

A State-Space Model for Abundance Estimation from Bottom Trawl Data with Applications to Norwegian Winter Survey

Peter M. MacKean* Geir Storvik† Sondre Aanes‡

September 10, 2015

Abstract

We study a hierarchical dynamic state-space model for abundance estimation. A generic data fusion approach for combining computer simulated posterior samples of catch output data with observed research survey indices using sequential importance sampling is presented. Posterior samples of catch generated from a computer software are used as a primary source of input data through which fisheries dependent information is mediated. Direct total stock abundance estimates are obtained without the need to estimate any intermediate parameters such as catchability and mortality. Numerical results of a simulation study show that our method provides a useful alternative to existing methods. We apply the method to data from the Barents Sea Winter survey for Northeast Arctic cod (*Gadus morhua*). The results based on our method are comparable to results based on current methods.

Keywords: Catchability, Importance Sampling, Sequential Monte Carlo

*Department of Mathematics, University of Oslo, email:maisha@math.uio.no

†Department of Mathematics, University of Oslo, email:geirs@math.uio.no

‡Norwegian Computing Center, Oslo, Norway, email:Sondre.Aanes@nr.no

1 Introduction

An important fisheries management goal is often to utilize information from observed data in designing sustainable harvesting strategies. The primary sources of data frequently encountered in fisheries may be broadly grouped into fishery dependent and fishery independent data. However, combining data from multiple sources usually presents some statistical challenges. This has been the motivation among statisticians and marine ecologists who have invested tremendous efforts in the development of different stock assessment models. These models are used to investigate dynamics in stock size as well as prediction of the impact of the harvesting process on future stock populations through simultaneous modeling of empirical data (??). State-space models are often used to incorporate process randomness in stock assessment. Much of the early statistical work in this area focused mainly on development of models using classical likelihood-based methods of inference (???). However, it is well known that this approach can be computationally limited by the need for high dimensional integration which requires unrealistic linearity and normality assumptions (??).

? proposed a fully Bayesian state-space modeling approach as a less restrictive alternative to likelihood based inference. Bayesian state-space models (BSSM) provide a powerful and flexible framework for modeling stochastic dynamics in animal populations (??). This class of models is now widely used in fisheries to simultaneously model both process and measurement errors for proper propagation and handling of commonly encountered data uncertainties (?). Several authors have recently investigated BSSM using the Markov chain Monte Carlo (MCMC) method which is considered gold standard particularly for marginal likelihoods (?????). However, sequential Monte Carlo methods may offer a computationally less expensive alternative (???).

??? present a Bayesian framework for estimating catch-at-age from different data sources. They combine age, length and weight data in a Bayesian hierarchical model for the estimation of catch-at-age with appropriate uncertainty and sampling scheme. ? used estimates of catch-at-age data and indices of relative abundance to investigate stock dynamics using an age-structured Bayesian state-space model with separable mortality causes. In this paper, we examine the fundamental issues of combining information from multiple data sources and error propagation in total abundance estimation. We study a hierarchical dynamic abundance estimation model using sequential Monte Carlo methods by extending the work of ?. The main difference of

our method is that unlike ?, we conceptualize true catch as unobserved and treat computer generated posterior output samples as our source of fishery dependent input data. The posterior data are combined with observed fishery independent survey indices using adaptive importance sampling algorithms. Furthermore, unlike current models, here we formulate the usual observation process model to eliminate the need to deal with nuisance “catchability” parameters.

Outline: In Section 2, a brief description of the type of data we consider is given. In Section 3, we motivate the Bayesian framework and the proposed sequential importance sampling method as well as describing the proposed dynamic model for abundance. We discuss a marine fisheries application example and present the results of the analysis in Section 4. We conclude the paper with some concluding remarks in Section 5.

2 Data

The primary sources of data that are frequently used in fisheries research may be broadly grouped into fishery dependent and fishery independent data. In this paper, fishery dependent data consists of landings data usually from commercial fleets which contains information about amount (biomass), size and the age-structure of the catches. These data are used to obtain estimates of catch-at-age numbers which are often reported observed catch data. The catch data are the primary input for fishery dependent information in models for the estimation of stock size and the overall stock dynamics. Nevertheless, the catch estimates are not directly observed but estimated from composite measures which are themselves observed in error.

