QLK5-CT1999-01609

---

dst^2

Development of structurally detailed statistically testable models of marine populations

---

Consolidated Progress Report: 1 January 2002 to 31 December 2002

Commencement date: 01-01-2000

Completion date: 31-12-2003

Co-ordinator:

Dr. G. Stefánsson
Marine Research Institute
PO Box 1390
Skúlagata 4
Reykjavík
Iceland

Phone: +354 552 0240
Fax: +354 562 3790
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

Project summary

Report structure

This third dst^2 progress report is structured in exactly the same manner as the first two reports, i.e. separated into this summary, a short description of the status of each workpackage and a collection of documents describing in more detail progress on individual components of each workpackage. The summary is intended to provide an overview, with an emphasis on the status of each deliverable, milestone etc.

As noted earlier, there were staffing problems at the MRI and FRS, leading to delays in a few deliverables and general prograss, but these particular issues have been resolved.

The workpackages are in 5 groups:

Number	WP group	Workpackages
1	Data warehouse	1.1-1.7
2	Structural models/programming	2.1-2.4
3	Estimation and inference w/programming	3.1-3.2
4	Estimation of parameters outside program	4.1-4.3
5	Case studies	5.1-5.3

Following these short summary subsections, the following sections summarise the current state of work for each workpackage, including whether it is complete or not. Each workpackage also contains an estimate of the amount of time initially intended for the workpackage along with the current estimate of time allotted to it, to date. Further, each workpackage lists the status of all deliverables and objectives in some detail.

Several annexes are included with this report. These annexes describe completed, current or planned work relating to the workpackages in somewhat more detail than the relatively short report body.

The numbering convention in this report is largely as in former reports, i.e. with slight changes in numbering of deliverables and WPs from the original Technical Annex.

Status of the project as a whole

Data bases (WP 1)

The data for implementing the models will be set up in the form of a data warehouse. All important aspects of development of the data warehouse (DW) are proceeding as planned. Thus, most prototype data are available in data warehouses and extraction modules are available, both for general use and specifically for Gadget input files. Work related to this group of WPs is largely done.

Models and implementation (WP 2 and 3)

Several model components were due at this time but somewhat fewer are ready. Work on migration models has proceeded adequately and several such models are being tested. Novel methods for including tagging data have been developed and are being tested.

The computer program, Gadget, implementing the models has consistently been ahead of schedule. All proposed modifications and additions have been implemented as soon as proposals have been made concrete.

A serious deficiency in WPs 2 and 3 is the state of affairs with models of recruitment, i.e. closure of the life cycle. Reasons for this include both resource limitations (manpower) and lack of coordination between biological knowledge and modelling. Implementation work has been initiated, however, and will be completed in 2003.

Other delays in these WPs are of minor consequence, since e.g. the growth modules were greatly enhanced earlier through the adoption of more flexible growth update mechanisms. In 2002 considerable testing and model validation was undertaken, particularly with regard to the growth mechanism. Thus, both the migration and growth modules are in reasonable shape, though more work is planned for both in 2003.

As mentioned in earlier reports, special programs have been specially developed for investigating sensitivity of solutions to perturbation in the parameter space, etc. Further programs will need to be developed to undertake explorations of the high-dimensional parameter spaces and also to undertake statistical analyses of goodness of fit and so on.

Estimation outside program (WP 4)

Investigations, independent of the assessment model itself, have been conducted at SCUI, UiB and DIFRES. Most of these have been completed. Some remaining issues of how those results can be applied within Gadget runs are being investigated.

Case studies (WP 5)

All case studies are well underway. Initial delays, mainly due to manpower problems, appear to have been solved.

Status of objectives

The 4 primary objectives driving the current project were described in the technical annex and are summarised here for clarity.

Objective 1 of the project is to collect relevant data and to provide objective means of analysing these. The intention is to assemble relevant data in a highly-disaggregated form in a database format which is designed to hold many different classes of data but which can provide summarised data extractions in a format amenable to the analytic routines.

This objective has been met: A data warehouse has been developed, which can be used in a general (international) fisheries context. The entire system is based completely on open source software and can be freely distributed. Most of the case study data have been entered into the various DWs, at least in prototype form.

Objective 2 is to validate present estimates of the stock sizes, exploitation histories and the associated uncertainties for a number of case studies, using statistically appropriate models that include detailed descriptions of growth, migration and predation.

Current work on this topic includes the development of appropriate statistical methods to undertake comparisons of different assessment models. It is, however, not obvious whether this objective can be met, since there are several statistical hurdles which need to be overcome. The final verdict will have to wait until the very end of the project, but several methods are being investigated.

Objective 3 is to evaluate whether, when and how increased complexity in models enhances the ability to provide management advice in: (a) Advising on effects of closed area restrictions (b) Advising on the state of the stocks and annual catch forecasts (c) Advising on small-scale fishery effects such as local depletion of forage species.

This will be addressed in case studies in 2003.

Objective 4 is to evaluate and validate some current perceptions of limit reference points for safe exploitation of key resources, using models which take spatial concerns and multispecies interactions into account. Current stock management advice is based on comparison of population parameters with biological reference points that are all evaluated on an aggregate-model, single-species basis. The objective will be met when a comparison of the consequences of stock management based on simple-model advice and management based on detailed-model advice has been drawn for at least one case study.

This objective has been met, as described in earlier reports.

Status of deliverables

A deliverable indicates completion of a specific piece of code, evaluation of methods, completion of data sets or test runs etc. Details on the completion of deliverables is indicated within each workpackage, a summary is provided in the following table.

Due year	Due now	Completed	In normal progress	Lagging	Total
2000-2002	49	43		6	49
2003	0		8		8
Total	49	43	8	6	57
%	86	75	14	11	

It is seen that 11% of the deliverables are somewhat behind schedule. The extent of this varies from a few months delay to complete rescheduling, as detailed elsewhere.

Further summary tables describing the state of deliverables according to their due date and completion status is given in Annex A

The lagging deliverables are of course somewhat of a concern, but this has been addressed through the following means: When looking through the detailed lists of deliverables, it is seen that of the 14 deliverables not yet completed, these are distributed among WP groups as follows

WP group	1	2	3	4	5	Total
Remaining deliverables (lagging or in normal progress)	0	2	1	4	8	14
Total # deliverables	17	8	8	10	14	57

In terms of numbers, the model implementations in case studies (WP group 5) contains the greatest number of remaining deliverables, partly because many are due in 2003, but some are lagging. This issue has been addressed through enhanced correspondence between staff at the institutes, involving considerably more travel than originally envisaged, but clearly needed in order to complete the task. There is full reason to expect this to result in the completion of the deliverables in WP group 5 in 2003.

However, within WP group 4, although some deliverables (4.1) are close to completion, others are simply difficult and testing techniques under development (4.2) have not been investigated before. It is, however, clear at this moment that this research will lead to methods to illustrate model adequacy, though it may not lead to the exact techniques initially desired.

Effort on WP group 2 has been redefined in order to ensure that progress will be made in terms of deliverables, in 2003.

Status of milestones

Milestones refer to completion of major tasks within the project. The milestones for dst^2 are:

M1: Table definitions. Complete definitions are available for all tables to be seen in the final data warehouse. This is a prerequisite for programming the data warehouse.

This milestone has been reached.

M2: Data entry for the case studies completed. This is a prerequisite for loading the data warehouse.

This milestone has been reached: All data entry has been completed to the extent needed to load the data warehouses, though some further data checks will no doubt be conducted. Initial data has been loaded into 2 of three DWs.

M3: Prototype structural elements and inference tools defined and programmed. This is a prerequisite for being able to test some of the detailed models.

This milestone has been reached.: All prototypes are available.

It is envisaged that the model will continue to be developed interactively as the case studies proceed, but the prototype is now complete.

M4: Prototypes implemented for the case studies and developed further.

Several prototypes are complete and are being developed further. One is still in an early stage but man-power has been allocated to speed it up.

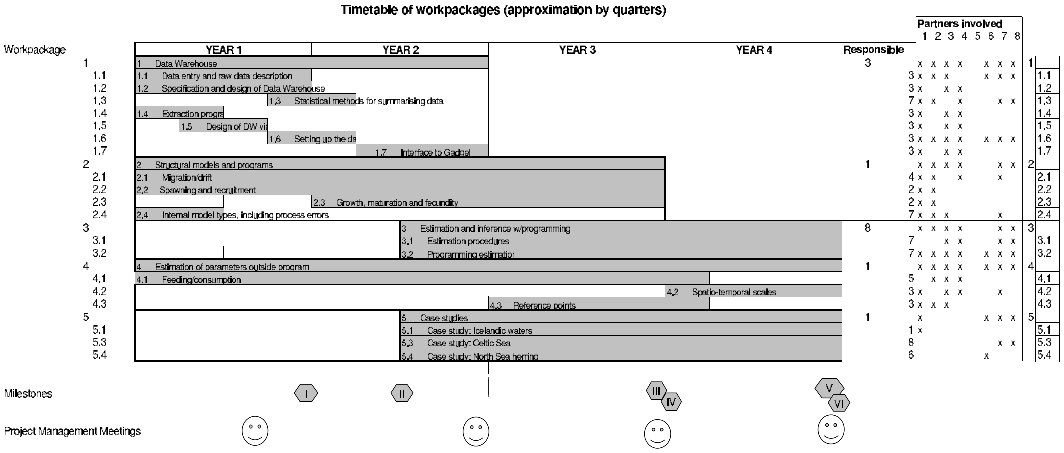
M5: Appropriate parameterization for case-study models chosen by statistical hypothesis-testing. In particular this implies that model fitting procedures are available.

This milestone is due at the end of 2003. Prototype model fitting procedures are available but likelihood functions need to be further developed in accordance with results from goodness-of-fit tests which indicate traditional probability distributions for fisheries data to be seriously invalidated.

M6: Evaluation of implications of new models compared with simpler models completed.

As reported earlier, this milestone is due at the end of 2003. Initial tests indicate that this will require the full length of time indicated, since this is a very complex issue indeed. Basically, this refers to comparing models where the likelihood function is a composite of many individual components, which is a non-trivial statistical problem. To add to the confusion, issues such as how to weight individual components are still unsolved. Approaches to this are being developed (see Annexes to both first and second reports and Annex F.1)

The time table for the project as a whole, including milestones, is given in Fig. 1. It is seen that milestones, M1-4 were due in 2000-2002 and they should all be considered largely complete. The minor deviations are under control.



Workpackage	Task	Responsible	1	2	3	4	5	6	7	8	
1	Data Warehouse	3	x	x	x	x		x	x	x	1
1.1	Data entry and raw data description	3	x	x	x			x	x	x	1.1
1.2	Specification and design of Data Warehouse	3	x		x	x					1.2
1.3	Statistical methods for summarising data	7	x	x		x			x	x	1.3
1.4	Extraction progra	3	x		x	x					1.4
1.5	Design of DW vi	3	x		x	x					1.5
1.6	Setting up the da	3	x	x	x	x		x	x	x	1.6
1.7	Interface to Gadget	3	x		x	x					1.7
2	Structural models and programs	1	x	x	x	x			x	x	2
2.1	Migration/drift	4	x	x		x			x		2.1
2.2	Spawning and recruitment	2	x	x							2.2
2.3	Growth, maturation and fecundity	2	x	x							2.3
2.4	Internal model types, including process errors	7	x	x	x				x		2.4
3	Estimation and inference w/programming	8	x	x	x	x			x	x	3
3.1	Estimation procedures	7			x	x			x	x	3.1
3.2	Programming estimatior	7	x	x	x	x		x	x	x	3.2
4	Estimation of parameters outside program	1	x	x	x	x		x	x	x	4
4.1	Feeding/consumption	5		x	x	x					4.1
4.2	Spatio-temporal scales	3	x		x	x			x		4.2
4.3	Reference points	3	x	x	x						4.3
5	Case studies	1	x					x	x	x	5
5.1	Case study: Icelandic waters	1	x								5.1
5.3	Case study: Celtic Sea	8							x	x	5.3
5.4	Case study: North Sea herring	6						x			5.4

Partners involved

Figure 1: Time table

Contents

1 Workpackage group: Data warehouse

1.1 Workpackage: Data entry and raw data description

Start date or starting event: February 2000

N^{o} of the partner responsible: 3

N^{o}s of other partners involved: 1,2,6,7,8

Table 1: Person-months by partner within workpackage

Partner	1	2	3	4	5	6	7	8
2000	24	1	0	.	.	12.6	0	0.0
2001	35	0	0.7	.	.	0.0	2	0.5
2002	30	0	0	.		8	3.42	0
Total to date	89	1	0.7	.	.	20.6	5.42	0.5
Total planned	120	2	6	3	.	14.0	3	0.5

Time to completion of workpackage (current estimate): 0

Objectives

- To enter, correct and document raw data in the institutional data bases in order to have the foundations required for compiling the data warehouse. *Status of objective:***Met**
- To create the basis for the data warehouse. *Status of objective:***Met**

Description of work conducted to date

As noted in the first dst^2 report, the main point and most important part of this workpackage is to "get the raw data in order" so as to enable its inclusion in the various data warehouses. This workpackage is now largely complete. All of the institutes have organized their data sufficiently to set up at least prototype data warehouses, and although some further cleanup is envisaged within dst^2 in 2003, this is relatively minor.

MRI (1):
Deliverable D1.1.1 is considered complete, as testing procedures have been developed; a sufficient number of years is available in a double-checked data base to set up a data warehouse for biological measurements, general survey and landings data. A description of the existing database is available at the MRI website, http://www.hafro.is (in Icelandic). Acoustic data have been aggregated to a level adequate for entry into the DW as a separate project, but will be used in dst^2.

IMR (2) and *DIFRES (3)*: Work in WP 1.1 was completed and reported in earlier reports.

FRS (6):
The data for the NS herring case study have been formatted in the ASCII Format Exchange Format into 6 differents files.

- a catch table
- 3 biological tables (catchsample table, lengthcell, age)
- environment table

Some time has been spent on correcting the data from FRS and finding more complete sources of data.

The lookup tables have been created but the DW is still not built. This will be done in 2003 through an exchange of personnel between FRS and MRI.

CEFAS (7) & IFREMER(8):
With the entering of 26,539 stomach content records (18,129 stomachs) for the Celtic Sea, this data entry phase of the project has been completed. Additional analyses of the entered data has been carried out (Annex I.3) in order to validate the data which was entered by hand from paper records and logbooks. The format of the raw data is described in Annex I.3 and all records are stored in MS Access but also in a less detailed format in the DST2 data warehouse.

The French stomach sampling data for the Celtic sea has been entered by partner 7 and a joint Celtic Sea stomach data base was created by that partner.

Deliverables

The outcome of this work will be corrected databases containing the raw data which form the bases for the data summaries used in the data warehouses.

D1.1.1: Corrected, documented data base for Icelandic waters. Q5 **Complete**

D1.1.2: Corrected, documented data base for North Sea herring. Q5 **Complete**

D1.1.4: Corrected, documented prototype data base for the Celtic Sea. Q5 **Complete**

Milestones and expected results

This is work towards milestone M2.

1.2 Workpackage: Specification and design of Data Warehouse

Start date or starting event: February, 2000

N^o of the partner responsible: 3

N^{o}s of other partners involved: 1,4

Table 2: Person-months by partner within workpackage

Partner	1	2	3	4	5	6	7	8
2000	2	.	2	1.5	.	.	.	.
2001	3	.	8.8	0	.	.	.	.
Total to date	5	.	10.8	1.5	.	.	.	.
Total planned	5	.	11	12	.	.	.	.

Time to completion of workpackage (current estimate): 0

Summary: As reported in the second progress report, this work package is considered complete.

CORBA development continues as a part of other WPs.

Objectives

- To design coordinated modules and structures of the DW along with programs for import and export. *Status of objective:***Met**

Description of work conducted to date See earlier reports.

Deliverables

D1.2.1: DW design. Q5 **Complete**

Milestones and expected results

This is work towards milestone M1.

1.3 Workpackage: Statistical methods for summarising data

Start date or starting event: October, 2000

N^o of the partner responsible: 7

N^{o}s of other partners involved: 1, 2, 4, 8

Table 3: Person-months by partner within workpackage

Partner	1	2	3	4	5	6	7	8
2000	2	0	.	0	.	.	0	0.5
2001	3	0	.	0	.	.	0	0.0
2002	1	0	.	0	.	.	0	0.0
Total to date	6	0	.	0	.	.	0	0.5
Total planned	6	3	.	3	.	.	2	0.5

Time to completion of workpackage (current estimate): 0

Summary: As reported in the second progress report, this work package is considered complete.

Objectives

- To define the computational methods to be used when computing the summary tables to be used in the data warehouses. *Status of objective:***Met**

Description of work conducted to date

See earlier reports.

Deliverables

D1.3.1: Algorithms and a set of SQL programs to be used for taking the raw data and converting the data into summaries useful for ecosystem modelling. Q7 **Complete**

Milestones and expected results

This is work towards milestone M2.

1.4 Workpackage: Extraction programs (prototypes)

Start date or starting event: February, 2000

N^o of the partner responsible: 3

N^{o}s of other partners involved: 1, 4

Summary: As reported in the second progress report, this work package is considered complete.

Time to completion of workpackage (current estimate): 0

Table 4: Person-months by partner within workpackage

Partner	1	2	3	4	5	6	7	8
2000	5	.	1	3	.	.	.	
2001	0	.	0.5	0	.	.	.	
Total to date	5	.	1.5	3	.	.	.	
Total planned	5	.	8	3	.	.	.	

Objectives

- To define and write prototype programs to extract data from raw data bases into data warehouse tables and Gadget data files. *Status of objective:***Met**

Description of work conducted to date See earlier reports.

Deliverables

D1.4.1: Prototype data sets for Gadget and prototype data warehouse views. Q3 **Complete**

Milestones and expected results

This is work towards milestones M1, M4.

1.5 Workpackage: Design of DW views and structures

Start date or starting event: April, 2000

N^o of the partner responsible: 3

N^{o}s of other partners involved: 1,4

Summary: This work package is considered complete. The structures are complete and several views have been developed. Some further views will be developed in 2003, but these are a minor effort and pertain mainly to offering a public view into the data warehouse, rather than that needed for the project.

Table 5: Person-months by partner within workpackage

Partner	1	2	3	4	5	6	7	8
2000	1	.	2	0.5	.	.	.	.
2001	1	.	0.5	0	.	.	.	.
2002	5	.	0	1.5	.	.	.	.
Total to date	7	.	2.5	2	.	.	.	.
Total planned	12	.	7	12.0	.	.	.	.

Time to completion of workpackage (current estimate): 0

Objectives

- To design the data structures of the data warehouse. *Status of objective:***Met**
- To develop a modern data warehouse using CORBA technology. *Status of objective:***Changed and met**
- To design an Internet based program that can report the content of the data warehouse in a predefined set of views. *Status of objective:***Met**

Description of work conducted to date

See earlier reports for the most part. Minor work continues within this workpackage on testing technologies (CORBA), but in essence this workpackage is considered complete.

MRI (1):
MRI has contributed to design and implementation of upload and extraction modules. Design issues are described in annexes containing a CORBA paper (Annex B.1), and a paper on the data warehouse structure (Annex B.2).

Deliverables

D1.5.1: Design of data warehouse views. Q4 **Complete**

D1.5.2: Design of the database structures. Q4 **Complete**

Milestones and expected results

This is work towards milestones M1, M4.

Milestone M1 is complete when this workpackage is complete.

1.6 Workpackage: Setting up the data warehouse

Start date or starting event: End of 1.5

N^o of the partner responsible: 3

N^{o}s of other partners involved: 1,2,4,6,7,8

Table 6: Person-months by partner within workpackage

Partner	1	2	3	4	5	6	7	8
2000	8	1	6.0	2	.	1.9	0.0	0.3
2001	1	0	3.9	0	.	0	0.5	0.2
2002	8	0	1.9	8	.	0	1	0.5
Total to date	17	1	11.8	10	.	1.9	1.5	1
Total planned	24	3	32	24	.	14	1	0.5

Time to completion of workpackage (current estimate): 6

Objectives

- To provide the data description, links and formats required to implement the DW programs for all case studies in a unified manner. *Status of objective:***Met**
- To implement and test the security software, browser- and e-mail-based import software and SQL based data warehouse export software to Gadget at selected sites. *Status of objective:***Met**
- To implement and test CORBA technology at selected sites. *Status of objective:***Work ongoing**
- To evaluate and select a technology for the DW. *Status of objective:***Met**
- To implement the DW at all sites. *Status of objective:***Work ongoing**

Description of work conducted to date

All essential parts of this workpackage are considered complete. The various techniques have been fairly extensively tested and appropriate techniques have been selected.

MRI (1):
MRI has set up the DW for Icelandic waters, as described in Annex C.1. The spatial hierarchy is defined in Annex H.1.

As described in earlier reports, a system of distributed data warehouses based on the public domain Postgresql data base system has been designed and implemented.

CORBA continues to be tested for connecting the data warehouses. Should this prove to be feasible, it will provide an additional and potentially very useful feature, not strictly required in the project (but see Annex B.1).

DIFRES (3): DIFRES input to the WP has been described in earlier reports.

SCUI (4):
An attempt has been made to upgrade the versions of the software used in developing the application. Earlier problems with the ORB implementation of choice (JacORB) with the latest version of Java are now being put to the test with the latest version of JacORB. Similarly, an attempt is being made to integrate the latest version of Postgres database into the application.

It is neccessary to stay up-to-date with the latest versions of all freeware distributions, both because of the greater support that the devlopers of the distributions give to the latest versions, and because of bugfixes. Usually, each new version of freeware has had a significant number of errors fixed, which will both contribute to the reliability and security of the application. Furthermore, older versions of these programs may become obsolete, making it even harder to implement any changes to the server/client application.

In light of the number of changes made from older versions to JacORB 1.4.1, some problems are expected during this phase.

Also, an effort to simplify the process of installing the application at the client site is being made. We believe that the installation of the java client should be very nearly automated once this is done. However, this relies on the correct installation of Java, Postgres and JacORB.

Some of the work under this workpackage was transferred from other workpackages (e.g. 1.2).

FRS (6):
Data concerning North Sea Herring for all tables, except Table 8, are ready to be loaded into a Warehouse. The related look-up tables are being prepared.

The status of the North Sea herring DW is given in Annex C.2.

Celtic Sea: CEFAS (7) & IFREMER (8)
The French and English Celtic sea data concerning survey data and commercial catch data have been uploaded into the Celtic sea data warehouse (Tables 1, 4, 5, 6, 7 and 18). An initial data warehouse version has been installed in Nantes. The stomach data are the only data missing in the Celtic Sea data warehouse as data entry had taken longer than planned and problems with the data format were encountered due to incomplete and missing information for many records. The data warehouse formats were adapted and incorporation of the stomach data should be completed soon. As this provides a consistent DW, D1.6.9 is considered complete.

Deliverables

D1.6.1: Each institute provides descriptions of the data storage format used for all data sets to be used as a basis for the DW. Q6 **Complete**

D1.6.2: Known techniques of DW implementation using XML are described in detail and tested. Q6 **Complete**

D1.6.3: CORBA technology described in detail and tested. Q6 **Complete**

D1.6.4: The results of using CORBA and XML compared to select the appropriate technology. Q7 **Completed** as a modified deliverable, see above.

D1.6.5: Final choice of DW software set up, tested and distributed. Q8 **Complete**

D1.6.6: Data warehouse for Icelandic waters. Q9 **Complete**

D1.6.7: Data warehouse for North Sea herring. Q9 **Ongoing - Input complete**

D1.6.9: Data warehouse for the Celtic Sea. Q9 **Complete**

Milestones and expected results

This is work towards milestones M4, M5, M6.

Milestone M4 is complete when this workpackage is complete.

1.7 Workpackage: Interface to Gadget

Start date or starting event: End of 1.6

N^o of the partner responsible: 3

N^{o}s of other partners involved: 1,4

Table 7: Person-months by partner within workpackage

Partner	1	2	3	4	5	6	7	8
2000	2	.	1	0	.	.	.	.
2001	2	.	0	0	.	.	.	.
2002	4	.	0	0	.	.	.	.
Total to date	8	.	1	0	.	.	.	.
Total planned	18	.	10	12	.	.	.	.

Time to completion of workpackage (current estimate): 0

Summary: This workpackage is considered complete. The Gadget input file structure was redefined so that Gadget configuration data is as separate from biological measurements as possible, and the biological data are stored in column format. This drastically simplified the effort needed to extract Gadget input data from the DW.

Objectives

- To define and write procedures to extract data from data warehouses into the file formats which are used by the assessments and hypothesis testing modules. *Status of objective:***Met**

Description of work conducted to date

MRI(1):
In accordance with definitions set out at a joint meeting, MRI staff have defined extraction filters to generate Gadget input files (Annex B.3).

Deliverables

D1.7.1: Software to set up all data and parameter files for Gadget. Q9 **Complete**

Milestones and expected results

This is work towards milestones M5, M6.

M1 needs to be completed before this workpackage can be completed.

2 Workpackage group: Structural models

2.1 Workpackage: Migration/drift

Start date or starting event: February, 2000

N^o of the partner responsible: 4

N^{o}s of other partners involved: 1, 2, 7

Table 8: Person-months by partner within workpackage

Partner	1	2	3	4	5	6	7	8
2000	1	0	.	10	.	.	0.0	.
2001	6	3	.	18	.	.	0.5	.
2002	0	8	.	14	.	.	1.0	.
Total to date	7	11	.	42	.	.	1.5	.
Total planned	8	12	.	45	.	.	6.0	.

Time to completion of workpackage (current estimate): 12

Objectives

- To develop a migration model for use in Gadget. *Status of objective:***Met**
- To evaluate the data needs for estimating migration parameters. *Status of objective:***Work ongoing**
- To identify potential environmental variables that might lead to a better understanding of fish migration and larval drift. *Status of objective:***Work ongoing**

Description of work conducted to date

Much of the work within this WP is fairly independent of the development of Gadget code. In order to complete at least the initial definitions of the migration components in Gadget it has been decided to allow the migration proportions within Gadget to be (nonlinear) regression-type functions of external variables such as water movement. This completes D2.1.2. Continuing work may well give alternative formulations.

MRI (1):
Earlier work described the incorporation of migration probabilities varying smoothly in time.

Current and immediate future work involves testing the estimability of these parameters, including estimation of the variance of the point estimates. The first steps to this (completed in 2002) include incorporating tagging data in Gadget, implemented as tagged sub-populations which are subject to the same processes as the underlying population being simulated (Annex D.3). Also incorporated are prototype likelihood components (a Poisson distribution). In 2003 the development of this Gadget component will be completed and tested.

Joint work between MRI and IFREMER involves the development of a bootstrap for estimating variances by repeated extractions of subsets of data from the data warehouse.

IMR (2):
Parameters in the migration model will be sought estimated using classical maximum likelihood estimation. Current fields and temperature fields for the period 1997-1999 have been implemented and

likelihood functions based on observed abundance distributions taken as the integrator values from the IMR demersal fish survey in February and on biological samples in March have been defined.

Preliminary results with a simple migration hypothesis where temperature is the only field in which the capelin navigates were presented to the dst2 meeting on fish migrations in Bergen, 17-18 November, 2002.

SCUI (4):
Completion of continuous density migration models (Annex D.1) incorporating a temperature field (and also other environmental variables if available); parameterization of the model. Setting up of the framework for application to Barents Sea capelin and Iceland capelin. Completed for Barents Sea (Annex D.2).

CEFAS (7):
Analysis of tagging data

A tagging study carried out in the Celtic and Irish Seas by CEFAS staff (Dunn and Pawson, 2002) used a total of 2788 plaice tag/recapture records over the period 1979 - 1996. This study demonstrated that extensive migrations are undertaken by plaice in this region as well as the existence of substocks within ICES areas VIIa and VIIf&g. Further work in 2003, will examine migration patterns in cod and sole using the small amount of tagging data available.

Dunn MR, Pawson MG (2002) The stock structure and migrations of plaice populations on the west coast of England and Wales. Journal of fish biology, 61 (2): 360-393.

Deliverables

D2.1.1: Initial, simple migration model. Q4. **Complete**

D2.1.2: A proposed mathematical model formulation for use in Gadget, based on comprehensive theoretical analysis of the biolological and physical properties of the system, e.g. currents and temperature. Q12 **Complete**

D2.1.3: Evaluation of reduction in variances as a result of reducing uncertainty in different data sources, including tagging data, survey indices and hydrographic information, using simulation with Gadget. Q12 **Ongoing**

Milestones and expected results

This is work towards milestones M5, M6.

Parts of this task provide input to milestone M3, almost complete.

2.2 Workpackage: Spawning and recruitment

Start date or starting event: February, 2000

N^o of the partner responsible: 2

N^{o}s of other partners involved: 1

Table 9: Person-months by partner within workpackage

Partner	1	2	3	4	5	6	7	8
2000	1	32	.	.	.	.	.	.
2001	0	28	.	.	.	.	.	.
2002	2	13	.	.	.	.	.	.
Total to date	3	73	.	.	.	.	.	.
Total planned	18	88	.	.	.	.	.	.

Time to completion of workpackage (current estimate): 12

Objectives

- To formulate a description of the recruitment process in order to have a model of the full life-cycle of each species. *Status of objective:***Work ongoing**

Description of work conducted to date

MRI (1):
Gadget includes a simple spawning module, developed at MRI, but not been tested yet. The next step in developing the biological modelling component of Gadget at the MRI will be to close the life cycle by allowing the spawning module to generate eggs which will drift (using the migration module) to nursery areas. As a first step towards obtaining realistic abundance values the mortality in each region will be estimated or set to give magnitudes corresponding to survey indices.

IMR (2):
It has been outlined how a closed life cycle should be implemented within Gadget (Bogstad, Implementation of closed life cycle within Gadget - Annex D.4).

A new generic formulation for cannibalism in the fleksibest formulation was implemented in the Gadget code. This allows for prey switching by relating the level of cannibalism mortality to the abundance of other prey (e.g. capelin). In addition, it is based on the model for size preferences given in Annex C.6 in the 2001 dst2 report.

Refinements to the length-based method that is being used to estimate total egg production (TEP) for Northeast Arctic cod over the full time period have continued. These include: · re-estimation of weight-at-length using new weight/length relationships · development of a new fecundity model that includes a term for relative condition. Thus, the length-specific fecundity varies according to the condition value that year.

This work allows for expressing the fecundity as a function of Gadget state variables (number and mean weight of fish in a given age-length cell), and is thus implementable in Gadget (see Bogstad, Annex D.4). The justification for this is that the correlation between TEP and recruitment is stronger than the correlation between SSB and recruitment for Northeast Arctic cod (ICES, 2002).

Knowledge from the published papers listed below can also be implemented in a closed life-cycle model.

References:

Fiksen, Ø. and Slotte, A. Stock-environment recruitment models for Norwegian spring spawning herring. Canadian Journal of Aquatic Sciences 59:211-217.

Helle, K., Pennington, M., Bogstad, B. and Ottersen, G., 2002. Some environmental factors that influence the growth of Arcto-Norwegian cod from the early juvenile to the adult stage. Environmental Biology of Fishes 65:341-348.

ICES, 2002. Report of the Study Group on the Incorporation of Process Information into Stock-recruitment Models. Lowestoft, 14-18 January 2002. ICES C. M. 2002/C:01.

Marshall, C.T., O'Brien, L., Tomkiewicz, J., Marteinsdóttir, G., Morgan., M-J., Saborido-Rey, F., Köster, F., Blanchard, J.L., Secor, D., Kraus, G., Wright, P., Mukhina, N.V., and Björnsson, H. in press. Developing alternative indices of reproductive potential for use in fisheries management: case studies for stocks spanning an information gradient. J. Northwest Atl. Fish. Sci.

Mukhina, N.V., Marshall, C.T., and Yaragina, N.A. in press. Tracking the signal in year-class strength of Northeast Arctic cod through multiple survey estimates of egg, larval and juvenile abundance. J. Sea Res.

Deliverables

D2.2.1: A module to describe the recruitment process. Q12 **Ongoing**

Milestones and expected results

This is work towards milestones M3, M5, M6.

2.3 Workpackage: Growth, maturation and fecundity

Start date or starting event: February, 2001

N^o of the partner responsible: 2

N^{o}s of other partners involved: 1, 4

Table 10: Person-months by partner within workpackage

Partner	1	2	3	4	5	6	7	8
2000	1	12	.	0	.	.	.	.
2001	11	8	.	2	.	.	.	.
2002	0	5	.	1	.	.	.	.
Total to date	12	25	.	3	.	.	.	.
Total planned	12	30	.	28	.	.	.	.

Time to completion of workpackage (current estimate): 12

In terms of deliverables, this workpackage is considered complete, but some issues need to be cleaned up before the objectives have been met. This will be done at IMR.

Objectives

- To evaluate different models of growth and maturation including dispersion of length at age, and models for effective fecundity related to nutritional state as well as maternal effects on recruitment. *Status of objective:***Work ongoing**
- To model the relationship between growth and consumption taking into account body size and metabolic costs, in order to make predictions of growth in the short term. *Status of objective:***Work ongoing**

Description of work conducted to date

Models which have been developed at the MRI during the past several years are now in place in Gadget. These models are fairly ad-hoc, but they are enough to complete the case studies.

MRI (1):
The growth models developed earlier by MRI and described in the dst^2 report for 2000 have been adopted as the current state-of-the-art in Gadget, along with maturity increment models, thus completing D2.3.1.

IMR (2):
A method was developed for disaggregating the length-at-age and weight-at-length signals in the growth, and was applied to Northeast Arctic cod. The method used Russian and Norwegian age/length keys that have recently become available for the full time period (1946-2001) to generate estimates of mean length-at-age and then paired these estimates with the corresponding values of mean weight-at-age. This allowed year-specific weight/length relationships to be estimated and these relationships were used to generate length-specific condition values for the stock. The index is referred to as a composite condition index (Kc) to underscore that it is derived from highly aggregated stock-level data rather than from data

collected at the individual-level. The time series of Kc values shows both year and cohort effects. This is relevant to construction/choice of growth models for Northeast Arctic cod in Gadget.

Maturity at length for Northeast Arctic cod has also been re-estimated for the full time series. This allows for constructing a maturation model, by taking into account the proportion of first-time spawners. The model could be made sex-specific as well as dependent on both length and weight.

References:

ICES 2003. Report of the Study Group on Growth, Maturity and Condition in Stock Projections, Copenhagen 5-10 December 2002. ICES C.M. 2003/D:01, 95 pp.

Oskarsson, G. J., Kjesbu, O. S. and Slotte, A. 2002. Predictions of realised fecundity and spawning time in Norwegian spring spawning herring (Clupea harengus L.). Journal of Sea Research 48(1):59-79.

SCUI (4):
Work on this package is still at the preliminary stage. Some investigations of the linking of "stage of maturity" and migrations has been carried out.

Deliverables

D2.3.1: Mathematical models of the growth and maturation process. Q12 **Complete** (Extensions may be developed at IMR).

Milestones and expected results

This is work towards milestone M3, the prototype of D2.3.1 towards M3 is done.

2.4 Workpackage: Internal model types, including process errors

Start date or starting event: February, 2000

N^o of the partner responsible: 7

N^{o}s of other partners involved: 1,2,3,8

Table 11: Person-months by partner within workpackage

Partner	1	2	3	4	5	6	7	8
2000	6	1	2.1	.	.	.	0.0	0.5
2001	0	2	0.0	.	.	.	0.5	1.0
2002	0	10	0.0	.	.	.	0.0	0
Total to date	6	13	2.1	.	.	.	0.5	1.5
Total planned	2	6	4.0	.	.	.	6	2.0

Time to completion of workpackage (current estimate): 0

Summary: This workpackage is almost complete. From a mathematical and programming point of view an approach has been developed where process error can easily be included in almost any model parameter. As reported earlier, considerable external work has been done on Bayesian approaches and new growth update functions have been implemented in Gadget. Some further enhancements will be developed, but this will be continued as part of other WPs.

Objectives

- To define appropriate internal population models. *Status of objective:***Met**
- To incorporate process error through time series models within Gadget *Status of objective:***Met**

Description of work conducted to date

MRI (1):
As reported earlier, MRI has completed the inclusion of process-error models in Gadget. Earlier work included the inclusion of a beta-binomial model for the growth update mechanism. Recent additions include modifications to growth models, partly in order to allow for easier debugging. Considerable work has been put into developing external simulation for the purpose of testing the Gadget code as well as the behaviour of algorithms.

IMR (2):
A matrix formulation of an age-length structured model which is single-species, but allows for multiple population groups and multiple areas, has been developed (Frøysa and Bogstad, 2003).

Frøysa, K. G., and Bogstad, B. 2003. Matrix formulation of age- and length structured population models. (in submission).

Deliverables

D2.4.1: Mathematical formulation of general populations dynamics models in state space form. Q3 **Complete**

D2.4.2: Corresponding program modules. Q4 **Complete**

D2.4.3: Prescription to incorporate relevant prior information into population dynamics models (model structure, parameter distributions, etc.). Q4 **Complete**

Milestones and expected results

This is work towards milestone M3 which has been reached in terms of requirements from this workpackage.

3 Workpackage group: Estimation and inference

3.1 Workpackage: Estimation procedures

Start date or starting event: July, 2001

N^o of the partner responsible: 7

N^{o}s of other partners involved: 1,2,3,4,8

Table 12: Person-months by partner within workpackage

Partner	1	2	3	4	5	6	7	8
2000	6	0	1.2	0	.	.	0	0
2001	17	2	3.7	2	.	.	0	3
2002	1	7	3.0	2	.	.	0	2
Total to date	24	9	7.9	4	.	.	0	5
Total planned	24	12	4	16	.	.	7	8

Time to completion of workpackage (current estimate): 12

Summary This package is on schedule. For example, though D3.1.1 is not complete in the form of a published paper, it is available in the form of a computer program. Similarly, work towards D3.1.2 has resulted in a battery of tests, which have indicated that a complete revision of statistical assumptions needs to be undertaken for fisheries data. D3.1.3 is considered complete since the programs have been developed.

Objectives

- To identify the probability distributions, likelihood functions and prior distributions appropriate to mathematical descriptions of fish population dynamics. *Status of objective:***Work ongoing**
- To specify the range of circumstances under which different methods of parameter estimation may be used, together with guidance on the calculation of appropriate confidence statements. *Status of objective:***Work ongoing**
- To model and program simulation procedures describing the marine ecosystem, using a detailed, structured approach. *Status of objective:***Met**
- To define and test methods to verify goodness-of-fit measures for general likelihood functions in a highly nonlinear framework. *Status of objective:***Work underway**
- To list, evaluate and propose methods for statistical testing of hypotheses in Gadget. *Status of objective:***Work ongoing**

Description of work conducted in to date

MRI (1):
Work on the identification of model problems has been continued as has the development of methods to estimate weights on likelihood components (F.1). The methodology for estimation of weights is considered complete but some further work is needed on the model diagnostics.

Maximum likelihood estimation is the method of choice for the problem at hand, in light of the computational requirements, which pretty much eliminates Bayesian methods. The suggested methods of (iterative) weighting (F.1) appear to work and are therefore the recommended method for adjusting default weighting, rather than using pure likelihood functions, as these have been shown to give considerably erroneous results when weights are incorrect and the likelihoods are not correctly chosen. This summary of MRI work describes the completion of D3.1.3.

Earlier MRI work has led to the completion of D3.1.1.

IMR (2):
Most of the IMR work done here relates to the objective ‘To model and program simulation procedures describing the marine ecosystem, using a detailed, structured approach’. This is a prerequisite for Deliverable 3.1.3, and involves work on the computer code as well as the documentation.

Have conducted an examination of various aspects of the Fleksibest model, looking both for bugs and for more general issues. Have identified and fixed a series of bugs in Gadget.

DIFRES (3):
Part of the model described in Annex D.3 in last years report has been changed or modified: 1. The model for other food has been changed in order to fit the stomach data. 2. Some of the vulnerability parameters have been extended to include quarterly variations. 3. The Dirichlet distribution has been included as an option for modelling the relative stomach content. How this can be done is described. The changes are described in the revised version of Lewy and Vinther (2003) in Annex G.1.

SCUI (4):
Further work on parameterization of a migration model and identification of parameters which can be estimated from field and tagging data. Investigations of available data on temperature, food density (not available as yet) and of observations - direct or indirect- of the spatial distribution of capelin in the Barents Sea and Central Atlantic. Work started on appropriate statistical methods for estimating parameters and calculating migration matrices.

IFREMER (8):
A literature review was carried out for bootstrap methods that could be used for incorporating data uncertainty into the parameter estimation process by Gadget. Some examples will be tested soon.

Deliverables

D3.1.1: Mathematical descriptions of estimation methods for model components. Q8 **Complete**

D3.1.2: Descriptions of goodness-of-fit tests for composite likelihoods. Q8 **Complete**

D3.1.3: Implementations and tests of different estimation procedures in different scenarios. Q12 **Complete**

D3.1.4: Algorithmic description of estimation methods and goodness- of-fit tests to facilitate programming both within and outside of Gadget. Q12 **Complete**

Milestones and expected results

This is work towards milestones M3, M5 and M6, with prototypes towards M3 reached earlier than half-way through the workpackage which was planned.

3.2 Workpackage: Programming estimation

Start date or starting event: July, 2001

N^o of the partner responsible: 7

N^{o}s of other partners involved: 1,2,3,4,6,8

Table 13: Person-months by partner within workpackage

Partner	1	2	3	4	5	6	7	8
2000	6	1	3.1	0	.	0	0	0.0
2001	26	2	3.5	0	.	0	0	1.0
2002	1	4	3.0	0	.	0	2.4	1
Total to date	33	7	9.6	0	.	0	2.4	2
Total planned	36	12	4.0	16	.	4	6	3.5

Time to completion of workpackage (current estimate): 12

Summary This workpackage is considered complete. Minor tests and specific parametrisation of minimisation algorithms will continue throughout 2003.

Objectives

- To program statistical (maximum likelihood) estimation of unknown parameters in these simulations. *Status of objective:***Met**
- To evaluate several different minimization algorithms and select algorithms for use with Gadget. *Status of objective:***Work ongoing**
- To obtain a general version of Bayesian analysis as an option in Gadget. *Status of objective:***Dropped**
- To incorporate processes formulated in Tasks 2 and 4 into the collection of modules. *Status of objective:***Work ongoing**
- To evaluate the possibilities of using parallel processing on a network of processors when estimating very many parameters in a complex multispecies spatially disaggregated model. *Status of objective:***Met**
- To evaluate the effects of incorporating automatic differentiation in the minimisation algorithms in Gadget and implement the possibility. *Status of objective:***Met - will not be used; replaced by parallel computing**

Description of work conducted to date

MRI (2):
The effects of parallel processing are being evaluated in order to estimate the improvements in run-times for models of different sizes on different-size networks. This work has been continued in 2002 and into 2003, with expected completion in fall, 2003.

Given how well the development of parallel methods has developed these are seen as a highly feasible method for estimating parameters in the nonlinear models under consideration. Having considered some of the possibilities of incorporating automatic differentiation within Gadget, the conclusion is that (a) the existing parallel solutions are adequate and (b) incorporating automatic differentiation would require extensive changes to the code which is already massive, without obvious gains, since possible speedup can not be guaranteed to exceed that obtained from the parallel implementation.

IMR (2):
Examined the suitability functions for the catch, and their interaction with the general Fleksibest model. Identified a number of problems. Fixed those which could cause the program to crash. Also identified a number of more subtle non-differentiability issues in Gadget. It is not clear which of these need fixing, and which are merely 'features' of the model. Prioritised the identified issues, with an explanation of their different possible impacts on Gadget. Consideration of these issues is ongoing at MRI.

DIFRES (3):
The likelihood part for the stomach content observations has been implemented and combined with the catch models which completes the programming of the stochastic multispecies model using AD Model Builder. The program includes three options for the likelihood function associated with stomach content observations: Independent normal distributions, independent log normal distributions and the Dirichlet distribution.

The single species part of the model has been tested by simulation. Analyses including the full model using data for the North Sea have been initiated. Catch data for ten species from 1975-2000 and stomach content data for 1991 have been analysed. It seems difficult to obtain the Hessian matrix for calculating uncertainties associated with estimates of the biological parameters. We are still working on this problem.

SCUI (4):
Not yet started.

FRS (6):
K. Patterson was supposed to do Bayesian Analysis. This subject has been abandoned since he left the project. No other suggestions have been made for this deliverable.

CEFAS(7) & IFREMER (8):
The Bayesian implementation of a single species single area length-based model was started using the freely available software BUGS (www.mrc-bsu.cam.ac.uk). The intention is to apply the model simulated data and then compare the parameter estimates with those obtained using Gadget. Compilation problems were encountered caused by the complexity of the model. BUGS works with conditional distributions and uses Markow Chain Monte Carlo sampling to obtain a sample from the joint posterior distribution. The use of conjugate priors is highly recommended as it speeds up the process. Further work on specifying likelihood functions and prior distributions is necessary.

Deliverables

D3.2.1: Program modules to carry out estimation for likelihood and Bayesian estimation. Q8 **Likelihood estimation complete, Bayesian dropped**

D3.2.2: A selected (composite) minimisation algorithm. Q12 **Complete**

D3.2.3: A program which can utilise parallel processing for improved performance. Q8 **Complete**

D3.2.4: An evaluation of the effect of parallelisation. Q12 **underway**

D3.2.5: A set of standard fisheries examples that can be used to check that algorithms are correctly implemented and that results are consistent with those obtained using commercially available software. Q10 **Complete**

Milestones and expected results

This is work towards milestones M3 and M6, with deliverables for M3 finished earlier than scheduled.

4 Workpackage group: Estimation of parameters outside program

4.1 Workpackage: Feeding/consumption

Start date or starting event: February, 2000

N^o of the partner responsible: 5

N^{o}s of other partners involved: 2, 3, 4

Table 14: Person-months by partner within workpackage

Partner	1	2	3	4	5	6	7	8
2000	.	1	1.3	0	2	.	.	.
2001	.	1	12.1	0	26	.	.	.
2002	.	1	12.3	1	26	.	.	.
Total to date	.	3	25.7	1	54	.	.	.
Total planned	.	3	8.0	4.0	58	.	.	.

Time to completion of workpackage (current estimate): 6

Objectives

- To develop a multi-species spatially explicit feeding/consumption model based on habitat and diet selection with an evolutionary fitness basis. *Status of objective:***Met**
- To estimate sampling error of stomach contents data in the NS Stomach Sampling data base. *Status of objective:***Met**

Description of work conducted to date

IMR (2):
IMR has prepared data (capelin, cod, temperature) for input to spatially explicit feeding and consumption model developed by UiB. IMR has also provided advice to UiB on how to use stomach content data to calculate the consumption of herring and capelin by cod in the Barents Sea, and taken part in the interpretation of the results of such calculations. This work relates to deliverables 4.1.1, 4.1.2 and 4.1.3.

DIFRES (3):
Estimation of the distribution of relative stomach content

The work with analysing the probability distribution and especially the variance structure of the mean stomach diet composition for five predator species available from the ICES Stomach Sampling Project 1991 has continued. This work is crucial in order to formulate the likelihood function of the stomach observations.

The distributions are obtained by bootstrapping the stomach observations. One main result is that several types of distributions for specified predator/prey combinations are found: Unimodal, bimodal, delta distributions with positive zero probabilities, symmetric and asymmetric distributions. The “dirty” distributions in this connection, i.e. the delta and the bimodal distributions are found in connection with seldom eaten prey.

Theoretically, stomach observations, mean stomach proportions for a predator, are correlated because they sum to one and if two prey species frequently are present in the same stomach. The analyses indicated that in most cases the correlations are insignificant. However, in the case of the “other group” prey species, which may amount up to 30% of the stomach content, significant correlations with other prey items were found.

The distributions implemented in the stochastic multispecies model (see WP 3) were the normal, the lognormal and the Dirichlet distribution. When using these models the “dirty” distributions are ignored. On might assume that the wrong distribution used in these cases do not seriously affect the parameter estimates because the stomach content is small. However, this would need a further investigation.

The work is presented in a draft paper (Vinther and Lewy 2003) in Annex G.2.

The analyses are almost completed and a draft paper is presented. The paper will be finalised.

Estimation of fish consumption by stomach content analysis

The bias in the method currently used to estimate food intake for use in the stochastic multispecies model for the North Sea has been investigated. Substanital bias was found to occur in a number of cases. In particular, simplifications in the calculation of total intake leads to overestimation of food intake. The method used to distribute total food intake on different prey species tends to overestimate large prey which occurs in a minor proportion of the stomachs, whereas the intake of small and common prey as the group ’other food’ is underestimated. The bias is caused by the effect of the unequal distribution of prey in the stomachs and other food in the stomach on the evacuation of the individual prey items. A new method is presented which takes account of these factors. This work is presented in draft (Rindorf 2003) in Annex G.3.

SCUI (4):
Not yet started.

UiB (5):
The work conducted to date relates to deliverables D4.1.1, D4.1.2, and D4.1.3. The work related to D4.1.1 and D4.1.2 is to develop a multi-species spatially explicit feeding/consumption model. The biological model is individual-based and relies on habitat and diet selection with an evolutionary fitness basis. Consumption is estimated explicitly through the interactions of the fish with its environment, and metabolic costs are estimated using bioenergetics models. Temperature is derived from observations as well as ocean circulation models. An important element in the predator-prey modelling using bottom-up approaches as here, is the modelling of movement behaviour. Different methods for incorporating migration have been attempted during the project including artificial neural networks (Huse and Giske, 1998; Huse et al., 1999), adapted random walk (Huse, 2001), and rule based approaches. In the previous year the focus has been on developing models for adaptive predator-prey interactions in general, and specific models for the interaction between cod, capelin and herring. The general predator-prey work has resulted in a manuscript that is currently in review (Huse, submitted). This study is motivated by the interaction between cod and capelin in the Barents Sea, but the realism is kept at a modest level, and instead the focus is put on developing individual-based models for studying predator-prey interactions. The results show that IBMs are indeed a valuable tool for studying predator-prey interactions, and for example the study illustrates the vulnerability of prey homing to known spawning areas, to which predators easily adapt their migrations. The predator-prey model has been further developed into a specific cod-capelin-herring model. This model uses a rule-based approach for modelling fish movements. The rules are a combination of known features about the migration patterns of the stocks, and fitness maximising rules form life history theory. This rule-based approach is applied in deliverables 4.1.1. and 4.1.2, see annex. The results show that cod consumption of capelin and herring is highly dependent on the degree of spatial overlap between the species. In the model this depends on the rules specified to direct movement of the species. The approach is promising, but more work is needed to study the effect

of different movement rules on the growth, mortality and spatial dynamics of the target species. The computer code used in the modelling is available upon request. It may be interfaced with GADGET by running the current program separately, but this connection has not yet been made.

The aim of D4.1.3 is to provide estimates of cod's consumption of capelin and herring. Consumption of herring is estimated using a new method for estimating consumption in piscivorous fish developed in this project (Johansen et al. 2003). Estimates of consumption of capelin are under processing. Levels of tmax (Johansen et al. 2003) are estimated for different size groups of capelin, giving reasonable results. The remaining part of the estimation procedure is calculation of consumption by areas and seasons. This will be done in the same way as for herring, but with extrapolation schemes specific to capelin. When the extrapolation schemes are decided on, total consumption estimates for 1992-1997 will be produced. This will be finished during the first half of 2003. The consumption estimates will eventually be related to abundance of prey, and climate. Temperature fields and abundance fields for cod, capelin and herring in the Barents Sea have already been produced for use in deliverables D4.1.1 and D4.1.2. These cover the period autumn 1990 to autumn 1996. The temperature and abundance fields are based on CTD and survey data, and are calculated in a regular grid of 20X20 km squares. For the temperature data, a spatial interpolation technique was applied. The temperature and abundance field will also be used in workpackage 2.1 for modeling migration of Barents Sea capelin.

During the previous annual DST2 meeting in Sete, a meeting between the different groups involved in studying fish migrations within the project was planned. This meeting was held in Bergen on 17 and 18 November 2002. At the meeting different approaches for modelling fish migrations were presented and discussed. The group agreed to perform a formal comparison of the different modelling approaches using the Barents Sea capelin as a case study. This species was chosen because of the good availability of capelin data, and forcing data such as temperature, and also because capelin has been the target of previous modelling studies. The output of the comparison should be a report describing the different approaches and recommendations for the implementation in GADGET. The agenda for the meeting and a description of the model comparison agreed at the meeting is enclosed.

Annexes:
1. Huse, G. (MS). Deliverables 4.1.1 and 4.1.2. of DST2. Annex G.4
2. Report from DST2 meeting in Bergen on fish migration. Annex K.3

References:
Huse, G. 2001. Modelling habitat choice in fish using adapted random walk. Sarsia 86: 477-483.
Huse, G. (submitted to The American Naturalist). Adaptive predator and prey migration in a large scale environment.
Huse, G., and Giske, J. 1998. Ecology in Mare Pentium: an individual-based spatio-temporal model for fish with adapted behaviour. Fish. Res. 37: 163-178.
Huse, G., Strand, E., and Giske, J. 1999. Implementing behaviour in individual-based models using neural networks and genetic algorithms. Evol. Ecol. 13: 469-483.
Johansen G.O., Bogstad B., Mehl S., and Ulltang Ø. 2003. Consumption of juvenile herring (Clupea harengus L.) by Northeast Arctic cod (Gadus morhua L) in the Barents Sea: a new approach to estimating consumption in piscivorous fish. In: Johansen, G.O. 2003. The predator-prey interaction between Northeast arctic cod (Gadus morhua L) and juvenile Norwegian spring-spawning herring Clupea harengus L.) in the Barents Sea. Dr. scient. Thesis, University of Bergen.

Deliverables

D4.1.1: Age- and size-dependent growth and predation mortality for cod, capelin and juvenile herring in the Barents Sea. Q9 **Complete**

D4.1.2: Distribution of cod, capelin and juvenile herring in the Barents Sea. Q9 **Complete**

D4.1.3: Estimates of cod's consumption of capelin and herring related to prey density, stock overlap and physical factors. Q9 **Ongoing, herring complete**

D4.1.4: Estimation of the distribution or the variance of the relative stomach contents. Q9 **Complete**

D4.1.5: Final vital statistics to Gadget: Q12 **Ongoing, UiB and SCUI**

Milestones and expected results

This was initially foreseen as work towards milestones M3 and M5, though this may not end up being required for completion of M3.

4.2 Workpackage: Spatio-temporal scales

Start date or starting event: January, 2003

N^o of the partner responsible: 3

N^{o}s of other partners involved: 1, 4, 7

Table 15: Person-months by partner within workpackage

Partner	1	2	3	4	5	6	7	8
2000	0	.	1.3	0	.	.	0.0	.
2001	0	.	0	0	.	.	0.0	.
2002	3	.	0	2	.	.	3.1	.
Total to date	3	.	1.3	2	.	.	3.1	.
Total planned	6	.	10.0	6	.	.	10.8	.

Time to completion of workpackage (current estimate): 12

Objectives

- To investigate the adequate spatial and temporal scales to be used in Gadget. *Status of objective:***On hold**
- To explore the importance of spatial inhomogeneity for the assessments North Sea fish stocks. *Status of objective:***Work ongoing**
- To identify the extent of the spatial and temporal distribution of species within the Celtic Sea. *Status of objective:***Work ongoing**

Description of work conducted to date

This workpackage is not due to start until 2003.

MRI (1):
Work has not started on this workpackage.

DIFRES (3):
As described in last years report the MSVPA has been used to estimate suitability coefficients for the Northern and Southern North Sea separately. Stock and catch predictions will be carried out both for the entire North Sea and for the two areas in order to quantify possible differences.

SCUI (4):
Initial investigations of appropriate spatial and temporal scales to use in a migration model and the sensitivity of results to these spatio-temporal resolution. Investigations of statisitcal measures to characterize spatial distributions and their dependence on scale.

CEFAS (7):
Spatio-temporal mapping and statistical analysis of English groundfish survey data were carried out for several fish species in the Celtic Sea (cod, haddock, whiting, megrim, hake, monkfish, Nephrops, blue whiting, horse mackerel and mackerel). The purpose of this work was to provide a description of species distributions (including different patterns for mature and immature fish based on length at 50% maturity), in order to evaluate the most appropriate spatial units for a spatially explicit Celtic Sea model (see Annex I.2).

Generalised Additive Models (GAM) were used to model the probability of presence, using a binomial family with a logit link function. This work is preliminary and will continue over the final year of the project; we hope to model the presence and absence of fish eggs and larvae in the Celtic Sea based on tri-annual fish egg surveys.

Deliverables

D4.2.1: Effects of different levels of disaggregation in Gadget. Q14

D4.2.2: Quantification of the importance for fish stock assessment of spatial disaggregation. Q14 **Complete**

D4.2.3: (This deliverable was a duplicate of 4.1.4 and has therefore been dropped)

D4.2.4: Estimates of biological parameters required as inputs to WP 5.2. Q10 **Complete**

Milestones and expected results

This is a part of the work towards several milestones (M3-M6).

4.3 Workpackage: Reference points

Start date or starting event: January, 2002

N^o of the partner responsible: 3

N^{o}s of other partners involved: 1, 2

Table 16: Person-months by partner within workpackage

Partner	1	2	3	4	5	6	7	8
2000	0	0	5.0	.	.	.	.	.
2001	0	0	1.1	.	.	.	.	.
2002	0	0	0	.	.	.	.	.
Total to date	0	0	6.1	.	.	.	.	.
Total planned	6	3	12.0	.	.	.	.	.

Time to completion of workpackage (current estimate): 6

Objectives

- To develop multispecies biological reference points. *Status of objective:***Met**
- To indicate how medium-term simulations can be conducted in Gadget. *Status of objective:***Work not started**

Description of work conducted to date

This workpackage started in 2002.

MRI (1):
Work on medium-term predictions will start at MRI in 2003.

IMR (2):
So far, no work directed towards dst2 has been done by IMR.

DIFRES (3):
Deliverable D4.3.1 is considered completed with the Canadian Journal paper of Collie and Gislason (2001) and the paper of Collie et al. submitted to ICES Journal of Marine Science. A draft version of the last paper was presented in last years report.

Collie, J.S. and Gislason, H. (2001). Biological reference points for fish stocks in a multispecies context. Can. J. Fish. Aquat. Sci. Vol. 58, no. 11, pp. 2167-2176.

Collie, J.S., Gislason, H. Using AMOEBAs to Integrate Multispecies, Multifleet Fisheries Advice.

Deliverables

D4.3.1: Definition of multispecies reference points for North Sea fisheries and sustainable multi fleet fisheries. Q12 **Complete**

Milestones and expected results

Results from this task include multispecies reference points and proposals on how to conduct medium-term simulations from Gadget.

5 Workpackage group: Case studies

5.1 Workpackage: Case study: Icelandic waters

Start date or starting event: July, 2001

N^o of the partner responsible: 1

N^{o}s of other partners involved:

Table 17: Person-months by partner within workpackage

Partner	1	2	3	4	5	6	7	8
2000	1	.	.	.	.	.	.	.
2001	18	.	.	.	.	.	.	.
2002	39	.	.	.	.	.	.	.
Total to date	58	.	.	.	.	.	.	.
Total planned	96	.	.	.	.	.	.	.

Time to completion of workpackage (current estimate): 12

Objectives

- To implement prototype models for Icelandic waters, based on a common program base and the data warehouse for the region. *Status of objective:* **Work ongoing, rough prototype ready**

Description of work conducted to date

MRI (1):
As reported earlier, prototype datasets have been generated and prototype multispecies model runs have

been developed, resulting in several requirements for changes to the base Gadget model, mainly with respect to the likelihood functions. Work on this WP in 2002 emphasized getting a prototype DW in place, now complete along with new prototype model runs based on extractions from the DW.

Work in 2003 will emphasize getting these models into shape, so that they do not get rejected by common goodness-of-fit tests.

Deliverables

D5.1.1. Prototype data sets for the area around Iceland. Q6 **Complete**

D5.1.2. Prototype model run. Q7**Complete**

D5.1.3. Proposed generic model improvements (to be programmed in workpackage 3). Q9 **Complete**

D5.1.4. Prototype data warehouse with data for this case study based on definitions in workpackage 1. Q9 **Complete**

D5.1.5. Adopted model run on adopted data set: Proposed model explains data adequately. Q16

D5.1.6. Answers to case study questions: Hypothesis tests conducted with Gadget and have obtained directions on importance of complexity in models. Q16

Milestones and expected results

This workpackage is a part of the work towards milestones M4, M5 and M6.

5.2 Workpackage: Case study: Celtic Sea

Start date or starting event: July, 2001

N^o of the partner responsible: 8

N^{o}s of other partners involved: 7

Table 18: Person-months by partner within workpackage

Partner	1	2	3	4	5	6	7	8
2000	.	.	.	.	.	.	0	3.0
2001	.	.	.	.	.	.	5	2.7
2002	.	.	.	.	.	.	3.86	9.7
Total to date	.	.	.	.	.	.	8.86	15.4
Total planned	.	.	.	.	.	.	11	15.0

Time to completion of workpackage (current estimate): 12

Objectives

- To implement models for selected species in the Celtic Sea that incorporate both spatial and temporal information on the stocks. *Status of objective:***Work ongoing**
- To validate model system components and incorporate into a prototype system for the evaluation of closed areas. *Status of objective:***Work ongoing**
- To consider management parameters such as total biomass by area, spawning stock biomass (SSB) by area, and catch by metier by area. *Status of objective:***Work ongoing**

Description of work conducted to date

Dates of interim meetings: Nantes (28-29 January 2002), Lowestoft (November 2002)

CEFAS (7) & IFREMER (8):
Analysis of French fishing activity (IFREMER) (Annex I.1)

The initial analysis of the spatial and temporal distribution of the French fishing effort (reported on last year) was published (Mahevas and Trenkel 2002a) and some results presented at the ICES Annual Science Conference (Mahevas and Trenkel 2002b). The data analysis was then continued in order to obtain homogenous areas and seasons using cluster analysis for the coefficients of the spatio-temporal model (see annex I.1). The result are a great number (16) of spatial and temporal (7) units. As these are too many units to be realistically taken account of in the multispecies model, it was decided to concentrate at the moment on one spatial unit, with the possibility to use nearshore and an offshore zone later on (see below).

Mahevas, S. et Trenkel, V.M. 2002a. Utilisation de modèles linéairs mixtes pour décrire la distribution spatio-temporelle de l'effort de pêche de la flottille française en mer Celtique. Journal de la Société Française de Statistique, 143, 177-186.

Mahevas, S. et Trenkel, V.M. 2002b. Using mixed models to describe the spatio-temporal allocation of effort of the French fleet fishing in the Celtic Sea. ICES Annual Science Conference, 2-5 Octobre 2002, ICES CM 2002/V:17, 5 pp.

Métaireau C., 2002. Déscription de la dynamique spatio-temporelle de l'effort de pêche en mer Celtique. D.E.S.S. Modélisation et Analyse Statistique de l'Information, Université de Bretagne Sud - Vannes. 43 p.

Comparison of French & UK survey data

The comparison of the consistency of the survey data obtained by the two countries, started in 2001, was completed and a manuscript was submitted (Trenkel et al submitted).

Trenkel V.M., Pinnegar, J.K., Rochet, M.-J. et Rackham, B. Why different surveys provide similar pictures of community structure but not of individual populations. ICES Journal of Marine Science (submitted)

Running GADGET
Initial single-species, single-area GADGET models within Linux (version 7) have been constructed for cod (UK) (Annex I.6), blue whiting (UK) (Annex I.5) and whiting (France) (Annex I.4). Lacking data input formats and relevant examples for multispecies Gadget models, the three individual models could not be combined in 2002. The absence of tools for easy analysis of results and the verification of model hypotheses, made the use of Gadget and the interpretation of the obtained results difficult. Some initial results from the fitted models are are presented for whiting in Annex I.4. This completes deliverable D5.2.1 and is a step towards D5.2.2-D5.2.3.

Stomach content analyses
Initial analyses of stomach content data were carried out (Annex I.3), including examination of feeding patterns of predators with increasing length (in cm). It is intended that analyses will continue on this data-set in order to calculate consumption rates and to evaluate multi-species "functional responses" (the way a predator's feeding rate is influenced by prey availability). This work will be of direct relevance in our efforts to incorporate biological interactions in the Celtic Sea GADGET model (based on some of the principles described for MULTSPEC by Bogstad et al. 1997).

Bjarte Bogstad, Kjellrun Hiis Hauge and Øyvind Ulltang (1997) MULTSPEC - A multi-species model for fish and marine mammals in the Barents Sea. J. North. Atl. Fish. Sci., Vol. 22: 317–341.

Deliverables

D5.2.1. Prototype data sets for the Celtic Sea. Q6 **Complete**

D5.2.2 Comparison of externally estimated parameters with values obtained using the GADGET modelling approach. Any difference in the estimates obtained to be explored. Q12 **Ongoing**

D5.2.3 Adopted model run for each species based upon agreement with existing historical data and knowledge of the fishery. Q16

D5.2.4 Comparative evaluation of simple model approach (single species) and complex model approach (multi-species). Q14

Milestones and expected results

This workpackage is a part of the work towards milestones M4, M5 and M6.

5.3 Workpackage: Case study: North Sea herring

Start date or starting event: Completion of North Sea Herring components of the Data Warehouse.

N^{o} of the partner responsible: 6

N^{o}s of other partners involved:

Table 19: Person-months by partner within workpackage

Partner	1	2	3	4	5	6	7	8
2000	.	.	.	.	.	0	.	.
2001	.	.	.	.	.		.	.
2002	.	.	.	.	.	2	.	.
Total to date	.	.	.	.	.	2	.	.
Total planned	.	.	.	.	.	50	.	.

Time to completion of workpackage (current estimate): 12

Objectives

- To implement prototype models for North Sea herring, based on a common program base and a data warehouse for the area and species in question, and to investigate the appropriateness of current short- and medium-term management approaches. *Status of objective:***Work ongoing**

Description of work conducted to date

This workpackage was due to start mid-year 2001 when prototype runs are due to start.

FRS (6):
The software has been installed on the machine in Aberdeen. A run with data from herring has been done with J. Begley in spring 2002. This run cannot be considered as a prototype model

Simulated data have been analysed using GADGET in order to test and get experience with the model. A presentation has been done in 2002 at the meeting in Sete and the results are available in Annex J.1. FRS has proposed certain methods of simulating data for testing purposes. These simulation methods are being developed at MRI, thus completing D5.3.2.

No run has already been done of the final database. FRS is waiting for the scripts that create GADGET file to be available.

Deliverables

D5.3.1. Prototype model run. Q7 **Delayed, ongoing**.

D5.3.2. Proposed generic Improvements (to be programmed in WP 3). Q9 **Complete**.

D5.3.3. Adopted model run on adopted data where the model explains data adequately. Q16

D5.3.4. Evaluations of key project objectives for case study:

- Validation of conventional estimates of stock size and exploitation rates using the ’gadget’ modelling approach, and exploration of main points of difference. Q16
- Comparative evaluation of complex-model and simple-model approaches to estimating short-term catch forecasts and stock sizes. Q16
- Validation of existing management approaches for North Sea herring. Q16

Milestones and expected results

This workpackage is a part of the work towards milestones M4, M5 and M6.

A Deliverables

This section provides summaries of all deliverables, ordered according to their due status, indicating their completion status.

It must be noted that in several cases the definition of “complete” is subjective. The approach taken is to list a deliverable as “complete” if a product is available. In some cases the product is a required prototype, in other cases a final product was originally envisaged but it has been found that a prototype is sufficient at present and advancement on other workpackages can proceed as planned. In some cases (e.g. deliverable 1.1.1) a prototype is available but is not enough and in these cases the delay or revised schedule is duly noted.

More detail on each deliverable in relation to the corresponding workpackage is given in the body of the report.

A.1 Status of due deliverables

The following table lists the deliverables due by the end the third year, not earlier reported as complete, along with their due quarter and completion status.

WP	D#	Deliverable	Due	Status
1.1	D1.1.1	Corrected, documented data base for Icelandic waters.	Q5	Complete
1.1	D1.1.2	Corrected, documented data base for North Sea herring.	Q5	Complete
1.6	D1.6.3	CORBA technology described in detail and tested.	Q6	Complete
1.6	D1.6.4	The results of using CORBA and XML compared.	Q7	Completed in revised form: Distributed Postgresql data warehouses established as basis.
1.6	D1.6.5	Final choice of DW software set up, tested and distributed.	Q8	Complete
1.6	D1.6.6	Data warehouse for Icelandic waters.	Q9	Complete
1.6	D1.6.7	Data warehouse for North Sea herring.	Q9	Ongoing - Input complete
1.6	D1.6.9	Data warehouse for the Celtic Sea.	Q9	Complete
1.7	D1.7.1	Software to set up all data and parameter files for Gadget.	Q9	Completed through the generation of DW extraction routines and modifications of Gadget input file formats. Revision: Not all parameter files will be generated as this is impractical. Rather, sets of standard/basic parameter files are available, but data files are generated.

2.1	D2.1.2	A proposed mathematical (migration) model formulation for use in Gadget.	Q12	Complete
2.1	D2.1.3	Evaluation of reduction in variances as a result of reducing uncertainty in data sources.	Q12	Started, using bootstrapping MRI and IFREMER.
2.2	D2.2.1	A module to describe the recruitment process.	Q12	Ongoing. Generic work on the reqruitment process ongoing at IMR. Mathematical modelling and implementation of closure of life cycle in Gadget ongoing at MRI, to be completed in Q15.
2.3	D2.3.1	Mathematical models of the growth and maturation process.	Q12	Complete
3.1	D3.1.1	Mathematical descriptions of estimation methods for model components.	Q8	Complete (see Annex E.2)
3.1	D3.1.3	Implementations and tests of different estimation procedures in different scenarios.	Q12	Complete
3.1	D3.1.4	Algorithmic description of estimation methods and goodness-of-fit tests.	Q12	Complete (Estimation methods completed in Annex E.2 and g.o.f. reported in Annex D2 in 2002 report)
3.2	D3.2.1	Program modules to carry out estimation for likelihood and Bayesian estimation.	Q8	Complete (Bayesian part dropped)
3.2	D3.2.2	A selected (composite) minimisation algorithm.	Q12	Complete (up to selection of parameters in deliverable 3.2.4)
3.2	D3.2.4	An evaluation of the effect of parallelisation.	Q12	Underway, to be completed summer 2003
3.2	D3.2.5	A set of standard fisheries examples that can be used to check that algorithms are correctly implemented.	Q10	Complete

4.1	D4.1.1	Age- and size-dependent growth and predation mortality for cod, capelin and juvenile herring in the Barents Sea.	Q9	Complete
4.1	D4.1.2	Maps of horizontal distribution of cod, capelin and juvenile herring in the Barents Sea.	Q9	Complete
4.1	D4.1.3	Estimates of cod's consumption of capelin and herring related to prey density, stock overlap and physical factors.	Q9	Ongoing, herring complete
4.1	D4.1.4	Estimation of the distribution or the variance of the relative stomach contents.	Q9	Complete (documentation pending)
4.1	D4.1.5	Final vital (consumption) statistics to Gadget.	Q12	Ongoing in a joint subproject between UiB and SCUI
4.2	D4.2.4	Estimates of biological parameters required as inputs to WP5.2	Q10	Complete
4.3	D4.3.1	Definition of multispecies reference points for North Sea fisheries.	Q12	Complete

5.1	D5.1.3	Proposed generic model improvements (to be programmed in task 3).	Q9	Complete
5.1	D5.1.4	Prototype data warehouse with data for this case study based on definitions in task 1.	Q9	Complete
5.2	D5.2.2	Comparison of externally estimated parameters with Gadget.	Q12	Ongoing
5.3	D5.3.1	Prototype model run for North Sea herring.	Q7	Delayed, revised pace.
5.3	D5.3.2	North Sea herring. Proposed generic Improvements (to be programmed in WP 3).	Q9	Complete

A.2 Earlier deliverables, now completed

The following table lists those deliverables which were due and completed before this year of reporting.

WP	D#	Deliverable	Due	Status
1.1	D1.1.4	Corrected, documented prototype data base for the Celtic Sea.	Q5	Complete
1.2	D1.2.1	DW design.	Q5	Complete
1.3	D1.3.1	Algorithms and a set of SQL programs to be used for taking the raw data and converting the data into summaries useful for ecosystem modelling.	Q7	Complete
1.4	D1.4.1	Prototype data sets for Gadget and prototype data warehouse views.	Q3	Complete
1.5	D1.5.1	Design of data warehouse views.	Q4	Complete
1.5	D1.5.2	Design of the database structures.	Q4	Complete
1.6	D1.6.1	Each institute provides descriptions of the data storage format used.	Q6	Complete.
1.6	D1.6.2	Known techniques of DW implementation using XML are described and tested	Q6	Complete
2.1	D2.1.1	Initial, simple migration model.	Q4	Complete
2.4	D2.4.1	Mathematical formulation of general populations dynamics models in state space form.	Q3	Complete
2.4	D2.4.2	Corresponding general population dynamics program modules.	Q4	Complete
2.4	D2.4.3	Prescription to incorporate relevant prior information into population dynamics models.	Q4	Complete
3.1	D3.1.2	Descriptions of goodness-of-fit tests for composite likelihoods.	Q8	Complete
3.2	D3.2.3	A program which can utilise parallel processing for improved performance.	Q8	Complete
5.1	D5.1.1	Prototype data sets for the area around Iceland.	Q6	Complete.
5.1	D5.1.2	Prototype model run (Iceland).	Q7	Complete
5.2	D5.2.1	Prototype data sets for the Celtic Sea area.	Q6	Complete

A.3 The next steps

The following table lists the remaining deliverables, due in 2003.

WP	D#	Deliverable	Due	Status
4.2	D4.2.1	Effects of different levels of disaggration in Gadget.	Q14	
4.2	D4.2.2	Quantification of the importance for fish stock assessment of spatial disaggregation of the North Sea	Q14	Complete
5.1	D5.1.5	Adopted model run on adopted data set.	Q16	
5.1	D5.1.6	Answers to case study questions.	Q16	
5.2	D5.2.3	Adopted model run for each species.	Q16	
5.2	D5.2.4	Comparative evaluation of simple model approach (single species) and complex model approach (multi-species).	Q14	
5.3	D5.3.3	North Sea herring. Adopted model run on adopted data set.	Q16	
5.3	D5.3.4	North Sea herring. Evaluations of key project objectives for case study.	Q16	

B Datawarehouse Technology

B.1 Demonstration of CORBA technology for a distributed database

Vojtěch Kupča (`vojtech@hafro.is`)
Marine Research Institute, Reykjavík, Iceland
July 2002 (update January 2003)

This paper discusses the status and problems of the current DST^2 DW CORBA approach, a possible solution and an alternative solution to this task is also proposed.

1 Introduction

At the beginning of the DST^2 project it was proposed that CORBA should be used to demonstrate the linking of several databases containing fish related data. CORBA is a specification that defines platform and language independent distributed application framework. Free implementations are for instance: *Orbit*, *MICO*, *TAO* or *JacORB*; the last is based on Java and is convenient to use because it reduces programming complexity compared to C/C++ based variants (first three on the list).

2 Problems of CORBA approach

There are several problems with the CORBA approach that were not thought of until recently when more detailed descriptions of the requirements became available. The main issues are:

1. design of a real distributed database (DD) system
2. transparent combination of queries
3. network failure handling
4. network traffic optimization
5. access to other institutes' networks
6. security

Within this project it is not possible to address all DD functionality. One can learn more on this topic at project Mariposa address:

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

Mariposa is a halted five year old project that was aimed at creating a distributed version of PostgreSQL and only reached the alpha version. Within DST^2 it may be possible to demonstrate database linking on a small subset of queries that solve a couple of particular tasks. Going further is not feasible due to the above stated technical obstacles. In more detail, the query combination problem is a process of combining several partial queries into the resulting query as if only one large database was queried. For possible DST^2 application only several query types can be handled: for instance `SELECT a, b, sum(c) as c FROM table WHERE ... GROUP BY a,b`. As can be seen, problems arise if it is necessary to combine queries that use aggregate functions e.g. `COUNT` or `SUM`. More complex queries involving subselects, unions, etc. will be most likely not be supported. Next it would be necessary to be able to handle network failures on the level of application (typical example is a link going down when sending data across the network). Also, it was estimated that based on a parameter setup and implementation details, there might occur substantial network traffic (anywhere from 10KB up to 1MB) sometimes causing a noticeable delay at the user interface. Finally political problems with access to other institutes' databases (for example allowing certain data ports in firewall setups) and security issues (as a default data would be passing Internet in the unencrypted form) have to be dealt with.

3 Solution

A possible solution that could fit within the mentioned limits seems to be a CORBA sever-client model where servers would query local databases and send results to the client that requested them. The client in turn would have to combine the results as if a single database was queried. Afterwards, the resulting data would have to be written to the socket using an agreed protocol. Then, using the same protocol the extraction php program would have to read the data from this socket, process them and display them in an appropriate form. The extraction php program would have to be changed so that it does not directly connect to the local database. It would only fetch the combined result from the socket.

Advantages:

- demonstration of CORBA
- platform and language independence

Disadvantages:

- implementation cost
- not a complete and general solution
- need to modify extraction php program
- potentially slow

4 Alternative Solutions

An elegant alternative solution seems to be the usage of asynchronous replication (implemented as a free program PostgreSQL Replicator) where the data could be asynchronously (using definable replication periods) replicated across the network from one site to another. Replication models supported are, apart from hybrid ones, also *master-slave* and *read/write on local partition*. The second one gives all involved institutes the opportunity to manage their own data, since noone else would be able to modify rows that were added by a given site.

Advantages:

- relatively easier to setup then CORBA
- strong security using `scp` (secure copy)
- compact and plannable network load (data transfered in bursts)
- client is fast and can be programmed as a local application
- client can be anywhere on the Internet
- general solution, works even when some sites are down
- easy backup

Disadvantages:

- legal issues concerning data availability from other institutes (it could be helpful to specify what operations are allowed with the non-local data)

If we stay with the distributed version solution and demonstrate only a subset of possible queries, then one may also consider only using remote querying of the database using classical PHP PostgreSQL API (or *dblink* which is available directly from PostgreSQL). This is possible as only one database system will be used (PostgreSQL) on one type of OS (Linux), where the advantages of CORBA are pushed back a little.

For a matter of completeness it is sufficient to say that ideas about extending database drives (either ODBC or libpq) were discarded because of the complexity of this task.

Finally, as a simple non-CORBA solution, one can also consider a setup in which a user queries remote databases using a web interface, whose responsibility is to contact and retrieve data from selected sites and present them to the user using HTTP protocol and sockets. This solution would be commonly used unless one wanted to test CORBA technology.

5 Priorities

The priority of the DW tasks in the project are as follows:

1. local export of DW data into file sets suitable for Gadget input
2. local institutional DW access (stock assessments, etc.)
3. **demonstration of distributed database linking**
4. local (later maybe distributed) public access to the DW data

The final implementation method and any other topic mentioned in this paper are subjects to further discussion.

6 Current status

It was decided that the approach described in section 3 will be implemented. It is necessary to bear in mind that the system was developed at the University of Iceland and has not yet been tested by MRI. At MRI there exists a php interface software ("interface") which talks to the CORBA part ("query system") over sockets using a simple communication protocol. This protocol uses the following commands:

- `RESET`: reset query system
- `READY`: query system is ready
- `HOSTLIST` *hostlist*: list of hosts the next query will be sent to
- `QUERY` *query*: query to be sent
- `RESULT` *result*: result of the query (combined from all hosts as if one system was queried)
- `CLOSE`: close connection to query system

For a complete example of communication see the following Appendix section. Testing of functionality of this protocol was possible using a C program which was simulating the query system in that it was sending dummy commands of the described protocol over the network. The result was then properly displayed as an HTML page. The current selection options of the interface include selection based on year, month, species, gearclass, vesselclass lengthcell, lengthcellmin, lengthcellmax, agemin, agemax. Possible output types for now are age-length keys (`ALK`), age-length distributions with average lengths at age (`ALD`) and catches in numbers, total landings in tons and catches in numbers by age (`CNR`).

To summarise the status of work as of January 2003:

- First version of query system exists but was not tested at MRI.
- Web interface part, whose task is to send user requests to query system and present returned data, exists and was tested at MRI to work with a C language program simulating query system behaviour.
- Since this system is not yet working as a whole, it has not been installed anywhere. Because of the restrictive FRS network policy it will most likely not be possible to install the system there. However, an attempt will be made to setup a functional system at IFREMER in Nantes. There are several technical issues that can prevent the query system from working correctly. Mainly these are open high communication ports of client-server network applications which could collide with restricted ranges of ports on local firewall systems. Security issues of sending data over the Internet should also be addressed.

Appendix

Example of communication between query system and web interface.

```
CORBA query system                  PHP interface program
-----------------------------------------------------------------
                                    connect();
                                    write('RESET');
read('RESET');
reset();
write('READY');
                                    read('READY');
                                    write('HOSTLIST h1 h2 ... hN');
                                    write('QUERY query\n');
read('HOSTLIST ...');
read('QUERY ...');
doQuery(hostlist, query);
write('RESULT rows columns
       colname1 ... colnameN
       data1 ... dataN\n', 4096);  read('RESULT ...', 4096);
                                    write('CLOSE\n');
                                    exit();
read('CLOSE');
listen();
```

Remarks:
The command `HOSTLIST` is followed by a space separated list of hosts which the next query should be sent to. `QUERY` can be either one that returns data or one that returns a true or false result (like query `DROP TABLE tmp;`). Based on this, `RESULT` can be of two forms: **a)** `RESULT rows columns colname1 ... colnameN data1 ... dataN`, where `rows` and `columns` are the number of rows and columns in result, `colnameN` is the name of the N-th column, and `data1 ... dataN` are data arranged in a sequence of rows. **b)** `RESULT 0` or `RESULT 1` for (un)successful boolean result.

B.2 DST^2 data warehouse structure and data import

Vojtěch Kupča (`vojtech@hafro.is`), Peter Sandbeck (`pes@dfu.min.dk`)

Marine Research Institute, Reykjavík, Iceland

January 2003

This paper describes the data warehouse (DW) structure, the procedure to import data from institutional databases and the way to access data through a web-based interface.

1 Introduction

To enable easier and potentially faster access to the institutional databases, containing scientific data on fish, a solution of a DW with aggregated statistical quantities was proposed. In the first place, it is necessary to briefly define our usage of the term *data warehouse*: in general DW is regarded as a special type of arrangement of one or more databases that contains 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 Appendix A). In the case of MRI, the extraction and transformation is done by querying the institutional database (with the help of views) and storing data in an ASCII file exchange format (AFEF). This ensures that other institutes can import data into the same structure of DW. The structure was agreed upon during previous DST^2 meetings and finetuned with respect to the DW in the last few months. After inspecting the structure of the DW (see next 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 as a rough guideline, it is fully acceptable to call our projected databases DWs, with the possibility of building real DMs on top of them, if required later on in the project.

2 Data warehouse structure

At present the DW is structured around 47 tables (3 'management' tables, 3 tables to support Gadget extraction, 23 lookup tables and 19 data tables), that split up as indicated in Table 1.

As can be seen from the summary table (Table 1) four new tables were added that contain stomach content data and two tables (18 and 19) for incomplete data that do not fit common DW structure. Stomach content tables create a data hierarchy similar to that of the biological samples.

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 (eg 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.

Not shown in the above list are tables `d_maxages, d_maxleng, l_areasofareas` which contain static data (on minimum and maximum lengths and ages for a given species and areas of DW divisions and subdivisions) used in Gadget output routines. A detailed description of the DW structure can be found at:
`http://www.hafro.is/~vojtech/dw`

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 (not used yet)
10	`TAG`	tagging	data on fish tagging
11	`JUV`	juvenile	juvenile data (not used yet)
12	`SPE`	species	info on species, stocks and market categories of catch or catchsample 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 and IFREMER

Table 1:

3 Data import

Data import is done by the PHP[1] program `dst2up.php`. This program recognizes AFEF and uses static column definitions from file `keydef.php` (which, in fact, represents a fixed mapping between AFEF and DW) to correctly insert or update rows in the DW.

General AFEF file characteristics:

- 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`
- after the table code a tab separated list of values corresponding to each table follows

For a complete AFEF specification see Appendix C. The upload of data has to follow the hierarchy of the DW. This, for example, means that data on age (lowest table in hierarchy) cannot be uploaded until the corresponding information in higher-level tables is available. In such a case the upload program will recognize the missing reference from the higher table and report an error. Therefore, the tables must be imported in the correct order, starting from eg. table `SAM`, followed by `CAS, LEC, AGE`, as

---

[1] PHP is an interpreted scripting language mainly aimed at creating dynamically generated webpages, but can also be used as an interpreter in a command line environment.

appropriate for the hierarchy of tables (see Appendix B). Standalone tables `MSC, TAG, ALK, STM` (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 constraint 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[2].
- bug in upload program (use switch `-v` and report an error)

As 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[3] 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:

```
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
```

Using `dst2up.php` it is also possible to delete data from whole tables. This process includes deleting data *and* deleting all management information from the `KEYS` table. When you are deleting data manually from the `psql` console, always be sure about what you are doing. With many records, deletion may take considerable time. This is most probably due to the continuous reorganizing of the affected indexes after every deletion operation. In the future, one can think about dropping and recreating all necessary indexes before and after the deletion. It is also intended that all deletion capabilities are moved to a separate script.

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

---

[2] 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 in that 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 $p = 10^{-8}$).

[3] 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.

```
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[4]. 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 a 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`.

4 Raw data 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 of complex multiline queries within seconds.

5 Current status

The following list shows the status of the technical part of the DW as of January 2003:

- **DW structure**: The DW tables and structure are finalized. If there should be a need for the future addition of table(s) it should be possible to extend the existing structure.
- **Upload program**: Upload program is complete. Based on experiments with a large number of AFEF data files it was possible to remove some bugs during the development phase.
- **Code** : Ideally, with a better understanding of functionality of the PostgreSQL database and its features and the capabilities of PHP, it would be desirable to clean or even recode parts of the upload program for better readability of the code.
- **Lookup tables**: Content of the lookup tables was being updated as new data files kept arriving. The last few stomach content related lookup tables are to be completed later. In other aspects, the lookup tables are complete.
- **Deletion**: The issue of deletion from the DW has to be considered as well. It is already possible to consistently delete whole tables from the DW. However deletion based on year, month, species, batchid, etc. will require more sophisticated method of deletion to be devised. Due to referential integrity constraints, the hierarchical structure of the DW, and the way of managing references among rows, this is not a trivial task.
- **Icelandic waters DW**: The DW has been installed and data uploaded – a full description is in Annex C.1

---

[4] 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.

- **Celtic Sea DW**: In May 2002, the DW was successfully installed in Nantes, however some issues still remain: **a)** unresolved question of remote access to Celtic Sea data[5]; **b)** undecided way for Lowestoft to access database.

---

[5] An ideal solution would seem to be access by secure shell (`ssh`), which enables logging into a remote machine and executing commands on a remote machine over an encrypted communication channel within an insecure network (like e.g. Internet).

Appendix A

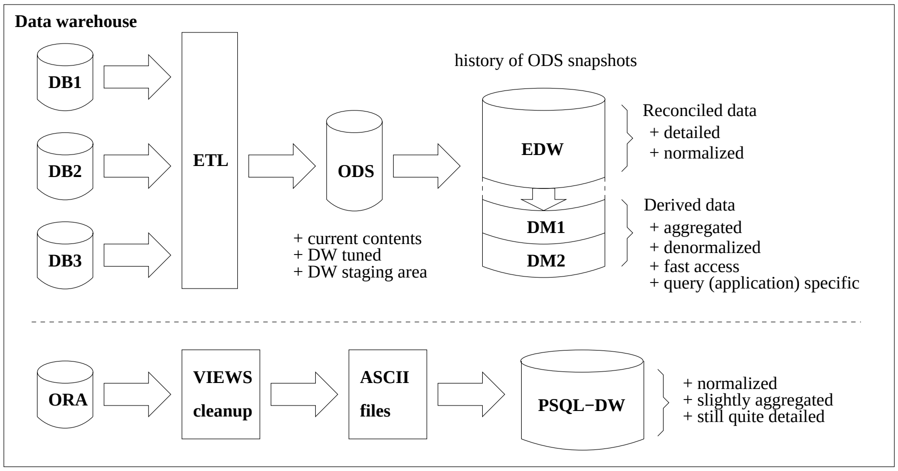


Figure 1: **General data warehousing architecture and the correspondence to the internal MRI structure.** Terms: ETL: extract-transform-load process, ODS: operational datastore, EDW: enterprise data warehouse, DM: data mart, ORA: Oracle database, PSQL-DW: PostgreSQL data warehouse.

Appendix B

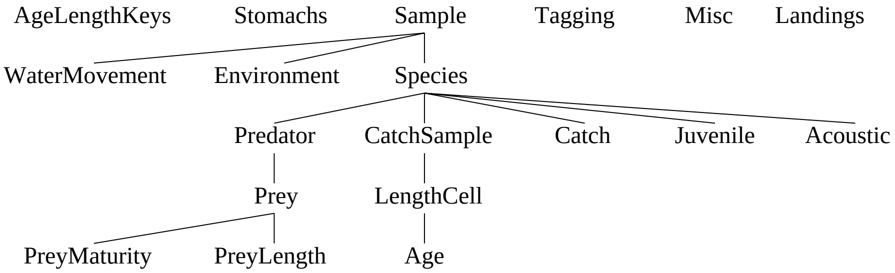


Figure 2: **DW tables hierarchy.** Note that table Species is never directly uploaded, but is implicitly populated on the upload of one of the tables one level lower in the hierarchy. This also means that to upload one of the tables Catch, CatchSample, Juvenile, Acoustic or Predator it is enough to have a corresponding row in table Sample.

Appendix C

ASCII file exchange format (AFEF)

DW tables can be split into several basic categories: **a) 1, 12**: tables holding position, time and species information; **b) 5, 6, 7**: biological samples data; **c) 2, 3**: environmental information; **d) 4**: catch data; **e) 13, 14, 15, 16**: stomach content data; **f) 8, 10, 17, 18, 19**: standalone tables for special purposes; **g) 9, 11**: acoustic, juvenile data.

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

ASCII file exchange format (AFEF) for DW

                                 01/2003

General remarks:

1) columns are separated by tabs
2) star (*) denotes key atribute set
   that has to be unique
3) NN denotes NOT NULL atribute
   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* region            NN
  6* division          NN
  7* subdivision
  8* gridcell
  9* vesselclass
 10* vesselsubclass
 11* gearclass
 12* gearsubclass
 13* week              NN
 14* depthstratum      NN
 15  temperature
 16  salinity
 17  zooplankton

TABLE 3: WaterMovement

  0  WAM
  1* institute         NN
  2* year              NN
  3* quarter           NN
  4* month             NN
  5* region            NN
  6* division          NN
  7* subdivision
  8* gridcell
  9* vesselclass
 10* vesselsubclass
 11* gearclass
 12* gearsubclass
 13* week              NN
 14* depthstratum      NN
 15  exchangecoef
 16  divisionfrom      NN
 17  divisionto        NN

TABLE 4: Catch

  0  CAT
  1* institute         NN
  2* year              NN
  3* quarter           NN
  4* month             NN
  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  tripnum
 18  powermean
 19  grtmean
 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             NN
  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             NN
  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

TABLE 7: Age

  0  AGE
  1* institute         NN
  2* year              NN
  3* quarter           NN
  4* month             NN
  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

TABLE 8: Misc

  0  MSC
  1* institute         NN
  2* year              NN
  3* quarter           NN
  4* month             NN
  5* region            NN
  6* division          NN
  7* subdivision
  8* gridcell
  9  species           NN
 10  stock
 11  sexcode
 12  age               NN
 13  sampleingtype     NN
 14  areaggregate      NN
 15  sourceyear
 16  marketcategory
 17  vesselclass
 18  vesselsubclass
 19  gearclass
 20  gearsubclass
 21  maturitystage
 22  fishmeasurednum   NN
 23  fishlengthmean    NN
 24  fishweightmean
 25  surveyindex
 26  surveyindexcv
 27  catchnum
 28  stocknum
 29  diseaserate

TABLE 9: Acoustic

  0  ACO
  1* institute         NN
  2* year              NN
  3* quarter           NN
  4* month             NN
  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  areaaggregate
 17  number            NN
 18  biomass           NN
 19  signal

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


TABLE 11: Juvenile

  0  JUV
  1* institute         NN
```

```
  2* year              NN
  3* quarter           NN
  4* month             NN
  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  areaaggregate     NN
 17  caughtnum
 18  m2num

TABLE 12: Species

  0  SPE
  1* institute         NN
  2* year              NN
  3* quarter           NN
  4* month             NN
  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             NN
  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
 26  proportionfeeding

TABLE 14: Prey

  0  PRE
```

```
  1* institute         NN
  2* year              NN
  3* quarter           NN
  4* month             NN
  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
 27  propbynumber
 28  propbyweight

TABLE 15: PreyLength

  0  PRL
  1* institute         NN
  2* year              NN
  3* quarter           NN
  4* month             NN
  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
 28  propldist

TABLE 16: PreyMaturity

  0  PRM
  1* institute         NN
  2* year              NN
  3* quarter           NN
  4* month             NN
  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  weightnumber      NN
 29  meanpreyweight
```

TABLE 17: Landings

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

TABLE 18: AgeLengthKeys

```
  0  ALK
  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* species           NN
 14* stock
 15* marketcategory
 16* samplingtype
 17* samplingstrategy
 18* lengthcell        NN
 19* sexcode
 20* maturitystage
 21* age               NN
 22  agenum            NN
 23  calcmethod
```

TABLE 19: Stomachs

```
  0  STM
  1* institute         NN
  2* startyear         NN
  3* endyear
  4* quarter
  5* month
  6* region            NN
  7* division
  8* subdivision
  9* gridcell
 10* predspecies       NN
 11* predlength        NN
 12* preyspecies
 13* preylength        NN
 14  prednumber
 15  preynumber
 16  propbynumber
 17  preyweight
 18  propbyweight
```

B.3 DST^2 datawarehouse output for Gadget

Vojtěch Kupča (`vojtech@hafro.is`)
Marine Research Institute, Reykjavik, Iceland February 2003

This report deals with extraction of data from the datawarehouse (DW), specifically with respect to input for the ecosystem simulation program *Gadget*. The most up-to-date information can be obtained from the MRI website [1].

1 Data flow

From the point of view of data flow in the DST^2 project, several dependent parts exist. 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 the DW (`dst2up.php`)
6. extraction of data into Gadget formats (GF) (`dst2ext.php`)
7. Gadget simulation using extracted GF data
8. output of simulation, postprocessing, presentation

Points 3-5 are realized by the php scripts `dst2exp.php` (data export the from institutional Oracle database) and `dst2imp.php` (data import into the DW PostgreSQL database). These scripts help the end user hide some of the technical aspects of data export/import part. This paper mainly contains updated and extended information on data extraction from the DW, with stress on point 6 in the above list. To access the DW one can choose from two basic possibilities:

a) since the DW is implemented in a PostgreSQL database it is possible to use the native command line interface *psql* as a filter. An example would be a query sent directly from Unix shell environment:

```
echo "SELECT * FROM sample" | psql -d oko1
```

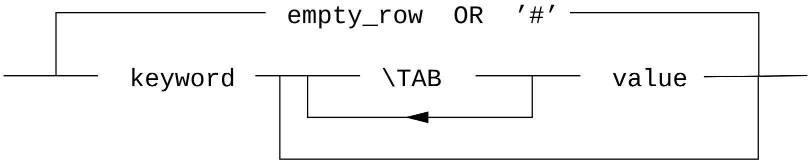
This, however, requires some prior knowledge of the structure of the DW tables. It is also necessary to know the format of the Gadget input files, to be able to modify the output of the previous command appropriately. This way of accessing DW can come in handy in certain situations such as checking correctness 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. `dst2ext.php` has a simple command line syntax:

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

2 Control file

As can be seen from above, the complexity of the 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. This is the syntax diagram of the control file:



If no value is assigned to a keyword, the default value is used. Table 1 contains a complete list of allowed keywords that can occur in the control file.

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 key output (for description of all Gadget filetypes see Appendix A), 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 a comma. However, certain parameters allow multiple *lists* (separated by a space) to be assigned to them. For example one may be interested in an age-length key split by commercial fishing season (spring, autumn). In such a case the output will have two parts, one for each 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 Appendix B. 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
```

3 Group file

Group file holds aliases for one or more lists of values of a keyword. Group file has a 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,`

keyword	brief description	example
General parameters		
`hostname`	hostname of the database server	`haffugl`
`database`	database name 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`
`years`	year or list of 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	
`preyspecies`	prey species ICES code or 3-letter abbrev.	`1`
`digestionstage`	digestion stage DW code	
Gadget-specific		
`alkeyscount`	used in initial files; reports alk numbers	`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 tag file	`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`
`tagid`	label used in tag file	`tagid1`
`tagloss`	tag loss rate used in tag file	`0`
`temperature`	temperature in area file	`5`
`zerotimestep`	switch to omit zero datasets in output	`1`

Table 1: Control file keywords Remark: Lookup tables for `preygroup` and `digestionstage` are yet to be defined.

`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]) and `GSTO` (`maturitystage`) filetypes (for information on filetypes see next section). 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:

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

On the other hand a 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
101,102,103,104,105,106,107,108,109,110,111,112,113,114
```

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. A complete example of a group file is in Appendix C.

4 Filetypes

As mentioned before, the keyword `filetype` specifies the type of output file. So far the following functions have been implemented:

a) for Gadget:
age-length keys (`GALK`), length distributions (`GLD`), age distributions (`GAD`), average length at age (`GAVL`), average weight at age (`GAVW`), age-length keys split by stock (`GSTO`), average weight by lengthcell (`GAVG`), area file (`GARE`), time file (`GTIM`), total landings (`GFLT`), initial age-length keys and length distributions (`GINI`), average length for recruited fish (`GREC`), prey biomass as proportions (`GPRE`), predators by lengthcell (`GPRD`), tagging file (`GTAG`), tagged fish by lengthcell (`GTAD`), recaptured fish (`GTAR`).

b) for stock assessment:
age-length table (`ALT`), age-length key matrix (`ALK`), age-length distribution matrix and length distribution (`ALD`), catches in numbers by age (`CNR`).

The difference between **a)** and **b)** is in the formating of the output. 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

---

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

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

`lengthcell, lengthcellmin, lengthcellmax, agemin` and `agemax`. For instance in a case of `GALK` filetype, setup

```
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       age7    len4    5
1998    1       1       age8    len1    0
1998    1       1       age8    len2    1
1998    1       1       age8    len3    0
1998    1       1       age8    len4    0
1998    1       1       age9    len1    0
1998    1       1       age9    len2    0
1998    1       1       age9    len3    0
1998    1       1       age9    len4    0
```

We see that 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 length group has added 'technical' 0.01 to ensure a non-empty interval:

```
len1    45      47
len2    47      49
len3    49      51
len4    51      51.01
```

Based on the type of outputo 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 6).

For filetypes `GALK, GLD, GAD, GSTO, GAVL, GAVW` a file `likelihood` is also created with likelihood component information. 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`.

