rate is combined with estimates of the proportion of the cod stock consuming the prey and area-specific abundance of cod, giving consumption of herring on a seasonal and yearly basis. The consumption estimates differ from those obtained using current methods. Predation mortality of herring is estimated directly from the consumption estimates by combining them with acoustic herring abundance data.

Johansen, G. O., B. Bogstad, S. Mehl, and Ø. Ulltang. 2004. Consumption of juvenile herring (Clupea harengus) by cod (Gadus morhua) in the Barents Sea: a new approach to estimating consumption in piscivorous fish. Canadian Journal of Fisheries and Aquatic Sciences 61:343-359.

Biological reference points for fish stocks in a multispecies context

Jeremy S. Collie and Henrik Gislason

Abstract: Biological reference points (BRPs) are widely used to define safe levels of harvesting for marine fish populations. Most BRPs are either minimum acceptable biomass levels or maximum fishing mortality rates. The values of BRPs are determined from historical abundance data and the life-history parameters of the fish species. However, when the life-history parameters change over time, the BRPs become moving targets. In particular, the natural mortality rate of prey species depends on predator levels; conversely, predator growth rates depend on prey availability. We tested a suite of BRPs for their robustness to observed changes in natural mortality and growth rates. We used the relatively simple Baltic Sea fish community for this sensitivity test, with cod as predator and sprat and herring as prey. In general, the BRPs were much more sensitive to the changes in natural mortality rates than to growth variation. For a prey species like sprat, fishing mortality reference levels should be conditioned on the level of predation mortality. For a predator species, a conservative level of fishing mortality can be identified that will prevent growth overfishing and ensure stock replacement. These first-order multispecies interactions should be considered when defining BRPs for medium-term (5-10 year) management decisions.

Collie, J. S. and Gislason, H. 2001. Biological reference points for fish stocks in a multispecies context Can. J. Fish. Aquat. Sci./J. Can. Sci. Halieut. Aquat. 58(11): 2167-2176 (2001)

Using AMOEBAs to display multispecies, multifleet fisheries advice

Jeremy S. Collie, Henrik Gislason, and Morten Vinther

Abstract: In multispecies fish communities, predation levels change dynamically in response to changes in the abundance of predator and prey species, as influenced by the fisheries that exploit them. In addition to community-level metrics, it remains necessary to track the abundance of each species relative to its biological reference point. In situations with many interacting species, exploited by multiple fishing fleets, it can be complicated to illustrate how the effort of each fleet will affect the abundance of each species. We have adapted the AMOEBA approach to graph the reference levels of multiple interacting species exploited by multiple fleets. This method is illustrated with 10 species and eight fishing fleets in the North Sea. We fit a relatively simple response-surface model to the predictions of a fully age-structured multispecies model. The response-surface model links the AMOEBA for fishing effort to separate AMOEBAs for spawning stock biomass, fishing mortality, and yield. Ordination is used to give the shape of the AMOEBAs functional meaning by relating fish species to the fleets that catch them. The aim is to present the results of dynamic multispecies models in a format that can be readily understood by decision makers. Interactive versions of the AMOEBAs can be used to identify desirable combinations of effort levels and to test the compatibility of the set of single-species biological reference points.

Collie, J. S., Gislason, H. and Vinther, V. Using AMOEBAs to display multispecies, multifleet fisheries advice. ICES Journal of Marine Science, 60: 709-720.

Estimation of fish consumption by stomach content analysis

A. Rindorf* and P. Lewy

*University of Copenhagen, c/o Danish Institute for Fisheries Research, Charlottenlund Castle, DK2920 Charlottenlund, Denmark.

Tel: +45 33 96 33 59; Fax: +45 33 96 33 33; email: ar@dfu.min.dk Abstract

Abstract: This study presents an analysis of the bias introduced by using simplified methods to calculate food intake from stomach contents. A frequently adopted approximation is the method suggested by Pennington (1985. Dana, vol. vol. 5, pp. 81-86) rather than the numerical solution of the underlying differential equation. This procedure provided results close to actual intake rate in a simulation study. In contrast to this, serious positive bias was introduced by calculating food intake from the contents of pooled stomach samples. An expression is given which can be used to correct analytically for this bias in the cases where the contents of the non-empty stomachs follow a gamma or log-normal distribution. A new method is suggested for the estimation of the intake of separate prey types. The method takes account of the distribution and evacuation of individual prey types as well as the effect of other food in the stomach on evacuation. A comparison between the new method and traditionally used methods revealed high bias in the traditional estimates.