Although most authors have previously ignored any uncertainties associated with these reported catch estimates, there have been recent efforts to formally model the bias using for example lognormal noise ???. ??? recently introduced a Bayesian hierarchical modeling framework for estimating catch-at-age data. They propose a Monte Carlo based inference procedure for obtaining posterior samples of catch-at-age conditional on the observed fishing process measures. This procedure has been implemented in a computer software recently developed at the Institute of Marine Research in Norway which is hereafter referred to as the ECA program. Here we utilize the computer generated output posterior samples from the ECA program as the source of fishery dependent input data.

Fishery independent information usually consists of survey data collected from research vessels. These data are in principle complex multivariate observations on composites measures which may include total catch biomass, otholiths and other biological samples, acoustics data, as well as assessments on geo-physical spatial variables as well. These measures are often used to compute indices of relative abundance which are assumed to be related to the stock abundance. In this paper we use winter survey data on demersal fish species in the Barents Sea. These data are available through the Norwegian Institute of Marine Research (IMR). A detailed description of these data may be found in ?.

3 Modelling and inference

Let the tuple $\{N, C\}$ denote the state process and $\{D^N, D^C\}$ be the observation process, where N is stock abundance, C is true catch, D^N denotes observed survey indices of relative abundance data and D^C the observed catch data. Note that both N and C are multidimensional including abundances/catches for many years and age groups. Following the Bayesian paradigm, our aim will be to perform inference based on the posterior distribution $p(N, C|D^N, D^C)$.

Suppose the assumption of conditional independence of D^N and D^C is reasonable, that is $p(D^N, D^C|N, C) = p(D^C|C) p(D^N|N)$. Then Bayes rule yields

$$\begin{aligned} p(N, C|D^N, D^C) &\propto p(C) p(D^C|C) p(N|C) p(D^N|N) \\ &\propto p(C|D^C) p(N|C) p(D^N|N), \end{aligned} \quad (1)$$

where $p(C|D^C)$ is the posterior catch distribution based on catch data D^C only, $p(D^N|N)$ is a likelihood for survey indices conditional on the state process and $p(N|C)$ is a dynamic state space process model. The decomposition in equation(1) conveniently permits us to incorporate fishery dependent information in our estimation using independently computer generated posterior samples from $p(C|D^C)$ which in this paper are obtained through the ECA program. So far, we can simulate the state process from the joint distribution of N and C conditional on the observed catch process through the dynamic process model described in section 3.2 below. However, catch data alone do not provide sufficient information to uniquely determine trends in stock abundance (?).

In this section, we present a generic data fusion approach for combining computer simulated posterior output samples of catch data with observed research survey indices data using a simple adaptive importance sampling algorithm (??). We use posterior samples generated from the ECA program conditional of the observed catch-at-age estimates as a source of input catch data with better resolution (??).

3.1 Importance Sampling algorithm

The proposed algorithm is a simple application of importance sampling:

Algorithm 1 The main algorithm

for $j = 1$ to B **do**

1. Generate $C^j \sim p(C|D^C)$, $j = 1, \dots, B$,

2. State $\theta \sim p(\theta)$

2. Simulate $N^j \sim q(N|C^j; \theta)$

3. Calculate importance sampling weights $W^j \propto \frac{p(N^j|C^j; \theta)p(D^N|N^j; \theta)}{q(N^j|C^j, D^N; \theta)}$,

$\sum_{j=1}^B W^j = 1$.

end for

Estimate $E[g(N)]$ by $\sum_{j=1}^B g(N^j)W^j$.

In this algorithm the first step samples from the posterior catch distribution based on DC only using a separate computer software. In the second step, hyperparameters (typically catchability and mortality parameters). The likelihood term in step 3, $p(D^N|N^j; \theta)$, is described in Section 3.4. Note that this procedure has a generic structure, hence many different specifications are possible for all the terms $p(C|D^C)$, $p(N|C; \theta)$ and $p(D^N|N; \theta)$. The specific choices of $p(N|C; \theta)$ and $p(D^N|N; \theta)$ are described in Section 3.2. In this paper, $p(C|D^C)$ will be given by the procedure described in ??? while specific choices of $p(N|C; \theta)$ and $p(D^N|N; \theta)$ are described in Section 3.2. Algorithm 1 can be seen as an application of ordinary sequential Monte Carlo (?), and even more so when we consider the time-dynamics of the N process in Section 3.3. Many refinements such as using other proposal distributions for N , resampling steps and so on, are then possible to incorporate.

