1 Title: A stochastic surplus production model in continuous-time

2 Alternative title 1: A stochastic surplus production model in continuous-time for data-

3 limited stock assessment

4 Alternative title 2: A stochastic surplus production model for data-limited stock assess-

5 ment

6 Authors: Martin W. Pedersen[1,4], Casper W. Berg[1]

7 [1] National Institute of Aquatic Resources, Technical University of Denmark, Charlotten-

8 lund Slot, Jægersborg Allé 1, 2920 Copenhagen, Denmark

9 [4] Corresponding author: map@aqua.dtu.dk

10 Running title: Stochastic surplus production model

11

Summary (Abstract)

Surplus production modelling has a long history as a method for managing data-limited fish stocks. Recent advancements have cast surplus production models as state-space models that separate random variability of stock dynamics from error in observed indices of biomass. We present a stochastic surplus production model in continuous-time (SPiCT), which in addition to stock dynamics also models the dynamics of the fisheries. This enables error in the catch process to be reflected in the uncertainty of estimated model parameters and management quantities. Benefits of the continuous-time state-space model formulation include the ability to provide estimates of exploitable biomass and fishing mortality at any point in time from data sampled at arbitrary and possibly irregular intervals. We show in a simulation that the ability to analyse sub-annual data can increase the effective sample size and improve estimation of reference points relative to discrete-time analysis of aggregated annual data. Finally, sub-annual data from five North Sea stocks are analysed with particular focus on using residual analysis to diagnose model insufficiencies and identify necessary model extensions such as robust estimation and incorporation of seasonality. We argue that including all known sources of uncertainty, propagation of that uncertainty to reference points and checking of model assumptions using residuals are critical prerequisites to rigorous fish stock management based on surplus production models.

Contents

Key words: Data-limited methods, fisheries management, maximum sustainable yield, seasonal population dynamics, stock assessment, Pella-Tomlinson model.

1 Introduction

Surplus production models are used to assess the biomass and exploitation level of marine populations in data-limited situations where age and size information are unavailable (Punt, 2003). By modelling the temporal evolution of the aggregated biomass targeted

3

by fishing, surplus production models do not require information about the age or size-structure of the stock. Surplus production models are therefore primarily applied to stocks where the only data available are observations of commercial catches together with some index of exploitable biomass such as commercial catch-per-unit-effort (CPUE) or as derived from scientific survey data (Polacheck *et al.*, 1993). Common applications of surplus production models include large migratory pelagic fishes such as tuna, sharks and billfishes (e.g. Brodziak & Ishimura, 2011; Carvalho *et al.*, 2014), but also crustaceans that are generally difficult to age (e.g. Smith & Addison, 2003).

The population dynamics represented by surplus production models builds on principles of logistic or the more general theta-logistic growth (Pedersen *et al.*, 2011) resulting in a dome-shaped relationship between biomass production and population size. In the formulation of Schaefer (1954) the maximum production occurs when the population size is at half its carrying capacity. This restriction is avoided in the generalised form of Pella & Tomlinson (1969), where skewness of the production function allows maximum production at any biomass below the carrying capacity. Thus, a population is optimally exploited in terms of biomass if it is harvested to remain at the biomass level that results in the maximum production defined as the maximum sustainable yield (MSY).

It is widely recognised that the model structure of surplus production models is too simple to adequately describe the population dynamics of a real-world stock subject to variability in size structure, species interactions, recruitment, catchability, selectivity, environmental conditions etc. (Pella & Tomlinson, 1969). To mitigate this, it is common to include a random error term in the equation governing the biomass dynamics as a proxy for unmodelled variability (process error). Similarly, it is often assumed that the biomass index is subject to error in sampling that causes the observed values to deviate from the true. This variability is incorporated by including an observation error term in the equation describing how the index data relate to the biomass. Models including random terms

4

are often referred to as stochastic.

While the basic deterministic formulation of surplus production models has remained unchanged for decades, the methods used for estimating model parameters from observations while acknowledging random variability have undergone major developments. Least-squares methods for fitting production curves (Pella & Tomlinson, 1969) developed into more explicit assignment of noise in process-error or observation-error models (Polacheck *et al.*, 1993). Key to the incorporation of random variability was the discrete-time formulation of Schnute (1977) achieved by approximating observations by annual averages. The majority of subsequent model extensions, with the exception of ASPIC (Prager, 1994), adopted the discrete-time form including state-space models that simultaneously estimate both process and observation error using Bayesian (Meyer & Millar, 1999) or frequentist approaches (Punt, 2003). ASPIC, while free of discrete-time average approximations, lacks flexibility owing to its deterministic population dynamics. Regardless of temporal form, observation error in previous models refer to error in the index observation with the often unstated implicit assumption that catch observations are error-free.

The increased flexibility of state-space models, which are the variants of surplus production models most commonly applied today, entails a need for informative data. Thus, data situations with short time series or limited contrast may require that some model parameters are constrained or fixed to obtain model identifiability. For example, the ratio of process to observation noise is often fixed (Ono *et al.*, 2012), while estimating the shape of the production curve is commonly avoided by assuming the symmetric form of Schaefer (1954). Constraining estimation by fixing model parameters affects estimates of remaining free parameters, some of which are directly related to management quantities. Thus, assessing the sensitivity of results to assumed constraints is an important aspect of surplus production modelling, which if ignored may lead to poor management.

Stochastic surplus production models aim to account for autocorrelation inherent in

time series data of catches and biomass indices. This implies that the recursive, or so called one-step-ahead, residuals of a model fit are assumed to be standard normally distributed and independent. While this is a well-known property of many time series models (Ljung & Box, 1978), assessing the validity of these assumptions is often ignored when fitting stochastic surplus production models. This is problematic as residual patterns can indicate violation of model assumptions, which potentially invalidates model estimates and associated confidence intervals thus conveying a misleading impression of stock status.

We present a stochastic surplus production model in continuous-time (SPiCT), which incorporates dynamics in both biomass and fisheries and observation error of both catches and biomass indices. The model has a general state-space form that as special cases contain process and observation-error models (Polacheck *et al.*, 1993; Prager, 1994) as well as state-space models that assume error-free catches (Meyer & Millar, 1999; Punt, 2003; Ono *et al.*, 2012). Seasonal extensions to the fisheries dynamics component of the state-space model are also developed facilitating the use of sub-annual data that contain seasonal patterns. Estimation performance, as a function of the number of available observations and under model misspecifications, is evaluated through simulation experiments. Differences in performance between discrete-time and continuous-time models are assessed both in a simulation experiment and by analysing the South Atlantic albacore data set of Polacheck *et al.* (1993) and comparing with previously published results from discrete-time models. Finally, we analyse five North Sea stocks using quarterly resolved data with emphasis on the use of one-step-ahead residuals for detecting possible lack of model fit.

2   Methods

2.1   Model formulation

Generalised surplus production models in the form of Pella & Tomlinson (1969) can be written as

$$\frac{dB_t}{dt} = \frac{r}{n-1} B_t \left(1 - \left[\frac{B_t}{K} \right]^{n-1} \right) - F_t B_t, \tag{1}$$

where B_t is the exploitable stock biomass, F_t is the instantaneous fishing mortality rate, r is the intrinsic growth rate of the population, K is the carrying capacity, and $n > 0$ is a unitless parameter determining the shape of the production curve. In case $n = 2$ the stock production reduces to $rB_t(1 - B/K)$ (Schaefer, 1954), while the limit of the stock production for $n \to 1$ is $rB_t \log(K/B_t)$ (Fox, 1970). The term $F_t B_t$ of Eqn. (1) represents the instantaneous catch while the remaining part of the right-hand-side of Eqn. (1) represents the instantaneous biomass surplus production of the stock following a theta-logistic growth function (Pedersen *et al.*, 2011). The intrinsic growth rate r models density-independent growth and natural mortality. The carrying capacity, K, is a density dependent growth penalty corresponding to the equilibrium B_t of an unexploited stock $(F_t = 0)$.

The parametrisation of Eqn. (1), while easy to interpret biologically, is difficult to estimate owing high correlation between r and K. Using geometric arguments related to the dome-shaped production curve Fletcher (1978) derived a more stable parametrisation

$$\frac{dB_t}{dt} = \gamma m \frac{B_t}{K} - \gamma m \left(\frac{B_t}{K} \right)^n - F_t B_t, \tag{2}$$

where $\gamma = n^{n/(n-1)}/(n-1)$ and

$$m = \frac{rK}{n^{n/(n-1)}} \tag{3}$$

$_{151}$ The case $n = 1$ is a removable singularity corresponding to the model of Fox (1970).