Rindorf, A. and Lewy, P. submitted. Estimation of fish consumption by stomach content analysis

Estimation of evacuation rates in the field

A. Rindorf*

Danish Institute for Fisheries Research, Charlottenlund Castle, DK2920 Charlottenlund, Denmark.

* Tel: +45 33 96 33 59; Fax: +45 33 96 33 33; email: ar@dfu.min.dk

Abstract: This study presents two methods to calculate evacuation rates based on observed diel changes in occurrence and mean weight of prey in predator stomachs. The methods do not require predators to exhibit prolonged non-feeding periods, but the ingestion of each particular prey type must be restricted to certain diel periods. Data from more than 7500 whiting collected at 5 locations in the North Sea are used to demonstrate the methods. The evacuation rates estimated from field data are similar to laboratory results, though a tendency for estimates to exceed literature values slightly is noted. Bias is introduced if a large proportion of the prey is evacuated completely in the interval between subsequent samples and if significant amounts of other food are present in the stomach together with the prey in question. In conclusion, the methods can be used to supplement laboratory estimates of evacuation rates or provide first estimates for species that are not easily maintained in the laboratory.

Rindorf, A. 2004. Estimation of evacuation rates in the field. Journal of Fish Biology 65(1): 262–281

Hafrannsóknastofnun. Fjölrit
Marine Research Institute. Reports

Þessi listi er einnig á Netinu
(This list is also on the Internet)