3.2 State Space Model

In this section, we formulate a State Space model in the context of age structured fisheries data. In general, a state–space modeling framework involves two parallel time series: a state process and an observation process (??). In our case the state process both include abundance N and catch C . However, in algorithm 1, the catch model is implicitly given through the availability of $p(C|D^C)$. Our model specification will therefore be concentrating on the conditional distribution of abundance *given* catch and on the likelihood $p(D^N|N)$.

3.3 The abundance process model

Usually, abundance is structured into age-groups and yearly components with a time-discrete specification, typically derived from continuous time differential equations (??). Here, we also include season, in order to take into account that both fisheries dependent and fisheries independent data are collected at relatively narrow time-windows.

Let Y, A and S respectively denote the current total number of years, number of ordinal fish age-groups and number of fishing seasons per year in a given fisheries time series survey. Let $N = \{N_{y,s,a}, y = 1, \dots, Y, s = 1, \dots, S, a = 1, \dots, A\}$ where $N_{y,s,a}$ denotes stock abundance of age a at the start of season s in year y and similarly $C = \{C_{y,s,a}, y = 1, \dots, Y, s = 1, \dots, S, a = 1, \dots, A\}$ where $C_{y,s,a}$ denotes the total catch-at-age of age a fish at the start of season s in year y . For both catch and abundance, age category A denotes a plus group.

Introduce further $M = \{M_{y,a,s}, y = 1, \dots, Y, s = 1, \dots, S, a = 1, \dots, A\}$ where $M_{y,s,a}$ is the total non-fishing mortality among age a fish over season s of year y . Although different forms of the population cohort equation are frequently encountered in the literature (??), here we consider the model

$$N_{y,s,a} = N_{y,s,a-1} e^{-M_{y,s,a-1}} - C_{y,s,a-1}, \quad (2)$$

$y = 1, \dots, Y, s = 2, \dots, S, a = 1, \dots, A$, for within year transitions, and

$$N_{y+1,a,1} = N_{y,a-1,S} e^{-M_{y,a-1,S}} - C_{y,a-1,S}, \quad (3)$$

$y = 1, \dots, Y, a = 2, \dots, A - 1$ for between year transitions. Since the last age group usually involves both between and within cohort transitions, we use

the following slightly modified between year equation

$$N_{y+1,A,1} = N_{y,A-1,S} e^{-M_{y,A-1,S}} - C_{y,A-1,S} + N_{y,A,S} e^{-M_{y,A,S}} - C_{y,A,S}, \quad (4)$$

$y = 2, \dots, Y$. These equations can be seen as a discretization of a continuous time model with a constant natural mortality rate within each season and all catch performed in the end of each season. Given the mortalities M , the process is deterministic, but stochasticity is incorporated through distributional assumptions about M , similar to e.g. ??, see below.

Since abundances need to be non-negative, constraints on the $N_{y,s,a}$'s and thereby indirectly on the $M_{y,s,a}$'s needs to be incorporated. Simplification of these constraints are obtained by considering the dynamic models above backwards, giving equations

$$N_{y,s,a-1} = [N_{y,s,a} + C_{y,s,a-1}] e^{M_{y,s,a-1}} \quad (5)$$

for $y = 1, \dots, Y$, $s = 2, \dots, S$, $a = 1, \dots, A$,

$$N_{y,a-1,S} = [N_{y+1,a,1} + C_{y,a-1,S}] e^{M_{y,a-1,S}}$$

for $y = 1, \dots, Y$, $a = 2, \dots, A - 1$, and

$$N_{y,A-1,S} = [N_{y+1,A,1} + C_{y,A-1,S} - (N_{y,A,S} e^{-M_{y,A,S}} - C_{y,A,S})] e^{M_{y,A-1,S}}. \quad (6)$$