$_{152}$ Estimating m instead of r results in a more well-defined optimum because of the separate

$_{153}$ roles of m and K in defining the production curve with K representing the width of the

$_{154}$ biomass range and $m = MSY^d$ representing the maximum sustainable yield (maximum

$_{155}$ attainable surplus production). The biomass and fishing mortality leading to maximum

$_{156}$ surplus production are $B_{MSY}^d = n^{1/(1-n)}K$ and $F_{MSY}^d = m/B_{MSY}$. The superscript d

$_{157}$ indicates that these are deterministic reference points that do not account for random

$_{158}$ variability. If $n = 1$ the reference points have the limits $B_{MSY}^d = K/e$, $MSY^d = rK/e$,

$_{159}$ and $F_{MSY}^d = r$.

$_{160}$ Eqn. (2) is a simplified and deterministic description of biomass dynamics. In real-

$_{161}$ ity many additional factors (e.g. species interactions, environmental variability) influence

$_{162}$ biomass dynamics. In the absence of specific data pertaining to these processes one can

$_{163}$ model their influence using a stochastic process noise term. Including process noise in

$_{164}$ Eqn. (2) results in a stochastic surplus production model in continuous-time (SPiCT)

$$dB_t = \left(\gamma m \frac{B_t}{K} - \gamma m \left[\frac{B_t}{K} \right]^n - F_t B_t \right) dt + \sigma_B B_t dW_t, \tag{4}$$

$_{165}$ where σ_B is the standard deviation of the process noise, and W_t is Brownian motion.

$_{166}$ Representing surplus production by a stochastic differential equation (SDE, Øksendal,

$_{167}$ 2013) acknowledges the presence of random and unmodelled process variability, while

$_{168}$ retaining the property that the process is defined continuously in time and not restricted

$_{169}$ to specific discrete time points.

$_{170}$ The process noise of Eqn. (4) is multiplicative owing to the presence of B_t in the noise

$_{171}$ term $\sigma_B B_t dW_t$. Multiplicative noise terms can, in terms of numerical implementation and

model fitting, lead to instability problems. By Lamperti transforming the model (Iacus, 2009) an additive noise term is obtained

$$dZ_t = \left(\frac{\gamma m}{K} - \frac{\gamma m}{K} \left[\frac{e^{Z_t}}{K} \right]^{n-1} - F_t - \frac{1}{2}\sigma_B^2 \right) dt + \sigma_B dW_t, \tag{5}$$

where $Z_t = \log(B_t)$. Interestingly, the standard deviation of the process noise, σ_B^2, now appears in the deterministic part of Eqn. (5) and therefore influences the reference points related to MSY. For $n > 1$, Bordet & Rivest (2014) derived the following stochastic reference points

$$B_{MSY} = B_{MSY}^d \left(1 - \frac{1 + F_{MSY}^d(n-2)/2}{F_{MSY}^d(2 - F_{MSY}^d)^2}\sigma_B^2 \right), \tag{6}$$

$$F_{MSY} = F_{MSY}^d - \frac{(n-1)(1 - F_{MSY}^d)}{(2 - F_{MSY}^d)^2}\sigma_B^2, \tag{7}$$

$$MSY = MSY^d \left(1 - \frac{n/2}{1 - (1 - F_{MSY}^d)^2}\sigma_B^2 \right). \tag{8}$$

Determining whether the stochastic reference points hold for $n < 1$ is an area of ongoing research and outside the scope of the current study. It is evident that $F_{MSY} < F_{MSY}^d$ when $\sigma_B^2 > 0$, which implies that the presence of process noise entails a conservative reduction in recommended fishing pressure relative to the deterministic case, however the opposite behaviour is observed if Eqn. (7) is assumed to hold for $n < 1$. We therefore use the stochastic reference points when $n > 1$ and the deterministic reference points when $0 < n \leq 1$.

Subject to constant F_t, B_t has a stationary distribution (Bordet & Rivest, 2014) with an expected equilibrium biomass, which can be approximated by

$$E(B_\infty|F_t) = K \left(1 - \frac{(n-1)}{n} \frac{F_t}{F_{MSY}^d} \right)^{1/(n-1)} \left(1 - \frac{n/2}{1 - \left(1 - nF_{MSY}^d + [n-1]F_t \right)^2} \sigma_B^2 \right).$$

(9)

As expected $E(B_\infty|F_t) = K$ in the absence of fishing ($F_t = 0$) and process noise ($\sigma_B^2 = 0$). In general $E(B_\infty|F_t)$ can be interpreted as a fished equilibrium, which can serve as a biomass predictor if current conditions remain constant.

The majority of existing production models leave the process of commercial fishing, F_t, unmodelled. Discrete-time models (e.g. Punt, 2003) commonly assume that $F_t = C_t/B_t$ where C_t is the observed aggregated catch in year t. This approach implicitly assumes that B_t and F_t represent annual averages of biomass and fishing mortality. Perhaps more importantly it is assumed that the catch is observed without error. If present, observation error in the catch will therefore propagate directly to F_t and influence conclusions regarding the current fishing pressure. Furthermore, previous models are only able to estimate F_t at times when a catch observation is available.

An alternative approach, which addresses the above issues, is to model F_t as a separate and unobserved process in the same sense that B_t is unobserved, which allows F_t to be estimated at any time even when a catch observation is unavailable. Our general model for F_t is the product of a random component G_t and a seasonal component S_t

$$F_t = S_t G_t \tag{10}$$

$$d \log G_t = \sigma_F dV_t \tag{11}$$

where dV_t is standard Brownian motion and σ_F is the standard deviation of the noise. If only annual data are available it is not possible to estimate within-year dynamics and therefore $S_t = 1$ and consequently $F_t = G_t$. Given sub-annual data we suggest two models for seasonal variation in the fishing. The first model represents seasonal variation using

10

a cyclic spline, while the second model uses a system of SDEs whose solutions oscillate periodically. In the case of the seasonal spline F_t follows the model

$$F_t = \exp(D_{s(t)})G_t \tag{12}$$

where $D_{s(t)}$ is a cyclic B-spline with a period of one year with $s(t) \in [0; 1]$ being a mapping from t to the proportion of the current year that has passed. The possible annual variation allowed by the cyclic B-spline is determined by a chosen number of so-called knots. The number of knots must be smaller than or equal to the number of catch observations per year (e.g. quarterly catches can at most accommodate four temporally equidistant knots). The values of the cyclic B-spline is defined by the parameter vector ϕ of length equal to the number of knots minus one. In the case of annual data (one knot) the cyclic B-spline reduces to a constant ($D_{s(t)} = 1$) and ϕ has zero length and is therefore not estimated. Note that the seasonal pattern represented by the spline remains constant in time. Thus, a spline-based model is not able to adapt to changes in amplitude and timing (phase) of the real seasonal fishing pattern. Such variations in the fishing pattern would, when fitted with a spline-based model, likely lead to autocorrelated catch residuals.

A way to overcome the potential problems of the spline-based seasonal model is to construct a model that uses the current state and the state a year ago to predict the future state. In discrete-time such a seasonal correlation structure is readily implemented by including the state at appropriate temporal lags, however in continuous-time obtaining such a correlation structure is more complicated. An approximation of this structure is to use a system of coupled SDEs, which perturb each other resulting in oscillating solutions. A simple form of such a system is

11

$$dU_{1,t} = -(\lambda U_{1,t} + \omega U_{2,t})dt + \sigma_U dX_{1,t} \tag{13}$$

$$dU_{2,t} = -(\lambda U_{2,t} - \omega U_{1,t})dt + \sigma_U dX_{2,t} \tag{14}$$

where $dX_{1,t}$ and $dX_{2,t}$ are standard Brownian motion, σ_U is the standard deviation of the noise, ω is a phase parameter set such that the period of the cycles is one year, and $\lambda > 0$ is a damping parameter that ensures that the process is stationary. The solution to the SDE system is known analytically to be oscillatory (Gardiner, 1985), thus $U_{1,t}$ can be used to represent seasonal variation in F_t by the following expression

$$F_t = \exp(U_{1,t})G_t. \tag{15}$$

One can therefore think of $U_{2,t}$ as a slave state that is not used directly to calculate F_t as it only enters through its coupling with $U_{1,t}$. The coupled SDE model is more flexible than the spline based model in that it is able to accommodate gradual shifts in amplitude and timing of the seasonal fishing pattern. This advantage comes at the cost of increased model complexity (as $U_{1,t}$ and $U_{2,t}$ are unobserved processes) and therefore potential difficulties in obtaining model convergence.