http://www.hafro.is/Bokasafn/Timarit/fjolr.htm

1. **Kjartan Thors, Þórdís Ólafsdóttir:** Skýrsla um leit að byggingarefnum í sjó við Austfirði sumarið 1975. Reykjavík 1975. 62 s. (Ófáanlegt - Out of print).
2. **Kjartan Thors:** Skýrsla um rannsóknir hafsbotnsins í sunnanverðum Faxaflóa sumarið 1975. Reykjavík 1977. 24 s.
3. **Karl Gunnarsson, Konráð Þórisson:** Áhrif skolpmengunar á fjöruþörunga í nágrenni Reykjavíkur. Reykjavík 1977. 19 s. (Ófáanlegt - *Out of print*).
4. **Einar Jónsson:** Meingunarrannsóknir í Skerjafirði. Áhrif frárennslis á botndýralíf. Reykjavík 1976. 26 s. (Ófáanlegt - *Out of print*).
5. **Karl Gunnarsson, Konráð Þórisson:** Stórþari á Breiðafirði. Reykjavík 1979. 53 s.
6. **Karl Gunnarsson:** Rannsóknir á hrossaþara (*Laminaria digitata*) á Breiðafirði. 1. Hrossaþari við Fagurey. Reykjavík 1980. 17 s. (Ófáanlegt - *Out of print*).
7. **Einar Jónsson:** Líffræðiathuganir á beitusmokk haustið 1979. Áfangaskýrsla. Reykjavík 1980. 22 s. (Ófáanlegt - *Out of print*).
8. **Kjartan Thors:** Botngerð á nokkrum hrygningarstöðvum síldarinnar. Reykjavík 1981. 25 s. (Ófáanlegt - *Out of print*).
9. **Stefán S. Kristmannsson:** Hitastig, selta og vatns- og seltubúskapur í Hvalfirði 1947-1978. Reykjavík 1983. 27 s.
10. **Jón Ólafsson:** Þungmálmar í kræklingi við Suðvestur-land. Reykjavík 1983. 50 s.
11. Nytjastofnar sjávar og umhverfisþættir 1987. Aflahorfur 1988. *State of Marine Stocks and Environmental Conditions in Icelandic Waters 1987. Fishing Prospects 1988*. Reykjavík 1987. 68 s. (Ófáanlegt - *Out of print*).
12. Haf- og fiskirannsóknir 1988-1992. Reykjavík 1988. 17 s. (Ófáanlegt - *Out of print*).
13. **Ólafur K. Pálsson, Björn Æ. Steinarsson, Einar Jónsson, Gunnar Jónsson, Gunnar Stefánsson, Sigfús A. Schopka:** Stofnmæling botnfiska á Íslandsmiðum. Reykjavík 1988. 76 s. (Ófáanlegt - *Out of print*).
14. Nytjastofnar sjávar og umhverfisþættir 1988. Aflahorfur 1989. *State of Marine Stocks and Environmental Conditions in Icelandic Waters 1988. Fishing Prospects 1989*. Reykjavík 1988. 126 s.
15. Ástand humar- og rækjustofna 1988. Aflahorfur 1989. Reykjavík 1988. 16 s.
16. **Kjartan Thors, Jóhann Helgason:** Jarðlög við Vestmannaeyjar. Áfangaskýrsla um jarðlagagreiningu og könnun neðansjávareldvarpa með endurvarpsmælingum. Reykjavík 1988. 41 s.
17. **Stefán S. Kristmannsson:** Sjávarhitamælingar við strendur Íslands 1987-1988. Reykjavík 1989. 102 s.
18. **Stefán S. Kristmannsson, Svend-Aage Malmberg, Jóhannes Briem:** *Western Iceland Sea. Greenland Sea Project. CTD Data Report. Joint Danish-Icelandic Cruise R/V Bjarni Sæmundsson, September 1987*. Reykjavík 1989. 181 s.
19. Nytjastofnar sjávar og umhverfisþættir 1989. Aflahorfur 1990. *State of Marine Stocks and Environmental Conditions in Icelandic Waters 1989. Fishing Prospects 1990*. Reykjavík 1989. 128 s. (Ófáanlegt - *Out of print*).
20. **Sigfús A. Schopka, Björn Æ. Steinarsson, Einar Jónsson, Gunnar Jónsson, Gunnar Stefánsson, Ólafur K. Pálsson:** Stofnmæling botnfiska á Íslandsmiðum 1989. Rannsóknaskýrsla. Reykjavík 1989. 54 s.
21. Nytjastofnar sjávar og umhverfisþættir 1990. Aflahorfur 1991. *State of Marine Stocks and Environmental Conditions in Icelandic Waters 1990. Fishing prospects 1991*. Reykjavík 1990. 145 s.
22. **Gunnar Jónsson, Björn Æ. Steinarsson, Einar Jónsson, Gunnar Stefánsson, Ólafur K. Pálsson, Sigfús A. Schopka:** Stofnmæling botnfiska á Íslandsmiðum 1990. Reykjavík 1990. 53 s. (Ófáanlegt - *Out of print*).
23. **Stefán S. Kristmannsson, Svend-Aage Malmberg, Jóhannes Briem, Erik Buch:** *Western Iceland Sea - Greenland Sea Project - CTD Data Report. Joint Danish Icelandic Cruise R/V Bjarni Sæmundsson, September 1988*. Reykjavík 1991. 84 s. (Ófáanlegt - *Out of print*).
24. **Stefán S. Kristmannsson:** Sjávarhitamælingar við strendur Íslands 1989-1990. Reykjavík 1991. 105 s. (Ófáanlegt - *Out of print*).
25. Nytjastofnar sjávar og umhverfisþættir 1991. Aflahorfur fiskveiðiárið 1991/92. *State of Marine Stocks and Environmental Conditions in Icelandic Waters 1991. Prospects for the Quota Year 1991/92*. Reykjavík 1991. 153 s. (Ófáanlegt - *Out of print*).
26. **Páll Reynisson, Hjálmar Vilhjálmsson:** Mælingar á stærð loðnustofnsins 1978-1991. Aðferðir og niðurstöður. Reykjavík 1991. 108 s.
27. **Stefán S. Kristmannsson, Svend-Aage Malmberg, Jóhannes Briem, Erik Buch:** *Western Iceland Sea - Greenland Sea Project - CTD Data Report. Joint Danish Icelandic Cruise R/V Bjarni Sæmundsson, September 1989. Reykjavík 1991*. Reykjavík 1991. 93 s.
28. **Gunnar Stefánsson, Björn Æ. Steinarsson, Einar Jónsson, Gunnar Jónsson, Ólafur K. Pálsson, Sigfús A. Schopka:** Stofnmæling botnfiska á Íslandsmiðum 1991. Rannsóknaskýrsla. Reykjavík 1991. 60 s.
29. Nytjastofnar sjávar og umhverfisþættir 1992. Aflahorfur fiskveiðiárið 1992/93. *State of Marine Stocks and Environmental Conditions in Icelandic Waters 1992. Prospects for*

the Quota Year 1992/93. Reykjavík 1992. 147 s. (Ófáanlegt - *Out of print*).

30. **Van Aken, Hendrik, Jóhannes Briem, Erik Buch, Stefán S. Kristmannsson, Svend-Aage Malmberg, Sven Ober:** *Western Iceland Sea. GSP Moored Current Meter Data Greenland - Jan Mayen and Denmark Strait September 1988 - September 1989*. Reykjavík 1992. 177 s.