The stochasticity in the model is then incorporated by making distributional assumptions about the $M_{y,s,a}$'s. Here, we assume constant annual natural mortality $M_{y,a}$, which is uniformly distributed across season within a year such that $M_{y,s,a} = \frac{1}{S} M_{y,a}$, where $M_{y,s,a}$ denotes mortality attributed to season s in year y . Following ?, we further assume the hierarchical mortality model

$$\log(M_{y,a}) = m_a + U_y + V_{y,a}, \quad (7)$$

where $m_a = \log(0.2)$ and

$$U_y \sim N(0, \xi), \quad V_{y,a} \sim N(0, \phi)$$

Millar and Meyer (2000) placed priors on recruitment in all years and on numbers-at-age all ages at year one. In a slight departure, here we specify

priors on numbers-at-age in the last year and for the terminal age at each year. We assume

$$N_{Y,a,S} = \frac{C_{Y,a,S}}{F_{Y,a,S}} e^{M_{Y,a,S}}, \quad (8)$$

$$N_{y,A,S} = \frac{C_{y,A,S}}{F_{y,A,S}} e^{M_{y,A,S}} \quad (9)$$

where the fishing Mortality rates $F_{y,A,S}$ are given priors

$$F_{y,A,S} \sim \text{Uniform}[0, 1] \quad (10)$$

for year Y and age A , where S denotes the number of seasons.

In equation (4) and (6), we account for the fact that age A stock size in year y is a mixture of survivors from age groups $A - 1$ and A in the previous year. We introduce a random mixing variable ω_y to account for indistinguishable contributions of survivors from different cohorts to age A stocks. Accordingly, we reparametrize equation (6) as

$$N_{y,A-1,S} = [\omega_y N_{y+1,A,1}^0 + C_{y,A1,S}] e^{M_{y,A1,S}}, \quad (11)$$

where $N_{y+1,A,1}^0 \sim \text{Uniform}[0, N_A^0 + C_A^0]$, for $y = 1, \dots, Y$. Equation (10), follows by substituting

$$N_{y,A,S} = [(1 - \omega_y) N_{y+1,A,1}^0 + C_{y,A,S}] e^{M_{y,A,S}},$$

in equation (6). Note that this implies that

$$\omega_y = \frac{N_{y,A-1,S} e^{-M_{y,A1,S}} - C_{y,A1,S}}{N_{y,A-1,S} e^{-M_{y,A1,S}} - C_{y,A1,S} + N_{y,A,S} e^{-M_{y,A,S}} - C_{y,A,S}}$$

the fraction from age group $A - 1$ that contributes to $N_{y+1,A,S}$. Hence, it is sufficient to specify a prior on y instead of placing priors on $N_{y,A,S}$. We consider a skewed beta prior

$$\omega_y \sim \text{Beta}(\alpha, \beta), \quad (12)$$

for $y = 1, \dots, Y$, where the hyper-parameters are chosen to allow for greater contribution from survivors of year $y1$ cohort $A1$ stocks relative to cohort $A+$ stocks, where $A+$ denotes indistinguishable age A or older stocks at year $y1$.

3.4 Likelihood for survey indices

Survey indices of relative abundance are used as the primary source of observations in this paper. We assume that the observed catches are related to the population process through a posterior catch distribution. Let $I_{y,a}$, denote the survey relative abundance index and $N_{y,a}^I$, be the numbers-at-age at the time of survey respectively. The log-normal model is frequently used as a sensible choice for modeling $I_{y,a}$, so that:

$$I_{y,s^*,a} = q_a N_{y,s,a} e^{\varepsilon_{y,s^*,a}}, \quad (13)$$

where q_a denotes age specific catchability, assuming the observed relative abundance is proportional to the numbers-at-age at the time of survey. The subscript $s^I \in \{1, \dots, S\}$, denotes the season at which the survey indices are observed. Here we assume indices are only available at one (the same) season per year, but the framework is easily extended to more general cases.