An observation of commercial catch is reported as the cumulative catch C_t over a time interval Δ_t. Given B_t and F_t the observed catch in log can therefore be written as an integral in continuous-time (Prager, 1994) plus noise

$$\log(C_t) = \log\left(\int_t^{t+\Delta_t} F_s B_s ds\right) + \epsilon_t, \tag{16}$$

where the catch observation errors $\epsilon_t \sim N(0, \sigma_C^2)$ are independent and σ_C is the standard deviation of the catch observation error. The formulation of Eqn. (16) allows the noise of the F_t process to be separated from the observation noise of C_t. The model furthermore handles catches sampled at any time and aggregated over any interval length Δ_t, e.g. a

12

₂₄₅ quarter of a year. For model fitting we assume to have N_{obsC} observations of C_t together

₂₄₆ with the associated time intervals Δ_t over which C_t were accumulated.

₂₄₇ In addition to catch observations we assume to have observations of N_i series of indices

₂₄₈ of exploitable biomass ($I_{t,i}$ for $i = 1, \ldots, N_i$) with series i containing $N_{obsI,i}$ observations.

₂₄₉ A biomass index could be commercial or scientific catch-per-unit-of-effort data or other

₂₅₀ biomass indicators e.g. as derived from acoustic surveys. Contrary to catch observations

₂₅₁ that are aggregated over a period of time, $I_{t,i}$ are regarded as "snapshots" related to the

₂₅₂ time point t given as

$$\log(I_{t,i}) = \log(q_i B_t) + e_{t,i}, \tag{17}$$

₂₅₃ where $e_{t,i} \sim N(0, \sigma_{I,i}^2)$ are independent normal deviates and $\sigma_{I,i}$ is the standard deviation

₂₅₄ of the ith index observation error, and q_i is a catchability parameter for the ith index. In

₂₅₅ the common situation where only one series of biomass indices is available ($N_i = 1$) the i

₂₅₆ index will be omitted for notational simplicity.

₂₅₇ We also define the ratios between observation and process errors $\alpha = \sigma_I / \sigma_B$ and

₂₅₈ $\beta = \sigma_C / \sigma_F$. In cases where it is not possible to separate process and observation error

₂₅₉ a common simplification is to assume process error of B_t and observation error of I_t to

₂₆₀ be equal (Ono *et al.*, 2012; Thorson *et al.*, 2013), i.e. to fix $\alpha = 1$. A similar relationship

₂₆₁ between the process error of F_t and the observation error of C_t could be envisioned, i.e.

₂₆₂ that $\beta = 1$, which we use when σ_C and σ_F cannot be estimated separately.

₂₆₃ Extreme observations or outliers in index and catch is a commonly encountered prob-

₂₆₄ lem in fisheries data (Chen *et al.*, 1994). Such outliers are poorly modelled when using the

₂₆₅ normal distribution for observation errors, which may lead to bias of parameter estimates.

₂₆₆ Common approaches to mitigate the influence of outliers include objective outlier detection

₂₆₇ and subsequent residual rescaling (Prager, 2002), or robust estimation using fat-tailed er-

ror distributions (Chen *et al.*, 2000). Here we take a simple approach to robust estimation where the observation error follow the mixture distribution $pN(0, \sigma^2) + (1-p)N(0, [w\sigma]^2)$, where p and w are parameters controlling the fatness of the tails. The parameters p and w should in principle be estimated from data, however as we include robust estimation mainly for illustrative purposes we fix $p = 0.95$ and $w = 15$ for simplicity.

In summary, SPiCT is a state-space model for surplus production containing unobserved processes for B_t and F_t and observed quantities C_t and $I_{t,i}$ which include observation noise. In addition to the usual parameters of deterministic production models, process- and observation variance parameters, σ_B, σ_F, $\sigma_{I,i}$, and σ_C, are estimated from data (if possible), while the unobserved processes B_t and F_t are treated as random effects. The spline based representation of seasonal fisheries requires ϕ as an additional parameter vector, while the coupled SDE system requires the parameters λ and σ_U as well as the unobserved processes $U_{1,t}$ and $U_{2,t}$, which are also treated as random effects.

2.2 Including prior information

Auxiliary information can, if available, be incorporated in a Bayesian estimation framework using so-called informative priors, which are probability distributions that narrow the range of the model parameters they target. Including priors typically stabilises model fitting and reduces uncertainty of estimated quantities. However, the opposite is true if information contained in the data contradicts prior information. In a data-limited situation where some model parameters may be difficult to estimate, vaguely informative priors may be specified to aid model convergence and identifiability. This approach can be regarded as a compromise between fixing parameters and estimating them unconstrained (Magnusson & Hilborn, 2007). It is, of course, imperative that priors are only included if their specification relies on a solid foundation such as meta-analyses or independent data. Particular caution is required if informative priors are specified for n, r, m, or K, as these

14

²⁹³ are the main parameters determining management quantities.

2.3 Estimates and intervals

²⁹⁵ Both frequentist and Bayesian inference of model parameters are possible. In a frequent-

²⁹⁶ ist framework, model parameters are estimated by maximising the log-likelihood function

²⁹⁷ (Pawitan, 2013). In case of available *a priori* information, prior distributions are multi-

²⁹⁸ plied with the likelihood function to obtain the posterior distribution. Bayesian maximum

²⁹⁹ *a posteriori* parameter estimates are thus located at the maximum of the posterior distri-

³⁰⁰ bution (Box & Tiao, 2011). While confidence intervals of the frequentist framework and

³⁰¹ credible intervals of the Bayesian framework have philosophical differences they are in this

³⁰² study both approximated using the curvature of the objective function at the optimum

³⁰³ and are therefore both abbreviated CI.

2.4 Model checking

³⁰⁵ An important step in fitting stochastic models is the post-hoc evaluation of the quality of

³⁰⁶ the model fit as expressed by the residuals. Generally, for a state-space model the so-called

³⁰⁷ one-step-ahead (OSA) residual r_{t_j} at time t_j is

$$r_{t_j} = \frac{Y_{t_j} - E(Y_{t_j}|Y^{t_{j-1}})}{\sqrt{\text{var}(Y_{t_j}|Y^{t_{j-1}})}}, \tag{18}$$

³⁰⁸ where $E(Y_{t_j}|Y^{t_{j-1}})$ is the OSA prediction of the observation Y_{t_j} given $Y^{t_{j-1}} = \{Y_{t_1}, \ldots, Y_{t_{j-1}}\}$,

³⁰⁹ and $\text{var}(Y_{t_j}|Y^{t_{j-1}})$ is the variance of this prediction (Harvey, 1990). We calculate r_{t_j} for

³¹⁰ $j \in \{2, \ldots, N_{obs}\}$, where $N_{obs} = N_{obsC} + \sum_i N_{obsI,i}$ is the total number of available obser-

³¹¹ vations. The vector $Y^{t_{N_{obs}}}$, which comprises all catch and index observations, is ordered

³¹² after the times the data points are observed. Thus, while indices are assumed to be ob-

³¹³ served as snapshots, catches are observed at the end of the interval they refer to (i.e. C_t

15

is observed at $t + \Delta t$). Residual checks are performed separately for each data series of catches and indices to enable identification of the problematic model component.

If the estimating model is equal to the data-generating model the OSA residuals (onwards simply referred to as residuals) should be independent and standard normally distributed. Model deficiencies can therefore be indicated by checking whether the residuals display these properties. Possible violation of the independence assumption can be checked by plotting the empirical autocorrelation function or by using the test of Ljung & Box (1978). Apparent autocorrelation in the residuals indicates that the model lacks the structure to appropriately describe underlying dependencies in the data. As a consequence the obtained parameter estimates and associated CIs cannot be trusted. Additionally, residuals should also be tested for normality, e.g. using the test of Shapiro & Wilk (1965), and bias (mean different from zero) using a standard t-test.

