Statistical aspects of fish stock assessment

Berg, Casper Willestofte; Madsen, Henrik; Thygesen, Uffe Høgsbro; Nielsen, Anders

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

Statistical aspects of fish stock assessment

Casper Willestofte Berg

PHD-2013-302

DTU Compute
Department of Applied Mathematics and Computer Science
Technical University of Denmark
Richard Petersens Plads
Building 324
2800 Lyngby, Denmark
Tel. +4525 3031
www.compute.dtu.dk

Summary (English)

Fish stock assessments are conducted for two main purposes: 1) To estimate past and present fish abundances and their commercial exploitation rates. 2) To predict the consequences of different management strategies in order to ensure a sustainable fishery in the future.

This thesis concerns statistical aspects of fish stocks assessment, which includes topics such as time series analysis, generalized additive models (GAMs), and non-linear state-space/mixed models capable of handling missing data and a high number of latent states and parameters. The aim is to improve the existing methods for stock assessment by application of state-of-the-art statistical methodology. The main contributions are presented in the form of six research papers.

The major part of the thesis deals with age-structured assessment models, which is the most common approach. Conversion from length to age distributions in the catches is a necessary step in age-based stock assessment models. For this purpose, GAMs and continuation ratio logits are combined to model the probability of age as a smooth function of length and spatial coordinates, which constitutes an improvement over traditional methods based on area-stratification. GAMs and delta-distributions are applied for the calculation of indices of abundance from trawl survey data, and different error structures for these are investigated.

Two extensions to the state-space approach to age-structured stock assessment modelling are presented. The first extension introduces multivariate error distributions on survey catch-at-age data. The second extension is an integrated

assessment model for overlapping sub-stocks subject to joint exploitation in the area of overlap. Estimation and inference is carried out using maximum likelihood.

Finally, a biomass dynamic model based on stochastic differential equations is presented. This work extends the classical approaches to biomass modelling by incorporating observation errors on the catches, and allowing for missing and non-equidistant samples in time.

Summary (Danish)

Bestandsvurderinger af fisk har to overordnede formål: 1) At estimere den historiske udvikling i fiskebestandene og graden af kommerciel udnyttelse af disse. 2) At forudsige konsekvenserne af forskellige forvaltningsstrategier med det formål at sikre et fremtidigt bæredygtigt fiskeri.

Denne afhandling beskæftiger sig med statistiske aspekter ved bestandsvurdering af fisk, hvilket inkluderer emner såsom tidsrækkeanalyse, generaliserede additive modeller (GAM'er) og ikke-lineære tilstands-/mixede modeller til håndtering af manglende datapunkter og et højt antal uobserverede tilstandsvariable og parametre. Formålet er at forbedre eksisterende metoder til bestandsvurdering ved at anvende avancerede statistiske metoder. De primære bidrag er præsenteret i form af seks videnskabelige artikler.

En overvejende del af denne afhandling omhandler alders-strukturerede modeller til bestandsvurdering, hvilket er den den mest almindelige tilgangsvinkel. Konvertering fra længde- til aldersfordelinger i fangsterne er et nødvendigt skridt i alders-strukturerede modeller. Til det formål kombineres GAM'er og fortsættelses-logitter for at modellere aldersfordelingen som en glat funktion af længde og geografisk position, hvilket udgør en forbedring i forhold til traditionelle metoder baseret på område-stratificering. GAM'er og delta-fordelinger bliver anvendt til beregning af relative bestandsestimater ud fra data fra videnskabelige bundtrawlstogter, og forskellige fejl-strukturer for disse undersøges.

To udvidelser til state-space tilgangen til aldersstrukturerede bestandsvurderingsmodeller bliver præsenteret. Den første udvidelse introducerer multivariate fejl-fordelinger for fangster per aldersgruppe fra videnskabelige togter. Den an-

den udvidelse er en integreret bestandsvurderingsmodel for overlappende underbestande, der udnyttes samlet kommercielt i området for overlap. Estimation of inferens baseres på princippet om maximum likelihood.

Slutteligt præsenteres en dynamisk biomasse-model baseret på stokastiske differentialligniner. Dette arbejde er en udvidelse a klassiske metoder til biomassemodellering idet det tillader observations-fejl på fangsterne samt manglende og ikke-ækvidistante målinger i tid.

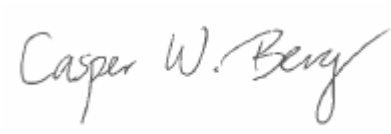
Preface

This thesis was prepared at the National Institute of Aquatic Resources (DTU Aqua) and the department of Informatics and Mathematical Modelling (DTU Compute, formerly DTU Informatics) at the Technical University of Denmark in partial fulfillment of the requirements for acquiring the Ph.D. degree in engineering. The work was carried out between November 2009 and April 2013 and was supported by a DTU Scholarship.

The thesis deals with statistical aspects of fish stock assessments, with special focus on age-based models and the state-space approach.

The thesis consists of a summary report and six scientific research papers created during the Ph.D.-period. Papers I to III are published in international peer-reviewed journals, IV is under review in an international journal, and the last two represent work in (far) progress.

Lyngby, 30-April-2013



Casper Willestofte Berg

Acknowledgements

I would like to thank my supervisors Henrik Madsen, Uffe H. Thygesen, and not the least Anders Nielsen for excellent supervision and continuous encouragement during this project.

I would also like to thank my "other supervisors" and fellow south-wing occupants of Charlottenlund Castle: Kasper Kristensen, Peter Lewy, and Anna Rindorf for their massive help, collaboration and sharing of ideas and insights about stock assessment.

I was so fortunate to visit the University of Hawaii at the Joint Institute for Marine and Atmospheric Research during this project. In this respect, I would like to thank Pierre Kleiber, John Sibert, Johnoel Ancheta, Dodie Lau, and Kevin Weng for making this stay so pleasant and for their insights in how to assess (and grill) tuna.

Finally I would like to thank Henrik Mosegaard, Karin Hüssy, Lotte Worsøe Clausen, Martin Wæver Pedersen, Rune Juhl, Jan Beyer, Niels Gerner Andersen, and Axel Temming (in random order) for their various contributions to this project.

List of Publications

This thesis is based on the following original research papers

I **Casper W. Berg** and Axel Temming. Estimation of feeding patterns for piscivorous fish using individual prey data from stomach contents. *Canadian Journal of Fisheries and Aquatic Sciences*, 68(5):834–841, 2011

II M.W. Pedersen, **C.W. Berg**, U.H. Thygesen, A. Nielsen, and H. Madsen. Estimation methods for nonlinear state-space models in ecology. *Ecological Modelling*, 222(8):1394–1400, 2011

III **Casper W. Berg** and Kasper Kristensen. Spatial age-length key modelling using continuation ratio logits. *Fisheries Research*, 129-130:119–126, 2012

IV **Casper W. Berg**, Anders Nielsen, and Kasper Kristensen. Evaluation of alternative age-based methods for estimating relative abundance from survey data in relation to assessment models. *Submitted to Fisheries Research*

V **Casper W. Berg** and Anders Nielsen. Joint assessment modelling of two herring stocks subject to mixed fisheries.

VI **Casper W. Berg** and Henrik Madsen. Continuous Time Stochastic Modelling of Biomass Dynamics.

The following list of publications present work which is either peripheral to the subject in this thesis or where the author of this thesis did not have a leading role in producing the manuscript or a combination of the two. The content of these papers will not be discussed further.

1. Claus Stenberg, Mikael van Deurs, Josianne Støttrup, Henrik Mosegaard, Thomas Grome, Grete E. Dinesen, Asbjørn Christensen, Henrik Jensen, Maria Kaspersen, **Casper W. Berg**, Simon B. Leonhard, Henrik Skov, John Pedersen, Christian B. Hvidt, Maks Klaustrup, Simon B. Leonhard, Claus Stenberg, and Josianne Støttrup. Effect of the Horns Rev 1 Offshore Wind Farm on Fish Communities. Follow-up Seven Years after Construction. *DTU Aqua Report*. DTU Aqua, 2011

2. Lotte Worsøe Clausen, Henrik Mosegaard, **Casper W. Berg**, and Clara Ulrich. The Gordian knot: managing herring (Clupea harengus) bridging across populations, fishery units, management areas, and politics, volume N:06. *International Council for the Exploration of the Sea (ICES)*, 2012

3. Clara Ulrich, Søren Lorentzen Post, Lotte Worsøe Clausen, **Casper W. Berg**, Mikael van Deurs, Henrik Mosegaard, and Mark Payne. Modelling the mix- ing of herring stocks between the Baltic and the North Sea from otolith data. *International Council for the Exploration of the Sea (ICES)*, 2012

4. Benjamin M. Bolker, Beth Gardner, Mark Maunder, **Casper W. Berg**, Mollie Brooks, Liza Comita, Elizabeth Crone, Sarah Cubaynes, Trevor Davies, Perry de Valpine, Jessica Ford, Olivier Gimenez, Marc Kery, Eun Jung Kim, Cleridy Lennert-Cody, Arni Magnusson, Steve Martell, John Nash, Anders Nielsen, Jim Regetz, Hans Skaug, and Elise Zipkin. Strategies for fitting nonlinear ecological models in R, AD Model Builder, and BUGS. *Methods in Ecology and Evolution (Published online April 22 2013)*, 2013

Contents

Chapter 1

Introduction

1.1 Background

Fish stock assessment models are statistically quite challenging because a) The main quantities of interest (population abundance and fishery pressure) are not directly observed. b) The ecological processes governing the system are often non-linear and far from fully understood. c) Many different types of high-dimensional data sources often needs to be combined, typically non-Gaussian and highly skewed, covering different time-spans and geographical regions, and containing intricate correlations, which may lead to substantially different conclusions about the stock depending on the model formulation and how the different data sources are weighted.

One of the starting points for this project was to develop a model for two sub-populations of Atlantic herring. Because there is an overlap in the spatial distribution of the two species, they are caught together by commercial fishing vessels in the area of overlap. However, the area of overlap is currently managed, but not analyzed, separately, causing problems for the managers. The model should therefore address this issue. The available data sources at that time for such a model were however not sufficient – survey indices of abundance for the area of overlap were missing, although the area had been surveyed. Hence, new survey indices had to be calculated, which led to rather complete re-examination

of the statistical methodology that had been used so far for this purpose. A major part of this thesis is therefore concerned with topics related to survey index calculations and their role in stock assessments, although other topics are touched upon as well.

1.2 Objectives

The thesis concerns statistical aspects of fish stocks assessment, which includes topics such as time series analysis, generalized additive models (GAMs), and non-linear state-space/mixed models with missing data and a high number of latent states and parameters. The overall purpose of this thesis is to develop improved methods for fish stock assessment.

More specifically, the objectives of this study are

- To develop an improved method for analyzing stomach data with individual identified prey items using mixed models to detect temporal patterns in feeding behavior. Stomach data is an important source for knowledge about predator/prey interactions and consumption rates, which can for instance be utilized in multi-species fisheries models, and patterns in feeding behavior may reveal insights about predator-prey interactions and aid sampling design.

- To investigate methods for estimation in non-linear state-space models for stock assessment.

- To improve the currently used methodology for calculating indices of abundance by age from trawl survey data. To achieve this, two main problems must be dealt with

    1. Conversion from numbers-at-length to numbers-at-age using subsamples of age and length taking geographical variations into account.
    2. Integrating numbers-at-age from individual hauls to a index of abundance for the whole population. This includes dealing with highly skewed data, unbalanced designs, and correction for external factors that may affect the catch rate other than abundance.

- To combine the developed methodology for calculating survey indices with the state-space approach to age-based stock assessment models, and to evaluate its impact on the precision of such assessments.

- To develop an integrated assessment model for North Sea and Western Baltic herring. Such a model will constitute an improvement over the currently used methodology, which involves splitting of data based on sampled proportions and carrying out separate assessments on each stock. Since this a problem that is not unique to the examined herring stocks, this model may generally applicable to other cases of mixing sub-stocks of the same species.
- To account for both process and observation error and to handle missing observations as well as varying sample times when estimating in biomass production models. Many current applications of biomass models ignore process error as well as errors on the observed commercial catches, which may lead to biased estimates of key management quantities such as the maximum sustainable yield (MSY).

1.3 Outline

Chapter 2 gives a brief overview of the data sources used in this thesis and of stock assessment in general. Chapter 3 provides a review of some of the past and current methodology used for fish stock assessment and describes how this is applied and extended in Papers I-VI. Some of the contents in this chapter is therefore also presented in the papers. Chapter 4 summarizes the main results and findings documented in Papers I-VI. Appendices A and B contain descriptions of two publicly available software packages related to stock assessment, which were applied in, and whose development benefitted from, this thesis.

Chapter 2

Data

2.1 Commercial Catch and Effort Data

Fish stock assessments are primarily conducted for commercially exploited populations, and one of their major purposes is to quantify the implications of different management strategies. A fish stock may be exploited by one or more fleets that differ in their type of fishing gear and nationality. The mass of the total catches is measured for each fleet, and subsamples of weight, length, and age are used to obtain estimates of the total number of fish caught in each length- and/or age-group. Some error due to the catch sampling must therefore be expected, mostly with respect to age samples, as these are much sparser collected due to the extra costs from analyzing for age. In some cases there are so-called effort data available for the commercial fleets. Fishing effort may be recorded as the number of days at sea, fishing hours or something similar.

For many EU stocks effort data are not used in the assessments, presumably due to poor or insufficient data. However, growing time-series of satellite based vessel monitoring systems (VMS), that automatically collect the position of all major fishing vessels, are beginning to allow for the use of high-resolution effort catch-effort data, especially when combined with logbook data on the catches (Gerritsen and Lordan, 2011). When effort data is not available, the total catches are sometimes known as "residual catches".

2.2 Surveys

Kimura and Somerton (2006) states: "Survey time series are the essential anchor that makes modern fish stock assessment modeling possible". This is because commercial catch data alone do not contain sufficient information to estimate the stock size and fishing mortality. A variety of sampling methods for fish exist such as capture with active fishing gears (e.g. trawls), passive gears such as baited hooks, plankton nets for capturing eggs and fish larvae, hydro-acoustics, and even video photography. The goal is common for all these methods: To produce indices of abundance that may be used in combination with a population model for the purpose of managing the fishery.

2.3 Trawl Surveys

Bottom trawl surveys is probably the most analyzed and widely used method to produce survey indices. Fish are sampled by towing a large conical net across the sea bottom, usually for a standard distance or duration of time. The International Council for Exploration of the Sea coordinates (among others) the International Bottom Trawl Survey (IBTS) in the North Sea which constitute a long and important time-series of survey data for this area. The data, which are very detailed, are stored in the DATRAS database (ICES, 2012a).

2.4 Acoustic Surveys

Hydro-acoustic surveys measure the acoustic reflection energy from fish schools using an echo sounder, which can be converted to biomass or numbers of individuals from calibration experiments (Misund, 1997). Acoustic surveys have the advantage, that echo integration data is collected continuously during the survey such that a much larger volume of water can be covered than with trawling. The downsides are among others that reflection energy may stem from other sources than schools of fish creating false positives in the signal, and that correct allocation of the integrated echo signal to species and size groups is difficult.

2.5 Stomach Contents

Since actual feeding events for fishes are rarely observed directly, analyses of the stomach contents are important for understanding and quantifying feeding preferences and consumption rates. These enable predation mortalities inflicted by fish on each other to be included in multi-species models (e.g. Gislason and Helgason (1985), Lewy and Vinther (2004)) Accurate models of gastric evacuation have been developed using aquarium experiments (Andersen and Beyer, 2005), which allows feeding rates to be determined from stomach samples from the wild. A review of the role of stomach data and gastric evacuation experiments is given in Bromley (1994).

2.6 Tagging Data

Classical tags involves releasing a large number of caught and marked individuals back into the population. The proportion of tagged individuals that are found in new catches can be used for population estimates and survival rates (Nielsen, 2004), and their geographical displacement for inferring movement patterns. Electronic archival tags can make continuous measurements of the environment around the fish such as depth and temperature, which allow for estimation of whole movement trajectories (a discipline known as geolocation, Nielsen and Sibert (2007)). Spatially resolved stock assessment models that incorporate movement of fish exist (e.g. Quinn et al. (1990),Hampton and Fournier (2001)), although they are still underused due to their complexity and high data demands. A review of methods for incorporating movement into stock assessment models is given by Goethel et al. (2011).

2.7 Age Data

Otoliths (or "earstones") can be used for determining the age of many species of fish. In analogy with trees, alternating periods of slow and fast growth in the winter and summertime lead to alternating dark and translucent zones in the otolith, which makes it possible to determine its age (see Figure 2.1). Time series of paired observations of age and length are key in age-structured stock assessment models.

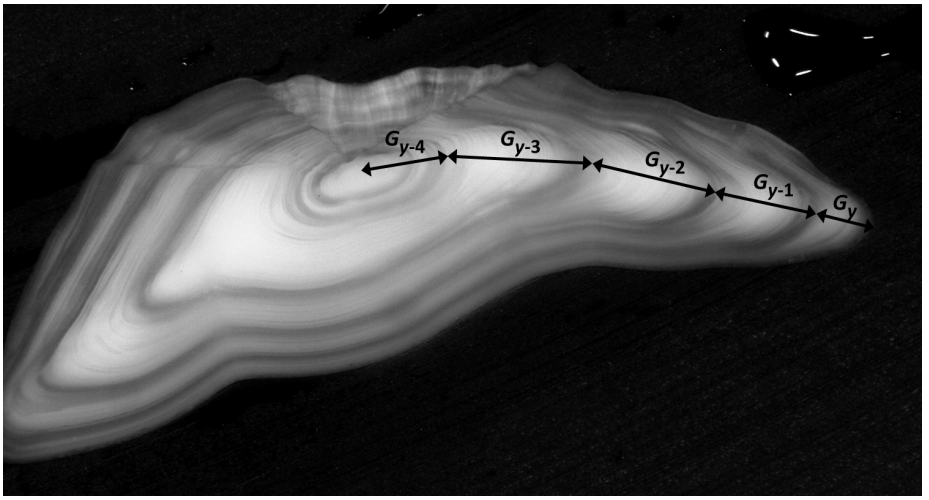


Figure 2.1: Otolith from an Atlantic Cod with 4 winter rings (Picture and graphics: Karin Hüssy, DTU Aqua).

2.8 Summary of Data Sources Used

Paper I combines models of gastric evacuation with mixed effect models to estimate feeding patterns from stomach data. Paper II is based on simulated data only. Paper III introduces the DATRAS-package for manipulating and analyzing trawl survey data, and combines continuation ratio logits with generalized additive models (GAMs) for analyzing age and length data containing spatial variations. Paper IV combines the methodology from paper III with Delta-GAM models for estimating indices of abundance from trawl survey data. Commercial catch data is also utilized in paper IV to evaluate the precision of different procedures for estimating the survey indices by carrying out full stock assessments. Paper V uses multiple sources of survey data (acoustic as well as trawl surveys) for the assessment of two herring stocks. Paper IV deals with estimation in biomass models, which require both commercial and survey catch data but without the use of age-samples.

Chapter 3

Methods

There are four main factors contributing to the dynamics of an exploited fish population (Beverton et al., 1957):

- Recruitment
- Growth
- Fishing mortality
- Natural mortality

Recruitment is defined as the number individuals entering the exploited part of the population. Eggs, larvae and juveniles (the pre-recruit stages) often occupy special nursery grounds away from the adult population, and they are too small to be retained by normal fishing gear, which explains why this definition is useful. Fishing mortality and natural mortality is the rate of deaths due to fishing and other (natural) causes respectively. A fifth factor, movement and migration, may also be considered, but data demands are high for models that incorporate this element, so this is often ignored.

In the simplest models the population is described only by its total mass (or numbers/density), at the expense that recruitment, growth, and natural mortality cannot be separated. A more precise description of the population is

therefore usually adapted by describing the population by its size or age distribution. This chapter contains a description of different models for fish stock assessment, with the main focus on the models that are applied in papers II-VI and statistical methods for estimation in such models.

3.1 Models Based on Total Weight

While age and length based models provide a detailed description of a fish population with respect to its size-composition, these require many costly samples of the catches, which are often not available for less commercially interesting stocks, in third-world countries, or simply not available as far back in time as records on total catch in weight. Therefore, simpler models based only on the total weight of catches as well as the population remain useful for fisheries management (Punt, 2003).

These models are known as *biomass models* or *(surplus) production models*, and a general discrete-time description of these models is given by

$$B_{y+1} = (B_y + g(B_y) - C_y)\,\mathrm{e}^{\epsilon_y} \tag{3.1}$$

$$I_y = qB_y\mathrm{e}^{\eta_y} \tag{3.2}$$

where B is the biomass, C_y is the catch during year y, I_y is a relative biomass index, q is a constant proportionality factor (catchability), r is the intrinsic rate of growth, and K is the carrying capacity (unexploited equilibrium biomass). The surplus production function $g()$ determines how the density dependence reduces the productivity at high population sizes. Common choices of surplus production function include the Schaefer model (Schaefer, 1954)

$$g(B) = rB\left(1 - \frac{B}{K}\right), \tag{3.3}$$

the Fox model (Fox, 1970)

$$g(B) = rB\left(1 - \frac{\log B}{\log K}\right) \tag{3.4}$$

and the Pella-Tomlinson model (Pella and Tomlinson, 1969)

$$g(B) = \frac{r}{p}B\left(1 - \left(\frac{B}{K}\right)^p\right) \tag{3.5}$$

, which is a generalized form that includes the Schaefer ($p = 1$) and the Fox ($p \to 0$) models. The power parameter p is often difficult to estimate from data. The

theta-logistic model examined in Paper II is a re-parameterization of the Pella-Tomlinson model (the first term $\frac{r}{p}$ is replaced with r). Production models have formed the basis of the concept of *Maximum Sustainable Yield* (MSY), which is the largest catch that can continuously removed from the population while keeping it in an equilibrium state ($g(B) = C$). The fishing mortality and biomass at this equilibrium point (F_{MSY} and B_{MSY}) can be derived analytically for all of the production models above, which is why these models have popularized the concept of MSY. Although the concept of MSY has been criticized by the scientific community for being overly optimistic, and ignoring several aspects such as the cost of fishing and multi-species effects (e.g. Larkin (1977)), it was written into the 1982 United Nations Convention for the Law of the Sea (United Nations, 1982), and is still used as a key concept for fisheries management in both the EU and the US.

3.1.1 An SDE formulation

In state-space terminology, (3.1) is a process (or system) equation while (3.2) is an observation equation, and hence $\epsilon_y \sim N(0, \sigma_\epsilon^2)$ describes process error due to imperfections of the model, and $\eta_y \sim N(0, \sigma_\eta^2)$ describes measurement error. Historically, one or the other type of error have usually been ignored (Polacheck et al., 1993) although this may lead to biased parameter estimates (and hence biased estimates of MSY) (Chen and Andrew, 1998). Furthermore, the catches C are also observations that may be more or less subject to observation noise as well. Paper VI introduces a formulation of the Schaefer model based on a set of stochastic differential equations (SDEs) rather than the typical discrete time formulation used above, which allows for estimation of both process and observation error (on Is as well as Cs), varying sample times, and missing observations. The model is formulated in the log-domain, i.e. using $Z_t = \log(B_t)$. This has several advantages. One is that it avoids negative biomasses that may arise otherwise (for instance if an observed catch exceeds the current biomass estimate). Another advantage is that the multiplicative error structure in (3.1) and (3.2) becomes additive in the log-domain. In the continuous time formulation, C must be replaced with the differential equivalent $\frac{dC_t}{dt} = F_t B_t$, where F_t is the fishing mortality. The SDE version of the process equations are:

$$dZ_t = \left(r - \frac{r}{K} e^{Z_t} - F_t - \frac{1}{2}\sigma_B^2 \right) dt + \sigma_B dW_t \tag{3.6}$$

$$d\log(F_t) = \sigma_F dV_t \tag{3.7}$$

where W_t and V_t are independent standardized Brownian motions.

Estimation in SDEs is complicated, but advanced statistical software packages have been developed to this end, and Paper VI applies the methodology in Kristensen et al. (2004) which is implemented in the `CTSM-R` package (Juhl et al., 2013).

Since the catch observations consist of accumulated catches over some period (typically a year), we the observation equation should ideally be based on the integral $\Delta C_y = \int_{y-1}^{y} F_t B_t dt = C_y - C_{y-1}$. However, the `CTSM-R` package requires observations that depends on the state vector at one point in time only. Hence, the following approximation is used

$$\log \Delta C_y \approx \log F_{y-\frac{1}{2}} + \log B_{y-\frac{1}{2}} \tag{3.8}$$

That is, we evaluate the F and B processes half-way through the year, and approximate the integral of the product of the two series simply by the product evaluated in the middle of the time step.

Using this approximation, the observation equations become

$$\log(i_t) = \log(I_t) + \log(q) + \epsilon_{I,t} \tag{3.9}$$

$$\log(\Delta c_y) = \log F_{y-\frac{1}{2}} + \log B_{y-\frac{1}{2}} + \epsilon_{C,y} \tag{3.10}$$

where $\epsilon_{I,t} \sim N(0, \sigma_I)$ and $\epsilon_{C,y} \sim N(0, \sigma_C)$.

3.2 Models Based on Numbers-by-age

Most analytical stock assessments are based on age-structured models. There are several reasons for this (including historical reasons), but a main reason is that biological processes such as growth and reproduction are inherently linked to yearly cycles.

Mortality is typically divided into fishing mortality and and natural mortality. This can be expressed through the following differential equation:

$$\frac{d}{dt} N_t = -\underbrace{(F_t + M_t)}_{Z_t} N_t \tag{3.11}$$

where N_t is the total number of individuals at time t, and the total mortality Z_t is the sum of fishing mortality F_t and natural mortality M_t. The natural mortality is usually considered to be known a priori, since it is hardly ever possible

to estimate temporal development in fishing mortality and natural mortality simultaneously from data.

The solution to 3.11, using one year as the time unit and assuming constant mortalities within a year, is known as the *stock equation*,

$$N_{a+1,y+1} = N_{a,y}e^{-Z_{a,y}}. \tag{3.12}$$

where $N_{a,y}$ is the number of alive fish at the beginning of the year. The number of individuals that died due to fishing is found using the *catch equation*

$$C_{a,y} = \frac{F_{a,y}}{Z_{a,y}}(1 - e^{-Z_{a,y}})N_{a,y}. \tag{3.13}$$

Equations (3.13) and (3.12) are fundamental in age-structured fish stock assessment models.

The recruitment is typically linked to the *spawning stock biomass* (SSB), which is key quantity in relation to fish stock assessments. It is defined as the total mass of sexually mature individuals in a population and is thus a measure of the reproductive capacity of a stock. In age-structured models it calculated as follows

$$SSB_y = \sum_{a=1}^{A} N_{a,y}w_{a,y}p_a \tag{3.14}$$

where w is the mean weight of a fish, p is the proportion of mature individuals, and N is the estimated number of fish. Subscripts a and y denote age group and year respectively.

Since the number of future recruits in a population is key in determining its future development, much research has been devoted to the development of models for predicting recruitment. Environmental forcings, predation, and food availability in the pre-recruit stages are often considered in this respect (e.g. Beaugrand et al. (2003)). Although there are many publications which have found correlations between the environment and recruitment success, few of these have stood the test of time and have failed upon later retests using longer time-series of data (Myers, 1998).

Hence, most stock assessment models assume a simple relationship between the spawning stock biomass in the previous year SSB_{y-1} and the next year's recruitment R_y, and the most frequently used are the Beverton-Holt model Beverton et al. (1957)

$$R_y = \frac{\alpha SSB_{y-1}}{1 + \frac{SSB_{y-1}}{K}}, \tag{3.15}$$

and the Ricker model Ricker (1954)

$$R_y = \alpha SSB_{y-1} \exp\left(-\beta SSB_{y-1}\right). \tag{3.16}$$

Compensatory density dependence is the term used to describe the phenomenon that at high population densities the mortality increases due to increased competition, predation, and disease transmission, which prevents the population from exploding. However, for most fish stocks there are only data available from periods where stock sizes are well below the carrying capacity, and recruitment is often highly variable. Hence there is often little hope to distinguish between these models. This is illustrated in Figure 3.1.

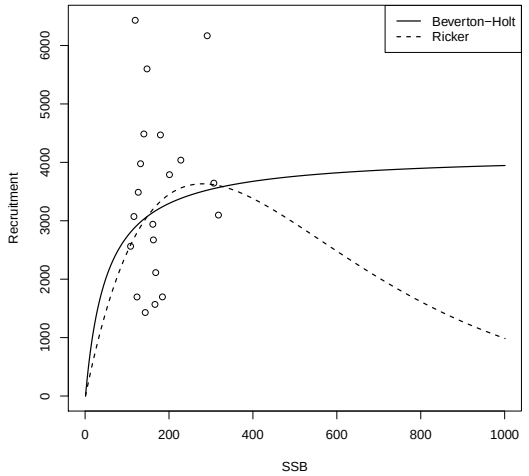


Figure 3.1: Stock-recruitment for Western Baltic Spring Spawning Herring (1991-2011).

3.2.1 Age-Length Keys

Converting observed catches from length to age distributions is a standard procedure when analyzing fisheries data using age-structured models, and to this end a so-called age-length key (ALK) is typically applied. While measuring for length is easy, age determination requires much more expensive laboratory work, so only a small fraction of the catch is analyzed for age. This implies, that many hauls may contain very few or no age samples, which means that samples must be pooled somehow under the assumption that the ALK can be

considered constant (or at least similar) for hauls that are close in space and time. The current ICES procedure is to use raw proportions of age by length group aggregated over larger areas (ICES, 2012c), and to fill in gaps in the ALK by "borrowing" samples from neighboring areas if no fish have been sampled for some length groups in the area under consideration.

Continuation ratio logits (CRLs, Agresti (2010)) is a statistical model for ordered categorical responses that have been applied as an objective and more robust way to obtain ALKs (Kvist et al. (2000), Rindorf and Lewy (2001)), and to permit statistical inference such as to test whether two ALKs can be considered identical (Gerritsen et al. (2006), Stari et al. (2010)).

Let $a = R \dots A$ be the age groups under consideration and l the length of a fish. The distribution of ages $P_a = \{p_R \dots p_A\}$ is modelled through A minus R models for the conditional probability of being of age a given that it is at least age a:

$$\pi_a = P(Y = a | Y \geq a) = \frac{p_a}{p_a + \ldots + p_A}, \quad a = R \dots A - 1$$

which relates to the unconditional probabilities p_a (the ALK) trough the following equations

$$\begin{aligned} p_R &= \pi_R \\ p_a &= \pi_a \left(1 - \sum_{j=R}^{a-1} p_j \right) = \pi_a \prod_{j=R}^{a-1} (1 - \pi_j), \quad a > R \end{aligned}$$

The CRLs based on area-stratification are usually just linear functions of length

$$\text{logit}(\pi_a) = \alpha_a + \beta_a l$$

While solving the problem of having gaps in the ALK, the problem of selecting appropriate areas in which the ALK can be considered equal is not addressed. To this end, Paper III proposes to replace area stratification with smooth functions s of the spatial coordinates (lon,lat):

$$\text{logit}(\pi_a) = \alpha_a + s_{1,a}(lon, lat) + s_{2,a}(lon, lat)l \tag{3.17}$$

This type of model is known as a varying-coefficients model (Hastie and Tibshirani, 1993) since the regression coefficient on length is now a smooth function of covariates. Other covariates may of course also be included.

3.2.2 Catch standardization of survey data

Changes in survey design represents a challenge when time-series of indices of abundance are to be calculated. If such changes are not accounted for, trends in the indices may be a result of the changes in design rather than changes in abundance. Catch standardization refers to procedures for removing effects from the indices (or catch-effort data) that are caused by changes in design or other covariates such as environmental conditions (Maunder and Punt, 2004).

Such analyses are statistically challenging because catch rates are notoriously variable in that they often contain a high proportion of zero values as well as occasional huge catches. Delta-distributions (e.g. Pennington (1983)), where zero values are modelled separately and the positive values are assumed to be log-normal (or Gamma) distributed, are a popular choice for describing highly skewed catch data. More recent studies have combined delta-distributions with generalized linear models (GLMs), generalized linear mixed models (GLMMs) and generalized additive models (GAMs) to correct for effects such as spatial position, depth, and time of day (Stefansson, 1996; Petrakis et al., 2001; Piet, 2002; Adlerstein and Ehrich, 2003; Beare et al., 2005). Discrete valued distributions such as the negative binomial (e.g. Kristensen et al. (2006)) have been investigated as well, but if an age-length key has been applied prior to the analysis, the response variable is no longer discrete, and models for continuous responses must be used instead. The Tweedie distribution (Tweedie, 1984) has recently been suggested as an alternative to delta-distributions (Candy, 2004; Shono, 2008). It has a nice interpretation as a compound Poisson distribution, because it is equivalent to the distribution of $Z = W_1 + ... + W_N$, where W_k are independent identically distributed Gamma variables, and N follows a Poisson distribution. Hence, it has a point mass in zero (corresponding to $N = 0$) but is continuous on the rest of positive part of the real axis.

Nevertheless, a simpler method based on arithmetic means and area stratification is still the standard method employed by ICES (ICES, 2012c).

Paper IV compares the method employed by ICES with the delta- and Tweedie distributions using GAMs to correct for changes in survey design. The delta- and Tweedie models uses the following relationship between the expected response μ, which is numbers-at-age or 1/0 for the binomial part of the delta models, and external factors

$$g(\mu_i) = \text{Year(i)} + \text{U(i)}_{\text{ship}} + f_1(\text{lon}_i, \text{lat}_i) + f_2(\text{depth}_i) + f_3(\text{time}_i) \qquad (3.18)$$

where Year(i) maps the ith haul to a categorical effect for each year, $\mathrm{U(i)}_{\mathrm{ship}} \sim N(0, \sigma_u^2)$ is a random vessel effect, f_1 is a 2-dimensional thin plate regression spline on the geographical coordinates, f_2 is a 1-dimensional thin plate spline for the effect of bottom depth, and f_3 is a cyclic cubic regression spline on the time of day.

The motivation for using a smooth function of geographical coordinates in place of area-stratification is the same as in the previous section on ALK estimation: The problem of selecting appropriate strata that have a sufficient number of data points in all the years, yet still being small enough to assume of homogeneity within strata, is replaced with an easier problem of smoothness selection for the splines, which can be solved more less automatically using modern software packages (Wood, 2006). In addition, related studies have shown that GAMs provide more accurate estimates (Maxwell et al., 2012).

3.2.3 Deterministic Assessment Methods

Virtual population analysis (VPA, Fry (1949); Gulland (1965)) is a purely deterministic method for stock assessment that unlike statistical methods does not provide any quantification of uncertainties of its estimates. Briefly, the procedure requires a guess on either the number of survivors or (more often) the fishing mortalities in the final year and in the plus group (i.e. the last column and the last row of either the N-matrix or the F-matrix). By recursive substitution into the catch and stock equations (3.13) and (3.12) it is possible to back-calculate the entire N-matrix and F-matrix numerically. Pope (1972) derived an approximation by pretending the catch is taken exactly in the middle of the year, which may be solved analytically. Further details about VPA are found in (Ulltang, 1977; Lassen and Medley, 2001).

VPA relies on catch data alone which does not provide much information about the fishing mortality in the terminal years. Extended survivors analysis (XSA, Shepherd (1999)) is an extension to VPA, which allow effort information for commercial fleets or survey CPUEs to be included in the analysis. Although not a statistical model, this method is still used as a standard method for stock assessment by ICES today.

3.2.4 Statistical Assessment Methods

Stochastic models for stock assessment (introduced by Doubleday (1976), and generalized to include auxiliary data by Fournier and Archibald (1982)) differs

from the deterministic methods in that they treat observations as stochastic variables subject to noise. The amount of observation noise is estimated from data as a part of the model, which allows for quantification of uncertainties and proper statistical hypothesis testing. These classical statistical assessment models, also known as statistical catch-at-age models (SCA), are fully parameterized: unknown quantities are parameters of the model and observations are random variables. However, more observations than parameters are needed for the model to be identifiable, which means that some constraints on the model is necessary. For instance, it is not possible to estimate an age- and year-specific fishery mortality $F_{a,y}$ for every combination of age and year. Hence, a separability assumption is usually made in which the fishing mortality is a product of a year and an age effect $F_{a,y} = F_a F_y$. This assumption is a pragmatic solution, but it does not allow for selectivity pattern to evolve over time. The state-space model presented in section 3.2.7 solves this problem.

3.2.5 State-space models

Modern statistical approaches for analyzing fisheries time-series attempts to meet the challenge of separating different sources of uncertainty such as process error, observation error, and uncertainty about model structure (de Valpine and Hastings, 2002). State-space models offer a formalized framework for analyzing systems where both process error and observation error exist. The perhaps simplest example is where the state η follow a random walk process

$$\eta_{t+1} = \eta_t + \epsilon_{t,\eta} \tag{3.19}$$

where $\epsilon_{t,\eta} \sim N(0, \sigma_\eta^2)$ are independent Gaussian random variables. In the context of fisheries models, η could represent the logarithm of the fishing mortality. The system equation (3.19) would then describe the variability in the temporal development of the fishing mortality. If we pretend we had direct observations Y of the fishery mortality, these would most certainly deviate from the true fishing mortality due to errors from the measurement process. The observation equation (3.20) describes this other source of error:

$$Y_t = \eta_t + \epsilon_{t,Y}. \tag{3.20}$$

where $\epsilon_{t,Y} \sim N(0, \sigma_Y^2)$ is the observation error. The true state of the unobserved fishery mortality process η can never be known, but using the state-space frame-

work it is possible to give an optimal (in a least-squares sense) reconstruction $\hat{\eta}$ and to quantify the uncertainty of that estimate. The above example is linear with Gaussian errors, which means that the Kalman filter gives the optimal solution (Harvey, 1989).

Since the basic equations for fish stock assessment models are non-linear, estimation must be based on other methods such as the Extended Kalman Filter (Gudmundsson, 1994). An alternative estimation approach consists of viewing the states as random effects in a mixed effects model. This involves specifying the joint likelihood $L(\theta, \eta, Y)$ as the product of the conditional density of observations given the random effects, $f_\theta(Y|\eta)$, and the marginal density of the random effects $h_\theta(\eta)$, which in the random walk example is,

$$L(\theta, \eta, Y) = \prod_{t=2}^{N} \phi\left(\eta_t - \eta_{t-1}, \sigma_\eta^2\right) \prod_{t=1}^{N} \phi\left(Y_t - \eta_t, \sigma_Y^2\right) \tag{3.21}$$

where ϕ denotes the Gaussian probability density function. The maximum likelihood estimate of θ is found be integrating out the random effects to obtain the marginal distribution of θ given Y. A method to evaluate this integral is outlined in the next section.

3.2.6 ADMB and the Laplace approximation

Automatic Differentiation Model Builder (ADMB, Fournier et al. (2012)) is a powerful open-source programming framework for solving complex non-linear optimization problems. This includes, and is in fact aimed at, maximum likelihood estimation in highly parameterized statistical models and random effect models. Efficient algorithms for optimization of a (likelihood) function of many variables involves computing its derivatives, which is typically handled by numeric finite-difference methods or by finding analytical expressions. Finite-difference methods are not exact and become slow when many parameters are involved, while deriving analytical expressions for a complex model with perhaps hundreds of parameters is impractical and error-prone. A clever method known as automatic differentiation allows derivatives to be calculated to machine precision for any objective-function expressed as computer code by successively applying the chain rule of calculus to every operation that affects the objective function. ADMB integrates automatic differentiation with a function minimizer as well as other functions for statistical analysis such as the Laplace approximation for random effect models, profile likelihood, and MCMC and the Delta method for estimation of uncertainties of estimated quantities.

In random effects models it is common to use the Laplace approximation (Wolfinger and Xihong, 1997) for evaluation of the marginal likelihood. The computational aspects of AD and the Laplace approximation are given in Skaug and Fournier (2006), and a short description of the Laplace approximation is given in the following.

Let θ be a vector of parameters (fixed effects) and let η be a vector of latent random effects; the states in the state-space framework. Let Y be a vector of observations, and let $g(\theta, \eta)$ be the joint log-likelihood.

The marginal likelihood is found by integrating out the random effects in the likelihood $\exp(g(\theta, \eta))$,

$$L(\theta|Y) = \int \exp(g(\theta, \eta))d\eta \tag{3.22}$$

but this integral is very difficult to evaluate numerically.

Let $\hat{\eta}_\theta$ be the argument that maximizes the joint likelihood for fixed θ, i.e.

$$\hat{\eta}_\theta = \underset{\eta}{\operatorname{argmax}}\, g(\theta, \eta) \tag{3.23}$$

The Laplace approximation replaces the integrand with a second order Taylor expansion of the likelihood function around the optimum $\hat{\eta}_\theta$:

$$g(\theta, \eta) \approx g(\theta, \hat{\eta}_\theta) + \frac{1}{2}\left(\eta - \hat{\eta}_\theta\right)^T H(\theta)\left(\eta - \hat{\eta}_\theta\right) \tag{3.24}$$

where $H(\theta)$ is the Hessian matrix (second order partial derivatives) of the log-likelihood function evaluated at the maximum $\hat{\eta}_\theta$. Notice, that since we are at the optimum, all first order partial derivatives are zero, which is why these are not present in the Taylor expansion. Also notice, that the second term is proportional to the probability density function for a multivariate normal distribution, which by definition has its integral equal to one. Hence,

$$L(\theta|Y) \approx \exp\left(g\left(\theta, \hat{\eta}_\theta\right)\right) \int g(\theta, \hat{\eta}_\theta) + \frac{1}{2}\left(\eta - \hat{\eta}_\theta\right)^T H(\theta)\left(\eta - \hat{\eta}_\theta\right) d\eta \tag{3.25}$$

$$\propto \exp\left(g\left(\theta, \hat{\eta}_\theta\right)\right) \frac{1}{\sqrt{\det H(\theta)}} \tag{3.26}$$

Estimation of θ by maximizing the Laplace approximation is known as the *outer* optimization problem, while the estimation of η (3.23) is known as the

inner problem. The joint estimation can be accomplished by iterating between optimizing η for fixed θ and vice versa (Skaug and Fournier, 2006).

In paper I random effects that follow independent truncated normal distributions are applied. In this case, the Laplace approximation may be inappropriate since the posterior distribution is probably not Gaussian, but it is also not needed since the high-dimensional integral (3.22) reduces to products of one-dimensional integrals.

3.2.7 SAM

This section introduces the state-space model for age-structured stock assessments. This model was initially developed by Anders Nielsen and extensions of this model appear in paper IV and V. A similar model is presented by Gudmundsson and Gunnlaugsson (2012).

As usual, the state-space model consists of a vector of states η describing the system at time t, a set of process equations describing the transition of states between time steps T, and finally a set of observation equations O relating the observations Y to the states.

$$\eta_t = T\left(\eta_{t-1}\right) + \varepsilon_{T,t} \tag{3.27}$$
$$Y_t = O\left(\eta_t\right) + \varepsilon_{O,t} \tag{3.28}$$

The vector of states, η, consist of log-transformed numbers-at-age $\log N_1, \ldots, \log N_A$ and fishing mortalities $\log F_{i_1}, \ldots, \log F_{i_n}$. The fishing mortalities may correspond to age-classes or to groupings of these. The general structure of the model is illustrated in Figure 3.2.

The process equations for $\log N_i$ using yearly time-steps (t is replaced with y) are:

$$\log N_{1,y} = \log R(\eta, \theta_R) + \varepsilon_{R,y} \tag{3.29}$$
$$\log N_{a,y} = \log N_{a-1,y-1} - F_{a-1,y-1} - M_{a-1} \; + \varepsilon_{S,a,y} \quad , 2 \le a \le A \tag{3.30}$$
$$\begin{aligned} \log N_{A,y} = \log(&\mathrm{e}^{\log N_{A-1,y-1} - F_{A-1,y-1} - M_{A-1}} \\ &+ \mathrm{e}^{\log N_{A,y-1} - F_{A,y-1} - M_A}) + \varepsilon_{S,A,y} \end{aligned} \tag{3.31}$$

where M_a is natural mortality at age a, and F is the total fishing mortality. The recruitment process is described by (3.29) which could be standard functions

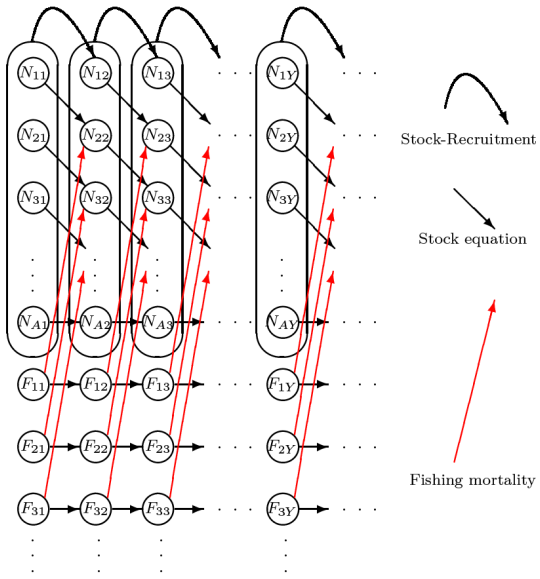


Figure 3.2: Illustration of state transitions in age-based stock assessment models (based on Nielsen (2004)).

of SSB such as the Ricker or Beverton-Holt models, whose parameters θ_R is estimated as part of the model. The process errors ε_R and ε_S are assumed to be zero-mean independent normal distributed with two separate variance parameters, one for recruitment σ_R^2, and one for survival σ_S^2.

The fishery mortalities, F, are assumed to follow either independent random walks:

$$\log F_{a,y} = \log F_{a,y-1} \ + \varepsilon_{F,a,y}, \qquad 1 \leq a \leq A \tag{3.32}$$

or a correlated random walk model (with scalars replaced by vectors)

$$\log F_{y} = \log F_{y-1} \ + \varepsilon_{F,y} \tag{3.33}$$

such that $\varepsilon_{F,y} \sim N(\mathbf{0}, \mathbf{\Sigma}_F)$ and $\mathbf{\Sigma}_{F,ij} = \sigma_F^2 \rho$ for $i \neq j$ and $\mathbf{\Sigma}_{F,ii} = \sigma_F^2$, where σ_F and ρ are parameters to be estimated. When $\rho = 1.0$ we have the special case of a multiplicative structure in $\log F$ whereas $\rho = 0$ allows for completely independent development by age group in fishery mortality over time.

Assuming independent observations, the observation equations become:

$$\log C_{a,y} = \log\left(\frac{F_{a,y}}{Z_{a,y}}(1 - e^{-Z_{a,y}})N_{a,y}\right) + \varepsilon^C_{a,y} \tag{3.34}$$

$$\log I^{(s)}_{a,y} = \log\left(Q^{(s)}_a e^{-Z_{a,y}\frac{D^{(s)}}{365}} N_{a,y}\right) + \varepsilon^{(s)}_{a,y} \tag{3.35}$$

where $Z_{a,y} = M_a + F_{a,y}$ is the total mortality rate, $D^{(s)}$ is the number of days into the year where the survey s is conducted, and $Q^{(s)}_a$ are catchability parameters, $\varepsilon^C_{a,y} \sim N(0, \sigma^2_C)$, and $\varepsilon^{(s)}_{a,y} \sim N(0, \sigma^2_s)$.

The joint likelihood is

$$L(\theta, \eta, Y) = \prod_{t=2}^{N} \phi\left(\eta_t - T\left(\eta_{t-1}\right), \Sigma_T\right) \prod_{t=1}^{N} \phi\left(Y_t - O\left(\eta_t\right), \Sigma_O\right) \tag{3.36}$$

where ϕ denotes the Gaussian probability density function. Notice how each factor depends on at most two random effect vectors (η_t and η_{t-1}). This a common feature of state-space models known as partial separability, and by wrapping the evaluation of each factor into a specially declared `SEPARABLE_FUNCTION` call in ADMB the computational costs can be greatly reduced, since the conditional independence between most of the random effects implies that the Hessian is a banded matrix containing many zeroes (Skaug and Fournier, 2006).

3.2.8 Incorporating within year correlations on survey indices

It is common to assume independence among ages in observations from scientific surveys like in (3.35), although it has previously been demonstrated that this assumption is not valid in many cases (Walters and Punt (1994),Pennington and Godø (1995),Myers and Cadigan (1995)).

Paper IV demonstrates that such correlations are also present for survey data on North Sea herring, sprat, and whiting, especially among the older age-groups. The SAM model is therefore extended to account for such correlations by changing the observation equation (3.35) to

$$\log I^{(s)}_{y} = \log\left(Q^{(s)} \circ e^{-Z_y \frac{D^{(s)}}{365}} N_y\right) + \varepsilon^{(s)}_{y} \tag{3.37}$$

where $\varepsilon_y^{(s)} \sim N(\mathbf{0}, \mathbf{\Sigma}_y)$ (note that scalars in eqn. 3.35 are replaced with vectors containing all age groups at once, and "$\circ$" denotes element-wise multiplication).

The covariance matrices $\mathbf{\Sigma}_y$ are found by bootstrapping entire hauls from the survey experiment and repeating the entire procedure that leads to numbers-at-age estimates from the surveys.

3.2.9 A mixed stock model extension

Many commercial fisheries exploit several stocks at once, which implies that the assessments of the stocks under exploitation ideally should be integrated. Integrated analysis involves analyzing all the available data in as raw a form as possible in a single analysis, and should be preferred over separate analyses when possible (Maunder and Punt, 2012).

Atlantic herring is an example of a species that is characterized by several local sub-stocks, which differ in their choice of spawning grounds and timing of spawning. These stocks may therefore be regarded as isolated with respect to reproduction, but they often appear together in samples of catches from both commercial and survey vessels, so their exploitation rates are linked in areas where such overlap occurs. Western Baltic spring spawning herring (WBSS) and North Sea autumn spawning herring (NSAS) are currently modelled separately by dis-aggregating catches according to sub-samples of spawner-type (ICES, 2012b). This split ignores spatial aspects of the exploitation pattern that may exist, since the area between the spawning grounds of the two species where the mixing occurs (known as area IIIa) is managed separately.

Such stock-complexes have previously been modelled using coupled differential (or difference) equations, also known as compartmental models, e.g.

$$\begin{aligned} \frac{dN_1}{dt} &= -(Z_1 + k_{12})N_1 + k_{21}N_2 \\ \frac{dN_2}{dt} &= k_{12}N_1 - (Z_2 + k_{21})N_2 \end{aligned}$$

where $k_{ij} \geq 0$ represents the transport from area i to area j. A schematic illustration of the described herring complex is shown in Figure 3.3.

These models require the estimation of a matrix with transport coefficients from and to each area (the ks) from tagging data (Quinn et al., 1990; Goethel et al., 2011). However, tagging data is not available for the mentioned herring stocks,

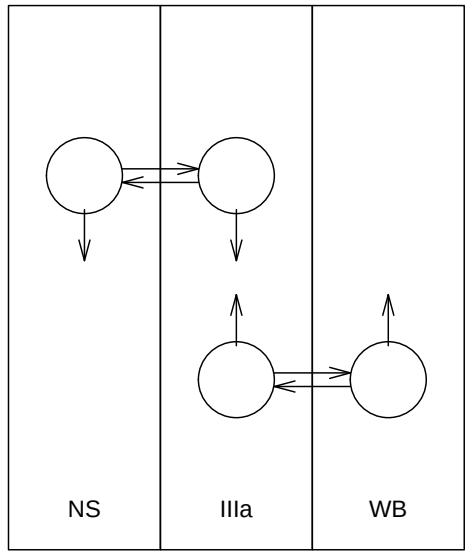


Figure 3.3: Compartmental model for North Sea and Western Baltic herring. Arrows between compartments indicate transport. Outgoing (vertical) arrows represent mortality.

so other approaches should be investigated. To this end, Paper V develops an integrated assessment model of WBSS and NSAS herring. The aim of this model is to give a separate estimate of the fishing mortality in the mixing area, and to allow inferences and predictions of the mixing dynamics.

The model comprises the two stocks, $s = 1, 2$, and three areas $k = 1 \dots 3$, where individuals belonging to stock $s = 1$ can be located in either area $k = 1$ or 2, but not $k = 3$, and individuals belonging to stock $s = 2$ can be located in area $k = 2$ or 3, but not $k = 1$.

The state vector defined in section 3.2.7 is extended accordingly to consist of the log-transformed number of individuals in each age class for each stock, $\log N_a^{(s)}$, the log-transformed fisheries mortalities in each area $\log F_a^{(k)}$, and finally the logit-proportion of age class a from stock s that is located in the mixing area (k=2), $\beta_a^{(s)} = \log\left(\frac{\pi_a^{(s)}}{1-\pi_a^{(s)}}\right)$.

The process equations (with age index omitted) for $\log N$ are made stock-specific, area specific for $\log F$, and a random walk process is assumed for the age and area-specific logit-proportions β to allow for temporal development in spatial distribution of stocks over areas:

$$\log N_{s,y} = \log\left(N_{s,y-1}e^{-Z_{s,y-1}}\right) + \epsilon_{N,s,y} \quad (3.38)$$

$$\log F_{k,y} = \log F_{k,y-1} + \epsilon_{F,k,y} \quad (3.39)$$

$$\beta_{s,y} = \beta_{s,y-1} + \epsilon_{\beta,s,y} \quad (3.40)$$

and the observation equations are:

$$\log C_{k,y} = \log\left(\sum_s \left[\frac{F_{k,y}}{Z_{s,k,y}}(1-e^{-Z_{s,k,y}})N_{s,k,y}\right]\right) + \epsilon_{C,k,y} \quad (3.41)$$

$$\log I_{f,k,y} = \log\left(Q_{f,k,y}e^{D_f/365}\left(\sum_s e^{-Z_{s,k,y}}N_{s,k,y}\right)\right) + \epsilon_{I,f,k,y} \quad (3.42)$$

$$\text{logit}(p_{AS,y}) = \text{logit}\left(\frac{\pi^{(1)}N^{(1)}}{\pi^{(1)}N^{(1)} + \pi^{(2)}N^{(2)}}\right) + \epsilon_{p_{AS},y} \quad (3.43)$$

Subscript k denote area, s stock, y year, and f is fleet. Besides the standard set of parameters from in the single stock case, we have also process noise for the temporal development in the logit-proportions described by σ_β, and the split proportion observation noise (σ_p).

3.3 Models Based on Length and Other Models

Reliable age data are not always available. For some species of fish the ring structures in the otoliths lack sufficient contrast to identify the age, and in other cases the data is simply missing. In such situations, the population may be more conveniently described by its length distribution rather than the age distribution. Gudmundsson (1995) developed a model based on catch-at-length and compared estimation accuracy with a model based on catch-at-age data of similar variability. Higher accuracy was obtained when the age-data was used.

Integrated assessment models that combine the statistical treatment of length and age observations in a single model (e.g. Fournier et al. (1998)) have not been considered in this thesis, although these represent a more sound albeit involved statistical approach.

Another recent advanced statistical approach to population modeling based purely on length data is described in Kristensen (2009). Here, a biological model for growth, recruitment and mortality (a so-called size-spectrum model) is combined with a statistical model based on the log-Gaussian Cox-process (LGCP) for survey observations, which incorporates size, space and time correlations.

Chapter 4

Results and Conclusion

This chapter summarizes the main results and findings documented in Papers I-VI.

Paper I re-analyzes the data set presented in Mergardt and Temming (1997), who were able to identify one nightly peak period of feeding, which constituted an improvement over analyses performed with methods based on mean stomach contents, which failed to identify any diel feeding pattern. However, their results did not agree with their expectations: "The opposite direction of the vertical migration routes of whiting and sandeels reduces the potential times of spatial overlap to two narrow periods during dusk and dawn, from which one would expect two feeding peaks.". The failure to identify two feeding peaks was attributed the limited precision of the back-calculation method. The new method presented in paper I is able to confirm the original hypothesis of a bimodal feeding pattern using the same data set. The problem is illustrated in figure 4.1: By disregarding varying amounts of uncertainty in the observations, the signal in data is blurred and the true fluctuations in the feeding pattern are not sufficiently captured.

The purpose of Paper II was to assess the estimation performance of three approaches to nonlinear state-space models: Hidden Markov Models (HMM) using a discretized state-space, ADMB, and a Markov Chain Monte Carlo method (BUGS, Spiegelhalter et al. (2003)). While BUGS has received the most atten-

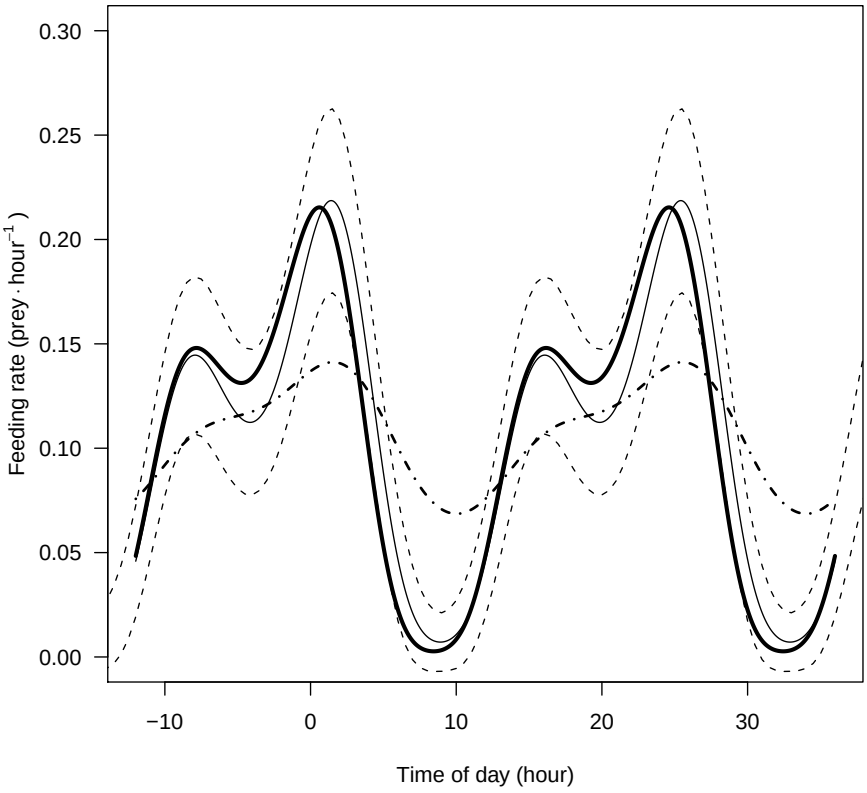


Figure 4.1: Example of a reconstructed feeding pattern the simulation study in Paper I. The true feeding intensity (thick solid), the estimated intensity (thin solid), and the marginal 95% confidence interval (thin dashed) are shown. The thick dashed line is based on a simple ML estimate that does not take varying amounts of observation error into account (from Paper I).

tion in the literature on ecological modelling, this must largely be accredited its user friendly implementation (winBUGS) rather than superiority with respect to parameter estimation. ADMB was found to be by far the fastest method owing to its use of automatic differentiation combined with the Laplace approximation. The only downside is, that the Laplace approximation requires approximate Gaussian posterior distributions, although it is possible to replace the Laplace approximation with importance sampling at the expense of the speed. The Laplace approximation is nevertheless applicable in most practical applications (Skaug and Fournier, 2006). The HMM method gave practically identical results to ADMB, but with much longer computation times. The HMM method and BUGS may however be applied to a larger range of problems due to their ability to handle discrete random effects and multi-modal posterior distributions.