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

```
; [ALT]
; 450   470     490     510
3       3       10      5
0       1       0       0
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, herring = 5, capelin = 5, other species = 10). This ensures that lengthgroups are never smaller than what the database has to offer. If `lengthcell` is not specified, 10mm is used as the default value.

5 Debugging

For the purpose of debugging and inspecting in detail all actions of `dst2ext.php` there exists an optional parameter `-v`. It displays all SQL queries that were sent to the database to obtain the 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.

6 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 include: `GSPE, GALKc, GALKs, GLDc, GLDs, GADc, GADs, GAVLc, GAVLs, GAVWc, GAVWs, GSTOc, GSTOs, GAVG, GREC, GINI, GARE, GTIM, GFLT, GPRE, GPRD`. The idea behind this 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 `c` (catch) or `s` (survey) appended to their names. `dst2input.sh` takes the specified control file and in the case of catch files substitutes the 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 rewrites `gearclass` and `gearsubclass` empty. With `GAVG` (average weights), `samplingtype 130,135` (IGFS, autumn survey) is taken. For other filetypes unchanged parameters are used. Corresponding filenames are created with suffixes `.survey_name` (for survey files) and `.fleet_name` (for catch files) which are substituted for the first value of control file keywords `surveynames` and `fleetnames` respectively.

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 Appendix D.

7 Installation and system requirements

For all parts of data extraction to work several general technical requirements have to be met:

1. installed latest PostgreSQL database [2]
2. installed latest PHP [3] (compiled with option `-with-pgsql`, that includes PostgreSQL database API by using `libpq.so` from PosgreSQL directory tree)
3. installed PHP data extraction program

MRI installation

At present it is possible to setup data extraction at MRI on Sun machines using the following procedure (directory `~/dw` can be changed to anything convenient for user):

1. `mkdir ~/dw`
2. `cp -R ~vojtech/dw/Gadget/* ~/dw`
3. `cd ~/dw`
4. `dst2input.sh`

Then it is possible to edit `Control/dst2ext.ctl` or create a new control file or group file as necessary. Locations of control and group files are freely adjustable.

8 References

[1] For description of available filetypes, lookup tables, AFEF and other information concerning DW see:
`http://www.hafro.is/~vojtech/dw`

[2] for the lastest version of PHP (as of writing 4.3.0) see:
`http://www.php.net`

[3] for the lastest version of PostgreSQL (as of writing 7.3.2) see:
`http://www.postgresql.org`

Appendix A

Gadget input filetypes

```
+-----------------------------------------------------------------------------+
| type  | action | description    | parameters      | joining | files         |
|-------+---------------------------------------------------------------------|
| GALK  |                                                                     |
|-------+---------------------------------------------------------------------|
|       |        | selects number |                 |         |               |
|       |        | of aged fish   |                 |         |               |
|       |        | to obtain      |                 |         |               |
|       |        | age-length     |                 |         |               |
|       |        | table, for     |                 | SAM >   | *.alkeys.*    |
|       |        | stock          |                 | SPE >   | len.agg       |
|       | alkeys | assessment     | all             | CAS >   | age.agg       |
|       |        | also           |                 | LEC >   | area.agg      |
|       |        | converting     |                 | AGE     | likelihood    |
|       |        | numbers to     |                 |         |               |
|       |        | proportions to |                 |         |               |
|       |        | get age-length |                 |         |               |
|       |        | keys           |                 |         |               |
|-------+---------------------------------------------------------------------|
| GLD   |                                                                     |
|-------+---------------------------------------------------------------------|
|       |        |                |                 | SAM >   | *.ldist.*     |
|       |        | selects length |                 | SPE >   | len.agg       |
|       | length | distributions  | all             | CAS >   | area.agg      |
|       |        |                |                 | LEC     | allage.agg    |
|       |        |                |                 |         | likelihood    |
|-------+---------------------------------------------------------------------|
| 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        |                 |         |               |
|       |--------+----------------+-----------------+---------|               |
|       | alkeys | see GALK       |                 |         | *.adist.*     |
|       |--------+----------------+-----------------+---------| age.agg       |
|       | sld    | see GALD       |                 |         | area.agg      |
|       |--------+----------------+-----------------+---------| alllength.agg |
|       |        | computing sums |                 |         | likelihood    |
|       | agesum | at ages from   |                 |         |               |
|       |        | sld output     |                 |         |               |
|-------+---------------------------------------------------------------------|
| GAVL  |                                                                     |
|-------+---------------------------------------------------------------------|
|       | alkeys | see GALK       |                 |         |               |
|       |--------+----------------+-----------------+---------|               |
|       | length | see GLD        |                 |         |               |
|       |--------+----------------+-----------------+---------|               |
|       | sld    | see GALD       |                 |         |               |
|       |--------+----------------+-----------------+---------|               |
|       |        | compute        |                 |         |               |
```

```
|       |        | average        |                 |         |                |
|       | avgald | lengths at age |                 |         |                |
|       |        | from sld       |                 |         |                |
|       |        | output         |                 |         | *.meanle.*     |
|       |--------+----------------+-----------------+---------| age.agg        |
|       |        | computes       |                 |         | area.agg       |
|       |        | standard       |                 |         | likelihood     |
|       |        | deviation at   |                 |         |                |
|       | stddev | age based on   |                 |         |                |
|       |        | data from sld  |                 |         |                |
|       |        | and avgald     |                 |         |                |
|       |        | outputs        |                 |         |                |
|       |--------+----------------+-----------------+---------|                |
|       |        | computes       |                 |         |                |
|       | number | number of fish |                 |         |                |
|       |        | at ages from   |                 |         |                |
|       |        | sld            |                 |         |                |
|-------+-----------------------------------------------------------------------|
| GAVW  |                                                                       |
|-------+-----------------------------------------------------------------------|
|       | alkeys | see GALK       |                 |         |                |
|       |--------+----------------+-----------------+---------|                |
|       | length | see GLD        |                 |         |                |
|       |--------+----------------+-----------------+---------|                |
|       | sld    | see GALD       |                 |         | *.meanwg.*     |
|       |--------+----------------+-----------------+---------| age.agg        |
|       |        | selects        |                 | SAM >   | area.agg       |
|       |        | average        |                 | SPE >   | likelihood     |
|       | avgwgt | weights and    | all             | CAS >   |                |
|       |        | variances at   |                 | LEC >   |                |
|       |        | age            |                 | AGE     |                |
|       |--------+----------------+-----------------+---------|                |
|       | number | see GAVL       |                 |         |                |
|-------+-----------------------------------------------------------------------|
| GSTO  |                                                                       |
|-------+-----------------------------------------------------------------------|
|       |        | see GALK,      |                 |         |                |
|       |        | includes       |                 |         | *.alkstock.*   |
|       |        | stocknames to  |                 |         | len.agg        |
|       | alkeys | produce alkeys |                 |         | age.agg        |
|       |        | output split   |                 |         | area.agg       |
|       |        | by maturity    |                 |         | likelihood     |
|       |        | stage and/or   |                 |         |                |
|       |        | sex groups     |                 |         |                |
|-------+-----------------------------------------------------------------------|
| GAVG  |                                                                       |
|-------+-----------------------------------------------------------------------|
|       |        | selects        | areas, species, | SAM >   |                |
|       |        | average        | samplingtype,   | SPE >   |                |
|       | weight | weights for a  | lengthcell,     | CAS >   | *.refweights   |
|       |        | given          | lengthcellmin,  | LEC     |                |
|       |        | lengthcell     | lengthcellmax   |         |                |
|-------+-----------------------------------------------------------------------|
| 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     | *.fleet        |
|       |        | on gearclass,  | marketcategory, |         | *.fleet.data   |
|       |        | gearsubclass;  | samplingtype    |         | *.predict.data |
|       |        | 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,       |                 |         | *.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 = 1 from   |                 |         |                |
|       |        | d_minmax       |                 |         |                |
|       |--------+----------------+-----------------+---------|                |
|       |        | selects an     |                 |         |                |
|       |        | average length |                 |         |                |
|       |        | and standard   |                 |         |                |
|       |        | deviation for  |                 |         |                |
|       |        | recruited fish |                 |         |                |
|       |        | at age =       | all except      |         | *.rec          |
|       |        | agemin,        | gearclass and   | SAM >   |                |
|       |        | "first"        | gearsubclass,   | SPE >   |                |
|       | recrui | timestep is    | agemax =        | CAS >   |                |
|       |        | determined by  | agemin,         | LEC >   |                |
|       |        | samplingtype;  | timestep = YEA  | AGE     |                |
|       |        | for years less |                 |         |                |
|       |        | then 1985,     |                 |         |                |
|       |        | average of     |                 |         |                |
|       |        | years          |                 |         |                |
|       |        | 1985-1989 is   |                 |         |                |
|       |        | taken          |                 |         |                |
|-------+----------------------------------------------------------------------|
| GPRE  |                                                                      |
|-------+----------------------------------------------------------------------|
|       |        | selects prey   |                 |         |                |
```

```
|       |        | biomass         |                 |         |                |
|       |        | specified by    |                 | SAM >   | *.prey         |
|       | preywg | preys;          | all (for        | SPE >   | len.agg        |
|       |        | computes        | stomach)        | PRD >   | area.agg       |
|       |        | proportions     |                 | PRE     | prey.agg       |
|       |        | per predator    |                 |         |                |
|       |        | length          |                 |         |                |
|-------+----------------------------------------------------------------------|
| GPRD  |                                                                      |
|-------+----------------------------------------------------------------------|
|       |        | see GPRE,       |                 |         |                |
|       |        | outputs         |                 |         | *.predator     |
|       | preywg | predator by     |                 |         | len.agg        |
|       |        | lengthcell      |                 |         | area.agg       |
|       |        | corresponding   |                 |         | prey.agg       |
|       |        | to GPRE         |                 |         |                |
|-------+----------------------------------------------------------------------|
| GTAG  |                                                                      |
|-------+----------------------------------------------------------------------|
|       |        |                 | years, species, |         |                |
|       |        |                 | timestep,       |         |                |
|       |        |                 | lengthcell,     |         |                |
|       |        | generates       | lengthcellmin,  |         |                |
|       | tag    | tagfile         | lengthcellmax,  |         | tagging        |
|       |        |                 | lengthcell,     |         |                |
|       |        |                 | tagid,          |         |                |
|       |        |                 | tagstock,       |         |                |
|       |        |                 | tagloss         |         |                |
|-------+----------------------------------------------------------------------|
| GTAD  |                                                                      |
|-------+----------------------------------------------------------------------|
|       |        | selects number  | years, species, |         |                |
|       |        | of tagged fish  | areas,          |         | tag.data       |
|       | tagrel | split by        | lengthcellmin,  | TAG     | len.agg        |
|       |        | lengthcell      | lengthcellmax,  |         |                |
|       |        |                 | tagid           |         |                |
|-------+----------------------------------------------------------------------|
| GTAR  |                                                                      |
|-------+----------------------------------------------------------------------|
|       |        | selects         | years, species, |         |                |
|       |        | individual      | areas,          |         | recap.txt      |
|       | tagrec | recaptured      | lengthcellmin,  | TAG     | len.agg        |
|       |        | fish            | lengthcellmax,  |         |                |
|       |        |                 | tagid           |         |                |
+------------------------------------------------------------------------------+
 *   optional type not used by Gadget
 >   natural join
```

Control file example

```
#
# dst2ext control file
#
# following keywords can contain multigroups (i.e. values separated by space):
# years, areas, timestep, gearclass, gearsubclass, vesselclass, vesselsubclass,
# samplingtype, maturitystage, stock, sex, predatormaturity, predatorsex,
# preygroup, digestionstage
#
# for gadget filetypes: except for gearclass/gearsubclass (with GFLT filetype)
# and maturitystage (GSTO filetype) multigroups should not be used in other
# parameters
#
# ---------------------------------------
#       database related info
# ---------------------------------------

hostname        haffugl
database        oko3
user            vojtech

# group file path

groupfile       Control/dst2ext.grp

# ---------------------------------------
#       general setup parameters
# ---------------------------------------

# gadget or stock assessment filetype:
# ALT, ALK, ALD, CNR, CPUE, GALK, GALD, GLD, GAD, GAVL, GAVW, GAVG, GARE,
# GTIM, GINI, GFLT, GREC, GSPE, GTAG, GTAD, GTAR, GSTO, GPRE, GPRD

filetype        GPRE

# years

years           1996

# divisions or subdivisions

areas           IN

# timestep

timestep        YEA

# ---------------------------------------
#       catch data parameters
# ---------------------------------------

# gears and vessels

gearclass
gearsubclass
vesselclass
vesselsubclass

# species
# COD, HAD, HER, CAP, PRA, PLE, WHG, WHB, MEG, SOL, MAC, HOM, MON, NEP, HKB,
# GAD, RED

species         COD

# stocks
# ISI (icelandic shrimp, inshore),
```

```
# ISO (icelandic shrimp, off-shore),
# ISH (icelandic herring),
# NOH (norwegian herring),
# NSH (north sea herring)

stock

# ----------------------------------------
#       survey data parameters
# ----------------------------------------

# samplingtype, maturitystage

samplingtype

# lengthcell info, lengthcell has to be >= step for a given species,
# step is: HER = 5mm, PRA = 0.5mm, other species = 10mm

lengthcell      100
lengthcellmin   350
lengthcellmax   750

# age info, '+' specifies a plus group

agemin          5
agemax          6

# maximum length or age for omitted lengthcellmax and/or agemax parameters
# (amax is also used as an upper bound of age plusgroup), if lmax or amax
# are empty and lengthcellmax or agemax are also empty then an extra query
# is sent to the database to get exact values for a given species

lmax            1500
amax            20

# maturity and sex parameters

maturitystage
sex

# marketcategory and samplingstrategy

marketcategory
samplingstrategy

# stomach contents related parameters

predatormaturity
predatorsex
preygroup
preyspecies
digestionstage

# ----------------------------------------
#       gadget parameters
# ----------------------------------------

#  + fleetnames correspond to gearclass*gearsubclass in numbers and order
#
#  + survey contains list of weights for surveys used in fleet file
#
#  + surveynames correspond to surveys in numbers and order
#
#  + switch is used in prediction file
#
#  + fleetscale is for rescaling gear-based landings using total landings
#
#  + zerotimestep removes (value 0) or keeps (value 1, default) empty data sets
#
#  + prey species list; '*' is used for other species: if included
```

```
#    biomass ratios in each length group have to add up to 1.
#
#  + prefix is used for file references in likelihood, fleet and stock files
#
#  + alkeyscount switches on and off (1 or 0, default 0) appending of an extra
#    column of alks numbers to initial population files

alkeyscount     1
fleetnames      net
fleetscale      1
outputdir       Input
prediction      2006
prefix          cod
preys           CAP HAD *
stockname       had
surveynames     IGFS shrimp
survey          30000 40000
switch          1
tagid           tagid1
tagloss         0
temperature     5
zerotimestep    1
```

Appendix C

Group file example

```
#
# dst2ext group file
#
# definition type, identifier, group(s)
#

division        NS      102,103,104,105,110,111,112,113 101,106,107,108,109,114
division        IN      101,102,103,104,105,106,107,108
division        NSl     102,103,104,105 101,106,107,108

timestep        HLF     1,2,3,4,5,6 7,8,9,10,11,12
timestep        TRD     1,2,3,4 5,6,7,8 9,10,11,12
timestep        QTR     1,2,3 4,5,6 7,8,9 10,11,12
timestep        IGFS    1,2,3

years           98-02   1998 1999 2000 2001 2002
years           78-02   1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988
1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002
years           85-02   1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995
1996 1997 1998 1999 2000 2001 2002
years           82-02   1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992
1993 1994 1995 1996 1997 1998 1999 2000 2001 2002

gearclass       COD-ALL 101,102,103,105,106
gearsubclass    TR      101
vesselsubclass  COM     102,103,104

samplingtype    SURVEY  130,131,133,134,135
```

Appendix D

dst2input.sh

```
Extraction of Gadget input data from DW

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

filetypes: ALL  : all files
           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.$
           GPRE : ^.prey
           GPRD : ^.predator
```

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

C Institute Datawarehouses

C.1 Datawarehouse for Icelandic waters

Lorna Taylor
Marine Research Institute, Iceland.

February 2003

Introduction

The datawarehouse (DW) for Icelandic waters has been established primarily to aid the development of GADGET models. The position of the DW in the dst^2 project is illustrated in fig. 1.

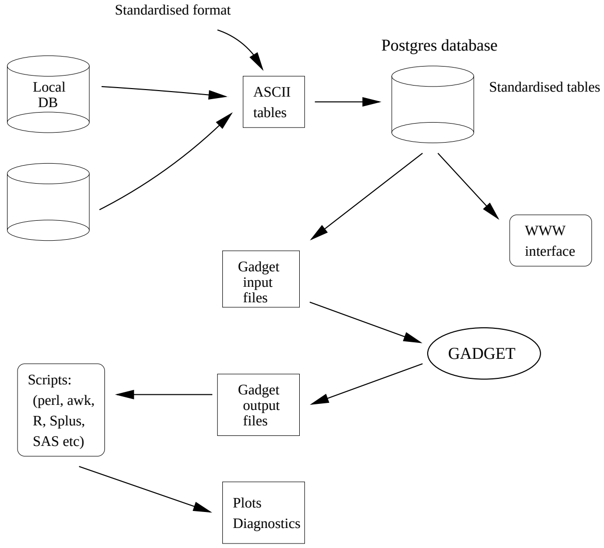


Figure 1: Structure of data processing for the datawarehouse and GADGET

Data are extracted from the institute database, aggregated, error checked and output as ASCII tables (full definition in Appendix C of Annex B.2). These tables are uploaded into the hierarchical datawarehouse (the tabular structure is defined in Annex B.2). From the datawarehouse data can be extracted using PostgreSQL (from the command line or using a browser interface) or using routines which can extract data either into the format required for GADGET or for stock assessment. Use of the DW to create input files for GADGET models simplifies the process of creating new GADGET models.

Database structure

The parts of the datawarehouse covered here are: biological sampling (Station → Species → Sample → Length → Age), stomach tables (Station → Species → Predator → Prey → Prey length/Prey maturity) and log book data (Station → Species → Catch) each of which comprises part of the hierarchical structure, along with mark-recapture (tagging) and landings (Landings) data which are separate tables. The structure of the datawarehouse currently containing data is illustrated in fig. 2. The hierarchical structure means that for data to be uploaded a corresponding record (ie identical unique key) must be available at the level above.

The landings table was not in the original DW definition but it was decided to include it, at least for Icelandic waters, as the existing record of landings is disaggregated by subdivision using log book data. It contains the official landings data by month, species and stock along with vessel and gear class

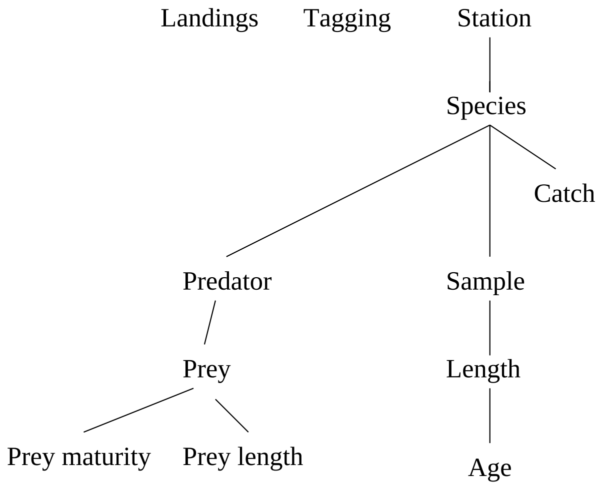


Figure 2: Datawarehouse structure

information.

Data sources

The areas of the database considered so far are: sampling data, log book data, official catch statistics, stomach data and tagging data.

Landings statistics

All data apart from the landings statistics are collected by the Marine Research Institute. Since September 1992, landings statistics have been collected by the Directorate of Fisheries, prior to that data were collected by Fiskifélag Íslands and those from 1982 to 1992 are available as ASCII files. The data collected by the Directorate of Fisheries are stored in an Oracle database.

MRI data

Data at the Marine Research Institute are stored in an Oracle database. This database system has been operating since 1985, with data prior to 1985 being input by year moving back in time. Data are currently being entered for 1975. These data are input into different tabular structures than the post 1985 data and are more thoroughly checked for errors.

The data storage format is similar for most species with capelin and herring the most important exceptions. In the tables recording biological samples (commercial and research) whenever SeaScale (an automatic system used to record and upload data into the institute DB) is used, capelin and herring are input in the standard format. Care needs to be taken that this is taken into account when extracting data.

Position is often known, both at the start and end points of each tow, but for older data sometimes only the statistical rectangle (1 degree longitude, 1/2 degree latitude) and subrectangle (1/2 degree longitude, 1/4 degree latitude), or the rectangle alone, are known. The position of shrimp data is identified by an additional system of area definitions.

Stomach data are stored in two formats: pooled data where the stomachs of fish within a length class were pooled and individual sampling when the length of each fish, often along with age, sex and maturity, are available as well as the stomach contents. Individual sampling became the standard method in 1993.

In 1991 log books were made mandatory for vessels larger than 10 tonnes and in September 1999 for all vessels. Limited log book data also exist for some trawlers back to the 1970's.

Extraction of the data into the ASCII file format

Data are extracted from the MRI Oracle database into the ASCII standard format using sequences of views and mapping tables.

Tables have been created in SQL which map the codes used in the institute DB with those used in the DW. The conversion of codes is then done by joining the appropriate tables. This makes it easy to see and alter the mappings. For the biological sampling, mark-recapture and stomach data all processing of the data is done using a sequence of SQL tables and views.

The codes used in the DW are defined in the Look up tables (http://www.hafro.is/∼vojtech/dw).

Datawarehouse definitions

Spatial scale

The finest spatial resolution at which the Icelandic waters DW stores data is subdivision. Gridcell is defined as a statistical sub-rectangle but no data are currently being stored at this level. The hierarchical spatial structure is described in Annex H.1.

Station positions in the institute database are assigned to a gridcell and mapped from the gridcell to subdivision. Where gridcell is not available, the statistical rectangle is known and an alternative mapping is available for rectangle to subdivision.

In the case of shrimp data, there is an additional set of mappings which can be used to identify subdivision from skíki and fjarðareitur which are spatial units used in the recording of thse data.

Temporal scale

Data are stored by year, quarter and month.

Vessel class and subclass

Vessel class is defined by length $<$ 12m, 12 – 24m and > 24m as for the Celtic Sea and North Sea.

At present, vessel subclasses are research, commercial and foreign. It would be useful to be able to identify some particular types of vessel, such as freezer trawlers and distinguish between boats which are controlled by number of days fishing and quotas, but this cannot be done directly from the database.

Gear class and subclass

Gear class refers to the type of fishing gear, with the range of gear types defined in the institute database considerably reduced. Table 1 contains the DW gear class definitions.

Gear subclass is only used for cod where it is mesh size (inches) for gillnets. For other gear types the mesh size is referred to (at least approximately) by the combination of gear type, year and sampling type.

Sampling type

Sampling type distinguishes between sampling of commercial catches, various surveys and research sampling not associated with a survey. Table 2 contains the DW sampling type definitions.

DW code	Gear type
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

Table 1: Datawarehouse gear class definitions for Icelandic waters

DW code	Sampling type
101	sea sampling
102	harbour sampling
103	sampling by fishermen
110	research, not survey
130	Icelandic ground fish survey
131	shrimp survey
133	0-group survey
134	gillnet survey
135	autumn survey

Table 2: Datawarehouse sampling type definitions for Icelandic waters

Species information

Length divisions

The accuracy at which fish lengths are stored is species dependent (Table 3).

Species	length interval
cod	10mm
haddock	10mm
herring	5mm
capelin	5mm
shrimp	0.5mm
redfish	10mm

Table 3: Length intervals

Maturity stages

Four maturity stages are normally used but shrimp have separate codes and are included as being immature or mature (Table 4):

Stocks

Herring is recorded as either being Icelandic summer spawning or Norwegian spring spawning.

Shrimp are identified as either being inshore or offshore. As the inshore stocks are limited to small

Code	description
1	immature
2	prespawning
3	spawning
4	resting
6	mature
10	immature shrimp
11	mature shrimp

Table 4: Maturity stages

areas, subdivision is sufficient for further definition of inshore stocks in GADGET.

Aggregation of data

The aim of the aggregation is to store data on a scale where the necessity of returning to individual records is minimal but to reduce the size of the database.

Biological sampling

The data are aggregated either by summing or taking the mean. For each record at month, subdivision, gear type, species etc the total number of stations, along with the total number of fish sampled, aged, weighed is recorded. In each lengthcell, sex and maturity stage the number of fish is recorded, along with the mean weight and standard deviation of weight. Similarly, mean weight is recorded for age, length, sex and maturity.

The survey index is calculated to be a scaling factor, taking into account the number of fish counted but not measured and the length of the tow. This is done for the Icelandic groundfish survey, the shrimp survey and the autumn survey.

When ungutted weight is not available but gutted weight is, then gutted weight is used multiplied by a scaling factor.

For capelin and herring data, information on the ship, survey and year is used to identify the data storage method used in the DB, to ensure the data are extracted in the correct form for the DW.

Despite maturity stage and sex being available with the length data in the institute DB, they were not used, as the hierarchical structure of the DW rejected many of these data as they do no agree with the data on maturity and sex in the age table. This is probably due to the data being entered inconsistently. For this reason all data pertaining to maturity stage and sex are taken from the age table in the MRI DB.

Stomach data

The pooled and individual samples are extracted separately but the aggregation methods are similar and the data are output in the same form. The samplingstrategy column is used to indicate the type of data, 105 and 106 indicating individual and pooled stomachs respectively.

The most common prey are identified to species level (Table 5), others are included in as a prey group eg fish, benthos. The grouping system is that used in the institute database.

For each predator cell (ie by subdivision, month, species, age, length etc) the total number and weight of each prey group/species is calculated and the corresponding proportion by weight and number.

When prey lengths are stored the total length distribution of each prey species (in numbers and as a

Latin name	English common name
Gadus morhua	cod
Melanogrammus aeglefinus	haddock
Sebastes marinus	golden redfish
Anarhichas lupus	Atlantic wolffish
Hippoglossus hippoglossus	Atlantic halibut
Hippoglossoides platessoides	long rough dab
-	flatfish
Clupea harengus	herring
Mallotus villosus	capelin
Micromesistius poutassou	blue whiting
Nephrops norvegicus	Norway lobster
Pandalus borealis	deep water prawn
Sebastes viviparus	Norway haddock
Myctophydae	lanternfish
Lumpenidae	blennies
Lycodes sp.	eelpout
Sebastes sp.	redfish
Gadidae sp.	gadoids
Ammodytidae sp.	sand eels

Table 5: Prey included to species or family level in the DW.

proportion) for each predator cell is calculated. Prey sex is also included in this table but is seldom recorded except for capelin in cod stomachs. The final table on prey maturity and weight only includes data on capelin from cod stomachs. The format of the Oracle tables, means that it is not possible to include the maturity and weight data below the length data in the table hierarchy.

Digestion stage cannot be included in the DW, despite being available in the DB. The hierarchy of the DW requires it to be available at all levels and it is limited in the DB.

Mark-recapture data

Unlike the other tables, mark-recapture data are stored as individuals and the spatial resolution is finer. The gridcell of release and recapture (when appropriate) of individual fish is recorded whenever available.

When fish are double tagged, only one tag is recorded in the DW. Only data where the release position is known and stations from which records on uncaptured fish are available are included in the DW.

Log book data

Log book records of catches are used to estimate the catches by subdivision. These are recorded in the table catch both as the values calculated directly from the log books and also scaled to the total landings (by gear type) for the year. When log books are not available, landing harbour indicates the subdivision small boats fished in. This cannot be assumed for larger vessels a longer time series of log books is available for them.

Landings

The landings data are aggregated by month, vesselclass, vesselsubclass, gearclass, gearsubclass, species and stock.

Data validation

An integral part of the data extraction procedure is to check the source data for potential errors. Addition checks are also done outwith the extraction routines. The aim is to notify the database administrator of errors in the institute database and have them changed at source rather than exclude them from the datawarehouse. This allows for data to be checked, improves the institute DB and prevents the exclusion of fish which are merely unusual. When potential errors are found, the original paper records are checked and if necessary the otolith.

Most of the error checks within the extraction routines are done using predefined database tables which contain the bounds for each species and data type.

Tests outside the extraction routines

The coordinates (start and end position) are compared with each other and with the statistical rectangle and subrectangle defined in the database whenever data fall outside the existing area definitions,

The timing and gear type of surveys are checked. This has indicated a problem with the current structure of the database and led to a modification of the institute DB which enables more accurate identification of surveys.

Maturity and sex codes are checked to ensure they only include defined values.

Any problems with the hierarchy are identified when the data are uploaded into the DW.

Generic error checks

Database tables of expected bounds for several data types for all species have been created. These are used for more than one data type.

- Minimum and maximum length - dependent on sampling type.
- Minimum and maximum age.
- Minimum and maximum length at age.
- Minimum and maximum weight at age.

Biological sampling tables

As the institute DB is not strictly hierarchical, it is possible for a length to be recorded at a station in the age file, but not be recorded in the length file. Such errors (which in the past were a side effect of the sampling procedure) are identified. The stations for which more fish have been entered into the age then the length table are also identified.

The problem with the hierarchy also affects the extraction of data relating to maturity and sex.

Stomach data tables

There are some instances where the biological data about predators does not agree with the corresponding data in the age table. It was found that this is a problem with the script to load the stomach data and these fish are not included in the DW.

The lengths of prey are checked and those outwith a set of bounds are excluded from the DW. More than one unit of measurement appears to have been used within a prey species causing a wide range of values to be entered into the DB.

Mark-recapture data

There are instances where the length of recaptured fish is less then the release length. Although these can be identified, in most cases the discrepancy cannot be resolved. The date of recapture is also checked against the date of release. With the herring data there are some problems: in most cases, only returned fish have been entered in the MRI DB and this has limited the number of data which have been included in the DW.

Uploading data

Upload of the ASCII file exchange format (AFEF) data into the DW is possible using the PHP program dst2up.php. This program recognises the AFEF format and DW structure and consistently uploads data into particular DW tables. Problematic lines that could not be uploaded for some reason are reported in an error file for later scrutiny. For further detail on data import see Annex B.2.

Status of data in the datawarehouse

The current status of data in the DW is described in tables 6 & 7. The tables are described in Appendix C of Annex B.2. Some minor alterations will be made to these data and they will be reuploaded as corrections are made to the MRI DB. Data input to the MRI DB will be uploaded regularly to the DW.

Data type	Tables	Species	Years
Biological sampling	1, 12, 5, 6, 7	cod, haddock, capelin, herring	1976 - 2002
Biological sampling	1, 12, 5, 6	shrimp, redfish	1976 - 2002
Stomachs	1, 12, 13, 14, 15, 16	cod, haddock, herring	1979 - 1999
Mark-recapture	10	cod, herring	1991 - 2002
Log book	1, 12, 4	cod, haddock, capelin, herring, shrimp	1982 - 2002
Landings	17	cod, haddock, capelin, herring, shrimp	1992 - 2002

Table 6: Status of data loaded in the DW for Icelandic waters

species	Table 6 LEC	Table 7 AGE	Table 10 TAG	Table 13 PRD
cod	640000	310000	130000	13922
haddock	250000	120000	59	1763
redfish	120000			73
shrimp	120000			
capelin	60000	60000		
herring	40000	40000		
total	1230000	530000	130000	20000

Table 7: Number of rows in the Icelandic waters DW (to the nearest 10000) by table and species.

Extracting data

Data can be extracted from the DW from the command line using PostgreSQL, using php data extraction routines and with a web browser. The extraction routines which are described more fully in Annex B.3 output data either in a format suitable for GADGET or for stock assessment.

GADGET

The extraction procedure for GADGET is described fully in Annex B.3, a description of the revised GADGET file format is in Annex E.1 and an example data set in Annex E.3.

Creating a GADGET model data set using the extraction package requires a set of scripts. Table 8 describes seven sets of commands (each with a filename as indicated) which can be used to output a single species cod model. Likelihood data are from one commercial fleet and two surveys. The likelihood data output from this example are more extensive than would be used in an actual model to illustrate the keywords required for the different data types.

The most important demarcation in the extraction of data is between input data and the likelihood data. In some cases all the input files, excepting the fleet data, can be output from a single script. For the likelihood data, however, a new file is required for each fleet or survey. The prefix is required in all scripts to name the output files.

The data groups used in tables 8 and 9 are described in table 10 and in the examples would be contained in a separate file cod.grp.

To extract data a typical command would be:

```
dst2input.sh single/Control/flt.ctl single/Data GFLT
```

Input files

`GSPE` is a template and requires editing by the user, it assumes that there is renewal and standard names are used for the initial value file, renewal file and aggregation files. Some default data types are used: for `GINI` and `GREC` `samplingtype` 130 is the default, although another survey can be selected; for `GAVG` data from `samplingtypes` 130 and 135 are used and in `GREC` the weight-length relationship coefficients are taken from a look up table. In `GREC` `agemin` is assumed to be the age of renewal, if a different age is required then another script would be created with the required `agemin` and in predicted years, renewal is assumed to equal that of the final year of data selected.

When the fleet data are output, the prediction file is created automatically, as is the fleet file which requires editing. The option `fleetscale` scales the catch of the selected commercial fleets to the total catch by year. The catch in kilogrammes for the surveys is input by the user in `survey` and is a dummy value which is essential as it allows the modelled data to be compared with the survey likelihood data.

As temperature data has not been entered into the DW, temperature for the area file is selected by the user and a single temperature is output for all areas and timesteps.

Likelihood files

Catch and survey likelihood files have different filetypes, the s suffix indicates a survey file and c a catch file. When a survey filetype is selected, the `gearclass, gearsubclass, vesselclass, vesselsubclass` keywords are not used and the extraction is based solely on the selected `samplingtype`. For catch filetypes, `samplingtype` is assumed to be from 101, 102 and 103.

The `zerotimestep` keyword is only used for likelihood data and removes all rows when no data are available for a year, area, timestep combination.

The likelihood file is automatically output along with the likelihood data with each subsequent request for likelihood data appended to the existing likelihood file. The appropriate aggregation files are also output with the likelihood data. As these can be overwritten by subsequent requests, files may need to be renamed during the data extraction procedure.

	Input data			Likelihood data			
filename	init.ext	flt.ext	rec.ext	catch1.ext	catch4.ext	igfs.ext	aut.ext
keyword	initial values	fleets	renewal	catch		survey	
hostname	haffugl	haffugl	haffugl	haffugl	haffugl	haffugl	haffugl
database	oko3	oko3	oko3	oko3	oko3	oko3	oko3
user	vojtech	vojtech	vojtech	vojtech	vojtech	vojtech	vojtech
groupfile	cod.grp	cod.grp	cod.grp	cod.grp	cod.grp	cod.grp	cod.grp
filetype	GSPE, GARE, GTIM, GAVG, GINI	GFLT	GREC	GLDc, GALDc, GAVWc	GALKc, GAVLc	GLDs, GALDs, GAVLs, GAVWs GALKs	GLDs, GALDs, GAVLs, GAVWs GALKs
years	82-02	82-02	82-02	82-02	82-02	85-02	85-02
areas	IN	IN	IN	IN	IN	IN	IN
timestep	QTR	QTR	QTR	QTR	QTR	IGFS	QTR
gearclass		COD-ALL		COD-ALL	COD-ALL		
vesselsubclass		102		102	102		
fleetscale		1					
samplingtype		130 135	130			130	135
species	COD	COD	COD	COD	COD	COD	COD
lengthcell	10		10	20	40	20	20
lengthcellmin	100		100	100	100	100	100
lengthcellmax	1400		1400	1400	1400	1400	1400
agemin	1		1	1	1	1	1
agemax	12+			12+	12+	12+	12+
lmax	1400		1400	1400	1400	1400	1400
amax	14			14	14	14	14
stocknames	cod	cod	cod	cod	cod	cod	cod
fleetnames		comm		comm	comm		
surveynames		igfs aut				igfs	aut
prediction	2006	2006	2006				
switch		1					
survey		30000 20000					
zerotimestep				0	0	0	0
temperature	5						
outputdir	Input	Input	Input	Input	Input	Input	Input
prefix	cod	cod	cod	cod	cod	cod	cod

Table 8: An example of the extraction routines for a simple single species GADGET model

Multistock likelihood data

There are two other types of likelihood data which can be used in multistock and multispecies models: stock distribution and stomach content. Examples of keywords to extract these data types are given in table 9.

Stock assessment

Data in a more appropriate form for stock assessment, such as length distributions, age length distributions and catch in numbers at age can be extracted from the DW using the same keywords as the GADGET data but with different filetypes ie ALK, LD etc.

	Stock distribution				Stomach contents
	Catch		Survey		
keyword	by age	by length	by age	by length	by length
hostname	haffugl	haffugl	haffugl	haffugl	haffugl
database	oko3	oko3	oko3	oko3	oko3
user	vojtech	vojtech	vojtech	vojtech	vojtech
groupfile	cod.grp	cod.grp	cod.grp	cod.grp	cod.grp
filetype	GSTOc	GSTOc	GSTOs	GSTOs	GPRE
years	82-02	82-02	85-02	85-02	82-02
areas	IN	IN	IN	IN	IN
timestep	QTR	QTR	IGFS	IGFS	HALF
gearclass	COD-ALL	COD-ALL			
vesselsubclass	102	102			
samplingtype			130	130	
species	COD	COD	COD	COD	COD
lengthcell	1400	40	1400	40	100
lengthcellmin	100	100	100	100	100
lengthcellmax	1400	1400	1400	1400	1400
agemin	1	1	1	1	1
agemax	12+	12+	12+	12+	12+
lmax	1400	1400	1400	1400	1400
amax	14	14	14	14	14
maturitystage	1 2,3,4	1 2,3,4	1 2,3,4	1 2,3,4	
preys					CAP COD *
stocknames	codi codm	codi codm	codi codm	codi codm	codm
fleetnames	comm	comm			
surveynames			igfs	igfs	
zerotimestep	0	0	0	0	0
outputdir	Input	Input	Input	Input	Input
prefix	cod	cod	cod	cod	cod

Table 9: Extraction routines for stock distribution and stomach content likelihood data

definition type	identifier	group(s)
years	82-02	1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002
division	IN	101,102,103,104,105,106,107,108
timestep	QTR	1,2,3 4,5,6 7,8,9 10,11,12
timestep	IGFS	3
timestep	HALF	1,2,3,4,5,6 7,8,9,10,11,12
gearclass	COD-ALL	101,102,103,105,106

Table 10: Group file commands

Future work

Immediate

At present there are some problems with access to the DW which need to be resolved eg vojtech is the only user with access. A more permanent set up of the DW needs to be established, possibly on a separate computer bought for that purpose.

Since the data were loaded into the DW, there have been many changes to the MRI DB so these data need to be reloaded. The method in which the surveys are dealt with will also be modified due to changes in the structure of the MRI DB.

Redfish (*Sebastes marinus*) data from stomach tables, log books and landings will be added. Other species, such as saithe (*Pollachius virens*) and Atlantic wolffish (*Anarhichas lupus*) will also be included. The addition of species to the DW will depend mainly on species which will be modelled using GADGET.

Stomach data after 1999 need to be accessed - they are stored in separate tables - and loaded. When stomach data from new predator species are added, it may be necessary to change the species included as prey types. The prey groups need to be described more fully and the definitions possibly changed.

Test the herring tagging data more thoroughly to ensure that only stations where all marked fish have been entered into the DB are included.

Work is being done to provide access to the DW through a web browser. This will be open to the public and data selection will be through menus and clickable maps. Initially the data available in this manner will be limited but access will be increased.

Longer term changes

The recent log-book data will be extracted and processed using SQL, rather than outputting data for processing with other software.

Landings data prior to 1992 needs to be accessed and added.

Environmental, zooplankton and acoustic data are still to be included in the DW. For environmental and zooplankton, this will only involve extracting and processing data from the MRI DB. Acoustic data however are not available in the DB. The raw data are stored on archive tape and will be time consuming and difficult to process. Another option would be to use the data which have already been processed and have been used in assessments. These data are more easily accessible, although preparing them for the DW could still be time consuming. The reference table (Table 8) will also be uploaded.

There are tagging data, including haddock, which have not been entered into the MRI DB. When these are they can be entered into the DW.

More work will be done on the extraction routines. This will involve the establishment of look up tables which will store information which is required in data extraction, but may be time consuming to extract directly from the data. These will include weight-length relationships for each species.

One of the new stand alone tables in the datawarehouse was introduced to enable age length keys to be stored for the Celtic Sea data. This table could also be used as a back up table when a request is made for age data but insufficient are available for that area/timestep. The user would be told that a stored ALK was being used in assigning ages to the length data.

C.2 The database for the North Sea herring case study

Marco Kienzle, FRS, Aberdeen

1 Introduction

The dst2 project includes the North Sea herring fishery as a case study in order to analyse available data using the Globally applicable Area Dis-aggregated General Ecosystem Toolbox (GADGET). The data has been organised into 3 differents categories of table each describing a particular aspect of the fishery: the herring fishing fleet, the biology of the species and the hydrographic condition of the North Sea.

In each table, the data are characterised by general and specific fields of information. The general fields are common to all the tables. They define data attributes like the position where data was collected, the time of sampling and the method of sampling. In contrast, some fields are found only in specific tables as they give specific information only related to one aspect of the fishery.

Particular care has been taken during the creation of this database to verify that the information stored corresponds to the one used for the stock assessment of the North Sea herring as reported by the Herring Assessment Working Group (HAWG).

2 General information

2.1 Sources of data

This database contains information provided by 4 national and 1 international institute. The marine research institutes of Denmark[1], Norway[2], the Netherlands[3] and UK (Scotland)[4] provided data about their national catches of herring and associated biological data (see Table 1). The International Council For the Exploration of the Sea (ICES) provided the biological information collected during the International Bottom Trawl Surveys (IBTS). This institute also provides the hydrographic information concerning the North Sea[5]. Finally, incomplete time series of catches of herring were filled in using data coming from the Herring Assessment Working Group reports (ICES 2001).

2.2 The North Sea area

This database holds information about the region encompassing ICES divisions IIIa, IVa, IVb, IVc and VIId (see map in page 83), referred to as the North Sea region. This region extends from $49.5^{0}5$ to 62^{o} North and from 2^{o} West to 13^{o} East. For this database this region is divided into 4 divisions, these correspond to ICES division IIIa, IVa, IVb, with IVc and VIId combined because landings of herring are not always available for each separatly. The whole North Sea region is further sub-divided into 240 ICES statistical squares (referred to as gridcells in the database).Table 2 shows the count of the gridcells in each divisions.

2.3 Species and stock

Two populations of herring (*Clupea harengus*) with different biometrics characteristics and spawning season are found in the North Sea (Cushing and Bridger 1966). They are commonly referred to as the spring spawners and the autumn spawners. Individuals from these population are identified by vertebral counts (ICES 1990). This database stores data only from autumn spawning herring. The autumn spawner population is composed of at least 2 stocks: one living in the Northern (IVa) and Central (IVb) North Sea known as the Bank stock and one wintering and spawning in the Southern North Sea (IVc) and East English Channel (VIId). Since 1971, the autumn spawning herring is managed as a single

---

[1] Danish Institute for Fisheries Research (DIFRES)

[2] Intitute of Marine Research (IMR)

[3] the Netherlands Institute for Fisheries Research (RIVO)

[4] Fisheries Research Services (FRS)

[5] see `http://www.ices.dk/ocean/`

stock by the HAWG (Burd 1985) but management advice is given for each stock separatly (ICES 2001).

2.4 Vessels and gears

The vessels fishing for herring in the North Sea are classified according to 3 characteristics: their length, nationality and type of activity. There are 3 length categories: less than 12 m, between 12 and 24 m and greater than 24 m. The vessels are also split into 3 categories according to their nationality and activity: scientific vessels, national commercial vessels, and foreign commercial vessel. Vessels are considered as national or foreign relative to the institute providing the information *e.g.* for FRS data, vessels registered in United Kingdom are considered as national.

The gears used for fishing have been classified into classes and sub-classes (Table 3). The gear class is based on 2 characteristics of the fishing process that define 7 categories. The first characteristic splits the gears accoriding to the movement involved in capturing the fish. A gear is classified either as passive if the capture is caused by the movement of the fish toward the gear, or active if the gear moves after the fish to capture it. The second characteristic is the position in the water column at which the gear operates. Gears that are in contact with the bottom are classified as demersal and all other gears are classified as pelagic. If there is uncertainty regarding a gear, it may be considered undetermined for one or both characteristics. As a consequence, this classification has 7 categories. The gear sub-classes group the gears into a further 34 categories (Table 3).

3 Specific information

The following sections describe the specific variables contain in each table of data.

3.1 Fleet information

The fleet information is stored in the catch table. This table gives detailed information regarding fishing activity by specifying information on the effort involved in fishing and the resulting catches. The data are organised into the following fields

1. number of distinct vessels operating in the area
2. number of fishing trips done by all vessels
3. average power (in kilowatts) of the vessels
4. average gross tonnage of the vessels
5. total number of hours spent by all the vessels
6. total weight (in kg) caught by all the vessels
7. measure of the catch per unit of effort (CPUE) as the ratio of the total weight of fishes caught by the total amount of hours fishing (in kg/hours).

These data were aggregated by month and ICES statistical squares where detailed information was available, otherwise the data are aggregated at a higher level of resolution in space and time. This database stores the official landings data of the 11 countries involved in the exploitation of the herring in the NS for the period from 1985 to 2001.

Some remarks It should be noted that detailed information on the fleet is available only from the Fisheries Research Services (FRS, Scotland-UK). The quality of these data is variable, for example it is not mandatory for Scottish fishermen to report time spent fishing. As a consequence it makes this variable and any further derived data, such as CPUE, of little use for further data analysis. Subject Matter

Experts (SME) notify that there are big distortion in the reported power of the vessel and advise not to use this variable for further analysis.

The intensity of fishing activity may give insight on the abundance and distribution of the fish population. It should be remembered that CPUE has never be very successful in past stock assessments of pelagic populations, especially as a measurement of abundance.

3.2 Biological information

Biological information collected from a sample of herring is either a length frequency distribution (LFD) or a maturity-age-length key (MALK). LFDs are collected using a random sampling strategy. They give the numbers of fish in each length bin. The MALK data describe the maturity, the age and the length of a proportion of individuals from the sample. These data are gathered using a length stratified protocol *i.e.* a fixed number of fish in each length bin are sampled.
The biological samples are caught by commercial or scientific vessels (see vessels description in section 2.4). Commercial samples come from market surveys while scientific samples are collected during pelagic or demersal surveys.
The database contains 3 tables dealing with the biological information, these are grouped by gridcell and month. The catchsample table gives a summary of the number of fishes sampled for length, age and weight. The lengthcell table gives the length frequency distribution. The age table provides the MALKs. The fields specific to the lengthcell and the age tables are:

1. length bin expressed in millimetres
2. sex
3. maturity, on a 4 stage scale
4. mean weight of the sample

3.3 Environmental information

The environment table contains hydrographic information about the NS for the period 1980 to 1999. It stores average temperature (in degree centigrade) and salinity (in practical salinity unit) by week and gridcell for different depths, which range from sea surface to a maximum of 2052 meters.

4 Verification and correction applied to the data

4.1 Fleet information

A correction is applied to the data from UK-Scotland in order to take into account catches made in division IVa but reported in VIa. The catches reported in the 4 ICES statistical squares between 46E5 and 49E5 were assigned to the adjacent statistical squares in division IVa *i.e.* 46E6 to 49E6.

The catch data provided by research institutes were compared with those available from the HAWG reports between 1985 and 2001, in order to (a) check that the data stored in the database have not been altered during their manipulation and (b) provide a database coherent with the one used by the ICES working group for the stock assessment of the NS herring. Figure 1 shows their trend for each division of the NS. The agreement between the 2 sources of data is often good, sometimes perfect *e.g.* Norwegian data from ICES division IVb. Nevertheless the circles in the graphs indicate that all time series have to be completed outside the period 1990 to 2000 and that Danish data from ICES division IIIa are completly missing. This information is used to complete the catch data in the database. The remaining discrepancies are corrected using a weighting factor in order to achieve a perfect match between the two sets of data. The weighting factor vary from 0.1 and 3.7 but is in between 0.64 and 1.23 for 90% of the data.

4.2 Biological information

A comparison of the mean weight at age between data stored in the DW and those reported in the HAWG has been performed to verify the consistency of the DW. The data in the HAWG's reports are calculated using commercial landing. The comparaison of the mean weight at age calculated using all the data in the DW (Fig. 2) shows that these data are generally in good agreement although the mean weight at age from the DW are often inferior, especially for those from 1990 and between 1998 and 2000. These differences may be due to grouping of commercial and scientific data. The same analysis made on data split into these two categories, which is possible only with Norwegian data, shows that

- both commercial and scientific data stored in the DW give higher value of mean weight at age for age group 0 and 1.
- mean weight from commercial data are in good agreement for age group 2 to 8.
- mean weight from scientific data are lower than those from commercial landing for age group 2 to 8.

Although, there are differences between particular set of data stored in the DW, the general trend in the data is in good agreement with those reported by the HAWG.

4.3 Environmental information

The data store in this table have not been compared to another source of hydrographic data.

Acknowledgement

I am grateful to S. Keltz for the corrections and comments he made to improve this document.

References

A. C. Burd. Recent changes in the central and southern North Sea herring stocks. *C*anadian Journal of fishery and aquatic sciences, 42(Suppl. 1):192 206, 1985.

D. H. Cushing and J. P. Bridger. *T*he stock of herring in the North Sea and changes due to fishing, volume 25 of *F*ishery Investigations 2. Ministry of Agriculture, Fisheries and Food, 1966.

ICES. Herring Assessment Working Group for the Area South of 62oN. Technical Report CM/ASSESS:14, ICES, 1990.

ICES. Herring Assessment Working Group for the Area South of 62oN. Technical Report CM/ACFM:12, ICES, 2001. 4

Tables

Institute	Fleet data	Biological data	Hydrographic data
DIFRES (Denmark)	Total landing by gridcell and month.	Length frequency and maturity-age-length key from the market sampling.	-
IMR (Norway)	Total landing by division and month.	Length frequency and maturity-age-length key from the scientific and market sampling.	-
RIVO (Netherland)	Total landing by gridcell and month.	Maturity-age-length key from scientific and market sampling.	-
FRS (UK-Scotland)	Landing by gridcell and day for each vessel and description of the vessels characteristics.	Length frequency and maturity-age-length keys from the market sampling	-
ICES	Total landing by division, country and year from the Herring Assessment Working Group	length frequency distribution and maturity-age-length key from the International Bottom Trawl Surveys	Temperature and salinity from the hydrographic database.

Table 1: Summary of the data provided by each national institute of research and ICES.

Division	Number of gridcell
IIIa	15
IVa	94
IVb	90
IVc+VIId	51

Table 2: Number of gridcell in each sub-division.

Classes	**Sub-classes**
Demersal passive	Demersal Gill Nets
Pelagic passive	Drift gillnets Pelagic Gill Nets
Undetermined passive	Gillnets and entangling nets (not specified) Small And Hand Lines
Demersal active	Beam Trawl Boat dredge Bottom otter trawl (side or stern not specified) Bottom otter trawl (stern) Queen Scallop Trawl Demersal Pair Trawl, Demersal Twin/Mult Trawl Demersal Trawl Bottom pair trawl - NB: operated by 2 vessels Pair seine, Demersal Pair Seine - NB: operated by 2 vessels Scottish seine, Demersal Seine Nephrops trawl, Nephrop Trawl (Single Rig) Nephrops trawl (twin/multiple rigged), Twin/Mult Rig Neph Trawl Twin trawls Otter twin multi trawls Light Trawl (under 90ft)
Pelagic active	Otter trawl (not specified) Midwater otter trawl (side or stern not specified) Midwater otter trawl (side) Midwater otter trawl (stern) Pusre seine operated by one vessel Purse seine operated by two vessels Purse Seine Shrimp trawl Single Boat Pelagic Trawl Industrial Trawl Pair Trawl Pelagic Midwater pair trawl - NB: operated by 2 vessels Mid trawls (not specified)
Undetermined active	Pair trawl (not specified) - NB: operated by 2 vessels
Demersal undetermined	Other Demersal Nets

Table 3: Categories used to classify the fishing gears.

Graphics

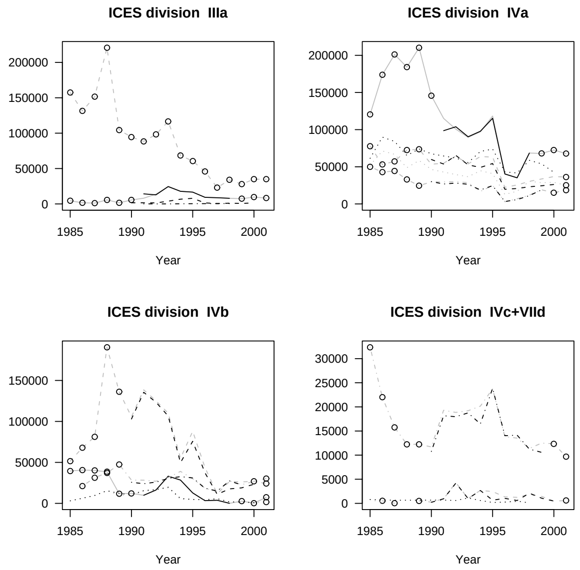


Figure 1: Comparison between trend of landing given by the raw data (in black) and the HAWG report (in grey). The landings, expressed in tons, have separated into divisions. Each country is represented by a particular line type: Norway (solid line), Denmark (dashed line), UK-Scotland (dotted line) and Netherlands (dashed and dotted line). The circles indicate where data are missing from the original files.

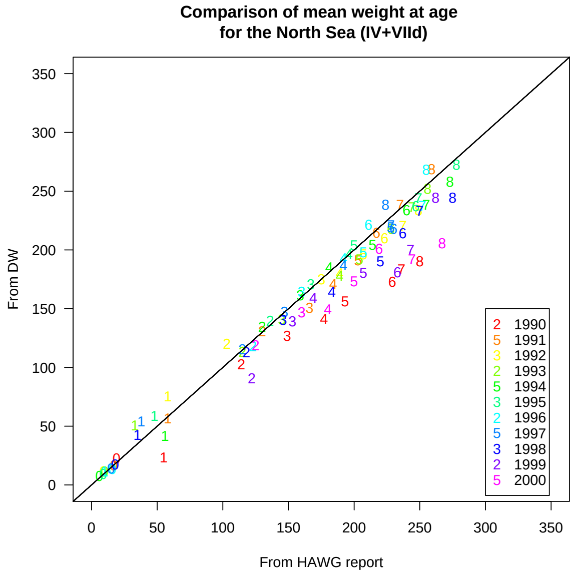


Figure 2: Comparison of the mean weight at age information stored in the database and provided by HAWG. The solid line descibes points for which $y = x$.

Maps

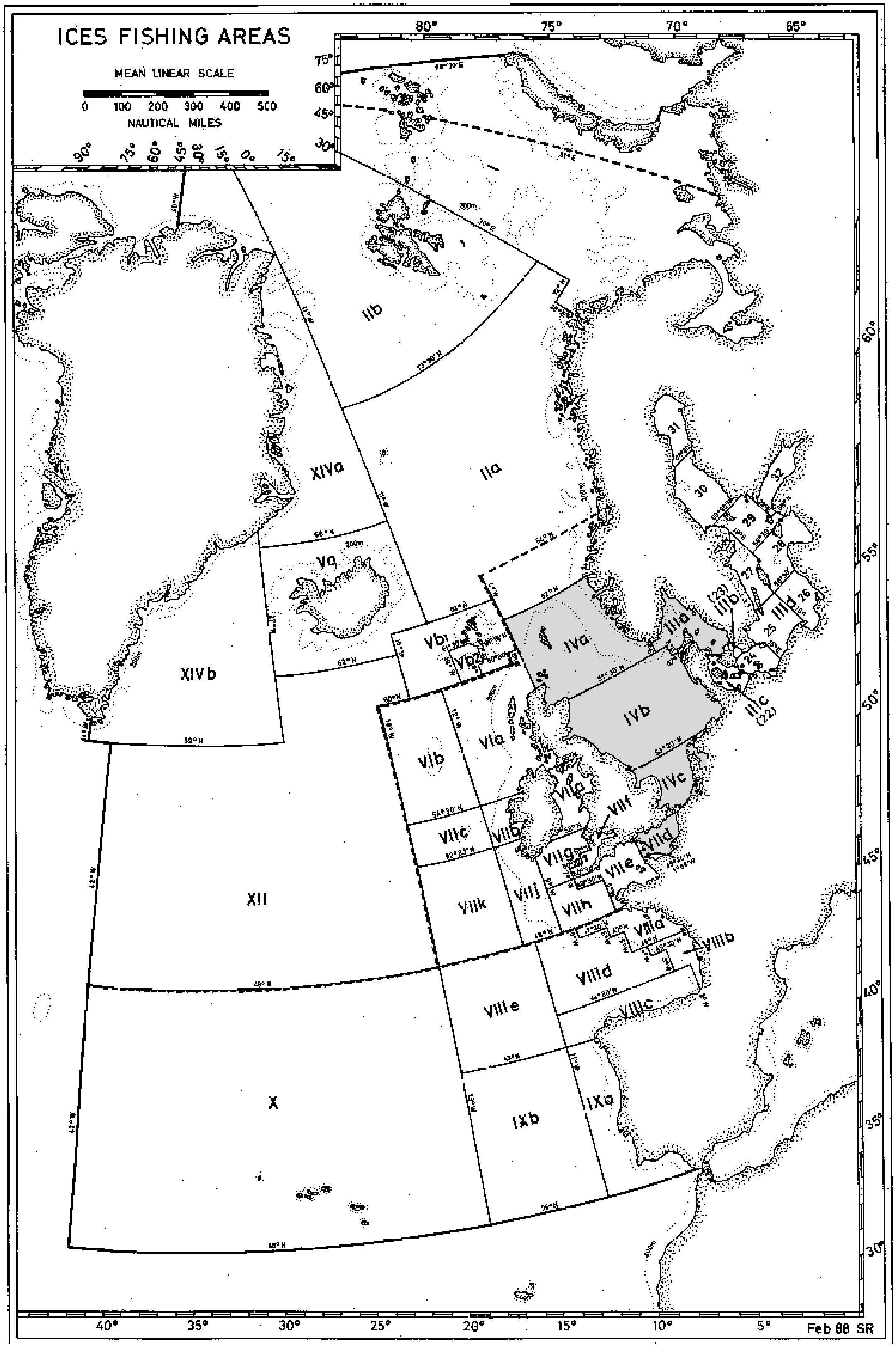


Figure 3: Boundaries of the ICES sub-area in the North-east Atlantic.

D Biological Modelling

D.1 Simulations with a dynamic continuous model of fish migration

Sven Þ. Sigurðsson, Kjartan G. Magnússon, Petro Babak,
Stefán F. Guðmundsson and Eva Hlín Dereksdóttir.

Science Institute, University of Iceland

Introduction

In the progress report 1 January 2001 to 31 December 2001 we presented a dynamic continuous single species model for the spatial distribution and migration for pelagic fish. It consists of a pair of partial differential equations describing balance of mass and momentum, and taking into account maturity distribution. The form of the latter equation is empirical and is guided by considerations of being able to control the velocity and density by external factors such as temperature distribution, distance to spawning grounds, and direction of movement and density of neighbouring fish. An integral part of the model is its realization by a finite element model that was described in some detail. Through this realization the model may also be viewed as a distributed compartmental model. Below we illustrate the role of the parameters of the model by using it to simulate movement of capelin in the Barents Sea. These are analogous to the simulations presented in this report using a continuous density Kolmogorov type model. Before presenting the simulations we review briefly the aim with this model and its basic form. We leave out dependence on maturity since it is not included in the simulations. Similarly we leave out drag effect of sea-currents. A fuller description is given in Sigurdsson et al. (2002).

Mathematcal Model

The aim with the construction of this model has been to:

- Aid in the assessment of what type of migration models might be included into a more general prediction model of changes in fish stock size (DST^2, 2001, DST^2, 2002)
- Test various assumptions on what external factors may influence fish migration.
- Provide a continuous counterpart to dynamic discrete particle migration models like those presented by us in an earlier report (Hubbard et al., 2001)
- Provide a dynamic continuous counterpart to a continuous density Kolmogorov type model that we have been developing concurrently (Magnússon et al., 2002), the main difference between the models being that in the former density describes actual density of fish but in the latter it describes the probability density of the location of the fish.
- Develop a distributed analogue to compartmental models like BORMICON where migration between compartments is described in terms of transition matrices, (BORMICON, 1997). Such a distributed model may in particular aid in the assesssment of the transition coefficients of the compartmental model.

The **values** that the model simulates are:

$\rho(t,x,y)$, the **density** of fishstock at time *t* and location (x,y)

$\boldsymbol{v}(t,x,y) = [u(t,x,y), v(t,x,y)]^T$, the **lateral velocity** of the fishstock at time *t* and location (x,y)

The **input** into the model, apart from initial values and boundary conditions around the region under consideration, are:

$\boldsymbol{p}(t,x,y)$, **preferred direction** whose role is eg. to align the direction of movement of the fish to that of its neighbours, turn the fish towards preferred food and/or temperature conditions, and towards some preferred location or path during spawning migration.

$v_p\,(t,x,y)$, **preferred speed** whose role is to control the speed of the fish and adjust it to the speed of its neighbours.

$\rho_p\,(t,x,y)$, **preferred density** whose role is to push the fish apart if the density becomes too high as well as possibly attract the fish to each other if the density becomes too low. This density may eg. depend on food density and hence on time and location.

The model consists of a pair of partial differential equations. Firstly, an equation of **mass balance:**

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{v}) = 0 \tag{1}$$

Here we are ignoring, for the sake of simplicity, any possible recruitment or mortality that would otherwise be incorporated as a source term on the right hand side.

Secondly, an empirical equation of **momentum balance:**

$$\frac{\partial (\rho \boldsymbol{v})}{\partial t} + \nabla \cdot (\rho \boldsymbol{v} \otimes \boldsymbol{v}) = \rho \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \boldsymbol{v} + d \nabla \rho \tag{2}$$

Here

$$a = \log \left(\left(\frac{v_p}{|\boldsymbol{v}|} \right)^{\frac{1}{\tau_a}} \right) \approx \frac{1}{\tau_a |\boldsymbol{v}|} \left(v_p - |\boldsymbol{v}| \right) \qquad \text{if} \qquad v_p \approx |\boldsymbol{v}| \tag{3}$$

where v_p is the preferred speed and

τ_a is the time constant of adjustment to this speed.

$$b = \frac{1}{\tau_b} \sin^{-1} \left(\frac{u p_y - v p_x}{|\boldsymbol{v}|} \right) \approx \frac{u p_y - v p_x}{\tau_b |\boldsymbol{v}|} \qquad \text{if} \qquad p \approx \frac{1}{|\boldsymbol{v}|} v \tag{4}$$

where $\boldsymbol{p} = \begin{bmatrix} p_x & p_y \end{bmatrix}$ is the preferred direction and

τ_b is the time constant of adjustment to that direction

$\sin^{-1}\left(\frac{u p_y - v p_x}{|\boldsymbol{v}|}\right)$ is the angle between $\boldsymbol{v}$ and $\boldsymbol{p}$

$$d = d' \left(\rho - \rho_p \right) \tag{5}$$

for some constant coefficient *d'* where ρ_p is the preferred density.

The preferred direction $\boldsymbol{p}$ may in general be composed of different components each associated with a specific time constant. Typically we may have a component that is in the average direction of the surrounding fish in some specified neighbourhood Ω with radius ϵ. The angle of this direction, θ_{align}, may be defined as

$$\theta_{align} = \tan^{-1}\left(\frac{\int\int_{\Omega}\rho(t,\xi,\eta)sin\left(\theta(t,\xi,\eta)\right)d\xi d\eta}{\int\int_{\Omega}\rho(t,\xi,\eta)cos\left(\theta(t,\xi,\eta)\right)d\xi d\eta}\right) \tag{6}$$

where

$$\left[\begin{array}{cc}\cos\left(\theta\left(t,\xi,\eta\right)\right) & \sin\left(\theta\left(t,\xi,\eta\right)\right)\end{array}\right] = \frac{1}{\left|\boldsymbol{v}\left(t,\xi,\eta\right)\right|}\boldsymbol{v}\left(t,\xi,\eta\right) \tag{7}$$

and the angle according to equation (6) is modified appropriately when $\cos\left(\theta\left(t,\xi,\eta\right)\right) \leq 0$.

Secondly, we may have a component in the direction of the gradient of a temperature or a food density function. Denote the angle of this direction θ_{grad}. Finally we may have a component that is in the direction towards some spawning ground. Denote the angle of this direction θ_{spawn}. Denoting the corresponding time constants with τ_{align}, τ_{grad}, and τ_{spawn} we then set

$$\frac{1}{\tau_b} = \frac{1}{\tau_{align}} + \frac{1}{\tau_{grad}} + \frac{1}{\tau_{spawn}} \tag{8}$$

and

$$\theta_b = \frac{1}{1+\frac{\tau_{align}}{\tau_{grad}}+\frac{\tau_{align}}{\tau_{spawn}}}\theta_{align} + \frac{1}{1+\frac{\tau_{grad}}{\tau_{align}}+\frac{\tau_{grad}}{\tau_{spawn}}}\theta_{grad} + \frac{1}{1+\frac{\tau_{spawn}}{\tau_{align}}+\frac{\tau_{spawn}}{\tau_{grad}}}\theta_{spawn} \tag{9}$$

i.e. a linear combination of the three angles where the angle associated with the shortest timeconstant carries the largest weight, and such that

$$\frac{\theta_b}{\tau_b} = \frac{\theta_{align}}{\tau_{align}} + \frac{\theta_{grad}}{\tau_{grad}} + \frac{\theta_{spawn}}{\tau_{spawn}} \tag{10}$$

In our implementation we use in fact a slightly different approach by taking the same linear combination of unit direction vectors as is done in Hubbard et al.(2001). The difference between these two approaches is most pronounced when two of the directions are opposite each other.

The term $d\nabla\rho$ in equation (2) acts effectively as a pressure force while d>0. By specifying d as is done in equation (5) so that d<0 when ρ<ρ_p in order to maintain the preferred density ρ_p and encourage group formation we get sharp fronts in the density values at $\rho = \rho_p$, causing difficulties in numerical simulations.

We use a spatial approximation of the continuous model above, using a finite element Galerkin approach with triangular elements. This gives considerable flexibility when treating irregular domains and boundaries and allowing variation in resolution. A second important feature is that we effectively have a local coordinate reference system within each element. This makes it possible to introduce local transformations if we want eg. to transform longitude and latitude into (x,y)-cordinates or rotate the system into a principal direction in the case of anisotropy in coefficients. Finally, the scheme has the important property of ensuring mass conservation. Stability considerations lead us to consider a so-called staggered approximation where the velocity, $\boldsymbol{v}$, is approximated by constants within the triangular elements and the density, ρ, is approximated by a continuous function, linear within elements, and thus specified by the approximate values at the corners of the element. Further details are given in Sigurdsson et. al (2002) and DST^2 (2002).

Simulations

We now show some results of using the model presented above to simulate the migration of capelin in the Barents Sea. The capelin stock in the Barents Sea performs extensive seasonal migrations (Gjøsæter, 1998). In winter and early spring a southward spawning migration takes place towards the coast of North Norway and Russia, where the fish die after spawning. The next generation – which hatched two years earlier- migrates north and north-eastwards to feeding grounds in summer and autumn. In our simulations the spawning and feeding migrations are represented as a one year cycle – i.e. a southward spawning migration followed by a northward feeding migration.

The area of simulation along with its finite element division is shown in figure 1. In the bottom left hand corner is the coast of North-Norway and Russia. Near the bottom at the right is Novaya Zemlya, Franz Josef Island is in the top right hand corner and Svalbard at the centre at the top. This area is approximately 1400x1300 km^2. Each triangular element in this simulation is of the same size 200 km^2 so the area is divided into approximately 8000 elements. The whole area has been projected onto an (x,y)-plane allowing us to work in those coordinates.

Figure 1: The finite element division of the Barents Sea region. The area of each triangular element is 200 km^2. The total area is approximately 1400x1300 km^2.

The preferred speed is set to be 40 km/day throughout the simulations. The timestep is 0.1 day. Thus we can cross an element in approximately 5 timesteps. The timeconstant for the speed is also set to 0.1 days. This implies that the speed is effectively set to its preferred value at the end of each timestep.

The preferred direction is composed of three components. The spawning direction component is the negative gradient of the shortest distance to the spawning region, that has been specified along the coast as shown in figure 2. In particular there is no spawning component within the spawning region. Moreover, this component is only active in the period December 20 - May 15. Since there is limited information on available zooplankton we specify a feeding region as shown in the upper part of figure 2, and then introduce a feeding direction in a completely analogous way to the spawning direction but which is active in the period May 15 - December 20.

Figure 2: The Barents Sea with the attracting spawning and feeding regions.

The gradient direction component is determined by the temperatue distribution based on twice yearly measurements made available by the Institute of Marine Research in Bergen, Norway. Figure 3 shows the temperature distribution in March and September 1991. We use linear interpolation in order to obtain temperature values at other times. On the feeding grounds the direction is that of the positive gradient if the temperature is below 0°C and that of the negative gradient if the temperature is above 3°C. In the interval between 0°C and 3°C, taken to be the preferred temperature, there is no gradient component. On the spawning grounds the directions are similar except that the interval of preferred temperature is shifted to the interval between 3°C and 6°C. This is in line with with observations about ambient temperature as discussed more fully in Magnússon et al. (2002). We implement this by letting the interval be between 0°C and 3°C on September 15 and between 3°C and 6°C on April 15, and interpolate linearly with time between these extremes at other dates. When both the spawning (or feeding) direction and the gradient component are present they have the same timeconstant i.e. the same weight. When neither of them is present (which may be the case in some elements within the spawning and feeding regions), the direction is given a random value within the element at the end of each timestep. Finally, we include in one simulation an alignment direction component.

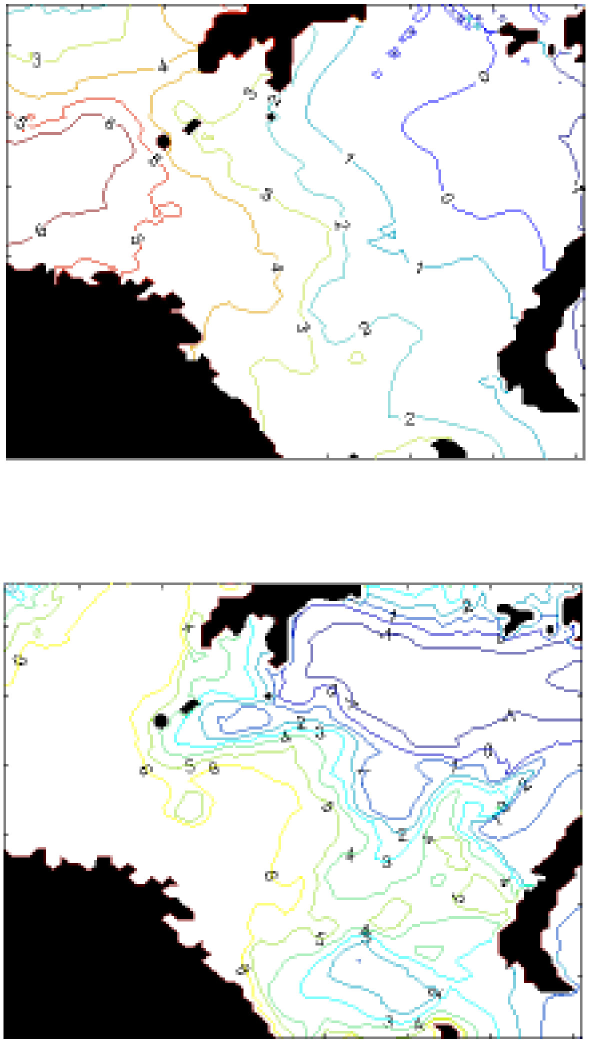


Figure 3: Temperature distribution in March 1991 (upper) and in September 1991 (lower).

The capelin density is treated as a dimensionless quantity and its actual value is arbitrary. In our simulations we set it as 250 within a single triangular element at the centre of the feeding region at December 15 1989, and 0 everywhere else. We then run the simulation for one year before presenting the results shown below. A considerable dispersion has taken taken place when the capelin returns to the same place a year later. In the second year, which we show, this dispersion does not, however, increase significantly, nor does it do so in subsequent years.

In the first simulation shown in figure 4 we have no alignment nor preferred density. The timeconstants for both the spawning (feeding) direction and the gradient direction are set as 0.13 days which effectively means that the direction of velocity is set as the average of these directions at the end of each timestep. The results are very close those given in figure 3 in Magnússon et al. (2002) for a continuous density Kolmogorov type model even if the interpretation of the density is not the same. In the Kolmogorov model the density refers to the probability density of location whereas in the model of this report it refers to actual density. Furthermore in the Kolmogorov model the velocity is set to a given value at each timestep, rather than the change in velocity being controlled by a differential equation. The choice of preferred speed and direction is however the same in both simulations so by choosing the timeconstants to be the same as the timestep the similarity in the results is not so surprising. We refer again to Magnússon et al. (2002) for an account of how these results relate to what is known about capelin migrations in the Barents Sea.

The lowest density shown in the figures 4 - 8 is 0.013, and the lightest shade corresponds to density up to 0.1, the next shade to density up to 0.5, and the one after that to density up to 1.0. If we have a uniform density of 1.0 in these simulations the total mass covers an area of 250 elements or 50000 km^2. The darkest shade shown in the figures amounts to density values in the interval from 4 to 5. The dark continuous curve in these figures is the track of the centre of the mass.

The second simulation shown in figure 5 uses identical parameters to the first one except that the time-constants for speed and direction are increased by a factor 10, meaning that the fish reacts slower to changes in the environment. This leads to a considerable increase in the dispersion of the mass. The effect is not unsimilar to that shown in figure 6 for the third simulation where we again use identical parameters to those in the first simulations except that we now set *d*'=1000 km^2/day^2 along with zero value for preferred density, thus only including the effect of pushing the fish apart. The dispersion effect is, however, somewhat smoother. The effect in this case is similar to introducing a diffusion term in the mass balance equation with a diffusion coefficient of size $O(\Delta t \cdot d')$, where Δt denotes the timestep, and thus of size $\approx 100\,\mathrm{km}^2/\mathrm{day}$ in these simulations. In this context one should also be reminded of the fact that in the finite element model we use a fairly simple upwinding scheme in order to avoid spurious oscillations and negative values of density. The effect of this is again similar to introducing a diffusion term in the mass balance equation with a diffusion coefficient of size $O\left(l \cdot |\boldsymbol{v}|\right)$, where *l* denotes a characteristic length of a finite element, and thus of size at least $\approx 100\,\mathrm{km}^2/\mathrm{day}$, although mainly in the direction of motion and at the trailing end of the moving mass.

In the fourth simulation shown in figure 7 we consider the effect of introducing alignment with a radius of influence of 27 km, which means that the velocity in a neighbouring element plays part if the distance between the centre of the element under consideration and the centre of the neighbouring element is less than 27 km. On average, velocities in 8 neighbouring elements play a part. The timeconstant of this direction component is set to be three times that of the other two components. The results are somewhat inconclusive although the increased dispersion is to be expected since the averaging of direction over a neighbourhood has a similar effect as introducing a diffusion term into the momentum equation. It is of course questionable whether it is meaningful to introduce the alignment effect when simulating on such a coarse scale.

Finally, in the fifth simulation shown in figure 8, we consider the effect of setting a preferred density of 10. If the mass were uniformly distributed with this density it would cover 25 elements. We set *d*' = 10 km^2/day^2 but the results are very similar to those of the first simulation in figure 6, slightly more patchy but the highest density value again only reaching 5. Thus one would like to increase the value of *d*' but with the present implementation of the model the program breaks down if one tries this, and the same holds true if the value of the preferred density is increased. This relates to the numerical problem of maintaining sharp density fronts while keeping the density positive.

December15, 1990

June 1, 1991

January 15, 1991

September 15, 1991

March 15, 1991

December 15, 1991

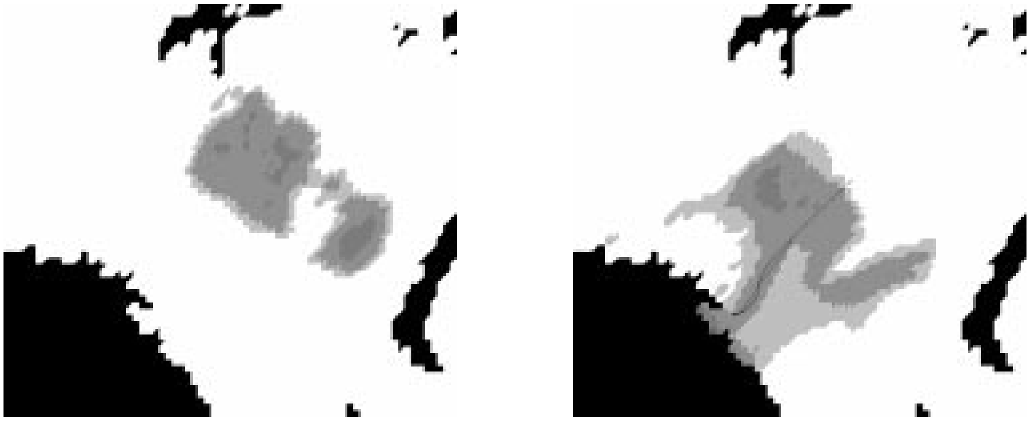
Figure 6. The simulated spatial distribution for Barents Sea capelin over a one year cycle. $\tau_{grad} = \tau_{spawn} = 0.13 \text{ days}$, no alignment, $\tau_a = 0.10 \text{ days}$, $d' = 0$

December15, 1990 June 1, 1991

January 15, 1991 September 15, 1991

March 15, 1991 December 15, 1991

Figure 7. The simulated spatial distribution for Barents Sea capelin over a one year cycle. The effect of increasing the time constants for direction and speed.
$\tau_{grad} = \tau_{spawn} = 1.3 \text{ days},$ no alignment, $\tau_a = 1.0$ days, $d' = 0$



December15, 1990

June 1, 1991

January 15, 1991

September 15, 1991

March 15, 1991

December 15, 1991

Figure 8. The simulated spatial distribution for Barents Sea capelin over a one year cycle. The effect setting *d*'>0 with zero preferred density.
$\tau_{grad} = \tau_{spawn} = 0.13$ days, $\tau_a = 0.10$ days, no alignment, $d' = 1000$ km^2/day^2, $\rho_p = 0$

December15, 1990

June 1, 1991

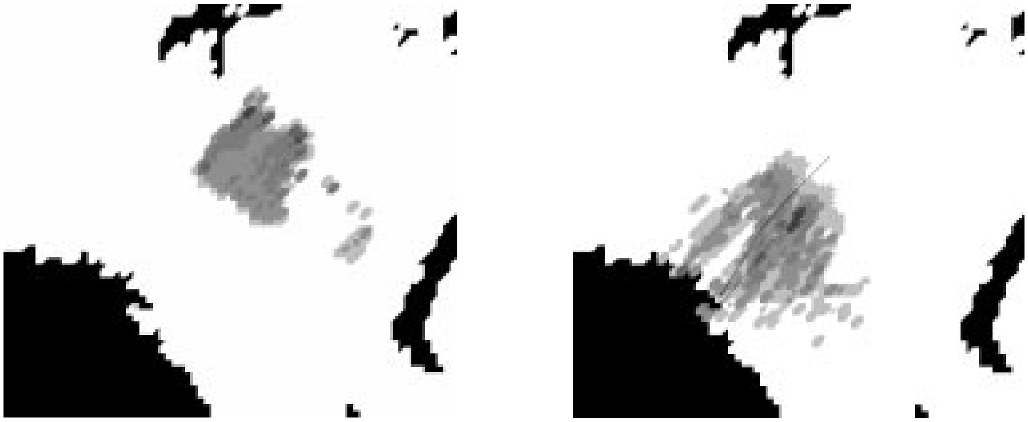
Januar 15, 1991

September 15, 1991

March 15, 1991

December 15, 1991

Figure 9. The simulated spatial distribution for Barents Sea capelin over a one year cycle. The effect of introducing alignment.
$\tau_{grad} = \tau_{spawn} = 0.13 \text{ days}, \quad \tau_{align} = 0.39$ days with radius of 27 km, $\tau_a = 0.10 \text{ days}, \quad d' = 0$



December15, 1990 June 1, 1991

Januar 15, 1991 September 15, 1991

March 15, 1991 December 15, 1991

Figure 10. The simulated spatial distribution for Barents Sea capelin over a one year cycle. The effect settin *d*'>0 with positive preferred density.
$\tau_{grad} = \tau_{spawn} = 0.13$ days, no alignment, $\tau_a = 0.10$ days, $d' = 10\ \mathrm{km}^2/\mathrm{day}^2$, $\rho_p = 10$

Discussion and further work

While the simulations presented above indicate that the model and its finite element realization provide us with a useful tool for simulating fish migration and testing assumptions on what external factors may influence such a migration, some work remains in order to fulfill the aims set out at the beginning. The problem of how to deal with sharp fronts that arise when introducing positive preferred density remains unsolved and attention has to be paid to the numerical dispersion at the trailing end of the moving mass. These problems relate more generally to the question about the relative advantages of continuous density models compared to discrete particle models both with respect to accuracy and computational efficiency that we have not properly investigated. In order to make the discrete model that we have developed (Hubbard et al., 2001) more efficient, we have considered the concept of superparticles that may combine or split up, as well as integrating it into the same triangular network that we use in our finite element realization. When that has been done a proper investigation can be carried out.

The comparisons that can be made between the simulations of the capelin movement in the Barents Sea and reality are mainly of a qualitative nature. In an effort to carry out a more quantitative comparison we are now turning our attention to the seas around Iceland. This also allows us to include the compartmental division of the BORMICON model (BORMICON, 1997), and thus provides us with a concrete example where the results of a distributed model can be related to an aggregated one. In particular, we want to address the problem of how a distributed model can aid in the estimation of transition coefficents describing movement between compartments in the compartmental model. Here the continuous density Kolmogorov type model, that we have developed concurrently (Magnússon et al., 2002), where the density is the probabilty density of location rather than actual density, may provide a more useful tool. However, in our experience so far, it has proved to be fruitful to be able to switch between all these different types of models.

Acknowledgements. This work is supported by the Science Research Council of Iceland and by EU grant QLK5-CT1999-01609. The permission of the Institute of Marine Research, Bergen, Norway, to use the temperature data is gratefully acknowledged as are dr. Geir Huse and dr. Geir Odd Johansen for preparing this data into a suitable form.

References

BORMICON. 1997. A Boreal Migration and Consumption model. Editors: Gunnar Stefansson and Olafur K.Palsson. Marine Research Institute, Reykjavik, Technical Report 58.

DST^2. 2001. Development of structurally detailed statistically testable models of marine populations QLK5-CT1999-01609. Progress Report for 1 January 2000 to 31 December 2000. Marine Research Institute, Reykjavik, Technical Report Nr. 78.

DST^2. 2002. Development of structurally detailed statistically testable models of marine populations QLK5-CT1999-01609. Progress Report for 1 January 2001 to 31 December 2001. Marine Research Institute, Reykjavik, Technical Report Nr. 87.

Gjøsæter, H. 1998. The population biology and exploitation of capelin (Mallotus villosus) in the Barents Sea, Sarsia 83, 453-496.

Gjøsæter, H., Dommasnes, A., Røttingen, B. 1998. The Barents Sea capelin stock 1972-1977. A synthesis of results from acoustic surveys. Sarsia 83, 497-510.

Hubbard, S., Magnússon, K.G., Babak, P., Sigurðsson, S. 2001. Formation of fish schools and migration of fish. Science Institute, University of Iceland, Report RH-28-2001.

Magnússon, K.G., Sigurðsson, S., Babak, P., Guðmundsson, S.F., Dereksdóttir, E.H. 2002. A continuous density Kolmogorov type model for a migrating fish stock. Science Institute, University of Iceland, Report RH-24-2002.

Sigurðsson, S., Magnússon, K.G., Babak, P., Guðmundsson, S.F., Dereksdóttir, E.H. 2002. Dynamic continuous model of fish migration. Science Institute, University of Iceland, Report RH-25-2002.

D.2 A continuous density Kolmogorov type model for a migrating fish stock

Kjartan G. Magnússon, Sven Þ. Sigurðsson, Petro Babak,
Stefán F. Guðmundsson and Eva Hlín Dereksdóttir

Abstract A continuous probability density model for the spatial distribution and migration pattern for a pelagic fish stock is described. The model is derived as the continuum limit of a random walk in the plane which leads to an advection-diffusion equation. The direction of the velocity vector is given by the gradient of a "comfort function" which incorporates factors such as temperature, food density, distance to spawning grounds, etc., which are believed to affect the behaviour of the capelin. An application to Barents Sea capelin is presented.

Introduction

Some pelagic fish species such as capelin (Mallotus villosus) and herring (Clubea harangus) migrate over long distances between feeding and spawning grounds. Knowledge is rather limited about the factors governing the migration patterns, such as the timing and the geographical route. However, temperature seems to be an important variable since migrating fish schools tend to follow boundaries between warm and cold water masses and certain isotherms which the fish are reluctant to cross often form barriers confining the movement of schools. Oceanic currents, salinity and bottom topography are also likely to influence the migration. Food density obviously plays a key role on feeding migrations and spawning migrations are likely to be controlled to a large extent by genetic factors. Furthermore, the migration patterns are known to depend on internal variables such as the physiological state or the state maturity of the fish. On a local scale, the motion of individual fish is governed by the motion of the neighbouring fish in the school (Partridge, 1982). We will attempt to incorporate some of these factors into the model described below.

Aggregations, movements of schools and migrations can be modelled in a number of ways: as an optimization problem where a fitness function is optimized (Fiksen, Giske and Slagstad, 1995); in an *ad hoc* way by transitions matrices whose elements $a(i,j,t)$ are the fraction of fish in sub-area i which move to sub-area j in time-step t (Tjelmeland and Bogstad, 1998); and as a dynamical system with either, equations for individual particles where the motion may be discrete or continuous in time leading to a system of difference equations or differential equations respectively, or as a continuous density model leading to partial differential equations (Toner and Tu, 1998; Okubo, 1986). Models of individual behaviour are sometimes referred to as Lagrangian models and continuum models as Eulerian models, see for example Grünbaum and Okubo (1994).

In a previous paper we presented an individual based model, discrete in time, as a first attempt to model migrations as a Self-Propelled Interacting Particle System moving in a "force field". The motion of individual particles (individual fish or schools of fish) is governed by environmental gradients and by the average motion of the neighbouring particles, as well as by temperature barriers (Hubbard et al., 2003). Here we will model the process as a biased random walk in the plane and take the continuum limit, thus going from a discrete individual based model to a continuum model, cf. Grünbaum (1994). The velocity and the diffusion – which will in general be anisotropic - will be modelled as functions of environmental conditions and the location of the spawning grounds which are taken to be a region which the fish are attracted towards; in effect, the fish experience a "force" pulling them towards the spawning grounds. The movements are therefore determined by non-local forces in addition to the local environmental "forces". Capelin migrations in the Barents Sea (Gjøsæter, 1998) will be taken as an example or a case study.

Model description

Consider a particle undergoing a random walk in the plane. Every τ time-units the particle can either stay in the same location or move a distance Δ in any direction. Denote by *q(x,y)* the probability that the particle moves when it is at *(x,y)* and assume that the conditional probability density function (i.e. the pdf given that the particle moves) for the direction angle θ is of the form

$$p(x,y;\theta) = a(x,y;\theta) + \Delta b(x,y;\theta) \qquad \theta \in [-\pi, \pi[$$

We will assume that $a(x,y;\theta+\pi) = a(x,y;\theta)$ and $b(x,y;\theta+\pi) = -b(x,y;\theta)$, i.e. *a* is the symmetric part of *p* and *b* the anti-symmetric part under a rotation by π but for simplicity of notation we will not show the *t*- dependence explicitly. In order for *p* to be a proper probability density function, i.e.

$$\int_{-\pi}^{\pi} p\left(x,y;\theta\right) d\theta = 1 \qquad \text{, for all } \Delta,$$

we must have $\int_{-\pi}^{\pi} b\left(x,y;\theta\right) d\theta = 0$. This is guaranteed by the anti-symmetry condition on *b*.
Let *u(x,y;t)* be the probability density function for the particle's location at time *t*.

Then

$$\begin{aligned} u(x,y;t+\tau) &= \left(1 - q\left(x,y\right)\right) u(x,y;t) + \\ &\quad + q(x,y) \int_{-\pi}^{\pi} p\left(x - \Delta\cos\theta, y - \Delta\sin\theta; \theta\right) u\left(x - \Delta\cos\theta, y - \Delta\sin\theta; t\right) d\theta \end{aligned}$$

Expanding in a Taylor series, and letting Δ and τ tend to zero in such a way that the limit

$$\lim_{\Delta\to 0, \tau\to 0} \frac{\Delta^2}{\tau} = D_0$$

exists, leads to the Kolmogorov forward equation for the probability density function *u(x,y,t)* for the location of the particle

$$\frac{\partial u}{\partial t} = -\frac{\partial}{\partial x}\left(c_1 u\right) - \frac{\partial}{\partial y}\left(c_2 u\right) + \frac{\partial^2}{\partial x^2}\left(D_1 u\right) + \frac{\partial^2}{\partial y^2}\left(D_2 u\right) + 2\frac{\partial^2}{\partial x \partial y}\left(D_{12} u\right)$$

where

$$c_1\left(x,y\right) = D_0 q\left(x,y\right) \int_{-\pi}^{\pi} \cos\theta \cdot b(x,y;\theta) d\theta \qquad c_2\left(x,y\right) = D_0 q\left(x,y\right) \int_{-\pi}^{\pi} \sin\theta \cdot b(x,y;\theta) d\theta$$

$$D_1\left(x,y\right) = \frac{D_0}{2} q\left(x,y\right) \int_{-\pi}^{\pi} \cos^2\theta \cdot a(x,y;\theta) d\theta \qquad D_2\left(x,y\right) = \frac{D_0}{2} q\left(x,y\right) \int_{-\pi}^{\pi} \sin^2\theta \cdot a(x,y;\theta) d\theta$$

$$D_{12}\left(x,y\right) = \frac{D_0}{2} q\left(x,y\right) \int_{-\pi}^{\pi} \cos\theta \sin\theta \cdot a(x,y;\theta) d\theta$$

The following identities – which are satisfied because $a(x, y; \theta)$ is π-periodic – are used in the derivation of the Kolmogorov equation

$$\int_{-\pi}^{\pi} \cos\theta \cdot a(x, y; \theta) d\theta = \int_{-\pi}^{\pi} \sin\theta \cdot a(x, y; \theta) d\theta = 0$$

.

The Kolmogorov equation can be written as

$$\frac{\partial u}{\partial t} = \nabla \cdot \left(-c_1 u + \frac{\partial}{\partial x}(D_1 u) + \frac{\partial}{\partial y}(D_{12} u), -c_2 u + \frac{\partial}{\partial y}(D_2 u) + \frac{\partial}{\partial x}(D_{12} u) \right)$$

or

$$\begin{aligned} \frac{\partial u}{\partial t} \quad &= \quad \nabla \cdot \left(\left(-c_1 + \frac{\partial D_1}{\partial x} + \frac{\partial D_{12}}{\partial y}, -c_2 + \frac{\partial D_{12}}{\partial y} + \frac{\partial D_{12}}{\partial x} \right) u \right) + \\ &\quad + \nabla \cdot \left(D_1 \frac{\partial u}{\partial x} + D_{12} \frac{\partial u}{\partial y}, D_2 \frac{\partial u}{\partial y} + D_{12} \frac{\partial u}{\partial x} \right), \end{aligned}$$

which means that there are additional velocity components resulting from the spatial heterogeneity in diffusion. Let the region in which the particle moves be denoted by S. Assume that it is bounded and that particles cannot escape through the boundary. Thus

$$\iint_S u(x, y, t) dxdy = 1 \qquad \forall t > 0.$$

Integrating the Kolmogorov equation over S and using Green's theorem, we get

$$\frac{\partial}{\partial t} \iint_S u dxdy = \int_{\partial S} \left[-c_1 u + \frac{\partial}{\partial x}(D_1 u) + \frac{\partial}{\partial y}(D_{12} u), -c_2 u + \frac{\partial}{\partial y}(D_2 u) + \frac{\partial}{\partial x}(D_{12} u) \right] \cdot \vec{n} ds$$

where $\vec{n} = (n_1, n_2)$ is the unit outward normal to ∂S. Since the left-hand side must be zero the natural boundary condition is

$$\left(-c_1 u + \frac{\partial}{\partial x}(D_1 u) + \frac{\partial}{\partial y}(D_{12} u) \right) n_1 + \left(-c_2 u + \frac{\partial}{\partial y}(D_2 u) + \frac{\partial}{\partial x}(D_{12} u) \right) n_2 = 0 \quad \text{on } \partial S$$

In addition, the initial distribution must be specified, i.e.

$$u(x, y, 0) = u_0(x, y), \quad \text{where} \iint_s u_0(x, y) dxdy = 1.$$

Returning briefly to the discrete model, it is easy to see that, since $\Delta x = \Delta \cos\theta$ and $\Delta y = \Delta \sin\theta$ are the displacements in the *x*- and *y*- directions respectively in one jump, the expected displacements are

$$E[\Delta x] = \Delta^2 q(x, y) \int_{-\pi}^{\pi} \cos\theta \cdot b(\theta) d\theta \quad \text{and} \quad E[\Delta y] = \Delta^2 q(x, y) \int_{-\pi}^{\pi} \sin\theta \cdot b(\theta) d\theta$$

Thus,

$$c_1\left(x,y\right)=\lim_{\tau\to 0}\frac{E\left[\Delta x\right]}{\tau}\quad \text{and} \quad c_2\left(x,y\right)=\lim_{\tau\to 0}\frac{E\left[\Delta y\right]}{\tau}.$$

In view of this we will take the velocity vector to be $\vec{c}(x,v)=(c_1(x,y),c_2(x,y))^T$.

Similarly,

$$E\left[\Delta x^2\right]=\Delta^2 q(x,y)\int\limits_{-\pi}^{\pi}\cos^2\theta\cdot a(\theta)d\theta+O\left(\Delta^3\right)$$

$$E\left[\Delta y^2\right]=\Delta^2 q(x,y)\int\limits_{-\pi}^{\pi}\sin^2\theta\cdot a(\theta)d\theta+O\left(\Delta^3\right)$$

$$E\left[\Delta x\Delta y\right]=\Delta^2 q(x,y)\int\limits_{-\pi}^{\pi}\cos\theta\sin\theta\cdot a(\theta)d\theta+O\left(\Delta^3\right)$$

and thus

$$D_1\left(x,y\right)=\lim_{\tau\to 0}\frac{E\left[\Delta x^2\right]}{2\tau},\quad D_2\left(x,y\right)=\lim_{\tau\to 0}\frac{E\left[\Delta y^2\right]}{2\tau},\quad D_{12}\left(x,y\right)=\lim_{\tau\to 0}\frac{E\left[\Delta x\Delta y\right]}{2\tau}$$

In order to fully specify the model, some further assumptions about the probability density $p(x,y;\theta)=a(x,y;\theta)+\Delta b(x,y;\theta)$ are required. Let a vector field of “preferred directions” be given. In the case when the movements and migrations of fish stocks are being modelled, this field may be defined as a combination of environmental gradients (food density, temperature), a direction towards a specified point or area, typically a spawning area, and the average direction of motion in a neighbourhood of the point *(x,y)*. Let the direction angle of the preferred direction be $\theta_0\left(x,y\right)$. We will assume that the mode of $p(x,y;\theta)$ is at $\theta_0\left(x,y\right)$ for all Δ. It follows that $a(\theta)$ and $b(\theta)$ both have a mode at θ_0. For the general case we can then choose any 2π-periodic functions $a(\theta)$ and $b(\theta)$, satisfying

- $a(x,y;\theta+\pi)=a(x,y;\theta)$ and $b(x,y;\theta+\pi)=-b(x,y;\theta)$
- $\int\limits_{-\pi}^{\pi}a\left(x,y;\theta\right)d\theta=1$
- $a(x,y;\theta)$ and $b(x,y;\theta)$ have a mode at $\theta=\theta_0\left(x,y\right)$

If we assume in addition that *b* is symmetric about $\theta_0\left(x,y\right)$ i.e.

- $b((\theta_0+\theta)=b\left(\theta_0-\theta\right)$

then it follows - in fact all that is required is this last condition and the 2π-periodicity of *b* (Appendix 1) - that $\theta_0\left(x,y\right)$ the direction angle of velocity vector $\left(c_1\left(x,y\right),c_2\left(x,y\right)\right)$, i.e.

$$\vec{c}\left(x,y\right)=\left(c_1\left(x,y\right),c_1\left(x,y\right)\right)^T=D_0 q(x,y)\int\limits_{-\pi}^{\pi}\cos\left(s\right)b\left(\theta_0+s\right)ds\left(\cos\left(\theta_0\right),\sin\left(\theta_0\right)\right)^T$$

Thus $(c_1,c_2)=s(x,y)\left(\cos\theta_0,\sin\theta_0\right)$ where *s(x,y)* is the speed of the particle at *(x,y)*. The “preferred angle” $\theta_0\left(x,y\right)$ is given at each point and if the speed *s(x,y)* is also specified, then (c_1,c_2) is fully determined. The diffusion coefficients can be written as (Appendix 2)

$$\begin{aligned} D_1 &= \frac{D_0}{4} q(x,y) + A\cos(2\phi) - B\sin(2\phi) \\ D_2 &= \frac{D_0}{2} q(x,y) - D_1 \\ D_{12} &= A\sin(2\phi) + B\cos(2\phi) \end{aligned}$$

where

$$A = \frac{D_0}{4} q(x,y) \int_{-\pi}^{\pi} a(\theta+\phi)\cos(2\theta)\,d\theta \qquad B = \frac{D_0}{4} q(x,y) \int_{-\pi}^{\pi} a(\theta+\phi)\sin(2\theta)\,d\theta$$

where ϕ is some specified angle. Making the additional assumption that $a(x,y;\theta)$ is symmetric about the angle $\phi(x,y)$ - which is typically $\theta_0(x,y)$ - i.e.

- $a(\phi+\theta) = a(\phi-\theta)$

then *B*=0 in the expressions for the diffusion coefficients. Furthermore, $a(\theta) = \bar{a}(\theta-\phi)$ where $\bar{a}$ is an even function. Thus, if $a(x,y;\theta) = \bar{a}(\theta - \phi(x,y))$, i.e. the position only affects $a(\theta)$ via the angle $\phi(x,y)$, then the coefficient *A* is independent of $\phi(x,y)$. Thus, if *q* is a constant, then *A* is a constant.

One obvious choice for $a(\theta)$ is

$$a(\theta) = b/\sqrt{1 - e^2\cos^2(\theta-\phi)};$$

the polar plot of $a(\theta)$ is simply an ellipse with the major axis in the direction of ϕ. Note that since $a(\theta)$ must integrate to one over $[-\pi,\pi]$, the only free parameter is *e*, the ecccentricity of the ellipse. Assuming that the two angles $\theta_0(x,y)$ and $\phi(x,y)$ are given at each point – usually $\phi(x,y) = \theta_0(x,y)$ - then the general model contains the following unspecified parameters:

- speed $s(x,y) = \|\vec{c}(x,y)\|$;
- *e*(*x*,*y*) – the eccentricity of the ellipse determining the degree of anisotropy in diffusion;
- $D(x,y) = \frac{D_0}{2} q(x,y)$ - the diffusion coefficient.

If *e*=0, then $a(\theta)$ is the density for a uniform distribution, i.e. $a(\theta) = (2\pi)^{-1}$. It follows that $D_{12} = 0$ and $D_1 = D_2$ and the model becomes simply

$$\frac{\partial u}{\partial t} = -\nabla \cdot (\vec{c}u) + \Delta(Du) \qquad (x,y) \in S,$$

$$u(\vec{c}\cdot\vec{n}) - \frac{\partial}{\partial \vec{n}}(Du) = 0 \qquad (x,y) \in \partial S,$$

where

$$D(x,y) = \frac{D_0}{4} q(x,y).$$

It only remains to be specify the preferred angle, $\theta_0(x,y)$ (we will normally take $\phi(x,y) = \theta_0(x,y)$). We define a “comfort” or “utility” function (Reed and Balchen, 1982), which incorporates all the factors which can affect the comfort or “wellbeing” of the fish, i.e. temperature, food density, distance from the spawning grounds, etc.

$$U_t(x,y) = u_t\left(T\left(x,y\right), f\left(x,y\right), d(x,y), ...\right).$$

Here, T denotes temperature, f food density, d the distance to the spawning grounds, and so on. The fish move on the time-varying “comfort surface” defined by u=U_t(x,y) and the gradient of U defines the preferred angle $\theta_0\left(x,y\right)$. The comfort function depends on time, since for example, the attraction towards the spawning grounds is only active for a part of the year and food density believed to be of relatively little importance on migrations to the spawning ground, since feeding is very limited during that time.

Numerical simulations

The capelin stock in the Barents Sea performs extensive seasonal migrations (see Gjøsæter, 1998). In winter and early spring a southward spawning migration takes place towards the coast of North Norway and Russia, where the fish die after spawning. The next generation - which hatched two years earlier – migrates north and north-eastwards to feeding grounds in summer and autumn. The distribution and the spawning locations depend on hydrographic conditions; the distribution extends further north and east in years with a strong inflow of Atlantic waters - i.e. “warm years”- while the capelin are found further to the south and west in “cold years” (Gjøsæter, 1998). Similarly, spawning takes place further east in warm years. We present here a few simulations which give a very simplified picture of capelin migrations in the Barents Sea, based on assumptions about capelin behaviour and reaction to hydrographic factors such as temperature. The spawning and feeding migrations will be represented as a one year cycle - i.e. a southward spawning migration followed by a northward feeding migration.

The comfort function used in the simulations is assumed to depend only on temperature, food density and location of spawning grounds and is taken to be a linear combination of corresponding functions of these three variables

$$U_t\left(x,y\right) = \alpha_1\left(t\right) r(T(x,y)) + \alpha_2\left(t\right) s(f(x,y)) - \alpha_3\left(t\right) d_S(x,y).$$

The functional form of the temperature function r is based on the observation that most fish species have a preferred temperature range; Barents Sea capelin, for example, are found in waters with ambient temperature from -1° C to 2° C in autumn (Gjøsæter, 1998) and Icelandic capelin do not enter into waters with temperatures below -1° C (Vilhjálmsson, 1994). However, Barents Sea capelin often spawn in considerably warmer waters, 5° - 6° C (Gjøsæter, 1998).

Let $[T_1, T_2]$ be the preferred temperature range. We will take

$$r(T) = \begin{cases} -\left(T - T_1\right)^4 & if \quad T \leq T_1 \\ 0 & if \quad T_1 \leq T \leq T_2 \\ -\left(T - T_2\right)^2 & if \quad T_2 \leq T \end{cases}$$

There is thus a tendency to move towards regions where the temperature is within the preferred range and this tendency is stronger in cold waters. The food density function is assumed to have the form

$$s(f) = \frac{f}{h+f},$$

where *h* is a constant. The functional form of *s* means that the tendency to move towards higher food densities decreases with increasing density and is virtually non-existent at high densities i.e. if there is enough food at the present location of the fish.

The distance function $d_S(x, y)$ is defined as follows: a specified region *S* is defined as the spawning grounds; $d_S(x, y)$ is then simply the distance of *(x,y)* from *S*, i.e.

$$d_S(x, y) = \inf_{(\xi,\eta)\in S} \sqrt{(\xi - x)^2 + (\eta - y)^2}$$

and is thus, in particular, zero if $(x, y) \in S$.

In the absence of reliable measurements of food density in the Barents Sea, we will define a feeding area *F*, based on various sources concerning the location of feeding capelin in summer, autumn and early winter (Gjøsæter, 1998), and define a distance function $d_F(x, y)$, analogous to $d_S(x, y)$. The function *s*(*f*(*x,y*)) in the comfort function *U*, is then replaced by the negative of $d_F(x, y)$. This can be regarded as an approximation to a food density field which increases towards *F* and is constant inside *F*.

Fig. 1 shows the Barents Sea with the coast of North-Norway and Russia in the bottom left hand corner and Novaya Zemlya to the east and Svalbard and Franz Josef Land in the northwest and northeast corners respectively. The spawning and feeding areas are shaded, the spawning area is off the Norwegian and Russian coast and the feeding area in the north between Novaya Zemlya and Svalbard. This map covers an area of approximately 1400x1300 km^2.

The temperature field used in the simulations, is based on twice yearly measurements 1990-1995 made available by the Institute of Marine Research in Bergen, Norway. Fig. 2 shows the temperature distribution in March and September 1991. Temperature distributions for other parts of the year were obtained by interpolation.

Figures 3-6 show four different simulation of the spatial distribution through a complete one-year cycle. The partial differential equation for the probability density is solved by a finite element scheme as described in Sigurdsson *et al* (2002). Constant speed – 40 km/day – and isotropic diffusion is assumed in all the simulation runs described below. The coefficients α_i in the expression for the utility function are taken to be zero or one, i.e. $\alpha_1(t) = 1$ for all t; $\alpha_2(t) = 1$ on the feeding migration and zero otherwise; and $\alpha_3(t) = 1$ on the spawning migration and zero otherwise. Each of the Figs 3-6 shows the spatial distribution at six different points in time over a one-year period. The system is first simulated through one cycle (i.e. a year) in order to reduce the effects from the initial spatial distribution. Also shown is the trajectory of the expected value $\{(\bar{x}(t), \bar{y}(t)) : t>0\}$ where

$$\bar{x}(t) = \iint_S xu(x, y, t)dxdy; \qquad \bar{y}(t) = \iint_S yu(x, y, t)dxdy$$

Fig. 3 shows the simulated distribution for what we refer to as the Base Case: in line with the observations about ambient temperature, the preferred temperature range is assumed to change linearly from [0,3] ° C when the fish are on the feeding grounds to [3,6] ° C when on the spawning grounds; the feeding area is attracting during feeding migration; and the diffusion coefficients are 0.672 km^2/day outside the spawning and feeding areas and 100.8 km^2/day inside, i.e. larger by a factor of 150. These values are not based on any data but are selected to give what looks like a “reasonable” picture of the spatial distribution and spread. We assume that diffusion is lower when the fish are actively migrating, than on the feeding grounds or spawning grounds, when they are either searching for food or for suitable spawning locations. This figure shows the movement from the feeding grounds between Svalbard and Novaya Zemlya to the spawning grounds of the coast of North Norway and Russia and the return journey.

In Fig. 4, everything is the same as for the Base Case except the preferred temperature range is constant at [0,3] ° C throughout the year. The consequence of this is that spawning takes place further east and a component of the stock does not reach the spawning grounds since its path is blocked by a tongue of warm water (cf. Fig.2). This can also be seen from the trajectory of the expected value or the centre of mass; it does not reach as far south as in Fig.3.

Fig. 5 shows the results when the diffusion coefficients are increased tenfold, but everything else is unchanged from the Base Case. As expected, the spatial patterns are of the same shape as in Fig. 3, but the spread is greater.

Finally, Fig. 6 shows what happens when there is no specified feeding area, i.e. no attraction to a feeding area. The temperature distribution and the preferred temperature range prevent the northward movement and the stock remains stuck near the spawning grounds, once they are reached. The temperature alone is unable to drive the stock to the feeding grounds. In fact, two of the coefficients in the comfort function, α_2 and α_3 are zero after spawning and if the fish happen to be in a region where the temperature is within the preferred range, the gradient of the comfort function is zero and there is therefore no movement, except diffusion.

In order to investigate how the spawning locations vary with temperature distribution, we simulated the model for the entire period for which temperature distribution is available, i.e. 1990-1995. The results are shown in Fig. 7. The starting point is December 15 1990, but the first year is not shown. The spawning locations are similar in most years, only slight variations, except 1994 when spawning takes place further west. The reason can clearly be seen in Fig. 8, which shows the area where the temperature is within the preferred range. The warm waters did not extend as far east in 1994 as in normal years. However, comparisons with known and assumed spawning areas in 1991 and 1994 (Gjøsæter, 1998) reveals that spawning in 1994 was confined to a relatively small area close to the eastern border of the overall spawning region, and extended much further west in 1991. This is not quite in accordance with the results shown here, demonstrating that further experimentation with the simulation model is required. It should be pointed out that the spawning stock size was very low in 1994-95 (Gjøsæter et al , 1998).

Discussion and further work

We have presented a simulation model of the migration of Barents Sea capelin. The model itself is reasonably general, but a number of simplifying assumptions have been made, mostly due to lack of knowledge about which biological and environmental factors influence migration patterns and how. The lack of understanding of the mechanisms, which govern migrations means that the model will inevitably be based more on what are believed to be reasonable assumptions about fish movements than on firm biological knowledge and data.

The full suite of possible behaviours, which the model can exhibit when the assumptions are less restrictive, needs to be explored further. Variable speeds and anisotropic diffusion are likely to be much closer to reality than the isotropic diffusion used in the simulation described here. The effect of including other factors such as salinity, location of the ice edge, etc must also be explored. Capelin appear in general to favour low temperatures as stated earlier and this preference is reflected in the form of the temperature part of the comfort function, but spawning often takes place in considerably warmer waters. An attempt was made to address this in the simulations described above by using a variable preferred temperature range. Preliminary comparisons of the simulation results with known and assumed spawning locations shows that some experimentation and modification of the preferred range are necessary. Possible inclusion of vertical migrations should be looked into, since fish can to some extent choose a more suitable temperature by changing depth. It is possible to introduce some interaction between the neighbouring fish/particles by adding a term in the comfort function whose gradient is in the direction of the average velocity in a neighbourhood of a point *(x,y)*, i.e. $\int\limits_{\Omega(x,y)} u(\xi,\eta;t)\vec{c}(\xi,\eta;t)d\xi d\eta$. This will make the governing equation non-linear. Finally, some of the parameters, such as start of spawning migration, speed, etc are likely to depend on some internal variable, such as energy reserves or state of maturity. This physiological variable should be modelled and the partial differential equation describing the motion should be linked - via some parameters such as speed – to the physiological model.

Since the probability density function $u(x,y;t)$ can also be regarded as giving the distribution of the density of the fish stock, the simulated distribution can be compared to observed distributions of the migrating fish stock and estimates of parameters obtained and the model tested. No attempt has been made to do this here except in a very crude way by visualization of the simulated and observed distributions. A proper estimation of all the unknowns in the model is clearly impossible due to the lack of data and a subset parameters which are suitable for estimation should therefore be selected and the others fixed. Parameters which are candidates for estimation are for example cruising speed, which should be variable within the year, and the preferred temperature range. A proper calibration of the model, estimation of parameters and testing of the model based on observations is the next task.

Acknowledgements. This work was supported in part by the Science Research Council of Iceland and by EU grant QLK5-CT1999-01609. The permission of the Institute of Marine Research, Bergen, Norway to use the temperature data is gratefully acknowledged as are dr. Geir Huse and dr. Geir Odd Johansen for preparing this data into a suitable form.

References

Ø. Fiksen, J. Giske and D. Slagstad. A spatially explicit fitness-based model of capelin migrations in the Barents Sea. Fish. Oceanogr. 4 (1995), 193-208.

H. Gjøsæter. The population biology and exploitation of capelin (Mallotus villosus) in the Barents Sea. Sarsia 83(1998):453-496.

H. Gjøsæter, A. Dommasnes and B. Røttingen. The Barents Sea capelin stock 1972-1997. A synthesis of results from acoustic surveys. Sarsia (1998) 83:497-510

D. Grünbaum. Translating stochastic density-dependent individual behaviour with sensory constraints to an Eulerian model of animal swarming. J. Math. Biol.(1994) 33: 139-161.

D. Grünbaum and A. Okubo. *Modelling social animal aggregations* in “Frontiers in Mathematical Biology” (ed. S. Levin). Springer 1994.

S. Hubbard, Petro Babak, Sven Sigurdsson and K.G. Magnússon. A model of the formation of fish schools and migrations of fish. Report RH-28-2001, Science Institute, University of Iceland. To appear in Ecological Modelling.

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.

A. Okubo. Dynamical aspects of animal grouping: swarms, schools, flocks, and herds. Adv. Biophys., 22 (1986), 1-94.

B.L. Partridge. The structure and function of fish schools. Scientific American June 1982, 90-99.

M.J. Reed and J.G. Balchen. A multidimensional continuum model of the fish dynamics and behaviour: Application to the Barents Sea capelin (*Mallotus villosus*) Modelling, Identification and Control 3(1982): 65-109.

S.Þ. Sigurdsson, Kjartan G. Magnússon, Petro Babak, Stefán F. Guðmundsson and Eva Hlín Dereksdóttir. Dynamic continuous model of fish migration.. Report RH-25-2002, Science Institute, University of Iceland.

S. Tjelmeland and B. Bogstad, B. *Biological Modelling*, in “Models for Multispecies Management”, (ed. T. Rødsethe). Physica-Verlag 1998.

J. Toner, J. and Y. Tu. Flocks, herds and schools: A quantitative theory of flocking. Physical Review E, 58(4,1998): 4828-4858.

H. Vilhjálmsson. The Icelandic capelin stock. Journal of the Marine Research Institute Reykjavik, Vol. XIII (1994), No. 1. 281 pp.

Appendix 1 – Formulation of the velocity vector $\vec{c}$.

Assume that $b(\theta)$ is 2π- periodic and symmetric about θ_0, i.e.

- $b(\theta + 2\pi) = b(\theta)$
- $b(\theta_0 + \theta) = b(\theta_0 - \theta)$

Then

$$\int_{-\pi}^{\pi} \cos\theta \cdot b(\theta)\, d\theta = \int_{-\pi-\theta_0}^{\pi-\theta_0} \cos(\theta_0 + \xi) \cdot b(\theta_0 + \xi)\, d\xi = \int_{-\pi}^{\pi} \cos(\theta_0 + \xi) \cdot b(\theta_0 + \xi)\, d\xi$$

since the integrand 2π- periodic.

This equals

$$\int_{-\pi}^{\pi} \cos\theta_0 \cos\xi \cdot b(\theta_0 + \xi)\, d\xi - \int_{-\pi}^{\pi} \sin\theta_0 \sin\xi \cdot b(\theta_0 + \xi)\, d\xi = \cos\theta_0 \int_{-\pi}^{\pi} \cos\xi \cdot b(\theta_0 + \xi)\, d\xi$$

since the integrand in the latter integral is an odd function of ?.

Similarly

$$\int_{-\pi}^{\pi} \sin\theta \cdot b(\theta)\, d\xi = \sin\theta_0 \int_{-\pi}^{\pi} \cos\xi \cdot b(\theta_0 + \xi)\, d\xi$$

and hence

$$(c_1(x, y), c_2(x, y))' = D_0 q(x, y) \int_{-\pi}^{\pi} \cos\xi \cdot b(\theta_0 + \xi)\, d\xi\, (\cos\theta_0, \sin\theta_0)'$$

$$\begin{aligned}
D_1\left(x,y\right) &= \frac{D_0}{2}q\left(x,y\right)\int\limits_{-\pi}^{\pi}\cos^2\theta\cdot a(x,y;\theta)d\theta \\
&= \frac{D_0}{2}q\left(x,y\right)\int\limits_{-\pi-\phi}^{\pi-\phi}\cos^2\left(\phi+\xi\right)\cdot a(\phi+\xi)d\xi \\
&= \frac{D_0}{2}q\left(x,y\right)\int\limits_{-\pi}^{\pi}\cos^2\left(\phi+\xi\right)\cdot a(\phi+\xi)d\xi \\
&= \frac{D_0}{2}q\left(x,y\right)\int\limits_{-\pi}^{\pi}\frac{1}{2}\left(1+\cos 2\left(\phi+\xi\right)\right)\cdot a(\phi+\xi)d\xi \\
&= \frac{D_0}{2}q\left(x,y\right)\cdot \\
&\quad \left(\int\limits_{-\pi}^{\pi}\frac{1}{2}\cdot a(\phi+\xi)d\xi+\cos 2\phi\int\limits_{-\pi}^{\pi}\frac{1}{2}\cos 2\xi\cdot a(\phi+\xi)d\xi-\sin 2\phi\int\limits_{-\pi}^{\pi}\frac{1}{2}\sin 2\xi\cdot a(\phi+\xi)d\xi\right) \\
&= \frac{D_0}{2}q\left(x,y\right)\left(\frac{1}{2}+\cos 2\phi\int\limits_{-\pi}^{\pi}\frac{1}{2}\cos 2\xi\cdot a(\phi+\xi)d\xi-\sin 2\phi\int\limits_{-\pi}^{\pi}\frac{1}{2}\sin 2\xi\cdot a(\phi+\xi)d\xi\right) \\
&= \frac{D_0}{4}q\left(x,y\right)+A\cos 2\phi-B\sin 2\phi
\end{aligned}$$

Similarly

$$\begin{aligned}
D_{12}\left(x,y\right) &= \frac{D_0}{2}q\left(x,y\right)\int\limits_{-\pi}^{\pi}\cos\theta\sin\theta\cdot a(x,y;\theta)d\theta \\
&= \frac{D_0}{2}q\left(x,y\right)\int\limits_{-\pi}^{\pi}\cos\left(\phi+\xi\right)\sin\left(\phi+\xi\right)\cdot a\left(\phi+\xi\right)d\xi \\
&= \frac{D_0}{4}q\left(x,y\right)\int\limits_{-\pi}^{\pi}\sin 2\left(\phi+\xi\right)\cdot a\left(\phi+\xi\right)d\xi \\
&= \frac{D_0}{4}q\left(x,y\right)\left(\sin 2\phi\int\limits_{-\pi}^{\pi}\cos 2\xi\cdot a\left(\phi+\xi\right)d\xi+\cos 2\phi\int\limits_{-\pi}^{\pi}\sin 2\xi\cdot a\left(\phi+\xi\right)d\xi\right) \\
&= A\sin 2\phi+B\cos 2\phi
\end{aligned}$$

Figure 1. The Barents Sea with the attracting spawning and feeding regions.

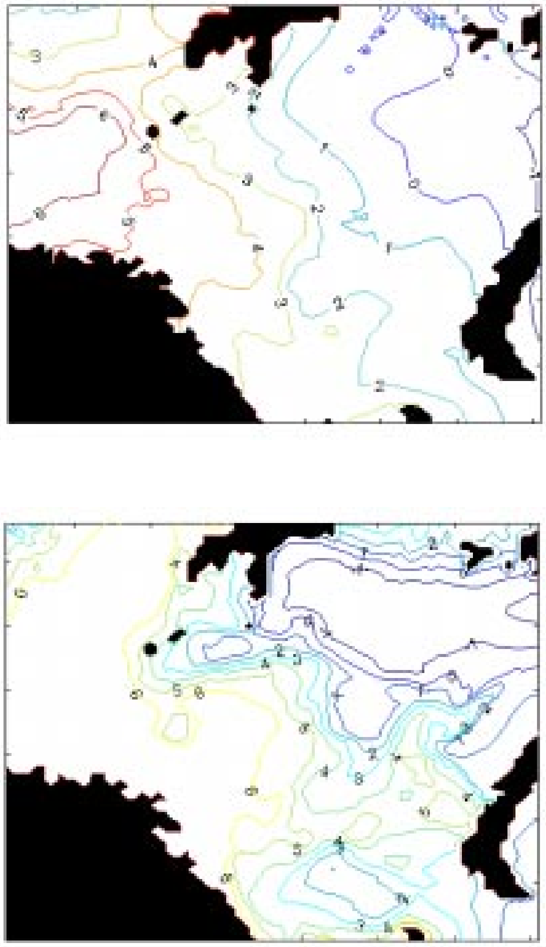


Figure 2. Temperature distribution in March 1991 (upper) and in September 1991 (lower).

December15, 1990

June 1, 1991

January 15, 1991

September 15, 1991

March 15, 1991

December 15, 1991

Figure 3. The simulated spatial distribution for Barents Sea capelin over a one year cycle. The preferred temperature range changes linearly from [0,3]°C to [3,6]°C. This is termed the base case simulation.

December15, 1990

June 1, 1991

January 15, 1991

September 15, 1991

March 15, 1991

December 15, 1991

Figure 4. The simulated spatial distribution for Barents Sea capelin over a one year cycle. The preferred temperature range is constant at [0,3]°C.

December15, 1990

June 1, 1991

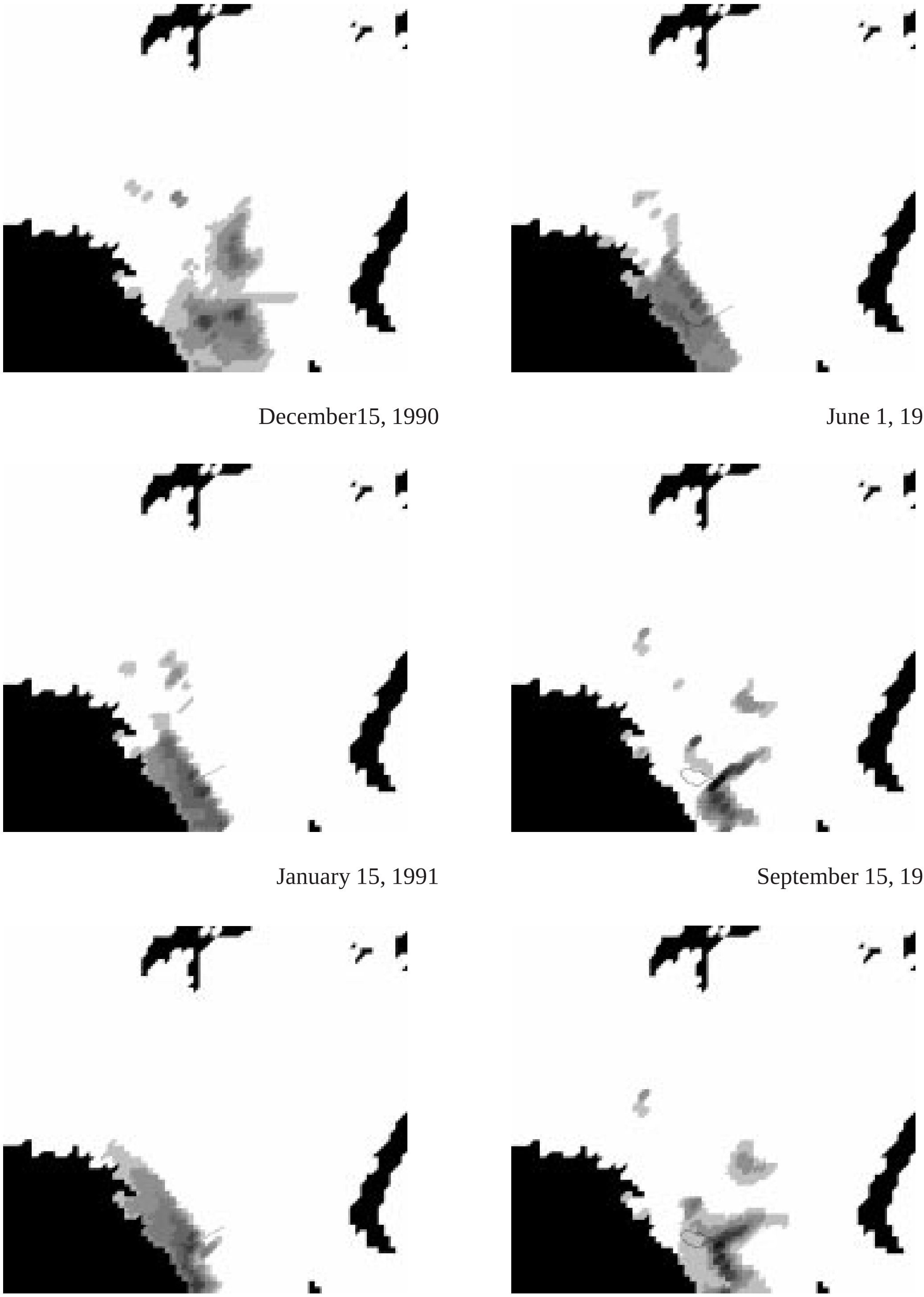
January 15, 1991

September 15, 1991

March 15, 1991

December 15, 1991

Figure 5. The simulated spatial distribution for Barents Sea capelin over a one year cycle. The preferred temperature range changes linearly from [0,3]°C to [3,6]°C. The numerical value of the diffusion constants is ten times higher than in the base case simulation.



December15, 1990

June 1, 1991

January 15, 1991

September 15, 1991

March 15, 1991

December 15, 1991

Figure 6. The simulated spatial distribution for Barents Sea capelin over a one year cycle. The preferred temperature range changes linearly from [0,3]°C to [3,6]°C. No feeding area is specified.

December15, 1990 June 1, 1991

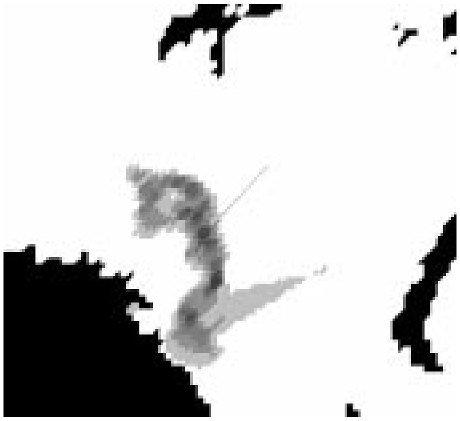
January 15, 1991 September 15, 1991

March 15, 1991 December 15, 1991

Figure 7. The simulated spatial distribution for Barents Sea capelin over a period of five years. The preferred temperature range changes linearly from [0,3]°C to [3,6]°C.

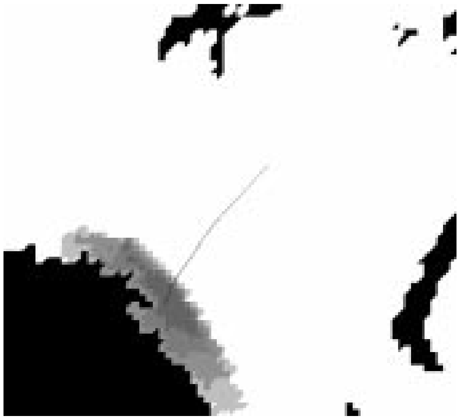
December15, 1990

June 1, 1991

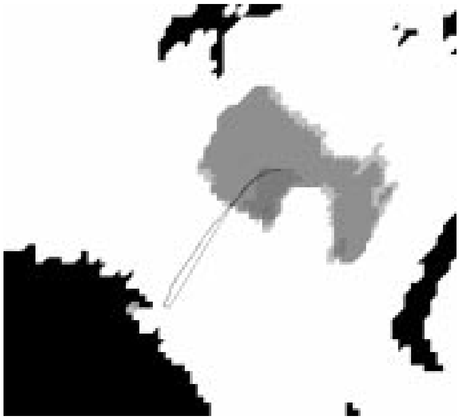


January 15, 1991

September 15, 1991



March 15, 1991



December 15, 1991

Figure 7. Cont'd.

December15, 1990

June 1, 1991

January 15, 1991

September 15, 1991

March 15, 1991

December 15, 1991

Figure 7. Cont’d.

December15, 1990

June 1, 1991

January 15, 1991

September 15, 1991

March 15, 1991

December 15, 1991

Figure 7. Cont'd.

December15, 1990

June 1, 1991

January 15, 1991

September 15, 1991

March 15, 1991

December 15, 1991

Figure 7. Cont’d.

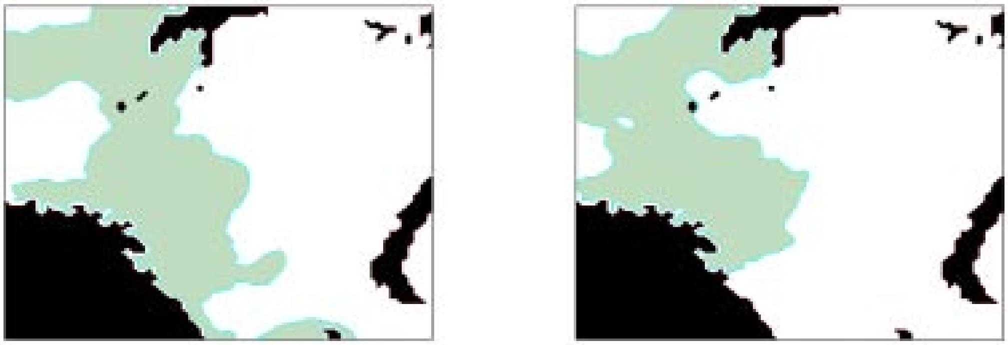


Figure 8. The shaded areas show where the temperature in March for the years 1991 (left) and 1994 (right) is within the preferred range.

D.3 Including tagging experiments in simulations of fish populations

Gunnar Stefansson[1,2] **Audbjorg Jakobsdottir**[1] **and Sigurdur Hannesson**[1]

[1]Marine Research Institute, Reykjavík, Iceland
[2]Univ. of Iceland,Reykjavík, Iceland

Abstract: This paper describes an implementation of tagged sub-population within simulations of of fish populations for the purpose of enhancing estimation of unknown parameters.

Population dynamics

The following describes in fairly generic terms the population dynamics which can be taken into account in spatially explicit multispecies models of fish populations. All of these processes are implemented in Gadget. These processes need to be taken into account when implementing tagged sub-populations.

Denote by N_{almsrt} the number (N) of fish of species s, age a, length l and maturity stage m, alive in region r and timestep t. This group of fish is a subset of the entire collection of fish of a given species. Such a collection will be referred to as a population (or 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 fish ($m = 0, 1$), though in the case of closed life-cycles, there will be 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_{ls}.

In order to distinguish between predator and prey species in an interaction, the following notation is used:

s	a general species.
p	a prey species.
P	a predator species.
l	a prey length group.
L	a predator length group.

Accordingly, N_{lp} denotes the number of individuals of a prey species p of length l and N_{LP} the numbers in length group L of a predator P. The notation is solely for convenience as there is no reason mathematically to exclude one predator's prey being another species' predator.

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

- The **migration** a population undertakes on a given time step is described by matrices, $A_{almst} = (a_{almsr_1r_2t})^{r_1=1,..R}_{r_2=1,...,R}$, containing the proportion $a_{almsr_1r_2t}$ of the population that moves from area r_2 to area r_1 and R is the number of areas. Hence, if $u = (u_1, ..., u_R)$ are abundance numbers by area for an age-length-maturity-species group in a population, the area distribution after the migration has taken place, is $A_{almst}u$.
- **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.
- **Growth** in length within population dynamics models may be implemented through growth in weight since the weight increase is more naturally linked to consumption, but alternatively growth

may be according to a fixed growth schedule in length or weight. Growth results in transfers of fish from one length group to other length groups in such a fashion that, when viewed as a whole, the fish population grows on average by the required amount.

- **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.
- **Aging**, 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.

In addition to all the processes described above, two new processes can affect the tagged sub-population:

- Tag loss
- Tagging mortality

Ignoring for the moment these two additional processes, 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 dramatically 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, $p'_{almsrte} = N'_{almsrte}/N_{almsrt}$, in a given group.

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 $p'_{almsrte}$ 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.

Likelihood components

A simulation 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. Since the same process applies to the tagged sub-population, the predicted number of tag returns is just the predicted tagged catch, $\hat{C}_{almsrfte}$. 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.

The treatment of these data is initially taken as Poisson counts. Since the Poisson parameters λ 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. 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 quite 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.

Discussion

The above mathematical components have been implemented in C++ within Gadget. Tests using real data will be performed in 2003.

The level of information available in tagging data is not clear, nor is the required number of tagging experiments in order to obtain a given reduction in variance of migration parameters. Bootstrapping methods will be developed to shed some light on these issues.

D.4 Modelling of the recruitment process with application to Gadget

Bjarte Bogstad, IMR, Bergen. November 2002.

Introduction

Gadget does at present not include a formulation of a closed life cycle for a stock. In this document I first review some studies of modelling of recruitment, and then discuss possible mathematical formulations of a closed life cycle within Gadget. It is assumed that the recruitment is calculated on basis of the mature stock. Modelling of the maturation process is not discussed in this document.

Modelling recruitment – general issues

Ulltang (1996) discussed the problem of predicting recruitment, using Northeast Arctic cod as an example. The paragraph below is taken from his paper, with minor modifications:

As discussed by Paulik (1973) and Rothschild (1986), further insight into the recruitment problem requires a more detailed look at mechanisms operating at different life stages. Both authors noted that a stock-recruitment function R=f(SSB) can be written in a nested form of the type

$$R = f3(f2(f1(SSB))) \tag{1}$$

where f1 might be the egg production as a function of the spawning stock biomass SSB, f2 larval production as a function of egg production, and f3 number of recruits as a function of larval production. The following problems can be identified:

(a) What determines population fecundity (total egg production)

(b) What are the magnitude of and causes of variations in survival from

- eggs to larvae
- larvae to 0-group
- 0-group to recruitment to the fishery

In a population model, not all these phases need to be modelled separately. Some of the functions in the nested formulation of the type given in (1) may be combined, depending on the knowledge available as well as the purpose of the modelling. One may model the relationship between number(N)/individual weight(W)/biomass of the mature stock(SSB) and recruitment R by using e.g. one of the stepwise approaches given below:

SSB -> R(age x) (2)
SSB -> egg -> larvae ->R(age 0) (3)
N, W-> egg -> R(age x) (4)
Number of mature individuals - >R(age x) (5)

Both for implementation purposes and for analyses of model properties, it is most convenient, but not necessary, to assume the recruitment process to be Markovian (like the rest of Gadget). This would mean that the number of recruits at time *t+1* depends only on the mature stock at time *t* and not on the mature stock at earlier time steps. The formulations given in equations 2, 4 and 5 can be implemented without violating the Markovian property by introducing recruits in the model as eggs/larvae in the time step after spawning takes place, and then treat them as a ‘dummy’ population with no interaction with other stocks until they reach the minimum ‘model age’ (x in equations 2, 4 and 5).

It seems reasonable to model recruitment in two phases: First to calculate the number of offspring actually produced at time *t*, and then calculate the number of recruits to be introduced in the model at time *t*+1 as a function of the number of offspring actually produced.

Number of offspring

It is common to use the spawning stock biomass (SSB) as a proxy for the recruitment potential. Let $N_{a,i,s}$ *(t)* and $W_{a,i,s}$ *(t)* be the number and individual weight of individuals in age group *a*, length group *i* and stage *s* at time *t*. The recruits are produced in the same area as the mature stock is found, and thus an area index is not needed. Let *u* be the stage index for mature fish. The spawning stock biomass is then given as:

$$SSB(t) = \sum_{a,i} N_{a,i,u}(t) W_{a,i,u}(t) \qquad (6)$$

As a first approximation, one may assume that the egg production is proportional to SSB:

$$TEP(t) = \chi_1 SSB(t) \qquad (7)$$

However, the number of eggs produced per unit biomass of mature fish will generally depend on the sex distribution, the fish size (length/weight) and possibly other factors. For Northeast Arctic cod, a formula for the number of eggs spawned (TEP) has been developed (Marshall et al., 2001):

$$TEP(t) = \chi_2 \sum_{a,i} N_{a,i,v}(t) l_i^{\alpha} W_{a,i,v}(t)^{\beta} \qquad (8)$$

where l_i is the mean length of fish in length group *i* and the subscript *v* denotes mature female fish.

For marine mammals, a first approach could be to model the number of pups/calves, *NR(t)* as proportional to the number of mature female individuals:

$$NR(t) = \sum_{a,i} \chi_3 N_{a,i,v}(t) \qquad (9)$$

where χ_3 is the number of offspring born per mature female individual. This could be refined by making χ_3 age- or size-dependent. In Bogstad et al. (1997), χ_3 was made age-dependent, but this was done because the marine mammal stocks were not divided into an immature and a mature part, and thus $\chi_3(a)$ represented the proportion mature at age (no length structure was used), multiplied by the number of offspring born per mature individual.

Equations 7-9 can be unified into one formula describing the number of offspring (or some proxy for it), *NO(t)*, as a function of the number and individual weight of each age and length group of the female mature stock:

$$NO(t) = \chi \sum_{a,i} N_{a,i,u}(t) l_i^{\alpha} W_{a,i,u}(t)^{\beta} \qquad (10)$$

Equation (7) can then be obtained by setting $\alpha = 0$ and $\beta = 1$ in (10), while equation (9) can be obtained by setting $\alpha = 0$ and $\beta = 0$ in (10). The stage index *u* may here denote either the mature stock or the female mature stock. This unification into one formula is very useful for implementation in Gadget.

From number of offspring to modelled recruitment

The transition from SSB to recruitment given in (2) as well as the steps in the transition (survival) from eggs to later stages given in (3) and (4) can be described by different mathematical formulations. Two formulae which are often used to describe such transitions are the Beverton-Holt and Ricker stock-recruit relationships. Together with the Shepherd model, they can be described by a common formula. If we replace spawner abundance in the Beverton-Holt and Ricker stock-recruit relationship with *NO(t)* from equation (10), combine the functional forms in a similar way as done by Myers et al., (1994), and assume the recruitment process to be Markovian, we get:

$$R(t+1) = \frac{\mu NO(t) e^{-\gamma NO(t)}}{\tau + NO(t)^{\kappa}} \qquad (11)$$

, where *R(t+1)* are the number of recruits at time *t+1*. If γ =0 and $\kappa = 1$ in equation (11), we have the Beverton-Holt type relationship. The recruitment then approaches R$_{max} = \mu$ asymptotically as *NO(t)* approaches infinity. τ is the value of *NO(t)* giving 50% of maximum recruitment (λ).

If $\kappa = 0$ and $\tau = 0$ in equation (11), we have the Ricker-type relationship. The maximum recruitment occurs when *NO(t)=1/*γ and is given by $R_{max} = \mu/\gamma e$. λ is given by

$$2e(\gamma\lambda)e^{-\gamma\lambda} - 1 = 0 \qquad (12)$$

The solutions of equation (12) are given approximately by

$$\lambda = \frac{0.231961}{\gamma} \qquad \text{or} \qquad \lambda = \frac{2.678347}{\gamma} \qquad (13)$$

$\gamma = \kappa = 0$ in equation (11) give recruitment proportional to the number of offspring.

The description of recruitment given here only concerns the number of recruits. Length distribution of recruits can be handled essentially in the same way as in the present implementation of Gadget, where the length distribution is calculated based on the mean length and the standard deviation of the mean length of recruits.

Implementation issues:

Recruitment should be possible in every time step.

Mortality and weight loss may also be associated with spawning. Should these processes be implemented as a part of the spawning process, or should they be included together with the other mortality/growth processes?

When running Gadget, one should be able to choose between using a stock/recruitment relationship and the present implementation of recruitment where the number of recruits for each year is a parameter, which may be estimated.

At present, Gadget includes a simple spawning module, where age-dependent weight loss and mortality due to spawning is implemented. The next step will be to allow the spawning module to generate offspring (eggs/larvae/pups/calves), which will drift/migrate to nursery areas using the migration module. Also, it should be considered whether weight loss and mortality due to spawning should be made functions of other state variables than age. There should be consistency (not equality) between the spawning (birth) weight loss and the biomass of offspring.

Temperature and other environmental factors could be introduced in equations (10) and (11), but this has not been discussed here since the environmental influence cannot be written on a generic functional form.

Acknowledgements

Thanks to Kristin Guldbrandsen Frøysa, IMR, for her valuable comments to the manuscript.

References:

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.

Marshall, C. T., Yaragina, N. A., and Thorsen, A. 2001. Evaluating the effectiveness of biological reference points in conserving reproductive potential. ICES C. M. 2001/V:17.

Myers, R. A., Rosenberg, A.A., Mace, P. M. Barrowman, N., and Restrepo, V. R. 1994. In search of threshold for recruitment overfishing. ICES J. mar. Sci. 51:191-205.

Paulik, G. J. 1973. Studies of the possible form of the recruitment curve. Rapports et Proces-verbaux des Reunions du Conseil International pour l'Exploration de la Mer, 164:303-315.

Rothschild, B. J. 1986. Dynamics of Marine Fish Populations. Harvard University Press. 277 pp.

Ulltang, Ø. 1996. Biological knowledge in stock assessment. ICES J. mar. Sci. 53:659-675.

E GADGET

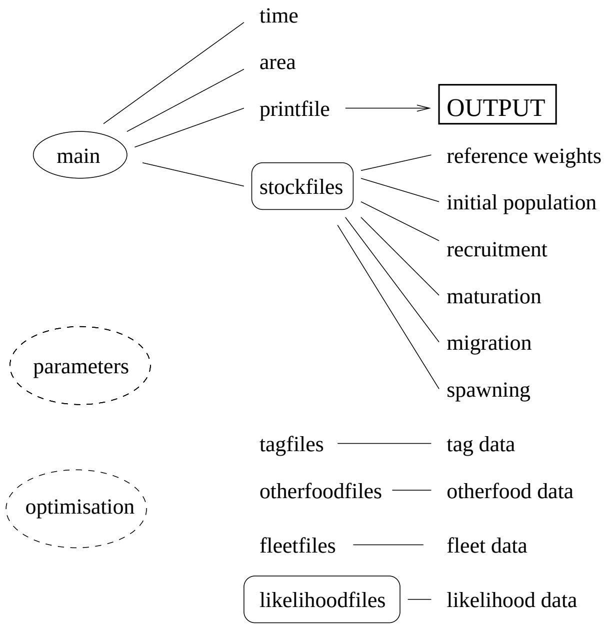
E.1 GADGET software - 2002 updates

James Begley,
Marine Research Institute, Iceland

February, 2003

1 Data Format

Gadget can run complicated statistical ecosystem 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 “real” 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. Gadget models are, therefore, both data- and computationally- intensive. An indication of the data input files required is shown in the diagram below.



Generating the required data files to run a Gadget model used to be a time consuming and error prone task. The old format for the input data was to give the data in a series of matrices, with the size of these matrices depending on the range and amount of data. Whilst this meant that the data files were small and easy to input into Gadget, it meant that they were difficult to generate automatically, and were not in a human-friendly format.

The input data is now in a column-based format, which means that the input data files can easily be generated using a simple extraction routine from a database. This means that the construction of a new Gadget model can be done reasonably quickly, and the models will be free of any formating errors in the Gadget input files that could prevent Gadget models from running. These data files are also much easier for the user to read, however they also much longer, and so require more storage space.

The basic format for the new data files is to display the data in a number of columns, with the first two columns giving time information (year and step), followed by a number of columns giving details about the data (eg. area, age or length), and then columns giving the data itself. The columns describing the data are treated as simple text labels, with the label defined in separate files (termed an "aggregation files") which are read into Gadget before the data files. Any data associated with a text label in the data files that has not been specified in the aggregation files is ignored.

The data format is explained in more detail in the Gadget User Guide (which is available from the MRI ftp site) and in the worked example, contained in Annex E.3 of this report. An examination of a section of a data file is given below. For this example (from a section of the length-based survey indices for Icelandic Haddock), the data file contains the following information:

```
; year  step  area   length  number
1985    1     north  50cm    13501
1986    1     north  50cm    8092
1987    1     north  50cm    25936
1988    1     north  50cm    30964
1989    1     north  50cm    48066
1990    1     north  50cm    35847
...
```

This section of data shows that there are 13501 samples taken from the area "north" with a length of "50cm". However, for this data to be used by Gadget, the labels "north" and "50cm" need to be defined in an aggregation file. The area aggregation file will specify how the area labels are mapped onto Gadget areas, and an example of this would be:

```
north   1  3
south   2
```

This has specified that the label "north" should be interpreted as a combination of Gadget areas 1 and 3, and the area "south" covers area 2 (which is not shown in the example data file). In a similar manner, the length aggregation file would specify how the label "50cm" maps onto the length groups used by Gadget, so an example of this would be

```
45cm  45.5  47.5
50cm  47.5  52.5
55cm  52.5  57.5
```

This has specified that the label "50cm" maps onto the length groups that cover the lengths from 47.5cm to 52.5cm that have declared in Gadget.

2 Tagging

It is now possible to model tagging experiments using the Gadget software. This is done by defining a set of tagging experiments, giving details of the time, location and length of the stock when they were tagged. This tagged sub-stock is then followed through the Gadget model, as they grow, eat, mature, migrate and die along with the rest of the stock. A likelihood component is then used, to compare the tagged stock within the model to the recaptures from the tagging experiment. This aspect of the Gadget software is still under developement.

3 Parameter Files

The format for the parameter files has changed, so that the format for both the input to, and output from, Gadget are the same. This means that the output from one run can be directly used as the input to a second run. The parameter input file specifies the name of the parameters, an initial value, lower and upper bounds and a flag to indicate whether the parameter is to be optimised or not. The final values of the parameters, along with the bounds and optimiser switch, are written to a file (called “params.out”) which can then be used as the starting point for a subsequent run, if required.

4 Paramin

The optimisation algorithms used in Gadget can be computationally intensive, so to speed up the processing time a parallel processing optimiser, called “Paramin” has been developed which can run on computers that have the PVM libraries installed on them. PVM (“Parallel Virtual Machine”) is a freely available library that can treat a network of computers as a single parallel processing machine. The PVM libraries are available as a free download from the PVM website (see `http://www.epm.ornl.gov/pvm/pvm_home.html` for more details).

Paramin can implement a combination of up to 3 minimisation algorithms. These are Simulated Annealing, Hooke & Jeeves, and a quasi-Newton algorithm BFGS, in that order. The input to, and output from paramin is in the same format as that for Gadget. In particular, the parameter input and output files are identical, so the output from a Paramin run can be used as the input to a Gadget run, and vice versa.

The Hooke & Jeeves and Simulated Annealing algorithms used in Paramin are parallel extensions of the algorithms used in Gadget. The BFGS algorithm is a iterative gradient method that starts at an initial value, and successively generate new points, so that the function value decreases in each iteration. This is done by choosing a direction that makes an angle greater than 90 degrees with the gradient of the function, using an inexact linesearch called the Armijo condition. This is done by trying at first an initial stepsize an than reducing the stepsize until a value that satisfies the Armijo condition is found. Since the objective function is not necessarily differentiable, the gradient is calculated numerically.

5 More Information

Gadget is a complex software tool, and it is recognised that there might be some difficulties in setting up models to run correctly, and interpreting the output from the models. To assist users in using Gadget, there is an email address that questions, comments and suggestions should be sent to. It is `mailto:gadgethelp@hafro.is`. Emails sent to this address should get answered within a couple of working days.

Alternatively, there is a Gadget web page set up which has some useful information that might help solve Gadget problems. The address for the website is `http://www.hafro.is/gadget`. From the Gadget website, there are links to the latest version of the software, a list of changes to the software and a link to the most up-to-date copy of the Gadget User Guide.

E.2 A composite parallel algorithm for minimizing expensive functions: A preliminary report

Gunnar Stefansson

Univ. of Iceland and Marine Research Institute
Reykjavik

Abstract

Minimization of 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 simulated annealing, the Hooke and Jeeves algorithm and BFGS. 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.

1 Introduction

The problem of minimizing complicated nonlinear functions is quite common. The present paper describes an algorithm for undertaking such minimizations 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 Stefansson and Palsson (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

$$min\{f(x) : x\epsilon R^n\}, \tag{1}$$

where f 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.

2 Form of algorithms

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 specialized for parallelizaton and nonportable and will therefore not be considered further.

Other parallel methods for minimization have emphasized using parallel techniques internally in the minimization algorithm, e.g. as a part of obtaining matrix decomposition. This is clearly not the bottleneck in the present situation where computations within the minimization algorithm are dwarfed by to the time taken for function evaluations.

The approach, therefore, will be to implement a minimization algorithm where the function evaluations are distributed across a computer network, evaluating each single function call on a single processor. The minimization 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.

3 The composite algorithm

3.1 Design principles

Given that there is no guarantee ahead of time that the function is unimodal, there is a need to involve and test a global optimizer, at least to get to the vicinity of the minimum. Simulated annealing is used 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 minimizer which does not require differentiability. Several such algorithms are available but the Hooke and Jeeves methods has been used here.

However, close enough to the optimum the usual properties of likelihood functions should apply so it is expected to exhibit behavior 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 is evaluated.

3.2 Simulated annealing in parallel

Simulated annealing (SA), as described in (Corana et al, 1987), checks randomly generated variations on each coordinate of the currently best point. The only change to make this parallel is to request all coordinates simultaneously.

Upon returning a new and improved (or accepted, rather) point, this is changed to the initial point and subsequent requests are made from the new location.

There may be many requests in the SA queue. Old requests are of course outdated as a new and improved point comes in and hence this queue is a LIFO queue.

Apart from these changes to accommodate parallel computation, the algorithm proceeds as in the Corana et al. (1987) simulated annealing algorithm for multimodal and robust optimization as implemented and modified by by Goffe et al. (1994).

3.3 The Hooke and Jeeves algorithm in parallel

The basic Hooke and Jeeves (HJ) algorithm (Hooke and Jeeves, 1961) searches from an initial point in discrete steps along each direction of each coordinate axis, moving the saerch to each improved point. Having looped through all coordinates, and thus obtained a new point, a pattern search is undertaken by taking a step in the direction from the initial point to the new point.

The parallel version presented here is derived from the sequential implementation of Johnson (1994). The main difference is to request as many coordinate searches as there are available computers, 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. However, there will be approximately a linear increase in performance.

When a positive result is obtained, the new requests must be issued from this revised intermediate point. If there are few processors and/or the step size is large, there will not be in conflict with the sequential algorithm. On larger networks or with smaller stepsizes, however, there will be many returns with improvements from different coordinates and there is therefore a conflict which needs to be resolved.

The approach taken is to re-issue coordinate searches from each new intermediate best point, though with a maximum number of calls for each coordinate.

Having thus obtained a previous and a new point, a pattern search using the difference between these as a direction vector is undertaken as in the sequential versiona. In the parallel setting there is no reason to request only one point and hence a sequence of points in this direction is requested, using multiples of the default stepsize.

In this implementation the number of requests are simply set so that the outstanding requests match the number of available processors.

3.4 The BFGS algorithm in parallel

The BFGS technique implemented here is as described e.g. in Kennedy and Gentle (1980). Notably, analytical derivatives are not available and hence numerical differentiation is used.

3.4.1 Derivatives

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

$$\frac{f(x) - f(x - \delta e_i)}{\delta}. \tag{2}$$

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

$$\frac{f(x + \delta e_i) - f(x - \delta e_i)}{2\delta} \tag{3}$$

is reverted to. 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 more accurately.

The requests for function evaluations are all made simultaneously. The gradient computation needs to wait for the last evalution 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 it is currently idle.

3.4.2 The line search

Define $g(\alpha) = f(x + \alpha h)$ where x is the current point, $h = H\nabla f(x)$ and H is the approximation to the inverse Hessian matrix in the current step.

The Newton step of $\alpha = 1$ is a natural first test in the line search (Kennedy and Gentle, 1980) and thus at least one single function evaluation needs to be undertaken at $g(\alpha) = f(x + \alpha h)$ for $\alpha = 1$. However, it would be a waste not to use the opportunity to obtain function evaluations for a range of α-values, since computing one function value per processor takes no longer. Thus, at the first linesearch evaluation, with only information on $g(0)$ and $g'(0)$, a judicial choice must be made on values of α to be used. This choice depends on a number of parameters which can be modified and performance should be investigated by modifying these on a case-by-case basis.

The goal of the initial choice of α-values is to have a very high probability of not needing another round of function evaluations. In the final phases of minimization, it is likely that $\alpha = 1$ will be sufficient. However, when the algorithm is in its initial phase, it is impossible to know *a priori* whether a suitable choice of α is close to 0 or for very high values. Obvious search mechanisms will search from $\alpha = 1$ through $\alpha = 2^{-n}$ towards zero or through $\alpha = 2^n$ towards infinity, as the case may be.

Once all the function values have been requested, the results will appear one by one and each can be subjected to testing using any available criterion. The simplest would be to verify whether the best value of α has been bracketed in a small enough interval. However, convergence may be improved upon by instead using tests which evaluate whether the decrease in function value is sufficient to guarantee convergence of the procedure as a whole. These tests must not use derivatives except $g'(0)$. The test used here is the Armijo criterion (Bertsekas, 1999).

The first modification to the serial Armijo method is to reqest several the α-values simultaneously. The α-values are generated of the form $\alpha = \beta^m s$ for $m = 1,\ 2,\ \ldots, p$ (where $\beta < 1$ is fixed and $s = 1$ by default). 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 m 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 $\alpha = \beta^m s$ and accept the one which satisfies the criterion and corresponds to the smallest such integer m. 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 m from those returned.

In the event that the first series of line search evaluations does not provide a sufficiently good α-value, a second phase must be started and this is done simply via continuing the sequence. Since $g\prime(0) < 0$ (see below), this process is guaranteed to terminate.

3.4.3 Hessian update

The inverse Hessian is updated using the BFGS algorithm, as described e.g. in Bertsekas (1999).

The Hessian matrix, H, should always be positive definite and hence the derivative of g, $g\prime(\alpha) = \nabla f(x + \alpha H \nabla f(x))^t H \nabla f(x)$ which satisfies $g\prime(0) = \nabla f(x)^t H \nabla f(x)$ should also satisfy $g\prime(0) > 0$. In cases when $g\prime(0) < 0$ the current H is clearly not positive definite, most likely due to numerical problems, and hence the BFGS algorithm is restarted.

Normally, H is initialized to I, but optionally in difficult problems, H can be initialized 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 $(x + \delta e_i f(x + \delta e_i))$.

3.5 Further implementational details

The performance of the combined algorithm is dominated by the relative number of function calls allowed, or accuracy requirement, for each method. These pairs of numbers are therefore defined as the primary parameters in the combined algorithm.

4 Preliminary, qualitative results

The SA algorithm selected here requires an order of magnitude more function calls than either of the others, even for the loosest convergence criterion. This is easily seen by considering the minimum number of loops involved in resetting the temperature criterion even just once. For this reason SA should only be used when obtaining initial values or starting points in a new problem. Only rarely will the SA be justified for general use as a front-end in its current form.

The HJ and BFGS algorithms are seen to perform fairly similarly in terms of getting to the final point. In fact, in the sequence SA-HJ-BFGS, HJ alone is in most cases quite adequate for finding the minimum. This conclusion does not hold in the multimodal situation nor does this give an estimate of the inverse Hessian at the solution.

Naturally, for a well-behaved problem, the BFGS method 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 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.

5 Discussion

Alternative implementations of SA should be tested, with an emphasis on using much smaller sample sizes for each temperature change. This might make SA a feasible front end in the generic procedure.

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.

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. This could have considerable potential in getting a better starting direction than using $H = I$.

Acknowledgements

The parallel mechanism was implemented using PVM (Geist et al., 1994) using a low-level C++ interface programmed by A. Jakobsdottir who also streamlined the code. The parallel versions of the various minimisation algorithms have been programmed mainly by Th. L. Thorarinsdottir and K. Y. Jonsdottir, who also rewrote the original FORTRAN simulated annealing program in C++. Considerable debugging and testing was done by D. Howell. This work was supported in part by EU grant QLK5-CT1999-01609.

References

Bertsekas, D. P. 1999. Nonlinear programming. Athena Scientific, 708pp.

Corana A, Marchesi M, Martini C, Ridella S, 1987. Minimizing Multimodal Functions Of Continuous-Variables With The Simulated Annealing Algorithm. ACM Transactions On Mathematical Software 13 (3): 262-280

Al Geist (Editor), Adam Beguelin, Jack Dongarra, 1994. PVM: Parallel Virtual Machine: A Users' Guide and Tutorial for Network Parallel Computing (Scientific and Engineering Computation). MIT Press, 299pp.

Goffe WL, Ferrier GD, Rogers J, 1994. Global Optimization Of Statistical Functions With Simulated Annealing J Econometrics 60 (1-2): 65-99

R. Hooke and T. A. Jeeves, "Direct Search Solution of Numerical and Statistical Problems", Journal of the ACM, Vol. 8, April 1961, pp. 212-229.

Johnson, M. G. 1994. Nonlinear Optimization using the algorithm of Hooke and Jeeves

Kennedy and Gentle, 1980. Statistical computing.

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.

E.3 Guide to a Simple Gadget Example

James Begley,
Marine Research Institute, Iceland

February, 2003

1 Introduction

This is a guide to the Gadget example that is currently available with the Gadget distribution. This example is intended to show how the basics of how Gadget works, and as such some elements of the model have been skipped for simplicity. The stock used in this example is Haddock in Icelandic waters, and the data files for this example can be downloaded from the Hafro ftp site, and extracts from these data files are included in this guide.

It is recommended that this guide is read in conjunction with the Gadget User Guide, which is also available from the Hafro ftp site. As explained in the user guide, comments in the data files are denoted by a semi-colon “;”, and parameters that can be optimised by Gadget are denoted by a hash “#”.

2 Main file

The main input file is called “main”. This file contains links to other files which make up the Gadget model. There is no data declared in this file, only links to other files.

```
;
; Main file for the haddock example
;
timefile    time        ; specifying years 1978-2006
areafile    area        ; specifying only one area
printfile   printfile   ; specifying the printer classes
;
[stock]          ; the description of the stock data
stockfiles       had
;
[tagging]        ; the description of the tagging data
; currently no tagging experiments in the haddock example
;
[otherfood]      ; the description of the food data
; currently no otherfood in the haddock example
;
[fleet]          ; the description of the fleet data
fleetfiles       fleet
;
[likelihood]     ; the description of the likelihood data
likelihoodfiles  likelihood
```

The first 2 lines of the main file for this haddock example lists the files that define the timesteps and are to be used for the example. The next line lists the file that is to be used to specify the printing that is required - that is output from the modelled data, not output from the model parameters.

In the [stock] section, there is one file listed which describes the haddock stock to be used for this example. There are then sections for the tagging experiments and the otherfood, which are blank for this example. The [fleet] section lists the file that is required to define the fleets for this example. Finally there is the [likelihood] section which defines the likelihood components for this example.

3 Time file

The time file defines the timesteps to be used by Gadget for this example. In this case, Gadget is to run a model from 1978 through to 2006, with 4 equal timesteps in each year.

```
;
; Time file for the haddock example
;
firstyear       1978
firststep       1
lastyear        2006
laststep        4
notimesteps     4 3 3 3 3
```

4 Area file

The area file defines the area data to be used by Gadget for this example. In this example, a single area has been defined, with a constant temperature for the duration of the model.

```
;
; Area file for the haddock example
;
areas  1
size   200000
temperature
;
; year  step  area  temperature
1978    1     1     5
1978    2     1     5
1978    3     1     5
1978    4     1     5
1979    1     1     5
1979    2     1     5
1979    3     1     5
1979    4     1     5
1980    1     1     5
1980    2     1     5
1980    3     1     5
1980    4     1     5
...
2005    3     1     5
2005    4     1     5
2006    1     1     5
2006    2     1     5
2006    3     1     5
2006    4     1     5
```

5 Aggregation files

There are a number of simple aggregation files that are required for this example.

5.1 Age aggregation files

There are 2 age aggregation files - one that lists the possible ages individually (ie. no aggregation) and one that groups all the ages together into one age group. These age aggregation files also define the text labels that are to be used when inputting and outputting the data for this example.

```
;
; Age aggregation file - no aggreagation
;
age1    1
age2    2
age3    3
age4    4
age5    5
age6    6
age7    7
age8    8
age9    9
age10   10
```

```
;
; Age aggregation file - all ages aggregated together
;
allages    1  2  3  4  5  6  7  8  9  10
```

5.2 Area aggregation file

Although there is only one area in this example, it is still necessary to define a area aggregation file. This is because it defines a text label that is to be used when inputting and outputting the data for this example.

```
;
; Area aggregation file - only one area
;
allareas   1
```

5.3 Length aggregation files

There are a total of 8 length aggregation files for this example. One aggregation file aggregates the stock into 2cm length groups (by combining the 1cm length groups that are declared for the stock), and a second file aggregates all the length groups together into one length group. There are also 6 aggregation files corresponding to the 6 survey index likelihood components that are declared in the likelihood file, which aggregate the stock into one or more 5cm length groups.

```
;
; Length aggregation file - 2cm length groups
;
len1     4.5     6.5
```

```
len2     6.5     8.5
len3     8.5     10.5
len4     10.5    12.5
len5     12.5    14.5
len6     14.5    16.5
len7     16.5    18.5
len8     18.5    20.5
len9     20.5    22.5
len10    22.5    24.5
...
len35    72.5    74.5
len36    74.5    76.5
len37    76.5    78.5
len38    78.5    80.5
len39    80.5    82.5
len40    82.5    84.5
len41    84.5    86.5
len42    86.5    88.5
len43    88.5    90.5


;
; Length aggregation file - all lengths aggregated together
;
all      4.5     90.5


;
; Length aggregation file for survey indices
;
mlen10   7.5     12.5


;
; Length aggregation file for survey indices
;
mlen15   12.5    17.5


;
; Length aggregation file for survey indices
;
mlen20   17.5    22.5


;
; Length aggregation file for survey indices
;
mlen25   22.5    27.5
mlen30   27.5    32.5
mlen35   32.5    37.5
mlen40   37.5    42.5
mlen45   42.5    47.5


;
; Length aggregation file for survey indices
;
mlen50   47.5    52.5
mlen55   52.5    57.5
mlen60   57.5    62.5


;
```

```
; Length aggregation file for survey indices
;
mlen65   62.5    67.5
mlen70   67.5    72.5
mlen75   72.5    77.5
```

6 Stock file

The stock file contains the various parameters that define the stock to be used in the Gadget model. The first section of this file gives the minimum and maximum age and length of the stock and the location of a reference weight file that specifies a reference weight for each length group for the stock.

The next section of this file cover the parameters required for the growth of the stock. The growth function used in this example is an expanded form of the Von Bertalanffy growth function, split so that the increase in weight is calculated first, and then the change in weight is used to calculate a change in length, as shown in equation 1 and equation 4 below:

$$\Delta W_i = \Delta t q_0 e^{q_1 T} \left(\left(\frac{W_i}{q_2} \right)^{q_4} - \left(\frac{W_i}{q_3} \right)^{q_5} \right) \tag{1}$$

$$r = \frac{W - (p_0 + p_8 (p_1 + p_2 p_8)) W_{ref}}{W} \tag{2}$$

$$f(x) = \begin{cases} 0 & \text{if } p_3 + p_4 x \leq 0 \\ p_5 & \text{if } p_3 + p_4 x \geq p_5 \\ p_3 + p_4 x & \text{otherwise} \end{cases} \tag{3}$$

$$\Delta L_i = \frac{\Delta W_i}{p_6 p_7 l^{p_7 - 1}} f(r) \tag{4}$$

where:
<Δt> is the length of the timestep
<T> is the temperature
<W_{ref}> is the reference weight

For this example, q_1 is set to zero, removing the temperature dependance from the equation, and q_2 is set equal to q_3, further simplifying the equation. Equations 2 and 3 introduce the concept of starvation to the Gadget model, by using a function of the weight and the reference weight when calculating the length increase due to the growth. For this example, p_0 is set to one and p_8 to zero, which considerably simplifies equation 2.

The next section of the stock file defines whether the stock acts as a prey or a predator, and specifies the initial conditions, which are used to calculate the stock that exists at the beginning of the first timestep. There then follows sections used to describe how the stock would migrate, mature, move, recruit and spawn, which are mostly unused for the example. However, the stock does have recruits to ensure that it doesn't die out, which are define in the recruitment file.

```
;
; Haddock stock file for the haddock example
;
```

```
stockname       had
livesonareas    1
minage          1
maxage          10
minlength       4.5
maxlength       90.5
dl              1
refweightfile   had.refweights
;
; the growth and consumption can be calculated on a finer scale
; so the length groups for this are specified here
;
growthandeatlengths   len.agg
;
; information about the growth of the stock
;
doesgrow              1
growthfunctionnumber  3
;
; the parameters required for this growth function are
;
wgrowthparameters     1#grq0  0  1#grq2  1#grq2  0.666  1
lgrowthparameters     1  0  0  1  2.2  1.4  8.85e-6  3.0257  0
;
power                 3
maxlengthgroupgrowth  20
beta                  1#bbeta
;
; information about the natural mortality of the stock
; age             1   2    3   4   5   6   7   8   9   10
naturalmortality  0.5 0.35 0.2 0.2 0.2 0.2 0.2 0.3 0.4 0.7
;
; information about the stock acting as a prey
;
iseaten           1 ; the fleet is considered as predator
preylengths       len.agg
;
; information about the stock acting as a predator
;
doeseat           0
;
; information about the initial conditions of the stock
;
initialconditions
numbers
minage            1
maxage            10
minlength         4.5
maxlength         90.5
dl                1
;
; age       1 2            3             4             5             ...
agemultiple 0 10000*1#age2 6065.3*1#age3 3678.8*1#age4 2231.3*1#age5 ...
;
initstockfile     had.init
;
; information about the migration of the stock
;
doesmigrate       0
;
; information about the maturation of the stock
```

```
;
doesmature        0
;
; information about the movement of the stock
;
doesmove          0
;
; information about the recruitment to the stock
;
doesrenew         1
renewaldatafile   had.rec
;
; information about the spawning of the stock
;
doesspawn         0
```

6.1 Reference weight file

The reference weight file gives a reference length-weight relationship for the stock in this example. This is used to generate the entries in the age-length cells for the initial conditions, and to modify the growth of the stock after starvation (so that the growth results in an increase in weight not length for the underweight fish). This files simply lists a "reference" mean weight for each length group for the stock.

```
;
; Reference length-weight relationship for the haddock example
;
;length weight
5       0.001307
6       0.002154
7       0.003285
8       0.004735
9       0.006538
10      0.008725
11      0.011328
12      0.013848
13      0.017795
14      0.021701
15      0.025792
...
83      5.088293
84      5.261409
85      5.438262
86      5.618887
87      5.803317
88      5.99159
89      6.183739
90      6.3798
```

6.2 Initial conditions file

The initial conditions file gives a Normal distribution for each area/age group combination. This will be used by Gadget to construct an initial population of 10,000 fish in each area/age group, with the length groups specified by a mean length and standard deviation. The mean weight for the initial population is calculated by multiplying the reference weight by the condition factor.

To change from a population with 10,000 fish in each area/age group to the initial population used in the model, each age group is multiplied by an area weighting factor (“distribution”), and this result is then multiplied by the age multiple (which is specified in the main stock file).

```
;
; Initial stock file for the haddock example
;
; age  area  distribution  meanlength  standdev  condition
1      1     100           16.41203    2.247188  1
2      1     100           27.15520    2.898219  1
3      1     100           36.98713    4.070510  1
4      1     100           43.77545    4.927558  1
5      1     100           49.43773    5.540416  1
6      1     100           53.76334    5.807182  1
7      1     100           58.64396    6.023261  1
8      1     100           66.10526    8.00      1
9      1     100           60.88235    9.00      1
10     1     100           63.00       9.00      1
```

6.3 Recruitment file

The recruitment file defines the number of the recruits that are to be added to the stock, along with information about the age, length and weight of these recruits. The number of these recruits is given, for each timestep/area combination, in units of 10,000 fish. The age is specified as minimum age of the stock.

These recruits are defined as a simple length based stock, with a Normal distribution around a mean length and standard deviation of the length given in the input file. The mean weight of the recruits is then calculated from the standard weight-length relationship, given in equation 5 below:

$$W = \alpha L^{\beta} \tag{5}$$

Note that in this example, the recruits are assumed to have the same weight and length distribution in each year. The number of recuits for years that there is data available (1978 - 1999) are parameters that the optimiser can adjust to try to get a better fit between the modelled data and the input data, where as for future years when there is no data available it is assumed that there is a constant number of recruits.

```
;
; Recruitment data for the haddock example
;
Normaldistribution
minlength 4.5
maxlength 28.5
dl        1
;
; year  step  area  age  number        meanlen  stdev  alpha    beta
1978    1     1     1    1000*1#rec78  16.41    2.25   8.85e-6  3.0257
1979    1     1     1    1000*1#rec79  16.41    2.25   8.85e-6  3.0257
1980    1     1     1    1000*1#rec80  16.41    2.25   8.85e-6  3.0257
1981    1     1     1    1000*1#rec81  16.41    2.25   8.85e-6  3.0257
1982    1     1     1    1000*1#rec82  16.41    2.25   8.85e-6  3.0257
1983    1     1     1    1000*1#rec83  16.41    2.25   8.85e-6  3.0257
1984    1     1     1    1000*1#rec84  16.41    2.25   8.85e-6  3.0257
1985    1     1     1    1000*1#rec85  16.41    2.25   8.85e-6  3.0257
1986    1     1     1    1000*1#rec86  16.41    2.25   8.85e-6  3.0257
1987    1     1     1    1000*1#rec87  16.41    2.25   8.85e-6  3.0257
```

```
1988    1     1     1    1000*1#rec88  16.41    2.25   8.85e-6  3.0257
1989    1     1     1    1000*1#rec89  16.41    2.25   8.85e-6  3.0257
1990    1     1     1    1000*1#rec90  16.41    2.25   8.85e-6  3.0257
1991    1     1     1    1000*1#rec91  16.41    2.25   8.85e-6  3.0257
1992    1     1     1    1000*1#rec92  16.41    2.25   8.85e-6  3.0257
1993    1     1     1    1000*1#rec93  16.41    2.25   8.85e-6  3.0257
1994    1     1     1    1000*1#rec94  16.41    2.25   8.85e-6  3.0257
1995    1     1     1    1000*1#rec95  16.41    2.25   8.85e-6  3.0257
1996    1     1     1    1000*1#rec96  16.41    2.25   8.85e-6  3.0257
1997    1     1     1    1000*1#rec97  16.41    2.25   8.85e-6  3.0257
1998    1     1     1    1000*1#rec98  16.41    2.25   8.85e-6  3.0257
1999    1     1     1    1000*1#rec99  16.41    2.25   8.85e-6  3.0257
2000    1     1     1    9000          16.41    2.25   8.85e-6  3.0257
2001    1     1     1    9000          16.41    2.25   8.85e-6  3.0257
2002    1     1     1    9000          16.41    2.25   8.85e-6  3.0257
2003    1     1     1    9000          16.41    2.25   8.85e-6  3.0257
2004    1     1     1    9000          16.41    2.25   8.85e-6  3.0257
2005    1     1     1    9000          16.41    2.25   8.85e-6  3.0257
2006    1     1     1    9000          16.41    2.25   8.85e-6  3.0257
```

7 Fleet file

The fleet file defines the fleets that are present in the Gadget model. The various fleets are defined by specifying the fleet type, name, area and length groups (which in this example are set to the minimum and maximum lengths of the stock). The fleets also have a suitability function, that describes how likely it is that the fleet will catch fish of a given length. The suitability function used is an exponential suitability function, given in equation 6 below:

$$S(l,L) = \frac{\delta}{1 + e^{-\alpha-\beta l-\gamma L}} \tag{6}$$

where:
<l> is the length of the prey
<L> is the length of the predator

Note that in this example, <γ> is set to 0 which removes any dependance on the length of the predator, and <δ> is always set to 1. <α> and <β> are parameters that the optimiser can adjust to try to get a better fit between the modelled data and the input data.

There are 3 fleets defined in this example. The commercial fleet (”comm”) covers all the commercial fishing activity, and all the available landings data is specified in the data file. The survey fleet (“survey”) covers all the government survey activity, and this fleet is assumed to land a constant amount of fish for all the years in the model. The third fleet (“future”) covers all the predicted commercial fishing activity from mid 1999 (when the commercial landing data stops being available) to the end of the models timesteps.

```
;
; Fleet data for the haddock example
;
; Details for the commercial fleet
;
[fleetcomponent]
totalfleet      comm            ; fleet name
livesonareas    1               ; areas for the fleet
```

```
; In this example, the length information for the fleets is
; meaningless so set them to min/max length for the stock
lengths         4.5 90.5        ; lengths for this fleet
multiplicative  1.0             ; scaling factor
suitability                     ; alpha    beta     gamma  delta
had    function ExpsuitfuncA    1#acomm    1#bcomm  0      1
amount          fleet.data      ; where the catch data is stored
;
; Details for the survey fleet
;
[fleetcomponent]
totalfleet      survey          ; fleet name
livesonareas    1               ; areas for the fleet
lengths         4.5 90.5        ; lengths for this fleet
multiplicative  1.0             ; scaling factor
suitability
had    function ExpsuitfuncA    1#asur    1#bsur    0      1
amount          fleet.data      ; where the catch data is stored
;
; Details for the predicted fleet
;
[fleetcomponent]
linearfleet     future          ; fleet name
livesonareas    1               ; areas for the fleet
lengths         4.5 90.5        ; lengths for this fleet
multiplicative  0.5             ; scaling factor
suitability
had    function ExpsuitfuncA    1#acomm    1#bcomm  0      1
amount          fleet.predict   ; where the catch data is stored
```

7.1 Fleet data files

The 2 fleet data files contain details of the landings made in each timestep/area/fleet combination for the fleets that have been declared in the main fleet file. The first data file is a list of the total weight of the landing data currently available (ie. all the survey data and the commercial landings data up to the first timestep of 1999) for each timestep/area/fleet combination. The second data file contains a list of the ratios to be used when calculating the amount that the fleet will catch, for the timestep/area combinations when commercial fleet effort is required in the future and no landings data is available (ie. from the second timestep of 1999).

```
;
; Fleet catch data in kilos for the haddock example
;
; year  step  area  fleet   amount
1978    1     1     comm    8444000
1978    2     1     comm    14834000
1978    3     1     comm    9985000
1978    4     1     comm    10184000
1979    1     1     comm    10753000
1979    2     1     comm    18893000
1979    3     1     comm    12717000
1979    4     1     comm    12971000
1980    1     1     comm    9933000
1980    2     1     comm    17451000
1980    3     1     comm    11747000
1980    4     1     comm    11981000
1981    1     1     comm    12352000
1981    2     1     comm    21701000
```

```
1981    3     1     comm    14608000
1981    4     1     comm    14899000
...
1997    2     1     survey  30000
1998    2     1     survey  30000
1999    2     1     survey  30000
2000    2     1     survey  30000
2001    2     1     survey  30000
2002    2     1     survey  30000
2003    2     1     survey  30000
2004    2     1     survey  30000
2005    2     1     survey  30000
2006    2     1     survey  30000

;
; Predicted fleet effort for the haddock example
;
; year  step  area  fleet   ratio
1999    2     1     future  1
1999    3     1     future  1
1999    4     1     future  1
2000    1     1     future  1
2000    2     1     future  1
2000    3     1     future  1
2000    4     1     future  1
...
2005    1     1     future  1
2005    2     1     future  1
2005    3     1     future  1
2005    4     1     future  1
2006    1     1     future  1
2006    2     1     future  1
2006    3     1     future  1
2006    4     1     future  1
```

8 Likelihood file

The likelihood file defines the various likelihood components that will be used to compare the data from within the Gadget model with external data. Each likelihood component calculates a “goodness of fit” between the 2 sets of data to give a a likelihood score, and there is then a weighted sum of these likelihood scores to give an overall likelihood score, which can be minimised if Gadget is performing an optimising run.

In this example, there are a total of 14 likelihood components defined to test the goodness of fit between the 2 sets of data. These components are ”BoundLikelihood”, “Understocking”, 2 “CatchStatistics”, 4 “CatchDistribution” and 6 “SurveyIndices” components.

BoundLikelihood
The BoundLikelihood component is used to apply a penalty weight to any parameter that goes outside the bounds during the optimsing process. Applying this penalty weight will force the parameter away from the bounds and back into the range of numbers that have been specified in the parameter file. In this example the penalty weights are specified in the file “penaltyfile”.

Understocking
The Understocking component is used to apply a penalty whenever there has been overconsumption by

the fleets, and there is insufficient stock for a fleet to catch. In this example this penalty is the sum of squares of the understocking that has occured in the model.

CatchStatistics
The CatchStatistics components are used to compare biological data sampled from the model with that sampled from landings data for the fleets. In this example there are 2 comparisons , one for data from the commercial fleet and one for the survey fleet. In each case a weighted sum of squares of the mean length at age is used to calculate the goodness of fit between the 2 sets of data. The data which will be compared to the results from within the Gadget model are given in the 2 data files that are specified.

CatchDistribution
The CatchDistribution components are used to compare distribution data sampled from the model with that sampled from landings data for the fleets. In this example there are 2 comparisons (one for the commercial fleet, one for the survey fleet) with the data aggregated into length groups and a further 2 comparisons with the data aggregated into age-length groups. In each case a multinomial function is used to calculate the goodness of fit between the 2 sets of data. The data which will be compared to the results from within the Gadget model are given in the 4 data files that are specified.

SurveyIndices
The SurveyIndices components are used to compare stock indices calculated within the Gadget model to indices calculated from a standardized survey for that stock. In this example the survey indices calculaed are based on a length group survey, and there are 6 comparisons for 6 different length group aggregations. The survey indices for the length groups are given in the data file “had.surveyindex”, and the length group aggregations are defined in the various length aggregation files.

```
;
; Likelihood file for the haddock example
;
; first specify the likelihood bounds
[component]
name               bounds            ; likelihood component name
weight             10                ; weight for component
type               Penalty           ; type of component
datafile           penaltyfile       ; data file for this component
;
; now specify when to check for understocking
;
[component]
name               understocking     ; likelihood component name
weight             1e-12             ; weight for component
type               Understocking     ; type of component
areaaggfile        allarea.agg       ; area aggregation file
fleetnames         comm              ; source of fleet data
yearsandsteps      all all           ; timesteps to check
;
; the mean length statistics data
;
[component]
name               meanl.sur         ; likelihood component name
weight             2e-6              ; weight for component
type               CatchStatistics   ; type of component
datafile           had.meanle.sur    ; data file for this component
functionnumber     2                 ; function type
overconsumption    1                 ; flag for overconsumption
areaaggfile        allarea.agg       ; area aggregation file
ageaggfile         age.agg           ; age aggreagation file
```

```
fleetnames         survey            ; source of fleet data
stocknames         had               ; source of stock data
;
[component]
name               meanl.catch       ; likelihood component name
weight             0.3e-6            ; weight for component
type               CatchStatistics   ; type of component
datafile           had.meanle.catch  ; data file for this component
functionnumber     2                 ; function type
overconsumption    1                 ; flag for overconsumption
areaaggfile        allarea.agg       ; area aggregation file
ageaggfile         age.agg           ; age aggreagation file
fleetnames         comm              ; source of fleet data
stocknames         had               ; source of stock data
;
; the length distribution data
;
[component]
name               ldist.sur         ; likelihood component name
weight             0.05e-6           ; weight for component
type               CatchDistribution ; type of component
datafile           had.ldist.sur     ; data file for this component
functionnumber     1                 ; function type
overconsumption    1                 ; flag for overconsumption
minimumprobability 20                ; used if outcome is improbable
areaaggfile        allarea.agg       ; area aggregation file
ageaggfile         allage.agg        ; age aggreagation file
lenaggfile         len.agg           ; length aggregation file
fleetnames         survey            ; source of fleet data
stocknames         had               ; source of stock data
;
[component]
name               ldist.catch       ; likelihood component name
weight             3e-6              ; weight for component
type               CatchDistribution ; type of component
datafile           had.ldist.catch   ; data file for this component
functionnumber     1                 ; function type
overconsumption    1                 ; flag for overconsumption
minimumprobability 20                ; used if outcome is improbable
areaaggfile        allarea.agg       ; area aggregation file
ageaggfile         allage.agg        ; age aggreagation file
lenaggfile         len.agg           ; length aggregation file
fleetnames         comm              ; source of fleet data
stocknames         had               ; source of stock data
;
; the age-length distribution data
;
[component]
name               alkeys.sur        ; likelihood component name
weight             7e-6              ; weight for component
type               CatchDistribution ; type of component
datafile           had.alkeys.sur    ; data file for this component
functionnumber     1                 ; function type
overconsumption    1                 ; flag for overconsumption
minimumprobability 20                ; used if outcome is improbable
areaaggfile        allarea.agg       ; area aggregation file
ageaggfile         age.agg           ; age aggreagation file
lenaggfile         len.agg           ; length aggregation file
fleetnames         survey            ; source of fleet data
stocknames         had               ; source of stock data
;
```

```
[component]
name               alkeys.catch      ; likelihood component name
weight             2e-6              ; weight for component
type               CatchDistribution ; type of component
datafile           had.alkeys.catch  ; data file for this component
functionnumber     1                 ; function type
overconsumption    1                 ; flag for overconsumption
minimumprobability 20                ; used if outcome is improbable
areaaggfile        allarea.agg       ; area aggregation file
ageaggfile         age.agg           ; age aggreagation file
lenaggfile         len.agg           ; length aggregation file
fleetnames         comm              ; source of fleet data
stocknames         had               ; source of stock data
;
; the survey index data
;
[component]
name          si10                   ; likelihood component name
weight        70e-4                  ; weight for component
type          SurveyIndices          ; type of component
datafile      had.surveyindex        ; data file for this component
sitype        lengths                ; survey index type
areaaggfile   allarea.agg            ; area aggregation file
lenaggfile    si10len.agg            ; length aggregation file
stocknames    had                    ; source of stock data
fittype       FixedSlopeLogLinearFit ; type of data fit
slope         1                      ; slope is fixed
;
[component]
name          si15                   ; likelihood component name
weight        100e-4                 ; weight for component
type          SurveyIndices          ; type of component
datafile      had.surveyindex        ; data file for this component
sitype        lengths                ; survey index type
areaaggfile   allarea.agg            ; area aggregation file
lenaggfile    si15len.agg            ; length aggregation file
stocknames    had                    ; source of stock data
fittype       FixedSlopeLogLinearFit ; type of data fit
slope         1                      ; slope is fixed
;
[component]
name          si20                   ; likelihood component name
weight        100e-4                 ; weight for component
type          SurveyIndices          ; type of component
datafile      had.surveyindex        ; data file for this component
sitype        lengths                ; survey index type
areaaggfile   allarea.agg            ; area aggregation file
lenaggfile    si20len.agg            ; length aggregation file
stocknames    had                    ; source of stock data
fittype       FixedSlopeLogLinearFit ; type of data fit
slope         1                      ; slope is fixed
;
[component]
name          si25to45               ; likelihood component name
weight        100e-4                 ; weight for component
type          SurveyIndices          ; type of component
datafile      had.surveyindex        ; data file for this component
sitype        lengths                ; survey index type
areaaggfile   allarea.agg            ; area aggregation file
lenaggfile    si2545len.agg          ; length aggregation file
stocknames    had                    ; source of stock data
```

```
fittype       FixedSlopeLogLinearFit ; type of data fit
slope         1                      ; slope is fixed
;
[component]
name          si50to60               ; likelihood component name
weight        100e-4                 ; weight for component
type          SurveyIndices          ; type of component
datafile      had.surveyindex        ; data file for this component
sitype        lengths                ; survey index type
areaaggfile   allarea.agg            ; area aggregation file
lenaggfile    si5060len.agg          ; length aggregation file
stocknames    had                    ; source of stock data
fittype       FixedSlopeLogLinearFit ; type of data fit
slope         1                      ; slope is fixed
;
[component]
name          si65to75               ; likelihood component name
weight        70e-4                  ; weight for component
type          SurveyIndices          ; type of component
datafile      had.surveyindex        ; data file for this component
sitype        lengths                ; survey index type
areaaggfile   allarea.agg            ; area aggregation file
lenaggfile    si6575len.agg          ; length aggregation file
stocknames    had                    ; source of stock data
fittype       FixedSlopeLogLinearFit ; type of data fit
slope         1                      ; slope is fixed
```

8.1 Penalty file

The penalty file contains the likelihood penalty that is to be applied when any of the parameters goes outside its bound, defined in the parameter input file. For this example, only a default setting is given which will be applied to each parameter that goes outside a bound. It is possible to assign different penalties to each parameter, by specifying the parameters individually, but that is not done in this example.

```
;
; Penalty file for the haddock example
; This file lists the penalties applied to the variables
; when the value reaches the bound specified in the inputfile
;
; switch   power  lowerW   upperW
default    2      10000    10000 ; default setting
```

8.2 Mean length files

The 2 mean length data files contain the number of samples, and the mean length and standard deviation of the length for these samples, in each timestep/area/age combination for the 2 fleets. For this likelihood component there is no area or age aggregation, as defined by the aggregation files declared in the main likelihood file.

The likelihood function that is used to compare the data from these files with the corresponding data from the model is a weighted sum of squares of the mean length, given in equation 7 below:

$$\ell = \sum_{time} \sum_{areas} \sum_{ages} \left(\frac{(x-\mu)^2}{s^2} N \right) \tag{7}$$

where:
<x> is the sample mean length from the data
<μ> is the mean length calculated from the model
<s> is the standard deviation of the length from the data
<N> is the sample size

```
;
; Mean length data for the survey catch for the haddock example
;
; year   step  area      age    number  mean   variance
1989     2     allareas  age1   60      15.35  2.0
1989     2     allareas  age2   124     28.01  3.7
1989     2     allareas  age3   238     36.14  6.2
1989     2     allareas  age4   829     45.64  5.0
1989     2     allareas  age5   336     54.41  5.5
1989     2     allareas  age6   106     63.42  6.3
1989     2     allareas  age7   8       67.88  7.1
1989     2     allareas  age8   8       72.88  6.6
1989     2     allareas  age9   4       70.75  5.8
1989     2     allareas  age10  3       75     5.6
1990     2     allareas  age1   235     15.63  2.0
1990     2     allareas  age2   227     27.51  3.7
1990     2     allareas  age3   192     37.13  6.2
1990     2     allareas  age4   267     44.58  5.0
1990     2     allareas  age5   620     51.01  5.5
1990     2     allareas  age6   299     59.47  6.3
1990     2     allareas  age7   38      67.29  7.1
1990     2     allareas  age8   9       72.11  6.6
1990     2     allareas  age9   2       71     5.8
1990     2     allareas  age10  0       0      5.6
...
1999     2     allareas  age1   310     14.67  2.0
1999     2     allareas  age2   312     27.63  3.7
1999     2     allareas  age3   148     36.2   6.2
1999     2     allareas  age4   174     44.36  5.0
1999     2     allareas  age5   24      50.21  5.5
1999     2     allareas  age6   15      55.53  6.3
1999     2     allareas  age7   4       63.25  7.1
1999     2     allareas  age8   6       60.33  6.6
1999     2     allareas  age9   1       69     5.8
1999     2     allareas  age10  0       0      5.6
```

```
;
; Mean length data for the commercial catch for the haddock example
;
; year   step  area      age    number  mean   variance
1979     1     allareas  age1   0       0      2.0
1979     1     allareas  age2   0       0      3.7
1979     1     allareas  age3   0       0      6.2
1979     1     allareas  age4   7       55.43  5.0
1979     1     allareas  age5   48      58.92  5.5
1979     1     allareas  age6   96      64.2   6.3
1979     1     allareas  age7   36      68.61  7.1
1979     1     allareas  age8   4       73.5   6.6
1979     1     allareas  age9   1       73     5.8
1979     1     allareas  age10  0       0      5.6
1979     2     allareas  age1   0       0      2.0
1979     2     allareas  age2   0       0      3.7
1979     2     allareas  age3   13      42.15  6.2
1979     2     allareas  age4   73      51.44  5.0
```

```
1979     2     allareas  age5   224     58.56  5.5
1979     2     allareas  age6   258     62.16  6.3
1979     2     allareas  age7   20      69.6   7.1
1979     2     allareas  age8   8       70.5   6.6
1979     2     allareas  age9   1       72     5.8
1979     2     allareas  age10  0       0      5.6
...
1998     3     allareas  age1   0       0      2.0
1998     3     allareas  age2   0       0      3.7
1998     3     allareas  age3   17      45     6.2
1998     3     allareas  age4   33      49.42  5.0
1998     3     allareas  age5   78      56.26  5.5
1998     3     allareas  age6   19      58.26  6.3
1998     3     allareas  age7   16      63.31  7.1
1998     3     allareas  age8   28      66.04  6.6
1998     3     allareas  age9   3       74.67  5.8
1998     3     allareas  age10  0       0      5.6
```

8.3 Length distribution files

The 2 length distribution data files contain the number of samples in each timestep/ area/age/length group combination for the 2 fleets. For this likelihood component, there is no area aggregation, all the age groups have been aggregated together into one age group, and the length groups have been aggregated into 2cm length groups, as defined by the aggregation files declared in the main likelihood file.

The likelihood function that is used to compare the data from these files with the corresponding data from the model is a multinomial function, given in equation 8 below:

$$\ell = 2 \sum_{time} \sum_{areas} \sum_{age} \left(\log N_{tra}! - \sum_{length} \log N_{tral}! + \sum_{length} \left(N_{tral} \log \frac{\pi_{tral}}{\sum \pi_{tral}} \right) \right) \qquad (8)$$

where:
<π> is the model sample size for that time/area/age/length combination
<N> is the data sample size for that time/area/age/length combination

```
;
; Length distribution of the survey data for the haddock example
;
; year   step  area       age      length   number
1985     2     allareas   allages  len1     0
1985     2     allareas   allages  len2     0
1985     2     allareas   allages  len3     4
1985     2     allareas   allages  len4     112
1985     2     allareas   allages  len5     1718
1985     2     allareas   allages  len6     4419
1985     2     allareas   allages  len7     3503
1985     2     allareas   allages  len8     1408
1985     2     allareas   allages  len9     403
1985     2     allareas   allages  len10    216
1985     2     allareas   allages  len11    880
1985     2     allareas   allages  len12    2423
1985     2     allareas   allages  len13    4311
1985     2     allareas   allages  len14    4699
1985     2     allareas   allages  len15    2663
```

```
1985     2     allareas   allages  len16    1204
1985     2     allareas   allages  len17    683
1985     2     allareas   allages  len18    962
1985     2     allareas   allages  len19    1588
1985     2     allareas   allages  len20    1935
1985     2     allareas   allages  len21    2508
1985     2     allareas   allages  len22    2222
...
1999     2     allareas   allages  len37    63
1999     2     allareas   allages  len38    39
1999     2     allareas   allages  len39    11
1999     2     allareas   allages  len40    16
1999     2     allareas   allages  len41    1
1999     2     allareas   allages  len42    1
1999     2     allareas   allages  len43    1
```

```
;
; Length distribution of the commercial data for the haddock example
;
; year   step  area       age      length   number
1979     1     allareas   allages  len1     0
1979     1     allareas   allages  len2     0
1979     1     allareas   allages  len3     0
1979     1     allareas   allages  len4     0
1979     1     allareas   allages  len5     0
1979     1     allareas   allages  len6     0
1979     1     allareas   allages  len7     0
1979     1     allareas   allages  len8     0
1979     1     allareas   allages  len9     0
1979     1     allareas   allages  len10    0
1979     1     allareas   allages  len11    0
1979     1     allareas   allages  len12    0
1979     1     allareas   allages  len13    0
1979     1     allareas   allages  len14    0
1979     1     allareas   allages  len15    1
1979     1     allareas   allages  len16    3
1979     1     allareas   allages  len17    11
1979     1     allareas   allages  len18    17
1979     1     allareas   allages  len19    32
1979     1     allareas   allages  len20    53
1979     1     allareas   allages  len21    97
1979     1     allareas   allages  len22    145
1979     1     allareas   allages  len23    248
1979     1     allareas   allages  len24    334
1979     1     allareas   allages  len25    375
1979     1     allareas   allages  len26    433
1979     1     allareas   allages  len27    444
1979     1     allareas   allages  len28    486
...
1999     1     allareas   allages  len37    22
1999     1     allareas   allages  len38    14
1999     1     allareas   allages  len39    12
1999     1     allareas   allages  len40    5
1999     1     allareas   allages  len41    0
1999     1     allareas   allages  len42    0
1999     1     allareas   allages  len43    1
```

8.4 Age-length distribution files

The 2 age-length distribution data files contain the number of samples in each timestep/area/age/length group combination for the 2 fleets. For this likelihood component, there is no area or age group aggregation, and the length groups have been aggregated into 2cm length groups, as defined by the aggregation files declared in the main likelihood file.

The only difference between the age-length distribution data files and the length distribution data files is the age aggregation that takes place for the length distribution. Note that by summing up all 10 entries for one timestep/length group in these files, it is possible to recreate the entries in the length distribution files.

The likelihood function that is used to compare the data from these files with the corresponding data from the model is a multinomial function, given in equation 8 above.

```
;
; Age-length distribution of the survey data for the haddock example
;
; year   step  area       age    length   number
1985     2     allareas   age1   len1     0
1985     2     allareas   age2   len1     0
1985     2     allareas   age3   len1     0
1985     2     allareas   age4   len1     0
1985     2     allareas   age5   len1     0
1985     2     allareas   age6   len1     0
1985     2     allareas   age7   len1     0
1985     2     allareas   age8   len1     0
1985     2     allareas   age9   len1     0
1985     2     allareas   age10  len1     0
1985     2     allareas   age1   len2     0
1985     2     allareas   age2   len2     0
1985     2     allareas   age3   len2     0
1985     2     allareas   age4   len2     0
1985     2     allareas   age5   len2     0
1985     2     allareas   age6   len2     0
1985     2     allareas   age7   len2     0
1985     2     allareas   age8   len2     0
1985     2     allareas   age9   len2     0
1985     2     allareas   age10  len2     0
1985     2     allareas   age1   len3     1
1985     2     allareas   age2   len3     0
1985     2     allareas   age3   len3     0
1985     2     allareas   age4   len3     0
1985     2     allareas   age5   len3     0
1985     2     allareas   age6   len3     0
1985     2     allareas   age7   len3     0
1985     2     allareas   age8   len3     0
1985     2     allareas   age9   len3     0
1985     2     allareas   age10  len3     0
1985     2     allareas   age1   len4     10
1985     2     allareas   age2   len4     0
...
1999     2     allareas   age6   len43    0
1999     2     allareas   age7   len43    0
1999     2     allareas   age8   len43    0
1999     2     allareas   age9   len43    0
1999     2     allareas   age10  len43    0
```

```
;
; Age-length distribution of the commercial data for the haddock example
;
; year   step  area       age    length   number
1979     1     allareas   age1   len1     0
1979     1     allareas   age2   len1     0
1979     1     allareas   age3   len1     0
1979     1     allareas   age4   len1     0
1979     1     allareas   age5   len1     0
1979     1     allareas   age6   len1     0
1979     1     allareas   age7   len1     0
1979     1     allareas   age8   len1     0
1979     1     allareas   age9   len1     0
1979     1     allareas   age10  len1     0
1979     1     allareas   age1   len2     0
1979     1     allareas   age2   len2     0
1979     1     allareas   age3   len2     0
1979     1     allareas   age4   len2     0
...
1998     3     allareas   age4   len43    0
1998     3     allareas   age5   len43    0
1998     3     allareas   age6   len43    0
1998     3     allareas   age7   len43    0
1998     3     allareas   age8   len43    0
1998     3     allareas   age9   len43    0
1998     3     allareas   age10  len43    0
```

8.5 Survey index file

The survey index data file contains the number of samples in each timestep/area/ length group combination for the 6 survey indices defined in the main likelihood file. For this likelihood component, there is no area or age group aggregation, and the length groups have been aggregated into 5cm length groups, as defined by the various length aggregation files declared in the main likelihood file.

The likelihood function that is used to compare the data from these files with the corresponding data from the model is a log linear regression function. For the regression line (specified in the main likelihood file), the slope is fixed and the intercept calculated by Gadget. This is given in equation 9 below:

$$\ell = \sum_{time} \Big(I_t - (\alpha + \beta N_t) \Big)^2 \tag{9}$$

where:
<I> is the survey index
<N> is the corresponding index calculated in the Gadget model
<α> is the intercept of the regression line
<β> is the slope of the regression line (which has been set to 1)

```
;
; Survey index data for the haddock example
;
; year  step  area      length  number
1985    1     allareas  mlen10  258
1986    1     allareas  mlen10  808
1987    1     allareas  mlen10  286
1988    1     allareas  mlen10  131
1989    1     allareas  mlen10  361
```

```
1990    1     allareas  mlen10  568
1991    1     allareas  mlen10  1163
1992    1     allareas  mlen10  1713
1993    1     allareas  mlen10  642
1994    1     allareas  mlen10  782
1995    1     allareas  mlen10  171
1996    1     allareas  mlen10  508
1997    1     allareas  mlen10  217
1998    1     allareas  mlen10  197
1999    1     allareas  mlen10  2690
1985    1     allareas  mlen15  20063
1986    1     allareas  mlen15  91563
1987    1     allareas  mlen15  20086
1988    1     allareas  mlen15  12051
1989    1     allareas  mlen15  8853
1990    1     allareas  mlen15  64132
...
1993    1     allareas  mlen75  948
1994    1     allareas  mlen75  533
1995    1     allareas  mlen75  371
1996    1     allareas  mlen75  553
1997    1     allareas  mlen75  318
1998    1     allareas  mlen75  505
1999    1     allareas  mlen75  674
```

9 Print file

The printfile defines the content of the output files that will be generated when a stochastic run of Gadget is performed (by specifying the “-s” command line option when Gadget is started). This output is defined by specifying details of the stock, area, age and length groups and the name of the output file that is to be generated.

In this example there are 4 output files to be generated. The first file (created by “StockStdPrinter”) contains an age-based summary of the stock, giving details of the number, length, mean weight and consumption for each timestep/area/age group combination. The next two files (created by “StockPrinter” and “StockFullPrinter”) give details of the number and mean weight for each timestep/area/age group/length group combination - the difference being the amount of aggregation that takes place. The final file (created by “PredatorPrinter”) contains information about the predator/prey combination (with the predator being the fleets) and gives details of the biomass consumed for each timestep/area/predator length group/prey length group combination.

```
;
; Print file for the haddock example
;
[component]
type            stockstdprinter
stockname       had              ; name of the stock
areaaggfile     allarea.agg      ; area aggregation file
printfile       out/had.std      ; name for the output file
yearsandsteps   all all          ; timesteps to print
;
[component]
type            stockprinter
stocknames      had              ; name of the stock (could be more than one)
areaaggfile     allarea.agg      ; area aggregation file
ageaggfile      allage.agg       ; age aggregation file
```

```
lenaggfile      len.agg          ; length aggregation file
printfile       out/had.print    ; name for the output file
yearsandsteps   all all          ; timesteps to print
;
[component]
type            stockfullprinter
stockname       had              ; name of the stock
areaaggfile     allarea.agg      ; area aggregation file
printfile       out/had.stock    ; name for the output file
yearsandsteps   all all          ; timesteps to print
;
[component]
type            predatorprinter
predators       comm future      ; name of the predators (fleets)
preys           had              ; name of the preys
areaaggfile     allarea.agg      ; area aggregation file
predlenaggfile  alllen.agg       ; length aggreagtion file for the predators
preylenaggfile  len.agg          ; length aggregation file for the preys
printfile       out/had.fleet    ; name for the ouuput file
yearsandsteps   all all          ; timesteps to print
```

9.1 StockStdPrinter output

The output file that is generated by the stockstdprinter printer class is given below. This class summarises the data available for the stock, giving the number, mean length, mean weight, length standard deviation, number consumed and biomass consumed for each timestep/area/age group combination.

```
; Gadget version 2.0.02 running on hafnasandur Mon Feb 10 14:47:43 2003
; Standard output file for the stock had
; year-step-area-age-number-mean length-mean weight-stddev length ...
1978    1    1    1 2.2377895e+08     16.41  0.0444454       2.25 ...
1978    1    1    2 1.0285317e+08   29.8986   0.239963    3.61989 ...
1978    1    1    3     13576603   39.4734   0.585367    4.57133 ...
1978    1    1    4     20759881   46.0429   0.940242    5.24777 ...
1978    1    1    5     46760341    51.507    1.31759     5.7698 ...
1978    1    1    6    1808507.5   55.7022    1.65875     6.0014 ...
1978    1    1    7    741870.25   60.4509      2.101    6.18222 ...
1978    1    1    8     11378927   67.6647    2.94619    7.95431 ...
1978    1    1    9     25158941   62.5636    2.38451    8.94957 ...
1978    1    1   10    771146.28   64.6152    2.61065     8.9165 ...
1978    2    1    1 1.9745062e+08   20.9104  0.0699871    3.39627 ...
1978    2    1    2     93984333    32.459    0.30937    4.01496 ...
1978    2    1    3     12697786   41.6399   0.693022     5.0295 ...
1978    2    1    4     18951432   48.1069    1.07244    5.57896 ...
1978    2    1    5     41693571   53.4303    1.46937    6.00262 ...
1978    2    1    6    1588074.2   57.5301    1.82684    6.19847 ...
1978    2    1    7    643576.57   62.1813    2.28764    6.34397 ...
1978    2    1    8    9554658.7   69.1972    3.15214    7.96021 ...
1978    2    1    9     20769063   64.1637    2.56871    8.96017 ...
1978    2    1   10    588655.64   66.1815    2.80272    8.91051 ...
1978    3    1    1 1.7420548e+08   22.8558   0.104058    3.71852 ...
1978    3    1    2     85845786   34.6412   0.388004    4.60412 ...
1978    3    1    3     11885496   43.7357   0.809641    5.42533 ...
1978    3    1    4     17430666     50.12    1.21441    5.88096 ...
1978    3    1    5     37795319   55.3349    1.63169    6.21883 ...
1978    3    1    6    1427157.2   59.3484    2.00496    6.37542 ...
1978    3    1    7    574605.43    63.898    2.48258    6.48176 ...
1978    3    1    8    8288087.5   70.7135    3.36434    7.94123 ...
1978    3    1    9     17647459    65.779    2.76316    8.94646 ...
```

```
1978    3    1   10    463204.88    67.7538    3.00382    8.87781 ...
1978    4    1    1 1.5367201e+08   25.7112   0.147325     4.0496 ...
1978    4    1    2     78258267    36.9107   0.476281    4.99275 ...
1978    4    1    3     11065950    45.7865   0.934106     5.7459 ...
...
2006    4    1    1     61764642    25.8222   0.147986    4.14854 ...
2006    4    1    2     42755447    35.5779   0.419947     5.6459 ...
2006    4    1    3     31965748    44.0572   0.827166    6.53605 ...
2006    4    1    4     20839220    51.5206    1.33984    7.15173 ...
2006    4    1    5     11839987    58.3771    1.95411    7.68658 ...
2006    4    1    6    6121559.6    64.9279    2.67567    8.09454 ...
2006    4    1    7      2965164     71.307    3.51046     8.2034 ...
2006    4    1    8      3623721    76.8593    4.39112    7.63606 ...
2006    4    1    9    486773.98    80.8823     5.2124    6.92879 ...
2006    4    1   10    157149.29    86.1675    6.76523    5.23193 ...
```

9.2 StockPrinter output

The output file that is generated by the stockprinter printer class is given below. This class gives a more detailed view of the information available for the stock, giving the number and mean weight for each timestep/area/age group/length group combination specified in the aggregation files. The labels displayed for the area, age group and length group come from those given in the aggregation files.

```
; Gadget version 2.0.02 running on hafnasandur Mon Feb 10 14:47:43 2003
; Output file for the following stocks had
; year-step-area-age-length-number-weight
1978    1 allareas  allages     len1    995.193 0.00191354
1978    1 allareas  allages     len2      43026 0.00454673
1978    1 allareas  allages     len3     860844 0.00886809
1978    1 allareas  allages     len4 8.01603e+06   0.015264
1978    1 allareas  allages     len5 3.49402e+07  0.0241086
1978    1 allareas  allages     len6 7.16301e+07  0.0357836
1978    1 allareas  allages     len7 6.92705e+07  0.0507034
1978    1 allareas  allages     len8 3.17711e+07   0.069303
1978    1 allareas  allages     len9 7.94557e+06  0.0916179
1978    1 allareas  allages    len10 5.26611e+06   0.116013
1978    1 allareas  allages    len11 1.16291e+07   0.146101
1978    1 allareas  allages    len12 1.96601e+07   0.181922
1978    1 allareas  allages    len13 2.32207e+07   0.223447
1978    1 allareas  allages    len14 2.01052e+07   0.270148
1978    1 allareas  allages    len15 1.38054e+07   0.322652
1978    1 allareas  allages    len16 8.6823e+06   0.384569
1978    1 allareas  allages    len17 6.18012e+06   0.463307
1978    1 allareas  allages    len18 5.68363e+06   0.557399
1978    1 allareas  allages    len19 6.24462e+06   0.657535
1978    1 allareas  allages    len20  7.249e+06   0.765544
...
2006    4 allareas  allages    len35 1.21218e+06    3.64103
2006    4 allareas  allages    len36 1.0435e+06    3.94543
2006    4 allareas  allages    len37     888892    4.26335
2006    4 allareas  allages    len38     743557    4.59495
2006    4 allareas  allages    len39     605777    4.94046
2006    4 allareas  allages    len40     477600    5.30022
2006    4 allareas  allages    len41     362886    5.67473
2006    4 allareas  allages    len42     263087    6.06073
2006    4 allareas  allages    len43     526692    7.20074
```

9.3 StockFullPrinter output

The output file that is generated by the stockfullprinter printer class is given below. This class gives a more detailed view of the information available for the stock, giving the number and mean weight for each timestep/area/age group/length group combination with no aggregation.

