Supplement

A stochastic surplus production model in continuous-time

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

[1] National Institute of Aquatic Resources, Technical University of Denmark, Charlottenlund Slot, Jægersborg Allé 1, 2920 Copenhagen, Denmark

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

	ASPIC			SPiCT1			SPiCT2		
	estimate	lower	upper	estimate	lower	upper	estimate	lower	upper
K	222.40	152.14	384.00	225.93	120.02	425.27	202.86	139.94	294.06
r	0.36	0.17	0.57	0.35	0.16	0.77	1.05	0.00	2725.41
$q \cdot 10^4$	26.12	13.19	40.28	26.03	11.51	58.84	34.47	18.87	62.96
B_{1990}	75.63			74.25	28.12	196.05	53.05	23.89	117.77
B_{1990}/K	0.34	0.25	0.50	0.33	0.21	0.52	0.26	0.13	0.51
MSY	20.22	15.82	22.14	19.67	15.81	24.46	22.30	17.06	29.14
E_{MSY}	69.60			67.00	49.80	90.13	100.73	31.39	323.23

Table S1: Comparison of estimated quantities using ASPIC version 7.02 and SPiCT including 95% CIs indicated by lower and upper bounds. Note that CIs are not provided by ASPIC for all quantities.

S1 South Atlantic albacore

Table S1 contains detailed results of fitting ASPIC version 7.02 and SPiCT to the South Atlantic albacore data set of Polacheck *et al.* (1993).

S2 Simulation study 1

S2.1 Methods

S2.1.1 Models with limited or no prior information

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 of parameter estimates), the coverage of 95% CIs (proportion containing the true parameter), and the median bias of estimates. These quantities were evaluated for eight variants of SPiCT (model A to H, Table S2), with particular focus on the influence of the parameters n, α, and β, which can be difficult to estimate.

The flexibility of a surplus production model including the exponent n, while biologically appealing, is known to cause estimation problems (Prager, 2002). The influence of n on estimation performance was therefore assessed for cases where n was estimated, where n was fixed correctly to the true value ($n_{true} = 2$), and where n was fixed wrongly ($n = 3$, misspecification error of 50%) (Table S2).

It is well documented that separating process and observation noise is difficult when fitting surplus production models (Polacheck *et al.*, 1993). The influence of α and β (i.e. the ratios between the noise of B_t and I_t, and F_t and C_t respectively) on performance was therefore assessed when the parameters were estimated, fixed correctly to the true values

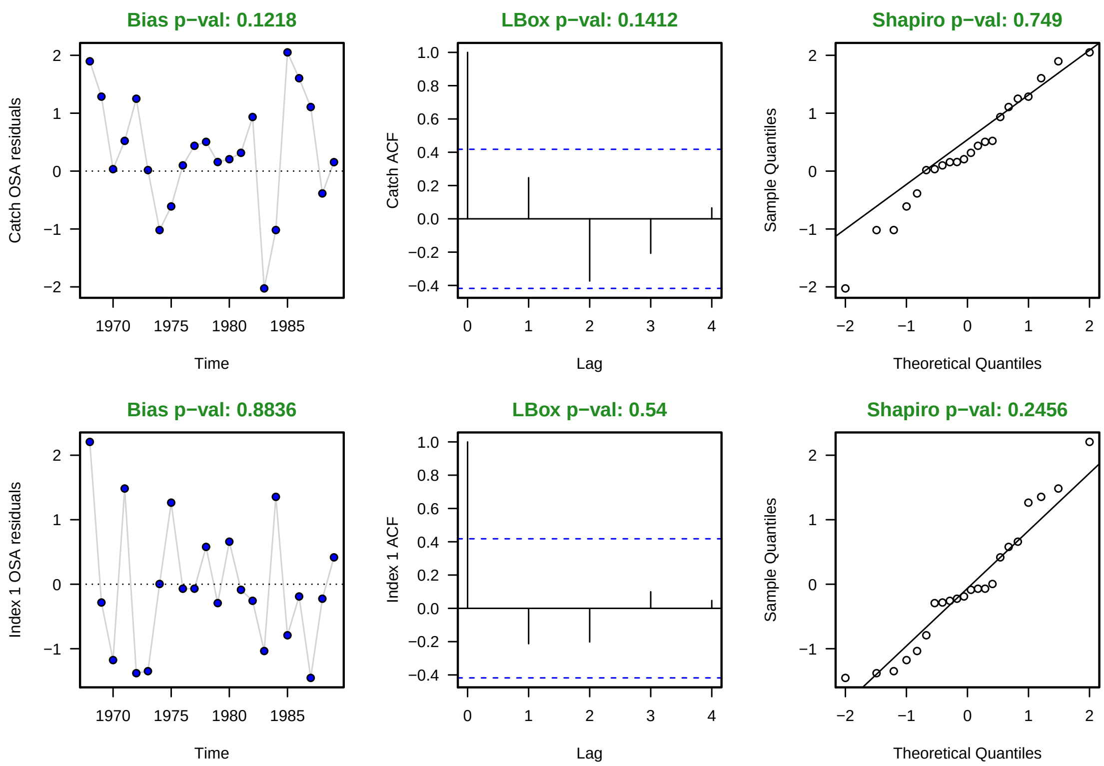

Figure S1: SPiCT1: Diagnostics represented by residual scatter plots (left column), empirical autocorrelation functions (ACFs, middle column), and QQ-plots (right column) of catch and index residuals obtained by fitting SPiCT to the South Atlantic albacore dataset of Polacheck *et al.* (1993). No significant violations of independence, bias and normality assumptions were found.

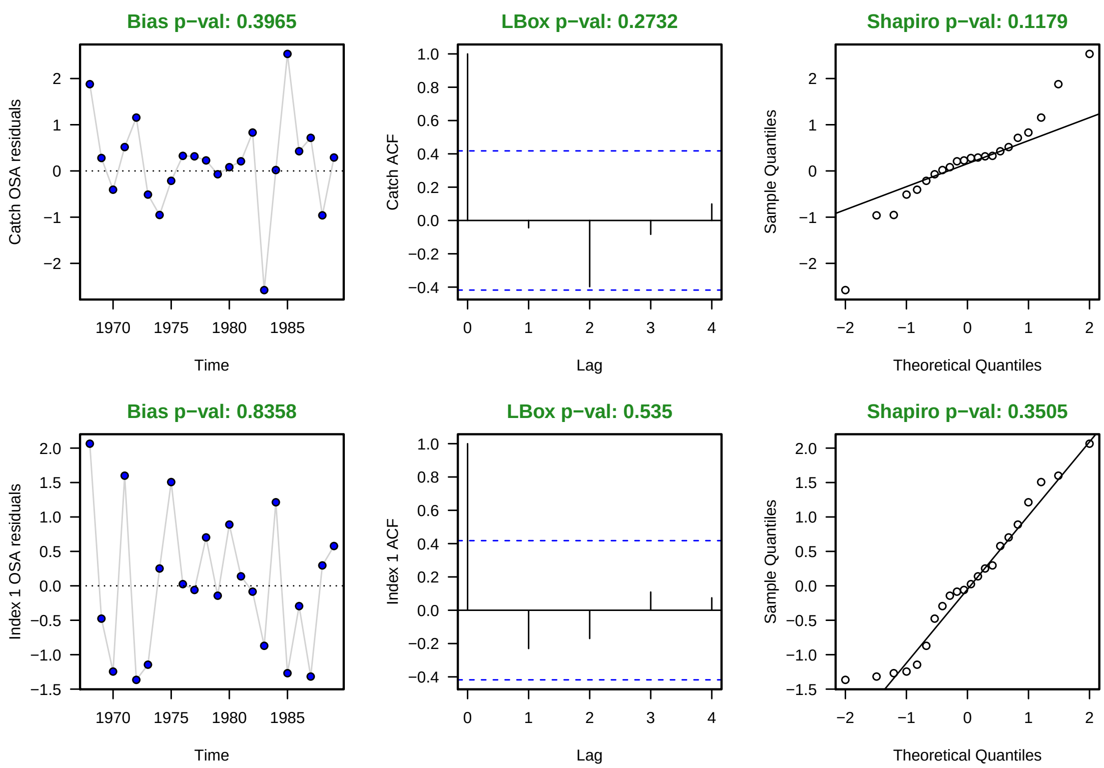

Figure S2: SPiCT2: Diagnostics represented by residual scatter plots (left column), empirical autocorrelation functions (ACFs, middle column), and QQ-plots (right column) of catch and index residuals obtained by fitting SPiCT to the South Atlantic albacore dataset of Polacheck *et al.* (1993). No significant violations of independence, bias and normality assumptions were found.