Paper III combines continuation ratio logits with non-parametric models (GAMs) to estimate age as smooth functions of length and spatial position for fish. The results confirmed the findings in similar studies, namely that significant spatial variation in the relationship between length and age exist for many fish stocks. This is illustrated in figure 4.2, which also illustrates the flexibility and smoothness obtained in the age-length keys by using this method. By replacing spatial stratification with thin-plate splines to account for this variability, a higher number of age-groups may be included in the analysis. Also, improved internal consistency was found in indices of abundance derived using this method indicating improved precision.

Paper IV is related to paper III in that it also applies GAMs to replace area stratification, although this time it is for the purpose of catch standardization of survey indices of abundance. The standard method for the species considered in this study (the stratified mean method) was compared to three more similar alternatives: Delta-Lognormal, Delta-Gamma, and Tweedie distributed observations, all of which accounted for spatial effects, unbalanced designs, and covariate effects (depth, time-of-day, vessel and gear) through GAMs. The Delta-Lognormal model proved to be the best choice. This was supported by information criteria (AIC/BIC), higher internal/external consistency in the produced survey indices, and finally by leading to smaller uncertainties on estimates of spawning stock biomass and fishery mortality from stock assessments. Significant positive correlations between the estimated numbers-at-age for the older age-groups were evident from the residual analyses from the stock assessment model. This was also confirmed by a bootstrapping procedure using the survey data only. The stock assessment model was therefore extended to account for these correlations. The perhaps most important conclusion in paper IV is, that the stratified mean method currently used for many stock assessments by ICES performs very poorly compared to the examined alternatives.

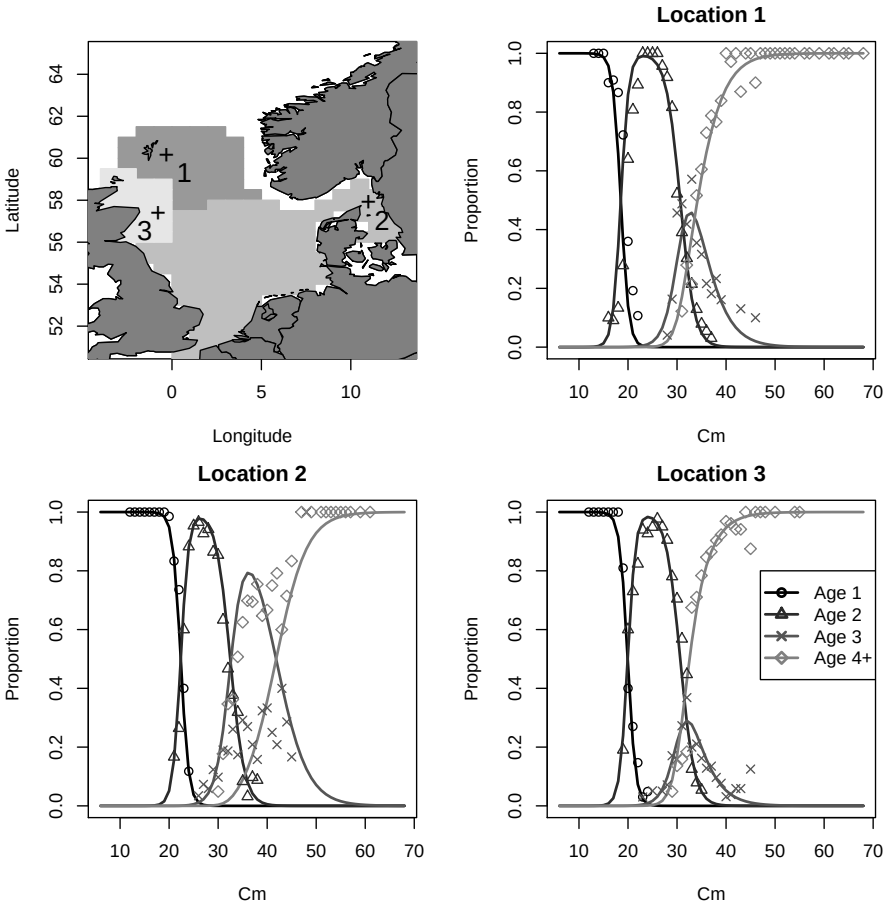


Figure 4.2: Illustration of the results of applying a spatially smooth age-length key to data on North Sea Haddock (solid lines) in three selected locations contrasted with the alternative of using raw observed proportions (points) within each of the boxed areas (from Paper III).

Paper V introduces integrated assessment modelling of two herring stocks. This model represents an improvement over the currently applied methodology, in which two separate single stock assessments are carried out using data that have been split by stock according to subsamples of the catches. There are (at least) three reasons why the integrated approach is to be preferred. 1) The two stocks cover three management areas each with separate associated catch quotas, and only the integrated approach permits estimation of the fishery mortality in all three areas. 2) The uncertainty in the splitting proportions is quantified and accounted for in the calculations of all quantities of interest. 3) Unusual large cohorts observed in either stock are accounted for in the predictions of future catches in the area where the two stocks mix. The model is simple compared to other models with spatial structure in that it does not explicitly model migration between the areas, but assumes homogeneity within areas and time-steps. This assumption is probably more or less violated, however more data is needed to facilitate more realistic models of the movement and exploitation dynamics.

Paper VI considers estimation in biomass models – an important class of models for stock assessment in data-poor situations where only the total mass of the catches is recorded. Most models of this type rely on estimators that only takes one type of error into account, namely observation error on the relative indices of abundance. The proposed model takes two additional sources of error into account: process error due to simple structure of the model compared to the real system, and observation error on the commercial catches. Failing to account for substantial sources of error will lead to biased results and wrong perceptions of the uncertainty related to key management quantities. The model is formulation in continuous time using stochastic differential equations rather than the usual discrete time approach. A distinct advantage by using the continuous time formulation is the ability to handle varying sample times. In other words, observed commercial landings or survey indices may be recorded at any time of the year and used as data for the estimation process without having to change the model. It was demonstrated that the maximum sustainable yield (MSY) depends on the process error, and that the estimates of MSY is rather uncertain in many cases.

4.1 Conclusions

This Ph.D. project has touched upon several statistical subjects related to fish stock assessment. Common among the subjects were the application of mixed effects models, with particular attention given to state-space models, and various methods for estimation in such models were considered.

A main result in this thesis is, that important information may be lost when preprocessing steps are applied with the purpose of reducing data dimensionality prior to subsequent statistical analyses. While the ideal solution is to integrate the preprocessing into the statistical model, this a complicated and computationally demanding task. However, a compromise solution, which was repeatedly applied in this thesis, consists of applying statistical methodology in the preprocessing step in order to obtain probability distributions rather than point estimates of the output quantities from the preprocessing step.

Special attention was given to the problem of calculating age-specific indices of abundance from trawl survey data. This involved the development of a new method for converting from length to age-distributions, and taking effects due to the survey design into account, such as high sampling variability, depth, daylight and vessel effects. It was demonstrated, that the developed methods gave significant different estimates of abundance than currently used methods, and that the new methods constituted an improvement.

Two extensions to a state-space approach to age-structured stock assessment modelling were presented. The first extension introduced multivariate error distributions on survey catch-at-age data. The second extension was an integrated assessment model for overlapping sub-stocks subject to mixed exploitation in the area of overlap. Both extensions are important first steps towards better handling of the uncertainties related to the input to, as well as the output of, stock assessment models.

Finally, a biomass dynamic model based on stochastic differential equations was developed. This work extended the classical approaches to biomass modelling by incorporating observation errors on the catches, and allowing for missing and non-equidistant samples.

Bibliography

Adlerstein, S. and Ehrich, S. (2003). Patterns in diel variation of cod catches in north sea bottom trawl surveys. *Fisheries Research*, 63(2):169 – 178.

Agresti, A. (2010). *Analysis of Ordinal Categorical Data*. Wiley Series in Probability and Statistics. Wiley.

Andersen, N. G. and Beyer, J. E. (2005). Gastric evacuation of mixed stomach contents in predatory gadoids: an expanded application of the square root model to estimate food rations. *J. Fish Biol.*, 67(5):1413–1433.

Beare, D., Needle, C., Burns, F., and Reid, D. (2005). Using survey data independently from commercial data in stock assessment: an example using haddock in ICES Division VIa. *ICES Journal of Marine Science: Journal du Conseil*, 62(5):996–1005.

Beaugrand, G., Brander, K. M., Alistair Lindley, J., Souissi, S., and Reid, P. C. (2003). Plankton effect on cod recruitment in the north sea. *Nature*, 426(6967):661–664.

Beverton, R. J. H., Holt, S. J., et al. (1957). On the dynamics of exploited fish populations. *Fishery Investigations Series 2: Sea Fisheries*, 19.

Bromley, P. J. (1994). The role of gastric evacuation experiments in quantifying the feeding rates of predatory fish. *Reviews in Fish Biology and Fisheries*, 4(1):36–66.

Candy, S. (2004). Modelling catch and effort data using generalised linear models, the tweedie distribution, random vessel effects and random stratum-by-year effects. *CCAMLR SCIENCE*, 11:59–80.

Chen, Y. and Andrew, N. (1998). Parameter estimation in modelling the dynamics of fish stock biomass: Are currently used observation-error estimators reliable? *Canadian Journal of Fisheries and Aquatic Sciences*, 55(3):749–760.

de Valpine, P. and Hastings, A. (2002). Fitting population models incorporating process noise and observation error. *Ecological Monographs*, 72:57–76.

Doubleday, W. (1976). A least squares approach to analyzing catch at age data. *Int. Comm. Northwest Atl. Fish. Res. Bull*, 12:69–81.

Fournier, D. and Archibald, C. P. (1982). A general theory for analyzing catch at age data. *Canadian Journal of Fisheries and Aquatic Sciences*, 39(8):1195–1207.

Fournier, D. A., Hampton, J., and Sibert, J. R. (1998). Multifan-cl: A length-based, age-structured model for fisheries stock assessment, with application to sout pacific albacore, thunnus alalunga. *Canadian Journal of Fisheries and Aquatic Sciences*, 55(9):2105–2116.

Fournier, D. A., Skaug, H. J., Ancheta, J., Sibert, J., Ianelli, J., Magnusson, A., Maunder, M. N., and Nielsen, A. (2012). AD Model Builder: Using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, 27(2):233–249.

Fox, W. W. (1970). An exponential surplus-yield model for optimizing exploited fish populations. *Transactions of the American Fisheries Society*, 99(1):80–88.

Fry, F. E. (1949). Statistics of a lake trout fishery. *Biometrics*, 5(1):27–67.

Gerritsen, H. and Lordan, C. (2011). Integrating vessel monitoring systems (vms) data with daily catch data from logbooks to explore the spatial distribution of catch and effort at high resolution. *ICES Journal of Marine Science: Journal du Conseil*, 68(1):245–252.

Gerritsen, H. D., McGrath, D., and Lordan, C. (2006). A simple method for comparing age-length keys reveals significant regional differences within a single stock of haddock (melanogrammus aeglefinus). *ICES Journal of Marine Science*, 63(6):1096–1100.

Gislason, H. and Helgason, T. (1985). Species interaction in assessment of fish stocks with special application to the north sea. *Dana*, 5(2):1–44.

Goethel, D. R., Quinn, Terrance J., I., and Cadrin, S. X. (2011). Incorporating spatial structure in stock assessment: Movement modeling in marine fish population dynamics. *Reviews in Fisheries Science*, 19(2):119–136.

Gudmundsson and Gunnlaugsson (2012). Selection and estimation of sequential catch-at-age models. *Canadian Journal of Fisheries and Aquatic Sciences*, 69(11):1760.

Gudmundsson, G. (1994). Time series analysis of catch-at-age observations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 43(1):117–126.

Gudmundsson, G. (1995). Time series analysis of catch-at-length data. *ICES Journal of Marine Science*, 52(5):781–795.

Gulland, J. (1965). Estimation of mortality rates. annex to arctic fisheries workshop group report. *Int. Counc. Explor. Sea CM*.

Hampton, J. and Fournier, D. A. (2001). A spatially disaggregated, length-based, age-structured population model of yellowfin tuna (thunnus albacares) in the western and central pacific ocean. *Marine and Freshwater Research*, 52(7):937–963.

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, London, England.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(4):pp. 757–796.

ICES (2012a). Datras: survey descriptions. `http://datras.ices.dk/Home/Descriptions.aspx`.

ICES (2012b). Report of the herring assessment working group for the area south of 62 deg n (hawg). *ICES Document CM 2012/ACOM:06*.

ICES (2012c). Report of the Workshop on Implementation in DATRAS of Confidence Limits Estimation of, 10–12 May 2006, ICES Headquarters, Copenhagen.

Juhl, R., Kristensen, N., Bacher, P., Kloppenborg, J., and Madsen, H. (2013). CTSM-R - Continuous Time Stochastic Modelling for R. `http://www.ctsm.info`.

Kimura, D. K. and Somerton, D. A. (2006). Review of statistical aspects of survey sampling for marine fisheries. *Reviews in Fisheries Science*, 14(3):245–283.

Kristensen, K. (2009). *Statistical Aspects of Heterogeneous Population Dynamics*. PhD thesis, Department of Mathematical Sciences, University of Copenhagen.

Kristensen, K. and Berg, C. W. (2012). Datras package for r. `http://rforge.net/DATRAS/`.

Kristensen, K., Lewy, P., and Beyer, J. E. (2006). How to validate a length-based model of single-species fish stock dynamics. *Canadian Journal of Fisheries and Aquatic Sciences*, 63(11):2531–2542.

Kristensen, N. R., Madsen, H., and Jørgensen, S. B. (2004). Parameter estimation in stochastic grey-box models. *Automatica*, 40(2):225–237.

Kvist, T., Gislason, H., and Thyregod, P. (2000). Using continuation-ratio logits to analyze the variation of the age composition of fish catches. *Journal of Applied Statistics*, 27(3):303–319.

Larkin, P. A. (1977). An epitaph for the concept of maximum sustained yield. *Transactions of the American fisheries society*, 106(1):1–11.

Lassen, H. and Medley, P. (2001). *Virtual population analysis: a practical manual for stock assessment*, volume 400. Food & Agriculture Org.

Lewy, P. and Vinther, M. (2004). A stochastic age–length-structured multispecies model applied to North Sea stocks. *ICES CM 2004/FF:20*.

Maunder, M. N. and Punt, A. E. (2004). Standardizing catch and effort data: a review of recent approaches. *Fisheries Research*, 70(2-3):141–159.

Maunder, M. N. and Punt, A. E. (2012). A review of integrated analysis in fisheries stock assessment. *Fisheries Research*.

Maxwell, D. L., Armstrong, M. J., Beggs, S., and Aldridge, J. N. (2012). Annual egg production estimates of cod (gadus morhua), plaice (pleuronectes platessa) and haddock (melanogrammus aeglefinus) in the irish sea: The effects of modelling choices and assumptions. *Fisheries Research*, 117-118(0):146 – 155. Egg Production Methods in Marine Fisheries.

Mergardt, N. and Temming, A. (1997). Diel pattern of food intake in whiting (*Merlangius merlangus*) investigated from the weight of partly digested food particles in the stomach and laboratory determined particle decay functions. *ICES J. Mar. Sci.*, 54(2):226–242.

Misund, O. A. (1997). Underwater acoustics in marine fisheries and fisheries research. *Reviews in Fish Biology and Fisheries*, 7(1):1–34.

Myers, R. A. (1998). When do environment–recruitment correlations work? *Reviews in Fish Biology and Fisheries*, 8(3):285–305.

Myers, R. A. and Cadigan, N. G. (1995). Statistical analysis of catch-at-age data with correlated errors. *Canadian Journal of Fisheries and Aquatic Sciences*, 52(6):1265–1273.

Nielsen, A. (2004). *Estimating Fish Movement*. PhD thesis, The Royal Veterinary and Agricultural University.

Nielsen, A. and Sibert, J. R. (2007). State-space model for light-based tracking of marine animals. *Can. J. Fish. Aquat. Sci.*, 64:1055–1068.

Pella, J. J. and Tomlinson, P. K. (1969). *A generalized stock production model*. Inter-American Tropical Tuna Commission.

Pennington, M. (1983). Efficient estimators of abundance, for fish and plankton surveys. *Biometrics*, 39.

Pennington, M. and Godø, O. R. (1995). Measuring the effect of changes in catchability on the variance of marine survey abundance indices. *Fisheries research*, 23(3):301–310.

Petrakis, G., MacLennan, D. N., and Newton, A. W. (2001). Day–night and depth effects on catch rates during trawl surveys in the North Sea. *ICES Journal of Marine Science*, 58(1):50–60.

Piet, G. J. (2002). Using external information and GAMs to improve catch-at-age indices for North Sea plaice and sole. *ICES Journal of Marine Science: Journal du Conseil*, 59(3):624–632.

Polacheck, T., Hilborn, R., and Punt, A. E. (1993). Fitting surplus production models: Comparing methods and measuring uncertainty. *Canadian Journal of Fisheries and Aquatic Sciences*, 50(12):2597–2607.

Pope, J. (1972). An investigation of the accuracy of virtual population analysis using cohort analysis. *Int Comm Northwest Atl Fish Res Bull*, 9:65–74.

Punt, A. E. (2003). Extending production models to include process error in the population dynamics. *Canadian Journal of Fisheries and Aquatic Sciences*, 60(10):1217–1228.

Quinn, T. J., Deriso, R. B., and Neal, P. (1990). Migratory catch-age analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, 47(12):2315–2327.

Ricker, W. E. (1954). Stock and recruitment. *Journal of the Fisheries Board of Canada*, 11(5):559–623.

Rindorf, A. and Lewy, P. (2001). Analyses of length and age distributions using continuation-ratio logits. *Canadian Journal of Fisheries & Aquatic Sciences*, 58(6):1141–1152.

Schaefer, M. B. (1954). *Some aspects of the dynamics of populations important to the management of the commercial marine fisheries*. Inter-American Tropical Tuna Commission.

Shepherd, J. (1999). Extended survivors analysis: An improved method for the analysis of catch-at-age data and abundance indices. *ICES Journal of Marine Science: Journal du Conseil*, 56(5):584–591.

Shono, H. (2008). Application of the Tweedie distribution to zero-catch data in CPUE analysis. *Fisheries Research*, 93(1-2):154–162.

Skaug, H. and Fournier, D. (2006). Automatic approximation of the marginal likelihood in non-gaussian hierarchical models. *Comput. Stat. Data An.*, 51(2):699–709.

Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003). WinBUGS user manual. *Version 1.4. MRC Biostatistics Unit, Cambridge, UK*.

Stari, T., Preedy, K. F., McKenzie, E., Gurney, W. S., Heath, M. R., Kunzlik, P. A., and Speirs, D. C. (2010). Smooth age length keys: Observations and implications for data collection on north sea haddock. *Fisheries Research*, 105(1):2–12.

Stefansson, G. (1996). Analysis of groundfish survey abundance data: combining the GLM and delta approaches. *ICES Journal of Marine Science*, 53(3):577–588.

Tweedie, M. C. K. (1984). An index which distinguishes between some important exponential families. In Ghosh, J. K. and Roy, J., editors, *Statistics: Applications and New Directions*, pages 579–604. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference, Calcutta: Indian Statistical Institute.

Ulltang, Ø. (1977). Sources of errors in and limitations of virtual population analysis (cohort analysis). *Journal du Conseil*, 37(3):249–260.

United Nations (1982). United Nations Convention on the Law of the Sea. `http://www.un.org/Depts/los/convention_agreements/texts/unclos/unclos_e.pdf`.

Walters, C. and Punt, A. (1994). Placing Odds on Sustainable Catch Using Virtual Population Analysis and Survey Data. *Canadian Journal of Fisheries and Aquatic Sciences*, 51(4):946–958.

Wolfinger, R. and Xihong, L. (1997). Two Taylor-series approximation methods for nonlinear mixed models. *Comput. Stat. Data An.*, 25(4):465–490.

Wood, S. (2006). *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC.

Appendix A

DATRAS-package

The DATRAS package (Kristensen and Berg, 2012) for the `R` programming language contains functions for reading and manipulating trawl survey data from the DATRAS database (datras.ices.dk). This database contains scientific survey data covering "The Baltic Sea, Skagerrak, Kattegat, North Sea, English Channel, Celtic Sea, Irish Sea, Bay of Biscay and the eastern Atlantic from the Shetlands to Gibraltar." and "Up to 45 years of continuous time series data". The raw "exchange" format is very comprehensive with detailed data about haul position, gear types and experimental conditions, numbers-at-length by species, sub-sampled fish with age determination data and many at times intricate as well as crucial details and codes for different reporting/sub-sampling schemes. Handling and interpreting data from DATRAS correctly from scratch takes a significant amount of effort and time, but this R package can reduce much of this workload to a few lines of code. The raw exchange format can be read into a `DATRASraw` object in `R` using the package. These data objects contain three components

1. Age data - one vector per individual fish
2. Hydro data - one vector per haul, position and experimental conditions.
3. Length data - numbers per length group by haul and species

One particular useful function in the DATRAS package is the `subset()` function. It allows subsetting over all three components at once, without the need for specifying for which component(s) the subset clauses apply, because the function will look for the variable names in all components and apply the clauses where appropriate. This is best illustrated by an example:

```
require(DATRAS)

## Example file
zipfile <- system.file("exchange","Exchange1.zip",package="DATRAS")

## Step 1. Read exchange data into R
d <- readExchange(zipfile)

## Take subset
d <- subset(d,lon>10,Species=="Gadus morhua")
```

The above commands will read the raw data into R and take subset over all three components leaving only data for Atlantic Cod (Gadus morhua) and excluding all data from hauls with a position east of 10 degrees longitude. In fact the "Species" variable is only present in component 1 and 3, whereas "lon" is only available in component 2, but the user need not worry about this, since the function will automatically perform the subsetting over the correct components. In addition the function will also remove empty factor levels after subsetting and ensure that these are consistent across all components.

The individuals components can be accessed by the list operator, e.g. the second component of `d` in the above example is accessed directly using `d[[2]]`. The special `$` operator in `R` has been redefined to work seamlessly with `DATRASraw` objects such that `d[[2]]$haul.id` is identical to `d$haul.id`.

Size spectra on haul level are conveniently analysed using the `addSpectrum()` function, which adds the numbers caught per length group (cm) in the variable `N` on component 2, an example:

```
> d<-addSpectrum(d)
> head(d$N[,10:15])
                          sizeGroup
haul.id                    [13,14) [14,15) [15,16) [16,17) [17,18) [18,19)
  2009:1:SWE:ARG:GOV:82:20       0       0       0       5       1       2
  2009:1:SWE:ARG:GOV:81:19       1       1       7       4       6       2
  2009:1:SWE:ARG:GOV:80:18       2       3       0       0       0       1
  2009:1:SWE:ARG:GOV:75:17       0       0       0       0       0       0
  2009:1:SWE:ARG:GOV:74:16       0       0       0       0       0       1
  2009:1:SWE:ARG:GOV:73:15       0       0       0       2       2       0
```

Smooth age-length keys can be also easily be fitted using the methodology in Paper III, and used to predict the number of individuals caught by age group rather than length group:

```
## fit age-length key and predict numbers-at-age
alk <- fitALK(d,minAge=1,maxAge=4)
d$Nage = predict(alk)
head(d$Nage)

## plot age length key (smooth and raw proportions by length group)
plotALKfit(alk,row=1)
plotALKraw(d,minAge=1,maxAge=4,add=TRUE)
```

The resulting plot is shown in figure A.1.

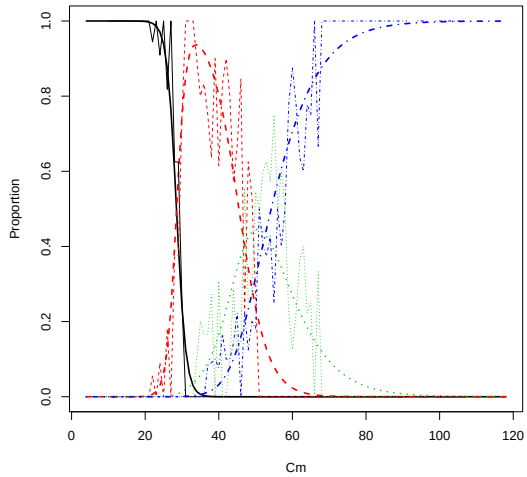


Figure A.1: Example of a raw (thin lines) and smooth (thick lines) age-length key.

Some additional useful functions for `DATRASraw` objects are `print(), plot(), summary()`, sample unit subset (e.g. `d[1:2]`), `c()` for combining data sets, and `as.data.frame()`.

The entire package including source code is available online at www.rforge.net/DATRAS.

Appendix B

Stockassessment.org

There are many stakeholders involved when it comes to the management of marine resources, and since stock assessments provide the basis for the scientific advice for managers, it is important that all the materials and methodology is open for peer review and that the results are reproducible by others. In practice however, it is often the case that only one or two scientist are able to carry out the assessment, because getting all the data and configuring a computer to reproduce the results takes quite a bit of work. This puts a tight barrier on how many alternative model/data configurations that can be explored during the review process, as this is limited by the number of persons who are able to do this. In 2009 Anders Nielsen and I made the first prototype of a 100% web-based tool for stock assessment under the domain www.stockassessment.org. It provided the users with a personal account from which they could configure, run and compare assessments with no other requirements than an internet connection and a browser. It quickly became apparent to us, that this tool was very useful and needed as it became the primary tool for many official stock assessments. During this Ph.D. this tool was continuously developed, and in the summer of 2012 the International Council for Exploration of Sea sponsored a major upgrade of this system and new server. Due to the great impact of this new tool and the fact the development of it has benefitted greatly from this Ph.D-project, I feel a short description is warranted in this appendix.

B.1 Basic Design

An assessment at stockassessment.org is organized into several folders, but the 4 essential ones each has its own page on the site: "Data", "Configuration", "Source Code", and "Results". The "Data", "Configuration", and "Source Code" folders contain plain text files, which are editable from the homepage. The "Results" contains the output from an assessment in shape of HTML-tables and picture files with graphs. These results are generated by an (editable) R-file called "plotscript.R" in the "Source Code" folder. The dependencies between data, configuration, source code and results are handled by the "Make" tool with a so-called Makefile, which is part of the source code. For instance, it is specified in the Makefile that the output graphs only depend on the plotting scripts and the model being run. This means that if e.g. only the plotscript.R file is modified since the current results were created, the Make tool knows that we do not need to estimate model parameters again, but only to update the plots by re-running the plotting script. Also, the Make tool can automatically run commands in parallel (using the `-j` option), such that the run time for e.g. leave-one-out analyses can be greatly reduced.

B.2 Multiple Users and Version Control

Each user has his/her own account, and all changes made by a user are local changes to that account. This effectively prevents any loss of work that has been saved. However, all files are stored in two locations: 1) in a local folder on the server where all changes are applied, and 2) in a version control repository (Subversion). Once a user is happy with some changes he made, he can submit the updated assessment to version control system, which will then track the changes made and save the updated version. This feature also facilitates the possibility of multiple users working on the same assessment, since locally modified files or entire assessment can always be synchronized with the latest version in the repository.

B.3 Creating New Assessments

All new assessments are based on a common core of source code, which is also stored in the repository. The main job in creating a new assessment is thus to set up the data and configuration files. Stockassessment.org has a "Data wizard" (figure B.1) with step-by-step uploading and validation of data files. Once all data files are present and validated, the system can generate a default configuration, which allows a first run to be made.

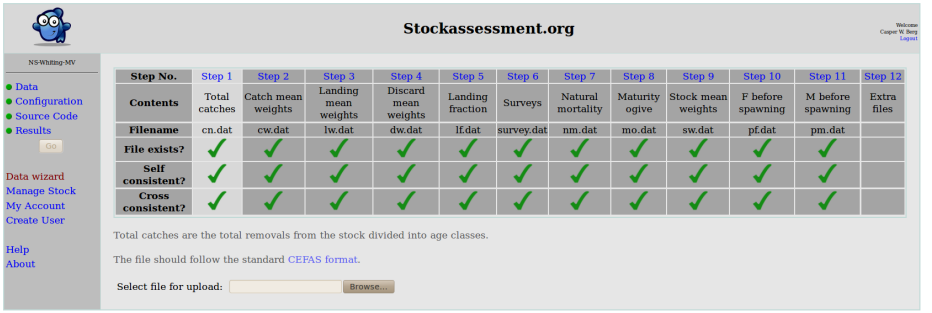


Figure B.1: Stockassessment.org data wizard screen shot.

B.4 Code editing

A unique and useful feature of stockassessment.org is the ability to alter, compile and run every bit of code associated with an assessment directly from a web-browser. This relieves the user from spending potentially many hours on setting up the complete list of tools needed on his/her own computer. However, allowing users to enter and run computer code on the server directly from the browser is associated with serious security problems, since if not handled correctly a user may e.g. (intentionally or not) wipe the discs, infect it with malware and so on. To prevent such issues each user are provided a secure shell (SSH) account on the server, and all executing of editable code is sent through this connection. This prevents users from writing to the server discs anywhere else but on their own account, and permits limitation of CPU resources on user level.

B.5 Comparing Runs

Like with any statistical analysis, the process of creating an assessment is iterative, and comparing two analyses with different configurations is thus a core job for the investigator. There is therefore a built-in functionality for comparing two runs graphically and formally through likelihood ratio tests. The graphical comparison feature is illustrated in figure B.2.

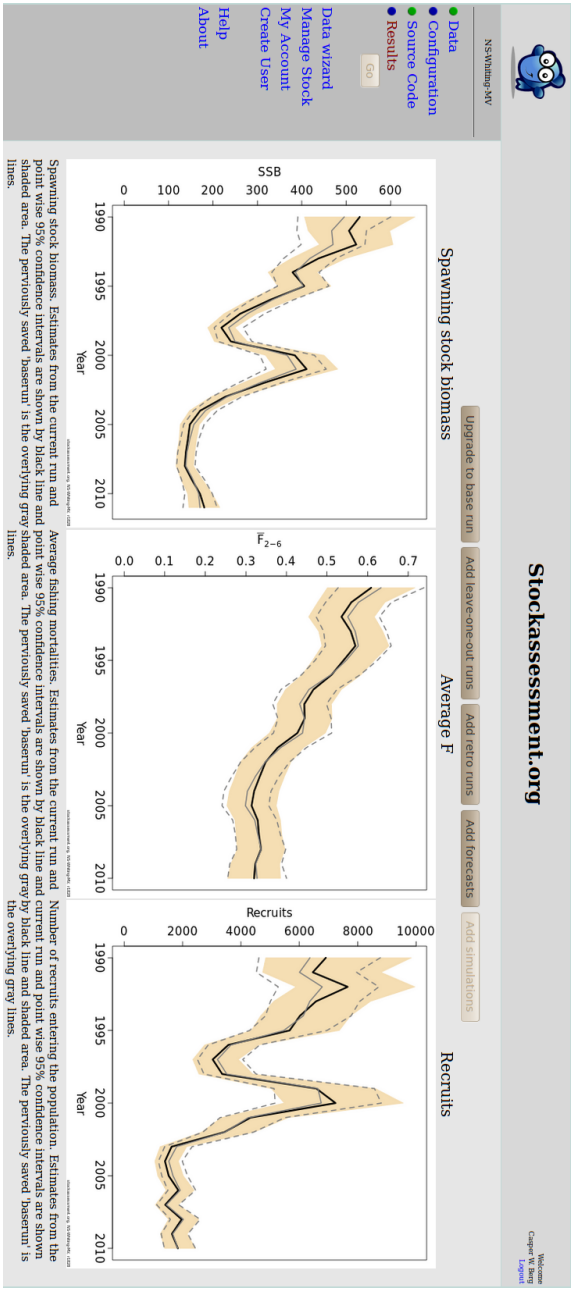


Figure B.2: Stockassessment.org results page screen shot.

Appendix C

Paper I

Estimation of feeding patterns for piscivorous fish using individual prey data from stomach contents

Casper Willestofte Berg and Axel Temming

Abstract: The problem of estimating temporal feeding patterns using stomach data is considered, where the time of ingestion for each prey item can be predicted through a gastric evacuation model. The arrival of prey is modelled as a nonhomogeneous Poisson process with known periodic intensity. A maximum likelihood approach is used to estimate the intensity, which is assumed to be the same for all predators, incorporating different uncertainties for the arrival time of each prey item. The method is applied to a case where a population of whiting (*Merlangius merlangus*) is feeding on sandeel (*Ammodytes marinus*), and peak feeding periods around dusk and dawn are identified.

Résumé : Nous examinons le problème de l'estimation des patrons temporels d'alimentation à l'aide de données de contenus stomacaux, lorsque le moment de l'ingestion de chaque proie individuelle peut être prédit à l'aide d'un modèle d'évacuation gastrique. L'arrivée des proies est modélisée comme un processus de Poisson non homogène avec une intensité périodique connue. Nous utilisons une méthode de vraisemblance maximale pour estimer l'intensité que nous présupposons être la même pour tous les prédateurs et nous incorporons différentes incertitudes concernant le moment d'arrivée de chaque proie individuelle. Nous proposons un exemple du modèle dans lequel une population de merlans (*Merlangius merlangus*) se nourrit de lançons nordiques (*Ammodytes marinus*) avec des maximums d'alimentation aux environs du coucher et du lever du soleil.

[Traduit par la Rédaction]

Introduction

Knowledge about the interactions between predatory fishes and their prey is important for understanding and modelling marine ecosystems. These interactions are rarely observed directly. Instead, many studies on feeding behaviour have been conducted using information derived from the stomach contents of sampled fish from the wild. The objective of these studies is often estimation of food consumption rates or feeding preferences or to obtain information about temporal patterns in feeding behaviour. Consumption rates and preferences can for instance be utilized in multispecies fisheries models, and patterns in feeding behaviour may reveal insights about predator–prey interactions and aid sampling design.

Analyses of consumption rates and feeding cycles are often based on average stomach contents (mass) (e.g., Sainsbury 1986; Adlerstein and Welleman 2000). In Adlerstein and Welleman (2000), a generalized additive model (GAM) was used to model the average stomach contents with time of day as one of the explanatory variables. Magnússon and Aspelund (1997) modelled the prey frequency instead of the mass, but the possibility of a diurnal feeding pattern is dismissed after fitting a GAM on average stomach mass with hour of sampling included. GAMs or generalized linear models (GLMs) have also been used to model the probability that a stomach contains a particular prey group (e.g., Stefansson and Palsson 1997 and Rindorf 2003). Both approaches required sampling in appropriate intervals throughout the (diel) cycle and provide only information about the distribution in the chosen sampling points. Hall et al. (1995) combined total stomach contents with a digestion model in a dynamic model, such that both the feeding intensity and the meal size distribution were estimated. The cost of this extra complexity is that the diel pattern is restricted to be piecewise constant with a maximum of four "switch points".

An alternative to modelling the total stomach contents is to model the ingestion times of individual prey items, a method that was used in this study. This idea was introduced by Griffiths (1976), and it involves back-calculating the ingestion time of partially digested prey items found in the stomach using a gastric evacuation model (GEM). This method requires stomach data, where each individual prey item has been identified, weighed, and measured for length. This is a demanding task, and it reduces the number of available data sets substantially, but the extra information permits deeper levels of analysis. Johansen et al. (2004) estimated consumption rates using back-calculated ingestion times, but assuming that the feeding rate follows a simple Poisson process disregarding any diel variation. The errors on the back-calculated times were also ignored in Johansen et al. (2004). The goal in this paper is to estimate diel feeding patterns from back-calculated ingestion times while taking their error distributions into account.

Andersen and Beyer (2007) showed that previously used

Received 20 May 2010. Accepted 17 February 2011. Published at www.nrcresearchpress.com/cjfas on 18 May 2011.
J21820

Paper handled by Associate Editor Stephen Smith.

C.W. Berg. National Institute of Aquatic Resources, Technical University of Denmark, Charlottenlund Castle, 2920 Charlottenlund, Denmark.
A. Temming. Institute for Hydrobiology and Fisheries Science, University of Hamburg, Olbersweg 24, 22767 Hamburg, Germany.

Corresponding author: C.W. Berg (e-mail: cbe@aqua.dtu.dk).

(purely mass-dependent) GEMs are biased in case of multiple meals present in the stomach. They developed a new surface-dependent model using mechanistic arguments that correctly predicts evacuation times in a wide range of experimental settings, including multiple meals. In this GEM, each food item was represented by a cylinder, whose exposed sides gradually peeled off while the length was preserved. Andersen and Beyer (2008*a*) derived the error distribution of the back-calculated ingestion times from the cylinder model and found it to be approximately Gaussian. Because the errors on these ingestion times can be substantial and of varying magnitude between prey items, it is important to take these errors into account. The method developed in this paper will weight each prey item according to the precision of its back-calculated ingestion time using the result in Andersen and Beyer (2008*a*) when estimating the diel feeding pattern. Also, we utilize the information that each prey item must have been ingested before the predator was caught by considering a truncated normal distribution.

Diel feeding patterns are investigated by letting the feeding rate follow a nonhomogeneous Poisson process (NHPP) with periodic intensity. The NHPP is a stochastic process, where the (feeding) intensity is allowed to change as a function of time, unlike the normal Poisson process for which the intensity is constant through time. The error distributions are accounted for by using mixed models and the method of maximum likelihood. The statistical framework allows us to formally test whether a population-level diel feeding pattern is present and to quantify the uncertainties on the estimated intensity curve.

The power of this new method will be exemplified by re-analyzing a data set that has previously been treated in Mergardt and Temming (1997) and Temming and Mergardt (2002). In these two studies, stomach data from a population of whiting (*Merlangius merlangus*) feeding on sandeel (*Ammodytes marinus*) were analysed. The authors noted that usual methods using average stomach contents fail to detect temporal feeding patterns in this case, where food particles are large but few and digestion is relatively slow. They expected to find feeding peaks at dusk and dawn because of documented behaviours of the two species. The authors could not confirm this hypothesis, but noted that it is unlikely to separate two peaks (by inspection of histograms) because of the noise on the hindcasted ingestion times. Our new analysis of this data set confirms the original hypothesis of a bimodal feeding pattern. Although this case involves only one prey type, it is straightforward to analyse cases involving different prey types using this technique.

Materials and methods

Gastric evacuation model (GEM)

Ingestion times are not directly observed but instead they are back-calculated from partly digested prey items using a GEM. The GEM used in this study was developed by Andersen and Beyer (2005*a* and 2005*b*). In case of a single prey item, the model is

$$\frac{\mathrm{d}S_t}{\mathrm{d}t} = -\rho\sqrt{S_t}; \quad \rho = \rho_{LHE}L^{\eta}\exp(\delta H)E^{-\gamma}$$

where S_t is the remaining prey mass at time t, and ρ is the rate parameter that depends on a prey-specific basic resistance to digestion ρ_{LHE}, predator length L, temperature H, and prey energy density E. The remaining parameters were estimated in Andersen (2001): $\eta = 1.44$, $\delta = 0.078$, and $\gamma = 0.86$. The prey is assumed to be homogeneous, such that ρ is constant in time. In case of multiple meals in the stomach, it is assumed that equal fractions of the surface area of the prey items are exposed to digestion. The evacuation rate of prey i can be written as a function of its own mass $S_{i,t}$ and its length l_i (standard length) (see for instance Andersen and Beyer 2008*a*):

$$\frac{\mathrm{d}S_{i,t}}{\mathrm{d}t} = -\rho_{i,t}d_{i,t}\sqrt{S_{i,t}}$$
$$d_{i,t} = \sqrt{S_t l_i}\left(\sum\sqrt{S_t l_i}\right)^{-1}$$

There is no general analytical solution to $S_{i,t}$, so instead first-order numerical integration can be used for projecting the masses of each prey back in time until their original mass is reached. Because suitable parts of prey lengths are preserved for a longer period in the stomach, an estimate of the original mass $\widehat{w}$ can be obtained from known length–mass relationships. It is assumed that the error is normally distributed with a constant coefficient of variation (CV):

$$\sqrt{w} \sim N\left[\sqrt{\widehat{w}}, \left(\sigma_m\sqrt{\widehat{w}}\right)^2\right]$$

Furthermore, it is assumed that there is predator-specific variation in the rate parameter ρ:

$$(1) \qquad \rho = \widehat{\rho}(1+\epsilon); \quad \epsilon \sim N(0, \sigma_e^2)$$

Andersen and Beyer (2008*a*) found that the distribution of the time of ingestion for the ith latest prey τ_i is approximately normal with variance

$$\mathbb{V}(\tau_i|S_{i,t_s}) \approx (\sigma_e\widehat{\tau}_i)^2 + \left[2(\widehat{\rho}_{i,t_i}d_{i,t_i})^{-1}\sigma_m\sqrt{\widehat{w}}\right]^2$$

Andersen and Beyer (2008*a*) provided the general estimates 0.1 for σ_e and 0.03 for σ_m. In this work, we will truncate the above normal distribution at the point in time where the predator was caught. This is done to account for the fact that any prey item must have been eaten by the predator before it was caught.

Likelihood function for NHPPs

We will now consider how to estimate a common time-varying feeding rate function, λ (the NHPP), for all predators, given that we have estimated the distribution of the ingestion time for each prey item observed using the model in the previous section.

Let us first recap the likelihood function for NHPPs when the arrival times (here, ingestion times) are known without error (see e.g., Alizadeh et al. 2008): An arrival process is observed for time interval $[t_{\text{start}}, t_{\text{end}}]$; a vector of arrival times $\boldsymbol{t} = (t_1, \ldots, t_n)$ is recorded ($t_{\text{start}} \le t_1 < \ldots < t_n \le t_{\text{end}}$), and we wish to estimate an unknown arrival rate function $\lambda(\cdot)$. For the moment, assume that some suitable parametrization has been chosen for λ.

Given $\boldsymbol{t}$ and an arbitrary (nonnegative) choice of λ, the joint probability density can be written

$$f(\boldsymbol{t}, \lambda) = \exp\Big(- \int_{t_n}^{t_{\text{end}}} \lambda(t)\,\mathrm{d}t\Big) \prod_{i=1}^{n} \lambda(t_i) \times \exp\Big(- \int_{t_{i-1}}^{t_i} \lambda(t)\,\mathrm{d}t\Big)$$

where we define $t_0 = t_{\text{start}}$. The log-likelihood function l is then

$$(2) \qquad l(\boldsymbol{t}, \lambda) = \sum_{i=1}^{n} \log \lambda(t_i) - \int_{t_{\text{start}}}^{t_{\text{end}}} \lambda(t)\,\mathrm{d}t$$

In this study, the arrival times are not directly observed, but instead we have some noisy observation $T_i = t_i + \varepsilon_i$. This leads to a nonlinear mixed effects model due to the random effect ε_i. Here, it is assumed that T_i follows a truncated normal distribution $T_i \sim N(T_i, \sigma_i^2, t_{\text{start}}, t_{\text{end}})$ with known σ_i for all i. Given λ, the data ($\mathbf{X} = \boldsymbol{T}$, $\boldsymbol{\sigma}$, t_{end}, t_{start}), and the true arrival times $\boldsymbol{s}$, the joint log-likelihood becomes

$$l(\mathbf{X}, \boldsymbol{s}, \lambda) = \sum_{i=1}^{n} \log \lambda(s_i) + \log \frac{1}{c_i} \phi\left(\frac{s_i - T_i}{\sigma_i}\right) - \int_{t_{\text{start}}}^{t_{\text{end}}} \lambda(t)\,\mathrm{d}t$$

where ϕ is the standard normal probability density function, and $c_i = \sigma_i \left[\Phi\left(\frac{t_{\text{end}} - T_i}{\sigma_i}\right) - \Phi\left(\frac{t_{\text{start}} - T_i}{\sigma_i}\right)\right]$ is a normalization constant due to the truncated normal distribution.

Now consider the situation where we have observed noisy arrival times for p independent predators in p time intervals $[t_{\text{start}}^k, t_{\text{end}}^k]$, $k = 1 \ldots p$, and assume that all predators consume prey arriving from the same nonhomogeneous Poisson process, $\lambda(\cdot)$. Let $\boldsymbol{T} = (\boldsymbol{T}^1, \ldots, \boldsymbol{T}^p)$ and $\boldsymbol{s} = (\boldsymbol{s}^1, \ldots, \boldsymbol{s}^p)$ denote the observed and true arrival times, respectively, for each of the p predators, and let n_k be the number of arrivals observed for the kth predator. The log-likelihood is then simply obtained by summing over all predators

$$l(\mathbf{X}, \boldsymbol{s}, \lambda) = \sum_{k=1}^{p} \left[\sum_{i=1}^{n_k} \log \lambda(s_i^k) + \log \frac{1}{c_i} \phi\left(\frac{s_i^k - T_i^k}{\sigma_i^k}\right) - \int_{t_{\text{start}}^k}^{t_{\text{end}}^k} \lambda(t)\,\mathrm{d}t \right]$$

By integrating out the latent true arrival times $\boldsymbol{s}$ in the likelihood and taking logarithms, we obtain

$$(3) \qquad l(\mathbf{X}, \lambda) = \sum_{k=1}^{p} \left[\sum_{i=1}^{n_k} \log \int_{t_{\text{start}}^k}^{t_{\text{end}}^k} \lambda(t) \frac{1}{c_i} \phi\left(\frac{t - T_i^k}{\sigma_i^k}\right) \mathrm{d}t - \int_{t_{\text{start}}^k}^{t_{\text{end}}^k} \lambda(t)\,\mathrm{d}t \right]$$

Note, that it is also possible to consider two nested levels of random effects: a predator effect on the digestion rate that affects all ingestion times from the same predator and a prey effect due to the prediction error on the original mass (this is described in Supplemental Appendix S1, available online[1]). We refrain from separating error terms because of the much greater numerical complexity and because it is likely that little will be gained from this, since each predator has eaten only a few food items in the case considered.

Observation interval

Unfortunately, we do not have precise intervals $[t_{\text{start}}^k, t_{\text{end}}^k]$ for which we can say that we have observed all arrivals, even if we take noisy arrival times into account. We effectively stopped to observe the predator when the digestion process stopped. There are indications that a caught fish stops digesting once in the trawl because of their struggling behaviour (Farrell et al. 2001). We therefore chose t_{end} as the midpoint of the time between release and withdrawal of the trawl. The associated uncertainty is ignored, as the time between release and withdrawal is relatively short (30 min) for the case considered.

The beginning of the observation interval, t_{start}, is not as straightforward to quantify. In Johansen et al. (2004), a similar variable is used ($t_{\max}$). This is interpreted as the upper time limit that a prey item (Atlantic herring, *Clupea harengus*) of a given length group can be digested in the stomach of a cod of a given mass at a given temperature, such that the total length is still measurable. Predation rate is then calculated as the number of observed prey with an estimated digestion time less than $t_{\max}$ divided by $t_{\max}$. The authors find $t_{\max}$ by inspection of the cumulative frequency of measurable herring as a function of estimated digestion time. This should be a straight line until the time when prey become immeasurable, assuming uniform distribution of the digestive stages of the prey. It is equivalent to increasing the length of a common observation interval for all predators until the mean predation rate starts to drop, because prey are digested beyond recognition within the observation window. The assumption of uniform digestive stages is easily violated in case of any patterns in feeding behaviour. One could also argue that $t_{\max}$ not only depends on the prey length, predator size, and temperature, but also on the stomach contents of the given predator — for an empty stomach, $t_{\max}$ should be lower than for a full stomach, because the total evacuation time is longer in the full stomach. One must however be careful using observation intervals that depend on observed stomach contents, as it will introduce bias if the difference between full and empty stomachs is not weighted correctly.

To circumvent the problems associated with choosing a suitable interval, we introduce a "smooth" horizon, ψ, which can be interpreted as the probability of observing a prey item of random size in the stomach $t_{\text{end}}^k - t$ hours after it has been ingested. A simulation study was performed to find a suitable parametrization of ψ. It was found that a sigmoid function whose "center" ($\psi = 0.5$) is inversely proportional to the (length- and temperature-specific) digestion rate ρ provided a reasonable fit:

$$\psi'(t, t_{\text{end}}^k, \rho^k, \alpha, \beta) = \psi'_k(t) = \frac{1}{1 + \exp\left[\alpha\left(t_{\text{end}}^k - t - \frac{\beta}{\rho^k}\right)\right]}$$

where $\alpha > 0$.

[1]Supplementary data are available with the article through the journal Web site (http://www.nrcresearchpress.com/cjfas).

We want to ensure that ψ attains the value of 1 at the time of catch, so we divide by the function evaluated at that point:

$$\psi_k(t) = \frac{\psi'_k(t)}{\psi'_k(t_{\text{end}})} \tag{4}$$

Now, we replace $\lambda(t)$ by $\psi_k(t)\lambda(t)$ in the likelihoods (eqs. 2 and 3) and integrate from minus infinity. For eq. 2 that ignores measurement errors, this gives

$$l(\mathbf{X}, \boldsymbol{\theta}) = \sum_{i=1}^{n} \log \psi_k(t_i)\lambda(t_i) - \int_{-\infty}^{t_{\text{end}}} \psi_k(t)\lambda(t)\,\mathrm{d}t \tag{5}$$

and for the mixed model that takes the errors into account (eq. 3), we obtain

$$l(\mathbf{X}, \boldsymbol{\theta}) = \sum_{k=1}^{p} \left[\sum_{i=1}^{n_k} \log \int_{-\infty}^{t^k_{\text{end}}} \psi_k(t)\lambda(t) \frac{1}{c_i} \phi\left(\frac{t - T_i^k}{\sigma_i^k}\right) \mathrm{d}t - \int_{-\infty}^{t^k_{\text{end}}} \psi_k(t)\lambda(t)\,\mathrm{d}t \right] \tag{6}$$

where $\boldsymbol{\theta} = (\lambda, \alpha, \beta)$ is the vector of parameters ($\alpha, \beta \in \mathbb{R}$ and λ is a function). Using the smooth horizon corresponds to an expectation of the distribution of evacuation times having the form of a dampened version of λ.

Prey items with unpredictable ingestion times

It may be that the time of ingestion cannot be determined for some prey items. For example, the assumption that equal fractions of the surface area of prey items are exposed to digestion may become problematic when many prey items are in the stomach together, as some might be shielded from digestion by others (10 prey items could be used as a reasonable threshold). A solution to that problem is to consider only the number of prey found in stomachs (n_k), where $n_k > 10$, but not the arrival times. The probability of such an observation is

$$\mathbb{P}(n_k, t^k_{\text{end}}, \rho^k, \alpha, \beta, \lambda) = \frac{1}{n_k!} \left(\int_{-\infty}^{t^k_{\text{end}}} \psi_k(t)\lambda(t)\,\mathrm{d}t \right)^{n_k} \times \exp\left[- \int_{-\infty}^{t^k_{\text{end}}} \psi_k(t)\lambda(t)\,\mathrm{d}t \right]$$

which yields the following contribution to the log-likelihood (for one predator):

$$l\left(n_k, t^k_{\text{end}}, \rho^k, \alpha, \beta, \lambda\right) = n_k \log\left(\int_{-\infty}^{t^k_{\text{end}}} \psi_k(t)\lambda(t)\,\mathrm{d}t \right) - \log(n_k!) - \int_{-\infty}^{t^k_{\text{end}}} \psi_k(t)\lambda(t)\,\mathrm{d}t$$

This is equivalent to the situation where we let the variances of the arrival times go to infinity in eq. 6, so unpredictable arrival times can therefore in practice be handled by setting the variances of these untrustworthy arrival times to a suitable large number.

Parametrization of arrival rate

The simplest choice would be to let λ be piecewise constant. As noted in Alizadeh et al. (2008), such functions are appropriate when discrete events affect all or a major subset of the population at once. The rise and set of the sun are not discrete events, although it might be a reasonable approximation to consider them as such. A smooth curve is, on the other hand, more visually appealing, and it permits the powerful technique of automatic differentiation for maximizing the likelihood.

The feeding process of a population is very often assumed to be influenced by sunrise and sunset and (or) perhaps the tide or other periodic events, which restricts our attention to periodic functions.

We need to ensure that the arrival rate is nonnegative over the entire interval considered, and when defined as a periodic function, that amounts to everywhere on the real axis. There are several possible ways of enforcing nonnegativity. One simple way is to use an exponentiated function, a solution employed in Kuhl et al. (1997), where λ has the form

$$\lambda(t; \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\omega}, \boldsymbol{\phi}) = \exp\left[\sum_{i=1}^{m} \alpha_i t^i + \sum_{j=1}^{p} \gamma_j \sin(\omega_j t + \phi_j) \right] \tag{7}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^m$ and $\boldsymbol{\gamma}, \boldsymbol{\omega}, \boldsymbol{\phi} \in \mathbb{R}^p$. This parametrization allows a general trend over time represented by the m-degree polynomial as well as multiple (unknown) periodicities represented by p trigonometric functions. For our purpose, we can disregard the possibility of a trend over time. Unlike Kuhl et al. (1997), we also assume that the period Q is known, because recurrent feeding patterns are most likely to be diurnal. In this case, eq. 7 is just a reparametrized Fourier series.

In this study, nonnegativity is enforced by squaring the Fourier series instead of exponentiating:

$$\lambda(t; \boldsymbol{a}, \boldsymbol{b}) = \left[\frac{1}{2} a_0 + \sum_{i=1}^{N} a_i \cos\left(\frac{2\pi t i}{Q}\right) + b_i \sin\left(\frac{2\pi t i}{Q}\right) \right]^2 \tag{8}$$

Squaring is fairly similar to exponentiating as long as the variations in λ are limited and the function we wish to approximate is not close to zero. However, zero exists only as the limit for the exponential function, which might lead to numerical problems for cases with a high proportion of empty stomachs. Also, it is possible to write the antiderivative in closed form for the squared Fourier, but not for the exponentiated Fourier. Likelihood ratio tests can be used to decide whether adding an extra harmonic gives a significant better fit.

Daily ration

Let the daily ration be the expected number of prey consumed per day. Given the arrival rate $\lambda(t)$, the expected number of arrivals N over a period of 24 h is

$$\mu \equiv \mathbb{E}(N) = \int_0^{24} \lambda(t)\,\mathrm{d}t$$

Numerical methods

The negative log-likelihood (eq. 6) is minimized using the open source Automatic Differentiation Model Builder (ADMB) software (http://www.admb-project.org). The inte-

gration in eq. 6, $\int_a^b f(x)\,dx$, is carried out with Romberg's method (using Bell 2005), which has error $O[(b-a)/2^{n-1}]^{2(p+1)}$. The values $n = 9$ and $p = 2$ were used, which were the lowest possible numbers not leading to any noteworthy visual difference in the estimated intensity curves. Minus infinity is replaced with the 1% fractile in eq. 4 for a given set of parameters. All standard deviations and confidence limits reported are found through the estimated Fischer information matrix. The ADMB code can be obtained by contacting the corresponding author.

Simulation

To test the method's ability to reconstruct feeding intensity curves, a number of simulated data sets were created. Two cases based on real data were simulated: cod feeding on sandeel and whiting feeding on sandeel. The former setting involves medium-sized cod (28–64 cm) eating a sandeel at a relatively high frequency (2.4 per day) and where digestion happens faster because of the larger predators at higher temperature (14 °C). The latter case was set to resemble the case study described in the Introduction, where evacuation is slower leading to higher uncertainties on the hindcasted ingestion times. Both sets of parameters were chosen such that two distinct modes were present (details on how the simulations were carried out can be found in Supplemental Appendix S2[1]). The number of stomachs for each simulation was set to 150 in the cod case (900 prey items on average) and 200 in the whiting case (600 prey items on average).

The potential bias introduced by unobserved prey evacuated prior to the time of sampling is discussed in Andersen and Beyer (2008*b*), and a bias of –3% to –22% was found to cover most situations for North Sea gadoids, but can be more or less extreme depending on the actual setting. The most important factor, however, is the time between ingestion and sampling (i.e., the time spent in the stomach). The amount of bias introduced by unobserved prey was examined by plotting the predicted ingestion times versus the truth for simulated data in a setting similar to the case study (the result is shown in Supplemental Fig. S1[1]). A second-order polynomial (with no intercept) was found to provide a good fit, and it is also shown. It is seen that the bias is relatively low for this particular case. To evaluate if the bias due to unobserved prey is problematic for our estimation problem, we need to consider the estimated smooth horizon also.

Performance measures

For the simulated data sets, we can measure the error in terms of deviation of the estimated arrival rate from the true arrival rate. This error can be compared with the same measure calculated for the simple maximum likelihood estimate (eq. 5).

Here we use a subset of the performance measures formulated in Johnson et al. (1994). These are based on the absolute errors for a series of K replications of the same simulation experiment ($m = 1, \ldots, K$). Let $\widehat{\lambda}_m(t)$ denote the estimated rate function for the mth replication. The average absolute error is then given by

$$\delta_k \equiv \frac{1}{Q}\int_0^Q \left|\widehat{\lambda}_m(t) - \lambda(t)\right| dt$$

The sample mean of errors from the K replications is denoted $\bar{\delta}$ and the corresponding CV is found in Johnson et al. (1994) as

$$V_{\bar{\delta}} = \left[\frac{1}{K-1}\sum_{m=1}^{K}(\delta_m - \bar{\delta})^2\right]^{1/2}\bar{\delta}^{-1}$$

Case study: whiting (*Merlangius merlangus*) feeding on sandeel (*Ammodytes marinus*)

In this section, we apply our methodology on stomach data from 690 whiting caught in the North Sea. This data set has previously been analysed in Mergardt and Temming (1997) and Temming and Mergardt (2002) and is thoroughly described in the former, so only a short description is provided here. The whiting were caught within a 10 × 10 nautical mile square (1 nautical mile = 1.852 km) over a period of 3 days (30 May to 1 June 1992). The sea temperature was 7.9 °C. Only one size group of whiting (25–29.9 cm) was sampled. The whiting population had been feeding almost exclusively on sandeel. All material other than identifiable sandeel was weighed for each stomach also, and we know that a high proportion of this category consisted of detached loose sandeel material. In the analysis, we therefore assume that all material in this category consists of detached loose sandeel. We need this assumption, because the precise contents of this category was not used in the original analyses of the data and was therefore not available. It is further assumed that this detached material stems from individuals other than the identified sandeel, or more precisely from a single unobserved sandeel. Alternatively, all (or some) of the detached material could stem from the identified sandeel, but this would lead to an unrealistic high proportion of individuals exceeding their predicted fresh mass, so this possibility is rejected. The number of sandeel found in a single stomach ranged between 0 and 4, which seems to fit the Poisson assumption (no immediate signs of overdispersion). The prediction of the mass of a sandeel at time of ingestion was found using the method and parameters from Mergardt and Temming (1997). The energy density for sandeel was set to vary linearly from 5.5 to 6.0 kJ·g^{-1} for sandeel of length 11–20 cm, as the energy density was found to increase with size in Pedersen and Hislop (2001). The basic resistance to digestion parameter, ρ_{LHE}, was set to 1.29×10^{-3} for sandeel, as found in Andersen (2001). 370 out of a total of 694 sandeels in the data set were too digested to measure the standard length. The length of these individuals was set to the mean length of the measurable ones (13.89), and their hindcasted ingestion times were left out of the subsequent analysis because of their excessive uncertainty.

