

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. Berg¹
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

12 **Summary (Abstract)**

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

31 **Contents**

32 **1 Introduction**

33 **2 Methods**

34 2.1 Model formulation

35 2.2 Including prior information

36 2.3 Estimates and intervals
37 2.4 Model checking
38 2.5 Implementation
39 2.6 Examples
40 2.6.1 South Atlantic albacore
41 2.6.2 Simulation study 1
42 2.6.3 Simulation study 2
43 2.6.4 North Sea stocks

44 **3 Results**

45 3.1 South Atlantic albacore
46 3.2 Simulation study 1
47 3.3 Simulation study 2
48 3.4 North Sea stocks

49 **4 Discussion**

50 **5 Acknowledgements**

51 **Key words:** Data-limited methods, fisheries management, maximum sus-
52 tainable yield, seasonal population dynamics, stock assessment, Pella-Tomlinson
53 model.

54

55 **1 Introduction**

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

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

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

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

85 are often referred to as stochastic.

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

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

110 Stochastic surplus production models aim to account for autocorrelation inherent in

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

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

132 2 Methods

133 2.1 Model formulation

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

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

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

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

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

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

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

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)

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

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

172 model fitting, lead to instability problems. By Lamperti transforming the model (Iacus,
173 2009) an additive noise term is obtained

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

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

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

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

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

178 Determining whether the stochastic reference points hold for $n < 1$ is an area of ongoing
179 research and outside the scope of the current study. It is evident that $F_{MSY} < F_{MSY}^d$
180 when $\sigma_B^2 > 0$, which implies that the presence of process noise entails a conservative
181 reduction in recommended fishing pressure relative to the deterministic case, however the
182 opposite behaviour is observed if Eqn. (7) is assumed to hold for $n < 1$. We therefore use
183 the stochastic reference points when $n > 1$ and the deterministic reference points when
184 $0 < n \leq 1$.

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

$$E(B_\infty|F_t) = K \left(1 - \frac{(n-1)}{n} \frac{F_t}{F_{MSY}^d}\right)^{1/(n-1)} \left(1 - \frac{n/2}{1 - (1 - nF_{MSY}^d + [n-1]F_t)^2 \sigma_B^2}\right). \quad (9)$$

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

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

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

$$F_t = S_t G_t \quad (10)$$

$$d \log G_t = \sigma_F dV_t \quad (11)$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $i = 1, \dots, N_i$) 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

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

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 i th index observation error, and q_i is a catchability parameter for the i th 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-

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

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

281 **2.2 Including prior information**

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

293 are the main parameters determining management quantities.

294 **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.

304 **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

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

308 where $E(Y_{t_j}|Y^{t_{j-1}})$ is the OSA prediction of the observation Y_{t_j} given $Y^{t_{j-1}} = \{Y_{t_1}, \dots, Y_{t_{j-1}}\}$,
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 $j \in \{2, \dots, N_{obs}\}$, 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

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

316 If the estimating model is equal to the data-generating model the OSA residuals (on-
317 wards simply referred to as residuals) should be independent and standard normally dis-
318 tributed. Model deficiencies can therefore be indicated by checking whether the residuals
319 display these properties. Possible violation of the independence assumption can be checked
320 by plotting the empirical autocorrelation function or by using the test of Ljung & Box
321 (1978). Apparent autocorrelation in the residuals indicates that the model lacks the struc-
322 ture to appropriately describe underlying dependencies in the data. As a consequence the
323 obtained parameter estimates and associated CIs cannot be trusted. Additionally, resid-
324 uals should also be tested for normality, e.g. using the test of Shapiro & Wilk (1965), and
325 bias (mean different from zero) using a standard t -test.

326 **2.5 Implementation**

327 The model is implemented using Template Model Builder (TMB, Kristensen *et al.*, 2015),
328 which is a recently developed estimation framework for R (R Core Team, 2015). TMB
329 is efficient in fitting models with many random effects and is therefore well-suited for es-
330 timation of state-space models. The temporal dimension of the continuous-time model is
331 resolved numerically using an Euler scheme (Iacus, 2009), which discretises time into in-
332 tervals of fixed length dt_{Euler} . The number of time intervals per year is therefore $1/dt_{Euler}$.
333 To stabilise parameter estimation and calculation of the initial one-step-ahead residuals we
334 impose a wide normal distribution on the initial states $\log(B_t/K)$ and $\log(F_t)$ with mean
335 $\log(0.8)$, on $\log(n)$ with mean $\log(2)$, and on $\log(\alpha)$ and $\log(\beta)$ with mean $\log(1)$. All
336 distributions have a standard deviation of 10. These constraints on model parameters can
337 also be interpreted as uninformative priors. Note, however, that these distributions are
338 imposed purely for numerical reasons and should not have other impact than improving

339 numerical stability and avoiding extreme nonsensical estimates.