```
; Gadget version 2.0.02 running on hafnasandur Mon Feb 10 14:47:43 2003
; Full output file for the stock had
; year-step-area-age-length-number-mean weight
1978    1    1    1    5     103.36 0.00115297
1978    1    1    1    6    891.833 0.00200168
1978    1    1    1    7     6315.8 0.00319122
1978    1    1    1    8    36710.2 0.00477994
1978    1    1    1    9     175129 0.00682645
1978    1    1    1   10     685715 0.00938952
1978    1    1    1   11 2.20365e+06  0.0125281
1978    1    1    1   12 5.81238e+06  0.0163013
1978    1    1    1   13 1.25829e+07  0.0207683
1978    1    1    1   14 2.23572e+07  0.0259886
1978    1    1    1   15 3.2604e+07  0.0320216
1978    1    1    1   16 3.90244e+07  0.0389268
1978    1    1    1   17 3.83368e+07  0.0467641
1978    1    1    1   18 3.09107e+07  0.0555932
1978    1    1    1   19 2.04557e+07  0.0654739
1978    1    1    1   20 1.11105e+07  0.0764662
1978    1    1    1   21 4.95298e+06  0.0886303
1978    1    1    1   22 1.81223e+06   0.102026
1978    1    1    1   23     544218   0.116714
1978    1    1    1   24     134136   0.132754
1978    1    1    1   25    27135.2   0.150207
1978    1    1    1   26     4505.4   0.169133
1978    1    1    1   27    613.969   0.189592
1978    1    1    1   28    68.6711   0.211646
1978    1    1    1   29          0          0
1978    1    1    1   30          0          0
1978    1    1    1   31          0          0
1978    1    1    1   32          0          0
1978    1    1    1   33          0          0
1978    1    1    1   34          0          0
1978    1    1    1   35          0          0
1978    1    1    1   36          0          0
1978    1    1    1   37          0          0
...
2006    4    1   10   81     4701.6    5.12461
2006    4    1   10   82    5213.33    5.29507
2006    4    1   10   83    5702.95    5.46909
2006    4    1   10   84    6155.36    5.64672
2006    4    1   10   85    6555.58    5.82804
2006    4    1   10   86    6891.08    6.01316
2006    4    1   10   87     7150.1    6.20224
2006    4    1   10   88    7112.11    6.38528
2006    4    1   10   89    6766.64    6.55577
2006    4    1   10   90    76473.4    8.17371
```

9.4 PredatorPrinter output

The output file that is generated by the predatorprinter printer class is given below. This class gives a detailed view of the information available for the predator/prey combination specified in the printfile, giving the biomass consumed for each timestep/area/predator length group/prey length group combination specified in the aggregation files. The labels displayed for the area, predator length group and prey length group come from those given in the aggregation files. Note that there is only one predator length group in this example, since the predator is a combination of the commercial fleet and the future fleet.

```
; Gadget version 2.0.02 running on hafnasandur Mon Feb 10 14:47:43 2003
; Predation output file for the following predators comm future
; Consuming the following preys had
; year-step-area-pred length-prey length-biomass consumed
1978    1 allareas      all     len1 1.90944e-10
1978    1 allareas      all     len2 2.76937e-09
1978    1 allareas      all     len3 7.90625e-08
1978    1 allareas      all     len4 6.4151e-06
1978    1 allareas      all     len5 0.000447578
1978    1 allareas      all     len6 0.0182648
1978    1 allareas      all     len7 0.480843
1978    1 allareas      all     len8   7.7407
1978    1 allareas      all     len9  79.9068
1978    1 allareas      all    len10  454.927
1978    1 allareas      all    len11  1645.09
1978    1 allareas      all    len12  3701.31
1978    1 allareas      all    len13  5278.53
1978    1 allareas      all    len14  5011.01
1978    1 allareas      all    len15  4426.19
1978    1 allareas      all    len16  6467.16
1978    1 allareas      all    len17  13664.1
1978    1 allareas      all    len18  28602.7
1978    1 allareas      all    len19  56642.7
1978    1 allareas      all    len20   107846
...
2006    4 allareas      all    len38   443284
2006    4 allareas      all    len39   382070
2006    4 allareas      all    len40   316995
2006    4 allareas      all    len41   252690
2006    4 allareas      all    len42   191934
2006    4 allareas      all    len43   445041
```

10 Parameter file

The parameter file is used to specify the initial values for the switches that are to be used in the Gadget model. This file is specified by a "-i <filename>" command line option when Gadget is started, and contains a list of all the switches, their initial value, the lower and upper bounds and a flag to note whether the optimiser should optimise that switch or not.

```
;
; Reference parameter file for the haddock example
;
switch   value        lower   upper  optimize
grq0     6.3421717    1       10     1 ; q0 in growth function
grq2     17.020301    5       20     1 ; q2, q3 in growth function
bbeta    11.792983    0.1     5000   1 ; beta in beta-binomial
age2     0.011234341  0.00001 1      1 ; initial number of age 2 fish
age3     0.0023677952 0.00001 1      1 ; initial number of age 3 fish
age4     0.0060429722 0.00001 1      1 ; initial number of age 4 fish
age5     0.022761935  0.00001 1      1 ; initial number of age 5 fish
age6     0.0014660210 0.00001 1      1 ; initial number of age 6 fish
age7     0.0010000000 0.00001 1      1 ; initial number of age 7 fish
```