Results

Simulations

Cod case

In the cod case, the errors were generally so small that it was possible to recognize the modes in histograms of the raw back-calculated ingestion times. The standard deviations ranged between 0.2 and 12 h with a mean of 4.5 h. It therefore made sense to fix the number of harmonics to $N = 2$ and

compare $\bar{\delta}$ for eqs. 5 and 6, which disregards and utilizes the associated uncertainties on the ingestion times, respectively. Convergence was achieved in all cases.

The performance measures for the simulated data sets are shown (Table 1). The average absolute error is seen to be halved when accounting for the noise. The corresponding CV, on the other hand, is substantially larger for eq. 6, but this was expected, as the noise flattens out the signal (i.e., disregarding the noise gives less variance but more bias). The estimated intensity curves and their associated confidence bounds were also inspected in the simulation experiment. The plots (see example in Fig. 1) revealed that the method is generally capable of reconstructing the generating curve from the noisy measurements and that the confidence bounds are reasonable.

Whiting case

For the whiting case, the uncertainties on the hindcasted arrivals were of such magnitude that the histograms of the raw back-calculated ingestion times generally appeared flat. Estimation of the arrival rate where the uncertainties were ignored and using $N = 2$ as in the cod case would therefore result in an almost constant curve, making it practically equivalent to $N = 0$, with which we will thus compare performance measures. Estimation was therefore performed using $N = 0\ldots2$, and likelihood ratio tests were used to determine how many harmonics that could be identified from the data (the resulting performance measures are shown in Supplemental Fig. S2[1]). All cases had a smaller average error than the constant curve, except one that is only slightly worse. In seven cases, a constant feeding rate could not be rejected, in which case the errors coincide. In 18 cases, a constant feeding rate was rejected in favour of using one harmonic, and in the last seven cases two harmonics were found to provide a significantly better fit. This suggests that the number of prey items observed in each simulation (about 600) is in the low end to be able to demonstrate bimodality in the feeding pattern for this particular case.

Case study

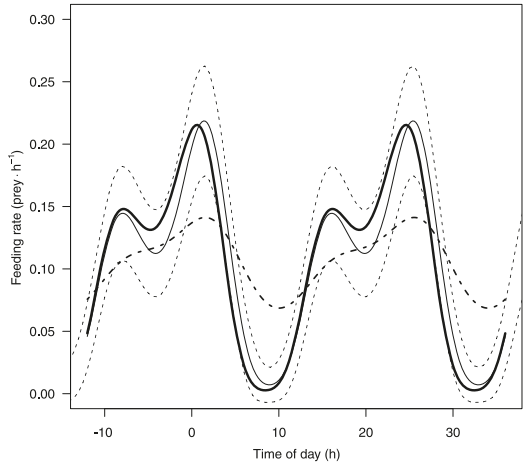
Mergardt and Temming (1997) have already noted that the mean mass of stomach contents revealed no clear pattern of diel food intake periodicity. Using the method of back-calculation (although a different GEM), inspection of histograms, and a simulation study that compared different scenarios with the observed frequencies in 2 h intervals, it was concluded that feeding peaked at night and had its minimum around midday. The present analysis shows some different results, which emphasizes the need for taking the errors into account.

The likelihood ratio tests for the number of harmonics (Table 2) show that both the homogeneous Poisson model and the one using a single harmonic is rejected in favour of using two harmonics, whereas using three does not lead to a significantly better fit. The estimated intensity curve using two harmonics is shown (Fig. 2). There are two clear peaks at dawn and dusk, which has a nice biological interpretation also mentioned in Mergardt and Temming (1997), namely that feeding occurs at the spatial overlap during dusk–dawn where whiting migrate away from the light and sandeel migrate towards it.

Table 1. Performance measures for 100 simulated data sets with cod feeding on sandeel.

Measure	Eq. 5	Eq. 6	Change
$\bar{\delta}$	0.037	0.019	–50%
V_δ	0.108	0.284	163%

Fig. 1. Example of a reconstructed feeding pattern from the simulation study (cod case). The true intensity (thick solid line), the estimated intensity (thin solid line), and the marginal 95% confidence interval (thin dashed lines) is shown. The simple maximum likelihood estimate (eq. 4) is also shown (thick dashed line).



The estimated probability of observing a sandeel in the stomach at time t since ingestion for a whiting of average size is shown (Supplemental Fig. S3[1]). It is seen that the probability is low after a couple of days, which means that the bias introduced by unobserved prey on sandeels with longer evacuation times is negligible (cf. Supplemental Fig. S1[1]).

The feeding rate μ is listed among the parameter estimates (Supplemental Table S1[1]) and is found to be one sandeel every 3.3 days with a quite low marginal standard error. In comparison, Temming and Mergardt (2002) found the mean time between meals to be 4 days for the same set of data. The reason for this discrepancy is not clear, as the applied methods are very different.

Discussion

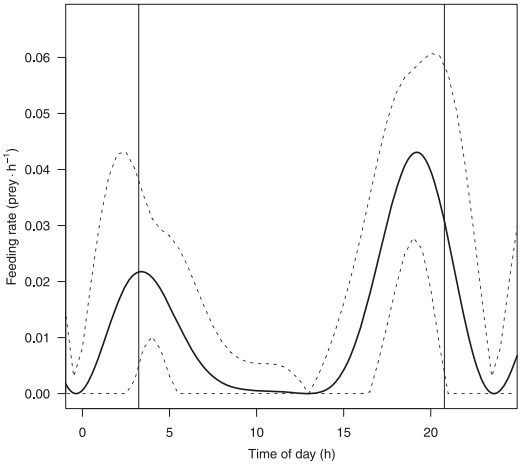
Maximum likelihood estimation of arrival rates from noisy observations has been considered, a problem that arises when ingestion times from partially digested food items are predicted through gastric evacuation models. This new method takes advantage of knowledge about the distribution of the individual error on each predicted ingestion time, which here is assumed to be a truncated normal distribution with known mean and variance (Andersen and Beyer 2008*a* and 2008*b*). As a result of including the error distributions, most emphasis is put on those observations with small errors for determining the shape of intensity curve. A key advantage of this

Table 2. Likelihood ratio test for the number of harmonics, N.

N	Log-likelihood	Test size	$\chi^2_{0.05,2}$	p
0	–1928.94			
1	–1917.78	22.32	5.99	<0.01
2	–1914.78	6.00	5.99	0.05
3	–1913.96	1.64	5.99	0.44

Note: The homogeneous Poisson process is obtained for $N = 0$.

Fig. 2. Intensity curve (eq. 6) with 95% confidence intervals for the case study. The vertical ines indicate sunrise and sunset.



method in comparison with those based on total stomach contents (e.g., Adlerstein and Welleman 2000; Rindorf 2003)) is that it is not necessary to sample throughout the daily cycle, and the analysis is not confined by those sampling points. This can be useful if catch rates vary with the time of day. Another advantage is the ability to handle situations where slow digestion of infrequent arriving prey leads to a flat signal with respect to mean total stomach contents.

The present analysis depends on a number of assumptions (population-level Poisson process, periodicity, and no satiation), which should be thought of as approximations needed to facilitate the analysis rather than representing the truth. The assumption of 24 h periodicity should therefore be regarded as a local approximation, but could in principle be tested by changing the period to some appropriate integer scaling $i \cdot 24$ covering the duration of sampling, followed by estimation of the arrival rate with i times the original number of harmonics and performing likelihood ratio tests (this will ensure that $i = 1$ is a submodel of, for example, $i = 3$; the two models are equal when the Fourier coefficients of all but frequency indices divisible by three are zero). Owing to the low signal-to-noise ratio encountered in this particular study and only 3 days of sampling, it must however be deemed unlikely that this will lead to rejection of our hypothesis of 24 h periodicity. Another assumption is that the population feeding pattern can be well approximated using a squared Fourier series. This is a convenient way to express periodicity, but it is a priori not very likely that data were generated from such a curve. Some care should therefore be taken when interpreting the estimated curve and its confidence bounds — deviations from a constant intensity such as two narrow spikes around dusk and dawn would require many harmonics and thus many parameters using this parameterization. Such a curve would therefore most likely be rejected in favour of a smoother curve.

Possible satiation effects have been ignored in this study, although such an effect has been documented for many species including whiting (see for instance Rindorf 2002). Ignoring satiation will lead to some false periods of zero arrivals, or equivalently, shorter effective observation intervals for satiated individuals. It is, however, not as straightforward to include satiation in this model as for a model based on total stomach contents, as it would be natural to let satiation depend on total stomach contents. On the other hand, including satiation will most probably not change the estimated locations of the peak feeding times, but only add to the uncertainty of the estimated curve.

An interesting but nontrivial question is how many samples are needed to demonstrate some predetermined amount of variation in the feeding rate for some scenario that one wishes to examine. Here, simulation studies can provide some insight as it did in our case, where it was found that 600 prey items was often not enough.

Acknowledgements

Thanks are extended to Kasper Kristensen, Niels G. Andersen, Jan E. Beyer, Uffe H. Thygesen, and Anders Nielsen for useful discussions and comments on earlier drafts. This work was sponsored by the REX project and DTU Aqua. The REX project was funded by the Ministry of Food, Agriculture and Fisheries and the European Fisheries Fund.

References

Adlerstein, S.A., and Welleman, H.C. 2000. Diel variation of stomach contents of North Sea cod (*Gadus morhua*) during a 24-h fishing survey: an analysis using generalized additive models. Can. J. Fish. Aquat. Sci. **57**(12): 2363–2367. doi:10.1139/cjfas-57-12-2363.

Alizadeh, F., Eckstein, J., Noyan, N., and Rudolf, G. 2008. Arrival rate approximation by nonnegative cubic splines. Oper. Res. **56**(1): 140–156. doi:10.1287/opre.1070.0443.

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**(5): 1198–1217. doi:10.1111/j.1095-8649.2001.tb00186.x.

Andersen, N.G., and Beyer, J.E. 2005*a*. Gastric evacuation of mixed stomach contents in predatory gadoids: an expanded application of the square root model to estimate food rations. J. Fish Biol. **67**(5): 1413–1433. doi:10.1111/j.0022-1112.2005.00835.x.

Andersen, N.G., and Beyer, J.E. 2005*b*. Mechanistic modelling of gastric evacuation in predatory gadoids applying the square root model to describe surface-dependent evacuation. J. Fish Biol. **67**(5): 1392–1412. doi:10.1111/j.0022-1112.2005.00834.x.

Andersen, N.G., and Beyer, J.E. 2007. How are prey fishes of multiple meals evacuated from the stomach of a piscivorous fish? J. Fish Biol. **71**(1): 219–234. doi:10.1111/j.1095-8649.2007.01486.x.

Andersen, N.G., and Beyer, J.E. 2008*a*. Precision of ingestion time and evacuation predictors for individual prey in stomachs of

predatory fishes. Fish. Res. **92**(1): 11–22. doi:10.1016/j.fishres.2007.12.007.

Andersen, N.G., and Beyer, J.E. 2008*b*. Predicting ingestion times of individual prey from information about stomach contents of predatory fishes in the field. Fish. Res. **92**(1): 1–10. doi:10.1016/j.fishres.2007.12.004.

Bell, B. 2005. CppAD: a package for differentiation of C++ algorithms. Available from http://www.coin-or.org/CppAD [updated 23 January 2011].

Farrell, A.P., Thorarensen, H., Axelsson, M., Crocker, C.E., Gamperl, A.K., and Cech, J.J.,, Jr. 2001. Gut blood flow in fish during exercise and severe hypercapnia. Comp. Biochem. Physiol. A Mol. Integr. Physiol. **128**(3): 549–561. doi:10.1016/S1095-6433(00)00335-4. PMID:11246044.

Griffiths, W.E. 1976. Feeding and gastric evacuation in perch (*Perca fluviatilis* L.). Mauri Ora, **4**: 19–34.

Hall, S.J., Gurney, W.S.C., Dobby, H., Basford, D.J., Heaney, S.D., and Robertson, M.R. 1995. Inferring feeding patterns from stomach contents Data. J. Anim. Ecol. **64**(1): 39–62. doi:10.2307/5826.

Johansen, G.O., Bogstad, B., Mehl, S., and Ulltang, Ø. 2004. Consumption of juvenile herring (*Clupea harengus*) by cod (*Gadus morhua*) in the Barents Sea: a new approach to estimating consumption in piscivorous fish. Can. J. Fish. Aquat. Sci. **61**(3): 343–359. doi:10.1139/f03-168.

Johnson, M.A., Lee, S., and Wilson, J.R. 1994. Experimental evaluation of a procedure for estimating nonhomogeneous Poisson processes having cyclic behavior. ORSA J. Comp. **6**(4): 356–368.

Kuhl, M., Wilson, J., and Johnson, M. 1997. Estimating and simulating Poisson processes having trends or multiple periodicities. IIE Trans. **29**(3): 201–211. doi:10.1080/07408179708966327.

Magnússon, K.G., and Aspelund, T. 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. **54**(4): 876–889. doi:10.1139/cjfas-54-4-876.

Mergardt, N., and Temming, A. 1997. Diel pattern of food intake in whiting (*Merlangius merlangus*) investigated from the weight of partly digested food particles in the stomach and laboratory determined particle decay functions. ICES J. Mar. Sci. **54**(2): 226–242. doi:10.1006/jmsc.1996.0190.

Pedersen, J., and Hislop, J.R.G. 2001. Seasonal variations in the energy density of fishes in the North Sea. J. Fish Biol. **59**(2): 380–389. doi:10.1111/j.1095-8649.2001.tb00137.x.

Rindorf, A. 2002. The effect of stomach fullness on food intake of whiting in the North Sea. J. Fish Biol. **61**(3): 579–593. doi:10.1111/j.1095-8649.2002.tb00897.x.

Rindorf, A. 2003. Diel feeding pattern of whiting in the North Sea. Mar. Ecol. Prog. Ser. **249**: 265–276. doi:10.3354/meps249265.

Sainsbury, K.J. 1986. Estimation of food consumption from field observations of fish feeding cycles. J. Fish Biol. **29**(1): 23–36. doi:10.1111/j.1095-8649.1986.tb04923.x.

Stefansson, G., and Palsson, O.K. 1997. Statistical evaluation and modelling of the stomach contents of Icelandic cod (*Gadus morhua*). Can. J. Fish. Aquat. Sci. **54**(1): 169–181. doi:10.1139/cjfas-54-1-169.

Temming, A., and Mergardt, N. 2002. Estimating the mean time between meals in the field from stomach content data and gastric evacuation functions of whiting (*Merlangius merlangus* L.) feeding on sandeel (*Ammodytes marinus* Raitt). ICES J. Mar. Sci. **59**(4): 782–793. doi:10.1006/jmsc.2002.1181.

Online Supplemental Materials for "Estimation of Feeding Patterns for Piscivorous Fish Using Individual Prey Data from Stomach Contents"

Casper W. Berg and Axel Temming

Supplemental Appendix S1

Separating Error Terms

Assuming a predator specific variation in the rate parameter ρ and a prey specific variation on the error on the predicted original mass $\hat{w}$, the error structure is correctly described by two nested levels of random effects. For two arrival times from the same predator this amounts to the following contribution to the likelihood in terms of the observed arrival times:

$$\int_{\rho=0}^{\infty} \phi(\rho, \sigma_e^2) \int_{u_1} \lambda(T_{1,\rho} + u_1) \frac{1}{c_1} \phi\left(\frac{u_1}{\sigma_{1,\rho}}\right) du_1 \times \int_{u_2} \lambda(T_{2,\rho} + u_2) \frac{1}{c_2} \phi\left(\frac{u_2}{\sigma_{2,\rho}}\right) du_2 d\rho$$

And thus the log-likelihood becomes

$$l(\mathbf{X}, \lambda) = \sum_{k=1}^{p} \left[\log\left(\int_{\rho=0}^{\infty} \phi(\rho, \sigma_e^2) \prod_{i=1}^{n_k} \int_{u_i} \lambda(T_{i,\rho} + u_i) \times \frac{1}{c_i} \phi\left(\frac{u_i}{\sigma_{i,\rho}}\right) du_i d\rho \right) - \int_{t_{start}^k}^{t_{end}^k} \lambda(t) dt \right].$$

The integration with respect to ρ must be performed numerically, as we do not have the ingestion time as a function of the evacuation rate ρ on closed form in case of multiple meals. It can perhaps be questioned whether separation of error terms is worth the extra trouble, especially if the mean number of prey per stomach low. In this model it will also be more complicated to estimate different arrival rates for different types of prey as you cannot simply divide into separate datasets for each prey type. Using separate datasets is only possible when the errors are independent.

Supplemental Appendix S2

Sampling Arrival Times for NHPPs

As described in [2], inversion sampling can be used to generate random samples of the time to the next arrival given the time of the current arrival and the NHPP rate function λ. This involves the solution of an integral for each sample which is undesirable. An alternative which avoids integrals is rejection sampling/thinning. To sample from a pdf f, we need an envelope function g with the restriction that $g(t) \geq f(t)$ everywhere. For this we use a simple Poisson process, which implies a constant hazard rate $g(t) = \lambda = c$. To generate samples in the interval $[t_1, t_2]$ we start by generating N proposal samples $\mathbf{t_p}$, where $N \backsim Pois(c(t_2 - t_1))$ uniform on $[t_1, t_2]$ and N acceptance thresholds $\mathbf{a}$ uniform on $[0, 1]$. For each proposal we accept it if $a_i < f(t_{p,i})/c$ and reject it otherwise.

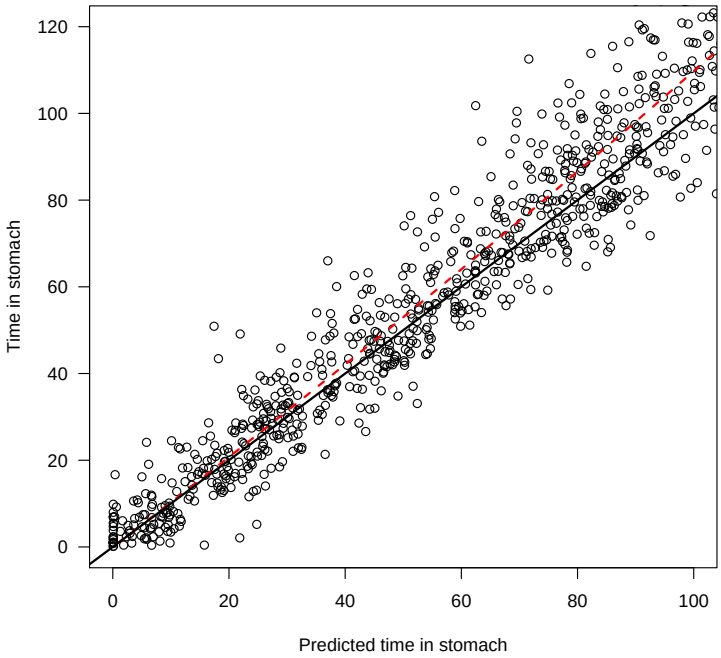
Simulating Stomach Data

A random time of catch was sampled for each predator, which must be sufficiently large to assume stationarity of the distribution of stomach contents. The catch time was chosen to be uniformly distributed over the daily cycle, and the temperature was held constant. Each individual predator got its own digestion rate ρ according to

$$\rho = \hat{\rho}(1 + \epsilon); \quad \epsilon \backsim N(0, \sigma_e^2)$$

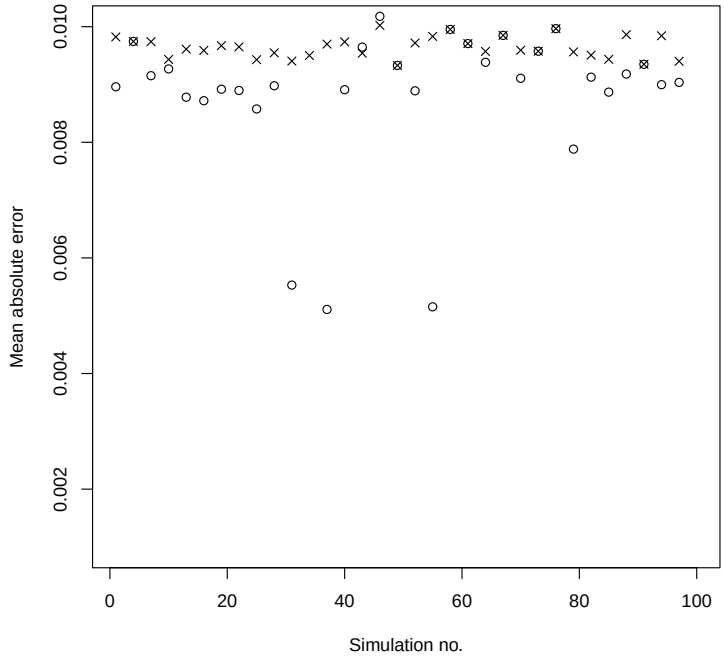
using the general estimate of 0.1 of the coefficient of variation σ_e found by [1]. The size of prey were sampled by sampling the length of the prey from a log-normal distribution with parameters from a fit to real data. The weight for each prey was then found by use of the power function $\sqrt{\hat{w}} = cL^d$, which describes the relationship between length L and body mass. The parameters c and d were found by linear regression on the log-transformed length, which is consistent with the assumption of constant coefficient of variation. A general estimate of 0.03 of σ_m found in [1] was used. Now a simulated stomach can be obtained by adding prey to the stomach according to the simulated arrival times and prey sizes and projecting the mass of each item forward in time until the time of catch. The forward projection was performed numerically by a simple Euler scheme.

Supplemental Figure S1



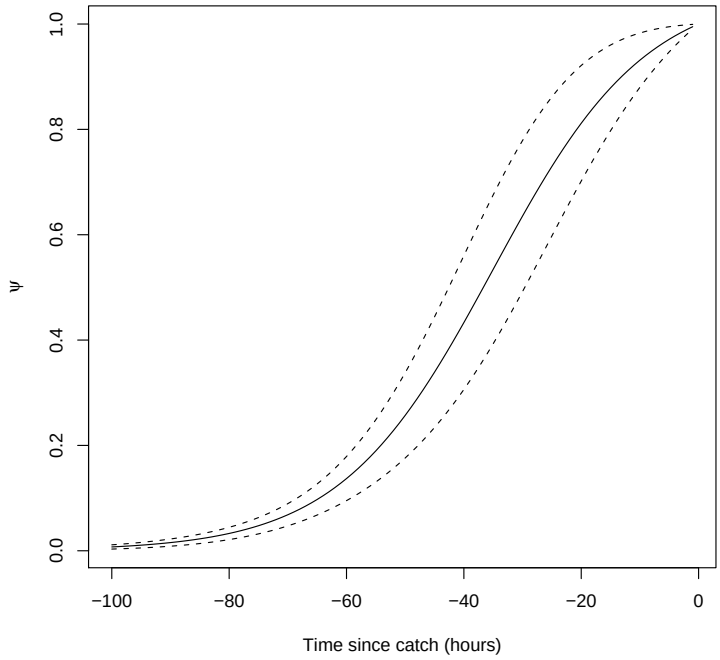
Supplemental Figure 1: Bias from unobserved prey for simulated data in a setting similar to the case study. The unbiased relationship ($y = x$) is represented by the solid line. A better fit is obtained by a second order polynomial (dashed line) indicating that some bias is present.

Supplemental Figure S2



Supplemental Figure 2: Performance measures for 32 simulated datasets with whiting feeding on sandeel. The resulting errors from estimating a constant intensity (crosses) and from using the chosen model (circles) are shown. The choice of model for each simulation is based on the likelihood-ratio test, using the constant intensity as the null hypothesis.

Supplemental Figure S3



Supplemental Figure 3: Estimated probability of observing a sandeel in the stomach at time t since ingestion for a whiting with average digestion rate. The 95% confidence interval is also shown.

Supplemental Table S1

Parameter	Estimate	Standard deviation
a_0	0.0275	0.039
a_1	-0.0052	0.043
a_2	-0.0330	0.040
b_1	-0.1437	0.019
b_2	-0.0545	0.020
$\log(\alpha)$	-2.568	0.095
$\log(\beta)$	2.266	0.124
μ	0.3014	0.029

Supplemental Table 1: Parameter estimates

References

[1] Andersen N.G., and Beyer J.E. 2008. Precision of ingestion time and evacuation predictors for individual prey in stomachs of predatory fishes. Fisheries Research, **92**: 11–22.

[2] Kuhl, M.E. and Wilson, J.R. and Johnson, M.A. 1997. Estimating and simulating Poisson processes having trends or multiple periodicities. IIE Transactions, **29**: 201–211.

Appendix D

Paper II

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Estimation methods for nonlinear state-space models in ecology

M.W. Pedersen[a,*], C.W. Berg[a,b], U.H. Thygesen[b], A. Nielsen[b], H. Madsen[a]

[a] Department for Informatics and Mathematical Modelling, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[b] National Institute of Aquatic Resources, Technical University of Denmark, 2920 Charlottenlund, Denmark

ARTICLE INFO

Article history:
Received 1 September 2010
Received in revised form 7 January 2011
Accepted 12 January 2011
Available online 19 February 2011

Keywords:
AD Model Builder
Hidden Markov model
Mixed model
Monte Carlo
Theta logistic population model
WinBUGS

ABSTRACT

The use of nonlinear state-space models for analyzing ecological systems is increasing. A wide range of estimation methods for such models are available to ecologists, however it is not always clear, which is the appropriate method to choose. To this end, three approaches to estimation in the theta logistic model for population dynamics were benchmarked by Wang (2007). Similarly, we examine and compare the estimation performance of three alternative methods using simulated data. The first approach is to partition the state-space into a finite number of states and formulate the problem as a hidden Markov model (HMM). The second method uses the mixed effects modeling and fast numerical integration framework of the AD Model Builder (ADMB) open-source software. The third alternative is to use the popular Bayesian framework of BUGS. The study showed that state and parameter estimation performance for all three methods was largely identical, however with BUGS providing overall wider credible intervals for parameters than HMM and ADMB confidence intervals.

1. Introduction

State-space models (SSMs) have become the favored approach in modeling time varying ecological phenomena such as population dynamics (Wang, 2007; Gimenez et al., 2007), animal movement (Patterson et al., 2008) and animal behavior (Morales et al., 2004). SSMs come in a variety of classes depending on the problem type (linear or nonlinear) and the error structure of the data (Gaussian or non-Gaussian). In the linear and Gaussian case an exact solution to the SSM can be found using the Kalman filter (KF), which is the optimal estimator (Madsen, 2008). In case of minor departures from linearity, KF variants, such as the extended KF or unscented KF, can be employed. Both methods are reviewed and discussed by Wang (2007). In cases where the state-space equations are highly nonlinear, it is inappropriate to use any KF variant. For ecological problems Markov chain Monte Carlo (MCMC) is perhaps the most common approach to accommodate model nonlinearities owing to its flexibility and general applicability. In addition, free software for MCMC analysis is available, for example the widely used WinBUGS (Gimenez et al., 2008). An example of non-WinBUGS MCMC population modeling is explained by Wang (2007).

We address three powerful methods for the analysis of nonlinear state-space models, two of which have only gained moderate attention previously within the field of ecology compared to the third. The idea of the first method we present is to discretize the continuous state-space and then reformulate the SSM as a hidden Markov model (HMM) (see Zucchini and MacDonald, 2009). A similar approach was described by Kitagawa (1987). The second method we consider is implemented in the open-source software AD Model Builder (ADMB-project, 2009a). In ADMB the SSM is formulated as a statistical model with mixed effects. A major advantage of ADMB is that it makes efficient use of available computer resources by so-called automatic differentiation. Thirdly, we apply OpenBUGS, which is the open-source version of WinBUGS (Spiegelhalter et al., 1996). BUGS is flexible and therefore widely used in modeling ecological systems (Gimenez et al., 2008).

To broaden the perspective of this study we apply the three methods to simulated data from the theta logistic population model, which is a nonlinear SSM. The same example was analyzed by Wang (2007). The performance of the three methods is summarized with respect to a range of aspects: complexity of implementation, computing time, estimation accuracy, limiting assumptions, and algorithmic design. Algorithmic design refers to the amount of subjective tuning required before actual estimation can begin. Because of reduced subjective influence, methods with fewer tuning parameters are often preferable. Finally, we discuss some differences between Bayesian (BUGS) and frequentist (HMM and ADMB) methods.

2. Methods

A state-space model describes the dynamics of a latent state ($\mathbf{X}_t$) and how data ($\mathbf{Y}_t$) relate to this state. An important feature of SSMs is their ability to model random variations in the latent state and in

* Corresponding author. Tel.: +45 4525 3095; fax: +45 4588 2673.
E-mail addresses: mwp@imm.dtu.dk, wpsgodd@gmail.com (M.W. Pedersen), cbe@aqua.dtu.dk (C.W. Berg), uht@aqua.dtu.dk (U.H. Thygesen), an@aqua.dtu.dk (A. Nielsen), hm@imm.dtu.dk (H. Madsen).

0304-3800/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2011.01.007

data. For $t \in \{1, \ldots, N\}$ the general system and observation equations of the SSM are respectively $\mathbf{X}_t = g(t, \mathbf{X}_{t-1}, \mathbf{e}_t)$, and $\mathbf{Y}_t = h(t, \mathbf{X}_t, \mathbf{u}_t)$, where $\mathbf{e}_t \sim N(\mathbf{0}, \mathbf{Q}_t)$ is the system error and $\mathbf{u}_t \sim N(0, \mathbf{R}_t)$ is the observation error. Here, "$\sim N(\cdot)$" means Gaussian distributed. Because of the possible nonlinearity of g and h, advanced filtering and smoothing methods must be employed to gain meaningful estimates of $\mathbf{X}_t$. In this respect, the extended Kalman filter, the unscented Kalman filter, and Bayesian filtering e.g. using Markov chain Monte Carlo (MCMC) sampling or BUGS are common approaches. Alternative methods for nonlinear state estimation are hidden Markov models (HMMs, Zucchini and MacDonald, 2009) and mixed effects models using the software AD Model Builder (ADMB). ADMB is freely available and open-source (ADMB-project, 2009a).

2.1. Benchmarking of estimation methods

The log-transformed theta logistic population growth model (Wang, 2007) was used as benchmark example for assessing the estimation performance of HMM, ADMB and BUGS. The system and observation equations for this model are

$$X_t = X_{t-1} + r_0\left(1 - \left(\frac{\exp(X_{t-1})}{K}\right)^{\theta}\right) + e_t, \quad (1)$$

$$Y_t = X_t + u_t, \quad (2)$$

where $e_t \sim N(0, Q)$ and $u_t \sim N(0, R)$.

Following Wang (2007), two different tests of the methods were carried out:

1. State estimation performance with known parameter values, i.e. the ability of the methods to estimate the population level x_t for all t. Obviously, this test is free of Bayesian prior assumptions on parameters.
2. Estimation of states and all five model parameters, $\lambda = (\log(\theta), \log(r_0), K, \log(Q), \log(R))$, simultaneously. This situation is common in practice if model parameters cannot be estimated from independent data. Notice that parameters that may yield estimates close to zero are log-transformed to avoid invalid parameter values.

Specifically for test 1, $T = 2000$ data replicates were simulated with $N = 200$, $K = 1000$, $Q = 0.01$, $R = 0.04$, and the initial state $x_0 = 3$ using 12 different sets of the θ and r_0 parameters (see Table 1). The performance of the methods was evaluated using an estimate of the state estimation error:

$$\mathrm{RMSE} = \frac{1}{T}\sum_{i=1}^{T}\left(\frac{1}{N}\sum_{t=1}^{N}(\widehat{x}_{i,t} - x_t)^2\right)^{1/2}, \quad (3)$$

Table 1
Performance of state estimation as defined by Eq. (3) for HMM, ADMB, and BUGS.

Sim. no.	r_0	θ	RMSE		
			HMM	ADMB	BUGS
1	0.1	0.5	0.100	0.100	0.100
2	0.5	0.5	0.099	0.099	0.100
3	0.75	0.5	0.097	0.097	0.097
4	1.0	0.5	0.095	0.095	0.095
5	0.1	1.0	0.100	0.100	0.100
6	0.5	1.0	0.095	0.095	0.095
7	0.75	1.0	0.091	0.092	0.092
8	1.0	1.0	0.090	0.090	0.090
9	0.1	1.5	0.100	0.100	0.100
10	0.5	1.5	0.092	0.092	0.092
11	0.75	1.5	0.091	0.091	0.091
12	1.0	1.5	0.096	0.096	0.096

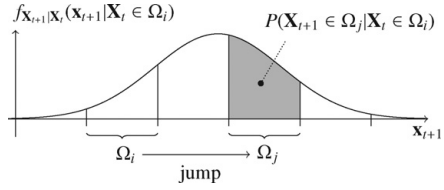


Fig. 1. Probability of a jump (transition) from the state Ω_i to the state Ω_j in the time interval from t to $t+1$ in a HMM. The shaded area corresponds to the integral in (4).

where $\widehat{x}_{i,t}$ is the state estimate for replicate i at time t, and x_t is the true state at time t.

Specifically for test 2, two datasets were simulated using two other sets of parameter values: $\lambda_1 = (\theta = 0.5,\ r_0 = 0.1,\ K = 900,\ Q = 0.01,\ R = 0.04)$ and $\lambda_2 = (\theta = 1.5,\ r_0 = 0.1,\ K = 900,\ Q = 0.01,\ R = 0.04)$ with the number of data points $N = 200$. Parameter estimates for these data using the three methods were found similarly to the study of Wang (2007). We further used these two parameter configurations to generate plots of the joint profile likelihood surfaces for r_0 and θ, which were transformed to confidence contours via a χ^2-distribution as in Polansky et al. (2009). The simulated data sets for λ_1 and λ_2 are available in the supplementary material to enable comparison of our results with future estimation methods. Additionally for test 2 we estimated all five model parameters along with 95% intervals using $T = 200$ of the data sets simulated for test 1. Inspired by Lambert et al. (2005), the purpose here was to evaluate the frequentist properties of the intervals provided by the three estimation methods.

2.2. Hidden Markov model with Matlab

The integrals involved in the prediction, filtering, and smoothing steps for nonlinear SSMs (see e.g. Eq. (2.2, 2.3 and 2.5) in Kitagawa, 1987) can, in general, not be solved analytically. However, by partitioning the continuous state-space uniformly into n parts the solution can be computed using hidden Markov models (HMMs) (Zucchini and MacDonald, 2009). See de Valpine and Hastings (2002) for an ecologically motivated study using a similar method. A state is denoted Ω_i, where $i \in \{1, 2, \ldots, n\}$. The probability distribution of the state given the observations $\mathcal{Y}_t$ available by time t is $P(\mathbf{X}_t \in \Omega_i|\mathcal{Y}_t) = p_t(i|\mathcal{Y}_t)$ which are collected in the row vector $\mathbf{p}_t(\mathcal{Y}_t) = \{p_t(i|\mathcal{Y}_t)\}$. The transition probability of jumping from Ω_i to Ω_j (see Fig. 1) is

$$p_t(i,j) = P(\mathbf{X}_{t+1} \in \Omega_j|\mathbf{X}_t \in \Omega_i) = \int_{\Omega_j} f_{\mathbf{x}_{t+1}|\mathbf{x}_t}(\mathbf{x}_{t+1}|\mathbf{X}_t \in \Omega_i)d\mathbf{x}_{t+1}. \quad (4)$$

For one-dimensional problems Ω_i are intervals on the line, in two dimensions Ω_i are areas, and analogously for higher dimensions. Note that the $n \times n$ probability transition matrix $\mathbf{P}_t = \{p_t(i,j)\}$ is not homogeneous, i.e. the transition probabilities may change as a function of time as indicated by (1). Now, the HMM prediction, filtering, and smoothing equations are respectively

$$\mathbf{p}_t(\mathcal{Y}_{t-1}) = \mathbf{p}_{t-1}(\mathcal{Y}_{t-1})\mathbf{P}_{t-1},$$
$$\mathbf{p}_t(\mathcal{Y}_t) = \psi_t^{-1}\mathbf{p}_t(\mathcal{Y}_{t-1}) \odot \mathbf{L}(y_t),$$
$$\mathbf{p}_t(\mathcal{Y}_N) = \mathbf{p}_t(\mathcal{Y}_t) \odot [\{\mathbf{p}_{t+1}(\mathcal{Y}_N) \oslash \mathbf{p}_{t+1}(\mathcal{Y}_t)\}\mathbf{P}_t^T]$$

where '$\odot$' and '$\oslash$' are elementwise matrix multiplication and division, respectively. The likelihood of the observations $\mathbf{L}(y_t)$ is a row vector with elements $p_t(y_t|i)$ and $\psi_t = \mathbf{p}_t(\mathcal{Y}_{t-1}) \cdot \mathbf{L}(y_t)^T$ is a normalization constant with '$\cdot$' denoting dot product. The estimate of the state given all N observations is simply the mean of the distribution $\mathbf{p}_t(\mathcal{Y}_N)$.

Using the above scheme we can estimate the unknown parameters (λ) of the SSM by maximizing the likelihood function

$$\mathcal{L}(\lambda|\mathcal{Y}_N) = f_{\mathcal{Y}N}(\mathcal{Y}_N|\lambda) = [\mathbf{L}(y_1)\cdot\mathbf{1}]\prod_{t=2}^{N}\psi_t, \quad (5)$$

as in Kitagawa (1987), where $\mathbf{1}$ is a column vector of ones. The maximum likelihood (ML) estimate of the model parameters $\widehat{\lambda}$ is found by optimizing (5) as a function of λ. The covariance matrix of $\widehat{\lambda}$ is approximated by the inverse Hessian of the likelihood function at the optimum $\widehat{\lambda}$. This approximation is appropriate because the ML estimate is asymptotically Gaussian under certain regularity conditions (Cappé et al., 2005). Thus, confidence intervals can be constructed using the approximated covariance matrix. Under parameter transformations it is important to construct the confidence intervals in the transformed parameters and then reverse transform the computed confidence limits.

When analyzing the theta logistic model we set $n = 251$. The bounds of the discrete state-space are chosen such, that the probability of the true state falling outside the grid is negligible. That is, we use the observation model (2) to determine bounds that envelope the true latent state with a probability close to 1. This approach is similar to the one used in de Valpine and Hastings (2002). Details on grid specification can be found in the supplementary material containing model code.

The HMM code provided in the supplementary material was written in Matlab, but the method is not language specific. Matlab was chosen because it is widely used and has a syntax which is relatively easy to understand even for non-Matlab users.

2.3. Mixed effects model with AD Model Builder

Hierarchical mixed effects models are an alternative framework for analyzing nonlinear SSMs. The states are the random effects of the model and are collectively referred to as $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$. Here, as in Madsen and Thyregod (2010), we specify a model for the data conditional on the unobserved random effects, $f_{\mathcal{Y}_N|\mathcal{X}}(\mathcal{Y}_N|\mathcal{X}, \lambda_a)$ which corresponds to (2). We also specify a model for the random effects, $f_{\mathcal{X}}(\mathcal{X}|\lambda_b)$ which corresponds to (1). The joint density of random effects and observations conditional on the parameters is

$$f_{\mathcal{X},\mathcal{Y}_N}(\mathcal{X}, \mathcal{Y}_N|\lambda) = f_{\mathcal{X}}(\mathcal{X}|\lambda_b) f_{\mathcal{Y}_N|\mathcal{X}}(\mathcal{Y}_N|\mathcal{X}, \lambda_a).$$

To obtain the marginal likelihood for estimating $\lambda = \{\lambda_a, \lambda_b\}$ we integrate over the unobserved random effects

$$\mathcal{L}(\lambda|\mathcal{Y}_N) = f_{\mathcal{Y}_N}(\mathcal{Y}_N|\lambda) = \int_{\mathbb{R}^N} f_{\mathcal{X},\mathcal{Y}_N}(\mathcal{X}, \mathcal{Y}_N|\lambda)\, d\mathcal{X}. \quad (6)$$

The N-dimensional integral in (6) is generally challenging to solve, and for nonlinear mixed models we must resort to numerical approximation methods for estimating the model parameters. An efficient and widely used method for this is the Laplace approximation (Wolfinger and Xihong, 1997), which replaces the integrand with a second order Taylor expansion around the optimum of the log-likelihood function. This allows for elimination of the integral, because the second-order Taylor expansion can be formulated as a known constant multiplied by a multivariate Gaussian density, which integrates to unity. For nonlinear models the distribution of the random effect may not be Gaussian. Then the Laplace approximation is not exact. In particular for multi modal distributions one should use the Laplace approximation with caution. Still, when analyzing nonlinear models with moderately skewed unimodal distributions good results can be obtained with the Laplace approximation (Vonesh, 1996; Mortensen, 2009). In any case it is important to investigate if the approximation is critically violated e.g. by Monte Carlo sampling from the random effects distribution.

Even with the Laplace approximation maximization of the marginal log-likelihood with respect to λ is challenging. A computationally efficient method is to combine the Laplace approximation with so-called automatic differentiation (AD Skaug and Fournier, 2006). AD is a technique for finding the gradient of a function h (in our case the log-likelihood), provided that h can be expressed in computed code. Evaluating h using AD gives the function value along with the gradient of h at the point of evaluation. The gradient is computed using the chain rule of calculus on every operation in the code that contributes to the value of h. For efficient maximization of the Laplace approximation of the marginal log-likelihood with respect to λ, up to third order partial derivatives must be found. Skaug and Fournier (2006) show how this can be accomplished by repeated use of AD.

The above procedure is implemented in AD Model Builder (ADMB), which we use to analyze the theta logistic model. ADMB is an open-source software package and programming language based on C++. It includes a function minimizer for ML parameter estimation and a random effects module, which utilizes the Laplace approximation for integration of random effects. Standard deviations for constructing confidence intervals are calculated using the delta method (Oehlert, 1992) and automatically reported on all estimated quantities. The covariance matrix for all states in an SSM is a banded matrix (Skaug and Fournier, 2006). ADMB can exploit this property by using the SEPARABLE FUNCTION construct (ADMB-project, 2009b) to gain significant speed improvements. Other than this useful property ADMB has no tuning parameters as such.

2.4. Monte Carlo estimation with BUGS

Finally, we analyze the theta logistic model using the Bayesian modeling language BUGS, which is an MCMC estimation method (Spiegelhalter et al., 1996). BUGS is a popular tool in ecological modeling (e.g. Gimenez et al., 2007; Jonsen et al., 2005; Schofield et al., 2009). BUGS is best known in the WinBUGS form which has a graphical user interface. Here, however, we use the open-source alternative OpenBUGS, yet the BUGS code provided in the supplementary material is compatible with WinBUGS.

A Bayesian analysis requires that prior distributions are specified for the model parameters. The type of prior distributions and parameter values related to these distributions should reflect the *a priori* knowledge that is available about the model parameters. BUGS then uses Gibbs sampling (Casella and George, 1992) to explore the posterior distribution of the parameter and state-space by incorporating the information specified by the priors, the state-space model, and the observed data. The Gibbs algorithm exploits that sampling the posterior is sometimes simpler via its conditional distributions rather than directly from the joint distribution. This is the case for state-space models where direct sampling of the posterior for states and parameters is difficult. Instead, sampling model parameters from priors and then sampling X_t conditional on model parameters and remaining states $(X_1, \ldots, X_{t-1}, X_{t+1}, \ldots, X_N)$ for all t is simple using (1). The ampling algorithm applied by BUGS in specific cases depends on the form and type of the conditional distribution, and also on the composition of priors on model parameters (see Spiegelhalter et al., 1996, 2003, for details).

We consider the common practical situation where *a priori* knowledge is unavailable and estimation therefore relies entirely on information in data. How to specify vague (or uninformative) priors is a topic of on-going research (Gelman, 2006; Lambert et al., 2005), which is outside the scope of this study. One suggested vague prior is a uniform distribution with wide support (Spiegelhalter et al., 1996). So, we choose a uniform prior for K, and uniform priors for $\log\theta$ and $\log r_0$ that were much wider than the natural biological bounds for the parameter values. By log-transforming θ and r_0

Table 2
Computing times for HMM, ADMB, BUGS1 (inverse-Gamma prior on variances), and BUGS2 (uniform prior on log-standard deviations). All times are for a single dataset run on the same computer.

	HMM	ADMB	BUGS1	BUGS2
State est.	6.12 s	0.49 s	58 s	58 s
Par. est.	225 s	2.5 s	118 s	614 s

biological meaningful (i.e. positive) parameter values are ensured. The state-space formulation implies that the variance parameters Q and R are non-zero and therefore also require prior distributions. It is common to assign vague inverse-gamma distributed priors to variance parameters (Spiegelhalter et al., 2003; Lambert et al., 2005). Gelman (2006), however, recommends using a uniform prior on the log-transformed standard deviation. Therefore, to asses the sensitivity of the estimation results to the choice of prior we perform BUGS estimation in two separate cases: BUGS1 using an inverse-gamma distribution for Q and R, and BUGS2 using a uniform distribution on the log-transformed standard deviation, i.e. $0.5\log(Q)$ and $0.5\log(R)$.

Estimation using BUGS involves a number of tuning parameters: the initial values for the sampling scheme can be found in the supplementary material online along with the specifics of the priors. The total number of generated samples was 100,000 with 50,000 used for burn-in. The appropriate number of samples was found iteratively by repeated application of Geweke Z score test for convergence (Geweke, 1992). The BUGS thinning rate was 50 (for reducing sample autocorrelation, which was apparent for θ and r_0 at lower thinning rates). With these values of the tuning parameters we get an effective sample size of 1000. For summarizing the estimation results the maximum *a posteriori* (MAP) parameter estimates along with 95% credible intervals are reported (where the lower bound equals the 2.5% quantile and the upper bound equals the 97.5% quantile of the posterior distribution).

3. Results

State estimation results for the three methods using known parameter values were practically identical (Table 1). ADMB was an order of magnitude faster than HMM, which, in turn, was an order of magnitude faster than BUGS (Table 2). State estimation using estimated parameter values also gave practically identical results for all three methods (Fig. 2). Regarding ML parameter estimation and confidence intervals (CIs) for λ_1 and λ_2, HMM and ADMB performed almost identically (Table 3). Likewise, MAP estimates and credible intervals provided by BUGS1 and BUGS2 were overall similar in the λ_2 case. In the λ_1 case, however, BUGS1 MAP estimates of θ and r_0 were markedly lower and higher respectively than the estimates provided by HMM, ADMB, and BUGS2. Perhaps most surprisingly was the upper limit of the credible interval for K seemingly quite sensitive to the choice of prior employed by BUGS, and in both cases considerably higher than the HMM and ADMB CI upper limits. Some notable differences between CIs and credible intervals were present for θ, K, and r_0 in the λ_1 case (Table 3), with BUGS generally being more conservative and providing wider intervals (in the log domain). Inspection of the joint profile likelihood surfaces for θ and r_0 revealed that contour lines closely approximated elliptical shapes for λ_2 (Fig. 4, panel B), thus indicating that the quadratic approximation used by HMM and ADMB was appropriate. For λ_1, on the other hand, the quadratic approximation was only appropriate until the 65% confidence limit where the contour shape started to diverge from the elliptical shape (Fig. 4, panel A). If comparing the limits of the intervals provided by all three methods for the λ_1 case (Table 3) with the extents of the likelihood surface (Fig. 4, panel A), it is clear that neither credible intervals nor CIs captured the actual range of plausible parameter values.

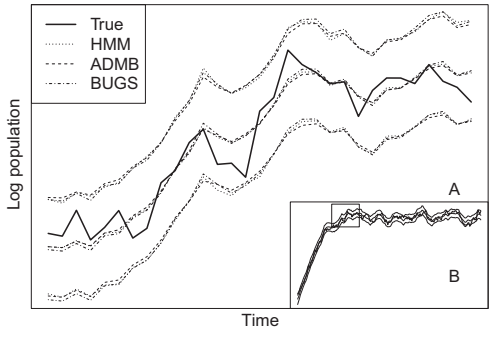


Fig. 2. State estimation of the theta logistic model with 95% intervals using the estimated parameter values in Table 3. True states were generated using $\lambda_2(\theta = 1.5)$. Panel A is a zoom of a part of the full time series indicated by the small box in panel B. Clearly in this case, HMM, ADMB, and BUGS gave close to identical state estimation results.

Visualizing the empirical distributions of the $T = 200$ parameter estimates (Fig. 3) showed largely identical results for all three methods. For all parameters the average 95% CIs provided by HMM and ADMB closely approximated the 2.5% and 97.5% quantiles of the corresponding empirical distribution. Similar results were observed for BUGS1 and BUGS2 for parameters R and Q. Regarding the three remaining parameters θ, K, and r_0, on the other hand, the average credible intervals were markedly wider than the corresponding quantiles of their empirical distribution, and therefore also wider than their CI counterparts. The difference in results between the two vague priors (BUGS1 and BUGS2) was minimal except for the credible intervals for K where BUGS2 gave wider intervals than BUGS1. Since both priors have been regarded in the literature as vague their influence on the resulting intervals is surprising. Computing times for parameter estimation showed that ADMB, again, was significantly faster than HMM and BUGS (Table 2). Interestingly, BUGS1 was considerably (six times) faster than BUGS2. This results can most likely be ascribed to BUGS using different sampling algorithms in the two cases.

4. Discussion

Dynamical processes are prevalent in ecology. State-space models are commonly used in the analysis of such nonlinear processes because they join separate models of the ecological system and the observation process. This paper assessed the performance of three methods for estimation in nonlinear state-space models: an approach using hidden Markov models (HMM), the open-source AD Model Builder framework (ADMB), and the BUGS language. HMM and ADMB are frequentist (non-Bayesian) methods, while BUGS is Bayesian. To facilitate a transparent comparison among available estimation methods we considered the theta logistic population model, which Wang (2007) analyzed with three other methods (extended Kalman filter, the unscented Kalman filter and a Metropolis–Hastings approach). To increase accessibility, the computer code for our three methods can be found in the online supplementary material.

The state estimation root mean square errors (RMSEs) of HMM, ADMB, and BUGS (Table 1) were lower than those for the three methods presented by Wang (2007), his Table 1. The 95% intervals for the parameter estimates of θ provided by our three methods all included the true values (Table 3). Note that they also included $\theta = 1$, which means that the models could not distinguish between

Table 3
Parameter values estimated by HMM, ADMB, BUGS1 (inverse-Gamma prior on variances), and BUGS2 (uniform prior on log-standard deviations) with related 95% intervals. Data were simulated with the listed true parameter values: $\lambda_1 = (\theta = 0.5,\ r_0 = 0.1,\ K = 900,\ Q = 0.01,\ R = 0.04)$ and $\lambda_2 = (\theta = 1.5,\ r_0 = 0.1,\ K = 900,\ Q = 0.01,\ R = 0.04)$ of the theta logistic model.

	HMM		ADMB		BUGS1		BUGS2	
	ML est.	95% conf. intv.	ML est.	95% conf. intv.	MAP est.	95% cred. intv.	MAP est.	95% cred. intv.
λ_1								
θ	0.588	0.134–2.588	0.583	0.129–2.640	0.374	0.0210–1.446	0.538	0.020–1.496
K	829.3	643.3-1015	829.5	639.2–1020	860.0	629.2–1900	834.0	638.3–4957
r_0	0.116	0.046–0.298	0.117	0.045–0.305	0.135	0.053–1.667	0.118	0.045–1.666
R	0.041	0.032-0.053	0.041	0.032-0.053	0.042	0.031-0.054	0.041	0.031-0.052
Q	0.0092	0.0052–0.016	0.0092	0.0051–0.017	0.011	0.0055–0.017	0.0099	0.0060–0.018
λ_2								
θ	1.098	0.412–2.926	1.079	0.402–2.902	1.006	0.043–2.551	1.037	0.043–2.869
K	886.9	792.7–981.0	887.0	790.5–983.5	891.3	769.3–1121	910.0	774.9–1097
r_0	0.128	0.082–0.201	0.129	0.081–0.203	0.127	0.078–1.136	0.134	0.074–1.032
R	0.043	0.032–0.056	0.043	0.032–0.056	0.043	0.031–0.056	0.044	0.032–0.056
Q	0.0082	0.0038–0.018	0.0081	0.0045–0.015	0.0094	0.0041–0.018	0.0086	0.0043–0.019

a concave and convex relation between population size and growth rate. This is in contrast with the credible intervals in Wang (2007), his Table 2, that excluded $\theta = 1$, however three out of six of his credible intervals also excluded the true parameter value, which is of some concern.

Recent studies have indicated that θ and r_0 of the theta logistic model (1) can be difficult to identify for certain data sets (Polansky et al., 2009). This is the case because given $\theta < 1$ similar model dynamics can be generated for different values of θ (Clark et al., 2010). Supporting this, a joint profile likelihood surface for logθ and logr_0 showed that combinations of different values for the two parameters may fit data equally well, i.e. result in practically identical model likelihoods (Fig. 4, panel A, data generated with $\theta = 0.5$). Still parameters estimated by HMM and AMDB were reasonably accurate (Table 3, case λ_1), however the confidence intervals (CIs) were too narrow when compared to the contours of the confidence regions in Fig. 4, panel A. This result underlines the importance of validating the quadratic approximation to the log-likelihood function employed by HMM and ADMB before using it to construct CIs. The credible intervals from BUGS were wider and therefore more realistic than the CIs provided by HMM and ADMB, yet the interval bounds were narrower than the range of plausible models indi-

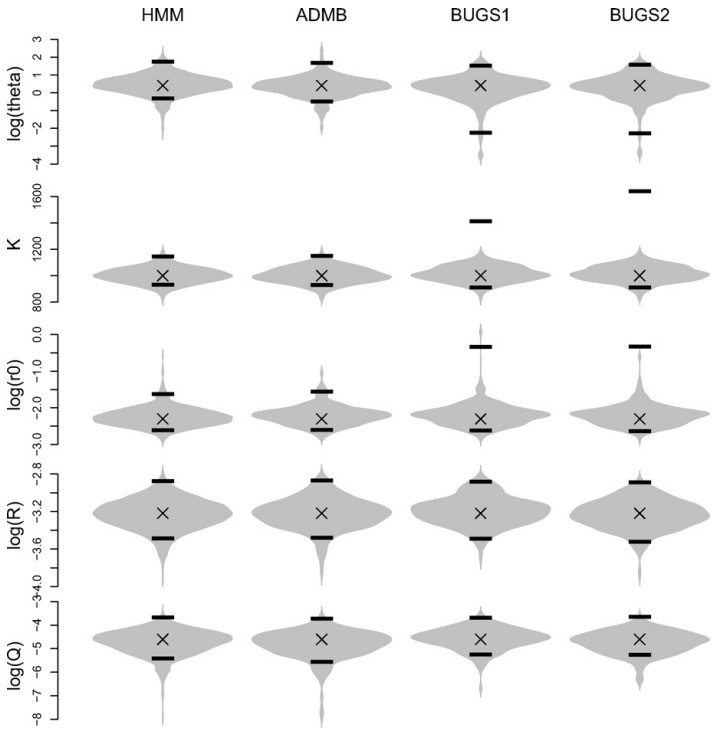


Fig. 3. Violin plots showing the empirical distribution of $T = 200$ parameter estimates. Data used for estimation were simulated with the parameter configuration $\lambda = (\theta = 1.5,\ r_0 = 0.1,\ K = 1000,\ Q = 0.01,\ R = 0.04)$. Crosses indicate the true parameter values, λ. Horizontal lines indicate the average limits of the 200 individual 95% intervals.

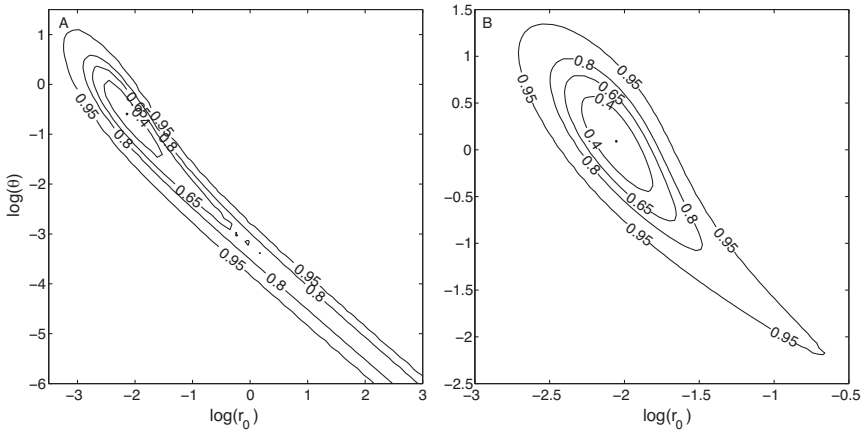


Fig. 4. Joint profile likelihood surfaces for two simulated data sets of the theta-logistic model (see also Table 3). Panel A: parameters used for simulation $\lambda_1 = (\theta = 0.5,\ r_0 = 0.1,\ K = 900,\ Q = 0.01,\ R = 0.04)$. Panel B: $\lambda_2 = (\theta = 1.5,\ r_0 = 0.1,\ K = 900,\ Q = 0.01,\ R = 0.04)$. Following (Polansky et al., 2009) the joint profile log-likelihood surfaces have been transformed to confidence contours via a χ^2-distribution of the profiled models versus the model where all five parameters are estimated. Dots indicate the minima of the transformed surfaces equivalent to the maximum likelihood (ML) point. Both surfaces have elliptically shaped contours in proximity to the ML point in which case a quadratic approximation as used by HMM and ADMB is appropriate. While the surface for λ_2 (panel B) is close to quadratic even at the 95% level, the surface for λ_1 (panel A) departs from the quadratic shape at the 65% level.

cated by the profile likelihood surface. A possible explanation for this difference is that a substantial Monte Carlo sample size may be required to fully explore the posterior distribution when two parameters are highly correlated (Gamerman, 1997). If complications with parameter identifiability as illustrated in Fig. 4, panel A, are encountered in practical situations it is recommended to switch to a simpler model with fewer parameters e.g. by setting $\theta = 1$ (Clark et al., 2010).

For the data set generated with $\theta = 1.5$, the joint profile likelihood surface for $\log\theta$ and $\log r_0$ was well approximated by a quadratic function (Fig. 4, panel B). Thus, log-transforming θ and r_0 in the theta-logistic model avoids a boomerang-shaped likelihood surface (see e.g. Figure 2 in Polansky et al., 2009), which deviates considerably from a quadratic function. Thus, the CIs computed for HMM and ADMB in the log-transformed parameter space (Table 3, case λ_2) corresponded well to the confidence contours in Fig. 4, panel B. For BUGS credible intervals the conclusion was the same.