The fundamental challenge of this formulation is often how to estimate the unknown catchability nuisance parameters. We circumvent this problem by considering the *relative indices* $\tilde{I}_{y,s,a^I} \approx I_{y,a} / \sum_{j=1}^Y I_{j,a}$ which we assume are related to the *relative abundances* $\tilde{N}_{y,s,a^I} \approx N_{y,a} / \sum_{j=1}^Y N_{j,a,s^I}$ through the model

$$\tilde{I}_{y,s^*,a} = \tilde{N}_{y,s^*,a} e^{\tilde{\varepsilon}_{y,s^*,a}}, \quad (14)$$

which does not explicitly depend on the estimates of the unknown nuisance parameters. Here, we assume $\varepsilon_{y,s^*,a} \sim N(-0.5\sigma_{y,s^*,a}^2, \sigma_{y,s^*,a}^2)$ so that $\tilde{I}_{y,s^*,a}$ is an unbiased estimate of $N_{y,s^*,a}$. Given bootstrap replicates of the survey indices, we specify $\tilde{I}_{y,s^*,a}$ as the mean of these replicates and $\sigma_{y,s^*,a}$ as their standard deviation.

3.5 Observation model

In this paper, we use indices of relative abundance from research surveys at the Barent sea as the primary source of Fisheries independent data. The other source of observations are observed catches data which are usually obtained by randomly collecting samples from landings by commercial fleets. Although, catch data is often used as the primary source of fisheries dependent data in most applications, here it constitutes a realization of the true but unobserved catch which is related to the population process through a

posterior catch distribution. Hence, fisheries dependent information is indirectly conveyed through the posterior catch distribution denoted by $p(C|D^C)$, where DC denotes the observed catch data. We use samples generated from the posterior distribution to estimate true catch.

3.5.1 Likelihood for survey indices

Let $I_{y,a}$ denote the survey relative abundance index and $N_{y,a}$ be the numbers-at-age at the time of survey respectively. The log-normal model is often considered as a sensible choice for modeling $I_{y,a}$ (Millar and Meyer, 2000). We similarly assume here that the survey indices of relative abundance are related to the population size at the time of survey through:

$$I_{y,a} = q_a N_{y,a} e^{\varepsilon_{y,a}}, \quad (15)$$

where q_a denotes age specific catchability, assuming the observed relative abundance is proportional to the numbers-at-age at the time of survey and $\varepsilon_{y,a}$ is a normally distributed mean zero random variable with variance σ_I^2 , assumed independent between ages and years (Aanes et al, 2007).

The fundamental challenge of this formulation is often how to estimate the unknown catchability nuisance parameters. Assuming constant age-specific catchabilities between surveys, we circumvent this problem by considering the normalized *relative indices* $\tilde{I}_{y,a} = I_{y,a} / \sum_{j=1}^Y I_{j,a}$. In Appendix A, we derive the likelihood for $I_{y,a}$ under the assumptions of a log-normal model. In this model, we show that the *relative indices* are related to *relative abundances* $\tilde{N}_{y,a,s} \approx N_{y,a,s} / \sum_{j=1}^Y N_{j,a,s}$ through

$$I_{y,a} = N_{y,a,s} e^{\varepsilon_{y,a}}, \quad (16)$$

which does not explicitly depend on the estimates of the unknown nuisance parameters, where the subscript $s \in \{1, \dots, S\}$ indicates that the relative abundances are computed for the season in which the survey indices are observed. Here we assume indices are only available at one (the same) season per year, but the framework can be easily extended to more general cases. By the independency of $I_{y,a}$ observations, we assume $I_{y,a}$ are independent between years and ages and lognormally distributed. Here, we assume $\varepsilon_{y,a} = \tilde{\varepsilon}_{y,a} 0.5\sigma_{\tilde{I}}^2$, where $\tilde{\varepsilon}_{y,a}$ is normally distributed with mean zero and variance $\sigma_{\tilde{I}}^2$, so that $\tilde{I}_{y,a}$ is an unbiased estimate of $N_{y,a,s}$. In this paper, we calibrate the prior for \tilde{I} based on the variability in the bootstrap distribution of the observed survey indices.

3.6 Priors

The model requires specification of prior distributions for parameters (m_a, ξ, ϕ) describing the variability in mortality. The lognormal distribution is usually considered a reasonable choice for modeling mortality. We assume quantities. We specify a hierarchical prior model

$$\xi \sim \text{IG}(\alpha_1, \beta_1), \phi \sim \text{IG}(\alpha_2, \beta_2),$$

In order to make the results comparable to ?, we make similar assumptions by setting $\alpha_1 = \alpha_2 = 3.78$ and $\beta_1 = \beta_2 = 0.478$ respectively (more details can be found in ? Appendix B). We set $m_a = \log(0.2)$ and choose a large value for σ^2 to represent our little belief in the fixed VPA value.