31. **Björn Æ. Steinarsson, Einar Jónsson, Gunnar Jónsson, Gunnar Stefánsson, Ólafur K. Pálsson, Sigfús A. Schopka:** Stofnmæling botnfiska á Íslandsmiðum 1992. Reykjavík 1993. 71 s. (Ófáanlegt - *Out of print*).

32. **Guðrún Marteinsdóttir, Gunnar Jónsson, Ólafur V. Einarsson:** Útbreiðsla grálúðu við Vestur- og Norðvesturland 1992. Reykjavík 1993. 42 s. (Ófáanlegt - *Out of print*).

33. **Ingvar Hallgrímsson:** Rækjuleit á djúpslóð við Ísland. Reykjavík 1993. 63 s.

34. Nytjastofnar sjávar 1992/93. Aflahorfur fiskveiðiárið 1993/94. *State of Marine Stocks in Icelandic Waters 1992/93. Prospects for the Quota Year 1993/94*. Reykjavík 1993. 140 s.

35. **Ólafur K. Pálsson, Björn Æ. Steinarsson, Einar Jónsson, Gunnar Jónsson, Gunnar Stefánsson, Sigfús A. Schopka:** Stofnmæling botnfiska á Íslandsmiðum 1993. Reykjavík 1994. 89 s.

36. **Jónbjörn Pálsson, Guðrún Marteinsdóttir, Gunnar Jónsson:** Könnun á útbreiðslu grálúðu fyrir Austfjörðum 1993. Reykjavík 1994. 37 s.

37. Nytjastofnar sjávar 1993/94. Aflahorfur fiskveiðiárið 1994/95. *State of Marine Stocks in Icelandic Waters 1993/94. Prospects for the Quota Year 1994/95*. Reykjavík 1994. 150 s.

38. **Stefán S. Kristmannsson, Svend-Aage Malmberg, Jóhannes Briem, Erik Buch:** *Western Iceland Sea - Greenland Sea Project - CTD Data Report. Joint Danish Icelandic Cruise R/V Bjarni Sæmundsson, September 1990*. Reykjavík 1994. 99 s.

39. **Stefán S. Kristmannsson, Svend-Aage Malmberg, Jóhannes Briem, Erik Buch:** *Western Iceland Sea - Greenland Sea Project - CTD Data Report. Joint Danish Icelandic Cruise R/V Bjarni Sæmundsson, September 1991*. Reykjavík 1994. 94 s.

40. Þættir úr vistfræði sjávar 1994. Reykjavík 1994. 50 s.

41. **John Mortensen, Jóhannes Briem, Erik Buch, Svend-Aage Malmberg:** *Western Iceland Sea - Greenland Sea Project - Moored Current Meter Data Greenland - Jan Mayen, Denmark Strait and Kolbeinsey Ridge September 1990 to September 1991*. Reykjavík 1995. 73 s.

42. **Einar Jónsson, Björn Æ. Steinarsson, Gunnar Jónsson, Gunnar Stefánsson, Ólafur K. Pálsson, Sigfús A. Schopka:** Stofnmæling botnfiska á Íslandsmiðum 1994. - Rannsóknaskýrsla. Reykjavík 1995. 107 s.

43. Nytjastofnar sjávar 1994/95. Aflahorfur fiskveiðiárið 1995/96. *State of Marine Stocks in Icelandic Waters 1994/95 - Prospects for the Quota Year 1995/96*. Reykjavík 1995. 163 s.

44. Þættir úr vistfræði sjávar 1995. *Environmental Conditions in Icelandic Waters 1995*. Reykjavík 1995. 34 s.

45. **Sigfús A. Schopka, Björn Æ. Steinarsson, Einar Jónsson, Gunnar Jónsson, Gunnar Stefánsson, Höskuldur Björnsson, Ólafur K. Pálsson:** Stofnmæling botnfiska á Íslandsmiðum 1995. Rannsóknaskýrsla. *Icelandic Groundfish Survey 1995. Survey Report*. Reykjavík 1996. 46 s.

46. Nytjastofnar sjávar 1995/96. Aflahorfur fiskveiðiárið 1996/97. *State of Marine Stocks in Icelandic Waters 1995/96. Prospects for the Quota Year 1996/97*. Reykjavík 1996. 175 s.