Similarly to Lambert et al. (2005), the frequentist properties of the three estimation methods were evaluated. To this end we used so-called violin-plots (Fig. 3), where the empirical distribution of 200 parameter estimates was compared with the average of the corresponding 200 95% interval bounds. In discussing our results it is important to stress that CIs provided by frequentist methods (HMM and ADMB) and credible intervals provided by Bayesian methods (BUGS) have fundamentally different interpretations. A 95% CI is an interval which contains the true parameter in 95% of a large number of repeated experiments. Conversely, a 95% credible interval is an interval which has a 95% posterior probability of containing the parameter for the experiment at hand. From Fig. 3 it was evident that the CIs were consistent with corresponding quantiles of the empirical distributions. This further confirms the validity of the quadratic approximation of the log-likelihood function. The empirical distributions of the BUGS parameter estimates under vague prior assumptions were largely identical to their HMM and ADMB counterparts. However, Fig. 3 showed that even when assigning vague priors it cannot be expected that credible intervals coincide with frequentist CIs, which by definition do not incorporate *a priori* knowledge. In addition, considerable differences in credible intervals were present between the two BUGS analyses using different vague priors (Fig. 3). Thus, it is crucial, when employing Bayesian methods in the absence of *a priori* knowledge, to assess the sensitivity of credible intervals to the choice of distribution for the vague prior.

ADMB uses automatic differentiation to estimate the states and parameters of the model, which is the main reason for its computing time superiority (Table 2). This advantage will only increase further as models become more complex and the number of parameters grows. The main disadvantage of ADMB is, that the Laplace approximation for the density of the random effects (here equivalent to the latent states) must be reasonable. In our test cases the latent state estimation of ADMB was close to identical to the HMM and BUGS results (Fig. 2), which justifies using the Laplace approximation. If results from alternative methods are not available, the quality of the approximation can be assessed using the built-in importance sampling functionality (p. 35, ADMB-project, 2009b). Another possible complication of ADMB is that some programming experience in C++ is required. The HMM approach, on the other hand, has the advantage of being language independent, i.e. the method can be implemented in any programming language, for which a function optimizer is available. The programming background of the modeler is therefore of minor concern. The computing speed of the HMM approach is, at worst, proportional to the number of grid cells squared, a number which grows rapidly with increasing state dimension. Thus, HMMs are best suited for one or two-dimensional problems. BUGS depends less on state dimension because it is Monte Carlo based and it requires no density approximations nor differentiability. Consequently, BUGS is flexible and applicable to the widest variety of problems of the three methods we have examined. In addition, WinBUGS (Spiegelhalter et al., 2003) can be used to view and produce BUGS code graphically. This further increases the accessibility of the method.

BUGS and Monte Carlo based methods in general have tuning parameters that cannot be estimated from data and therefore require subjective input from the modeler. The tuning parameters include the number of samples, burn-in time, thinning rate, convergence assessment, and choice of prior distribution, all of which

influence the estimation results significantly. This fact is underlined in the BUGS manual (Spiegelhalter et al., 1996, p. 1), and it is emphasized that the modeler using BUGS must have a sound understanding of the Gibbs sampler. Our results supported this in that computing times (Table 2) and interval estimation (Fig. 3) were significantly influenced by the choice of prior. In contrast, ADMB has no tuning parameters as such, but it does have certain options that are more or less relevant depending on the type of problem, for example the SEPARABLE FUNCTION construct. HMM has two tuning parameters: the extent of the grid and the grid resolution. Limiting the state-space involves a risk of truncating the latent state path. To minimize this risk the approach of de Valpine and Hastings (2002) was followed, where bounds are chosen so wide that the probability of latent path truncation is negligible. Naturally, wider grid extents and higher grid resolution entail an increase in computation time. Thus, determining the value of these parameters is a trade-off between computing speed and accuracy of results. Generally, if one is uncertain about the grid specifications, we recommend to start with a wide and coarse grid to get preliminary results, and then adapt extents and refine the grid accordingly if needed. If the conclusion is unchanged on the adapted grid there is strong evidence that the latent path is enclosed and properly resolved by the discretization.

5. Conclusion

In summary, the three methods considered in this paper are all powerful approaches to nonlinear state-space modeling of ecological systems. ADMB is by far the fastest method owing to its use of the Laplace approximation and automatic differentiation. This limits ADMB to problems where the state distributions are unimodal, which, however, is the case in the majority of practical examples. In contrast, HMM and BUGS are more general and are able to handle arbitrary state distributions. HMM requires specification of a spatial grid and is limited to problems with low state dimensions, say below four. BUGS has fewest model restrictions, but requires specification of prior information and other subjective input from the modeler in the form of algorithmic tuning parameters.

State-space methods provide a natural paradigm for ecosystem modeling. Thus, it is imperative that the ecological community is alert to progress in other scientific fields where state-space models are used and developed. This paper evaluated the performance, with respect to estimation accuracy and speed, of three advanced methods for state-space analysis. The study showed that state and parameter estimation performance for all three methods was largely identical, however with BUGS providing overall wider credible intervals for parameters than HMM and ADMB confidence intervals.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ecolmodel.2011.01.007.

References

ADMB-project, 2009a. AD Model Builder: Automatic Differentiation Model Builder. Developed by David Fournier and freely available from admb-project.org.

ADMB-project, 2009b. Random Effects in AD Model Builder: ADMB-RE User Guide.

Cappé, O., Moulines, E., Ryden, T., 2005. Inference in Hidden Markov Models. Springer.

Casella, G., George, E., 1992. Explaining the Gibbs sampler. American Statistician 46 (3), 167–174.

Clark, F., Brook, B., Delean, S., Reşit Akçakaya, H., Bradshaw, C., 2010. The theta-logistic is unreliable for modelling most census data. Methods in Ecology and Evolution 1 (3), 253–262.

de Valpine, P., Hastings, A., 2002. Fitting population models incorporating process noise and observation error. Ecological Monographs 72, 57–76.

Gamerman, D., 1997. Sampling from the posterior distribution in generalized linear mixed models. Statistics and Computing 7 (1), 57–68.

Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 1 (3), 515–533.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Barnardo, J., Berger, J., David, A., Smith, A. (Eds.), Bayesian Statistics, vol. 4. Clarendon Press, Oxford, UK, pp. 169–193.

Gimenez, O., Bonner, S., King, R., Parker, R., Brooks, S., Jamieson, L., Grosbois, V., Morgan, B., Thomas, L., 2008. WinBUGS for population ecologists: Bayesian modeling using Markov Chain Monte Carlo methods. In: Thomson, David L., C., E.G., Conroy, M.J. (Eds.), Modeling Demographic Processes In Marked Populations. Vol. 3 of Mathematics and Statistics. Springer, US, pp. 883–915.

Gimenez, O., Rossi, V., Choquet, R., Dehais, C., Doris, B., Varella, H., Vila, J., Pradel, R., 2007. State-space modelling of data on marked individuals. Ecological Modelling 206 (3–4), 431–438.

Jonsen, I.D., Flemming, J.M., Myers, R.A., 2005. Robust state-space modeling of animal movement data. Ecology 86 (11), 2874–2880.

Kitagawa, G., 1987. Non-Gaussian state-space modeling of nonstationary time series. Journal of the American Statistical Association 82 (400), 1032–1041.

Lambert, P., Sutton, A., Burton, P., Abrams, K., Jones, D., 2005. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. Statistics in Medicine 24 (15), 2401–2428.

Madsen, H., 2008. Time Series Analysis. Chapman & Hall/CRC London, London.

Madsen, H., Thyregod, P., 2010. Introduction to General and Generalized Linear Models. Chapman & Hall/CRC, London.

Morales, J., Haydon, D., Frair, J., Holsinger, K., Fryxell, J., 2004. Extracting more out of relocation data: building movement models as mixtures of random walks. Ecology 85 (9), 2436–2445.

Mortensen, S.B., 2009. Markov and mixed models with applications. Ph.D. Thesis, Technical University of Denmark (DTU), Kgs. Lyngby, Denmark.

Oehlert, G.W., 1992. A note on the delta method. American Statistician 46 (1), 27–29.

Patterson, T., Thomas, L., Wilcox, C., Ovaskainen, O., Matthiopoulos, J., 2008. State-space models of individual animal movement. Trends in Ecology & Evolution 23 (2), 87–94.

Polansky, L., De Valpine, P., Lloyd-Smith, J., Getz, W., 2009. Likelihood ridges and multimodality in population growth rate models. Ecology 90 (8), 2313–2320.

Schofield, M., Barker, R., MacKenzie, D., 2009. Flexible hierarchical mark-recapture modeling for open populations using WinBUGS. Environmental and Ecological Statistics 16 (3), 369–387.

Skaug, H., Fournier, D., 2006. Automatic approximation of the marginal likelihood in non-gaussian hierarchical models. Computational Statistics & Data Analysis 51 (2), 699–709.

Spiegelhalter, D., Thomas, A., Best, N., Gilks, W., 1996. Bayesian Inference Using Gibbs Sampling. Version 0.5 (Version ii). MRC Biostatistics Unit, Cambridge.

Spiegelhalter, D., Thomas, A., Best, N., Lunn, D., 2003. WinBUGS User Manual. Version 1.4. MRC Biostatistics Unit, Cambridge, UK.

Vonesh, E.F., 1996. A note on the use of laplace's approximation for nonlinear mixed-effects models. Biometrika 83 (2), 447–452.

Wang, G., 2007. On the latent state estimation of nonlinear population dynamics using Bayesian and non-Bayesian state-space models. Ecological Modelling 200 (3–4), 521–528.

Wolfinger, R., Xihong, L., 1997. Two Taylor-series approximation methods for nonlinear mixed models. Computational Statistics & Data Analysis 25 (4), 465–490.

Zucchini, W., MacDonald, I., 2009. Hidden Markov Models for Time Series. Chapman & Hall/CRC, London.

Appendix E

Paper III

Fisheries Research 129–130 (2012) 119–126

Contents lists available at SciVerse ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres

Spatial age-length key modelling using continuation ratio logits

Casper W. Berg[a,*], Kasper Kristensen[b]

[a] National Institute of Aquatic Resources, Technical University of Denmark, Charlottenlund Castle, 2920 Charlottenlund, Denmark
[b] Department of Informatics and Mathematical Modeling, Technical University of Denmark, Richard Petersens Plads, Building 305, 2800 Lyngby, Denmark

ARTICLE INFO

Article history:
Received 27 March 2012
Received in revised form 21 June 2012
Accepted 22 June 2012

Keywords:
Age length key comparison
Continuation ratio logits
Generalized Additive Models

ABSTRACT

Many fish stock assessments are based on numbers at age from research sampling programmes and samples from commercial catches. However, only a small fraction of the catch is typically analyzed for age as this is a costly and time-consuming process. Larger samples of the length distribution and a so-called age-length key (ALK) is then used to obtain the age distribution. Regional differences in ALKs are not uncommon, but stratification is often problematic due to a small number of samples. Here, we combine generalized additive modelling with continuation ratio logits to model the probability of age given length and spatial coordinates to overcome these issues. The method is applied to data gathered on North Sea haddock (*Melanogrammus aeglefinus*), cod (*Gadus morhua*), whiting (*Merlangius merlangus*) and herring (*Clupea harengus*) and its implications for a simple age-based survey index of abundance are examined. The spatial varying ALK outperforms simpler approaches with respect to AIC and BIC, and the survey indices created using the spatial varying ALK displays better internal and external consistency indicating improved precision.

1. Introduction

Estimation of catch at age from combined samples of length and age is standard procedure in analyses of fisheries data. Only a small fraction of the catch is typically analyzed for age as this is a costly and time-consuming process. Larger samples of the length distribution and a so-called age-length key (ALK) are then used to obtain the age distribution.

The ALK is typically estimated from length-stratified subsamples of the catch that are analyzed for age by examining the annual ring structure in the otholiths. Missing or few data points for a given combination of strata such as age, length group, and geographical area frequently occur due to unreadable otholiths or simply because no fish were caught. In this case the raw observed proportions of age-at-length are unsuitable for assigning age to fish in these length groups. A solution to this problem is to use a statistical model to create a smooth distribution of age given length and possibly other covariates, such that missing values can be interpolated in an objective and robust way, and the uncertainty due to the sampling variability can be taken into account. A statistical model also has the advantage of allowing formal testing of hypotheses such as whether two ALKs can be considered identical.

Continuation ratio logits (CRLs) is a type of model for ordered categorical responses (such as age groups) and it has previously been used for modelling ALKs (Kvist et al., 2000; Rindorf and Lewy, 2001). In addition to ALKs, Rindorf and Lewy (2001) also applied CRLs for estimating smooth length distributions.

CRLs have also been used to investigate spatial differences in ALKs (Gerritsen et al., 2006; Stari et al., 2010). In both cases, significant spatial differences were found in ALKs for North Sea haddock. Gerritsen et al. (2006) divided their data into 3 depth strata and examined the differences between using a single ALK and ALKs calculated for each stratum. The shallow stratum was significantly different from the deeper strata, with higher probabilities for younger fish in the shallow stratum. Using a combined ALK for all the strata resulted in nearly twice as many 1-year olds compared to a survey index calculated from the stratified ALKs. Stari et al. (2010) found significant differences between geographical areas, mature and immature fish, commercial and survey data, and fleets using different fishing gear.

In all previous applications of CRLs to ALK modelling, Generalized Linear Models (GLMs) have been used for estimation, and stratification has been used to model the effect of regional differences. Any type of stratification will exacerbate the problems with missing data, and the choice of strata will often be a somewhat subjective decision made by the modeller. In situations where detailed information about the geographical origin of the age samples is available, it is possible to consider alternatives to stratification by area.

One such alternative is to use Generalized Additive Models (GAMs) in place of GLMs. GAMs is a non-parametric tool for non-linear modelling, which allows smooth functions of the explanatory

* Corresponding author.
E-mail address: cbe@aqua.dtu.dk (C.W. Berg).

0165-7836/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.fishres.2012.06.016

variables in the specification of the mean value distribution, and numerous studies have used GAMs for modelling spatial effects. Toscas et al. (2009) used GAMs to fit spatio-temporal models of prawn catches with 2D thin plate regression splines for modelling spatial variation. The GAMs interpolated the data well, but extrapolation beyond data coverage was found to be problematic. Maxwell et al. (2012) compared GAMs to a stratified mean method (stratification in time and space) for estimating egg production of cod, plaice, and haddock in the Irish Sea. The methods gave relatively consistent estimates, but the GAM methodology offered higher precision and was better suited for handling missing observations. For a thorough introduction to GAMs see Wood (2006).

In this study we describe how GAMs can be used for fitting CRLs to model age as a smooth function of length and geographical position. In addition to the advantages offered by the GLM approach, it eliminates the need for spatial stratification by providing an ALK that varies smoothly with geographical position.

The methodology has been fully implemented in the DATRAS software package (Kristensen and Berg, 2012) for R (R Development Core Team, 2012), which offers an accessible way to create ALKs from all the data available in the DATRAS database (www.datras.ices.dk) as well as other data using this format.

Using ten years of survey data the new method is compared to the traditionally applied regional stratification of ALKs to determine whether a significantly better fit to data is obtained. Furthermore, internal and external consistencies are calculated to examine whether the new method leads to improved precision when used to create indices of abundance by age.

2. Methods

The response variable is the age group of a fish, $a = R \ldots A$, i.e., ordered categorical also known as ordinal response, where R denotes the youngest age category and A is the oldest. The latter category is often defined to be a "plus group" which consists of fish of age A or older. For each fish where the age has been determined, a set of covariates $\mathbf{x}$ is also available, which in this study includes the length l of the fish and the spatial coordinates of the fishery.

The continuation ratio model (Agresti, 2010) is well suited to model the distribution of ages $P_a(\mathbf{x}) = \{p_R \ldots p_A\}$. This is accomplished through A minus R models for the conditional probability of being of age a given that it is at least age a. That is, let

$$\pi_a = P(Y = a | Y \geq a) = \frac{p_a}{p_a + \cdots + p_A}, \quad a = R \ldots A - 1$$

be those conditional probabilities and let our set of continuation ratio logits be given by GAMs of the following type:

$$\text{logit}(\pi_a[\mathbf{x}_i]) = \mathbf{x}_i^* \boldsymbol{\theta}_a + f_{1a}(x_{1i}) + f_{2a}(x_{2i}, x_{3i}) + f_{3a}(x_{4i}) x_{5i} + \cdots, \quad a = R \ldots A - 1$$

where $\mathbf{x}_i$ is a vector of covariates, $\mathbf{x}_i^*$ is a subset of the covariates entering linearly in the model, $\boldsymbol{\theta}$ is the corresponding parameter vector, and f_j denotes some smooth function of the covariates x_k, which may be of one or more dimensions and also multiplied by known covariates. Given the set of $A - R$ models, we can calculate the estimated unconditional probabilities $\hat{p}_a$ from the conditional probabilities $\hat{\pi}_a$ (the dependence on covariates is omitted here):

$$\hat{p}_R = \hat{\pi}_R$$

$$\hat{p}_a = \hat{\pi}_a \left(1 - \sum_{j=R}^{a-1} \hat{p}_j\right) = \hat{\pi}_a \prod_{j=R}^{a-1} \left(1 - \hat{\pi}_j\right), \quad a > R$$

We choose to consider the following six formulations of the CRLs:

1. A single common ALK for the whole North Sea fitted using GLM methodology.
2. A stratified approach having separate ALKs within 3 subareas of the North Sea (see Fig. 1), also fitted using GLMs.
3. A smooth spatial varying ALK fitted using GAMs with smoothness selection by AIC. Only the intercept in the models are allowed to vary with location.
4. Same as model 4, but with smoothness selection by BIC instead of AIC.
5. A GAM where both the intercept and the regression coefficient on length are allowed to vary with geographical coordinates.
6. Like model 4, but with the same spatial effect in all years, as opposed to estimating a set of parameters for each year.

Using mathematical notation these six models can be written as follows:

$$\text{logit}(\pi_{ayq}[\mathbf{x}_i]) = \alpha_{ayq} + \beta_{ayq} l_i \tag{1}$$

$$\text{logit}(\pi_{ayq}[\mathbf{x}_i]) = \alpha_{ayq} + \delta_{ayq}(Area_i) + \omega_{ayq}(Area_i) l_i \tag{2}$$

$$\text{logit}(\pi_{ayq}[\mathbf{x}_i]) = \alpha_{ayq} + \beta_{ayq} l_i + s_{ayq,AIC}(lon_i, lat_i) \tag{3}$$

$$\text{logit}(\pi_{ayq}[\mathbf{x}_i]) = \alpha_{ayq} + \beta_{ayq} l_i + s_{ayq,BIC}(lon_i, lat_i) \tag{4}$$

$$\text{logit}(\pi_{ayq}[\mathbf{x}_i]) = \alpha_{ayq} + s_{ayq,BIC}(lon_i, lat_i) l_i + s_{ayq,BIC}(lon_i, lat_i) \tag{5}$$

$$\text{logit}(\pi_{ayq}[\mathbf{x}_i]) = \alpha_{ayq} + \beta_{ayq} l_i + s_{aq,BIC}(lon_i, lat_i) \tag{6}$$

where i denotes the ith fish, l denotes the length of the fish, (lon, lat) the geographical coordinates where the haul was taken (longitude and latitude), $\delta_a(Area_i)$ maps the ith observation to one of 3 categorical effects for a division of the North Sea into 3 areas (see Fig. 1), and similarly denotes ω_a a regression parameter for each of the 3 areas, s_a is a thin plate spline in two dimensions, where subscripts AIC and BIC denote which criterion is used for smoothness selection, and (α_a, β_a) are ordinary regression parameters to be estimated. Subscripts y and q have been included here to indicate that each combination of year and quarter should have a distinct set of parameters to account for differences in population structure.

Note, that model 2 is equivalent to dividing the data set according to the 3 areas and fitting model 1 with individual parameters for each area. Models 3 and 4 include a spatial varying intercept for each continuation ratio logit but a common regression parameter on length, whereas model 5 is a varying-coefficients model (Hastie and Tibshirani, 1993), where both the intercept and the regression parameter are allowed to vary with geographical coordinates. Model 6 is like model 4 except that the spatial effect is constrained to be identical for all years. All the parameters in the model has the subscript a indicating that each logit has a distinct set of parameters. This implies that the likelihood equation can be partitioned into separate terms for each logit (Agresti, 2010; Kvist et al., 2000), and hence each logit can be fitted separately. Similarly, the total deviance for the model is simply the sum of deviances from the individual fits. This feature makes it possible to fit the continuation ratio logit model using standard software that can handle binomial responses.

Our GAM models are based on the implementation in the mgcv package for R (Wood, 2006), which offers a variety of types including multi-dimensional splines and automatic smoothness selection. We follow the recommendation in Wood (2006), who suggests using thin plate regression splines for inputs on same scale and where isotropy is relevant such as spatial coordinates. All the thin plate splines used in this study for geographical effects are splines with shrinkage smoothing (Wood, 2006, p. 160), which

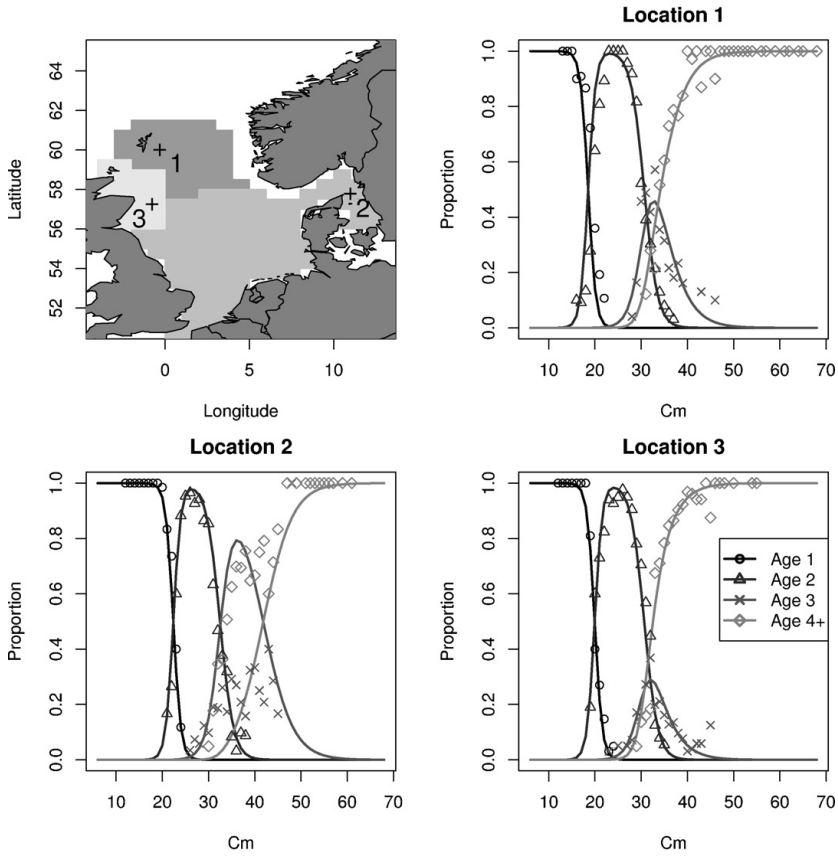


Fig. 1. Map of the three areas used in model 2 and of three selected locations 1,2 and 3 (top left). Predicted probabilities of age given length using model 4 at each location (solid lines) as well as the raw observed proportions (points) in each of the boxed areas for the year 2011 Q1 (top right and bottom figures).

allows them to be completely eliminated from the model in the sense of having all the parameters estimated to be zero.

2.1. Evaluating the implication of the ALK

Previous works (Gerritsen et al., 2006; Stari et al., 2010) have utilized the generalized likelihood ratio test for measuring whether two ALKs could be considered identical. This test requires that the smaller model is nested within the larger, which is not the case for all our models. Instead, we use the AIC and BIC values to investigate which models are more appropriate. For the GAMs, the number of parameters, which is needed to calculate the AIC and BIC, is replaced by the *effective degrees of freedom* (edf), see Wood (2006) for details.

However, since the AIC and BIC only applies to the age data, it does not tell us whether applying the estimated ALKs to all the length data will result in significant changes in an index of abundance by age. We will therefore create such an index to investigate the implications of our proposed method for creating ALKs. If a spatial ALK, in addition to providing a better fit to the age data, also results in improved precision for a derived index of abundance, this can be seen as further evidence that the spatial ALK is more appropriate. If the spatial effect in the ALKs were really noise rather than a true signal, one would expect the precision of an index of abundance to deteriorate when applying a spatial ALK as opposed to a non-spatial ALK.

We choose one of the simplest estimates of abundance:

$$I_{ayq} = \frac{1}{h_{yq}} \sum_{i=1}^{n_{yq}} \hat{p}_a(\mathbf{x}_i) \tag{7}$$

where I_{ayq} is the average predicted number of fish caught in age group a per haul in year and quarter (y, q), n is the total number of fish caught, and h is the number of hauls.

An appropriate way to test whether one index of abundance is more accurate than another would be to run full assessment models using the different indices as well as commercial catch data and compare their estimated observation variances. However, since this is a quite complicated task we choose a simpler way of comparing our different indices of abundance based on the concepts of internal and external consistency (e.g., Payne et al., 2009). Under the assumptions that an index is proportional to the abundance without error and of constant catchability and constant total mortality over time, the logarithm of the abundance at time t should be perfectly correlated with the logarithm of abundance of the same cohort at time $t + \Delta t$. Although all these assumptions are not

correct, we should still be able to obtain significant correlations for values of Δt within the range of a year, given that the signal in the time-series outweighs the variability from sampling noise and violations of our assumptions. Recapping from Payne et al. (2009), if we assume that we have a survey index with a log-normal error structure and substitute this into the Baranov catch equation we get

$$I_{a(t)} = q_{a(t)} N_{a(t)} \epsilon_{a(t)}, \quad \epsilon \sim LN(0, \sigma_{a(t)})$$

$$\log(I_{a(t)}) = \log(I_{a(t+\Delta t)}) + \log\left(\frac{q_{a(t+\Delta t)}}{q_{a(t)}}\right) + \log(\epsilon_{a(t+\Delta t)}) - \log(\epsilon_{a(t)}) - Z_a(t, t+\Delta t)$$

where $I_{a(t)}$ refers to the index of abundance for some age group a at time t, q denotes catchability, Z the total mortality over the considered time interval, and ϵ is a random log-normal distributed component.

Internal consistency refers to correlations between Is within the same survey index (e.g., Age 1 in quarter 1 year y versus Age 2 in quarter 1 in year $y+1$), whereas external consistency refers to comparing two independent survey indices, such as those for quarter 1 and 3 (Q1 and Q3). We will refer to internal consistency between age a and $a+1$ ($\Delta t = 1$ year) in quarter q as $IC(q, a)$ and external consistency between the same age classes in Q1 and Q3 ($\Delta t = 0.5$ years) as $EC(a)$.

3. Case studies

In this section the method will be applied to ten years (2001–2011) of data from the International Bottom Trawl Survey (IBTS) obtained from the DATRAS database (www.datras.ices.dk). The samples are collected in the first and third quarters of the year and all samples are caught using the same gear type. For further details about the IBTS survey see (ICES, 2012). In section 3.1 we will investigate an application of models 1 through 6 on North Sea haddock data. Section 3.2 will deal with a less detailed rerun of models 1 and 4 on multiple species focusing on consistencies only.

3.1. Haddock using models 1–6

For the area stratified model 2 we divide the North Sea into 3 areas (Fig. 1) with roughly the same number of age samples per year (see tables in online supplemental material[1]). Area 2 is much larger than the others, but this is due to the fact that haddock is primarily caught in the northern parts of the North Sea. For all models except model 2, it was possible to consider up to age group 8 without estimation problems. However, for simplicity we consider the age groups 1 to 4+ for all models, where the last group consists of fish of age 4 or older. As age 0 appears for the first time in the IBTS survey in Q3, it must also be included when creating the ALKs for this time-series, but results of this estimation are not included in the further analysis.

Table 1 shows the AIC and BIC calculated for each combination of model and quarters. Since lower values of AIC and BIC are to be preferred, model 2 is consistently better than model 1, implying that there is significant geographical variation in the ALKs. Model 3 is consistently best with respect to AIC while models 6 and 5 are respectively best with respect to BIC for Q1 and Q3, but the differences are much smaller between models 3–5 than the rest. These values provide strong evidence against a null hypothesis of no spatial effect in the ALKs, and also that the stratified GLM approach did not sufficiently capture the spatial variation.

Fig. 1 shows the fitted distribution (model 4) of age given length at three selected locations, as well as the raw observed proportions within each stratum. The observed proportions seem to differ between areas, and the fit in the three chosen locations resembles the raw observations in the three strata. We should note, that we cannot expect the fitted distributions to be the best interpolation of the raw proportions since the raw proportions are calculated over the entire stratum, but the shown fitted distributions applies only to the points in space marked by the numbers on the map, and the fits will therefore vary over the strata due to the significant spatial effect in the model.

Table 1
Haddock: summary of models 1–6. The columns 'ΔAIC' and 'ΔBIC' contain the decrease in AIC and BIC from model 1, and the best values are shown in bold face. The column 'edf' contains the effective number of parameters.

Model	Quarter	edf	ΔAIC	ΔBIC
1	1	66	0	0
2	1	198	3028.36	1940.47
3	1	770.10	**10,155.53**	4352.64
4	1	361.13	9252.17	6819.83
5	1	396.28	9600.90	6878.87
6	1	125	7659.33	**7173.11**
1	3	66	0	0
2	3	198	2966.02	1858.38
3	3	936.62	**10,605.83**	3300.25
4	3	384.29	9268.28	6597.43
5	3	421.89	9720.27	**6733.91**
6	3	130.88	6685.59	6141.19

Fig. 2 shows the spatial pattern in the probability of being older than one year given a length of 20 cm in 2011 Q1. The figure illustrates that there is spatial contrast in the data with a peak east of the Scottish coast. Given that a 20 cm haddock is caught in this region, it is more likely to be 2 years or older than being 1 year old, whereas the opposite is true in the south-eastern parts of the North Sea.

In order to illustrate the differences between models 1–6, the estimated age probabilities for a 30 cm haddock along a selected route (Fig. 3) from each model in year 2001 Q3 are shown in Fig. 4. The same plots for all the years and quarters can be found in the

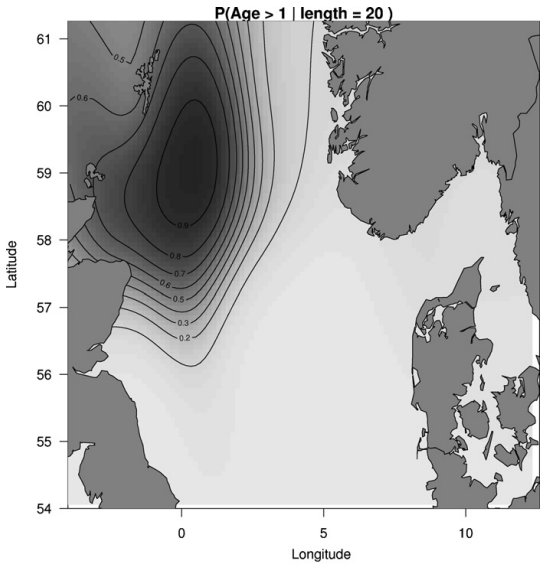


Fig. 2. Contour plot of the estimated probability (model 4) of being older than 1 year given a length of 20 cm in year 2011 Q1.

[1] See Appendix A.

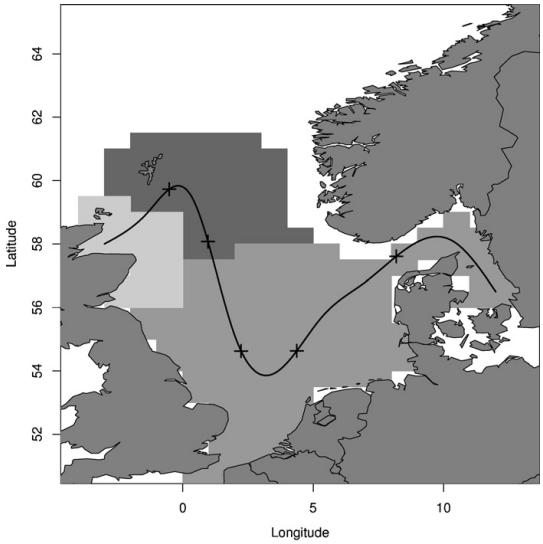


Fig. 3. A selected route through the North Sea and some selected points marked by '+'.

online supplemental materials. The models based on GAMs (3–6) all show a steep increase in probability for observing younger individuals on last part of the route around the Skagerrak region. Although there is considerable variation between years and quarters in the estimated probabilities, the spatial pattern seems to be relatively consistent. This is also supported by the fact, that model 6, which has the same spatial effect over all the years, was chosen as the best model by the BIC criterion for Q3. Models 4 and 5 display very similar results, while model 3 in some years estimates some more wiggly curves in comparison, due to the AIC criterion being less restrictive than BIC in terms of the amount of smoothing.

To illustrate the implications of using the different models for our simple index of abundance we have plotted $\log(I_{2yq})$ and $\log(I_{3yq})$ in Fig. 5. There seems to be very high consistencies between the series, both internally and externally, for all ALKs. This implies, that even though significant differences were found between the ALKs, the resulting indices of abundance turned out to be quite similar. The uncertainties on the indices of abundance were further investigated using bootstrapping (not shown), and these analyses confirmed that the difference between the calculated indices were generally not statistically different.

The internal and external consistencies are shown in Table 2, which confirms the apparent high correlation observed in Fig. 5,

Table 2
Haddock: internal and external consistencies for models 1–6. Internal consistency between age a and $a+1$ in quarter q is referred to as $IC(q, a)$ and external consistency between the same age classes in Q1 and Q3 as $EC(a)$. Best average consistency is shown in bold face.

Type\Model	1	2	3	4	5	6
IC(Q1, 1)	0.961	0.955	0.956	0.954	0.954	0.961
IC(Q1, 2)	0.910	0.919	0.918	0.917	0.916	0.918
IC(Q3, 1)	0.951	0.950	0.949	0.968	0.973	0.966
IC(Q3, 2)	0.970	0.976	0.944	0.963	0.968	0.969
EC(1)	0.972	0.973	0.969	0.968	0.969	0.955
EC(2)	0.985	0.993	0.980	0.992	0.992	0.994
EC(3)	0.921	0.948	0.954	0.956	0.945	0.963
Avg	0.953	0.959	0.953	0.960	0.959	**0.961**

Table 3
Haddock: internal and external consistencies for models 1 and 4. Internal consistency between age a and $a+1$ in quarter q is referred to as $IC(q, a)$ and external consistency between the same age classes in Q1 and Q3 as $EC(a)$. Best average consistencies are shown in bold face.

	IC(Q1, x)		IC(Q3, x)		EC(x)	
x\Model	1	4	1	4	1	4
1	0.96	0.95	0.93	0.97	0.97	0.97
2	0.91	0.92	0.95	0.97	0.99	0.99
3	0.95	0.95	0.97	0.96	0.92	0.96
4	0.93	0.95	0.97	0.96	0.93	0.99
5	0.94	0.97	0.98	0.99	0.95	0.96
6	0.88	0.95	0.92	0.93	0.91	0.95
7	0.73	0.68	0.94	0.92	0.88	0.90
Avg	0.90	**0.91**	0.95	**0.96**	0.94	**0.96**

Table 4
Cod: internal and external consistencies for models 1 and 4. Internal consistency between age a and $a+1$ in quarter q is referred to as $IC(q, a)$ and external consistency between the same age classes in Q1 and Q3 as $EC(a)$. Best average consistencies are shown in bold face.

	IC(Q1, x)		IC(Q3, x)		EC(x)	
x\Model	1	4	1	4	1	4
1	0.56	0.68	0.86	0.85	0.91	0.93
2	0.71	0.88	0.19	0.31	0.77	0.75
3	0.87	0.83	0.36	0.57	0.43	0.49
4	0.66	0.63	0.30	0.27	0.42	0.48
5	0.37	0.33	0.40	0.37	0.58	0.64
Avg	0.63	**0.67**	0.42	**0.47**	0.62	**0.66**

which implies a very strong signal in data. On average, models 4–6 have higher consistencies than the rest, which validates our conclusion that there is a spatial effect and that the GAM framework outperforms the stratified approach.

3.2. Models 1 and 4 on more species

Tables 3–6 show internal and external consistencies for models 1 and 4 for cod, haddock, whiting and herring in the North Sea. The choice of model 4 among the different GAM formulations was rather arbitrary, although it can be considered the more conservative choice with respect to the amount of spatial variation in the ALKs, as it uses the fewest number of effective parameters of the GAMs. Since we do not consider model 2, we can include a higher number of age groups without worrying about years with no observations of older age groups. For all species except herring, model 4 is consistently better than model 1 with respect to average consistency over age groups. While haddock has very high consistencies even in older age classes, herring has appalling consistencies for Q1 (some are even negative). Whiting and Cod have fairly good consistencies, perhaps with the exception of IC(Q3) for cod (4). These

Table 5
Whiting: internal and external consistencies for models 1 and 4. Internal consistency between age a and $a+1$ in quarter q is referred to as $IC(q, a)$ and external consistency between the same age classes in Q1 and Q3 as $EC(a)$. Best average consistencies are shown in bold face.

	IC(Q1, x)		IC(Q3, x)		EC(x)	
x\Model	1	4	1	4	1	4
1	0.79	0.76	0.70	0.72	0.84	0.86
2	0.96	0.98	0.83	0.82	0.85	0.84
3	0.86	0.87	0.76	0.78	0.88	0.90
4	0.63	0.65	0.85	0.85	0.67	0.67
5	0.37	0.47	0.85	0.84	0.57	0.57
Avg	0.72	**0.75**	0.80	0.80	0.76	**0.77**

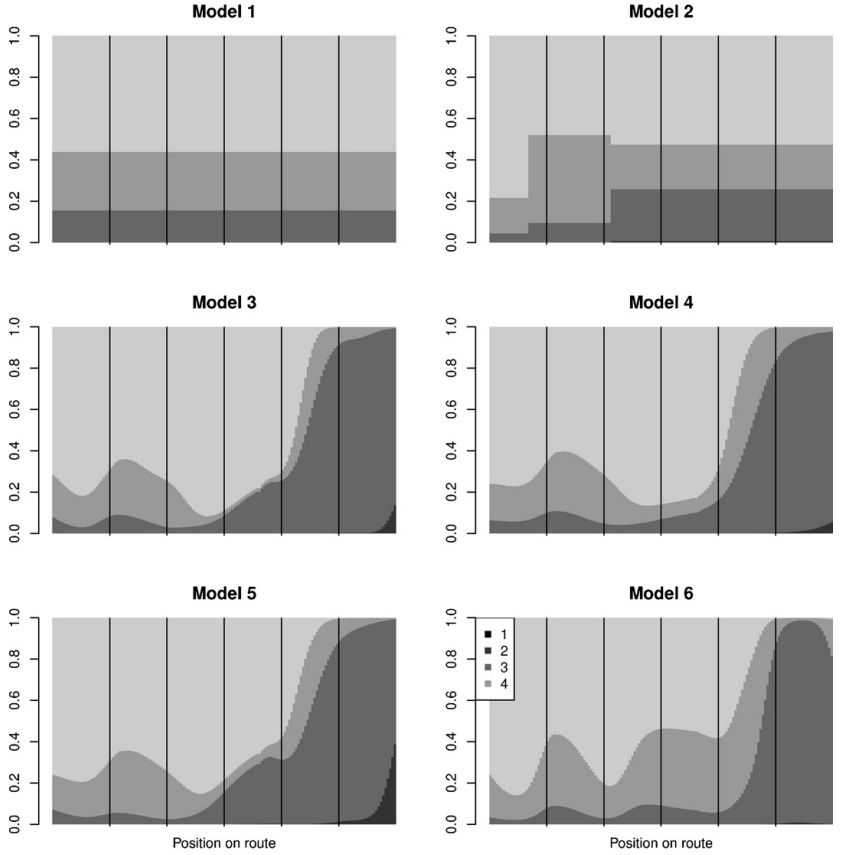


Fig. 4. Estimated age probabilities for a 30-cm haddock along the route shown in Fig. 3 from models 1 to 6 for year 2001 Q3. The x-axis corresponds to the position on the route from west to east, and the vertical lines indicate the positions marked with a '+' on the map.

results emphasize the results found for haddock, namely that there generally is spatial variation in length-at-age, and that improved precision in indices of abundance can be obtained by including this variation in the ALKs.

4. Discussion

Several studies have suggested using continuation ratio logits for modelling the age distribution in catch data from length-stratified subsamples of age in place of raw proportions of age given length. Two studies have also shown regional as well as other effects using CRLs for North Sea haddock (Gerritsen et al., 2006; Stari et al., 2010), a result that is confirmed in this study. While these studies used a number of parameters proportional to the number of boxed areas using GLM methodology, we propose to use GAM methodology to model spatial effects as a smooth surface and thereby be able to predict numbers-at-age at the haul level, whenever the required information is available. This removes the problem of having to select appropriate boxes for the data, and the problem of missing data whenever a too fine-grained stratification is chosen. This effect is comparable to the result found in Maxwell et al. (2012), who compared GAMs with a stratified mean method for modelling egg production in fishes. Also, the ALKs based on GAMs provided a much better fit to data than the GLM based methods examined in this study, and they were also superior in terms of both AIC and BIC. Our proposed model allows for a higher number of age groups than usual to be considered when an age based index of abundance is to be created, and, although there were only small differences in the survey indices between ALK methods, our results indicated that including spatial variation in ALKs seemed to improve the precision of the indices. It is straightforward to expand the number of covariates used in this study, using the same

Table 6
Herring: internal and external consistencies for models 1 and 4. Internal consistency between age a and $a+1$ in quarter q is referred to as $IC(q, a)$ and external consistency between the same age classes in Q1 and Q3 as $EC(a)$. Best average consistencies are shown in bold face.

	$IC(\text{Q1}, x)$		$IC(\text{Q3}, x)$		$EC(x)$	
x\Model	1	4	1	4	1	4
1	0.22	0.36	0.58	0.56	0.58	0.57
2	−0.14	−0.09	0.78	0.77	0.28	0.34
3	0.07	−0.07	0.71	0.69	0.44	0.55
4	0.23	0.05	0.76	0.75	0.61	0.49
5	0.24	0.35	0.80	0.83	0.63	0.54
Avg	0.12	0.12	**0.73**	0.72	**0.51**	0.50

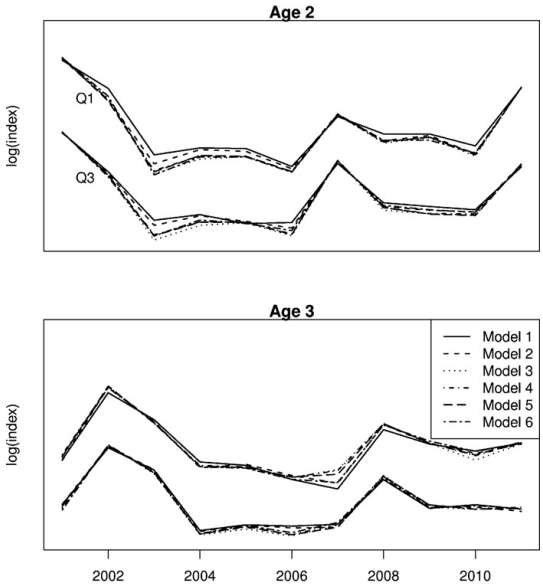


Fig. 5. Index of abundances for age groups 2 and 3 in Q1 and Q3. The series have been rescaled to prevent overlap between Q1 and Q3 for better overview, so only relative comparisons of the time-series are meaningful.

technique. While a spatial smoother is a convenient way of modelling the observed differences in ALKs between areas, it does not offer us an explanation for the observed effects. Possible explanations could be regional differences in growth, but also local variation in relative abundance of age-classes, which can occur due to migration, local differences in natural mortality, or even effects due to the data collection such as different laboratories used for ageing. In other words, the observed differences might be equally well explained by other covariates not included in our models, but given adequate spatial overlap between for instance different ageing labs, it will be possible to test for such an effect within our model framework while still accounting for residual unexplained spatial correlations by including a thin plate regression spline. Very high internal and external consistencies were found for all the age classes examined for haddock. While lack of consistency points to problems with some of the usual assumptions made for survey indices, strong consistencies are not proof of an excellent survey index, e.g., a constant index, which could hardly be informative, would yield perfect consistencies. We found fairly good consistencies for whiting and cod, but poorer consistencies for herring.

To ensure that changes in catch rates are due to changes in the population size rather than changes in survey design or other factors, survey indices should be standardized in some way to make them representative for the stock and comparable between years, a process which is sometimes called catch-rate standardization (e.g., Maunder and Punt, 2004). We should however keep in mind, that we did not perform a proper catch-rate standardization, but instead used a very simple index based on average numbers per haul. Also, catch rates for herring are generally much more variable than for the other species considered in this study, which can explain why our simple index performs so poorly for herring.

We should note, that even though many stock assessment models use age-structured indices of abundance as input, alternatives exist such as purely length-based models (e.g., Kristensen et al., 2006) or integrated stock assessments (e.g., Fournier et al., 1998) in which the separation into age-classes is performed within the stock assessment model, such that the associated uncertainty is included in the estimation. For stock assessments it should certainly be preferred to include the uncertainties due to the ALK estimation, either by integrating the ALK estimation within the stock assessment model, or to estimate the uncertainties on the derived indices of abundance by age outside the model, and provide these uncertainties as input to the stock assessment model along with the indices. The latter approach could be accomplished by bootstrapping, and is possible to carry out using the DATRAS-package.

Another useful aspect of ALKs is to combine them with the distributions of length and apply Bayes formula to get the probability of length given age (as opposed to age given length in ALKs), which for instance can be used to examine growth or differences in length distributions between regions. This idea was pursued in Rindorf and Lewy (2001) where CRLs were used for both the ALKs as well as the length distributions. The idea is, that since length distributions suffer from the same problems as age distributions, namely being patchy when small areas or individual hauls are considered, CRLs can be used to obtain smooth length distributions. Rindorf and Lewy (2001) used a seventh degree polynomial to obtain the length distributions on different locations, but noted that other types of smooth functions could be considered. GAMs could be considered in this respect, and this could be an interesting area for future research.

Fisheries data can be very complex, and the data sets available from DATRAS are certainly no exception to this rule. Producing an age-based survey index, which includes the application of an ALK, is therefore often a challenging task, and reproducing them by other people even more so. We have provided a software package for R that allows for manipulation of data from the DATRAS database, and easy generation and application of robust ALKs without the need for area stratification. The software package and all its source code is publicly available (Kristensen and Berg, 2012), which allows for adaptation to other data sets than those from the DATRAS database, including samples from commercial fisheries. Example code showing how to reproduce the models found in this paper is included in the online supplemental material. We have shown, that our approach is superior to the stratified approach with respect to AIC and BIC, and that it generally leads to better internal and external consistencies for age based survey indices.

Acknowledgements

The authors wish to thank all staff involved in the IBTS data collection, ICES for providing easy access to the data and answering our questions about how to interpret them, two anonymous reviewers for their constructive comments, and finally we wish to thank Anna Rindorf and Peter Lewy for valuable input to this manuscript.

Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fishres.2012.06.016.

References

Agresti, A., 2010. Analysis of Ordinal Categorical Data. Wiley Series in Probability and Statistics. Wiley.
Fournier, D.A., Hampton, J., Sibert, J.R., 1998. Multifan-cl: a length-based, age-structured model for fisheries stock assessment, with application to sout Pacific albacore, *Thunnus alalunga*. Can. J. Fish. Aquat. Sci. 55, 2105–2116.
Gerritsen, H.D., McGrath, D., Lordan, C., 2006. A simple method for comparing age-length keys reveals significant regional differences within a single stock of haddock (*Melanogrammus aeglefinus*). ICES J. Mar. Sci. 63, 1096–1100.
Hastie, T., Tibshirani, R., 1993. Varying-coefficient models. J. Roy. Stat. Soc. B 55, 757–796.

ICES, 2012. Datras: Survey Descriptions. http://datras.ices.dk/Home/Descriptions.aspx.
Kristensen, K., Berg, C.W., 2012. Datras Package for R. http://rforge.net/DATRAS/.
Kristensen, K., Lewy, P., Beyer, J.E., 2006. How to validate a length-based model of single-species fish stock dynamics. Can. J. Fish. Aquat. Sci. 63, 2531–2542.
Kvist, T., Gislason, H., Thyregod, P., 2000. Using continuation-ratio logits to analyze the variation of the age composition of fish catches. J. Appl. Stat. 27, 303–319.
Maunder, M.N., Punt, A.E., 2004. Standardizing catch and effort data: a review of recent approaches. Fish. Res. 70, 141–159.
Maxwell, D.L., Armstrong, M.J., Beggs, S., Aldridge, J.N., 2012. Annual egg production estimates of cod (*Gadus morhua*), plaice (*Pleuronectes platessa*) and haddock (*Melanogrammus aeglefinus*) in the Irish sea: the effects of modelling choices and assumptions. Fish. Res. 117–118, 146–155. Egg Production Methods in Marine Fisheries.
Payne, M.R., Clausen, L.W., Mosegaard, H., 2009. Finding the signal in the noise: objective data-selection criteria improve the assessment of western Baltic spring-spawning herring. ICES J. Mar. Sci. 66, 1673–1680.
R Development Core Team, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0.
Rindorf, A., Lewy, P., 2001. Analyses of length and age distributions using continuation-ratio logits. Can. J. Fish. Aquat. Sci. 58, 1141–1152.
Stari, T., Preedy, K.F., McKenzie, E., Gurney, W.S., Heath, M.R., Kunzlik, P.A., Speirs, D.C., 2010. Smooth age length keys: observations and implications for data collection on north sea haddock. Fish. Res. 105, 2–12.
Toscas, P.J., Vance, D.J., Burridge, C.Y., Dichmont, C.M., Zhou, S., Venables, W.N., Pendrey, R.C., Donovan, A., 2009. Spatio-temporal modelling of prawns in albatross bay, karumba and mornington island. Fish. Res. 96, 173–187.
Wood, S., 2006. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC.

Online Supplemental Materials for "Spatial age-length key modelling using continuation ratio logits"

Casper W. Berg and Kasper Kristensen

Example of fitting ALKs using the DATRAS package

```
## Install and load DATRAS-package and data
## The data can be downloaded from the DATRAS homepage:
## http://datras.ices.dk/Data_products/Download/Download_Data_public.aspx
install.packages("DATRAS",,"http://rforge.net/",type="source")
library(DATRAS)
dAll <- readExchange("/exchange/IBTS/2001-2011.zip")

## create 3 areas by merging Roundfish areas.
dAll$rfpool=NA
conv=c("1"=1,"2"=2,"3"=3,"4"=2,"5"=2,"6"=2,"7"=2,"8"=2,"9"=2)
dAll$rfpool=as.factor( conv[as.character(dAll$Roundfish)] )

## choose North Sea haddock in quarter 1,
## only valid hauls
##  and only hauls where all standard species have been identified.
dQ1=subset(dAll,Species=="Melanogrammus aeglefinus",Roundfish %in% 1:9,
           Quarter==1,Year %in% 2001:2011,HaulVal=="V",StdSpecRecCode==1,Gear=="GOV")

## prepare data by adding spectrum, use 1 cm groups.
dQ1=addSpectrum(dQ1,by=1)

## Split data by year
dQ1.ysplit = split(dQ1,dQ1$Year)

## Declare settings for models 1-5:
mf = list( "cra~LngtCm","cra~LngtCm*rfpool",NULL,NULL,NULL)
gammas=c(NA,NA,1.4,NA,NA)
ack=c(FALSE,FALSE,TRUE,TRUE,TRUE)
useBICs=c(FALSE,FALSE,FALSE,TRUE,TRUE)
varCofs=c(FALSE,FALSE,FALSE,FALSE,TRUE)
maxKs=c(49,49,49,49,49)

## Fit model 1-5:
models=list()
for(m in 1:5){
  models[[m]] = lapply(dQ1.ysplit,fitALK,minAge=1,maxAge=4,model=mf[[m]],
                       gamma=gammas[m],autoChooseK=ack[m],
                        useBIC=useBICs[m],varCof=varCofs[m],maxK=maxKs[m])
}


## Model 6 must be fitted on all year simultaneously
## and 'Year' must be included in the model formula
models[[6]]=fitALK(dQ1,minAge=1,maxAge=8,
               model="cra~Year*LngtCm+s(lon,lat,k=25,bs='ts')",gamma=4)

## Predict numbers at age for model 4
Nage.model4=lapply(models[[4]],predict)

## plot simple survey index for model 4
sIndexModel4=sapply(Nage.model4,function(x) colSums(x)/nrow(x))
matplot(2001:2011,t(log(sIndexModel4)),type="b",ylab="log(I)")
```

Supplemental Tables

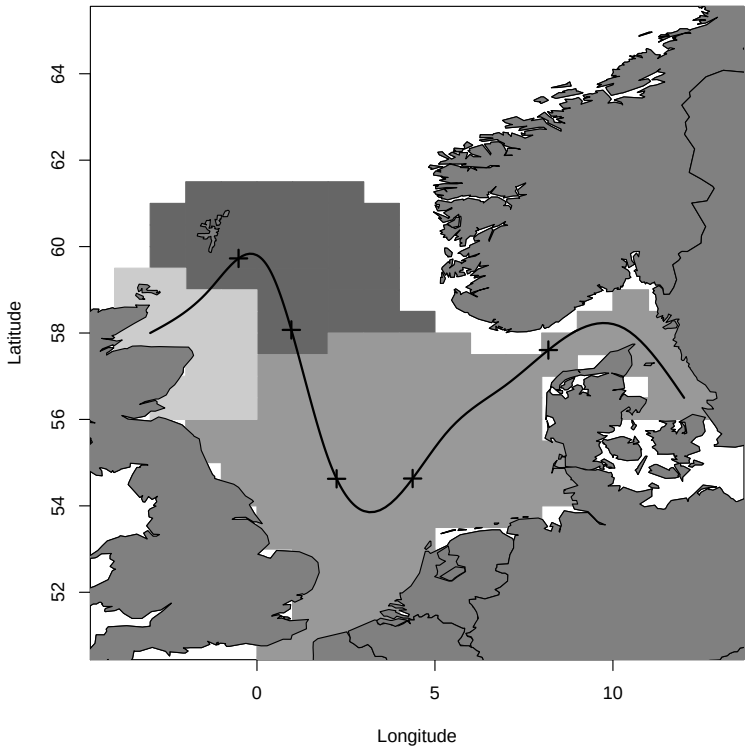
Year \ Area	1	2	3
2001	862	2909	917
2002	646	1891	744
2003	744	1747	1054
2004	730	1417	1089
2005	1061	1056	779
2006	492	743	306
2007	398	1215	490
2008	934	1388	854
2009	910	1174	861
2010	995	1432	936
2011	1064	1637	912

Table 1: Number of age samples in Q1

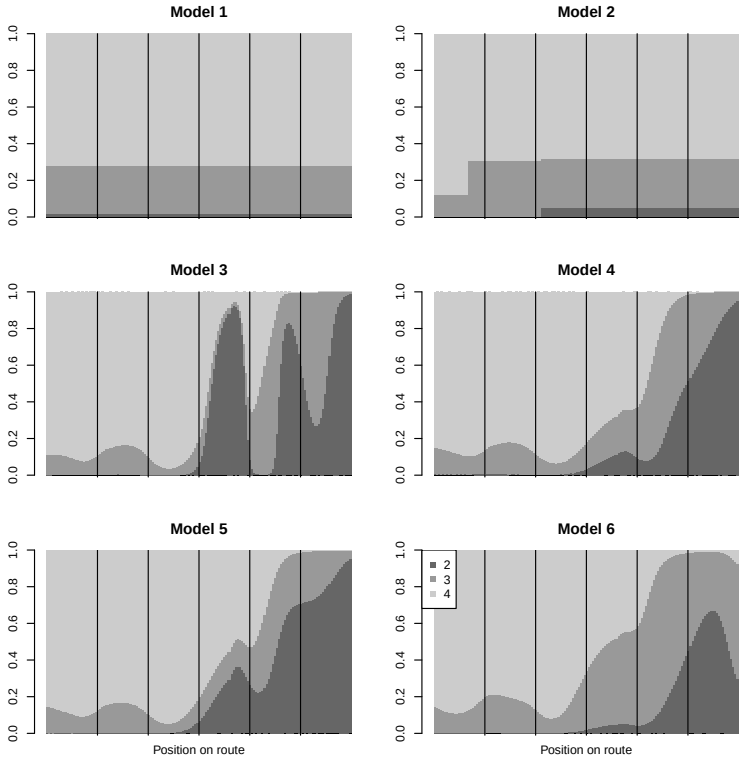
Year \ Area	1	2	3
2001	1472	1735	957
2002	1268	1497	793
2003	1153	1681	962
2004	1218	1274	903
2005	868	1588	630
2006	1566	1868	863
2007	1343	1452	772
2008	1685	1248	729
2009	932	1270	691
2010	1683	1720	794
2011	1468	1364	640

Table 2: Number of age samples in Q3

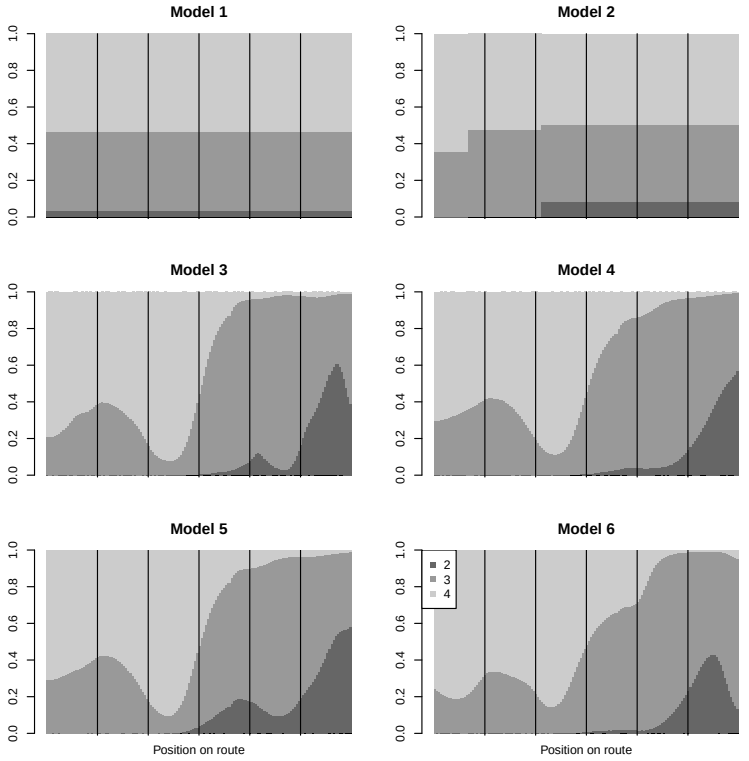
Supplemental Figures



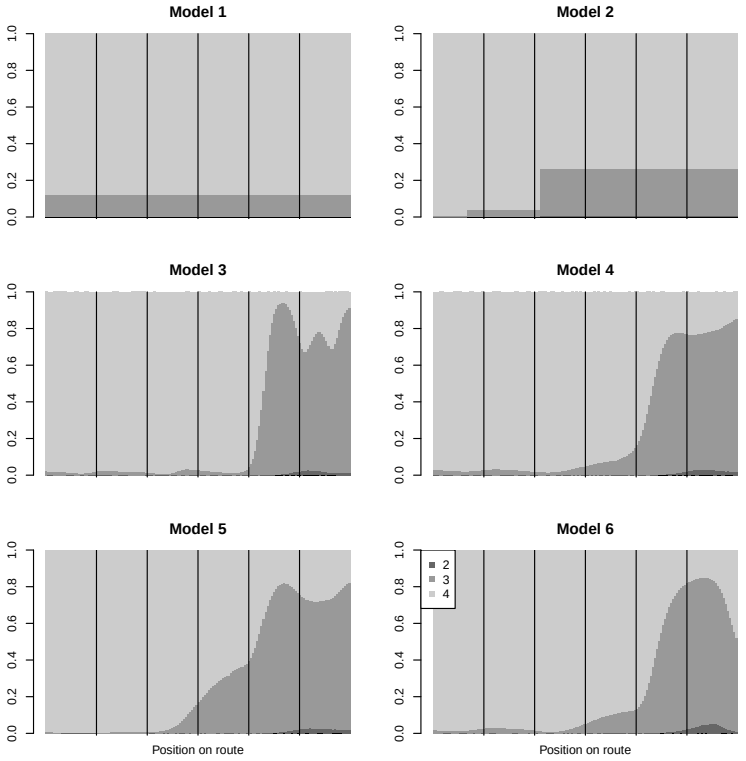
Supplemental Figure 1: A selected route through the North with some selected points marked by '+'.