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

343 2.6 Examples

344 2.6.1 South Atlantic albacore

345 To facilitate a comparison with previously published surplus production models and es-
346 timation approaches we fitted SPiCT to the South Atlantic albacore (*Thunnus alalunga*,
347 scombridae) dataset of Polacheck *et al.* (1993). The dataset comprise $N_{obsC} = N_{obsI} = 23$
348 years of catch and index pairs. The data contained too little information to allow uncon-
349 strained estimation of parameters α , β and n . First, we therefore fixed $\alpha = 4$ similar to
350 the estimates found by Meyer & Millar (1999), $\beta = 1$, and $n = 2$ resulting in a quadratic
351 production curve (SPiCT1). Second, as an alternative to fixing parameters (SPiCT2),
352 we applied vague normally distributed priors to α , β , and n with mean parameters equal
353 to the fixed values of the SPiCT1 case and standard deviations of 2 in the log domain
354 (Table 1). In both SPiCT1 and SPiCT2 we set $dt_{Euler} = 1/16$ year. Finally, we fitted
355 ASPIC version 7.02 (Prager, 1994), which is a continuous-time observation error model, to
356 contrast the discrete-time observation error and process error models of Polacheck *et al.*
357 (1993) and the state-space model of Meyer & Millar (1999). Confidence intervals were
358 obtained using the bootstrap module of ASPIC with 1000 samples.

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

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

368 **2.6.2 Simulation study 1**

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

378 **2.6.3 Simulation study 2**

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

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.

410 2.6.4 North Sea stocks

411 To illustrate the utility of SPiCT we analysed sub-annual data from five North Sea fish
412 stocks in the period 1975-2006: Cod (*Gadus morhua*, gadidae) with $N_{obsC} = 124$ and
413 $N_{obsI} = 61$, whiting (*Merlangius merlangus*, gadidae) with $N_{obsC} = 124$ and $N_{obsI} =$
414 60 , haddock (*Melanogrammus aeglefinus*, gadidae) with $N_{obsC} = 124$ and $N_{obsI} = 60$,
415 herring (*Clupea harengus*, clupeidae) with $N_{obsC} = 124$ and $N_{obsI} = 60$, and Norway pout
416 (*Trisopterus esmarkii*, gadidae) with $N_{obsC} = 119$ and $N_{obsI} = 57$.

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

421 The fish stocks of the North Sea are subject to substantial fishing pressure and are
422 therefore surveyed extensively by scientific vessels. We gathered survey data for the time
423 period 1975-2006 from the ICES DATRAS database (ICES, 2012) and calculated indices
424 of exploitable stock biomass (ESB) by weighting age-structured survey catches with the
425 ratio of age-specific commercial selectivity to survey selectivity. Specifically, the index is
426 $I_{ESB} = \sum_a P_{c_a}/P_{s_a} N_{s_a} W_{s_a}$, where P_{c_a} and P_{s_a} are commercial and survey selectivities
427 respectively, and N_{s_a} and W_{s_a} are numbers and mean weight at age a in the surveys re-
428 spectively (time index omitted for simplicity). Thus, calculating I_{ESB} requires knowledge
429 or assumptions about selectivities. Here we approximate the ratio P_{c_a}/P_{s_a} by the ratio
430 Q_{c_a}/Q_{s_a} , where Q_{c_a} is the proportion of the commercial catches at age a and Q_{s_a} is the
431 proportion of the survey catches at age a . For small sample sizes this is a crude approx-
432 imation in which case simple assumptions regarding selectivities may be more robust. If
433 available, sampling CVs of the index can be used to weight index observations, however
434 this information was not part of the North Sea dataset.