47. **Björn Æ. Steinarsson, Gunnar Jónsson, Hörður Andrésson, Jónbjörn Pálsson:** Könnun á flatfiski í Faxaflóa með dragnót sumarið 1995 - Rannsóknaskýrsla. *Flatfish Survey in Faxaflói with Danish Seine in Summer 1995 - Survey Report*. Reykjavík 1996. 38 s.

48. **Steingrímur Jónsson:** *Ecology of Eyjafjörður Project. Physical Parameters Measured in Eyjafjörður in the Period April 1992 - August 1993*. Reykjavík 1996. 144 s.

49. **Guðni Þorsteinsson:** Tilraunir með þorskgildrur við Ísland. Rannsóknaskýrsla. Reykjavík 1996. 28 s.

50. **Jón Ólafsson, Magnús Danielsen, Sólveig Ólafsdóttir, Þórarinn Arnarson:** Næringarefni í sjó undan Ánanaustum í nóvember 1995. Unnið fyrir Gatnamálastjórann í Reykjavík. Reykjavík 1996. 50 s.

51. **Þórunn Þórðardóttir, Agnes Eydal:** *Phytoplankton at the Ocean Quahog Harvesting Areas Off the Southwest Coast of Iceland 1994*. Svifþörungar á kúfiskmiðum út af norðvesturströnd Íslands 1994. Reykjavík 1996. 28 s.

52. **Gunnar Jónsson, Björn Æ. Steinarsson, Einar Jónsson, Gunnar Stefánsson, Höskuldur Björnsson, Ólafur K. Pálsson, Sigfús A. Schopka:** Stofnmæling botnfiska á Íslandsmiðum 1996. Rannsóknaskýrsla. *Icelandic Groundfish Survey 1996. Survey Report*. Reykjavík 1997. 46 s.

53. Þættir úr vistfræði sjávar 1996. *Environmental Conditions in Icelandic Waters 1996*. Reykjavík 1997. 29 s.

54. **Vilhjálmur Þorsteinsson, Ásta Guðmundsdóttir, Guðrún Marteinsdóttir, Guðni Þorsteinsson og Ólafur K. Pálsson:** Stofnmæling hrygningarþorsks með þorskanetum 1996. *Gill-net Survey to Establish Indices of Abundance for the Spawning Stock of Icelandic Cod in 1996*. Reykjavík 1997. 22 s.

55. Hafrannsóknastofnunin: Rannsókna- og starfsáætlun árin 1997-2001. Reykjavík 1997. 59 s. (Ófáanlegt - *Out of print*).

56. Nytjastofnar sjávar 1996/97. Aflahorfur fiskveiðiárið 1997/98. *State of Marine Stocks in Icelandic Waters 1996/97. Prospects for the Quota Year 1997/98*. Reykjavík 1997. 167 s.

57. Fjölstofnarannsóknir 1992-1995. Reykjavík 1997. 410 s.

58. **Gunnar Stefánsson, Ólafur K. Pálsson** (*editors*): *BORMICON. A Boreal Migration and Consumption Model*. Reykjavík 1997. 223 s. (Ófáanlegt - *Out of print*).

59. **Halldór Narfi Stefánsson, Hersir Sigurgeirsson, Höskuldur Björnsson:** *BORMICON. User's Manual*. Reykjavík 1997. 61 s. (Ófáanlegt - *Out of print*).

60. **Halldór Narfi Stefánsson, Hersir Sigurgeirsson, Höskuldur Björnsson:** *BORMICON. Programmer's Manual. Reykjavík 1997*. 215 s. (Ófáanlegt - *Out of print*).

61. **Þorsteinn Sigurðsson, Einar Hjörleifsson, Höskuldur Björnsson, Ólafur Karvel Pálsson**: Stofnmæling botnfiska á Íslandsmiðum haustið 1996. Reykjavík 1997. 34 s.

62. **Guðrún Helgadóttir:** *Paleoclimate (0 to >14 ka) of W and NW Iceland: An Iceland/USA Contribution to P.A.L.E. Cruise Report B9-97, R/V Bjarni Sæmundsson RE 30, 17th-30th July 1997*. Reykjavík 1997. 29 s.

63. **Halldóra Skarphéðinsdóttir, Karl Gunnarsson:** Lífríki sjávar í Breiðafirði: Yfirlit rannsókna. *A review of literature on marine biology in Breiðafjörður*. Reykjavík 1997. 57 s.