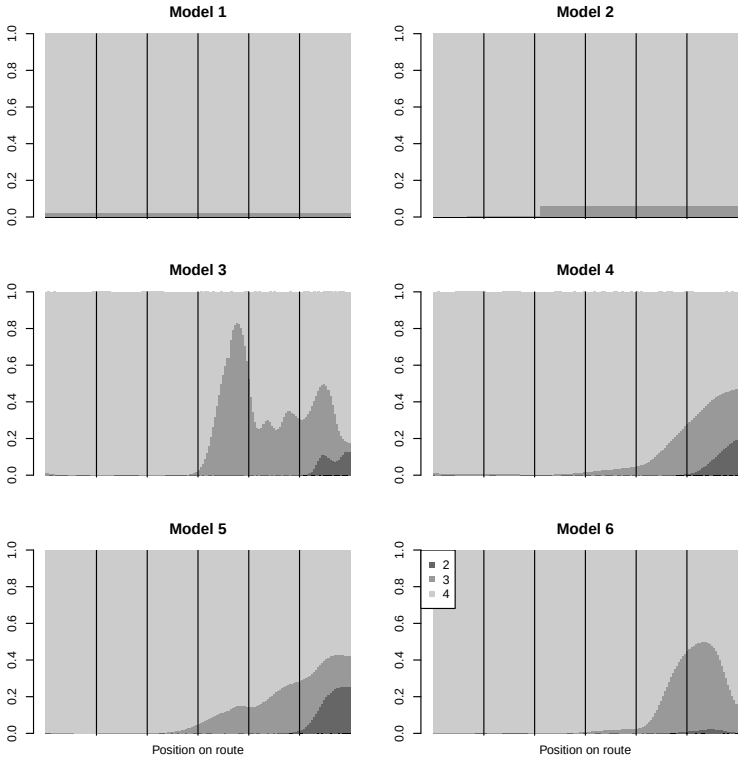
Supplemental Figure 2: Estimated age probabilities for a 30 cm haddock along the route shown in figure 1 from models 1-6 for year 2001 Q1. The x-axis corresponds to the position on the route from west to east, and the vertical lines indicate the positions marked with a '+' on the map.



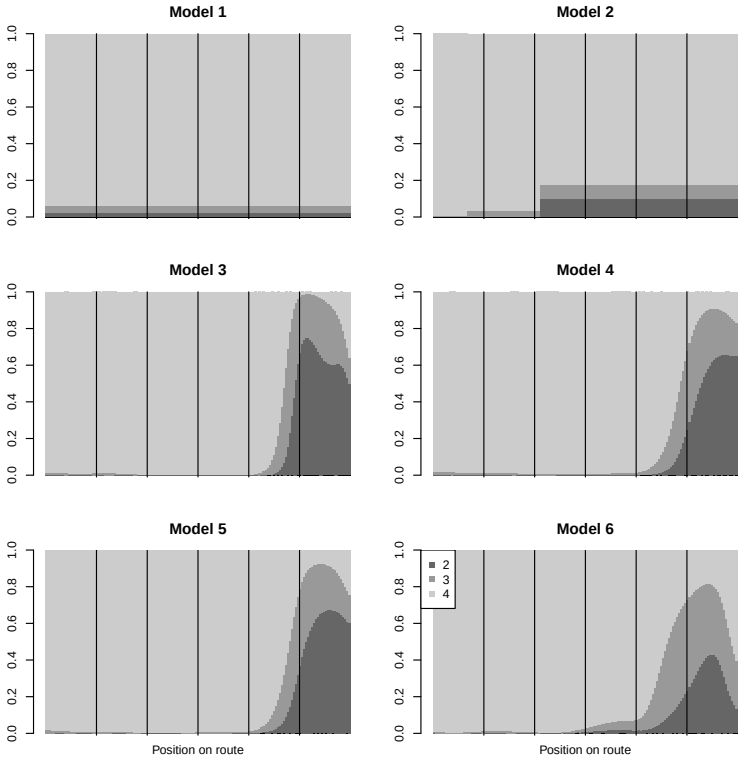
Supplemental Figure 3: As figure 2, but for 2002 Q1.



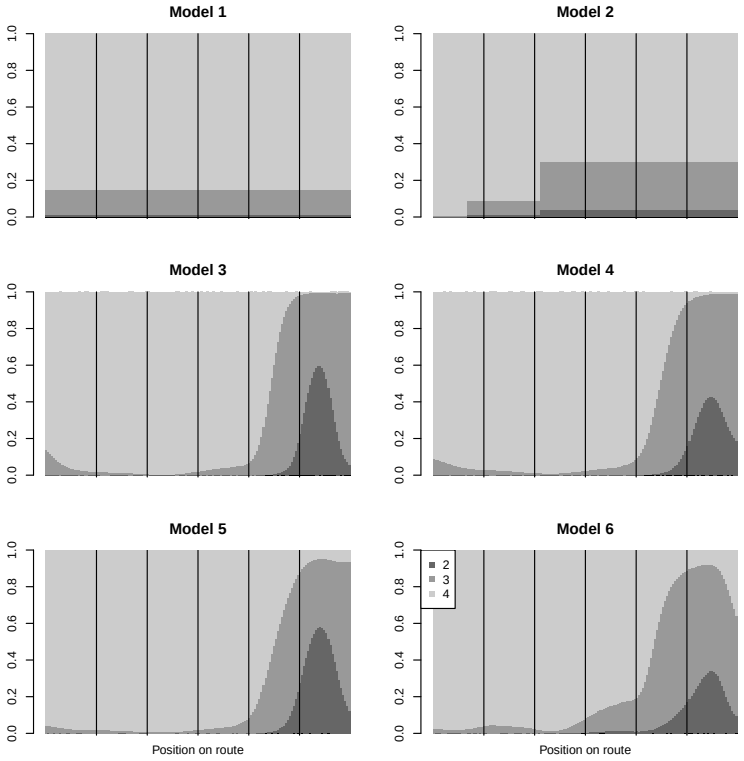
Supplemental Figure 4: As figure 2, but for 2003 Q1.



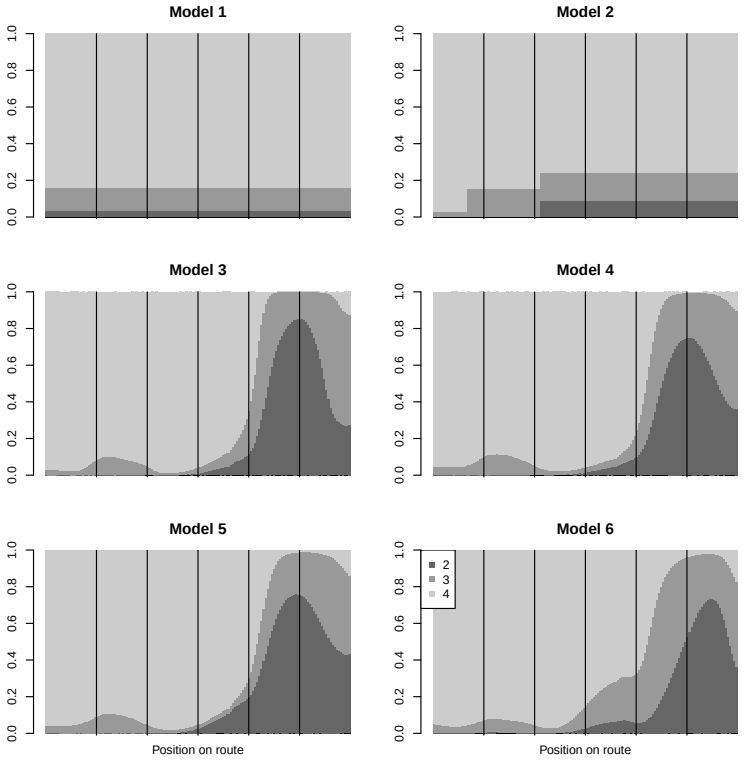
Supplemental Figure 5: As figure 2, but for 2004 Q1.



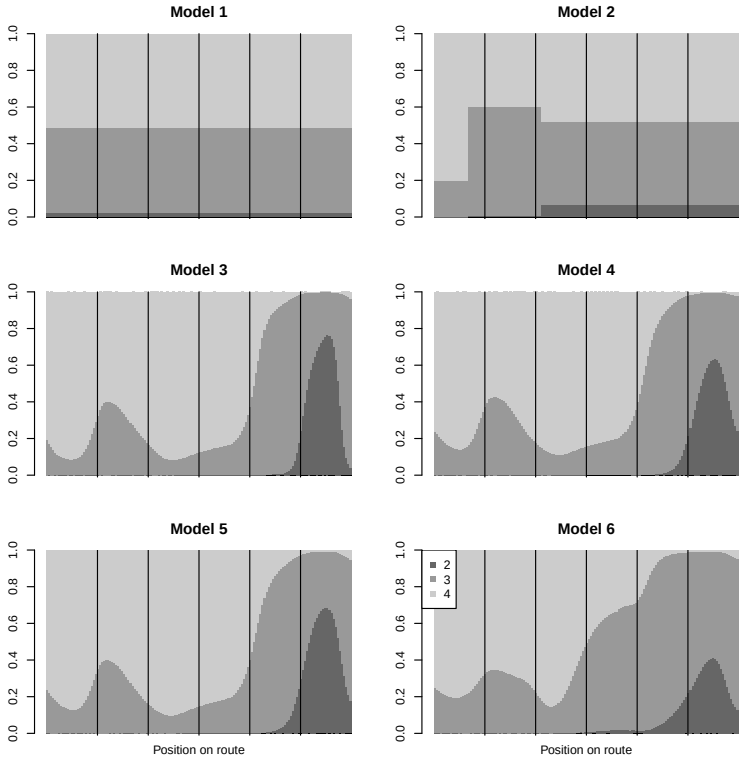
Supplemental Figure 6: As figure 2, but for 2005 Q1.



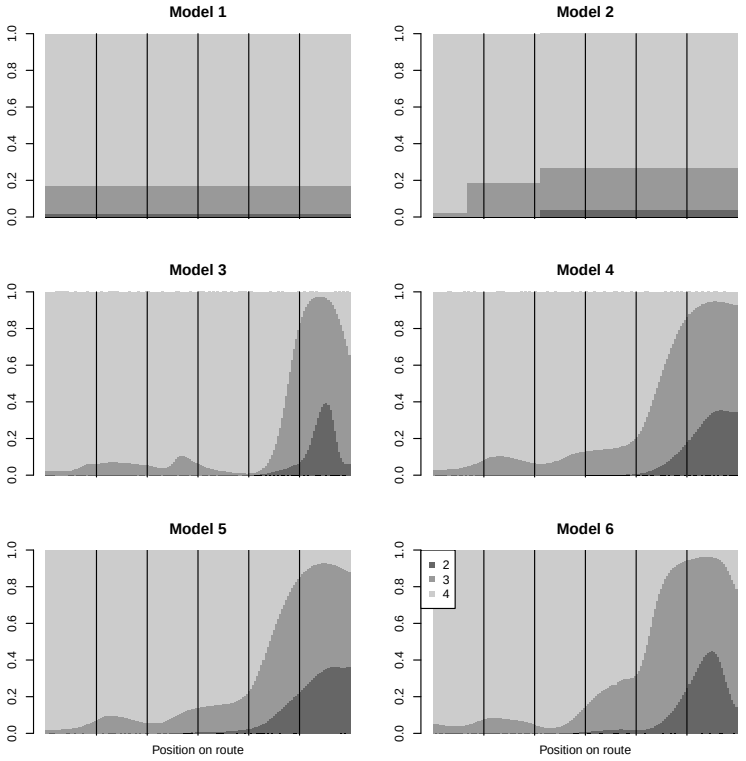
Supplemental Figure 7: As figure 2, but for 2006 Q1.



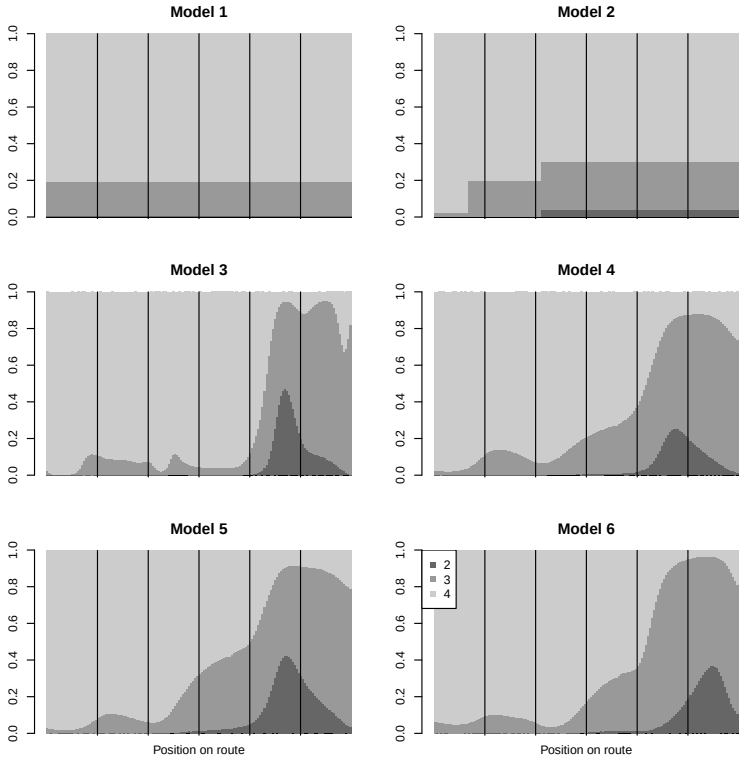
Supplemental Figure 8: As figure 2, but for 2007 Q1.



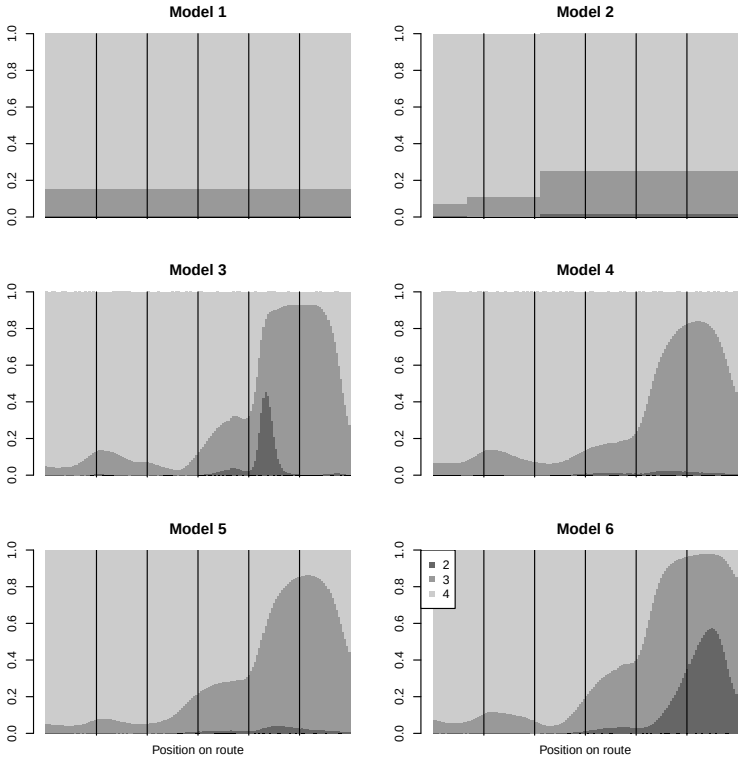
Supplemental Figure 9: As figure 2, but for 2008 Q1.



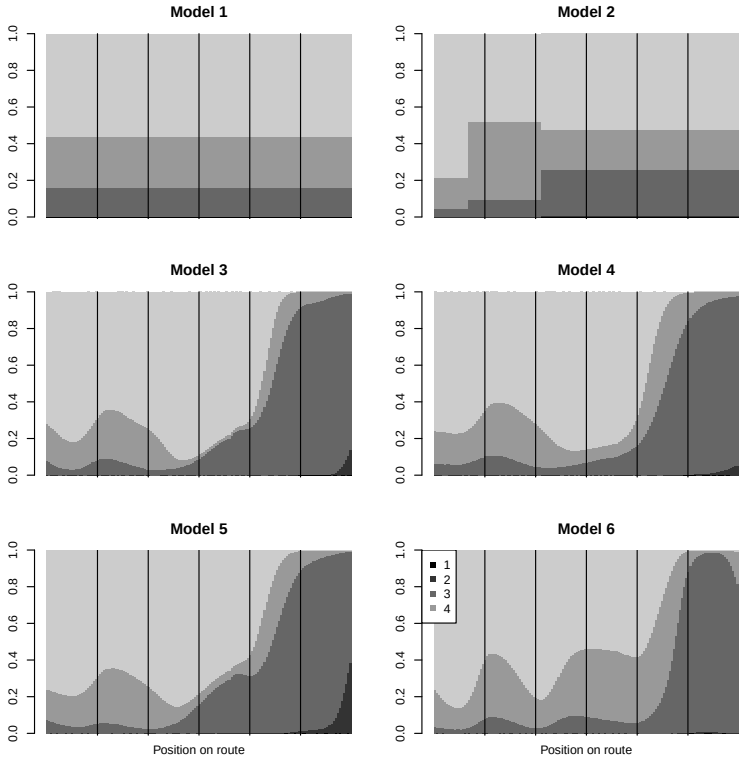
Supplemental Figure 10: As figure 2, but for 2009 Q1.



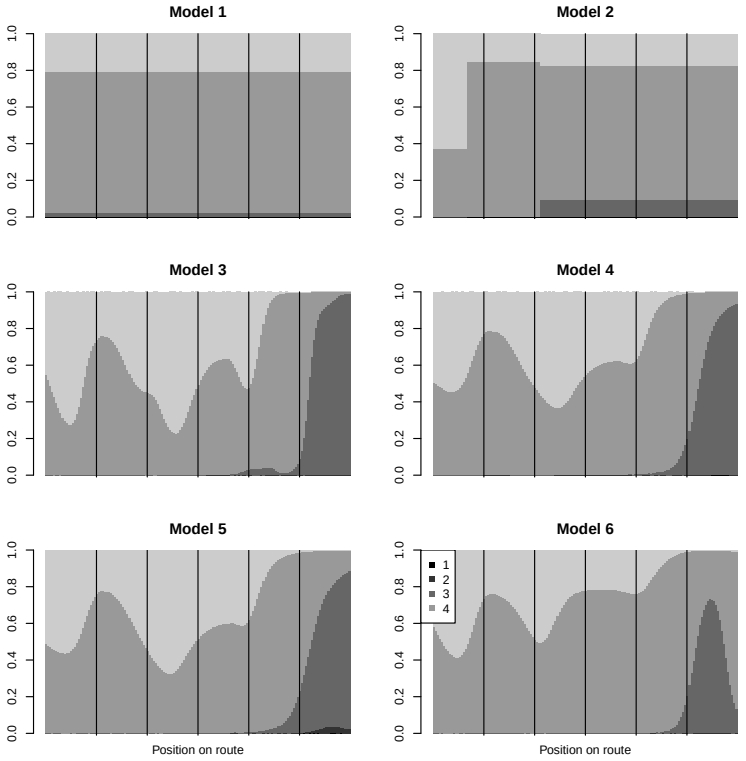
Supplemental Figure 11: As figure 2, but for 2010 Q1.



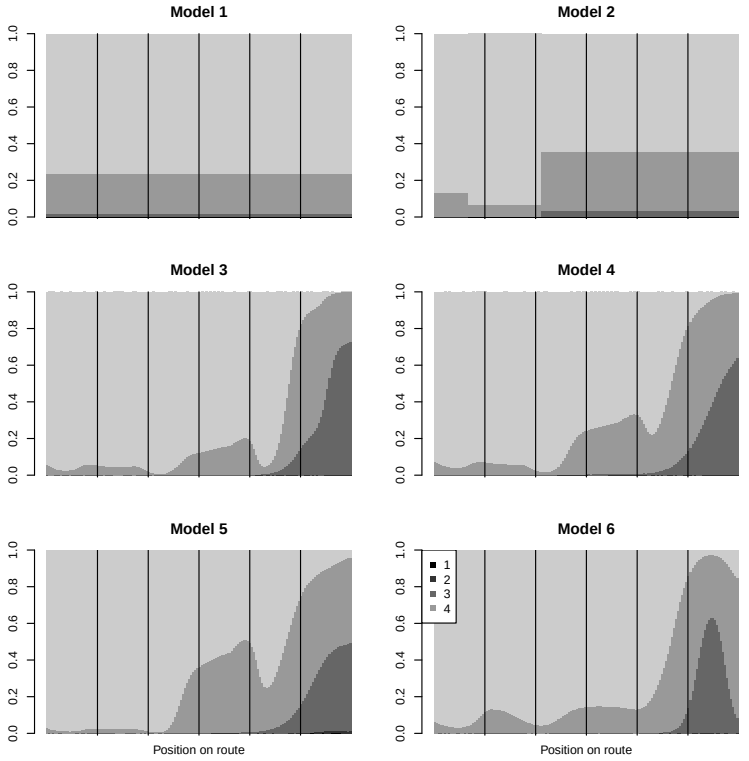
Supplemental Figure 12: As figure 2, but for 2011 Q1.



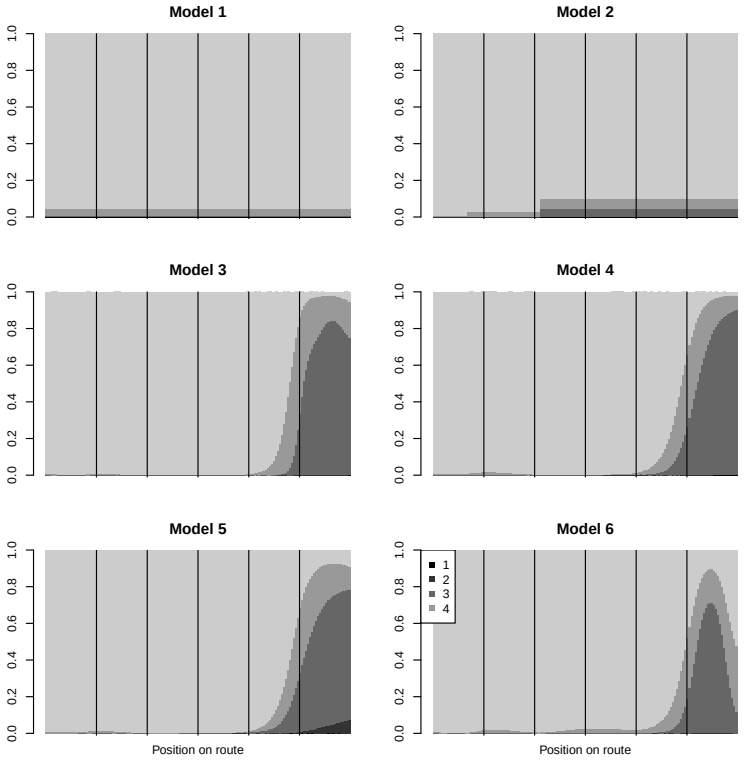
Supplemental Figure 13: As figure 2, but for 2001 Q3.



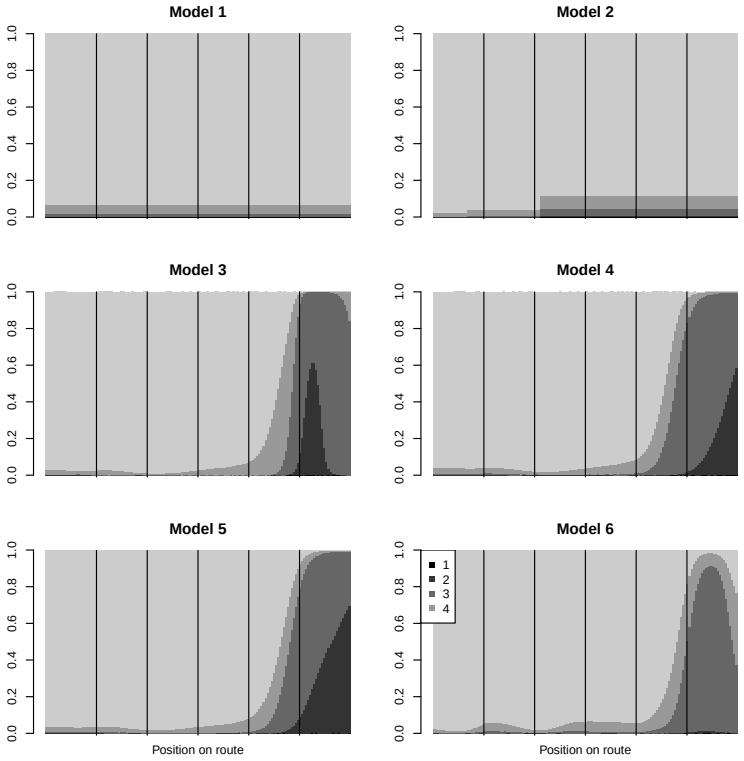
Supplemental Figure 14: As figure 2, but for 2002 Q3.



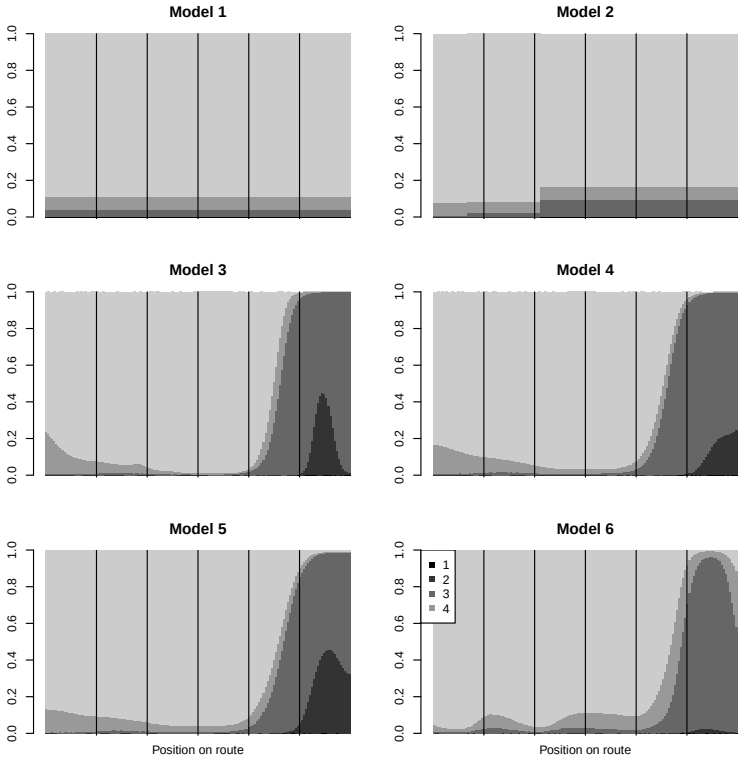
Supplemental Figure 15: As figure 2, but for 2003 Q3.



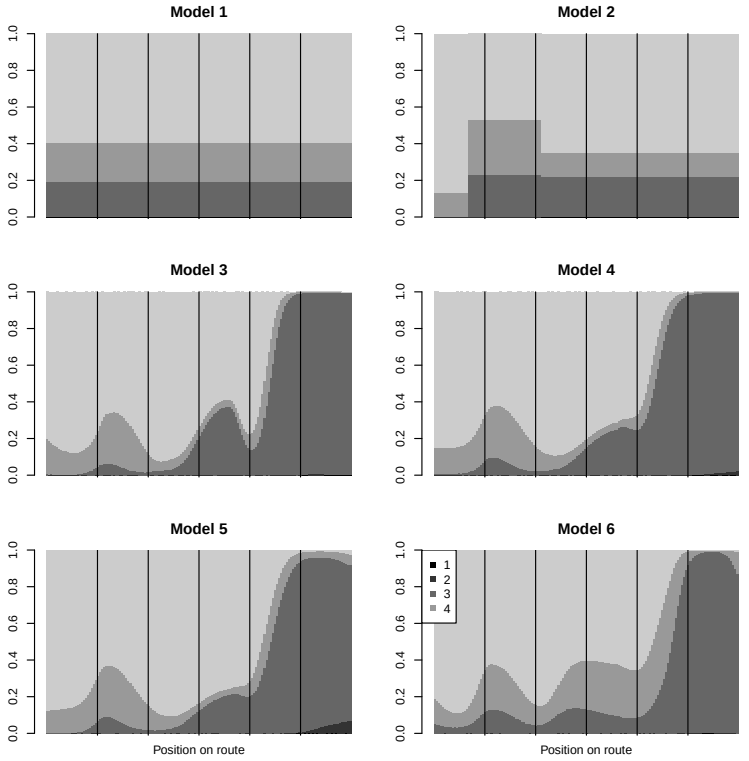
Supplemental Figure 16: As figure 2, but for 2004 Q3.



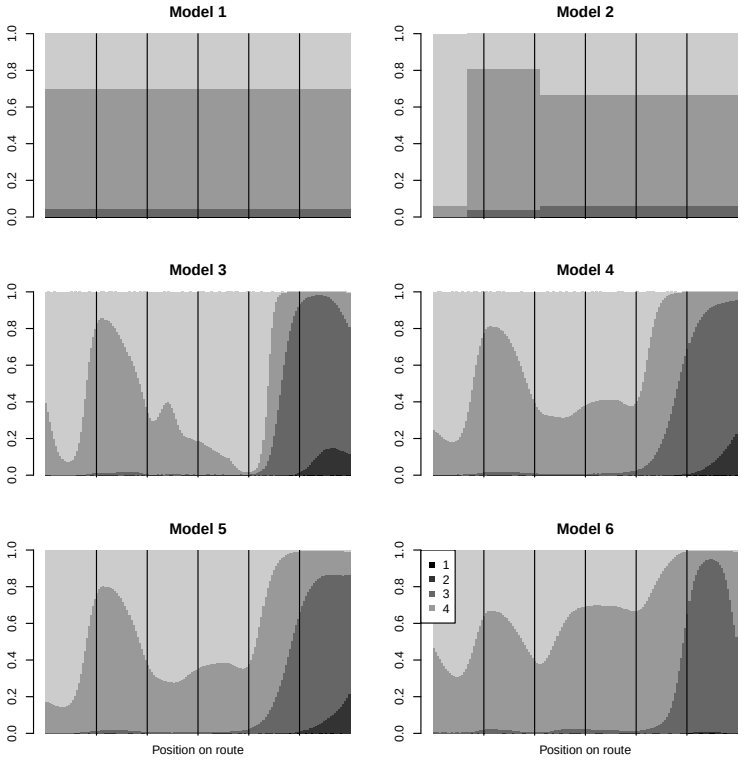
Supplemental Figure 17: As figure 2, but for 2005 Q3.



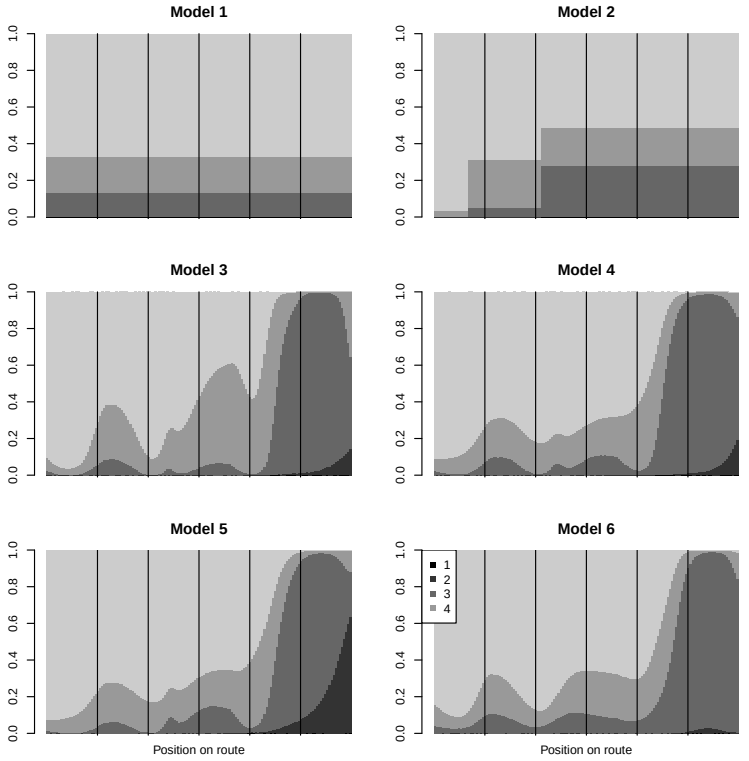
Supplemental Figure 18: As figure 2, but for 2006 Q3.



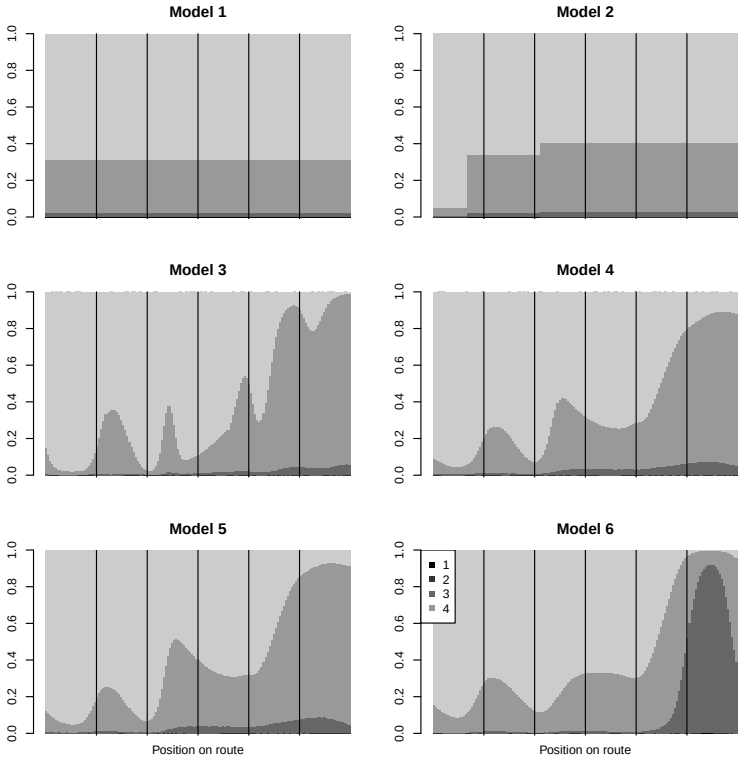
Supplemental Figure 19: As figure 2, but for 2007 Q3.



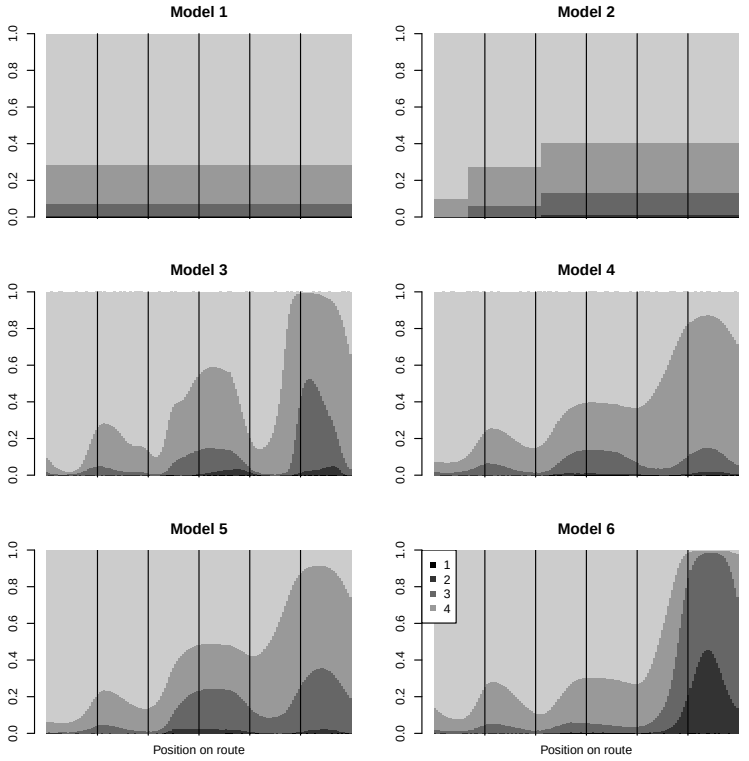
Supplemental Figure 20: As figure 2, but for 2008 Q3.



Supplemental Figure 21: As figure 2, but for 2009 Q3.



Supplemental Figure 22: As figure 2, but for 2010 Q3.



Supplemental Figure 23: As figure 2, but for 2011 Q3.

Appendix F

Paper IV

Evaluation of alternative age-based methods for estimating relative abundance from survey data in relation to assessment models

Casper W. Berg[a,b,*], Anders Nielsen[a], Kasper Kristensen[b]

[a]*National Institute of Aquatic Resources, Technical University of Denmark, Charlottenlund Castle, 2920 Charlottenlund, Denmark.*
[b]*Department of Informatics and Mathematical Modeling, Technical University of Denmark, Richard Petersens Plads, Building 305, 2800 Lyngby, Denmark*

Abstract

Indices of abundance from fishery-independent trawl surveys constitute an important source of information for many fish stock assessments. Indices are often calculated using area stratified sample means on age-disaggregated data, and finally treated in stock assessment models as independent observations.

We evaluate a series of alternative methods for calculating indices of abundance from trawl survey data (Delta-Lognormal, Delta-Gamma, and Tweedie using Generalized Additive Models) as well as different error structures for these indices when used as input in an age-based stock assessment model (time-constant vs time-varying variance, and independent versus correlated age groups within years).

The methodology is applied to data on North Sea herring (*Clupea harengus*), sprat (*Sprattus sprattus*), and whiting (*Merlangius merlangus*), and the Delta-Lognormal model is found best in terms of internal and external consistencies as well as AIC/BIC. Finally, stock assessments using the derived indices are presented, and the importance of selecting the most appropriate method for calculating indices as well as their error structure is highlighted.

Keywords: Age-based stock assessment, Survey Indices, Generalized Additive Models

1. Introduction

Many fish stock assessments are based on two key sources of input data: 1) The removals from the population due to commercial fishing and 2) Indices of abundance based either on catch and effort data from commercial or recreational fisheries, or from independent scientific surveys (Maunder and Punt, 2004). An index of abundance is a relative measure of e.g. the biomass or number of individuals in a population and most often proportionality is assumed:

$$I_y = qN_y$$

*corresponding author
Email address: cbe@aqua.dtu.dk (Casper W. Berg)

In this paper the focus will be on the analysis of fishery-independent survey data to create age-disaggregated indices of abundance, as well as on the subsequent use of these as input for a stock assessment model.

Several quite different approaches to the analysis of survey data exist depending on the design of the experiment, see Kimura and Somerton (2006) for a review. A popular method is based on stratified analysis, where the region of interest is divided into smaller strata and assuming that abundance is homogeneous within strata. The mean abundance is then calculated for each stratum and summed to give an index for the whole region. The probably simplest procedures uses the arithmetic mean within strata (e.g. ICES (2012c)). A slightly more refined alternative is the use of delta-distributions (e.g. Pennington (1983)), where zero values are modelled separately and the positive values are assumed to be log-normal (or Gamma) distributed. The mean in the delta-distribution is a more efficient estimator when the nonzero values are well approximated by a lognormal distribution, specifically it is less sensitive to the occasional huge catches that are often found in marine data sets (Pennington, 1996). Discrete valued distributions such as the negative binomial (Kristensen et al., 2006) and the Log-Gaussian Cox Process (LGCP) (Lewy and Kristensen, 2009) have also been applied, but age-disaggregated indices are typically not discrete valued, so these will not be considered in this study. More recently the Tweedie distribution (Tweedie, 1984) has been suggested as an alternative to delta-distributions (Candy, 2004; Shono, 2008).

When external factors other than changes in abundance affect the catch rate, these need to be corrected for in order to obtain an unbiased index (Maunder and Punt, 2004). To this end, more advanced methods such as generalized linear models (GLMs), generalized linear mixed models (GLMMs) and generalized additive models (GAMs) have previously been applied to correct for effects such as spatial position, depth, and time of day (Stefansson, 1996; Petrakis et al., 2001; Piet, 2002; Adlerstein and Ehrich, 2003; Beare et al., 2005). GAMs permit non-linear smooth relations between the response and the explanatory variables, so spatial stratification can conveniently be replaced by smooth functions of geographical coordinates (so-called splines). When stratification is used, there will be a trade-off between loss of spatial resolution due to assumption of homogeneity within strata and problems with few or missing values when a too fine-grained stratification is used. When using GAMs, this trade-off problem is replaced with an easier problem of smoothness selection for the splines, which can be solved more less automatically using modern software packages (Wood, 2006a).

Although useful on their own, one of the main uses of indices of abundance is to use them as input to an assessment model in combination with commercial catch data to obtain estimates of biomass and fishery mortality. The way that trawl survey data enters into many stock assessment models, can roughly be described as follows: 1) Numbers-at-length data from individual hauls are preprocessed and reduced to one matrix of numbers-at-age (the index of abundance). 2) Each number in this matrix is taken as an observation of the relative abundance-at-age in the stock assessment model (often assumed independent and with constant CV through time). Although often separated for convenience, we will demonstrate that more information can be extracted from the data by combining the analyses: Instead of reducing the information from a survey to one matrix, we will use three matrices (by adding standard deviations and correlations), and by actually carrying out the stock assessments with different distributional assumptions about the survey indices, we are provided with additional means to evaluate the impact of changes in the preprocessing step. This paper is therefore organized as follows: The first part of this paper deals with a comparison of the stratified mean method (SMM) with three variations of the Delta-GAM approach for calculating indices of abundance from trawl survey data, and second

part of this paper deals with performing stock assessments using different assumptions on the error structure for the survey indices derived in the first part. Using bootstrap methodology we will show, that there can be considerable positive correlations between abundance indices by age within the same year, and that including these correlations in a stock assessment model improves the model.

2. Materials and methods

2.1. Data

The data sets consist of 20 years (1992-2011) of biannual (Q1 and Q3) trawl survey data from the International Bottom Trawl Survey (IBTS) in the North Sea, downloaded from the DATRAS database (www.datras.ices.dk, data downloaded 2012-04-30). The commercial catch-at-age data, natural mortalities, proportion mature, and weight-at-age used in the stock assessments are taken from ICES (2012a) for sprat and herring and from ICES (2012b) for whiting, and the number of age-groups used in the analyses are also the same as in these two sources. Numbers-at-length from the trawl surveys are first converted to numbers-at-age using the method described Berg and Kristensen (2012) and implemented in Kristensen and Berg (2012), see online supplemental materials for details.

2.2. Stratified Mean Method

The North Sea is divided into so-called statistical rectangles each of size 1° longitude x 0.5° latitude, where (ideally) two hauls should be taken each quarter (each by different countries), each separated by at least 10 n.m. (ICES, 2010). The survey index is calculated using the stratified mean method (SMM) by taking the mean catch per rectangle, and then the mean over all rectangles in the North Sea. This method is similar to the current way that survey indices for use in assessment are calculated for stocks in the North Sea (ICES, 2012c), and is thus a relevant baseline to compare with.

2.3. Delta-GAM and Tweedie

The delta-models consist of two parts: one that describes the probability for a non-zero catch (binomial response), and another that describes the distribution of a catch given that it is non-zero (positive continuous). We will consider both the lognormal distribution and the Gamma distribution for the positive values. We assume the following relationship in both parts of the model between the expected response (μ, which is numbers-at-age or 1/0 for positive/non-positive catch depending on the model) and external factors:

$$g(\mu_i) = \text{Year(i)} + \text{U(i)}_{\text{ship}} + f_1(\text{lon}_i, \text{lat}_i) + f_2(\text{depth}_i) + f_3(\text{time}_i) \tag{1}$$

where Year(i) maps the ith haul to a categorical effect for each year, $\text{U(i)}_{\text{ship}} \sim N(0, \sigma_u)$ is a random effect for the vessel collecting haul i, f_1 is a 2-dimensional thin plate regression spline on the geographical coordinates, f_2 is a 1-dimensional thin plate spline for the effect of bottom depth, and f_3 is a cyclic cubic regression spline on the time of day (i.e. with same start end end point). The function g is the link function, which is taken to be the logit function for the binomial model, and the logarithm for the strictly positive responses in the Gamma and Tweedie models. The lognormal part of the delta-lognormal model is fitted by log-transforming the response and

using the Gaussian distribution with a unit link. Each combination of quarter and age group are estimated separately.

The length to age conversion may produce numbers that are very close to zero, which poses problems for the log-normal distribution and Gamma distribution when the mean is far from zero (Myers and Pepin, 1990; Kimura and Somerton, 2006). This can be remedied by simply treating values below some small chosen constant k as zero, and thereby move these from the positive component of the delta-distribution to the zero component (Folmer and Pennington, 2000). A preliminary analysis using histograms of residuals from the positive part of the delta models indicated that $k = 0.01$ was a reasonable choice (not the often ad-hoc chosen value of $k = 1$, which resulted in clearly non-Gaussian residuals in positive part of the delta-lognormal model).

The likelihood of the delta-distributions, can be found by fitting the model for the zero and positive observations independently, and utilizing that the full likelihood is given by

$$L = \prod_{i;y_i=0} \mathbb{P}(y_i = 0; \theta_Z) \prod_{i;y_i>0} \mathbb{P}(y_i > 0; \theta_Z) f(y_i; \theta_P) \tag{2}$$

However, since the delta-lognormal model is fitted using the log-transformed response, we cannot compare likelihoods with models for the untransformed response directly, but since the log-transformation is monotone, we can apply the formula for the change of variable $Z = \log(Y)$ to get

$$L = \prod_{i;y_i=0} \mathbb{P}(y_i = 0; \theta_Z) \prod_{i;y_i>0} \mathbb{P}(y_i > 0; \theta_Z) f(\log(y_i); \theta_P) \frac{1}{y_i} \tag{3}$$

which enables the comparison of the delta-lognormal models with the delta-gamma and Tweedie models, when the former has been fitted on the log-transformed response.

The Tweedie distribution has been proposed as a interesting alternative to delta-distributions (Tweedie, 1984; Candy, 2004; Shono, 2008) due to its nice interpretation as a compound Poisson-gamma distribution, and its ability to handle both zero and positive values simultaneously.

The Tweedie distribution has three parameters and is a member of the exponential family with variance $Var[y_i] = \phi\mu_i^p$. For $1 < p < 2$ this distribution has support on all non-negative real number, i.e. a continuous density on the positive reals with a point mass in zero. In this case it is also known as a compound Poisson distribution, because it is equivalent to the distribution of $Z = W_1 + ... + W_N$, where W_k are independent identically distributed Gamma variables, and N follows a Poisson distribution (Candy, 2004). The fitting of GAMs with a Tweedie distribution for $1 < p < 2$ can be accomplished with the `mgcv`-package in `R`, which uses the series evaluation by Dunn and Smyth (2005) for fixed values of p. We use the same strategy as Candy (2004) and Shono (2008) to fit p, which is by optimizing the profile likelihood for this parameter. The mean value specification for the Tweedie model is chosen to be identical to that of the Delta-distributions. The thin plate splines are estimated with shrinkage smoothing (Wood, 2006a, pg. 160), and smoothness selection is carried out with the marginal likelihood method (Wood, 2011).

2.4. Extracting the index of abundance from the models

The usual procedure for GLMs is to use the estimated year effects as the indices of abundance (Maunder and Punt, 2004). For the Delta-models this is not possible, and instead we must integrate the fitted abundance surface to obtain the index (Stefansson, 1996).

However, when covariates such as depth, that are measured as part of the sampling procedure, are included in the model, it is necessary to obtain the same covariates for each point in space that we would be integrating over. Although bathymetri maps could be used for depth, other variables might not be as easily obtainable. Another possible problem with a fine-grained integration of the abundance surface, is that we might be extrapolating to areas where the model is invalid. This could be extremely deep or shallow areas, where the estimated depth effect is inappropriate, or in-trawlable areas, where the abundance is unknown. To overcome these problems, we choose the following procedure to obtain the abundance estimates for each year: 1) Divide the survey area into small subareas of approximately equal size. 2) For each sub-area where at least one haul has been taken, choose one haul position to be representative of this sub-area, e.g. the one closest to the centroid of all hauls in the given sub-area. 3) Take the sum over all predicted abundances using the same reference vessel (or zero in case of a random vessel effects) in the chosen haul positions. This approach has the advantage, that all covariates are immediately available given that they were collected at the chosen haul positions, and it avoids the problems that could be associated with extrapolation in space. The resulting grid is shown in supplemental figure 1.

2.5. Estimating the statistical distribution of the indices

Often only the point estimates of the indices are used as input to stock assessment models, and additional knowledge about the distribution of the indices is thus discarded. However, various methods exist for obtaining approximations of the probability distribution of the indices from the Delta-GAM and Tweedie models, e.g. Wood (2006a,b), as well as for the SMM ICES (2012c). Bootstrapping is a widely used technique when analytic methods are infeasible, and has previously been applied to fisheries survey data to deal specifically with the variability due to sub-sampling of age and length (e.g. Cervino and Saborido-Rey (2006)). We therefore apply bootstrapping to estimate the distribution of the indices.

In the following we assume, that the length distribution is known without error, and that all variability is due to sampling variability from the hauls and the age-sampling procedure. This is reasonable since the number of length samples is much greater than the number of age samples. If we sample entire hauls and thereby also the age samples in the bootstrap procedure, we will automatically incorporate the uncertainty due to the age sampling and possible correlations due to clustering of similar age groups. We can also obtain estimates of the correlation between the different age groups in our estimated indices of abundance from the bootstrap samples. The bootstrapping is carried out as follows:

Let n_y denote the number of hauls in a given year, and let a haul consist of both its associated length distribution as well as the age samples taken from that particular haul.

1. Create a bootstrap sample of the hauls, i.e. for each year sample n_y hauls with replacement from the data from the given year.
2. For each year, use the bootstrapped age data to estimate an ALK, and convert from length to age for each haul.
3. Estimate indices of abundance by age from the bootstrapped data set.

We choose to use 400 bootstrap replicates per survey to estimate standard deviations and correlation matrices for every vector of log-indices by age for a given year, $\log I_y$.

Although it would be possible to use the estimated standard deviations on $\log I_y$ directly in the stock assessment model, these typically underestimates variability between I_y and the estimated values of qN_y from the stock assessment model (Maunder and Punt, 2004). This can be due to

violations of some of the assumptions in the catch standardization model, the stock assessment model, or both, e.g. that that catchability q is not constant. We will discuss further the choice error structure in section 2.6.

2.6. Stock assessment model

While internal and external consistencies provide some means to evaluate survey indices, they rely on assumptions of constant total mortality over time (Berg and Kristensen, 2012). Hence, a more appropriate way to evaluate the survey indices when mortality varies over time is to carry out full stock assessments using a statistical model, such that we utilize the extra information we have from the commercial catches and thereby estimate a time-varying mortality. State-space models allows separation of process and observation errors, which leads to an objective way of weighting each data source in the estimation process. Briefly, a state-space model consists of a set of states describing the complete system at time t, a set of transition (or system) equations describing how the states at time t relate to those at time $t+1$, and a set of observation equations relating states with observations. All of these relations must be defined in terms of probability distributions whose parameters should be estimated from the data. For a more thorough discussion of the state-space approach to stock assessment models, see e.g. Gudmundsson and Gunnlaugsson (2012). For this application, the states consist of log-transformed numbers-at-age $\log N_1, \ldots, \log N_A$ and fishing mortalities $\log F_{i_1}, \ldots, \log F_{i_n}$ corresponding to different age classes and total commercial catches.

Using yearly time-steps (and replacing t with y) the transition equations for $\log N_i$ are:

$$\log N_{1,y} = \log N_{1,y-1} + \varepsilon_{R,y} \tag{4}$$

$$\log N_{a,y} = \log N_{a-1,y-1} - F_{a-1,y-1} - M_{a-1} + \varepsilon_{S,a,y} \qquad 2 \le a \le A \tag{5}$$

$$\log N_{A,y} = \log\left(\mathrm{e}^{\log N_{A-1,y-1} - F_{A-1,y-1} - M_{A-1}} + \mathrm{e}^{\log N_{A,y-1} - F_{A,y-1} - M_A}\right) + \varepsilon_{S,A,y} \tag{6}$$

where M_a is natural mortality at age a, which is assumed to be known a priori, $F_{a-1,y-1}$ is the total fishing mortality. A simple random walk model was chosen for the recruitment process (eq. 4), because it is our experience that there most often is no gain from switching to functions of SSB such as Ricker or Beverton-Holt (Gudmundsson and Gunnlaugsson (2012) made the same observation and used a similar simplification). The process errors on $\log N$ are assumed to be zero-mean independent normal distributed with two separate variance parameters, one for recruitment σ_R^2, and one for survival σ_S^2. For the fishery mortalities, F, a correlated random walk model is assumed (note, that the age index is now omitted indicating that we are dealing with vectors rather than scalars as above):

$$\log F_y = \log F_{y-1} + \varepsilon_{F,y} \tag{7}$$

such that $\varepsilon_{F,y} \sim N(\mathbf{0}, \mathbf{\Sigma}_F)$ and $\mathbf{\Sigma}_{F,ij} = \sigma_F^2 \rho$ for $i \neq j$ and $\mathbf{\Sigma}_{F,ii} = \sigma_F^2$, where σ_F and ρ are parameters to be estimated. When $\rho = 1.0$ we have the special case of a multiplicative structure in $\log F$ whereas $\rho = 0$ allows for completely independent development by age group in fishery mortality over time.

Assuming independent observations, the observation equations become:

$$\log C_{a,y} = \log\left(\frac{F_{a,y}}{Z_{a,y}}(1 - e^{-Z_{a,y}})N_{a,y}\right) + \varepsilon^{C}_{a,y} \tag{8}$$

$$\log I^{(s)}_{a,y} = \log\left(Q^{(s)}_{a} e^{-Z_{a,y}\frac{D^{(s)}}{365}} N_{a,y}\right) + \varepsilon^{(s)}_{a,y} \tag{9}$$

where $Z_{a,y} = M_a + F_{a,y}$ is the total mortality rate, $D^{(s)}$ is the number of days into the year where the survey s is conducted, and $Q^{(s)}_a$ are catchability parameters, $\varepsilon^{C}_{a,y} \sim N(0, \sigma^2_C)$, and $\varepsilon^{(s)}_{a,y} \sim N(0, \sigma^2_s)$.

Now, to accommodate correlated observations of $\log I_y$ we change (9) to

$$\log I^{(s)}_{y} = \log\left(Q^{(s)} \circ e^{-Z_y \frac{D^{(s)}}{365}} N_y\right) + \varepsilon^{(s)}_{y} \tag{10}$$

where $\varepsilon^{(s)}_{y} \sim N\left(\mathbf{0}, \mathbf{\Sigma}_y\right)$ (note again, that the scalars in eqn. 9 are replaced with vectors containing all age groups at once, and "$\circ$" denotes element-wise multiplication).

The covariance matrices $\mathbf{\Sigma}_y$ are the empirical covariance matrices from the bootstrapped log-indices described in section 2.5. The effect of using a multivariate normal distribution rather than the independent normal distribution for each age group can be examined by inspection of the scaled residuals (e.g., Myers and Cadigan, 1995) of the survey indices:

$$\mathbf{\Sigma}_y^{-\frac{1}{2}}\left(I^{(s)}_y - \hat{I}^{(s)}_y\right)$$

Rather than working directly with the estimated covariance matrices, we choose instead to parameterize them in terms of correlation matrices and vectors of standard deviations $\mathbf{\Sigma} = \text{diag}(\sigma)\mathbf{R}\text{diag}(\sigma)$. In order to account for additional uncertainty on the survey indices other than that which is accounted for in the bootstrapping procedure, we examine the following error structures:

$$\sigma_t = \sigma, \qquad \mathbf{\Sigma}_t = \sigma_t \mathbf{I} \tag{11}$$

$$\sigma_t = w_t \sigma, \qquad \mathbf{\Sigma}_t = \sigma_t \mathbf{I} \tag{12}$$

$$\sigma_t = \sqrt{\sigma^2 + w_t^2}, \qquad \mathbf{\Sigma}_t = \sigma_t \mathbf{I} \tag{13}$$

$$\sigma_t = \sigma, \qquad \mathbf{\Sigma}_t = \text{diag}\left(\sigma_t\right)\mathbf{R}_t\text{diag}\left(\sigma_t\right) \tag{14}$$

$$\sigma_t = w_t \sigma, \qquad \mathbf{\Sigma}_t = \text{diag}\left(\sigma_t\right)\mathbf{R}_t\text{diag}\left(\sigma_t\right) \tag{15}$$

$$\sigma_t = \sqrt{\sigma^2 + w_t^2}, \qquad \mathbf{\Sigma}_t = \text{diag}\left(\sigma_t\right)\mathbf{R}_t\text{diag}\left(\sigma_t\right) \tag{16}$$

where the observation variance at time t, σ_t is either constant through time (σ), time-varying and proportional to the estimated standard deviation from the bootstrapping procedure, w_t, or time-varying with a total variance given by the sum of w_t and an additional constant variance component. In all cases σ is a parameter to be estimated, while w_t and the correlation matrices $\mathbf{R}_t$ are assumed to be known without error from the bootstrap procedure, and $\mathbf{I}$ is the identity matrix.

2.7. Model evaluation

2.7.1. AIC/BIC

To evaluate which distribution (log-normal, Gamma, or Tweedie) that provide the best fit to the individual haul data we compare the AIC and BIC values for each distribution. Since we are using GAMs, we replace the number of observations with the effective degrees of freedom (edf, see Wood (2006a)). Since age groups are estimated independently, we can simply add the log-likelihoods and edfs for each age group to obtain one the AIC/BIC for all age groups combined.

2.7.2. Internal and external consistency

We define the internal consistency (IC) as the correlation between $logI_{y,q,a}$ and $logI_{y+1,q,a+1}$ and external consistency as the correlation between $logI_{y,q_1,a}$ and $logI_{y,q_3,a}$, where y,q, and a denotes year, quarter and age respectively. Positive consistencies implies that we can "follow the cohorts" within (IC) and between (EC) surveys, (see e.g. Berg and Kristensen (2012) for details).

