Review of State-Space Models for Fisheries Science

3 authors, including:

William H. Aeberhard
Stevens Institute of Technology
9 PUBLICATIONS   37 CITATIONS

SEE PROFILE

Anders Nielsen
Technical University of Denmark
62 PUBLICATIONS   2,907 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Saddlepoint Tests for Accurate and Robust Inference on Overdispersed Count Data View project

Macroecological patterns of macrofaunal diversity in sandy beaches View project

Review of State-Space Models for Fisheries Science

William H. Aeberhard,[1] Joanna Mills Flemming,[1] and Anders Nielsen[2]

[1]Department of Mathematics and Statistics, Dalhousie University, Canada

[2]National Institute for Aquatic Resources, Technical University of Denmark, Denmark

Abstract

Fisheries science is concerned with the management and understanding of the raising and harvesting of fish. Fish stocks are assessed using biological and fisheries data with the goal of estimating either their total population or biomass. Stock assessment models also make it possible to predict how stocks will respond to varying levels of fishing pressure in the future. Such tools are essential with overfishing now reducing stocks and employment worldwide, with in turn many serious social, economic and environmental implications. Increasingly, a state-space framework is being used in place of deterministic and standard parametric stock assessment models. These efforts have not only had considerable impact on fisheries management but have advanced the supporting statistical theory and inference tools as well as the required software. An application of such techniques to the North Sea cod stock highlights what should be considered best practices for science-based fisheries management.

1 Introduction

The Food and Agriculture Organization of the United Nations (FAO) describes a fishery as a unit determined by an authority (or other entity) that is engaged in raising and/or harvesting fish. These fish are often grouped into stocks as defined by the species or type of fish, geographic area, method of fishing and people involved. Fisheries science is concerned with the management and understanding of fisheries and is multidisciplinary: biology, conservation, ecology, economics, limnology, management, oceanography and statistics all play important roles.

Fisheries scientists collect biological and fisheries data in order to perform stock assessments (typically annually) that provide fisheries managers with information required to regulate fish stocks. Biological data describing the stock include the age structure, age at first spawning, fecundity, ratio of males to females, natural mortality (M), fish growth rate, spawning behavior, critical habitats, migratory habits and food preferences. Fisheries data often include the type of fishery (e.g. commercial versus recreational), gear type (longline, rod and reel, nets, etc.), pounds of fish caught, fishing effort and the time and geographic location of the best catches. **Stock assessment models** combine both biological and fisheries data to estimate either the total population or total biomass of the stock. These models

make it possible to assess the current status and condition of the stock as well as predict how stocks will respond to varying levels of fishing pressure in the future. The rationale is that if one can reliably estimate the biomass of the stock and understand its biology then one can estimate how many fish can be safely removed from the stock in order to ensure a sustainable resource.

Overfishing, including the taking of fish beyond sustainable levels, is a global problem that is reducing fish stocks and employment, with many serious social, economic and environmental implications. The goal of fisheries management is to reduce the level of overfishing that occurs and restore stocks that have been overfished. Costello et al. (2016) report that the general state of global fish stocks is poor and declining. Of 4,714 fisheries assessed in the year 2012, only 32% remained at or above the biomass target that supports maximum sustainable yield, the rest have slipped below that critical threshold. Meanwhile, among the major stocks the FAO estimates that 29% currently are overexploited or depleted (Food and Agriculture Organization of the United Nations, 2014). A 2014 report by the Prince of Wales's International Sustainability Unit and the New York-based Environmental Defense Fund (Holmes et al., 2014) estimated that while global fisheries were adding $270 billion a year to global gross domestic product, the full implementation of sustainable fishing could increase this figure by an estimated $50 billion.

Stock assessment models are required for prediction purposes, to integrate a variety of data sources, and to provide the best advice possible. Stationary, single-stock models have long been viewed as overly simplistic (Larkin, 1977) and there has been increasing emphasis on ecosystem-based management (e.g. Link et al., 2011). Amongst other factors, multi-species and ecosystem interactions, as well as temporal and spatial variation, are often found to be important. Assumptions of stationarity are highly questionable in light of impacts of greenhouse gas emissions and other anthropogenic forcing factors. Furthermore, all models are only approximations of reality and fisheries data are often noisy such that one should consider stochasticity (i.e. an error term) not only when describing the dynamics of a fish stock or ecosystem but also in linking such dynamics to actual observed data. Doing so can provide more realistic predictions of the impacts of future management actions. In fact, models that allow for both process error in the population dynamics and observation error in the data used to estimate model parameters are increasingly used in fisheries science. Such models are generally referred to as **state-space models** (SSMs), since they combine stochastic assumptions about both observed quantities and unobserved states driving a dynamic system under study (such as a fish stock), and are considered by many to be the best practice for fish stock assessment (e.g. Aanes et al., 2007; Gudmundsson and Gunnlaugsson, 2012) and more generally for ecological modelling (e.g. Buckland et al., 2007; Minto et al., 2013; Pedersen et al., 2011).

In Section 2 we review how fisheries scientists have so far developed tools for stock assessment in order to fully utilize the SSM framework. These efforts have had considerable impact not only on fisheries management but also on general SSM theory and inference tools as is evidenced in Section 3. In Section 4, particular attention is paid to fisheries science requirements, including the prediction of nonlinear functions of fixed and random effects as well as supporting software. In Section 5 we provide a real data example highlighting what we consider to be best practices for science-based fisheries management. We conclude with a discussion of the current challenges facing fisheries science globally.

2 Evolution of Stock Assessment Models

Basic fisheries research began in North America and Northern Europe in the 1850s motivated by beliefs that fisheries resources had significantly declined (Royce, 1988). At the time fisheries were vital for providing food and employment much like they still are today in many developing countries. The United States Commission on Fish and Fisheries was created in 1871. The International Council for the Exploration of the Sea (ICES; www.ices.dk) was formed in Europe and in 1901 it endorsed scientific inquiry. The Biological Board of Canada was established in 1912.

Fedor Baranov is considered by many to have laid the foundation for quantitative fisheries science with his Baranov catch equation (Baranov, 1918):

$$C = \frac{F}{F + M}(1 - e^{-(F+M)T})N_0. \tag{1}$$

The catch equation describes the situation where an initial cohort of N_0 fish is subject to two constant mortality rates: fishing mortality F and natural mortality M (all causes except fishing). The surviving cohort size will develop over time according to $\frac{dN}{dt} = -(F + M)N$, which means that at time T the surviving size of the population is $e^{-(F+M)T}N_0$ and the dead population size is $(1 - e^{-(F+M)T})N_0$. The resulting catch C is the fraction $\frac{F}{F+M}$ of the dead population, which died from fishing mortality.

Derzhavin (1922) applied the Baranov catch equation (1) to catch-at-age data in order to assess population size while assuming negligible mortality from sources other than fishing. A similar method was used in Fry (1949), where a "virtual population" was introduced to represent the initial population size (assumed to be at least as big as what was later caught). Notice that this concept of a virtual population size is reminiscent of state-space terminology, as it refers to an unobserved random variable.

Beverton (1954) and Beverton and Holt (1957) partitioned the fishing mortality into (yearly) fishing effort and fishing selectivity for the different age groups. The latter is referred to as the separability assumption (Doubleday, 1976), where the age- and year-specific fishing mortality $F_{a,t}$ is expressed as the products of a yearly component f_t and an age specific component S_a, as $F_{a,t} = f_t S_a$.

In Murphy (1965) and Gulland (1965) an iterative approach was used to solve the non-linear equations describing the fish population dynamics. At the time the iterative solver required considerable manual effort, so the methods became more tractable when Pope (1972) presented a simple analytic approximation for solving systems of non-linear equations.

The early algorithms were deterministic, as the catches (at age) were simply subtracted as known values, step-by-step within each cohort. This yielded flexible estimates that matched the data. However these deterministic methods did not provide any measure of uncertainty, or a consistent way of predicting stock sizes in future years (known as forecasts).

A standard parametric statistical model (without random effects) must be formulated with fewer model parameters than observations in order to be identifiable. The deterministic iterative procedures had one $F_{a,t}$ and one $N_{a,t}$ "parameter" for each catch observation. To cast the fish stock assessment problem into a standard parametric statistical framework it was necessary to reduce the number of parameters. Incorporating a separability assumption led to many different parametric statistical assessment models (e.g. Fournier and Archibald, 1982; Lewy, 1988; Paloheimo, 1980; Pope and Shepherd, 1982). These models used deterministic population dynamics equations with all stochasticity attributed to observation error.

The separability assumption was easy to work with and reduced estimation to a standard minimization problem with a clear convergence criterion. It also made the standard toolbox for statistical inference and model validation available. The down side was that the

assumption was not always valid. The fishing gear could evolve as could regulative measures (e.g. mesh size) and spatial fish availability. These in turn could change the relative selection pattern of the fishery. Within the parametric statistical model framework this was dealt with by dividing the time period into time blocks of constant separability (e.g. Methot and Wetzel, 2013), using splines to obtain a more flexible selection (e.g. Aarts and Poos, 2009; Butterworth, Ianelli, and Hilborn, 2003; Jardim et al., 2015) while keeping the number of parameters low, or allowing penalized deviances with fixed penalties (e.g. Methot and Wetzel, 2013).

The penalized deviance approach assumes a parametric structure for fishing mortalities (e.g. separability), but extends the structure by allowing the fishing mortalities to deviate according to extra parameters (e.g. $F_{a,t} = f_t S_a e^{\delta_{a,t}}$). The deviance parameters $\delta_{a,t}$ are then penalized (e.g. by assuming that $\delta_{a,t} \sim \mathcal{N}(0, \sigma_\delta^2)$). The key is that the penalty parameter (here σ_δ^2) is fixed and entirely subjective. Similar subjectivity is applied to the degree of smoothness in the spline approaches.

