

# Robust state space models for estimating fish stock maturities

Ximing XU<sup>1\*</sup>, Eva CANTONI<sup>2</sup>, Joanna Mills FLEMMING<sup>1</sup> and Chris FIELD<sup>1</sup>

<sup>1</sup>*Department of Mathematics and Statistics, Dalhousie University, 6316 Coburg Road, PO Box 15000, Halifax, Nova Scotia, Canada B3H 4R2*

<sup>2</sup>*Research Center for Statistics and Geneva School of Economics and Management, University of Geneva, Geneva, Switzerland*

*Key words and phrases:* Automatic Differentiation Model Builder; Laplace approximation; robustness; state space models; weighted likelihood.

*MSC 2010:* Primary 62F35; secondary 62M10.

*Abstract:* Here, we formulate robust state space models (SSMs) and develop inference tools in the context of fisheries science and management. Our prototype model concerns the maturity of fish by age over time, knowledge of which is fundamental to understanding the dynamics and productivity of fish stocks, a key component in fish stock assessment. Our SSM incorporates dynamics over time and yields robust estimates of the proportion of fish mature at various ages for a collection of cohorts of interest. The estimates are obtained using an iterative weighted likelihood where the high-dimensional unobserved dynamics of the SSM and the optimization over the fixed parameters are handled with the Automatic Differentiation Model Builder. The data used to both demonstrate and validate our new approach comes from the annual sampling program of Fisheries and Oceans Canada, the primary scientific and regulatory body responsible for stock assessments. In addition, we carried out a comprehensive simulation study to demonstrate the robustness of our approach to realistic contamination. *The Canadian Journal of Statistics* 43: 133–150; 2015 © 2015 Statistical Society of Canada

*Résumé:* Les auteurs conçoivent un modèle à espace d'états robuste et élaborent des outils d'inférence dans le contexte des sciences halieutiques et de la gestion des pêches. Leur prototype de modèle porte sur la maturité des poissons selon l'âge en fonction du temps dont la connaissance est fondamentale afin de comprendre la dynamique et la productivité des stocks halieutiques, des éléments clés de leur évaluation. Le modèle à espace d'états proposé incorpore la dynamique temporelle et produit des estimés robustes de la proportion de poissons matures à différents âges pour un ensemble de cohortes d'intérêt. Les auteurs obtiennent leurs estimés par la vraisemblance pondérée itérée où la dynamique non observée en haute dimension du modèle à espace d'états et l'optimisation des paramètres fixes sont gérées à l'aide du logiciel *Automatic Differentiation Model Builder*. Ils illustrent et valident leur approche à l'aide de données provenant du programme d'échantillonnage annuel de Pêches et Océans Canada, le principal organisme de réglementation responsable de l'évaluation des stocks halieutiques. Les auteurs ont également procédé à une simulation d'envergure démontrant la robustesse de leur approche dans des scénarios de contamination réalistes. *La revue canadienne de statistique* 43: 133–150; 2015 © 2015 Société statistique du Canada

## 1. INTRODUCTION

In the management of fisheries, fish stock assessment plays an important role. A fish stock is a group of fish of the same species that live in the same geographic area and mix enough to breed with each other when mature. Stock assessments provide fisheries managers with the information

---

\* Author to whom correspondence may be addressed.  
E-mail: Ximing.Xu@dal.ca

that is used in the regulation of a fish stock. In the marine waters of Canada, Fisheries, and Oceans Canada (DFO) is the primary scientific and regulatory body responsible for stock assessment. As such, DFO runs annual biological sampling programs collecting data on the number of fish examined as well as the number found to be mature for a wide range of ages in various stocks. Knowledge of maturity is fundamental to understanding the dynamics and productivity of fish stocks and is key to estimating spawning stock biomass (the proportion of the stock capable of reproducing), as an example.

Data collected by DFO may exhibit considerable variability and yet fisheries scientists must rely on them to provide sound management advice. Models have been developed to describe how a fish stock changes over time in response to fishing and other sources of mortality but these are only approximations of reality. Fortunately, models that explicitly include both process error in the dynamics and observation error in the data are increasingly being used in fisheries science. Such models are often referred to as “state space models” (SSMs), and are considered by many to be the best practice for fish stock assessment (e.g., Schnute, 1994; Meyer & Millar, 1999; Lewy & Nielsen, 2003; Aanes et al., 2007; Gudmundsson & Gunnlaugsson, 2012) and more generally in ecological modelling (e.g., Buckland et al., 2007; Pedersen et al., 2011; Minto et al., 2013).

Here, we develop robust SSMs to be used for fisheries science and management. They allow for both errors and misspecification in the observation process and in the state/dynamic process. These robust statistical methodologies are essential because the data being modelled in fisheries science and management may be subject to large measurement error and the state and observation processes are only approximations with the consequence that there may well be outlying observations or deviating substructures. Much of the recent research on robustness for SSMs (Roh et al., 2013; Calvet, Czellar, & Ronchetti, 2014) has focussed on building in robustness against errors in the observation process without much attention to robust estimation of the parameters. The parameters are important here because the interest is in estimates of the proportion mature for a specific year and cohort. Our method adds robustness weights at the likelihood level as an appropriate way of allowing for deviations in the observation process. There has been almost no research whatsoever addressing robustness in the state process. Since the state equation involves errors most often modelled as Gaussian, it is important to ensure robustness against error distributions in the neighbourhood of the core distribution as our model allows.

Our robust formulation is based on weighting the likelihood directly rather than weighting the score function as is more usual in robust statistics (Huber, 1981; Hampel et al., 1986; Heritier et al., 2009). Specially, we build on the proposal by Field & Smith (1994) in which observations are downweighted if they fall in the upper or lower tails of the distribution based on the current parameter estimates. Dupuis & Morgenthaler (2002) proposed an approximated bias correction based on the weighted score function as in Field & Smith (1994). In contrast our approach maximizes the weighted likelihood subject to constraints on the weights and avoids the use of the score function (see Section 3). Our decision to weight at the likelihood level (rather than at the score function level) was based on the availability of software like Automatic Differentiation Model Builder (ADMB, Fournier et al., 2012) which can maximize complex likelihoods, and, also the fact that for the model we develop, the score function is complex and actually more difficult to work with than the likelihood itself. We demonstrate the effectiveness of this method for robustly estimating fish stock maturities.

ADMB is a powerful software package for the development of state-of-the-art nonlinear statistical models. ADMB utilizes reverse mode automatic differentiation to provide fast optimization and, as such, is highly suitable for fitting complex and highly parameterized models. In addition, for those models that include random effects it utilizes the Laplace approximation to integrate out the random effects. Advantages of ADMB have recently been documented in a large study of diverse ecological models (Bolker et al., 2013), and especially for SSMs (Pedersen et al., 2011). In

our model ADMB has several advantages including its speed which enables us to use a bootstrap method for constructing confidence intervals and to do the necessary iterative computations along with the accuracy of the optimization achieved through the use of the automatic differentiation feature.

Making tools that reliably estimate fish stock maturities (and other population parameters) available to fisheries scientists both within and outside of Canada will no doubt help to sustain some of our most valuable renewable resources and contribute to more prosperous fisheries economies in the future. To our knowledge we are the first to consider these issues within a formal robustness SSM framework.

Section 2 gives an introduction to our SSM for estimating fish stock maturities as well as the inference for this model. In Section 3, we present the details of our weighted likelihood and the iterative algorithm developed for its evaluation. The performance of our approach is demonstrated and compared with the classical likelihood approach through simulation study in Section 4 and with real data in Section 5. We conclude with a discussion and directions for future research in Section 6.

## 2. METHODOLOGY

A SSM involves two random processes: an observation density or probability distribution that describes the behaviour of the observed phenomenon at a given time  $t$ , and a state density or probability distribution that describes the (hidden/latent) dynamics through time.