2.7.3. Areas of confidence ellipses

Since we cannot use any standard tests for comparing different data sets, we will compare the precision with which we can estimate the spawning stock biomass (SSB) and average fishing mortality using the SMM and GAM approaches respectively. It is intuitively clear, that consistent data sources with a low amount of noise will lead to smaller confidence ellipses and hence higher precision than inconsistent data sources with a larger amount of noise given time-series of equal length. Confidence ellipses for pairs of parameters are constructed from the corresponding marginals of the estimated covariance matrix for the parameters, and from pairs of derived quantities using the delta method (Oehlert, 1992). The area of a confidence ellipsis is proportional to $\sqrt{e_1 e_2}$, where e_1 and e_2 are the eigenvalues of the corresponding covariance matrix.

2.7.4. Likelihood comparison

For selection between error structures, the above criteria is not appropriate, since change to a more appropriate model for the observations will not necessarily lead to smaller confidence ellipses. Instead, likelihood based criteria such as AIC or BIC are appropriate, but since all the error structures under consideration has the same number of parameters in the stock assessment model, the choice of penalty due to the number of parameters does not matter and direct comparison of likelihoods are possible.

2.8. Software

The trawl data was handled in `R` (R Development Core Team, 2012) using the DATRAS package (Kristensen and Berg, 2012), and the GAM models were fitted using the `mgcv` package (Wood, 2006a). The source code for the GAM models can be obtained by contacting the corresponding author. The stock assessment models were fitted using AD model builder (Fournier et al., 2012), and the entire source code as well as data for the models are available online at www.stockassessment.org (look for the assessments named "NS-Sprat-MV","NS-Herring-MV", and "NS-Whiting-MV").

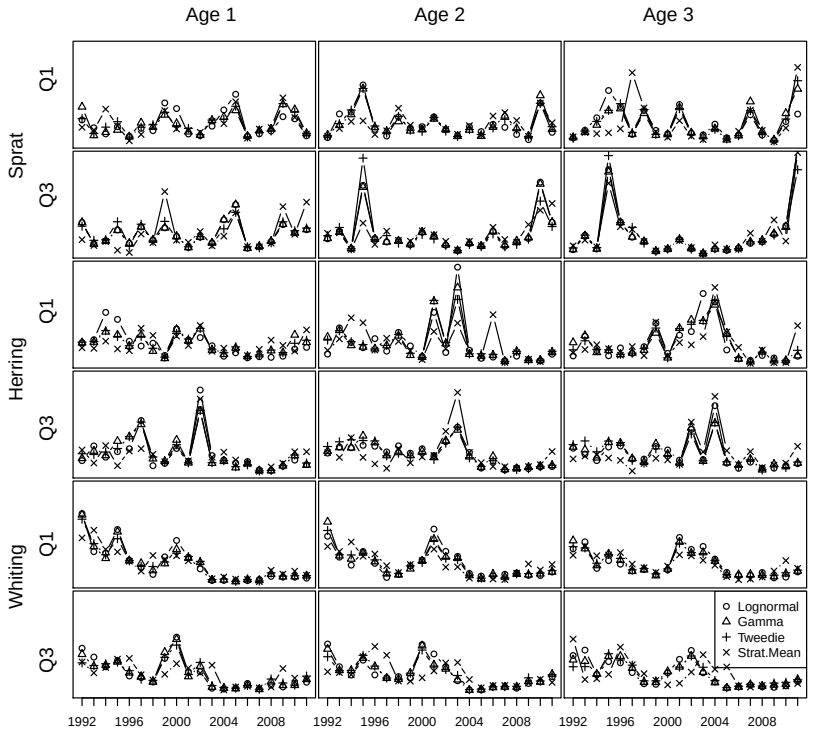


Figure 1: Estimated relative survey indices by species, quarter, age and method. Scaling on y-axis is irrelevant since indices are relative.

3. Results

3.1. Calculation of indices

3.1.1. Sprat

The delta-lognormal model (DLN) gives the best overall fit to the haul data when compared to the delta-gamma and the Tweedie models using both AIC as well as BIC as the criterion (Supplemental table 1). The estimated depth effects are rather weak, but there seems to be a consistent tendency for the catch rate to decline on depths below 50 m. With the exception of age 1 in Q1, catch rates during daylight are significantly larger than in the night time. Sprat are mainly caught in the southern parts of the North Sea both in Q1 and Q3. Age group 1 seems to be confined to a smaller area than the older age groups both in Q1 and Q3. In Q1 high catch rates

Lognormal	Gamma	Tweedie	Lognormal 2	Strat.Mean	Quarter	Species
0.27	0.18	0.21	0.13	0.08	IC(Q1)	Sprat
0.60	0.50	0.32	**0.61**	0.20	IC(Q3)	Sprat
0.52	**0.62**	0.59	0.47	0.46	EC	Sprat
0.61	0.56	0.48	0.60	0.06	IC(Q1)	Herring
0.68	0.63	**0.71**	0.62	0.66	IC(Q3)	Herring
0.59	0.54	0.53	0.50	0.46	EC	Herring
0.73	0.76	0.82	0.80	**0.88**	IC(Q1)	Whiting
0.87	0.86	0.85	0.83	0.85	IC(Q3)	Whiting
0.83	0.82	0.82	**0.84**	0.78	EC	Whiting

Table 1: Average consistencies (average over ages 1-5). The columns "Lognormal", "Gamma", and "Tweedie" correspond to indices calculated using equation 1 with the respective distributions, whereas "Lognormal 2" is a delta-lognormal model with year effects only in equation 1. Best consistencies are shown in bold face.

of older age groups are found near the English channel, which constitutes the southern boundary of the assessment area (Supplemental figure 2). This could cause problems with the Q1 index, if substantial parts of the population may be found south of the assessment area in Q1.

The delta-models are clearly much better than the stratified mean model in terms of overall consistencies, and the DLN the best overall (table 1). The stratified mean method seems to have the most problems with the older age-groups in terms of bad consistency, and the same pattern can be found in the DLN model with year effects only. This could indicate, that the spatial distribution of age 1 is less consistent than that of the older age groups, so the time-constant spatial effect might not be completely appropriate for age group 1. The Q3 indices have much higher consistencies than the Q1 indices, although the estimated standard deviations on the log indices are lower on average in Q1 than Q3 (figures 14 and 16, supp. mat). This could for instance be linked to the problems with high catch rates near the southern boundary in Q1.

3.1.2. Herring

As for sprat, the DLN model gives a better fit to the haul data than the delta-gamma and the Tweedie model (Supp. table 5). The estimated depth effects differ clearly between both age-groups and quarters, but with a overall trend of older age groups being mainly caught in deeper waters than age group 1. Catch rates for herring are higher in the day time than during night, with the biggest contrast found in Q3. The spatial distribution of herring catches varies greatly between age groups and quarters. In Q1, there are very high catch rates of older age groups on the boundaries of the stock assessment area, both in the English channel and near the Norwegian deep. There are no such apparent problems with the catches in Q3.

The stratified mean model performs very poorly compared to the GAMs in Q1 with respect to internal consistency, but in Q3 the differences are smaller. Overall, the DLN model provides the best consistencies.

In the final year (2011) there are some large discrepancies between the SMM and the GAMs with the former indicating a much higher abundance, which may influence the assessment substantially. In 2003 Q1 the situation is reversed, the GAMs suggest a much higher abundance than the SMM. Some light is shed on this behavior when the distributions of catch in numbers (regardless of age) are examined closer: In 2003 we have the observed statistics [Proportion of zeroes; Median; Mean; Max] = [9.0%; 153; 1499; 64440] whereas the same statistics in 2011

are [12.0%; 36.5; 2152; 174400], i.e. there is a large maximum catch in 2011 which results in a larger mean value than in 2003, although e.g. the median observation in 2011 is much lower than in 2003.

3.1.3. Whiting

Again, the DLN model fits the haul data for whiting better than the two other delta distributions. In both quarters there are significant depth effects for all age groups. Like in the herring case younger age groups tends to be caught in shallower waters than the older age groups. The maps of their spatial abundance reveal that the older ages are farther from the English and Scottish coast, but there are no alarmingly high catch rates near near the boundaries of the assessment area in either quarter.

We choose to look at some catch statistics in two years (2000 and 2009) with substantial discrepancies between the SMM and GAMs. In the year 2000 we have [Proportion of zeroes; Median; Mean; Max] = [2.5%; 400; 1547; 20390] compared to [10.6%; 47.5; 679.2; 27510] in 2009. In both years the distribution are highly skewed to the right, but the large maximum in 2009 drives the mean value to be not that far from the mean over the whole series, although the median value and proportion of zeroes suggests a low abundance year compared to the whole series, where the corresponding statistics are [4.1%; 124.0; 884.9; 170100].

3.2. Stock assessments

The stratified mean method differs quite drastically from the DLN methods in the last years (e.g. Sprat in Q3, Herring in Q1, cf. figure 1). Since the stratified mean estimates are much larger than the corresponding DLN estimates, the former yields higher estimates of SSB and lower estimates of F compared to the latter (change in [SSB,$\bar{F}$] in final year: [+37%,-11%] for sprat, [+49%,-73%] for herring, and [+29%,-43%] for whiting when using SMM, see figure 2). The confidence ellipses of the the SSB and $\bar{F}$ estimates in the final year (figure 3) also illustrate the substantial differences between the SMM and the DLN model. For all three stocks the areas of these confidence ellipses are smaller when using input from DLN model compared to the SMM under the assumption of independent observations in each age group for both methods. This implies, that there is more consistency among the data sources based on the DLN approach, since highly inconsistent and noisy data sources will result in more uncertain estimates about the state of the stock. Including correlations between age groups within years gives slightly larger confidence ellipses for all stocks (see figure 3). This is not too surprising, since having correlated data reduces the effective number of observations and hence should give larger uncertainties. It is worth also to notice, that the confidence ellipses from the model with correlations included are still smaller than the ones from the SMM.

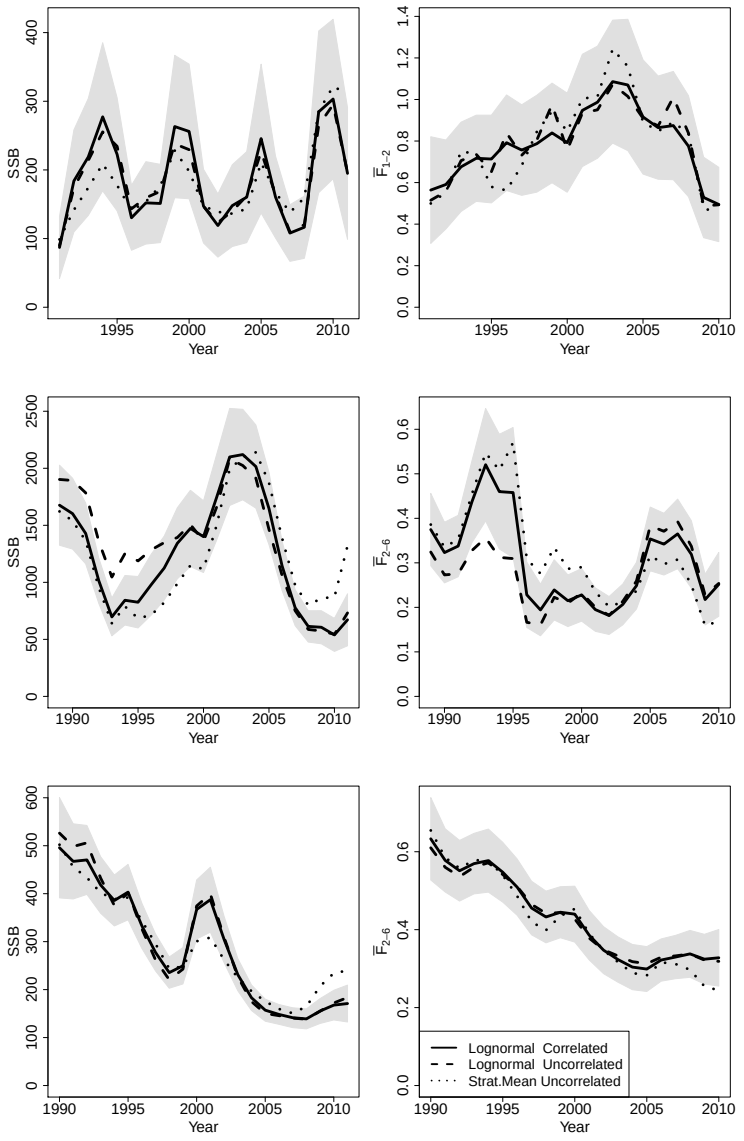


Figure 2: Estimated spawning stock biomasses (left column) and average fishing mortalities (right column) for sprat (top row), herring (middle row), and whiting (bottom row). The shaded areas represent 95% marginal confidence intervals calculated using the lognormal indices with correlations.

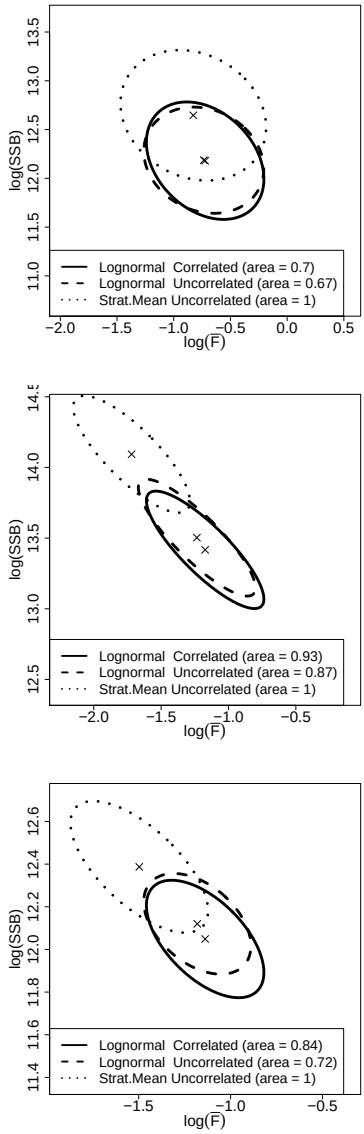


Figure 3: 95% contour ellipses for the joint distribution of $\log \bar{F}$ and log(SSB) in the last data year (2011) for sprat (top row), herring (middle row), and whiting (bottom row).

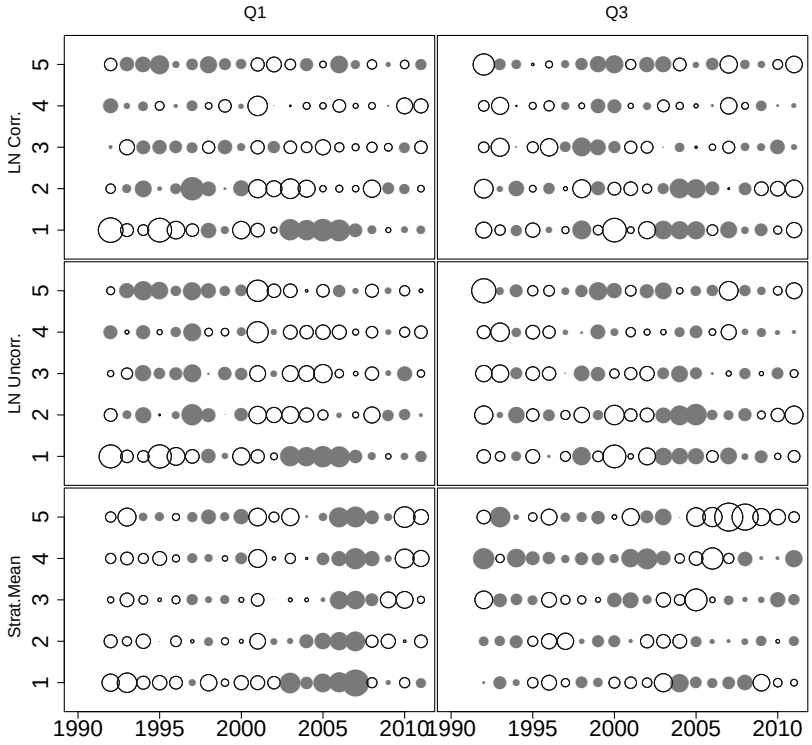


Figure 4: Whiting: survey standardized residual plots from the fitted assessment models.

The joint negative log-likelihoods for the different error structures (eqns 11–16) are shown in table 2. Since smaller values are to be preferred, the errors on the logarithm of the survey abundance indices are more likely to be distributed with the assumed multivariate normal structure than to be independent among age groups within years. The change in likelihood by changing from constant variance (eqn. 11) to time-varying (eqns. 12 or 13) is minor compared to the differences between uncorrelated and correlated errors, though there is a slight improvement for sprat and herring, but for whiting the equal variance assumption has the slightly better likelihood.

The standardized residuals for the survey data obtained from the stock assessment for whiting using the SMM, DLN model and eq. 11, and DLN model and eq. 16 are shown in figure 4. The SMM residuals are clearly problematic, there seems to be problems with correlations across ages as well as time, and maybe even cohorts for the Q3 index. The DLN residuals are somewhat better, although the Q1 index seems to have an overweight of negative values in the beginning

Model	(11)	(12)	(13)	(14)	(15)	(16)
Correlated	No	No	No	Yes	Yes	Yes
σ_t	σ	$w_t\sigma$	$\sqrt{\sigma^2 + w_t^2}$	σ	$w_t\sigma$	$\sqrt{\sigma^2 + w_t^2}$
sprat	267.79	266.05	266.25	255.18	**253.63**	253.73
herring	257.66	255.44	255.07	218.44	217.63	**216.34**
whiting	163.17	165.44	163.24	131.87	134.04	**131.76**

Table 2: Negative log likelihood from the stock assessment models using different survey data observation error structures.

of the time-series for age 2+, and consequently an overweight of positive residuals in the end, but vice versa for age group 1. The differences between the correlated and uncorrelated DLN residuals are more subtle than between the SMM and DLN, but there are some small improvements. The corresponding residuals from the sprat and herring assessments can be found in the supplemental materials. For these two species there are generally less differences between the patterns in the residuals between the three models than for whiting. The Q3 herring index has a series of positive residuals in the beginning of the time-series followed by negative ones for age 2+ regardless of the model for the indices. The SMM indices seem to suffer from variance inhomogeneity with larger residuals occurring in the last five years in Q1 and the first six years in Q3. The DLN indices look better with respect to homogeneity of variance. The residuals from the sprat assessments look very similar between methods, and there are no immediately apparent problems. The similarity between using uncorrelated versus correlated observations is expected here, since the correlations found for sprat were minor compared to those for herring and whiting (Supp. mat.).

The distribution maps and internal/external consistencies suggested that the Q3 indices described the sprat and herring stocks better than the Q1 indices. This can be investigated further by comparing the estimated precisions (inverse variances) from the stock assessment model. For sprat, the ratio of observation variances (using eq. 11) is estimated to be $\frac{\sigma^2_{Q1}}{\sigma^2_{Q3}} = 1.15$, for herring the same number is 1.4, and for whiting 2.4, and thus we must conclude that the Q3 survey indeed provides a more precise index of abundance for these three species than the one in Q1.

4. Discussion

We have considered statistical aspects in the joint process of calculating indices of abundance by age and using them as input for a stock assessment model. The first problem was how to derive the indices from individual hauls from trawl survey data. To this end we compared the method of stratified sample means, which is currently used for the stocks considered, with a Delta-Lognormal, Delta-Gamma, and Tweedie model, and the Delta-Lognormal was found to provide the best fit. This is in contrast with the results found for silky shark and yellowfin tuna by Shono (2008), where a Tweedie model gave the best results. More importantly, the SMM method examined in this study is similar to the one used by most survey indices used as basis for management advice in ICES, including the three species considered.

The advantage of using the proposed DLN model over the SMM was further confirmed by carrying out stock assessments using the indices – higher precision on estimates of SSB and $\bar{F}$ was obtained with the DLN indicating that it provided the more accurate index.

Although indices from the SMM are simple to compute and do not need to be updated back in time once new data are available, it has certain shortcomings compared to a proper statistical model such as the DLN model. One problem, which has long been known, is the sensitivity to extraordinary large catches (Pennington, 1996). When such huge catches occur at the end of the time-series, as they did in our examples, it will lead to overly optimistic biomass estimates. Another problem is the inability of the SMM to control for effects such as depth, gear, environmental factors etc. on the catch rate, although these effects were, although statistically significant, minor compared to the effect of large catches for the cases in this study. These problems are all well described in the literature (Pennington, 1996; Stefansson, 1996; Maunder and Punt, 2004), although the majority have applied area stratification and/or GLM methodology in place of GAMs to deal with spatial heterogeneity and uneven sampling effort. Recent studies (Berg and Kristensen, 2012; Maxwell et al., 2012) have advocated the use of GAMs and 2D-splines to replace area stratification, which, in addition to providing more accurate estimates using fewer parameters, alleviates the modeller of problems with selecting the strata and possibly the subsequent problems of having few or missing data points for some combinations of years and strata. The suitability of GAMs to interpolate missing values is of particular relevance to the IBTS survey considered in this study, since in the later years the survey area was expanded to the south. There are (at least) three possible options if one wishes to include these additional hauls in the index calculations: 1) Define one big southern stratum that includes at least one haul in all years and assume homogeneity 2) extrapolate abundances in the years with missing data from neighboring sampled regions, or 3) if there is a consistent spatial pattern between years, then estimate a time-invariant spatial effect as in the present study and use this to predict the missing values. The choice between these options should be based on the validity of the associated assumptions. For instance, if there is a high degree of spatial correlation but with varying positions of high abundance areas one of the two first options might be more appropriate, whereas given a highly consistent spatial pattern option three should be preferred. If the expansion of the surveyed area is substantial, extrapolation should be done with caution.

The second problem addressed in this paper was whether it was reasonable to assume uncorrelated observations within years and/or time-constant variances for the indices. Utilizing the estimated standard deviations on the log-indices from the DLN model in the assessment model did not change the results much regardless of the variance parameterization. One obvious reason could be, that the estimated SDs were rather constant through time. Also, there is a fundamentally different interpretation of the SDs from the DLN models and those from the assessment model (Maunder and Starr, 2003), i.e. the main source of the assessment SD is the discrepancy between the measured survey abundance and the true population, and not the measuring uncertainty from the experiment itself. In contrast to the estimated SDs, including the estimated correlations improved the likelihood of the data substantially, and did prove to have some impact on the assessment results. Although our motivation is the same as in Walters and Punt (1994) and Myers and Cadigan (1995), our method differ in that we are using bootstrap estimates from DLN model to estimate the correlations as opposed to estimating the correlation structure within the assessment model. When estimated within the assessment model using indices only, the information from the individual hauls is not included, and therefore the only option is to assume a simple correlation structure with few extra parameters due to the limited number of data points. An important reason why including correlations among age groups can improve assessments is given in Myers and Cadigan (1995): Say we observe higher abundances than expected in the final year of the older age groups and these are known to be positively correlated with the youngest age group, then we can utilize this information to heighten the estimates of the youngest age

group even though this cohort has only been observed once. However, the estimated correlations in this study were found to be greatest between the oldest age groups whereas the estimates of the youngest age group were often nearly uncorrelated with those of the older age groups. Assuming a common correlation among age groups as in Walters and Punt (1994) and Myers and Cadigan (1995) in this situation would be problematic – it could improve the estimates of older age groups at expense of the youngest, which is undesirable for forecasting purposes.

The second problem addressed in this paper was whether it was reasonable to assume uncorrelated observations within years and/or time-constant variances for the indices. Utilizing the estimated standard deviations on the log-indices from the DLN model in the assessment model did not change the results much regardless of the variance parameterization. One obvious reason could be, that the estimated SDs were rather constant through time. Also, there is a fundamentally different interpretation of the SDs from the DLN models and those from the assessment model (Maunder and Starr, 2003), i.e. the main source of the assessment SD is the discrepancy between the measured survey abundance and the true population, and not the measuring uncertainty from the experiment itself. In contrast to the estimated SDs, including the estimated correlations improved the likelihood of the data substantially, and did prove to have some impact on the assessment results. Although our motivation is the same as in Walters and Punt (1994) and Myers and Cadigan (1995), our method differ in that we are using bootstrap estimates from DLN model to estimate the correlations as opposed to estimating the correlation structure within the assessment model. When estimated within the assessment model using indices only, the information from the individual hauls is not included, and therefore the only option is to assume a simple correlation structure with few extra parameters due to the limited number of data points. An important reason why including correlations among age groups can improve assessments is given in Myers and Cadigan (1995): Say we observe higher abundances than expected in the final year of the older age groups and these are known to be positively correlated with the youngest age group, then we can utilize this information to highten the estimates of the youngest age group even though this cohort has only been observed once. However, the estimated correlations in this study were found to be greatest between the oldest age groups whereas the estimates of the youngest age group were often nearly uncorrelated with those of the older age groups. Assuming a common correlation among age groups as in Walters and Punt (1994) and Myers and Cadigan (1995) in this situation would be problematic – it could improve the estimates of older age groups at expense of the youngest, which is undesirable for forecasting purposes.

We should note, that several alternative methods exist for dealing with the problems discussed in this study. In so-called integrated assessment models (Maunder and Punt, 2004) the raw data is handled within the assessment model, i.e. age-length relationships and index calculations are computed as part of the assessment instead of as a preprocessing step like in the present study. The advantage of this is that the uncertainty from the preprocessing step can be easily propagated onto the final assessment output (at least when this is formulated as a maximum likelihood problem). However, assessment models are often, as in this study, quite complex and computer intensive, so reducing the dimensionality of the data in advance by preprocessing can be necessary. This study shows, that in order to avoid too much loss of information, it is important to consider not only point estimates from the preprocessing step, but to include standard deviations and correlations as well, and to continuously re-evaluate such preprocessing steps. Also, we must recommend that the SMM is no longer used to create indices of abundance for the highly aggregated stocks considered in this study.

5. Acknowledgements

The authors wish to thank all staff involved in the IBTS data collection, ICES for providing easy access to the data, and finally Anna Rindorf and Peter Lewy for valuable input to this manuscript.

References

Adlerstein, S., Ehrich, S., 2003. Patterns in diel variation of cod catches in north sea bottom trawl surveys. Fisheries Research 63, 169 – 178.

Beare, D., Needle, C., Burns, F., Reid, D., 2005. Using survey data independently from commercial data in stock assessment: an example using haddock in ICES Division VIa. ICES Journal of Marine Science: Journal du Conseil 62, 996–1005. http://icesjms.oxfordjournals.org/content/62/5/996.full.pdf+html.

Berg, C.W., Kristensen, K., 2012. Spatial age-length key modelling using continuation ratio logits. Fisheries Research 129-130, 119–126.

Candy, S., 2004. Modelling catch and effort data using generalised linear models, the tweedie distribution, random vessel effects and random stratum-by-year effects. CCAMLR SCIENCE 11, 59–80.

Cervino, S., Saborido-Rey, F., 2006. Using the bootstrap to investigate the effects of varying tow lengths and catch sampling schemes in fish survey. Fisheries Research 79, 294 – 302.

Dunn, P.K., Smyth, G.K., 2005. Series evaluation of Tweedie exponential dispersion model densities. Statistics and Computing 15, 267–280.

Folmer, O., Pennington, M., 2000. A statistical evaluation of the design and precision of the shrimp trawl survey off West Greenland. Fisheries Research 49, 165–178.

Fournier, D.A., Skaug, H.J., Ancheta, J., Sibert, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., 2012. AD Model Builder: Using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods and Software 27, 233–249.

Gudmundsson, Gunnlaugsson, 2012. Selection and estimation of sequential catch-at-age models. Canadian Journal of Fisheries and Aquatic Sciences 69, 1760.

ICES, 2010. Manual for the International Bottom Trawl Surveys Revision VIII.

ICES, 2012a. Report of the Herring Assessment Working Group for the Area South of 62 deg N (HAWG). ICES Document CM 2012/ACOM:06 .

ICES, 2012b. Report of the Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK). ICES Document CM 2012/ACOM:13 .

ICES, 2012c. Report of the Workshop on Implementation in DATRAS of Confidence Limits Estimation of, 1012 May 2006, ICES Headquarters, Copenhagen.

Kimura, D.K., Somerton, D.A., 2006. Review of statistical aspects of survey sampling for marine fisheries. Reviews in Fisheries Science 14, 245–283.

Kristensen, K., Berg, C.W., 2012. DATRAS package for R. http://rforge.net/DATRAS/.

Kristensen, K., Lewy, P., Beyer, J.E., 2006. How to validate a length-based model of single-species fish stock dynamics. Canadian Journal of Fisheries and Aquatic Sciences 63, 2531–2542. http://www.nrcresearchpress.com/doi/pdf/10.1139/f06-135.

Lewy, P., Kristensen, K., 2009. Modelling the distribution of fish accounting for spatial correlation and overdispersion. Canadian Journal of Fisheries and Aquatic Sciences 66, 1809–1820.

Maunder, M., Starr, P., 2003. Fitting fisheries models to standardised CPUE abundance indices. Fisheries Research 63, 43–50.

Maunder, M.N., Punt, A.E., 2004. Standardizing catch and effort data: a review of recent approaches. Fisheries Research 70, 141–159.

Maxwell, D.L., Armstrong, M.J., Beggs, S., Aldridge, J.N., 2012. Annual egg production estimates of cod (gadus morhua), plaice (pleuronectes platessa) and haddock (melanogrammus aeglefinus) in the irish sea: The effects of modelling choices and assumptions. Fisheries Research 117-118, 146 – 155. Egg Production Methods in Marine Fisheries.

Myers, R.A., Cadigan, N.G., 1995. Statistical analysis of catch-at-age data with correlated errors. Canadian Journal of Fisheries and Aquatic Sciences 52, 1265–1273.

Myers, R.A., Pepin, P., 1990. The robustness of lognormal-based estimators of abundance. Biometrics 46, 1185–1192+.

Oehlert, G.W., 1992. A Note on the Delta Method. American Statistician 46, 27–29.

Pennington, M., 1983. Efficient estimators of abundance, for fish and plankton surveys. Biometrics 39.

Pennington, M., 1996. Estimating the mean and variance from highly skewed marine data. Fishery Bulletin (Washington D C) 94, 498–505.

Petrakis, G., MacLennan, D.N., Newton, A.W., 2001. Daynight and depth effects on catch rates during trawl surveys in the North Sea. ICES Journal of Marine Science 58, 50–60.

Piet, G.J., 2002. Using external information and GAMs to improve catch-at-age indices for North Sea plaice and sole. ICES Journal of Marine Science: Journal du Conseil 59, 624–632. http://icesjms.oxfordjournals.org/content/59/3/624.full.pdf+html.

R Development Core Team, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0.

Shono, H., 2008. Application of the Tweedie distribution to zero-catch data in CPUE analysis. Fisheries Research 93, 154–162.

Stefansson, G., 1996. Analysis of groundfish survey abundance data: combining the GLM and delta approaches. ICES Journal of Marine Science 53, 577–588.

Tweedie, M.C.K., 1984. An index which distinguishes between some important exponential families, in: Ghosh, J.K., Roy, J. (Eds.), Statistics: Applications and New Directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference, Calcutta: Indian Statistical Institute, pp. 579–604.

Walters, C., Punt, A., 1994. Placing Odds on Sustainable Catch Using Virtual Population Analysis and Survey Data. Canadian Journal of Fisheries and Aquatic Sciences 51, 946–958.

Wood, S., 2006a. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC.

Wood, S.N., 2006b. On confidence intervals for generalized additive models based on penalized regression splines. Australian & New Zealand Journal of Statistics 48, 445–464.

Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73, 3–36.

Online Supplemental Materials for "Evaluation of alternative age-based methods for estimating relative abundance from survey data in relation to assessment models".

Casper W. Berg, Anders Nielsen, and Kasper Kristensen

Calculation of age-length keys

Age-length keys are estimated using the spatially varying continuation ratio logit model described in [Berg and Kristensen(2012)]. Briefly, the procedure consists of estimating the probability distribution of age as a smooth function of length and geographical coordinates for each combination of year and quarter. Given K age groups, the continuation ratio model can be estimated by combining $K-1$ separately fitted binomial models, which models the conditional probability, π of a fish being of age a given that it is at least of age a. The model used for the conditional probabilities is

$$\text{logit}(\pi_{ayq}[\mathbf{x}_i]) = \alpha_{ayq} + \beta_{ayq} l_i + s_{ayq}(lon_i, lat_i) \tag{1}$$

where i denotes the ith fish, l denotes the length of the fish, (lon, lat) the geographical coordinates where the haul was taken (longitude and latitude), s_a is a thin plate spline in two dimensions, and (α_a, β_a) are ordinary regression parameters to be estimated. Subscripts y and q have been included here to indicate that each combination of year and quarter should have a distinct set of parameters to account for differences in population structure. The amount of smoothness imposed on the spline is selected using the BIC criterion.

Supplemental Tables

	Lognormal Q1	Gamma Q1	Tweedie Q1	Lognormal Q3	Gamma Q3	Tweedie Q3
LogLik	-72644.72	-74605.70	-80790.58	-38616.95	-39517.39	-44946.36
edf	714.91	889.02	606.87	596.88	707.59	533.42
AIC	146719.25	150989.45	162794.91	78427.65	80449.97	90959.56
BIC	152574.06	158270.18	167764.95	83201.18	86108.96	95225.63

Table 1: Sprat: Statistics for model selection.

	Lognormal	Gamma	Tweedie	Lognormal 2	Strat.Mean
1	0.29	0.04	0.16	0.30	0.17
2	0.32	0.35	0.32	0.03	-0.01
3	0.21	0.15	0.15	0.06	0.07
Avg	0.27	0.18	0.21	0.13	0.08

Table 2: Sprat: Internal consistencies for Q1

	Lognormal	Gamma	Tweedie	Lognormal 2	Strat.Mean
1	0.59	0.37	0.33	0.72	0.25
2	0.55	0.41	0.37	0.48	0.17
3	0.68	0.72	0.26	0.65	0.17
Avg	0.60	0.50	0.32	0.61	0.20

Table 3: Sprat: Internal consistencies for Q3

	Lognormal	Gamma	Tweedie	Lognormal 2	Strat.Mean
1	0.77	0.70	0.68	0.73	0.60
2	0.48	0.57	0.49	0.36	0.27
3	0.54	0.62	0.60	0.51	0.48
4	0.29	0.57	0.57	0.26	0.51
Avg	0.52	0.62	0.59	0.47	0.46

Table 4: Sprat: External consistencies between Q1 and Q3

	Lognormal Q1	Gamma Q1	Tweedie Q1	Lognormal Q3	Gamma Q3	Tweedie Q3
LogLik	-92717.73	-98242.00	-108218.05	-84566.42	-87351.84	-94750.61
edf	1105.15	1475.23	975.95	1111.25	1404.63	942.28
AIC	187645.76	199434.46	218388.00	171355.35	177512.93	191385.77
BIC	197144.57	212114.13	226776.30	180693.18	189316.02	199303.70

Table 5: Herring: Statistics for model selection.

	Lognormal	Gamma	Tweedie	Lognormal 2	Strat.Mean
1	0.78	0.77	0.80	0.75	0.45
2	0.60	0.58	0.52	0.63	-0.06
3	0.56	0.50	0.37	0.58	-0.16
4	0.51	0.41	0.25	0.43	0.00
Avg	0.61	0.56	0.48	0.60	0.06

Table 6: Herring: Internal consistencies for Q1

	Lognormal	Gamma	Tweedie	Lognormal 2	Strat.Mean
1	0.66	0.63	0.71	0.71	0.60
2	0.78	0.73	0.78	0.71	0.60
3	0.71	0.64	0.73	0.54	0.74
4	0.58	0.51	0.60	0.53	0.71
Avg	0.68	0.63	0.71	0.62	0.66

Table 7: Herring: Internal consistencies for Q3

	Lognormal	Gamma	Tweedie	Lognormal 2	Strat.Mean
1	0.60	0.63	0.67	0.49	0.47
2	0.78	0.68	0.71	0.74	0.41
3	0.62	0.48	0.49	0.52	0.31
4	0.57	0.43	0.38	0.49	0.54
5	0.37	0.50	0.40	0.28	0.54
Avg	0.59	0.54	0.53	0.50	0.46

Table 8: Herring: External consistencies between Q1 and Q3

	Lognormal Q1	Gamma Q1	Tweedie Q1	Lognormal Q3	Gamma Q3	Tweedie Q3
LogLik	-116092.89	-119248.16	-127048.47	-86692.51	-88571.49	-94994.59
edf	1091.92	1201.12	852.29	1074.62	1141.80	798.69
AIC	234369.61	240898.55	255801.52	175534.27	179426.58	191586.56
BIC	243754.68	251222.18	263126.98	184564.30	189021.09	198297.98

Table 9: Whiting: Statistics for model selection.

	Lognormal	Gamma	Tweedie	Lognormal 2	Strat.Mean
1	0.72	0.79	0.85	0.80	0.87
2	0.82	0.87	0.90	0.87	0.95
3	0.70	0.68	0.78	0.79	0.93
4	0.69	0.68	0.73	0.74	0.77
Avg	0.73	0.76	0.82	0.80	0.88

Table 10: Whiting: Internal consistencies for Q1

	Lognormal	Gamma	Tweedie	Lognormal 2	Strat.Mean
1	0.88	0.84	0.84	0.89	0.87
2	0.91	0.92	0.90	0.89	0.85
3	0.93	0.91	0.88	0.83	0.85
4	0.78	0.76	0.79	0.72	0.83
Avg	0.87	0.86	0.85	0.83	0.85

Table 11: Whiting: Internal consistencies for Q3

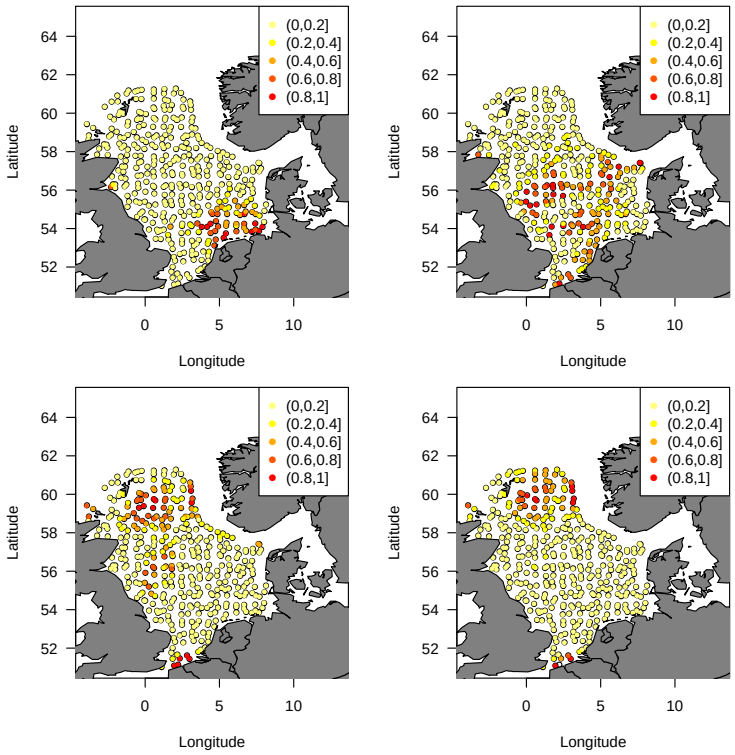
	Lognormal	Gamma	Tweedie	Lognormal 2	Strat.Mean
1	0.94	0.92	0.91	0.94	0.88
2	0.77	0.84	0.86	0.80	0.89
3	0.85	0.83	0.85	0.87	0.82
4	0.86	0.83	0.84	0.86	0.74
5	0.75	0.70	0.66	0.75	0.57
Avg	0.83	0.82	0.82	0.84	0.78

Table 12: Whiting: External consistencies between Q1 and Q3

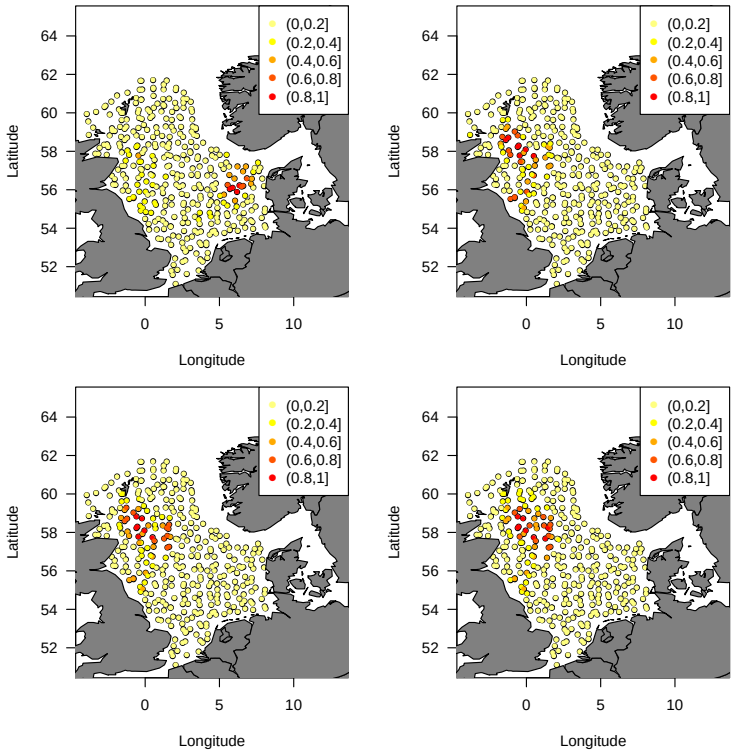
Supplemental Figures

The following distribution maps (or "concentration maps") should be interpreted as follows: The haul positions associated with each bin contain 20% of the total catch, and the (0.8,1.0] bin contains the *minimum* number of hauls that will account for the 20%, the (0.8,1.0] and (0.6,0.8] bins combined contain the *minimum* number of hauls that will account for 40% of the catch, and so on. For a more precise mathematical definition see [Lewy and Kristensen(2009)].

Herring



Supplemental Figure 1: Herring: Distribution map for ages 1 (top left), 2 (top right), 3 (bottom left), and 4 (bottom right) in Q1.



Supplemental Figure 2: Herring: Distribution map for ages 1 (top left), 2 (top right), 3 (bottom left), and 4 (bottom right) in Q3.

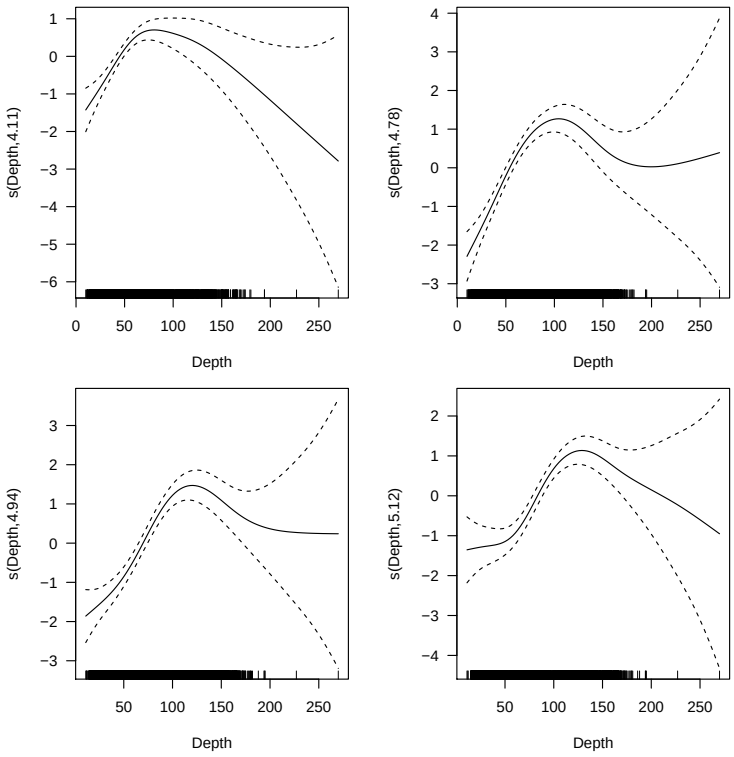


Figure 1: Herring: Depth effect for ages 1 (top left), 2 (top right), 3 (bottom left), and 4 (bottom right) in Q1

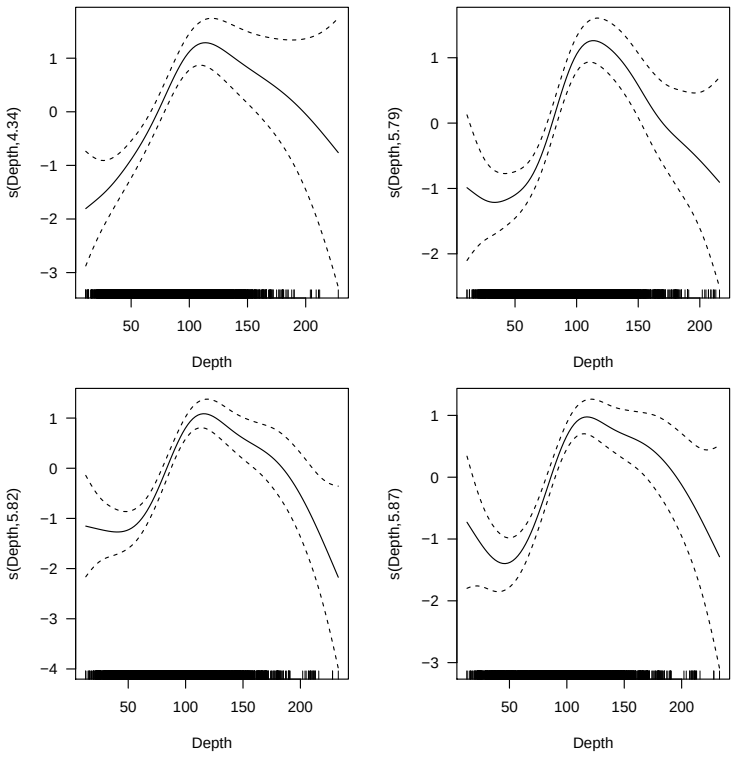


Figure 2: Herring: Depth effect for ages 1 (top left), 2 (top right), 3 (bottom left), and 4 (bottom right) in Q3

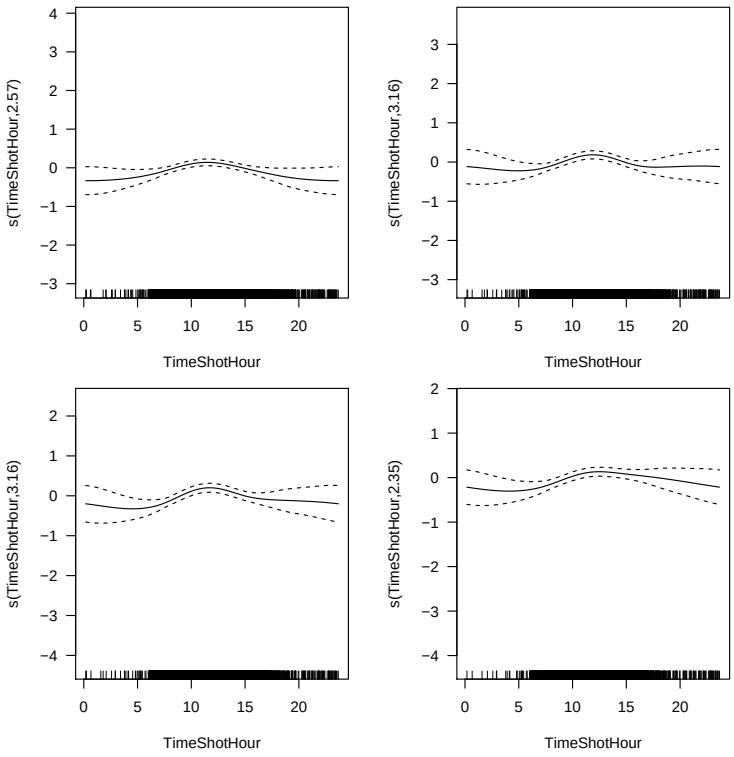


Figure 3: Herring: Time of day effect for ages 1 (top left), 2 (top right), 3 (bottom left), and 4 (bottom right) in Q1

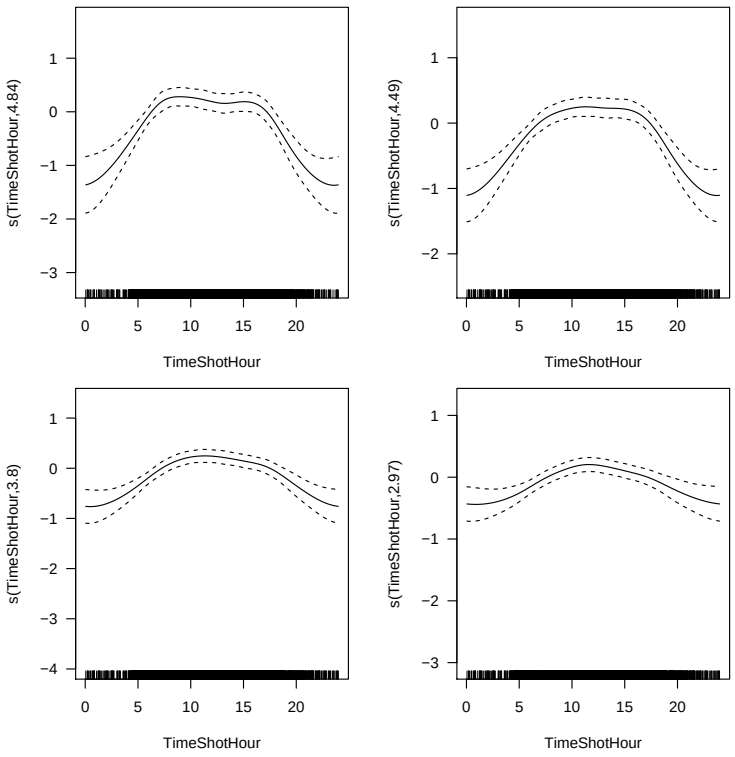


Figure 4: Herring: Depth effect for ages 1 (top left), 2 (top right), 3 (bottom left), and 4 (bottom right) in Q3

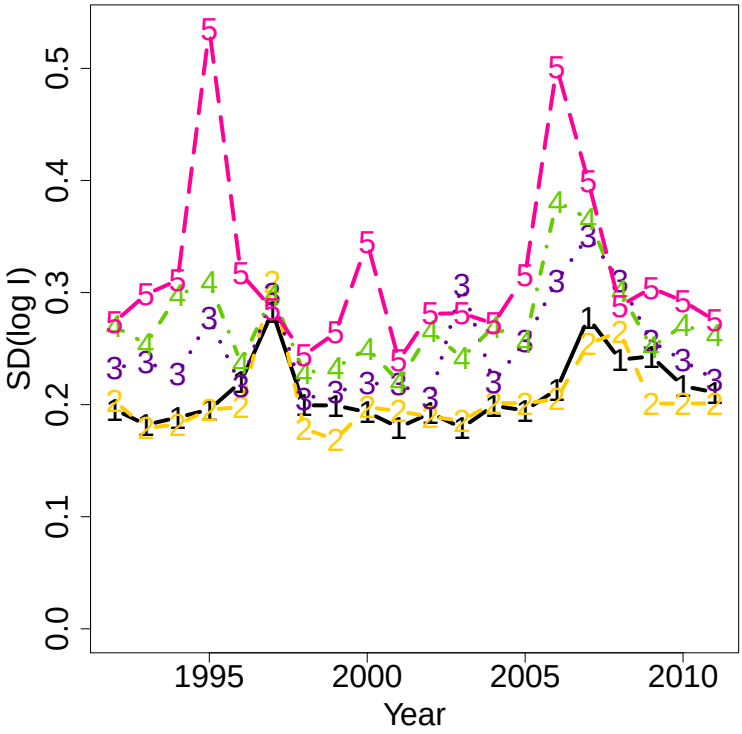


Figure 5: Herring: Estimated standard deviations on log(I) from bootstrapping Q1.

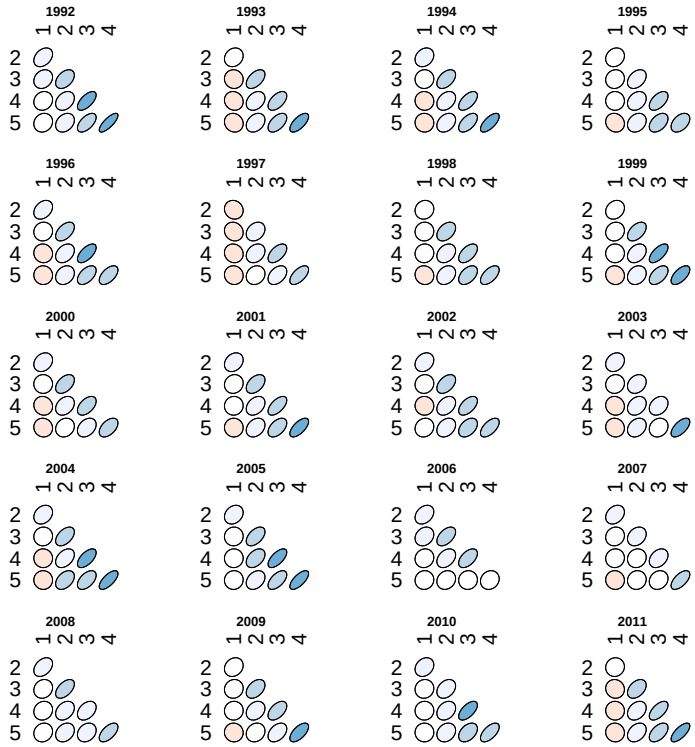


Figure 6: Herring: Estimated correlation matrices on log(I) from bootstrapping Q1.

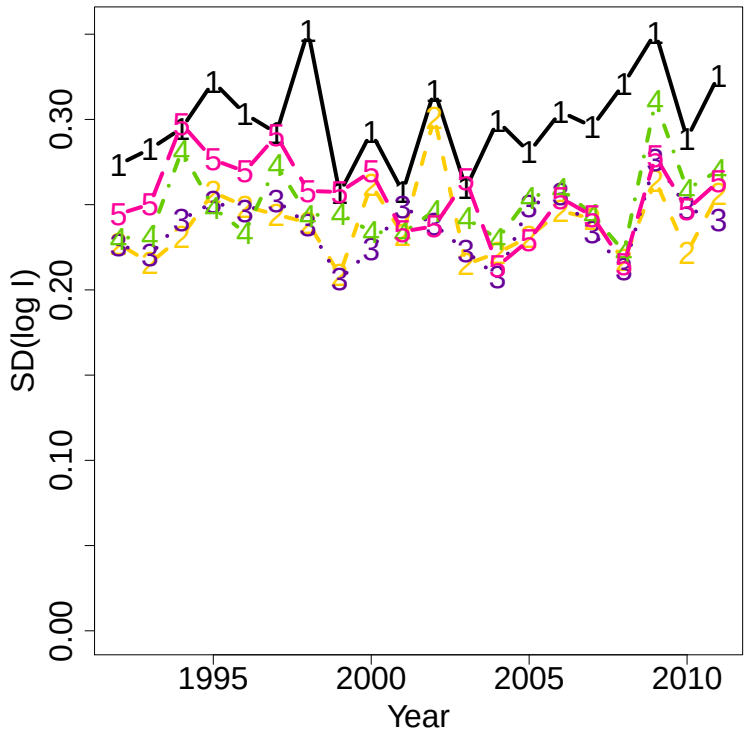


Figure 7: Herring: Estimated standard deviations on log(I) from bootstrapping Q3.

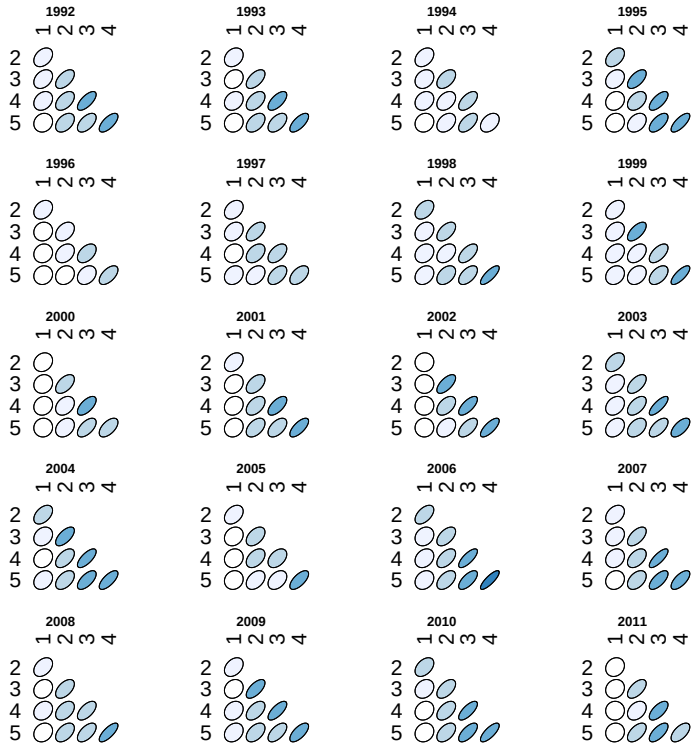


Figure 8: Herring: Estimated correlation matrices on log(I) from bootstrapping Q3.