Model	Estimated	Fixed to true	Fixed to wrong	Priors
A	α, β, n	–	–	–
B	α, β	n	–	–
C	α	n, β	–	–
D	–	n, β, α	–	–
E	–	β, α	n	–
F	–	n, α	β	–
G	–	n, β	α	–
H	α, β, n	–	–	Vague on α, β, n

Table S2: Variants of SPiCT used in the simulation study. For all models the parameters m, K, q, σ_B, and σ_F were estimated. Depending on the model the remaining parameters (α, β, n) were either estimated, fixed to the true value, or fixed to an incorrect value. Models A-D had decreasing complexity (fewer parameters to be estimated). Models E, F, and G had misspecified parameters (n, β, and α respectively).

($\alpha_{true} = 1$, $\beta_{true} = 1$), and fixed wrongly ($\alpha = 2$, $\beta = 2$, misspecification error of 100%) (Table S2).

As a trade-off between unconstrained estimation of a parameter and fixing a parameter we fitted a model with vague priors imposed on α, β, and n (Table S2). Specifically we imposed:

- $\log(\alpha) \sim N(\log[1], 2)$.
- $\log(\beta) \sim N(\log[1], 2)$.
- $\log(n) \sim N(\log[2], 2)$.

If information about α, β, and n is present in data this model should provide similar estimates to the same model without priors, while if data contain little information about α, β, and n the model will default to the prior estimates instead of failing estimation.

The simulation protocol was to first fit SPiCT to the dataset of Polacheck *et al.* (1993) while fixing $n = 2$, $\alpha = 1$, $\beta = 1$. Using the resulting parameters estimates (Table S1), batches of 1000 datasets of catch and biomass index pairs were then simulated using Eqns. (5, 11, 16, 17) with the number of observations per dataset in a batch given by $N_{obsC} = N_{obsI} = \{15, 30, 60, 120, 240\}$ resulting in a total of 5000 simulated datasets. The eight model variants (Table S2) were then each fitted to all datasets resulting in a total of 40000 SPiCT fits. For all simulations and model fits we set $dt_{Euler} = 1/8$ year. This choice is a trade-off between obtaining a fine temporal resolution while limiting computational requirements to maintain tractability of the simulation. For this study the choice of dt_{Euler}

is arbitrary as the same value is used for both simulation and estimation and therefore should not have significant impact on results.

For comparison with a standard observation error method without priors we fitted also ASPIC 7.02 (Prager, 1994) following the above stated protocol. However, as ASPIC is limited to a maximum of 150 observations the trial containing 240 observation was omitted. Furthermore, as ASPIC does not calculate uncertainty of B_{last} this quantity was omitted from ASPIC results. The bootstrap module of ASPIC was used with 1000 samples to obtain 95% CIs.

S2.1.2 Models with informative prior information

We also fitted the model of Meyer & Millar (1999) following the above stated protocol with data simulated in continuous-time (CT) and compared with SPiCT using identical priors for the two models. In addition, to obtain a more fair comparison of the two models as the Meyer & Millar (1999) model is formulated in discrete-time (DT) and does not incorporate observation error on catches, we also simulated data with $dt_{Euler} = 1$ and $\sigma_C = 0.001$.

Each model was fitted to 200 datasets. Priors were imposed on the same parameters as in Meyer & Millar (1999):

- $K \sim$ log-normal$(5.16, 0.516)$.
- $r \sim$ log-normal$(-0.68, 0.51)$.
- $q \sim$ inverse-gamma$(0.000794, 0.000282)$.
- $\tau^2 \sim$ inverse-gamma$(0.510, 0.00476)$.
- $\sigma^2 \sim$ inverse-gamma$(0.510, 0.00476)$.
- $B_{t_0}/K \sim$ log-normal$(-0.223, \sigma)$, where t_0 is the initial time.

The BUGS (Lunn *et al.*, 2000) code used is identical the code used in Meyer & Millar (1999) with the addition that negative values of biomass are not allowed. We also used identical BUGS settings: burn-in period of 25000 iterations, total iterations were 225000, with thinning of 25. As the model of Meyer & Millar (1999) does not report F_{last}/F_{MSY} it was not possible to include this quantity in the results.

S2.1.3 Performance statistics

For each combination of model and number of observations several performance statistics were collected: 1) the proportion of converged model fits. For SPiCT a fit was defined as converged if the optimiser nlminb (R Core Team, 2015) reports a successful completion and if CIs were successfully calculated and finite. For ASPIC a fit was defined as converged if the optimiser reported the error code "0" indicating normal convergence. For the model of Meyer & Millar (1999) a fit was defined as converged if it successfully passed the Heidelberger and Welch stationarity and halfwidth tests and Geweke's Z-scores test as in Meyer & Millar (1999).

For converged fits we then calculated 2) the median coefficient of variation (CV) as an indicator of estimation precision, 3) the proportion of the calculated 95% CIs that contained the true value as an indicator of CI coverage, and 4) the median bias of estimates, i.e.

$$\text{median bias} = \text{median}\left(\frac{\widehat{\theta} - \theta}{\theta}\right). \tag{1}$$

as in Magnusson & Hilborn (2007). The performance statistics were calculated for the following parameters: F_{MSY}, MSY, B_{MSY}, F_{last}/F_{MSY}, B_{last}, B_{last}/B_{MSY}, where B_{last} and F_{last} refer to the estimated biomass and fishing mortality at the time of the last observation.

S2.2 Results

S2.2.1 Models with limited or no prior information

The convergence properties of the eight model variants (Table S2) improved for increasing number of observations (Fig. S3). Estimating all parameters (model A) resulted in reasonable performance (convergence proportion of 0.7-0.8) for short data sets ($N_{obsC} = N_{obsI} < 60$), and improved performance (convergence proportion 0.9-0.95) for larger datasets ($N_{obsC} = N_{obsI} \geq$ 60-240) albeit never reaching 100% convergence. All remaining SPiCT models without priors performed approximately identically with 0.95 convergence proportion for short data sets and 100% convergence for larger datasets. Estimating all parameters with vague priors imposed on difficult parameters (model H) resulted in improved convergence performance relative to the same model without priors (model A). The convergence performance of ASPIC were....

For all combinations of models and parameters CVs decreased with increasing number of observations (Fig. S4). Model A was less precise (higher median CV) in estimating reference points (F_{MSY} and B_{MSY}) and stock status (F_{last}/F_{MSY} and B_{last}/B_{MSY}), however showed similar performance for MSY and B_{last} relative to simpler models. For models B-G, MSY and B_{last}/B_{MSY} were estimated with highest precision (lowest CV) relative to other parameters even for short datasets. For F_{MSY} and F_{last}/F_{MSY} intermediate precision was obtained while estimation precision of B_{MSY} and absolute biomass, B_{last}, were slightly worse for short data sets. For large datasets (60 - 240 observations) model A and model H performed similarly, while for smaller datasets (15-30 observations) model H

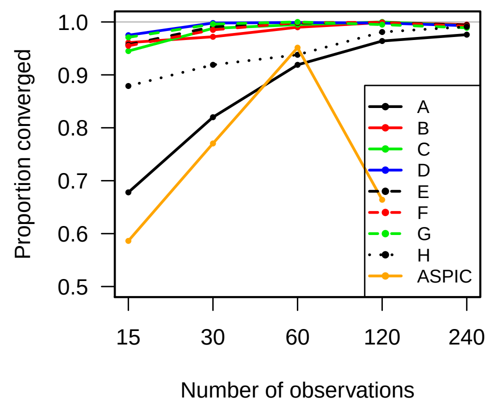

Figure S3: Estimation convergence of simulated datasets as a function of the model and the number of observations with solid lines representing correctly specified models and dashed lines representing misspecified models. Model A showed markedly poorer convergence properties relative to models B-G that overall performed identically and converged toward 100% for increasing number of observations.

obtained lower median CVs compared to model A as a result of the stabilising information from the vague priors on difficult parameters. Overall, median CVs produced by ASPIC performed similarly to CVs produced by SPiCT models B-G for all quantities except MSY and B_{MSY} for which median CVs were inflated (Fig. S4).