### 2.1. Model Formulation

We wish to formulate a SSM that allows us to reliably estimate the proportion of fish mature in each cohort of a stock through time (along with other relevant population parameters which model the dynamics over time). Each year fish are sampled to determine their age and maturity status (a binary variable). The cohort is the birth year which is the sample year minus the age of the fish (e.g., a 5-year-old fish assessed in 2000 belongs to the 1995 cohort). In a time series of samples, early cohorts are those older fish at the start of sampling, and recent cohorts are those young fish sampled at the end of the time series (see Cadigan, Morgan, & Brattey, 2014, for a graphical illustration). We define  $Y_{ac}$  to be the number of mature females out of a total number of sampled female fish  $n_{ac}$  for cohort  $c$  and age  $a$ . We denote the first cohort as  $C_1$  and the last cohort as  $C_l$ . We assume age  $a$  ranges from 1 to  $A$  with the mean age of the observed fish denoted by  $\bar{a}$ . We assume that  $Y_{ac}$  follows a binomial distribution with parameters  $p_{ac}$  and  $n_{ac}$ . The observed values are denoted by  $y_{ac}$ . We assume that, within a cohort, maturity increases monotonically with age, and as such the linear logistic model is a simple option. Furthermore, the maturity curves are assumed to be similar for adjacent cohorts but to differ more for distant cohorts. To account for this correlation, the AR(1) processes are used to model the cohort-specific intercepts and slopes. That is, following Cadigan, Morgan, & Brattey (2014), we model the probability  $p_{ac}$  of observing a mature fish in cohort  $c$  of age  $a$  as

$$p_{ac} = \exp(\eta_{ac})/[1 + \exp(\eta_{ac})], \tag{1}$$

with  $\eta_{ac} = \beta_0 + \beta_1(a - \bar{a}) + \tilde{\beta}_{0c} + \tilde{\beta}_{1c}(a - \bar{a})$ , where  $\beta_0$  and  $\beta_1$  are fixed parameters and  $\tilde{\beta}_{0c}$  and  $\tilde{\beta}_{1c}$  (for  $c = C_1, \dots, C_l$ ) are dynamic parameters described by the following AR(1) state process

$$\tilde{\beta}_c = \Phi \tilde{\beta}_{c-1} + \epsilon_c \text{ for } c = C_2, \dots, C_l, \tag{2}$$

where  $\tilde{\beta}_c = (\tilde{\beta}_{0c}, \tilde{\beta}_{1c})^T$ ,  $\Phi = \begin{pmatrix} \varphi_0 & 0 \\ 0 & \varphi_1 \end{pmatrix}$ , and  $\epsilon_c = (\epsilon_{0c}, \epsilon_{1c})^T$ . We assume that the errors  $\epsilon_c$  are independent with bivariate normal distribution  $N(\mathbf{0}, \Sigma)$  where  $\mathbf{0} = (0, 0)^T$  and  $\Sigma = \begin{pmatrix} \sigma_0^2 & 0 \\ 0 & \sigma_1^2 \end{pmatrix}$ . As pointed out by Cadigan, Morgan, & Bratney (2014),  $\tilde{\beta}_{0c}$  and  $\tilde{\beta}_{1c}$  will usually correlate with each other, but their correlations will be greatly reduced after standardizing the age and the assumption of independence works well in practice. It is easy to show that the stationary distribution of  $\tilde{\beta}_{0c}$  is  $N\left(0, \frac{\sigma_0^2}{1-\varphi_0^2}\right)$  and the stationary distribution of  $\tilde{\beta}_{1c}$  is  $N\left(0, \frac{\sigma_1^2}{1-\varphi_1^2}\right)$ . Note that these normal distributions can be replaced by a  $t$ -distribution or Huber's least favourable distribution. Throughout the paper we assume that  $\tilde{\beta}_{0C_1}$  and  $\tilde{\beta}_{1C_1}$  follow the stationary distributions.

We define  $y_c = (y_{1c}, \dots, y_{Ac})^T$ ,  $\tilde{\beta}_c = (\tilde{\beta}_{0c}, \tilde{\beta}_{1c})^T$ , for  $c = C_1, \dots, C_l$  and  $\theta = (\beta_0, \beta_1, \varphi_0, \varphi_1, \sigma_0^2, \sigma_1^2)^T$ . The observation density  $P(y_c | \tilde{\beta}_c, \theta)$  is then

$$\begin{aligned} P(y_c | \tilde{\beta}_c, \theta) &= \prod_{a=1}^A P(y_{ac} | \tilde{\beta}_c, \theta) \\ &= \prod_{a=1}^A \binom{n_{ac}}{y_{ac}} p_{ac}^{y_{ac}} (1 - p_{ac})^{n_{ac} - y_{ac}} \\ &= \prod_{a=1}^A \exp \left\{ y_{ac} \eta_{ac} - n_{ac} \log[1 + \exp(\eta_{ac})] + \log \binom{n_{ac}}{y_{ac}} \right\}. \end{aligned} \tag{3}$$

The marginal likelihood of the SSM model defined by Equations (1)–(3) is

$$\begin{aligned} L(\theta; y) &= \int \dots \int \prod_{c=C_1}^{C_l} P(y_c | \tilde{\beta}_c, \theta) P(\tilde{\beta}_c | \tilde{\beta}_{c-1}, \theta) d\tilde{\beta}_{C_1}, \dots, \tilde{\beta}_{C_l} \\ &= \int \dots \int \prod_{c=C_1}^{C_l} \left[ \prod_{a=1}^A P(y_{ac} | \tilde{\beta}_c, \theta) \right] P(\tilde{\beta}_c | \tilde{\beta}_{c-1}, \theta) d\tilde{\beta}_{C_1}, \dots, \tilde{\beta}_{C_l}, \end{aligned} \tag{4}$$

where we define  $P(\tilde{\beta}_{C_1} | \tilde{\beta}_{C_1-1}, \theta) = P(\tilde{\beta}_{C_1} | \theta)$ , the stationary density of  $\tilde{\beta}_c$ .  $P(y_c | \tilde{\beta}_c, \theta)$  is as given in Equation (3) and  $P(\tilde{\beta}_c | \tilde{\beta}_{c-1}, \theta)$  is the product of two normal densities:  $N(\varphi_0 \tilde{\beta}_{0(c-1)}, \sigma_0^2) \times N(\varphi_1 \tilde{\beta}_{1(c-1)}, \sigma_1^2)$ . Note that the SSM can also be considered as a mixed effects model with  $\tilde{\beta}_c$ 's as random effects.

### 2.2. Inference

Analytical calculation of the marginal likelihood (Eq. (4)) is impossible due to the high dimensional integrals involved (as with most SSMs). Inference for SSMs therefore usually relies on recursive forms of filtering and smoothing the latent states that define the temporal structure of the observations. However, as pointed out recently by Ruiz-Cárdenas et al. (2012), this is unnecessary if the data are fixed, that is, completely available before analysis as is the case here. Hence rather than take a filtering approach we instead prefer a direct approach for evaluating the marginal likelihood and, as such, have implemented our SSM using the ADMB as noted in the Introduction.

ADMB uses the Laplace approximation of the high-dimensional integral (Eq. (4)) and our marginal likelihood is approximated by

$$L(\theta; \mathbf{y}) \simeq \exp \left\{ \log(L(\theta | \mathbf{y}, \tilde{\boldsymbol{\beta}})) \right\} \Big|_{\tilde{\boldsymbol{\beta}}=\hat{\mathbf{b}}(\theta)} - 0.5 \log(|H|) \Big|_{\tilde{\boldsymbol{\beta}}=\hat{\mathbf{b}}(\theta)},$$

where  $L(\theta | \mathbf{y}, \tilde{\boldsymbol{\beta}}) = \prod_{c=C_1}^{C_I} P(y_c | \tilde{\boldsymbol{\beta}}_c, \theta)P(\tilde{\boldsymbol{\beta}}_c | \tilde{\boldsymbol{\beta}}_{c-1}, \theta)$  is the joint likelihood function of the observations  $\mathbf{y}$  and the latent states  $\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\beta}}_1^T, \dots, \tilde{\boldsymbol{\beta}}_C^T)^T$ ,  $\hat{\mathbf{b}}(\theta)$  is the maximizer of  $\log(L(\theta | \mathbf{y}, \tilde{\boldsymbol{\beta}}))$  with respect to  $\tilde{\boldsymbol{\beta}}$ , and  $H$  is the matrix of second derivatives of  $\log(L(\theta | \mathbf{y}, \tilde{\boldsymbol{\beta}}))$  with respect to  $\tilde{\boldsymbol{\beta}}$ . The parameters  $\theta$  are estimated by maximizing the Laplace approximated marginal likelihood and the predicted random effects (latent states)  $\tilde{\boldsymbol{\beta}}$  are obtained by maximizing the joint likelihood  $L(\hat{\theta} | \mathbf{y}, \tilde{\boldsymbol{\beta}})$ , given the parameter estimates  $\hat{\theta}$ . More details on the two-stage parameter estimation approach can be found in Fournier et al. (2012).

The main quantities of interest in this paper, the proportions mature,  $p_{ac}$ , can be estimated as a function (Eq. (1)) of the estimated parameters and random effects.