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

439 We assessed to which extent models of different complexity were identifiable given the
440 data by first estimating a baseline model with a spline-based seasonal representation while
441 fixing $\alpha = 1$ and $n = 2$, which are standard assumptions (Ono *et al.*, 2012; Thorson *et al.*,
442 2013) and fixing $\beta = 1$. All remaining model parameters ($K, m, q, \sigma_B = \sigma_I, \sigma_F = \sigma_C$)
443 were estimated. We then released the fixed parameters for estimation in the following
444 order: α, β, n resulting in three models to be estimated in addition to the baseline. We
445 set $dt_{Euler} = 1/32$ year for all model fits.

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

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

461 exploitation for decades it seems fair to assume that current biomass is not many orders of
462 magnitude larger than the observed catches. As prior distribution for $\log(B_y)$ we therefore
463 used $N(\mu_{B_y}, \sigma_{B_y}^2)$, with $\mu_{B_y} = \log(k \max C_t)$ and $\sigma_{B_y} = 3$. Note that a standard deviation
464 of 3 in the log domain is a rather uninformative prior. The specific choice of k has minimal
465 impact on the results as long as μ_{B_y} is in the same order of magnitude as the maximum
466 catch ($1 \leq k \leq 9$) and the prior is uninformative ($\sigma_{B_y} \geq 3$). In the current study we set
467 $k = 3$.

468 **3 Results**

469 **3.1 South Atlantic albacore**

470 Fitting SPiCT1 to the South Atlantic albacore dataset of Polacheck *et al.* (1993) produced
471 results comparable to those obtained using alternative approaches (Fig. 1). The point
472 estimates of SPiCT1 were particularly similar to those obtained using observation error
473 estimators, however the point estimates of alternative approaches were contained in the
474 95% CI of the SPiCT1 point estimates. SPiCT2 produced a wider CI of E_{MSY} while CIs
475 of K and B_{1990} were narrower compared to CIs of SPiCT1. CIs of remaining parameters
476 were largely similar between SPiCT1 and SPiCT2. Overall, among all methods estimates
477 of MSY were nearly identical and associated with the narrowest CIs.

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

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

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

492 Predictions of B_t , B_t/B_{MSY} , F_t , F_t/F_{MSY} and C_t can be obtained by propagating
493 the model beyond the time span of the data (Fig. 2), however uncertainty increases for
494 each predicted year as a result of the lack of data. Furthermore, as the model is a highly
495 simplified version of reality, it should not be relied on for generating long-term predictions
496 (> 2 years) even when residuals pass all tests.

497 [Figure 1 about here.]

498 [Figure 2 about here.]

499 [Table 1 about here.]

500 **3.2 Simulation study 1**

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

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

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

521 See supplement S2 for detailed description of results.

522 **3.3 Simulation study 2**

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

534 **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

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

563 In identifying the process to observation noise relationships, two of four estimates of α
564 were significantly different from the value one (Table 3), which is the commonly assumed
565 value when estimation is not possible (Ono *et al.*, 2012; Thorson *et al.*, 2013). In contrast,
566 only one of four estimates of β were significantly different from the value one. Similarly,
567 the 95% CI of n for Norway pout did not exclude the commonly assumed value of two
568 indicating a lack of evidence to deviate from a symmetric production function (Schaefer,
569 1954).

570 [Table 3 about here.]

571 [Figure 3 about here.]

572 [Figure 4 about here.]

573 4 Discussion

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

748 5 Acknowledgements

749 We wish to thank two anonymous reviewers for their helpful comments. This work is a
750 deliverable of the Nordic Centre for Research on Marine Ecosystems and Resources under
751 Climate Change (NorMER), which is funded by the Norden Top-level Research Initiat-
752 ive sub-programme “Effect Studies and Adaptation to Climate Change”. Furthermore,
753 this work has partly been conducted under the national funds to DTU Aqua from the
754 Danish Ministry of Environment and Food as well as under a contract with AZTI Tech-
755 nalia, Spain, to support development of the fisheries research and fisheries advice in Saudi
756 Arabia as support to the King Fahd University of Petroleum and Minerals. Finally, a
757 minor part of the work has also been conducted under the EU Tender MARE/2014/44
758 Study on the approaches to management for data-poor stocks in mixed fisheries (EU
759 EASME/EMFF/2014/1.3.2.4/SI2.721116).