2.5 Implementation

The model is implemented using Template Model Builder (TMB, Kristensen *et al.*, 2015), which is a recently developed estimation framework for R (R Core Team, 2015). TMB is efficient in fitting models with many random effects and is therefore well-suited for estimation of state-space models. The temporal dimension of the continuous-time model is resolved numerically using an Euler scheme (Iacus, 2009), which discretises time into intervals of fixed length dt_{Euler}. The number of time intervals per year is therefore $1/dt_{Euler}$. To stabilise parameter estimation and calculation of the initial one-step-ahead residuals we impose a wide normal distribution on the initial states $\log(B_t/K)$ and $\log(F_t)$ with mean $\log(0.8)$, on $\log(n)$ with mean $\log(2)$, and on $\log(\alpha)$ and $\log(\beta)$ with mean $\log(1)$. All distributions have a standard deviation of 10. These constraints on model parameters can also be interpreted as uninformative priors. Note, however, that these distributions are imposed purely for numerical reasons and should not have other impact than improving

numerical stability and avoiding extreme nonsensical estimates.

R and TMB code for fitting the model is bundled in an R package included as part of the supplementary information. The package also includes routines for postprocessing, model checking and plotting.

2.6 Examples

2.6.1 South Atlantic albacore

To facilitate a comparison with previously published surplus production models and estimation approaches we fitted SPiCT to the South Atlantic albacore (*Thunnus alalunga*, scombridae) dataset of Polacheck *et al.* (1993). The dataset comprise $N_{obsC} = N_{obsI} = 23$ years of catch and index pairs. The data contained too little information to allow unconstrained estimation of parameters α, β and n. First, we therefore fixed $\alpha = 4$ similar to the estimates found by Meyer & Millar (1999), $\beta = 1$, and $n = 2$ resulting in a quadratic production curve (SPiCT1). Second, as an alternative to fixing parameters (SPiCT2), we applied vague normally distributed priors to α, β, and n with mean parameters equal to the fixed values of the SPiCT1 case and standard deviations of 2 in the log domain (Table 1). In both SPiCT1 and SPiCT2 we set $dt_{Euler} = 1/16$ year. Finally, we fitted ASPIC version 7.02 (Prager, 1994), which is a continuous-time observation error model, to contrast the discrete-time observation error and process error models of Polacheck *et al.* (1993) and the state-space model of Meyer & Millar (1999). Confidence intervals were obtained using the bootstrap module of ASPIC with 1000 samples.

Similar to Polacheck *et al.* (1993) results are presented in terms of model parameters K, r, and q, B_{1990} (the predicted biomass in 1990), B_{1990}/K (the biomass depletion relative to K in 1990), the estimated MSY, and the estimated optimal effort $E_{MSY} = F_{MSY}/q$. For SPiCT the stochastic reference points are reported, the residuals were checked for

autocorrelation and tested for bias and normality, and a short-term forecast under a constant F_t scenario was included to illustrate the ability of the model to forecast catch including uncertainty. To facilitate comparison of the estimation methods we calculate estimated quantities relative to the estimates of the observation error model of Polacheck *et al.* (1993).

2.6.2 Simulation study 1

The purpose of simulation study 1 study was to quantify the estimation performance of SPiCT in terms of estimation stability (proportion of converged runs), estimation precision (expressed by the coefficient of variation, CV, of estimates), the coverage of 95% CIs (proportion containing the true value), and the median bias of estimates. These quantities were evaluated for eight variants of SPiCT (Table S2), ASPIC version 7.02 (Prager, 1994), with particular focus on the influence of fixing and misspecifying the parameters n, α, and β, which can be difficult to estimate. Subsequently, we also fitted the model of Meyer & Millar (1999) and compared with SPiCT using identical informative priors for the two models. See supplement S2 for detailed description.

2.6.3 Simulation study 2

The purpose of simulation study 2 was to assess the difference in estimation performance between a continuous-time model fitted to quarterly data containing within-year seasonal variation and a discrete time model ($dt_{Euler} = 1$ year) fitted to annual data obtained by aggregating the quarterly data. In practice catch and biomass index data are often resolved by annual time steps, however for some stocks data with a higher temporal resolution are available. While quarterly data contain four times the number of observations of the corresponding aggregated annual data, they do not necessarily contain four times the information. This is because increasing the sampling frequency of a process typically

18

387 results in increased autocorrelation of data and is unlikely to substantially increase contrast

388 compared to annual data.

389   For simulation we used parameter values found by fitting SPiCT to the South Atlantic

390 albacore dataset of Polacheck *et al.* (1993) while fixing $\alpha = 1$, $\beta = 1$ and $n = 2$. Seasonal

391 variation in F_t was included using a spline-based model with $\phi = (0.05, 0.1, 1.8)$ resulting

392 in low values of F_t in quarters two and three, and high values of F_t in quarter four relative

393 to quarter one. We simulated 30 years of biomass and fishing dynamics using a fine

394 time step ($dt_{Euler} = 1/64$ year) and collected quarterly ($N_{obsC,Q} = N_{obsI,Q} = 120$) and

395 corresponding annual ($N_{obsC,A} = N_{obsI,A} = 30$) datasets of catches and biomass index.

396   We fitted SPiCT using $dt_{Euler} = 1/32$ year to the quarterly data with $n = 2$ and $\beta = 1$

397 fixed to their true values while σ_B and σ_I were estimated separately. The continuous-time

398 model estimated ϕ to fit the seasonal variation in F_t using the spline-based approach. We

399 also fitted the model using $dt_{Euler} = 1$ year (i.e. a discrete-time model) to the annual

400 data with the same parametrisation (excluding ϕ). The procedure was repeated 1000

401 times. To assess the difference in general utility of the models the proportion of converged

402 estimations for each model was calculated. Estimation performance was then summarised

403 for the datasets for which both models converged in terms of median CV and coverage of

404 95% CIs of F_{MSY}, B_{MSY}, MSY, σ_B, and σ_I. As the CV scales with the inverse square

405 root of the number of independent observations we also calculated $J = (CV_A/CV_Q)^2$ for

406 each of the five quantities, to express the increase in effective sample size. The squared

407 ratio of the CVs (J) is proportional to the ratio of the effective sample sizes (i.e. the

408 number of independent observations) of quarterly data to annual data and therefore an

409 indicator of information gain.

2.6.4   North Sea stocks

To illustrate the utility of SPiCT we analysed sub-annual data from five North Sea fish stocks in the period 1975-2006: Cod (*Gadus morhua*, gadidae) with $N_{obsC} = 124$ and $N_{obsI} = 61$, whiting (*Merlangius merlangus*, gadidae) with $N_{obsC} = 124$ and $N_{obsI} = 60$, haddock (*Melanogrammus aeglefinus*, gadidae) with $N_{obsC} = 124$ and $N_{obsI} = 60$, herring (*Clupea harengus*, clupeidae) with $N_{obsC} = 124$ and $N_{obsI} = 60$, and Norway pout (*Trisopterus esmarkii*, gadidae) with $N_{obsC} = 119$ and $N_{obsI} = 57$.

Quarterly information on total catch was obtained from ICES (2005). As quarterly catches have not been compiled in the more recent years, and some of the included stocks have minor data issues requiring in-depth scrutiny that are beyond the scope of this paper, our results for these stocks should be viewed solely as illustrative.

The fish stocks of the North Sea are subject to substantial fishing pressure and are therefore surveyed extensively by scientific vessels. We gathered survey data for the time period 1975-2006 from the ICES DATRAS database (ICES, 2012) and calculated indices of exploitable stock biomass (ESB) by weighting age-structured survey catches with the ratio of age-specific commercial selectivity to survey selectivity. Specifically, the index is $I_{ESB} = \sum_a Pc_a/Ps_a Ns_a Ws_a$, where Pc_a and Ps_a are commercial and survey selectivities respectively, and Ns_a and Ws_a are numbers and mean weight at age a in the surveys respectively (time index omitted for simplicity). Thus, calculating I_{ESB} requires knowledge or assumptions about selectivities. Here we approximate the ratio Pc_a/Ps_a by the ratio Qc_a/Qs_a, where Qc_a is the proportion of the commercial catches at age a and Qs_a is the proportion of the survey catches at age a. For small sample sizes this is a crude approximation in which case simple assumptions regarding selectivities may be more robust. If available, sampling CVs of the index can be used to weight index observations, however this information was not part of the North Sea dataset.

Owing to variable survey frequency index time series contain annual, bi-annual and quarterly time steps between observations. With the continuous-time formulation of SPiCT it is possible to analyse the variably sampled data directly without the need to aggregate observations into fixed time steps.