```
age8     0.026070485  0.00001 1      1 ; initial number of age 8 fish
age9     0.096904951  0.00001 1      1 ; initial number of age 9 fish
rec78    22.377895    0.2     34     1 ; number of recruits in 1978
rec79    9.9845325    0.2     34     1 ; number of recruits in 1979
rec80    1.2187327    0.2     34     1 ; number of recruits in 1980
rec81    11.258073    0.2     34     1 ; number of recruits in 1981
rec82    4.8916410    0.2     34     1 ; number of recruits in 1982
rec83    3.5328430    0.2     34     1 ; number of recruits in 1983
rec84    8.2212940    0.2     34     1 ; number of recruits in 1984
rec85    9.0904755    0.2     34     1 ; number of recruits in 1985
rec86    24.934956    0.2     34     1 ; number of recruits in 1986
rec87    9.0852408    0.2     34     1 ; number of recruits in 1987
rec88    5.0242726    0.2     34     1 ; number of recruits in 1988
rec89    4.3285368    0.2     34     1 ; number of recruits in 1989
rec90    11.906707    0.2     34     1 ; number of recruits in 1990
rec91    20.386092    0.2     34     1 ; number of recruits in 1991
rec92    11.129591    0.2     34     1 ; number of recruits in 1992
rec93    9.1155295    0.2     34     1 ; number of recruits in 1993
rec94    11.856131    0.2     34     1 ; number of recruits in 1994
rec95    6.5406113    0.2     34     1 ; number of recruits in 1995
rec96    14.777468    0.2     34     1 ; number of recruits in 1996
rec97    4.2329148    0.2     34     1 ; number of recruits in 1997
rec98    7.9890151    0.2     34     1 ; number of recruits in 1998
rec99    24.796618    0.2     34     1 ; number of recruits in 1999
acomm    -10.342371   -100    -1     1 ; alpha in fleet suitability
bcomm    0.21405130   0.1     10     1 ; beta in fleet suitability
asur     -4.3391371   -100    -1     1 ; alpha in survey suitability
bsur     0.22464659   0.1     10     1 ; beta in survey suitability
```

F Estimation

F.1 Issues in multispecies models

Gunnar Stefansson[1]
Dept. of Mathematics
University of Iceland and Marine Research Institute

Abstract

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 this paper and some potential solutions are discussed.

1 Introduction

Models of the dynamics of fish populations are usually developed in response to certain questions of interest. The greatest emphasis has 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.

Each of these three issues, multispecies interactions, statistical models and uncertainty estimation can have a major impact on the percieved or likely range of population abundance and potential or safe harvest from the resource (Patterson et al, 2000, Patterson et al, 2001, Restrepo et al, 2000 and Gavaris et al, 2000).

This paper discusses specific modelling issues which arise in large exercises in statistical models of multispecies ecosystems, such as BorMiCon (Stefansson and Palsson, 1997a) and variants thereof. In accordance with specifications set out by Stefansson and Palsson (1998), this model has been developed as a forward simulation model using statistical estimation through weighted combinations of several log-likelihood criteria. Although that particular framework is a case-specific multispecies model, the statistical and modelling issues that arise are perfectly general to any multispecies model or, in fact many complex nonlinear statistical models which incorporate several data sources. The issues discussed in this paper have thus become sources of problems when developing, implementing and verifying models based on Gadget (Anon, 2002).

---

[1] Supported in part by EU grant QLK5-CT1999-01609.

2 Questions relating to management or general biology

Traditional fisheries management has been based on one or more regulatory mechanisms. Typically these involve quotas on individual species, effort limitations, closed areas and controls on the types of gear used.

Each of these mechanisms leads to different questions which need to be answered by the models used. Common to all systems is a need to evaluate the long-term yield potential, usually of each individual stock. In addition, system-specific information is required, e.g. a quota system will need advice on absolute catches for each stock, whereas an effort-control system only needs a suggestion for an overall effort allocation.

The presence of biological and technical interactions implies a second round of questions, regardless of whether the basic system is effort- or quota-based. Thus in the effort system, there is a need to evaluate the effects of a fleet switching from one species to another and in the quota system issues of coordinating quotas in mixed fisheries can become important. Under any management system effects of prey harvests on predator growth can be important as can the impact on prey abundance of reduced fishing of a predator.

These same and several other multispecies-oriented questions arise when closed areas are used. In particular migration models become important in this case. The migration models need to incorporate or estimate migration, drift or dispersal parameters which are crucial in predicting the effects of closed areas.

In the multispecies context it has been found that the form of prey selection can be quite important, as related to management. Thus, as a prey species diminishes in abundance, the predator may switch disproportionately to new prey. At low prey abundance such effects can have considerable impact on predictions of the effect of fishing, rebuilding predictions etc.

It is therefore seen that management issues give rise to a variety of multispecies-oriented questions which brings a need for a whole new breed of models. These new models need to incorporate much more detailed biology than has been considered in assessment models before.

Many traditional models of fish populations used a single mean length at age or were length-based without considering age. In most applications recruitment is highly variable and this induces a need to track yearclasses through age/time but the selectivity of the fishing gear as well as food selection is length-dependent and the food supply will also affect the growth of the predator. For these reasons modern models need to be able to account for variability in both age and length. Thus there is a need to incorporate growth models where the development of the distribution of length at age is made explicitly dependent on food supply. These models clearly become an order of magnitude more complex than models which are only age-based.

When models are made age- and length-based a growth update mechanism needs to be included. Several approaches exist for this (Bogstad et al, 1992; Stefansson and Palsson, 1997a), but the lack of decent data on individual growth has hampered the validation of such models. It would seem, however, that overly simple models of growth updates can fail to adequately explain the length distribution at age in samples and therefore parametric models with more flexibility are needed (Stefansson 2001).

Dynamic multispecies models which have been used most extensively for assessment purposes (e.g. MSVPA) have not included estimation of a recruitment function or attempted to close the life cycle. Closure of the life cycle is important, however, when attempting to estimate medium- or long-term yield potential. This has been investigated in the context of limit and precautionary reference points and the multispecies effects have been found not only to be quite important but also nontrivial (Collie and Gislason 2001).

On rare occasions migration modules have been incorporated in multispecies models. These need to be linked to tagging data and the tagged fish need to be affected by the same (or similar) processes as the remainder of the population. This leads to a considerable exercise in modelling and computing resources but it is not clear initially how well such parameters can be estimated and new methods of assessing estimability and uncertainty estimation are needed.

Several methods have been used to describe species interactions. In particular different functional forms have been used to describe (local) consumption, with several functional responses available. In rare cases these have been compared formally through adequacy of fits to data (Magnusson and Palsson, 1989) but usually a single model is assumed or an alternative model is tested as a deviant from a baseline (Mohn and Bowen, 1994; Danielsson et al, 1997; Stefansson et al, 1997).

These biological issues are brought into models through the inclusion of model complexity and more data sources. This leads to entire new classes of problems relating to statistical and numerical estimation. As will be seen below, new methods are needed to enhance the ability to detect inconsistencies between models and several data sets, the ability to detect the appropriate functional relationships to be included in different model components, measures to indicate deviations from data, describe estimability of parameters and so on. These methods to address the above biological issues are quite general and statistical in nature and need to be developed as such.

3 Statistical issues

3.1 Basic model assumptions

There are many cases when basic model assumptions used in single species assessments are clearly violated, yet there is no easy way to fix the problem. For example it is common in several assessment models that a nonlinear relationship between an abundance index and modelled population size is obtained. In principle the power in this relationship should be the constant 1 for surveys which are designed to provide abundance indices. Obtaining a power which is significantly different from 1 is not a problem per se, but when the estimated powers are large for some age groups and small for others without regularity (apparent e.g. in Stefansson, 1992, as noted by Butterworth, Pers. Comm.) there would appear to be some effect present which is not captured by the model but is being picked up by such parameters.

Such unquantifiable effects are somewhat worrisome since there is no clear reason why they should appear. In some cases it appears that changing a parameter from being a constant to being a time series effect gives a considerably improved model fit. It is clear that very many concepts included in fisheries models are not constant in reality but should be allowed to vary smoothly with time. This is known to be true of such simple things as the annual selection pattern (Gudmundsson, 1994) which clearly violates the commonly used seperability assumption on fishing mortality (e.g. Deriso et al, 1985). Similarly, migration rates will also change with time (as modelled in MULTSPEC, cf Bogstad et al, 1997), though probably not completely at random or independently from year to year but with some time series structure.

It is quite possible that the real problem which manifests itself in “strange” parameter estimates is really due to the lack of flexibility and lack of such time series parametrization. The truly worrisome situation is when a simple assessment model is used with no diagnostics to indicate that there is a problem of this form.

Section 3.4 below addresses the issue of detecting some of these problems when different data sets allow the identification of model inconsistencies. The generic model identification issue (defining the mean function) is somewhat different. In simple cases such as simple linear regression with repeated measurements it is possible to use lack-of-fit tests (Neter et al 1996) to identify whether the assumption of linearity is inappropriate. In constrained linear or nonlinear optimisation it is possible to use “shadow values” to investigate the effects of alleviating the constraints.

If the nonlinear model is statistical it may be possible to use factors to investigate whether the functional form is incorrect. Although, as above, in complex models it is sometimes possible to change a parameter to a time series effect, there is a more general need to identify where such model changes will work to provide a better fit. The multispecies models become quite complex with modules to describe growth, migration, consumption, fishing, etc. The equivalent of "contraints" is the linking mechanism between these modules. Basically, what is needed is a generic method to indicate how much the fit will improve if some of the links between these complex modules are broken or made more flexible.

The methods described in Richards (1991) and the weighting methods described below provide one approach to investigating such problems but more generic approaches are needed as the incorrect links may not correspond directly to data sets but only to pairs of biological modules.

3.2 Weighting data sources

The typical statistical multispecies model is based on several data sources. Although this phenomenon is common in fishery science, the number of data sources tends to go up more than linearly as the number of species increases. Thus, there are data sets on length measurements, age readings, tagging experiments, stomach contents, landings, survey abundance and so on, all of which play a role in determining parameters of these models. The issue of combining all these data sources through a single model can be examined using a much simpler model.

Considering first the simple independent-Gaussian assumption on probability distribution (possibly after log-transformation), it is more common than not that data come from several different sources in which case the total likelihood function becomes a product of the likelihoods of the different components (assuming independent sampling from the various data sets). The objective function to be minimized for estimation purposes is the negative log likelihood which becomes the sum of the log components (or sum of squares if the variances are assumed known).

When this Gaussian assumption is used for the different data sources, it is in principle possible to estimate all variances involved using maximum likelihood. In practice, however, this can be quite tricky. This is particularly the case when the model can not explain all the data sources simultaneously.

The actual problem in this case consists of two parts. Firstly, the weights (usually inverse variances) to be given to each data source are unknown and secondly the model is incorrect. A large model is like a black box, where parameters are put in and out come values of likelihood components or sums of squares. The effect of an incorrect black box in this case is that a simultaneous good fit to all data sources can not be obtained. This has been a known problem in fishery science for some time (Richards, 1991; Schnute and Hilborn, 1993).

To mimic this issue, take the simple true state of nature (the data generator) to consist of independent measurements around two straight lines $Y_{ij} \sim n(\alpha_i + \beta_i x_{ij}, \sigma_i^2), \quad j = 1, \ldots, n_i \;\; i = 1, 2.$

As in reality the black box model assumed will not know the variances, but what is worse, the model is typically incorrect in reality so here it does not know that the slopes are different, but assumes that $\beta_1 = \beta_2$. In this scenario, maximum likelihood estimations of all parameters including the variances can be used, but the values of the log-likelihood surfaces are plotted in Fig. 1 as individual curves for each data set as well as the sum for simulated data with n_i=50 data points in each set. This trivially simple example suffices to illustrate the problem of apparently incompatible data sets.

Of course when the slopes are similar there is no problem, but as they move apart, the situation becomes as documented for real data where the model can not fit both data sets. It is seen in Fig. 1 that when the true variances are similar, the maximum likelihood estimate can flip between being close to each of the two true slopes and correspondingly the variance estimators flip from one or the other being correct. This is in accordance with the results found by Richards (1991). It is also seen in this case that it is not particularly feasible to obtain (local) diagnostics in terms of e.g. the slope of each objective function since the slope can be close to zero at the minimum of the other (this is due to the logarithmic nature of the likelihood as a function of the variance).

In terms of variance estimation there is light at the end of the tunnel, however, since it is technically feasible to estimate the variance in each data set under a Gaussian assumption. This can be done by

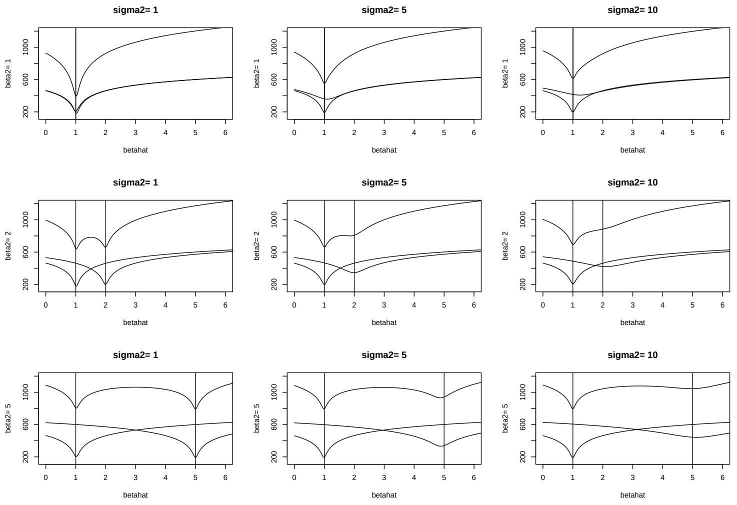


Figure 1: Minimum negative log-likelihood value from simulated data using assumed model, as a function of $\hat{\beta}_2$, for fixed true values of $\beta_1 = 1$ and $\sigma_1 = 1$ but different true values of σ_2 and β_2.

first reducing any single variance towards zero (i.e. diminishing the influence of any other data sets) and using the corresponding resulting sum of squares in estimating that variance. Using this technique iteratively across data sets will result in unbiassed estimates of the appropriate weighting factors, even if the model can not fit all data sets simultaneously, as long as each individual data set can be explained. A slight problem with degrees of freedom will remain but in cases when the number of data sources and samples is large it will be minor.

For example, a multispecies model may not be able to reconcile growth data on a predator with abundance data on a prey. In this case, fitting to the predator growth data alone may force modelled prey abundance to take on unrealistic values, but estimates of the variance of growth measurements will be obtained.

To stabilize estimation in real applications it is likely that the various alternative data sources can not be completely omitted but included with very low relative weights compared to the one where the variance is being estimated.

Since it is thus possible to obtain estimates of the variability of each data set, it remains to develop methods to elucidate the “incorrectness” of the black box, to investigate alternative likelihood functions and investigate the effects of non-independent sources of measurements.

3.3 Likelihood functions

Likelihood functions in fishery science are commonly based on traditional probability distribution such as the Gaussian, multinomial, negative binomial or gamma distributions.

The negative binomial, lognormal and gamma distributions are sometimes used for abundance data. It has frequently been noted that these distributions may not be appropriate for these data, even when restricting to tows with positive numbers of fish (e.g. Steinarsson and Stefansson, 1986, Stefansson, 1996 and Brynjarsdottir, 2002). Similarly, it has been found that even the simple Bernoulli assumption is not an appropriate model for the extremely simple presence/absence of fish in each tow (Stefansson and Palsson, 1997b). Complex models have also been used, combining several distributions (Lo et al, 1992) but these have also been found wanting (Stefansson, 1996).

In fish stock assessment models it is not uncommon to assume the catches in numbers at age or length to be lognormal, but in other settings a multinomial distribution is preferred. Now, it is indeed clear that even if the entire catch was sampled for age, the distribution would not be lognormal, but the multinomial assumption seems quite popular for counts, particularly for length distributions (e.g. MacDonald and Pitcher, 1979). Unfortunately this assumption is usually not valid for real data sets.

As an example, consider the simplest case of length distributions from a groundfish survey. To standardise the measurements and simplify binomial variances it has been proposed to randomly subsample a fixed number (n) of length-measured fish (here 25) from each station (Hrafnkelsson and Stefansson, 2002). If these measurements were from a multinomial distribution, then the variance of the number of fish, X, per length cell would be $V[X] = npq$ where p is the true proportion of fish in that length cell and $q = 1 - p$. On the other hand the theoretical mean is $E[X] = np$. Within a small area (a statistical sub-rectangle of 30' by 15') the data from different tow locations can be used to compute the observed means and variances. These data and theoretical curve are given in Fig. 2.

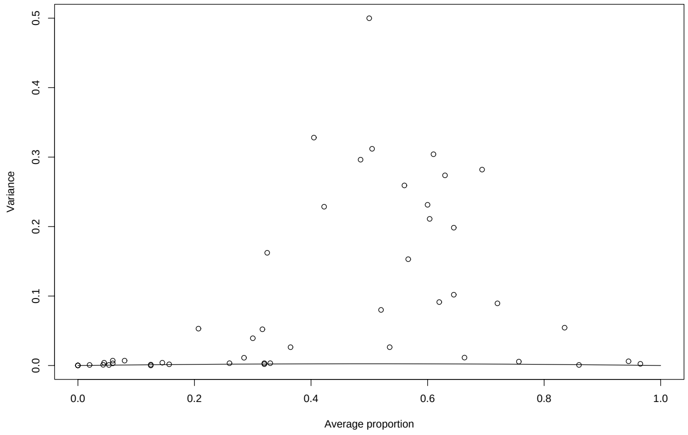


Figure 2: Variance-mean relationship for proportion (p) of cod per tow below 45 cm, theoretical (solid curve, $Var[\hat{p}] = pq/n$) and observations (circles). Each point represents the mean and variance of the number of $n = 100$ randomly selected fish below 45 cm caught at different tow locations within a small area.

It is clear from this analysis that the multinomial assumption is completely incorrect in terms of the variance function. Simple further analysis along the same lines indicates that the correlations between length cells is similarly incorrect. Since this is in fact fairly obvious when the biology is considered (the correlation between the abundance of fish in adjacent size groups is bound to be higher than between large and small size groups), the general use of a multinomial is questionable to say the least.

At present there is no easy way to amend the multinomial to incorporate the general correlation structure. Using a multivariate Gaussian density might suffice but this needs to be formally evaluated.

3.4 Apparent incompatibilities

As noted above, in many real applications the pictures indicated by the different data sources appear to be inconsistent. Thus time trends of catch-per-unit-effort (CPUE) from different fishing gear can indicate different trends in abundance. The true reasons for this can be simply that there is a time trend in catchability of the commercial fleet, which is not incorporated in the model. In terms of fitting (incorrect) models what is seen is an incompatibility between data sources. Methods are required to detect such apparent incompatibilities.

Sometimes time series components can be added to the models to alleviate problems such as these. Thus, simple random walk effects in catchabilities have the potential to describe persistent changes in catchability. Naturally, generic time series effects are no subsitute for an appropriate physical model of the process, but in the absence of such information the time series approach has considerable potential.

The generality of these apparent inconsistencies between data sources is unknown but several examples have been documented where these effects are important enough to give quite different results. These effects are not restricted to complex models as they have been found in simple single-species models, e.g. different abundance trends or different results solely due to different weighting of individual age goups in the same survey (Stefansson, 1998).

It is therefore seen that at least in some cases fitting to more than one data set gives very different indications of important quantities such as population trends. Some approaches to this issue have been proposed, such as fitting to each data series separately (Polacheck et at, 1993). A more general approach is to investigate series of weightings to each data set (Stefansson, 1998), as detailed in section 3.2. This can potentially give estimates of appropriate weightings for each data set.

Having obtained the weights, a selected parameter can be estimated using only one data source and then all the remaining ones. If there are enough data in the “remaining” data sources, then one can treat that estimate as fixed and test whether the estimate using the single data source deviates significantly from the fixed value using a standard t-test based on the Hessian matrix or a likelihood ratio test. Each approach depends on assumptions, but this would be a way forward to investigate the inconsistencies quantitatively and objectively. Alternative methods based on testing the entire parameter vectors are under development.

It follows that the question of apparently incompatible data sets can to some extent be detached from the issue of weights to be given to each data set. However, the issue of selection of likelihood functions remains problematic and intertwined with apparent incompatibility: Incompatibility is detected if a likelihood component increases “too much” (i.e. significantly) when a parameter is fixed at the value estimated by the remainder of the data. If the likelihood function is inappropriate then it is quite possible that the value of “too much” is inappropriate, i.e. rejection occurs too frequently. The overdispersion issue in the multinomial abve is precisely such a case.

In fisheries it is quite common that most standard likelihood functions are rejected as inappropriate. As more complex models are considered these issues will become steadily more important and means to resolve them need to be developed.

As indicated in this section, to obtain an orderly development of these complex models a new methodology is needed to verify model adequacy. In simple linear regression with repeated measurements this is done through lack-of-fit tests (Neter et al 1996). In the present setting these verifications need to be formalized by comparing point estimates through different weighting schemes. Such statistical comparisons are outlined in Stefansson (2004).

3.5 Goodness-of-fit tests

As mentioned above, several likelihood functions are rejected when statistical measures of goodness-of-fit are applied. The reasons for this are manyfold, but they certainly include overdispersion due to patchiness, correlation due to the animals’ behavior, a model’s basic inability to fit the data (i.e. error in the mean function) and so on.

Assuming that the issues above of selecting appropriate likelihood functions are alleviated, the functional relationships are properly modelled, appropriate weights are estimated for each data set and so on, then the goodness-of-fit tests should not reject.

The multispecies model BorMiCon can use a large data set, as seen in the first implementation described in Stefansson and Palsson (1997a). Initial goodness-of-fit tests were carried out for the outputs from this 3-species, multiarea model. The tests included tests for normality (for assumed Gaussian residuals), tests for the multinomial assumptions for length distributions and so on. For some of the classes of tests, most tests rejected the appropriateness of the distributional assumptions. In no cases was the proportion rejected less than (or even close to) 5% when using a level of 5%. It follows that there is considerable work that needs to be undertaken to make these models such descriptions of the ecosystems that they actually explain the measurements.

Work on goodness-of-fit tests will thus become crucial as the likelihood functions, model identification schemes and weighting methods are improved. At present, the apparently incompatible data sets, inadequate likelihood functions and not-fully-developed weighting mechanisms completely override any goodness-of-fit test development.

The goodness-of-fit tests to be used in the fishery science of the future will have to be able to handle generic likelihood functions. Ideally a goodness-of-fit test should be able to deal with deviations from different likelihood components in a generic manner, rather than the distribution-specific manner used in today's tests. At present this is an open problem.

4 Discussion: Estimability and uncertainty

In multiple linear regression the concept of an "estimable function" or parameter is a well-defined mathematical concept (Scheffe, 1959) as it basically denotes a parameter which is uniquely defined in the model. Similarly, but not equivalently, the structure of the $\mathbf{X}$-matrix defines what parameters are well estimated etc. In the typical non-linear highly complex spatially disaggregated multispecies statistical fisheries models, this is not so simple.

It is clear within several models that some subsets of parameters are more difficult to estimate or disentangle than others. Notably, the selection pattern in a survey is closely linked to the growth curve, particularly if only mean length at age is available.

For some parameters, it is not clear whether data exist to estimate them properly. Migration parameters are an example where this may be difficult, though of course some bounds are placed on these parameters through regional catch history alone and some minimal information will be obtained through tagging experiments.

It is not a priori clear at all just how accurately migration parameters can be estimated, but it is clear that they are linked to very many other parameters of population models. For example different areas are intrinsically associated with different growth rates through variable food supply and temperature. In this case modelled or real changes in migration rates will also manifest themselves in corresponding changes in regional body size and weight. It follows that the migration rates form a crucial part of the multispecies models and resolving estimation issues for these parameters is of major importance.

In order to clear up questions of estimability in complex models simulated data sets will most likely be required. In principle these can be used to determine what quantities stand a remote chance of being estimated.

Estimating uncertainty is related to the question of estimability. It is not clear how uncertainty should be estimated in complex models with many data sources. However, bootstrapping all the various data sources may eventually provide the uncertainty estimates required.

To ensure some level of objectivity, the way forward with complex fisheries models needs to be statistically based. Statistical methods need to be developed to ensure orderly buildups of these nonlinear models through increasing complexity as well as selection of appropriate functional relationships and statistical tests are needed to evaluate whether models explain all available data sets adequately.

References

Anon. 2002. Development of structurally detailed statistically testable models of marine populations(dst2). QLK5-CT1999-01609. Progress report for 1. January 2001 to 31 December 2001. MRI Tech. Report 87.

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

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.

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

Brynjarsdottir, J. 2002. Statistical analysis of cod catch data from Icelandic groundfish surveys. MRI Tech. Report 91, Reykjavík. 81pp.

Collie JS and Gislason H, 2001. Biological reference points for fish stocks in a multispecies context Can. J. Fish. Aquat. Sci. 58 (11): 2167-2176

Daníelsson, Á, Stefánsson, G., Baldursson, F.M., Thorarinsson, K. (1997). Utilization of the Icelandic Cod Stock in a Multispecies Context. Marine Resource Economics 12 329-244

Deriso RB, Quinn TJ and Neal PR, 1985. Catch-age analysis with auxiliary information Can J Fish Aquat Sci 42 (4): 815-824

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. Stefánsson (2000). Comparison of Uncertainty Estimates in the Short Term Using Real Data. CM 2000/V:03

Gudmundsson, G. 1994. Time-series analysis of catch-at-age observations. Applied Statist. J. Royal Statist. Soc. Ser. C 43 (1): 117-126

Hrafnkelsson and Stefansson, 2002. A model for categorical length data from groundfish surveys. Submitted.

Lo NCH, Jacobson LD, Squire JL, 1992 Indexes of relative abundance from fish spotter data based on delta-lognormal models. Can J Fish Aquat Sci 49 (12): 2515-2526

MacDonald, P. D. M. and Pitcher, T. J. (1979). Age-groups from size-frequency data : a versatile and efficient method of analyzing distribution mixtures. J. Fish. Res. Board Can. 36 987-1001

Magnússon, K.G., Pálsson, Ó.K. (1989). Trophic ecological relationships of Icelandic cod Rapp. P.-v. Réun. Cons.int. Explor.Mer, 188 206-224.

Mohn, R. and Bowen, W.D. (1994). A model of grey seal predation on 4VsW cod in its effects on the dynamics and potential yield of cod. DFO Atlantic Fisheries Research Document 94/64.

Neter, J., Kutner, W. H., Nachtsheim C. J. and Wasserman, W. 1996. Applied linear statistical models. McGraw-Hill Boston. 1408pp.

Patterson, K.R., R.M. Cook, C.D. Darby, S. Gavaris, V.R. Restrepo, A.E. Punt, B. Mesnil, D. W. Skagen, G. Stefánsson, 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.

Polacheck, T., Hilborn, R. and Punt, A.E. (1993). Fitting Surplus Production Models: Comparing Methods and Measuring Uncertainty. Can J. Fish. Aquat Sci. 50 2597-2607

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. Stefánsson (2000). Do Different Methods Provide Accurate Probability Statements in the Short Term? ICES CM 2000/V:08

Richards LJ, 1991. Use of contradictory data sources in stock assessments. Fish Res 11 (3-4): 225-238

Scheffe, H. 1959. The analysis of variance. John Wiley&Sons. New York. 477 pp.

Schnute, J. T., and R. Hilborn. 1993. Analysis of contradictory data sources in fish stock assessment. Canadian Journal of Fisheries and Aquatic Sciences 50:1916-1923.

Stefansson, G. (1992). Notes on the stock-dynamics and assessments of the Icelandic cod. ICES C.M. 1992/G:71

Stefansson, G. (1996). Analysis of groundfish survey abundance data: combining the GLM and delta approaches. ICES Journal of Marine Science 53 577-588

Stefansson, G. and Palsson, O. K. (eds.) 1997a. Bormicon: A Boreal Migration and Consumption Model. Marine Research Institute technical report. 253 pp.

Stefansson, G. and Palsson, Ó.K. (1997b). Statistical evaluation and modelling of the stomach contents of Icelandic cod (Gadus morhua) Can J. Fish. Aquat. Sci. 54 169-181

Stefánsson, G. and Pálsson, Ó. K. (1998). A framework for multispecies modelling of Boreal systems. Reviews in Fish Biology and Fisheries, 8:1, 101-104.

Stefánsson, G., Sigurjónsson, J. and Víkingsson, 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. (1998). Comparing different information sources in a multispecies context. In Fishery Stock Assessment Models. Alaska Sea Grant College Program. AK-SG-98-01. 741-758.

Stefansson, G. 2001. An implementation of growth. In Anon. 2002. Development of structurally detailed statistically testable models of marine populations (dst2). QLK5-CT1999-01609. Progress report for 1. January 2000 to 31 December 2000. MRI Tech. Report 78.

Stefansson, G. 2004. Consistent combination of likelihood components. In prep.

Steinarsson, B. Æ. and Stefansson, G. (1986). Comparison of random and fixed trawl stations in Icelandic groundfish surveys and some computational considerations. ICES C.M.1986/D:13

G External biological models

G.1 Modeling stochastic age-length-structured multispecies stock dynamics.

PETER LEWY and MORTEN VINTHER

Draft January 2003

Introduction

Development of a stochastic multi-species model

The model developed is supposed to include a historical part, “VPA”, and a prediction part. The model assumes that the area considered is homogeneous with respect to growth and mortality. Migration within the area will not be considered.

Stock dynamics including fishing and predation mortality of commercially important stocks will be described using commercial and survey catch-at-age data and stomach contents data. The model will be based on ICES MSVPA (reference). However, the food preferences of the predators, the suitability parameters, will be modelled further to reduce the number of parameters, using the formulation of Andersen and Ursin or other formulations.

The model will be implemented as modules such that separate processes will be clearly separated and exchangeable. The following modules be included:

- Data input/database module
- Modules for predation and growth models.
- A module for definition of the likelihood function for the catch-at-age observations.
- A module for definition of the likelihood function for the stomach contents observations
- An output module enabling graphical and other presentations.

The prediction model may be fleet based and include technical interactions, i. e. that the catch composition by species and age for each fleet is accounted for by the model.

Notation

s denotes the species
prey denotes the prey species
of denotes other food
pred denotes the predator species
a,b denotes the age
q denotes the season
y denotes the year
C(s,a,y,q) denotes observed catch in numbers
$\hat{C}(s,a,y,q)$ denotes the expected catch in numbers
$\sigma(s,a)$ denotes the standard deviation of log catches
CPUE(survey,s,a,y,q)
$\sigma(survey,s,a)$ denotes the standard deviation of log CPUE
N(s,a,y,q) denotes the stock numbers in the sea
Z(s,a,y,q) denotes total mortality rate
F(s,a,y,q) denotes fishing mortality rate
Y denotes the number of years available
A(s) denotes the number of age groups
NOS denotes the number of seasons
RS(*s*) denotes the season where the fish is recruited to the fishery

M1(s,a,y,q) denotes natural mortality excluding predation
M2(prey,a,y,q) denotes predation mortality
SUIT(prey,l_{prey}*,pred,*l_{pred}*,q)* denotes suitability for given prey/length and predator/length groups
l_{prey} denotes a length interval for a prey species
$l_{predator}$ denotes a length interval for a predator species
$\eta(pred)$ denotes the parameter expressing the log "mean" of the preferred prey size
$\sigma_{pref}(pred)$ denotes the "standard deviation" in the food preference function
$\rho(prey, pred)$ denotes vulnerability parameters
STOM(prey,l_{prey}*,pred,*l_{pred}*,y,q)* denotes the observed average weight proportion of a prey/length group to the total weight of the stomach contents of a predator/length group by year and season.
$AVAILB(pred, l_{pred}, y, q)$ denotes total available biomass
$w(s, a, y, q)$ denotes the mean weight at age
Food(pred,b,y,q) denotes the food intake

Modelling total catch-at-age observations

Catch-at-age is considered a stochastic variable subject to sampling and process variation. The probability model for these observations is a multiplicative model e.g. as defined by Lewy (200?):

Catch at age is assumed to be log normal distributed with log mean equal to log of the standard catch equation, i.e. ln*(C(s,a,y,q))* $\sim$ Normal(*E(*ln*(C(s,a,y,q))*, $\sigma^2_{catch}(s,a)$). The variances for the fully exploited fish may be assumed having the same variance. Thus, the likelihood function, L_C, associated with the catches is

$$L_C = \prod_{s,a,y,q} \frac{1}{\sigma_{catch}(s,a)\sqrt{2\pi}} \exp(-(\ln(C(s,a,y,q)) - E(\ln C(s,a,y,q))))^2/(2\sigma^2_{catch}(s,a))) \quad (1)$$

Where

$$E(\ln(C(s,a,y,q)) = \ln(\frac{F(s,a,y,q)}{Z(s,a,y,q)}(1 - \exp(-Z(s,a,y,q))N(s,a,y,q)) \quad (2)$$

The negative log-likelihood for total catches then becomes:

$$\begin{aligned} l_C = -\ln(L_C) \propto 4Y \sum_{s,a} A(s) \ln(\sigma_{catch}(s,a)) + \\ \sum_{s,a,y,q} (\ln(C(s,a,y,q)) - E(\ln(C(s,a,y,q))))^2/(2\sigma^2_{catch}(s,a))) \end{aligned} \quad (3)$$

Modelling survey indices

CPUE(survey,s,a,y) in analogy with the commercial catches the survey indices are assumed lognormal distributed with mean

$E(\ln(CPUE(survey, s, a, y)) = \ln(Q(survey, s, a, y, q)\bar{N}(s, a, y, q))$

where *Q* indicates catchability of the survey and *q* is the quarter in which the survey takes place. Log variance, $\sigma(survey, s, a)$ for the fully exploited fish is assumed to the same. The log-likelihood is on the same form as equation (3).

Modelling fishing mortality

Total fishing mortality is for each species, *F(s,a,y,q)*, is modelled as a partly separable model (the species index, *s*, is left out for convenience):

$F(a, y, q) = F_1(a, y)F_2(y)F_3(a, y, q)$

This model assumes fishing mortality by age, year and quarter can be split up into age/year, year and age/year/quarter effects.

The range of year considered may be divided into a number of periods, T_{year} , such that

$F_1(a, y) = F_1(a, period)$ for each *y* in the *period* = 1,2,. . . ,T_{year}

Correspondingly

$F_3(a, y, q) = F_3(a, period, q)$

The age groups, for instance the fully exploited, may be combined as well:

$F_1(a, period) = F_1(a_{limit}, period)$ for $a \geq a_{limit}$ and *period*=1,2,. . . T_{year}

$F_3(a, period, q) = F_3(a_{limit}, perid, q)$ for $a \geq a_{limit}$ and *period*=1,2,. . . T_{year}

The age interval from the first age up to a_{limit} may be further divided into several intervals.

In order to ensure unique parameters two ties have to be set. We have chosen to fix F_2*(y=1)*=1 and F_3*(a,period=*T_{year}*,q=4)*=1.

In general if we have *Y* years, *A* age groups we 4*YA* observations and 5$a_{limit}T_{year}$ - 2 parameters. If *Y*=20, *A*=10, T_{year}=2 and a_{limit}=3 we have 800 observations and 28 parameters.

Survival of the stocks

The survival of the stock in the sea is described by the usual exponential decay equation.

$$N(s, a, y, q + 1) = N(s, a, y, q) \exp(-Z(s, a, y, q)) \tag{4}$$

or

$$N(s, a, y + 1, q = 1) = N(s, a, y, q = \text{last season}) \exp(-Z(s, a, y, q = \text{last season}) \tag{5}$$

Initial stock size, i.e. the stock in the first year and recruitment over years are considered as parameters in the model while the remaining are considered as functions of the parameters recursively determined by equations (4) and (5).

The stock parameters are $N(s, a \geq NOS - RS(s), y = 1, q = 1), N(s, a = 0, y, q = RS(s))$ for all species, years and age groups $a \geq 1$, where *NOS* is the number of seasons and *RS*(*s*) is the season where species *s* is recruited to the fishery.

In a multispecies model including fish predation total mortality, $Z(s, a, y, q)$ are divided into three components, predation mortality, other mortality and fishing.

$$Z(s, a, y, q) = M1(s, a, q) + M2(s, a, y, q) + F(s, a, y, q) \tag{6}$$

If M1 and M2 are assumed known the model described so far is a stochastic single species assessment model (e.g. Lewy ???) where each of the species can be treated independently of the interactions due to predation.

Stomach contents models

The observations considered for modelling predator food preference and estimation of parameters scaled to the total North Sea is average relative stomach content for the entire stock. These are calculated as weighted averages of prey proportion by Roundfish area weighted with the density of the predators. (reference til MV paper). These observations, *STOM(prey,*l_{prey}*,pred* l_{pred}*,y,q)*, are considered for each length group, l_{prey} and l_{pred} and are assumed to be stochastic variables subject to sampling and process variations. A probability distribution for these observations has to be specified to calculate the likelihood function.

In spite of that the catch models are age structured the stomach contents models considered will – in contrast to deterministic multispecies models - be length based. This decision is based on results of bootstrapping the stomach content observations (Vinther 2001), which show that for given predator and prey species the observed weight proportions in the stomach in some cases have correlations close or equal to one and hence are close to be linearly dependent for a range of age groups. This applies for instance to weight proportions of the 2, 3 and 4 years old sandeel in the stomach of 2 year old cod, *STOM*(*sandeel,a,cod,2,y,q*) for a=2,3,4. In most cases this is caused by that externally given age/length keys are used to translate a length groups to age groups for both predator and prey. Caused by these linear dependencies we have chosen to model food preference based on length.

For a given predator/length the observations, proportions of the prey species/length groups included, may further be correlated of two reasons: 1. The proportions are summing to one. 2. Specific prey items may occur in the stomachs in a systematic pattern due to the combined effect of the preference of the predator and the spatial overlap of prey species and the predator. Thus, for a given predator the observations of all prey species, (*STOM(prey,prey length,pred,pred length,y,q)*)$_{prey,length}$, need to be described by a multivariate distribution which includes correlation between observations.

Three possibilities have been considered: A Dirichlet and a multivariate normal and a multivariate log normal distribution. Basis for the stomach contents modelling is the expected value and the variance/covariance matrix of the observed proportions in the diet, *STOM*, which has to be specified.

The expected value of the stomach observations is modelled using the theory developed by Andersen and Ursin (1977) and the modifications made by Gislason and Helgason (1985), which resulted in the deterministic Multispecies VPA, MSVPA.

Assuming that the quarterly stomach sampling has taken place uniformly over the quarters the expected value of log observations has been set to:

$$E(STOM(prey, l_{prey}, pred, l_{pred}, y, q)) =$$

$$\frac{\bar{N}(prey,l_{prey},y,q)w(\bar{l}_{prey}(y,q))(SUIT(\bar{l}_{prey}(y,q),\bar{l}_{pred}(y,q))}{\sum\limits_{prey,l_{prey}} \bar{N}(prey,l_{prey},y,q)w(\bar{l}_{prey}(y,q))SUIT(\bar{l}_{prey}(y,q),\bar{l}_{pred}(y,q))+B(of,y,q)*SUIT(of,\bar{l}_{pred}(y,q)}$$

where l_{prey} and l_{pred} are specified length groups, where $SUIT(\bar{l}_{prey}(y,q), \bar{l}_{pred}(y,q))$ and $SUIT(of, \bar{l}_{pred}(y,q))$ are parameters describing the combined effect of food availability and preference of specified prey species and other food (*of*), where $\bar{l}_{prey}$ and $\bar{l}_{pred}$ are mean length of the fish in the length groups and where $\bar{N}(prey, l_{prey}, y, q)$ is the mean stock number during the quarters for specified preys and prey length groups. As stock numbers are age based in the catch model stock numbers by length in the above equation need to be expressed by age. This is done using length/age keys which for each age group gives the proportion by length group: If $\beta(l_{prey}, a, y, q)$ denotes the proportion of fish in length group l_{prey} of *a* year old fish ($\sum\limits_{l_{prey}} \beta(l_{prey}, a, y, q) = 1$) we have that $N(prey, l_{prey}, y, q) = \sum\limits_{a} \beta(l_{prey}, a, y, q) N(prey, a, y, q)$.

Using this formula one can show that

$\bar{N}(prey, l_{prey}, y, q) = \sum\limits_{a} \beta(l_{prey}, a, y, q) \bar{N}(prey, a, y, q)$

Insertion of the latter formula into the expected value of ln *STOM* gives

$$\begin{gathered} E(STOM(prey, l_{prey}, pred, y, q) = \\ \frac{w(\bar{l}_{prey}(y,q)) SUIT(\bar{l}_{prey}(y,q), \bar{l}_{pred}(y,q)) \sum_{a} \beta(l_{prey}, a, y, q) \bar{N}(prey, a, y, q)}{AVAILB(pred, l_{pred}, y, q)} \end{gathered} \tag{7}$$

where

$$\begin{aligned} &AVAILB(pred, l_{pred}, y, q) = \\ &\sum_{prey, l_{prey}} w(\bar{l}_{prey}(y, q)) SUIT(\bar{l}_{prey}(y, q), \bar{l}_{pred}(y, q)) \sum_{a} \beta(l_{prey}, a, y, q) \bar{N}(prey, a, y, q) \\ &+ B(of, y, q) \times SUIT(of, \bar{l}_{pred}(y, q)) \end{aligned}$$

where *of* is other food and $B(of, y, q)$ is biomass of other food.

In the ICES MSVPA the suitability by age are estimated directly without further modelling assuming that they are independent of the year. This is not convenient in a stochastic model because of the large number of parameters. Thus, suitability is modelled assuming that food preference can be split into species dependent and size dependent components as suggested by Andersen and Ursin (1977):

$$\begin{aligned} SUIT(\bar{l}_{prey}(y, q), \bar{l}_{pred}(y, q)) &= \rho(prey, pred, q) \exp(-\frac{(\ln \frac{\bar{\mathrm{l}}_{\mathrm{pred}}(y,q)}{\bar{\mathrm{l}}_{\mathrm{prey}}(y,q)} - \eta(pred))^2}{2\sigma^2_{pref}(pred)}) \\ &= \rho(prey, pred, q) g(prey, pred, y, q) \end{aligned} \tag{8}$$

where ρ are prey and predator species dependent vulnerability parameters, where $\eta(pred)$ denote the relative size preference of the predator and $\sigma^2_{pref}(pred)$ the “variance” of preferred length size and where

$$g(prey, pred, y, q) = \exp(-\frac{(\ln \frac{\bar{\mathrm{l}}_{\mathrm{pred}}(y,q)}{\bar{\mathrm{l}}_{\mathrm{prey}}(y,q)} - \eta(pred))^2}{2\sigma^2_{pref}(pred)}) \tag{9}$$

As the size of other food is not known suitability for this prey is modelled as follows:

$$SUIT(of, pred, \bar{l}_{pred}(y, q)) = \rho(of, pred, q) \exp(-\alpha(pred)(l - l_{\min(pred)})) \tag{10}$$

Consequently the other food term in equation (7) can be written as

$$B(of,y,q)*SUIT(of,\bar{l}_{pred}(y,q)) = B(of,y,q)\rho(of,pred,q)\exp(-\alpha(pred)(l-l_{\min(pred)})) \tag{10a}$$

In contrast to the biomass of prey species estimated from catch observations the unknown biomasses of other food plays a special role because no observations are available for estimation of this parameter. However, unique determination of $B(of,y,q)$ and $\rho(of,pred,q)$ requires that one parameter should be fixed. If we further assume that the biomass of other food is assumed to be constant over years and quarters ($B(of) = B(of,y,q)$) we choose to fix *B*(*of*) to a value for which $\rho(of,pred,q)$ are close to one.

Still a number of ties between parameters exist requiring specification of some parameters unique estimation. The fraction on the right hand side of equation (7) can be written as

$$\frac{B(py,l_{py},y,q)\rho(py,pd,q)g(py,pd,y,q)}{\sum\limits_{preu,l_{py}} B(py,l_{py},y,q)\rho(py,pd,q)g(py,pd,y,q) + B(of,y,q)\rho(of,pd,q)\exp(-\alpha(pd)(l-l_{\min(pd)}))} \tag{11}$$

where pd is $pred$ and py is $prey$.

Equation (11) indicates tat for each predator and for given values of biomasses,
$B(prey,l_{prey},y,q)\mathrm{and}B(of,pred,y,q)$ there is a tie between vulnerability parameters, $\rho(prey,pred,q)$. We choose for each predator in the first quarter that $\rho(of,pred,q=1)=1$.

For given values of biomasses the parameters in the food preference model now include the relative vulnerability parameters, $\rho(prey,pred,q)$ (other food not included), $\eta(pred)$ and $\sigma(pred)$.

The size dependent model for the suitable part of other food considered, $\exp(-\alpha(pred)(l-l_{\min(pred)}))$, has been chosen because the observations indicate that other food is less suitable for larger predators. In general the model means that the length of the predators are greater than the predator length corresponding to the preferred predator/prey size.)

As in Gislason and Helgason (1985) the age-structured predation mortality is determined by

$$M2(prey,a,y,q) = \sum_{pred,b} \frac{\bar{N}(pred,b,y,q)Food(pred,b,y,q)SUIT(a_{prey}(y,q),b_{pred}(y,q))}{AVAIL(pred,b,y,q)} \tag{13}$$

where

$$\begin{aligned} AVAIL(pred,b,y,q) = &\sum_{prey,a} \bar{N}(prey,a,y,q)w((prey,a,y,q)SUIT(a_{prey}(y,q),b_{pred}(y,q)) \\ &+OF(pred,b,y,q)SUIT(OF,pred,b,y,q) \end{aligned} \tag{14}$$

The age based suitability parameters in equations (12) and (13) are estimated using the length based formulations, (8) and (9):

$$SUIT(a_{prey}(y,q), b_{pred}(y,q)) = \rho(prey, pred, q) \exp(-\frac{(\ln \frac{\bar{\mathrm{l}}_{\mathrm{pred}}(b_{pred}(y,q))}{\bar{\mathrm{l}}_{\mathrm{prey}}(a_{prey}(y,q))} - \eta(pred))^2}{2\sigma^2_{pref}(pred)})$$

$$SUIT(of, pred, \bar{l}_{pred}(y,q)) = \rho(of, pred, q) \exp(\alpha(pred, q)\bar{l}_{pred}(b(y,q)))$$

where $\bar{l}_{pred}(b_{pred}(y,q))$ and $\bar{l}_{prey}(a_{prey}(y,q))$ denote the mean length of the age groups. As the suitability parameters appear in the likelihood functions referring to both catch-at-age and stomach content observations the estimated parameters are affected of both sources of information.

Food intake may taken as input from the literature or may be modelled by:

$$Food(pred, b, y, q) = \alpha(pred, q) weight(pred, b, y, q)^{\beta(pred)} \tag{15}$$

Analyses of the variation diet of North Sea predators, (Vinther and Lewy, 2002) indicate that the variance may formulated as

$$\begin{aligned} &VAR(STOM(prey, l_{prey}, pred, l_{pred}, y, q) = \\ &\frac{1}{V_{pred}U_{pred,l_{pred},q}} E(STOM(prey, l_{prey}, pred, l_{pred}, y, q))(1 - E(STOM(prey, l_{prey}, pred, l_{pred}, y, q)) \end{aligned} \tag{16}$$

where $U_{pred,l_{pred},q}$ is the number of hauls containing the predator/length in the stomach sampling programme and V_{pred} are parameters.

Using the Dirichlet distribution as distribution for the stomach diet in connection with the structured described by the equations (7)-(10) and (16) this can be done with the following parameterisation of the parameters in the (Dirichlet) distribution:

Assume for a given predator, *pred/l*$_{pred}$ year and quarter that the observed diet, $STOM(prey, l_{prey}, pred, l_{pred}, y, q)$ where $\sum\limits_{prey,l_{prey}} STOM(prey, l_{prey}, pred, l_{pred}, y, q) = 1$ follows a Dirichlet distribution with parameters $p_{prey,l_{prey}}$ with density function

$$\begin{aligned} &f_{pred,l_{pred}}(y,q) = f((STOM(prey, l_{prey}, pred, l_{pred}, y, q))_{prey,l_{prey}} | (p_{prey,l_{prey}})_{prey,l_{prey}}) \\ &= \frac{\Gamma(p)}{\prod\limits_{prey,l_{prey}} \Gamma(p_{prey,l_{prey}})} \prod_{prey,l_{prey}} STOM(prey, l_{prey}, pred, l_{pred}, y, q)^{p_{prey,l_{prey}}-1} \end{aligned} \tag{17}$$

where

$$p = \sum_{prey,l_{prey}} p_{prey,l_{prey}}$$

As the predation models are expressed through the mean values of each of the prey/prey length combinations (equation (7)) these values have to replace the parameters $p_{prey,l_{prey}}$. This is done in the following way:

The mean and the variance of the Dirichlet distribution, (17), is

$$E_{prey,l_{prey}} = E(STOM(prey,l_{prey},pred,l_{pred},y,q)) = \frac{p_{prey,l_{prey}}}{p}$$

(18)

$$VAR_{prey,l_{prey}} = \frac{1}{p+1} E_{prey,l_{prey}} (1 - E_{prey,l_{prey}})$$

(19)

Equation (18) implies that we replace the parameters $p_{prey,l_{prey}}$ by $pE_{prey,l_{prey}}$. Let *m* denote the number of prey/length groups. As $\sum_{prey,l_{prey}=1}^{m} E_{prey,l_{prey}} = 1$ the *m* parameters $p_{prey,l_{prey}}$ have been replaced by the *m* parameters $(p, E_{prey,l_{prey}=1}, ..., E_{prey,l_{prey}=m-1})$. Finally we do not use the many mean values $E_{prey,l_{prey}}$ as parameters but replaces them by the food preference parameters, $\rho(prey,pred)$, $\eta(pred)$, $\sigma_{pref}(pred)$ and $\bar{N}(prey,a,y,q)$.

The variance is modelled by using equations (19) and (16), which implies that

$p = V_{pred}U_{pred,l_{pred},q} - 1$

such that the parameter, *p*, is replaced by the parameters, V_{pred}.

Assuming that the diet observations are independent for the predator/length groups the likelihood function for including all predators/length groups are:

$$L_{STOM} = \prod_{pred,l_{pred},y,q} f_{pred,l_{pred}}(y,q)$$

(20)

In the case of using a multivariate normal distribution, it is assumed that the correlation between prey/length groups are known for each predator/length group, such that the variance/covariance matrices consist of variances as expressed in equation (19) and the known correlations. The multivariate normal distribution used is the just the product of multivariate normal distributions with mean as formulated by equation (7) and the variance/covariance as described.

The covariances used are the values obtained from the bootstrapping procedure described by Vinther and Lewy (2002).

The likelihood function, L_{STOM}, for the relative stomach contents observations now can be expressed by the standard density function of the multivariate normal distribution:

$$\begin{aligned}&L_{STOM} \cong \prod_{pred,l_{pred},y,q} |A(pred,l_{pred},y,q)|^{1/2} \times \\ &\exp(-0.5RES'(pred,l_{pred},y,q)A(pred,l_{pred},y,q)RES(pred,l_{pred},y,q)\end{aligned}$$

where the positive definite matrix *A*=*D* $^{-1}$ and *D* is the variance/covariance matrix of the observations, $\ln STOM(pred,l_{pred},y,q)_{pred,l_{pred}}$, where *RES* $'$ denotes the transposed vector

$$RES(pred,l_{pred},y,q) = \ln STOM(pred,l_{pred},y,q) - E(\ln STOM(pred,l_{pred},y,q)$$

where

$$\ln STOM(pred, l_{pred}, y, q) = \begin{pmatrix} \ln STOM(prey1, l^{1}_{prey_1}, pred, l_{pred}, y, q) \\ . \\ \ln STOM(prey1, l^{n_1}_{prey_1}, pred, l_{pred}, y, q) \\ . \\ . \\ \\ \ln STOM(prey\mathrm{m}, l^{1}_{prey_m}, pred, l_{pred}, y, q) \\ . \\ \ln STOM(prey\mathrm{m}, l^{n_m}_{prey_m}, pred, l_{pred}, y, q) \end{pmatrix}$$

$$E(\ln STOM(pred, l_{pred}, y, q)) = \begin{pmatrix} E(\ln STOM(prey1, l^{1}_{prey_1}, pred, l_{pred}, y, q)) \\ . \\ E(\ln STOM(prey1, l^{n_1}_{prey_1}, pred, l_{pred}, y, q)) \\ . \\ . \\ \\ E(\ln STOM(prey\mathrm{m}, l^{1}_{prey_m}, pred, l_{pred}, y, q)) \\ . \\ E(\ln STOM(prey\mathrm{m}, l^{n_m}_{prey_m}, pred, l_{pred}, y, q)) \end{pmatrix}$$

and where n_i denotes the number of length groups of prey species *i*.

Stock-recruitment relationship

Stock-recruitment is modelled using a parameterised model. Assuming recruitment is lognormal distributed the parameters for species *s*, $\beta(s)$ and $\gamma(s)$, are estimated by multiplying the likelihood function with a penalty function, l_R derived from

$R(s, y) = g(SSB(s, y), \beta(s), \gamma(s)) \exp(\sigma(s)\varepsilon(y))$

where *R* denotes the recruits, *s* denotes the species, *g* the stock-recruitment relation chosen, $\beta \mathrm{and} \gamma$ S-R parameters, σ the standard deviation and ε a standardized normal distribution.

Total likelihood function

The total likelihood function, *L*, then becomes

$L = L_C * L_{STOM} * L_{S-R}$

The parameters in the model are:

$N(s, a = RS(s) + i * NOS, y = 1, q = 1), i = 1, ..., A(s)$
$N(s, a = 0, y, q = RS(s)), y = 1, ..., Y$

M1(s,a)
$F_1(s,a), F_2(s,y), F_3(s,q)$
$\rho(prey, pred)$
$\eta(pred)$
$\sigma^2_{catch}(s)$
$\sigma^2_{pref}(pred)$

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.

Lewy 200?

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

Vinther, M., and Lewy, P. 2002. Analysis of the variation of the diet of North Sea demersal roundfish species. In prep.

G.2 Analysis of the variation of the diet of North Sea demersal roundfish species.

Morten Vinther and Peter Lewy
Draft January 2003

Introduction:

Modeling of stochastic population dynamics including predation requires modeling of food selection and data containing information on these processes. Stomach contents data describing the diet of the predators are important data in this connection. In stochastic population models stomach contents data should be used together with other data like catch per unit effort and total catch at age over a period. Both data sources contribute to estimating the biomass of predators and prey species and other parameters. As a consequence it is important to find probability distributions describing the variation of predators diet enabling formulation of the simultaneous likelihood function for estimation of parameters. If the variance/covariance can be estimated in the model the weighting problem with the different data sources thereby will be solved.

Very few studies of feeding behaviour include precision of the estimated diet (Ferry and Cailliet, 1996) investigations of the statistical distributional aspects of relative stomach contents have been carried out. Tirasin and Jørgensen (1999) have investigated the precision of estimated species composition in the diet of mackerel in the North Sea using the normal distribution as approximation. In a review study Cortés (1996) investigates whether weight proportions or normalizing transformations of prey in stomachs of lemon sharks follows normal distributions and concludes that the distributions significantly deviates from normality. Stefánsson and Pálsson (1997) analyzed total stomach contents of Icelandic cod and applying the so-called delta-gamma distribution to describe its distribution.

The use of stomach data in fish population models including species interactions have been limited (the MSVPA (ICES, 1997; ICES 2002), Fleksibest (Frøysa et al. 2002). As far as the authors know the only attempt to apply stomach contents data in a stochastic model is in Bormicon (Stefánsson and Pálsson, 1997) where the weight proportions of prey species in the stomach are assumed to follow independent normal distributions. This approach ignores possible correlation between prey species and the problem that some prey species sometimes are and sometimes are not included in the diet.

The purpose of the present work is to analyze the relative species composition by weight in the stomachs of predator species and identify multivariate probability distributions describing the data. The analyses will focus on the variance structure regarding factors as predator/predator size, prey/prey size and seasonality. Possible correlation between prey species will be investigated as well. The mean of prey species composition for a give predator, which depends on the food preference of the predator, will not be analyzed here but will treated in the context of a multi-species model simultaneously describing stock dynamics and food selection.

Observed relative stomach contents by weight, i.e. the weight proportion of prey items, have been considered as basis for modeling food preference instead of the absolute stomach weight. The reason is that twe find the functional response curve connected with relative weights in the stomach more realistic. (The functional response corresponding to relative weights is steepest at low prey densities and increasing to a maximum while the response for absolute weigtsis a simple linear increasing model).

Data from the International Council for the Exploration of the Sea (ICES) North Sea Stomach Sampling Project (ICES 1996**)** will be analyzed. Combining the stomach data with survey data from IBTS providing information on the spatial distribution of the predators estimates of the overall North Sea diet of a predator will be estimated. The probability distributions of the overall diet of cod, haddock, whiting and saithe in the North Sea will be studied by the bootstrapping. Various distribution functions and transformations will be considered.

Methods and material

An international stomach sampling program was initiated in 1981 to collect stomach contents data from economical important piscivorous fish species in the North Sea. The sampling program was under the auspices of ICES with the purpose to collect data on "who eats who" of the exploited fish in the North Sea for use in fish stock assessment. Stomachs were sampled from saithe (*Pollachius virens*), cod (*Gadus morhua*), haddock *(Melanogrammus aeglefinus*), whiting (*Merlangius merlangus*) and mackerel (*Scomber scombrus*). Stomach sampling continued in the period 1981 to 1991 with the highest sampling intensity in 1981 and 1991. Further information on the background for the ICES stomach sampling project are given in Daan (1989) **REF** and *Anon.* (1997) **REF**. In this paper we use stomach data from 1991 for cod, haddock, saithe and whiting.

Sampling activity given by the annual number of hauls and number of stomachs sampled is presented in Table 1 while Fig. 2 correspondingly gives the seasonal distribution by predator species and length classes.

Compilation of stomach contents data

For 1991 quarterly surveys have been conducted, such that samples of the species composition in the stomachs are given by year, quarter, predator, predator length/age, prey and prey length/age. The diet of a predator could be analyzed in different levels of spatial and temporal aggregations. In the present paper it has been chosen to consider quarterly averages of the diet of the North Sea population because these stomach data are used as observations in the stochastic multi-species population model. The compilation of the individual stomach samples from a trawl haul into average diet of the North Sea follows the technique given by *Anon.* (1997) and is briefly described below. Most stomachs have been pooled separately for each of the predator length groups considered.

For each haul the stomach samples for a given species and length class include the information on the number of a) empty stomachs; b) stomach with skeleton remains only; c) stomach with food and d) stomach with food, but regurgitated. Only stomach contents from the feeding, non-regurgitated stomachs were recorded and later bulked to save time. In the calculation of the average stomach content, it was assumed that the regurgitated stomachs had similar stomach content as the (valid) feeding fish.

First the average stomach content per ICES roundfish area is calculated using data for the ICES rectangles available. 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 weights. 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 ICES square weighted by the square root of the arithmetic mean of the observed CPUEs within a rectangle.

Partly digested prey items are in some cases not fully identified to species level or size class. In such cases a species or size redistribution of unidentified items was made accordingly to the observed diet (see *Anon.*, 1997 for details).

The length based observations were optionally transformed into age-groups using an age-length-key (ALK) given by quarter and roundfish area. The ALKs were derived from quarterly surveys or alternatively from commercial catches.

For a given predator the average North Sea stomach contents by quarter were finally calculated as a weighted mean of the average stomach contents by roundfish area. The quarterly proportions of the stock in the roundfish areas of the total North Sea stock of a given predator were used as weighting factors. The spatial distribution of the predators and age length keys by roundfish area were derived from quarterly surveys or commercial catches.

Analyzing the probability distribution of the overall North Sea diet

An analytical analysis of the distribution of the average diet would be difficult to carry out because of the complex data sources and imperfect data handling (e.g. prey items without length information or partly identified prey items). Instead the bootstrapping method (Efron and Tibshirani, 1993) was used.

5 predator and 7 prey species divided into length groups were included in the analysis. Besides the 7 prey species an "other food" group was considered consisting of the remaining other species lumped together. Items from the "other food" group were separated into 18 taxonomic prey species groups without considering length.

Bootstrapping was done separately for each stratum defined as the combination of quarter, roundfish area and predator length class. Simple Monte Carlo sampling with replacement was used to draw the individual samples (hauls) within the stratum. The number of resamples by stratum is the same as for the original data. For the bootstrapped data the quarterly average of the North Sea population species composition was calculated for using the method outlined above. In addition, the same procedure was used to calculate stomach contents given by age group of predator and preys.

The whole procedure was repeated 1000 times giving 1000 replicates of the North Sea population diets from which the histogram of the probability distribution, the mean, variance and covariance of individual prey groups were derived for the different predator/length groups.

The Kolmogorov-Smirnov D-statistic was calculated for the bootstrapped data for individual prey groups in order to classify the distributions into symmetric/non-symmetric distributions. Instead for a given predator/length and prey/length group the number the proportion of hauls for which the prey/length is not present in a haul was calculated.

Model for variance

General Linear Models (GLM) were used to analyze the bootstrap estimates of the variance of the prey proportions. The analysis was done assuming that variance structure was the same as for the Dirichlet distribution even that the Dirichlet distribution not was used to describe the relative diet composition of predators. This assumption was based on visual inspection of plots indicating a linear relation between log CV and logit mean of prey proportions. This relation can be justified as the variance of a prey proportion in the Dirichlet distribution equals

$$VAR(preyproportion_j) = \alpha_1 E(1-E)$$

where α_1 is a parameter (the sum of parameters in the distribution) and *E* is the mean of the prey proportion. This relation implies that

$$\ln(CV(proportion)) = \alpha - 0.5\text{logit}(E)$$

where *CV* is the coefficient of variation, α is a parameter and logit is the function $\text{logit}(E) = \ln(E/(1-E))$. The last formula indicates that there is linear relation between the logarithm of *CV* of a proportion and the logit of the mean proportion. Thus the model

$$\ln(CV) = \alpha + \beta\text{logit}(E)\beta \leq 0$$

was used as basis for modeling the variance of the mean prey proportions. The intercept was assumed to depend on effects of predator species (*pred*), the predator length (pred *l*), the prey species (*prey*) and the quarter (*q*). Furthermore, the sampling level was included in the model using either the number of hauls for which a given predator occurred in the samples, $n_{pred,l,q}$, or the number of hauls for which a specific prey was found in a predator, $o_{pred,l,prey,q}$. Finally, the prey mean weight (W) was included in the model in order to account for the prey size.

Thus the model investigated is given by:

$$\ln\left(CV_{pred,predl,prey,preyl,q}\right) = \alpha + \alpha_{pred} + \alpha_{predl} + \alpha_{prey} + \alpha_q + \gamma \ln(n_{pred,predl,q}) + \\ \phi \ln(o_{pred,l,prey,q}) + \varphi \ln(W_{pred,l,prey,preyl,q}) + \beta \text{logit}(E_{pred,predl,prey,preyl,q})$$

where

$\gamma \leq 0 \qquad \phi \leq 0 \qquad \beta \leq 0$

As the number of hauls, $n_{predl,q}$, and the number of hauls for which a specific prey was actually found, $o_{pred,l,prey,q}$ are highly correlated, only $n_{predl,q}$ was included.

Non-significant parameters were left out of the model assuming a 5% level of significance. The models were further reduced by considering the contribution of the various effects to total variance.

Model for covariance/correlation.

For a given quarter consider a given predator and predator length.

Let n denote the number of hauls for including he predator

Let $(h_{i,1}, ..., h_{i,m})$ denote species proportion in haul *i* of *m* species in the stomachs. As $\sum_{j=1}^{m} h_{ij} = 1$ it is sufficient to consider the distribution of for instance the first *m*-1 variable, $(h_{i1}, h_{i2}, ..., h_{im-1})$.

Some stomachs may not contain a specified prey species included in other stomachs. This means that some h_{ij} may be zero.

Let $I_j = \begin{cases} 0 \text{if} h_j = 0 \\ 1 \text{if} h_j > 0 \end{cases}$

where *j* denotes the prey species.

Let $n(I_1, ..., I_{m-1})$ denote the number of hauls for which the event $(I_1, ..., I_{m-1})$.

We have that $\sum_{(I_1,...,I_{\text{Im}-1})} n(I_1, ..., I_{m-1}) = n$

For a given *n* the set of $((n(I_1, ..., I_{m-1}))|n)$ is multinomially distributed.

Let $P_j = P(I_j = 0), j = 1, ..., m-1$ denote the probability that species *j* is not found in the stomach. $P_j = \sum_{i=1}^{n} P(I_{i1}, ..., I_{ij-1} I_j = 0, I_{ij+1}, ... I_{im-1})$.

The Delta-Dirichlet distribution

The multivariate Delta distribution

This distribution consists of a multivariate lognormal distribution with singularities at 0.

Let

$z_{ij} = \ln(h_{ij}|h_{ij}{>}0)\text{for} i = 1, ..., n j = 1, ..., m - 1$

$\mu_j = E(\ln(h_{ij}|h_{ij}{>}0))\text{for} i = 1, ..., n j = 1, ..., m - 1$

$c_{jk} = COV(z_{ij}, z_{ik})\text{for} i = 1, ..., n j = 1, ..., m - 1 k = 1, ..., m - 1$

$\alpha_j = Median(h_{ij}) = (1 - P_j)\exp(\mu_j)\text{for} i = 1, ..., n j = 1, ..., m - 1$

r_i denotes the number of prey species included in haul *i*, i.e $r_i = r(I_{i1}, ...I_{im-1}) = \sum_{j=1}^{m-1} I_{ij}$

Z_i denotes the log observation vector of dimension r_i which consists of the positive elements of $(z_{i1}I_{i1}, ..., z_{im-1}I_{im-1})$

M_i denotes the vector of dimension r_i which consists of the positive elements of $(\mu_1 I_{i1}, ..., \mu_{m-1}I_{im-1})$.

Σ denotes the variance/covariance matrix, $(_{jk})_{j,k}$, then define

Σ_i as the square matrix of dimension r_i for which the zero rows and columns of the dispersion matrix, Σ been have removed.

$\mu_j = \ln(\alpha_j) - \ln(1 - P_j)\text{for} i = 1, ..., n j = 1, ..., m - 1$

Let Q_i and R_i denote the vectors of dimension r_i which consist of the positive elements of $(\ln(\alpha_1)I_{i1}, ..., \ln(\alpha_{m-1})I_{im-1})$ and $(\ln(1-P_1)I_{i1}, ..., \ln(1-P_{m-1})I_{im-1})$ respectively. We thus have that

$M_i = Q_i - R_i$

The likelihood function is

$$\begin{aligned} L &= \begin{pmatrix} n \\ \\ (n(I_1, ..., I_{m-1})) \end{pmatrix} \times (2\pi)^{-n/2} \times \\ & \prod_{(I_1,...,I_m)} \left\{ P(I_1, ..., I_{m-1})^{n(I_1,...,I_{m-1})} \times |\Sigma_i|^{-n(I_1,...,I_{m-1})/2} \right\} \times \\ & \exp\left(-0.5 \sum_i \left\{ (Z_i - Q_i + R_i)^T \Sigma_i^{-1} (Z_i - Q_i + R_i) \right\} \right) \end{aligned}$$

If we parameterize by substituting M_i with Q_i and R_i minus log likelihood becomes:

$$\begin{aligned} l &\cong -\ln \begin{pmatrix} n \\ \\ (n(I_1, ..., I_{m-1})) \end{pmatrix} + n/2 \ln(2\pi) - \\ & \sum_{(I_1,...,I_m)} \{ n(I_1, ..., I_{m-1})(\ln(P(I_1, ..., I_{m-1})) - 0.5 \ln(|\Sigma_i|)) \} + \\ & 0.5 \sum_i \left\{ (Z_i - Q_i + R_i)^T \Sigma_i^{-1} (Z_i - Q_i + R_i) \right\} \end{aligned}$$

Finally we insert that

$$\begin{aligned}\alpha_j &= \alpha(prey, l_{prey}, pred, l_{pred}, y, q) \\ &= \frac{\bar{N}(prey, l_{prey}, y, q) w(l_{prey}(y,q))(SUIT(l_{prey}(y,q), l_{pred}(y,q))}{\sum\limits_{prey, l_{prey}} \bar{N}(prey, l_{prey}, y, q) w(l_{prey}(y,q)) SUIT(l_{prey}(y,q), l_{pred}(y,q))}\end{aligned}$$

where prey species/length group j corresponds to prey/length/predator/length group.

The variances c_{jj} should include parameters to be estimated.

Results

Selected histograms for the distributions of stomach contents proportions of the bootstrap replicates are given in Figure 2. The prey items for this figure were selected to illustrate the various types of the distributions occurring. It is clear from the examples that the distribution of replicates can be far from the normal distribution, and there is a clear link between the frequency of hauls a prey is occurring in the stomach and the shape of the distribution. Very few observations may give rise to delta- (distributions with positive probability typically of a the zero value) or bimodal distributions while more observations result in unimodal and more and less symmetric distributions. The occurrence of the delta distributions may be interpreted as missing spatial overlap between predator and prey.

The D-statistics from the Kolmogorov-Smirnov test of data against a normal distribution with mean and variance equal to the sample mean and variance is included in Figure 3 as well to give an impression of the values in relation to the shape of the distribution. Roughly speaking, a D-value below 0.1 indicates a unimodal and relatively symmetric distribution. Approximately 40% of all prey types belong to that category (Figure 3).

Even though many of the distributions of the bootstrap replicates are far from normal we simple the use the mean and empirical variance of the replicates.

The bootstrap estimates of prey mean proportion and coefficient of variation is shown for cod in Figure 4. In general CV decreases with increasing proportion of a prey. For preys contribution less than 5% of the diet the CV is in the range 10-150%. Cod as predator has been shown as an example, but a similar picture is found for the other predators. The average CVs are 61-65% for cod, haddock and whiting and higher (87%) for saithe.

In the GLM for variance of prey items the parameters predator, predator-length, quarter and mean proportion are highly significant in the GLMs where number of hauls with a specific prey as parameter (Table 2). Prey species and prey mean weight are, however, not significant. The model degree of freedom can be reduced substantially without a serious decrease in R-square by excluding the parameters predator*length and quarter. The simplest model including just the frequency of hauls with a specific prey has a surprisingly high R-square at 0.49.

The family of models where number of hauls is used as proxy for sampling level (Table 3) has in general a lower R-square and a different parameter selection than the family of models including the frequency of hauls with a specific prey species. With a model degree of freedom at five the GLMs (Table 3) include the parameters prey species, mean proportion and number of hauls as parameters. A further reduction in parameters will first exclude the prey species parameter and afterwards the number of haul parameter, leaving the mean proportion parameters as the single most important parameter.

Preferred model:

$$\ln(CV_{pred,predl,prey,preyl,q}) = \alpha_1 + \gamma \ln(n_{pred,preyl,q}) + \beta \text{logit}(m_{pred,predl,prey,preyl,q}) \quad (1)$$

where

$\gamma \leq 0 \qquad \beta \leq 0$

Equation (1) corresponds to that

$$CV^2_{pred,predl,prey,preyl,q} = \alpha(\tfrac{m_{pred,predl,prey,preyl,q}}{1-m_{pred,predl,prey,preyl,q}})^{\beta} n^{\gamma} \quad \alpha \geq 0$$

Now assume that the proportion in the stomach, *proportion*, is log normal distributed, i.e.

$$\ln(proportion) \sim N(\mu, \sigma^2)$$

We then know that

$$\begin{aligned} \sigma^2_{pred,predl,prey,preyl,q} &= \ln(1 + CV^2(proportion_{pred,predl,prey,preyl,q})) \\ &= \ln(1 + \alpha(\frac{m_{pred,predl,prey,preyl,q}}{1 - m_{pred,predl,prey,preyl,q}})^{\beta} n^{\gamma}) \end{aligned}$$

where

$\alpha \geq 0 \beta \leq 0 \gamma \leq 0$

Fig Correlations (length and age groups) frequency histogram

Results: model for variance ∼mean (length)
- Advise on sampling level
- Advise on stratification?

Results: Model for covariance?

Discussion

Suitable distributions and transformations.
Dirichlet, binomial, log-normal
Arc-sinus, logit, log

Diet in stochastic model, our suggestion
- Review of existing model implementations

References

Cortés, E. 1997. A critical review of methods of studying fish feeding based on analysis of stomach contents: application to elasmobranch fishes. Can. J. Fish. Aquat. Sci. **54**: 726-738.

Efron, B., and Tibshirani, R.J. 1993. An introduction to the Bootstrap. Chapman & Hall, London.

Frøysa, K.G., Bogstad, B., and Skagen, D.W. 2002. Fleksibest - an age-length structured fish stock assessment model. Fish. Res. **55:** 87-101.

ICES. 1997. Report of the Multispecies Assessment Working Group. ICES CM 1997/Assess:16.

ICES. 1996. Database report of the Stomach Sampling Project 1991. Coop. Res. Rep. **219**: 1-442.

Rice, C.R. 1988. Repeated cluster analysis of stomach contents data: method and application to diet of cod in NAFO division 3L. Environ. Biol. Fishes. **21**:263-277.

Stefánsson, G., and Pálsson, Ó.K. 1997. Statistical evaluation and modeling of the stomach contents of Icelandic cod (*Gadus morhua*). Can. J. Fish. Aquat. Sci. **54**: 169-181.

Stefánsson, G., and Pálsson, Ó.K. 1997. BORMICON, A Boreal Migration and Consumption Model. Mar. Res. Inst. Rep. **58**: 1-223.

Tirasin, E.M., and Jørgensen, T. 1999. An evaluation of the precision of diet description. Mar. Ecol. Prog. Ser. **182**: 243-252.

Table 1. Number of samples and number of stomachs by species and quarter, 1991

	Cod		Haddock		Saithe		Whiting	
Quarter	no. haul	no. stom	no. haul	no. stom	no. haul	no. stom	no. haul	no. stom
1	296	2182	105	2442	69	798	190	6152
2	277	3174	188	2768	76	1191	306	11330
3	201	2373	208	4971	39	402	279	11543
4	176	1999	115	2927	53	858	216	9406
All	950	9728	616	13108	237	3249	991	38431

Table 2. Model statistics using number of hauls, where a given prey item was found, as proxy for sampling level.

Significant model parameters	Model DF	R-square
$\alpha + \beta^S_s + \beta^{SL}_{sl} + \beta^Q_q + \beta_1 \log\mathrm{it}(m_{sl,q,p}) + \beta_3 \log(N_{sl,q,p})$	43	0.61
$\alpha + \beta^S_s + \beta^Q_q + \beta_1 \log\mathrm{it}(m_{sl,q,p}) + \beta_3 \log(N_{sl,q,p})$	8	0.58
$\alpha + \beta^S_s + \beta_1 \log\mathrm{it}(m_{sl,q,p}) + \beta_3 \log(N_{sl,q,p})$	5	0.58
$\alpha + \beta_1 \log\mathrm{it}(m_{sl,q,p}) + \beta_3 \log(N_{sl,q,p})$	2	0.53
$\alpha + \beta_3 \log(N_{sl,q,p})$	1	0.49

Table 3. Model statistics using number of hauls as proxy for sampling level.

Significant model parameters	Model DF	R-square
$\alpha + \beta^S_s + \beta^{SL}_{sl} + \beta^Q_q + \beta^P_p + \beta_1 \log\mathrm{it}(m_{sl,q,p}) + \beta_2 \log(N_{sl,q}) + \beta_4 \log(w_{sl,q,p})$	51	0.50
$\alpha + \beta^S_s + \beta^Q_q + \beta^P_p + \beta_1 \log\mathrm{it}(m_{sl,q,p}) + \beta_2 \log(N_{sl,q})$	15	0.43
$\alpha + \beta^S_s + \beta^P_p + \beta_1 \log\mathrm{it}(m_{sl,q,p}) + \beta_2 \log(N_{sl,q})$	12	0.42
$\alpha + \beta^P_p + \beta_1 \log\mathrm{it}(m_{sl,q,p}) + \beta_2 \log(N_{sl,q})$	5	0.37
$\alpha + \beta_1 \mathrm{logit}(m_{sl,q,p}) + \beta_2 \log(N_{sl,q})$	2	0.34
$\alpha + \beta_1 \mathrm{logit}(m_{sl,q,p})$	1	0.21

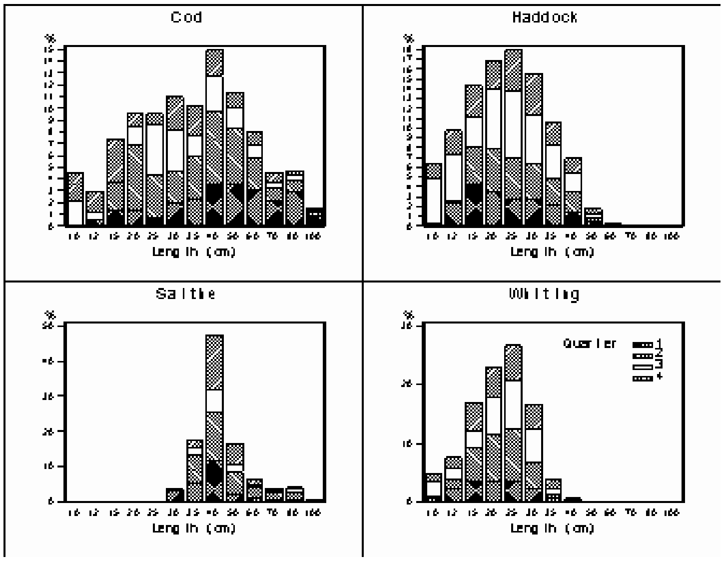


Figure 1: Relative number of sampled stomachs by quarter and length group(specified by lower cm range) and species.

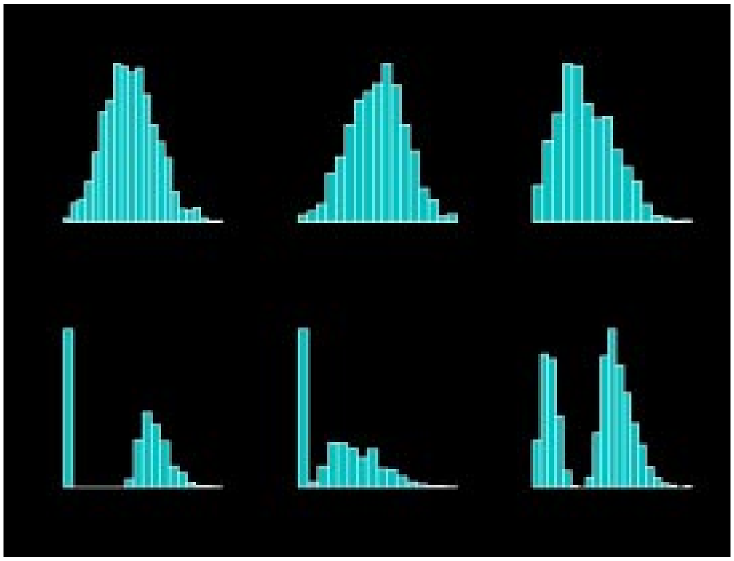


Figure 2: Distribution of bootstrap replicates of relative stomach contents for selected prey categories eaten by whiting 25-30 cm, 3 quarter of 1991. For each prey item, prey occurrence is given by number of hauls (h) and the number of ICES roundfish areas (a) where the item was found. The D value is Kolmogorov-Smirnov D-statistics

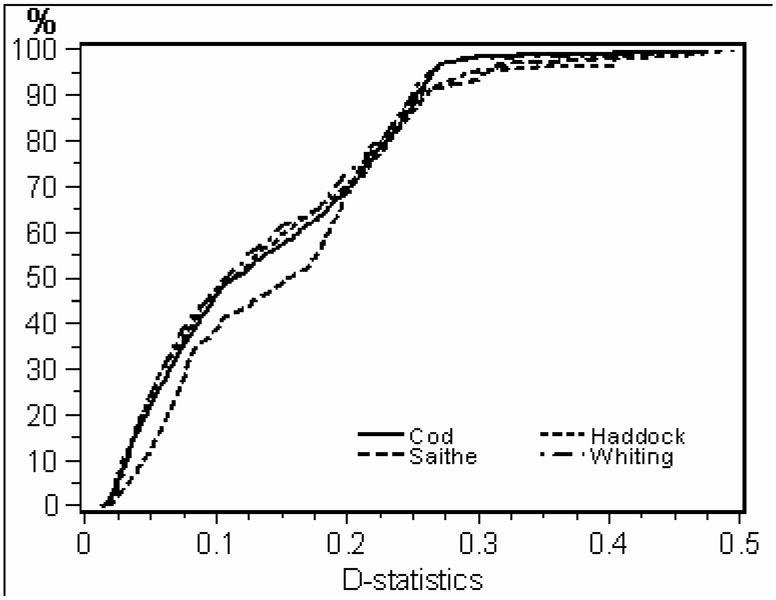


Figure 3: Cumulated number of prey item groups against Kolmogorov-Smirnov D-statistics by predator.

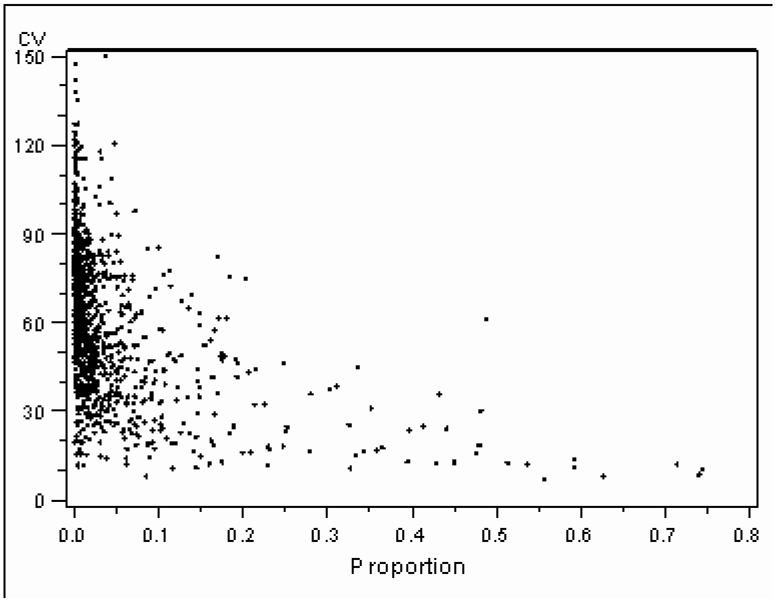


Figure 4: Mean proportion and Coefficient of Variation of preys calculated from of bootstrap replicates of cod stomach contents by length class and quarter.

G.3 Estimation of fish consumption by stomach content analysis

A. Rindorf* 2003

University of Copenhagen, c/o Danish Institute for Fisheries Research, Charlottenlund Castle, DK2920 Charlottenlund, Denmark.

Draft – not to be cited without prior reference to the author.

* Tel: +45 33 96 33 59; Fax: +45 33 96 33 33; email: ar@dfu.min.dk

Abstract

This study presents an analysis of the bias introduced by simplifying calculations of food intake from stomach contents. Typical simplifications include the adoption of the approximation suggested by Pennington (1985) rather than numerical solution of the underlying differential equation and the estimation of the average stomach content to a potency α as $E\left[S^{\alpha}\right] \cong \left(\frac{1}{N}\sum_{i=1}^{N} S_i\right)^{\alpha}$ where S is the contents of the ith stomach and N is the number of stomachs in the sample. The method suggested by Pennington (1985) for calculation of average food intake is found to produce results close to estimates derived when the underlying differential equation is approximated numerically. Serious positive bias was introduced by estimating $E\left[S^{\alpha}\right]$ as $\left(\frac{1}{N}\sum_{i=1}^{N} S_i\right)^{\alpha}$. In a data example, the positive bias was around 50% when $E\left[S^{\alpha}\right]$ was calculated in this way. A simple expression is suggested by which the bias can be corrected for. The correction requires the estimation of the mean and variance of the stomach contents as well as the underlying distribution.

When calculating the intake of separate prey types, a common procedure is to distribute the total intake on different prey types according to the weight percentage of each prey in the stomach content. The calculation is based on the implicit assumption that all prey types occur in all stomachs and are evacuated at the same rate. This method results in greatly biased estimates of intake. An alternative method is suggested which takes account of the uneven distribution of prey as well as the effect of other food in the stomach on evacuation. Simplifying the analyses by disregarding the change in evacuation induced by other food in the stomach results in moderate overestimation of the intake of all prey types.

Introduction

Food intake by fish is frequently estimated by combining knowledge of evacuation rates with field studies of stomach contents. Differing methods have been suggested, both depending on the shape of the relationship between stomach content and evacuation rate and on the assumption made on the progress of feeding between consecutive sampling points (Bajkov 1935, Eggers 1977, Elliot and Persson 1978, Pennington 1985, Sainsbury 1986, dos Santos and Jobling 1995, Hall et al. 1995). The evacuation of stomach contents of fish is generally well described by the relationship

$$\frac{dS}{dt} = -\rho S^{\alpha}$$

where ρ is the evacuation rate, S is stomach content and α is a constant (Jones 1974, Jobling 1981, Andersen 2001). If the predator ingested food at a constant rate in the interval between sampling points, a prerequisite to the estimation of food intake rate, C$_r$, is the solution of the differential equation

$$\frac{dS}{dt} = -\rho S^{\alpha} + C_r \tag{1}$$

This equation can be solved analytically for α=0 (linear evacuation) and α=1 (exponential evacuation) (Eggers 1977, Elliott and Persson 1978). However, for $\alpha \neq 0$ and $\alpha \neq 1$ the equation does not have an analytical solution. To estimate food intake in these cases, the expression must be approximated numerically. This is rarely done and usually food intake is estimated by methods resembling that suggested by Pennington (1985)(Jones 1978, Armstrong 1982, Patterson 1985, dos Santos and Jobling 1995) or even simpler methods (Bajkov 1935, Daan 1973, Lockwood 1980, Basimi and Grove 1985). The degree to which the estimates are affected by simplifying the calculations in this way has hitherto only been investigated for α=1 (Eggers 1977, Elliot and Persson 1978).

The estimate of food intake presented by Pennington (1985) calculates average hourly consumption between two sampling points as

$$C_s = \frac{\rho}{N} \sum_{i=1}^{N} E\left[S_i^{\alpha}\right] + \frac{1}{T}\left(E\left[S_T\right] - E\left[S_0\right]\right) \tag{2}$$

where N is the number of samples taken in the sampling interval t=0 to t=T, $E\left[S_0\right]$ and $E\left[S_T\right]$ is the average stomach content at times 0 and T, respectively and $E\left[S_i^{\alpha}\right] = \frac{1}{M_i} \sum_{j=1}^{M_i} S_{i,j}^{\alpha}$ where M$_i$ is the number of stomachs in the ith sample and S$_{i,j}$ is the contents of the jth stomach in the ith sample (Pennington 1985). This calculation requires an estimate of the average value of S_t^{α}. As sampling and analysing individual stomachs is labour intensive, S_t^{α} is usually estimated as $\left(E\left[S_t\right]\right)^{\alpha}$, where $E\left[S_t\right]$ denotes the average stomach content (Jones 1978, Armstrong 1982, Patterson 1985). However, as noted by several authors, this procedure introduces a bias as $E\left[S_t^{\alpha}\right] \neq \left(E\left[S_t\right]\right)^{\alpha}$ for $\alpha \neq 0$ and $\alpha \neq 1$ (Ursin et al. 1985, dos Santos and Jobling 1995, Andersen 2001).

Additional problems arise when several prey types are eaten and the amount eaten of each type has to be calculated. This may be done by distributing the total intake on different prey types according to the weight percentage of the prey in the total stomach content (Armstrong 1982, Hawkins et al. 1985, Seyhan and Grove 1998). The implicit assumption is that all prey occur in the average amount in all stomachs and that the evacuation rate of all prey is equal. These assumptions are rarely, if ever, met (dos Santos and Jobling 1995, Stefánsson and Pálsson 1997, Andersen 1999). Alternatively, the intake of each prey type has been calculated for each prey separately (Armstrong 1982, Patterson 1985). This allows evacuation rate to be varied between prey types. However, the method does not take the effect of other food in the stomach on evacuation into account (Jones 1974, Andersen 2001). Calculating the intake in this way is therefore based on the assumption that only one prey type is present in each stomach. This may hold in some cases (Magnússon and Aspelund 1997), but in many cases, each stomach contains several prey types (Rice 1988, Albert 1995). An exception to the assumption of separate evacuation is the equation given by dos Santos and Jobling (1995). However, their model is currently applicable only to cod and does not allow changes in the average stomach content over time.

In summary, an estimate of both total consumption and that of individual prey types should use an unbiased estimate of S_t^{α}. The difference between numerically integrating eq. 1 and the estimate derived by eq. 2 require further investigation and there is a need for an estimate of the intake of different prey types that does not require unwarranted assumptions on separate evacuation or the distribution of prey in the stomachs. Evacuation rates of different prey should be allowed to differ, the prey should not be assumed to occur in the stomachs where it is not present, and the effect of other food in the stomach on evacuation should be included.

This study presents a calculation of the difference between average hourly food intake estimated by numerical integration and the estimate suggested by Pennington (1985). The effect of the time elapsed between sampling points and evacuation rate on the difference between the two estimates is examined. Further, an analytical expression for calculating an unbiased estimate of $E\left[S_t^{\alpha}\right]$ from the mean and variance of the stomach contents is presented. The bias introduced by calculating the food intake of individual prey as a weight proportion of the total intake is examined as is the bias introduced by assuming no effect on evacuation of other prey in the stomach. The differences between the methods are exemplified by estimating food intake of whiting of 25 to 30 cm caught at 5 different locations in the North Sea (Rindorf 2002).

Methods

Comparison of numerical estimates of constant intake rate and the estimate derived by Pennington

The two estimates were compared by first using equation (2) and a known combination of S_0 and C_r to estimate S_T numerically, and then using equation (1) to estimate average consumption C_s. The relative difference between the two was then calculated as

$$\frac{C_s - C_r}{C_s}$$

This exercise was performed for all combinations of S_0 and C_r in the interval 0.05 g to 3.05 g. The coefficient α was fixed at 0.5, as this value is appropriate for a number of fish species (Jobling 1981, Andersen 2001). Evacuation rates of 0.04, 0.14 and 0.24 were used with a fixed time interval of 4 hours to examine the effect of evacuation rate. These correspond to evacuation rate of slowly evacuated prey in whiting stomachs at 8^oC and of rapidly evacuated prey at 16 ^{o}C. The length of the interval between consecutive samples were examined by fixing evacuation rate at 0.24 and estimating the difference for T=1, 4 and 10.

Calculation of $E\left[S_t^\alpha\right]$ from mean and variance of the stomach content

To correct the data for the inequality $E\left[S_t^\alpha\right] \neq \left(E\left[S_t\right]\right)^\alpha$, Andersen proposed a correction factor which describes the relative difference between the average square root of the stomach content and the square root of the average stomach content (Andersen 2001). This correction factor was calculated directly from a data set consisting of single saithe stomachs. However, if the distribution of the stomach content follows a known distribution, the correction factor can be calculated directly from the mean and variance of this distribution. The distribution of stomach contents in non-empty stomachs have often been shown to be highly skewed, and the log-normal- and gamma-distributions have been mentioned as appropriate for describing the distribution of non-empty stomachs (Stefánsson and Pálsson 1997). An estimate of the square root of the stomach content can be derived analytically for both distributions. The required parameters are the mean and variance of the contents of non-empty stomachs as well as the proportion of non-empty stomachs, p. If the contents of the non-empty stomachs are log-normal distributed, the analytical expression is simple. Let the weight of the contents of a non-empty stomach be described by X, where X is log-normal distributed with mean μ and variance σ^2 of the log stomach contents:

$X \in LN\left(\mu, \sigma^2\right)$

then

$X^\alpha \in LN\left(\alpha\mu, \alpha^2\sigma^2\right)$

The mean of X^α, $E\left[X^\alpha\right]$, can be calculated directly from the mean and variance of the log-normal distribution as

$E\left[X^\alpha\right] = e^{-\frac{\alpha^2\sigma^2}{2}} E\left[X\right]^\alpha$

Correcting for the proportion of empty stomachs, the mean of the of the contents of all stomachs in the potency α, $E\left[S^\alpha\right]$, becomes

$$E\left[S^\alpha\right] = pe^{-\frac{\alpha^2\sigma^2}{2}} E\left[X\right]^\alpha \tag{3}$$

The variance of S^α is

$V\left[X^\alpha\right] = e^{2\alpha\mu+\alpha^2\sigma^2}\left(e^{\alpha^2\sigma^2} - 1\right)$

$V\left[S^\alpha\right] = V\left[p\right] e^{-\alpha^2\sigma^2} E\left[X\right]^{2\alpha} + p^2\alpha^2\sigma^2 + V\left[p\right]\alpha^2\sigma^2$

If X is gamma distributed with parameters k and β where k is the shape parameter and β is the scale:

$X \in \Gamma\left(k, \beta\right)$

the mean of X^{α} is

$E\left[X^{\alpha}\right] = \frac{\Gamma(k+\alpha)}{\Gamma(k) k^{\alpha}} E\left[X\right]^{\alpha}$

and the variance of X^{α} is

$V\left[X^{\alpha}\right] = \frac{\beta^{2\alpha}}{\Gamma(k)} \left(\Gamma\left(k+2\alpha\right) - \frac{\left(\Gamma(k+\alpha)\right)^2}{\Gamma(k)}\right)$

$E\left[S^{\alpha}\right]$ is then

$$E\left[S^{\alpha}\right] = p \frac{\Gamma(k+\alpha)}{\Gamma(k) k^{\alpha}} E\left[X\right]^{\alpha} \tag{4}$$

and the variance of $E\left[S^{\alpha}\right]$ is

$$\begin{aligned} V\left[S^{\alpha}\right] &= V\left[p\right] \left(\frac{\Gamma(k+\alpha)}{\Gamma(k) k^{\alpha}}\right)^2 E\left[X\right]^{2\alpha} + p^2 \frac{\beta^{2\alpha}}{\Gamma(k)} \left(\Gamma(k+2\alpha) - \left(\frac{\Gamma(k+\alpha)}{\Gamma(k)}\right)^2\right) + \\ & \quad V\left[p\right] \frac{\beta^{2\alpha}}{\Gamma(k)} \left(\Gamma(k+2\alpha) - \left(\frac{\Gamma(k+\alpha)}{\Gamma(k)}\right)^2\right) \end{aligned}$$

The variance of the estimated consumption is readily calculated (Pennington 1985).

Estimating the intake of individual prey

The presence of other food in the stomach alters the evacuation of each particular prey (Jones 1974). The proportion of the evacuated material which consists of a particular prey is identical to the weight proportion of the prey in the stomach contents (Andersen 2001). The evacuation rate of the total stomach content lies between that of the prey and that of the other material in the stomach (Andersen 2001). In this study, it will be assumed that the evacuation rate of the total content can be estimated as the average of the evacuation rates of the separate prey weighted by the weight of each prey, though this may not be entirely correct (Andersen 2001). The differential equation describing the stomach content is then:

$\frac{dS}{dt} = -\pi_j \left(\rho_j \pi_j + \left(1 - \pi_j\right) \rho_{tot}\right) \left(S_{tot}\right)^{\alpha} + C_j$

where π_j denotes the weight proportion of prey species j and S$_j$ and S$_{tot}$ denotes the weight of prey j and the total weight of food in the stomach, respectively. The parameters all refer to stomachs where prey j is found as the prey is obviously not evacuated from stomachs where it is absent. ρ_j is the evacuation rate of the prey in question and ρ_{tot} is the evacuation rate of the total stomach content. As

$\pi_j = \frac{S_j}{S_{tot}}$

the expression can also be written as

$\frac{dS}{dt} = -\pi_j^{1-\alpha} \left(\pi_j \rho_j + \left(1 - \pi_j\right) \rho_{tot}\right) S_j^{\alpha} + C_j$

The expression $\pi_j^{1-\alpha} \left(\pi_j \rho_j + \left(1 - \pi_j\right) \rho_{tot}\right)$ can be replaced by a transformed evacuation rate, ρ_j’:

$\frac{dS}{dt} = -\rho_j' S_j^{\alpha} + C_j$

C$_j$ can then be estimated by eq. 2. π_{φ} is the proportion of the stomach contents which consist of j in the stomachs where j is found. The value of ${\pi_j}^{1-\alpha}$ should rightfully be estimated as the average of all observed values of $\pi_j^{1-\alpha}$ rather than from the average value of π_j. However, this difference was not taken into account in the present study and $\pi_j^{1-\alpha}$ was simply estimated from the average value of π_j.

When calculating the intake by eq. 2, the relative difference between the intake of prey j as calculated by the suggested method and by multiplying the weight percentage of j in the total stomach content, $\pi_{tot,j}$, to the total intake is

$$bias_{t,j}=\frac{\pi_{tot,j}}{\pi_j^{1-\alpha}}\frac{\sum\limits_{i=1}^{N}S_{tot,i}^{\alpha}}{\sum\limits_{i=1}^{N}S_{j,i}^{\alpha}}-1=\frac{\pi_{tot,j}}{\pi_j^{1-\alpha}}\frac{p_{tot}}{p_j}\frac{\frac{\Gamma(k_{tot}+\alpha)}{\Gamma(k_{tot})k_{tot}^{\alpha}}}{\frac{\Gamma(k_j+\alpha)}{\Gamma(k_j)k_j^{\alpha}}}\frac{\sum\limits_{i=1}^{N}E\left[X_{tot}\right]^{\alpha}}{\sum\limits_{i=1}^{N}E\left[X_j\right]^{\alpha}}-1$$

where $\pi_{tot,j}$ is the weight proportion of prey j in the contents of all stomachs and the subscript *tot* denotes the parameters relating to the total stomach content. It is assumed that $\pi_{tot,j}$ and π_j remains constant over the period, no change occurs in the average stomach content over the sampling period and ρ_j=ρ_{tot}=ρ. The values of p_j, etc. refer to the proportion of stomachs containing prey j etc. This bias is positive if the prey occurs in a lower proportion of the stomachs than food in general, if the shape parameter of the total stomach content exceeds that of the prey or if the mean weight of the prey in stomachs where the prey is present is greater than the mean weight of the total stomach content (holding all other parameters constant). Negative bias results from $\pi_{tot,j}$<$\pi_j^{1-\alpha}$ and as this holds true for all values of $\pi_{tot,j}$ and $\pi_j^{1-\alpha}$ between 0 and 1, positive bias can only be found when the positive effect of low occurrence or large mean weight of prey j exceeds the negative effect of the ratio between the π's. Negative bias may also occur when k_{tot}<k_j or when the average weight of the prey in the stomachs where the prey is present is less than the average weight of food in the stomachs. Assuming k_{tot}=k_j and inserting $E\left[X_{tot}\right]=\frac{p_j}{p_{tot}}\frac{E[X_j]}{\pi_{tot,j}}$, the bias can be reduced to

$$bias_{t,j}=\left(\frac{\pi_{tot,j}}{\pi_j}\frac{p_{tot}}{p_j}\right)^{1-\alpha}-1$$

The bias introduced by disregarding the effect of other food in the stomach on evacuation, *bias*$_{o,j}$, is

$$bias_{o,j}=\frac{\frac{\rho}{N}\sum\limits_{i=1}^{N}S_i^{\alpha}-\frac{\rho\pi_j^{1-\alpha}}{N}\sum\limits_{i=1}^{N}S_i^{\alpha}}{\frac{\rho\pi_j^{1-\alpha}}{N}\sum\limits_{i=1}^{N}S_i^{\alpha}}=\pi_j^{\alpha-1}-1$$

Again it is assumed that no change in the average stomach content occurs over the sampling period and ρ_j=ρ_{tot}=ρ.

The bias introduced by each of the simplifications (assuming $E\left[S_t^{\alpha}\right]$ to be equal to $\left(E\left[S_t\right]\right)^{\alpha}$, assuming prey to be present in all stomachs and assuming no effect on evacuation of other prey in the stomach) was calculated for α=0.5. To estimate the extent to which bias was introduced when the different methods were applied to real data, the food intake of whiting at five locations in the North Sea was estimated.

Data example

The data consisted of whiting of 25 to 30 cm length gathered around the clock at five locations in the North Sea. Trawl hauls were performed with four hour intervals for a total of 48 to 72 hours. The sampling procedure has been described in detail elsewhere as has the calculation of the occurrence and weight of different prey types (Rindorf, 2002, 2003, subm.). The total weight present in non-empty stomachs and the occurrence of non-empty stomachs were analysed by models similar to the analyses total weight and occurrence of the single prey described in Rindorf (2003, subm.). Briefly, the effect of time of day, location and sampling time on occurrence and mean weight of single prey and total stomach content was analysed. The weight of both single prey and total stomach content in stomachs were this exceeded zero was assumed to be gamma distributed. This procedure was used to minimize errors introduced by random variation. The average hourly food intake was calculated from the estimated average stomach content in non-empty stomachs, the estimated shape parameter, *k*, and the occurrence of non-empty stomachs, p.

The effect of the different methods for calculating of the intake of individual prey was investigated by estimating the intake of three different prey types. Benthic invertebrates were chosen as an example of a prey that is generally small, occurs frequently in the stomachs and has a high evacuation rate. Herring was chosen as an infrequent large prey with a low evacuation rate. Finally, crabs were selected to exemplify prey, which occurs frequently and posses a hard exoskeleton and thus is evacuated at a low rate. The calculation of the different estimates of food consumption is summarized in table 1.

Estimation of evacuation rates

The evacuation rate of the total stomach content was estimated as a weighted average of the evacuation rates of the individual prey. The average of the energy densities of all prey weighted by the weight percentage of the particular prey, E, was first calculated for prey without exoskeleton as suggested by Andersen (2001). Energy densities were assumed to be equivalent to values given by Andersen (2001) and Pedersen and Hislop (2001). The energy density of benthic invertebrates was assumed to be 3.7 kJ/g based on the evacuation rate given by Rindorf (subm.). The evacuation rate of the total stomach content at each location was then calculated as a weighted average of the evacuation rate of non-exoskeleton prey (calculated from the average energy density of these prey) and the evacuation rate of exo-skeleton prey (taken to be equal to the value given for brown shrimp in whiting of 27.5 cm by Andersen (2001)):
$\rho_{tot} = (0.0025\pi_{exo} + 0.152E^{-0.86}\pi_{endo})e^{0.078temp}$
where π_{exo} is the proportion of the stomach content which consists of exoskeleton prey at each location, π_{endo} is the proportion of the stomach content which consists of prey without an exoskeleton and *temp* is ambient temperature at the location. This is a modified version of the equation given by Andersen (2001) for whiting at 27.5 cm. The size and temperature specific evacuation rates, ρ_{LT}, of the three prey types were: benthic invertebrates: 0.00042 (Rindorf subm.), crabs: 0.00021 (value given for brown shrimp in Andersen (1999)) and herring: 0.00027 (derived from the relationship given by Andersen (2001) and an energy density of 6.1 kJ/g (Pedersen and Hislop 2001)). With a predator length of 27.5 cm, the evacuation rates become 0.050e$^{0.078temp}$, 0.025 e$^{0.078temp}$ and 0.032 e$^{0.078temp}$ for benthic invertebrates, herring and crabs, respectively. Evacuation rates at all locations are given in table 2.

Estimation of $\pi_{tot,j}$ and π_j

The percentage of the total stomach contents which consisted of each prey, $\pi_{tot,j}$, was calculated at each location and time of day. In addition, the percentage of the stomach content which was made up by prey j in the stomachs where prey j occurred, π_j, was calculated for each prey, location and time of day.

Results

Comparison of intake based on numerical approximation and Penningtons method

The intake estimates are generally similar for short intervals. However, increasing the time interval leads to dramatic differences when the stomach content or intake is low (Fig. 1). The difference rose to 40% at a 10-hour sampling interval and a very low food intake rate. The difference between the two estimates increase with evacuation rate, being virtually zero for an evacuation rate of 0.04 rising to up to a maximum of 10% during a four hour interval at an evacuation rate of 0.24 (Fig. 2). For the large majority of combinations of C_r and S_0, the difference is less than 5% if the sampling interval is 4 hours or less.

Difference between $E\left[\sqrt{X}\right]$ and $\sqrt{E\left[X\right]}$
The difference between average square root and square root of the average increases dramatically as the variance of the log-normal distributed data is increased (Fig. 3). For gamma distributed variables, the difference is small when k>1 and empty stomachs are excluded from calculation of the average stomach content. Failing to separate empty stomachs from stomachs containing food increases the difference between $E\left[\sqrt{X}\right]$ and $\sqrt{E\left[X\right]}$. The consumption calculated from the whiting data was estimated to be 51% higher, on average, when assuming $E\left[\sqrt{X}\right]$ to be equal to $\sqrt{E\left[X\right]}$ than was the case when $E\left[\sqrt{X}\right]$ was corrected for the inequality between the two (Table 3). The shape parameter of the distribution of the total stomach contents was 0.269 and the proportion of stomachs which contained food was around 0.8 (Table 3).

Effect of distributing the total intake on prey type according to weight percentage

The estimates derived by distributing the total intake on prey types according to weight percentage are seriously negatively biased if the prey does not occur in all stomachs (Fig. 4). The area in which the negative bias is less than 10% constitutes a very small part of the possible parameter combinations. If the prey is found in most of the stomachs where food is found, the bias is generally negative. As the occurrence of the prey in the stomachs which contain food decrease, positive bias becomes more likely. The data example demonstrates this as the estimate C_5 is lower than C_4 when the prey occurs frequently in the stomachs (Table 4). The estimate C_5 exceeds C_4 at low occurrences, in particular where the prey constituted less than 3 quarters of the stomach contents in the stomachs where it occurred. C_5 was estimated at 64% and 56% of C_4 on average for benthic invertebrates and crabs, respectively, when intake exceeded 3 mg/h. Herring estimated by C_5 was positively biased at location 3 though the occurrence was similar to that at location where a slight negative bias was found. This was presumably caused by the high mean weight of herring at location 3.

Effect of disregarding other food in the stomach

Failing to include the amount of other food in the stomach when calculating evacuation introduces a positive bias on estimates of consumption. When the prey in question constitutes half the stomach content weight, consumption is estimated to be c. 140% of the true consumption (Fig. 5). However, as the prey often constituted the majority of the stomach content when present in the whiting stomachs, the positive effect on consumption was only 26%, 1% and 10% on average for benthic invertebrates, herring and crabs, respectively (Table 4). The relative size of bias of the three groups reflects that expected by the size of the evacuation rate of the prey and the proportion of the stomach content made up by this prey.

Discussion

The difference between the estimates obtained by numerical integration as opposed to Penningtons method was minor for sampling intervals of 4 hours or less, and decreased with decreasing evacuation rate. The effect of this choice was much less than the effect of the technical assumptions made to simplify calculations, data collection or both. Serious bias was introduced both by estimating $E\left[S_t^\alpha\right]$ as $\left(E\left[S_t\right]\right)^\alpha$ and by making unwarranted assumptions regarding distribution of prey in the stomachs.

The estimation of $E\left[S_t^\alpha\right]$ by $\left(E\left[S_t\right]\right)^\alpha$ let to an overestimation of consumption. This effect was also noted by Ursin et al. (1985), dos Santos and Jobling (1995) and Andersen (2001) using different methods. However, neither of these authors attempted to estimate the effect directly from the mean and variance of the weight of the stomach contents. With the relationships presented here, calculation of $E\left[S_t^{\alpha}\right]$ is straight forward when the mean and variance as well as the underlying distribution is known. The estimation of these require the examination of individual stomachs, a labour intensive task. However, reasonable estimates of at least the coefficient of variation and distribution can probably be obtained from a minor proportion of all the stomachs collected. The remaining non-empty stomachs can then be pooled prior to weighing. Estimating total intake as suggested here should therefore not require extensive additional work compared to traditional methods.

The estimates of consumption of the three prey types were in most cases greatly underestimated when the traditional method of distributing the total consumption on prey types using the weight percentages was used. However, the intake of prey which occurred infrequently or did not constitute the large majority of the stomach content when present was positively biased. In general, larger fish prey occur only in a minor percentage of the stomachs collected (Albert 1995, Pillar and Barange 1997) and so consumption of fish is likely to be overestimated whereas that of invertebrates will be underestimated. Note that this conclusion holds even when no difference in evacuation rate exists between the two prey types.

The bias introduced by ignoring the effect of other prey on evacuation was minor in most cases. Only the intake of benthic invertebrates was seriously biased by this assumption and if simplifications of the calculations are to be made, the assumption of evacuation of each prey separately is far superior to the assumption of all prey being present in all stomachs. Unfortunately, studies calculating the intake of single prey by this method frequently neglect to correct for the bias introduced by assuming $E\left[S_t^{\alpha}\right]$ by $\left(E\left[S_t\right]\right)^{\alpha}$ (Armstrong 1982, Patterson 1985) and their results are therefore likely to be severely positively biased.

Estimating the intake of each prey type in the stomachs in the data set and summing these to obtain the total intake led to estimates of total intake of 60 to 108% of C_2. This difference was presumably due to a combination of random errors and a tendency to overestimate the mean weight of highly digested prey as prey is not identified during the last stages of digestion. It may therefore be necessary to adjust the intake of the individual prey types to obtain a summed intake equal to C_1.

The method suggested here has the advantage of incorporating the known sources of bias in food intake estimation while allowing average stomach content to change during the sampling period. The latter problem was not addressed by the model proposed by dos Santos and Jobling (1995). A major disadvantage of the method is, however, the expenses of collecting detailed information on the contents of individual stomachs. Making the assumption of constant CV of the stomach content may cut down costs as this would limit the need for information to the average weight and occurrence of each prey type. The bias introduced by assuming prey to be evacuated independently of other food in the stomach appears to be minor in most cases. This effect can either be ignored or alternatively, the average amount of other food present together with each prey can be determined from a smaller sub-sample of the data. The relative amount of other food present in the remaining stomachs may then be assumed to be identical to the amount in the sub-sample. The present study further demonstrates the importance of collecting samples throughout the diel cycle: The average intake estimated from daytime samples alone was only 25-50% of that estimated by night time samples in the data at hand. Sampling during daytime only would therefore lead to gross underestimation of the total intake of this prey type. In conclusion, reliable estimates of food consumption by fish require extensive sampling or at the very least some investigation of the assumptions underlying the calculations.

References

Albert, O. T. (1995). Diel changes in food and feeding of small gadoids on a costal bank. ICES J. Mar. Sci., vol 52, pp. 873-885.

Andersen, N. G. (1999). The effects of predator size, temperature, and prey characteristics on gastric evacuation in whiting. J. Fish Biol. (1999), vol. 54, pp. 287-301

Andersen, N. G. (2001). A gastric evacuation model for three predatory gadoids and implications of using pooled field data of stomach contents to estimate food rations. J. Fish Biol. 59, pp. 1198-1217

Armstrong, M. J. (1982). The predator-prey relationships of Irish Sea poor-cod (*Trisopterus minutes* L.), pouting (*Trisopterus luscus* L.), and cod (*Gadus morhua* L.). J. Cons. Int. Explor. Mer, 40, pp. 135-152

Bajkov, A. D. (1935). How to estimate the daily food consumption of fish under natural conditions. Trans. Am. Fish. Soc., 65, pp. 288-289

Basimi, R. A. and Grove, D. J. (1985). Estimates of daily food intake by an inshore population of Pleuronectes platessa L. off eastern Anglesey, North Wales. J. Fish Biol. 27, pp. 505-520