64. **Valdimar Ingi Gunnarsson og Anette Jarl Jörgensen:** Þorskrannsóknir við Ísland með tilliti til hafbeitar. Reykjavík 1998. 55 s.

65. **Jakob Magnússon, Vilhelmína Vilhelmsdóttir, Klara B. Jakobsdóttir:** Djúpslóð á Reykjaneshrygg: Könnunarleiðangrar 1993 og 1997. *Deep Water Area of the Reykjanes Ridge: Research Surveys in 1993 and 1997*. Reykjavík 1998. 50 s.

66. **Vilhjálmur Þorsteinsson, Ásta Guðmundsdóttir, Guðrún Marteinsdóttir:** Stofnmæling hrygningarþorsks með þorskanetum 1997. *Gill-net Survey of Spawning Cod in Icelandic Waters in 1997. Survey Report*. Reykjavík 1998. 19 s.

67. Nytjastofnar sjávar 1997/98. Aflahorfur fiskveiðiárið 1998/99. *State of Marine Stocks in Icelandic Waters 1997/98. Prospects for the Quota year 1998/99*. Reykjavík 1998. 168 s.

68. **Einar Jónsson, Hafsteinn Guðfinnsson:** Ýsurannsóknir á grunnslóð fyrir Suðurlandi 1989-1995. Reykjavík 1998. 75 s.

69. **Jónbjörn Pálsson, Björn Æ. Steinarsson, Einar Hjörleifsson, Gunnar Jónsson, Hörður Andrésson, Kristján Kristinsson:** Könnum á flatfiski í Faxaflóa með dragnót sumrin 1996 og 1997 - Rannsóknaskýrsla. *Flatfish Survey in Faxaflói with Danish Seine in Summers 1996 and 1997 - Survey Report*. Reykjavík 1998. 38 s.

70. **Kristinn Guðmundsson, Agnes Eydal:** Svifþörungar sem geta valdið skelfiskeitrun. Niðurstöður tegundagreininga og umhverfisathugana. *Phytoplankton, a Potential Risk for Shellfish Poisoning. Species Identification and Environmental Conditions*. Reykjavík 1998. 33 s.

71. **Ásta Guðmundsdóttir, Vilhjálmur Þorsteinsson, Guðrún Marteinsdóttir:** Stofnmæling hrygningarþorsks með þorskanetum 1998. *Gill-net survey of spawning cod in Icelandic waters in 1998*. Reykjavík 1998. 19 s.

72. Nytjastofnar sjávar 1998/1999. Aflahorfur fiskveiðiárið 1999/2000. *State of Marine Stocks in Icelandic Waters 1998/1999. Prospects for the Quota year 1999/2000. Reykjavík 1999*. 172 s. (Ófáanlegt - *Out of print*.)

73. Þættir úr vistfræði sjávar 1997 og 1998. *Environmental Conditions in Icelandic Waters 1997 and 1998*. Reykjavík 1999. 48 s.

74. **Matthías Oddgeirsson, Agnar Steinarsson og Björn Björnsson:** Mat á arðsemi sandhverfueldis á Íslandi. Grindavík 2000. 21 s.

75. Nytjastofnar sjávar 1999/2000. Aflahorfur fiskveiðiárið 2000/2001. *State of Marine Stocks in Icelandic Waters 1999/2000. Prospects for the Quota year 2000/2001. Reykjavík 2000*. 176 s.

76. **Jakob Magnússon, Jútta V. Magnússon, Klara B. Jakobsdóttir:** Djúpfiskarannsóknir. Framlag Íslands til rannsóknaverkefnisins EC FAIR PROJECT CT 95-0655 1996-1999. *Deep-Sea Fishes. Icelandic Contributions to the Deep Water Research Project. EC FAIR PROJECT CT 95-0655 1996-1999*. Reykjavík 2000. 164 s. (Ófáanlegt - *Out of print*).

77. Þættir úr vistfræði sjávar 1999. *Environmental Conditions in Icelandic Waters 1999*. Reykjavík 2000. 31 s.

78. *dst^2 Development of Structurally Detailed Statistically Testable Models of Marine Populations. QLK5-CT1999-01609. Progress Report for 1 January to 31 December 2000*. Reykjavik 2001. 341 s. (Ófáanlegt. - *Out of print*).

79. *Tagging Methods for Stock Assessment and Research in Fisheries*. Co-ordinator: Vilhjálmur Þorsteinsson. Reykjavik 2001. 179 s.