We assessed to which extent models of different complexity were identifiable given the data by first estimating a baseline model with a spline-based seasonal representation while fixing $\alpha = 1$ and $n = 2$, which are standard assumptions (Ono *et al.*, 2012; Thorson *et al.*, 2013) and fixing $\beta = 1$. All remaining model parameters (K, m, q, $\sigma_B = \sigma_I$, $\sigma_F = \sigma_C$) were estimated. We then released the fixed parameters for estimation in the following order: α, β, n resulting in three models to be estimated in addition to the baseline. We set $dt_{Euler} = 1/32$ year for all model fits.

Residuals were analysed to diagnose model insufficiencies. If significant residual auto-correlation was detected in catches as an indication of potential shifts in the seasonal fishing pattern we ran the more flexible coupled SDE seasonal model and reexamined residuals. If residuals deviated from normality possibly as a result of outlying observations we reanalysed data using robust observation error distributions for either or both catch and index observations. Models were selected as the ones with best residual diagnostics (fewest violations in terms of autocorrelation, bias and non-normality). If models performed equally in terms of diagnostics we selected models with highest complexity (most free model parameters) to give examples of estimated values of α, β and n. Aspects of formal statistical model selection are outside the scope of this study.

When fitting production models a constraint is often required to estimate absolute levels of biomass and fishing mortality (B_t and F_t respectively). This was the case for cod and whiting for which a bound was imposed on B_t in the year (y) the maximum catch was observed by applying a vague prior distribution based on information in the catch data. For stocks that have been subjected to sustained and relatively high commercial

21

exploitation for decades it seems fair to assume that current biomass is not many orders of magnitude larger than the observed catches. As prior distribution for $\log(B_y)$ we therefore used $N(\mu_{B_y}, \sigma^2_{B_y})$, with $\mu_{B_y} = \log(k \max C_t)$ and $\sigma_{B_y} = 3$. Note that a standard deviation of 3 in the log domain is a rather uninformative prior. The specific choice of k has minimal impact on the results as long as μ_{B_y} is in the same order of magnitude as the maximum catch $(1 \leq k \leq 9)$ and the prior is uninformative $(\sigma_{B_y} \geq 3)$. In the current study we set $k = 3$.

3 Results

3.1 South Atlantic albacore

Fitting SPiCT1 to the South Atlantic albacore dataset of Polacheck *et al.* (1993) produced results comparable to those obtained using alternative approaches (Fig. 1). The point estimates of SPiCT1 were particularly similar to those obtained using observation error estimators, however the point estimates of alternative approaches were contained in the 95% CI of the SPiCT1 point estimates. SPiCT2 produced a wider CI of E_{MSY} while CIs of K and B_{1990} were narrower compared to CIs of SPiCT1. CIs of remaining parameters were largely similar between SPiCT1 and SPiCT2. Overall, among all methods estimates of MSY were nearly identical and associated with the narrowest CIs.

SPiCT1 produced 95% CIs of B_t and F_t that were wide relative to the 95% CIs of B_t/B_{MSY} and F_t/F_{MSY} (Fig. 2). The reduction in CIs of relative estimates was a result of high correlation between estimates of absolute levels and reference points. For SPiCT2 this correlation was less marked and reference point estimates were uncertain resulting in inflated CIs of relative levels of biomass and fishing mortality making stock status less clear.

In terms of parameter estimates SPiCT2 obtained a significantly lower value of β

22

than 1 (Table 1) resulting in reduced estimated catch observation error (Fig. 2). Overall, estimates of α, β and n were uncertain with standard deviations in log domain of 1 (Table 1). Model checking of both SPiCT1 and SPiCT2 did not indicate any violation of assumptions regarding independence, bias and normality of residuals (Fig. S1, S2) for catch and index observations. Formal model selection would be a natural step to objectively choose between multiple models, however this topic is beyond the scope of the current study and is left to be explored in future work.

Predictions of B_t, B_t/B_{MSY}, F_t, F_t/F_{MSY} and C_t can be obtained by propagating the model beyond the time span of the data (Fig. 2), however uncertainty increases for each predicted year as a result of the lack of data. Furthermore, as the model is a highly simplified version of reality, it should not be relied on for generating long-term predictions (> 2 years) even when residuals pass all tests.

[Figure 1 about here.]

[Figure 2 about here.]

[Table 1 about here.]

3.2 Simulation study 1

Simultaneous estimation of α, β, and n (model A) was less stable than estimation of models with fixed $n = 2$, however, while parameter estimates provided by model A were uncertain, coverage of the resulting CIs was generally close to the expected 0.95. Simultaneous estimation of α, β, and n using vague priors (model H) improved both model convergence rate, median CV, and median bias, while CI coverage was unchanged relative to model A. Models with fixed $n = 2$ had lower CVs but also lower CI coverage for shorter datasets. CI coverage was generally reduced when misspecifying n, while misspecification of α and β influenced CI coverage less, in particular estimates of MSY and biomass relative to B_{MSY}

were unaffected. Misspecified models were generally biased when estimating absolute quantities (F_{MSY}, B_{MSY}, B_t) and unbiased when estimating relative quantities (F_t/F_{MSY}, B_t/B_{MSY}) and MSY. Despite its simpler model structure, ASPIC showed comparable performance to SPiCT for these three quantities.

Using informative priors SPiCT had a convergence rate close to one, while the model of Meyer & Millar (1999) had a convergence rate of 0.6 on average. Median CVs of both models decreased as the number of observations increased, however while CIs of SPiCT generally converged to the expected 95% the CIs of the Meyer & Millar (1999) model diverged for increasing number of observations. SPiCT estimates of absolute quantities were slightly biased ($\sim$5%) while relative quantities and MSY were unbiased. The Meyer & Millar (1999) model generally produced biased estimates of absolute quantities while MSY and B_{last}/B_{MSY} were largely unbiased.

See supplement S2 for detailed description of results.

3.3 Simulation study 2

The convergence rate was 99.9% for both the discrete-time model fitted to the annual data and the continuous-time model fitted to quarterly data. The results of the runs for which both models converged (Table 2) did not show a significant difference in median CVs of estimated MSY, while the continuous-time model obtained significantly improved CVs for the remaining quantities. The improvement in CV of B_{MSY} corresponded to a doubling of the effective sample size, while CVs of process and observation error standard deviations were reduced even further as a result of using quarterly data (Table 2). In terms of CI coverage significant differences between the discrete-time and continuous-time model were found for MSY and σ_I with the most substantial improvement in coverage found for MSY.

[Table 2 about here.]

3.4   North Sea stocks

535 Of the five selected model fits α and β were estimated for all stocks except herring, while

536 n was estimated for Norway pout only (Table 3).

537   The selected fit to Norway pout data detected a clear seasonal pattern in F_t (Fig. 3).

538 Residual analysis of this fit showed no violations in terms of autocorrelation, bias or

539 non-normality of residuals when using the coupled SDE seasonal model (Table 3). Fits to

540 Norway pout data using the spline-based seasonal representation resulted in autocorrelated

541 catch residuals, however the more flexible correlation pattern allowed by the coupled SDE

542 seasonal model was able to eliminate this autocorrelation (Table S3). This illustrates

543 how an observed pattern in residuals can aid in identifying a lacking model component.

544 Similarly, the analysis of whiting resulted in no violations when using the spline-based

545 seasonal model with robust observation error for index. This model was selected as a result

546 of non-normality of index residuals (apparent outliers in index observations) detected in

547 the baseline model (Table S3).

548   The selected model for haddock used the coupled SDE seasonal representation (Table 3)

549 as a remedy to detected autocorrelation in catch residuals when using a spline-based sea-

550 sonal model (Fig. 4). Non-normality detected in index residuals could not be remedied

551 by neither a robust observation error distribution nor a spline-based seasonal model

552 (Table S3). In analysing the cod data significant non-normality was detected in catch

553 residuals. Normality could be achieved using a robust catch observation error model,

554 however at a cost of introducing autocorrelation in catch residuals, which could not be

555 eliminated using the coupled SDE seasonal model. Finally, for herring a clear seasonal

556 pattern in F_t was estimated (Fig. 3) resulting in seasonal fluctuations in B_t. Data covered

557 contrasting periods of low and high levels of biomass resulting in reference point estimates

25

with higher precision relative to those of the other stocks (Fig. 3, right column). While a robust observation error distribution for the index eliminated non-normality of index residuals, non-normality of catch residuals and autocorrelation could not be eliminated within the current model suite indicating that unmodelled variability in the herring data remains.