Daan, N. (1973). A quantitative analyses of the food intake of North Sea cod, *Gadus morhua*. Netherlands Journal of Sea Research, 6, pp. 479-517

Dos Santos, J. and Jobling, M. (1995). Test of a food consumption model for the Atlantic cod. ICES J. Mar. Sci., vol 52, no. 2, pp. 209-220

Eggers, D. M. (1977). Factors in interpreting data obtained by diel sampling of fish stomachs. J. Fish. Res. Board Can., vol. 34, pp. 290-294

Elliott, J. M. and Persson, L. (1978). The estimation of daily rates of food consumption for fish. J. Anim. Ecol. vol. 47, pp. 977-991

Hall, S. J., W. S. C. Gurney, H. Dobby, D. J. Basford, S. D. Heaney & M. R. Robertson (1995). Inferring feeding patterns from stomach contents data. J. Anim. Ecol. (1995), vol. 64, pp. 39-62

Hawkins et al, 1985. Growth and feeding of juvenile cod, Gadus morhua (L). J. Cons. Int. Explor. Mer, vol 42, pp. 11-32
Jobling, 1981. Mathematical models of gastric emptying and the estimation of daily rates of food consumption by fish. J. Fish Biol., vol. 19, pp. 245-257

Jones, R. (1974). The rate of elimination of food from the stomachs of haddock *Melanogrammus aeglefinus*, cod *Gadus morhua*, and whiting *Merlangius merlangus*. J. Cons. int. Explor. Mer. (1974), vol. 35, no. 3, pp. 225-243

Jones, R. (1978). Estimates of food consumption of haddock (Melangrammus aeglefinus) and cod (*Gadus morhua*). J. Cons. Int. Explor. Mer, vol. 15, pp. 225-243

Lockwood, S. J. (1980). The daily food intake of 0-group plaice (*Pleuronectes platessa* L.) under natural conditions. J. Cons. Int. Explor. Mer, vol. 39, no. 2, pp. 154-159

Magnússon, K. G., & T. Aspelund (1997). A model for estimating meal frequency and meal size from stomach data with an application to Atlantic cod (Gadus morhua) feeding on capelin (Mallotus villosus). Can. J. Fish. Aquat. Sci. (1997), vol. 54, pp. 876-889

Pedersen, J., & Hislop, J. R. G. (2001). Seasonal variations in the energy density of fishes in the North Sea. *Journal of Fish Biology* **59**, 380-389. doi: 10.1006/jfbi.2001.1649.
Pennington, M. (1985). Estimating the average food consumption by fish in the field from stomach contents data. Dana, vol. 5, pp. 81-86

Pillar, S. C. and Barange, M. (1997). Diel variability in bottom trawl catches and feeding activity of the

Cape hakes off the west coast of South Africa. ICES J. Mar. Sci., vol. 54, no. 3, pp. 485-499

Rice, J. C. (1988). Repeated cluster analysis of stomach contents data: method and application to diet of cod in NAFO division 3L. Env. Biol. Fish., vol. 21, no. 4, pp. 263-277

Rindorf, A. (2002). The effect of stomach fullness on food intake of whiting in the North Sea. J. Fish Biol., no. 61 (3), pp. 579-593

Rindorf, A. (2003). Diel feeding pattern and prey choice of whiting (*Merlangius merlangus*) in the North Sea. *Marine Ecology Progress Series*. In press.

Rindorf, A. (subm.). Estimation of *in situ* evacuation rates. J. Fish Biol.

Sainsbury, K. J. (1986). Estimation of food consumption from field observations of fish feeding cycles. Journal of Fish Biology (1986), vol. 29, no. 1, pp. 23-36

Seyhan, K. and Grove, D. J. (1998). Food consumption of whiting, *Merlangius merlangus*, in the Eastern Irish Sea. Fish. Res., vol. 38, pp. 233-245

Stefánsson, G., & O. K. Pálsson (1997). Statistical evaluation and modelling of the stomach contents of Icelandic cod (*Gadus morhua*). Can. J. Fish. Aquat. Sci. (1997), vol. 54, pp. 169-181

Ursin, E., Pennington, M., Cohen, E. B. and Grosslein, M. D. (1985). Stomach evacuation rates of Atlantic cod (*Gadus morhua*) estimated from stomach contents and growth rates. Dana vol. 5, pp. 63-80

Table 1: Calculation of estimates of average hourly consumption. Notation as in the text. ρ_{tot} and ρ_p is the evacuation rate of the total stomach content and the prey specific evacuation rate, respectively.

Estimate	Calculation
C_1	S_t derived by numerical solution of $\frac{dS}{dt} = -\rho S^{0.5} + C_r$, in a standard software package
C_2	$C_2 = \frac{\rho_{tot}}{N} \sum_{i=1}^{N} p_i \frac{\Gamma(k_{tot}+0.5)}{\Gamma(k_{tot})\sqrt{k_{tot}}} \sqrt{E\left[X_i\right]} + \frac{1}{T}\left(p_T E\left[X_T\right] - p_0 E\left[X_0\right]\right)$
C_3	$C_3 = \frac{\rho_{tot}}{N} \sum_{i=1}^{N} \sqrt{p_i E\left[X_i\right]} + \frac{1}{T}\left(p_T E\left[X_T\right] - p_0 E\left[X_0\right]\right)$
C_4	$C_4 = \frac{1}{N} \sum_{i=1}^{N} \sqrt{\pi_j}\left(\pi_j \rho_j + (1-\pi_j)\rho_{tot}\right) p_{j,i} \frac{\Gamma(k_j+0.5)}{\Gamma(k_j)\sqrt{k_j}} \sqrt{E\left[X_{j,i}\right]} + \frac{1}{T}\left(p_{j,T} E\left[X_{j,T}\right] - p_{j,0} E\left[X_{j,0}\right]\right)$
C_5	$C_5 = \pi_{tot,j} C_2$
C_6	$C_6 = \frac{\rho_j}{N} \sum_{i=1}^{N} p_{j,i} \frac{\Gamma(k_j+0.5)}{\Gamma(k_j)\sqrt{k_j}} \sqrt{E\left[X_{j,i}\right]} + \frac{1}{T}\left(p_{j,T} E\left[X_{j,T}\right] - p_{j,0} E\left[X_{j,0}\right]\right)$

Table 2: Evacuation rates of total stomach contents and individual prey types.

Location	1	2	3	4	5
Temperature (°C)	16.7	13.7	10.6	8.2	8.6
ρ_{tot}	0.120	0.098	0.087	0.075	0.077
$\rho_{benthicinv}$	0.183	0.145	0.113	0.094	0.097
$\rho_{herring}$	0.118	0.094	0.074	0.061	0.063
ρ_{crabs}	0.091	0.072	0.056	0.047	0.048

Table 3: Comparison of average food intake (mg/h) estimated by correcting for the fact that $E\left[S_t^{\alpha}\right] \neq \left(E\left[S_t\right]\right)^{\alpha}$ (C_2) and by assuming $E\left[S_t^{\alpha}\right]$ to be equal to $\left(E\left[S_t\right]\right)^{\alpha}$ (C_3). The proportion of stomachs, which contained food, p, is also given.

Location	C_2	C_3	p
1	132	187	0.826
2	66	100	0.779
3	112	157	0.844
4	30	49	0.746
5	69	109	0.634

Table 4: Consumption (mg/h) of individual prey estimated by three different methods. Correcting for the unequal distribution of prey and the effect of other food on evacuation (C_4), applying weight percentages to total consumption (C_5) and correcting for the unequal distribution of prey but neglecting the effect of other prey on evacuation (C_6).

Location	1	2	3	4	5
Benthic Invertebrates					
average π_j	0.72	0.81	0.67	0.73	0.68
average p	0.23	0.39	0.40	0.08	0.02
average X	0.74	0.74	0.74	0.74	0.74
C_4	17.2	37.0	24.2	3.7	1.3
C_5	7.1	32.6	16.1	2.3	3.2
C_6	22.4	41.6	31.1	4.7	1.7
Herring					
average ϕ_j	0.96	0.84	0.92	1.00	0.98
average P	0.04	0.03	0.04	0.001	0.01
average X	6.7	6.6	14.1	3.4	5.8
C_4	30.7	6.3	24.4	0.2	2.3
C_5	30.4	2.3	31.1	0.5	2.8
C_6	30.9	6.6	24.6	0.2	2.3
Crabs					
average ϕ_j	0.50	0.57	0.48	0.60	0.50
average p	0.23	0.18	0.34	0.16	0.003
average X	0.29	0.73	0.29	0.37	0.22
C_4	10.3	15.2	8.6	3.3	0.07
C_5	6.6	2.9	7.3	5.5	0.13
C_6	12.0	16.4	9.6	3.4	0.08

Fig. 1. The relative difference between average consumption rate calculated by numerically integrating eq. 1 and the estimate suggested by Pennington (eq. 2). Effect of interval between sampling points: T=1 (A), T=4 (B) and T=10 (C). Relative difference: 0-1% (≤), 1-3% (′), 3-5% (′), 5-7% (′), >7% (′). Positive differences are hatched. Evacuation rate was fixed at 0.24.

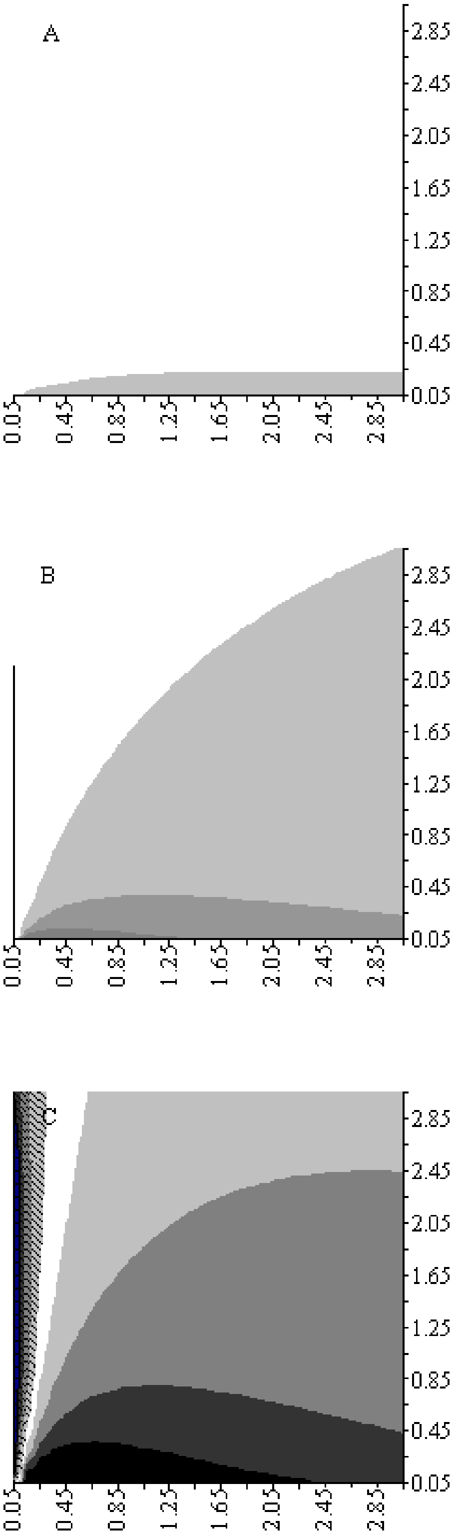


Fig. 2. The relative difference between average consumption rate calculated by numerically integrating eq. 1 and the estimate suggested by Pennington (eq. 2). Effect of evacuation rate: ρ=0.04 (A), ρ=0.14 (B) and ρ=0.24 (C). Relative difference: 0-1% (≤), 1-3% (′), 3-5% (′), 5-7% (′), 7-9% (′). Positive differences are hatched. Interval between sampling points was fixed at 4 hours.

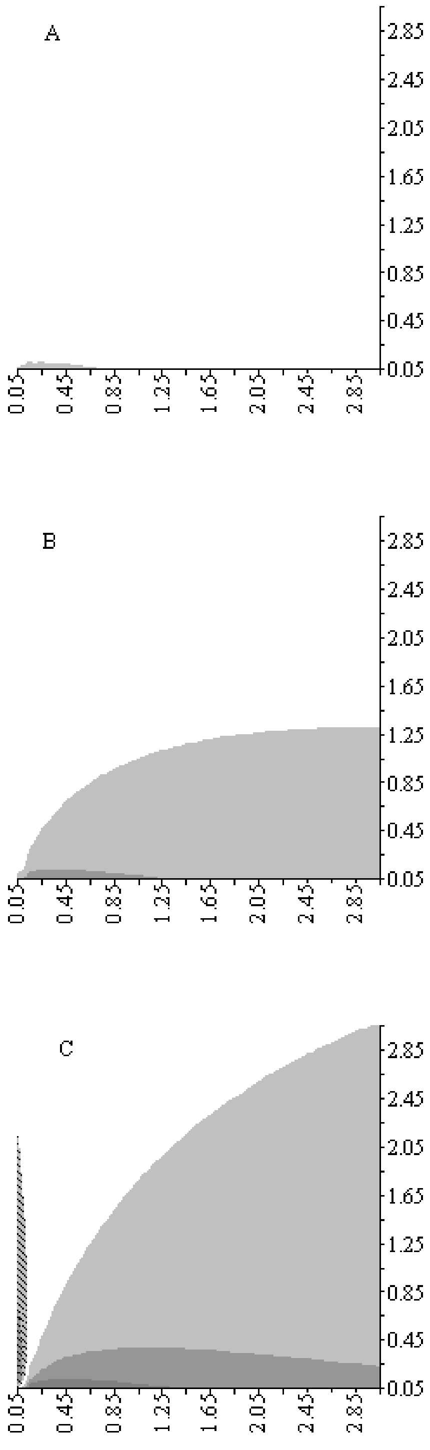


Fig. 3. Bias introduced by calculating $E\left[\sqrt{S_t}\right]$ as $\sqrt{E\left[S_t\right]}$ when S is gamma distributed (A) or log-normal distributed (B) as a function of the proportion of stomachs which contain food, p, and the shape parameter of the gamma distribution, k, and variance of the log-normal distribution, σ^2, respectively. Bias 0 to -20% ($\leq$), -20 to –40% ($'$), -40 to -60% ($'$), -60 to -80% ($'$) –80 to -100% ($'$).

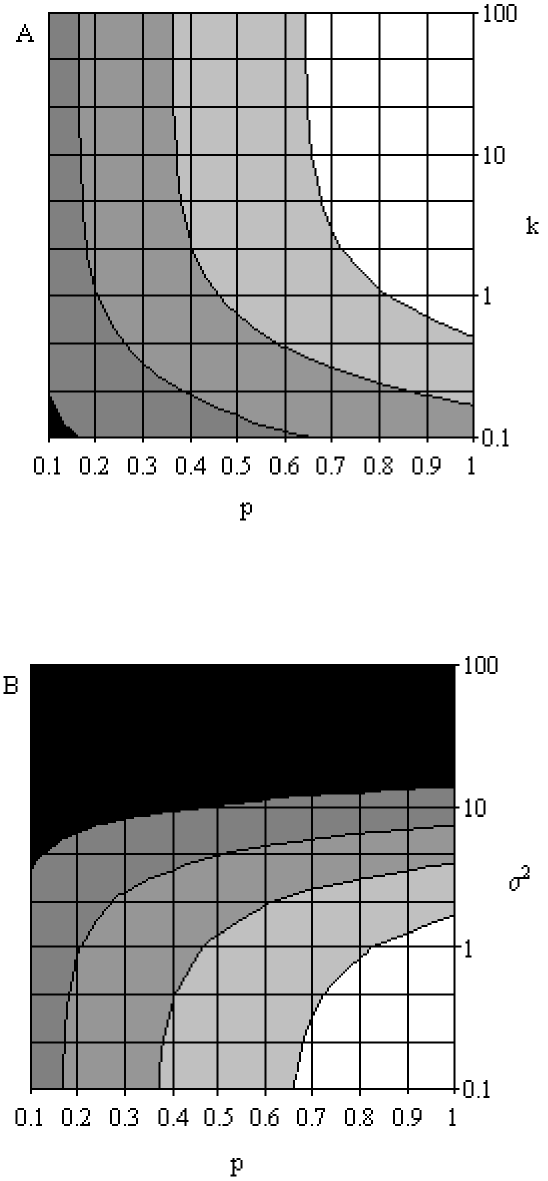


Fig. 4. Bias introduced by assuming all prey to occur in all stomachs in average amount as a function of the proportion of stomachs with food which contain prey j, p_j/p_{tot}, and the ratio between the proportion of the prey in the total stomach content and in the contents of the stomachs in which the prey is found, π_{tot}/π_p. Bias 0 to 10% ($\leq$), 10 to 30% ($'$), 30 to 50% ($'$), 50 to 70% ($'$) 70 to 90% ($'$). Positive differences are hatched.

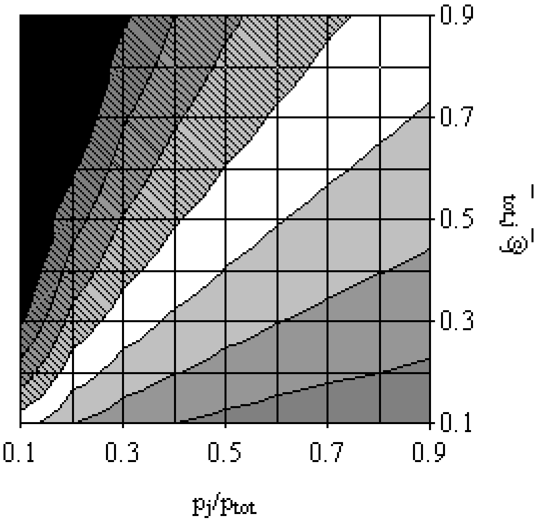
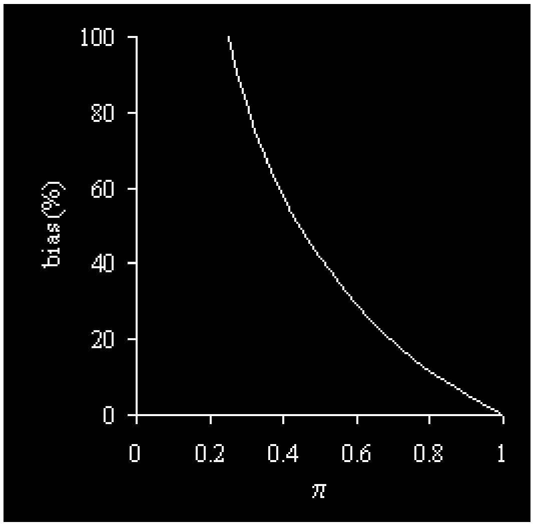


Fig. 5. Bias introduced by assuming no other food is present in the stomach as a function of the proportion of the stomach content which consists of the given prey, π.



G.4 Deliverables 4.1.1 and 4.1.2 of DST2

Not to be cited without prior permission from the author

By Geir Huse

Institute of Marine Research, Box 1870 Nordnes, N-5817 Bergen, Norway. Email: geir.huse@imr.no

Introduction

This report describes deliverables 4.1.1 and 4.1.2 of the EU funded project "Development of structurally detailed statistically testable models of marine populations" (QL K5 – CT 1999 – 01609). The work related to D4.1.1 and WP 4.1.2 is to develop a multi-species spatially explicit feeding/consumption model for the interaction between cod, herring and capelin. The spatial context is presented in Figure 1 which illustrates the potential for interaction between the populations.

Specifically the deliverables are:

4.1.1 Age and size dependent growth and predation mortality of cod herring and capelin in the Barents Sea.

4.1.2 Spatial distribution of cod, herring and capelin in the Barents Sea.

The basic approach is an individual-based model where habitat choice is determined by rules founded on evolutionary fitness maximisation and characteristic movement patterns of the target stocks. A central aspect has been the predator-prey interaction between cod and capelin, and cod and herring. For reviews of modelling spatial dynamics of fish stocks based on fitness maximising principles see Tyler and Rose (1994), and Giske *et al.*, (1998). Below the model initially is described before the results of the model simulations are presented and discussed.

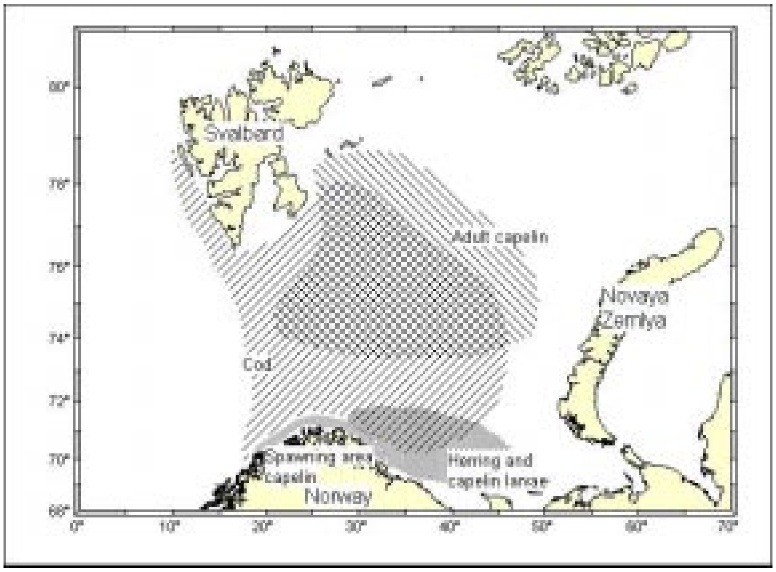


Figure 1: The Barents Sea with the main features of the distributions of cod, capelin and herring. Modified from Huse and Gjøsæter (1999).

The model

Temperature

The model domain consists of a 90 (x-axis) by 100 (y-axis) grid of 20x20 km squares (Fig. 2). Temperatures in the Barents Sea have been gathered by the Institute of Marine Research, Bergen, at quarterly intervals. Daily temperatures are generated by linear interpolation of the quarterly values. Furthermore to fill the entire grid with temperature values, linear spatial interpolation between observations was performed. In addition temperature fields generated by an ocean circulation model (Ådlandsvik and Sundby, 1994) were used for interpolation outside the area where temperature observations has been made. The simulated fields are shown in Fig. 2, which illustrates the magnitude of seasonal changes in temperature in the Barents Sea.

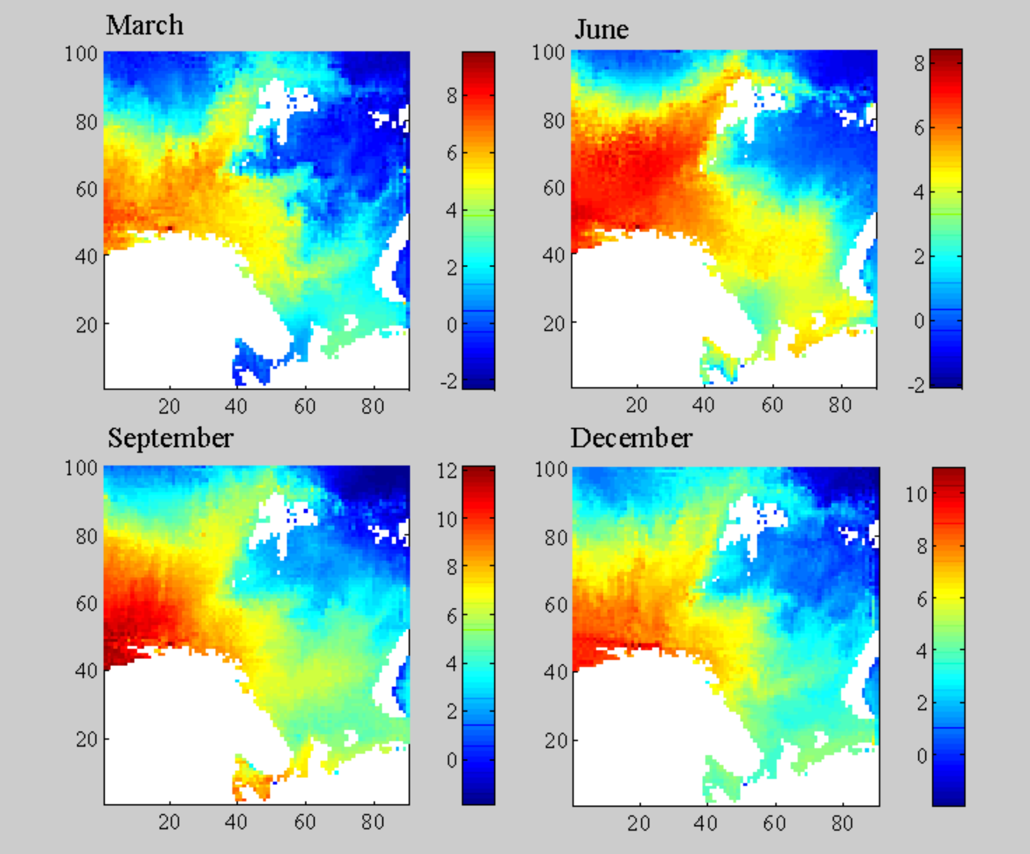


Figure 2: The model domain with simulated temperature distribution at four different months during 1990-91. The values given in the legend are temperatures in °C, and the white areas are land.

The Individual-Based Model

There are different ways of constructing individual-based models (IBMs), and for reviews see Grimm (1999), Railsback (2001), or Huse et al. (2002). One way to deal with fish populations that have high mortality is to use the super-individual approach (Scheffer *et al.*, 1995). A super-individual represents many identical individuals, here referred to as siblings (Fig. 3). This is an efficient way of maintaining 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 here 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*:*

$$AV = (a, l, w, n, x, y\{) \qquad (1)$$

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 at time *t*. The attribute vector was similar for all the species. The attribute vectors were initially generated from observations of capelin, cod and herring. Data from a survey in September-October (Gjøsæter *et al.*, 1998) was used to initiate the capelin super-individuals. Similarly data from a survey in August was used to initiate the cod. As juvenile herring is only targeted on a survey in May-June, these data were used to initiate the herring super-individuals. For cod and capelin the observations were given as average length and weight, and number of each age group in the occupied squares. For cod and capelin, spatially resolved biological data were available to initiate the super-individuals, while in the case of herring such data were pooled by age group for the entire cruise. The herring super-individuals were therefore initiated with the average weights and lengths of the age group. The observations were then divided into about 16000 super-individuals in proportion to the number of fish in each square. The age and size were based on biological samples. Capelin and herring abundance was determined by acoustic estimation (Gjøsæter *et al.*, 1998; Toresen *et al.*, 1998) while cod abundance was determined by using a combination of acoustic estimation and the swept area method. Age groups 1-3 were used for capelin, 1-2 for herring, and 2-10 for cod.

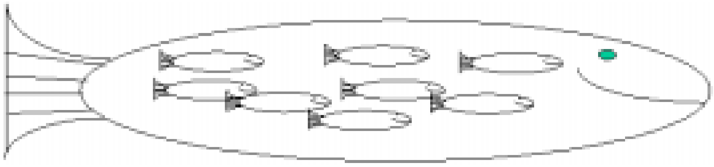


Figure 3: The super-individual approach illustrated.

Migration

Movement of the super-individuals was performed using rules. The rules are species specific and separated by ontogenetic stage. Thus both the spawning migration of mature individuals as well as the movements of the younger fish are implemented in this manner. The movement rules are listed in Table 1. The rules were based partly on specific knowledge of the species and partly on life history departure rules. In particular the “minimise mortality/growth” rule was used to direct movement during the feeding season (Werner and Gilliam, 1984). This rule was implemented by searching the neighbourhood, consisting of eight surrounding squares and the one currently occupied, and directing movement towards the best square. In addition to the directed movement (D_X and D_Y) resulting from the rules for movement along the X and Y axes, a stochastic term (*E*, Table 2) was added:

$$\begin{aligned} \Delta X &= D_X + E \\ \Delta Y &= D_Y + E \end{aligned} \tag{2}$$

where ΔX and ΔY is movement along the x and y axes respectively. Movement is then calculated by adding the movement to the current position as illustrated in Fig. 4. Capelin and cod was assumed to mature at a length of 14 and 75 cm respectively.

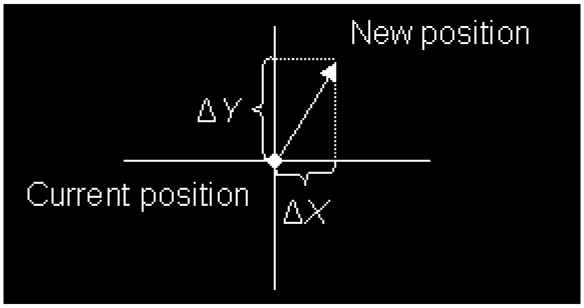


Figure 4: The calculation of movement in the model.

Table 1: The movement rules used to simulate the movement of cod, capelin and herring. Prey refers to herring and capelin.

Species	Immature	Mature
Cod	-Choose the square in the neighbourhood with the highest prey concentration -If no prey is found move north during winter and south during summer	-Migrate to Lofoten to spawn during winter -Outside the spawning season move towards the square in the neighbourhood with the highest prey concentration -If no capelin is found move north during winter and south during summer
Capelin	-Move southwards to the 3°C thermocline during winter -Minimise mortality/growth during summer/fall with a bias towards moving north	-Migrate southwards to spawn during spring
Herring	-Minimise mortality/growth during -Bias towards moving north during summer and south during winter	

Growth and Mortality
Growth was modelled using bioenergetics models (Hewett and Johnson, 1992). A basic balanced energy budget is the foundation of such models:

$$G = C - (R + S + F + U) \tag{3}$$

where *G* is growth, *C* is consumption, *R* is respiration, *S* is specific dynamic action, *F* is egestion, and *U* is excretion. Each of the variables on the right hand side of the equation were calculated by species specific equations and parameters. While there exists bioenergetics models for herring and cod, such models have previously not been developed and validated for capelin. Some work on the current project has therefore been performed to modify the herring model into a capelin bioenergetics model. This work has been reported earlier (Huse and Gjøsæter, 2002), and the modified herring model was used for calculating capelin growth here. While the equations for the metabolic costs were taken from the literature (Hewett and Johnson, 1992; Hansson *et al.*, 1996), consumption was modelled explicitly. A linear functional response with satiation due to stomach fullness was assumed for all the species. A zooplankton model based on time of the year and temperature was generated. Capelin growth was assumed proportional to the zooplankton biomass of the patch, reduced by the local abundance of competing individuals:

$$C_L = a_L \cdot Z \cdot w^{b_L} \cdot e^{-\frac{N}{K}} \tag{4}$$

where Z is the zooplankton biomass, N is the number of planktivorous fish (herring and capelin) in the patch, w is the body weight of the planktivorous fish, b_L is an allometric constant (Table 2), and K is the carrying capacity of a patch. While the biomass differs among patches, the carrying capacity was assumed to be similar (Table 2). A similar consumption model was used for herring, but with a different attack rate parameter (Table 2). Herring and capelin are assumed to have a stomach capacity of 4% of body weight per day.

Table 2: Parameters and variables used in the model. Parameter values in brackets are values for predators in cases where these differ from prey parameter values.

Symbol	Description	Value	Unit
a_C	Attack rate of cod	0.2	Prey pred^{-1}
a_L	Attack rate of capelin	0.042	g g^{-1}
a_S	Attack rate of herring	0.053	g g^{-1}
b_L	Capelin/herring consumption constant	-0.256	-
C$_L$, C$_S$, C$_C$	Capelin, herring and cod consumption	…	kg day^{-1}
E	Stochastic movement component	0.4	-
K	Carrying capacity for planktivores	5x10^{110}	Individuals
N	Abundance of planktivorous fish	…	Individuals
S	Number of siblings in super-individuals	…	Individuals
T	Temperature	…	°C
W	Average weight of planktivorous fish	…	kg
Z	Zooplankton biomass	…	g
ì	Daily mortality rate for cod	0.0015	Individuals day^{-1}

Note: Ellipses indicate no fixed value.

Cod growth is limited to maximally 4% of body weight per day, due to stomach satiation. Cod growth was calculated from the number of prey eaten, and the number of siblings in capelin super-individuals is reduced in accordance with predator feeding. Cod is assumed to have a linear functional response:

$$C_C = a_C \cdot N \cdot W \tag{5}$$

where a_C is the attack rate (Table 2), and W is the average body weight of the planktivorous fish in the patch. Cod foraging is calculated in two steps. First the feeding rate of each cod is calculated by sequential feeding according to the local capelin abundance, taking into account the reduction caused by previous predator feeding that day. Then the total predation in each patch is calculated by adding together this predation. Finally the average feeding rate is calculated and used as the realised feeding rate of cod in a patch. Upon cod feeding, the internal number of capelin super-individuals in the patch is reduced in proportion to the feeding of the predator. If several capelin and/or herring reside in one patch the mortality is distributed equally among them. In addition to feeding on fish, cod ingest many invertebrates including shrimp, amphipods, and euphausiids (Mehl, 1989). In order to account for this, a constant ration of invertebrates was added to the fish consumption.

The survival of cod siblings at time t is calculated by:

$$S_{Ct+1} = S_{Ct} \cdot e^{-\mu} \tag{6}$$

where S_{Ct} is the number of cod siblings in a super-individual at time t.

The computer code used in the modelling is available upon request. It may be interfaced with GADGET, but this connection has of yet not been made.

Deliverable 4.1.1: Age and size-dependent growth and predation mortality for cod, capelin and juvenile herring in the Barents Sea.

As mentioned above the model is initiated in fall and then run for a period of a year. References to years are made for the final year of the simulation.

Cod Growth

The model simulations predict that cod growth should vary strongly during the study period. In the first half of the period there was a high abundance of capelin in the Barents Sea, but towards the end, the stock has virtually collapsed (Gjøsæter, 1998). There is also a trend towards cod of about 75 cm having the highest growth relative to other size groups (Fig. 5). The growth is somewhat smaller than that seen for Northeast Arctic cod (Toresen, 2000). This probably reflects a too low abundance of alternative prey in years with low capelin abundance. There was also a seasonal pattern in growth, which reflects the seasonal changes in predation on capelin. Thus the condition of cod generally peaked in April and dropped off towards the fall.

Table 4: Age dependent growth (cm year^{-1}) of cod during 1991-1996. Age group refers to age at initiation.

	Year					
Age group	1991	1992	1993	1994	1995	1996
2	7.04	3.62	3.83	3.94	3.11	0.45
3	3.99	3.3	5.52	4.76	3.57	0.14
4	5.8	4.81	10.84	6.58	5.08	0.53
5	9.01	6.54	13.91	9.48	5.47	0.47
6	10.6	8.97	11.32	11.17	3.53	0.75
7	9.95	12.55	12.37	12.2	3.68	1.67
8	11.68	10.07	13.99	7.2	3.86	1.15
9	15.64	8.33	13.68	10.35	5.29	1.99
10	17.31	7.8	10.04	11.26	4.06	2.08

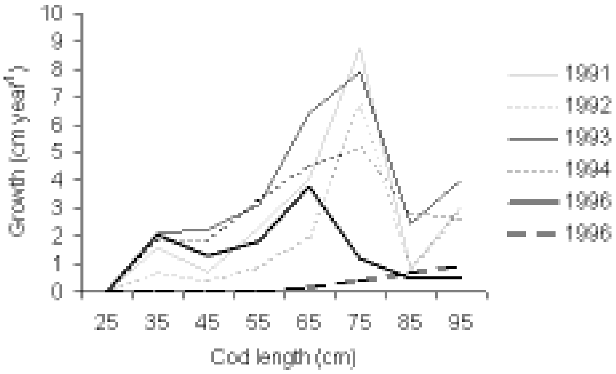


Figure 5: Size dependent growth of cod during 1991-1996.

Capelin Growth

Capelin growth varied both inter-annually (Table 5) and within the year (Fig. 6). The predicted weight at age 1 compared with the observed values (Toresen, 2000) produced at R^2 of 0.4. In 1995 the model predictions compare relatively poorly with observations while in 1991 and 1992 they compare very well. There is a general pattern with growth being positive during fall and late summer and negative for the rest of the year (Fig. 5).

Table 5: Age dependent length growth of capelin during 1991-1996. Age group refers to age at initiation.

	Year					
Age group	1991	1992	1993	1994	1995	1996
1	5.4	5.2	-0.5	9.5	17.3	4.7
2	1.7	8.1	-1.5	15.4	31.6	13.9

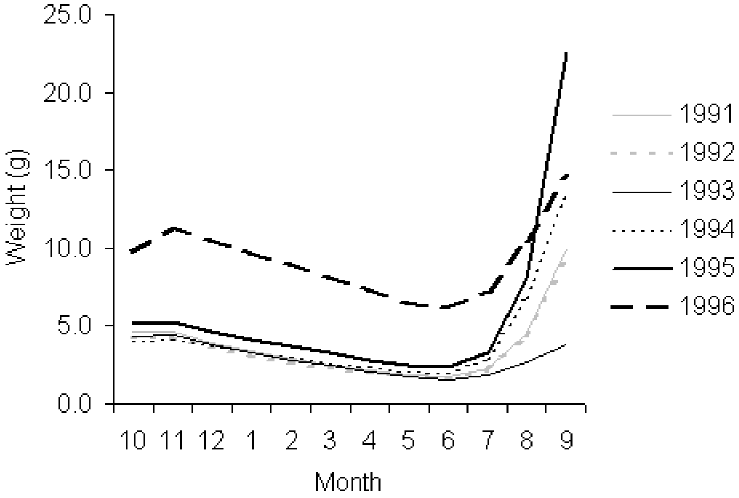


Figure 6: Seasonal growth of capelin during 1991-1996 from October to September.

Herring Growth

Herring growth was predicted to be greater during 1993 and 1994 than in 1995 (Table 6). The pattern in growth was similar to that seen for capelin (Figs. 6 and 7). Compared with real herring growth (Toresen and Østvedt, 2000), the model predictions are good for 1993 and 1994, but for 1995 the model predicts too low growth.

Table 6: Age dependent length growth of herring during 1993-1995. Age group refers to age at initiation.

		Year	
Age group	1993	1994	1995
1	3.89	5.38	0.7
2	3.41	4.3	0.74

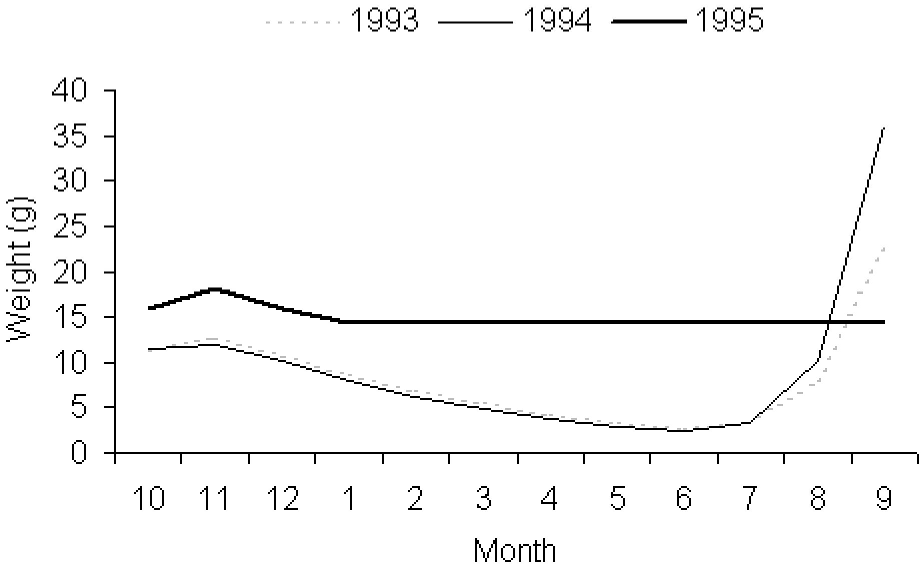


Figure 7: Growth of 1-year old herring over the year during 1993-1995.

Capelin Mortality
Capelin mortality was generally higher in the 2-year olds than in the 1-year olds. This was caused mainly by the maturation of parts of the cohort as 3-year olds, and this was pronounced in 1991, 1995 and 1996 (Table 7). The mortality attributed to cod predation in the 1-year old capelin was around 30% per year. This contributed to a predicted maximum annual capelin consumption by cod of about 1.5 mill. tonnes in 1992. This is less than previous estimates of consumption, which is about one million tonnes higher (Bogstad *et al.*, 2000).

Table 7: Age dependent mortality of capelin during 1991-1996. Age group refers to age at initiation.

	Year					
Age group	1991	1992	1993	1994	1995	1996
1	0.18	0.29	0.22	0.58	0.17	0.71
2	0.91	0.31	0.33	0.39	0.84	1.00

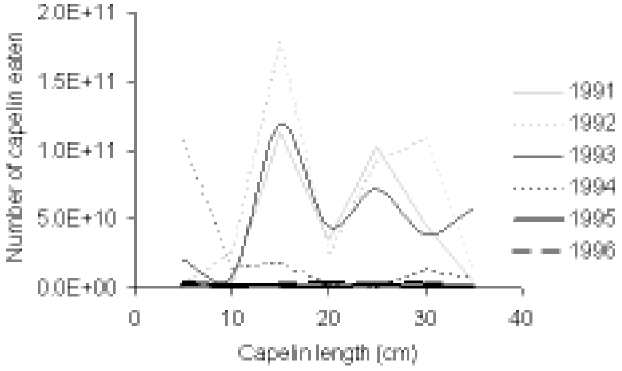


Figure 8: Length specific number of capelin eaten by cod.

Mortality of Juvenile Herring
The mortality of juvenile herring was about 30% per year (Table 8), and most of the mortality took place during the last months of the initiation year (Fig. 9). The mortality is predicted to be about the same as that for capelin (Tables 7 and 8), which might be expected since both species here are equally available to cod predation. The model overestimates cod consumption of herring compared with previous consumption estimates (Bogstad *et al.*, 2000; Johansen, 2003). However, the predicted mortality is lower than for juvenile herring since the model does not take into account other causes of mortality such as other predation and starvation.

Table 8: Age dependent mortality of herring during 1993-1995. Age group refers to age at initiation.

		Year	
Age group	1993	1994	1995
1	0.31	0.36	0.21
2	0.33	0.33	0.21

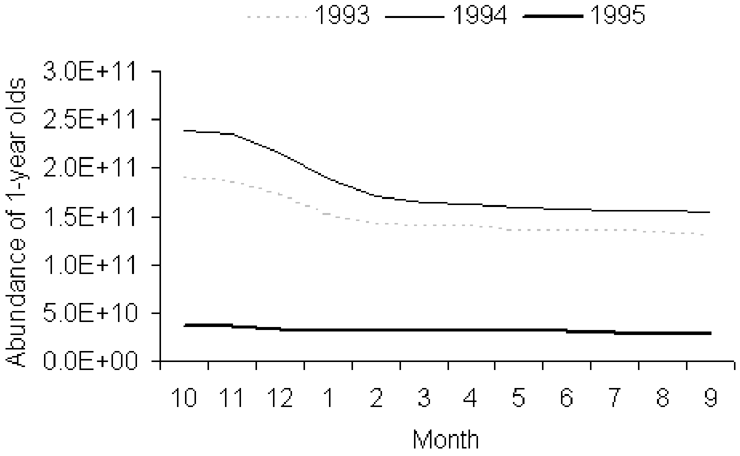


Figure 9: Size specific number of herring eaten by cod.

Deliverable 4.1.2: Distribution of cod, capelin and juvenile herring in the Barents Sea.

Temperature has been shown to have a great effect on the distribution of cod and capelin in the Barents Sea (Ozhigin and Luka, 1985; Gjøsæter, 1998; Ottersen *et al.*, 1998), with fish being distributed towards the west in cold years and relatively more towards the east in cold years. Even though there was little climatic variation within the study period, maps for a warm period 1990-1991 and a relatively colder period 1993-1994 are presented.

Spatial distribution of cod
The simulated spatial distribution of cod is shown in Figs. 10 and 11. At initiation the cod is spread over a large area, but as a result of the movement rules, they aggregate in certain areas where the prey availability is good. During February and March, the spawning migration of the adult stock is seen. In summer some of the cod pursues the capelin to the frontal areas in the northern Barents Sea, while most of the stock remains further south (Fig. 10). A pronounced feature is the association of the immature cod stock with the capelin stock during capelin spawning. This repeats a well known aspects of cod migrations in the Barents Sea.

Spatial distribution of capelin
Examples of the simulated spatial distribution of capelin are shown in Figs. 12 and 13. A notable features are the separation of the immature and mature components of the stock as seen in March and April. While the overlap between cod and capelin is relatively high during the first part of the year, there

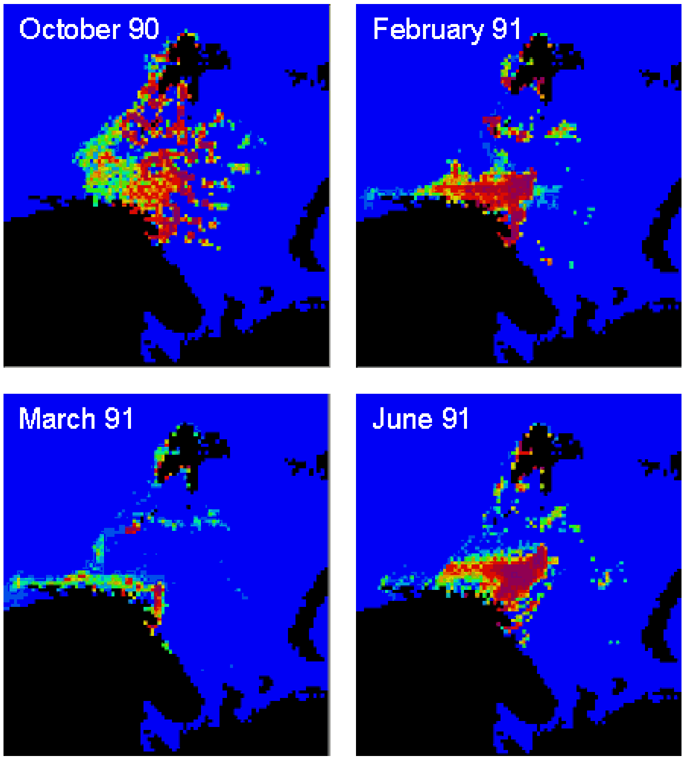


Figure 10: Spatial distribution of cod from fall of 1990 to summer of 1991.

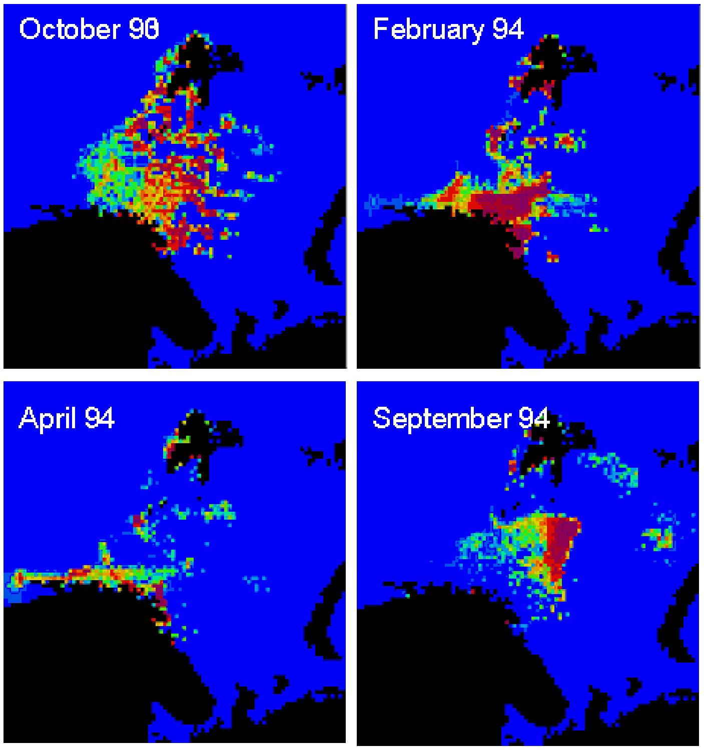


Figure 11: Spatial distribution of cod from fall of 1993 to fall of 1994.

is little overlap during the fall. The capelin distribution in 90-91 is more to the east and to the east than in 93-94. The former period is slightly colder than the latter, and such differences in spatial distribution caused by climatic state are well known in capelin (Ozhigin and Luka, 1985; Gjøsæter, 1998). The predicted spatial distributions of capelin resembles the observed capelin distribution quite well.

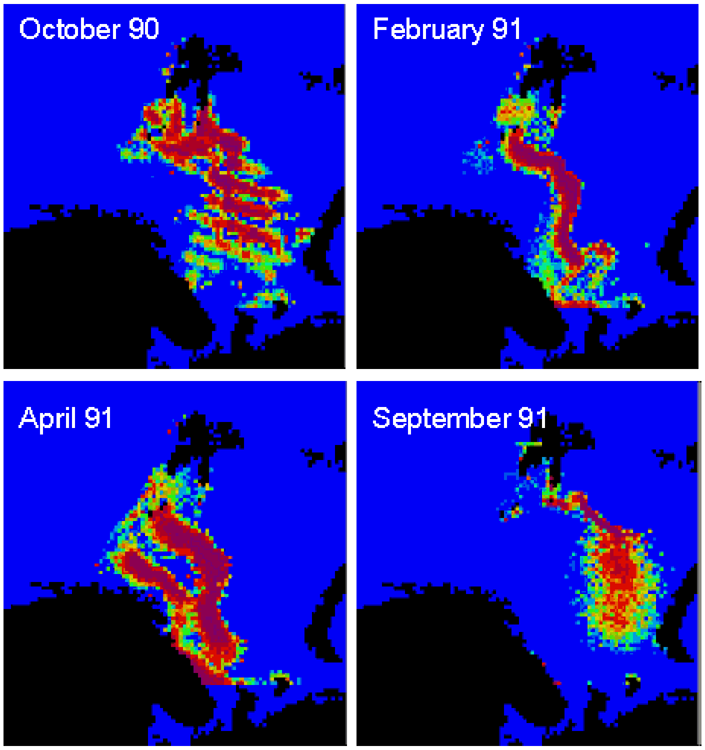


Figure 12: Spatial distribution of capelin from fall of 1990 to fall of 1991. Note the separation of the mature and immature fish in April.

Spatial distribution of herring

While cod and capelin stay in the Barents Sea over their entire ontogenetic life span, the herring only spends its adolescence in the Barents Sea. Consequently the herring does not undertake any pronounced seasonal migration like the other two species, although there are some minor changes in the distribution between winter and summer (Johansen, 2002). The herring migration model was therefore based on minimising mortality/growth, but with seasonally determined biases. As shown in Fig. 14 the herring distribution remained relatively compact throughout the simulation. One reason for this is the rule used which limits herring from going into water colder than 2 °C. The herring moves away from the coast in April, probably as a consequence of the onward movement of capelin, which attracts cod and thus increases the predation risk.

Discussion

The results 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 also known to vary inter-annually as a function of climatic conditions and density dependencies (Gjøsæter and Loeng, 1987; Skjoldal *et al.*, 1992). There was a notable variation in the simulated capelin growth. Maturation is an important process regarding capelin size at age since most capelin die following spawning. Thus if conditions are favourable for growth, a large part of the three year olds will mature, spawn and die. Similarly in case of unfavourable conditions maturity will be delayed. The

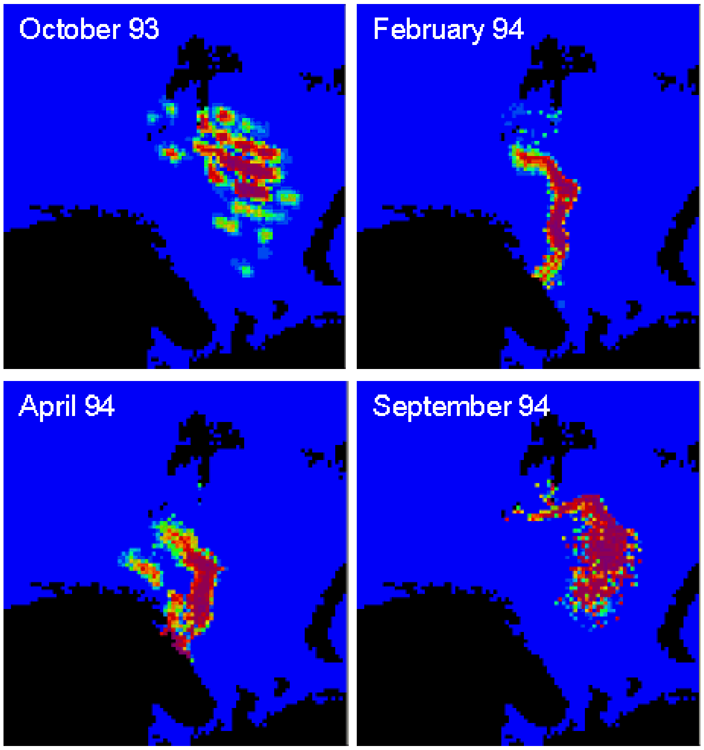


Figure 13: Spatial distribution of capelin from fall of 1993 to fall of 1994.

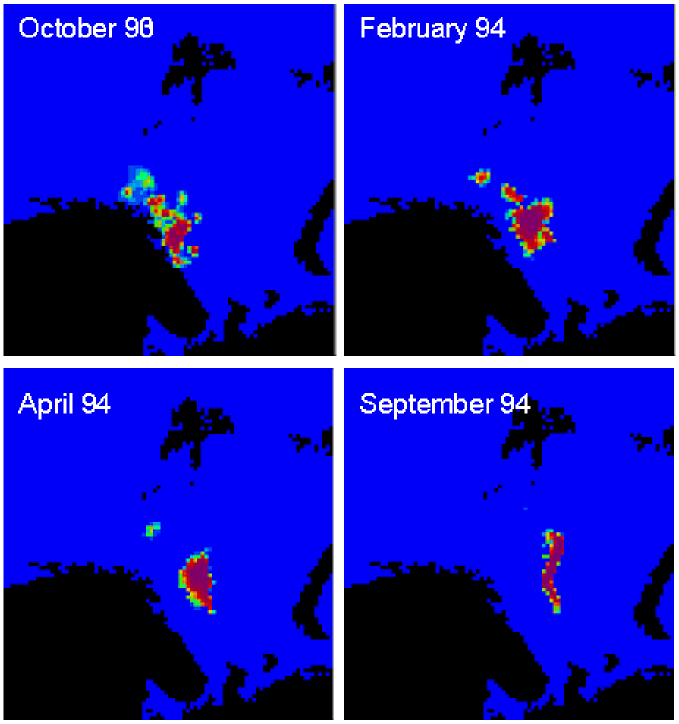


Figure 14: Spatial distribution of herring from fall of 1993 to fall of 1994.

population dynamic of capelin is sensitive to this factor, and this was reflected in the model simulations as well.

Emphasis here has been on developing a new modelling approach in fisheries assessment. The preliminary validation is promising although there are some problems. For example there is some inter annual variability in predicted growth that differs from observations, while in other years the fit is very good. These issues will be worked on during the remaining project period. The model predicted capelin and spatial distributions that are in accordance with observations. Although the cod distribution was predicted to be in accordance with observation, some more work is needed on the inclusion of invertebrate prey, which cod also feed on. The herring movement model needs to be validated better towards data from other years in order to better understand the spatial dynamics in the Barents Sea of this species. In addition to a general refinement of the model, the consumption output will be compared with consumption estimates calculated within DST^2.

The model presented here can be a valuable tool in fisheries management in different ways. It yields the basic rates of growth, mortality and consumption of the target stocks, which are essential components of population models including fisheries assessment models. The fishery is left out of the current version of the model, but this can easily be implemented. It also captures spatial processes, which to a large degree have been neglected in fisheries assessment. This implementation of spatial processes is promising, and may produce new insight into how different biological and physical factors impact on the spatial and population dynamics of fish stocks. This type of model is therefore valuable as a tool for testing hypotheses about fish migrations, for example the possible impact of climate changes on fish stock. Rather than seeing this contribution as an end result, it should be viewed as a starting point for future explorations of spatial IBMs in fisheries assessment.

Acknowledgements

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., Haug, T., and Mehl, S. 2000. Who eats whom in the Barents Sea. NAMMCO Scientific Publications 2: 98-119.

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.

Giske, J., Huse, G., and Fiksen, Ø. 1998. Modelling spatial dynamics of fish. Rev. Fish Biol. Fish. 8: 57-91.

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 Rottingen, B. 1998. The Barents Sea capelin stock 1972-1997. A synthesis of results from acoustic surveys. Sarsia 83: 497-510.

Gjøsæter, H., and Loeng, H. 1987. Growth of the Barents Sea capelin *Mallotus villosus* in relation to climate. Envir. Biol. Fish. 20: 293-300.

Grimm, V. 1999. Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future? Ecol. Modell. 115: 129-148.

Hamre, J. 1994. Biodiversity and the exploitation of the main fish stocks in the Norwegian-Barents Sea ecosystem. Biodiv. Cons. 3: 473-492.

Hansson, S., Rudstam, L. G., Kitchell, J. F., Johnson, B. L., and Peppard, P. E. 1996. Predation rates by North Sea cod (*Gadus morhua*) - Predictions from a models on gastric evacuation and bioenergetics. ICES J. Mar. Sci. 53: 107-114.

Hewett, S. W., and Johnson, B. L. 1992. Fish Bioenergetics Model 2. University of Wisconsin Sea Grant Institute, Madison

Huse, G., Giske, J., and Salvanes, A. G. V. 2002. Individual-based models. Handbook of Fish and Fisheries. P. J. B. Hart and J. Reynolds Blackwell Science, Oxford 2:228-248Handbook of Fish and Fisheries

Huse, G., and Gjøsæter, H. 1999. A neural network approach for predicting stock abundance of the Barents Sea capelin. Sarsia 84: 457-464.

Huse, G., and Gjøsæter, H. 2002. An evaluation of a bioenergetics model for capelin. Progress report for DST2 January 2002.:9

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 J. Mar. Sci. 59: 270-292.

Johansen, G. O. 2003. The predator-prey interaction between Northeast Arctic cod (*Gadus morhua* L.) and juvenile Norwegian spring spawning herring (*Clupea harengus* L.) in the Barents Sea. PhD dissertation, University of Bergen

Mehl, S. 1989. The Northeast Arctic cod stock's consumption of commercially exploited prey species in 1984-1986. Rapp. Proc. Verb. Réun. Cons. Int. l'Exp. Mer 188: 185-205.

Ottersen, G., Michalsen, K., and Nakken, O. 1998. Ambient temperature and distribution of north-east Arctic cod. ICES J. Mar. Sci. 55: 67-85.

Ozhigin, V. K., and Luka, G. I. 1985. Some peculiarities of capelin migrations depending on the thermal conditions in the Barents Sea. The Soviet-Norwegian symposium on the Barents Sea capelin. H. Gjøsæter Institute of Marine Research, Bergen, Norway, Bergen:135-148The Soviet-Norwegian symposium on the Barents Sea capelin

Railsback, S. F. 2001. Concepts from complex adaptive systems as a framework for individual-based modelling. Ecol. Modell. 139: 47-62.

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. Ecol. Modell. 80: 161-170.

Skjoldal, H. R., Gjøsæter, H., and Loeng, H. 1992. The Barents Sea ecosystem in the 1980s: ocean climate, plankton and capelin growth. ICES Mar. Sci. Symp. 195: 278-290.

Toresen, R. 2000. Havets ressurser 2000. Institute of Marine Research, Bergen, Norway

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. Fish. Res. 34: 27-37.

Toresen, R., and Østvedt, O. J. 2000. Variation in abundance of Norwegian spring-spawning herring (*Clupea harengus*, *Clupeidae*) throughout the 20th century and the influence of climatic fluctuations. Fish Fish. 1: 231-256.

Tyler, J. A., and Rose, K. A. 1994. Individual variability and spatial heterogeneity in fish population models. Reviews in Fish Biology Fisheries 4: 91-123.

Werner, E. E., and Gilliam, J. F. 1984. The ontogenetic niche of reproduction and species interactions in size-structured populations. Annual Review of Ecology and Systematics 15: 393-425.

Ådlandsvik, B., and Sundby, S. 1994. Modelling the transport of cod larvae from the Lofoten area. ICES Mar. Sci. Symp. 198: 379-392.

H Icelandic Waters Case Study

H.1 Definition of areas in Icelandic waters

Introduction

A hierarchical spatial structure is required which represents spatial areas suitable for use in Gadget and for storage of the data in the datawarehouse. These areas should reflect hydrography, bathymetry, species distributions and to a lesser extent fishing controls. The data should be aggregated at a level where use of the original data is seldom required.

The areas originally defined for Bormicon are described in Stefánsson et al (1997). For the datawarehouse it was decided that the smallest unit, the gridcell, should be a statistical subrectangle (1/2 degree longitude and 1/4 degree latitude). Subdivisions and divisions are composed of these gridcells.

For Gadget, the Bormicon areas, mapped onto gridcells, become the divisions and the boundaries between the subdivisions correspond to the 200m, 1000m and 1500m depth contours. Some modifications to this scheme were made to take into account fishing regulations, hydrography and the size of the resulting areas.

The appropriateness of this spatial structure, with repect to species composition, was tested using cluster analysis of species assemblage data.

Bormicon Areas

The original Bormicon areas, as described in Stefánsson et al (1997), are shown in fig. 1.

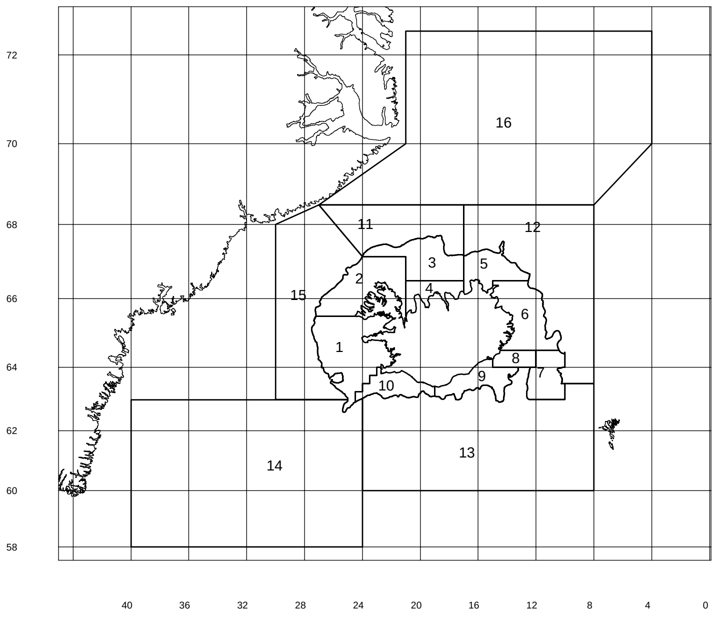


Figure 1: Bormicon areas

The GADGET Spatial Hierarchy

The numbering system for Gadget is slightly different to that for Bormicon. Divisions in Icelandic waters start at 100. So division 1 becomes 101 and the subdivisions of division 101 are 1011, 1012 etc.

The extent of the current spatial hierarchy has been partly determined by the data which have been included in the datawarehouse. As more species are input, the area may be extended.

Divisions

The divisions within the 500m depth contour are derived from the Bormicon areas (fig. 2, with gridcells being used as the smallest spatial unit. Some changes result from using a slightly different 500m depth contour. Other deviations from the original scheme are:

- the northern boundary between areas 2 and 3 is further south to follow the 200m depth contour.
- the southernmost section of the boundary between areas 2 and 4 is slightly further east to avoid splitting a shrimp fishery.
- areas 3 and 4 have been combined into a single division which is divided into subdivisions.
- area 8 has been incorporated into divisions 5 and 7. The reassignment of area 8 was based on the observed length distributions of cod and haddock.

Off the Icelandic coastal shelf, there are more differences between the Bormicon and Gadget spatial divisions. The Gadget divisions are predominantly derived from the bathymetry:

- the Reykjanes ridge separates divisions 109 and 114, which represent the Irminger Basin and Iceland Basin respectively.
- the Iceland-Faeroe Rise separates divisions 113 and 114.
- division 110 represents the Greenland coastal shelf to 500m depth.

Subdivisions

The coastal shelf subdivisions are based on the 200m depth contour (fig. 4). Exceptions to this are subdivisions 1011 and 1012 which represent regions with gear restrictions and 1054 (formerly area 8) which covers a frontal region. As some scientists prefer to analyse data from this region (1054) separately it has been included as a subdivision.

Subdivisions off the coastal shelf, are mainly in the vicinity of the Reykjanes ridge (south west) and the Iceland-Faeroe Rise (south east).

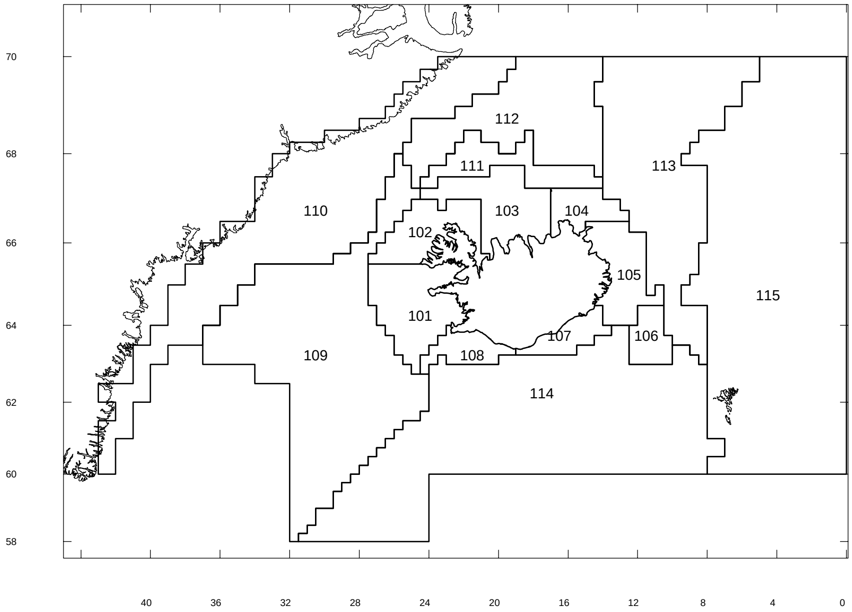


Figure 2: Gadget divisions

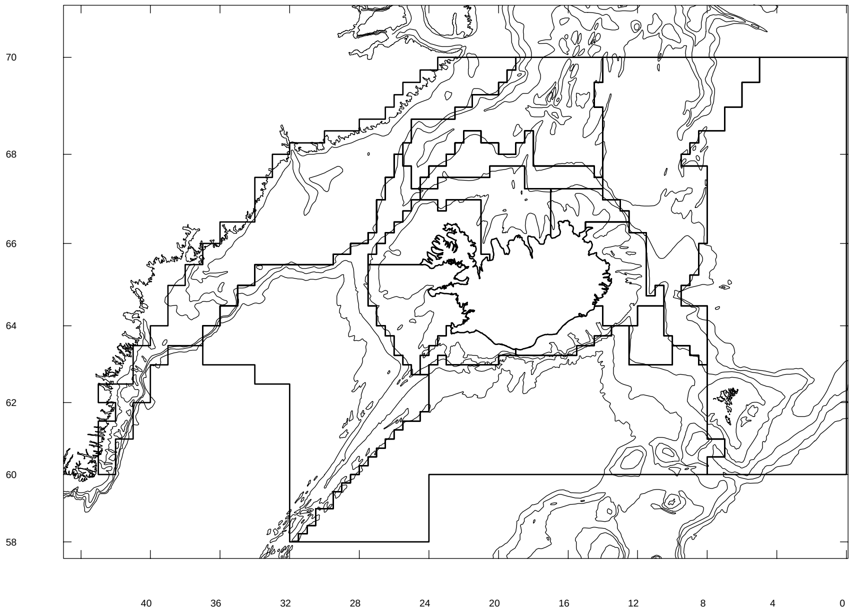


Figure 3: Gadget divisions with the 500m, 1000m and 1500m depth contours.

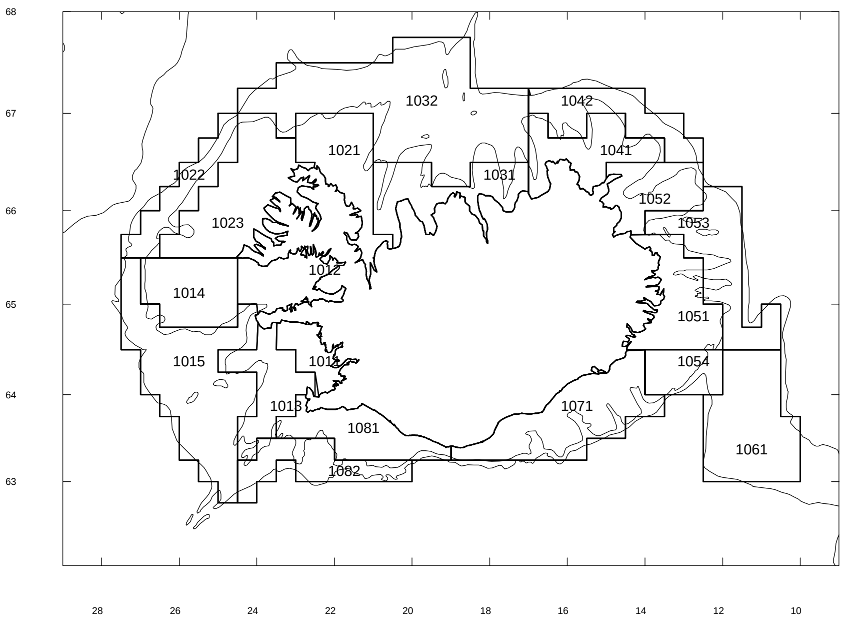


Figure 4: Gadget subdivisions within the 500m depth contour, with the 200m and 500m depth contours.

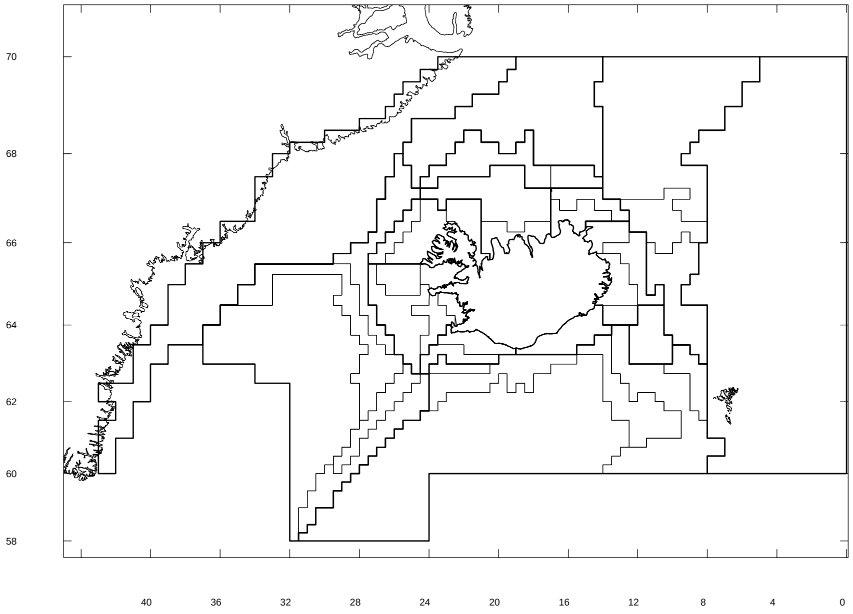


Figure 5: Gadget subdivisions

Data

Data from two surveys are used in the analyses: the Icelandic (Spring) Groundfish Survey from 1989 to 2002 and the Autumn Survey for the years 1996 to 2001. As the Groundfish Survey was designed for cod it is mostly on the Icelandic shelf covering depths to 500m. This area is represented by divisions 101-108. There are few data off the shelf and these were excluded from the analyses. The survey takes place in March each year and years prior to 1989 were excluded as species identification was inconsistent. The Autumn Survey includes some areas off the coastal shelf but 2002 was excluded as the number of species identified was considerably lower than previous years.

The species considered from both surveys are: cod, haddock, saithe, whiting, redfish (*Sebastes marinus*), Atlantic wolffish, starry ray, Greenland halibut, dab, long rough dab, moustache sculpin and *Sebastes viviparus*. Additional species in the analyses of the Autumn Survey data are roundnose grenadier, greater Argentine and deep water redfish (*Sebastes mentella*). Pelagic species such as capelin and herring were not included.

The abundance data were aggregated over all available years and within subrectangles, generating the total for each species and subrectangle. The fourth root of the total was then scaled to a mean of zero and variance equal to one. The surveys were analysed separately.

Cluster Analysis

Cluster analysis was used to define similar subrectangles. The partioning around medoids method “pam” in Splus 6.0 was applied to the scaled abundance data. No spatial restrictions were placed on clustering of the subrectangles and the spatial proximity of was not taken into account in the analyses.

Partioning around medoids operates on the dissimilarity matrix, minimizing the sum of disimmilarities. The required number of clusters is predefined, and an equivalent number of *medoids* are computed. Each observation is then assigned to the nearest *medoid* minimizing the sum of dissimilarites of each observation to the associated medoid.

A silhouette score is calculated for each observation which represents the difference between its average dissimilarity to other observations in the cluster it is assigned to and its average dissimilarity to observations in the second best cluster. The silhouette score is therefore an indication of the appropriateness of the classification. Average silhouette scores can be calculated for all observations providing an indication of the most appropriate number of clusters and by cluster to indicate which clusters are poorly or well defined.

Clustering was carried out on all the data together and then for the northern and southern regions separately.

Results

March groundfish survey

Cluster analysis of the entire region into 3 clusters indicates differences between the north (102, 103, 104, 105), south (101, 107, 108) and division 106 (table 1, fig. 6). Cluster 1 predominantly represents 106, cluster 2 the southern area and cluster 3 the north.

The main effect of further splitting the area into 5 clusters is a separation in the north, with divisions 102 consisting predominantly of cluster 5 and 103 and 104 consisting predominantly of cluster 4 (table 2, fig. 7).

Clustering the northern divisions (102, 103, 104, 105) into 4 clusters reveals some inshore-offshore differences (fig. 8). Cluster 4 is coastal (subdivisions 1021, 1022, 1031) and cluster 1 is predominantly

	101	102	103	104	105	106	107	108	North	South	106
1	11	5	12	3	6	16	6	0	26	17	16
2	47	9	1	0	4	0	20	24	10	91	0
3	0	34	41	23	40	1	0	0	138	0	1

Table 1: March groundfish survey data in 3 clusters by division and by north and south

	101	102	103	104	105	106	107	108
1	11	5	9	3	6	16	5	0
2	37	1	1	0	1	0	14	17
3	4	1	2	0	0	0	6	7
4	0	6	30	15	20	1	0	0
5	6	35	12	8	23	0	1	0

Table 2: March groundfish survey data in 5 clusters by division

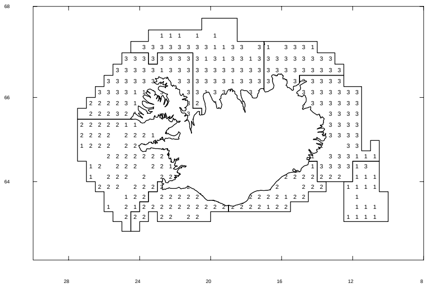


Figure 6: March groundfish survey data in 3 clusters

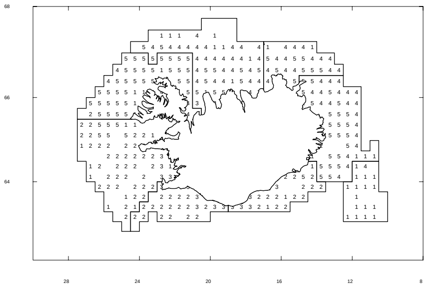


Figure 7: March groundfish survey data in 5 clusters

the offshore parts of division 102 and subdivision 1054 (previously area 8) (table 8).

	1021	1022	1023	1031	1032	1041	1042	1051	1052	1053	1054
1	4	12	9	1	7	2	0	3	1	2	6
2	1	2	0	1	12	1	2	2	1	5	1
3	6	6	1	7	18	10	10	6	10	10	0
4	2	5	0	8	0	1	0	2	0	1	0

Table 3: March groundfish survey data: northern divisions in 4 clusters by subdivision

Clustering the southern divsions (101, 107, 108) into 3 clusters also indicates an inshore-offshore pattern of species composition (fig. 9) with cluster 3 only being found along the coast (subdivisions 1011 and inshore parts of 107 and 1081) and cluster 2 mainly found in subdivision 1015 the most offshore cluster (table 4).

	1011	1012	1013	1014	1015	1071	1081	1082
1	1	8	3	10	13	16	8	10
2	2	2	1	2	11	4	0	0
3	3	1	1	0	0	6	6	0

Table 4: March groundfish survey data: southern divisions in 3 clusters by subdivision

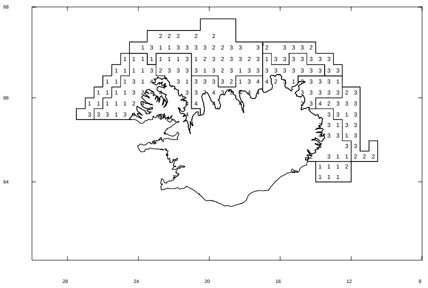


Figure 8: March groundfish survey data from the northern divisions in 4 clusters

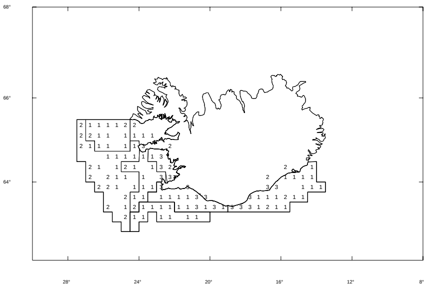


Figure 9: March groundfish survey data from the southern divisions in 3 clusters

Autumn survey

Cluster analysis of the entire region into 2 clusters, separates the area into on-shelf and off-shelf (fig. 10). With 5 clusters (table 5, fig. 11) the off-shelf area is split into 2 areas. There is a southwest off-shelf cluster (cluster 1 in table 5) incorporating divisions 109 and 114 (along with some of 101 and 108) and in the north cluster 2, which mainly included 111 and 113 with the outer area of 103. The deeper part of division 101 is represented by cluster 3, with the shallow part of 101 along with 108 and 107 in cluster 4. Cluster 5 represents most of divisions 102, 103, 104 and 105.

	101	102	103	104	105	106	107	108	109	110	111	112	113	114
1	4	1	0	0	0	2	3	6	31	0	0	0	0	10
2	0	1	17	6	5	4	0	0	9	3	22	3	16	3
3	27	6	0	0	2	2	2	4	1	0	0	0	0	2
4	8	1	1	0	0	0	5	8	0	0	0	0	0	0
5	3	21	26	13	19	0	3	0	0	1	0	0	0	0

Table 5: Autumn survey data in 5 clusters by division

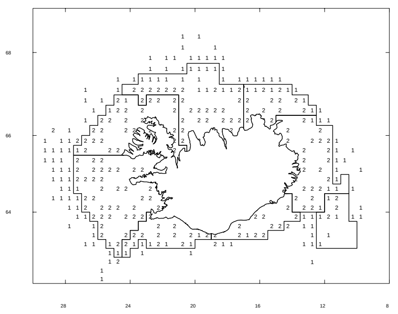


Figure 10: Autumn survey data in 2 clusters

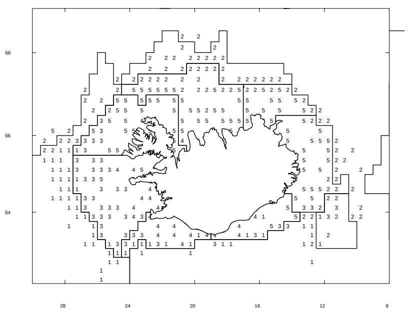


Figure 11: Autumn survey data in 5 clusters

Cluster analysis of the northern area into 5 clusters mainly indicates differences related to distance from shore/depth (tables 6 & 7 and figure 12). Cluster 5 is closest to shore, with clusters 2, 4, 3 and 1 tending to move progressively further offshore.

	102	103	104	105	110	111	112	113
1	1	2	0	0	2	13	3	9
2	23	5	4	10	0	0	0	0
3	1	15	6	4	1	9	0	7
4	1	14	7	10	1	0	0	0
5	4	8	2	2	0	0	0	0

Table 6: Autumn survey data: north divisions in 5 clusters

	1021	1022	1023	1031	1032	1041	1042	1051	1052	1053	1054
1	0	1	0	1	1	0	0	0	0	0	0
2	6	8	9	2	3	4	0	5	1	0	4
3	0	0	1	0	15	0	6	0	1	2	1
4	0	0	1	1	13	3	4	0	2	8	0
5	1	3	0	8	0	2	0	2	0	0	0

Table 7: Autumn survey data: northern on-shelf subdivision in 5 clusters (off-shelf divisions as in table 6)

Clustering the southern region into 6 clusters also produces clusters indicating differences in species composition relating to distance from shore/depth (fig. 13).

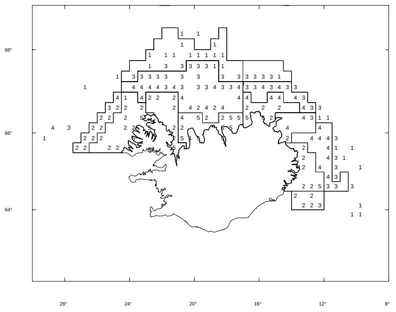


Figure 12: Autumn survey data from northern divisions in 5 clusters

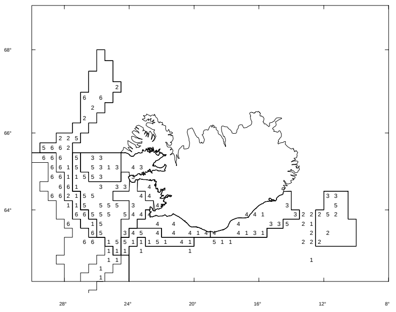


Figure 13: Autumn survey data from the southern divisions in 6 clusters

Discussion

Distinct differences in species composition are indicated between the northern divisions (divisions 102, 103, 104, 105), the southern divisions (101, 107, 108), division 106 and those off the Icelandic shelf (deeper than 500m). Within the northern and southern regions, differences in species composition are related to distance from shore or depth. The clusters opbtained from the survey data are in sufficient concordance with the divisions and subdivisions for these to be considered appropriate for data storage and area disaggregated modelling.

I Celtic Sea Case Study

I.1 Defining spatial and seasonal units of fleet dynamics for the French Demersal fleet fishing in the Celtic Sea

S. Mahevas, V. Trenkel and C. Métaireau
Ifremer, Laboratoire Maerha, Nantes, France

Introduction

Using generalised mixed linear models (GLMM) to describe the observed variation in space (statistical rectangles) and time (month) of the fishing time of the French trawlers operating in the Celtic Sea, we had quantified month and rectangle effects by the model coefficients (Mahévas and Trenkel 2002, DST2 Annual progress report 2001, E.4). Here we report on the continuation of the work which led to the definition of homogenous areas and seasons for the fishing activity (Métaireau 2002).

Methods

We construct homogeneous fishing areas and fishing seasons by a cluster analysis on the estimated month and rectangle coefficients obtained from a GLMM (Piñheiro and Bates 2000). We employed an ascending hierarchical cluster analysis based on ordered subtrees (Lebart *et al.* 1997). To prevent non adjacent rectangles or distant months months from being grouped, we imposed the constraint of spatial (respectively temporal) contiguity using as neighbours of a rectangle the eight adjacents rectangles and the previous and following months as neighbours for a given month.

Given the maximum likelihood method used to fit the GLMM models (Mahévas and Trenkel 2002), we can use a Wald test not only to test the nullity of the estimated coefficients β_i but also to test the equality of two or more coefficients (e.g. $\beta_i = \beta_j$). Hence, the distance metric chosen for clustering two rectangles (or two months) is 1-pvalue of the Wald test, where the test is for the equality of the two corresponding estimated coeficients:

$$\mathrm{H}_0{:}L\beta = c, \chi^2_{\mathrm{Wald}} = (\mathrm{L}\hat{\beta} - \mathrm{c})'(\mathrm{LVar}(\hat{\beta})\mathrm{L}')^{-1}(\mathrm{L}\hat{\beta} - \mathrm{c}) \sim \chi^2(\mathrm{rank(L)})$$

where L is the vector (1,-1) and Beta is the vector (β_i, β_j) and c=0

Clustering algorithm

1. Define *n* initial clusters D_0=β_1,…,β_n with their corresponding matrix L_i.
2. Iterate on *i* (*i*=1,…*n*) until only one cluster remains (i.e. D=β_1,…,β_n)
    (a) Define the set of all possible couples of neighbouring clusters
    (C_l,C_r) $\in$ (C_l,C_r) : C_l,C_r $\in$ D_i ,l $\geq$ r , C_r $\in$ neighbours(C_l)
    (b) For the set of couples (C_l,C_r) test whether their coefficients are different
    H_0 : β_i = β_j $\forall$ (i,j) $\in$ (C_l,C_l).
        - for each individual couple (β_i, β_j) $\in$ (C_l,C_r) (for instance the k^{th}), complete the rows of L (0,…,0,1,0,…,0,-1,0,…,0), L_{ki}=1, L_{kj}=-1,

- calculate the Wald statistic and the p-value :
  χ^2_{Wald} = $\hat{\beta}'$ L$'$(L Var($\hat{\beta}$)L$'$)$^{-1}$L$\hat{\beta}$
  P(χ^2(rang(L))> $\chi^2{}_{Wald}$)

(c) Aggregate the two clusters (C$_l{}^{max}$,C$_r{}^{max}$) with the highest p-value.
Then D$_i$=D$_{i-1}$ C$_l{}^{max}$,C$_r{}^{max}\cup$ C$_l{}^{max}\cup$C$_r{}^{max}$.

Results

The application of the clustering algorithm to the estimated rectangle coeficients allowed us to define 16 fishing zones (Figure 1) and seven seasons (Figure 2) at a probability level of 0.1 (=pvalue in Wald test). The size of the fishing zones varies substantially: the costal zones appear to be smaller than the off-shore zones. The statistical rectangle 31E3 is on its own; it is characterised by consistently the largest fishing times. Concerning the seasons, fishing activity is heterogenous in winter and was characterised by four seasons. It is more homogenous at the other times of the years, thus a spring, summer and autumn season were identified. December stands out from the other months as it is the period with the lowest fishing times.

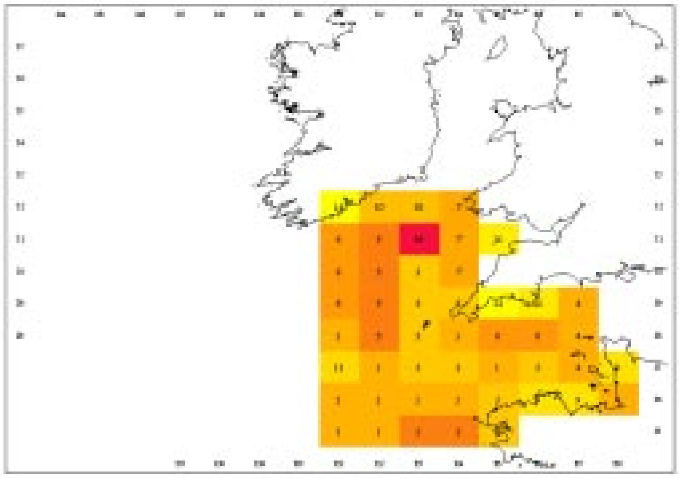


Figure 1: Homogenous fishing zones for fishing activity of French trawlers in the Celtic Sea (test level 1-α=0.9)

Figure 2: Homogenous fishing seasons for fishing activity of French trawlers in the Celtic Sea (test level 1-α=0.99)

References

Lebart, L., A. Morineau, et al. (1997). Statistique exploratoire multidimensionnelle. Paris, Dunod.

Mahévas, S. and V. Trenkel (2002). “Utilisation de modèles mixres pour décrire la distribution spatio-temporelle du temps de pêche e la flottille française en mer Celtique.” Journal de la Socitété Française de Statistique **143**: 177-186.

Métaireau C., 2002. Description de la dynamique spatio-temporelle de l’effort de pêche en mer Celtique. D.E.S.S. Modélisation et Analyse Statistique de l’Information, Université de Bretagne Sud - Vannes. 43 p.

Piñheiro, J. and D. Bates (2000). Mixed-effects models in S and S-plus. Berlin, Springer Verlag.

I.2 Presence/absence mapping and modelling

Julia L Blanchard, David Maxwell and Brian Rackham

CEFAS Laboratory, Lowestoft, United Kingdom

Introduction

Spatio-temporal mapping and statistical analyses were carried out for several fish species present in the Celtic Sea Gadget model. The purpose of this work package was descriptive rather than predictive with an aim to evaluate whether spatial units should be considered as part of the model structure. The relevant dst^2 workpackage for this annexe was “4.2 : Spatio-temporal scales”. This work is preliminary and will continue over the next few months.

Methods

Data collected from the annual Celtic Sea groundfish surveys were used to map the presence and absence of species for each year. The species included in this analysis were: cod, haddock, whiting, megrim (2 species), monkfish (2 species), Nephrops, blue whiting, mackerel and horse mackerel. The demersal species (cod, haddock whiting megrim and monkfish) were divided into two size groups (above and below length at 50% maturity (L_{50})) whereas the pelagic species and Nephrops were not divided into separate size groups.

Lengths at 50% maturity were predicted from maturity ogives constructed from maturity and length data collected over the 1986-2001 period for cod, haddock, whiting and megrim using logistic generalised linear models.

Generalised additive models were used to model the probability of presence using a binomial family with a logit link function. Probability of presence was predicted by longitude and latidude as an interaction term. Several types of year effects were also tested, expressing the year: linearly, with a loess smoother and as a factor (i.e. each year considered explicitly).

Results and Discussion

Maturity ogives predicted for cod, haddock, whiting and megrim (females only) are shown in Fig.1. L_{50} values (to the nearest cm) for cod, haddock and whiting were: 61 (0.56 s.e.), 24 (0.44 s.e.), 21 (0.35 s.e.) cm, respectively (Fig. 1). For species with strong sexual dimorphism (monkfish and megrim) female L_{50} was used. L_{50} for megrim was calculated as 24 (0.2 s.e.) cm and for monkfish was 73 cm based on information reported in Hislop et al. (2001). However, since very few monkfish were observed above this length they were not divided into separate size groups.

Observations of spatial presence-absence for each year and species are shown in Figures 2 -16. Survey stations were fairly consistent from 1985 to 2002 with the exception of 1994 when far fewer stations were sampled. Haddock, cod and whiting and megrim appeared to be present at both the shelf edge and central areas of the Celtic Sea although megrim tended to be closer to the shelf edge areas compared to the other species. Megrim had a much higher frequency of presence compared to cod, haddock and whiting. Mature fish were nearly always present at fewer stations than immature fish, reflecting the more frequent occurrence of smaller fish in the area. This is especially noticeable for cod. However, the L_{50} value use to separate cod size groups could have been overly high. Although it was a mean value there was a large degree of variation in cod length at maturity during the period covered by the maturity data (1986-2001). Also, the standard error for the L_{50} estimate was highest for cod (0.56).

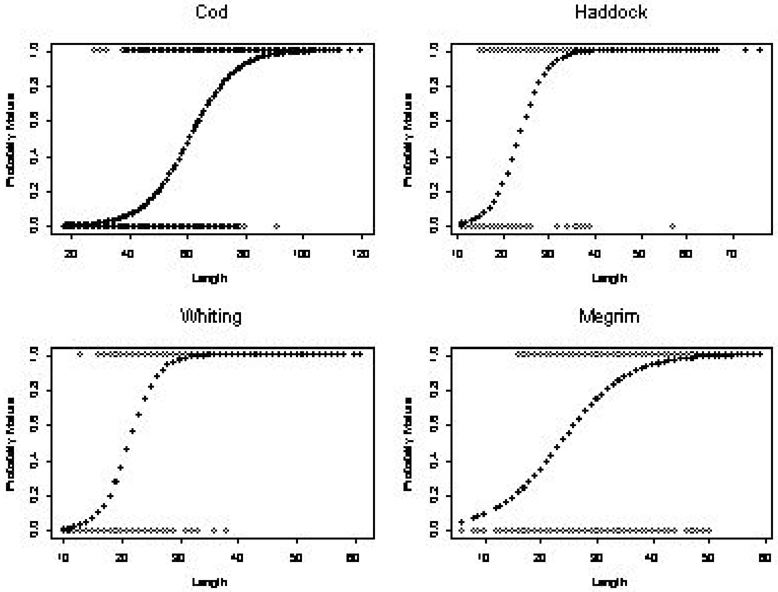


Figure 1: Maturity ogives used to predict L50 for cod, haddock, whiting and female megrim.

Monkfish (*Lophius piscatorius*) were present at both shelf-edge and shelf areas but white anglerfish (*Lophius budegassa*) were present more often on the shelf edge. Four- spotted megrim also was more prominently a shelf-edge species. Pelagic species: mackerel, horse mackerel and blue whiting had a much greater degree of presence compared to the other species and were also most noticeable along the shelf edge. Nephrops appeared to be less present but with a spatial prevalence in the central and northeastern part of the Celtic Sea.

There were noticeable differences in the spatial patterns of presence/absence across years for many of the species and these may reflect changes in the total population abundances. However, this is not possible to detect with presence/absence data alone.

Generalised additive models were used to predict the probability of presence using a loess smooth function with latitude and longitude expressed as an interaction term. Models were based on the time series for which sufficient sampling replication existed (1985-2002) and year effects were also tested. Unsurprisingly, year effects were significant for most of the species considered ($p<0.05$, ANOVA). However, the overall spatial pattern (all years combined) for each species is shown by predictive GAM surfaces (section A). These plots show a distinct dominance of pelagic species (such as mackerel, horse mackerel and blue whiting) in the Celtic Sea compared to demersal species. According to these models, there appears to be clear shelf edge species (megrim, four spotted megrim, white anglerfish, mackerel, horse mackerel and blue whiting) compared to those more frequently found throughout the Celtic Sea and towards the northeastern area (haddock, cod, whiting, monkfish, nephrops).

Figure 2: UK (E & W) Surveys Haddock Mature

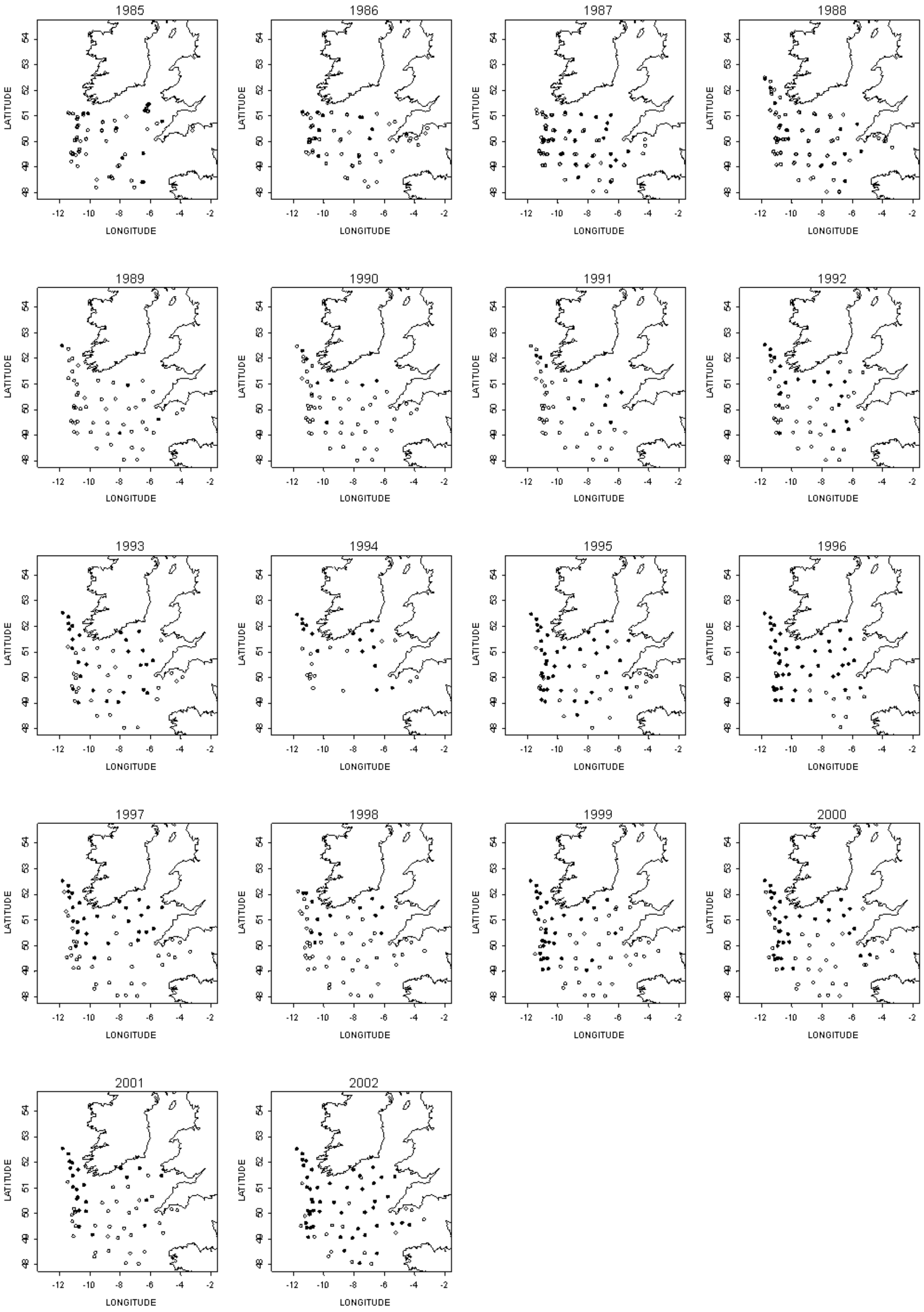


Figure 3: UK (E & W) Surveys Haddock Mature

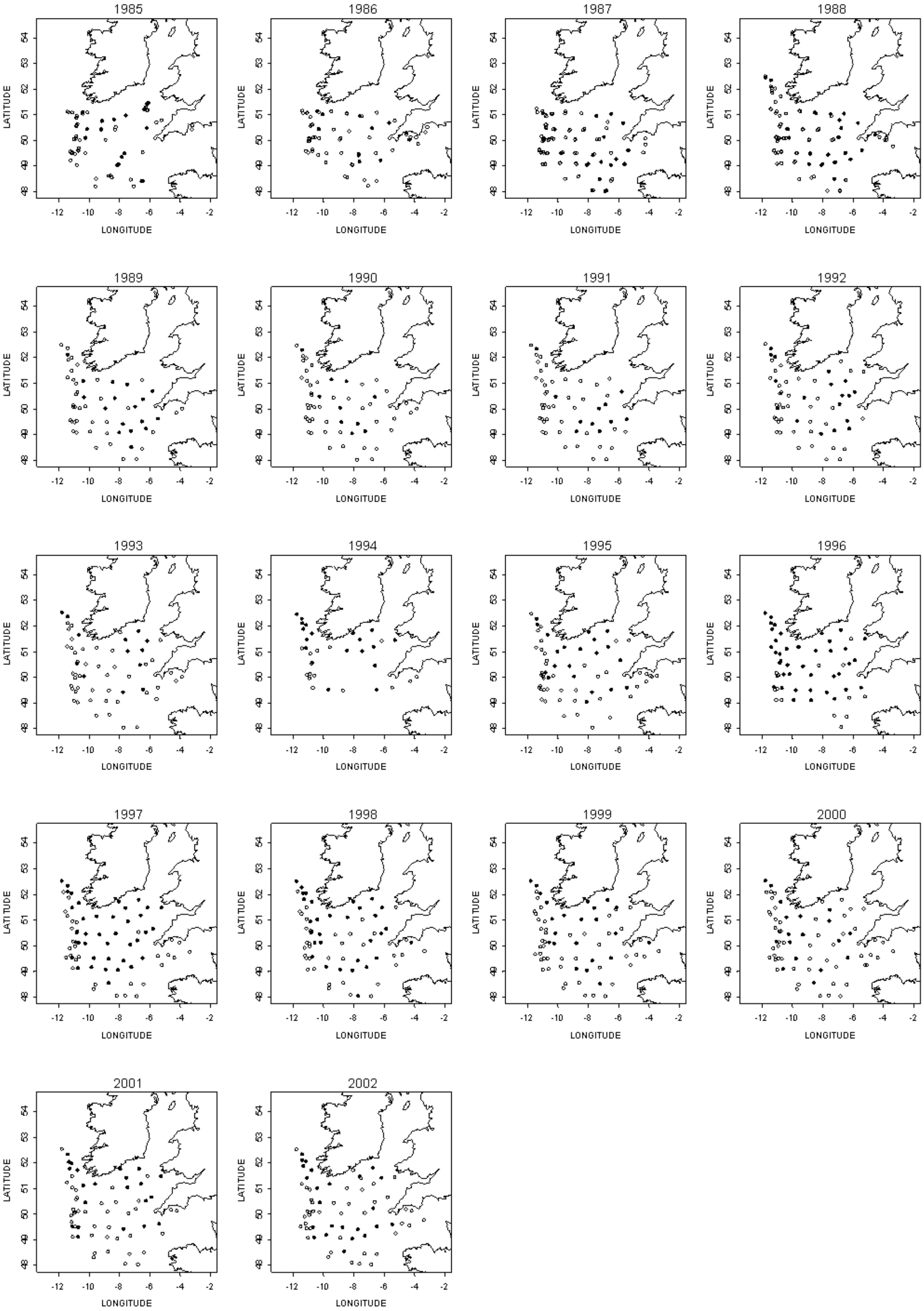


Figure 4: UK (E & W) Surveys Cod Immature

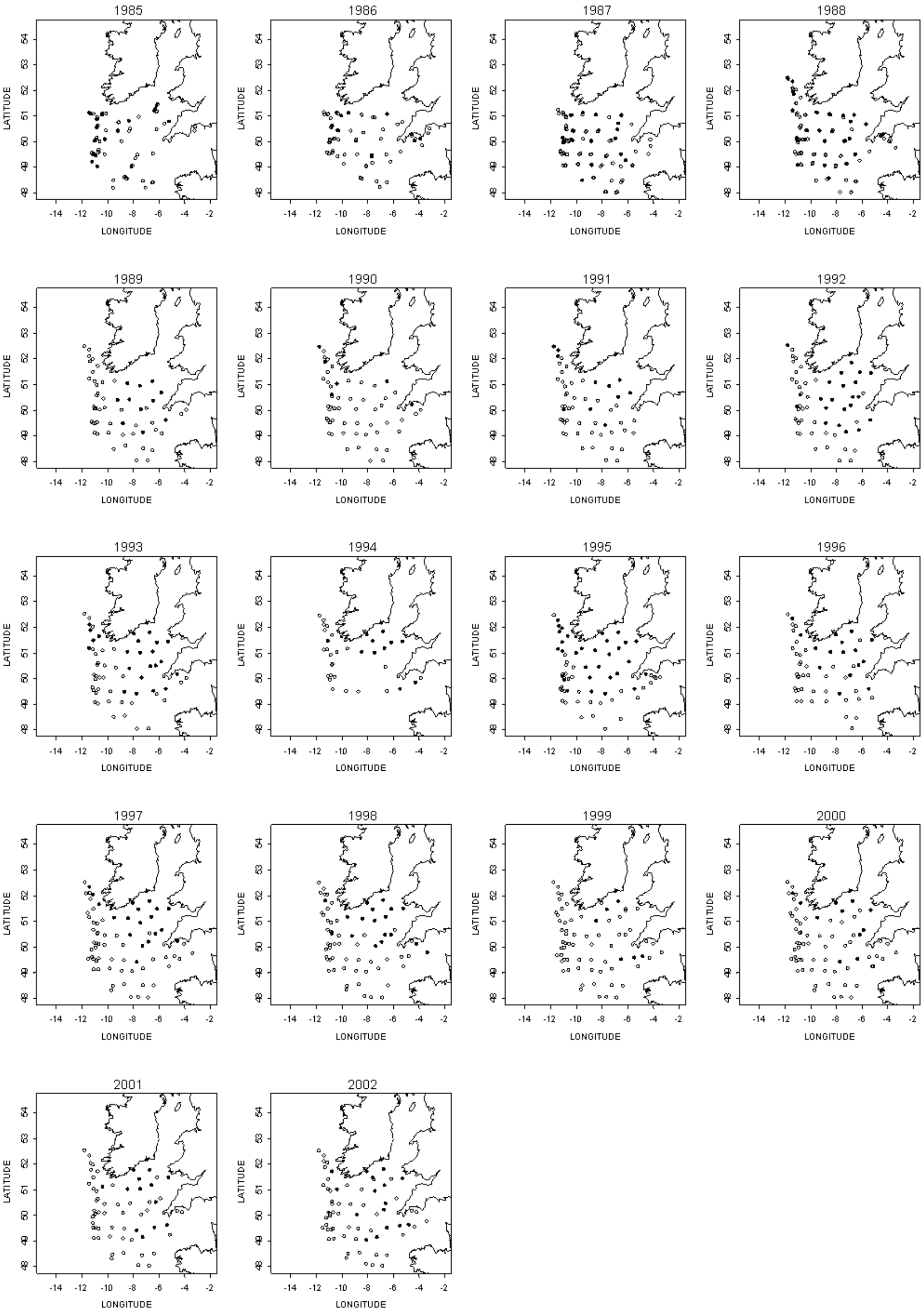
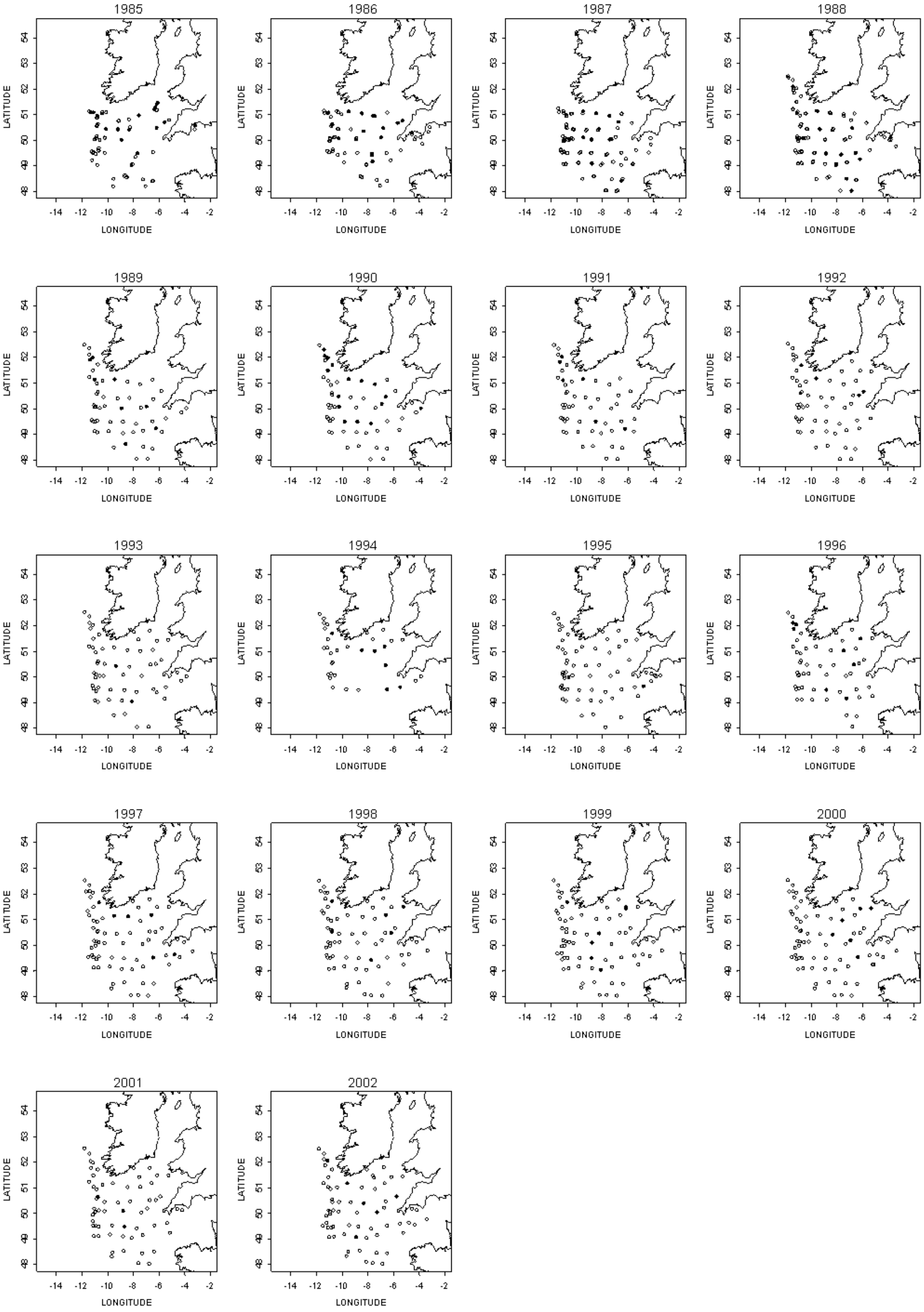


Figure 5: UK (E & W) Surveys Cod Mature



1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002

Figure 6: UK (E & W) Surveys Whiting Immature

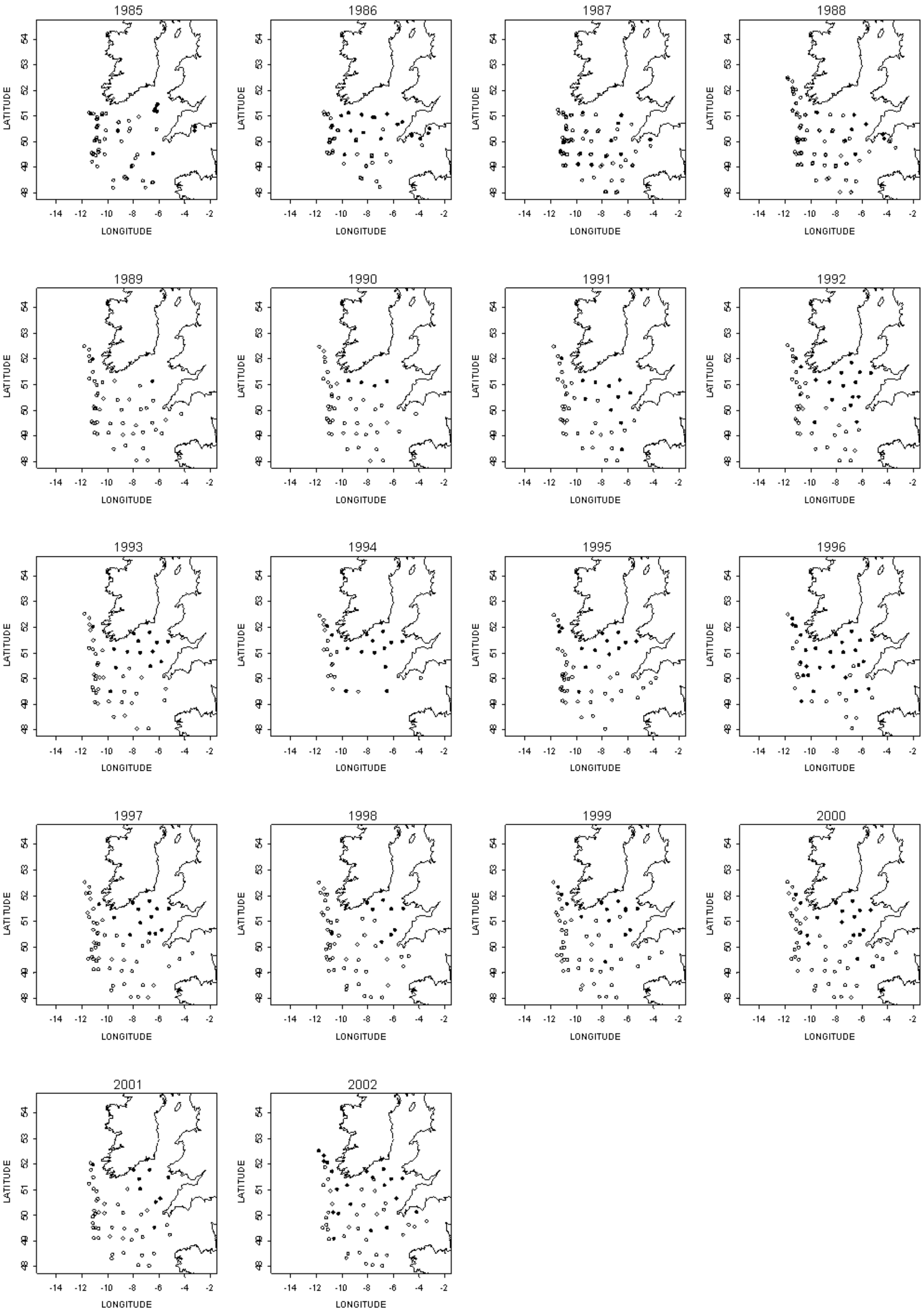


Figure 7: UK (E & W) Surveys Whiting Mature

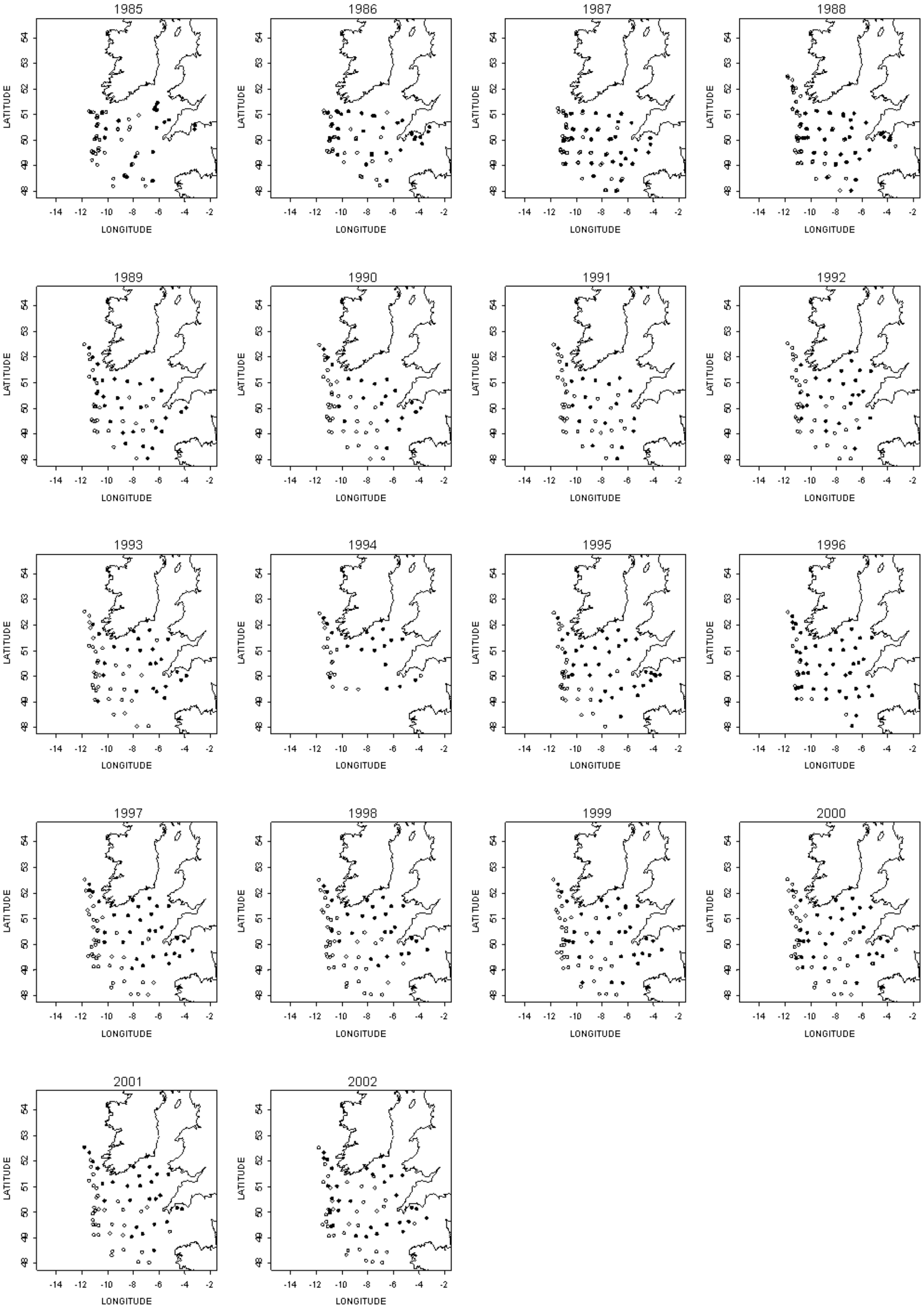


Figure 8: UK (E & W) Surveys Megrim Immature

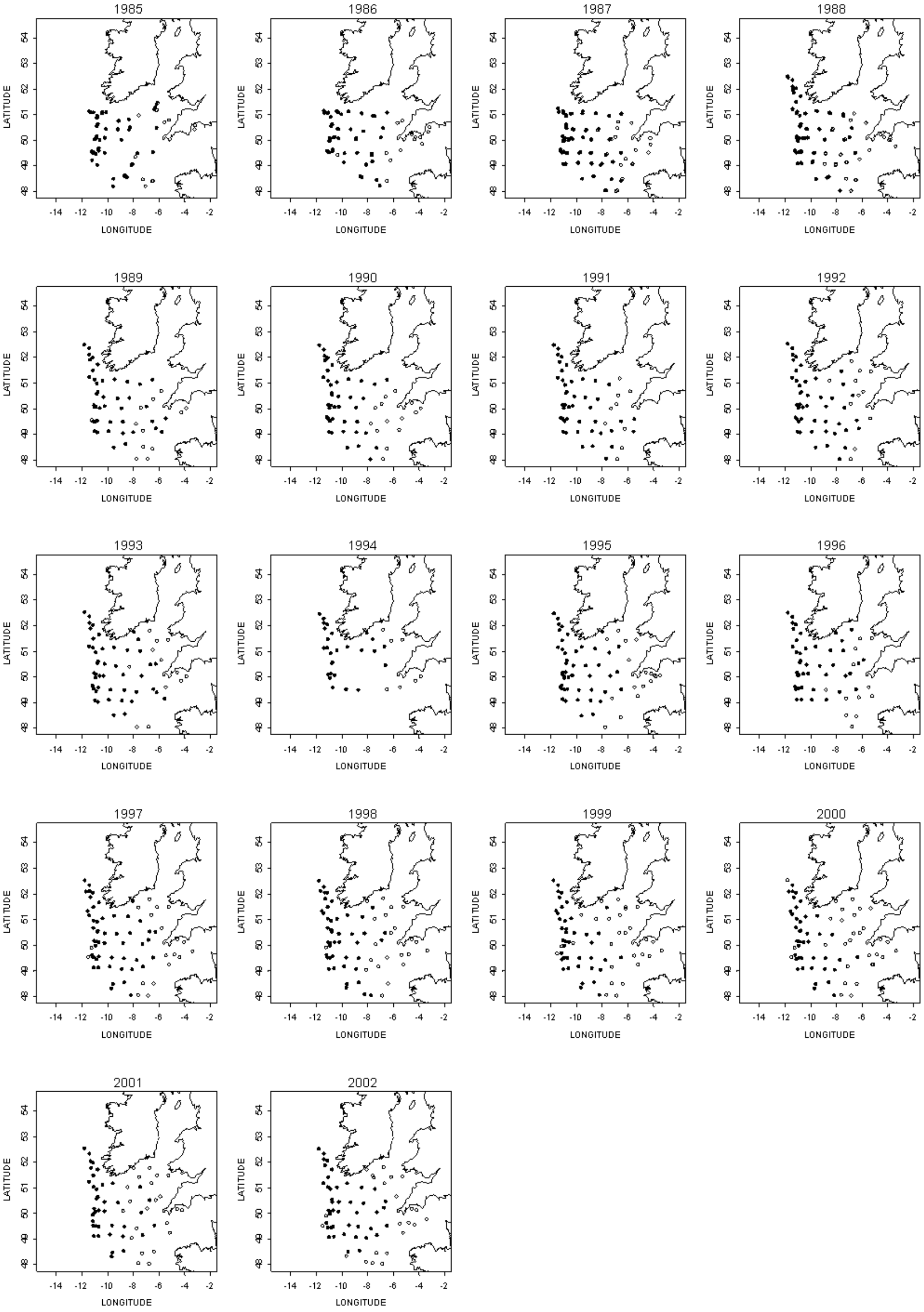


Figure 9: UK (E & W) Surveys Megrim Mature

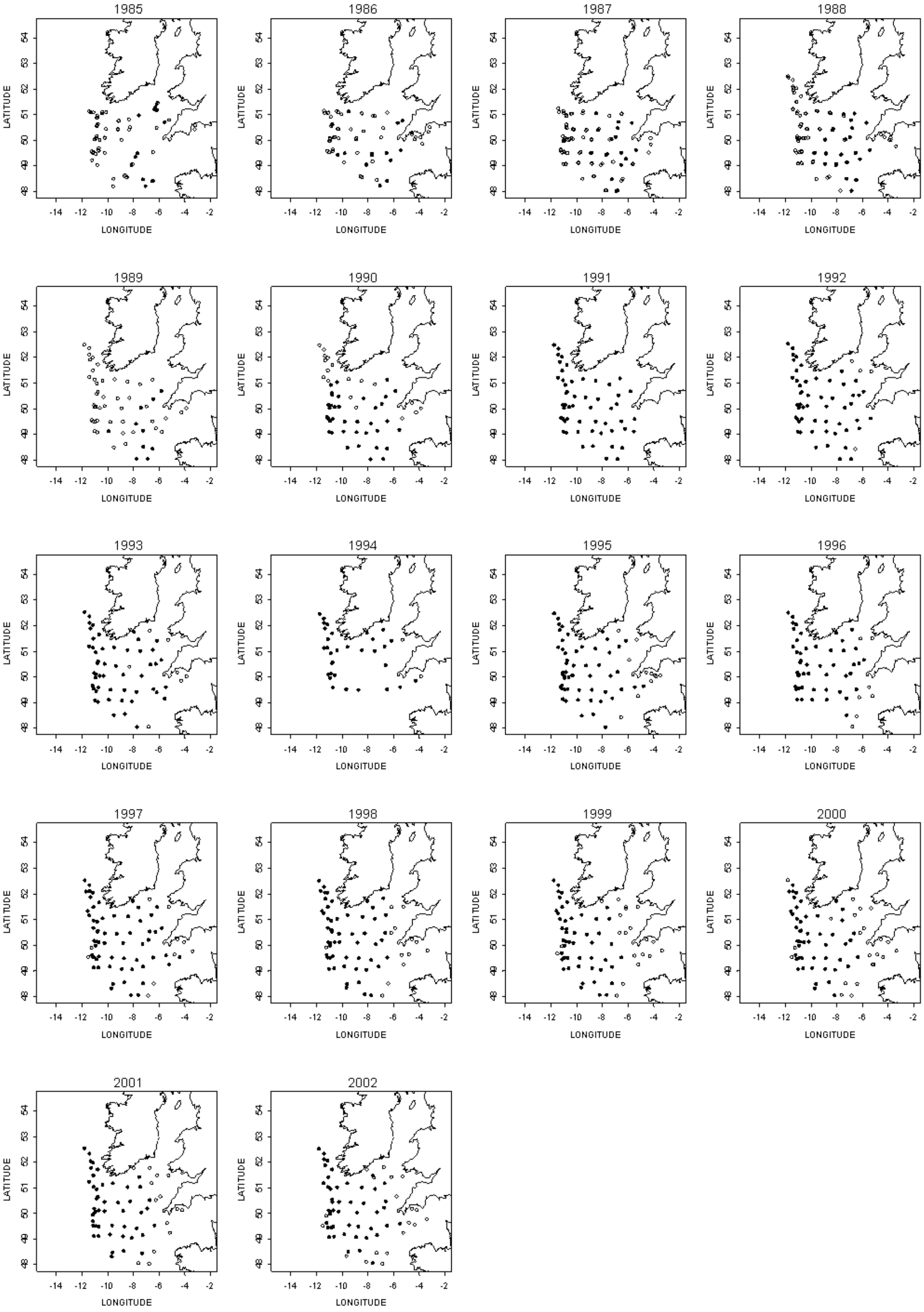


Figure 10: UK (E & W) Surveys Monkfish (Lophius piscatorius)

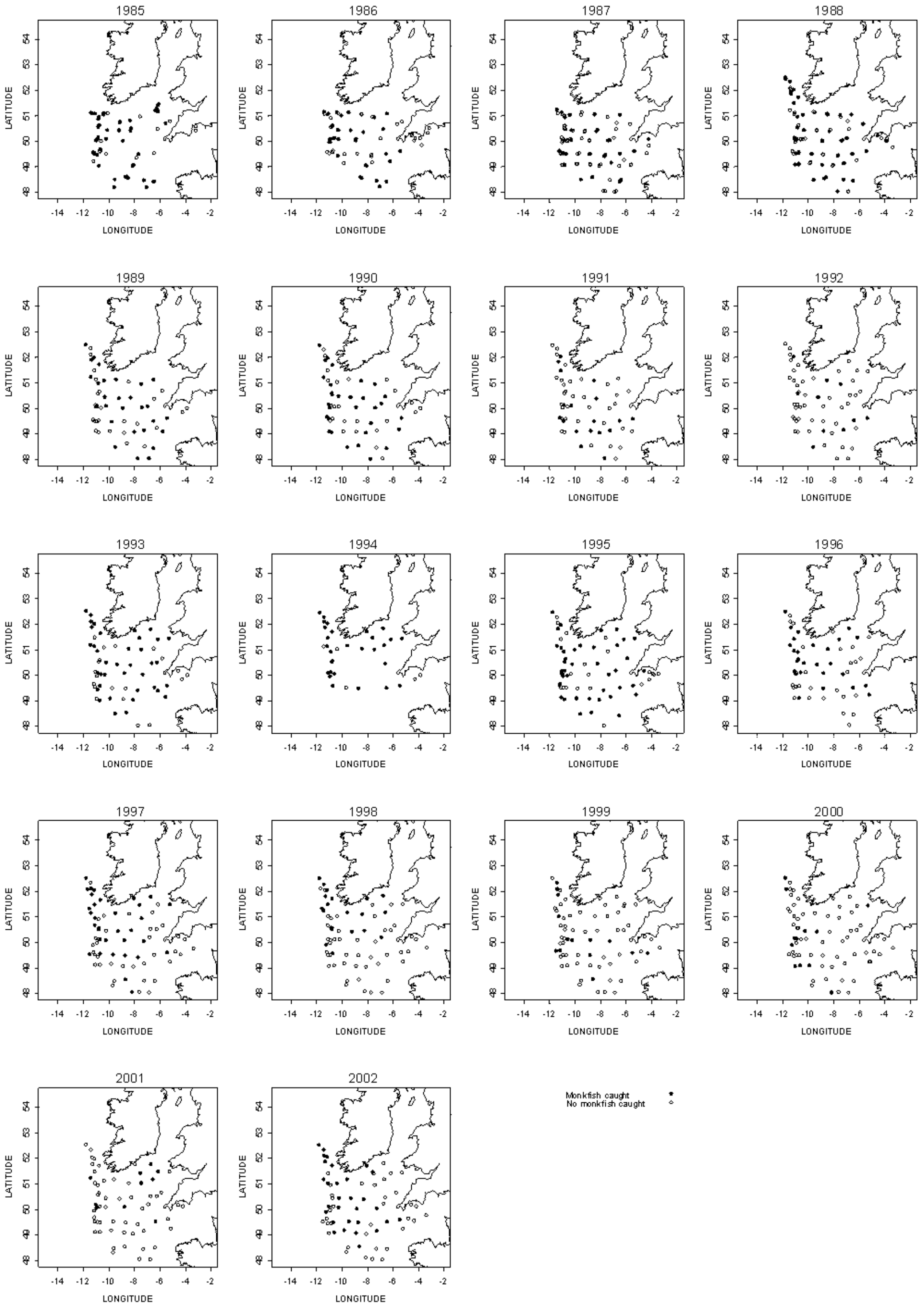


Figure 11: UK (E & W) Surveys White anglerfish (Lophius budegassa)

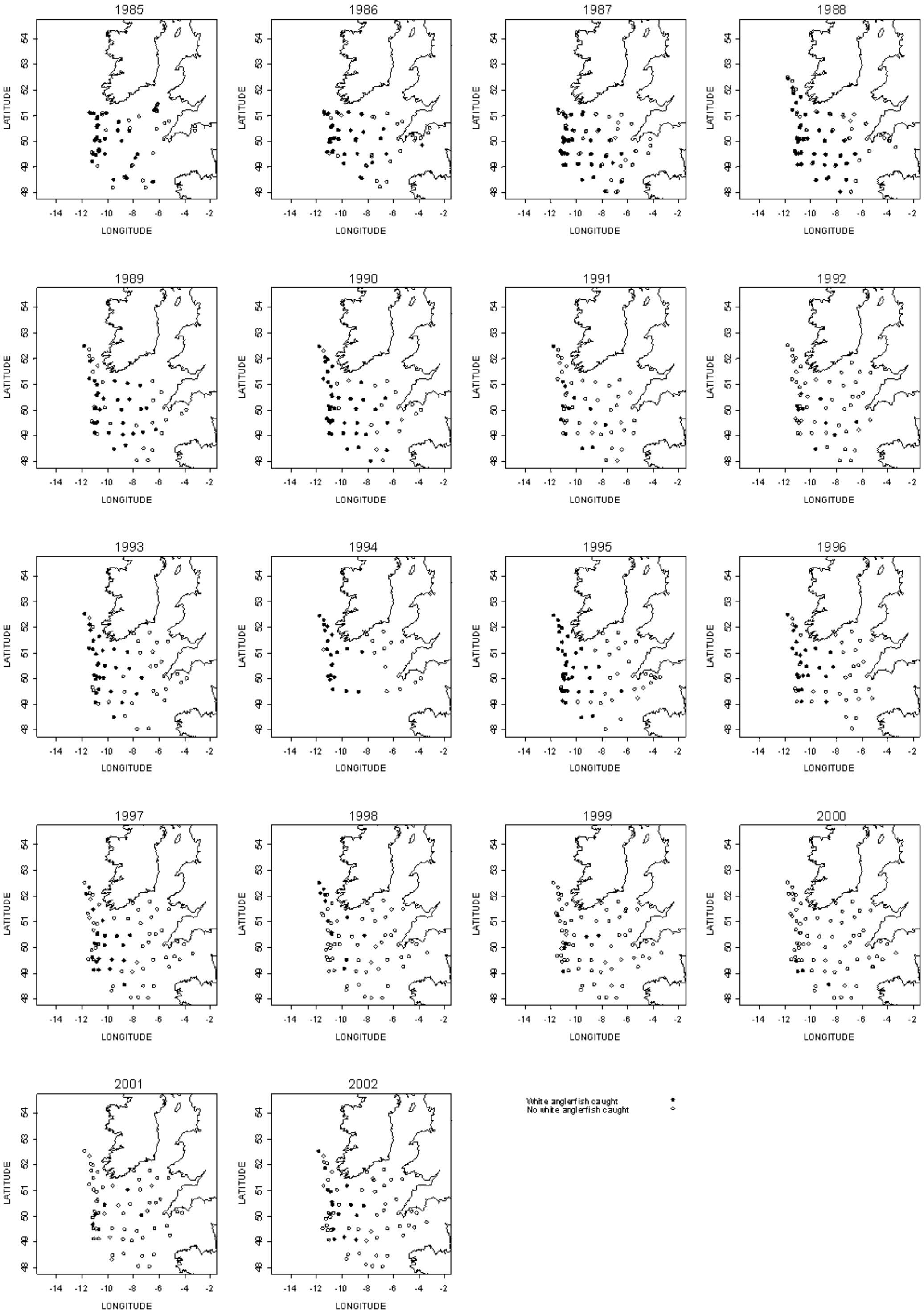


Figure 12: UK (E & W) Surveys Four spotted megrim (Lepidorhombhus boscii)

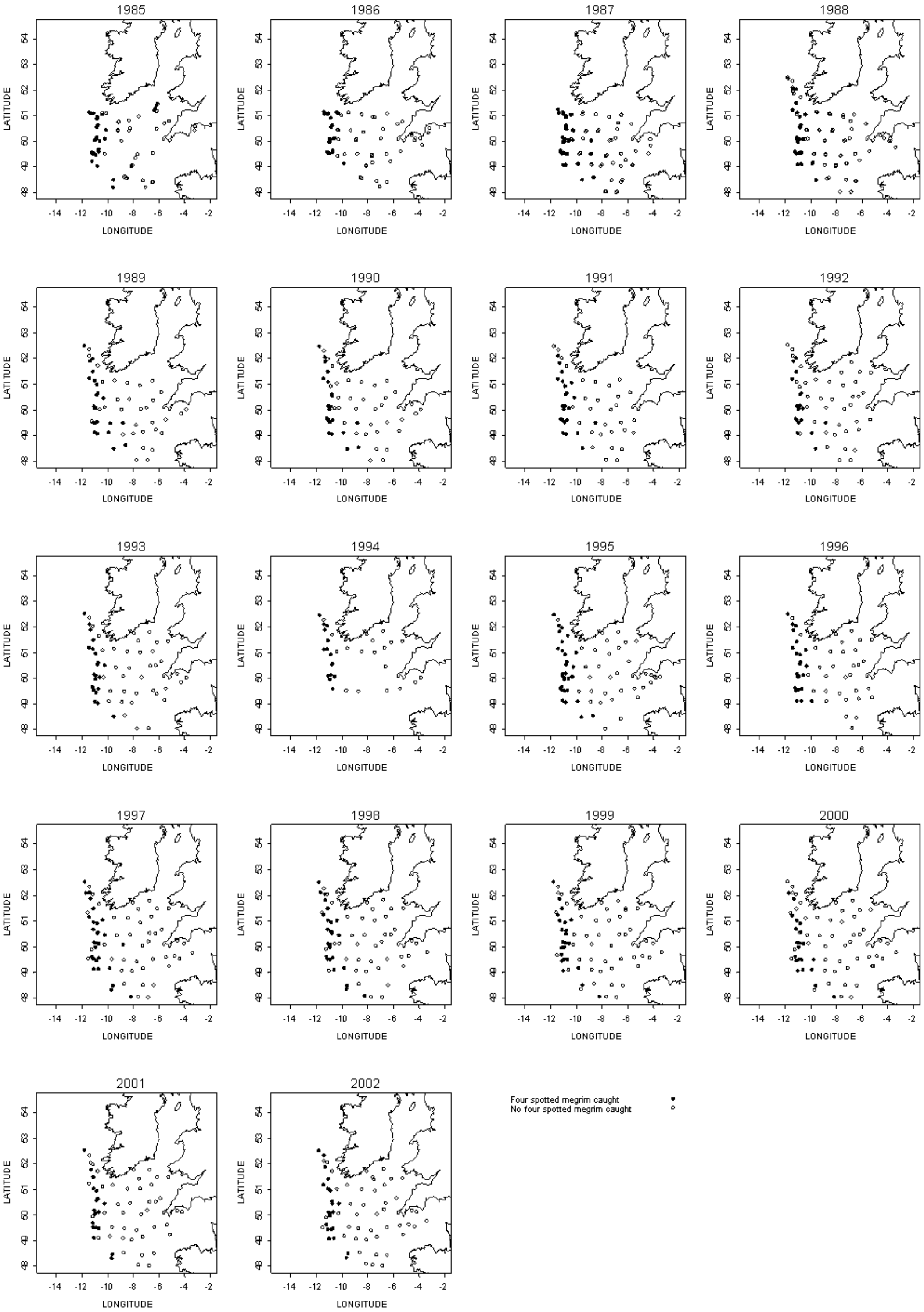


Figure 13: UK (E & W) Surveys Mackerel

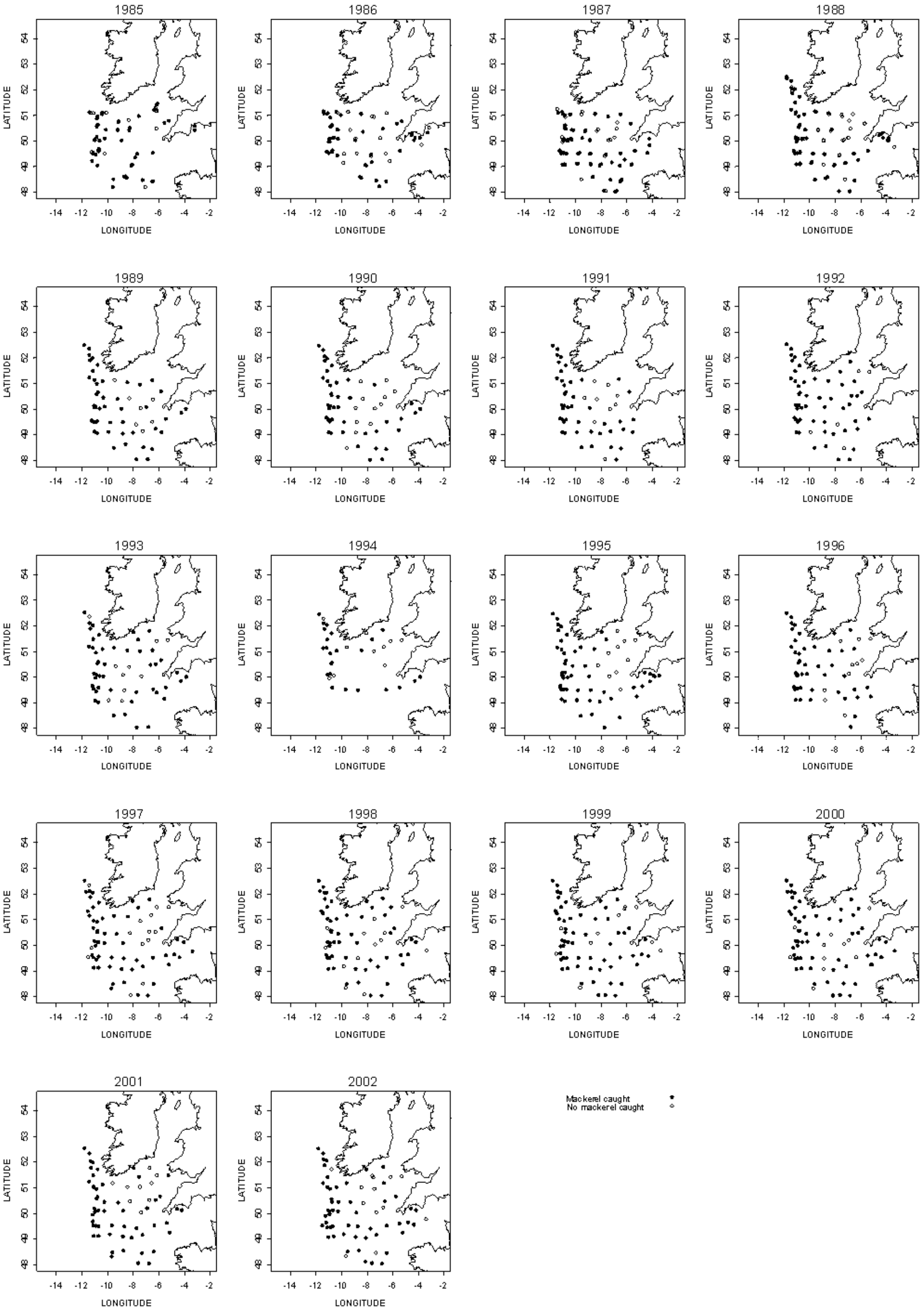


Figure 14: UK (E & W) Surveys Horse Mackerel

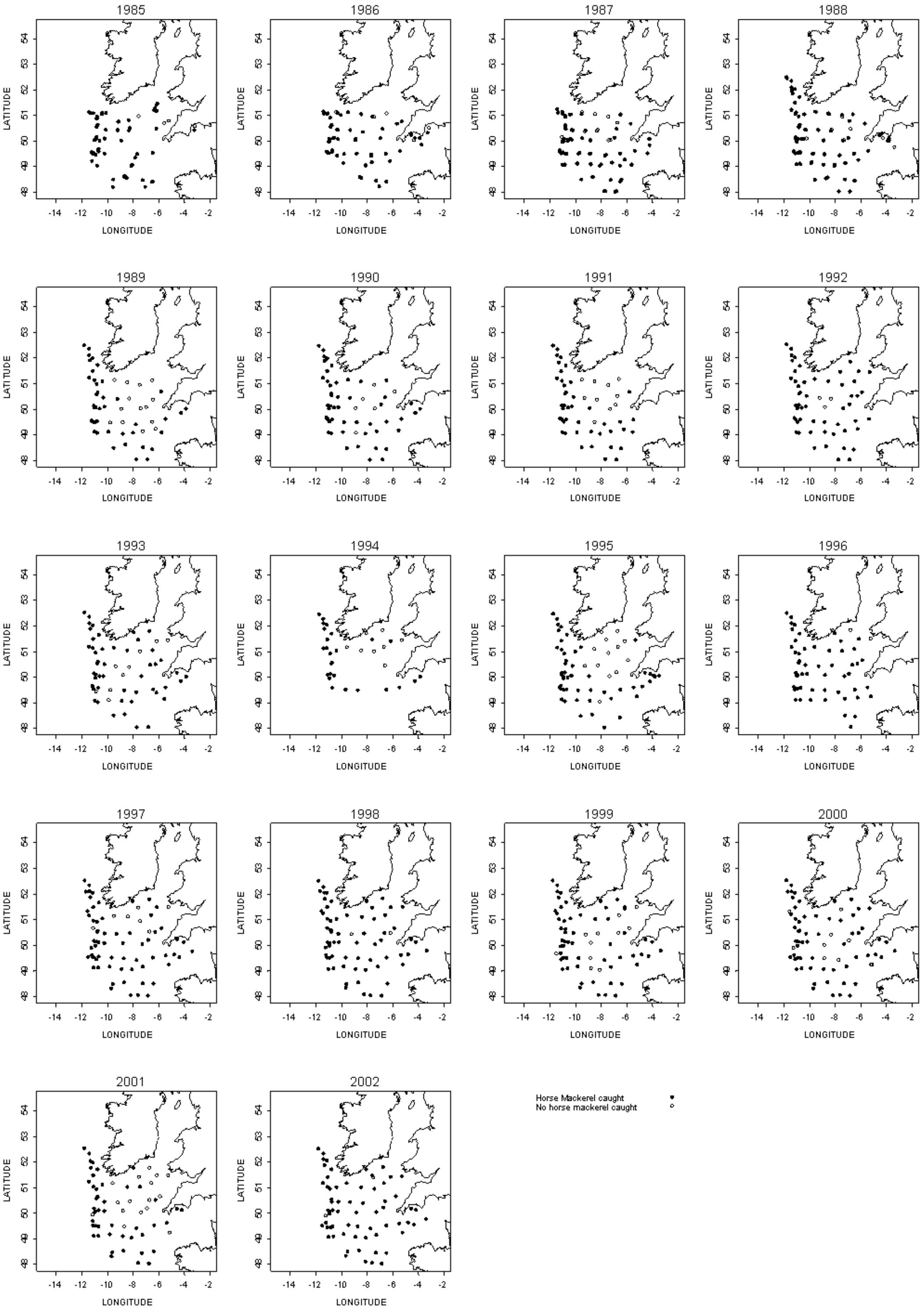


Figure 15: UK (E &W) Surveys Blue whiting

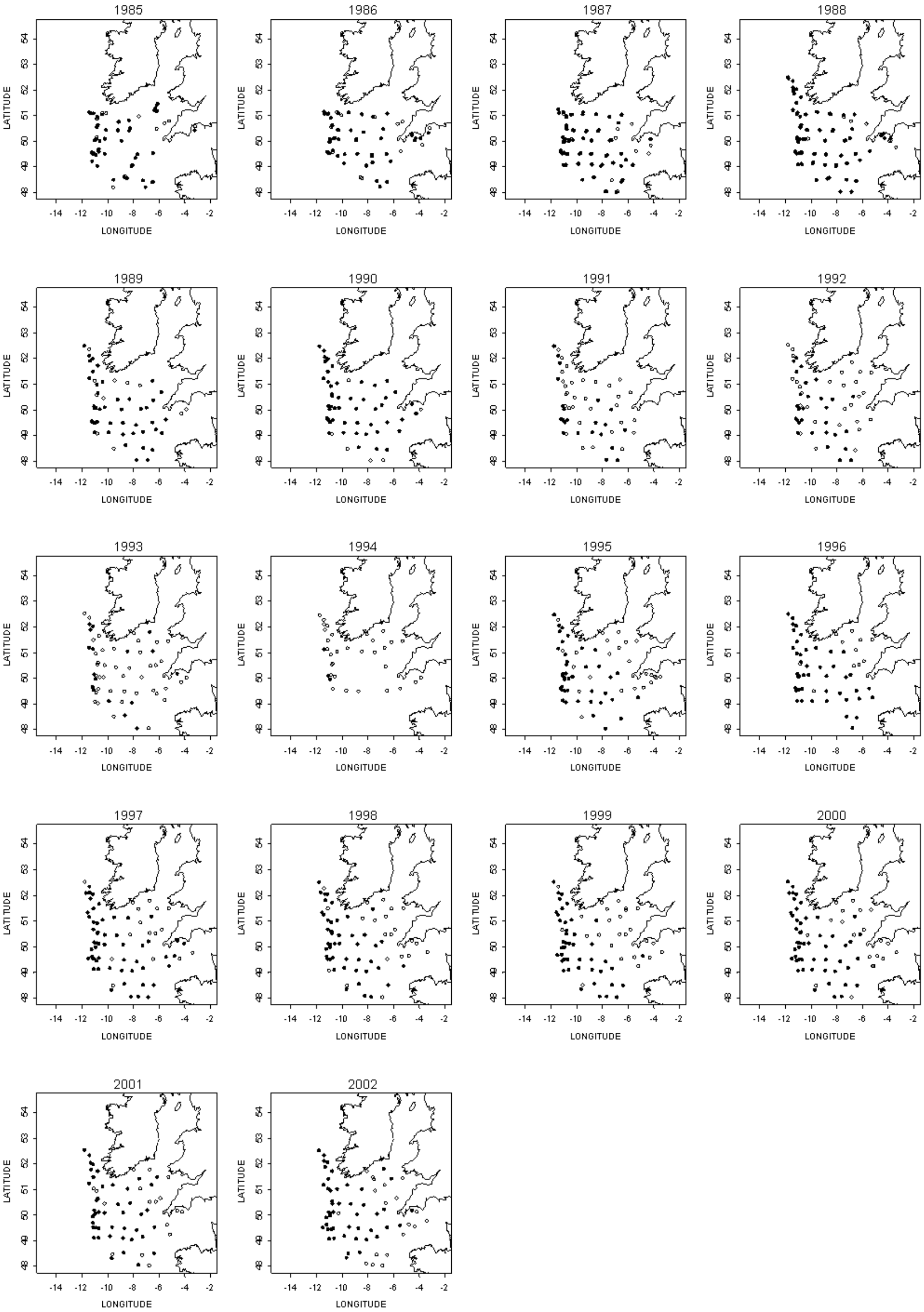
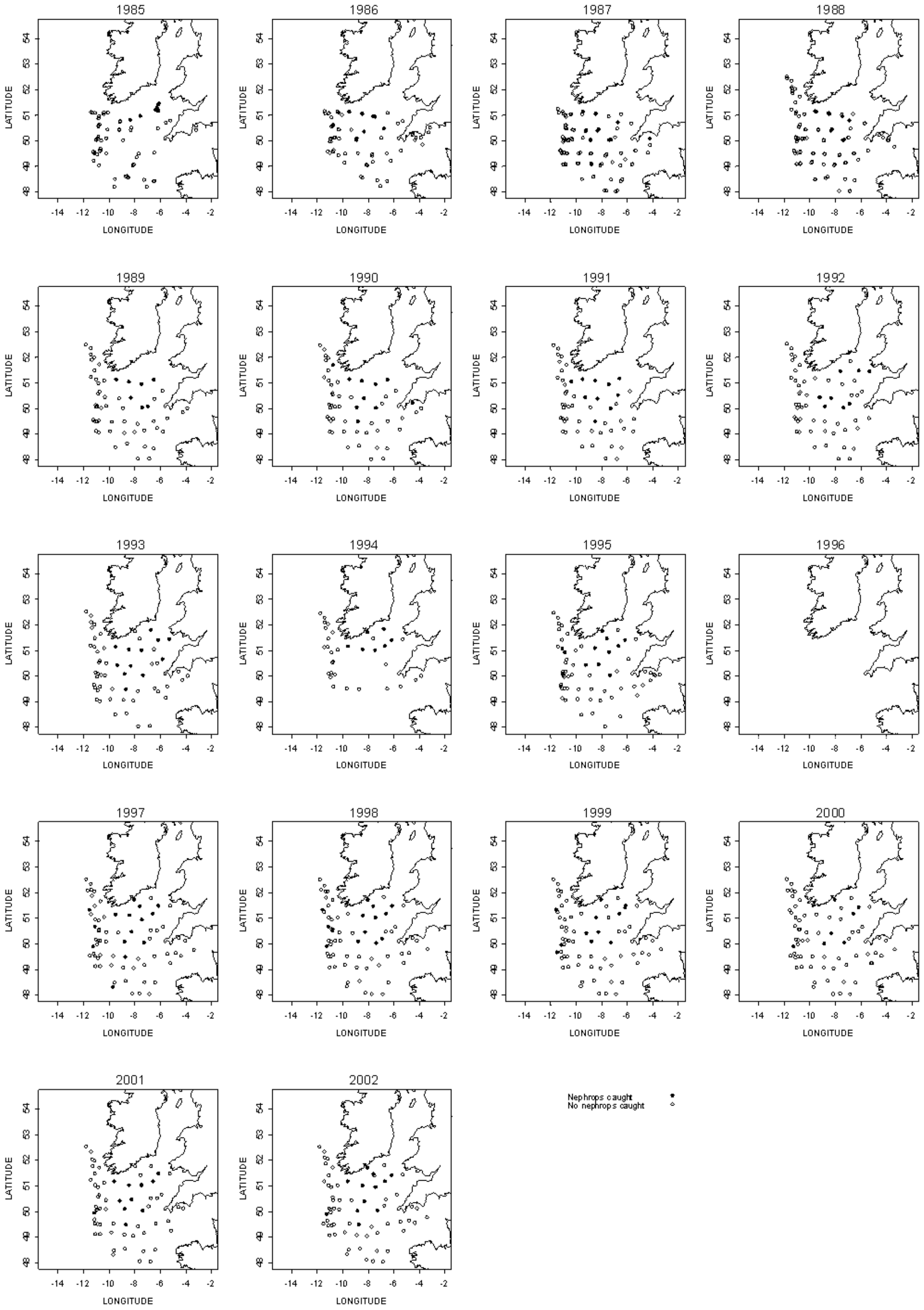
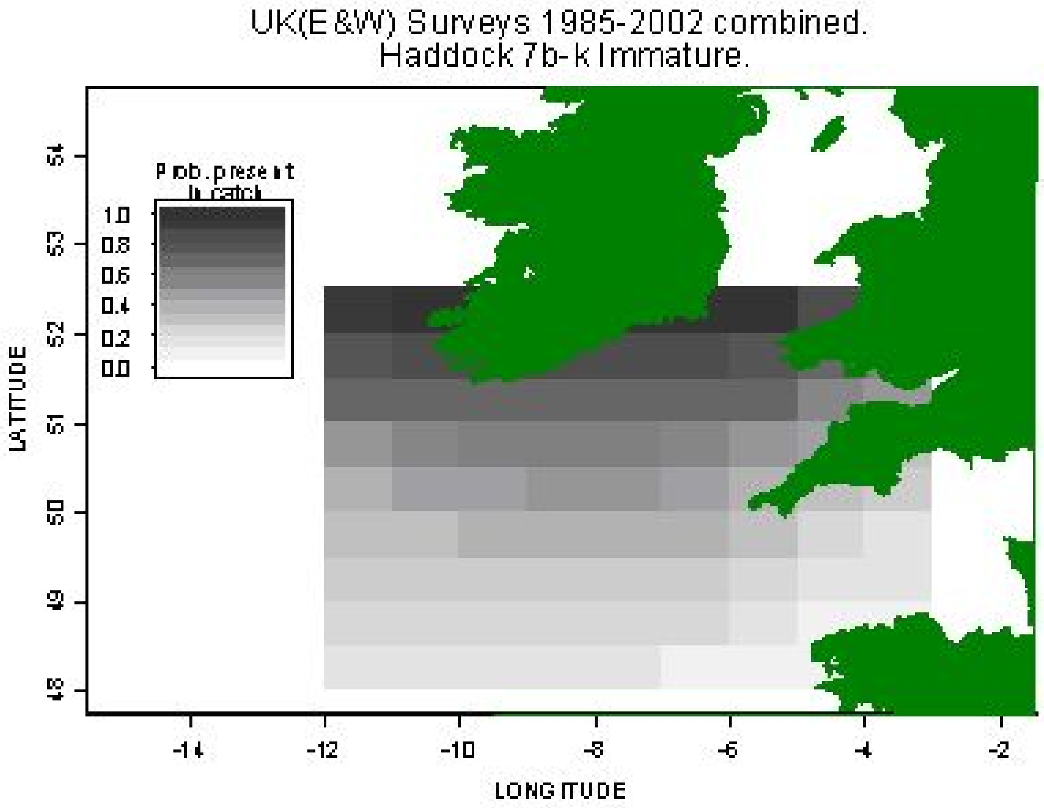
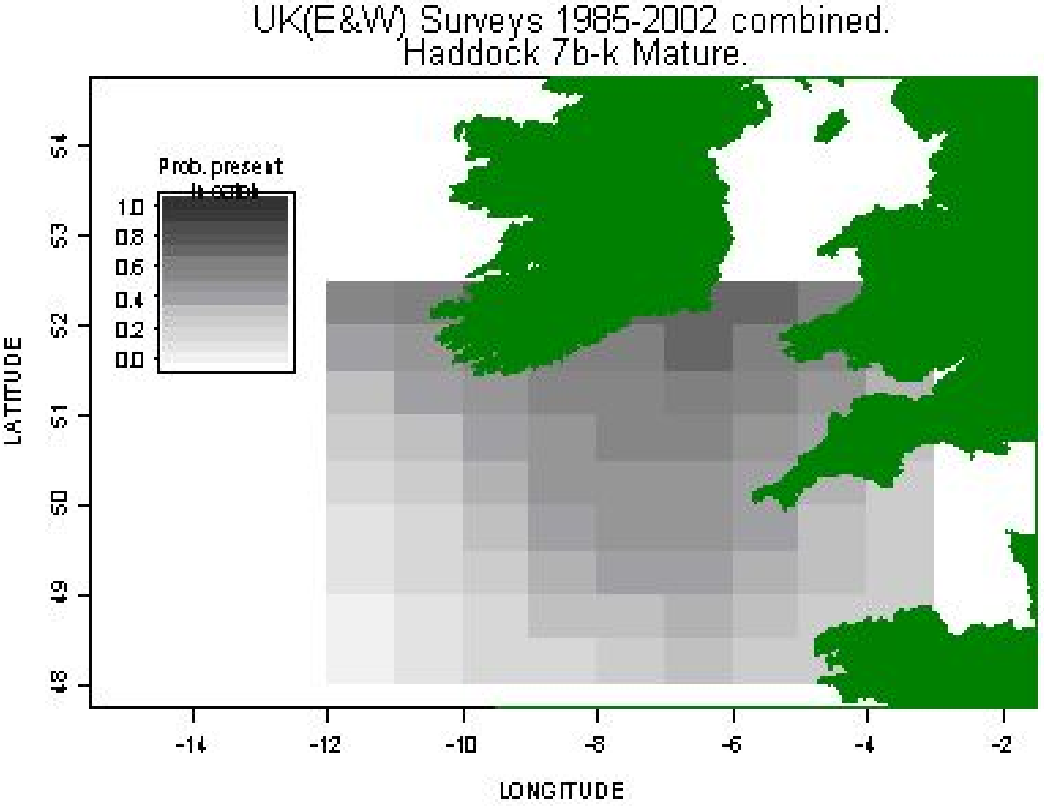
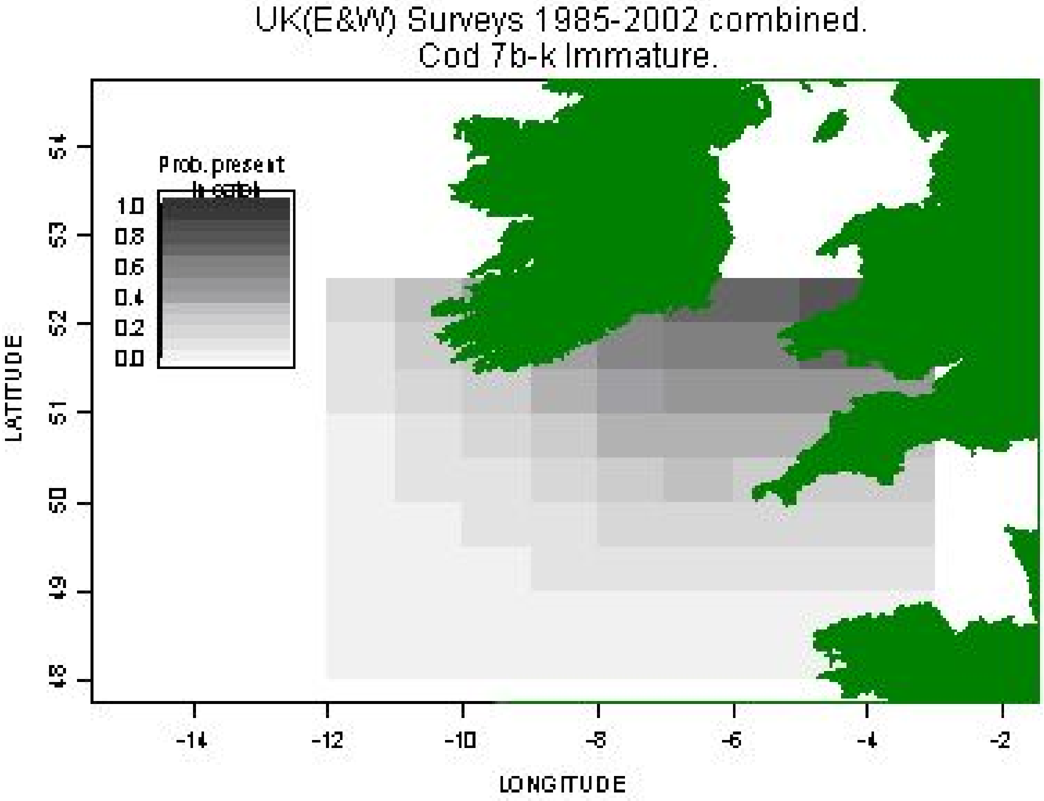


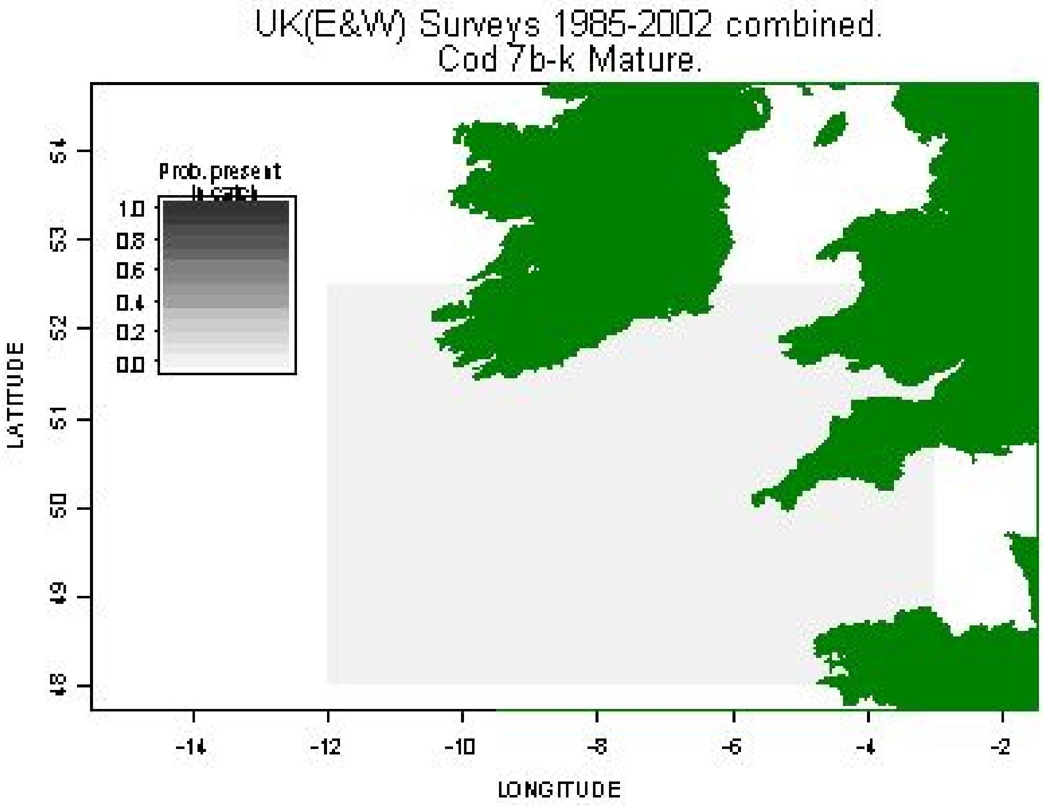
Figure 16: UK (E & W) Surveys Nephrops

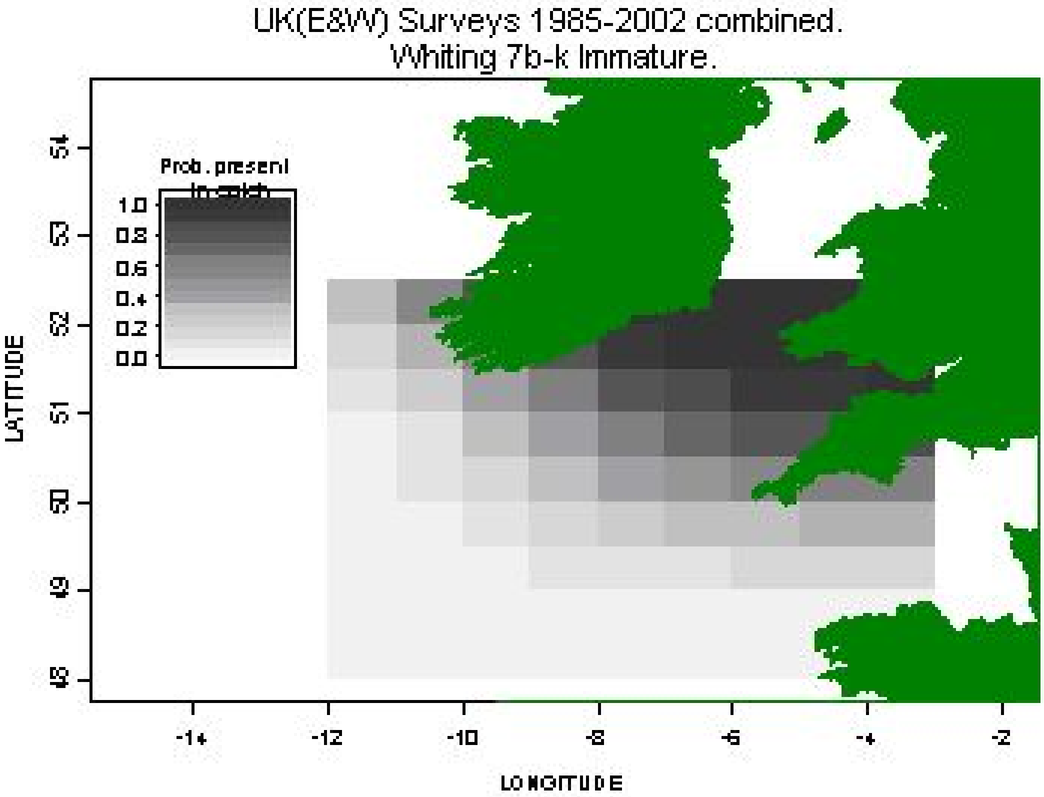


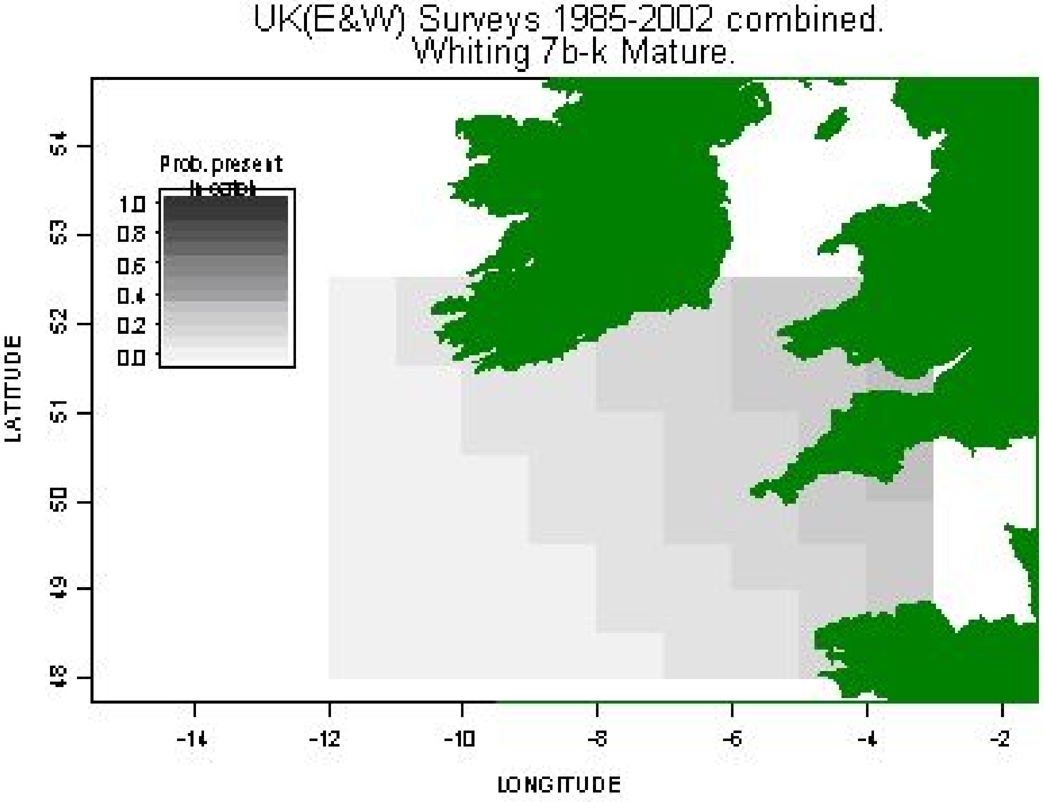

UK(E&W) Surveys 1985-2002 combined.
Haddock 7b-k Immature.
Prob. present in catch
1.0
0.8
0.6
0.4
0.2
0.0
LATITUDE
54
53
52
51
50
49
48
LONGITUDE
-14
-12
-10
-8
-6
-4
-2

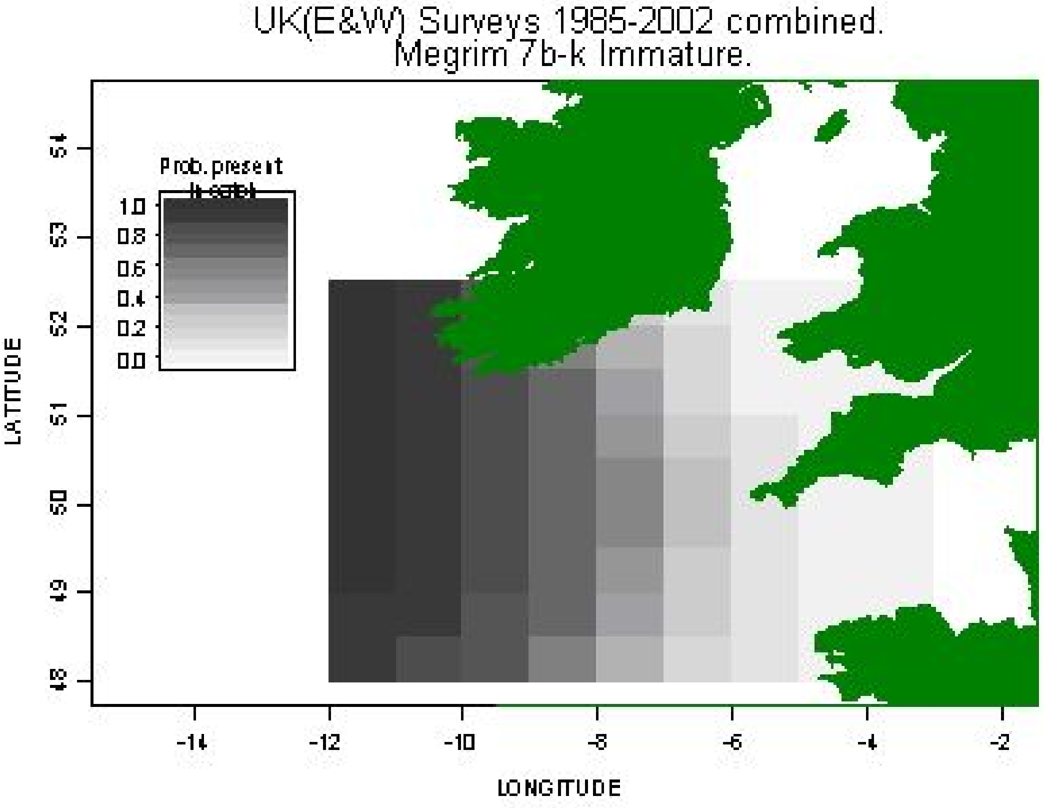

UK(E&W) Surveys 1985-2002 combined.
Haddock 7b-k Mature.

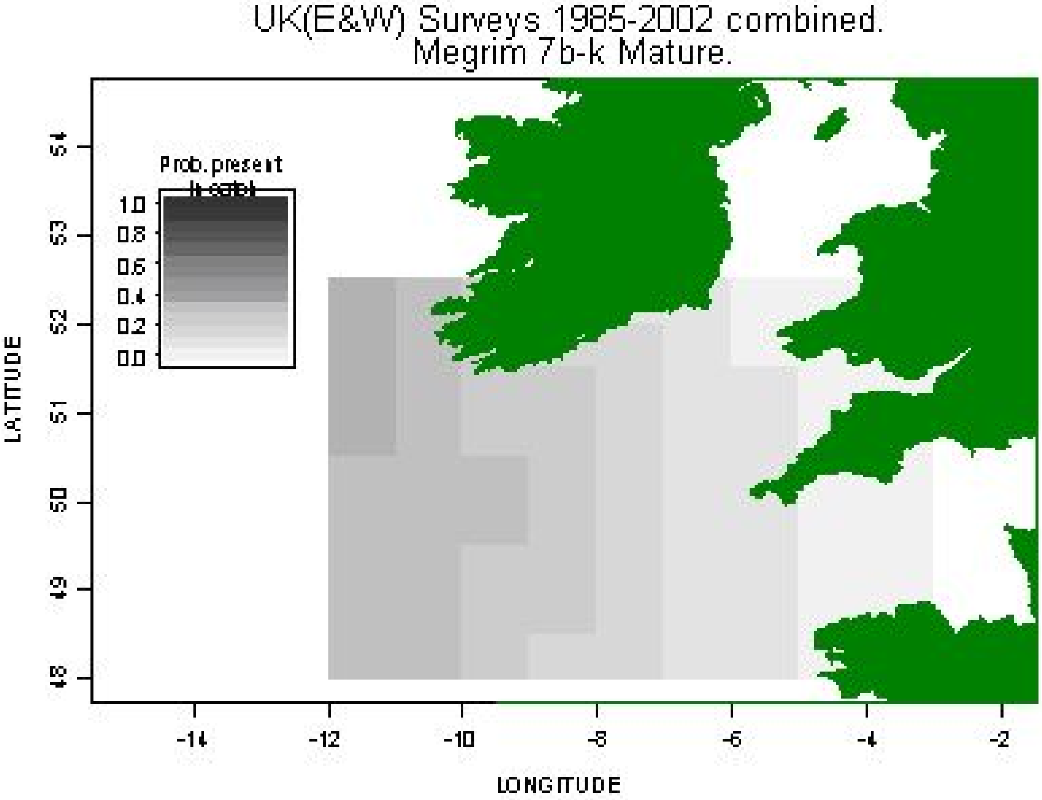

UK(E&W) Surveys 1985-2002 combined.
Cod 7b-k Immature.

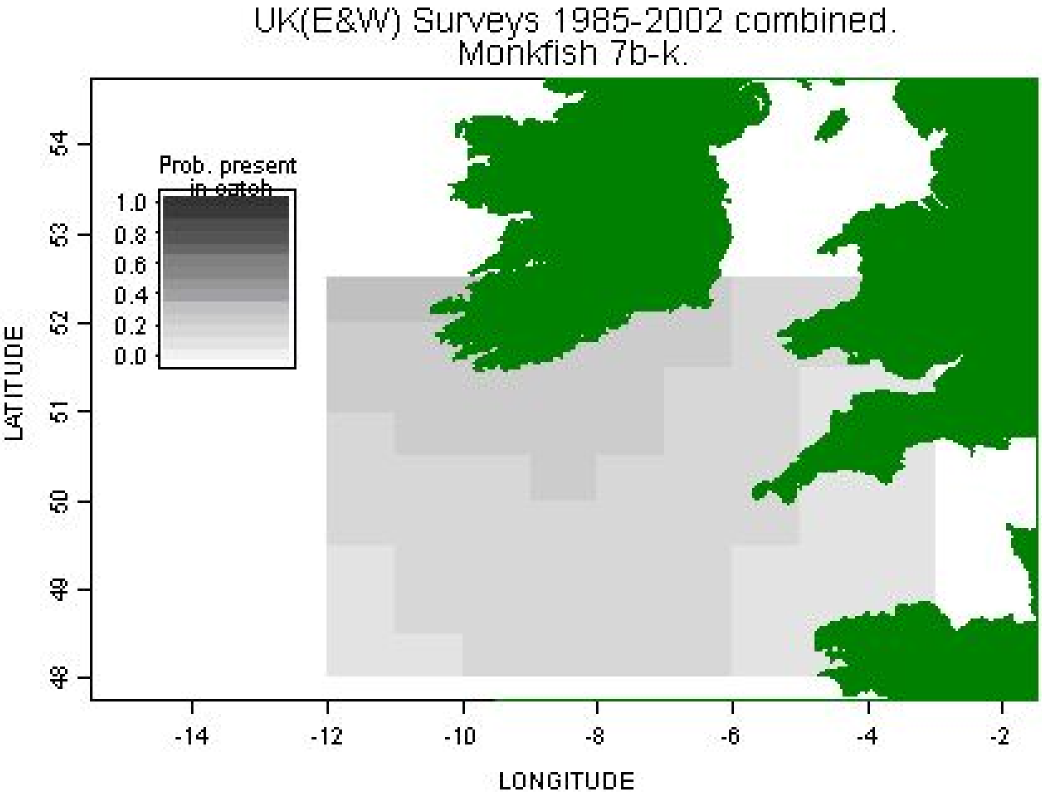

UK(E&W) Surveys 1985-2002 combined.
Cod 7b-k Mature.
Prob.present
in catch
1.0
0.8
0.6
0.4
0.2
0.0
LATITUDE
54
53
52
51
50
49
48
LONGITUDE
-14
-12
-10
-8
-6
-4
-2

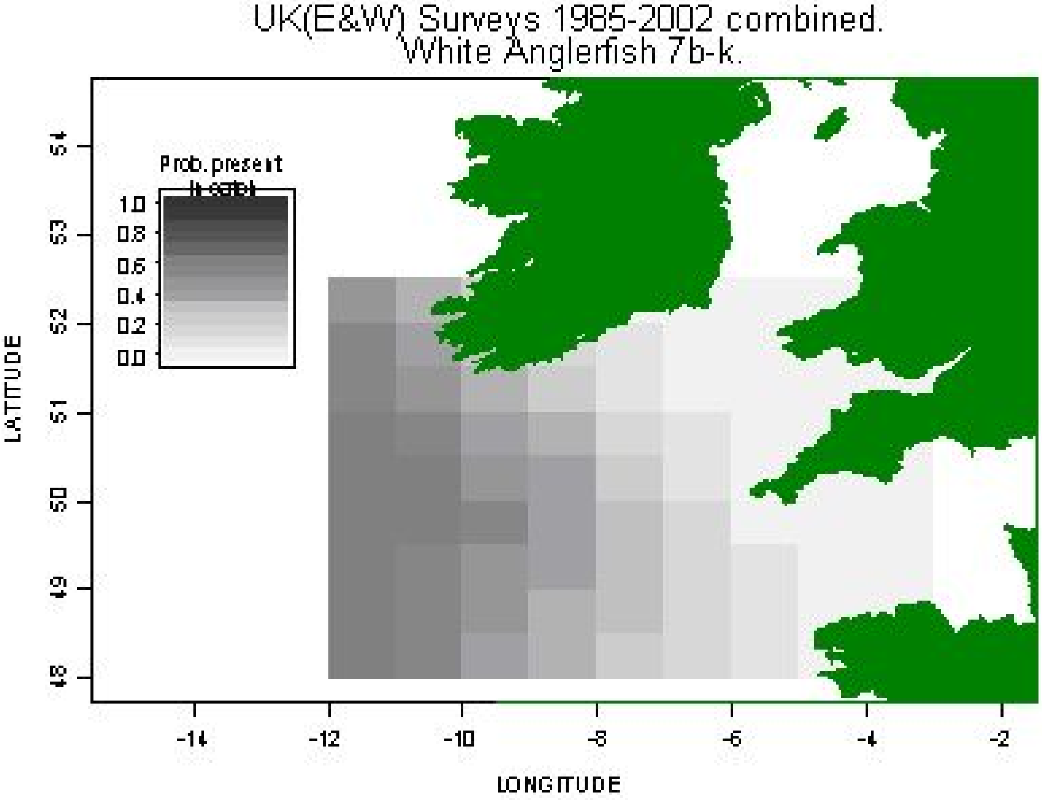

UK(E&W) Surveys 1985-2002 combined.
Whiting 7b-k Immature.
Prob.present
in catch
1.0
0.8
0.6
0.4
0.2
0.0
LATITUDE
54
53
52
51
50
49
48
LONGITUDE
-14
-12
-10
-8
-6
-4
-2

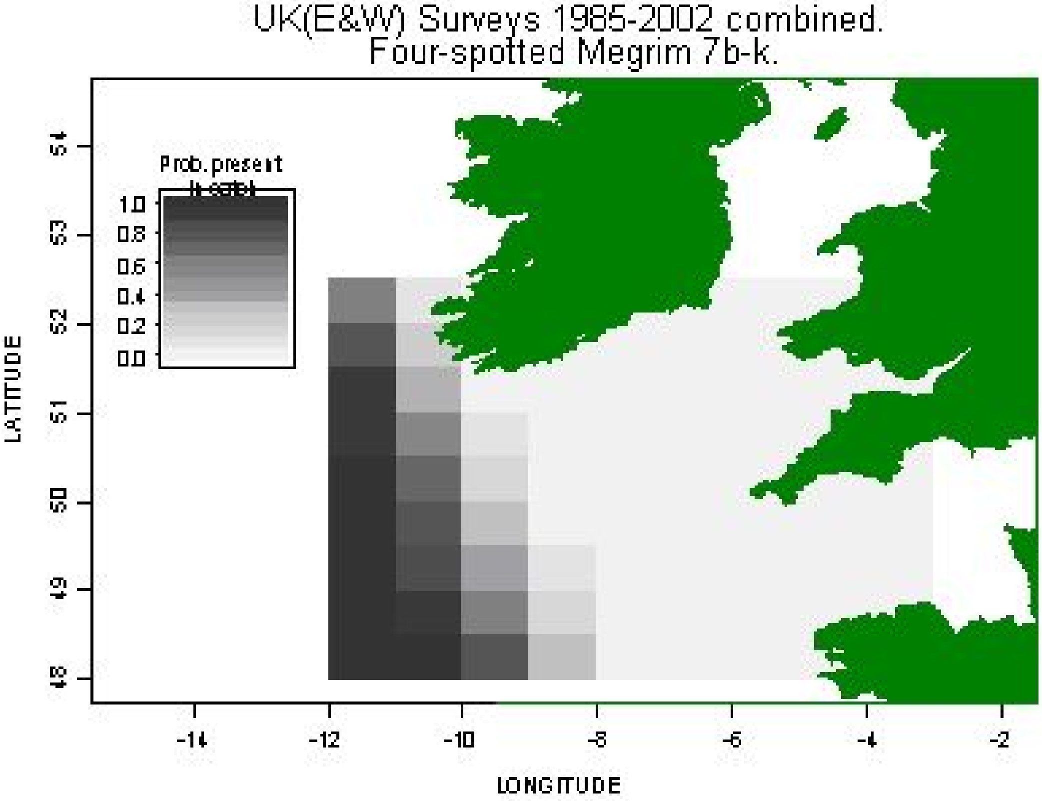

UK(E&W) Surveys 1985-2002 combined.
Whiting 7b-k Mature.
Prob. present
1.0
0.8
0.6
0.4
0.2
0.0
LATITUDE
54
53
52
51
50
49
48
LONGITUDE
-14
-12
-10
-8
-6
-4
-2

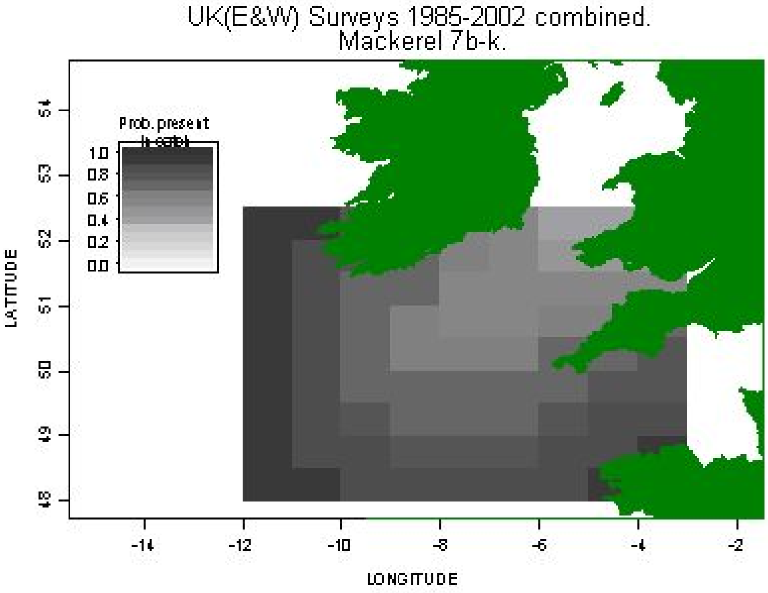

UK(E&W) Surveys 1985-2002 combined.
Megrim 7b-k Immature.
Prob. present
1.0
0.8
0.6
0.4
0.2
0.0
LATITUDE
54
53
52
51
50
49
48
LONGITUDE
-14
-12
-10
-8
-6
-4
-2

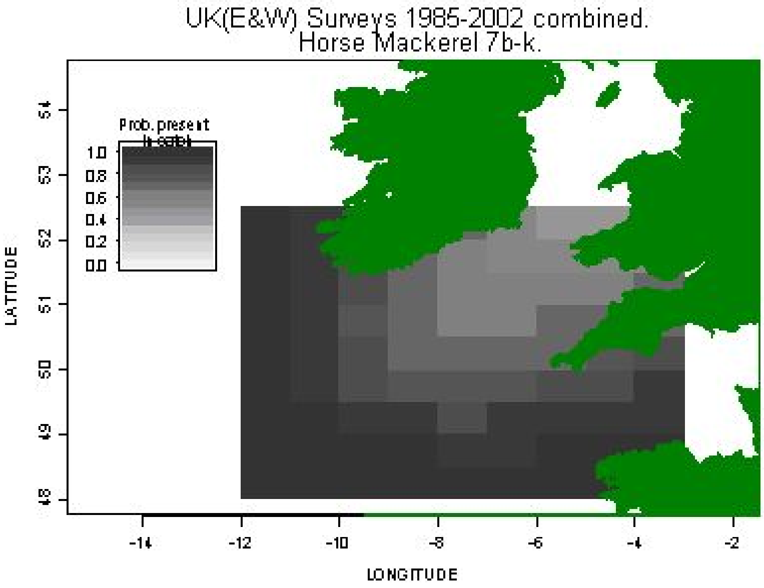

UK(E&W) Surveys 1985-2002 combined.
Megrim 7b-k Mature.
Prob. present in catch
1.0
0.8
0.6
0.4
0.2
0.0
LATITUDE
54
53
52
51
50
49
48
LONGITUDE
-14
-12
-10
-8
-6
-4
-2

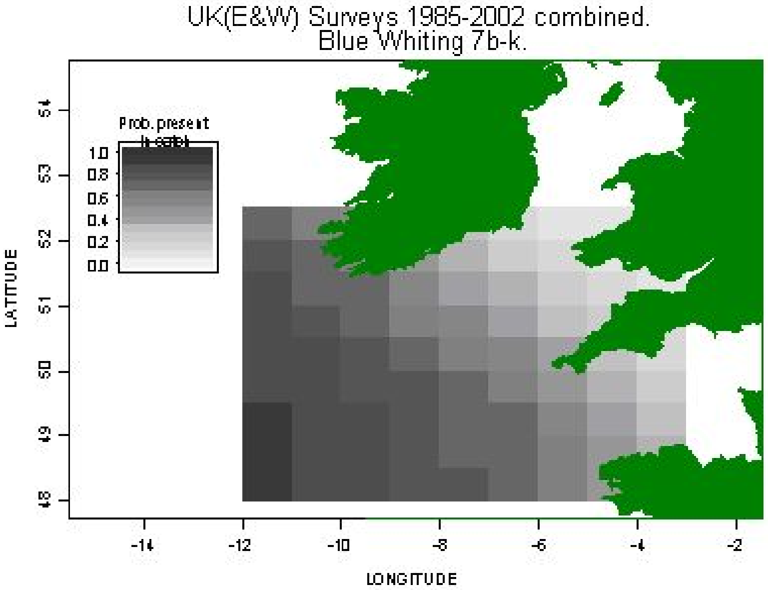

UK(E&W) Surveys 1985-2002 combined.
Monkfish 7b-k.
Prob. present in catch
1.0
0.8
0.6
0.4
0.2
0.0
LATITUDE
54
53
52
51
50
49
48
LONGITUDE
-14
-12
-10
-8
-6
-4
-2

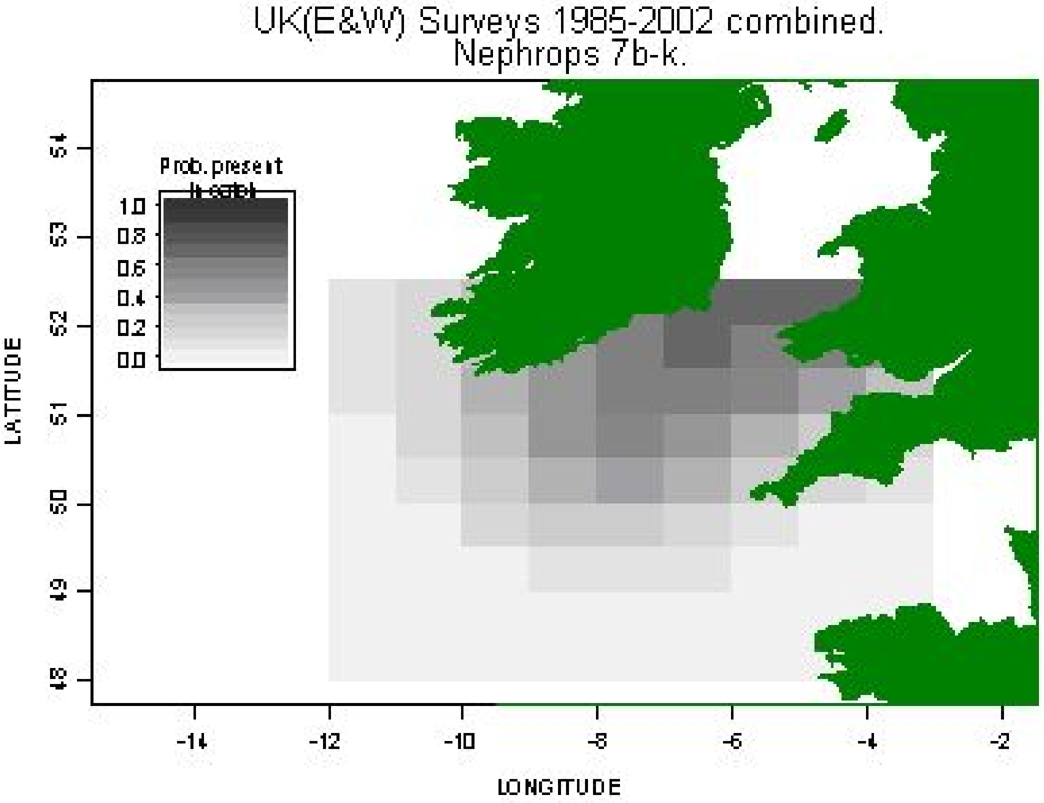

UK(E&W) Surveys 1985-2002 combined.
White Anglerfish 7b-k.
Prob.present (scaled)
1.0
0.8
0.6
0.4
0.2
0.0
LATITUDE
54
53
52
51
50
49
48
LONGITUDE
-14
-12
-10
-8
-6
-4
-2


UK(E&W) Surveys 1985-2002 combined.
Four-spotted Megrim 7b-k.
Prob.present (scaled)
1.0
0.8
0.6
0.4
0.2
0.0
LATITUDE
54
53
52
51
50
49
48
LONGITUDE
-14
-12
-10
-8
-6
-4
-2


UK(E&W) Surveys 1985-2002 combined.
Mackerel 7b-k.
Prob. present
1.0
0.8
0.6
0.4
0.2
0.0
LATITUDE
54
53
52
51
50
49
48
-14
-12
-10
-8
-6
-4
-2
LONGITUDE


UK(E&W) Surveys 1985-2002 combined.
Horse Mackerel 7b-k.
Prob. present
1.0
0.8
0.6
0.4
0.2
0.0
LATITUDE
54
53
52
51
50
49
48
-14
-12
-10
-8
-6
-4
-2
LONGITUDE


UK(E&W) Surveys 1985-2002 combined.
Blue Whiting 7b-k.
Prob.present
1.0
0.8
0.6
0.4
0.2
0.0
LATITUDE
54
53
52
51
50
49
48
-14
-12
-10
-8
-6
-4
-2
LONGITUDE


UK(E&W) Surveys 1985-2002 combined.
Nephrops 7b-k.
Prob.present
1.0
0.8
0.6
0.4
0.2
0.0
LATITUDE
54
53
52
51
50
49
48
-14
-12
-10
-8
-6
-4
-2
LONGITUDE

I.3 Does diet reflect prey availability in Celtic Sea fish?

(In press: Journal of Fish Biology)

J.K. Pinnegar[1], V. Trenkel[2], A. N. Tidd,[1] W.A. Dawson[1] and M.H. Dubuit[3]

[1]Centre for Environment, Fisheries and Aquaculture Science, Lowestoft Fisheries Laboratory, Lowestoft, NR33 0HT, UK.

[2]Laboratoire MAERHA, IFREMER, Rue de l'Ile d' Yeu, BP 21105, 44311 Nantes cedex 3, France.

[3]CNRS, Collège de France, Laboratoire de Biologie Marine, 29110 Concarneau, France.

Correspondence:

John K. Pinnegar

Tel. +44 1502 524229
Fax. +44 1502 513865
e-mail j.k.pinnegar@cefas.co.uk

Running Headline: Predator diet and prey availability.

Abstract

Feeding preferences of Celtic Sea fishes were investigated using a hitherto unpublished database of stomach content records, collected by French and English researchers between 1977 and 1994. The diet of cod, hake, megrim, whiting and saithe changed markedly as the animals grew larger, and although large predators generally chose larger bodied prey, the range of prey sizes targeted also increased. Large predators continued to select small, low value, benthic prey (e.g. *Callionymus* spp. and *Trisopterus* spp.) which were easier to catch, rather than larger, more energy lucrative prey (e.g. mackerel), even though these pelagic prey-fishes were nearly always available and were often very abundant. There is some evidence for density-dependence in predator preferences, with predators choosing more of a particular prey whenever they were more available in the environment. ICES stock estimates and UK groundfish survey catches were used as indices of prey abundance. Blue-whiting and other small pelagic fishes (*Argentina* spp. and clupeoids) were identified as being particularly important, and were consumed by some predators more often than would be expected given the abundance of these prey. There was no evidence for density-dependent feeding by predators on mackerel and only hake exhibited density-dependent preferences for horse-mackerel.

Key Words: Celtic Sea, diet, stomach contents, optimal foraging, preference, availability.

Introduction

The importance of predation as a regulatory process in marine systems has been well documented and piscivorous fishes are known to have a dramatic influence on population and community level dynamics (see Juanes *et al.* 2002). However, although a relatively large amount is known about whom eats who in marine systems, virtually nothing is known of the dynamics, and specifically how diets of piscivorous fish relate to changes in the abundance of their prey (Greenstreet *et al.*1998).

Most fish are selective foragers; they prefer to feed on some prey types but not on others (Mittelbach 2002). Much of the theoretical development of foraging theory has revolved around trying to explain why predators choose the items they do and whether a predator should theoretically choose to pursue a particular prey item that it has encountered (Mittelbach 2002). Ecologists have used optimisation criteria to address this question, arguing that natural selection should result in predator behaviours that maximise the rate of energy gain, which is a component of fitness. Charnov (1976) developed one of the first optimal diet models, and three basic predictions stem from this work:

1. predators should prefer prey that yield more energy per unit handling time.
2. as abundance of higher value prey increases in the environment, lower value prey should be dropped from the diet and predators should become more selective.
3. foragers should obey a quantitative threshold rule for when specific prey types should be included or excluded from an optimal diet (Stephens & Krebs 1986, Sih & Christensen 2001).

Prey encounter rate is dependent on the abundance of that prey in the environment but, in times when a focal prey is not available, predators can satisfy their nutritional and energy requirements, to some extent through adjustments in selection for prey quality. Prey fishes can exhibit a ten-fold difference in lipid content and a five-fold difference in energy density (Anthony *et al.* 2000). Thus as prey vary in abundance, resource value, catchability and handling time, optimally foraging predators must trade-off costs and benefits to optimise their own survival and reproductive fitness (Stephens & Krebs 1986).

Over the past 30 years a massive amount of stomach sampling effort has been expended in the North Sea (Pope 1991), but there has been little concerted research elsewhere in the northeast Atlantic (e.g. the Celtic sea). Marked changes are known to have occurred in the Celtic Sea ecosystem in recent years (Pinnegar et al. 2002), and given such changes it might be expected that the diet of predatory species will have varied in response. (Greenstreet *et al.* 1998).

The main aims of the present study were to (1) explore how the diet of Celtic Sea fish change with body-size, (2) examine how the size range of the targeted prey change with increasing body-size, (3) explore whether predators select prey in accordance with availability in the environment.

Methods

The Celtic Sea

The Celtic Sea is an area of continental shelf bordered by Ireland in the North, the UK in the East and the Bay of Biscay (47° N Latitude line) in the South (Fig. 1). It supports a diverse range of fishing fleets (métiers), characterised by the use of different fishing gear types and different target species (Marchal & Horwood 1996). Recent expansion of Celtic Sea fisheries has prompted concern about the present and future state of fish stocks, the scale of fishery discards/by-catch and possible implications for ecosystem functioning as a whole (Pinnegar *et al.* 2002).

Stomach-content database

Fish were sampled by UK and French researchers between 1977 and 1994. A total of 26,539 prey items were recorded from 66 predator species (18,129 stomachs).

UK samples (16,049 data-points) were taken as part of annual research cruises aboard the vessels *Cirolana*, *Clione* and *Scotia*. These surveys were aimed at providing fishery-independent estimates of fish abundance (Warnes & Jones, 1995). Location, predator length, stomach fullness (on a 10 point scale), prey identity, number and digestion state (on a 4 point scale), were recorded along with the number of empty stomachs. Where possible, prey length was recorded to the nearest cm.

French samples (10,492 data-points) were collected aboard commercial trawlers (*Agora, Elsinor, Galaxie, Madiana, Melodie, Opera, Peoria, Symphonie, Valériane*) during routine fishing operations in the Celtic Sea. These data have been largely documented by Du Buit (1982, 1992, 1995, 1996) but not in the context of prey availability. Predator length, prey identity, prey abundance, prey weight (in grams), and digestion state (on a 5 point scale) were recorded. No data relating to non-identifiable prey remains or empty stomachs were available. Prey length data only existed for megrim *Lepidorhombus whiffiagonis* (Walbaum), whiting *Merlangius merlangus* (L.) and saithe *Pollachius virens* (L.).

Of the 66 fish species for which stomach-content data existed, the number of records was greatest for cod *Gadus morhua* L., whiting, megrim, hake *Merluccius merluccius* (L.), haddock *Melanogrammus aeglefinus* (L.), and saithe. Initial analyses were carried out to determine which of these predators fed mainly on fish (in terms of total number of prey items). It was determined (Fig. 2) that haddock consume very little fish material (<2% of all items) and thus this species was excluded from further analysis. Cod, also consumed surprisingly little fish material ($\sim$22% of all items), but was retained in the analyses, largely on the basis of the commercial importance of this species.

The temporal coverage of the available data is detailed in Appendices 1 and 2. Sampling varied greatly from year to year, with the largest number of stomachs collected in 1984, 1985 and 1991 (Appendix 1).

Prey abundance

Mackerel *Scomber scombrus* L., horse-mackerel *Trachurus trachurus* and blue-whiting *Micromesistius poutassou* (Risso) were identified as being key fish prey for the predators considered here (cod, whiting, megrim, hake and saithe). Stocks of these pelagic prey species are assessed on an annual basis by ICES (International Council for the Exploration of the Sea), and thus population estimates were available for comparison with fish-prey numbers from stomach contents (available vs. consumed prey). Population numbers (in millions) were extracted from the 2002 reports of the ‘Northern Pelagic and Blue-whiting Working Group’ (WGNPBW) and the ‘Working Group on the Assessment of Mackerel, Horse-mackerel, Sardine and Anchovy’ (WGMHSA). The geographic units considered by ICES, were generally much larger than the Celtic Sea, but it was assumed that any temporal patterns apparent at the whole stock level would also be reflected at the Celtic Sea sub-stock level. The abundance of each pelagic species was compared with the stock size of the other two species using the non-parametric Kendall tau test (Conover, 1980).

Abundance estimates for all fish prey species were available from 1982-2003 through the annual groundfish survey of CEFAS (Centre for Environment, Fisheries & Aquaculture Science). Samples are collected each year using a Portuguese High Headline Trawl (PHHT), hauled by the research vessel *Cirolana*. During the early years of the survey (1982-1989) sampling was conducted in both spring (March and April) and Autumn (October and November), in later years however (1990 onwards), only a spring survey was carried out (Warnes & Jones, 1995). The distance travelled by the vessel whilst hauling, together with the gear geometry (monitored using acoustic transponders) were used to calculate ‘swept area’ and consequently fish numbers per unit area. Only haul-stations from the central Celtic Sea were included in the analyses (Fig. 1), and the estimated fish density (in numbers per km^2) was raised to the level of this whole central zone (124,505 km^2).

Prey Length

Scatter diagrams were plotted for each predator species and least-squares regression was carried out to determine the relationship between mean prey-size and predator size. To estimate changes in median, ‘minimum’, and ‘maximum’ prey size with increasing predator body-length, the available data was also analysed using quantile regression techniques (Scharf *et al.* 2000; Scharf *et al.* 1998), based on sums of the absolute values of residuals. Quantile regression estimates were obtained through minimisation of the quantity:

$$\sum_i \left| y_i - \sum_j \beta_j x_{ij} \right| h_i$$

where h_i is a multiplier equal to the chosen quantile value (i.e., 0.50 for the median). The quantiles generated by this method are generally robust to outlying values along the y-axis (prey length) and the procedure is ideally suited to the examination of data with heteroscedastic error distributions (Scharf *et al.* 1998). Correlation analyses were performed among 10% quantile, median (50% quantile) and 90% quantile slope estimates to determine whether increases in median prey size were primarily a result of changes in 'minimum' or 'maximum' prey sizes. In order to carry out quantile regression, BLOSSOM software was used, as described and developed by Cade *et al.* (1999) and Cade & Richards (1996).

Prey-Preference Indices

Many prey-selection indices have been proposed, and different indices seem appropriate for answering different ecological questions (Pearre 1982, Confer & Moore 1987). In the present paper values from two different indices were compared. The notation used in this paper is as used in Pearre (1982): a_d is the number of prey animals of species *a* in the predator's diet, b_d is the number of all other prey animals in the diet, a_e is the number of prey animals of species *a* in the environment and b_e is the number of all other prey animals in the environment. *d* is the total number of all animals in the diet, *e* is the total number of all animals in the environment and *n* is the sum *d* + *e*.

The still popular 'forage ratio', cited by Ivlev (1961) can be calculated as:

$$FR_a = \frac{r_a}{p_a} = \frac{a_d}{d} \Big/ \frac{a_e}{e}$$

where r_a is the proportion of prey species *a* in the diet, and p_a is the proportion in the environment. A value of 'one' denotes that the prey in the diet is in exactly the same proportions as in the environment, otherwise known as 'no selection' or 'random feeding'. Ingestion which is less than proportional to abundance in the environment ('negative selection') is denoted by values between 0 and 1, while 'positive selection' is represented by values ranging from 1 to infinity. Ivlev (1961) offered a refinement on the 'forage ratio' concept (and specifically dealt with the problem of asymmetry); Ivlev's index has undoubtedly been the most popular ever since. Ivlev's solution is known as 'electivity' (E_a), ranging from –1 to +1:

$$E_a = \frac{r_a - p_a}{r_a + p_a} = \frac{a_d}{d} - \frac{a_e}{e} \Big/ \frac{a_d}{d} + \frac{a_e}{e}$$

This index has, however, a major shortcoming; as shown by several authors (e.g., Jacobs (1974), the Ivlev index is not independent of prey density.

Chesson (1978, 1983) proposed an index (α_a) based on the constant preference coefficient discussed by O'Neill (1969). It can be expressed as a forage ratio normalized by the sum of all forage rations. For a two prey system:

$$\alpha_a = \frac{r_a}{p_a} \Big/ \sum_{i=1}^{2} \left(\frac{r_i}{p_i}\right)$$

$$= \frac{e}{d} \cdot \frac{a_d}{a_e} \Big/ \frac{e}{d} \left(\frac{a_d}{a_e} + \frac{b_d}{b_e}\right)$$

This index (also known as the 'standardized forage ratio' s_i) has become popular because of its use in food-web modelling packages (e.g. Christensen *et al.* 2000). The index is independent of prey availability, and in form is broadly similar to the index (β_{Na}) proposed by Manly *et al.* (1972).

Here, E_a and α_a were calculated based on all available French and English stomach data, irrespective of year. Small sample sizes (Appendices 1 and 2) precluded us from calculating meaningful indices on an annual basis. α_a was calculated for a 'portfolio' of seven fish-prey species (*Argentina* spp., *Callionymus* spp., *M. poutassou*, *S. scombrus*, *T. trachurus*, *Trisopterus* spp. and clupeidae) since these were found to be the key prey for cod, hake, whiting, megrim and saithe. Some of these prey 'species' are actually groups of closely related animals which could not be distinguished in stomach contents.

Preferences vs. changing prey availability

Fish-prey abundance, whether characterised on the basis of survey data or from ICES stock assessments, were plotted against the proportion of identifiable fish prey for each year that stomachs were collected. All data for non-identifiable fish-prey were excluded from analyses. A trend line was fitted using a variable-span smoother (Friedman 1984) and tests for significance were based on Kendall's tau (Conover 1980). The rank-based Kendall tau test is a non-parametric method which is particularly robust to outliers. In addition, in order to obtain robust correlation coefficients, 20% trimming was employed, whereby 20% of observations were removed – starting with the most extreme (using a routine within the statistical package S-Plus).

Results

Changes in predator diet with fish length

The diet of all species were found to change markedly with increasing predator body-length. In cod, whiting and hake there was a marked transition from consumption of crustaceans in smaller predators to a higher proportion of fish in larger animals. Fish never represented more than 40% of the diet in cod, and polychaetes always represented a small but consistent proportion.

In haddock, fish only ever represented a very minor component in the diet, which was primarily dominated by crustaceans, echinoderms, molluscs and polychaetes, even in large fish. Since abundance data were only available for fish-prey animals (invertebrates were not monitored on a regular and systematic basis), haddock were excluded from further analyses.

Unfortunately, no diet data were available for small saithe below the size of 40cm, probably because of distributional differences between adult and juvenile animals in the Celtic Sea. Diets of large saithe were overwhelmingly dominated by fish prey, but with some cephalopod and pteropod molluscs. Megrim appeared to consume similar proportions of crustaceans and fish throughout their lives, whilst cephalopods also provided a small but consistent contribution.

Based on the UK data it was possible to estimate that 64.7% of all megrim, 11.7% of hake, 20.4% of whiting, 19.4% of saithe, 6.1% of cod and 19.6% haddock stomachs were empty upon examination.

Predator length vs prey length

The range of absolute prey sizes eaten expanded with increasing body size for all five predators (Fig. 3), thus resulting (particularly in megrim) in highly heteroscedastic error distributions. Upper and lower limits of these relationships (90% and 10% quantiles) changed at different rates (Table I), and in every case the slope of the relationship between predator length and maximum (90% quantile) prey size was greater than the slope of the relationship with minimum prey size (10% quantile) (Table I). This indicated that large predators continued to target disproportionally small prey, throughout their lives.

Overall, for each predator – prey size increased as predator body size increased, and at different size thresholds predators targeted different prey types (Fig. 3). For hake (Fig. 3a) horse-mackerel and clupeoids (sardine *Sardina pilchardus* (Walbaum)) were targeted by larger predators (>50cm), whilst blue-whiting, *Argentina* spp., *Trisopterus* spp., and small clupeoids (mostly sprat *Sprattus sprattus* (L.)) were exploited by smaller predators (<50cm).

Whiting continued to consume a wide diversity of fish prey throughout their lives (Fig. 3b), although apparently targeting more clupeoids when they were smaller (<40cm) and only eating mackerel when they themselves were greater than 40cm long and the mackerel were 15-20cm long.

Megrim consumed a wide diversity and size-range of prey (Fig. 3c), including many dragonettes (*Callionymus* spp.) particularly when predators were smaller than 40cm, and *Trisopterus* spp. in larger predators. Blue-whiting and *Argentina* spp. were consumed over the whole predator size range, but were supplemented with 'other' fish (e.g. gobies, small flatfish etc.) in smaller predators.

The data for cod (Fig. 3d) were relatively sparse (only 110 data points), but it would seem that mackerel (in excess of 20cm) were selected by large predators (>70cm), whilst dragonettes and *Trisopteus* spp. were favoured by predators <80cm. Clupeoids (18-25cm) were only targeted by cod measuring 45-60cm.

Data were only available for large saithe (40-110cm), which selected mackerel in excess of 15cm long and clupeoids/*Trisopterus* of 6-20cm. Predators of 80-95cm also targeted blue-whiting (Fig. 3e).

The average prey length to predator length ratio was lowest for hake and whiting respectively (Table I). The lower the ratio, the bigger the prey (on average) chosen by the predator. Thus saithe, megrim and cod tended to choose smaller prey relative to their own size, than did hake and whiting. There was no significant correlation between the slopes of the 90% quantile relationship and the slopes of the median (50% quantile), across the five predators examined (r = 0.560). Similarly there was no significant correlation between the slopes of the 10% quantile line and that of the median (r = 0.394).

Preference vs. prey availability

For all five predators examined, Ivlev's Electivity Index E_a (Table II) proved to always be negative for horse-mackerel and (with the exception of saithe) for mackerel. This was undoubtedly a reflection of the extremely high abundance of these species in the environment in contrast to other potential fish prey. By contrast E_a for all five predators feeding on *Argentina* spp., *Callionymus* spp. and clupeoids were almost always highly positive, reflecting the rarity of these prey in the environment (or at least in the survey representation of the ecosystem). Taken at face value, it would appear that cod exhibits strong preferences for *Argentina* spp., *Trisopterus* spp. and *Callionymus* spp but a clear aversion for horse-mackerel. Hake exhibited particularly strong preferences for *Argentina*, clupeoids and blue-whiting. Megrim exhibited apparently similar feeding preferences to cod. Saithe exhibited a greater preference for mackerel than any other predator, but also for *Aregntina* and clupeoid species. Whiting apparently preferred *Trisopterus* spp., *Argentina* spp., *Callionymus* spp. and clupeoids.

In terms of Chesson's standardized forage ratio α_a (Table III), relative only to a portfolio of the seven key fish prey; if it is accepted that anything greater than 0.143 ($\alpha_a = 1/k$) represents positive selection, then mackerel, horse-mackerel and *Trisopterus* spp. were never positively selected. Cod and megrim showed a particular preference for dragonettes (i.e. small benthic fishes), whiting chose clupeoids, saithe chose argentines and clupeoids, whilst hake exhibited a particular preference for all types of small pelagic prey (i.e. blue-whiting, argentines and clupeoids). Pearre (1982) has demonstrated that estimates of E_a are particularly sensitive to the quantification of 'other species' in the environment whilst α_a is not. E_a is also sensitive to the absolute count of individuals in the environment whilst α_a is not. α_a is robust to changes in environmental counts but very sensitive to counts or proportions assumed from the stomach contents data.

Correlation analyses of ICES stock numbers for mackerel, horse-mackerel and blue-whiting using Kendall's tau test, revealed no significant relationship between horse-mackerel and mackerel stocks (r

= 0.341, p = 0.056). Similarly there was no significant correlation between blue-whiting and mackerel stocks (r = -0.127, p = 0.076). Blue-whiting and horse-mackerel numbers were positively related and exhibited a near-significant correlation coefficient (r =0.657, p = 0.056), illustrating that horse-mackerel and blue-whiting stocks exhibited broadly similar temporal dynamics, whilst temporal patterns were very different in the mackerel stock.

When the proportion of mackerel, horse-mackerel and blue-whiting in predator stomachs was plotted against the availability of these prey as defined by ICES stock assessments (Fig. 4), there were many positive trends, demonstrating that predators do sometimes select certain preys in proportion to their availability in the environment. This was not however, true for whiting which always exhibited negative relationships with these particular prey. Kendall's tau indicated that the only significant or near significant correlations were for megrim and hake feeding on blue-whiting. These plots contained many outliers and much variability, largely related to the very limited number of stomachs sampled in some years (Appendices 1 and 2). Many of these outliers were effectively removed by the 20% 'trimming' procedure employed prior to the application of Kendall's tau test.

When the proportion a particular prey represented in stomachs was plotted against the 'availability', as determined by CEFAS spring survey data (spanning 1982-1994), there were again many positive relationships (Fig. 5). Yet, because of the marked variability in the data, few of these proved to be significant. Despite the relatively large number of whiting stomachs sampled, there were no instances where the proportion in the stomach and the availability in the environment were significantly correlated (Fig. 5). Megrim appeared to consume significantly more dragonettes and gobies in years when these prey were more abundant. Cod consumed more blue-whiting when these fish were abundant and hake chose more horse-mackerel and blue-whiting in years when these prey were more abundant.

In the analyses conducted here, only positive corelations were tested for. There was a clear relationship between consumption of *Trispoterus* spp. by cod, and the abundance of *Trisopterus* spp. in the environment (Fig 5.). However because this was a strong negative relationship, the test for a positive correlation was rejected.

Discussion

Examination of stomach contents remains one of the few tools available for understanding the linkages which occur within ecosystems. However, stomach content analyses have a number of limitations (Deb 1997), for example they tend to provide mere snapshots of diets at particular points in time and space, and results depend extensively on the size of samples, the subjectivity and taxonomic knowledge of the investigator. Rarely are all links which occur in natural systems adequately quantified (Cohen & Newman, 1988) and for piscivorous species, an additional problem stems from the fact that many animals regurgitate food upon capture (Bowman, 1986) or only feed intermittently (Righton *et al.* 2002). This might explain the relatively high proportion of empty stomachs observed for megrim, whiting, saithe and haddock.

The analyses presented here demonstrated that for several predator species the proportion of the diet represented by fish greatly increased as the animals became larger. For cod this was true but the proportion represented by crustaceans remained relatively high, even in large animals. This is unlike the pattern exhibited by the same predator in the North Sea, where large animals almost exclusively feed on fish, mostly sandeels and herring (Høines & Bergstad 1999). For cod in the Irish Sea, the decapod *Nephrops norveguicus* is a very important prey (Armstrong 1982), and diets seem on the whole, to be similar to those in the Celtic Sea.

For a fish to be successful over a large geographic range, it helps to have very generalistic feeding behaviour. This is certainly true for cod (Mattson 1990), and probably for most of the other piscivore species considered here. Most boreal fishes are able to switch prey when their traditional focus has become less abundant or a more preferable alternative presents itself. Incorporation of this complex behaviour into useful fisheries models has caused some consternation and yet is crucial for realistic model predictions (Magnusson & Palsson 1991). As a predator becomes bigger, the level in the food-web at which it feeds (mean trophic level) steadily increases, both as result of selecting species of prey which themselves feed at a higher trophic level (e.g. the shift observed from crustacean to fish prey), or as a result of selecting larger prey (but of the same species), which again tend to feed at a higher trophic level (Cohen *et al.* 1993). The body mass of many piscivorous fishes species can increase by five orders of magnitude during their lives, and a fast-growing animal may begin life as a prey item, only to become the main predator on the same group of species within one year (Jennings *et al.* 2002). Consequently, it is perhaps not too surprising that changes were observed in the fish prey targeted by predators at different lengths.

Many studies have demonstrated that fish of varying taxonomic groups have an ‘optimal’ prey size, which is selected if the predator is given a choice (see Hart & Connellan, 1984). Fish tend to grow faster when feeding on their ‘optimal’ prey, and models based on particulate feeding fishes tend to suggest that the optimal prey size, should be the largest size that a predator can handle. In the present study it was demonstrated that at various size thresholds (of both the predator and prey), the feeding preference of the predator changes. Thus for example, only above a predator size of 40cm and a prey size of 15cm, do whiting start to consume mackerel. The existence of these apparent size thresholds might be seen as supporting one of the predictions of Optimal Foraging Theory (OFT), that ‘foraging should obey a quantitaive threshold rule for when specific prey should be included or excluded from the diet’ (Sih & Christensen 2001). However, in each case predator size – prey size distributions were highly asymmetric and predators did not always select the largest prey possible, which would presumably provide the best return for each feeding event. Large predators continued to eat small prey, which could be construed as being against one of the other key principles of OFT, that ‘predators should prefer prey that yield more energy per unit handling time’ (Scharf *et al.* 2000).

Sih & Christensen (2001) have observed that OFT often provides a relatively poor fit for predators foraging on mobile prey, including large fish feeding on smaller fish. OFT addresses ‘active predator choice’ (the probability of attack given an encounter with prey), while diet studies (including this one) typically measure diet preferences. For immobile prey and plankton feeders these typically concur, but for large, more mobile prey, they often differ (Sih & Christensen 2001). By ‘diet preference’ we mean a tendency for predators to consume prey more frequently than would be expected given the relative abundances of alternative prey in the environment. By ‘active predator choice, we mean a tendency for predators to attack some prey more than others given an encounter with each prey type. ‘Active predator choice’ is only one of several factors which can lead to non-random diets.

Many factors can influence the size and type of prey chosen by a predator (Juanes *et al.* 2002), these include the energy content of the prey, the effort required to find, pursue and handle the prey, the morphology or visual acuity of the predator the behaviour and habitat use of the prey (which in-turn affect the encounter rate). In the present study, the observation that small prey were retained in the diets of larger predators was not unusual. Scharf *et al.* (2000) demonstrated similar patterns in 18 species of marine fish (including some of the same species listed here). These authors hypothesised that the combination of high relative abundance and high capture probability for small (often benthic) prey, relative to large (often pelagic) prey, may lead to consistently high vulnerability to predation for small fishes. Because predator handling times also increase rapidly with prey size, they further suggested that the retention of small prey in the diet, may reflect profitable feeding decisions because search, capture and handling costs remain lower.

In balancing costs and benefits, optimally foraging predators can compensate for changes in prey availability by adjusting their preferences to take into account prey quality (Anthony *et al.* 2000). A diet high in lipid provides sufficient metabolizable energy for maintenance, so dietary protein can be allocated to tissue synthesis and growth. Pelagic fishes tend to have a higher lipid content than do demersal or benthic species, hence the observed inclusion of small benthic fishes in the diets of large predators may represent a trade-off between the low-value nature of the prey, and the energy saved in pursuit and capture. Interestingly, smaller pelagic species (e.g. sandeels, but presumably also blue-whiting and argentines), tend to be more lipid-rich than larger pelagic species (e.g. mackerel) (Van Pelt *et al.* 1997). As a consequence it might be expected that predators would exhibit a greater preference for these ‘small pelagics’ when they are available, rather than larger pelagic species which are faster-swimming and therefore more difficult to catch. Preference indices seem to suggest that hake and whiting did prefer small pelagic prey (*Aregentina* spp., clupeoids and blue-whiting), and there was very little evidence which would suggest that either mackerel or horse-mackerel (i.e. larger pelagics) were preferentially selected by any of the five predator considered.

Megrim and cod seemed to exhibit strong preferences for small benthic fishes, i.e. *Callionymus* spp. and *Trisopterus* species. However, caution must be exercised in interpreting E_a and α_a because the UK survey data used as an estimate of numbers of prey in the environment (i.e. availability) do not necessarily represent a true picture of the ecosystem. All trawl gears are selective and the choice of sampling method used can greatly impact on the perception of the structure and dynamics of the ecological community in a given area (e.g. Merret *et al.* 1991). The UK spring groundfish survey (using a modified Portuguese High-Headline otter trawl) probably under-estimates the abundance of small benthic fishes, even though it is equipped with bobbins, rubber disks and tickler chains on the groundrope, aimed at specifically increasing the capture efficiency for such species (Engås & Godø 1989). A small experimental beam trawl survey carried out in the same area caught many more gobies, dragonettes and flatfishes per unit area, than did the standard UK sampling gear (cod-end mesh 20mm) (V. Trenkel, unpublished manuscript). The overall implications for prey preference indices of not adequately sampling small benthic fishes, would be to apparently inflate the suggested preference for these rarer species, at the expense of large-pelagics (mackerel, horse-mackerel, blue-whiting) which dominate and perhaps were over-represented in the survey. The UK survey was originally established specifically to investigate the distribution and biology of mackerel (Warnes & Jones 1995); it was only widened to cover other species at a later stage.

In the present paper the proportion of blue-whiting, mackerel and horse-mackerel in the diet of predators were compared with their abundance, as determined by ICES stock assessments. This was completely independent of the UK spring groundfish survey and its associated problems, yet there was no evidence for density-dependent feeding by predators on either mackerel or horse-mackerel. Again from these analyses blue-whiting emerged as a favoured prey (for megrim and hake), with consumption relative to other prey animals, increasing in years when blue-whiting stocks were high. The importance of blue-whiting for predators in the Celtic Sea and Bay of Biscay has also been noted by (Du Buit 1982; Velasco & Oliva-Telas 2000).

UK spring survey catches are known to have changed markedly in their composition since 1982 (Pinnegar *et al.* 2002), inferring that great changes have occurred in the underlying fish assemblage. Consequently, it was interesting to see whether the variability observed in fish populations was echoed by changes in the importance of particular prey species in predator stomachs. Even though small benthic fishes were undoubtedly under-estimated by the UK survey, in years when dragonettes and gobies were apparently more abundant, more of these were consumed by megrim. Density-dependent consumption was observed for hake feeding on blue-whiting and horse-mackerel, and cod feeding on blue-whiting. Thus, despite the limited number of stomachs collected in some years, analyses have revealed some clear patterns regarding the feeding behaviour of commercially important fish species in the Celtic Sea.

Acknowledgements

This work was funded by the European Community under Framework V, project contract QLRT-1999-01609 (Development of Structurally Detailed Statistically Testable models of marine populations, DST^2), and the UK Department for Environment, Food & Rural Affairs contract MF0316.

References

Anthony, J. A., Roby, D. D. & Turco, K. R. (2000). Lipid content and energy density of forage fishes from the northern Gulf of Alaska. *Journal of Experimental Marine Biology and Ecology,* **248,** 53-78.

Armstrong, M. J. (1982). The predator-prey relationships of Irish Sea poor-cod (*Trisopterus minutus* L.), pouing (*Trisopterus luscus* L.), and cod (*Gadus morhua* L.). *Journal du Conseil International pour l'Exploration de la Mer,* **40,** 135-152.

Bowman, R. E. (1986). Effect of regurgitation on stomach content data of marine fisheries. *Environmental Biology of Fishes,* **16,** 171-181.

Cade, B. S. & Richards, J. D. (1996). Permutation tests for least absolute deviation regression. *Biometrics,* **52,** 886-902.

Cade, B. S., Terrell, J. W. & Schroeder, R. L. (1999). Estimating effects of limiting factors with regression quantiles. *Ecology,* **80,** 311-323.

Charnov, E. L. (1976). Optimal foraging: attack strategy of a mantid. *American Naturalist,* **110,** 141-151.

Chesson, J., (1978). Measuring preference in selective predation. *Ecology* **59,** 211-215.

Chesson, J., (1983). The estimation and analysis of preference and its relationship to foraging models. *Ecology* **64,** 1297-1304.

Christensen, V, Walters, C. J. & Pauly, D. (2000). *Ecopath with Ecosim: a Users Guide,* October 2000 Edition. Fisheries Centre, University of British Columbia, Vancouver, Canada and ICLARM, Penang, Malaysia. 130 p.

Cohen, J. E. & Newman, C. M. (1988). Dynamic basis of food web organization. *Ecology* **69,** 1655-1664.

Cohen, J. E., Pimm, S. L. Yodzis, P. & Saldaña, J. (1993). Body sizes of animal predators and animal prey in food webs. *Journal of Animal Ecology,* **62,** 67-78.

Confer, J. L. & Moore, M. V. (1987). Interpretting selectivity indexes calculated from field data or conditions of prey replacement. *Canadian Journal of Fisheries and Aquatic Science,* **44,** 1529-1533.

Conover, W. J. (1980). *Practical Nonparametric Statistics*. 2^{nd} edition, New York, Wiley.

Deb, D. (1997). Trophic uncertainty vs parsimony in food web research. *Oikos,* **78,** 191-194.

Du Buit, M. H. (1982). Essai d'évaluation de la prédation de quelques téléostéens en Mer Celtique. *Journal du Conseil International pour l'Exploration de la Mer*, **40,** 37-46.

Du Buit, M.H. (1992) Alimentation de la Cardine, *Lepidorhombus whiffiagonis* en Mer Celtique. *Cah. Biol. Mar.* **33,** 501-514.

Du Buit, M. H. (1995). Food and feeding of cod (*Gadus morhua* L.) in the Celtic Sea. *Fisheries Research,* **22,** 227-241.

Du Buit, M. H. (1996). Diet of hake (*Merluccius merluccius*) in the Celtic Sea. *Fisheries Research,* **28,** 381-394.

Engås, A. & Godø, O. R. (1989). Escape of fish under the fishing line of a Norwegian sampling trawl and its influence on survey results. *Journal du Conseil International pour l'Exploration de la Mer,* **45,** 269-276.

Greenstreet, S. P. R., McMillan, J. A. & Armstrong, E. (1998). Seasonal variation in the importance of pelagic fish in the diet of piscivorous fish in the Moray Firth, NE Scotland: a response to variation in prey abundance? *ICES Journal of Marine Science,* **55,** 121-133.

Friedman, J. H. (1984). *A variable span smoother.* Technical Report No. 5, Laboratory for Computational Statistics, Department of Statistics, Stanford University, California.

Hart, P. J. B. & Connellan, B. (1984). Cost of prey capture, growth rate and ration size in pike, *Esox lucius* L., as functions of prey weight. *Journal of Fish Biology,* **25,** 279-292.

Høines, Å. S. & Bergstad, O. A. (1999). Resource sharing among cod, haddock, saithe and pollock on a herring spawning ground. *Journal of Fish Biology,* **55,** 1233-1257.

Ivlev, V. S. (1961). *Experimental ecology of the feeding of fishes.* Yale University Press, New Haven, CT. 302pp.

Jennings, S., Grrenstreet, S. P. R., Hill, L., Piet, G. J., Pinnegar, J. K. & Warr, K. J. (2002). Long-term trends in the trophic structure of the North Sea fish community: evidence from stable-isotope analysis, size-spectra and community metrics. *Marine Biology,* **141:** 1085-1097,

Jacobs, J. (1974). Quantitative measurement of food selection, A modification of the forage ratio and Ivlev Electivity Index. *Oecologia,* **14,** 413-417.

Juanes, F., Buckel, J. A. & Scharf, F. S. (2002). Feeding ecology of piscivorous fishes. In *Handbook of Fish and Fisheries* (Hart P.J.B & Reynolds, J.D. eds.) Volume 1, pp267-283. Oxford, Blackwell Science.

Magnusson, K. G. & Palsson, O. K. (1991). Predator-prey interactions of cod and capelin in Icelandic waters. *ICES Marine Science Symposium*, **193,** 153-170.

Manly, B. F. J., Miller, P. & Cook, L. M. (1972). Analysis of a selective predation experiment. *American Naturalist* **106,** 719-736.

Marchal, P. & Horwood, J. (1996). Long-term targets for the Celtic Sea mixed-species multi-metiers fisheries. *Aquatic Living Resources*, **9,** 81-94.

Mattson, S. (1990). Food and feeding habits of fish species over a soft sublittoral bottom in the Northeast Atlantic. 1. Cod (*Gadus morhua* L.) (Gadidae). *Sarsia*, **75,** 247-260.

Merrett, N. R., Gordon, J. D. M., Stehmann, M. & Haedrich, R. L. (1991). Deep demersal fish assemblage structure in the Porcupine seabight (Eastern North Atlantic): Slope sampling by three different trawls compared. *Journal of the Marine Biological Association, U.K.*, **71,** 329-358.

Mittelbach, G.C. (2002). Fish foraging and habitat choice: a theoretical perspective. In *Handbook of Fish and Fisheries* (Hart P.J.B & Reynolds, J.D. eds.) Volume 1, pp251-266. Oxford, Blackwell Science.

O'Neill, R.V. (1969). Indirect estimation of energy fluxes in animal food webs. *Journal of Theoretical Biology*, **22,** 284-290.

Pearre, S. (1982). Estimating prey preference by predators: uses of various indices, and a proposal of another based on χ^2. Canadian Journal of Fisheries & Aquatic Science, **39,** 914-923.

Pinnegar, J.K., Jennings, S., O'Brien, C.M. & Polunin, N.V.C. (2002). Long-term changes in the trophic level of the Celtic Sea fish community and fish market price distribution. *Journal of Applied Ecology*, **39,** 377-390.

Pope, J. G. (1991). The ICES multispecies assessment working group: evolution, insights, and future problems. *ICES Marine Science Symposia* **193,** 22-33.

Righton, D., Metcalfe, J. & Connolly, P. (2001). Different behaviour of North and Irish Sea cod. *Nature*, **411,** 156.

Scharf, F..S., Juanes, F. & Rountree, R..A. (2000). Predator size-prey size relationships of marine fish predators: interspecific variation and effects of ontogeny and body size on trophic niche breadth. *Marine Ecology-Progress Series*, **208,** 229-248.

Scharf, F. S., Juanes, F. & Sutherland, M (1998). Inferring ecological relationships from the edges of scatter diagrams: comparison of regression techniques. *Ecology*, **79,** 448-460.

Sih, A. & Christensen, B. (2001). Optimal diet theory: when does it work, and when and why does it fail? *Animal Behaviour*, **61,** 379-390.

Stephens, D. W. & Krebs, J. R. (1986). *Foraging Theory.* Princeton, New Jersey: Princeton University Press.

Van Pelt, T. I., Piatt, J. F., Lance, B. K. & Roby, D. D. (1997). Proximate composition and energy density of some North Pacific forage fishes. *Comparative Biochemistry & Physiology, Series A.* 118A, 1398.

Velasco, F. & Oliva-Teles, A. (2000). Hake food consumption in the southern Bay of Biscay estimated from a gastric evacuation model. ICES-CM-2000/Q:11.

Warnes, S. & Jones, B. W. (1995). *Species Distributions from English Celtic Sea Grounfish Surveys, 1984-1991.* Fisheries Research Technical Report No. 58. Ministry of Agriculture, Fisheries and Food, Directorate of Fisheries Research, Lowestoft, UK.

Table I. Regression equations relating mean, median, maximum (90% quantile) and minimum (10% quantile) prey length (in cm) to predator length. Estimates for maximum, median and minimum prey sizes were generated using quantile regression techniques, estimates for mean prey sizes are least-squares estimates.

Predator Species	cmMean (Least-squares)	Median (Quantile regression)	90% Quantile	10% Quantile	Mean predator-prey size ratio (SD)	*n*
Hake	cmy=0.205x+8.143	y=0.197x+8.508	y=0.309x+10.036	y=0.189x+2.513	2.60 1.14	267
Megrim	cmy=0.256x+0.468	y=0.222x+1.666	y=0.333x+1.667	y=0.188x-1.313	4.29 1.95	690
Cod	cmy=0.216x+3.986	y=0.240x+3.440	y=0.250x+8.500	y=0.129x+1.226	4.07 1.71	110
Whiting	cmy=0.216x+4.277	y=0.208x+4.208	y=0.286x+5.714	y=0.167+2.667	3.18 0.99	277
Saithe	cmy=0.161x+5.874	y=0.143x+7.714	y=0.184x+9.211	y=0.167x-0.333	4.57 1.61	294

Table II. Electivity E_a. (Ivlev 1961). Based on all identifiable fish-prey items in stomachs (irrespective of year), and all fish caught in spring-surveys (years in which no diet data was collected were excluded). A value of ‘zero’ denotes that the prey are consumed in exactly the same proportions as in the environment i.e. no selection or ‘random-feeding’. Values ranging from -1 to 0 indicate ‘negative selection’ and values from 0 to +1 indicate ‘positive selection’.

	Prey name						
Predator name	*Argentina* spp.	*Callionymus* spp.	Clupeoids	Blue-whiting	Mackerel	Horse-mackerel	*Trisopterus* spp.
Cod	0.7265	0.9960	0.6451	0.0669	-0.5218	-0.8353	0.7399
Hake	0.8803	0.3530	0.8145	0.7345	-0.1471	-0.6624	0.6195
Megrim	0.9235	0.9930	0.8552	0.4149	-0.7920	-0.7456	0.7650
Saithe	0.9619	-1.0000	0.9626	0.2588	0.2633	-0.9559	0.3564
Whiting	0.8429	0.8030	0.9638	0.0131	-0.5794	-0.7606	0.8104

Table III. Standardised forage ratio, α_a. Preference in relation to a 'portfolio' of seven key-prey types (irrespective of year). The standardised forage ratio as presented, ranges between 0 and 1, with α_a = 0 representing avoidance and α_a = 1 exclusive feeding. α_a = 0.143 denotes 'random-feeding'.

	Prey name						
Predator Name	*Argentina* spp.	*Callionymus* spp.	Clupeoids	Blue-whiting	Mackerel	Horse-mackerel	*Trisopterus* spp.
Cod	0.0122	0.9629	0.0090	0.0022	0.0006	0.0002	0.0129
Hake	0.3995	0.0532	0.2488	0.1662	0.0189	0.0052	0.1083
Megrim	0.0759	0.8546	0.0387	0.0073	0.0004	0.0004	0.0227
Saithe	0.4706	0.0000	0.4788	0.0155	0.0157	0.0002	0.0193
Whiting	0.1363	0.1064	0.6297	0.0119	0.0031	0.0016	0.1110

FIGURE CAPTIONS

Fig. 1. Area covered by this study, with UK sampling sites indicated by open circles.

Fig. 2. Changes in the composition of Celtic Sea fish stomach contents with increasing predator body length (in cm), based on the number of total prey items.

Fig. 3. Length of predator (in cm) plotted against the length of prey (fish only) in the stomachs of whiting, hake, cod, megrim and saithe.

Fig. 4. The proportion of total fish prey represented by mackerel, horse-mackerel and blue-whiting in the stomachs of predators, plotted against the abundance (total stock numbers) as determined by ICES stock assessments.

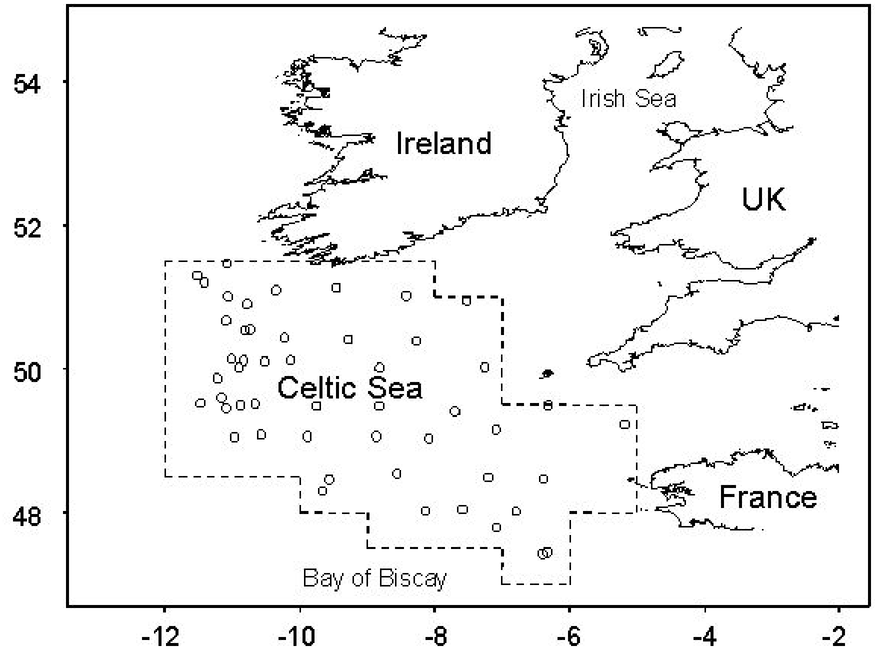
Fig. 5. The proportion represented by different prey types in the stomachs of predators, plotted against the abundance of these prey as determined by UK spring-groundfish surveys (expressed as a proportion of the total number of individuals). Prey codes are: CALL-SPP = *Callionymus* spp., CLUP-SPP = clupeidae, MICR-POU = *Micromesistius poutassou*, SCOM-SCO = *Scomber scombrus*, TRACH-TRU = *Trachurus trachurus*, TRIS-SPP = *Trisopterus* spp., ARGE-SPP = *Argentina* spp., GOBI-SPP = gobidae. Probability (p) level of Kendall's tau correlation tests given (above each panel), significant corelations ($p < 0.05$) indicated by *.

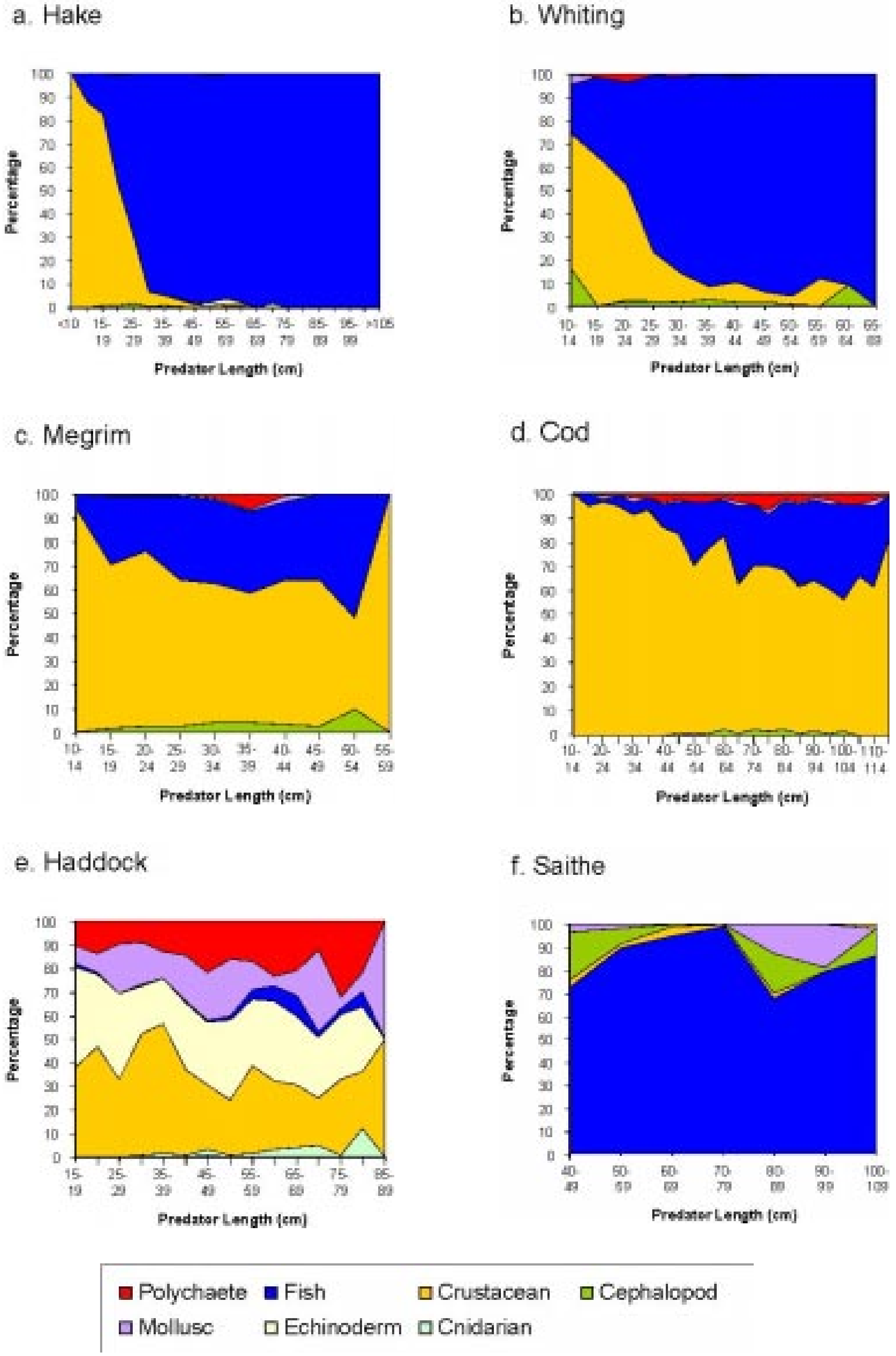
Appendix 1. Number of predator stomachs containing identifiable fish prey

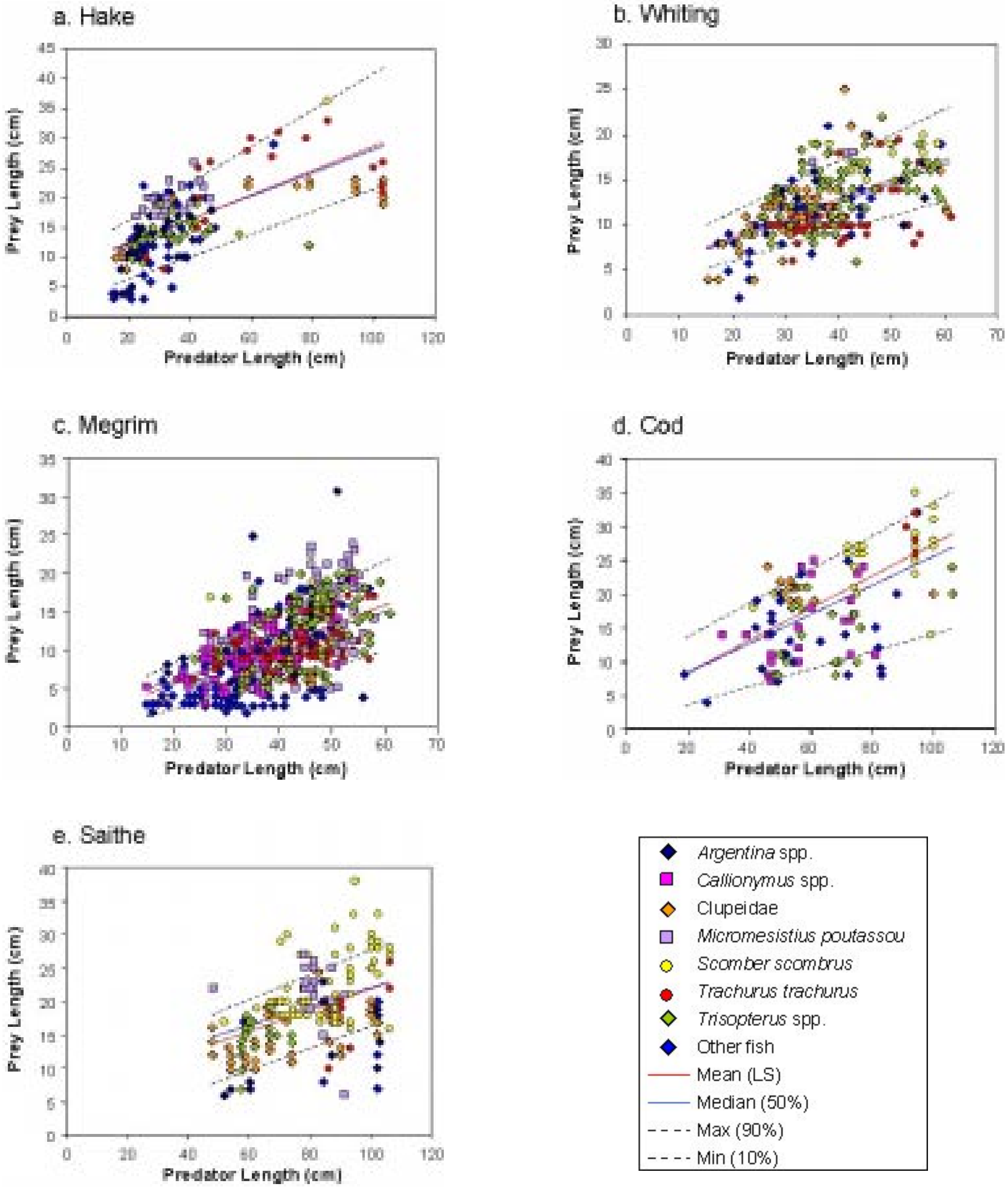
	Predator Species				
Year	Cod	Hake	Megrim	Saithe	Whiting
1977	5	19	17	0	45
1978	0	25	16	3	64
1979	12	4	8	16	2
1980	0	0	0	0	0
1981	33	57	22	35	38
1982	43	25	12	0	0
1983	3	202	223	0	66
1984	669	187	146	0	627
1985	46	365	317	42	69
1986	0	126	54	0	53
1987	7	5	1	13	0
1988	0	59	29	0	0
1989	0	0	0	0	0
1990	0	0	0	0	0
1991	61	232	106	186	92
1992	26	48	43	6	65
1993	43	28	93	17	32
1994	1	3	11	0	0
Total	**949**	**1385**	**1098**	**318**	**1153**

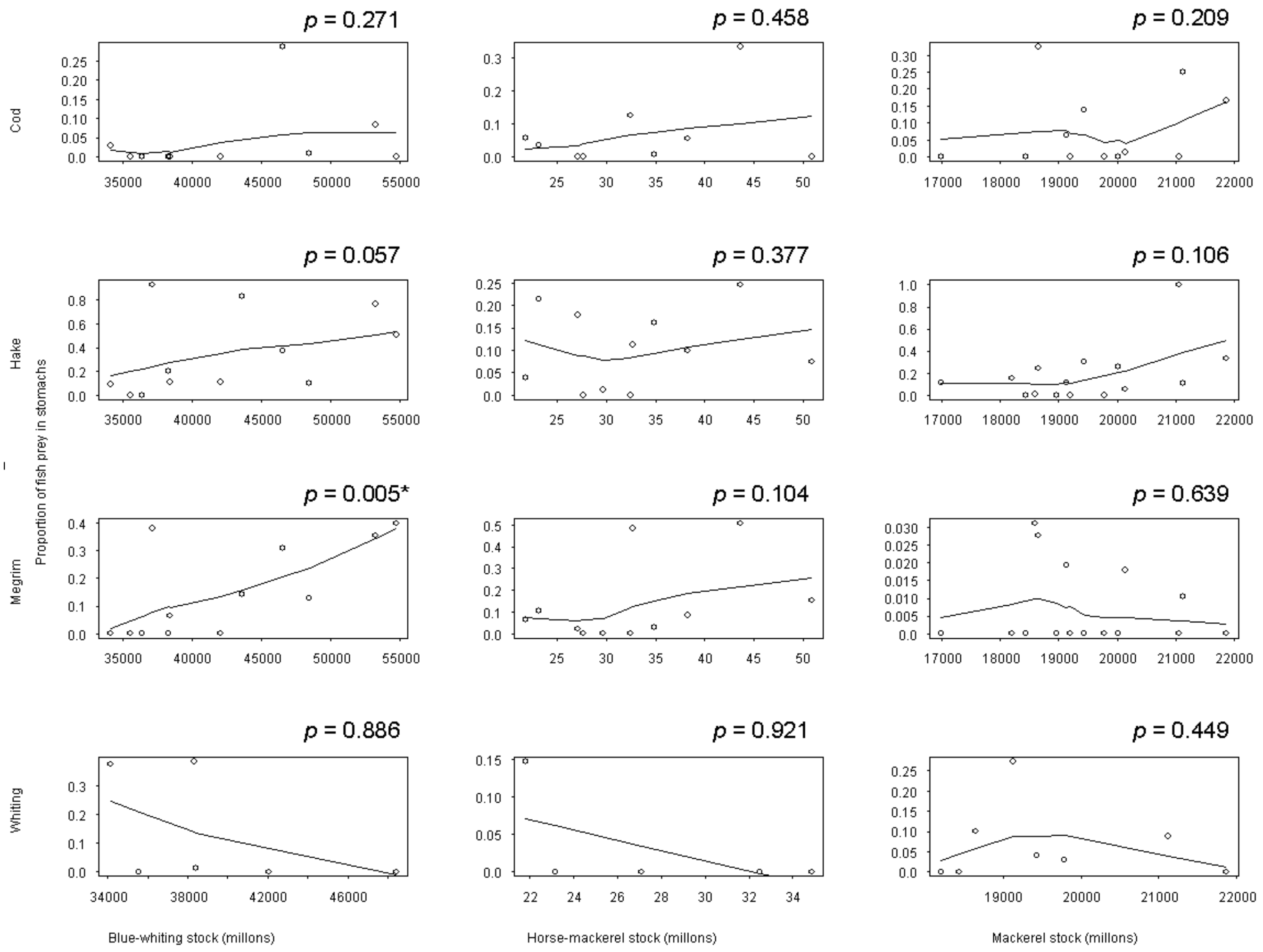
Appendix 2. Total number of fish prey items that have been identified in predator stomachs

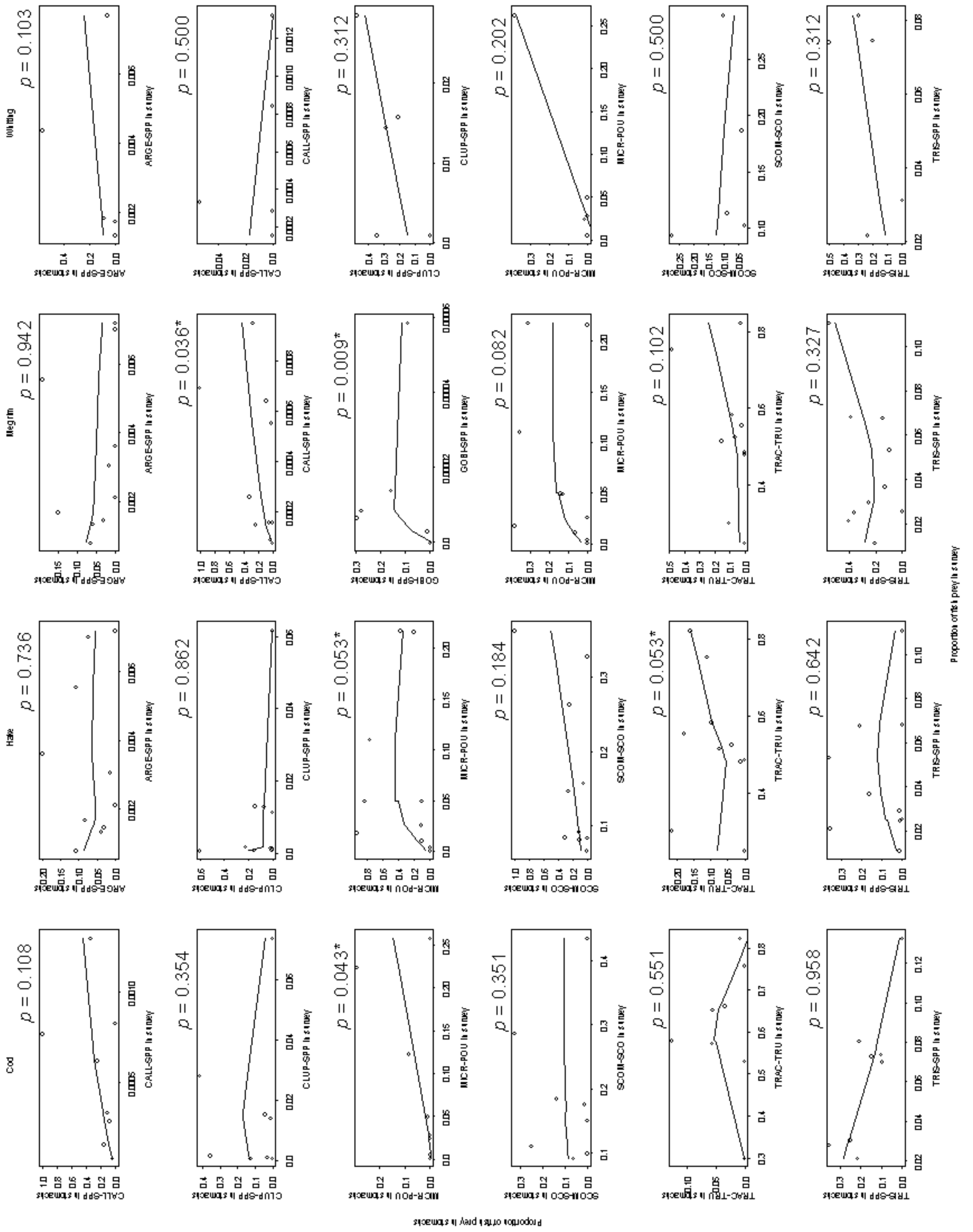
	Predator Species				
Year	Cod	Hake	Megrim	Saithe	Whiting
1977	5	26	26	0	62
1978	0	26	19	4	77
1979	12	4	8	16	2
1980	0	0	0	0	0
1981	36	66	23	59	44
1982	73	27	13	0	0
1983	3	256	354	0	100
1984	1104	222	167	0	847
1985	126	410	362	143	90
1986	0	215	64	0	69
1987	8	5	1	34	0
1988	0	81	29	0	0
1989	0	0	0	0	0
1990	0	0	0	0	0
1991	89	242	112	204	109
1992	29	56	47	6	74
1993	48	28	95	25	34
1994	1	3	11	0	0
Total	**1534**	**1667**	**1331**	**491**	**1508**


Irish Sea
Ireland
UK
Celtic Sea
France
Bay of Biscay
54
52
50
48
-12
-10
-8
-6
-4
-2


a. Hake
Percentage
100
90
80
70
60
50
40
30
20
10
0
<10
15-19
25-29
35-39
45-49
55-59
65-69
75-79
85-89
95-99
>105
Predator Length (cm)
b. Whiting
Percentage
10-14
15-19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
Predator Length (cm)
c. Megrim
Percentage
10-14
15-19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
Predator Length (cm)
d. Cod
Percentage
10-14
20-24
30-34
40-44
50-54
60-64
70-74
80-84
90-94
100-104
110-114
Predator Length (cm)
e. Haddock
Percentage
15-19
25-29
35-39
45-49
55-59
65-69
75-79
85-89
Predator Length (cm)
f. Saithe
Percentage
40-49
50-59
60-69
70-79
80-89
90-99
100-109
Predator Length (cm)
Polychaete
Fish
Crustacean
Cephalopod
Mollusc
Echinoderm
Cnidarian


a. Hake
Prey Length (cm)
Predator Length (cm)
b. Whiting
Prey Length (cm)
Predator Length (cm)
c. Megrim
Prey Length (cm)
Predator Length (cm)
d. Cod
Prey Length (cm)
Predator Length (cm)
e. Saithe
Prey Length (cm)
Predator Length (cm)
Argentina spp.
Callionymus spp.
Clupeidae
Micromesistius poutassou
Scomber scombrus
Trachurus trachurus
Trisopterus spp.
Other fish
Mean (LS)
Median (50%)
Max (90%)
Min (10%)


p = 0.271
p = 0.458
p = 0.209
p = 0.057
p = 0.377
p = 0.106
p = 0.005*
p = 0.104
p = 0.639
p = 0.886
p = 0.921
p = 0.449
Cod
Hake
Megrim
Whiting
Proportion of fish prey in stomachs
Blue-whiting stock (millons)
Horse-mackerel stock (millons)
Mackerel stock (millons)
35000
40000
45000
50000
55000
34000
38000
42000
46000
25
30
35
40
45
50
22
24
26
28
30
32
34
17000
18000
19000
20000
21000
22000


Cod
p = 0.108
p = 0.354
p = 0.043*
p = 0.351
p = 0.551
p = 0.958
Hake
p = 0.736
p = 0.862
p = 0.053*
p = 0.184
p = 0.053*
p = 0.642
Megrim
p = 0.942
p = 0.036*
p = 0.009*
p = 0.082
p = 0.102
p = 0.327
Whiting
p = 0.103
p = 0.500
p = 0.312
p = 0.202
p = 0.500
p = 0.312
Proportion of fish prey in stomachs
Proportion of fish prey in survey

I.4 Whiting model for Celtic Sea

V. Trenkel, Ifremer, France

A gadget has been set up for the case of whiting in the Celtic sea. Below are summarised the model choices, some initial results and the Gadget input files.

General model choices
ages: 1 to 10 years
length classes: 6 to 68 cm (1 cm length classes)
area: Celtic Sea ICES areas 7e-k
years: 1991-1998
time step: quarter

Fleets

- French grondfish survey (autumn) • French commercial landings
- Other catch (same selectivity as French trawlers); total landings from ICES data base excluding English and French landings
- English commercial landings
- English groundfish survey (spring)

Trawl selectivity function
logistic model

Population dynamics
growth function
von Bertalanffy expressed as weight growth with fixed weight length relationship
W= 0.00276 L^3.358 cm -> gram from Dorel (Ifremer internal database)
Linf= 66.2; t0=-0.95; k=0.175 from Dorel

reference weights
calculated using above weight-length relationship

natural mortality
fixed: different per age; values are guesses

recruitment
is estimated

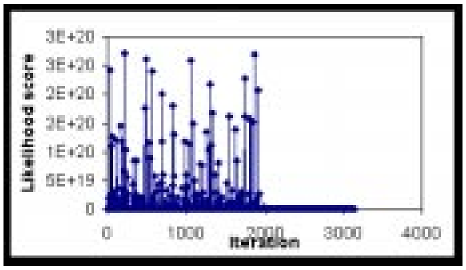
Likelihood components
Survey index : Evohe (French groundfish survey) numbers per haul by 5cm length classes multiplied by 50000; linear relationship of Population numbers and survey index.

Results

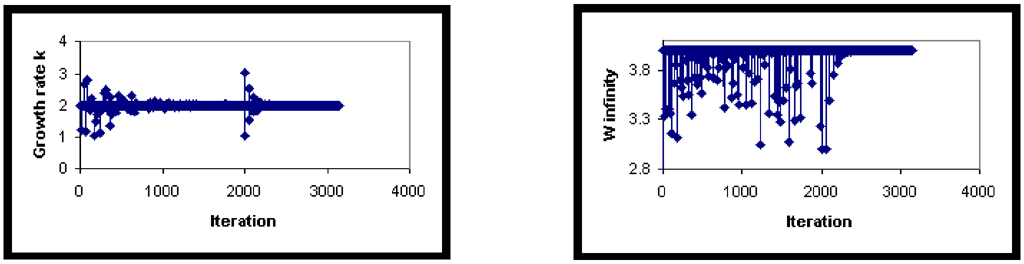
After running the likelihood optimisation option of Gadget several times, the whiting model has still not converged. The overall likelihood score stabilised somewhat (Figure 1a). However, both parameters of the von Bertalanffy biomass growth function got stuck at their respective boundaries (Fig 1b). Similar problems were encountered for other parameters (examples in Fig 1c).

Figure 1. Results of maximum likelihood parameter estimation with Gadget for the Celtic Sea whiting model. a) Overall likelihood score at each iteration step. b) Parameters of von Bertalanffy biomass growth function (k and W infinity). c) Initial numbers at age 2 and numbers of recruits in 1991.

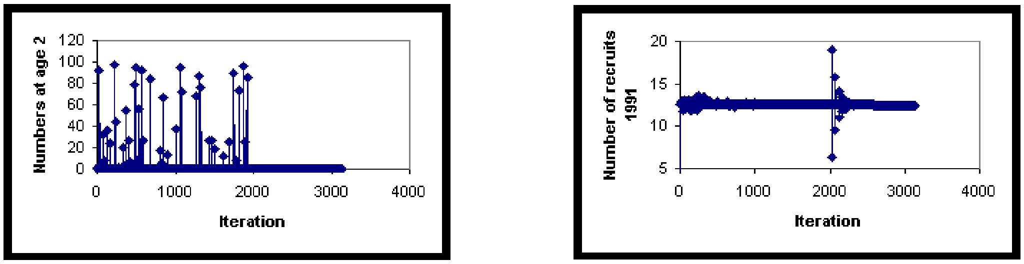
a)



b)



c)



In conclusion it can be said that the maximum likelihood estimation has failed so far. More investigations into the possible reasons for this failure need to be carried out. A closer look at the parameter boundary definitions and the model specification seems to be required.

Gadget input file: main