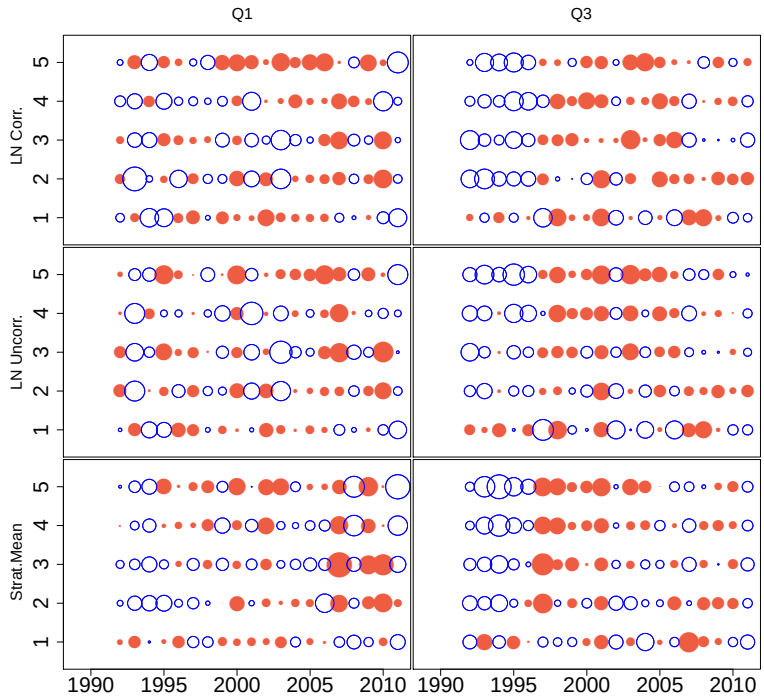
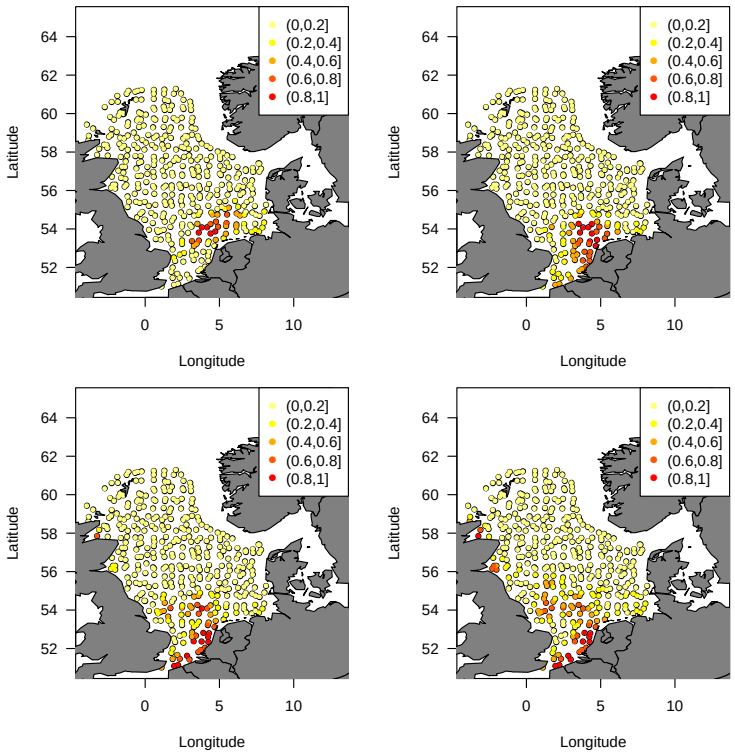
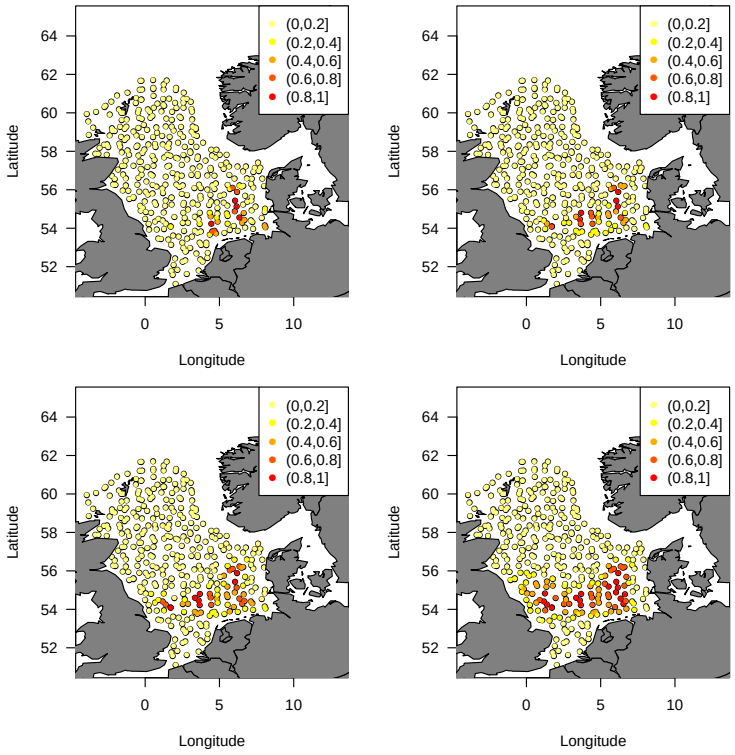


Figure 9: Herring: survey standardized residual plots from the fitted assessment models.

Sprat



Supplemental Figure 3: Sprat: Distribution map for ages 1 (top left), 2 (top right), 3 (bottom left), and 4 (bottom right) in Q1.



Supplemental Figure 4: Sprat: Distribution map for ages 1 (top left), 2 (top right), 3 (bottom left), and 4 (bottom right) in Q3.

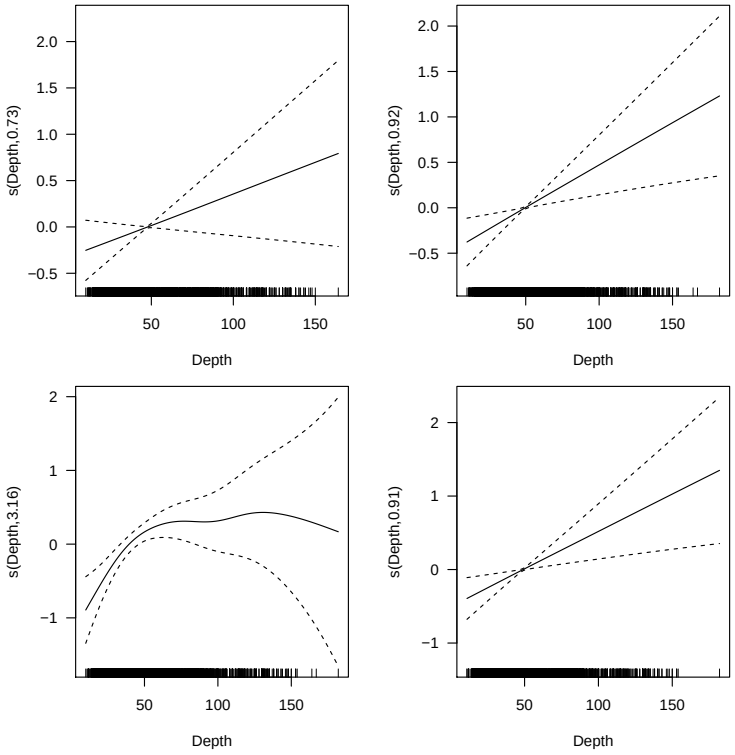


Figure 10: Sprat: Depth effect for ages 1 (top left), 2 (top right), 3 (bottom left), and 4 (bottom right) in Q1

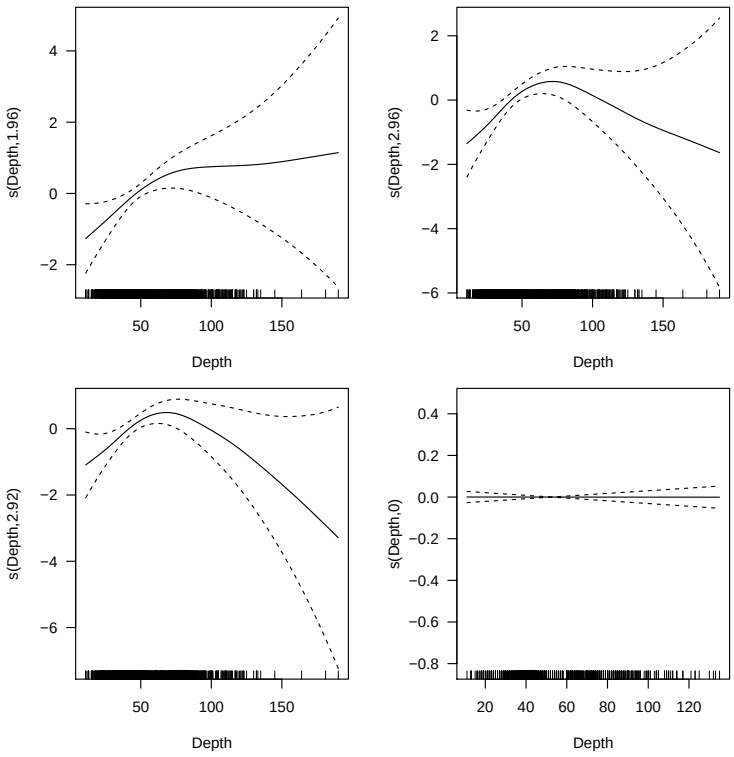


Figure 11: Sprat: Depth effect for ages 1 (top left), 2 (top right), 3 (bottom left), and 4 (bottom right) in Q3

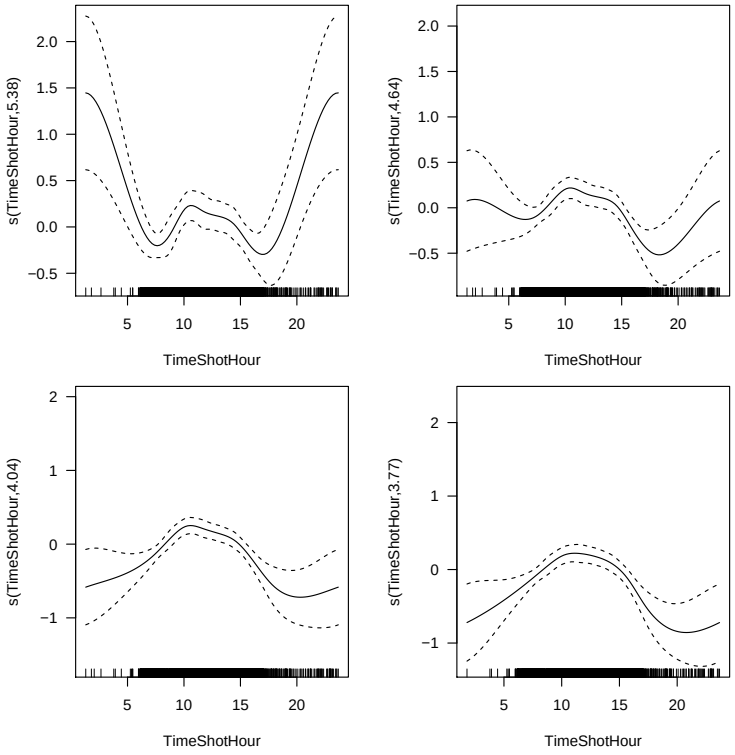


Figure 12: Sprat: Time of day effect for ages 1 (top left), 2 (top right), 3 (bottom left), and 4 (bottom right) in Q1

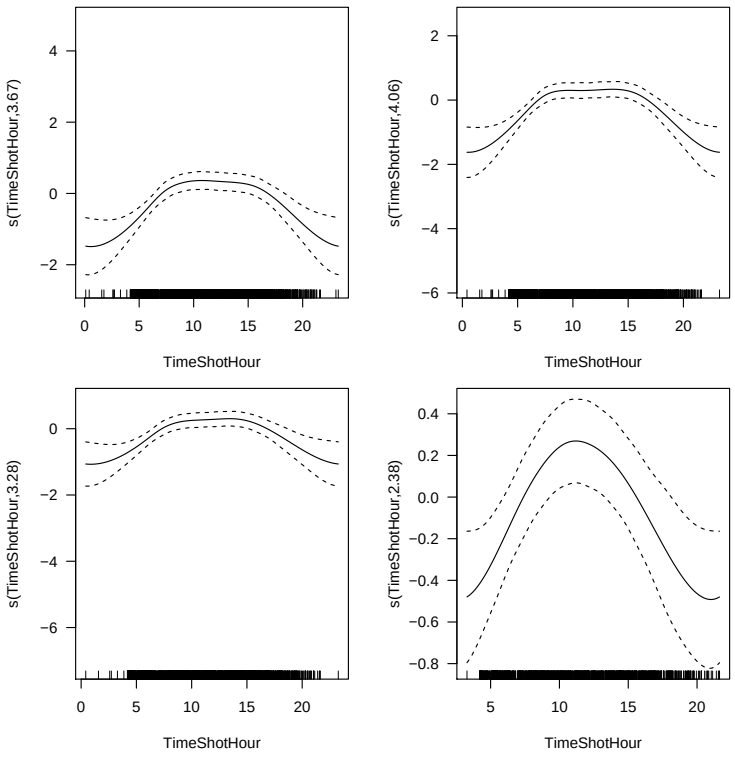


Figure 13: Sprat: Depth effect for ages 1 (top left), 2 (top right), 3 (bottom left), and 4 (bottom right) in Q3

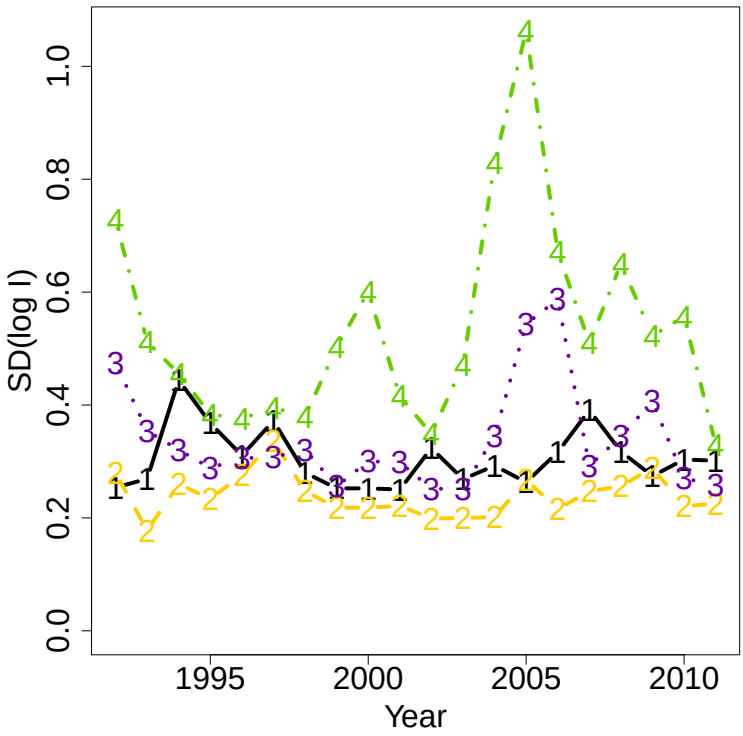


Figure 14: Sprat: Estimated standard deviations on log(I) from bootstrapping Q1.

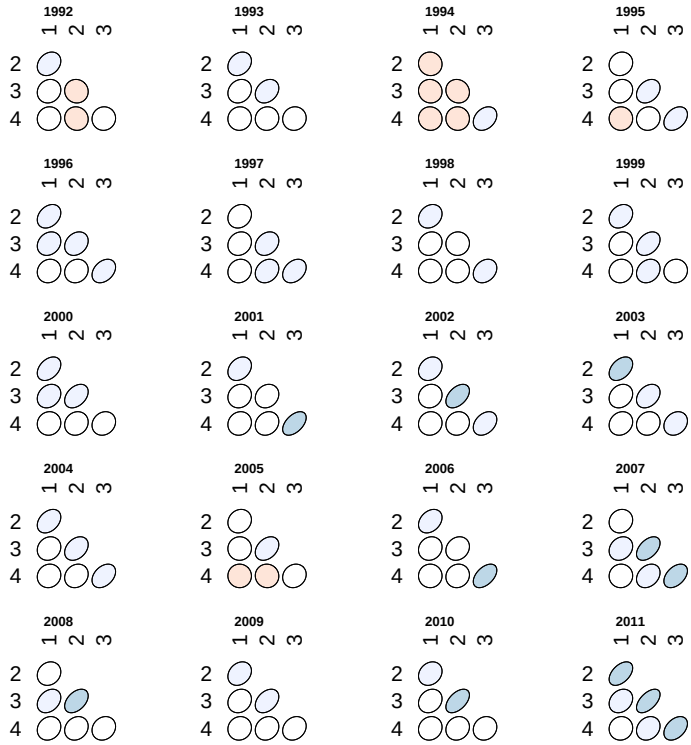


Figure 15: Sprat: Estimated correlation matrices on log(I) from bootstrapping Q1.

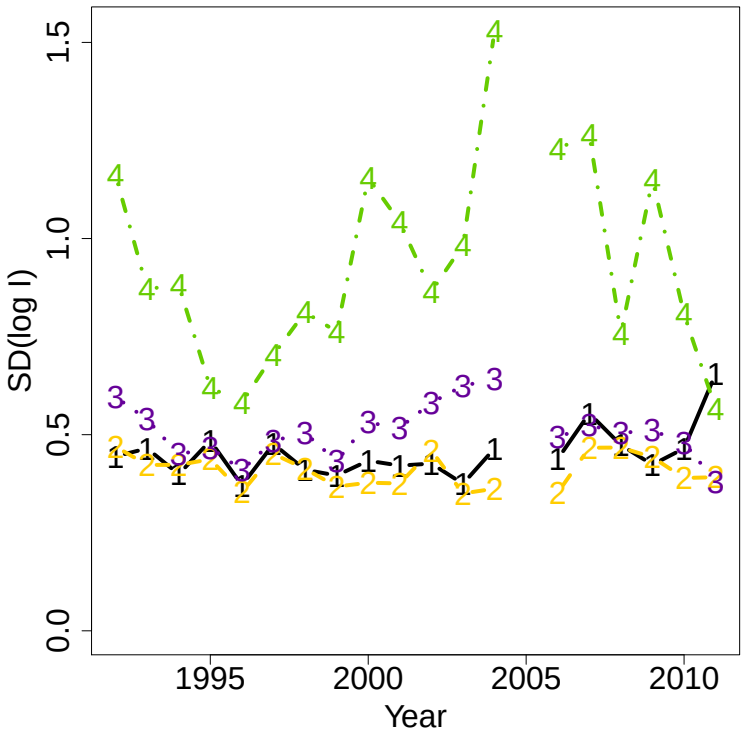


Figure 16: Sprat: Estimated standard deviations on log(I) from bootstrapping Q3.

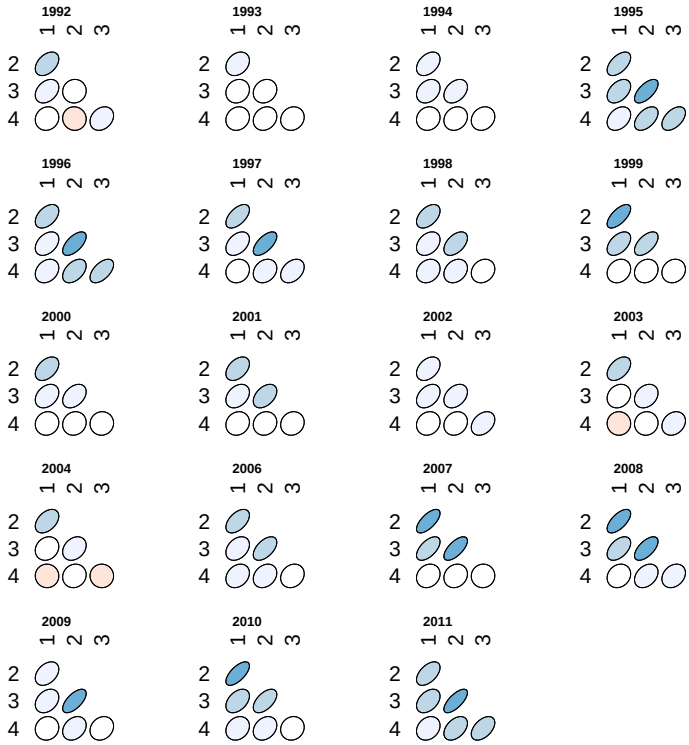


Figure 17: Sprat: Estimated correlation matrices on log(I) from bootstrapping Q3.

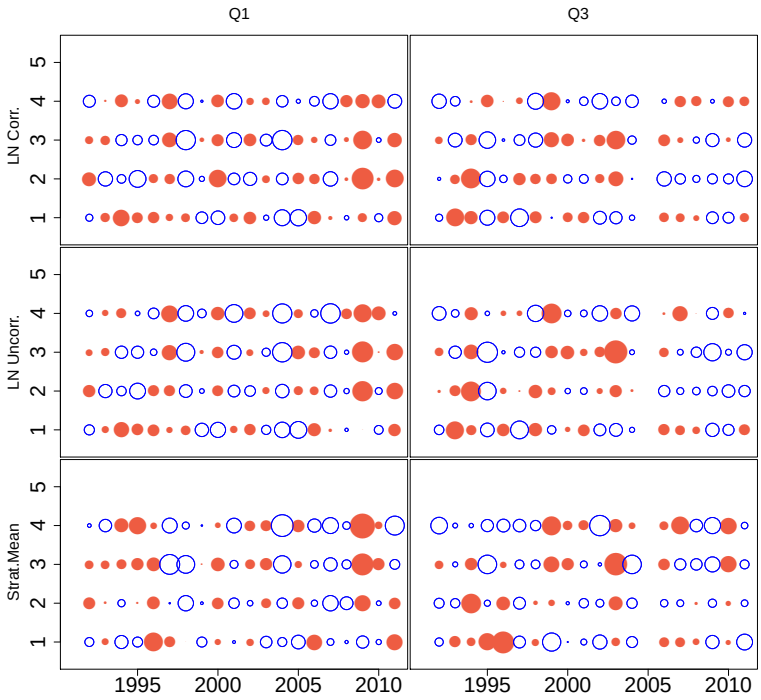
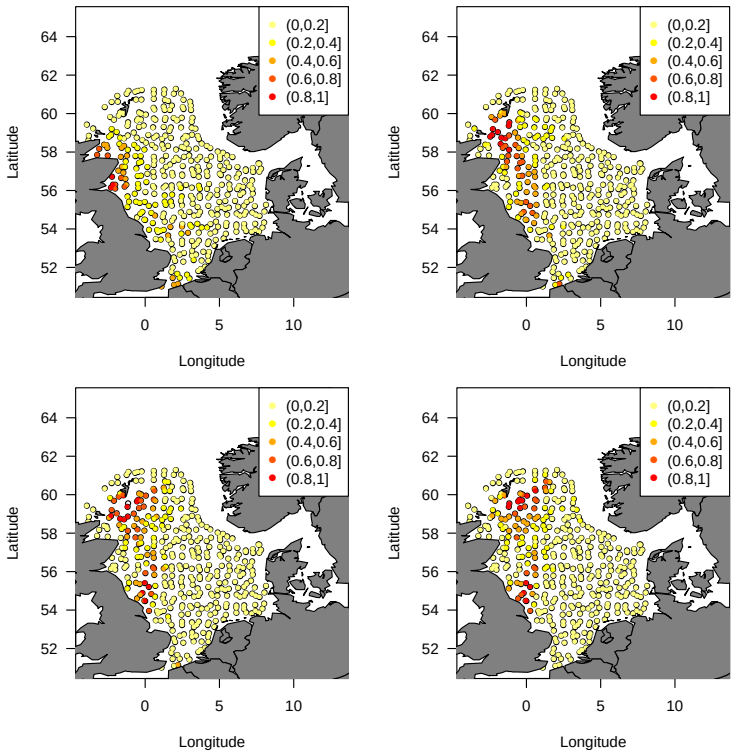
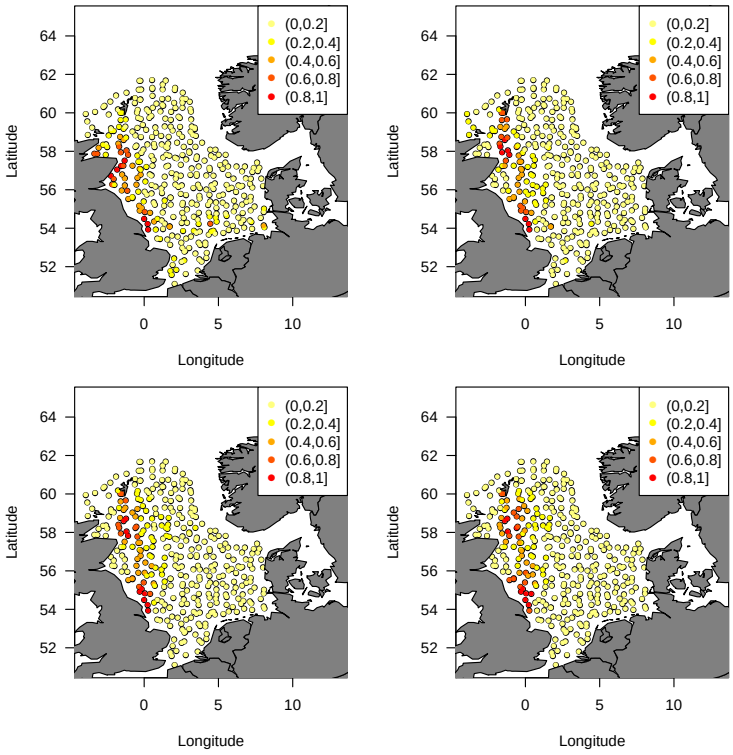


Figure 18: Sprat: survey standardized residual plots from the fitted assessment models.

Whiting



Supplemental Figure 5: Whiting: Distribution map for ages 1 (top left), 2 (top right), 3 (bottom left), and 4 (bottom right) in Q1.



Supplemental Figure 6: Whiting: Distribution map for ages 1 (top left), 2 (top right), 3 (bottom left), and 4 (bottom right) in Q3.

References

[Berg and Kristensen(2012)] Berg, C.W., Kristensen, K., 2012. Spatial age-length key modelling using continuation ratio logits. Fisheries Research 129-130, 119–126.

[Lewy and Kristensen(2009)] Lewy, P., Kristensen, K., 2009. Modelling the distribution of fish accounting for spatial correlation and overdispersion. Canadian Journal of Fisheries and Aquatic Sciences 66, 1809–1820.

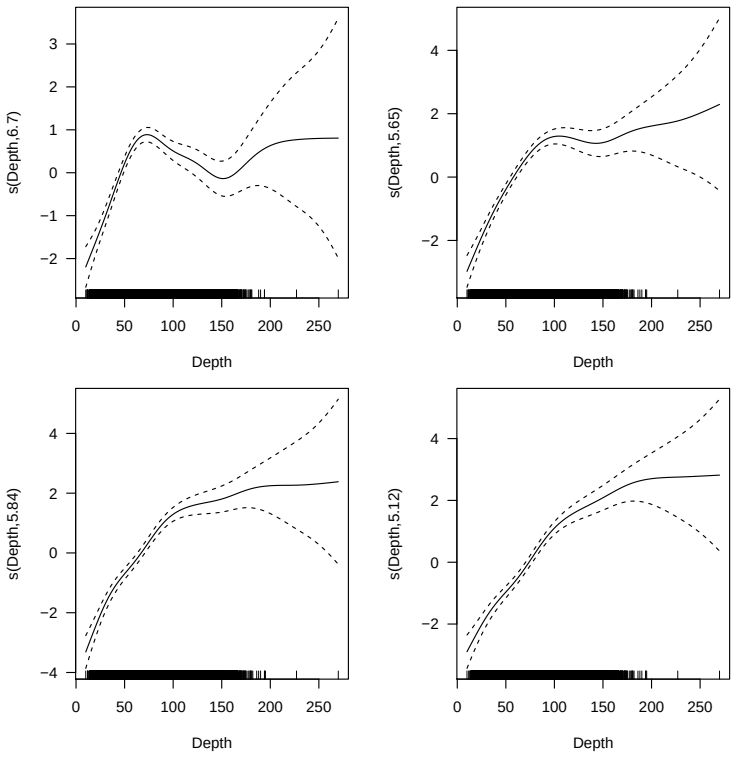


Figure 19: Whiting: Depth effect for ages 1 (top left), 2 (top right), 3 (bottom left), and 4 (bottom right) in Q1

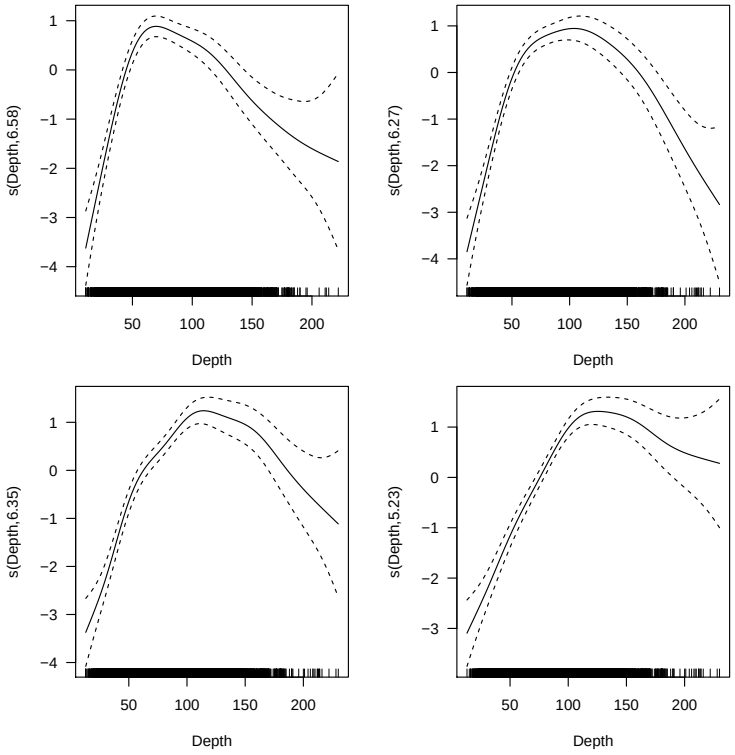


Figure 20: Whiting: Depth effect for ages 1 (top left), 2 (top right), 3 (bottom left), and 4 (bottom right) in Q3

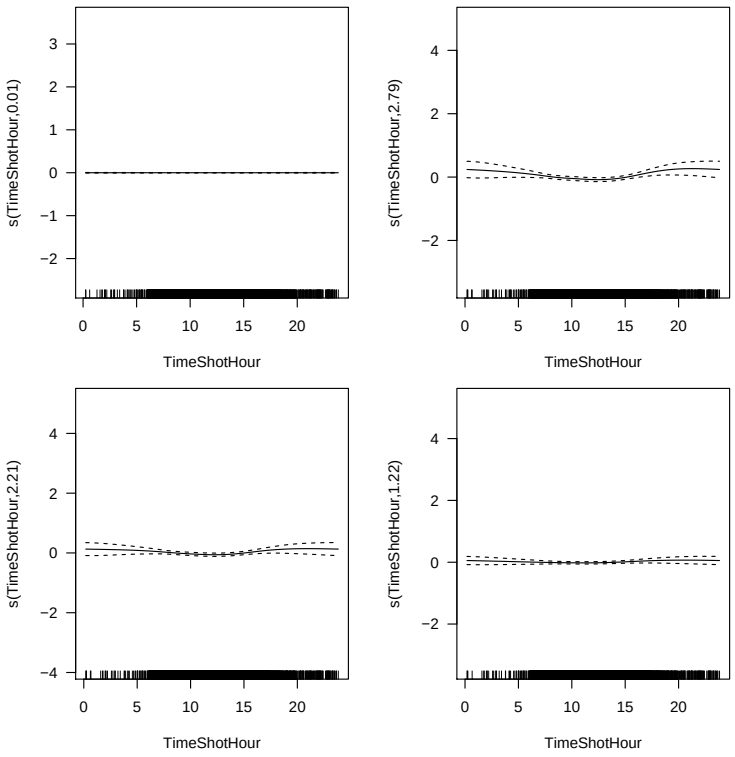


Figure 21: Whiting: Time of day effect for ages 1 (top left), 2 (top right), 3 (bottom left), and 4 (bottom right) in Q1

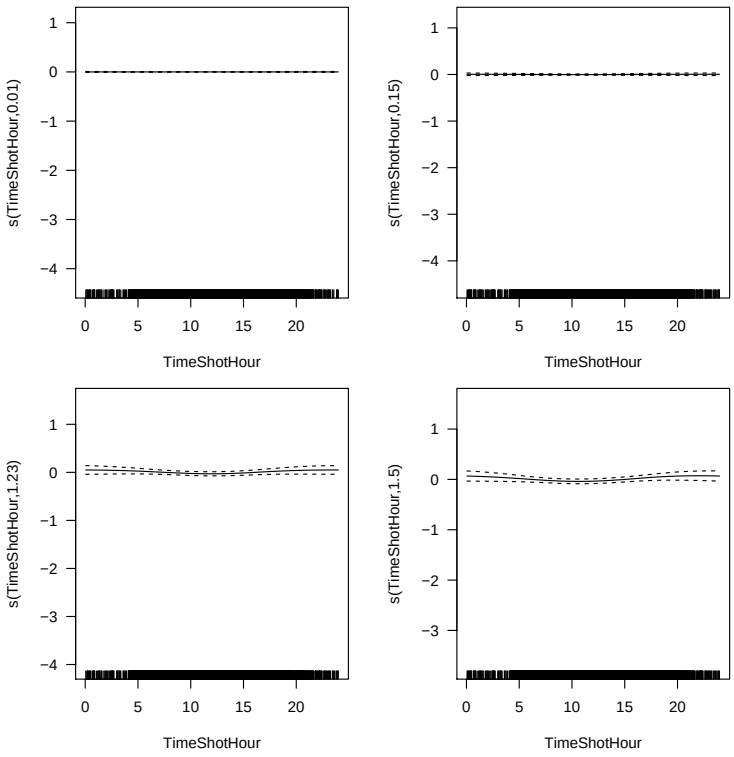


Figure 22: Whiting: Depth effect for ages 1 (top left), 2 (top right), 3 (bottom left), and 4 (bottom right) in Q3

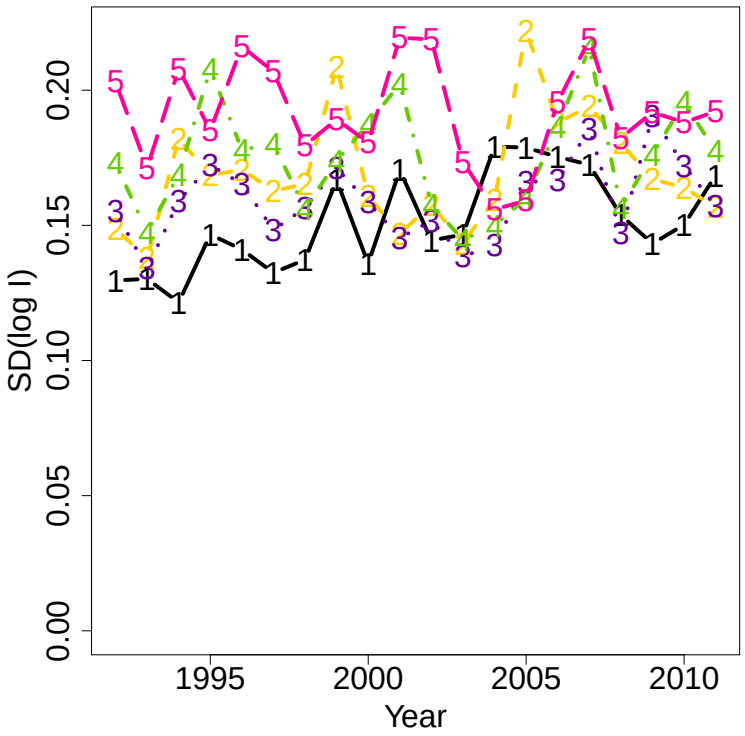


Figure 23: Whiting: Estimated standard deviations on log(I) from bootstrapping Q1.

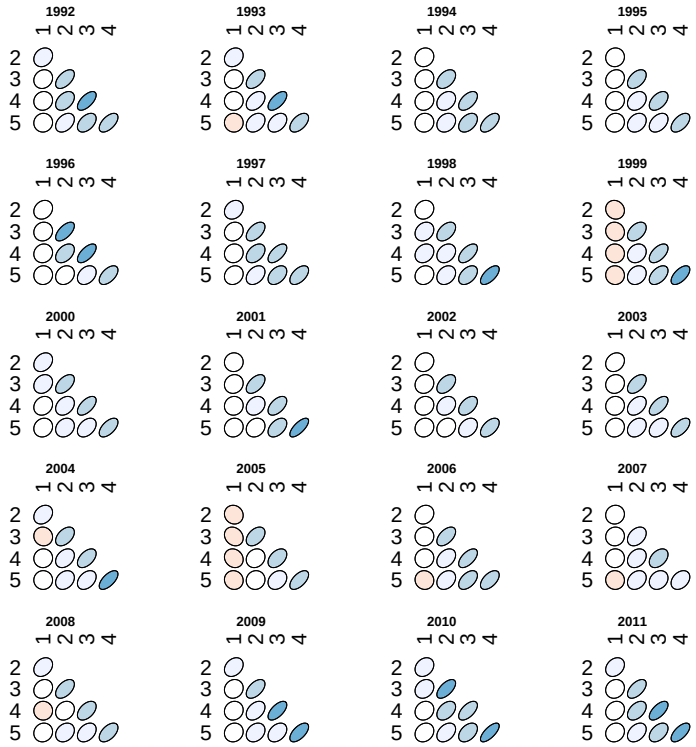


Figure 24: Whiting: Estimated correlation matrices on log(I) from bootstrapping Q1.

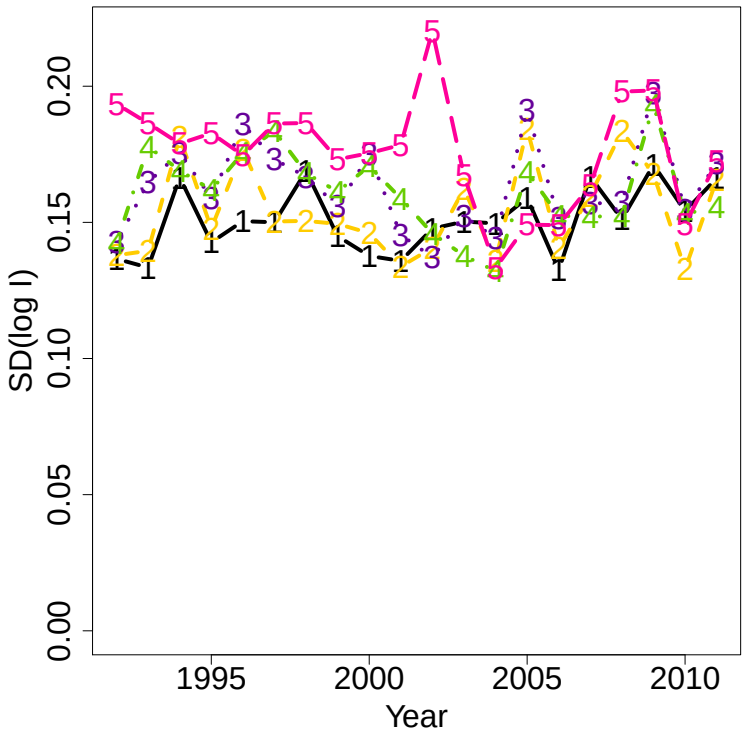


Figure 25: Whiting: Estimated standard deviations on log(I) from bootstrapping Q3.

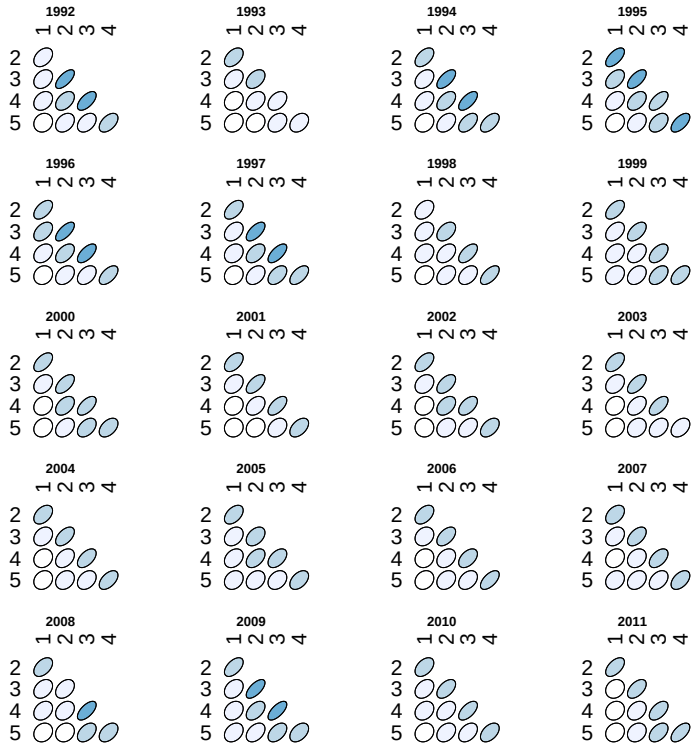


Figure 26: Whiting: Estimated correlation matrices on log(I) from bootstrapping Q3.

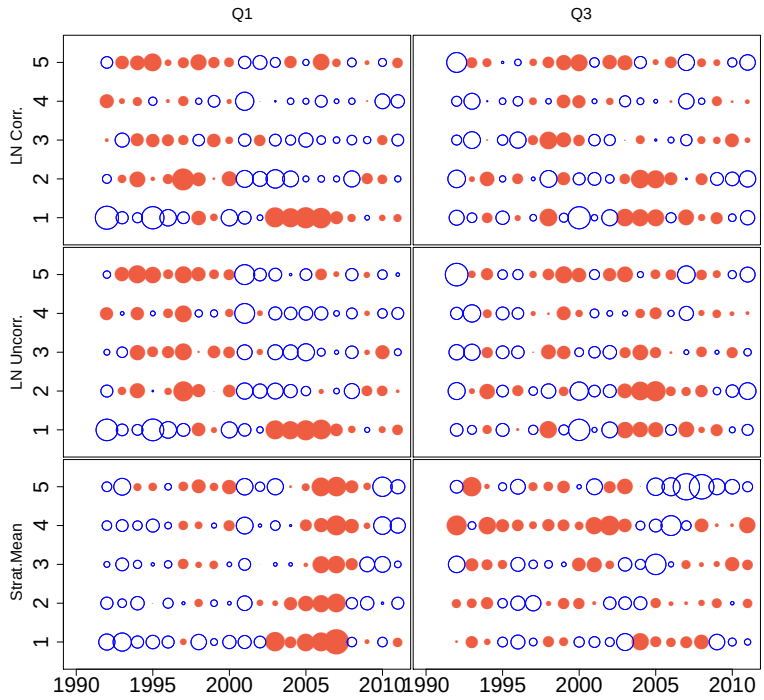


Figure 27: Whiting: survey standardized residual plots from the fitted assessment models.

Appendix G

Paper V

Joint assessment modelling of two herring stocks subject to mixed fisheries.

Casper Willestofte Berg Anders Nielsen

April 24, 2013

Abstract

We consider the problem of joint stock assessment of two stocks of herring with similar visual appearance, but with different origins and genes. There is a geographic zone in the middle separating the two stocks, where specimens of both stocks are found in the same hauls in what can be referred to as mixed fisheries.

We perform simulation and estimation of fishing mortalities and biomasses in three areas (two homogeneous and one mixed) using a state-space model formulation, where transport between areas occurs implicit between time-steps. The data used for the analysis consists of area-disaggregated commercial and survey catch-at-age numbers, as well as samples of the stock proportions from the mixing area.

Surveys covering multiple areas, such that catchability parameters could be assumed equal across areas, were important in estimating the abundance in the mixing area. Variation in relative year-class abundances between stocks explained the main variation in the observed composition in the mixing area, as opposed to changes in the spatial distribution of the two stocks between the three areas. The main advantages of our approach are 1) It does not require knowledge about movement rates e.g. from tagging experiments, 2) It requires fewer parameters than models which explicitly incorporate movement, 3) Unusual large cohorts observed in either stock are accounted for in the predictions of future catches in the mixing area.

1 Introduction

This work describes a stock assessment model featuring two stocks and three areas, where mixing between the two stocks occurs in the area in the middle. The purpose of this model is to provide an improvement over running two

separate single stock assessments using data that have been split by stock according to subsamples of the catches (the current practice for the case study, ICES (2012)) as well as a simple alternative to models that explicitly incorporate movement. Also, unlike the split single stock assessment approach, this model provides estimates of fishing mortality and numbers-at-age by stock in the mixing area, which can aid management decisions.

Ignoring sub-populations and movement in stock assessments may lead to local overexploitation (Kell et al., 2009; Montenegro et al., 2009; Goethel et al., 2011), but these are often hampered by the lack of high resolution data that are needed to facilitate the analyses.

Atlantic herring in the North Sea and Western Baltic are currently assessed as two separate stocks (ICES, 2012), although several smaller substock exist (Mariani et al., 2005; Bekkevold et al., 2007). These two stocks are characterised by differing spawning locations and their timing of spawning, the North Sea stock are mainly autumn spawners whereas the Western Baltic stock are spring spawners. North Sea autumn-spawning herring (NSAS) larvae drift into Skagerrak and the northern part of Kattegat (ICES area IIIa, see figure 1), which acts as nursery area until they reach sexual maturity at around age 2, at which time they leave the nursery area to spawn in the North Sea. The Western Baltic spring spawners (WBSS) perform age and season dependent feeding migrations into area IIIa, which leads to the appearance of both stocks in the same hauls (Ulrich et al., 2012). This puts this stock complex in the category of "natal homing with spatial overlap" using the classification by Goethel et al. (2011). NSAS and WBSS herring can be discriminated with a relatively high degree of confidence by otolith analysis (Mosegaard and Madsen, 1996; Clausen et al., 2007), and samples are routinely collected from the fisheries in order to apportion the catches to either stock. The two stocks are currently assessed separately using partitioned data on the basis of these samples, although restrictions on total allowable catch (TAC) are set for each of the areas shown in figure 1 (ICES, 2012). This procedure makes it quite complicated to give advice on the TAC in IIIa (ICES, 2010). No estimate of fishery mortality exist for this area, and the two separate assessments provide no direct means of predicting the future catch composition in IIIa, so the most recent yearly average split proportions are used for short-term projections.

This paper presents an integrated assessment of these two herring stocks, where the fishery and mixing process are explicitly accounted for. By doing this, we can present estimates of fishing mortality and numbers-at-age in all three areas (i.e. including the mixing area), as well as an alternative framework for providing predictions of the future catch composition here.

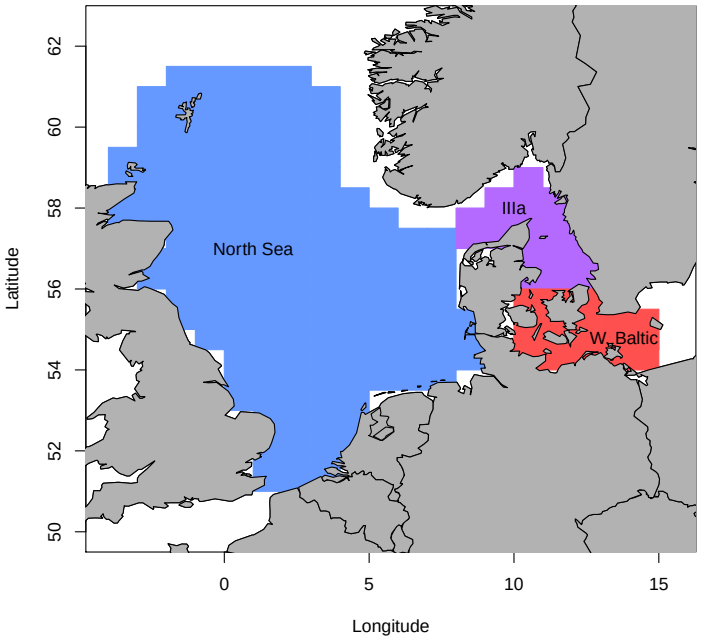


Figure 1: Map of areas defining the North Sea stock (blue), Western Baltic stock (red), and the mixing area, IIIa (purple).

Also, the uncertainty in the splitting proportions is quantified and accounted for in the calculations of all quantities of interest.

We examine the hypothesis, that the variations in observed split proportions in the mixing area can be explained by differences in relative cohort strength between the two stocks. The idea is illustrated in figure 2: let $(\pi^{(s)},N^{(s)})$ denote the proportion π of stock s with a total abundance of N, that is located inside the mixing area during a time-step. The values of $\pi^{(s)}$ are the same in a) and b), but the size of the NSAS cohort is doubled from a) to b) resulting in a substantial change in the proportion of autumn

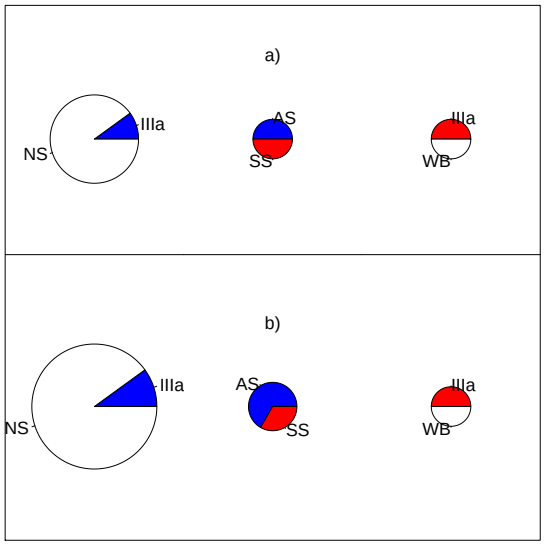


Figure 2: Illustration of how observed proportions in the mixing area (middle column) can be driven by changes in the larger population [left column, doubled from a) to b)].
a) $\pi^{(1)} = 0.1, \pi^{(2)} = 0.5, N^{(1)} = 5, N^{(2)} = 1, p_{AS} = 0.5$.
b) $\pi^{(1)} = 0.1, \pi^{(2)} = 0.5, N^{(1)} = 10, N^{(2)} = 1, p_{AS} = 0.67$.

spawners, p_{AS}. An alternative hypothesis, that changes in $\pi^{(s)}$ over time is responsible for the changes in split proportions, is also examined.

The assessment model is an extension of the single stock state-space model described in Berg et al.. State-space models are appealing because they can separate process from observation errors and are well suited to handle missing observations, which has led to their increased use within the field of stock assessment and ecology in general (Gudmundsson and Gunnlaugsson, 2012; Pedersen et al., 2011).

2 Materials and Methods

2.1 Data

The data consist of survey indices and commercial catches by age and area, as well as samples of the catch composition in the mixing area. Auxiliary information about weight, the proportion of sexually mature individuals, and natural mortality by age and stock is also needed. The auxiliary information is assumed to be known in advance and without error, whereas the data is assumed to contain random noise. This data are similar to the current assessments, so we refer the reader to ICES (2012) and the supplementary materials for a more detailed description of these. However, time-series of survey data calculated by area using the areas defined in figure 1 were not immediately available, so these were calculated using the methodology described in Berg and Kristensen (2012) and Berg et al..

2.2 Assessment model

We consider two stocks, $s = 1, 2$, each having A age classes $a = 1 \dots A$. Furthermore, we consider three areas $k = 1 \dots 3$, where individuals belonging to stock $s = 1$ can be located in either area $k = 1$ or 2, but not $k = 3$, and individuals belonging to stock $s = 2$ can be located in area $k = 2$ or 3, but not $k = 1$. The state vector consists of the log-transformed number of individuals in each age class for each stock, $\log N_a^{(s)}$, the log-transformed fisheries mortalities in each area $\log F_a^{(k)}$, and finally the logit-proportion of age class a from stock s that is located in the mixing area (k=2), $\beta_a^{(s)} = \log\left(\frac{\pi_a^{(s)}}{1-\pi_a^{(s)}}\right)$. The logit transformation is useful here, since the unit interval for π is stretched to cover the whole real axis for β. Some age classes may be grouped together with respect to F and β to reduce the number of parameters.

The system equations are (with age index omitted):

$$\log N_{s,t} = \log\left(N_{s,t-1}e^{-Z_{s,t-1}}\right) + \epsilon_{N,s,t} \quad (1)$$

$$\log F_{k,t} = \log F_{k,t-1} + \epsilon_{F,k,t} \quad (2)$$

$$\beta_{s,t} = \beta_{s,t-1} + \epsilon_{\beta,s,t} \quad (3)$$

and the observation equations are:

$$\log C_{k,t} = \log\left(\sum_s \left[\frac{F_{k,t}}{Z_{s,k,t}}(1-e^{-Z_{s,k,t}})N_{s,k,t}\right]\right) + \epsilon_{C,k,t} \quad (4)$$

$$\log I_{f,k,t} = \log\left(Q_{f,k,t}e^{D_f/365}\left(\sum_s e^{-Z_{s,k,t}}N_{s,k,t}\right)\right) + \epsilon_{I,f,k,t} \quad (5)$$

$$\text{logit}(p_{AS,t}) = \text{logit}\left(\frac{\pi^{(1)}N^{(1)}}{\pi^{(1)}N^{(1)}+\pi^{(2)}N^{(2)}}\right) + \epsilon_{p_{AS},t} \quad (6)$$

where C denotes catch in numbers, I denotes indices of abundance from scientific surveys carried out at day D_f, and p_{AS} is the proportion of autumn spawners in the mixing area. Subscript k denote area, s stock, t time, and f is fleet. $Z = F + M$ is the total mortality (fishing + natural). The model parameters consist of process variances $(\sigma_N^2, \sigma_F^2, \sigma_\beta^2)$, observation variances $(\sigma_C^2, \sigma_I^2, \sigma_p^2)$, and catchability parameters Q. All model parameters may be grouped together with respect to fleets, areas, ages or stocks, but only older age groups within the same survey fleet and survey fleets operating in multiple areas were considered as candidates for grouping. Model reduction by grouping of parameters was tested using likelihood ratio tests and the final model was validated by inspection of standardized residuals.

2.3 Simulation study

Since the state-space model describes how the latent states change trough time and how data are related to the states through probability distributions, it is straight-forward to simulate artificial data given the parameters and the initial state. Simulation of data followed by estimation can be used to test, whether it is even possible estimate the quantities of interest given the data, i.e. whether the model is identifiable. To this end, we simulated numerous data sets of same size and similar characteristics to the real data.

2.4 Software

All data preprocessing was handled in R (R Development Core Team, 2012), and trawl survey data was handled using the DATRAS package (Kristensen and Berg, 2012). The stock assessment model was fitted using ADMB (Fournier et al., 2012). The source code can be obtained by contacting the corresponding author.

3 Results

3.1 Simulation study

The simulation study revealed, that it was difficult to achieve convergence unless catchabilities could be assumed equal across two surveys covering the mixing area and the Western Baltic respectively. When this assumption was made, it was generally (but not always) possible achieve convergence of the model, and in that case the model was able to reconstruct the true latent states from the data and the coverage of the estimated confidence bounds appeared reasonable.

3.2 Estimation results

The estimated time-series of log-numbers-at-age for the three areas are shown in figure 3. We note, that the largest NSAS cohort was recruited in 2001, and that this cohort remained strong through all age groups, but that recruitment has been low in the following time period.

The estimated proportions of the stocks inside the mixing area, π, (figure 4) reveal, that the only significant part of the NSAS population that occupy the mixing area is the 1 and 2-year-olds, whereas ages 0 and 3+ of NSAS origin in IIIa constitute close to zero percent of the total NSAS population (top row). There is generally little temporal development in the proportions of the two stocks occupying the mixing area (top and bottom rows), except for age 1 of the WBSS stock (bottom row, second column), where a slight positive trend is observed. Despite the small changes in π over time, the observed mixing proportions, p_{AS}, exhibit pronounced temporal variation (figure 4, middle row). The large 2001 NSAS cohort and the subsequent poor recruitment is clearly reflected in the estimates of p_{AS} as well as in the observations, which supports the hypothesis that the variations in observed split proportions in the mixing area can be explained by differences in relative cohort strength between the two stocks.

The estimated average fishing mortalities ($\bar{F}$) by area display different trends over time (5). While $\bar{F}^{(1)}$ and $\bar{F}^{(2)}$ (North Sea and mixing area) have dropped substantially from the beginning to the end of the time-series, the same is not true for $\bar{F}^{(3)}$. The level of fishing pressure is estimated to be substantial higher in the mixing area than the two other areas. The estimates are, however, also quite uncertain and dependent on the assumption of equal catchabilities across area 2 and 3, so this might be an artifact of this assumption.

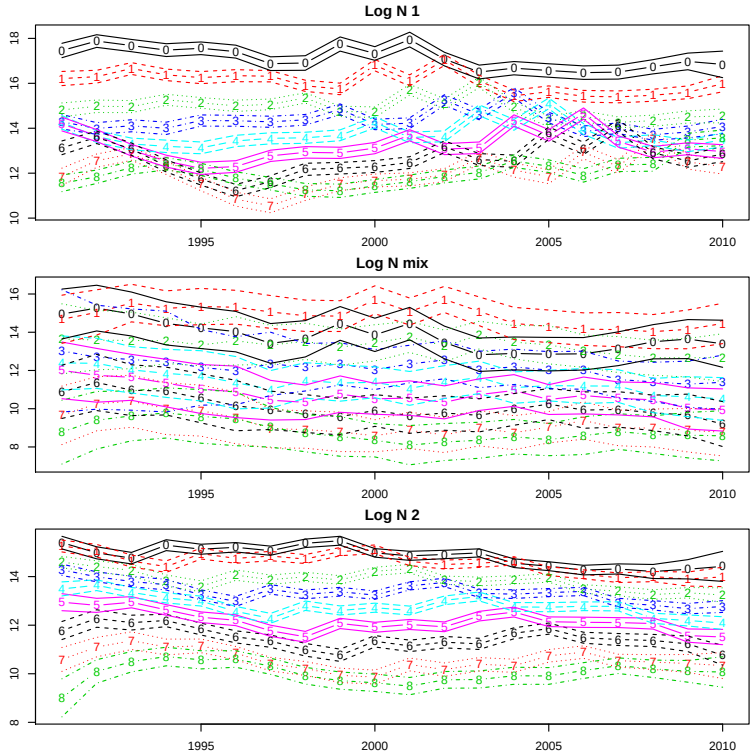


Figure 3: Estimated time-series of log-numbers-at-age for the three areas with 95% marginal confidence limits. Notice the distinct M-shaped recruitment pattern around year 2000 in the top figure (NSAS) rippling through the age-groups, and that the same pattern can be found in the mixing area (middle figure).

The estimated observation errors for each of the survey time-series reveal, that the acoustic surveys (NS-Acous and GERAS) contain significantly less noise than all of the trawl surveys (figure 6).

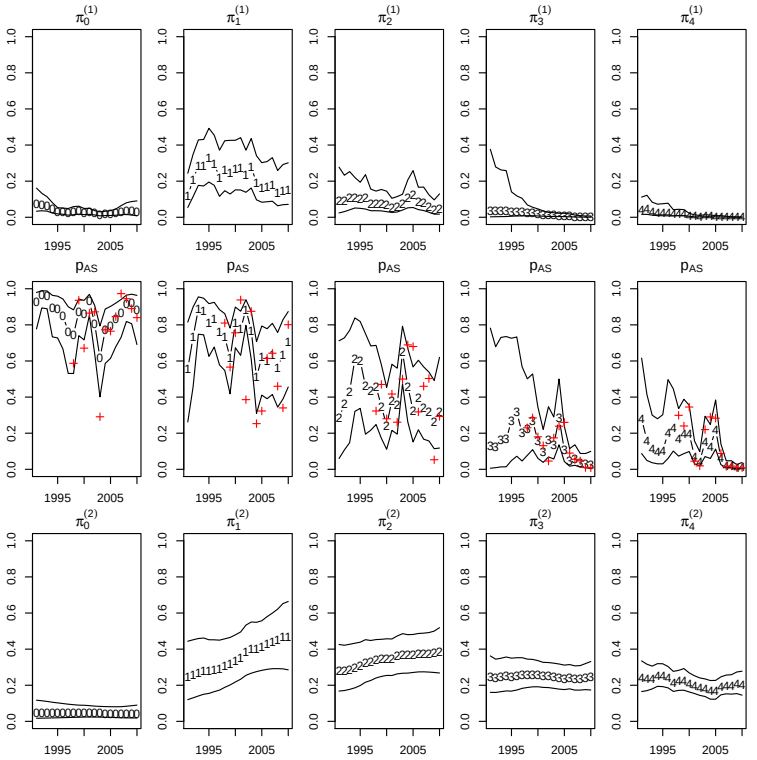


Figure 4: Top row: Estimated proportions of stock 1 (NSAS) inside the mixing area by age. Middle row: Estimated (black) proportions of NSAS in the mixing area. Red crosses indicate observed values. Bottom row: Estimated proportions of stock 2 (WBSS) inside the mixing area by age. Columns represent age groups and Solid lines indicate 95% confidence intervals.