The proportion of 95% CIs containing the true value converged for correctly specified models (A-D) to the expected 0.95 for all estimated quantities for increasing number of observations (Fig. S5). Model A and H provided CIs with a consistent coverage close to 0.95 for data sets of any size for F_{MSY}, B_{MSY}, F_{last}/F_{MSY}, B_{last}/B_{MSY}. Furthermore, correctly specified models had CI coverage of estimated reference points (F_{MSY}, B_{MSY}, MSY) and of relative fishing pressure (F_{last}/F_{MSY}) in the range 0.85-0.95, while CIs of absolute and relative biomass had somewhat poorer coverage with a range of 0.8-0.95. Misspecified models (E-G) generally produced CIs with poorer coverage than correctly specified models for increasing number of observations. Misspecification of n (model E) resulted in diverging CIs of F_{MSY}, MSY and B_{last}/B_{MSY}, somewhat unreliable CIs for F_{last}/F_{MSY} and B_{MSY} with a constant proportion of 0.85 containing the true values, while CIs of B_{last} converged toward a coverage of 0.95. Misspecification of β and α (models F and G) mainly affected CIs of F_{MSY} and B_{MSY} while CIs of remaining quantities generally followed the trend of correctly specified models. Coverage of ASPIC CIs was in the best

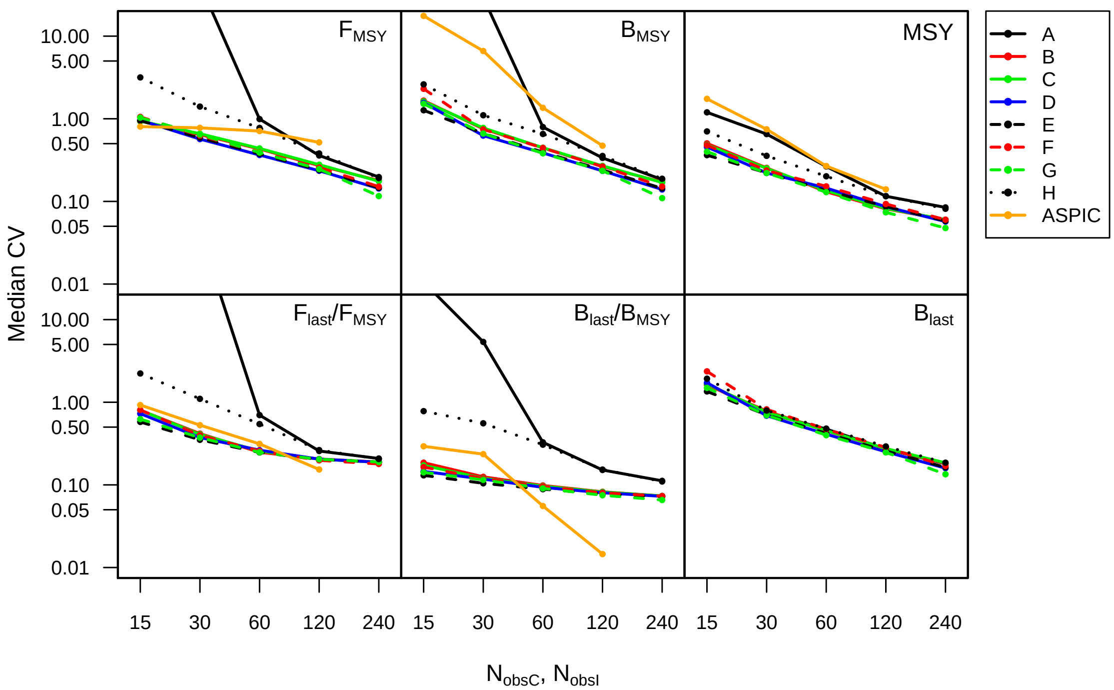

Figure S4: Median CV of estimated quantities for simulated datasets as a function of the model and the number of observations with solid lines representing correctly specified models, dashed lines representing misspecified models, and dotted lines representing models with vague priors. Convergence toward an asymptote was seen for all combinations of models and parameters for increasing number of observations. Model A generally produced larger median CVs than models B-G. For models B-G median CVs were generally similar. Model H produced lower median CVs than model A for small datasets (15 - 30 observations), while performance of the two models was identical for larger datasets (60 - 240 observations). Median CVs of ASPIC were similar to median CVs of SPiCT models B-G for all quantities except for MSY and B_{MSY}.

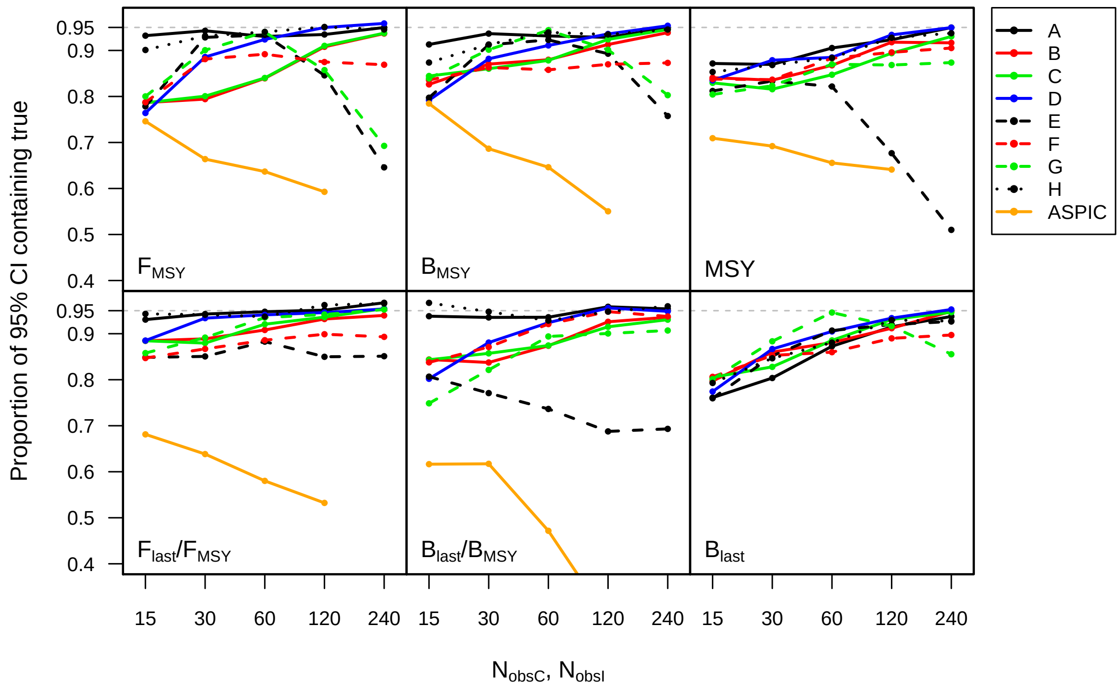

Figure S5: Proportion of 95% CIs that contained the true parameter value as a function of the model and the number of observations with solid lines representing correctly specified models, dashed lines representing misspecified models, and dotted lines representing models with vague priors. Correctly specified models converged to the expected 0.95 for all estimated quantities. CIs of misspecified models generally performed poorer than correctly specified models, particularly for misspecified n (model E) CIs only converged toward the expected 0.95 for B_{last}.

case (B_{MSY}) at 0.8 for datasets of 15 observations, but all ASPIC CIs diverged in coverage for increasing number of observations.