```
;This is the main file. It specifies which other files need to
be read
;
timefile     time    ; which years will the model run for
areafile     area    ; which areas wil the model run on
printfile    printfile  ; if commented out there will be no printing
;
[stock]                 ; description of the stockdata
stockfiles   whg
;
[tagging]               ; no tagging experiments
[otherfood]             ; food sources not otherwise accounted for in the
model
;
[fleet]                 ; description of the fleet
```

```
fleetfiles   fleets
;
[likelihood]            ; description of the likelihood components
likelihoodfiles likelihood
```

Gadget input file: whg

```
stockname     whg    ; the name of this stock
livesonareas    1    ; There is only one area in this run
minage          1    ; The stock is assumed to have ages from 1 to 10
maxage         10
minlength       6    ; Length range for 1-10 year old whiting
maxlength      74    ; lower bounds of length category!!!
dl              1
refweightfile   refweights.whg ; Keeps $W_ref$
; Minimum and maximum allowed length for each age group
; age    1    2     3     4     5     6     7     8     9     10
;
; Growth and consumption are possibly calculated on a coarser scale
; than the population itself. Therefore the lengthgroups for that
; have different endpoints than usual.
; growthandeatlengths   6  8 10 12 14 16 18 20 22 24 26 28 30 32
34 36 38 40 42 44 46 48 50 52 54 56 58 60  62 64 66 68
growthandeatlengths    whg.len.agg
;
; The individuals of the stock exhibit growth, so the following
defines the growth function
;
doesgrow     1
growthfunctionnumber 3 ; von Bertalanffy growth
Wgrowthparameters 5#whg.growth.k 0 7#whg.growth.winf 7#whg.growth.winf
0.666 1
;p_{0} p_{1} p_{2} p_{3} p_{4} p_{5} p_{6} p_{7} p_{8}
Lgrowthparameters 1 0 0 1 0 1 0.00885e-3 3.0257 0.5
power 3.358           ; The power in the length-weigth relationship
;;; taken from Dorel
maxlengthgroupgrowth 20 ; Binomial-style n for beta-binomial
beta     1#whg.betabin ; Beta-parameter in beta-binomial distribution
;
; Natural mortality is a vector with one number per age group
; - this is residual M (i.e. M1)
;
; age                1 2 3 4 5 6 7 8 9 10
naturalmortality     0.5 0.35 0.2 0.2 0.2 0.2 0.2 0.3 0.4 0.7
iscaught       0     ; The fleet is considered as predator
; so it eats instead of catches
iseaten          1
;
; The endpoints of the lengthgroups used when acting as a prey
preylengths whg.len.agg
doeseat          0
;doeseat 1
; suitability
; preyname1 fn p1 p2... pN1 ; (page 29 in users manual)
; preyname2 fn p1 p2... pN2
; whg 2 1.0#pred.a0 1.0#pred.a1 1.0#pred.a2 1.0#pred.a3 1.0#pred.a4
; maxconsumption m0 m1 m2 m3 ; (page 71 in users manual)
; halffeedingvalue hf ; feeding level half value
; Initial conditions are specified for each area, so only once here
```

```
initialconditions
numbers
minage       1          ; The fish comes into the model at age 1
maxage      10          ; Assume they only get 10 years old
minlength    6
maxlength   68
dl           1
; age       1  2    3     4     5     6     7     8     9    10
agemultiple 0 10000*0.0043#whg.2 6065.3*0.0043#whg.3
3678.8*0.0043#whg.4 2231.3*0.0043#whg.5 1353.4*0.0043#whg.6
820.8*0.0043#whg.7 497.9*0.0043#whg.8 302*0.0043#whg.9
10*0.0043#whg.9
initstockfile  whg.init
doesmigrate      0      ; Only one area in this example,
;so there is no migration
doesmature       0      ; Maturation is not included here
doesmove         0      ; There is only one stock,
; so no movements between stocks
doesrenew        1      ; Recruitment is included in this example
renewaldatafile whg.rec ; Get recruitment information from whg.rec
doesspawn        0      ; As they don’t maturate, they don’t spawn
;
```

Model input file: likelihood

```
;Likelihood file
;
;
[component]
name            meanl.surfr         ; name for this component
weight          2e-6                ; weight for this component
type            CatchStatistics     ; type of component
datafile        whg.meanle.surfr    ; datafile for component
functionnumber     2
; refer to book - function 2 is SUM(lbar-lhat)^2/(s^2)
overconsumption     0
areaaggfile     allarea.agg
ageaggfile      age.agg
fleetnames      whg.surveyfr
stocknames      whg
;
;
[component]
name            meanl.catchfr      ; name for this component
weight          0.3e-6             ; weight for this component
type            CatchStatistics    ; type of component
datafile        whg.meanle.catchfr     ; datafile for component
functionnumber       2             ; The log-likelihood function used
overconsumption      1             ; overconsumption of the stock
;is to be taken into account
areaaggfile     allarea.agg
ageaggfile      age.agg
fleetnames      whg.commercialcatchfr
stocknames      whg
;
;
[component]
name            ldist.surfr           ; name for this component
weight          0.05e-6               ; weight for this component
```

```
type            CatchDistribution     ; type of component
datafile        whg.lengthdist.surfr  ; datafile for component
functionnumber      1                 ; distribution of the data
; is multinomial
overconsumption     1                 ; Overconsumption of the stock
; is to be taken into account
minimumprobability     20    ; This is used if the outcome that occurs
;is very inprobable
areaaggfile     allarea.agg
ageaggfile      allage.agg
lenaggfile      whg.len.agg
fleetnames      whg.surveyfr
stocknames      whg
;
;
[component]
name            ldist.catchfr         ; name for this component
weight          3e-6                  ; weight for this component
type            CatchDistribution     ; type of component
datafile        whg.lengthdist.catchfr   ; datafile for component
functionnumber     1                  ; The distribution of the data
; is multinomial
overconsumption     1                ; Overconsumption of the stock is to
;be taken into account
minimumprobability     20            ; This is used if the outcome
; that occurs is very inprobable
areaaggfile     allarea.agg
ageaggfile      allage.agg
lenaggfile      whg.len.agg
fleetnames      whg.commercialcatchfr
stocknames      whg
;
;
[component]
name            alkeys.fr           ; name for this component
weight          7e-6                ; weight for this component
type            CatchDistribution   ; type of component
datafile        whg.alkeys.fr       ; datafile for component
functionnumber     1                ; The distribution of the data
; is multinomial
overconsumption     1               ; Overconsumption of the stock is to
; be taken into account
minimumprobability     20           ; This is used if the outcome
; that occurs is very inprobable
areaaggfile     allarea.agg
ageaggfile      age.agg
lenaggfile      whg.len.agg
fleetnames      whg.surveyfr whg.commercialcatchfr
stocknames      whg
;
;
[component]
name          Understocking       ; name for this component
weight        1e-12               ; weight for this component
type          Understocking       ; type of component
areaaggfile   allarea.agg
fleetnames    whg.commercialcatchfr
YearsAndSteps all all
;
;
[component]
```

```
name          sifrnewThalassa     ; name for this component
weight        70e-4               ; weight for this component
type          SurveyIndices       ; type of component
datafile      surveyindex         ; datafile for component
sitype        lengths             ; type of survey index component
areaaggfile   allarea.agg         ;area aggregation file
lenaggfile    si10to68fr.dat.agg
stocknames    whg
fittype       FixedSlopeLinearFit
slope          1                  ; beta is fixed
;
;
[component]
name          bounds          ; name for this component
weight          1             ; weight for this component
type          Penalty         ; type of component
datafile      penaltyfile     ; datafile for component
```

Model input file : refinput.file

```
switch              value          lower       upper    optimize
whg.growth.k        0.20439453     0.01        2           1 ;
whg.growth.winf     3.0000000      3           4           1 ;
whg.betabin         2.9482457      0.001       50          1 ;
whg.2               0.43585183     0.0005      100         1 ;
whg.3               0.027550016    0.0005      100         1 ;
whg.4               0.021510385    0.0005      100         1 ;
whg.5               0.019383210    0.0005      100         1 ;
whg.6               0.005383045    0.0005      100         1 ;
whg.7               0.000507812    0.0005      100         1 ;
whg.8               0.000500000    0.0005      100         1 ;
whg.9               6.15113e-05    0.00005     100         1 ;
whg.R091            12.296461      0.001       50          1 ;
whg.R092            7.3437500      0.001       50          1 ;
whg.R093            36.126549      0.001       50          1 ;
whg.R094            47.625732      0.001       50          1 ;
whg.R095            49.577637      0.001       50          1 ;
whg.R096            49.375000      0.001       50          1 ;
whg.R097            0.032226562    0.001       50          1 ;
whg.R098            33.250000      0.001       50          1 ;
whg.R099            8.2000000      0.001       50          1 ;
whg.selalphasurfr   -50.340770     -100        50          1 ;
whg.selbetasurfr    0.18954556     0.001       50          1 ;
whg.selalphacatchfr -15.513556     -60         50          1 ;
whg.selbetacatchfr  0.37458978     0.001       50          1 ;
```

I.5 Blue-Whiting model for Celtic Sea

John K. Pinnegar & Alex Tidd, CEFAS, UK

A gadget has been set up for blue-whiting (*Micromesistius poutassou*) in the Celtic sea. Below the model configuration is described and the Gadget (Version 2) input files are outlined.

This particular model draws heavily from ICES working group reports, and data for the Bay of Biscay. Very little data is available from French or UK sources, since most of the data collected for the ICES 'southern sub-stock' are assembled by Spain and Portugal.

General model choices
ages: 1 to 10 years
length classes: 9 to 41 cm (2 cm length classes)
area: Celtic Sea ICES areas 7e-k
years: 1984-1998
time step: quarter

Fleets

- English groundfish survey (spring and autumn)
- Portuguese and Spanish Commercial (age and length) data [from ICES reports]
- ICES overall blue-whiting catches for area 7e-k.

Trawl selectivity function
logistic model

Population dynamics
growth function
von Bertalanffy expressed as weight growth with fixed weight length relationship
W = 0.0038 L^3.082 cm - > gram from UK Survey data (1984-1998)
Linf= 39.9; t0 = -3.15; k=0.15 from (Bay of Biscay)

reference weights
calculated using above weight-length relationship

natural mortality
fixed: different per age; values are guesses

recruitement
is estimated

Likelihood components Currently no ALK data implemented but Spanish data has now been provided for 1995-2001 and will be added as soon as possible. (Also currently no survey index data)
Mean length data (survey): From UK Groundfish Survey (1984-1998) [Entered 29/05/02]
Length distribution (catch): From Spanish and Portuguese catches (reported in ICES WG report)
Length distribution (survey): From UK Groundfish Survey (1984-1998) [Entered 9/10/02]

WHB Stock file

```
stockname    whb      ; the name of this stock Blue-Whiting Model 25/02/03
livesonareas   1   ; There is only one area in this run
minage         1   ; The stock is assumed to have ages from 1 to 14
maxage        10
minlength      9   ; Length range for 1-10 year old whb (according to
UK Survey and commercial data)
```

```
maxlength     41
dl             1
refweightfile   refweights.whb ; Keeps $W_{ref}$
; Minimum and maximum allowed length for each age group
; age        1    2    3    4    5    6    7    8    9    10
;
; Growth and consumption are possible calculated on a coarser scale
; than the population itself. Therefore the lengthgroups for that
; have different endpoints than usual.
growthandeatlengths whblen.agg
;
; The individuals of the stock exhibit growth, so the following
; defines the growth function
;
doesgrow    1
growthfunctionnumber    3    ; Growth functions number 3 are:
; $\Delta w = \Delta tq_{0}e_{q_{1}T}(( \frac{W}{q_{2}} )^q_{4}
; - ( \frac{W}{q_{3}} )^q_{5})$,
; where W is mean weigth, T is temperature and q is vector of parameters.
; Now we let
; $r := \frac{W - ( p_{0} + p_{8} (
p_{1} + p_{2}p_{8} ) ) W_{ref}}{W}$,
; where $W_{ref}$ is the reference weigth and
; $f(x) := \begin{array}{11} 0 \textrm{if p_{3} + p_{4}x \leq 0,} p_{5}
; \textrm{if p{\o}{3} + p_{4}x \geq p_{5}, \\ p_{3} + p_{4}x
\textrm{else.} \end{array}$
; Then we let
; $\Delta l = \frac{\Delta w}{p_{6}p_{7}l_{p_{7}-1}} f(r)$,
; where p is vector of parameters.
;                $q_{0}   q_{1}   q_{2}   q_{3}  q_{4} q_{5}$
;[winf =p6*Linf^p7 q0=k*p7*winf q4=(p7-1)p7][Linf =
39.9, k= 0.15, a=3.8e-6, b=3.082 From FB961 & FB6014]
;
Wgrowthparameters   #whbgrowth.k  0  #whbgrowth.winf  #whbgrowth.winf
0.6755  1
;$p_{0}    p_{1}   p_{2}   p_{3}   p_{4}   p_{5}   p_{6}   p_{7}   p_{8}$
Lgrowthparameters       1   0    0   1   0   1   3.8e-6   3.082   0.5
power          3          ; The power in the length-weigth relationship
maxlengthgroupgrowth 10 ; Binomial-style n for beta-binomial
[How many length groups out of 16 possible, is a fish allowed
to grow in any given year]
beta         1#whb567 ; Beta-parameter in beta-binomial distribution
;
; Natural mortality is a vector with one number per age group
; - this is residual M (i.e. M1)
;
; age           1    2    3    4    5    6    7    8    9    10
naturalmortality    0.5   0.35   0.2   0.2   0.2   0.2   0.2   0.3  0.4  0.7
iscaught        0    ; The fleet is considered a predator, so it eats instead
; of catches
iseaten         1
;
; The endpoints of the lengthgroups used when acting as a prey
preylengths whblen.agg
doeseat         0
; Initial conditions are specified for each area, so only once here
initialconditions
numbers
minage          1    ; The fish comes into the model at age 1
maxage         10    ; Assume they only get 10 years old
minlength       9
```

```
maxlength      41
dl              1
; age           1    2    3    4    5    6    7    8    9    10
agemultiple 0  0.0043#whb2 0.0043#whb3 0.0043#whb4 0.0043#whb5
0.0043#whb6 0.0043#whb7 0.0043#whb8 0.0043#whb9 0.0043#whb9
initstockfile whb.init
doesmigrate     0    ; Only one area in this example, so there is no
migration
doesmature      0    ; Maturation is not included here
doesmove        0    ; Only one stock, so no movements between stocks
doesrenew       1    ; Recruitment is included in this example
renewaldatafile whb.rec     ; Get recruitment information from had.rec
doesspawn       0    ; As they don’t maturate, they don’t spawn
```

WHB Likelihood file

```
;Blue Whiting Micromesistius poutassou [26/02/03]
;
;
;[component]
;name          meanl.sur           ; name for this component
;weight        2e-6                ; weight for this component
;type          CatchStatistics     ; type of component
;datafile      whb.meanle.sur      ; datafile for component
;functionnumber   2    ; refer to book -
;                        function 2 is SUM(lbar-lhat)$^$2/(s$^$2)
;overconsumption  0
;areaaggfile   whballarea.agg
;ageaggfile    whbage.agg
;fleetnames    whbsurvey
;stocknames    whb
;
;
;[component]
;name          meanl.catch         ; name for this component
;weight        0.3e-6              ; weight for this component
;type          CatchStatistics     ; type of component
;datafile      whb.meanle.catch    ; datafile for component
;functionnumber   2     ; The loglikelihood function used, see above
;overconsumption  1     ; Overconsumption of the stock is to be taken
into account
;areaaggfile   whballarea.agg
;ageaggfile    whbage.agg
;fleetnames    whbcommercialcatch
;stocknames    whb
;
;
[component]
name           ldist.sur         ; name for this component
weight         0.05e-6           ; weight for this component
type           CatchDistribution        ; type of component
datafile       whb.lengthdist.sur       ; datafile for component
functionnumber    1     ; The distribution of the data is multinomial
overconsumption   1     ; Overconsumption of the stock is to be taken
;into account
minimumprobability    20 ; used if the outcome that occurs is improbable
areaaggfile    whballarea.agg
ageaggfile     whballage.agg
lenaggfile     whblen.agg
```

```
fleetnames     whbsurvey
stocknames     whb
;
;
[component]
name           ldist.catch         ; name for this component
weight         3e-6                ; weight for this component
type           CatchDistribution   ; type of component
datafile       whb.lengthdist.catch        ; datafile for component
functionnumber      1    ; The distribution of the data is multinomial
overconsumption     1    ; Overconsumption of the stock is to be taken
; into account
minimumprobability  20    ; used if the outcome that occurs is improbable
areaaggfile    whballarea.agg
ageaggfile     whballage.agg
lenaggfile     whblen.agg
fleetnames     whbcommercialcatch
stocknames     whb
;
;
;[component]
;name          alkeys.sur         ; name for this component
;weight        7e-6               ; weight for this component
;type          CatchDistribution        ; type of component
;datafile      whb.alkeys.sur     ; datafile for component
;functionnumber     1    ; The distribution of the data is multinomial
;overconsumption    1    ; Overconsumption of the stock is to be taken
; into account
;minimumprobability    20 ; used if the outcome that occurs is improbable
;areaaggfile   whballarea.agg
;ageaggfile    whbage.agg
;lenaggfile    whblen.agg
;fleetnames    whbsurvey
;stocknames    whb
;
;
[component]
name           alkeys.catch         ; name for this component
weight         2e-6                 ; weight for this component
type           CatchDistribution       ; type of component
datafile       whb.alkeys.catch        ; datafile for component
functionnumber      1    ; The distribution of the data is multinomial
overconsumption     1    ; Overconsumption of the stock is to be taken
; into account
minimumprobability    20 ; used if the outcome that occurs is improbable
areaaggfile    whballarea.agg
ageaggfile     whbage.agg
lenaggfile     whblen.agg
fleetnames     whbcommercialcatch
stocknames     whb
;
;
[component]
name           Understocking         ; name for this component
weight         1e-12                 ; weight for this component
type           Understocking         ; type of component
areaaggfile    whballarea.agg
fleetnames     whbcommercialcatch
YearsAndSteps  all all
;
;
```

```
;[component]
;name          whbsurveyindex         ; name for this component
;weight        70e-4                  ; weight for this component
;type          SurveyIndices          ; type of component
;datafile      whbsurveyindex         ; datafile for component
;sitype        lengths            ; type of survey index component
;areaaggfile   allarea.agg        ;area aggregation file
;lenaggfile    whb.si.dat.agg
;stocknames    whb
;fittype       FixedSlopeLinearFit  ; THIS WILL NEED TO BE CHANGED LATER!
;slope         1                      ; $\beta$ is fixed
;
;
;[component]
;name          nvpa                   ; name for this component
;weight        0                      ; weight for this component
;type          SurveyIndices          ; type of component
;datafile      whbsurveyindex         ; datafile for component
;fittype       FixedLogLinearFit
;slope         1
;
;
[component]
name           bounds                 ; name for this component
weight         1                      ; weight for this component
type           Penalty                ; type of component
datafile       penaltyfile            ; datafile for component
```

WHB Fleets file

```
;
[fleetcomponent]
totalfleet whbsurvey ; name of fleet
livesonareas 1 ;areas for the fleet
; UK Survey Data, Entered 09/10/02
;
; Number of areas
;
; Length of predator, has no meaning here
lengths 20 30
multiplicative 1.0 ; scaling factor
;
; The suitability function used here is of type
; S(l,L) = 1/( 1 + exp( -alpha - beta*l - gamma*L ) ),
; where l is length of prey and L is length of predator.
; The upperbound for the suitability function is not used here.
suitability
;    stock    type    alpha     beta    gamma    upperbound
whb function ExpsuitfuncA -20#whb403 1#whb404 0 1
amount whbfleet.data ; where the catch data is stored
;
[fleetcomponent]
totalfleet whbcommercialcatch ; name of fleet
livesonareas    1 ;areas for the fleet
; Number of areas (Data entered 09/10/02, Based on ICES but split
into quarters in accordance with 2002 WG Report)
;
; Length of predator, has no meaning here
lengths 20 30
```

```
multiplicative 1.0 ; scaling factor
;
; The suitability function used here is of type
; $S(l,L) = \frac{1}{( 1 + exp( -alpha - beta*l - gamma*L ) )}$,
; where l is length of prey and L is length of predator.
; The upperbound for the suitability function is not used here
suitability
;    stock    type    alpha    beta     gamma    upperbound
whb function ExpsuitfuncA -20#whb401 0.65#whb402 0 1
;
amount whbfleet.data ; where the catch data is stored
;
[fleetcomponent]
linearfleet predc ; name of fleet
livesonareas 1 ;areas for the fleet
; Number of areas
;
; Length of predator, has no meaning here
lengths 20 30
multiplicative 0.5 ; scaling factor
;
; The suitability function used here is of type
; S(l,L) = 1/( 1 + exp( -alpha - beta*l - gamma*L ) ),
; where l is length of prey and L is length of predator.
; The upperbound for the suitability function is not used here
suitability
;    stock    type    alpha    beta     gamma    upperbound
whb function ExpsuitfuncA -20#whb401 0.65#whb402 0 1
;
; The amount number is multiplied by this to get the real amount used
amount whbfleet.data ; where the catch data is stored
```

Penalty File

```
default 2    10000    10000; default setting
;
; variable is given
; variable is given
;
;marked lowerbound - upperbound
whbgrowth.k    2    10000    10000
whbgrowth.winf    2    10000    10000
whb567 2 10000 10000
whb2 2 10000 10000
whb3 2 10000 10000
whb4 2 10000 10000
whb5 2 10000 10000
whb6 2 10000 10000
whb7 2 10000 10000
whb8 2 10000 10000
whb9 2 10000 10000
whb384 2 10000 10000
whb385 2 10000 10000
whb386 2 10000 10000
whb387 2 10000 10000
whb388 2 10000 10000
whb389 2 10000 10000
whb390 2 10000 10000
whb391 2 10000 10000
```

```
whb392 2 10000 10000
whb393 2 10000 10000
whb394 2 10000 10000
whb395 2 10000 10000
whb396 2 10000 10000
whb397 2 10000 10000
whb398 2 10000 10000
whb399 2 10000 10000
whb403 2 10000 10000
whb404 2 10000 10000
whb401 2 10000 10000
whb402 2 10000 10000
```

WHB Recruitment File

```
Normaldistribution
minlength 5    ; minimum length of the recruitment
maxlength 65 ; maximum length of the recruitment
dl 1 ; length steps for recruitment
;exponents 0.00632e-3     3.13248, for mean sizes at age 1 see ’cod’
stock file
;The weight of the recruitment is given by W = alpha * length$^$beta
;
;yr step area age number mean len stdev alpha beta
1984 1 1 1    1000*2#whb384    19.2627 1.8814 3.8e-6 3.082
1985 1 1 1    1000*2#whb385    19.2627 1.8814 3.8e-6 3.082
1986 1 1 1    1000*2#whb386    19.2627 1.8814 3.8e-6 3.082
1987 1 1 1    1000*2#whb387    19.2627 1.8814 3.8e-6 3.082
1988 1 1 1    1000*2#whb388    19.2627 1.8814 3.8e-6 3.082
1989 1 1 1    1000*2#whb389    19.2627 1.8814 3.8e-6 3.082
1990 1 1 1    1000*2#whb390    19.2627 1.8814 3.8e-6 3.082
1991 1 1 1    1000*2#whb391    19.2627 1.8814 3.8e-6 3.082
1992 1 1 1    1000*2#whb392    19.2627 1.8814 3.8e-6 3.082
1993 1 1 1    1000*2#whb393    19.2627 1.8814 3.8e-6 3.082
1994 1 1 1    1000*2#whb394    19.2627 1.8814 3.8e-6 3.082
1995 1 1 1    1000*2#whb395    19.2627 1.8814 3.8e-6 3.082
1996 1 1 1    1000*2#whb396    19.2627 1.8814 3.8e-6 3.082
1997 1 1 1    1000*2#whb397    19.2627 1.8814 3.8e-6 3.082
1998 1 1 1    1000*2#whb398    19.2627 1.8814 3.8e-6 3.082
1999 1 1 1    1000*2#whb399    19.2627 1.8814 3.8e-6 3.082
2000 1 1 1    9000    19.2627 1.8814 3.8e-6 3.13248
2001 1 1 1    9000    19.2627 1.8814 3.8e-6 3.13248
2002 1 1 1    9000    19.2627 1.8814 3.8e-6 3.13248
2003 1 1 1    9000    19.2627 1.8814 3.8e-6 3.13248
2004 1 1 1    9000    19.2627 1.8814 3.8e-6 3.13248
2005 1 1 1    9000    19.2627 1.8814 3.8e-6 3.13248
2006 1 1 1    9000    19.2627 1.8814 3.8e-6 3.13248
```

WHB Initial stock (INIT) file [Based on Spanish ALK, 1995-2001]

```
; age    area    distribution    meanlength    stddeviation    condition
1    1    100    19.2627    1.8814    1
2    1    100    20.9527    1.8087    1
3    1    100    22.2602    1.7740    1
4    1    100    23.5560    1.9192    1
5    1    100    24.6504    2.0597    1
```

```
6    1    100    25.8049    2.2335    1
7    1    100    26.8903    2.4060    1
8    1    100    27.3575    2.3194    1
9    1    100    27.6305    2.9344    1
10    1    100    29.4437    3.1084    1
```

I.6 Cod model for Celtic Sea

John K. Pinnegar & Alex Tidd, CEFAS, UK

A gadget has been set up for cod (*Gadus morhua*) in the Celtic sea. Below the model configuration is described and the Gadget (Version 2) input files are outlined.

This particular model suffers problems associated with the survey data. Cod are simply so rare that there are gaps in the survey index, length-distribution and ALK data.

General model choices ages: 1 to 10 years
length classes: 5 to 120 cm (5 cm length classes)
area: Celtic Sea ICES areas 7e-k
years: 1984-1998
time step: quarter

Fleets

- English groundfish survey (spring and autumn), French groundfish survey (autumn)
- English and French commercial data (ALK and length-distribution)
- ICES overall cod catches for area 7e-k.

Trawl selectivity function logistic model

Population dynamics growth function
von Bertalanffy expressed as weight growth with fixed weight length relationship
W= 0.00632e-3 L^3.132 cm -> gram from UK Survey data (1984-1998)
Linf= 123; k=0.323 from LeFranc 1970 (Calais to Boulogne) [FB 929]

reference weights
calculated using above weight-length relationship

natural mortality
fixed: different per age; values are guesses

recruitement
is estimated

Likelihood components Mean length data (survey): From CEFAS Cirolana data (1984-1998) [Entered 29/05/02]
Mean length data (catch): UK & French commercial data (combined)

Length distribution (catch): UK & French commercial data (combined) [Entered 17/05/02]
Length distribution (survey): Combined UK & French Groundfish Survey [Entered 24/05/02]
ALK Catch: Combined French (commercial and survey) and UK commercial [Entered 27/05/02]

Survey ALK and Survey index not currently used as data too sketchy (need to revise size divisions, within Gadget version 2). Also need to standardise in accordance with number of hauls.

Cod Stockfile

```
stockname     cod ; the name of this stock
livesonareas    1 ; There is only one area in this run
minage          1 ; The stock is assumed to have ages from 1 to 14
maxage         10
minlength       5 ; Length range for 1-10 year old Cod (according to
```

```
UK Survey and commercial data)
maxlength     120
dl              1
refweightfile   refweights.cod ; Keeps $W_{ref}$
; Minimum and maximum allowed length for each age group
; age          1     2     3     4     5     6     7     8     9     10
;
; Growth and consumption are possible calculated on a coarser scale
; than the population itself. Therefore the lengthgroups for that
; have different endpoints than usual.
growthandeatlengths codlen.agg
;
; The individuals of the stock exhibit growth,
; so the following defines the growth function
;
doesgrow        1
growthfunctionnumber     3     ; Growth functions number 3 are:
; $\Delta w = \Delta tq_{0}e_{q_{1}T} (( \frac{W}{q_{2}} )^q_{4}
; - ( \frac{W}{q_{3}} )^q_{5} )$,
; where W is mean weigth, T is temperature and q is vector of parameters.
; Now we let
; $r := \frac{W - ( p_{0} + p_{8} (
; p_{1} + p_{2}p_{8} ) ) W_{ref}}{W}$,
; where $W_{ref}$ is the reference weigth and
; $f(x) := \left\{ \begin{array}{11}
0 \textrm{if p_{3} + p_{4}x \leq 0,} \\
p_{5} \textrm{if p{\o}{3} + p_{4}x \geq p_{5},
\\ p_{3} + p_{4}x \textrm{else.} \end{array}$
; Then we let
; $\Delta l = \frac{\Delta w}{p_{6}p_{7}l_{p_{7}-1}} f(r)$,
; where p is vector of parameters.
;                    $q_{0} q_{1} q_{2} q_{3} q_{4} q_{5}$
; winf=(p6*linf^p7)[assumed Linf 123, k 0232 from Fishbase 929]
; [q0=k*p7*winf][q4=(p7-1)/p7]
Wgrowthparameters      #cod.growth.k    0    #cod.growth.winf
#cod.growth.winf
0.666      1
;   p_{0} p_{1}  p_{2}  p_{3}  p_{4}  p_{5}   p_{6}   p_{7}   p_{8}
Lgrowthparameters         1       0       0       1       0       1
0.00632e-3    3.13248      0
power         3.13248     ; The power in the length-weigth relationship
maxlengthgroupgrowth 10   ; Binomial-style n for beta-binomial
beta        1#cod.betabin ; Beta-parameter in beta-binomial distribution
;
; Natural mortality is a vector with one number per age group
; - this is residual M (i.e. M1)
;
; age               1   2    3    4    5    6    7    8    9    10
naturalmortality   0.5 0.35  0.2  0.2  0.2  0.2  0.2  0.3  0.4  0.7
iscaught      0     ; The fleet is considered as predator,
; so it eats instead of catches
iseaten       1
;
; The endpoints of the lengthgroups used when acting as a prey
preylengths codlen.agg
doeseat       0
; Initial conditions are specified for each area, so only once
here
initialconditions
numbers
minage        1     ; The fish comes into the model at age 1
```