760 References

- 761 Bolker, B. M., Gardner, B., Maunder, M., Berg, C. W., Brooks, M., Comita, L., Crone,
762 E., Cubaynes, S., Davies, T., Valpine, P. *et al.* (2013). Strategies for fitting nonlin-
763 ear ecological models in R, AD Model Builder, and BUGS. *Methods in Ecology and*
764 *Evolution*, **4**, 501–512.
- 765 Bordet, C. & Rivest, L.-P. (2014). A stochastic Pella Tomlinson model and its maximum
766 sustainable yield. *Journal of theoretical biology*, **360**, 46–53.
- 767 Box, G. E. & Tiao, G. C. (2011). *Bayesian inference in statistical analysis, Volume 40*.
768 John Wiley & Sons.
- 769 Brodziak, J. & Ishimura, G. (2011). Development of Bayesian production models for
770 assessing the North Pacific swordfish population. *Fisheries Science*, **77**, 23–34.
- 771 Carvalho, F., Ahrens, R., Murie, D., Ponciano, J. M., Aires-da Silva, A., Maunder, M. N.
772 & Hazin, F. (2014). Incorporating specific change points in catchability in fisheries stock
773 assessment models: an alternative approach applied to the blue shark (*Prionace glauca*)
774 stock in the south Atlantic Ocean. *Fisheries Research*, **154**, 135–146.
- 775 Chen, Y., Breen, P. & Andrew, N. (2000). Impacts of outliers and mis-specification
776 of priors on Bayesian fisheries-stock assessment. *Canadian Journal of Fisheries and*
777 *Aquatic Sciences*, **57**, 2293–2305.
- 778 Chen, Y., Jackson, D. & Paloheimo, J. (1994). Robust regression approach to analyzing
779 fisheries data. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**, 1420–1429.
- 780 de Valpine, P. & Hilborn, R. (2005). State-space likelihoods for nonlinear fisheries time-
781 series. *Canadian Journal of Fisheries and Aquatic Sciences*, **62**, 1937–1952.

- 782 Fletcher, R. (1978). On the restructuring of the Pella-Tomlinson system. *Fish. Bull.*, **76**,
783 515–521.
- 784 Fox, W. W. (1970). An exponential surplus-yield model for optimizing exploited fish
785 populations. *Transactions of the American Fisheries Society*, **99**, 80–88.
- 786 Gardiner, C. W. (1985). *Handbook of stochastic methods, Volume 4*. Springer Berlin.
- 787 Harvey, A. C. (1990). *Forecasting, structural time series models and the Kalman filter*.
788 Cambridge university press.
- 789 Hilborn, R., Walters, C. J. *et al.* (1992). Quantitative fisheries stock assessment: choice,
790 dynamics and uncertainty. *Reviews in Fish Biology and Fisheries*, **2**, 177–178.
- 791 Iacus, S. M. (2009). *Simulation and inference for stochastic differential equations: with R*
792 *examples, Volume 1*. Springer.
- 793 ICES (2005). Report of the Study Group on Multi Species Assessment in the North Sea.
794 *ICES Document CM 2005/D:06*.
- 795 ICES (2012). DATRAS: survey descriptions. [http://datras.ices.dk/Home/](http://datras.ices.dk/Home/Descriptions.aspx)
796 [Descriptions.aspx](http://datras.ices.dk/Home/Descriptions.aspx).
- 797 Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H. J. & Bell, B. (2015). TMB: Auto-
798 matic Differentiation and Laplace Approximation. ArXiv e-print; in press, *Journal of*
799 *Statistical Software*.
- 800 Ljung, G. M. & Box, G. E. P. (1978). On a measure of lack of fit in time series models.
801 *Biometrika*, **65**, 297–303.
- 802 Magnusson, A. & Hilborn, R. (2007). What makes fisheries data informative? *Fish and*
803 *Fisheries*, **8**, 337–358.