Median biases of correctly specified models (A-D, H) all converged to zero, indicating that the estimation method underlying SPiCT is asymptotically unbiased (Fig. S6). All models had the lowest bias when estimating MSY, F_{last}/F_{MSY} and B_{last}/B_{MSY}, and even misspecified models (E-G) converged to zero suggesting that inference regarding these estimated quantities is robust. In contrast median biases of model E-G remained biased even for large datasets for the quantities F_{MSY}, B_{MSY}, and B_{last}. Notably for these three quantities, ASPIC showed comparable performance to SPiCT despite its simpler model structure.

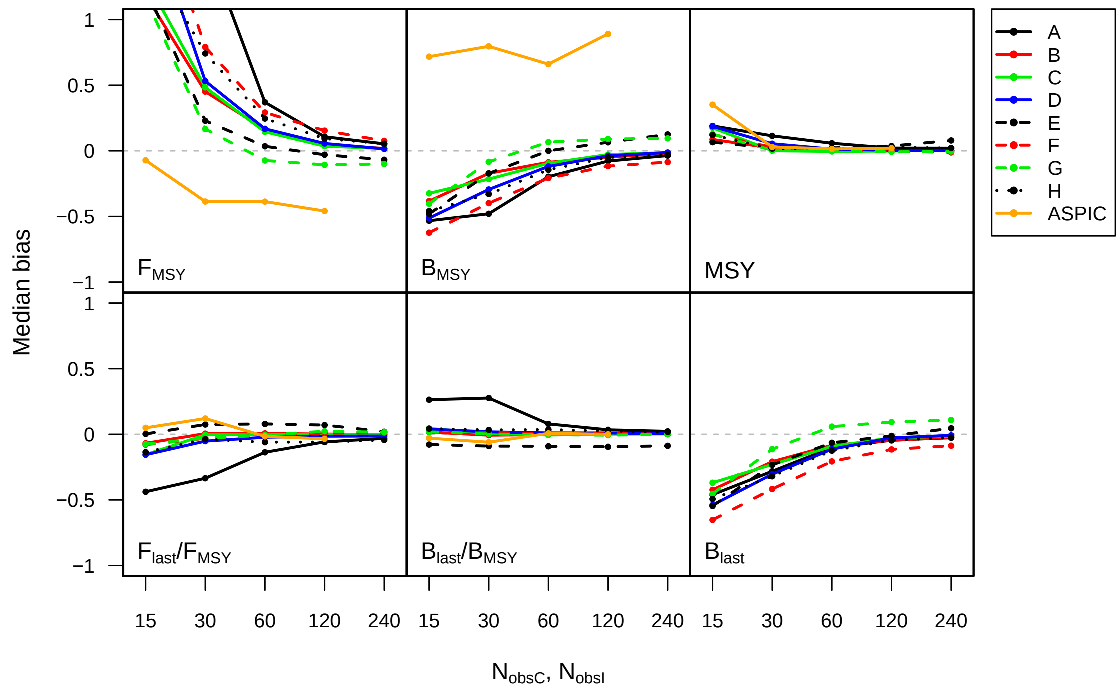

Figure S6: Median bias of estimated quantities as a function of the model and the number of observations with solid lines representing correctly specified models, dashed lines representing misspecified models, and dotted lines representing models with vague priors.

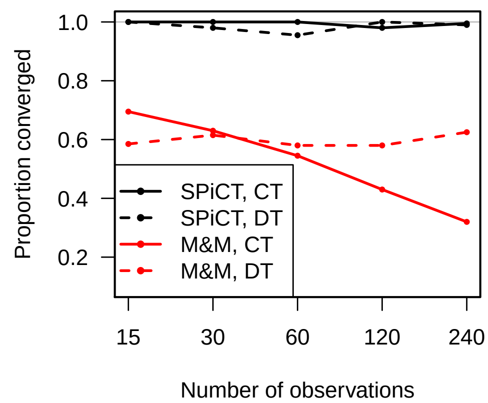

Figure S7: Estimation convergence of simulated datasets as a function of the model and the number of observations with "M&M" referring to the method of Meyer & Millar (1999).

S2.2.2 Models with informative prior information

Using informative priors SPiCT had a convergence rate close to one (Fig. S7), which is an improvement relative to estimation without informative priors (Fig. S3). The convergence rate of the Meyer & Millar (1999) model applied to data simulated in continuous-time decreased as a function of number of observations, indicating that the model setup (number of samples, burn-in, and thinning) may require manual adaptation depending on the dataset, however for data simulated in discrete time the convergence proportion was constant around 0.6 (Fig. S7).

Median CVs of both models decreased as the number of observations increased with SPiCT generally producing lower median CVs (Fig. S8) except for B_{last}/B_{MSY} for which median CVs were indistinguishable. Estimates of models fitted to DT data generally had lower median CVs than models fitted to CT data. This was likely because data simulated in DT had negligible observation error on catches.

The coverage of SPiCT CIs exceeded 0.95 for all estimated quantities and low number of observations, but converged to the expected 0.95 for MSY, F_{last}/F_{MSY}, B_{last}/B_{MSY}, and B_{last}, while convergence was less clear for the remaining quantities. The Meyer & Millar (1999) model CIs diverged for all parameters for increasing number of observations. This result was expected in the CT case as the estimation model differs from the data generating model. It was expected that the model would show improved performance for DT data, however that was not the case for CI coverage. A possible explanation is that the model of Meyer & Millar (1999) does not use the reparameterisation of Fletcher (1978),

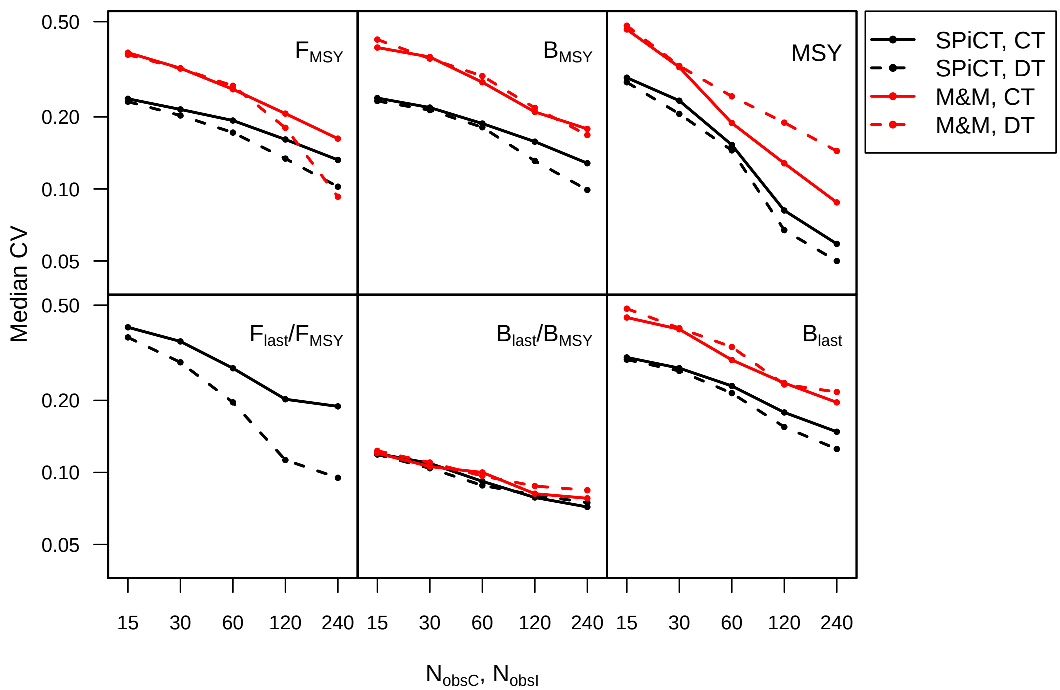

Figure S8: Median CV of estimated quantities for simulated datasets as a function of the model and the number of observations with "M&M" referring to the method of Meyer & Millar (1999). All median CVs decreased as the number of observations increased with SPiCT generally producing lower median CVs. Estimates of models fitted to discrete-time (DT) data generally had lower median CVs than models fitted to continuous-time (CT) data. This was likely because data simulated in DT had negligible observation error on catches.