4 Application

The proposed method was applied to data for Northeast Arctic cod from the Norwegian section of the Barents Sea for the period 1985-2003. In this analysis, we use catch data on biological samples and reported landings data obtained from commercial fleets which was provided by the Norwegian management authorities (?). To ensure the catch data is representative of the Barents Sea winter survey sampling frame, data believed to represent coastal cod were omitted from the analysis.

A Bayesian hierarchical model developed in ? was fitted to the resulting data according to the standard definition based on area and season to estimate catch-at-age. The model was implemented in ECA, a comprehensive computer software recently developed by the Norwegian Computing Center, the Institute of Marine Research and the University of Oslo in Norway. This program was then used to generate independent posterior catch-at-age samples from the full marginal catch distributions (??).

The other data used in this analysis are the annual research survey indices of relative abundance. The survey indices data are provided by the Norwegian Institute of Marine Research (IMR). The IMR, has been conducted a Winter Survey annually in January, February and March since 1981 at the Barents Sea (?). Independent relative abundance indices were obtained as the mean catch per trawled distance (?). To preserve the original survey design considerations, stratified bootstrap samples of observed abundance

indices were generated by resampling the observed data according to the survey sampling scheme.

The Norwegian Winter Survey only covers the section of the Barents Sea excluding the Russian waters. In this analysis however, the results have been scaled-up to the entire Barents Sea for comparison purposes with results in ? as well as the VPA analysis by the Arctic Fisheries Working Group respectively. A scaling factor was computed based on the data used in ? assuming total Norwegian catches are directly proportional to overall Winter catches from the Barents Sea.

The analysis results are summarized in Table 1. The estimated annual total catch based on our model is comparable to the VPA estimate as reported in the 2004 Report by the Arctic Fisheries Working Group (Figure 1). In the Figure, we compare our method to results of ? relative to the VPA estimate as gold standard. The results show that the proposed method performs consistently better than the ? estimator as evidenced by tighter estimated 95% credibility sets each year. It can be inferred from the figure that unlike our method, the ? approach appears to consistently overestimate total stock abundance.

The results further illustrates that although our estimate is virtually indistinguishable from the VPA estimate, earlier in the time series, it exhibits increasingly greater variability in later years. This pattern may be associated with the backward simulation in the dynamic model such that the influence of initial state prior distributions diminishes back in time from year Y as we incorporate more data.

5 Conclusions and discussion

Discuss the implications of doing modelling backwards, in particular what it means that we assume all mortalities to be independent of later (in time) abundances. This is in contrast to e.g. ? in which modelling is performed forwards.

Also discuss alternative modelling of likelihood for indices taking a more proper Bayesian approach.

Compared to ?, the proposed method provides a direct estimate of stock abundance which does not involve the estimation of intermediate parameters, achieves tighter intervals of the 95% credible set (included later) and is

Table 1: The DSSM estimated VPA-type (numbers-at-age) population size of ages 3 – 13⁺ years for the annual Barents Sea Winter survey for Northeast Arctic cod (*Gadus morhua*) for the years 1985–2003.