### 3. A WEIGHTED LIKELIHOOD APPROACH TO ROBUSTNESS

We introduce robustness at the likelihood level as a mechanism for allowing for deviations in the observation process. More precisely, the likelihood (4) is modified in the following way

$$WL(\theta; \mathbf{y}) = \int \dots \int \prod_{c=C_1}^{C_I} \left[ \prod_{a=1}^A P(y_{ac} | \tilde{\boldsymbol{\beta}}_c, \theta)^{w_{ac}(y_{ac} | \tilde{\boldsymbol{\beta}}_c, \theta)} \right] P(\tilde{\boldsymbol{\beta}}_c | \tilde{\boldsymbol{\beta}}_{c-1}, \theta) d\tilde{\boldsymbol{\beta}}_{C_1}, \dots, \tilde{\boldsymbol{\beta}}_{C_I}, \quad (5)$$

where the weights are based on those originally proposed by Field & Smith (1994), namely

$$w_{ac}(y_{ac} | \tilde{\boldsymbol{\beta}}_c, \theta) = \begin{cases} P(Y_{ac} \leq y_{ac} | \tilde{\boldsymbol{\beta}}_c, \theta)/p & \text{if } P(Y_{ac} \leq y_{ac} | \tilde{\boldsymbol{\beta}}_c, \theta) < p \\ P(Y_{ac} \geq y_{ac} | \tilde{\boldsymbol{\beta}}_c, \theta)/p & \text{if } P(Y_{ac} \geq y_{ac} | \tilde{\boldsymbol{\beta}}_c, \theta) < p \\ 1 & \text{otherwise.} \end{cases} \quad (6)$$

The value of  $p$  is usually set to be a small number between 0.01 and 0.1. With this definition of the weights, the observations in the upper and lower  $100p$  percentile will be downweighted even if they do not violate the model specification and in general, the estimation based on the resulting weighted likelihood will not be Fisher consistent unless  $p \rightarrow 0$ . Field & Smith (1994) also pointed out that the solution to the weighted likelihood may not be unique and suggested choosing the one which maximizes the sum of the weights. In particular we have to avoid the situation in which the majority of the observations have zero weight. To achieve consistent parameter estimation under the core model we propose the following adjustment to the weights defined by the Equation (6) and develop an iterative algorithm with which to maximize the adjusted-weighted likelihood.

Suppose that the number of observations with weights smaller than one according to (6) is  $m$  and denote the smaller-than-one weights from smallest to largest as  $w_{ac(1)}, \dots, w_{ac(m)}$ . For  $i = 1$  to  $m$ , the adjusted weights are then defined as

$$\tilde{w}_{ac(i)} = \begin{cases} w_{ac(i)} & \text{if } \frac{1}{i} \sum_{j=1}^i w_{ac(j)} < \gamma, \\ 1 & \text{otherwise.} \end{cases} \quad (7)$$

The tuning parameter  $\gamma$  allows us to control the tradeoff between robustness and efficiency. When  $\gamma = 1$ , the adjusted weights (7) becomes the original weights (6); when  $\gamma = 0$ , the weighted likelihood function (5) reduces to the ordinary unweighted likelihood. The value of  $\gamma$  adopted in

this paper is the expected value of the smallest weight  $w_{ac(1)}$ , which can be calculated based on the following results.

**Theorem 1.** *Under the correctly specified model the weights  $w_{ac} < 1$  defined by (6) are independently and identically distributed as  $\text{Unif}(0, 1)$ .*

*Proof.* Define  $U_{ac} = F(y_{ac}) = P(Y_{ac} \leq y_{ac} \mid \tilde{\beta}_c, \theta)$ . We know  $U_{ac} \sim \text{Unif}(0, 1)$ . It is easy to show that  $1 - U_{ac} = P(Y_{ac} \geq y_{ac} \mid \tilde{\beta}_c, \theta) \sim \text{Unif}(0, 1)$ . Given  $\tilde{\beta}_c$  and  $\theta$ , the  $Y_{ac}$  are independent for all  $a$  and  $c$  and hence all  $U_{ac}$  are independently distributed as  $\text{Unif}(0, 1)$ . We now only need to show that conditioning on  $U_{ac} < p$ , the random variable  $U_{ac}/p$  is distributed as  $\text{Unif}(0, 1)$ :

$$P\left(\frac{U_{ac}}{p} < u \mid U_{ac} < p\right) = P(U_{ac} < pu \mid U_{ac} < p) = \frac{\int_0^{pu} du_{ac}}{\int_0^p du_{ac}} = u, \quad (8)$$

i.e.,  $w_{ac} = U_{ac}/p \sim \text{Unif}(0, 1)$  given  $U_{ac} < p$ . ■

According to the definition of weights (6), the expected number of weights ( $< 1$ ) is  $2pn$  for uncontaminated data where  $n$  is the total number of  $y_{ac}$ 's. From the theory of order statistics, the smallest weight  $w_{ac(1)}$  has Beta distribution with the parameters  $\alpha = 1$  and  $\beta = 2pn$ , and its mean is  $E[w_{ac(1)}] = 1/(2pn + 1)$ .

Under the correctly specified model, as  $n \rightarrow \infty$  (i.e., the number of cohorts  $\rightarrow \infty$  since the age of fish is bounded),  $P\left(w_{ac} < \frac{1}{2pn+1}\right) \rightarrow 0$ . Hence the proportion of downweighted observations will converge to 0 and the asymptotic consistency is achieved. The value  $\gamma = \frac{1}{2pn+1}$  for the adjusted weights (7) is shown to work well both in simulation studies and with real data.

### 3.1. Iterative Algorithm

Since the weighted likelihood is complex with the weights depending on the unknown parameters  $\theta$  and random effects  $\tilde{\beta}$ , we use an iterative procedure in which we optimize the weighted likelihood using ADMB with fixed weights, recompute the weights based on the estimated parameters and iterate. We need to have initial weights to start with and we consider two possible approaches. The first is to use the weights derived from an initial optimization of the (unweighted) likelihood using ADMB. It is well known that likelihood estimation is sensitive to outlying observations and hence the resulting weights (6) may not be reliable for robust estimation. Without contamination, asymptotically this approach will not downweight any of the observations (based on our choice of  $\gamma = 1/(2pn + 1)$ ). In the presence of contamination, there is no guarantee that it will downweight appropriately the deviating data points, so that an alternate set of initial weights robust to outliers is critical for the performance of our iterative algorithm. In this paper, we consider using the robust GLM approach (Cantoni & Ronchetti, 2001) to obtain the initial weights. Through simulation studies we found that it would not work well to apply the robust GLM by cohort especially for the first and last few cohorts which do not contain sufficient information. As a result, we grouped adjacent cohorts together form a "bigger cohort." By doing that we assume the proportions mature-at-age do not vary much within adjacent cohorts (Cadigan, Morgan, & Bratney, 2014). There is a tradeoff between including more information and smoothness within a group, and we found that partitioning the cohorts into 4 or 5 groups usually works well in our data analysis. Our algorithm is therefore as follow:

1. Obtain the initial weights (6) based on the ordinary likelihood function or robust GLM.
2. Calculate the adjusted weights  $\tilde{w}_{ac}$  in (7).
3. Run ADMB on the weighted likelihood (5) with the adjusted weights.

4. Calculate the weights (6) based on the estimates of  $\theta$  and  $\beta_c$  from Step 3.
5. Repeat Steps 2–4 until the change in the logarithm of the weighted likelihood is smaller than a threshold value (0.01 is used in this paper).

We denote the above iterative algorithm with likelihood initial weights by WL\_ml and with robust GLM initial weights by WL\_r.

### 3.2. Choosing Between WL\_ml and WL\_r

WL\_ml and WL\_r commence with different choices for the initial weights and, as such, may produce different final estimates of the proportions mature. In our simulation study the true proportions mature ( $p_{ac}$ ) are known and hence we can calculate the mean of the absolute differences between the true proportions mature and the fitted ones  $\hat{p}_{ac}$ ,  $MAE = \sum_{ac} |\hat{p}_{ac} - p_{ac}|/n$ , which is used to evaluate the quality of the fit. In real data applications,  $p_{ac}$  is unknown and hence we propose to select between WL\_ml and WL\_r based on the  $\alpha$ -trimmed mean of the absolute differences between the observed proportions of maturity  $y_{ac}/n_{ac}$  and the fitted ones  $\hat{p}_{ac}$  excluding the largest 100 $\alpha$  percent of the differences,

$$\widehat{MAE}_{trim} = \frac{1}{n(1 - \alpha)} \sum_{i=1}^{n(1-\alpha)} |\hat{p}_{ac} - y_{ac}/n_{ac}|_{(i)}. \tag{9}$$

We note that in our simulation study when we take  $\alpha = 0.1$  the method with smaller  $\widehat{MAE}_{trim}$  also has smaller  $MAE$  in most cases especially when WL\_ml and WL\_r produce very different values of  $MAE$ . We denote the weighted likelihood approach with smaller  $\widehat{MAE}_{trim}$  as WL\_s.