SSMs for stock assessment were introduced by Sullivan (1992) and Gudmundsson (1994). **State-space assessment models** (SSAMs) are based on the same basic equations as earlier stock assessment models, but incorporate the idea of dynamic unobserved states. Their main advantage is that they avoid subjectivity with respect to time-varying selectivity (e.g. choice of time blocking, choice of spline smoothing degree, or choice of deviance penalties). SSAMs further partition the stochasticity into process error and observation error, which is particular useful for fishery management, as an important part of the managing process is to predict the stock size for future years.

Recently, SSAMs have received renewed attention (e.g. Brinch, Eikeset, and Stenseth, 2011; Cadigan, 2015; Nielsen and Berg, 2014), because new software (Fournier et al., 2012; Kristensen et al., 2016) has made it possible to use the Laplace approximation to implement and estimate these models efficiently. In fact these recent implementations have finally made SSAMs fully operational and they are now routinely used to manage important fish stocks around the world.

3 State-Space Framework

The origins of the SSM and related hidden Markov model can be traced back to the works of Kalman (1960) and Kalman and Bucy (1961). Motivated by engineering applications, the focus was mainly on "online" computations where live stream data required efficient implementation of recursive predictions and updates. In this section, we formally define SSMs in a general setting, but deliberately restrict ourselves to "offline" estimation and prediction as most applications in fisheries science posit the full data to be available at once. We defer to Künsch and Fearnhead (this issue) for a detailed discussion about the online methods of filtering and smoothing.

3.1 Observation and process equations

A SSM is defined by two stochastic processes $\boldsymbol{X}_t$ and $\boldsymbol{Y}_t$, with the index $t = 1, 2, \ldots$ denoting distinct time steps. The q-dimensional $\boldsymbol{X}_t$ process is often referred to as the unobserved state sequence since it represents dynamic features of the system under study (such as a fish stock) which are not directly observed. Closely related concepts include latent variables (e.g. Jöreskog, 1969) and random effects (e.g. McCulloch and Searle, 2000). $\boldsymbol{X}_t$ takes values in a state space X which we will consider uncountable, i.e. X $\subseteq \mathbb{R}^q$, since in fisheries applications $\boldsymbol{X}_t$ typically involves the fish stock biomass, its abundance and the fishing mortality rate

F. Note that by considering F as unobserved (random) states, its original deterministic connection to the observed catch C in Equation (1) does not hold anymore; F now realistically relates to the unobserved probability of a fish getting caught, whereas C represents the actual observed catch with potential error and misreporting. We will not cover the "discrete" case where X is countable, commonly referred to as a hidden Markov model and generally studied separately (see MacDonald and Zucchini, 1997). The r-dimensional $\boldsymbol{Y}_t$ process is directly observed and plays the role of a (vector) response variable. We will assume that the observation space $\mathsf{Y} \subseteq \mathbb{R}^r$ since in fisheries applications $\boldsymbol{Y}_t$ often includes the reported total commercial catch and survey indices corresponding to a standardized fishing effort. In addition to $\boldsymbol{Y}_t$ and $\boldsymbol{X}_t$, other variables may be observed; they usually play the role of fixed covariates and thus will be omitted in our notation.

Let the subscript $i : j$ denote the inclusive range of indices between i and j, so that $\boldsymbol{X}_{1:t} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_t)$, with the convention that $\boldsymbol{X}_{t:t} = \boldsymbol{X}_t$. The SSM relies on two main structural assumptions. The first assumption is that $\boldsymbol{X}_t$ is a Markov process of a given finite order $k > 0$. This means that the conditional distribution of $\boldsymbol{X}_t | \boldsymbol{X}_{(t-1):1}$ equals that of $\boldsymbol{X}_t | \boldsymbol{X}_{(t-1):(t-k)}$. Hereafter, we set $k = 1$ for simplicity of notation. The second main assumption is that the observations are conditionally independent: $\boldsymbol{Y}_t | \boldsymbol{X}_{1:\infty}$ is independent of $\boldsymbol{Y}_s | \boldsymbol{X}_{1:\infty}$, for $t \neq s$. This implies that any (marginal) dependence between the responses is entirely due to the unobserved states.

The SSM is a hierarchical model, and it can thus be conveniently represented by a set of equations, each representing a different level of hierarchy. The top level is represented by an **observation equation** (also known as a **measurement equation**) that links the response to the unobserved states. The second level is represented by a **process equation** that specifies the underlying dynamics of the states. Deeper levels of hierarchy can be added, such as prior distributions of model parameters if the SSM is to be cast in a Bayesian framework; we discuss this further in Section 3.4. If the observation and process equations characterize conditional expectations, as is most often the case in fisheries applications, the SSM can then be written as:

$$\mathrm{E}[\boldsymbol{Y}_t | \boldsymbol{X}_t] = g(\boldsymbol{X}_t, \boldsymbol{\theta}), \tag{2}$$
$$\mathrm{E}[\boldsymbol{X}_t | \boldsymbol{X}_{t-1}] = h(\boldsymbol{X}_{t-1}, \boldsymbol{\theta}), \tag{3}$$

for functions g and h satisfying some regularity conditions, $t = 1, 2, \ldots$ with the initial states $\boldsymbol{X}_0$ to be estimated (or predicted, see below), and where $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$ is the vector of all unknown model parameters. These parameters are sometimes called **fixed**, to contrast them with $\boldsymbol{X}_t$ seen as a **random** effect; in the same mixed effects terminology, $\boldsymbol{\theta}$ is said to be **estimated** while $\boldsymbol{X}_t$ is to be **predicted**.

3.2 The joint likelihood

The system of Equations (2)–(3) only specifies some deterministic features of the system (in terms of conditional expectations), while the whole model includes stochasticity as well. Hence, the SSM is better represented by the (conditional) distribution at each level:

$$\boldsymbol{Y}_t | \boldsymbol{X}_t \sim P_{\boldsymbol{\theta}}(\boldsymbol{y}_t | \boldsymbol{x}_t) \text{ with density } p_{\boldsymbol{\theta}}(\boldsymbol{y}_t | \boldsymbol{x}_t),$$
$$\boldsymbol{X}_t | \boldsymbol{X}_{t-1} \sim P_{\boldsymbol{\theta}}(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}) \text{ with density } p_{\boldsymbol{\theta}}(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}),$$

where, by a slight abuse of notation, we respectively use P and p to denote any distribution and probability density function, with the arguments lifting any ambiguity. A notable special case is the linear Gaussian SSM, going back to Kalman (1960), in which both functions g and

h, respectively in (2) and (3), are linear and both $P_{\boldsymbol{\theta}}(\boldsymbol{y}_t|\boldsymbol{x}_t)$ and $P_{\boldsymbol{\theta}}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ are Gaussian distributions with covariance matrices depending only on elements of $\boldsymbol{\theta}$.

For a sample $\boldsymbol{y}_{1:T}$ of size T, the two main assumptions allow us to write the joint density function of $\boldsymbol{Y}_{1:T}$ and $\boldsymbol{X}_{0:T}$, also known as the **joint likelihood** or the complete data likelihood, as:

$$L_{\text{joint}}(\boldsymbol{\theta}, \boldsymbol{y}_{1:T}, \boldsymbol{x}_{0:T}) = p_{\boldsymbol{\theta}}(\boldsymbol{x}_0) \prod_{t=1}^{T} p_{\boldsymbol{\theta}}(\boldsymbol{y}_t|\boldsymbol{x}_t) p_{\boldsymbol{\theta}}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}). \tag{4}$$

The joint likelihood is the main starting point for estimation, inference and prediction as it gives a comprehensive picture of the SSM. However it is not practical on its own, since the states sequence $\boldsymbol{X}_{0:T}$ is unobserved. This motivates the construction of other objective functions; in the next two subsections, we present the two most popular alternatives in the fisheries science literature, namely the marginal likelihood as used in a frequentist approach and the posterior density in a Bayesian approach.

3.3   Frequentist approach and the marginal likelihood

In a frequentist approach, that is, considering $\boldsymbol{\theta}$ as a fixed unknown parameter at the population level, the **marginal likelihood** is defined as

$$L_{\text{marginal}}(\boldsymbol{\theta}, \boldsymbol{y}_{1:T}) = \int L_{\text{joint}}(\boldsymbol{\theta}, \boldsymbol{y}_{1:T}, \boldsymbol{x}_{0:T}) \, d\boldsymbol{x}_{1:T}, \tag{5}$$

where the unobserved states have been integrated out, leaving a function depending only on the observed $\boldsymbol{y}_{1:T}$ and the unknown parameter $\boldsymbol{\theta}$. The maximum likelihood estimator (MLE) for $\boldsymbol{\theta}$ is then defined as

$$\hat{\boldsymbol{\theta}}_{\text{ML}} = \arg \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \log L_{\text{marginal}}(\boldsymbol{\theta}, \boldsymbol{y}_{1:T}).$$

Its consistency has been recently demonstrated to hold in a general setting by Douc et al. (2011); however, to the best of our knowledge, its asymptotic normality has only been formally proved in the "discrete" case of the hidden Markov model (Bickel, Ritov, and Ryden, 1998) and for SSMs with Gaussian observation and process distributions.