4 Discussion

The incorporation of movement, sub-populations and spatial structure in stock assessments is still the exception rather than the rule. Lack of long time-series of data in high spatial and temporal resolution is a major obstacle in this respect, since this is a prerequisite for many spatially explicit models

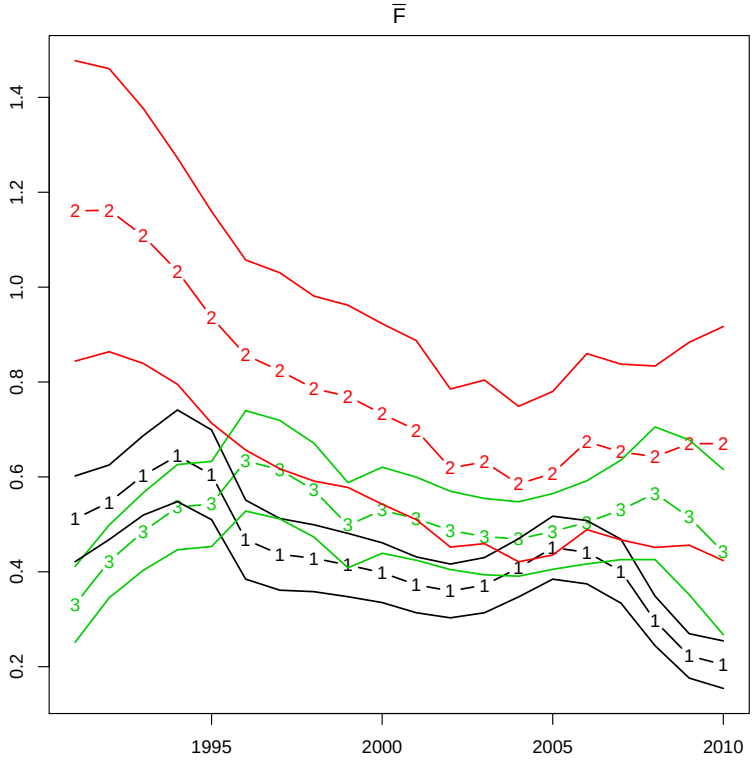


Figure 5: Average fishing mortality by area (black=NSAS, red=IIIa (mix), green=WBSS).

(Kleiber and Hampton, 1994). Tagging data provide a good source of information about movement rates, and these have successfully been integrated into stock assessment models (Sibert et al., 1999; Hampton and Fournier, 2001; Goethel et al., 2011). However, tagging data are not always available, so alternative methods are still needed. The present analysis should be seen as a first step in a transition from a methodology using two single stock assessments using partitioned data, and as an option when tagging data are not available. The proposed model is able to produce an estimate of the fish-

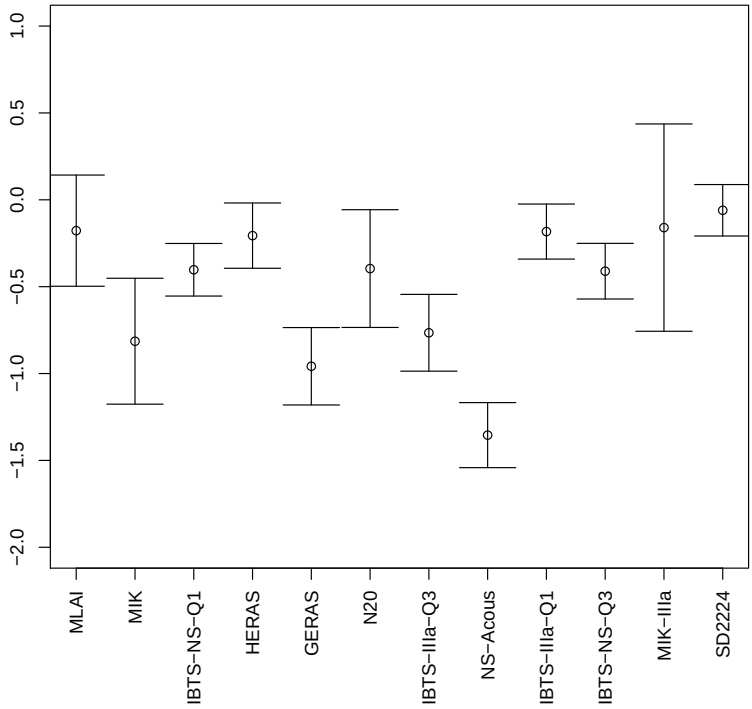


Figure 6: Estimated values of the logarithm of the observation standard deviations including 95% confidence intervals.

ing mortality time-series as well as numbers-at-age in the mixing area, which has not been done before. It captures the main dynamics of the signal in the observed mixing proportions, which stems from large recruitment events in the more abundant of the two species considered. The maximum likelihood framework enables us to draw inferences about changes in the state-vector over time, and the state-space formulation of the model permits separation of process and observation error, such that different data sources and be combined and weighted according to the ML-estimate of their observation error. This is important in many applications besides this one, where quality of different data sources varies greatly. Many stock assessment models, and

especially spatially explicit ones, often ignore either process or observation error (Quinn et al., 1990; Goethel et al., 2011) due to the added complexity. Advanced software packages that can effectively handle both types of error are in rapid development (Pedersen et al., 2011; Fournier et al., 2012), and the present analysis would have been nearly impossible without the ADMB tool used – more than 50 parameters were estimated (as well as several hundred latent states) in less than one hour on a standard laptop.

It was often not possible to achieve proper convergence of the model, unless two surveys covering two different areas could be assumed to have equal catchabilites. This result underlines the importance of having overlapping replicates in space and time when different research vessel types are used, as this proved essential in this example. Such data overlaps permits estimation of the relative efficiency of different vessels/gear types, and since survey data are relative measures of abundance, equal catchability may be assumed after rescaling the catches by the efficiency of the vessel that caught it. Designers of survey experiments should therefore aim for such overlaps whenever possible.

We did not perform single species assessments to compare results with, so we cannot infer anything from this analysis about how the method compares to split single species assessments with respect to estimation accuracy. However, this model can, unlike the two single stock assessments, capture and predict radical changes in the mixing proportions by age, as a result of the eventual strong cohort a from large recruitment event is rippling down the catch-at-age matrix for one of the stocks.

The problem of jointly assessing two spatially overlapping populations with natal homing is not uncommon, nor is the situation where no or little tagging information is available to estimate movement rates. Our proposed model may therefore prove useful in many other cases than the one examined.

References

Bekkevold, D., Clausen, L.A.W., Mariani, S., Andre, C., Christensen, T.B., Mosegaard, H., 2007. Divergent origins of sympatric herring population components determined using genetic mixture analysis. Marine Ecology Progress Series 337, 187–196.

Berg, C.W., Kristensen, K., 2012. Spatial age-length key modelling using continuation ratio logits. Fisheries Research 129-130, 119–126.

Berg, C.W., Nielsen, A., Kristensen, K., . Evaluation of alternative age-based method for estimating relative abundance from survey data in relation to assessment models. Unpublished .

Clausen, L.A.W., Bekkevold, D., Hatfield, E.M.C., Mosegaard, H., 2007. Application and validation of otolith microstructure as a stock identification method in mixed atlantic herring (clupea harengus) stocks in the north sea and western baltic. ICES Journal of Marine Science: Journal du Conseil 64, 377–385. http://icesjms.oxfordjournals.org/content/64/2/377.full.pdf+html.

Fournier, D.A., Hampton, J., Sibert, J.R., 1998. Multifan-cl: A length-based, age-structured model for fisheries stock assessment, with application to sout pacific albacore, thunnus alalunga. Canadian Journal of Fisheries and Aquatic Sciences 55, 2105–2116.

Fournier, D.A., Skaug, H.J., Ancheta, J., Sibert, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., 2012. Ad model builder: Using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods and Software 27, 233–249.

Goethel, D.R., Quinn, Terrance J., I., Cadrin, S.X., 2011. Incorporating spatial structure in stock assessment: Movement modeling in marine fish population dynamics. Reviews in Fisheries Science 19, 119–136.

Gudmundsson, Gunnlaugsson, 2012. Selection and estimation of sequential catch-at-age models. Canadian Journal of Fisheries and Aquatic Sciences 69, 1760.

Hampton, J., Fournier, D.A., 2001. A spatially disaggregated, length-based, age-structured population model of yellowfin tuna (thunnus albacares) in the western and central pacific ocean. Marine and Freshwater Research 52, 937–963.

ICES, 2010. Report of the workshop on procedures to establish the appropriate level of the mixed herring tac (spring western baltic (wbss) and autumn spawning north sea (nsas) stocks) in skagerrak and kattegat (division iiia). ICES Document CM 2010/ACOM:64 .

ICES, 2012. Report of the herring assessment working group for the area south of 62 deg n (hawg). ICES Document CM 2012/ACOM:06 .

Kell, L.T., Dickey-Collas, M., Hintzen, N.T., Nash, R.D.M., Pilling, G.M., Roel, B.A., 2009. Lumpers or splitters? evaluating recovery and management plans for metapopulations of herring. ICES Journal of Marine Science: Journal du Conseil 66, 1776–1783. http://icesjms.oxfordjournals.org/content/66/8/1776.full.pdf+html.

Kleiber, P., Hampton, J., 1994. Modeling effects of fads and islands on movement of skipjack tuna (katsuwonus pelamis): Estimating parameters from tagging data. Canadian Journal of Fisheries and Aquatic Sciences 51, 2642–2653.

Kristensen, K., Berg, C.W., 2012. Datras package for r. http://rforge.net/DATRAS/.

Mariani, S., Hutchinson, W., Hatfield, E., Ruzzante, D., Simmonds, E., Dahlgren, T., Andre, C., Brigham, J., Torstensen, E., Carvalho, G., 2005. North sea herring population structure revealed by microsatellite analysis. MARINE ECOLOGY PROGRESS SERIES 303, 245–257.

Montenegro, C., Maunder, M.N., Zilleruelo, M., 2009. Improving management advice through spatially explicit models and sharing information. Fisheries Research 100, 191–199.

Mosegaard, H., Madsen, K., 1996. Discrimination of mixed Herring stocks in the North Sea using vertebral counts and otolith microstructure. ICES Document CM 1996/H: 17 .

Pedersen, M., Berg, C., Thygesen, U., Nielsen, A., Madsen, H., 2011. Estimation methods for nonlinear state-space models in ecology. Ecological Modelling 222, 1394–1400.

Quinn, T.J., Deriso, R.B., Neal, P., 1990. Migratory catch-age analysis. Canadian Journal of Fisheries and Aquatic Sciences 47, 2315–2327.

R Development Core Team, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0.

Sibert, J.R., Hampton, J., Fournier, D.A., Bills, P.J., 1999. An advection-diffusion-reaction model for the estimation of fish movement parameters from tagging data, with application to skipjack tuna (katsuwonus pelamis). Canadian Journal of Fisheries and Aquatic Sciences 56, 925–938.

Ulrich, C., Post, S., W. Clausen, L., Berg, C., Deurs, M., Mosegaard, H., Payne, M., 2012. Modelling the mixing of herring stocks between the Baltic and the North Sea from otolith data (Extended abstract). International Council for the Exploration of the Sea (ICES).

Online Supplemental Materials for "Joint assessment modelling of two herring stocks subject to mixed fisheries"

Casper W. Berg and Anders Nielsen

Input data

Surveys

This section describes the scientific surveys conducted in the different areas of interest.

North Sea

IBTS Q1 and Q3

The International Bottom Trawl Survey (IBTS) has been carried out in a standardized way since in the first quarter of the year since 1984, although indeces are available back to the early 1970s. From 1991 to 1996 the survey was carried out in all four quarters, and since 1997 in the first and third quarter only, Simmonds (2009).

MLAI

The Multiplicative Larvae Abundance Indeces (MLAI) is based on different larvae surveys conducted during the spawning season. Due to changes in the sampling design this index is based on a model described in Groeger et al. (2001). This index has been used for many years as an indicator of spawning-stock biomass (SSB), Simmonds (2009).

MIK

During the IBTS Q1 sampling for large herring larvae is also carried out, which is used to create the so-called MIK-index (Methot Isaacs-Kidd). It provides an estimate of the number of larvae from the autumn spawning that have survived through the winter, and has been shown to be a good indicator of recruitment, ICES (2009).

Acoustic survey

Since 1989 a standardized acoustic survey has been carried out covering

the North Sea, Skagerrak and Kattegat (IIIa). This dataset can be accessed through the FishFrame database and includes information about the proportion of spring and autumn spawners, ICES (2009) Annex 3. This survey has been found to be the most consistent of the North Sea surveys, Simmonds (2009).

IIIa

MIK
Same as for the North Sea. **HERAS**
The Herring acoustic survey (HERAS) from 26 June to 10 July covers the area in the Skagerrak and the Kattegat as well as an "Norwegian" HERAS area (see figure 1 in Payne et al. (2009)). This index is used in the WBSS assessment after the data has been split into a spring-spawner and autumn-spawner components. In Payne et al. (2009) it was found that this survey is internally consistent for 3-6 ringers, so only these values are used in the official WBSS assessment from 1993 and onwards.
IBTS Q1
Takes place in January/February. This survey is not used in the current WBSS-assessment, because the WBSS population is at the spawning grounds/nursery areas outside IIIa at that time of year, Payne et al. (2009).
IBTS Q3
Takes place in late July-late September. Even though this survey should cover the WBSS population reasonably well, it is not used in the current WBSS assessment. No significant correlations among consecutive age classes for the same cohort were found in Payne et al. (2009).

Western Baltic

A joint German-Danish acoustic survey, (GERAS) is carried out between 2 and 21 October in the Western Baltic covering Subdivisions 21, 22, 23 and 24. All individuals in this survey are assumed to be of WBSS origin. Only ages 1-3 from 1994 and onwards (except 2001) are used in the assessment.
N20
The N20 larval survey is conducted in weekly intervals during the spawning season (March to June) in the Greifswalder Bodden near Rügen, and is available from 1992 and onwards.

BITS in SD2224

This Baltic trawl survey is not used for the current assessment of Western Baltic herring, but since there is an overlap in time and space with the BITS survey in Q1 and the IBTS in IIIa, it is possible to estimate a vessel independent survey index for both regions (i.e. the catchability may be assumed equal for the IIIa and SD2224 sub-areas). This means, that we have information about the relative abundance between these two areas, which greatly enhances the utility of these two data sources.

Natural mortality

The natural mortality is assumed to be known, as it usually not possible to estimate both natural and fishery mortality.

For WBSS, natural mortality was assumed constant over time and equal to 0.3, 0.5, and 0.2 for 0- ringers, 1-ringers, and 2+ -ringers respectively. The estimates of natural mortality were derived as a mean for the years 1977-1995 from the Baltic MSVPA (ICES (2009) p. 166).

For NSAS, the natural mortalities have been held constant from 1957 to date in the ICES Assessments (ICES (2009) p. 521) at 1.0, 1.0, 0.3, 0.2 and 0.1 from ages 0 to 4+ respectively.

It seems natural to let mortalities be specific to areas instead of stocks. But this creates some problems, as mortalities used by the current assessments are quite different, and we therefore need to come up with some common mortalities in the mixing area. One might also argue, that we cannot use the same mortalities for the same ringer-group between stocks, because the two stocks have quite different sizes in the same ringer-group – a 1-ringer NSAS has approximately the same weight as a 2-ringer WBSS.

Mixing samples

Before 1998, herring were identified as SS or AS by vertebral counts. From 1998 and up til today stock identification is done by examinig otolith microstructure, which is less error prone and requires fewer samples Mosegaard and Madsen (1996).

There are two sources of mixing samples: Samples of commercial landings taken at the habours, and samples taken by research vessels. The latter is only available at the times of the acoustic surveys in the summer, whereas commercial samples are taken througout the year.

As a simple start, we will consider the stock proportions constant within a year, although this a very crude assumption. Therefore we will only consider commercial samples of spawner type.

From the proportions used to divide the catch by stock in the usual assessments, which are performed on landings by quarter, we can calculate the landed proportions of autumn spawners p_{AS} by year and age by a simple weighted mean:

$$p_{AS,a,y} = \sum_{q=1}^{4} \frac{L_{a,y,q} p_{AS,a,y,q}}{L_{a,y}} \quad (1)$$

References

Groeger, J., Schnack, D., Rohlf, N., 2001. Optimisation of survey design and calculation procedure for the international herring larvae survey in the north sea. Archive of Fishery and Marine Research 49, 103–116.

ICES, 2009. Report of the herring assessment working group for the area south of 62 n. ICES Document CM 2009/ACOM:03 .

Mosegaard, H., Madsen, K., 1996. Discrimination of mixed Herring stocks in the North Sea using vertebral counts and otolith microstructure. ICES Document CM 1996/H: 17 .

Payne, M.R., Clausen, L.W., Mosegaard, H., 2009. Finding the signal in the noise: objective data-selection criteria improve the assessment of western baltic spring-spawning herring. ICES Journal of Marine Science 66, 1673–1680.

Simmonds, E.J., 2009. Evaluation of the quality of the North Sea herring assessment. ICES J. Mar. Sci. 66, 1814–1822. http://icesjms.oxfordjournals.org/cgi/reprint/66/8/1814.pdf.

Appendix H

Paper VI

Continuous Time Stochastic Modelling of Biomass Dynamics

Casper W. Berg and Henrik Madsen

24 April 2013

Abstract

A continuous time stochastic version of the Schaefer biomass dynamic model is developed, which allows for estimation of both process and observation error even given varying sample times. The model relies on data on catches (removals) and (relative) indices of abundance, and a unique feature of this model is that observation errors as well as missing observations are allowed for both data types. The model is applied to simulated data as well as real data on South Atlantic Albacore and 8 North Sea fish stocks. The results show that it is most often not possible to separate process from observation noise, unless hundreds of years of data are available. The examined stocks all appear to have been more or less over-exploited in the past, leading to biomasses well below that of maximum productivity. However, recent drops in the exploitation rates have led to all North Sea stocks except cod being currently harvested close to F_{MSY}, and hence the stocks are predicted to rebuild given that the current exploitation rates are maintained.

1 Introduction

Biomass models (sometimes called surplus production models, or simply production models) distinguish themselves from other models of populations of fish stocks by characterizing the current state of the entire population by a single number (typically the biomass of the population). This is a quite crude characterization of a population, since it does not specify the age/size composition of the population. Nevertheless, these models are useful in data-poor situations where only the mass of the catches is recorded, and are still frequently used (Punt and Szuwalski (2012)). The models have also formed the basis for the concept of maximum sustainable yield (MSY), which is used as a key aim for fisheries management all over the world (United Nations (1982)).

A minimum of two time-series is needed for each population under examination:

1. The removals from the population, i.e. the amount of commercial catches (C).
2. A biomass index I, which is a relative measure of the total population, or estimates of the absolute biomass B.

These models have typically been formulated in discrete time, and often considering observation-error or process-error only (Polacheck et al. (1993)). The models are of the form

$$B_{y+1} = (B_y + g(B_y) - C_y)\,\mathrm{e}^{\epsilon_y} \tag{1}$$

$$I_y = qB_y\mathrm{e}^{\eta_y} \tag{2}$$

and the most widespread choice of the surplus production function is $g(B) = rB(1-B/K)$ (Schaefer, 1954), as well as the Fox $g(B) = rB(1 - \log B/\log K)$ (Fox, 1970), and Pella-Tomlinson $g(B) = \frac{r}{p}B\left(1-(B/K)^p\right)$ (Pella and Tomlinson, 1969). Here, B is the biomass, C_y is the catch during year y, I_y is a relative biomass index, q is a constant proportionality factor (catchability), r is the intrinsic rate of growth, and K is the carrying capacity (unexploited equilibrium biomass). The process error $\epsilon_y \sim N(0, \sigma_\epsilon^2)$ describes random deviations from the expected development in biomass, which is to be distinguished from observation error, $\eta_y \sim N(0, \sigma_\eta^2)$, on the biomass index I. Observation error (or measurement error) describes the discrepancy between the underlying true state and the observed data, whereas process error (or system error) describes how the system in future states will randomly deviate from process equation, regardless of whether the states are known without error at the present. We will consider only the Schaefer model in the following.

The total catch over a small time step may be further described as the product of the fishing mortality and biomass $\frac{dC_t}{dt} = F_t B_t$. Setting dB_t/dt equal to zero and solving for B gives the equilibrium biomass $K - \frac{KF_t}{r}$, which multiplied by F gives the equilibrium catch (or yield). Observation error on the total catch is usually ignored in this type of model (although Chen and Andrew (1998) is one exception).

Taking the derivative with respect to F in the yield expression, setting it equal to zero and solving for F maximizes the yield and hence gives us the reference point $F_{MSY} = \frac{r}{2}$, and by plugging F_{MSY} into the expression for the equilibrium biomass, $B_{MSY} = \frac{K}{2}$.

Process error estimators (i.e. assuming $\sigma_\eta = 0$) are easy to compute since this is simply a multiple linear regression problem, but these have been shown to be less precise than observation error estimators ($\sigma_\epsilon = 0$) (Polacheck et al., 1993). State-space models are a class of models where both process and observation error may be present, although estimation in these models is much more difficult, especially in non-linear or non-Gaussian cases (Pedersen et al., 2011). In state-space model states (in this context B_t) as well as observations are random variables, and the likelihood is the probability of the observations, for which all possible state trajectories have been integrated out. A compromise between observation-error only and the state-space approach is the total-error method (Ludwig et al., 1988), also known as errors-in-variables. In contrast, the total-error method considers only one specific trajectory of states, and requires prior specification of the ratio of the variances, i.e. $\frac{\sigma_\epsilon}{\sigma_\eta} = w$, where w is some chosen constant (Punt, 2003; de Valpine and Hilborn, 2005). The total error method emerged as an improvement over methods that ignore either process or observation error, although it is known to give asymptotically biased parameter estimates and inaccurate confidence regions (de Valpine and Hilborn, 2005). The state-space approach should therefore be preferred over the others, and advanced statistical software tools such as the BUGS language (Spiegelhalter et al., 2003) and ADMB (Fournier et al., 2012) have made it possible to estimate parameters embedded in state-space models with relative ease (see Pedersen et al. (2011) for an example using both). These high level languages relieves the modeller of having to program his own (extended or unscented) Kalman filter (Reed and Simons, 1996; Punt, 2003) or Gibbs sampler, and thus helps minimize the risk of errors.

Although the state-space model may by identifiable, i.e. given increasingly large amounts of data the parameter estimates will converge to their true values, there may often not be enough data points to separate the two types of error. For example, when Meyer and Millar (1999) analyzed the data set presented in Polacheck et al. (1993) using a Bayesian approach and the winBUGS software, the posterior distribution of the amount of process error was nearly identical to the prior distribution,

indicating very little information in the data. de Valpine and Hilborn (2005) found similar results for the hake data set also presented in Polacheck et al. (1993). Regardless, it is important to attempt to separate the two types of error to obtain reliable point estimates as well as confidence intervals. In fact, when using the common assumption that the noise is multiplicative (Punt, 2003), the reference points F_{MSY} and B_{MSY} depend on the amount of process error (Bousquet et al., 2008). Although based on a model with a product-of-betas process error, Bousquet et al. (2008) note that for small amounts of process error ($\sigma \leq 0.25$) this model is approximately equal to a lognormal process error with the same variance, and give the following corrected reference points:

$$F_{MSY} = \frac{r}{2} - \frac{2(2-r)}{(4-r)^2}\sigma^2 + O(\sigma^3) \tag{3}$$

$$E(B_t) = \frac{K(r-F)}{r}\left(1 - \frac{(r-F)^{-1}\sigma^2}{2-r+F} + O(\sigma^3)\right) \tag{4}$$

This result states that the equilibrium biomass and yield are *decreasing* functions of the process error σ, and hence underlines the importance of separating the two types of error.

In this paper we will formulate the Schaefer model using a stochastic differential equation (SDE). That is, we will shift from a discrete time formulation to using a continuous time formulation. A distinct advantage by using the continuous time formulation is the ability to handle varying sample times. In other words, observed commercial landings or survey CPUEs may be recorded at any time of the year and used as data for the estimation process without having to change the model. Discrete time models can in principle be adjusted to time-varying sample times by changing the size of the time steps appropriately, but this typically involves much more coding and may change the interpretation of some parameters, which is undesirable for purposes of comparison. We perform parameter estimation in our model using `CTSM-R` (Juhl et al., 2013; Kristensen et al., 2004), which makes it possible to reproduce our results using minimal coding effort. `CTSM-R` is still relatively unknown within the field of ecology (although one example is (Møller et al., 2011)). In a time with increasing public and political demands for sustainable harvest of renewable resources, and where concepts such as MSY is being written into legislation everywhere, it is important to have easy-to-use tools for quantifying these reference points. In particular, it is important to correctly quantify the uncertainty on these reference points, such that the precautionary principle can be applied – i.e. in face of great uncertainty the strongest protective measures should be chosen. This work provides such a tool for ecologist while emphasizing the importance of recognizing all potential sources of error.

2 Materials and Methods

We formulate the Schaefer model as an SDE, such that the biomass dynamics as well as the catches occur in continuous time and are both subject to random perturbations. To enable separation of process error and observation error on the catch, we need an assumption of about the development of fishing mortality F over time. For simplicity we choose a random walk process to represent this development, such that the collected process equations in our continuous time stochastic version are:

$$dB_t = B_t\left(r - \frac{r}{K}B_t - F_t\right)dt + \sigma_B B_t dW_t$$
$$d\log(F_t) = \sigma_F dV_t$$

where W_t and V_t are independent standardized Brownian motions.

It is however more convenient to consider the log-transformed SDE, as it ensures a positive population, by introducing

$$Z_t = \log(B_t). \tag{5}$$

Using this we get the following (Lamperti)-transformed SDE,

$$dZ_t = \left(r - \frac{r}{K}e^{Z_t} - F_t - \frac{1}{2}\sigma_B^2\right)dt + \sigma_B dW_t, \tag{6}$$

for which the noise term does not depend on the state, B_t. The term $-\frac{1}{2}\sigma^2$ reflects the result from Bousquet et al. (2008), i.e. that the mean of the stationary distribution of Z_t given a constant F_t depends on σ. This is seen by taking expectation on both sides of the transformed SDE:

$$E\left(\frac{dZ_t}{dt}\right) = E\left(r - \frac{r}{K}e^{Z_t} - F_t - \frac{1}{2}\sigma^2\right) + E\left(\frac{\sigma dW_t}{dt}\right)$$
$$0 = E\left(r - \frac{r}{K}e^{Z_t} - F_t - \frac{1}{2}\sigma^2\right)$$
$$E(B_t) = \frac{K(r - F - \frac{1}{2}\sigma^2)}{r}$$

However, the expression by Bousquet et al. (2008) in (4) is a more precise approximation than the above.

In addition to the process equations, we need to relate the processes to observations through some observation equations. For survey indices the observation equation is identical to the logarithm of (2). The observation equation for the catches (removals) is less straightforward. We could imagine the catch process being described by another SDE:

$$dC_t = \exp\left(\log B_t + \log F_t\right)dt + \sigma_C dW_{C,t}$$

However, we do not observe the accumulated catches from the beginning of the time-series up to time t, i.e. C_t, directly, but rather accumulated catches over shorter time periods, often a year. That is, if we denote the true accumulated catch from time $y-1$ to y by ΔC_y we have $\Delta C_y = \int_{y-1}^{y} F_t B_t dt = C_y - C_{y-1}$.

Since neither B nor F are constant over time in the continuous time formulation, the catch integral over a whole year depends on all the states in that time frame. Although it is possible to project the states in small time steps and accumulate the catches, it is only possible to relate observations to the state vector at one point in time in `CTSM-R`, so instead we choose to approximate the catch integral with

$$\log \Delta C_y \approx \log F_{y-\frac{1}{2}} + \log B_{y-\frac{1}{2}} \tag{7}$$

That is, we evaluate the F and B processes half-way through the year, and approximate the integral of the product of the two series simply by the product evaluated in the middle of the time step. The approximation is illustrated in Figure 1. This approximation is exact when $\frac{dB_t}{dt} = 0$ and $\frac{dF_t}{dt} = 0$, since B and F are constants in that case and can therefore moved outside the integral. This is appealing since MSY calculations are based upon precisely the scenario where these two differentials are zero. The approximation here is of course also applicable to observations of catches that are accumulated over other periods than a year – the midpoint is simply changed and the catch equation is multiplied by dt. We may even have changing periods, e.g. catches by year in the beginning of the time-series and quarterly catches at the end, without changing the model at all thanks to the continuous time formulation.

Using this approximation, the observation equations become

$$\log(B_t) = \log(I_t) + \log(q) + \epsilon_{I,t} \tag{8}$$

$$\log(\Delta C_y) = \log F_{y-\frac{1}{2}} + \log B_{y-\frac{1}{2}} + \epsilon_{C,y} \tag{9}$$

where $\epsilon_{I,t} \sim N(0, \sigma_I)$ and $\epsilon_{C,y} \sim N(0, \sigma_C)$.

Due to the relatively short time-series of fisheries data that are typically available, it is often not possible to estimate both process and observation error. When this is the case, the likelihood function will be flat in a banana shaped region containing equally likely combinations of values for σ_B and σ_I, ranging from observation error only to process error only (de Valpine and Hilborn, 2005). Rather than proceeding with parameter estimates on the boundary of their feasible space (e.g. nearly zero variances), it is preferable to fix some parameters. In analogy with the total-error method we will therefore fix the ratio of the variances:

$$\sigma_I = \alpha\sigma_B \tag{10}$$

$$\sigma_C = \beta\sigma_F \tag{11}$$

where α, β are fixed positive constants.

2.1 Uncertainties on F_{MSY} and B_{MSY}

When dealing with estimated quantities it is essential to provide an assessment of the uncertainty of the estimates, and these will also facilitate any subsequent statistical tests. The variance-covariance matrix of the parameters is approximated in `CTSM-R` as in most other applications, that is by evaluating the Hessian of the likelihood function at the maximum (Kristensen et al., 2004). The normal approximation is more appropriate for parameters that are not strictly positive such as r, K, and variances, so these should be log-transformed prior to estimation (see source code in the Appendix). The variance-covariance matrix for $\log r$, $\log K$ and $\log \sigma$ can then be used in combination with the Delta method (Oehlert, 1992) to obtain a confidence ellipsis for $(\log B_{MSY}, \log F_{MSY})$, and hence also the "confidence banana" for (B_{MSY}, F_{MSY}).

2.2 Model validation

One-step ahead residuals were tested for autocorrelation at lag 1 using the Ljung-Box test (Ljung and Box, 1978).

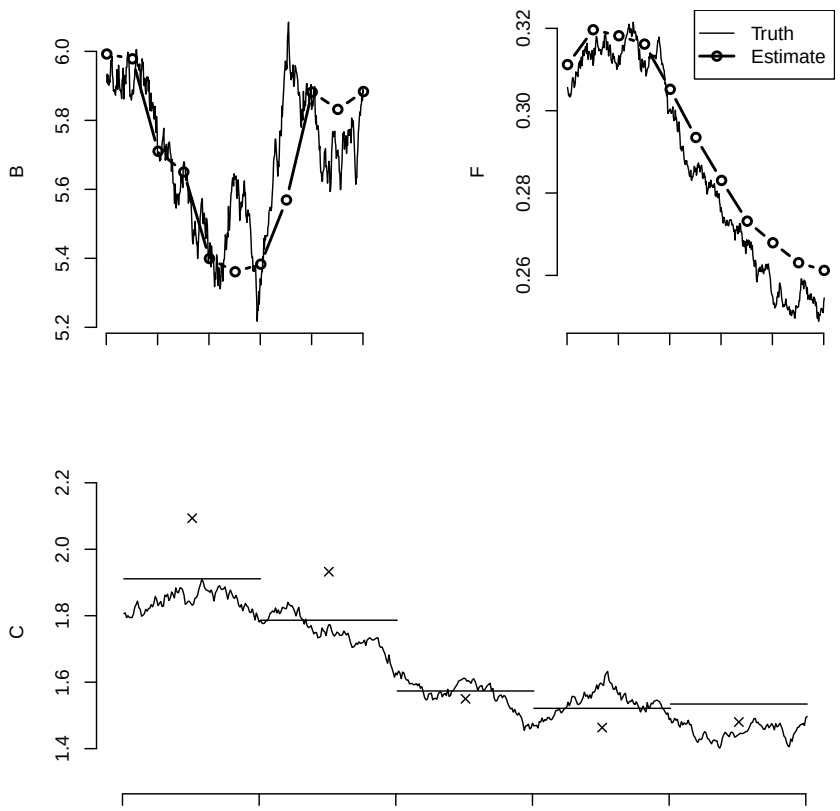


Figure 1: Excerpt of five years from simulated trajectories of the continuous time processes B and F (top panel) as well as the instantaneous catch rate C (bottom row). The top panels are overlaid with the estimated trajectories (thick lines). An approximation to the integral of the instantaneous catch over a year is given by $F \cdot B$ in the middle of a year, which is illustrated by the vertical lines in the bottom panel. The observed catches are shown as crosses.

2.3 Applications

We apply the proposed SDE formulation of the Schaefer model to *i*) The South Atlantic albacore data set presented and analyzed in Polacheck et al. (1993), and Meyer and Millar (1999), *ii*)

simulated data, and *iii*) Data from 8 stocks in the North Sea. In the last application biomass indices are replaced with estimates of total biomass from the SMS model (ICES, 2011), which is a stochastic age-based multi-species model (Lewy and Vinther, 2004). The simulations (*ii*) are carried out using an Euler scheme and the parameters estimated in (*i*) to ensure realistic values. Estimated quantities are compared with their true values using absolute relative errors as in Punt (2003); Punt and Szuwalski (2012). The following number of data points are tested 30, 60, 120, 240, and 480, using 100 simulations for each value.

3 Results

3.1 Case study 1: South Atlantic Albacore

Polacheck et al. (1993) analyzed this data set using using estimators that assume observation error or process error only, while Meyer and Millar (1999) used a Bayesian discrete time state-space model, where informative priors were specified for all parameters except q, which facilitated separation of the two types of error. A preliminary analysis using the methodology presented in this paper revealed that it is not possible to separate process from observation error (parameter estimates were on their bounds with very large associated uncertainties). Hence, we fix the process to observation error ratio by setting $\alpha = 4$ (similar to the results obtained in Meyer and Millar (1999)) and $\beta = 1$. To test the sensitivity of these assumptions we repeat the analysis using other values for (α, β). Table 1 compares our results with those from Polacheck et al. (1993) and Meyer and Millar (1999). Our estimates lie somewhere between the observation- and process error only estimators. The results are further illustrated in Figure 2. We note that the stock was close to the carrying capacity at the beginning of the time-series, which explains the declining biomass even though the fishing mortality was below F_{MSY} most of the time. In the last years the fishing mortality approached the upper red line $F = r - 0.5\sigma^2$, which marks the maximum F that will not lead to eventual extinction of the species.

The lower panel indicates, that the relative quantities B_{1990}/B_{MSY} and F_{1990}/F_{MSY} are not very sensitive to the choice of α, β, while the rather large confidence regions are somewhat affected. Due to the low process error ($\sigma_B = 0.054$) the differences between the deterministic reference points and (3) and (4) are negligible. No significant autocorrelation were found in the residuals.

3.2 Simulation study

The simulation study showed that the model is identifiable, i.e. that the parameter estimates converge to their true values given an increasing amount of data points. Figure 3 illustrates this for the two key quantities F_{MSY} and B_{MSY}, although it is also apparent, that quite long time-series of observations are needed in order to obtain relatively precise estimates. Similar results were obtained for the remaining parameters. Table 2 suggests, that several hundred years (or time-steps) of data are often needed for all parameters to be estimable. When estimation failed, one or more of the variance parameters could not be identified, which means that process- and observation error could not be separated with any reasonable confidence.

	Observation error model	Process error model	Bayesian State-space model	CTSM-R
K	239.60	153.40	279.80	200.65
r	0.33	0.62	0.29	0.40
$q \cdot 10^4$	26.71	43.72	23.89	30.04
B_{1990}	75.51	50.04	83.97	63.47
P_{1990}	0.32	0.33	0.30	0.32
MSP	19.65	23.78	19.26	20.01
E_{MSP}	61.40	70.90	60.86	66.42

Table 1: Comparison of parameter estimates and key management quantities: The first three columns are taken from Meyer and Millar (1999) (Table 2), whereas the last column is our CTSM-R solution (with $\alpha = 4, \beta = 1$. B1990 is the estimated biomass in 1990, P1990 is the same number divided by K, and $MSP = rK/4$ and $E_{MSP} = r/2q$.

No. data points	30	60	120	240	480
Estimable proportion	0.22	0.40	0.56	0.89	1.00

Table 2: Proportion of successful estimations from the simulation study. An estimation was deemed successful when no parameter bound had any significant effect on the final parameter estimates.

3.3 Case Study 2: North Sea

The results from the North Sea stocks are illustrated in Figures 4 and 5. Unlike the results from the South Atlantic Albacore data set, all stocks appear to have been depleted well below the carrying capacity (and hence also B_{MSY} - the green line) from the beginning of the time series, and most stocks appear to have been overfished in the past. The carrying capacity could not be reliably estimated for many of the stocks, indicated by the very wide confidence ellipses in the B_{MSY} direction (x-axis, column 4). Despite the large uncertainties on (B_{MSY}, F_{MSY}), the current state of the stocks are still outside the confidence ellipses. That is, although the fishing mortalities have dropped to levels around F_{MSY} for all stocks except cod and sprat in the recent years, all stocks are still depleted well below B_{MSY}. In other words, the North Sea stocks have been seriously overfished in the past leading to biomasses well below the point of maximum yield, but recent drops in the exploitation rates have led to sustainable levels of fishing mortality for most stocks, and hence the stocks are expected to rebuild given the current fishing pressure.

The results are somewhat sensitive to whether the catchability q can be assumed known or not (figures 4 and 5, column 4). Particularly the sole stock have an estimate of q that is substantially different from 1, which is the expected value since we are using absolute biomass estimates from the SMS model rather than relative indices. The assumption that $q = 1$ relies on another set of assumptions about natural mortality in the SMS assessment model used to produce the total biomass input for our model, so this discrepancy could be caused by wrong assumptions about natural mortality in the SMS model. We should also note, that except for haddock, the residuals (either from catches, biomass indices, or both) failed the test for independence for all of the North Sea stocks. This is somewhat disturbing, as this means that there is autocorrelation present in

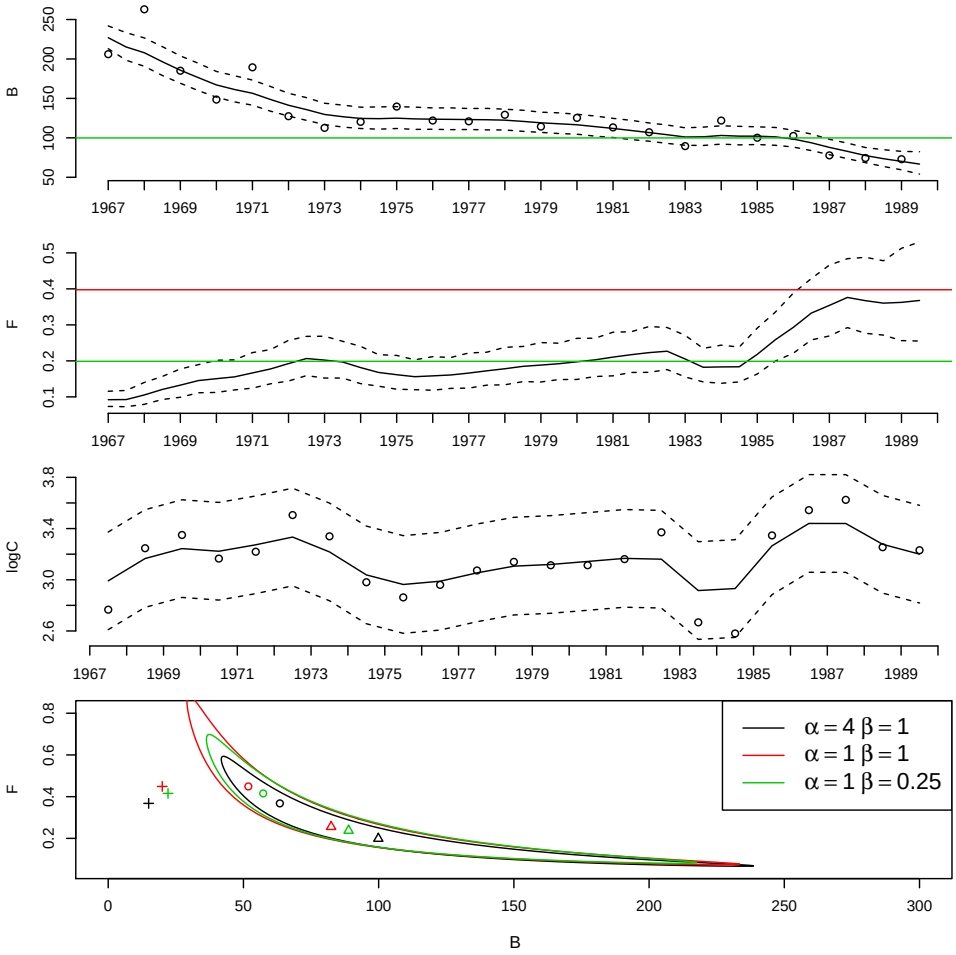


Figure 2: Estimated/observed biomass (B, top row, green line indicates B_{MSY}), Fishery mortality (F, 2. row, green line indicates F_{MSY}, red line is $r - 0.5\sigma^2$), log catches ($\log C$, row 3), and row 4 contains the estimates of (B_{MSY}, F_{MSY}) (triangle) with the 95% confidence ellipsis, (B_{1990}, F_{1990}) (circle), and (B_{eq}, F_{1990}) (plus symbol). α is the assumed ratio between process and observation error, i.e. $\alpha = 4$ corresponds to observation error being the predominant error source.

the residuals and hence the models cannot be formally validated. It is however not too surprising that some autocorrelation is present considering the simple model structure, which completely ignores the age composition in the population. Different age compositions will lead to different rates of growth/reproduction and may hence explain why autocorrelation is present. Another

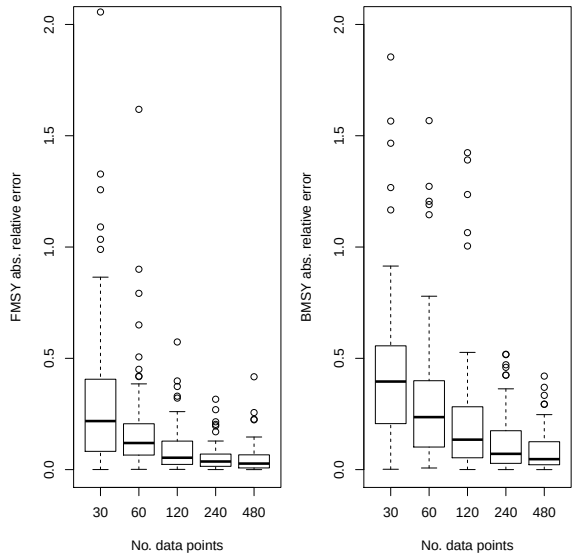


Figure 3: Box plots of the absolute relative errors on F_{MSY} and B_{MSY} as a function of the number of data points available for estimation.

possible source of autocorrelation could be missing covariates such as environmental forcings and multispecies effects which were not considered in this model, although the latter were actually included in the SMS model, which produced the biomass input data. Finally, the bias resulting from the approximation in (9) will be substantial when sudden changes in biomass or fishing mortality are present. Such changes are apparent for e.g. North Sea herring, so these results should be interpreted with caution.

	q	Cod	Had	Her	Pla	Sai	Sol	Spr	Whi
σ_B	1	0.09	0.27	0.05	0.10	0.10	0.07	0.19	0.16
σ_F	1	0.13	0.11	0.43	0.08	0.11	0.15	0.26	0.11
σ_F	$\hat{q}$	0.12	0.10	0.31	0.08	0.11	0.17	0.25	0.11
σ_B	$\hat{q}$	0.09	0.27	0.10	0.10	0.10	0.04	0.19	0.14

Table 3: Estimated process standard deviations.

4 Discussion

We have presented a continuous time stochastic formulation of the Schaefer biomass dynamic model. Using an approximation to the catch equation we have shown how to perform parameter estimation

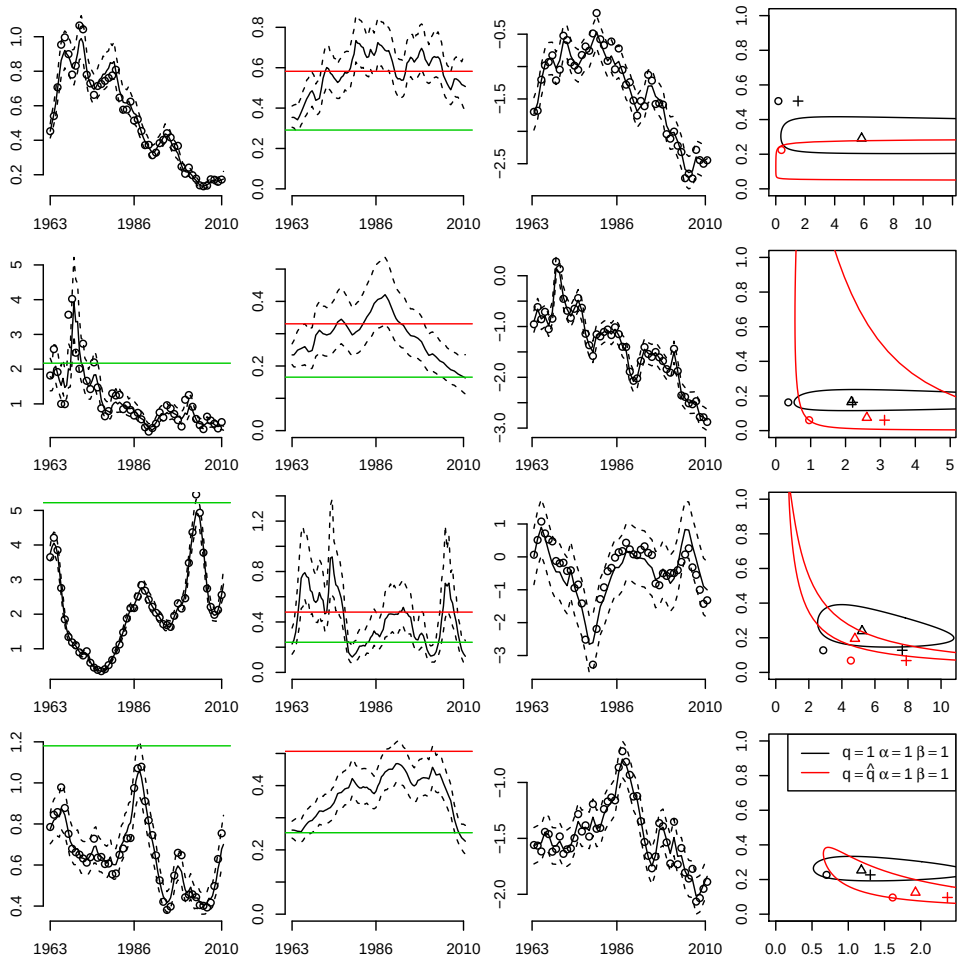


Figure 4: Estimated/observed biomass (B, column 1), Fishery mortality (F, column 2), log catches ($\log C$, column 3), and column 4 contains the estimate of (B_{MSY}, F_{MSY}) with the 95% confidence ellipsis (triangle), (B_{2011}, F_{2011}) (circle), and (B_{eq}, F_{2011}) (plus symbol). Rows 1 through 4 represent Cod, Haddock, Herring and Plaice.

in this model using the `CTSM-R` software package, which is a powerful tool for maximum likelihood estimation in continuous time stochastic state space models. The model was validated through a simulation study. Varying sample times and missing observations are naturally handled within this framework unlike for the most typically used estimation techniques. The model can be represented in a few lines of code using `CTSM-R` (see Appendix), which makes it easy for others to reproduce our

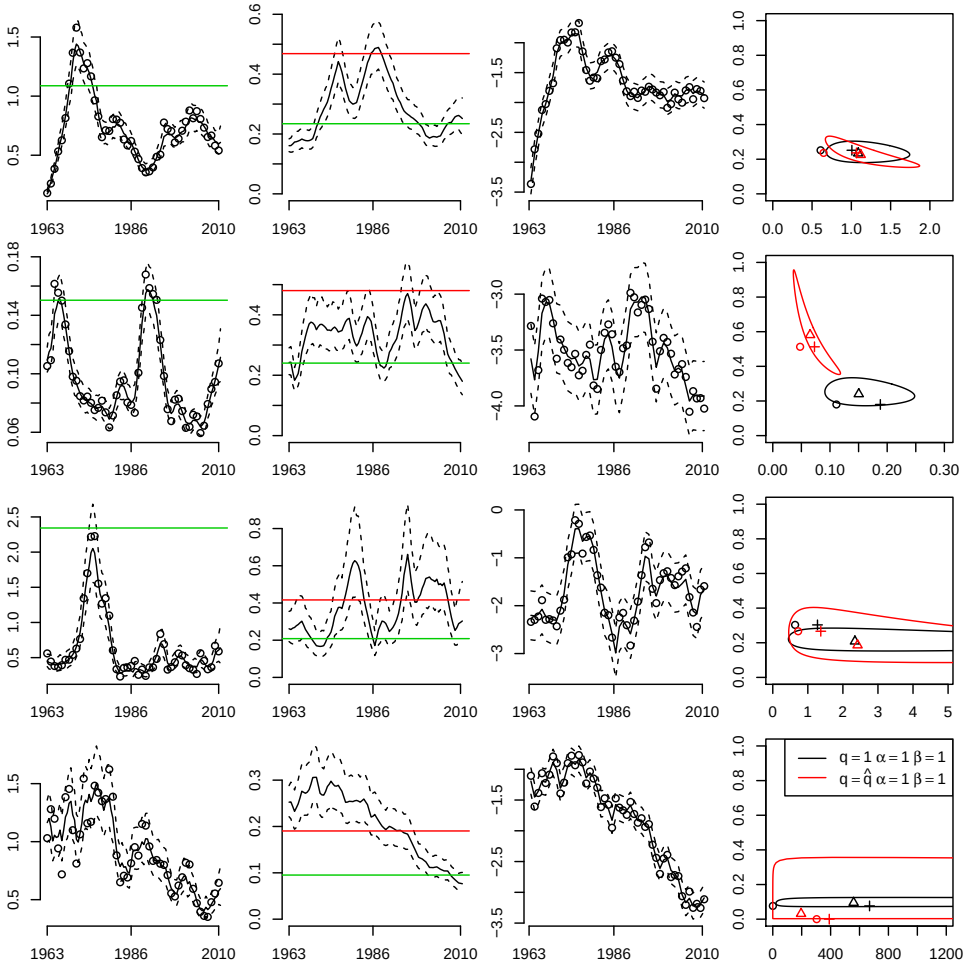


Figure 5: Same as figure 4 but for Saithe, Sole, Sprat and Whiting (rows 1 through 4).

results or adapt the code to other data sets. Another distinct feature of this model is the ability to handle observation error on catches as well as biomass indices, as errors on the catches (removals) are usually ignored. Since fishing mortality is included as a state variable, our model also allows for missing observations of biomass as well as of catches. The latter is unique for this model, and not possible in models where the catch is assumed to be known without error.

The fishing mortality was assumed to follow a random walk, which suffices for the situation where catches are aggregated over a year. If catches are available on a finer time-scale and there is a

seasonal pattern in exploitation, the random walk assumption could be replaced with some periodic function or functions of covariates such as commercial effort. If pulse harvest events are present semi-discrete models should be considered (Colvin et al., 2012).

While the simulation studies showed that the model is identifiable, separation of process and observation error was difficult and often required several hundred data points. This is in agreement with the results found in de Valpine and Hilborn (2005). In this case, rather than just assuming no process error, we must recommend a sensitivity analysis and application of the precautionary principle, since using the wrong estimator may result in serious bias (Chen and Andrew, 1998). However, we found that all our estimates were lying between previously found estimates assuming either no process or no observation error.

The estimates of the key management reference points B_{MSY} and F_{MSY} were found to be quite uncertain for many scenarios, which is in line with the results found in Punt (2003) and Punt and Szuwalski (2012). Nevertheless, for data-poor stocks where age- and size-composition data are not available, biomass production models remain useful tools for fisheries management. Given the many possible choices of estimator, shape of production function, and assumptions about variance parameters, it is useful to have different well tested off-the-shelf tools to test the validity/sensitivity of all assumptions made, and to quantify the uncertainties associated with any quantity used for management. By formulating the Schaefer model using a set of SDEs and using state of the art software for parameter estimation (`CTSM-R`), we have provided a fast and flexible framework for estimation in biomass dynamic models.

References

Bousquet, N., Duchesne, T., Rivest, L.P., 2008. Redefining the maximum sustainable yield for the Schaefer population model including multiplicative environmental noise. Journal of Theoretical Biology 254, 65–75.

Chen, Y., Andrew, N., 1998. Parameter estimation in modelling the dynamics of fish stock biomass: Are currently used observation-error estimators reliable? Canadian Journal of Fisheries and Aquatic Sciences 55, 749–760.

Colvin, M.E., Pierce, C.L., Stewart, T.W., 2012. Semidiscrete biomass dynamic modeling: an improved approach for assessing fish stock responses to pulsed harvest events. Canadian Journal of Fisheries and Aquatic Sciences 69, 201710–1721.

Fournier, D.A., Skaug, H.J., Ancheta, J., Sibert, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., 2012. AD Model Builder: Using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods and Software 27, 233–249.

Fox, W.W., 1970. An exponential surplus-yield model for optimizing exploited fish populations. Transactions of the American Fisheries Society 99, 80–88.

ICES, 2011. Report of the Working Group on Multispecies Assessment Methods (WGSAM). ICES Document CM 2011/SSGSUE:10 .

Juhl, R., Kristensen, N., Bacher, P., Kloppenborg, J., Madsen, H., 2013. CTSM-R - Continuous Time Stochastic Modelling for R. http://www.ctsm.info.

Kristensen, N.R., Madsen, H., Jørgensen, S.B., 2004. Parameter estimation in stochastic grey-box models. Automatica 40, 225–237.

Lewy, P., Vinther, M., 2004. A stochastic age–length-structured multispecies model applied to North Sea stocks. ICES CM 2004/FF:20 .

Ljung, G.M., Box, G.E.P., 1978. On a Measure of Lack of Fit in Time Series Models. Biometrika 65, 297–303.

Ludwig, D., Walters, C., Cooke, J., 1988. Comparison of two models and two estimation methods for catch and effort data. Natural Resource Modeling 2, 457–498.

Meyer, R., Millar, R.B., 1999. BUGS in Bayesian stock assessments. Canadian Journal of Fisheries and Aquatic Sciences 56, 1078–1086.

Møller, J.K., Madsen, H., Carstensen, J., 2011. Parameter estimation in a simple stochastic differential equation for phytoplankton modelling. Ecological Modelling 222, 1793–1799.

Oehlert, G.W., 1992. A Note on the Delta Method. American Statistician 46, 27–29.

Pedersen, M., Berg, C., Thygesen, U., Nielsen, A., Madsen, H., 2011. Estimation methods for nonlinear state-space models in ecology. Ecological Modelling 222, 1394–1400.

Pella, J.J., Tomlinson, P.K., 1969. A generalized stock production model. Inter-American Tropical Tuna Commission.

Polacheck, T., Hilborn, R., Punt, A.E., 1993. Fitting surplus production models: Comparing methods and measuring uncertainty. Canadian Journal of Fisheries and Aquatic Sciences 50, 2597–2607.

Punt, A.E., 2003. Extending production models to include process error in the population dynamics. Canadian Journal of Fisheries and Aquatic Sciences 60, 1217–1228.

Punt, A.E., Szuwalski, C., 2012. How well can FMSY and BMSY be estimated using empirical measures of surplus production? Fisheries Research 134-136, 113–124.

Reed, W.J., Simons, C.M., 1996. Analyzing catch-effort data by means of the Kalman filter. Canadian Journal of Fisheries and Aquatic Sciences 53, 2157–2166.

Schaefer, M.B., 1954. Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. Inter-American Tropical Tuna Commission.

Spiegelhalter, D., Thomas, A., Best, N., Lunn, D., 2003. WinBUGS user manual. Version 1.4. MRC Biostatistics Unit, Cambridge, UK .

United Nations, 1982. United Nations Convention on the Law of the Sea. http://www.un.org/Depts/los/convention_agreements/texts/unclos/unclos_e.pdf.

de Valpine, P., Hilborn, R., 2005. State-space likelihoods for nonlinear fisheries time-series. Canadian Journal of Fisheries and Aquatic Sciences 62, 1937–1952.

5 Appendix

```
library(ctsmr)

## read in data
data=scan("pola.dat",comment.char="#")

N=data[1];
catch=data[2:(2+N-1)]
cpue=data[(2+N):(2+2*N-1)]


## initial states
B0 = exp(5.5)
F0 = exp(-2.75)

## initial parameters
logK=5.4
logr=-1
sig11=0.05 ## B proc. std. dev.
sig22=0.01 ## F proc. std. dev.


model <- ctsm$new()


model$addSystem( dB1 ~ (exp(logr) - exp(logr-logK+B1) - exp(F1) - 0.5*exp(2*sig11))*dt + exp(sig11)*dw1 )
model$addSystem( dF1 ~ exp(sig22)*dw2 )

model$addObs(b1 ~ fakein*B1 + logq)
model$addObs(c1 ~ B1+F1)

## ctsmr requires at least one input, so make a fake one consisting of a vector of ones.
model$addInput(fakein)

model$setVariance(b1b1 ~ alpha*exp(2.0*sig11))
model$setVariance(c1c1 ~ beta*exp(2.0*sig22))

## set initial parameters and bounds (init,lower,upper)
model$setParameter(B1=c(log(B0),log(B0/1000),log(B0*1000)),
                   F1=c(log(F0),log(F0/1000),log(F0*1000)),
                   logr=c(logr,logr-log(100),logr+log(100)),
                   logK=c(logK,logK-log(1000),logK+log(1000)),
                   sig11=c(log(sig11),log(sig11/1000),log(sig11*1000)),
                   sig22=c(log(sig22),log(sig22/1000),log(sig22*1000)),
                   alpha=c(1),
                   beta=c(1),
                   logq=c(0,-10,10)
                   )


model$setOptions(con=list(solutionMethod="adams",nIEKF=1))

## rearrange data for catch approximation
tt=seq(1,N+0.5,length.out=2*N)
nn=length(tt)
obspadC = rep(NA,nn);
obspadC[seq(2,nn,by=2)]=log(catch)
obspadB = rep(NA,nn);
obspadB[seq(1,nn,by=2)]=log(cpue);
obs2=cbind(tt,obspadB,obspadC)

colnames(obs2)<-c("t","b1","c1")
```

```
data1 <- as.data.frame(cbind(obs2, fakein=1))
data1=rbind(data1,c(24,NA,NA,1))


## estimate
res <- model$estimate(data1, threads=2)
print(summary(res,extended=T))
smooth=smooth.ctsmr(res)
```