### 3.3. Bootstrap Confidence Intervals

In order to compute the confidence intervals for  $p_{ac}$  as well as the standard errors for  $\hat{\theta}$ , we use a moving blocks bootstrap method (Künsch, 1989) in an effort to deal with the dependence structure in our model. We denote  $b$  as the block size, and the observations in the cohort  $c$  as  $Z_c$ ,  $c = C_1, \dots, C_l$ . Define  $B_i = (Z_i, \dots, Z_{i+b-1})$  to be the observations in  $b$  consecutive cohorts. The moving blocks bootstrap is based on sampling with replacement from the collection of the overlapping blocks  $\{B_{C_1}, \dots, B_{C_l-b+1}\}$ . The total number of moving blocks is  $N_{mb} = (C_l - b + 1) - C_1 + 1$ . To have the same number of cohorts as the original data, the number of blocks  $k = (C_l - C_1 + 1)/b$  for each bootstrap sample and the  $k$  sampled blocks  $B_1^*, \dots, B_k^*$  are aligned in the order in which they were picked. In this paper, we consider only the simplest bootstrap confidence interval method i.e., the percentile interval for the estimated proportions mature,  $\hat{p}_{ac}$ .

## 4. SIMULATION STUDY

Here, we perform a comprehensive simulation study so as to fully understand the performance of our robust SSM framework for estimating fish stock maturities. To best mimic typical DFO biological sampling data, we consider the years of assessment to be from 1970 to 2009, and the age of the fish to be from 1 to 15. The corresponding cohorts are then from 1955 to 2008. Note that for cohort 2008 we will only have observations for fish of age 1. In each year, we suppose that 100 fish are sampled and classified as mature (or immature) for each age. The total number of observed proportions mature is then  $n = 15 \times 40 = 600$ . The average age of the observed fish is  $\bar{a} = 8$ . The values of the fixed parameters are taken to be the following:  $\beta_1 = 1.2$ , and  $\beta_0 = 4$ ; and the parameters describing the random effects distributions taken to be:  $\varphi_0 = \varphi_1 = 0.8$ ,  $\sigma_0 = 0.8$ , and  $\sigma_1 = 0.2$ . The probability  $p$  for the weights (6) is set to be 0.05 although the proportion

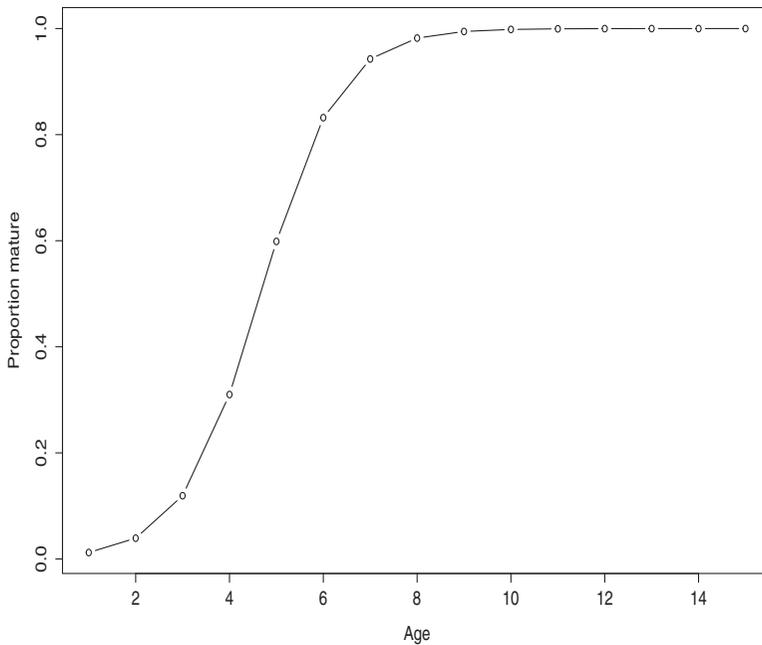


FIGURE 1: True proportions mature for ages 1 through 15.

of contaminated data may be smaller or greater than 0.05. The threshold value for stopping the iterative algorithm is 0.01. Without dynamics i.e.,  $\sigma_0 = \sigma_1 = 0$ , the plot of the proportion mature as a function of age is shown in Figure 1. Notice that fish older than 9 are almost always mature and those younger than 3 are almost always immature.

We now consider four different scenarios which reflect both different mechanisms and different proportions of contamination. For each scenario we generate 300 datasets and compare the weighted likelihood method with the ordinary likelihood method (ML). The initial weights are obtained either based on ordinary likelihood (WL\_ml) or using the robust GLM approach (WL\_r) described earlier. To apply the WL\_r method, the cohorts are partitioned into 4 groups with 15 consecutive cohorts in the first three groups and 9 cohorts in the last group.

#### 4.1. No Contamination

We first consider the case of no contamination. The estimated parameters obtained via the different approaches are shown in Figure 2, parameter estimates are similar and all tend to underestimate  $\sigma_0$ ,  $\sigma_1$ ,  $\varphi_0$ , and  $\varphi_1$ . As we are primarily interested in estimating the proportions mature for each age-cohort combination we plot the mean of the absolute differences between the estimated and true proportions (*MAE*) in Figure 3. As expected the estimated proportions based on the ordinary likelihood are more accurate than those based on the weighted likelihood, but the performance of WL\_ml is very similar to that of ML. If we increase the number of cohorts the difference between WL\_ml and ML will vanish.

#### 4.2. Contamination

There are several types of errors that can occur in the study of fish stock maturities. One possible error is that a fish's age could be estimated incorrectly which would affect (overestimate or underestimate) both the proportion at the age recorded and at the age that should have been recorded. Another common error is misclassification i.e., the mature fish is labelled as immature (or vice versa). Both sorts of errors will result in the overestimation or underestimation of the

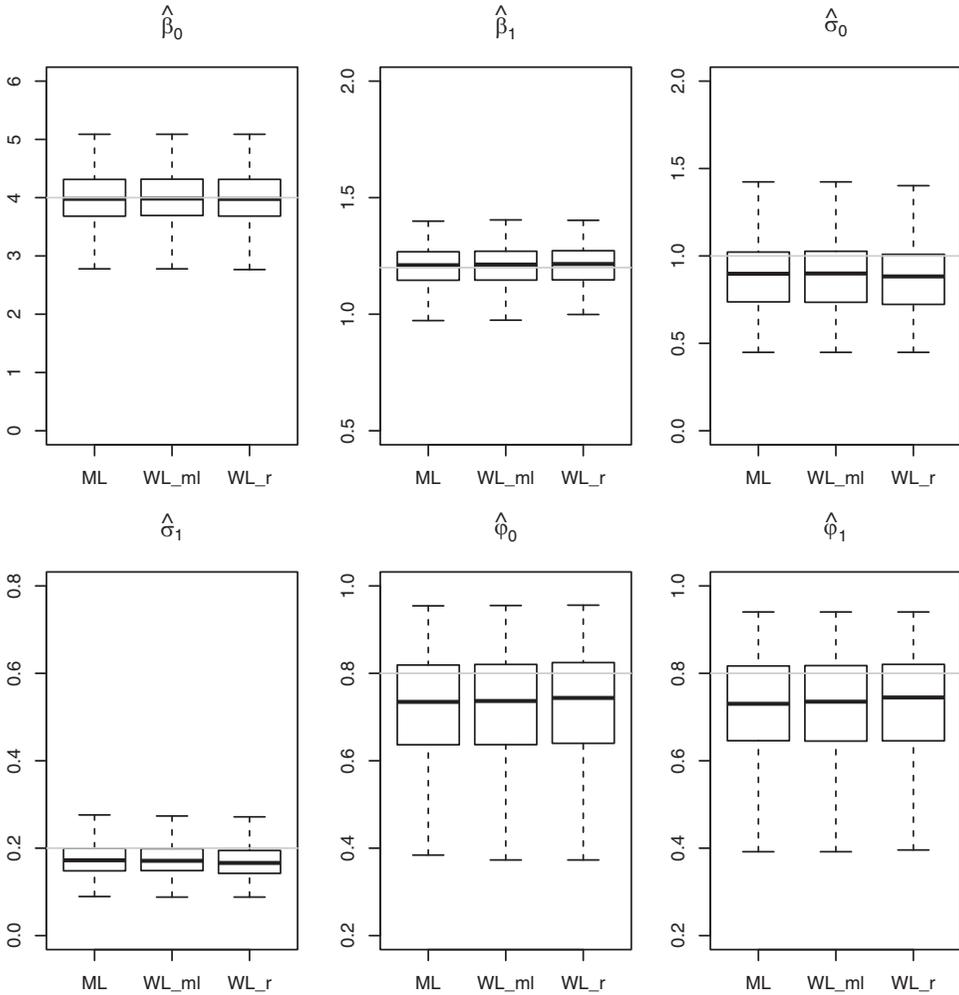


FIGURE 2: Estimated parameters in the case of no contamination. Grey lines denote the true parameter values.

proportion mature at a specific age  $a$  especially when  $a$  is between 3 and 9. The proportions mature vary most in the middle age groups and tend to be 0 for age 1–2 and 1 for age  $> 9$ . In the first error scenario, for each simulation we randomly select two ages from  $(3, 4, \dots, 9)$  across all cohorts and shift their true proportion mature ( $p_{\text{true}}$ ) to be either  $\max(0, p_{\text{true}} - 0.2)$  or  $\min(1, p_{\text{true}} + 0.2)$ . The proportion of contaminated data is  $2/15 \approx 0.13$  and hereafter referred to as age-specific contamination.