The evaluation (and maximization) of $L_{\text{marginal}}(\boldsymbol{\theta}, \boldsymbol{y}_{1:T})$ requires the computation of the integral in (5), whose dimension increases with the sample size T. Apart from the linear Gaussian case, where the well-known Kalman filter (Harvey, 1989; Kalman, 1960) gives a recursive formula for evaluating the marginal likelihood, this high-dimensional integral needs to be approximated, yielding approximated MLEs; see Section 3.5.

As a second step, the prediction of the unobserved states is typically done by maximizing the joint log-likelihood (4), given a consistent estimate for $\boldsymbol{\theta}$ such as the MLE:

$$\hat{\boldsymbol{x}}_{1:T,\text{ML}} = \arg \max_{\boldsymbol{x}_{1:T} \in \mathsf{X}^T} \log L_{\text{joint}}(\hat{\boldsymbol{\theta}}_{\text{ML}}, \boldsymbol{y}_{1:T}, \boldsymbol{x}_{1:T}), \tag{6}$$

where $\boldsymbol{x}_0$ is estimated and thus part of $\hat{\boldsymbol{\theta}}_{\text{ML}}$. This maximization is performed as if $\boldsymbol{x}_{1:T}$ was an ordinary parameter and $L_{\text{joint}}(\boldsymbol{\theta}, \boldsymbol{y}_{1:T}, \boldsymbol{x}_{1:T})$ was an ordinary likelihood with a known $\boldsymbol{\theta}$; this is related to the hierarchical (h-)likelihood of Lee and Nelder (1996) and is analogous to the pseudo maximum likelihood estimation of Gong and Samaniego (1981). Although $\hat{\boldsymbol{\theta}}_{\text{ML}}$ appears to be "plugged-in", its estimation variability can be taken into account in the prediction of the unobserved states: Thorson and Kristensen (2016) propose a first-order bias correction for the prediction of non-linear functions of random effects. This correction, based on the Laplace approximation of the moment generating function suggested by Tierney, Kass, and Kadane (1989), is particularly useful in fisheries management since harvest control strategies often require quantities derived as non-linear functions of fixed and random effects.

3.4 Bayesian approach and the posterior density function

The Bayesian paradigm posits that $\boldsymbol{\theta}$ is itself a random variable, with its distribution representing uncertainty independent of the sampling variability inherent to data. Any a priori knowledge about $\boldsymbol{\theta}$ (i.e. before collecting data) is represented by the **prior distribution** $\pi(\boldsymbol{\theta})$ specified by the data analyst. As data are collected, new information becomes available through the joint likelihood $L_{\text{joint}}(\boldsymbol{\theta}, \boldsymbol{y}_{1:T}, \boldsymbol{x}_{0:T})$, and the **posterior density**

$$p(\boldsymbol{\theta}, \boldsymbol{x}_{0:T} | \boldsymbol{y}_{1:T}) = \frac{L_{\text{joint}}(\boldsymbol{\theta}, \boldsymbol{y}_{1:T}, \boldsymbol{x}_{0:T}) \pi(\boldsymbol{\theta})}{\int \int L_{\text{joint}}(\boldsymbol{\theta}, \boldsymbol{y}_{1:T}, \boldsymbol{x}_{0:T}) \pi(\boldsymbol{\theta}) \, d\boldsymbol{x}_{0:T} d\boldsymbol{\theta}} \tag{7}$$

represents the updated knowledge about $\boldsymbol{\theta}$ and $\boldsymbol{X}_{0:T}$. Note that in the Bayesian approach $\boldsymbol{X}_{0:T}$ and $\boldsymbol{\theta}$ are not qualitatively different, as they are both random variables; considering them together is related to data augmentation (Tanner and Wong, 1987).

The denominator of (7) involves high-dimensional integrals that again are not tractable in general and need to be approximated. This is a noteworthy feature of SSMs: both frequentist and Bayesian approaches face the same computational challenges of approximating non-trivial integrals of large dimensions. However, contrary to the frequentist ML approach, the approximation here will not be carried out with a maximization in mind: the Bayesian paradigm does not rely on point estimates, but rather uses the whole posterior density for inference. This is why Bayesian SSMs are invariably fitted by simulation techniques that yield draws from the posterior density function; see Section 3.5. We note the converse is not true: simulation-based approximations to integrals are not restricted to a Bayesian framework.

3.5 Approximations to high-dimensional integrals

This section presents an overview of the methods most commonly used in fisheries science for dealing with the high-dimensional integrals inherent to the fitting of SSMs. They can be roughly classified as simulation-based methods and direct approximations of the integrand.

Simulation-based approximation methods have been of central interest in the implementation of non-linear and non-Gaussian SSMs since the 1960s. The reason is that the conditional density of the states given the observations

$$p_{\boldsymbol{\theta}}(\boldsymbol{x}_{0:T} | \boldsymbol{y}_{1:T}) = \frac{L_{\text{joint}}(\boldsymbol{\theta}, \boldsymbol{y}_{1:T}, \boldsymbol{x}_{0:T})}{\int L_{\text{joint}}(\boldsymbol{\theta}, \boldsymbol{y}_{1:T}, \boldsymbol{x}_{0:T}) \, d\boldsymbol{x}_{0:T}}$$

and in particular the **filtering density** $p_{\boldsymbol{\theta}}(\boldsymbol{x}_T | \boldsymbol{y}_{1:T})$ played an important role in the early engineering applications of SSMs, where $\boldsymbol{\theta}$ was either assumed known or its estimation of secondary importance. Standard Markov chain Monte Carlo (MCMC; Gilks, Richardson, and Spiegelhalter, 1995; Robert and Casella, 2005), while a general method for drawing from posterior distributions in a Bayesian context, is not well suited for such high-dimensional problems and was quickly superseded by more efficient methods tailored for recursive estimations (e.g. Kantas et al., 2015). We defer to Doucet, de Freitas, and Gordon (2001) for an overview of the evolution of importance sampling into sequential importance sampling and the advent of modern sequential Monte Carlo methods, which are also known as particle filters (see Fearnhead and Künsch, 2018); see also Hürzeler and Künsch (2001) and Liu and West (2001). That said, variants of MCMC were developed in parallel; the ones based on Hamiltonian dynamics (see Neal, 2011) proved to be highly efficient with SSMs and are available in contemporary MCMC software such as Stan (Stan Development Team, 2016). More recently, particle MCMC (Andrieu, Doucet, and Holenstein, 2010) was introduced as

a hybrid between a particle filter and a standard MCMC algorithm and was successfully applied in ecology (see e.g. Knape and de Valpine, 2012) and fisheries science (Hosack, Peters, and Ludsin, 2014).

Direct approximations of the integrand are mostly based on Laplace's method for approximating general integrals (see Barndorff-Nielsen and Cox, 1989, Ch. 6.2). Assuming a global maximum exists (e.g. over $\boldsymbol{X}_{1:T}$ in the case of the marginal likelihood), a second-order Taylor expansion of the logarithm of the integrand about this maximum yields the density kernel of a multivariate Gaussian distribution, which can be conveniently integrated following a standardization involving the Hessian determinant. The marginal likelihood in (5) can thus be approximated by

$$\tilde{L}_{\mathrm{marginal}}(\boldsymbol{\theta}, \boldsymbol{y}_{1:T}) = L_{\mathrm{joint}}(\boldsymbol{\theta}, \boldsymbol{y}_{1:T}, \hat{\boldsymbol{x}}_{1:T})(2\pi)^{qT}\big| -\boldsymbol{H}(\hat{\boldsymbol{x}}_{1:T}, \boldsymbol{\theta})\big|^{-1/2},$$

where

$$\hat{\boldsymbol{x}}_{1:T} = \hat{\boldsymbol{x}}_{1:t}(\boldsymbol{y}_{1:T}, \boldsymbol{\theta}) = \arg\max_{\boldsymbol{x}_{1:T} \in \mathsf{X}^T} \log L_{\mathrm{joint}}(\boldsymbol{\theta}, \boldsymbol{y}_{1:T}, \boldsymbol{x}_{1:T}),$$

and

$$\boldsymbol{H}(\boldsymbol{x}_{1:T}, \boldsymbol{\theta}) = \frac{\partial^2 \log L_{\mathrm{joint}}(\boldsymbol{\theta}, \boldsymbol{y}_{1:T}, \boldsymbol{x}_{1:T})}{\partial \boldsymbol{x}_{1:T}\partial \boldsymbol{x}_{1:T}^{\top}},$$

with $|\cdot|$ and $^{\top}$ denoting matrix determinant and transpose, respectively. An interesting byproduct of the Laplace approximation is the ML prediction for the unobserved states $\hat{\boldsymbol{x}}_{1:t}(\boldsymbol{y}_{1:T}, \boldsymbol{\theta})$, for a given value of $\boldsymbol{\theta}$. The maximization of this approximated marginal likelihood with respect to $\boldsymbol{\theta}$ is not trivial as the dependence of $\hat{\boldsymbol{x}}_{1:T}$ upon $\boldsymbol{\theta}$ needs to be accounted for. Furthermore, standard errors of the resulting (Laplace-approximated) MLE of $\boldsymbol{\theta}$ are typically obtained by (a generalized version of) the delta method and require derivatives of $\tilde{L}_{\mathrm{marginal}}$ (Kristensen et al., 2016; Skaug and Fournier, 2006, see). It is the development of automatic differentiation (AD; e.g. Griewank and Walther, 2008) that made the implementation of this procedure possible. Many popular statistical software packages in fisheries science make use of the Laplace approximation with built-in AD, such as AD Model Builder (ADMB; Fournier et al., 2012; Skaug and Fournier, 2006) and the R (R Core Team, 2017) package Template Model Builder (TMB; Kristensen et al., 2016). The integrated nested Laplace approximation approach (Rue, Martino, and Chopin, 2009), with the corresponding R package R-INLA, utilizes Laplace's method for similar goals but within a Bayesian framework, although it does not employ AD. In being limited to low dimensions of non-linear fixed effects because of a grid search algorithm, this approach is not much used in fisheries science. As a final remark, we note here that Laplace's method is the simplest case of the family of adaptive Gaussian-Hermite approximations; more sophisticated versions (with more than one node) seem to be limited to small dimensions (see e.g. Huber, Ronchetti, and Victoria-Feser, 2004) and are thus not readily suitable for the fitting of SSMs.