In identifying the process to observation noise relationships, two of four estimates of α were significantly different from the value one (Table 3), which is the commonly assumed value when estimation is not possible (Ono *et al.*, 2012; Thorson *et al.*, 2013). In contrast, only one of four estimates of β were significantly different from the value one. Similarly, the 95% CI of n for Norway pout did not exclude the commonly assumed value of two indicating a lack of evidence to deviate from a symmetric production function (Schaefer, 1954).

[Table 3 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

4   Discussion

The stochastic surplus production model in continuous-time (SPiCT) presented here is a full state-space model in that both biomass and fishing dynamics are modelled as states, which are observed indirectly through biomass indices and commercial catches sampled with error. A wide range of previously published surplus production models are nested within SPiCT: Observation error and process error estimators (Polacheck *et al.*, 1993) emerge if eliminating process or observation noise respectively. State-space models that assume catches are observed without error (Meyer & Millar, 1999; Punt, 2003; Ono *et al.*,

26

$_{581}$ 2012) are obtained by fixing the variance of the F_t-process to a large value and eliminating

$_{582}$ the observation noise on catches. Generally, discrete-time models (Meyer & Millar, 1999;

$_{583}$ Punt, 2003) are obtained by setting the temporal time step of the numerical solver (dt_{Euler})

$_{584}$ to 1 year, while continuous-time models (Schaefer, 1954; Pella & Tomlinson, 1969; Prager,

$_{585}$ 1994) arise when reducing dt_{Euler} to a value where parameter estimates do not change

$_{586}$ qualitatively if dt_{Euler} is refined further.

$_{587}$ An advantage of the continuous-time formulation is the ability of the model to ac-

$_{588}$ commodate arbitrary and irregularly sampled data without a need for catch and index

$_{589}$ observations to match temporally. It is therefore straightforward to fit the model to data

$_{590}$ containing a mix of annual, biannual and quarterly data as demonstrated in our ana-

$_{591}$ lysis of North Sea stocks. The additional information contained in quarterly observations

$_{592}$ relative to the corresponding annually aggregated data can be exploited to improve cov-

$_{593}$ erage of confidence intervals and reduce uncertainty of parameter estimates (Table 2).

$_{594}$ Furthermore, the substantial sample size provided by quarterly observations from North

$_{595}$ Sea stocks enabled estimation of process noise, observation noise and for one stock the

$_{596}$ production shape parameter (n), which are notoriously difficult to estimate using annual

$_{597}$ observations (Prager, 2002; Ono *et al.*, 2012).

$_{598}$ Explicit modelling of the fishery dynamics (F_t) as a latent process in continuous-time

$_{599}$ allows F_t to be estimated at arbitrary times without requiring temporal overlap with a

$_{600}$ catch observation. The F_t-process further allows catches to be predicted with uncertainty

$_{601}$ over any time interval by temporal integration of the product of the fishing and biomass

$_{602}$ processes. Catch predictions can then be compared with observed catches in an observation

$_{603}$ equation, which incorporates catch observation error (Eqn. 16). In this model formulation,

$_{604}$ the absence of the catch in the equation describing biomass dynamics avoids the risk of

$_{605}$ large catches leading to negative biomass estimates, which can be a problem when fitting

$_{606}$ discrete time models. The F_t and C_t components of the model are novel relative to previous

27

production models, and are first steps toward a framework that more realistically represent the uncertainty of fishing and catch observation processes.

The current model formulation, where catches are given by the temporal integral of the product of fishing and biomass, implies that B_t represents the biomass of the exploitable part of the stock and that B_{MSY} is the level of exploitable stock biomass (ESB) that maximises production. It is therefore important when calculating the index of B_t from scientific survey data to only include the size-classes targeted by the fishery. It is furthermore important to distinguish ESB from other biomass representations when comparing B_t or B_{MSY} to absolute estimates from alternative models. If the commercial selectivity curve matches the maturity curve of the stock then ESB can be interpreted as the spawning stock biomass (SSB), which is commonly reported from e.g. age-based models. Generally, the relative biomass, B_t/B_{MSY}, is less sensitive to the choice of biomass representation making it a robust estimator of stock status.

Currently, the main mechanism driving the fishing dynamics is seasonal variations in effort, which is modelled by imposing an annual correlation structure on the fishing mortality. The analyses of North Sea stocks showed that estimation of seasonal variation in fishing is possible if sub-annual catch observations are available (Figure 3). A potential extension of the fishing dynamics model could include economic components and allow estimation of bioeconomic quantities (Thorson *et al.*, 2013), however at a cost of increased data requirements.

In fisheries management it is useful to evaluate the implications of management decisions on stock status and future catches. Such evaluations can be made by predicting catches using Eqn. 16 under different fishing scenarios or by fixing future catches and predicting the corresponding levels of fishing mortality and biomass including associated prediction intervals. However, as the model is a highly simplified version of reality, it is not suited for generating long-term predictions (> 2 years).

28

While catch observations are aggregated over a time period, biomass index observations are assumed to be instantaneous snapshots. For survey based indices, where data are often gathered within weeks, this assumption seems reasonable. However, if the data collection period spans months or perhaps the whole year as for commercial CPUE data the assumption can be questioned. In such cases, an alternative to using biomass index data is to use commercial effort data directly in the model as an indicator of F_t. Such an extension could be readily implemented within the presented modelling framework.

In statistical modelling it is customary to conduct model checking by inspecting residuals for violation of independence and distributional assumptions. Deviating residual patterns can be used to diagnose model insufficiencies. Non-normal or biased residuals indicate presence of an unmodelled trend in data, e.g. a temporal shift in catchability, or presence of extreme outlying observations. If outliers are detected one may apply methods for outlier removal (Prager, 2002) or alternatively, as illustrated in the analysis of North Sea stocks (Table 3), shift to a robust model formulation e.g. by assuming a heavy-tailed distribution for the data series where outliers occur. Seasonal autocorrelation in residuals may be accounted for by including a seasonal component in the fisheries model. Here we represented seasonal dynamics using either a cyclic spline (Eqn. 12) or the more flexible oscillatory system of coupled SDEs (Eqn. 15), which can adapt to shifts in timing and amplitude of the seasonal fishing pattern and reduce residual autocorrelation (Fig. 4).

Persisting residual patterns may be caused by violations of assumptions related to observations, i.e. differences in biomass index and fisheries data in terms of spatial coverage or uncorrected mismatch between commercial selectivity and survey selectivity. Residual patterns can also be a result of significant changes in biomass and catchability due to migration, shifts in technology or fishing technique, trends in growth rate imposed by environmental changes etc. Extending the biomass dynamic model e.g. to a stage-based form could capture, more realistically, the selectivity of the fishery and the lagged dynamics

29

induced by interactions between juveniles and adults through recruitment and maturation. Inclusion of environmental covariates such as temperature, nutrient, or oxygen information may be used to induce longer-term trends in growth, which, if unmodelled, would result in biased estimates. Complete treatment of the course of action when model assumptions appear violated is outside the scope of this study. However, it is important to include results of residual analyses to enable an honest presentation of model results. If possible, it is furthermore preferable to compare results with those of alternative model classes (catch-only, age-structured, length-based etc.) in particular if residuals indicate a critical lack of fit.

Formal checking of statistical models is key in obtaining valid parameter estimates, reliable confidence intervals, and useful quantities for model selection. In general, model checking has philosophical as well as practical aspects. Idealists would argue that a model producing residuals that violate assumptions lacks the model components required to completely describe data. However, as all models are simplifications of reality and therefore wrong, they can never be expected to fully explain observations. Indeed, surplus production models represent a highly simplified reality where many poorly understood mechanisms such as species interactions and environmental effects are modelled as random variability. Thus adopting a pragmatic approach to model checking allowing mild departures from assumptions may be required.

The flexibility of the presented model is obtained by explicit modelling of both the biomass process (B_t) and the fisheries process (F_t). Estimation of the potentially large number of resulting random effects is facilitated by TMB via the Laplace approximation (Kristensen et $al.$, 2015). In calculating confidence or credible intervals TMB therefore relies on the assumption that the objective function is quadratic around the optimum, which in the frequentist framework is the asymptotic shape under certain regularity conditions (Wasserman, 2013). Thus, in a data-limited scenario asymptotic assumptions may not

30

hold resulting in decreased coverage of confidence intervals. In our simulation the Laplace approximation provided 95% confidence intervals ranging from 0.8-0.9 coverage when 15 observations of catch and index were available to the expected 0.95 coverage at 120-240 observations (Fig. S5). When analysing data containing a limited number of observations it is thus advisable to verify confidence interval coverage using bootstrap or likelihood profiles (de Valpine & Hilborn, 2005).