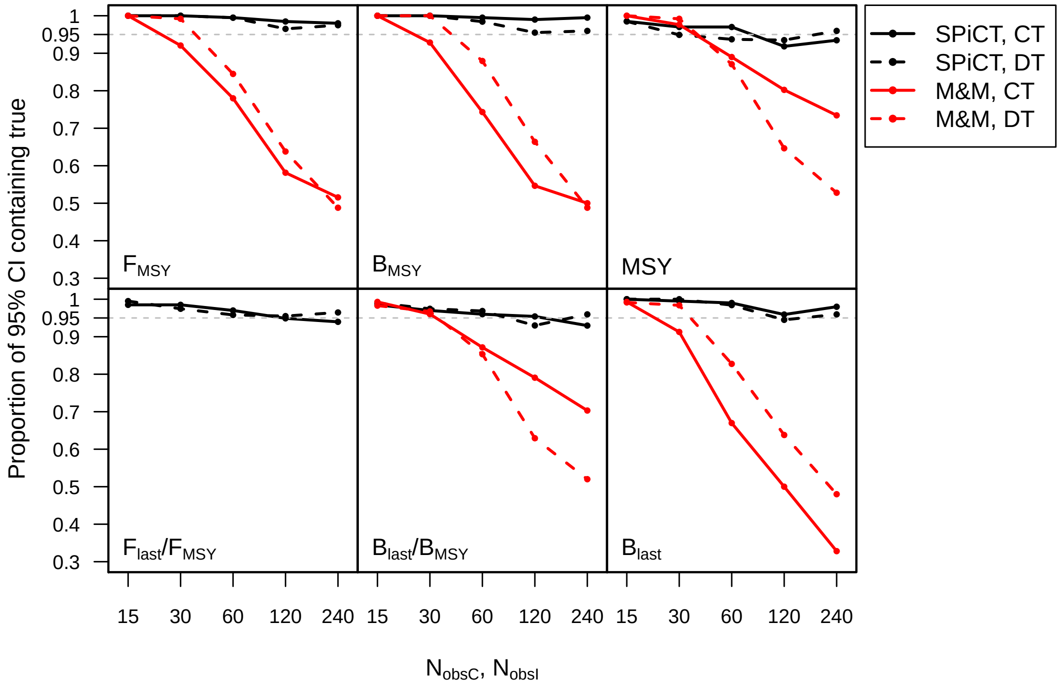

Figure S9: Proportion of 95% CIs that contained the true parameter value as a function of the model and the number of observations with "M&M" referring to the method of Meyer & Millar (1999). The coverage of SPiCT CIs exceeded 0.95 for low number of observations, but converged to the expected 0.95 for MSY, F_{last}/F_{MSY}, B_{last}/B_{MSY} and B_{last} while convergence was less clear for the remaining quantities. The Meyer & Millar (1999) model CIs diverged for all parameters as a result of the estimation model differing from the model generating data.

which potentially could improve the convergence and mixing properties of the Markov chain Monte Carlo estimator.

SPiCT produced unbiased estimates of MSY, F_{last}/F_{MSY}, and B_{last}/B_{MSY}, while estimates of the remaining quantities had biases of about 0.05. The Meyer & Millar (1999) model was unbiased in estimating B_{last}/B_{MSY}, while the biases of the remaining quantities had less clear properties.

S3 North Sea stocks

S3.1 Summary of results and diagnostics

All model fits are summarised in Table S3. Input columns:

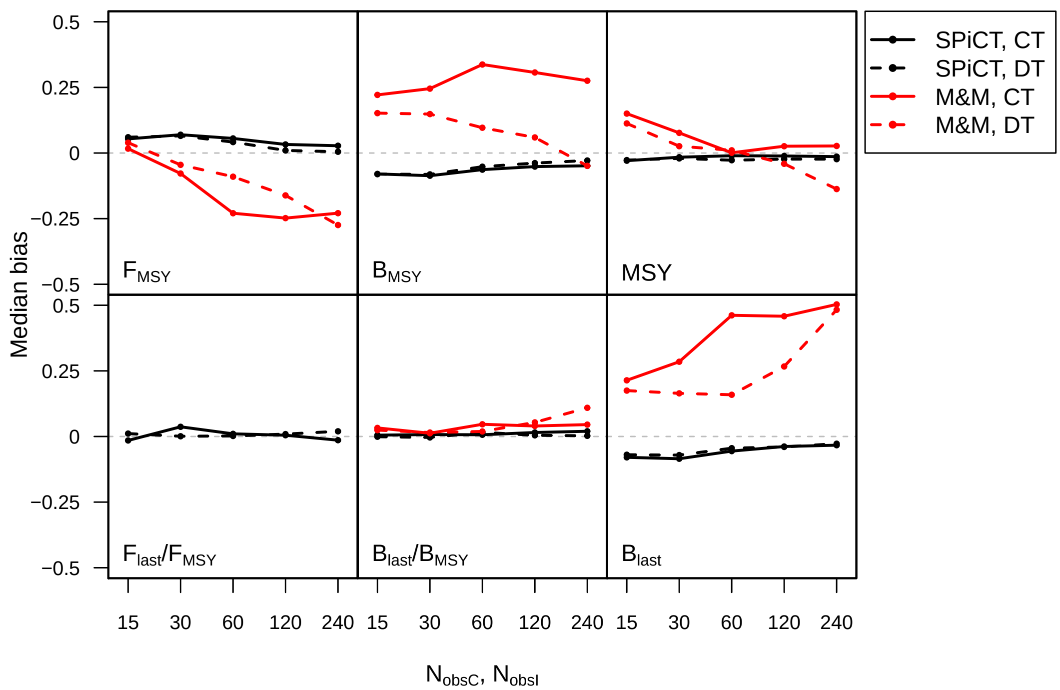

Figure S10: Median bias of estimated quantities as a function of the model and the number of observations with "M&M" referring to the method of Meyer & Millar (1999). SPiCT produced unbiased estimates of MSY, F_{last}/F_{MSY}, and B_{last}/B_{MSY}, while estimates of the remaining quantities had biases of about 0.05. The Meyer & Millar (1999) model was unbiased in estimating B_{last}/B_{MSY}, while the biases of the remaining quantities had less clear properties.

- nm: Species name.
- sd: Model for seasonal dynamics. Either spline or ssde1 (coupled SDE).
- robc: Use robust estimation for catches?
- robi: Use robust estimation for indices?
- prior: Use prior to stabilise estimation? (see main text for details).

Output columns containing results of residual analysis for the four models:

- small (fixed α, β, and n).
- rel.alpha (α estimated).
- rel.beta (α and β estimated).
- rel.n (α, β and n estimated).

Results are given as a character string of length six with each character representing statistical significance of a residual test (1: significant violation, 0: no violation) in the following order: catch autocorrelation function (ACF), catch normality, catch bias, index ACF, index normality, index bias.

S3.2 North Sea: residual diagnostic plots of selected model fits

Detailed diagnostic plots of the selected model fit for each of the five species are presented in Figs. S11-S15.