4 Evolution of Software

Available software and computational power has long dictated the capabilities and limits of assessment models. However, the need for more realistic models has also pushed the development of software tools forward, notably in order to handle general non-linear, high dimensional, mixed effects models like SSAMs.

Early algorithms for performing fish stock assessments were calculated by hand, and later via desk calculators. The development of general-purpose programming languages (e.g. FORTRAN, PASCAL, C/C++) then allowed for automation and thus unlocked the possibility to test and compare various modelling approaches or variants of the same model (Pope, 1972).

The first formal statistical fitting procedures to be programmed for fish stock assessments were based on linearizations of desired models (e.g. Gulland, 1961; Schnute, 1977). Linearization was convenient as it involved closed-form formulas, but was clearly problematic from a modelling point of view because of the unrealistic mechanisms it implied. Hence, efficient and stable non-linear optimization of objective functions was needed to express standard parametric statistical fish stock assessment models.

AD Model Builder (Fournier et al., 2012; Skaug and Fournier, 2006) was developed in the fish stock assessment community by David A. Fournier in the late 1980s. It was the first model development environment to combine the technique of AD (Griewank and Walther, 2008) with a quasi-Newton optimizer, and a simplified template language for model development. The combined result was that non-linear models with hundreds and even thousands of model parameters could be easily expressed and, more importantly, optimized in a fast and robust way (Bolker et al., 2013).

Hundreds of peer-reviewed publications cite ADMB, and almost all general-purpose stock assessment tools are written in its framework (e.g. Methot and Wetzel, 2013; Miller and Legault, 2015; Williams and Shertzer, 2015). Most of these are not SSMs, but strictly parametric fixed effects models. The influence of AD Model builder in the fish stock assessment community cannot be exaggerated. In 2008 David A. Fournier received the William E. Ricker Award from the American Fisheries Society. Two quotes from the support letter, that underline the importance of ADMB are:

> *If Ricker is considered the grandfather of modern fisheries stock assessment, then Fournier should be considered the father.* — Ray Hilborn, Prof., University of Washington.

> *It is no exaggeration to state that Dr. Fournier's development of AD Model Builder has empowered an entire generation of fishery stock assessment scientists.* — Richard D. Methot, Jr., NOAA Fisheries.

Although prior to the work of Skaug and Fournier (2006) AD Model builder did not include automatic features to deal with SSMs, it was nonetheless tremendously successful at providing the fisheries community with a suitable tool for expressing their parametric models. In 2008 AD Model Builder was then, via a grant from the Gordon and Betty Moore Foundation, transferred to the non-profit ADMB-Foundation (`admb-foundation.org`) who continues to provide it as open-source software.

Standard parametric models require a trade-off between maintaining a reasonably small number of model parameters while still expressing sufficiently flexible models. More flexible models (whether Bayesian or frequentist) can also be obtained via the introduction of random effects in order to allow for the expression of complex time-varying processes. This approach requires the addition of relatively few model parameters.

With the advent of faster computers, MCMC techniques (e.g. Gilks, Richardson, and Spiegelhalter, 1995) became a practical option with which to formulate models with random effects. Fish stock assessment models were developed with custom code (e.g. Virtala, Kuikka, and Arjas, 1998) and in specialized tools for MCMC simulations (e.g Meyer and Millar, 1999). The early implementations (e.g. Fryer, 2002; Gudmundsson, 1994; Sullivan, 1992) of SSAMs were based on the Kalman filter and required linearization of the equations, and custom code to setup the filtering algorithms. Later approaches used sampling and MCMC algorithms (e.g. McAllister et al., 1994; Meyer and Millar, 1999).

Tools for MCMC simulation, BUGS/WinBUGS (Lunn et al., 2000), and later OpenBUGS (`www.openbugs.net`) and JAGS (Plummer, 2003), allow the user to specify their model in

an intuitive model language. The common language BUGS (Bayesian inference Using Gibbs Sampling) is more flexible than the standard formula interfaces used to specify linear models in standard statistical packages, but less flexible than a general purpose programming language. A model is specified simply by assigning distributions to parameters and observations. Non-linear functions can relate different quantities. The BUGS language and model fitting via MCMC has become very popular in many fields (e.g. Bolker et al., 2013). However, for estimating model parameters these tools have not become widely used in fish stock assessments. This is because these approaches may still take hours or days to run in a realistic assessment setting, and convergence is more difficult to judge than for a simple optimization.

Early frequentist SSAMs (e.g. Fryer, 2002; Gudmundsson, 1994; Sullivan, 1992) were implemented in a general purpose programming language (FORTRAN) with changes to the model requiring changes to the filtering equations in the code.

AD Model Builder was later extended to allow for random effects (Skaug and Fournier, 2006). Random effects are integrated out via the Laplace approximation, and the Laplace approximation is made efficient by using automatic differentiation. This extension of AD Model Builder made it possible to formulate SSAMs without considering the details of the filtering algorithms (Brinch, Eikeset, and Stenseth, 2011; Nielsen and Berg, 2014).

TMB (Kristensen et al., 2016) is a later implementation of the same ideas upon which AD Model Builder is based, and is developed and maintained by Kasper Kristensen. TMB builds on well-supported open-source software projects such as CppAD (Bell, 2012), Eigen (Guennebaud et al., 2010) and CHOLMOD (Chen et al., 2008). TMB has been rigorously tested against AD Model Builder for a wide set of example models and shown to yield identical results. However, it is much faster for models that include random effects, such as SSAMs. The use of sparseness, matrix libraries, and the way in which the computational graph is stored and optimized, allows for much faster computation (Kristensen et al., 2016).

SSAMs are now being increasingly applied. More than twenty official fish stock assessments in ICES are now conducted with variations of the model described in Nielsen and Berg (2014) and in the next section. In Canada, the SSAM described in Cadigan (2015) is now used to assess the northern cod stock. The availability of software to easily express and efficiently estimate these models has made all the difference. For example, the model in Nielsen and Berg (2014) can be optimized in 3 seconds in TMB and 3 minutes in AD Model Builder.

5   Application to the North Sea Cod Stock

We present here a contemporary application of SSMs for fish stock assessments, namely the SSAM for cod (*Gadus morhua*) in the North Sea, the Skagerrak and the English Channel; see **Figure 1** for a map of the assessment area defining the stock. This model is now the main one used by ICES for its management advice and predictions for North Sea cod (ICES, 2015, p. 703–792); see Nielsen and Berg (2014) for additional details and variants of this SSAM. The full data are freely available on www.stockassessment.org and in the R package "stockassessment" available on the GitHub page of the third author (https://github.com/fishfollower/samex).

The response vector $\boldsymbol{Y}_t$ consists of the total commercial catches $(C_{a,t})_{a=1,...,6+}$ (as numbers) and indices of relative abundance (unitless) coming from two surveys $(I_{a,t}^{(1)})_{a=1,...,5+}$ and $(I_{a,t}^{(2)})_{a=1,...,4+}$. All three variables are indexed by the age class a (with the largest class being a "plus group" including older fish), and the time index t refers to the year ranges 1963–2014, 1983–2015, and 1992–2014, corresponding to the commercial catch, first, and second surveys,
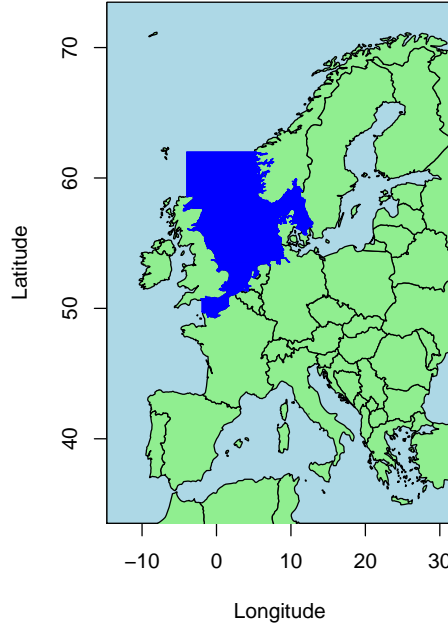
Figure 1: The assessment area for the North Sea cod stock, corresponding to Sub-area IV, the northern section of Division IIIa and Division VIId as defined in ICES (2015).

respectively. The first survey is conducted in the first quarter of the year while the second one is conducted in the third quarter of the year. The dimension of $\boldsymbol{Y}_t$ is thus $r = 15$ for the overlapping years. Missing values are not of concern here as only the available data contribute to the joint likelihood; in effect $\log C_{a,t}$, $\log I_{a,t}^{(1)}$ and $\log I_{a,t}^{(2)}$ are assumed independent conditional on the unobserved states.