We also consider the possibility of systematic errors occurring within a specific year perhaps due to observer inexperience. Here, in each simulation we randomly select 2 years from  $[1970, 2009]$ . For each age  $\in (2, 3, \dots, 10)$  within these 2 years, we shift the true proportion mature  $p_{\text{true}}$  to either  $\max(0, p_{\text{true}} - 0.2)$  or  $\min(1, p_{\text{true}} + 0.2)$ . The proportion of contaminated data is approximately  $2/40 = 0.05$  and hereafter referred to as year-specific contamination.

Our final type of contamination allows for the possibility that a specific cohort differs from other cohorts due to unusual ocean conditions in their first year of life, for example. Here the proportion mature  $p_{ac}$  is modelled as  $p_{ac} = \exp(\tilde{\eta}_{ac})/[1 + \exp(\tilde{\eta}_{ac})]$  with  $\tilde{\eta}_{ac} = \eta_{ac} + \epsilon \xi_{ac}$  where  $\epsilon \sim \text{Bernoulli}(0.05)$  and the extra error term  $\xi_{ac} \sim N(0, 3)$ .

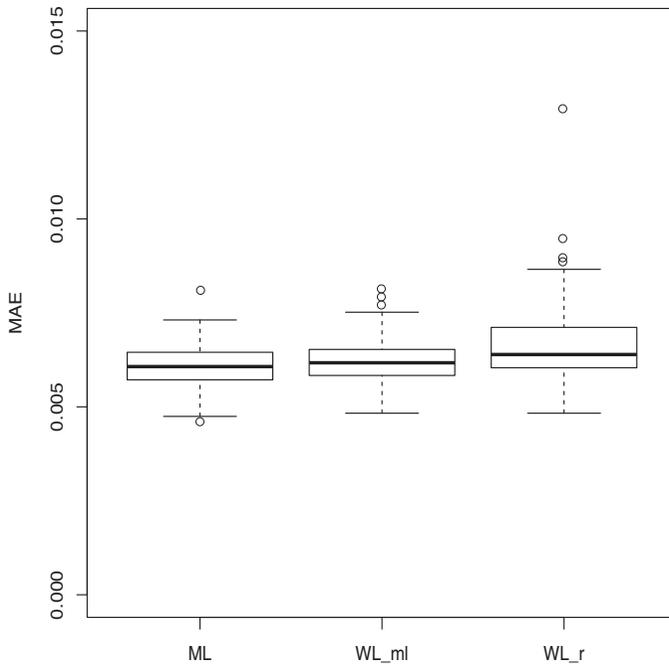


FIGURE 3: Mean absolute differences between the estimated and true proportions in the case of no contamination.

Parameter estimates obtained via each of our different approaches are shown in Figure 4. We see that for all types of contamination the maximum likelihood estimates of  $\beta_0$  and  $\beta_1$  are slightly biased while  $\varphi_0$  and  $\varphi_1$  are heavily underestimated. The weighted likelihood approaches provide more robust estimates of the parameters than the ordinary likelihood method. Interestingly the presence of contamination seems to help alleviate the problem of underestimating the latent state parameters  $\sigma_0$  and  $\sigma_1$ . The mean absolute differences between the estimated and true proportions ( $MAE$ ) are plotted in Figure 5. The estimated proportions based on the weighted likelihood are much more accurate than those based on the ordinary likelihood (ML). WL\_ml performs better than WL\_r. The criteria  $\widehat{MAE}_{\text{trim}}$  with  $\alpha = 0.1$  works well and the performance of WL\_s is very close to WL\_ml. In fact, in the case of age-specific contamination in 93% of the simulations the approach selected by  $\widehat{MAE}_{\text{trim}}$  agrees with the results from the  $MAE$ , especially when the difference between the two  $MAE$ 's is large. For year-specific contamination this statistic is 82% and for cohort-specific contamination it is 70%. We also randomly selected a proportion shift from (0.1, 0.2, 0.3) instead of using the fixed 0.2 and observed very similar results. When we increased the proportion shift to 0.4, the ordinary likelihood method was worse while the weighted likelihood approach got better since the outliers could be identified more easily and downweighted. The boxplots of the  $MAE$  with a proportion shift of 0.4 are shown in panel (b) of Figure 5.

To investigate the impact of the number of outliers on the robustness of our weighted likelihood, we also ran simulations with 4 and 8 years of data contaminated. As the number of contaminated years increases, the weighted likelihood approaches yield more and more robust parameter estimates compared with the ordinary likelihood method. The  $MAE$  in these cases are summarized in Table 1. We can see that the weighted likelihood approaches outperform the maximum likelihood method in all cases and the WL\_ml performs better than the WL\_r. Similarly we randomly selected 2, 4, or 8 cohorts and perturbed the proportions mature by  $\pm 0.2$ . The mean  $MAE$  for the different approaches are summarized in Table 2. We again see that the weighted

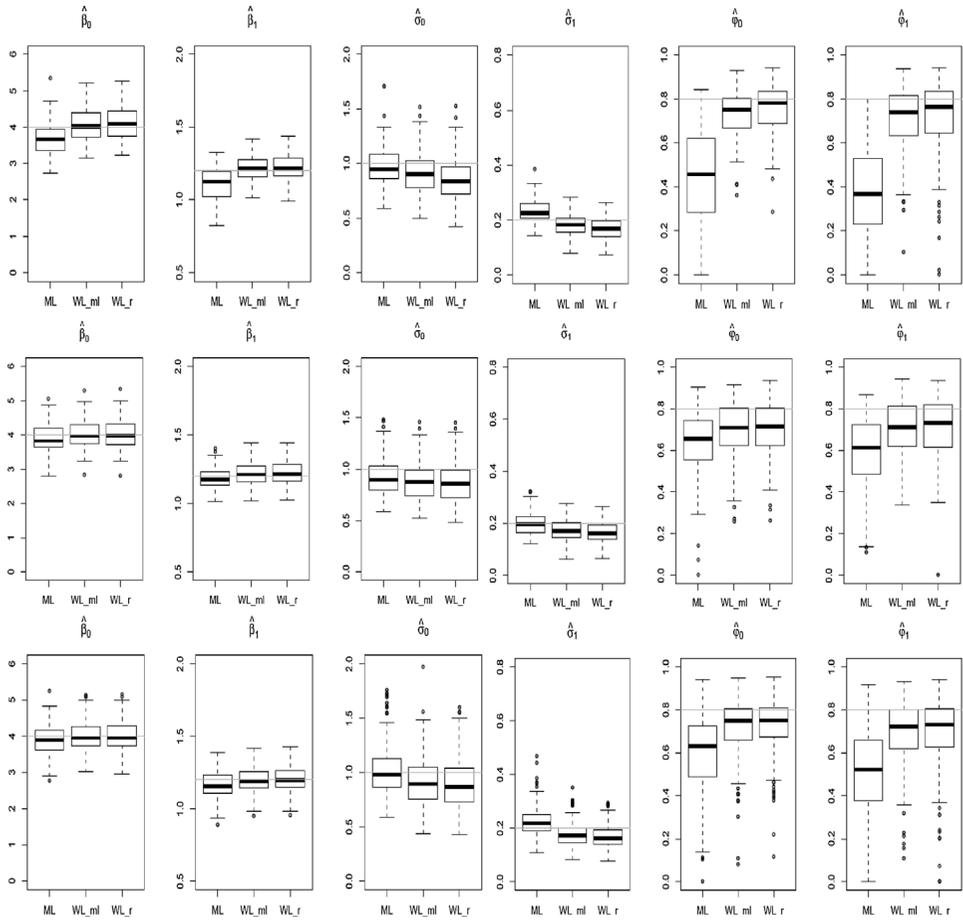


FIGURE 4: Estimated parameters in the case of contamination. The first row corresponds to age-specific contamination, the second row to year-specific contamination and the third row to cohort-specific contamination. Grey lines denote the true parameter values.

likelihood approaches outperform the maximum likelihood method in all cases and the WL<sub>r</sub> performs better than the WL<sub>ml</sub>.