References

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

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

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

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

	nm	sd	robc	robi	prior	small	rel.alpha	rel.beta	rel.n
1	herring	spline	FALSE	FALSE	FALSE	101101	<NA>	<NA>	<NA>
2	herring	spline	FALSE	TRUE	FALSE	101001	<NA>	<NA>	<NA>
3	herring	spline	TRUE	FALSE	FALSE	<NA>	<NA>	<NA>	<NA>
4	herring	spline	TRUE	TRUE	FALSE	<NA>	<NA>	<NA>	<NA>
5	herring	ssde1	FALSE	FALSE	FALSE	101101	<NA>	<NA>	<NA>
6	herring	ssde1	FALSE	TRUE	FALSE	<NA>	<NA>	<NA>	<NA>
7	herring	ssde1	TRUE	FALSE	FALSE	101101	<NA>	<NA>	<NA>
8	herring	ssde1	TRUE	TRUE	FALSE	101001	<NA>	<NA>	<NA>
9	npout	spline	FALSE	FALSE	FALSE	001000	001000	001000	001000
10	npout	spline	FALSE	TRUE	FALSE	<NA>	001000	<NA>	001000
11	npout	spline	TRUE	FALSE	FALSE	<NA>	<NA>	<NA>	<NA>
12	npout	spline	TRUE	TRUE	FALSE	<NA>	<NA>	<NA>	<NA>
13	npout	ssde1	FALSE	FALSE	FALSE	100000	000000	000000	000000
14	npout	ssde1	FALSE	TRUE	FALSE	100000	000000	000000	000000
15	npout	ssde1	TRUE	FALSE	FALSE	<NA>	000000	000000	000000
16	npout	ssde1	TRUE	TRUE	FALSE	000000	000000	000000	000000
17	haddock	spline	FALSE	FALSE	FALSE	011100	011100	001100	001100
18	haddock	spline	FALSE	TRUE	FALSE	011100	011100	001100	001100
19	haddock	spline	TRUE	FALSE	FALSE	011100	<NA>	<NA>	<NA>
20	haddock	spline	TRUE	TRUE	FALSE	<NA>	<NA>	<NA>	<NA>
21	haddock	ssde1	FALSE	FALSE	FALSE	010100	010100	000100	010100
22	haddock	ssde1	FALSE	TRUE	FALSE	010100	010100	000100	010100
23	haddock	ssde1	TRUE	FALSE	FALSE	010100	010100	000100	010100
24	haddock	ssde1	TRUE	TRUE	FALSE	000100	010100	000100	<NA>
25	cod	spline	FALSE	FALSE	TRUE	100000	101000	100000	<NA>
26	cod	spline	FALSE	TRUE	TRUE	100000	101000	100000	<NA>
27	cod	spline	TRUE	FALSE	TRUE	001000	101000	101000	<NA>
28	cod	spline	TRUE	TRUE	TRUE	001000	001000	101000	<NA>
29	cod	ssde1	FALSE	FALSE	TRUE	<NA>	<NA>	<NA>	<NA>
30	cod	ssde1	FALSE	TRUE	TRUE	<NA>	<NA>	<NA>	<NA>
31	cod	ssde1	TRUE	FALSE	TRUE	<NA>	<NA>	001000	<NA>
32	cod	ssde1	TRUE	TRUE	TRUE	<NA>	<NA>	001000	<NA>
33	whiting	spline	FALSE	FALSE	TRUE	000000	<NA>	<NA>	<NA>
34	whiting	spline	FALSE	TRUE	TRUE	000000	000000	000000	<NA>
35	whiting	spline	TRUE	FALSE	TRUE	<NA>	<NA>	<NA>	<NA>
36	whiting	spline	TRUE	TRUE	TRUE	<NA>	000000	000000	<NA>
37	whiting	ssde1	FALSE	FALSE	TRUE	000000	<NA>	<NA>	<NA>
38	whiting	ssde1	FALSE	TRUE	TRUE	<NA>	000000	000000	<NA>
39	whiting	ssde1	TRUE	FALSE	TRUE	<NA>	<NA>	<NA>	<NA>
40	whiting	ssde1	TRUE	TRUE	TRUE	<NA>	<NA>	<NA>	<NA>

Table S3: Diagnostics summary of all model fits to North Sea stocks. NA indicates non-convergence of the fit. See supplement text for explanation of columns.

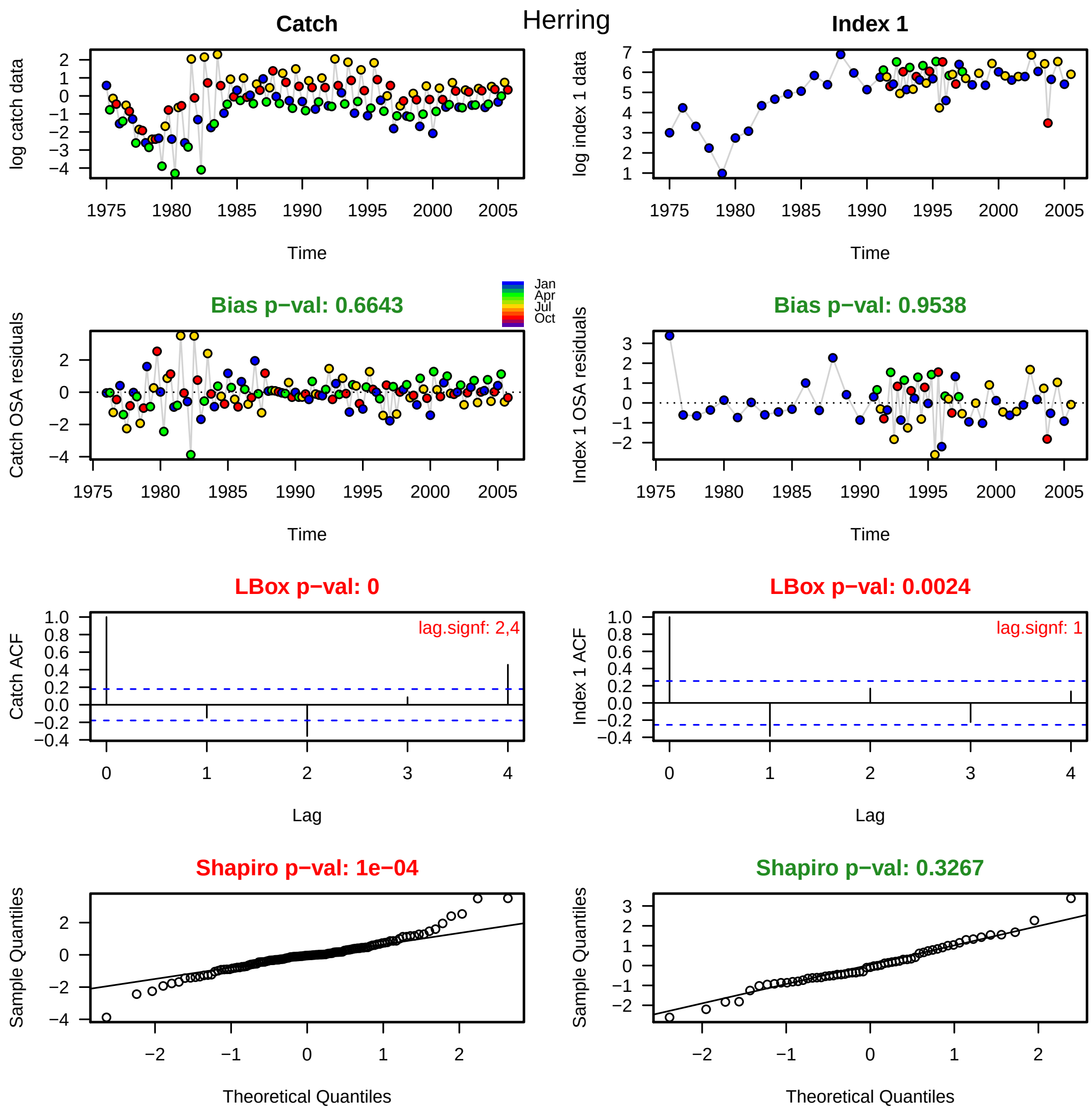

Figure S11: Residual diagnostics of SPiCT fit to quarterly data from North Sea herring, model of row 3 (rel.beta) in Table S3.