The random effects vector $\boldsymbol{X}_t$ is composed of the abundance of fish of age a in the stock, denoted $N_{a,t}$, and the fishing mortality rate (seen as an instantaneous hazard rate) $F_{a,t}$. Here, both a and t span their entire range, $a = 1, \ldots, 6+$ and $t = 1963, \ldots, 2015$, i.e. six age-classes over 53 years, so that the dimension of $\boldsymbol{X}_t$ is $q = 12$.

Other variables enter the model as fixed covariates, see the Appendix for further details and equations. The model can be summarized as follows: additive Gaussian error appears on the log scale in both observation and process equations; the Gaussian error terms are all independent except between age classes for fishing mortality at any given time point t, where a first order autoregressive-like covariance structure is assumed across ages; both g and h functions, respectively from the observation and process equations (2) and (3), are non-linear with components of $\boldsymbol{X}_t$ appearing in various places; $\boldsymbol{\theta}$ is of dimension $p = 34$ with location and scale parameters of the Gaussian distribution; finally, the dimension of the integral defining the marginal likelihood in (5) is $qT = 636$.

In this exposition, we take a frequentist approach and fit the model by maximizing the Laplace-approximated marginal log-likelihood using the R package TMB. We deliberately omit a full Bayesian analysis here since such stock assessments rarely assume the existence of prior knowledge. However we note that MCMC runs with uninformative priors (not reported here) yield posterior distributions whose modes agree with our point estimates and generally confirm our results. Simulation studies (also not reported here) show that all model parameters are identifiable and estimable given the amount of data.

The estimates and standard errors of the model parameters are given in **Table 1**. We

notice that estimated variances pertaining to the recruits ($a = 1$) are systematically larger than the other variances, potentially hinting that some processes such as the assumed random walk for $\log N_{1,t}$ are too simplistic. We also note that some of the catch scaling parameters τ_t (see the Appendix for details) definitely seem useful for accounting for unexplained variability in the catches in the years 1993–2005.

The random effects are predicted by maximizing the joint log-likelihood given the Laplace-approximated MLE for $\boldsymbol{\theta}$, as in (6). **Figure 2** presents quantities derived from the predicted $\boldsymbol{X}_t$ which are of direct interest to stock managers: the spawning stock biomass (SSB) is the aggregated biomass of mature fish, obtained by a weighted sum of the predicted abundance; the average fishing mortality $\bar{F}_{2-4}$ is computed as the empirical mean of the predicted $F_{a,t}$ for age classes 2, 3 and 4 at each time point t; and the number of recruits is the predicted abundance for age class 1. The error bars represent out-of-sample predictions with pointwise 95% confidence intervals. The fishing pressure has steadily increased until reaching a plateau between the 1980s and late 1990s, driving the stock to an all-time low SSB in the mid 2000s. Since then, the stock seems to gradually recover along with a decreasing fishing mortality. The early spikes in the recruits indicate episodic changes in the productivity of the stock, a phenomenon observed at the ecosystem level (ICES, 2015, p. 703). Since the early 2000s, the cod stock seems to have stabilized at a low productivity regime. Out-of-sample predictions indicate a rather constant number of recruits, but thanks to fairly stable fishing mortality the SSB may well continue this upward trend.

The TMB package allows one to compute one-observation-ahead (or out-of-sample prediction) residuals, which are displayed in **Figure 3**. These residuals are computed as the standardized difference between an observed value and the corresponding fitted value based on all past observations, i.e. for past t and a indices. Contrary to residuals based on the overall fit, these should not be auto-correlated if the model fits the data well (see Thygesen et al., 2017). The residuals for all three response components are shown as bubbles in **Figure 3**: no clear trend or variability pattern emerges, indicating an overall satisfactory fit.

The SSAM for North Sea cod is simple, yet realistic enough, to show how the SSM framework is successfully used in contemporary fisheries applications. In particular, the hierarchical structure which identifies multiple sources of error seems crucial to appropriately quantify uncertainty. The dimensionality and the intricacies of nested optimizations may be daunting at first, but the active development of modern statistical softwares, such as the R package TMB, definitely alleviates some problems by efficiently running computations in the background while remaining completely transparent thanks to its open-source code. As an example, fitting the above model to the cod data takes about 15 seconds on a 2.3 GHz 6-year-old laptop computer, only requires basic R programming skills, and the one-observation-ahead residuals are built-in since the latest major release.

6   Current Challenges

The complex SSMs required in the field of fish stock assessment make it an interesting and fruitful field for applied statisticians. In fact there is a pressing need for further development of both the underlying theory and supporting statistical software. It is notable that fisheries science has pushed the limits of existing statistical software packages, developed exciting new ones, and is to the best of our knowledge among the only academic disciplines wherein the very useful combination of automatic differentiation and Laplace approximation is routinely used.

Goodness-of-fit statistics, including residual analysis, remain a challenge. It is now possible to compute the classical one-observation-ahead residuals, which are independent and have

a standard Gaussian distribution for a correct model, by successively including observations to predict the following observation (Thygesen et al., 2017). This procedure can be fully automated and is now available in TMB. The recent development of goodness-of-fit statistics for mixed models (Nakagawa and Schielzeth, 2013) may be adaptable to the more general state-space framework.

Many assumptions required by state-space models are unverifiable in practice, potentially casting doubt on any inference drawn from stock assessment models. In response to this concern, robust estimation techniques which remain reliable even when distributional assumptions are not entirely satisfied have been recently developed (Aeberhard et al., 2017; Xu et al., 2015).

Fully parametric fish stock assessment models (not state-space models) often struggled to express sufficiently flexible models. Often the models were only borderline identifiable. To guard against drawing incorrect conclusions it has become standard practice to validate identifiability of models. This is done using simulation studies with jittering (starting the optimization from, say, 100 different initial values and confirming that they reach the same minimum). SSAMs have fewer model parameters, but identifiability can still be an issue, so it is important to continue this validation practice. Data cloning (Lele, Nadeem, and Schmuland, 2012) is certainly one avenue worth exploring further.

The approximation of high-dimensional integrals, regardless of the estimation method, involves a computational cost and additional assumptions that need to be assessed. In particular, the accuracy of the Laplace approximation used by ADMB/TMB should be rigorously studied. Validating the Laplace approximation can be done by comparing to more computational demanding methods, such as MCMC, particle filtering, or importance sampling, but often this is only practical in scaled down versions of the models.

What is considered data (or observations) in fish stock assessment models are not always data, but rather output from other models. Consider for instance an index of stock abundance. It is not possible to measure abundance directly, so it is based on catches from scientific survey vessels. These catches are then processed via a model accounting for a number of factors (e.g. gear type, time of day, spatial changes in fishing) which are expected to affect catchability. The goal is to distill the most precise yearly index of abundance. The resulting index of abundance, and possibly its estimated uncertainty, are then used as observations in the model. A lot of information is potentially lost in this process. It should be the goal of future assessment models to include the raw observations directly.

An important part of fish stock assessment and management is to be able to predict future stock sizes and catches given certain management actions. Deterministic and fully parametric assessment models require new model parameters for each year, so predictions require additional assumptions external to the model. A key feature of SSAMs is that their prediction mechanism (the process equation and its respective stochastic assumptions) is directly estimated, and can be used to give model-consistent predictions. This convenient ability could be extended to the biological parameters (e.g. weights, maturity proportions, natural mortality), which are currently treated as fixed covariates in the models. Extending current assessment models to include states and prediction models for these quantities would make these predictions (and associated uncertainties) more objective and realistic.

Appendix: North Sea Cod Data and Model Specification

In addition to the response variable

$$\boldsymbol{Y}_t = (\log C_{a=1,t}, \dots, \log C_{a=6+,t}, \log I^{(1)}_{a=1,t}, \dots, \log I^{(1)}_{a=5+,t}, \log I^{(2)}_{a=1,t}, \dots, \log I^{(2)}_{a=4+,t})^\top$$

and the random effects

$$\boldsymbol{X}_t = (\log N_{a=1,t}, \dots, \log N_{a=6+,t}, \log F_{a=1,t}, \dots, \log F_{a=6+,t})^\top,$$

an estimate of the natural mortality rate $M_{a,t}$ enters the model as a fixed covariate. $M_{a,t}$ is often derived from stomach sampling, or from additional knowledge about the ecosystem. Moreover, for computing derived quantities such as the spawning stock biomass, other auxiliary covariates are needed such as the mean weight of fish in the stock and the proportion of mature females at a given age.

The process equation describes the dynamics in the unobserved states, and is based on the conditional expectation of the current states given the previous states:

$$\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{X}_{t-1}] = \begin{cases} \log N_{1,t} &= \log N_{1,t-1} \\ \log N_{a,t} &= \log N_{a-1,t-1} - F_{a-1,t-1} - M_{a-1,t-1}, & 2 \le a < A \\ \log N_{A,t} &= \log \big[N_{A-1,t-1}\exp(-F_{A-1,t-1} - M_{A-1,t-1}) \\ & \quad + N_{A,t-1}\exp(-F_{A,t-1} - M_{A,t-1})\big] \\ \log F_{a,t} &= \log F_{a,t-1}, & 1 \le a \le A, \end{cases}$$

where A denotes the largest age class. These equations assume a random walk for $\log N_{1,t}$ and for the whole vector $(\log F_{1,t}, \dots, \log F_{A,t})^\top$, a survival process for $\log N_{a,t}$ where the combination of F and M represents total mortality, and a modified survival process for the "plus group" in $\log N_{A,t}$. The corresponding distribution $P_{\boldsymbol{\theta}}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ is a multivariate Gaussian with zero mean vector. The first A Gaussian error components are independent, while we enforce an AR(1)-like correlation structure for the others:

$$\mathrm{Cor}\big[\log(F_{a,t}), \log(F_{\tilde{a},t})\big] = \rho^{|a-\tilde{a}|},$$

where the between-age correlation ρ is an element of $\boldsymbol{\theta}$. Other fixed parameters include five separate variances: one for recruitment ($\sigma^2_{N_{a=1}}$); one for survival ($\sigma^2_{N_{a>1}}$); one for fishing mortality at age 1 ($\sigma^2_{F_{a=1}}$); and one for fishing mortality at older ages ($\sigma^2_{F_{a\ge 2}}$).