- 804 Maunder, M. N. (2003). Is it time to discard the Schaefer model from the stock assessment
805 scientist's toolbox? *Fisheries Research*, **61**, 145–149.
- 806 Meyer, R. & Millar, R. B. (1999). BUGS in Bayesian stock assessments. *Canadian Journal*
807 *of Fisheries and Aquatic Science*, **1086**, 1078–1086.
- 808 Øksendal, B. (2013). *Stochastic differential equations: an introduction with applications*.
809 Springer Science & Business Media.
- 810 Ono, K., Punt, A. E. & Rivot, E. (2012). Model performance analysis for Bayesian biomass
811 dynamics models using bias, precision and reliability metrics. *Fisheries Research*, **125**,
812 173–183.
- 813 Pawitan, Y. (2013). *In all likelihood: statistical modelling and inference using likelihood*.
814 OUP Oxford.
- 815 Pedersen, M. W., Berg, C. W., Thygesen, U. H., Nielsen, A. & Madsen, H. (2011). Estim-
816 ation methods for nonlinear state-space models in ecology. *Ecological Modelling*, **222**,
817 1394–1400.
- 818 Pella, J. J. & Tomlinson, P. K. (1969). A generalized stock production model. *Bulletin of*
819 *the Inter-American Tropical Tuna Commission*, **13**, 421–458.
- 820 Polacheck, T., Hilborn, R. & Punt, A. E. (1993). Fitting Surplus Production Models:
821 Comparing Methods and Measuring Uncertainty. *Canadian Journal of Fisheries and*
822 *Aquatic Science*.
- 823 Prager, M. H. (1994). A suite of extensions to a nonequilibrium surplus-production model.
824 *Fishery Bulletin*, **389**, 374–389.
- 825 Prager, M. H. (2002). Comparison of logistic and generalized surplus-production models

826 applied to swordfish, *Xiphias gladius*, in the north Atlantic Ocean. *Fisheries Research*,
827 **58**, 41–57.

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

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

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

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

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

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

841 Thorson, J. T., Cope, J. M., Branch, T. A., Jensen, O. P. & Walters, C. J. (2012).
842 Spawning biomass reference points for exploited marine fishes, incorporating taxonomic
843 and body size information. *Canadian Journal of Fisheries and Aquatic Sciences*, **69**,
844 1556–1568.

845 Thorson, J. T., Minto, C., Minto-Vera, C. V., Kleisner, K. M., Longo, C. & Jacobson,
846 L. (2013). A new role for effort dynamics in the theory of harvested populations and
847 data-poor stock assessment. *Canadian Journal of Fisheries and Aquatic Sciences*, **70**,
848 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.

851 List of Figures

- 852 1 Comparison of estimated quantities obtained from fits to the South Atlantic
853 albacore dataset relative to estimates obtained using an observation error
854 model (Polacheck *et al.*, 1993). Error bars are 95% CIs. Note that ASPIC
855 does not report uncertainty on all quantities. Overall, estimates were similar
856 with MSY and the relative estimate B_{1990}/K having lowest uncertainty.
857 Note that estimating α , β , and n (SPiCT2) lead to decreased uncertainty
858 of K and B_{1990} , while the uncertainty of E_{MSY} increased substantially.
- 859 2 Fits to the South Atlantic albacore dataset of Polacheck *et al.* (1993) with
860 α , β , and n fixed (SPiCT1) and α , β , and n estimated (SPiCT2). Solid blue
861 lines are estimated values, vertical grey lines indicate the time of the last
862 observation beyond which dotted lines indicate forecasts, dashed lines are
863 95% CI bounds for absolute estimated values, shaded blue regions are 95%
864 CIs for relative estimates (B_t/B_{MSY} or F_t/F_{MSY}), grey regions are 95%
865 CIs for estimated absolute reference points (horizontal lines), solid circles
866 are observations with index plotted as I_t/q . Notably, SPiCT1 estimated
867 narrower CIs of reference points and relative levels while SPiCT2 estimated
868 narrower CIs of absolute levels of B_t , F_t and catch.
- 869 3 Estimated biomass and annually averaged fishing mortality (solid blue lines)
870 of North Sea stocks relative to estimated reference points with 95% CI (blue
871 shaded region). Fishing mortality including within-year variation (shaded
872 blue lines in middle column) show the estimated seasonal pattern, which is
873 particularly marked for herring. Uncertainty of estimated reference points
874 is represented by 95% confidence regions on the relative scale (grey shaded
875 region in right column).
- 876 4 Significant lags are detected in the autocorrelation function (ACF) of North
877 Sea haddock catch residuals using a spline-based seasonal model (a). This
878 violation is remedied when using the coupled SDE seasonal model (b). The
879 detected autocorrelation is caused by the constant amplitude of the spline-
880 based model, which is insufficient compared to the adaptive representation
881 of the coupled SDE seasonal model (c).