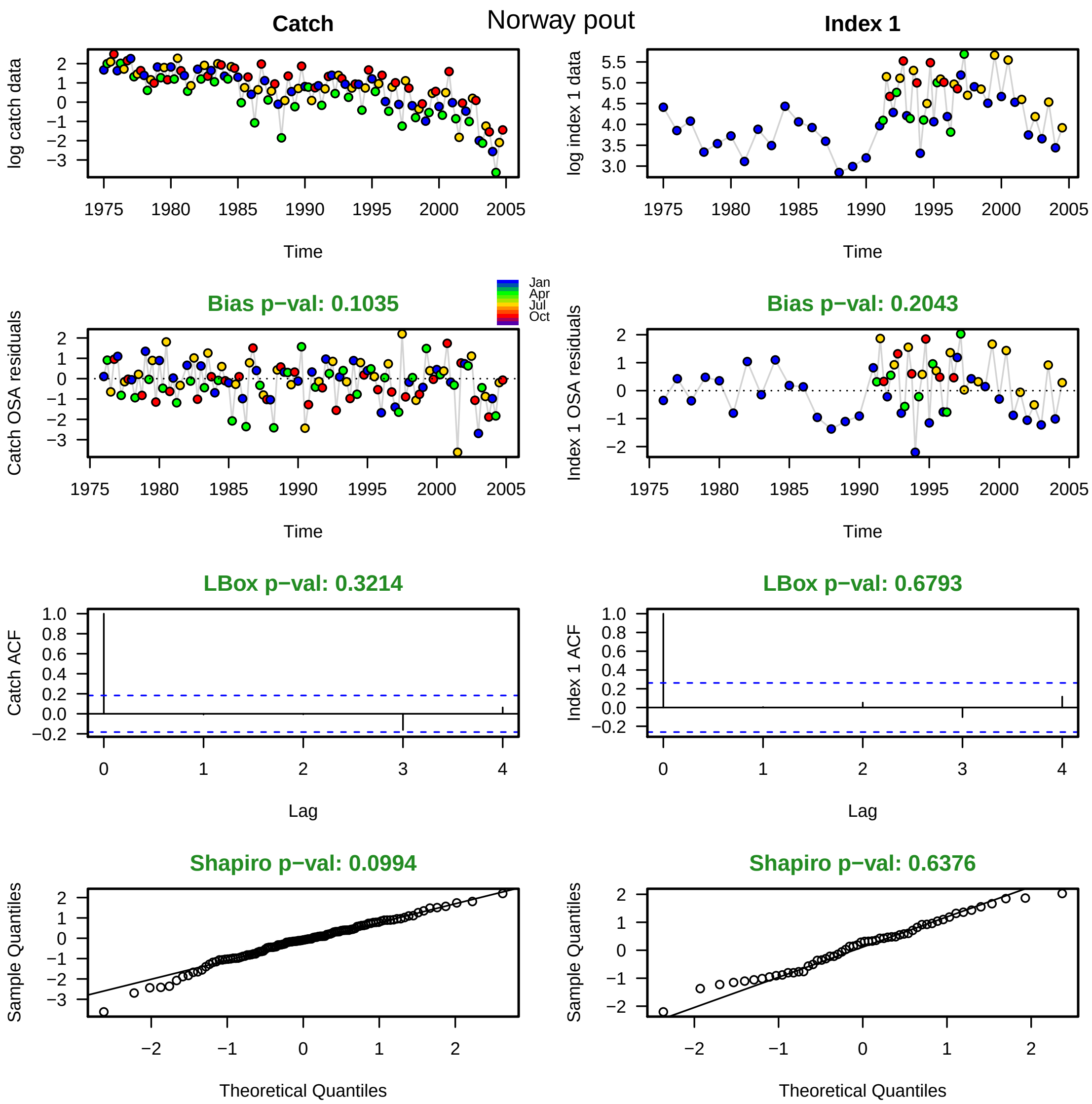

Figure S12: Residual diagnostics of SPiCT fit to quarterly data from North Sea Norway pout, model of row 13 (rel.n) in Table S3.

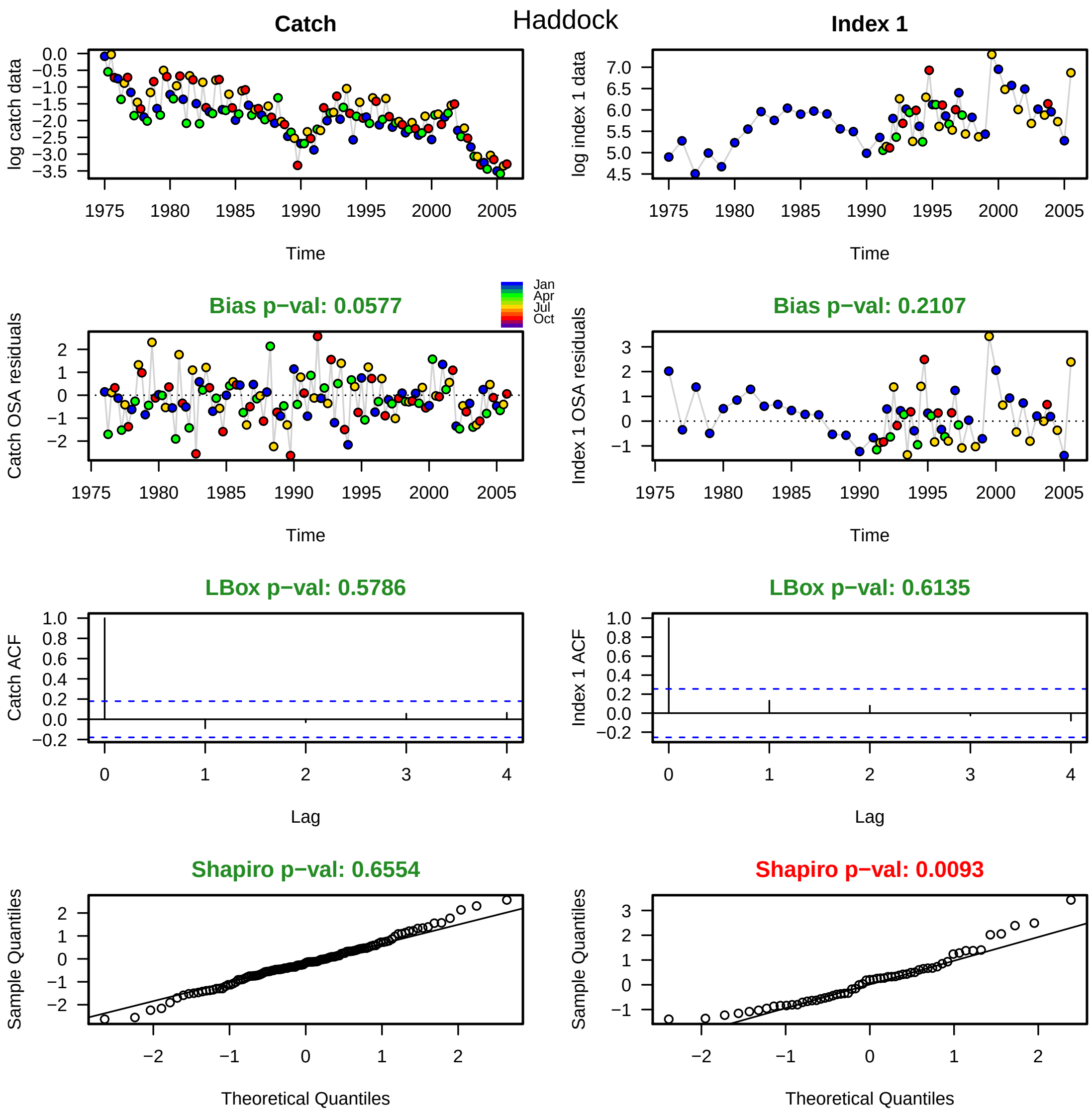

Figure S13: Residual diagnostics of SPiCT fit to quarterly data from North Sea haddock, model of row 21 (rel.n) in Table S3.