The observation equation relates the unobserved states to the observed response variables through a conditional expectation:

$$\mathrm{E}[\boldsymbol{Y}_t|\boldsymbol{X}_t] = \begin{cases} \log C_{a,t} &= \log\left[\frac{F_{a,t}}{Z_{a,t}}(1 - \exp(-Z_{a,t}))N_{a,t}\right] \\ \log I^{(s)}_{a,t} &= \log\left[Q^{(s)}_a \exp(-Z_{a,t}\frac{D^{(s)}}{365})N_{a,t}\right], \end{cases} \quad 1 \le a \le A,$$

where $s = 1, 2$ identifies the surveys, the largest age class A is 5+ for $s = 1$ and 4+ for $s = 2$, $Z_{a,t} = M_{a,t} + F_{a,t}$ is the total mortality rate, $D^{(s)}$ is the number of days into the year when survey (s) was conducted, $Q^{(s)}_a$ are so-called catchability coefficients which scale the survey relative indices to the stock abundance. The catchabilities are unknown parameters that need to be estimated, there are nine of them as they are distinct for each age class and each survey. Auxiliary information and expertise from fisheries scientists cast doubt on the reliability of the absolute level of the catches between 1993 and 2005, hence extra catch scaling parameters τ_t are added (and estimated) for these years:

$$\log C_{a,t} = \log\left[\frac{1}{\tau_t}\frac{F_{a,t}}{Z_{a,t}}(1 - \exp(-Z_{a,t}))N_{a,t}\right], \quad t \in \{1993, \dots, 2005\}.$$

The corresponding distribution $P_{\boldsymbol{\theta}}(\boldsymbol{y}_t|\boldsymbol{x}_t)$ is multivariate Gaussian with zero mean vector, all components are conditionally independent. Distinct variances are specified for different age groups for the three data sources: commercial catch has a variance for age 1 ($\sigma^2_{C_{a=1}}$), for

age 2 ($\sigma^2_{C_{a=2}}$), and a common one for older ages ($\sigma^2_{C_{a\geq3}}$); the two surveys have a variance for age 1 ($\sigma^2_{I^{(s)}_{a=1}}$ for $s=1,2$) and another one for older ages ($\sigma^2_{I^{(s)}_{a\geq2}}$ for $s=1,2$). This amounts to a total of seven observation variance parameters. We can finally summarize all fixed parameters which need to be estimated:

$$\boldsymbol{\theta} = (\rho, \sigma^2_{N_{a=1}}, \sigma^2_{N_{a\geq1}}, \sigma^2_{F_{a=1}}, \sigma^2_{F_{a\geq1}}, \sigma^2_{C_{a=1}}, \sigma^2_{C_{a=2}}, \sigma^2_{C_{a\geq3}}, \sigma^2_{I^{(1)}_{a=1}}, \sigma^2_{I^{(1)}_{a\geq1}}, \sigma^2_{I^{(2)}_{a=1}}, \sigma^2_{I^{(2)}_{a\geq1}},$$
$$\tau_{1993}, \ldots, \tau_{2005}, Q^{(1)}_{a=1}, \ldots, Q^{(1)}_{a=5+}, Q^{(2)}_{a=1}, \ldots, Q^{(2)}_{a=4+})^\top.$$

Disclosure Statement

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

Acknowledgements

References

Aanes, S., S. Engen, B.-E. Sæther, and R. Aanes (2007). Estimation of the parameters of fish stock dynamics from catch-at-age data and indices of abundance: can natural and fishing mortality be separated? *Canadian Journal of Fisheries and Aquatic Sciences* 64 (8), 1130–1142.

Aarts, G. and J. J. Poos (2009). Comprehensive discard reconstruction and abundance estimation using flexible selectivity functions. *ICES Journal of Marine Science* 66 (4), 763–771.

Aeberhard, W. H., E. Cantoni, C. A. Field, J. Mills Flemming, and X. Xu (2017). Robust and Consistent Estimation of Fixed Parameters of General State-Space Models. Working Paper.

Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B* 72 (3), 269–342.

Baranov, F. I. (1918). On the question of the biological basis of fisheries. *Nauchnye Issledovaniya Ikhriologicheskii Institutu Izvestiya* 1 (8), 81–128.

Barndorff-Nielsen, O. E. and D. R. Cox (1989). *Asymptotic techniques for use in statistics*. London, UK: Chapman & Hall.

Bell, B. M. (2012). CppAD: A package for C++ algorithmic differentiation. *Computational Infrastructure for Operations Research*. URL: http://www.coin-or.org/CppAD.

Beverton, R. J. H. (1954). *Notes on the use of theoretical models in the study of the dynamics of exploited fish populations*. Vol. 2. Miscellaneous Contributions. Beaufort, NC: US Fishery Laboratory.

Beverton, R. J. H. and S. J. Holt (1957). On the dynamics of exploited fish populations. In: *Fisheries Investigation Series 2*. Vol. 19. London, UK: Ministry of Agriculture, Fisheries and Food.

Bickel, P. J., Y. Ritov, and T. Ryden (1998). Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *The Annals of Statistics* 26 (4), 1614–1635.

Bolker, B. M., B. Gardner, M. Maunder, C. W. Berg, M. Brooks, L. Comita, E. Crone, S. Cubaynes, T. Davies, P. de Valpine, J. Ford, O. Gimnez, E. J. Kim, C. Lennert-Cody, A. Magnusson, S. Martell, J. Nash, A. Nielsen, J. Regetz, H. Skaug, and E. Zipkin (2013). Strategies for fitting nonlinear ecological models in R, AD Model Builder, and BUGS. *Methods in Ecology and Evolution* 4 (6), 501–512.

Brinch, C. N., A. M. Eikeset, and N. C. Stenseth (2011). Maximum likelihood estimation in nonlinear structured fisheries models using survey and catch-at-age data. *Canadian Journal of Fisheries and Aquatic Sciences* 68 (10), 1717–1731.

Buckland, S. T., K. B. Newman, C. Fernández, L. Thomas, and J. Harwood (2007). Embedding population dynamics models in inference. *Statistical Science* 22 (1), 44–58.

Butterworth, D. S., J. N. Ianelli, and R. Hilborn (2003). A statistical model for stock assessment of southern bluefin tuna with temporal changes in selectivity. *African Journal of Marine Science* 25 (1), 331–361.

Cadigan, N. G. (2015). A state-space stock assessment model for northern cod, including under-reported catches and variable natural mortality rates. *Canadian Journal of Fisheries and Aquatic Sciences* 73 (2), 296–308.

Chen, Y., T. A. Davis, W. W. Hager, and S. Rajamanickam (2008). Algorithm 887: CHOLMOD, supernodal sparse Cholesky factorization and update/downdate. *ACM Transactions on Mathematical Software* 35 (3), 1–22.

Costello, C., D. Ovando, T. Clavelle, C. K. Strauss, R. Hilborn, M. C. Melnychuk, T. A. Branch, S. D. Gaines, C. S. Szuwalski, R. B. Cabral, D. N. Raderb, and A. Leland (2016). Global fishery prospects under contrasting management regimes. *Proceedings of the National Academy of Sciences* 113 (18), 5125–5129.

Derzhavin, A. N. (1922). The stellate sturgeon (Acipenser stellatus Pallas), a biological sketch. *Byulleten Bakinskoi Ikhtiologicheskoi Stanstsii* 1, 1–393.

Doubleday, W. G. (1976). A least squares approach to analyzing catch at age data. *International Commission for the Northwest Atlantic Research Bulletin* 12, 69–81.

Douc, R., E. Moulines, J. Olsson, and R. Van Handel (2011). Consistency of the maximum likelihood estimator for general hidden Markov models. *The Annals of Statistics* 39 (1), 474–513.

Doucet, A., N. de Freitas, and N. J. Gordon (2001). An introduction to sequential Monte Carlo methods. In: *Sequential Monte Carlo Methods in Practice*. Ed. by A. Doucet, N. de Freitas, and N. J. Gordon. New York, NY: Springer-Verlag, 3–14.

Fearnhead, P. and H. R. Künsch (2018). Particle Filters and Data Assimilation. *Annual Review of Statistics and Its Application* 5, 421–449.

Food and Agriculture Organization of the United Nations (2014). *The State of World Fisheries and Aquaculture: Opportunities and Challenges*. Report. Rome, Italy: FAO.

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

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

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

Fryer, R. J. (2002). *TSA: Is it the way?* Appendix D in Report of Working Group on Methods of Fish Stock Assessment. ICES CM 2002/D:01, 86–93.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter, eds. (1995). *Markov chain Monte Carlo in Practice.* London, UK: Chapman & Hall/CRC.

Gong, G. and F. J. Samaniego (1981). Pseudo maximum likelihood estimation: Theory and applications. *The Annals of Statistics* 9 (4), 861–869.

Griewank, A. and A. Walther (2008). *Evaluating derivatives: principles and techniques of algorithmic differentiation.* 2nd ed. Philadelphia, PA: Society for Industrial and Applied Mathematics.