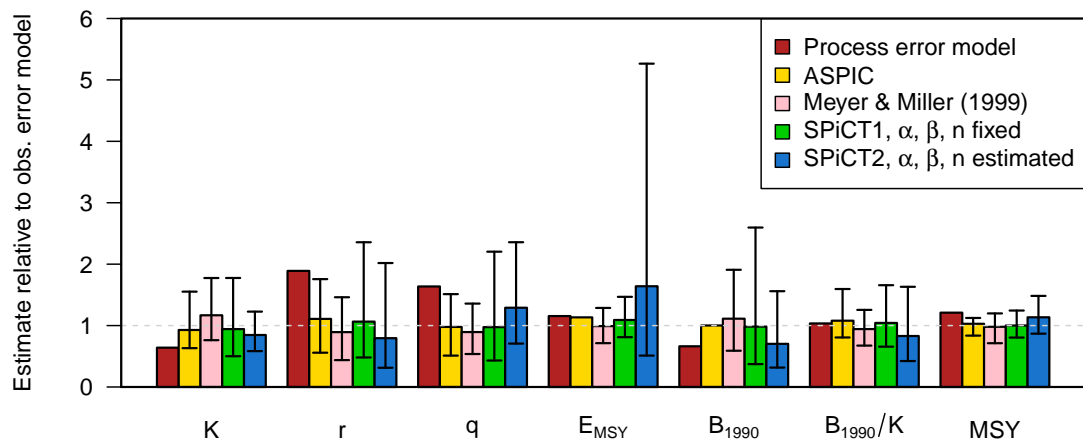


Figure 1