Comparing a deterministic model (ASPIC, Prager, 1994) and a discrete-time model (Meyer & Millar, 1999) with a stochastic model in continuous-time (SPiCT) requires simplifying assumptions to become tractable. The presented comparison assumes that a realistic representation of population dynamics evolves continuously in time and involves some degree of random variability. Fitting SPiCT to data generated by such a system generally produced unbiased parameter estimates (Fig. S6) with reliable confidence intervals (Fig. S5), while intervals produced by alternative methods were unreliable in particular for larger datasets (60-240 observations). In a data-limited situation (15-30 observations), all models produced biased estimates of F_{MSY} and B_{MSY}. In contrast, estimates of stock status (F/F_{MSY} and B/B_{MSY}) and MSY performed well across models and are thus the quantities for which production models in general provide the most robust inference.

There are several well-known complications associated with fitting surplus production models: First, as the input data are often collected annually it is common to have less than say 50 years of data available for estimation. Data scarcity can lead to problems with estimation instability and model identifiability. Second, reliable parameter estimation require sufficient contrast in data (Hilborn *et al.*, 1992). Lack of contrast can result in identifiability problems and high correlation among model parameters thus making the estimation unstable. Third, estimating both process and observation noise is notoriously difficult (Polacheck *et al.*, 1993; de Valpine & Hilborn, 2005) and may necessitate switching to either a process error or observation error model, or assume that the ratios of process

31

error to observation error (α and β) are known (Ono *et al.*, 2012; Thorson *et al.*, 2013). Fourth, the three mentioned points apply to the case where $n = 2$ (Schaefer, 1954). Thus, the increased nonlinearity of the generalized form (Pella & Tomlinson, 1969) likely exacerbates the mentioned estimation complications.

We do not claim to have solved the above mentioned problems. However, to mitigate potential stability issues we have used the improved parameterisation of Fletcher (1978) and have used TMB for model estimation, which relies on analytical derivatives of the objective function to make estimation as stable as possible (Kristensen *et al.*, 2015). Extensive simulation testing was facilitated by the efficiency of TMB enabling computing times of parameter estimates in the order of seconds. These aspects are key to obtaining stable fits of nonlinear models (Bolker *et al.*, 2013).

Estimating the shape of the production curve (determined by n) is critical to management because the resulting reference points and associated stock status are sensitive to the value of n (Eqns. 6-8). Our simulation experiment showed stable performance and high coverage of confidence intervals when simultaneously estimating process noise, observation noise and n even for limited data (Fig. S5). Additionally, the model was able to estimate n in one of five North Sea stocks (Table 3). These results are based on unconstrained estimation of n. Our simulation results demonstrated that estimation can be stabilised by translating meta-analyses (e.g. Thorson *et al.*, 2012) or knowledge of stock dynamics into a prior distribution for n (Fig. S3). Using a vague prior enabled n to be estimated for the South Atlantic albacore dataset of Polacheck *et al.* (1993) resulting in differences with potential implications for management relative to results using fixed $n = 2$ (Fig. 2). The improved stability of the presented modelling framework enables stock assessment scientists to explore the possibility of estimating n rather than relying on the common assumption that $n = 2$, which may result in biased reference point estimates and poor management decisions (Maunder, 2003).

Given an MSY-based approach to management, the modelling framework developed here focuses on honest reporting of model results through uncertainty quantification and model checking. By relaxing the common assumption that catches are known without error, a more realistic quantification of uncertainty on all reported quantities is obtained while residual diagnostics clarifies whether model assumptions are significantly violated. As a benefit of the continuous-time formulation quarterly resolved data may, if available, aid in reducing estimation uncertainty (Table 2) and allow seasonal predictions and management. Using the presented model, we have also demonstrated that while estimates of B_t, F_t, B_{MSY} and F_{MSY} are often highly uncertain, the relative quantities F_t/F_{MSY} and in particular B_t/B_{MSY} may have considerably less uncertainty (Fig. S4) and bias (Fig. S6) and should therefore form the primary basis of management decisions.

5 Acknowledgements

33

References

Bolker, B. M., Gardner, B., Maunder, M., Berg, C. W., Brooks, M., Comita, L., Crone, E., Cubaynes, S., Davies, T., Valpine, P. *et al.* (2013). Strategies for fitting nonlinear ecological models in R, AD Model Builder, and BUGS. *Methods in Ecology and Evolution*, **4**, 501–512.

Bordet, C. & Rivest, L.-P. (2014). A stochastic Pella Tomlinson model and its maximum sustainable yield. *Journal of theoretical biology*, **360**, 46–53.

Box, G. E. & Tiao, G. C. (2011). *Bayesian inference in statistical analysis*, *Volume 40*. John Wiley & Sons.

Brodziak, J. & Ishimura, G. (2011). Development of Bayesian production models for assessing the North Pacific swordfish population. *Fisheries Science*, **77**, 23–34.

Carvalho, F., Ahrens, R., Murie, D., Ponciano, J. M., Aires-da Silva, A., Maunder, M. N. & Hazin, F. (2014). Incorporating specific change points in catchability in fisheries stock assessment models: an alternative approach applied to the blue shark (Prionace glauca) stock in the south Atlantic Ocean. *Fisheries Research*, **154**, 135–146.

Chen, Y., Breen, P. & Andrew, N. (2000). Impacts of outliers and mis-specification of priors on Bayesian fisheries-stock assessment. *Canadian Journal of Fisheries and Aquatic Sciences*, **57**, 2293–2305.

Chen, Y., Jackson, D. & Paloheimo, J. (1994). Robust regression approach to analyzing fisheries data. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**, 1420–1429.

de Valpine, P. & Hilborn, R. (2005). State-space likelihoods for nonlinear fisheries time-series. *Canadian Journal of Fisheries and Aquatic Sciences*, **62**, 1937–1952.

34

Fletcher, R. (1978). On the restructuring of the Pella-Tomlinson system. *Fish. Bull*, **76**, 515–521.

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

Gardiner, C. W. (1985). *Handbook of stochastic methods*, *Volume 4*. Springer Berlin.

Harvey, A. C. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge university press.

Hilborn, R., Walters, C. J. *et al.* (1992). Quantitative fisheries stock assessment: choice, dynamics and uncertainty. *Reviews in Fish Biology and Fisheries*, **2**, 177–178.

Iacus, S. M. (2009). *Simulation and inference for stochastic differential equations: with R examples*, *Volume 1*. Springer.

ICES (2005). Report of the Study Group on Multi Species Assessment in the North Sea. *ICES Document CM 2005/D:06*.

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

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H. J. & Bell, B. (2015). TMB: Automatic Differentiation and Laplace Approximation. ArXiv e-print; in press, *Journal of Statistical Software*.

Ljung, G. M. & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, **65**, 297–303.

Magnusson, A. & Hilborn, R. (2007). What makes fisheries data informative? *Fish and Fisheries*, **8**, 337–358.

Maunder, M. N. (2003). Is it time to discard the Schaefer model from the stock assessment scientist's toolbox? *Fisheries Research*, **61**, 145–149.

Meyer, R. & Millar, R. B. (1999). BUGS in Bayesian stock assessments. *Canadian Journal of Fisheries and Aquatic Science*, **1086**, 1078–1086.

Øksendal, B. (2013). *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media.

Ono, K., Punt, A. E. & Rivot, E. (2012). Model performance analysis for Bayesian biomass dynamics models using bias, precision and reliability metrics. *Fisheries Research*, **125**, 173–183.

Pawitan, Y. (2013). *In all likelihood: statistical modelling and inference using likelihood*. OUP Oxford.

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.

Pella, J. J. & Tomlinson, P. K. (1969). A generalized stock production model. *Bulletin of the Inter-American Tropical Tuna Commission*, **13**, 421–458.

Polacheck, T., Hilborn, R. & Punt, A. E. (1993). Fitting Surplus Production Models: Comparing Methods and Measuring Uncertainty. *Canadian Journal of Fisheries and Aquatic Science*.

Prager, M. H. (1994). A suite of extensions to a nonequilibrium surplus-production model. *Fishery Bulletin*, **389**, 374–389.

Prager, M. H. (2002). Comparison of logistic and generalized surplus-production models

applied to swordfish, Xiphias gladius, in the north Atlantic Ocean. *Fisheries Research*, **58**, 41–57.

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

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Schaefer, M. B. (1954). Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. *Bulletin of the Inter-American Tropical Tuna Commission*, **1**, 25–56.