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

Gudmundsson, G. and T. Gunnlaugsson (2012). Selection and estimation of sequential catch-at-age models. *Canadian Journal of Fisheries and Aquatic Sciences* 69 (11), 1760–1772.

Guennebaud, G., B. Jacob, P. Avery, A. Bachrach, and S. et al. Barthelemy (2010). *Eigen v3.* URL: http://eigen.tuxfamily.org.

Gulland, J. A. (1961). Fishing and the stocks of fish at Iceland. *Fishery Investigations Series 2* 23 (4), 1–52.

— (1965). *Estimation of mortality rates.* Annex to Arctic Fisheries Working Group Report. ICES CM 1965, 71–79.

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

Holmes, L., C. K. Strauss, K. de Vos, and K. Bonzon (2014). *Towards investment in sustainable fisheries: A framework for financing the transition.* Discussion document. Environmental Defense Fund and The Prince of Wales's International Sustainability Unit.

Hosack, G. R., G. W. Peters, and S. A. Ludsin (2014). Interspecific relationships and environmentally driven catchabilities estimated from fisheries data. *Canadian Journal of Fisheries and Aquatic Sciences* 71 (3), 447–463.

Huber, P., E. M. Ronchetti, and M.-P. Victoria-Feser (2004). Estimation of generalized linear latent variable models. *Journal of the Royal Statistical Society, Series B* 66 (4), 893–908.

Hürzeler, Ma. and H. R. Künsch (2001). Approximating and maximising the likelihood for a general state-space model. In: *Sequential Monte Carlo Methods in Practice.* Ed. by A. Doucet, N. de Freitas, and N. J. Gordon. New York, NY: Springer-Verlag, 159–175.

ICES (2015). *Report of the Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK).* ICES CM 2015/ACOM:13.

Jardim, E., C. P. Millar, I. Mosqueira, F. Scott, G. C. Osio, M. Ferretti, N. Alzorriz, and A. Orio (2015). What if stock assessment is as simple as a linear model? The a4a initiative. *ICES Journal of Marine Science* 72 (1), 232–236.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika* 34 (2), 183–202.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* 82 (1), 35–45.

Kalman, R. E. and R. S. Bucy (1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering* 83 (3), 95–108.

Kantas, N., A. Doucet, S. S. Singh, J. Maciejowski, and N. Chopin (2015). On particle methods for parameter estimation in state-space models. *Statistical science* 30 (3), 328–351.

Knape, J. and P. de Valpine (2012). Fitting complex population models by combining particle filters with Markov chain Monte Carlo. *Ecology* 93 (2), 256–263.

Kristensen, K., A. Nielsen, C. W. Berg, H. Skaug, and B. M. Bell (2016). TMB: Automatic Differentiation and Laplace Approximation. *Journal of Statistical Software* 70 (5), 1–21.

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

Lee, Y. and J. A. Nelder (1996). Hierarchical Generalized Linear Models. *Journal of the Royal Statistical Society, Series B* 58 (4), 619–678.

Lele, S. R., K. Nadeem, and B. Schmuland (2012). Estimability and likelihood inference for generalized linear mixed models using data cloning. *Journal of the American Statistical Association* 105 (492), 1617–1625.

Lewy, P. (1988). Integrated stochastic virtual population analysis: estimates and their precision of fishing mortalities and stock sizes for the North Sea whiting stock. *ICES Journal of Marine Science* 44 (3), 217–228.

Link, J. S., A. Bundy, W. J. Overholtz, N. Shackell, J. Manderson, D. Duplisea, J. Hare, M. Koen-Alonso, and K. D. Friedland (2011). Ecosystem-based fisheries management in the Northwest Atlantic. *Fish and Fisheries* 12 (2), 152–170.

Liu, J. and M. West (2001). Combined parameter and state estimation in simulation-based filtering. In: *Sequential Monte Carlo Methods in Practice*. Ed. by A. Doucet, N. de Freitas, and N. J. Gordon. New York, NY: Springer-Verlag, 197–223.

Lunn, D. J., A. Thomas, N. Best, and D. J. Spiegelhalter (2000). WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing* 10 (4), 325–337.

MacDonald, I. L. and W. Zucchini (1997). *Hidden Markov and other models for discrete-valued time series*. London, UK: Chapman & Hall/CRC.

McAllister, M. K., E. K. Pikitch, A. E. Punt, and R. Hilborn (1994). A Bayesian approach to stock assessment and harvest decisions using the sampling/importance resampling algorithm. *Canadian Journal of Fisheries and Aquatic Sciences* 51 (12), 2673–2687.

McCulloch, C. E. and S. R. Searle (2000). *Generalized, Linear and Mixed Models*. New York, NY: Wiley.

Methot, R. D. and C. R. Wetzel (2013). Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research* 142, 86–99.

Meyer, R. and R. B. Millar (1999). BUGS in Bayesian stock assessments. *Canadian Journal of Fisheries and Aquatic Sciences* 56 (6), 1078–1087.

Miller, T. J. and C. M. Legault (2015). *Technical Details for ASAP version 4*. Northeast Fisheries Science Center Reference Document 15-17. U.S. Department of Commerce, National Oceanic and Atmospheric Administration.

Minto, C., J. Mills Flemming, G. L. Britten, and B. Worm (2013). Productivity dynamics of Atlantic cod. *Canadian Journal of Fisheries and Aquatic Sciences* 71 (2), 203–216.

Murphy, G. I. (1965). A solution of the catch equation. *Journal of the Fisheries Board of Canada* 22 (1), 191–202.

Nakagawa, S. and H. Schielzeth (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4 (2), 133–142.

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. In: *Handbook of Markov Chain Monte Carlo*. Ed. by S. Brooks, A. Gelman, G. Jones, and X.-L. Meng. Boca Raton, FL: Chapman & Hall/CRC, 113–162.

Nielsen, A. and C. W. Berg (2014). Estimation of time-varying selectivity in stock assessments using state-space models. *Fisheries Research* 158, 96–101.

Paloheimo, J. E. (1980). Estimation of mortality rates in fish populations. *Transactions of the American Fisheries Society* 109 (4), 378–386.

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

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*. Vol. 124, 125.

Pope, J. G. (1972). An investigation of the accuracy of virtual population analysis using cohort analysis. *International Commission for the Northwest Atlantic Research Bulletin* 9 (10), 65–74.

Pope, J. G. and J. G. Shepherd (1982). A simple method for the consistent interpretation of catch-at-age data. *ICES Journal of Marine Science* 40 (2), 176–184.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Robert, C. P. and G. Casella (2005). *Monte Carlo Statistical Methods*. 2nd ed. New York, NY: Springer-Verlag.

Royce, W. F. (1988). Centennial Lecture IV: The historical development of fishery science and management. *Marine Fisheries Review* 50 (4), 30–40.

Rue, H., S. Martino, and N. Chopin (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B* 71 (2), 319–392.

Schnute, J. T. (1977). Improved estimates from the Schaefer production model: Theoretical considerations. *Journal of the Fisheries Board of Canada* 34 (5), 583–603.

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

Stan Development Team (2016). *Stan Modeling Language Users Guide and Reference Manual*. Version 2.14.0. URL: http://mc-stan.org.

Sullivan, P. J. (1992). A Kalman filter approach to catch-at-length analysis. *Biometrics* 48 (1), 237–257.

Tanner, M. A. and W. H. Wong (1987). The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association* 82 (398), 528–540.

Thorson, J. T. and K. Kristensen (2016). Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples. *Fisheries Research* 175, 66–74.

Thygesen, U. H., C. M. Albertsen, C. W. Berg, K. Kristensen, and A. Nielsen (2017). Validation of state space models fitted as mixed effects models. Submitted.

Tierney, L., R. E. Kass, and J. B. Kadane (1989). Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *Journal of the American Statistical Association* 84 (407), 710–716.

Virtala, M., S. Kuikka, and E. Arjas (1998). Stochastic virtual population analysis. *ICES Journal of Marine Science* 55 (5), 892–904.

Williams, E. H. and K. W. Shertzer (2015). *Technical Documentation of the Beaufort Assessment Model (BAM)*. NOAA Technical Memorandum NMFS-SEFSC-671. U.S. Department of Commerce, National Oceanic and Atmospheric Administration.

Xu, X., E. Cantoni, J. Mills Flemming, and C. Field (2015). Robust state space models for estimating fish stock maturities. *Canadian Journal of Statistics* 43 (1), 133–150.

Table 1: North Sea cod data, estimates and standard errors of fixed parameters.