```
maxage       10     ; Assume they only get 10 years old
minlength     5
maxlength   120
dl            1
; age       1     2     3     4     5     6     7     8     9     10
agemultiple 0 0.0043#cod.2 0.0043#cod.3 0.0043#cod.4 0.0043#cod.5
0.0043#cod.6 0.0043#cod.7 0.0043#cod.8 0.0043#cod.9 0.0043#cod.9
initstockfile cod.init
doesmigrate    0     ; Only one area in this example, so no migration
doesmature     0     ; Maturation is not included here
doesmove       0     ; Only one stock, so no movements between stocks
doesrenew      1     ; Recruitment is included in this example
renewaldatafile cod.rec      ; Get recruitment information from had.rec
doesspawn      0     ; As they don’t maturate, they don’t spawn
```

Cod Likelihood file

```
;
;
[component]
name             meanl.sur           ; name for this component
weight           2e-6                ; weight for this component
type             CatchStatistics     ; type of component
datafile         cod.meanle.sur      ; datafile for component
functionnumber   2               ; refer to book -
;  function 2 is SUM(lbar-lhat)^2/(s^2)
overconsumption  0
areaaggfile      codallarea.agg
ageaggfile       codage.agg
fleetnames       codsurvey
stocknames       cod
;
;
[component]
name             meanl.catch         ; name for this component
weight           0.3e-6              ; weight for this component
type             CatchStatistics     ; type of component
datafile         cod.meanle.catch    ; datafile for component
functionnumber   2              ; The loglikelihood function used, see above
overconsumption  1              ; Overconsumption of the stock is to be
;  taken into account
areaaggfile      codallarea.agg
ageaggfile       codage.agg
fleetnames       codcommercialcatch
stocknames       cod
;
;
[component]
name             ldist.sur           ; name for this component
weight           0.05e-6             ; weight for this component
type             CatchDistribution   ; type of component
datafile         cod.lengthdist.sur  ; datafile for component
functionnumber   1              ; The distribution of the data is multinomial
overconsumption  1              ; Overconsumption of the stock is to be
                                ; taken into account
minimumprobability     20       ; used if the outcome that occurs
                                  ; is very improbable
areaaggfile      codallarea.agg
ageaggfile       codallage.agg
```

```
lenaggfile       codlen.agg
fleetnames       codsurvey
stocknames       cod
;
;
[component]
name             ldist.catch          ; name for this component
weight           3e-6                 ; weight for this component
type             CatchDistribution    ; type of component
datafile         cod.lengthdist.catch ; datafile for component
functionnumber   1              ; The distribution of the data is multinomial
overconsumption  1              ; Overconsumption of the stock is to be
                                ; taken into account
minimumprobability     20       ; used if the outcome that occurs
                                 ; is very improbable
areaaggfile      codallarea.agg
ageaggfile       codallage.agg
lenaggfile       codlen.agg
fleetnames       codcommercialcatch
stocknames       cod
;
;
;[component]
;name            alkeys.sur           ; name for this component
;weight          7e-6                 ; weight for this component
;type            CatchDistribution    ; type of component
;datafile        cod.alkeys.sur       ; datafile for component
;functionnumber  1                ; The distribution of the data is multinomial
;overconsumption 1                ; Overconsumption of the stock is to
be
                                   ; taken into account
;minimumprobability     20        ; used if the outcome that occurs
                                    ; is very improbable
;areaaggfile     codallarea.agg
;ageaggfile      codage.agg
;lenaggfile      codlen.agg
;fleetnames      codsurvey
;stocknames      cod
;
;
[component]
name             alkeys.catch         ; name for this component
weight           2e-6                 ; weight for this component
type             CatchDistribution    ; type of component
datafile         cod.alkeys.catch     ; datafile for component
functionnumber   1              ; The distribution of the data is multinomial
overconsumption  1              ; Overconsumption of the stock is to
be
                                 ; taken into account
minimumprobability     20       ; used if the outcome that occurs
                                    ; is very improbable
areaaggfile      codallarea.agg
ageaggfile       codage.agg
lenaggfile       codlen.agg
fleetnames       codcommercialcatch
stocknames       cod
;
;
[component]
name             Understocking       ; name for this component
weight           1e-12               ; weight for this component
```

```
type              Understocking       ; type of component
areaaggfile       codallarea.agg
fleetnames        codcommercialcatch
YearsAndSteps     all all
;
;
;[component]
;name             si10                ; name for this component
;weight           70e-4               ; weight for this component
;type             SurveyIndices       ; type of component
;datafile         codsurveyindex      ; datafile for component
;sitype           lengths             ; type of survey index component
;areaaggfile      codallarea.agg      ;area aggregation file
;lenaggfile       codsi10.dat.agg
;stocknames       cod
;fittype          FixedSlopeLinearFit ; THIS MAY HAVE TO BE CHANGED LATER!
;slope            1                   ; $\beta$ is fixed
;
;
[component]
name              bounds              ; name for this component
weight            1                   ; weight for this component
type              Penalty             ; type of component
datafile          penaltyfile         ; datafile for component
;
;
;
;;Stomach content likelihood function
;[component]
;name             CodStomach          ;or something
;weight           1e-05               ;choose a weighting
;type             2                   ;this means its a "ratios" type
;datafile         stomach.data
;numberfile       stomach.number
;minimumprobability     20            ;seems to be used elsewhere
;areaaggfile      codallarea.agg
;predators        cod                 ; the stock that does the eating
;predatorages              ;the predators can be grouped by age or by length
;                          ;to get a length based stomach content change
                           ;the keyword "predatorages" to "predatorlengths"
;ageaggfile      codage.agg  ;this line would read
                               ; "lenaggfile len.agg" for length
;preyaggfile      codprey.agg
```

Fleets file

```
[fleetcomponent]
totalfleet codsurvey ; name of fleet
livesonareas 1 ;areas for the fleet
; French \& UK Survey Data, Entered 24/5/02
;
; Number of areas
;
; Length of predator, has no meaning here
lengths 20 30
multiplicative 1.0 ; scaling factor
;
; The suitability function used here is of type
```

```
; S(l,L) = 1/( 1 + exp( -alpha - beta*l - gamma*L ) ),
; where l is length of prey and L is length of predator.
; The upperbound for the suitability function is not used here.
suitability
;     stock     type     alpha      beta     gamma     upperbound
cod function ExpsuitfuncA -20#cod.selsurv.alpha 1#cod.selsurv.beta 0 1
amount codfleet.data ; where the catch data is stored
;
[fleetcomponent]
totalfleet codcommercialcatch ; name of fleet
livesonareas     1 ;areas for the fleet
; Number of areas
;
; Length of predator, has no meaning here
lengths 20 30
multiplicative 1.0 ; scaling factor
;
; The suitability function used here is of type
; $S(l,L) = \frac{1}{( 1 + exp( -alpha - beta*l - gamma*L ) )}$,
; where l is length of prey and L is length of predator.
; The upperbound for the suitability function is not used here
suitability
;     stock     type     alpha     beta      gamma     upperbound
cod function ExpsuitfuncA -20#cod.selfleet.alpha 0.65#cod.selfleet.beta 0 1
;
amount codfleet.data ; where the catch data is stored
;
[fleetcomponent]
linearfleet predc ; name of fleet
livesonareas 1 ;areas for the fleet
; Number of areas
;
; Length of predator, has no meaning here
lengths 20 30
multiplicative 0.5 ; scaling factor
;
; The suitability function used here is of type
; S(l,L) = 1/( 1 + exp( -alpha - beta*l - gamma*L ) ),
; where l is length of prey and L is length of predator.
; The upperbound for the suitability function is not used here
suitability
;     stock     type     alpha     beta      gamma     upperbound
cod function ExpsuitfuncA -20#cod.selfleet.alpha 0.65#cod.selfleet.beta 0 1
;
; The amount number is multiplied by this to get the real amount used
amount codfleet.data ; where the catch data is stored
```

Penalty File

```
default 2     10000     10000; default setting
;
; variable is given
; variable is given
;
;marked lowerbound - upperbound
cod.growth.k     2     10000     10000
cod.growth.winf     2     10000     10000
cod.betabin 2 10000 10000
```

```
cod.2 2 10000 10000
cod.3 2 10000 10000
cod.4 2 10000 10000
cod.5 2 10000 10000
cod.6 2 10000 10000
cod.7 2 10000 10000
cod.8 2 10000 10000
cod.9 2 10000 10000
cod.R084 2 10000 10000
cod.R085 2 10000 10000
cod.R086 2 10000 10000
cod.R087 2 10000 10000
cod.R088 2 10000 10000
cod.R089 2 10000 10000
cod.R090 2 10000 10000
cod.R091 2 10000 10000
cod.R092 2 10000 10000
cod.R093 2 10000 10000
cod.R094 2 10000 10000
cod.R095 2 10000 10000
cod.R096 2 10000 10000
cod.R097 2 10000 10000
cod.R098 2 10000 10000
cod.R099 2 10000 10000
cod.selsurv.alpha 2 10000 10000
cod.selsurv.beta 2 10000 10000
cod.selfleet.alpha 2 10000 10000
cod.selfleet.beta 2 10000 10000
```

Recruitment File

```
Normaldistribution
minlength 5     ; minimum length of the recruitment
maxlength 65 ; maximum length of the recruitment
dl 1 ; length steps for recruitment
;exponents 0.00632e-3  3.13248, for mean sizes at age 1 see ’cod’ stock file
;The weight of the recruitment is given by W = alpha * length^beta
;
;yr step area age number mean len stdev alpha beta
1984 1 1 1     1000*2#cod.R084     39.78704 3.703158 0.00632e-3 3.13248
1985 1 1 1     1000*2#cod.R085     39.78704 3.703158 0.00632e-3 3.13248
1986 1 1 1     1000*2#cod.R086     39.78704 3.703158 0.00632e-3 3.13248
1987 1 1 1     1000*2#cod.R087     39.78704 3.703158 0.00632e-3 3.13248
1988 1 1 1     1000*2#cod.R088     39.78704 3.703158 0.00632e-3 3.13248
1989 1 1 1     1000*2#cod.R089     39.78704 3.703158 0.00632e-3 3.13248
1990 1 1 1     1000*2#cod.R090     39.78704 3.703158 0.00632e-3 3.13248
1991 1 1 1     1000*2#cod.R091     39.78704 3.703158 0.00632e-3 3.13248
1992 1 1 1     1000*2#cod.R092     39.78704 3.703158 0.00632e-3 3.13248
1993 1 1 1     1000*2#cod.R093     39.78704 3.703158 0.00632e-3 3.13248
1994 1 1 1     1000*2#cod.R094     39.78704 3.703158 0.00632e-3 3.13248
1995 1 1 1     1000*2#cod.R095     39.78704 3.703158 0.00632e-3 3.13248
1996 1 1 1     1000*2#cod.R096     39.78704 3.703158 0.00632e-3 3.13248
1997 1 1 1     1000*2#cod.R097     39.78704 3.703158 0.00632e-3 3.13248
1998 1 1 1     1000*2#cod.R098     39.78704 3.703158 0.00632e-3 3.13248
1999 1 1 1     1000*2#cod.R099     39.78704 3.703158 0.00632e-3 3.13248
2000 1 1 1     9000     39.78704 3.703158 0.00632e-3 3.13248
2001 1 1 1     9000     39.78704 3.703158 0.00632e-3 3.13248
2002 1 1 1     9000     39.78704 3.703158 0.00632e-3 3.13248
2003 1 1 1     9000     39.78704 3.703158 0.00632e-3 3.13248
```

```
2004 1 1 1     9000     39.78704 3.703158 0.00632e-3 3.13248
2005 1 1 1     9000     39.78704 3.703158 0.00632e-3 3.13248
2006 1 1 1     9000     39.78704 3.703158 0.00632e-3 3.13248
```

Initial stock (INIT) file

```
; age     area     distribution     meanlength     stddeviation     condition
1     1     100     39.78704     3.703158     1
2     1     100     56.86063     6.835298     1
3     1     100     74.83688     5.200927     1
4     1     100     84.75750     4.579793     1
5     1     100     91.94226     3.851679     1
6     1     100     96.46679     5.249746     1
7     1     100     99.22684     5.028542     1
8     1     100     102.99690     3.632177     1
9     1     100     97.75000     3.774917     1
10     1     100     105.00000     0     1
```

J North Sea Case Study

J.1 A preliminary validation attempt: a user's point of view

Marco Kienzle, Alessandro Gimona

Fisheries Research Services, PO Box 101 Victoria Road, Aberdeen AB11 9DB, Scotland, UK

1 Introduction

Models are descriptions of objects or processes occurring in the real world. Their purpose is to understand how systems function and eventually predict how they will behave in the future. In science, models are derived by logic from accepted truth. They are used by successive generations of scientists if their hypotheses and predictions are consistent with field observations. Like Newton's law of gravity, models can be falsified by new measurements. Still they can remain useful and accurate descriptions of natural phenomena under certain conditions. In this regard, they are never accepted to be true but are constantly compared to data in order to verify that they do not provide a false description of the real world.

During the last decades, programming has developed tremendously in science. Large amounts of computer models have been written to describe systems for which direct experimentation is impossible or too expensive. Although these models are more complex than a simple equation model, they still have to fulfil the same requirement of falsifiability if they are to be trusted and used. Law and Kelton (2000) and Balci (1998) stress the importance of verification of a simulation model but recognise that it is not an easy task and can even be impossible according to Oreskes et al. (1994). In the following, we will use the definition given by Oreskes et al. (1994) for the words verification[1] and validation[2]

In our particular case, GADGET aims at describing the functioning of the fish population and the fleet of a fishery and their interactions. Besides modelling all the essential component of this system, it also integrates a fitting algorithm that enables the modeller to estimate free parameters of the model by comparing output of the simulation module to field data. This aspect of GADGET makes it more similar to a statistical package specifically designed to estimate biological population parameters than a model built to test hypotheses. The assessment of the adequacy of GADGET as an instrument to infer reliable estimation of particular parameters is a key step in determining its actual and future usefulness.

2 Approach to conduct a successful simulation study

A computer simulation is an implementation of a conceptual or a mathematical model, therefore its verification and validation must be done at both levels. Law and McComas (2001) describe a seven steps approach to conduct a successful simulation study which is summarised in Figure 1. Since GADGET is in an advanced stage of development, actions described by steps 1 to 4 (included) have already been realised. As a user team, we will focus on step 5.

---

[1] **Verification**: (from Latin, *verus*, meaning true) is the assertion or establishment of truth. Verification is possible only for a so-called closed system *e.g.* logic statement or mathematical equation. Therefore its impossible to assess the veracity of a natural system model unless it takes into consideration all possible events *i.e.* is a closed system.

[2] **Validation** "denotes the establishment of legitimacy, typically given in terms of contracts, arguments, and methods. A valid argument is one that does not contain obvious error of logic. By analogy, a model that does not contain known or detectable flaws and is internally consistent can be said to be valid" taken from Oreskes et al. (1994).

2.1 Is the programmed model valid? (step 5)

Before using any computer program to estimate parameters, it is highly recommended to check that it is able to perform correctly the calculations that competing software are able to do (or what we are able to do by hand). The validation of GADGET implies that the software is suitable for parameter estimation *i.e.* is able to converge to an a-priori known solution. We suggest to test the software running it with a synthetic dataset representing a hypothetical, albeit plausible, fish population.

2.2 An hypothetical dataset

2.2.1 Biological parameters

This hypothetical dataset represents a single species fishery taking place in a single area. The time series of the data is arbitrarily chosen to be of 10 years *i.e.* long enough to let the simulation module stabilise. Each year is divided in four time steps (or seasons). The number of age groups is fixed to five, so that fish age is between 1 and 5 years.
Natural mortality (M) is fixed to 0.2 year^{-1} for all age groups except the 5^{th} one for which M=1000 year^{-1} in order to ensure that fishes older than 5 years will disappear from the simulation.
The length-weight relationship (LW) is assumed to be isometric with a Fulton condition factor (Ricker, 1975) fixed to 1×10^{-4} kg cm^{-3}.
Individuals are assumed to grow according to the Von Bertalanffy growth function (VBGF). Details about this function and its implementation in GADGET are described in appendix I, page 308. The following parameters are used

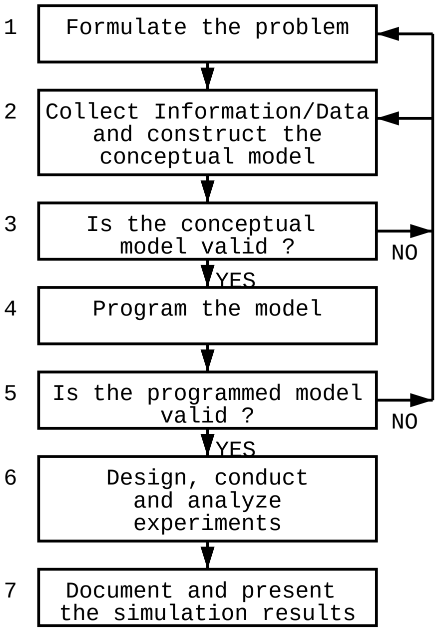


Figure 1: Step wise description of the development of a computer model. Diagram redrawn from Law and McComas (2001).

- $L_\infty = 100$ cm ($\equiv W_\infty = 100$ kg)
- $K = 0.51$ years^{-1}
- $t_0 = 0$ years

According to the two previous parameterisations, fishes follow the length and weight at age presented in Table 1.

Age (years)	**Length** (cm)	**Weight** (kg)
1	39.35	6.09
2	63.21	25.26
3	77.69	46.89
4	86.47	64.65
5	91.79	77.34

Table 1: Summary of the length and weight at age.

Age at recruitment is assumed to be 1 year with a magnitude of 1000. Initial abundance is fixed for each age group to 100.

2.2.2 Fishing fleet parameter

It is assumed that the hypothetical stock is fished by a commercial and a scientific fleet, both fishing during the first time step of each year. The catches of the commercial fleet are fixed to $10,000$ tons while those of the scientific survey are 100 tons.
Age-length keys of both fleets are created with 1000 values per age group distributed according to a Gaussian function with mean equal to the length-at-age given by VBGF (Table 1) and standard deviation equal to 1.
Suitability of both fleets is set to 1 by setting all parameters (α, β and γ) to 2000.

2.2.3 Validation of parameter estimation

An evaluation of the estimation given by GADGET is made using the asymptotic weight parameter (W_∞) of the VBGF (eq.(7) page 309) as an unknown parameter. This is done by enabling W_∞ and β (the parameter of the beta-binomial distribution (Stefánsson, 2000) to vary freely with the constraint of being positive. Results are compared to the expected value *i.e.* the value used to build the hypothetical set of data ($W_\infty = 100$ kg). Starting values are set to $W_\infty = 100$ kg and β =1.

3 Results

3.1 Variation of abundance and catch

Figure 2 shows the variation of the number of fish by age group as a function of time. A stable age structure is reached after a number of years equal to the number of age groups simulated (5 in this case). The abundance of individuals along a cohort decreases smoothly during the year when natural mortality acts alone and sharply at the beginning of each year when the fishing mortality is applied. (Note in Figure 2 that the number of recruits is 10 millions when it has been fixed to a thousand: input values are scaled by a factor 1×10^4.)
In the simulation, fishing mortality is applied before the recruitment during time step one. Therefore, the 1^{st} age group is never present in the catch (Figure 3 and Figure 4). The proportion of number of individuals of each age group in the catch reflects their respective proportions in the population (Figure 3) as a result of the fishing fleet being non-selective, whereas the proportion in weight of each age group in the catch is a trade-off between their abundance in the population and their mean weight-at-age (Figure 4).

3.2 Estimation of growth parameters

The algorithm returns a minimum likelihood of 2.2 corresponding to the following estimations:

- W_∞=50 kg
- β=1.06

Figure 5 shows an asymptotical increase of mean weight as a function of age. For ages 1 and 2, the mean weight is close to the input data while for older individuals it is less than the expected weight.

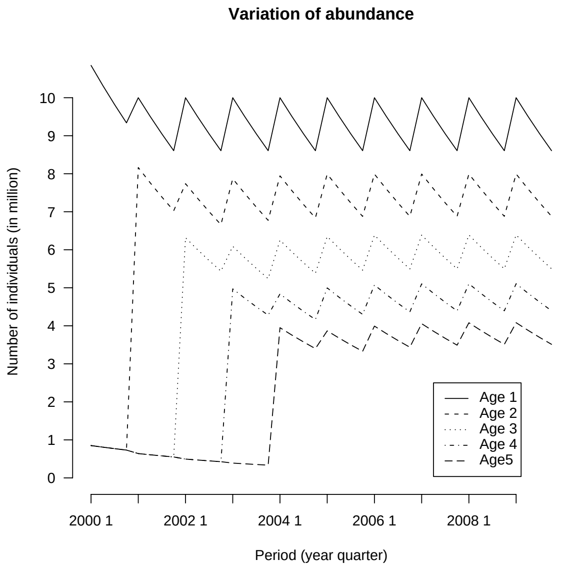


Figure 2: Variation of the abundance of individuals by age group produced by a run of GADGET.

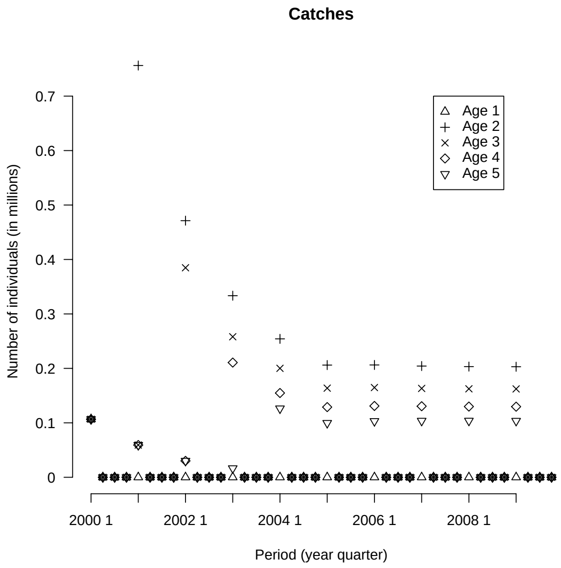


Figure 3: Yearly catches produced by GADGET.

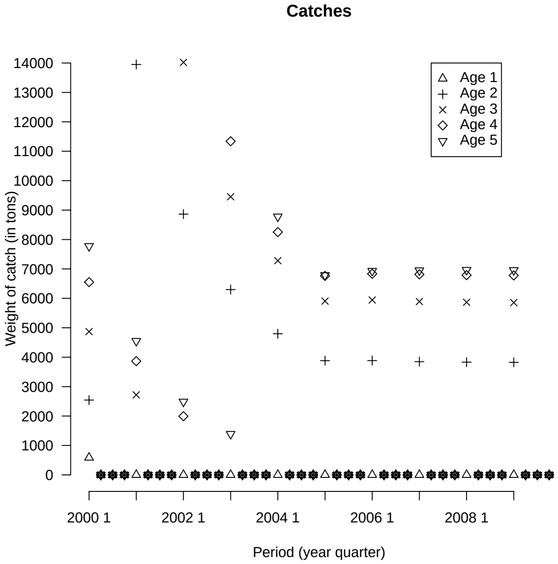


Figure 4: Yearly catches produced by GADGET.

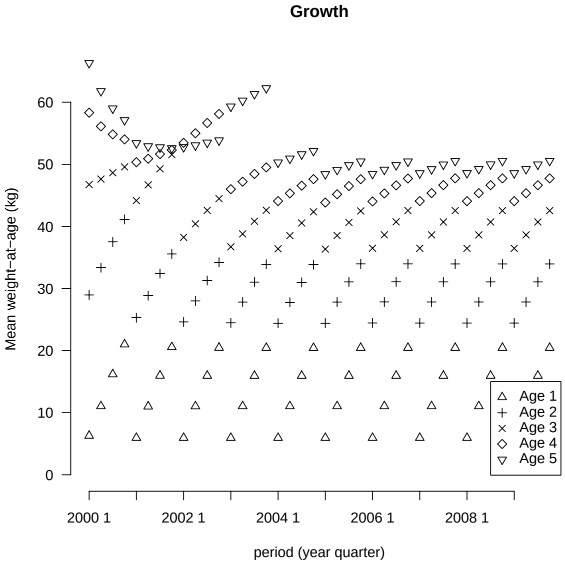


Figure 5: Variation of weight at age produce by GADGET for each age group (W_∞=50 kg, K=0.51 year^{-1}).

3.3 Optimisation of the fit

The previously estimate of W_∞ is small compared to that used to build the set of data. It has been suggested that this may come from an excessively small value of the growth velocity parameter of the VBGF (K). This explanation is tested by "freeing" the parameter θ_0 in the growth function (eq.(7) page 309). The minimisation algorithm is run giving a minimum likelihood value of 0.037 and the following estimations:

- W_∞=109.38 kg
- β=3.56
- θ_0=400.00 years^{-1}

The pattern of growth with increased W_∞ and K is presented in Figure 6. Mean weight-at-age has increased and is greater than the expected weight for all age groups except age 1.

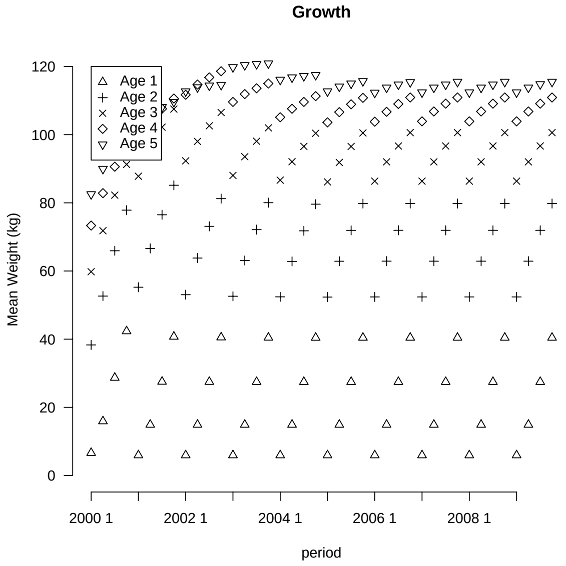


Figure 6: Variation of weight at age produce by GADGET for each age group (W_∞=109.38 kg, K=1.22 year^{-1}).

4 Conclusion

This example of the use of GADGET on an hypothetical set of data highlights some characteristics of this tool for simulation and estimation of biological population parameters. As a simulation tool, GADGET's output corresponds to the current knowledge of the behaviour of an exploited fish population and conforms to other modelling approaches used in fishery research (*e.g.* VPA). For example, natural mortality reduces continuously the abundance of each cohort according to the exponential decay model while fishing mortality acts instantaneously. The codification of the biological information into the software is straight forward and interpretation of the output is easy apart the growth function which requires more effort, in particular for a fishery biologist trained to use only VBGF.

As an estimation tool, the output of GADGET is more difficult to use. First, the output provides a point estimation for each parameter with no measure of its error. In this situation it is impossible to assess the quality of the estimation : at best we can conclude that the estimation differs from the expected result.

Second, GADGET's users could be tempted (as shown in section 3.3) to relax constraints on one or more parameters in order to achieve a better match between the prediction and the expectation. In fact this strategy decreases the optimum likelihood value by nearly a factor of 100 and returns a prediction of W_∞ which differs only by 10% from the expected value. Although this tuning process improves the estimation of W_∞, it provides an over-estimation of the growth velocity (K) by nearly a factor of 3. This problem underlines the risk of compensation between parameters which may become a major one when the number of parameter to be estimated is large.

The results of this simplistic trial might be due to one or more of the following reasons :

1. The dataset is not representative of a realistic fishery situation and therefore is outside the range of application of GADGET.
2. The translation of biological knowledge of the population into GADGET contains errors of codification and/or interpretation.
3. GADGET is not able to converge to an a-priori known solution.

5 Proposed improvements

A solution for the pitfalls described above is the creation of a plausible set of data that can be used to validate the software. According to Law and Kelton (2000), this process of validation will increase both the quality of GADGET and people's confidence to use it (see Law and Kelton (2000) for further discussion on methods).

This dataset could be used for didactic purposes by using it in a tutorial. This should facilitate the learning of GADGET by new users.

Given the necessity to compare the output of GADGET to other model or field data, we suggest that the calculation of error for each of the parameter's estimates should be implemented. This will enable one to check that GADGET is consistent with the prediction of other tools for fish stock assessment.

References

O. Balci. *Handbook of Simulation*, chapter 10. John Wiley & Sons, 1998.

A. M. Law and W. D. Kelton. *Simulation modeling and analysis*. Mc Graw Hill, third edition, 2000.

A. M. Law and M. G. McComas. How to build valid and credible simulation models. In B.A. Peters, J. S. Smith, and M. W. Rohrer, editors, *Proceedings of the Winter Simulation Conference*, 2001.

N. Oreskes, K. Shrader-Frechette, and K. Belitz. Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, 263:641 647, 1994.

W. E. Ricker. Computation and interpretation of biological statistics of fish populations. *Bulletin of the Fisheries Research Board of Canada*, 191:382 p., 1975.

G. Stefánsson. Development of structurally detailed statistically testable models of marine populations. Technical Report QLK5-CT1999-01609, Marine Research of Reykjavík (Iceland), 2000.

APPENDICES

1 Von Bertalanffy growth function in GADGET

1.1 Von Bertalanffy's growth function (VBGF)

The VBGF in length is

$$L(t) = L_\infty \times (1 - exp[-K \times (t - t_0)]) \tag{1}$$

If we are interested in the variations of length through time, we can calculate the derivative of the VBGF

$$\frac{dL}{dt} = K \times (L_\infty - L(t)) \tag{2}$$

If we want to express the variation of the size of the fish using weight instead of length, we can convert length to weight using the length-weight (L/W) relationship. The L/W is [3]

$$W = a \times L^b \;\Leftrightarrow\; L = \sqrt[b]{\frac{W}{a}} = \left(\frac{W}{a}\right)^{\frac{1}{b}} \tag{3}$$

Which imply the following derivative relation,

$$\frac{dW}{dt} = a \times b \times L^{b-1} \times \frac{dL}{dt} \tag{4}$$

Replacing eq.(2) and (3) into eq.(4) we have

$$\frac{dW}{dt} = a \times b \times \left(\frac{W}{a}\right)^{\frac{b-1}{b}} \times K \times \left(\left(\frac{W_\infty}{a}\right)^{\frac{1}{b}} - \left(\frac{W}{a}\right)^{\frac{1}{b}}\right)$$

Re-arrangement of the expression gives

$$\frac{dW}{dt} = a \times b \times K\left(\frac{W_\infty^{\frac{1}{b}} \times W^{\frac{b-1}{b}}}{a} - \frac{W}{a}\right)$$

Simplifying by a and dividing both sides by W_∞ we obtain

$$\frac{1}{W_\infty}\frac{dW}{dt} = b \times K \times \left(\left(\frac{W}{W_\infty}\right)^{\frac{b-1}{b}} - \frac{W}{W_\infty}\right) \tag{5}$$

Assuming the L/W relationship is isometric (*i.e.* b=3), eq. 5 become

$$\frac{1}{W_\infty}\frac{dW}{dt} = 3 \times K \times \left(\left(\frac{W}{W_\infty}\right)^{\frac{2}{3}} - \frac{W}{W_\infty}\right) \tag{6}$$

---

[3] in the following notation $L(t)$ and $W(t)$ will be abbreviate to L and W in order to simplify the notation.

1.2 Growth model in GADGET

The description of growth in GADGET is more general since it includes the effect of temperature, of the year, of the step and of the area :

$$\left.\frac{dW}{dt}\right|_{\substack{Age\,=\,i\\Year\,=\,t\\Step\,=\,s\\Area\,=\,a}} = \left[|\theta_0| \times e^{\theta_1 \times T} \times \left(\left(\frac{W(i)}{|\theta_2|}\right)^{\theta_4} - \left(\frac{W(i)}{|\theta_3|}\right)^{\theta_5}\right)\right] \times E_{Year(t)} \times E_{Step(s)} \times E_{Area(a)} \quad (7)$$

Where
θ is the vector of growth parameters (contains 6 values)
T is the temperature
$W(i)$ is the weight at age i
E stands for the effect of particular variable (here is considered year, step and area)

Eq.(7) collapse into VBGF eq.(6) if the parameters are set to the following values:
$\theta = (b \times K \times W_\infty, 0, W_\infty, W_\infty, \frac{2}{3}, 1)$
$E_{Year} = E_{Step} = E_{Area} = 1$

K Minutes

K.1 Programmers' meeting: Reykjavík, January 2002

A meeting of programmers in the dst^2 project was held at MRI headquarters in Reykjavik, Iceland, 22-25 January, 2002. This (draft) document contains proposals for programming changes which are needed to the current (1.0.03) version of Gadget.

The main output from the meeting is a specification of how and when to implement a major change in input and output file formats for Gadget. It is proposed that a major change in the program be introduced as Gadget 2.0.00 by June 15, along with conversion programs for converting old file formats to the new format.

Biological programming

Length and Weight growth

In charge: Höskuldur
Suggest how to make growth a function of both age and length. Programming task subsequently to be allocated.

Min and Maxlength, growth and eat length and prey length are not required in the stock files. Default is to use all length groups.

1. All stocks should store both number and weight. The reference length-weight relationship is input to the stock and relative condition equal to 1 means that the fish follow that relationship. Relative condition will always be referred to that file.
2. Relative condition as function of time and area will be input from a file. If the file is missing the relative condition will always be 1.

If "growth and eat length" disappear, InterpWgrowth and such things will disappear from the growth classes, simplifying them. Prey lengths and "Growth and eat length" were originally put in to reduce computation in multispecies mode but experience has shown that they are not terribly important for computation time at least in the examples so far. The suggestion here is therefore to let "Growth and eat lengths" and "prey lengths" remain but the default is not to use this so the normal user does not have to know about this. (The same with the banded age-length matrices).

Growth can be modelled as function of weight or length. If only growth in length is modelled, growth in weight is not considered at all, except all the weight will at each timestep be multiplied by the appropriate relative condition (if it changes). The Grower will ask for condition file.

GrowthCalc will be unchanged but new functions added. Growth functions now have names which are:

```
Vonbert-Weight
Vonbert-Length
ExtendedVonbert-Weight
ExtendedVonbert-Length
Jones&Hislop
```

The so-called Multspec growth function (GrowthCalcA) should disappear.

The function GrowthCalc takes the parameters

```
  virtual void GrowthCalc(int area,
                          doublevector& Lgrowth,
                          doublevector& Wgrowth,
                          const popinfovector& GrEatNumber,
                          const AreaClass* const Area,
                          const TimeClass* const TimeInfo,
                          const doublevector& Fphi,
                          const doublevector& Consumption,
                          const LengthGroupDivision* const LgrpDiv) const = 0;
};
```

This can be unchanged between versions as Wgrowth is not active in the functions based on length growth.

After GrowthCalc is calculated the next step in the “Growth process” is to call the function GrowthImplement which generates the “matrices” used to update the age-length matrices. This function can be unchanged.

The next step in the Process is to call the function Agebandmatrix::Grow which updates the length distributions but also lets the growth in weight fit. Here there need to be two instances of the function identified by the additional argument Grower::OnlyLength.

```
1.  Agebandmatrix::Grow(const doublematrix & Lgrowth,
        const doublematrix & Wgrowth );
```

The old function:

```
2.   Agebandmatrix::Grow(const doublematrix & Lgrowth,
      const doublematrix & Wgrowth,int OnlyLengthGrowth);
```

only length growth.

And the one related to maturation:

```
void Agebandmatrix::Grow(const doublematrix & Lgrowth,
                    const doublematrix & Wgrowth,
                    Maturity* const Mat,
                    const TimeClass* const TimeInfo,
                    const AreaClass* const Area,
                    const LengthGroupDivision* const GivenLDiv,
                    int area)
```

the old version

And the one related to maturation:

```
void Agebandmatrix::Grow(const doublematrix & Lgrowth,
                    const doublematrix & Wgrowth,
                    Maturity* const Mat,
                    const TimeClass* const TimeInfo,
                    const AreaClass* const Area,
                    const LengthGroupDivision* const GivenLDiv,
                    int area,int OnlyLengthGrowth)
```

the old version

The calls to this function are now:

```
Alkeys[AreaNr[area]].Grow(grower->LengthIncrease(area),
                                  grower->WeightIncrease(area)
                                grower->OnlyLengthGrowth());
```

The simple memberfunction OnlyLengthGrowth needs to be added to grower. The changes to the function are simple. There is an if statement in the beginning to check if OnlyLengthGrowth is true and the same function without the lines referring to weights are then called. The code would be shorter if the if statements were put inside the loops but that would slow the program down.

When weight increase is not calculated, weights must be calculated based on relative condition. As described earlier, relative condition will be defined as the proportion of the reference length-weight relationship.

In the end of the member function:

```
Stock::Grow(int area,
            const AreaClass* const Area,
            const TimeClass* const TimeInfo)
```

We could have Weight like:

```
        if(Grower->OnlyLengthGrowth())
        Stock::Conditions(Grower->Weight(area,TimeInfo),
                Grower->ConditionChanges(area,TimeInfo));

        Grower->ConditionChanges(area,TimeInfo)
```

is only to reduce calculations so new weights are only calculated when the condition changes. It would be best if condition was only based on maturity stage and time, but the Norwegian code now uses fleet base condition which is probably more realistic but it can partly be approximated by maturity stage and time.

The condition will be read with the growth function.

The options are:

```
year step area condition
1980 1    1    1
1980 1    2    1 ....
```

```
        conditionfile   ‘‘filename’’
```

or

```
        yeareffect      1 1 1 1 1 1 1 1
        stepeffect      1 1 1 1 1 1 1 1
        areaeffect      1 1 1 1 1 1 1 1
```

The conditionfile will have the format:

```
year step area condition
1980 1    1    1
1980 1    2    1
.
.
.
```

The question is if only the years and steps where condition changes should be in the file or all years and steps. Another option would be to have the both the condition file and:

```
yeareffect      1 1 1 1 1 1 1 1
stepeffect      1 1 1 1 1 1 1 1
areaeffect      1 1 1 1 1 1 1 1
```

at the same time and the conditionfile would only list deviations from the main model. yeareffect*stepeffect*areaeffect.

As a byproduct of this work the weight files in initial conditions should disappeared and the condition for each area (and possibly age) should be in the stockfile. The condition should then as before be deviations from the reference length weight relationship.

All this work is base on the reference length weight relationship which is read in the constructor for GrowthCalc as follows

```
  // Read information on reference weights.

  keeper->AddString("reference_weights");
  infile >> text;
  if (strcasecmp(text,"refweightfile") != 0)
    handle.Unexpected("refweightfile", text);
  infile >> text;
  ifstream subfile(text, ios::in);
  CheckIfFailure(subfile, text);
  handle.Open(text);
  CommentStream subcomment(subfile);
  //Read information on length increase.
  doublematrix tmpRefW;
  if (!ReadMatrixNx2(subcomment, tmpRefW))
    handle.Message("Wrong format");
  // Interpolate the reference weights. First there are some error checks.
  for( i = 0; i < tmpRefW.Nrow()-1;i++) {
    if( (tmpRefW[i+1][0] - tmpRefW[i][0] ) <= 0)
      handle.Message("Lengths must be strictly increasing");
  }
  if (LgrpDiv->Meanlength(0) < tmpRefW[0][0] ||
      LgrpDiv->Meanlength(LgrpDiv->NoLengthGroups()-1) >
      tmpRefW[tmpRefW.Nrow()-1][0] )
    handle.Message("Lengths must span the range of Growthlengths");

  handle.Close();
  subfile.close();

  int pos = 0;
  for ( int j = pos; j < LgrpDiv->NoLengthGroups();j++) {
    for( i = pos; i < tmpRefW.Nrow()-1;i++) {
      if(LgrpDiv->Meanlength(j) >= tmpRefW[i][0] && LgrpDiv->
Meanlength(j) <= tmpRefW[i+1][0]) {
        double ratio = (LgrpDiv->Meanlength(j)-tmpRefW[i][0])/
(tmpRefW[i+1][0]-tmpRefW[i][0]);
        Wref[j] = tmpRefW[i][1] + ratio*(tmpRefW[i+1][1]-tmpRefW[i][1]);
        pos = i;
      }
    }
  }
```

This bit of code is repeated in all the constructor for GrowthCalcA, GrowthCalcB etc and should be moved to the constructor for the baseclass GrowthCalcBase. the new member function Grower->Weight(area,TimeInfo) will be based on GrowthCalc:Wref and GrowthCalc->Condition. GrowthCalc is probably not accessible by stock so the member functions would have to be defined as member functions of Grower which used corresponding memberfunctions in GrowthCalc.

In charge: Gunnar
Suggest how to include density dependent mortality in 0-group fish. Programming task subsequently to be allocated.

In charge: Höskuldur
Suggest how to include fleet discard - perhaps as a fleet "digestion" or as two selection functions.

In charge: Auðbjörg
Propose method for inclusion of tagged subpopulations.

Code cleanup

In charge: Morten
Changes need to be done to likelihood and printer classes to make better use of inheritance and clean up the code. This work is in progress

In charge: Auðbjörg
The suitability functions which can be used in Gadget are identified by a number. This should be changed to a string value to make the data files clearer. This number is also dependent on the order in which the suitability functions are declared in the program which makes adding and especially deleting functions difficult. This should be changed so that adding and deleting suitability functions will be easier and minimal changes need to be done to existing data files when suitability functions are added/deleted.

In charge: Höskuldur
Propose a list of names of suitability functions, maturity functions and growth functions.

In charge: Höskuldur
Suggest how to clean up code where remnants of FORTRAN and C linkages still exist.

In charge: Auðbjörg
Linking Gadget classes with paramin causes problems. At least, the input classes should be able to be linked. Need to clean up the routines and classes which read the parameter values, names and bounds.

In charge: Höskuldur
Propose removal of unnecessary suitability functions if any.

Documentation

Suggestions on how to update documentation for both the User's and Programmer's Gadget manuals.

The User's manual (and possibly the Programmer's manual) needs to be updated to ensure that it matches the model. An introductory guide to sit alongside the reference manual would also be helpful. Sample models (model definition and data files) need to be prepared that can be distributed with the model. In addition to the haddock example, an example with more than one species and more than one area is required. This could be a cod-capelin model, or a 'fake' model. As we have specified that we would use fisheries examples a 'real' model would be preferable.

User’s Manual

In charge: NN
The introduction needs to be rewritten to introduce the concepts in a more rational order. The description of the model files will need major rewriting after the changes to the input format. *Delayed until after modification to input file formats.*

Programmer’s manual

In charge: NN
This is a useful, if specialised, document and should help ensure things such as coding standards get a wide distribution. It is already somewhat out of date, and this will become worse following the re-write. The manual is partially automatically generated, with the resulting file edited by hand. There is a potential problem in that this manual can only be automatically generated if the correct procedures have been followed in writing and modifying the code, which may mean updating this manual is more difficult than it should be. *Delayed until after modification to input file formats.*

Introduction to Gadget

In charge: Daniel
This is to contain an introduction to using Gadget program and a guide to model building, both aimed at first time users of the program. To be useful, it needs to refer to specific features of the files, and therefore will need to wait for the release of version 2.0.00 before it can be issued (although draft copies have been sent to those currently getting started with Gadget).

Haddock manual

In charge: Daniel
This uses the up to date example files and is included in the distribution with each new release. The shell script to generate them lives in the .../xmpl/had/doc directory along with the tex files, and refers to ../data, i.e. the correct distribution. Again this will need to be adjusted after the re-write.

In charge: Auðbjörg
Document pvm and paramin, both the installation procedure and simple instructions of use.

Data warehouse

In charge: Vojtech with James
Routines are required to output data from the datawarehouse into the new input file format.

Fleksibest-Bormicon code unification

In charge of defining project: Höskuldur

Class Structure. There are effectively 2 models currently within Gadget, drawing on the historically different approaches the Icelandic and Norwegian models have used. This can, and should, be simplified to ensure that there is only one model, capable of running either an “Icelandic” model or a “Norwegian” one, or possibly some combination of the two. The main difference is in the use (or otherwise) of fish weights for the simulation. This process should allow for the best part of each set of routines to be combined into a simplified code structure.

The cod cannibalism implementation in Fleksibest needs to be generalised by including prey switching as a general mechanism in Gadget.

Here is a list of what I have now on this matter.

lennaturalm and cannibalism

The stockfiles in Bormicon contained age depended natural mortality while in Fleksibest it was length based. In principle the natural mortality should not exist but should be implemented as a predator. This applies especially to length based natural mortality which fits very well with the predators which all eat according to length. The age dependent mortality does not fit with predators but is probably biologically wrong. The suggestion is to remove all natural mortality from the stock files and natural mortality should be a predator called “nature” with some specified suitability function.

Mixing different types of predators

TotalPredators, Poppredator,StockPredator,LinearPredators, MortPredator, MortPopPredator

Here the Mortality predators and Prey have to fit into the framework.

At present the connections in Bormicon are:

```
Fleet <-> LinearPredator
Stock <-> StockPredator
```

LinearPredator inherits from length predator and StockPredator from Poppredator which all inherit from:

```
StockPredator->Eat
```

Calculated how much each stock wants to eat. This consumption is then added to the preys by:

```
Preys(prey)->AddConsumption(area, cons[Iarea][prey][length]);.
```

In mort predator the Eat function would calculate the mortality that the predator wants to yield. The function

```
Preys(prey)->AddConsumption(area, mort[Iarea][prey][length])
```

would add the mortality and Adjust consumption would tell the predator how much he obtained.

After the consumption has been calculated all the member function returning consumption and number consumed should be the same so the same likelihood and print classes can be used. These functions could are:

```
predators[g]->Consumption(area, preys[h]->Name())
&predators[g]->NumberPriortoEating(area,preys[h]->Name());
```

Both used by **PredatorAggregator->Sum** and **PredatorAggreagor->NumberSum**

The class `fleetpreyaggregator` might have to be investigated for problems in this context.

A must either be mortality prey or function as in Bormicon. The same predator could in principle function as mortality predator on one prey and ordinary predator on another prey but that would involve some extra programming and the question is if it is worth doing so.

How mortalities sum makes it very clumsy to have natural mortality in the stockfiles but it was simple (but wrong) in Bormicon if natural mortality was 0.2 the stock was just reduced by exp(-natm*stepsize)

More than one type of fleetfiles can be allowed. At present the format in Bormicon is:

```
header with suitability
year    step    area1   area2   area3
```

And the numbers represent effort, catch depending on the type of fleet. They could also represent mortality.

Other forms of fleet files more suitable for the Norwegian approach could be defined as:

```
header with suitability

yeareffect
stepeffect
areaeffect
```

Prey switching

In Fleksibest the cod cannibalism relates to capelin biomass

```
M = f(cod)/capelin^b
```

This needs to be generalised so the model can implement Prey Switching.

Input formats

Deadline: Have operable version with new input data files along with perl scripts for converting from old to new file formats by June 15.

A rationalisation of the input file format has been decided upon which considerably simplifies the files, making the construction of new models easier. Outputting from the datawarehouse into the new file format will also be easier than before.

Optimisation files

In charge: James

Gadget: Change the `optinfo` file to the following format:

```
Simann
MaxIter 1000 ...  other simann parameters,
                  any not given will use default values.


Hooke
MaxIter 1000 ...  other hooke parameters
```

The order that the optimisation routines are declared will be the order they are run in. If you want to run just one, then only declare one - comment out the rest. This way will be expandable in the future

Paramin: Clean up the constants files.

Parameter input files

In charge: Auðbjörg

Have 2 input files:

A. Inputfile

This contains the parameters and the bounds of the parameters, eg:

```
switch value lowerbound upperbound optimise
fred   0.2   0.01       10         1
378    0.74  -2         50         0
```

For stochastic runs, parameter input files include a line with parameter names.

B. Penaltyfile

This contains the penalty weights to be applied if the bounds are met, eg:

```
switch  power lowerweight upperweight
default 2     1000        10000 ;this is the default penalty weight
378     2     0.1         1000 ;this is the penalty applied to parameter 378
```

“Main” Gadget file

In charge: James

This contains details of the time, area, stock and likelihood files needed to run Gadget. This will just be a list of files, as is the case now. eg

```
timefile time              ;use the file ’time’
areafile area
printing printfile
stockfiles haddock cod     ;use the files ’haddock’ and ’cod’
fleetfiles survey fleet    ;use the files ’survey’ and ’fleet’
likelihood likelihoodfile ;this is a new file that will contain
                           ;details of the likelihood files
```

Model files

In charge: James

time file - no change required
area file - change the temperature data to a column format, eg:

```
year step area temp
1980 1    1    5
1980 1    2    4
1980 2    1    5 ....
```

print file - no change required

Stock files

In charge: James

No change required, except for the initial values. NB: the min/max length at age should be made optional, with the default values being the full min/max range.

Initial values: the meanlengths and standard deviation of the lengths should be made into a separate file, with the following format:

```
age area meanlength standarddeviation
1   1    5          0.3
1   2    4.7        0.5
2   1    8          0.6
```

The file containing the initial weights should be replaced by one containing an initial condition (see Höskuldur).

```
year step area condition
1980 1    1    1.0 ;multiply the condition value by the ref weights
                   ;value to get the initial weight
```

Fleet files

Combine all the fleet data into just 2 files - a “header” file and a “data” file.

Header file
Keep the data from the existing file as far as the suitability function definition. A multiplicative constant is to be used - similar to that already used in the ‘linearfleet’ file - to allow crude scaling of the fleet catch. Each fleet will need an identifier.

Data file
The fleet catch data or expected fleet effort is to be kept in a simple column format. eg:

```
year step area fleet amount
1970 1    1    1     0
1970 1    2    1     0
1970 2    1    1     30000 ...
```

Likelihood files

In charge: James
Create a new likelihood file, containing the following information:

```
Component      ;key word to define a new likelihood component
name           ;Gadget name for the likelihood component
weight         ;weight
datafile       ;filename for the likelihood data - see below
type           ;eg.  SurveyIndices
stock          ;eg.  haddock - from above ...
               ;other things required, depending on type
```

Component:

The data files will be in the following format:

Details of the stock/area/age/length . . . and then the data, in a column format.

So, for the mean length likelihood component, the data file will look like:

```
year step area age number mean variance
1979 1    1    1   2      4.1  0.5
...
```

for the length distribution likelihood component, the data file will look like:

```
year step area lengthgroup number
1979 1    1    5           2
...
```

and for the age-length keys distribution likelihood component, the data file will look like:

```
year step area age lengthgroup number
1979 1    1    1   5           2
....
```

Configuration File

In charge: Morten
A configuration file should be added to Gadget, to avoid cluttering the command line with options. Existing command line options that should be included are: -print1 and -print2. In addition options for debug level.

The input format should be a file containing keyword - value pairs, comments are introduced by ; lasting til end of line. Keyword and value are separated by whitespace.

If both a command line option and config file sets and option, the command line option is used.

Example:

```
 likinterval 20            ; print interval 20 for -o output
 columninterval 2          ; print interval 2 for -co output
 minprintlevel 3           ; print extensive information on optimization
 debuglevel 1              ; print only critical warnings
 appendlogs 1              ; dont overwrite previous logfile
 logfile .gadgetlog        ; print log to .gadgetlog
 likfile lik.out           ; print likelihood to lik.out
 minfile minimization.out  ; print information on progress during
                           ; minimization to minimization.out
 parameterfile params.out  ; print parameters to params.out when finished
```

Implementation:

Extend the `PrintInfo` class, to read options from a file in addition to the command line. Search for configuration file in current directory and home directory, read from current directory if both exists. Use commentstream to read the file. For each non empty line in the file read the first word, if a valid keyword, read the second word according to type (int, string, etc.).

Then parse the command line as usual, overwriting options from the config file if they where also given on the command line.

Miscellaneous

Low priority: A formula syntax for parameters has been suggested, but the need for this has not been made clear.

Auðbjörg: Default parameter values should no longer be given in data- or model files.

Daniel: Suggest how all bounds can be specified in a single location.

Low priority: Suggest how to include log-transform on variables which have to be positive or logit transforms on variables which have to be bounded. To be implemented as additional columns in the parameter initial-value file to paramin or Gadget. Add columns of switches in the same file to indicate which parameters are to be optimised in each of several runs.

Low priority: Suggest how to vary simann and H&J parameters for each iterated run.

James: Some optimisation switches in paramin are hard-coded but all should be input.

Output formats

In charge: James

Output files for modelled stock numbers and biomass are redefined to have a fixed number of tab delimited columns, rather than the current setup. The format of the output files will be similar to that of the input file.

Model output files

There are 3 types of model output files, depending on how the data are aggregated (by length group, by age group and by length and age group). There will be a single output file for each type, with the data in a column format. There will be a header to each file, containing details of the model run (runid, date, time).

An example of the data in column format is:

```
year step area age number meanweight
1980 1    1    2   1000   4.703
1980 1    1    3   4456   7.789 ...
```

Likelihood output files

There are currently 2 types of likelihood output files. It is proposed to introduce a third file, which will only contain the parameters at the end of the optimisation run.

A. Parameters file.
This file will always be created (default name “params.out”) and will contain a header showing details of the optimisation run (date/optimisation used/likelihood score/number of iterations) and the final parameters in the same format as the input files.

B. Column output
This file will be in a similar format to the parameters file, but will also include information on the progress of the optimisation (number of iterations).

C. Likelihood output
This file will be similar to the current file, with the addition of the header information (giving runid/date/time/

...).
The section giving the parameters will be changed to a column format, with the parameters listed in the same order as they are used (which is not necessarily the order that they are given).
The section giving the likelihood components will also be changed to a column format and re-ordered as necessary.
The section giving the parameters values and likelihood at each iteration will be left as it is, with the addition of the iteration number at the beginning, and the change to using 2 tab characters to separate the sections (as opposed to one tab character at the moment).

Log files

In charge: James

The log file should contain names of all input, model, parameter and output files from the run along with their directories and last modification times of each file. The log file should contain a reference to a file containing the screen output from the program run. A program option should include whether the log file should be overwritten or not. The log file should also include start and end times for the run and indicate which parameters were optimised.

The log file includes a run-id line, which will be included in many (most) output files.

James: Header lines including weights etc should be moved from the likelihood output file to the log file.

Miscellaneous

Output from single-processor simulated annealing should include optional information on convergence results.

Likelihood output (monitoring) file: Header lines to be reorganised.

Pre & post-processing

Höskuldur: Think about output format for diagnostics of likelihood components.

Deadline: Have an example post-processing (Perl) program with graphics available by June 15, along with a diagnostics program.

Model diagnostics and Perl libraries

In charge: Daniel
Guidelines and model diagnostics are required to help users investigate model behaviour and structure. This will be based on a combination of statistical output such as the Hessian matrix, sensitivity analyses, and visual and numerical comparisons between modelled and actual catch and survey data. The minimum requirement is to produce summary plots and statistics from the data and the model results. It should be possible to analyse input data, model output or both together. This needs to have sufficient flexibility to allow the user to determine what is examined and should (initially at least) be aimed at meeting the needs of a stock assessment working group. It should also be possible to examine the details of the optimization process via such tools.

Beyond these diagnostic tools we also need a set of routines to add flexibility to the Gadget program by automating much of the required file manipulation and conducting multiple similar runs. For instance sensitivity tests, converting between old and new data formats and converting output formats into human readable tables.

At present a variety of these tools exist, in a variety of languages with sporadic coverage of the important areas and with limited distribution. These should be replaced by a widely distributed ‘tool box’, all of which should be written in a single language. The obvious candidate for this is Perl, which is powerful and flexible and also sufficiently standard to allow advanced users to modify the scripts to suit specialist purposes. These tools need to be carefully documented. Once the tool box exists users should be encouraged to submit suggestions (or code!) for new tools.

Further diagnostics may become apparent as we progress with the work described above.

Statistical programming

In charge: Lorna with Gunnar
The existing likelihood functions need to be enhanced, so that the real full likelihood functions are used. A method of estimating the variance and assigning weights to each component is required. It is probably also necessary to extend the multinomial functions.

In charge: Gunnar
A least one method to verify the existing Hessian computations in BFGS is required and alternatives would be preferable. This task could possibly be undertaken by a summer student.

In charge: Gunnar with Lorna
Methods of dealing with tagging data likelihood components need to be developed.

Lower priority: Bayesian Techniques. An investigation into the use of Bayesian techniques is required, and the Gadget code extended to use Bayesian theory.

Model testing and implementation

Life cycle model

NN: A proposal of how to set up an initial full life cycle model is required. At present no Gadget model fully simulates the complete life cycle from spawning through to maturity. It should currently be possible to construct such a model by adding extra stocks for newly spawned and immature fish and estimating migration/drift and mortality for both groups, in such a way as to match the survey data. This would meet the minimum requirements for the project and would not require major changes to the code. However, we would in effect be ‘fudging’ the numbers to match the survey data, rather than actually modelling the processes. We should therefore investigate doing this in a more sophisticated and realistic manner.

Reducing model ‘tightness’

Daniel: Reduce "tightness" within the model. The degree of flexibility in the model is dependent on the construction of the likelihood components and on the model formulation. We need to be able to increase the flexibility of the model without serious over-parameterization. – "solve" by describing how to address the issue e.g. in the haddock model.

The systems to be modelled are likely to vary in a number of ways, including biology, complexity, size, data availability and reliability, and degree of accuracy required. There is therefore no general ’solution’ to the question of how to build sufficient flexibility into a given model, rather this issue needs to be addressed separately for each model constructed. It is therefore important that this issue be covered in the documentation, to alert users to the importance of this issue, and to provide guidelines on how to adjust their model formulation to fit the different problems encountered. At the moment an attempt at this is being included in the ’introduction to Gadget’ guide, though an expanded version may eventually want to be included elsewhere , perhaps in the users manual.

Minimization test

A useful test of the Gadget optimisation would be to test the model on ‘perfect’ data. The best way to do this is to use the output stock ‘data’ from a model run as likelihood data for a second run. We know that the model is capable of producing this exact result, so we can see how well the different optimizers approach this ideal. This will much easier once we have reformatted the input and output data formats, so this should wait until after the re-write.

Estimability of migration parameters

Gunnar: Describe problem of estimability of migration parameters.

User interfaces

User Interface

Reorganising the input and output formats will do much to improve user friendliness. Having a standard set of tools (scripts) for plotting and diagnostics, along with documentation will also help. Adding a graphical user interface could make it easier to run Gadget for people unfamiliar with Unix.

The User Interface would consist of one or more of these parts:

1. Integrating the scripts for plotting and diagnostics into a graphical user interface.
2. Interface for changing starting values, likelihood component weight, which parameters to optimize, and start an optimization with these settings.
3. Interface for setting up the whole model. Should link with the data warehouse, to retrieve data according to the model setup.

1. and 2. are the easiest parts to implement and also the most useful.

Warnings

NN: During execution of Gadget it could be interesting to know if on bounds. Possibly easier to use Splus rather than put into model. Add the option of having the actual parameter value not the bound value output.

In charge: James
Asserts are currently turned off in some implementations due to incorrect asserts which need to be corrected. Asserts may in many or most cases be replaced by a combination of
`#define debug and if(debuglevel over x) fprintf(...`

In charge: James
Internal limits are sometimes bounding without any notification. Need some form of logging/warning messages.

Done/moved

xdone 4. Done: Check which files need to be included in the Gadget distribution for it to run on different platforms. E.g. the f2c.h file seems to be missing or in the "wrong directory" as looked up by the Gadget Makefile while compiling under Linux.

xdone 5. Overtaken by automake: Makefile. Decide on which compilation flags appropriate for each platform and try to make sure that the compilations on each platform are compatible. F.ex. the Makefile included with Gadget.1.0.00 has the flag NDEBUG defined for the compilation on a Solaris platform but not on a Linux platform. This means that when you compile the code on Solaris you turn off all assert statements but not on Linux.

xdone 6. Look into use of autoconf, and automake to facilitate portability of Gadget.

xdone 7. DONE: Have all file names in Gadget begin with lower case letters. Notably there is a file named Grow.cc and a file named grower.cc.

Minutes

```
Status reports:

Daniel:
    Tested and verified the haddock example

    Multi-stage optimisation:
        Tested H&J
        Paramin with all three algorithms

    Need documentation (user-friendly)

    Current and immediate users:
        MRI
        IMR
        IFREMER
        Lowestoft
        Aberdeen
        Japan
Þórdís:
    Simulated annealing followed by H&J in single-machine version.

Morten:
    Mostly assistance in running under Windows/Cygwin.
    Norway needs graphical user interface for Windows.
    Use existing comment options in parameter files to provide
    descriptive comments on parameters in popup boxes.

James:
    Gone through code & cleaned up in accordance with format in dst2
    report 1.
    Variables declared properly - Gadget 1.03.
    Work done on configure/automake, but needs more work to find pvm
    libraries.
    CVS implementation under consideration.

Höskuldur:
    Shrimp assessments.  Note:  Do not have any alternative assessments and
    hence little basis for evaluation of results.
```

Modifying code to combine Fleksibest and Bormicon models.

Auðbjörg:
Adding alphanumerical switches to input files.
Working on implementing same format for Gadget and paramin input files for parameters.

K.2 Minutes: Sète, June 2002

The third full dst^2 meeting took place in Séte at IFREMER from the 18th to 21st June, 2002.

These minutes contain summaries of some topics along with the main conclusions from the meeting.

Current status of workpackages/deliverables

Barents Sea: The Barents Sea deliverables will be completed in 2002.

Multispecies reference points: Work on multispecies reference points is complete. Published paper by DIFRES.

Celtic Sea: Celtic Sea GLM and GAM modelling, tagging, and stock-recruit relationships. No work is being done on this. Other methods could also be applied to obtain the appropriate biological information.

Are multispecies results significantly different from simple single species models? Some suggestions but nothing currently implemented.

Comparing externally estimated parameters with model results - not available this year. Delayed.

Feedback from herring model will provide information to improve the GADGET model.

Biological modelling in GADGET

Growth

Growth in Icelandic version of GADGET - theory

The current model is based on von Bertalanffy growth using a Taylor approximation. Weight is linked to length growth through $w = al^b$ with a, b fixed. Growth is by weight rather than length to deal directly with consumption and to allow fish which have lost weight to regain weight before growing.

Implementation of growth in GADGET

Growth in GADGET is by age and length rather than length alone as this enables growth to be modelled by cohort. Individual based models also indicate that each fish follows it’s own growth curve.

In GADGET spawning fish can die, bigger fish lose more weight when spawning and there is higher mortality in old big spawning fish.

There is a problem in that fish lengths can exceed L_∞.

Density dependent growth can be included by limiting other food or having a fixed prey biomass. For the lowest trophic levels which do not eat a ‘dummy’ prey could be implemented to introduce density dependence for species without prey.

Tagging data have shown that very large spawning fish have high mortality and such data should be able to provide information on growth. Although lengths from tagging data can be unreliable.

Temperature effects are possible and have been implemented, a problem is that knowledge of temperature is limited.

Discretisation and growth

Discretisation leads to problem with the relationship between length groups and time steps. How well is it possible to model length growth discretely? Is it possible to develop theoretical guidelines on this relationship? Analytical investigation would be very useful. Until now the relationship has only been considered by trial and error with case studies. Discretisation affects the variance, not the mean length. With narrower length groups (closer to a continuous function), the variance decreases but this is partly dependent on the availability of data. The problem is with discretisation, not with the beta binomial distribution.

An analytical approach to time steps and discretisation could start by considering a continuous function.

Work being done by MRI this summer with a Splus model replicating the growth model in GADGET will involve testing a range of L_∞ and k. Although this is not species or area dependent it is still not analytical.

Beta-binomial

Distribution of length around mean affects the resulting mean length. Fishing and growth not linear. There are some problems in length distributions not fitting the beta-binomial.

(Size dependent mortality affects the distribution leading to long tail of small fish - as these fish don't recruit to the fishery. Gudmundur Gudmundsson.)

Norwegian implementation of growth

Length updates are linear rather than von Bertalanffy. Observed weight is available (in a look up table) by length group and time step - although it would also be possible to use a weight-length relationship. Compare with observed length, update length and read in weight. It would be possible to relate annual growth to external factors. Condition factor should be included as it important in Barents Sea. Temperature is also not currently included.

Maturation

Current implementation. cf Bormicon manual

Behaviour and position are dependent on maturity status. Fishing behaviour varies by area but is also targeted more at the mature part of a cohort.

Immature fish are made mature by the model. Possible variables determining when this takes place are: length, age, growth, temperature, time (month) and weight. Where growth refers to the same period as maturation as the growth history is not stored.

The current methods were developed in the initial stages of Bormicon and this is an area which could be developed further as the current method does not model the maturation process. Although the focus is on length and growth, time of year is also critical for maturation. It has been suggested for salmon that the amount of food available at the appropriate time of the year triggers maturations.

In the adaptation for capelin (Icelandic case study), there is a length L_m after which all fish with length longer than L_m in May are mature. A steep sigmoid would produce similar results and also have the advantage of being differentiable.

The steep maturation slope could be due to biological requirements for migration and this length is the minimum required. Summer spawning herring also have a sharp cutoff between length of immature and mature fish and this has been stable over a long time period.

Cod can only mature at certain times of the year.

When fish move between length groups (growth update) mature fish are moved to a mature stock. If maturation is sex independent then sex is allocated when they are moved to the mature stock. If maturation sex dependent then they will have been separate stocks while immature.

Maturity should also be a function of age and feeding level could also be included.

Consumption

cf Bormicon manual

Functional relationships for feeding:

Type I	$C \propto B$ (until ceiling)
Type II	C (per pred) $\propto B$ + constant
	where the constant represents handling and digestion time
	asymptote
Type III	sigmoid
	$C \propto B^n/(B^n + A^n)$
	or $C \propto arctan B$
	where B prey biomass
	this represents prey switching
	at low prey biomass don't consume, possibly change area
	with n = 1 then type III = type II

where C is consumption, B prey biomass.

GADGET currently implements type II. There is no prey switching (type III feeding) where consumption of a prey depends on the density of the prey. Migration as a function of food supply would be one way of dealing with type III feeding with the predator migrating to another area at low prey densities.

As certain prey lengths are more suitable for a predator of a given length the suitability function is dome shaped, although not not necessarily symmetric.

The feeding level is the relationship between the maximum a predator can eat and the level it does eat, where maximum consumption is a function of predator length, predator weight and temperature.

In each time step, calculate level predator wants to eat, then need to calculate the amount of each prey eaten. For each prey need to keep track of amount eaten. It is possible for predators to want to eat more than exists but it is not possible for more than 90% of a prey to be eaten in one time step. The prey biomass is checked and if not enough then predator feeding is scaled down. This indicates that the time step is too long. It is possible to calculate consumption and migration on sub time steps (time steps shorter than data available).

2 other types of predator:

- total consumption is specified in tonnes, which is then split by prey and length groups. Called total predator in model.
- effort specified, consumption of each prey = effort x suitable biomass. This is called linear predator and almost identical to fishing effort on short time steps.

Residual natural mortality should be length, rather than age based. Age based at present as this is the standard in stock assessment. Formulating natural mortality as length based predation is more suitable.

The current implementation needs to be extended to type III. Consumption should depend on (prey biomass)^X and possibly prefer to treat different length groups of prey as separate prey.

There are 2 places where there is a type II, III dichotomy: (i) when calculating total consumption (type II but could use type III) and (ii) when dividing into different prey components

Difficult to get biological interpretation of power eg handling time, easier to parameterise if there is a biological meaning.

Spatial scale may take care of prey searching, if the scale is fine enough and migration a function of prey density or at least related to prey in some way.

Cod and capelin, capelin migrate through the predator area and are consumed. In this situation the predator is limited by max consumption and handling time. High prey density is associated with low cod density and vice versa.

Need suitability functions for consumption of prey.

Consumption could depend on the density of prey as it is only worth feeding at sufficient prey density.

Likelihood functions

There is a problem with comparing consumption in the model with observed stomach content data which do not relate to consumption in a straightforward manner.

Consumption could be related to growth through energy content and conversion efficiency from each prey species.

Including too much realism, particularly when comparing with modelled data, is a problem.

Cannibalism in the Barents Sea cod model

Single species model with cannibalism. Cannibalism is estimated and the resulting natural mortality calculated. Mortality rather than consumption is modelled although the implementation is similar. As it is a single species model some ad hoc measures are employed.

Cannibalism is a function of capelin biomass, with cannibalism low when capelin biomass is high and vice versa. Capelin is included to determine cod cannibalism despite not being included in the model.

Need to go from biomass to stomach content which requires the data to be transformed. A better way to relating stomach content to growth, e.g. using energy content of prey, needs to be found. In MULTISPEC a table via “Conversion efficiency” was partly implemented for marine mammals.

Immature and mature cod are included, both of which may be predators but only immature cod are prey.

In the first quarter, no small cod are in the spawning area and no cannibalism is observed.

As an alternative to the current formulation, capelin could be included as a prey which is only eaten but prey switching would need to be implemented.

The Fleksibest mortality formulation is different from Bormicon as the model is based on mortality.

Fishing/predation

Icelandic fishing

In the Icelandic version, fishing is a predator where target consumption is set equal to catch. Consumption is scaled down, when necessary as with any other predator and an understocking penalty imposed. This does not involve the assumption of accurate knowledge of catches. Catches are not accurate as there is misreporting and accuracy will vary by area. For future simulations fishing is modelled as a linear fleet.

If understocking is common it may indicate that the time steps are too long or that there is a problem with migration.

It could be possible to take reported catches as a lower bound (rather than as a target), particularly where discard data are available.

Catch at age models have two sources of error: age sampling and misreporting of catches.

<u>Predator - Linear fleet</u>

Linear fleet is used where effort is constant eg surveys and in predictions. There is a linear relationship and the slope = fishing mortality.

Type ‘0’: specify catch and remove prey regardless of prey suitable biomass

With linear fleet there is a linear relationship and the slope = fishing mortality.

For fleets the length of the predator does not matter. A dummy number has been used for most fleets, except for gill nets where length represents mesh size.

Selection functions are normally estimated in the simulation part of the model. But they could be defined and include effects such as year, season, area.

Norwegian fishing

The Norwegian implementation uses a statistical catch at length model. Catches are as a desired consumption by a predator.

Where the fleet can fish on either immature fish or mature fish or both. Catches are modelled by timestep, fleet and length group. There can be a difference been observed and modelled catches. The formulation of a suitable likelihood function for this is not trivial.

The standard catch equation assumes fishermen fish at random (as a type I predator). This is used in the Norwegian version but not in the Icelandic version of the model where fishing is type II except when predicting (linear fleet).

Likelihood functions

In the Icelandic version, assume sources of error in age, length and catch at age (so these have likelihood functions) but assume total catch is known and used directly not just as a likelihood.

There are measurement errors in total catch, and even if total catch were accurate, allocation to areas is a source of error. Errors in catch are also biased. A likelihood function could treat landings as a lower bound (with constraints). There is also political demand for such a model.

This is also linked to length based discarding. If catch in tonnes is not believed, then age and length distributions from the catch should not be used directly as discarding is size selective. Mortality, due to fishing but not included in total catch, also involves indirect mortality from fishing, especially for haddock. This has a similar effect on the population as discarding, as do black landings.

A major problem in this area is a lack of knowledge and observed data. Time series of discards could indicate years where there is a problem with catch data and help find year effects.

Absolute catch is currently only a penalty, and is not included as a likelihood function. A likelihood function for catch is necessary.

Future work

Growth

What should be done about fish over L_∞ has not been decided.

It was suggested that a log file could be generated to indicate the number of fish which exceed bounds.

Consumption

Implement type III consumption.

Closure of life cycle

Development of GADGET should include the closure of the life cycle.

For the inclusion of density dependent mortality of 0 group fish another aspect which would need to be considered is larval drift. Assuming a stock-recruitment relationship would be an easier option of including juvenile mortality.

Including larval drift will increase the number of stocks required but the extent to which this would happen depends on how the stocks are to be thought about biologically.

Further work is also required on growth and maturation processes within GADGET.

Migration models

Barents Sea

(deliverables 4.1.1, 4.1.2)

The objective is to develop a multispecies spatially explicit feeding/consumption model based on evolutionary fitness. It will involve the interaction of cod, capelin and juvenile herring.

The period from 1990 to 1993 was chosen as there was a considerable number of capelin, cod and juvenile herring in the Barents Sea at that time.

Model inputs are: temperature (map of Barents Sea at quarterly intervals), currents (from ocean circulation models), capelin (survey data), cod and herring.

An individual based model with cod and capelin currently exists - herring will be similar to capelin. Age structure not yet been implemented. Bioenergetics models have also been developed for the 3 species.

Cod has a given probability of finding a school of capelin, if they find it they feed, but with an upper limit. The model allows the examination of the impact of predator-prey interactions, can be used to investigate interactions between two highly mobile populations and can be used to consider the effect of different homing strategies.

Behaviour is adapted using a genetic algorithm. The model starts with random populations and the best individuals are selected. The best individuals are those with the highest fitness where, fitness = survival x weight.

Either an adapted random walk or artificial neural network (including day length, temperature etc) is used to determine the direction of migration.

When the model is run without homing (ie no spawning migration) prey aggregate along the ice in the Barents Sea in winter, spread out in summer (but not in south). The predators follow prey but there is more spatial overlap at some times of the year than others. With homing (ie prey migrate to Norwegian coast in March and April to spawn), predators are less successful. With two spawning areas prey are less successful and predators more so.

In the adapted random walk, prey have no spatial knowledge and with the adapted neural network (ANN) predators are better suited to the prey.

Predators affect the fitness of prey by eating and scaring them. Cold water is safer for capelin as it is associated with lower cod abundance. Prey can suffer a cost (‘cost of fear’) of avoiding predators with poorer growth and condition.

Future work and linking to GADGET

Output from the model can be compared with observed data (which are available on a coarser scale than from the model) by sampling model predictions on the scale of the observed data. Occurrence of capelin in cod stomachs can also be used for information on the spatial distribution.

Output data could be used for GADGET to model migration by estimating the percentage moving between areas using the ANN formulation of the model. Simulated spatial distributions of cod, capelin and herring along with growth and predation of these species can also be generated. If other data are required for GADGET then this should be discussed now, so that work can be done to look at the possibility of the IBM being able to generate them. The migration process is decided by temperature selection with capelin moving into the 3^oC thermocline and staying there. This is specific to year in terms of start position and temperature field.

Predicting the spawning migration of the Barents Sea capelin would be useful for a GADGET Barents Sea capelin model. It would be possible to simulate movement from autumn to spring by attempting to model migration through temperature maps etc

Barents Sea capelin model II

IMR are also working on a migration model which should be able to produce migration matrices for capelin (using temperature and current fields) in the Barents Sea.

Icelandic capelin

The aim is to generate a spatial distribution based on environmental conditions, ie temperature, food density, oceanic currents, using a mechanistic model. No data have been included.

A grid is specified (triangular) with input and output by square (numbers or mass) which are migration matrices.

There are two versions of the discrete model and a continuous model.

In the continuous model, each particle has preferred direction and speed (while trying to align itself with it's neighbours) along with preferred temperature.

In discrete model I, the school size is fixed and there is alignment between schools, a food density gradient, temperature boundaries, an attracting region (spawning ground) + directional noise and a food dependent energy level. Migration starts when there is sufficient energy. Discrete model II also includes the merging and splitting schools and the probability of a school splitting. Once schools reach the spawning ground they tend to split.

The models are entirely mechanistic, with work now being done on parameterisation. The distribution of capelin is very sensitive to the specified environment.

Icelandic capelin cannot have fixed migration matrices as the direction varies between years, however, the direction can be identified when it starts. The hypothesis on direction of migration around Iceland is that it depends on the position of the 2^oC temperature boundary.

Future work and linking to GADGET

To link this work to GADGET and the project deliverables, it is possible to generate migration matrices from model runs based on observed zooplankton and temperature for every year.

To start with an ‘average year’ could be produced based on average temperature and food density distributions.

It would be possible to compare the different migration models (Barents Sea and Icelandic capelin) by modelling the same area and reparameterising a model.

The SCUI model has not been parameterised for Iceland yet so could also be parameterised for the Barents Sea. There are possibly more data available for the Barents Sea which could be interesting.

Stomach content data

Consumption of juvenile herring in the the Barents Sea

cf report

Aim is the estimation of cod consumption of capelin, herring and relate it to prey density, stock overlap and physical factors. Relative estimates of consumption are better as data are not available for all seasons.

The stomach database is a collaboration between IMR and PINRO (Murmansk) and includes data from 1984 to the present. An average of 8000 cod stomachs are sampled per year. This project concentrates on 1992 - 1997. Samples mainly from bottom trawl surveys and the spatial and temporal distributions are variable. To get total estimates data are extrapolated.

Started with a gastric evacuation model with individual prey, rather than biomass, considered. (Temming and Anderson, 1994 (after Jones 197?)).

This model is not dependent on meal size, but on the amount remaining in the stomach and the equation can be rearranged to get digestion time. Parameter estimates were taken from Temming and Anderson and temperature data from IMR.

Consumption (number of prey consumed per cod per hour) is calculated by prey age and cod length group (mean predation rate). Time and area specific consumption takes into account the proportion feeding.

Calculated consumption of juvenile herring between 1992 and 1997 was compared with natural mortality of juvenile herring during this time period. Estimates of natural mortality were derived from from acoustic surveys as there is no fishing on these fish. It was found that more herring are consumed when the capelin stock is low. There is a belief that acoustic estimates of young herring may be poor.

Estimated consumption can be compared with prey density, stock overlap and physical factors. The consumption estimated will also be compared with the results of the IBM, at least to some extent using GIS. The comparison of consumption and prey density may also provide suitability functions for GADGET.

Consumption of capelin by cod will be calculated by the same method.

Stomach content data in a stochastic multispecies model

Preliminary analyses of North Sea stomach content data on how to model the distribution of relative stomach content observations. The data are from 1991 with cod, haddock, saithe and whiting as predators and herring, sprat, sand eel and Norway pout as the main prey species. The data were used to obtain total species composition by weight for each quarter and roundfish area, these were combined with IBTS as area weights (method from Daan, 1989).

One of the problems with stomach content data is that the data are proportions and correlated. Bormicon assumes independent normal distributions.

As it is not possible to calculate variance directly the data were bootstrapped. Hauls were drawn at random, with replacement, from a stratum obtaining the same number of hauls as available. Species composition in stomachs was then calculated for the whole North Sea. 1000 trials were conducted.

The variance of observations attributable to quarter, predator, prey, length, etc was looked at.

The Dirichlet distribution is a possibility for calculation of the variance as the data are proportions.

Correlation analysis indicates that suitable distributions could be: multivariate normal, or log normal, with variance model and simple or known covariance structure OR the Dirichlet distribution with the same variance model.

A problem is that “other species” constitute a significant proportion of the stomach contents.

How should stomach data likelihood function be implemented using this? It is difficult to estimate correlation parameters as they are correlated with so many other things.

GADGET

New version

A new version of GADGET will be available at the end of July 2002.

The GADGET input format has been changed to simplify the creation of new implementations. The data files are now columnar rather than tabular making than easier to create and to read into other programs such as Splus and Excel making them easier to assess. The files are, however, much longer than before. Length and age data are alpha numeric with another file specifying definitions of the ages and lengths, eg young fish = age1, age2, age3. Missing data are zeros, but if an entire area and time step is missing then no data are required.

The input files consist of the ‘main’ file, models files and data files. Rather than having model definitions, parameters and data in the same file.

The new format also outputs log files with a record of the model run, including the gadget version.

An updated users’ manual will be available later this year but an introductory guide is available now.

Migration matrices have not been changed. They contain the probability of moving from area A to area B in a particular timestep and a matrix format is the most intuitive. Some work needs to be done on the implementation of migration and after this it may be easier to redesign the migration input file.

The datawarehouse is able to output data in the GADGET input format. and Perl scripts are available to convert single area, single species models to the new format.

Paramin will also be updated and a new version will be distributed with an automake file which should make installation easier. You need PVM for paramin to work but for multispecies, multiarea models it’s necessary to run on several computers.

With the next release there will be another directory to do sensitivity analyses. As the new output format is not available there are no plotting options, but these will be developed.

Help in using GADGET and with the new file structure is available at gadgethelp@hafro.is

Variables that can change with time

Instead of a parameter value in the input file there is a filename. The file indicates the years in which the parameter value changes and enables a variety of parameter estimates, which can be calculated from a simple formula (eg regression). This could apply to any parameter in the model.

At present the data used in the formula are in the file with the formula but, to be consistent with the new version of GADGET, these should be in another file. This file could then be referred to on several occasions.

Future

Input files

It would be helpful if input files could be generated automatically as this would give an idea of suitable values. This may not be possible for some parameters but others will be straightforward.

Datawarehouse tables could store parameter values. These could either be values known from other work, or those already used in GADGET runs.

Recording model runs

When stock assessments have been done, the data and version of the model being used need to be stored so that results can be replicated if necessary.

Tagging in GADGET

At present tagged fish can only be included as separate stocks, which have the same parameters as their untagged equivalents. This would be very time consuming and untidy in a multiple area model or if many tagging experiments experiments were conducted simultaneously.

GADGET needs to be developed to make tagging experiments an attribute, although they still need to be separate in order to trace them. Then it would only be required to list the 2 stocks and tagging would just be an extra column in the stock description. As each tagged stock needs to have growth etc applied to it, this will still be slow. It is possible to have tagging active only for short time periods as tagged fish should be caught within 5 years. Nevertheless, tagging will make the model run far more slowly. It will also be a lot of work to implement. If computer time becomes a problem, it would be possible to not include age structure in the tagged stock.

Uncertainty estimates

At present only point estimates are available and confidence intervals would be useful in assessing model runs, interpreting results and estimating uncertainty in input data.

The current strategy is to use a Hessian matrix and some work has been and is being done on this.

Another possible approach would be to re-sample from input data eg length distributions. This should be straightforward with tagging data and some work is being done on this. If raw station data from surveys were being used then stations could be re-sampled, but the datawarehouse data are aggregated. For parametric bootstrapping to be undertaken, information on the underlying distribution should be stored in the datawarehouse.

There is insufficient time remaining for a full Bayesian implementation of GADGET to be developed but there will be a limited Bayesian implementation.

Graphical description

It was suggested that GADGET should be able to print a graphical version of the model. This could be used to check that the model is the one you want and it would also be easier to compare different models. For GADGET you could have the links along with the relevant growth function etc

Spatial and temporal scales

Guidelines on adequate spatial and temporal scales would be useful for those developing new GADGET models, providing information on what can and can't be done on certain scales.

There is a difference between biological scales ie spatial (which are case study and species specific) and scales important for modelling ie length (which is a general issue).

Length scales are important for the running of the model and there is a relationship between length groups and length steps for growth. Small length groups are always as good or better than longer groups, but computational time is affected and insufficient data may be available on short time steps for small length groups.

While work on case studies provides information on specific systems, general conclusions/guides are important for other people starting to use GADGET.

There is no analytical work planned on spatial and temporal scales, although SCUI are interested. If work is done, it will still be specific to the area the individual based model applies to.

The paper on multispecies reference points (DIFRES) cited in the 2001 report includes work on the effect of different levels of aggregation.

GUI

Bergen have requests from their stock assessment scientists for a GUI front end.

Example models

Icelandic cod

To indicate some uses of GADGET, an Icelandic cod example with one stock and two areas. The model was run with a timestep of one month and the fleets were bottom trawl, long line and gill nets (by mesh size).

This model can be used to check the catch in age data used in stock assessment.

When predicting, the model can be run with a range of F to assess the effect on the stock. On short time steps fishing effort is equivalent to fishing mortality. It can be seen that when fishing mortality increases there is a reduction in mean weight at age in the stock as removing the largest individuals from the stock implies removing the fastest growing individuals. This indicates that less fishing would lead to catching larger fish.

The model can also be used to estimate catchability of surveys for different areas, in the case of Iceland, it can be seen that catchability is greater in the north than the south.

Norwegian cod model - ICES

North East Arctic cod has been modelled for several years as a single area, single species GADGET model. In the past this has been criticised by ICES for not having an optimum (Hooke and Jeeves was the only optimiser used) this problem has been solved using by simulated annealing. Model runs have been done +-5% of the optimum as sensitivity analyses and the model has been found to be most sensitive to L_{50} of the fishing fleet and to be more sensitive to mean length of a cohort than the abundance. IMR are presenting a full assessment with prediction to the Arctic fisheries working group. Output routines enable comparison with other assessment methods and at the last assessment results were similar to those from XSA, with Fleksibest slightly more pessimistic.

Work been done on looking at different fishing patterns etc. Closing the life cycle would allow longer term predictions.

Celtic Sea whiting model

GADGET model

The whiting model has one area, ten ages, von Bertalanffy growth and 21 parameters area estimated.

It doesn’t work, but it does compile and run generating negative population numbers.

Initial values were taken from working group reports when possible. There are problems with the choice of boundaries (they were being hit), the choice of observation distributions and the choice of weights for likelihood components.

In general there is a problem with overparameterisation and many parameters are confounded.

Bayesian model

As there were problems with the GADGET model a Bayesian model was built to study problems with initial values, confounded parameters etc.

The model contains five ages, 1 area, is not dynamic and was created using BUGS (MCMC).

Formulating the Bayesian model helped in understanding GADGET.

The aim is to extend this to an equivalent of the single species, single area model to explore posterior distributions. This should help in studying the data and generating suitable initial values for GADGET. The visual representation of the model in BUGS also helps in understanding the structure of the model and a graphical model description would a useful addition to GADGET.

GADGET with simulated data

A simulated dataset was created to help learn and test GADGET.

Deterministic input data so the parameters of the data are known exactly. It was found that the number of years required for the model to stabilise was the same as the number of age groups.

GADGET was then used to estimate weight$_\infty$. When only estimating weight$_\infty$ it was half the value of the input data. When estimating k and weight$_\infty$, weight$_\infty$ was close to the expected value but k was then more than twice as large as in the input data.

This indicates a problem with parameter estimation, which will only be increased when many parameters are estimated. Error estimation of the parameter estimates would be helpful in assessing the reliability of parameters.

The parameter estimation was only done with Hooke and Jeeves (rather than a combination of simulated annealing and Hooke and Jeeves) which stops at max number of iterations. If it doesn’t converge you should start again with the final value. The input data were, however, simulated and deterministic.

The likelihood function was not continuous which may indicate that the bounds were incorrectly defined as if there are problems or hit a bound then the likelihood will not be smooth.

It has been found that if only a few good years of data, at the end of the time period, are available, it is better to use poor data from before that too, in order to get the model started.

As it is possible for several combinations of parameters, with different biological meanings, to produce the same model results confidence intervals for the parameter estimates would be useful in assessing which combinations are more realistic.

It was be useful for new users to have a “standard” data set to test the model and to learn how to use GADGET.

GADGET examples

Examples are required to help people create new models. Currently existing example models are:

GADGET	single species, single area haddock	real data
GADGET	cod-capelin, 3 areas being worked on	real data
GADGET	single species, single area	simulated data
Splus	single species, single area	simulated data

Datawarehouse

The structure has been updated, new features added to increase robustness and some checks have been included.

Some tables have been restructured and some new tables added, including a series of tables for stomach data. Lookup tables have been created, most of which are ready and easily expandable.

There are views into the datawarehouse which convert numeric codes in the datawarehouse (through the lookup tables) into more easily recognizable codes.

Input data is through the ascii file exchange format as decided. The first column of each table should be the code of the table it is to be loaded into. Missing data are underscores and columns are tab separated.

The new upload program is command line but it is possible to extend this to a web interface. When this is uploading data a ‘U’ reports an update if the row was added earlier (ie identical key attributes) and replaced. If there is no corresponding row higher in the hierarchy then it will not load.

The web interface, with the same capabilities as psql, displays data in the window. The number of lines per page can be specified but some knowledge of sql is required. Command line postgresql can also be used. There is also a command line interface (php program) from which you can call a control file to extract data. There are scripts to output data in a form suitable for stock assessment eg age-length distributions, length distributions, mean length at age etc. Data can also be extracted for the new gadget input file structure.

There is a deletion program to delete rows on conditions such as species, year, batch number, time.

CORBA

CORBA could be used to access data from more than one site in a single query over the internet. For this to work all sites must have the same database structure.

The procedure would be to:

1. log-on to common server
2. select data request
3. input parameters (area, time period etc)
4. select database site
5. submit input form

The results would then be displayed in the browser and data could be saved on the client machine.

The HTTP solution would be similar but would only use php at each site rather than CORBA and C++ and it would only be possible to link to one database at a time.

Current status of databases

Iceland

Icelandic data has been loaded into the datawarehouse for four species in the sampling, stomach and tagging tables. Catch data is readily available but not yet loaded but the inclusion of acoustic data will be delayed.

Celtic Sea

Some of the Celtic Sea data has been imported.

There are age length data (collated into ALKs) for the Celtic Sea which cannot be included in the hierarchy. A separate table will be created for age length data which can be used for the official ALKs and also for reference ALKs. These could be used whenever there are insufficient age data in an area but an ALK is required. The user should make a conscious decision to use such data rather than it being a default option without warnings.

A separate stomach table will also be available for data which is lacking hierarchical information.

Remote access

Remote access to data would be useful for Nantes and Lowestoft as at present all the data are stored in Nantes. This could be achieved by accessing the data remotely through a secure shell or by copying the data. This requires permission from local administrators. Currently declared people can access the machine.

Aberdeen will have a machine outside the firewall.

Documentation

Full descriptions are required of the data, including information on the method of collection and they can and cannot be used. There should be links or pointers from the table to descriptions of the data.

Celtic Sea case study

Data being considered are from 1984 to ?? with the aim of predicting 1999 onward. The food web initially proposed was then found to constitute a small part of the biomass and other species (particularly pelagic species) have been added.

So far single species, single area models of cod, whiting, blue whiting have been created.

There are some problems with the data. Although only French and English data are fully available other nations fish in the Celtic Sea and data for other nations has been obtained from ICES. An additional problem is that French and English data are not compatible, particularly stomach data.

As the Celtic Sea is very diverse the prey species in the model do not account for much of the stomach content data.

Coherence of survey data

cf paper on coherence of survey data in the 2001 report.

The UK and French surveys have been compared so see whether the different surveys give similar pictures of the populations. The surveys have been conducted at different times, with the French only in the autumn and the UK in spring and autumn. Other differences are that the UK survey has fixed stations whereas the French is random stratified, the gear is different and the French do 30 minute hauls compared to 60 minutes by the UK.

Abundance estimates from the different surveys were compared. There is a significant season effect (ANOVA) between the UK spring and autumn surveys, with a Wilcox paired rank test finding that 8 of 24 species have a season effect. Comparing UK with French data (season + gear + year) the greatest variability is interannual but spring can be negatively correlated with previous autumn. As recruits enter the fishery in autumn there are more small fish then which contributes to the seasonal differences in selectivity. There are also differences in size selectivity between the survey gear. These conflicting results cause a problem when fitting data.

Both nations show that over the time period the number of non-commercial species has increased and the trends of catching benthic/pelagic species are similar between nations.

Likelihood functions

At present many likelihood functions in GADGET are very simple, often only sums of squares, including those for stomach data. Goodness of fit analyses have indicated a number of problems with these likelihood functions.

The multinomial distribution is often considered suitable (and better than lognormal) for length data but there are problems. Variance and covariance are not multinomial for survey data as adjacent length groups are positively correlated.

Problems with the likelihood functions are the main reason why GADGET and Bormicon has not been used for assessments.

Multivariate normal is possibly a suitable choice for the likelihood function. The Dirichlet distribution in a Bayesian context has been proposed but it is far too complex to be implemented in GADGET and problems with likelihood functions for stomach data will be considerably greater than for length data.

There has been a paper published about modelling benthic grab samples using a likelihood function. A hierarchical model was used with additive Normal and multivariate Normal distributions. This could possibly be a way of including an overdispersed multivariate normal.

Billheimer, Dean (1997) Journal of Environmental Statistics also PhD thesis 1995

Two issues in weighting likelihood functions are (i) obtaining the correct weights and (ii) identifying when the model is seriously wrong. If the model is incorrect, then different weights will give different results. This is a problem in stock assessment, particularly with oldest and youngest ages.

By comparing results with different weight values, it is possible to identify areas with problems. Inconsistent results are often due to insufficient flexibility in the model. To a certain extent GADGET can deal with this eg suitability functions changing slowly over time.

Future

Future: summary

In terms of work on the functioning of GADGET, in past 2 years, a working parallel version of GADGET has been developed, a lot of work on diagnostics has resulted in changes to the code and made the model more stable. Other developments include the new input format which will make use of the model easier.

It has been decided to add formulae for variables, tagging will be implemented along with bootstrapping for the tagging. Bootstrapping for other variables may be implemented directly from the datawarehouse.

Planned work not mentioned so far includes: closure of the life cycle, code unification between Icelandic and Norwegian models, work on goodness of fit and new likelihood functions.

UiB model output needs to be linked with GADGET and a comparison of consumption made between modelled consumption from data and IBM estimates of consumption.

Results from the SCUI and UiB migration models need to be compared, there will be a meeting in Bergen to discuss this. The SCUI model should also be compared with GADGET.

UiB will try to estimate migration matrices for the Barents Sea.

The effect of parallelisation needs to be evaluated and this will probably be completed next year.

Other areas of future work are included earlier in the document.

Next meeting

Bergen August 2003 (week starting 18th).

K.3 DST^2 meeting in Bergen on fish migrations

17 and 18 November, 2002

Location

Department of Fisheries and Marine Biology, University of Bergen, The High Technology Centre, 3. floor.

Participants

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

Topics

The meeting will consist of presentations and informal discussion. A projector will be available for presentations. We will have a hub available to connect laptops together, so please bring network cables. This will allow easy exchange of data and code. Lunch will be provided. Discussions and presentations will cover, but not necessarily be limited to (randomised order):

1 What is known about the capelin in the Barents Sea (especially as regards migration patterns); what environmental factors are important (temperature, food density, salinity, light, oceanic currents, location of the polar front, physiological status etc)? What is the “preferred” temperature range?
2 Methods and philosophies for modelling fish migrations.
3 Process specification and data requirements.
4 Case study: The Barents Sea capelin during 1990-1996.
5 Implementation of migration in fisheries assessment models (Bormicon/Fleksibest).
6 Deliverables in DST^2.
7 What data is available for parameter estimation?
8 Model validation (how?)
9 What are the similarities and differences between the different models?
10 Importance of physiological structuring: One structured model or different models for different ages (stages)
11 Connection to Gadget. Should the migration models be used as an external tool to Gadget (e.g. providing migration matrices) or as an integral part of Gadget?

Agenda

Sunday

1000 Start

-Theme of the day: Modelling fish movements/migrations.
-Interactions between capelin and other species in the Barents Sea:

Bjarte
Geir Odd

-The case study – data set details.

Geir Odd

-Presentations of the various migration modelling approaches along with results achieved:

Kjartan
Thomas Philip
Sven
Eva Hlin
Stefan
Sigurd
Geir

1300 -1400 Lunch

-*Presentations continued.*

-Discussion and exchange of code and data.

1700 End

Monday

0900 Start

-***Theme of the day: Implementation of migration in fisheries assessment.***

-Migration matrices.

-Implementation of migration in Bormicon/Fleksibest.

-Deliverables in DST^2.

1200-1300 Lunch

-*Continuing discussion.*

1600 Summing up

1930 Dinner?

Comparison of different spatial modelling approaches for generating migration matrices in GADGET

At the meeting in Bergen on 17 and 18 November on modelling the spatial dynamics of fish, it was decided to use the different models to generate output on a form suitable for GADGET. This will allow production of migration matrices as well as a formal comparison of different modelling approaches. Below the format of this comparison is specified.

1 Rationale behind temperature as the sole forcing factor

It was decided that the best approach to the comparison was to use temperature as the main explanatory variable. This was founded partly based on temperature as being important for the distribution of capelin (see Dommasnes & Røttingen 1985; Gjøsæter 1998; Gjøsæter & al. 1998), and partly on the good availability of temperature data. As seen in the figure below there is a positive relationship, although weak, between temperature at the Kola section and the northerly extent of the capelin distribution during fall. In addition to being important for the feeding migration in summer, temperature is important for the spawning migration (Ozhigin & Luka 1985). In addition to temperature, the stock size and composition used to initiate the simulations will be important since this determines the spreading of the population as well as the proportion of the stock dying after spawning.

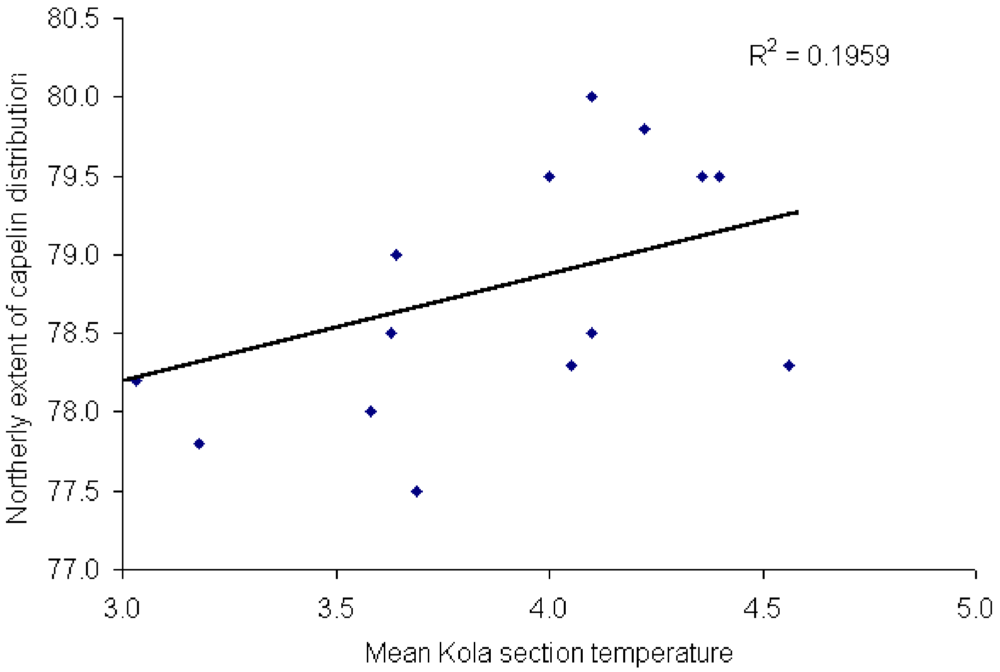


Figure 1: The relationship between the mean annual temperature at the Kola section and the northerly extent of the capelin distribution in September/October.

2 The test format

It was decided to perform the comparison for three different scenarios: one cold, one warm and one intermediate year, represented by 1981, 1991, and 1993 respectively. 1991 and 1993 is within the case study, but data for 1981 needs to be made available. Bjarte Bogstad will be responsible for getting the capelin, and temperature data while Geir Odd Johansen will interpolate the temperature data as for the case study period. The comparison will be made by initiating the capelin distribution from survey data the year ahead of the target years listed above. The starting date is October 1, which corresponds approximately to the timing of the capelin survey. Then the simulation should be made over the one year. Distributions valid for the 15. of each month should be used in the comparisons. Statistical comparisons of the simulated distributions will be made using spatial correlation. In addition to the model comparisons, the simulated distributions will be compared with survey observations of capelin in February, April (assumed spawning area, qualitative data), and September. This will be our reality check! In addition to distributions, migration matrices should be generated at bi-monthly intervals starting with January-February, March.-April, etc. This will allow us to more easily relate the results to GADGET.

3 Report

A report will be written to communicate the results. This will be used in the final project report of DST2. In addition a peer-review paper will be attempted. In order to facilitate the production of these papers, each model should be accompanied by a description and presentation of the results at the meeting summer 2003 in Bergen.

4 Meeting

The migration group will get together at the DST2 meeting in Bergen next August. Preferably the group should meet a day before the main meeting to present the results and write up the report/paper.

References

Dommasnes A, Røttingen I. 1985. Acoustic stock measurements of teh Barents Sea capelin 1972-1984 a review. In: Gjøsæter H, editor. *The proceedings of the Soviet-Norwegian symposium on the Barents Sea capelin*. Bergen: The Institute of Marine Research. p 45-108.

Gjøsæter H. 1998. The population biology and exploitation of capelin (*Mallotus villosus*) in the Barents Sea. *Sarsia* 83:453-496.

Gjøsæter H, Dommasnes A, Rottingen B. 1998. The Barents Sea capelin stock 1972-1997. A synthesis of results from acoustic surveys. *Sarsia* 83:497-510.

Ozhigin VK, Luka GI. Some peculiarities of capelin migrations depending on the thermal conditions in the Barents Sea. In: Gjøsæter H, editor; 1985; Bergen. Institute of Marine Research, Bergen, Norway. p 135-148.

Hafrannsóknastofnun. Fjölrit
Marine Research Institute. Reports

Þessi listi er einnig á Netinu
(This list is also on the Internet)

http://www.hafro.is/Timarit/fjolr.html

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 Norðvestur-land 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.