### 5. APPLICATION

In the Northwest Atlantic Ocean, cod are distributed from Greenland to Cape Hatteras and are managed as 12 stocks. The 3Ps stock off southern Newfoundland extends from Cape St. Marys to just west of Burgeo Bank, and over St. Pierre Bank and most of Green Bank. Female cod from this stock are generally thought to be maturing at younger ages in recent years. For example, about 35% of the females are mature by age 5 (47 cm) in recent cohorts, compared to only about 10% at age 5 (55 cm) amongst cohorts present in the 1970s-early 1980s. Catches from this stock have supported an inshore fixed gear fishery for centuries and are of vital importance to the area. Fish are caught offshore by mobile and fixed gear, and inshore by fixed gear only. Spanish and other non-Canadian fleets heavily exploited the stock in the 1960s and early 1970s. French catches increased in the offshore throughout the 1980s. A moratorium on fishing initiated in August 1993 ended in 1997.

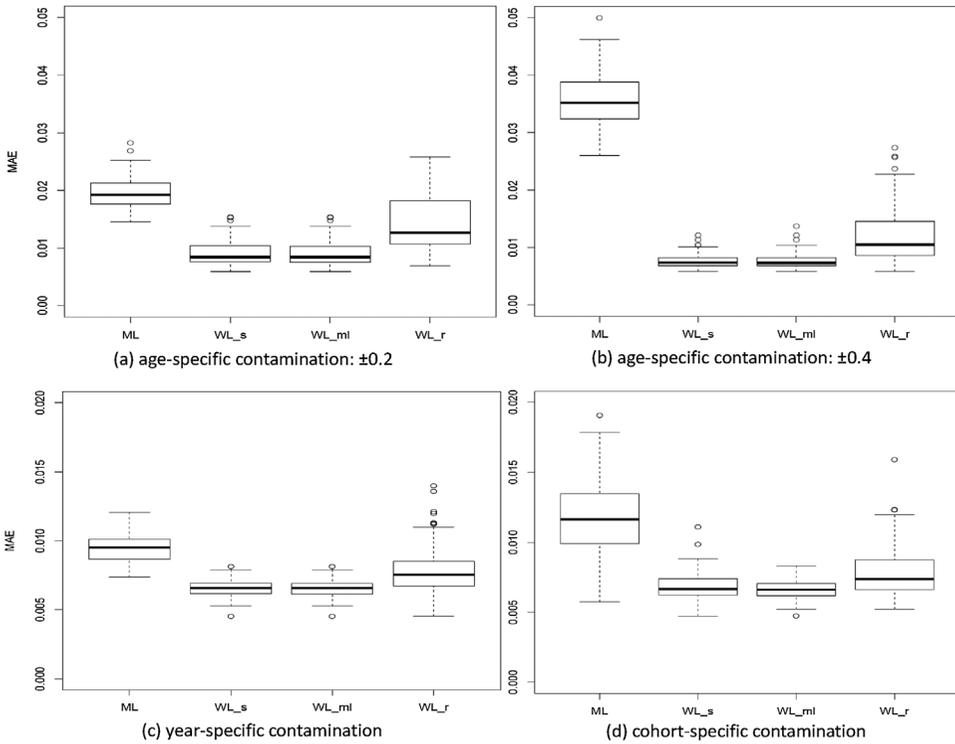


FIGURE 5: Mean absolute differences between the estimated and true proportions in the case of contamination. Age-specific contamination with a proportion shift of 0.2 in panel (a), age-specific contamination with a proportion shift of 0.4 in panel (b), year-specific contamination in panel (c), and cohort-specific contamination in panel (d).

The biological sampling data we consider here concerns the 3Ps cod stock and covers assessment years 1960 through 2011. The ages range between 1 and 24 and the cohorts are from 1946 to 2010. The total number of the observed proportions,  $y_{ac}/n_{ac}$ , is  $n = 689$ . We fitted the SSM to this data as specified in Section 2.1. The moving block method with block size  $b = 5$  is used to compute the confidence intervals for  $p_{ac}$  as well as the standard errors for  $\hat{\theta}$ .

In Table 3, we present the parameter estimates based on the ordinary likelihood and weighted likelihood methods from ADMB. We find that the WL\_ml and ML methods yield more similar parameter estimates than those obtained using the WL\_r method, in particular for the state process parameters. The standard errors based on 1,000 bootstrap samples are also shown in Table 3. The asymptotic standard errors returned by ADMB for the ML estimates are 1.25, 0.13, 0.78, 0.09,

TABLE 1: The mean MAE for the different approaches when the proportion shift is 0.2 and the number of contaminated years is 2, 4, and 8.

	2 years contaminated	4 years contaminated	8 years contaminated
ML	0.0094	0.0127	0.0191
WL_s	0.0066	0.0070	0.0088
WL_ml	0.0065	0.0069	0.0088
WL_r	0.0080	0.0095	0.0133

TABLE 2: The mean MAE for different approaches when the proportion shift is 0.2 and the number of contaminated cohorts is 2, 4, and 8.

	2 Cohorts contaminated	4 Cohorts contaminated	8 Cohorts contaminated
ML	0.0103	0.0151	0.0228
WL <sub>s</sub>	0.0090	0.0123	0.0179
WL <sub>ml</sub>	0.0094	0.0134	0.0208
WL <sub>r</sub>	0.0088	0.0102	0.0142

0.04, and 0.10, respectively. These results are in good agreement with the bootstrap standard errors for four of the six parameters and are larger for  $\beta_0$  and  $\sigma_0$ . At this point, it is somewhat unclear why the asymptotic results are larger for these two parameters but may be related to some uncertainty about the intercept value in the dynamics.

Figure 6 depicts the estimated proportions mature for age  $a = 7$ , as a function of cohort. The WL<sub>ml</sub> curve is very close to the fitted curve based on ordinary likelihood. The WL<sub>r</sub> curve is smoother and is less influenced by extreme data points. The curves for age= 3, 4, 5, 6, 8, 9 are plotted in Figure 7. The number of downweighted observed maturities  $y_{ac}$  (689 in total) is 6 for the WL<sub>ml</sub> approach and is 23 for WL<sub>r</sub>. The  $\widehat{MAE}_{trim}$  for the ML, WL<sub>ml</sub> and WL<sub>r</sub> methods are 0.0113, 0.0110, and 0.0101, respectively when  $\alpha = 0.1$  and hence WL<sub>s</sub>=WL<sub>r</sub>. In the remainder of this section we focus on the curves obtained via our WL<sub>r</sub> approach.

To construct the confidence intervals for the proportions mature,  $p_{ac}$ , 1,000 moving blocks bootstrap samples were drawn and the estimates of  $p_{ac}$  obtained for each sample using the WL<sub>r</sub> approach. The resulting 90% confidence intervals  $(p_{ac}^{*(51)}, p_{ac}^{*(950)})$ , are shown in Figures 8 and

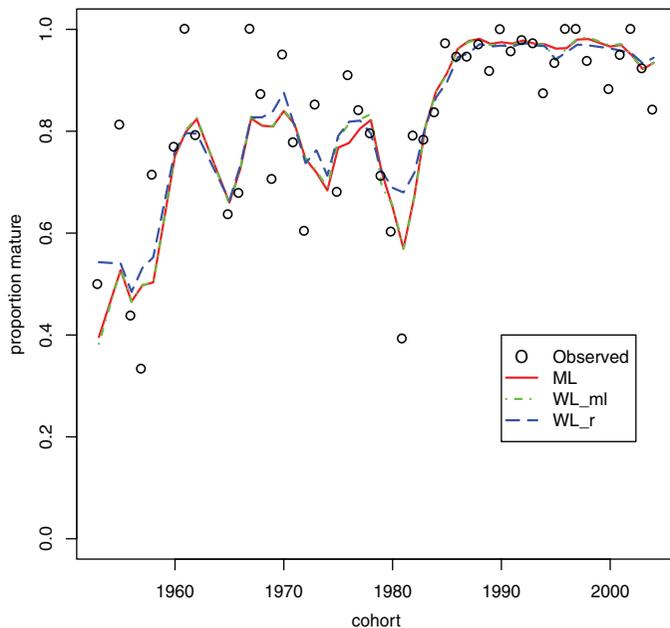


FIGURE 6: The proportion mature for age 7 over cohorts.