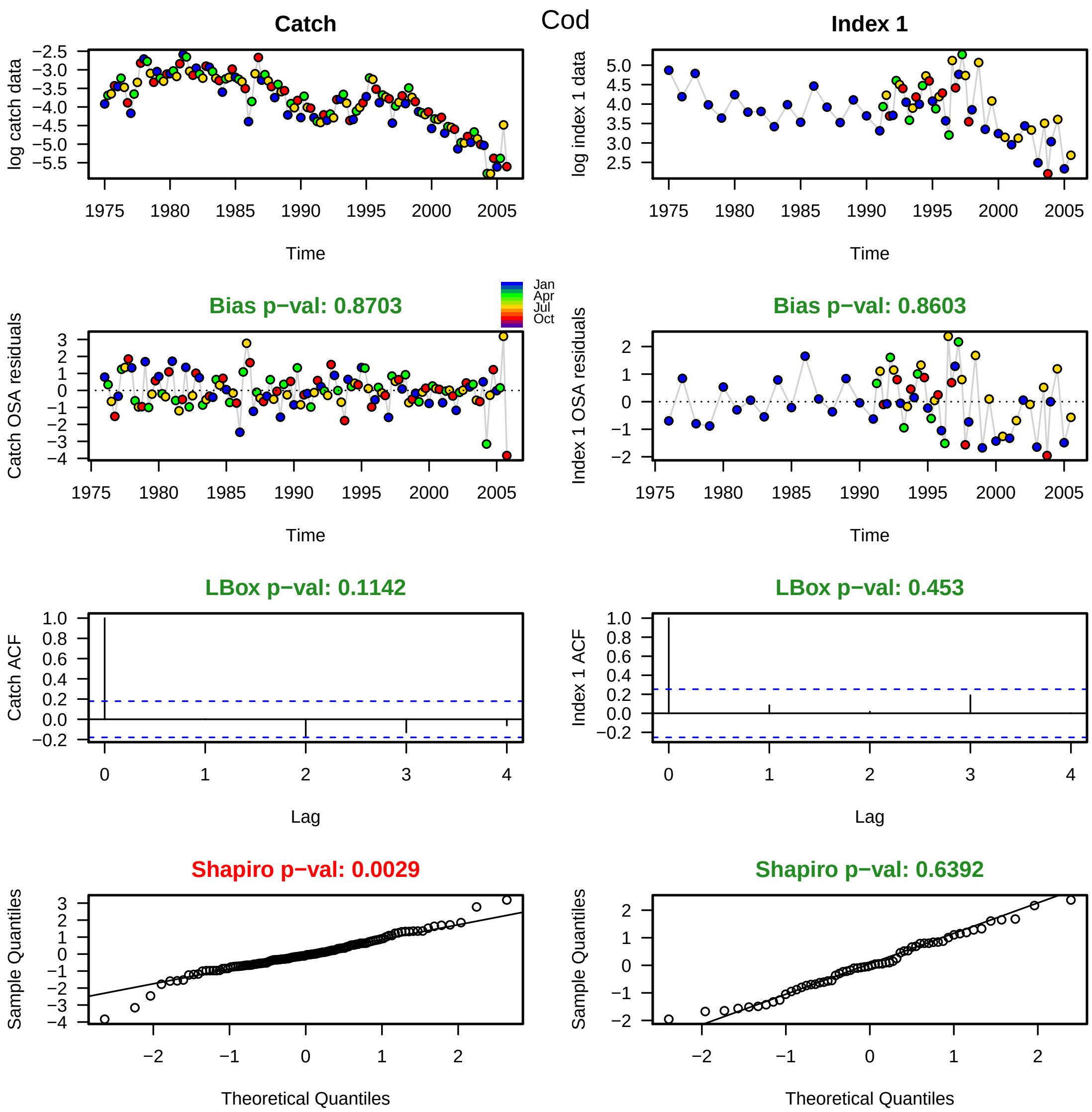

Figure S14: Residual diagnostics of SPiCT fit to quarterly data from North Sea cod, model of row 28 (rel.n) in Table S3.

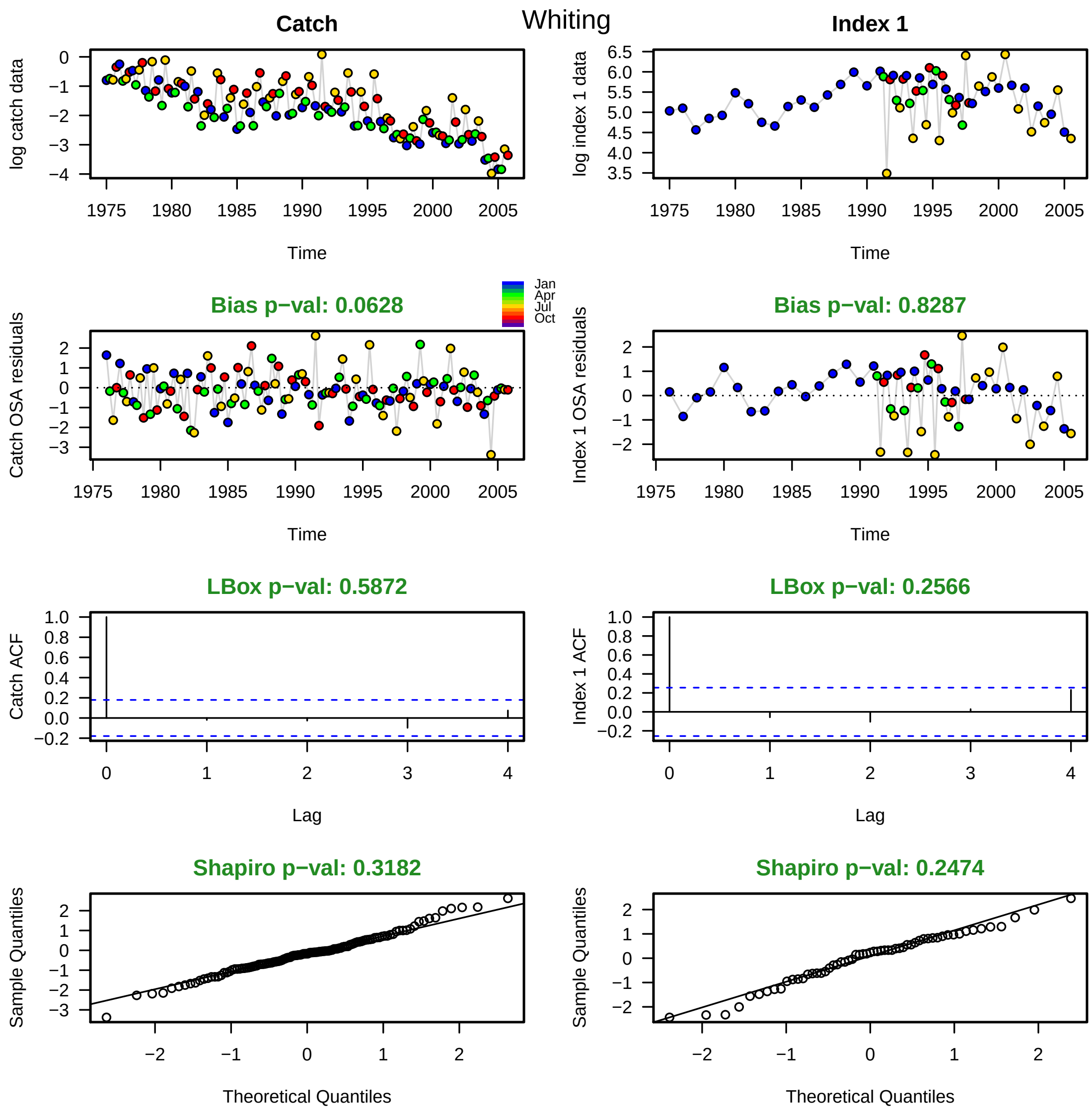

Figure S15: Residual diagnostics of SPiCT fit to quarterly data from North Sea whiting, model of row 34 (rel.n) in Table S3.

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.

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