Year	$N_{y,a} (\times 10^6)$											N_y (Total)
	3	4	5	6	7	8	9	10	11	12	13 ⁺	
1985.0	696.9	338.0	103.2	44.2	19.5	6.3	3.0	1.9	0.5	0.4	0.5	1214.3
1986.0	1250.1	584.6	255.0	63.2	19.3	6.8	1.9	1.2	0.9	0.3	0.6	2183.8
1987.0	563.2	1006.6	403.9	130.3	27.5	6.6	2.0	0.8	0.3	0.3	0.4	2141.9
1988.0	261.0	393.7	619.8	212.8	43.8	9.6	2.5	0.7	0.2	0.1	0.3	1544.7
1989.0	180.3	176.5	221.1	284.5	89.1	13.2	4.0	0.7	0.3	0.1	0.2	970.1
1990.0	280.4	137.4	132.3	100.1	137.6	31.7	5.5	1.9	0.4	0.2	0.2	827.6
1991.0	578.6	210.3	103.6	91.7	61.8	89.1	18.2	3.9	1.4	0.2	0.2	1159.0
1992.0	821.3	357.4	149.4	68.8	48.7	31.9	49.3	9.1	2.8	1.1	0.5	1540.3
1993.0	864.3	620.5	259.9	99.9	37.5	23.4	15.1	27.3	5.3	2.1	1.2	1956.7
1994.0	563.8	672.1	481.8	148.6	48.8	17.6	11.2	6.1	11.2	3.0	2.2	1966.3
1995.0	315.1	410.9	422.1	234.5	63.6	12.9	4.8	3.1	2.0	2.9	3.3	1475.2
1996.0	264.3	232.8	267.5	241.0	107.0	19.3	4.6	1.6	0.9	0.8	3.1	1142.9
1997.0	478.3	195.0	158.3	165.4	117.0	41.1	6.5	1.9	0.7	0.4	1.8	1166.4
1998.0	763.6	390.9	133.8	76.9	69.5	44.0	10.6	1.6	0.5	0.2	0.7	1492.4
1999.0	576.4	546.7	243.8	61.8	28.4	25.3	12.4	2.5	0.6	0.3	0.7	1498.9
2000.0	628.9	473.7	375.9	116.2	24.1	10.7	6.8	2.6	0.5	0.3	0.8	1640.5
2001.0	347.4	455.9	344.4	229.0	49.9	9.3	2.6	1.9	0.6	0.1	0.7	1441.9
2002.0	440.5	293.3	345.0	218.5	129.1	21.5	3.3	0.8	0.6	0.2	0.4	1453.3
2003.0	25.1	373.8	215.5	201.6	108.5	65.0	8.7	1.4	0.3	0.3	0.2	1000.4

relatively computationally fast. Therefore, the method provides a useful and robust extension of the work of ?.