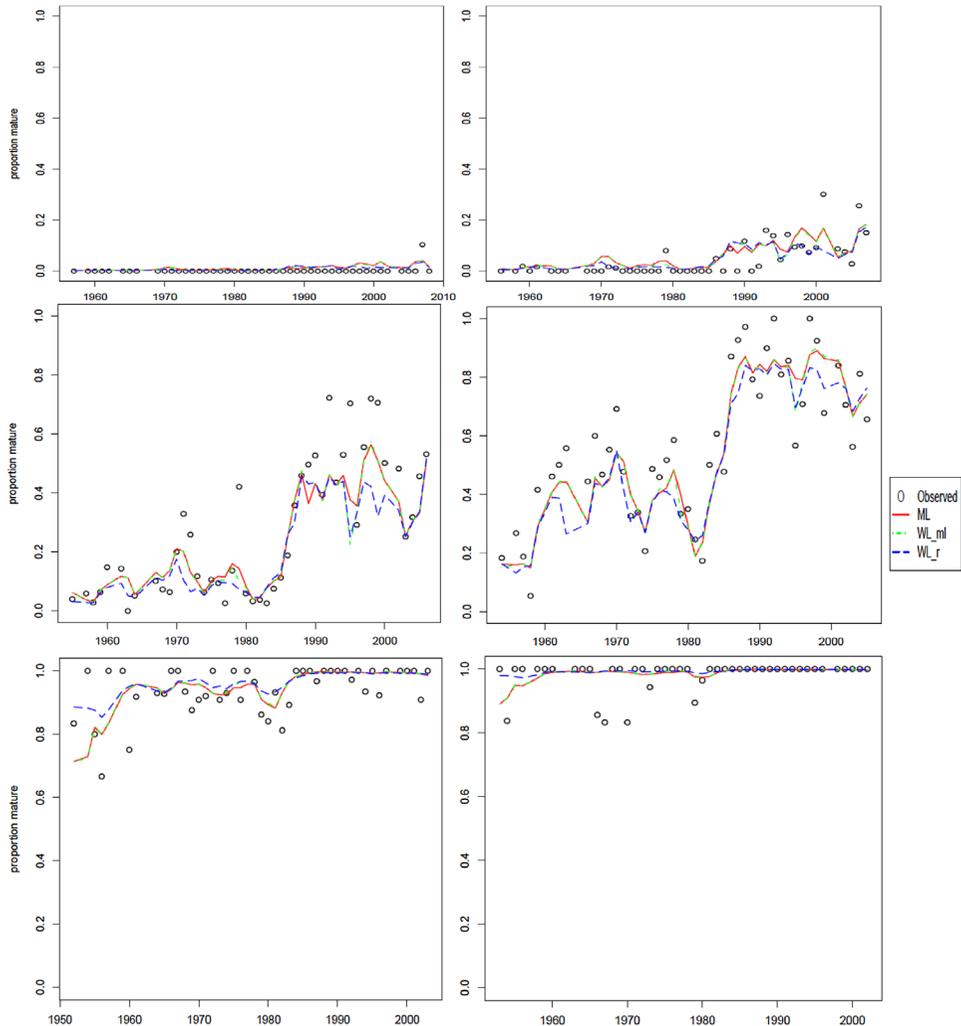


FIGURE 7: The proportions mature for age 3 (top left), age 4 (top right), age 5 (middle left), age 6 (middle right), age 8 (bottom left), and age 9 (bottom right) over cohorts.

9. Via simulation, using the  $WL_r$  parameter estimates in Table 3 as true parameter values, we find that the average pointwise coverage probabilities of the moving block bootstrap percentile intervals is between 85% and 90%, which is satisfactory and validates our bootstrap procedure.

## 6. DISCUSSION

In this paper, we have successfully incorporated robustness into the SSM framework used for assessing maturity at a given age  $a$  in a fish stock. This has been accomplished using a weighted likelihood approach with constraints on the weights. The evaluation of the weighted likelihood which involves high dimensional integration is accomplished via a Laplace approximation as implemented in the ADMB package (Fouriner et al., 2012). Our simulation results show that in the presence of contamination, the weighted likelihood approach performs very well.

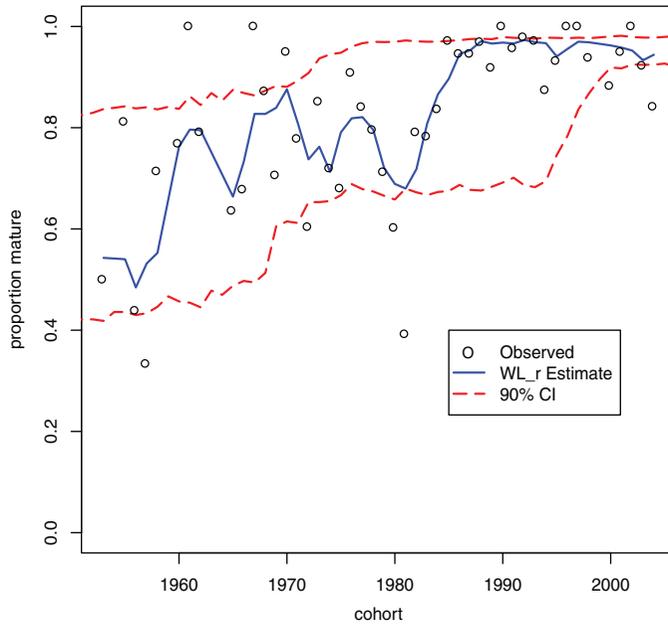


FIGURE 8: The 90% bootstrap percentile intervals for age 7.

In contrast to particle filter methods, the use of ADMB to optimize the weighted likelihood is preferable as we avoid some of the well known issues of particle filters where the evaluation of the likelihood is subject to significant Monte Carlo variation making it difficult to use gradient methods for optimization. Since ADMB does automatic differentiation, the gradient can be computed exactly subject to the approximation of the integral by the Laplace approximation. The results of our simulation study suggest that the Laplace approximation provides the necessary accuracy with which to obtain good parameter estimates.

There are several points to be made relating to our procedure.

1. As noted earlier, we could replace the normal distribution assumed for the states with a  $t$ -distribution or Huber’s least favorable. We carried out some trials where our model was fitted using a  $t_5$  both in the case where the actual dynamics were generated with normal errors and where they were  $t_5$  errors. Our experience was that fitting with normal errors for the states worked as well as fitting with the  $t_5$  even when the actual dynamics were generated from the  $t_5$ . It is possible that the Laplace approximation is not as accurate with the  $t_5$  as with the normal. For the Huber least favorable there is the issue of being non-differentiable at the transition point making it difficult to implement in ADMB (which requires differentiability). Of course, one can approximate Huber’s least favorable with a smoothly differentiable approximation which we did not do given the results obtained from the  $t_5$ . This is certainly an area of possible future research.
2. A more complex model like the beta-binomial that allows for over-dispersion may be more appropriate for modelling maturity data (Cadigan, Morgan, & Bratney, 2014). When the over-dispersion model is correctly specified we would expect that we can make more accurate inference than using robust statistical methods. However, our approach with the weighted likelihood will provide robustness not only against over-dispersion but also against other types of outlying behaviour in the data. Model selection is not a primary concern in this paper, and our approach should be applicable for a wide range of models for which the likelihood function is available.