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

Shapiro, S. S. & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, pp. 591–611.

Smith, M. T. & Addison, J. T. (2003). Methods for stock assessment of crustacean fisheries. *Fisheries Research*, **65**, 231–256.

Thorson, J. T., Cope, J. M., Branch, T. A., Jensen, O. P. & Walters, C. J. (2012). Spawning biomass reference points for exploited marine fishes, incorporating taxonomic and body size information. *Canadian Journal of Fisheries and Aquatic Sciences*, **69**, 1556–1568.

Thorson, J. T., Minto, C., Minte-Vera, C. V., Kleisner, K. M., Longo, C. & Jacobson, L. (2013). A new role for effort dynamics in the theory of harvested populations and data-poor stock assessment. *Canadian Journal of Fisheries and Aquatic Sciences*, **70**, 1829–1844.

849 Wasserman, L. (2013). *All of statistics: a concise course in statistical inference*. Springer

850 Science & Business Media.

Parameter	SPiCT1, fixed	SPiCT2, prior	SPiCT2, posterior
α	4.0	4.0 (0.073, 218)	12.3 (1.300, 117)
β	1.0	1.0 (0.018, 54.6)	0.12 (0.018, 0.78)
n	2.0	2.0 (0.037, 109)	0.75 (0.073, 7.75)

Table 1: SPiCT estimates of α, β and n using the South Atlantic albacore dataset of Polacheck *et al.* (1993) with 95% CIs in parentheses.

		F_{MSY}	B_{MSY}	MSY	σ_B	σ_I
Median CV	Annual	0.516	0.852	0.469	0.605	0.216
	Quarterly	0.500	0.605	0.480	0.289	0.095
	MW, P-value	0.010	$< 10^{-3}$	0.817	$< 10^{-3}$	$< 10^{-3}$
	J	1.066	1.980	1.000	4.391	5.182
CI coverage	Annual	0.866	0.882	0.735	0.966	0.930
	Quarterly	0.866	0.875	0.870	0.949	0.960
	Prop., P-value	1.000	0.681	$< 10^{-3}$	0.076	0.004

Table 2: Results of simulation 2 summarised by median CVs with P-values of Mann-Whitney (MW) rank sum tests of identical CVs, implied information gain (J) as given by the squared CV ratio, and coverage of 95% CIs of estimates with P-values of equal-proportions test. Results are based on the continuous-time model using 30 years of quarterly observations ($N_{obsC} = N_{obsI} = 120$) and the discrete-time model using the corresponding aggregated annual observations ($N_{obsC} = N_{obsI} = 30$).

Stock	Seasonal model	Robust error	α	β	n	Residual analysis
Herring	Spline	I_t	–	–	–	AC, NN
Norway Pout	Coupled SDE	–	0.59 (0.22, 1.58)	0.84 (0.44, 1.59)	1.27 (0.52, 3.07)	–
Haddock	Coupled SDE	–	1.53 (0.85, 2.77)	0.66 (0.42, 1.05)	–	NN
Cod	Spline	–	2.36 (1.31, 4.24)	0.27 (0.14, 0.52)	–	NN
Whiting	Spline	I_t	2.45 (1.03, 5.87)	1.32 (0.75, 2.3)	–	–

Table 3: Results of the selected fits of SPiCT to the five North Sea stocks with 95% CIs given in parentheses. Residual analyses showed no significant violation of assumptions for Norway pout and whiting, while autocorrelation (AC) was detected for herring, and non-normality (NN) was detected for herring, haddock and cod.
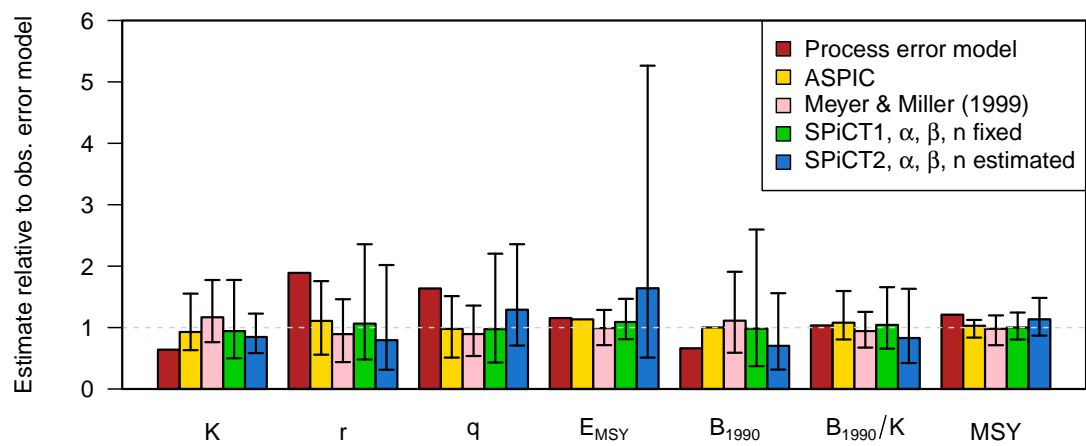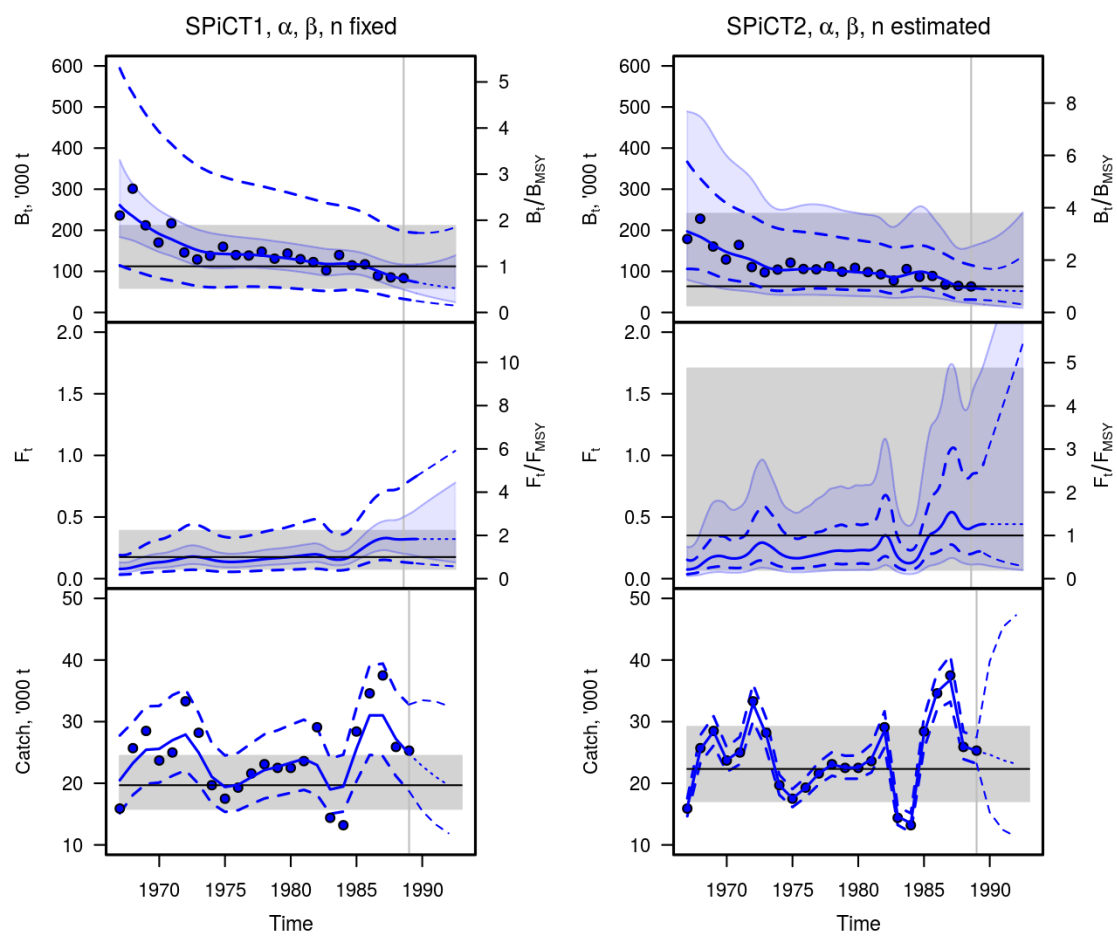
List of Figures

Figure 1

Figure 2

Figure 3

45

Figure 4

46