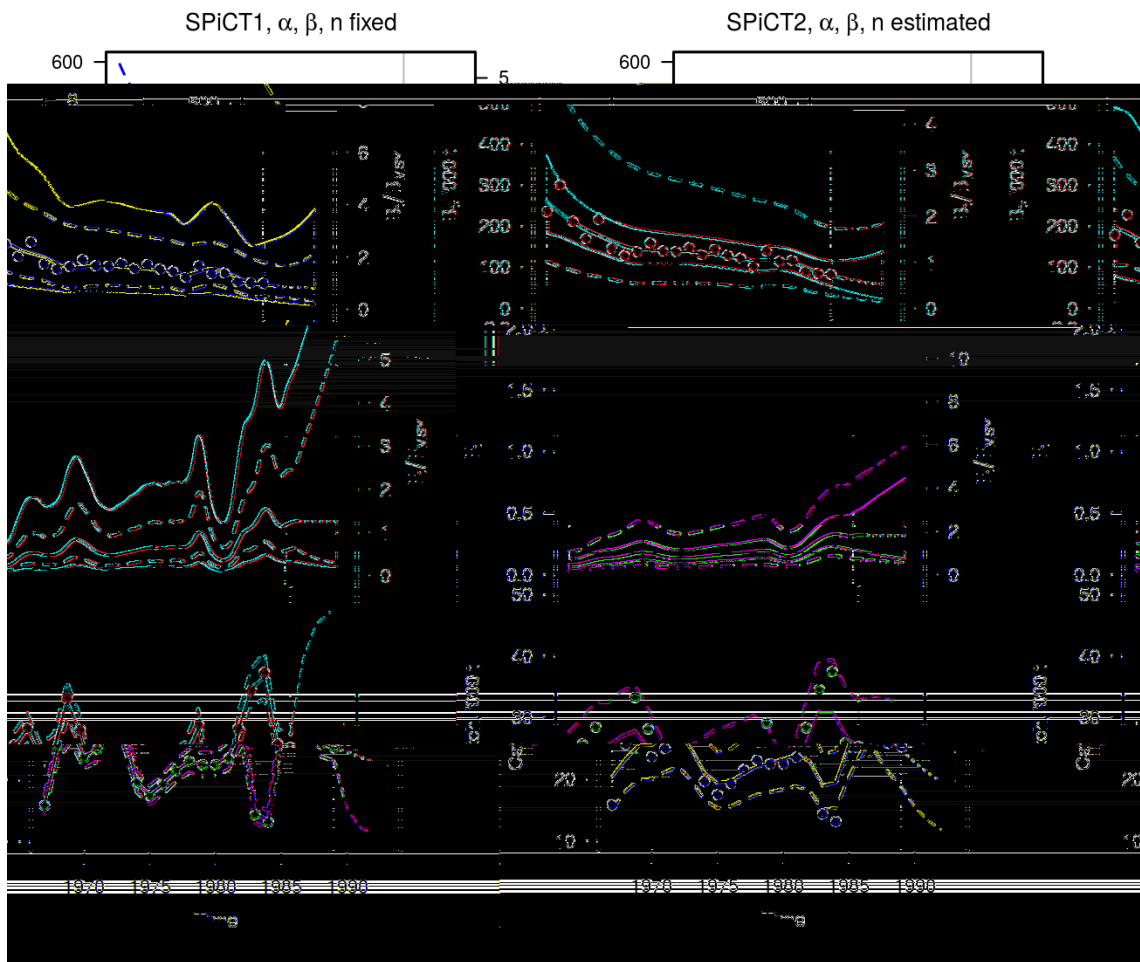


Figure 2

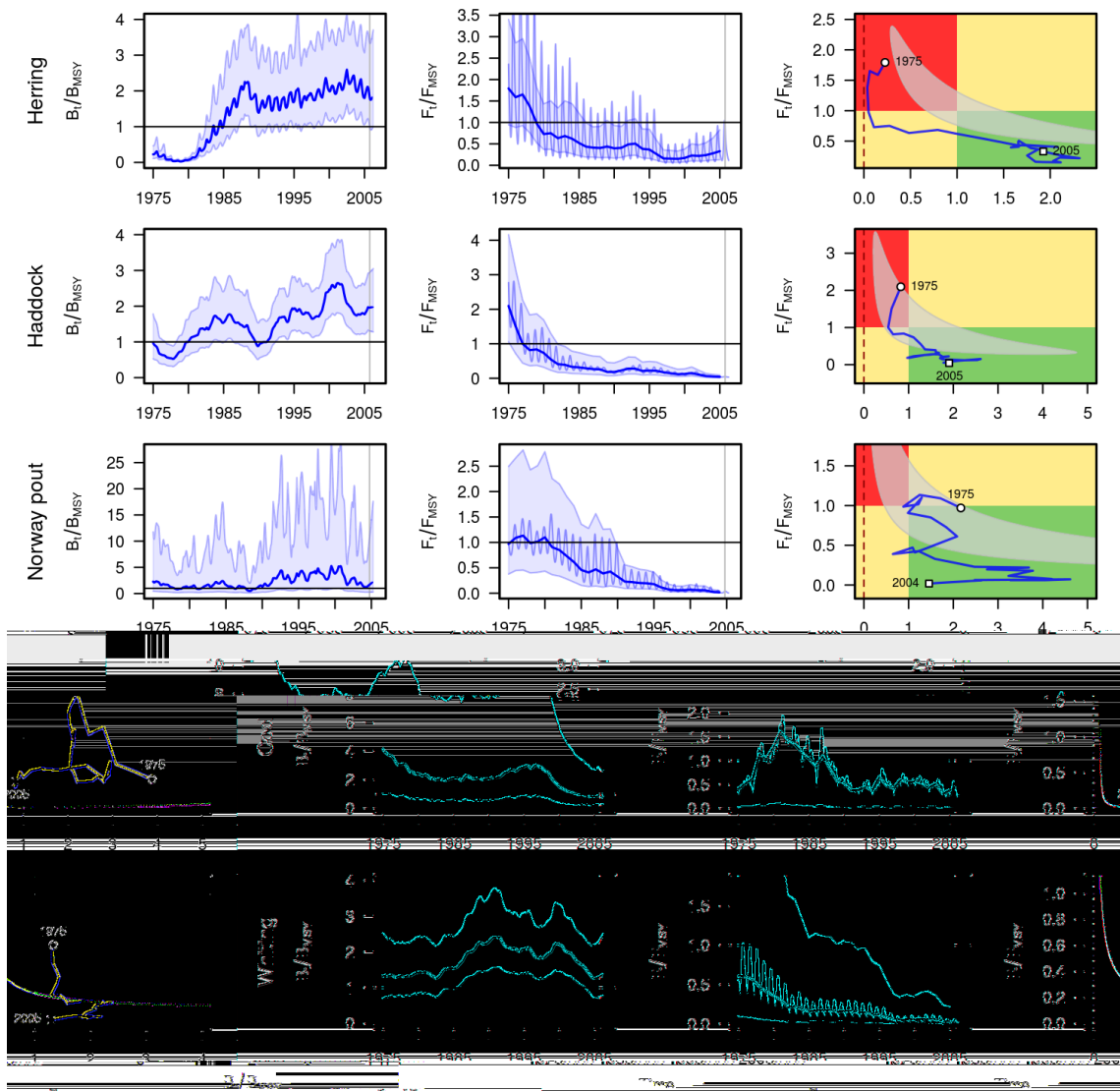