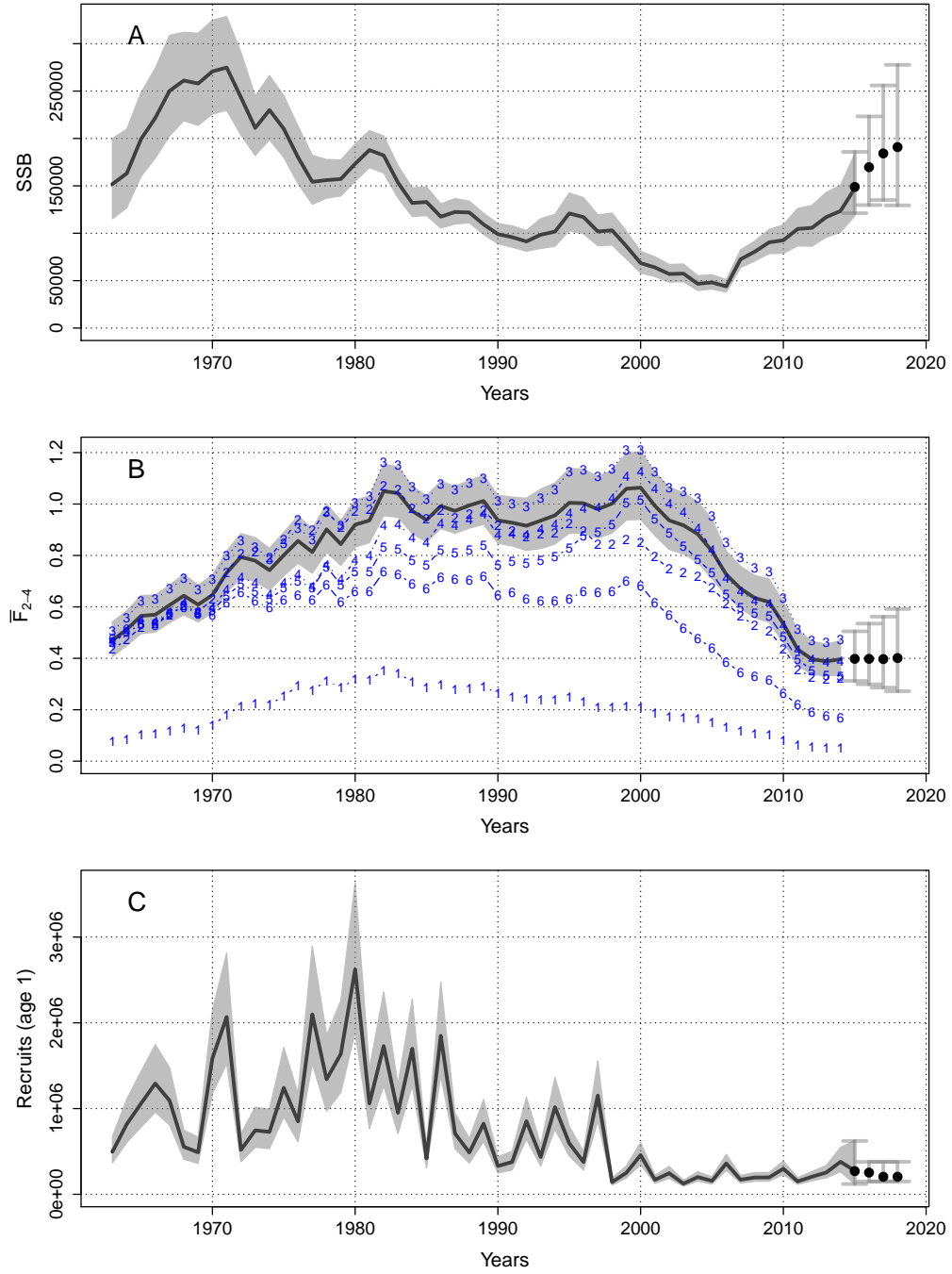| Parameter	Estimate	S.E.
$Q^{(1)}_{a=1}$	0.0076	0.0007
$Q^{(1)}_{a=2}$	0.0352	0.0018
$Q^{(1)}_{a=3}$	0.0661	0.0033
$Q^{(1)}_{a=4}$	0.0664	0.0038
$Q^{(1)}_{a=5+}$	0.0910	0.0065
$Q^{(2)}_{a=1}$	0.0279	0.0025
$Q^{(2)}_{a=2}$	0.0502	0.0034
$Q^{(2)}_{a=3}$	0.0653	0.0046
$Q^{(2)}_{a=4+}$	0.0656	0.0054
$\sigma_{F_{a=1}}$	0.1572	0.0361
$\sigma_{F_{a\geq 2}}$	0.0944	0.0125
$\sigma_{N_{a=1}}$	0.7483	0.0849
$\sigma_{N_{a\geq 2}}$	0.1113	0.0169
$\sigma_{C_{a=1}}$	0.6169	0.0707
$\sigma_{C_{a=2}}$	0.2411	0.0320
$\sigma_{C_{a\geq 3}}$	0.0644	0.0170
$\sigma_{I^{(1)}_{a=1}}$	0.4731	0.0649
$\sigma_{I^{(1)}_{a\geq 2}}$	0.2278	0.0164
$\sigma_{I^{(2)}_{a=1}}$	0.3727	0.0639
$\sigma_{I^{(2)}_{a\geq 2}}$	0.2586	0.0255
ρ	0.8862	0.0596
τ_{1993}	0.9388	0.0850
τ_{1994}	1.0344	0.1008
τ_{1995}	1.1744	0.1172
τ_{1996}	1.0155	0.1012
τ_{1997}	0.8589	0.0840
τ_{1998}	0.7148	0.0701
τ_{1999}	0.8573	0.0862
τ_{2000}	0.9437	0.0947
τ_{2001}	1.2900	0.1270
τ_{2002}	0.8660	0.0846
τ_{2003}	1.4899	0.1484
τ_{2004}	1.1323	0.1117
τ_{2005}	0.9627	0.0878

Figure 2: North Sea cod data, derived quantities of direct interest to the stock management: (A) spawning stock biomass; (B) average fishing mortality for age 2–4 (solid line and envelope) with the predicted fishing mortalities overlaid for all six age classes (numbers and dashes); (C) predicted number of recruits $N_{a=1,t}$. All shaded envelopes are (interpolated pointwise) 95% confidence intervals, while the error bars are out-of-sample predictions with respective 95% confidence intervals.
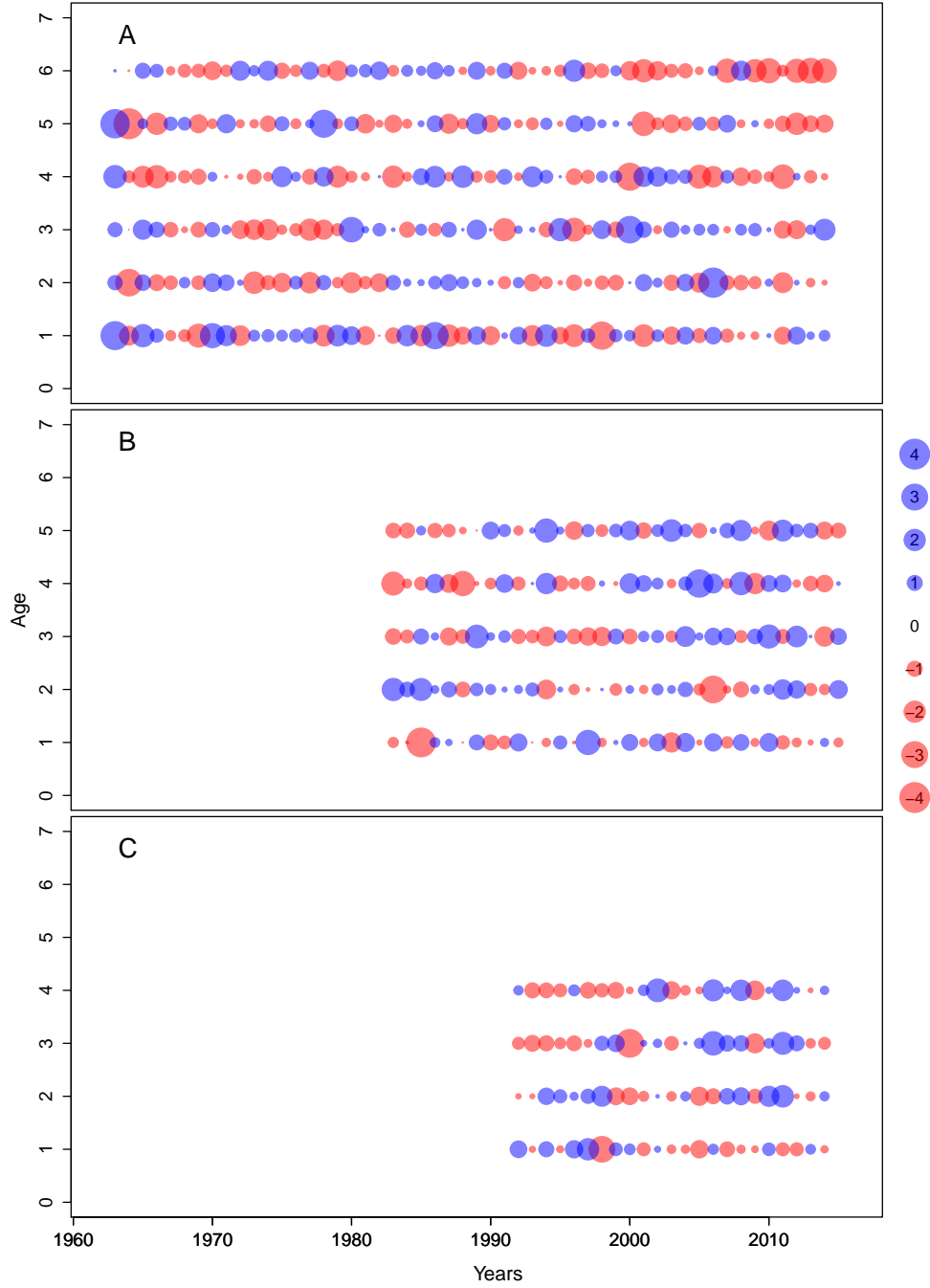
Figure 3: North Sea cod data, one-observation-ahead residuals for the three components of the response: (A) $C_{a,t}$; (B) $I_{a,t}^{(1)}$; (C) $I_{a,t}^{(2)}$. The magnitude of the residual is proportional to the area of the bubble. Blue bubbles represent positive residuals while red ones represent negative residuals.

23