80. Nytjastofnar sjávar 2000/2001. Aflahorfur fiskveiðiárið 2001/2002. *State of Marine Stocks in Icelandic Waters 2000/2001. Prospects for the Quota year 2001/2002*. Reykjavík 2001. 186 s.

81. **Jón Ólafsson, Sólveig R. Ólafsdóttir**: Ástand sjávar á losunarsvæði skolps undan Ánanaustum í febrúar 2000. Reykjavík 2001. 49 s.

82. **Hafsteinn G. Guðfinnsson, Karl Gunnarsson:** Sjór og sjávarnytjar í Héraðsflóa. Reykjavík 2001. 20 s.

83. Þættir úr vistfræði sjávar 2000. *Environmental Conditions in Icelandic Waters 2000*. Reykjavík 2001. 37 s.

84. **Guðrún G. Þórarinsdóttir, Hafsteinn G. Guðfinnsson, Karl Gunnarsson:** Sjávarnytjar í Hvalfirði. Reykjavík 2001. 14 s.

85. Rannsóknir á straumum, umhverfisþáttum og lífríki sjávar í Reyðarfirði frá júlí til október 2000. *Current measurements, environmental factors and biology of Reyðarfjörður in the period late July to the beginning of October 2000*. Hafsteinn Guðfinnsson (verkefnisstjóri). Reykjavík 2001. 135 s.

86. **Jón Ólafsson, Magnús Danielsen, Sólveig R. Ólafsdóttir, Jóhannes Briem:** Ferskvatnsáhrif í sjó við Norðausturland að vorlagi. Reykjavík 2002. 42 s.

87. *dst^2 Development of Structurally Detailed Statistically Testable Models of Marine Populations. QLK5-CT1999-01609. Progress Report for 1 January to 31 December 2001* Reykjavik 2002. 300 s.

88. Nytjastofnar sjávar 2001/2002. Aflahorfur fiskveiðiárið 2002/2003. *State of Marine Stocks in Icelandic Waters 2001/2002. Prospects for the Quota year 2002/2003. Reykjavík 2002*. 198 s.

89. **Kristinn Guðmundsson, Ástþór Gíslason, Jón Ólafsson, Konráð Þórisson, Rannveig Björnsdóttir, Sigmar A. Steingrímsson, Sólveig R. Ólafsdóttir, Öivind Kaasa:** *Ecology of Eyjafjörður project. Chemical and biological parameters measured in Eyjafjörður in the period April 1992-August 1993*. Reykjavík 2002. 129 s.

90. **Ólafur K. Pálsson, Guðmundur Karlsson, Ari Arason, Gísli R Gíslason, Guðmundur Jóhannesson, Sigurjón Aðalsteinsson:** Mælingar á brottkasti þorsks og ýsu árið 2001. Reykjavík 2002. 17 s.

91. **Jenný Brynjarsdóttir:** *Statistical Analysis of Cod Catch Data from Icelandic Groundfish Surveys. M.Sc. Thesis*. Reykjavík 2002. xvi, 81 s.

92. Umhverfisaðstæður, svifþörungar og kræklingur í Mjóafirði. Ritstjóri: Karl Gunnarsson. Reykjavík 2003. 81 s.

93. **Guðrún Marteinsdóttir** (o.fl.): *METACOD: The role of sub-stock structure in the maintenance of cod metapopulations*. METACOD: Stofngerð þorsks, hlutverk undirstofna í viðkomu þorskstofna við Ísland og Skotland. Reykjavík 2003. vii, 110 s.

94. **Ólafur K. Pálsson, Guðmundur Karlsson, Ari Arason, Gísli R. Gíslason, Guðmundur Jóhannesson og Sigurjón Aðalsteinsson**: Mælingar á brottkasti botnfiska 2002. Reykjavík 2003. 29 s.

95. **Kristján Kristinsson**: Lúðan (*Hippoglossus hippoglossus*) við Ísland og hugmyndir um aðgerðir til verndunar hennar. Reykjavík 2003. 33 s.

96. Þættir úr vistfræði sjávar 2001 og 2002. *Environmental conditions in Icelandic water 2001 and 2002*. Reykjavík 2003. 37 s.

97. Nytjastofnar sjávar 2002/2003. Aflahorfur fiskveiðiárið 2003/2004. *State of Marine Stocks in Icelandic Waters 2002/2003. Prospects for the Quota year 2003/2004.* Reykjavík 2003. 186 s.