Figure 3

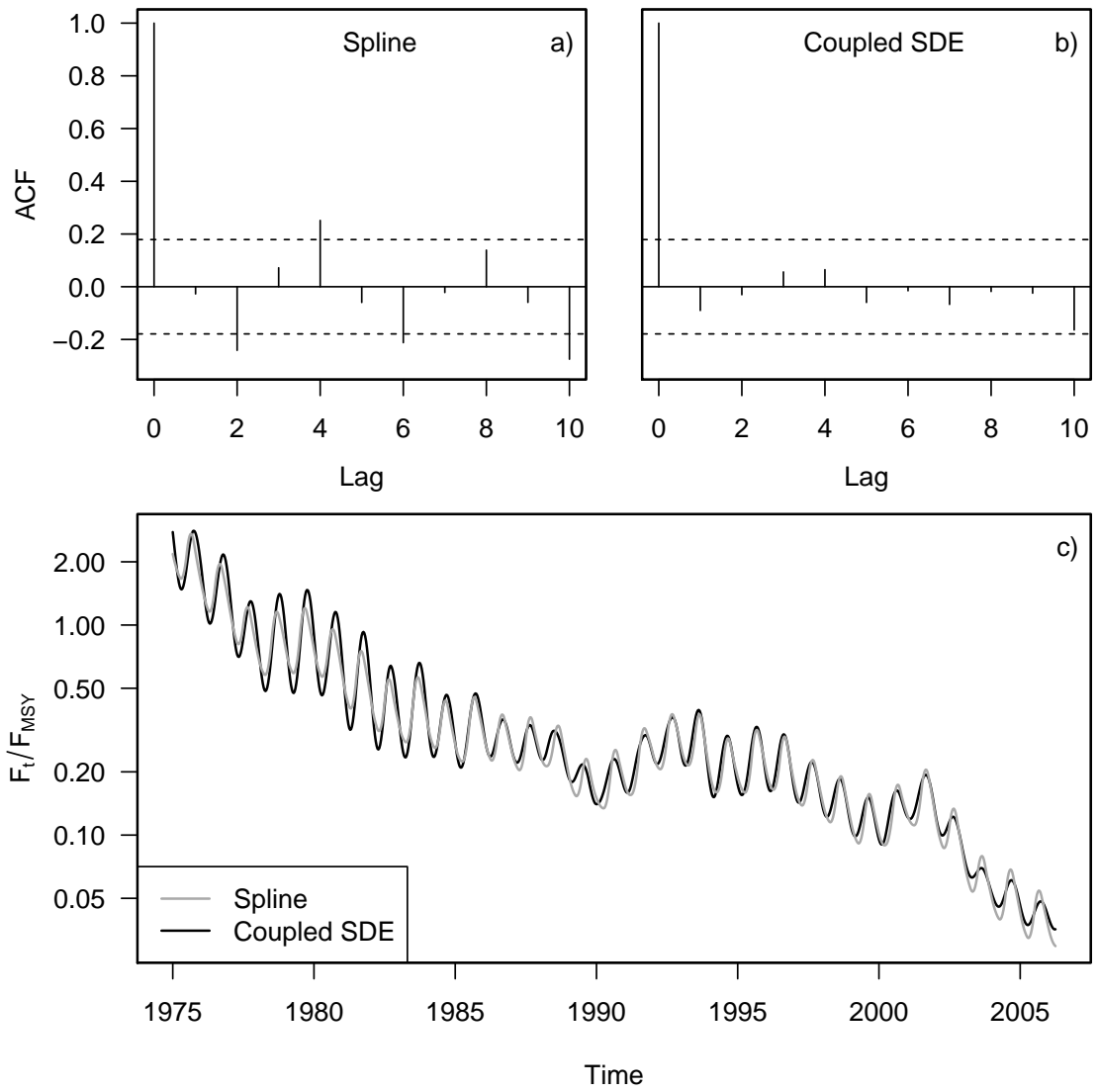


Figure 4