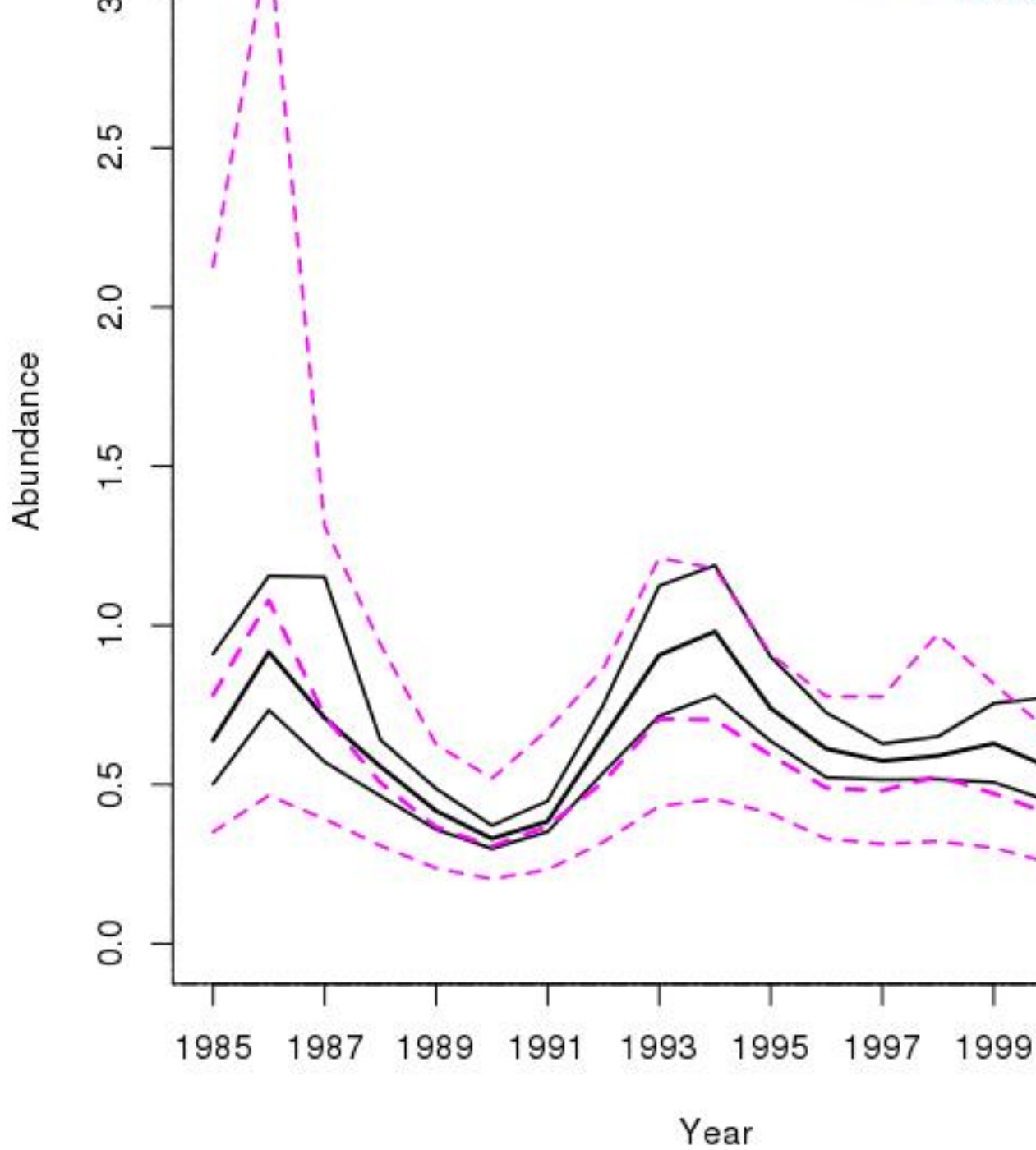


Figure 1: The 1985–2003 DSSM estimated Northeast Arctic cod (*Gadus morhua*) stock size assuming a minimum fishing age of 3-years (solid line). The dashed line (green) gives the estimates from ?. The dotted and broken lines gives the VPA and XSA estimates from the Arctic Fisheries Working Group respectively.

A Importance Sampling

Although we can simulate the states N^j , $j = 1, \dots, B$, from the joint distribution of N and C given observed catch data through the dynamic process model described in section (3.2), catch data alone do not provide sufficient information to uniquely determine trends in stock abundance (?). We can use Sequential Importance Sampling (SIS) to combine the fisheries dependent catch data with survey indices data which provides fisheries independent information on stock dynamics.

Let $C^j \sim p(C|D^C)$ be a sample from the posterior catch distribution conditional on observed catch data and $N^j \sim p(N|C)$ be a simulated state process conditional on sample C^j , for $j = 1, \dots, B$. Define $N = \{N_0|N_{0-}\}$, where $N_0 = \{N_{Y,a,S}^0, N_{y,A,S}^0\}$ denotes the initialized rows and columns of N and N_{0-} denotes N excluding the elements of N_0 respectively. Suppose that it is not possible to sample directly from the target density $p(N|C, D^N)$, but we can easily draw samples from some proposal density $q(N|C, D^N)$. Then, we can construct importance weights as

$$\begin{aligned} W^j &= p(D^N|N^j) p(N^j|C^j) / q(N^j|C^j) \\ &= \frac{q(D^N|N^j) q(N_{0-}^j|N_0^j, C^j, D^N) q(N_0^j|C^j)}{q_0(N_0^j|C^j, D^N) q(N_{0-}^j|N_0^j, C^j, D^N)} \\ &= p(D^N|N^j) p(N_0^j|C^j) / q_0(N_0^j|C^j, D^N) \end{aligned} \quad (17)$$

for $j = 1, \dots, B$, where the third equality follows by letting, $p(N_{0-}^j|N_0^j, C^j, D^N) = q(N_{0-}^j|C^j, D^N)$. Further simplification may be achieved by choosing a flat prior for $p(N_0^j|C^j)$. Normalized weight can be computed as

$$w^j = \frac{e^{\{\log(W^j) - \max[\log(W^j)]\}}}{\sum_{j=1}^B e^{\{\log(W^j) - \max[\log(W^j)]\}}}, \quad (18)$$

for $j=1, \dots, B$. The expected total abundance estimate \hat{N} is obtained as a function of the normalized importance sampling weights as

$$E[N|D^N, D^C] = \sum_{j=1}^B w^j f(N^j; C^j), \quad (19)$$

where w^j are normalized W^j weights for $j = 1, \dots, B$.