98. *dst^2 Development of Structurally Detailed Statistically Testable Models of Marine Populations. QLK5-CT1999-01609. Progress Report for 1 January to 31 December 2002*. Reykjavik 2003. 346 s.

99. **Agnes Eydal**: Áhrif næringarefna á tegundasamsetningu og fjölda svifþörunga í Hvalfirði. Reykjavík 2003. 44 s.

100. **Valdimar Ingi Gunnarsson** (o.fl.): Þorskeldiskvóti: Yfirlit yfir föngun og áframeldi þorsks á árinu 2002. Reykjavík 2004. 26 s.

101. Þættir úr vistfræði sjávar 2003. *Environmental conditions in Icelandic waters 2003.* Reykjavík 2004. 43 s.

102. Nytjastofnar sjávar 2003/2004. Aflahorfur fiskveiðiárið 2004/2005. *State of Marine Stocks in Icelandic Waters 2003/2004. Prospects for the Quota Year 2004/2005.* Reykjavík 2004. 175 s.

103. **Ólafur K. Pálsson** o.fl.: Mælingar á brottkasti 2003 og Meðafli í kolmunnaveiðum 2003. Reykjavík 2004. 37 s.

104. **Ásta Guðmundsdóttir, Þorsteinn Sigurðsson**: Veiðar og útbreiðsla íslensku sumargotssíldarinnar að haust- og vetrarlagi 1978-2003. Reykjavík 2004. 42 s.

105. **Einar Jónsson, Hafsteinn Guðfinnsson**: Ýsa á grunnslóð fyrir Suðurlandi 1994-1998. Reykjavík 2004. 44 s.

106. **Kristinn Guðmundsson, Þórunn Þórðardóttir, Gunnar Pétursson**: *Computation of daily primary production in Icelandic waters; a comparison of two different approaches.* Reykjavík 2004. 23 s.

107. **Kristinn Guðmundsson og Kristín J. Valsdóttir:** Frum-framleiðnimælingar á Hafrannsóknastofnuninni árin 1958-1999: Umfang, aðferðir og úrvinnsla. Reykjavík 2004. 56 s.

108. **John Mortensen**: *Satellite altimetry and circulation in the Denmark Strait and adjacent seas*. Reykjavík 2004. 84 s.

109. **Svend-Aage Malmberg**: *The Iceland Basin. Topography and oceanographic features*. Reykjavík 2004. 41 s.

110. **Sigmar Arnar Steingrímsson, Sólmundur Tr. Einarsson:** Kóralsvæði á Íslandsmiðum: Mat á ástandi og tillaga um aðgerðir til verndar þeim. Reykjavík 2004. 39 s.

111. **Björn Björnsson, Valdimar Ingi Gunnarsson (ritstj.):** Þorskeldi á Íslandi. Reykjavík 2004. 182 s.

112. **Jónbjörn Pálsson, Kristján Kristinsson:** Flatfiskar í humarleiðangri 1995-2003. Reykjavík 2005. 90 s.

113. **Valdimar I. Gunnarsson o.fl.**: Þorskeldiskvóti: Yfirlit yfir föngun og áframeldi þorsks á árinu 2003. Reykjavík 2005. 58 s.

114. **Kristján Kristinsson, Björn Ævarr Steinarsson og Sigfús Schopka**: Skyndilokanir á þorskveiðar í botnvörpu á Vestfjarðamiðum. Reykjavík 2005.

115. **Erlingur Hauksson** (ritstj). Sníkjuormar og fæða fiska, skarfa og sela. Reykjavík 2005. 47 s.

116. Þættir úr vistfræði sjávar 2004. *Environmental conditions in Icelandic waters 2004.* Reykjavík 2005. 46 s.

117. **Ólafur K. Pálsson** o.fl.: Mælingar á brottkasti 2004 og Meðafli í kolmunnaveiðum 2004. Reykjavík 2005. 37 s.

118. *dst^2 Development of Structurally Detailed Statistically Testable Models of Marine Populations. QLK5-CT1999-01609. Final report: 1 January 2000 to 31 August 2004. Volume 1.* Reykjavík 2005. 324 s.

119. *dst^2 Development of Structurally Detailed Statistically Testable Models of Marine Populations. QLK5-CT1999-01609. Final report: 1 January 2000 to 31 August 2004. Volume 2.* Reykjavík 2005. 194 s.

120. **James Begley**: *Gadget User Guide.* Reykjavík 2005. 90 s.