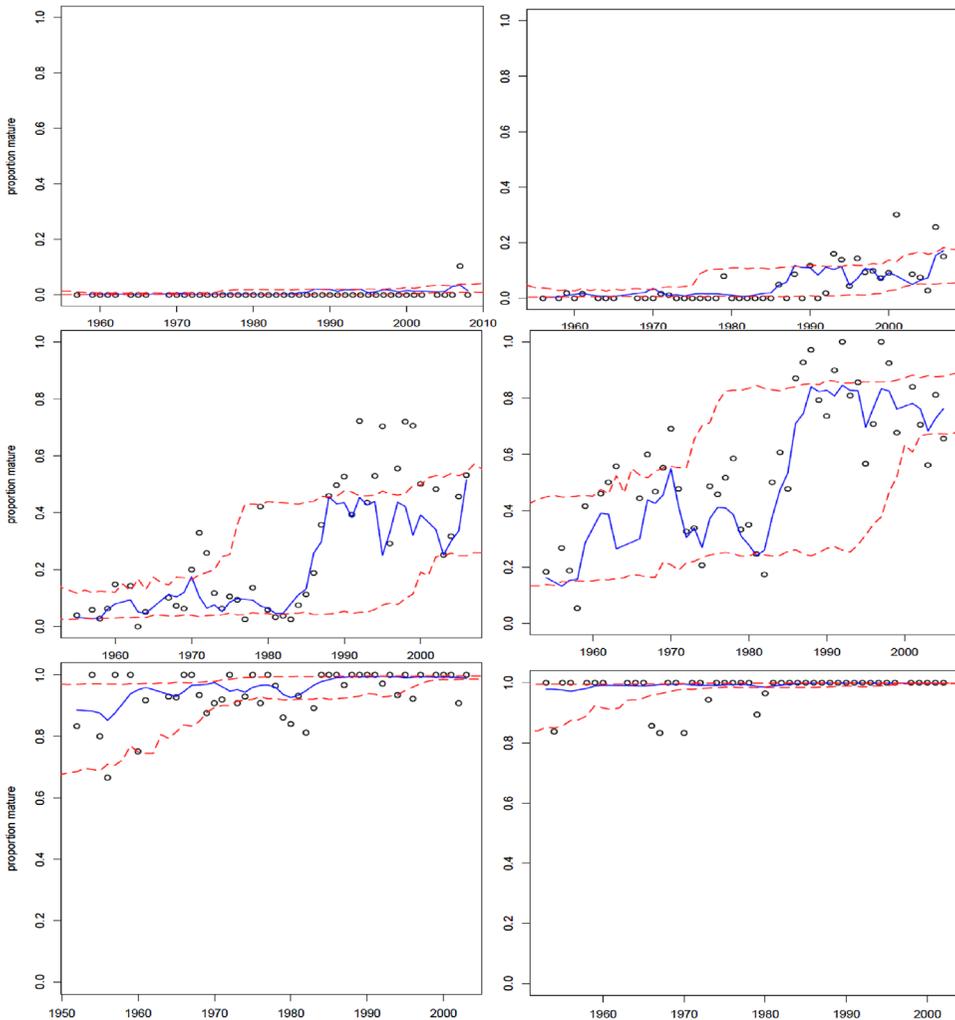


FIGURE 9: The 90% bootstrap percentile intervals for age 3 (top left), age 4 (top right), age 5 (middle left), age 6 (middle right), age 8 (bottom left), and age 9 (bottom right).

3. We have proposed a criterion for determining which of the initial weights gives better estimates by comparing the observed and predicted probabilities (see Eq. (9)). In our simulation study, we can replace the observed probabilities by the true probabilities and assess how well our procedure actually works. In more than 70% of the cases our criterion was also best when used with the true probabilities.

TABLE 3: Parameter estimates and bootstrap standard errors (in parentheses) for different methods.

	$\beta_0$	$\beta_1$	$\sigma_0$	$\sigma_1$	$\varphi_0$	$\varphi_1$
ML	3.24 (0.38)	1.73 (0.12)	1.76 (0.29)	0.27 (0.10)	0.96 (0.03)	0.84 (0.12)
WL <sub>ml</sub>	3.28 (0.41)	1.76 (0.12)	1.74 (0.29)	0.25 (0.10)	0.95 (0.03)	0.82 (0.13)
WL <sub>r</sub>	3.69 (0.43)	1.86 (0.13)	1.33 (0.30)	0.11 (0.12)	0.95 (0.04)	0.48 (0.35)

4. We also compared a one-step optimization procedure with the iterative optimization scheme that we recommend. In a small simulation, the results from the one-step optimization procedure were significantly worse than those obtained with our procedure.
5. We also considered a parametric bootstrap but observed that it is not straightforward for time-series data with dependent observations since we still have to integrate out the random effects. While this is certainly possible we decided to use the block bootstrap which has proven to be successful in many time series applications (e.g., Vogel & Shallcross, 1996). There are other block bootstrap methods available for dependent data (Lahiri, 2003). One of our future research directions is to develop robust bootstrap methods in the presence of deviating observations, by utilizing our weighted likelihood.

## ACKNOWLEDGEMENTS

The authors are grateful to the anonymous reviewer for insightful comments that substantially improved the paper. We would like to thank Noel Cadigan for providing the data and introducing us to ADMB and to Anders Nielsen for his assistance with ADMB. This research was supported by Discovery Grant to JMF and CF from the Natural Sciences and Engineering Research Council of Canada. Support was also provided to all authors by the Canadian Statistical Sciences Institute through a Collaborative Research Team Project titled “Advancements to State-Space Models for Fisheries Science.”

## BIBLIOGRAPHY

- Aanes, S., Engen, S., Sæther, B. E., & Aanes, R. (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, 1130–1142.
- Buckland, S. T., Newman, K. B., Fernandez, C., Thomas, L., & Harwood, J. (2007). Embedding population dynamics models in inference. *Statistical Science*, 22, 44–58.
- Bolker, B. M., Gardner, B., Maunder, M., Berg, C. W., Brooks, M., Comita, L., Crone, E., Cubaynes, S., Davies, T., de Valpine, P., Ford, J., Gimenez, O., Kry, M., Kim, E. J., Lennert-Cody, C., Magnusson, A., Martell, S., Nash, J., Nielsen, A., Regetz, J., Skaug, H., & Zipkin, E. (2013). Strategies for fitting nonlinear ecological models in R, AD Model Builder, and BUGS. *Methods in Ecology and Evolution*, 4, 501–512.
- Cadigan, N. G., Morgan, M. J., & Bratney, J. (2014). Improved estimation and forecasts of stock maturities using generalised linear mixed models with autocorrelated random effects. *Fisheries Management and Ecology*, 21, 343–356.
- Calvet, L. E., Czellar, V., & Ronchetti, E. (2014). Robust filtering. *Journal of the American Statistical Association*, forthcoming, DOI: 10.1080/01621459.2014.983520.
- Cantoni, E. & Ronchetti, E. (2001). Robust inference for generalized linear models. *Journal of the American Statistical Association*, 96, 1022–1030.
- Dupuis, D. J. & Morgenthaler, S. (2002). Robust weighted likelihood estimators with an application to bivariate extreme value problems. *The Canadian Journal of Statistics*, 30, 17–36.
- Field, C. & Smith, B. (1994). Robust estimation: A weighted maximum likelihood approach. *International Statistical Review*, 62, 405–424.
- Fournier, A. D., Skaug, J. H., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, N. M., Nielsen, A., & Sibert, J. (2012). AD Model Builder: Using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, 27, 233–249.
- Gudmundsson, G. & Gunnlaugsson, T. (2012). Selection and estimation of sequential catch-at-age models. *Canadian Journal of Fisheries and Aquatic Sciences*, 69, 1760–1772.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*, John Wiley & Sons, New York.

- Heritier, S., Cantoni, E., Copt, S., & Victoria-Feser, M.-P. (2009). *Robust Methods in Biostatistics*, Wiley, Chichester.
- Huber, P. J. (1981). *Robust Statistics*, John Wiley&Sons, New York.
- Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17, 1217–1241.
- Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*, Springer, New York.
- Lewy, P. & Nielsen, A. (2003). Modelling stochastic fish stock dynamics using Markov Chain Monte Carlo. *ICES Journal of Marine Science: Journal du Conseil*, 60, 743–752.
- Meyer, R. & Millar, R. B. (1999). Bayesian stock assessment using a statespace implementation of the delay difference model. *Canadian Journal of Fisheries and Aquatic Sciences*, 56, 37–52.
- Minto, C., Mills Flemming, J., Britten, G. L., & Worm, B. (2014). Productivity dynamics of Atlantic cod. *Canadian Journal of Fisheries and Aquatic Sciences*, 71, 203–216.
- Pedersen, M. W., Berg, C. W., Thygesen, U. H., Nielsen, A., & Madsen, H. (2011). Estimation methods for nonlinear state-space models in ecology. *Ecological Modelling*, 222, 1394–1400.
- Roh, S., Genton, M., Jun, M., Szunyogh, I., & Hoteit, I. (2013). Observation quality control with a robust ensemble Kalman filter. *Monthly Weather Review*, 141, 4414–4428.
- Ruiz-Cárdenas, R., Krainski, E., & Rue, H. (2012). Direct fitting of dynamic models using integrated nested Laplace approximations: INLA. *Computational Statistics & Data Analysis*, 56, 1808–1828.
- Schnute, J. T. (1994). A general framework for developing sequential fisheries models. *Canadian Journal of Fisheries and Aquatic Sciences*, 51, 1676–1688.
- Vogel, R. M. & Shallcross, A. L. (1996). The moving blocks bootstrap versus parametric time series models. *Water Resources Research*, 32, 1875–1882.

---

Received 5 August 2014

Accepted 13 December 2014