Revised performance measures for the simulation testing of the DPC standardization

Henning Winker

Summary

Revised performance evaluation methods for the simulation testing of the DPC standardization procedure is given below (compare to section 2.5 in MARAM IWS/DEC13/Linefish CPUE/P1). The revised performance evaluation measures include two measures of accuracy in the form of root-mean-squared errors (RMSE) for the abundance trend and the year effect, respectively, and a bias estimate for the estimated abundance trend. The revised results are presented in Fig. 1 and Tables 1-2.

Performance evaluation

The performance of the DPC standardization models was evaluated in terms of the ability to estimate the true abundance trend r_i for species i in comparison to the nominal CPUE indices. Estimates of $\hat{r}_i$ for species i were obtained from a simple log-linear regression of the form:

$$\log(\overline{\text{CPUE}}_{i,y}) = \alpha + \hat{r}_i y \qquad y = 1, 2, \ldots, 20 \qquad (8)$$

where $\overline{\text{CPUE}}_{i,y}$ is the estimated mean CPUE for species i in year y based on either the nominal or standardized CPUE, and a is the intercept term.

Performance was evaluated for each model scenario using root-mean-squared error (RMSE) and bias when estimating a log-linear trend in abundance (Thorson et al., 2012a):

$$\mathrm{RMSE(t)}_i = \sqrt{\frac{1}{n_k}\sum_{k=1}^{n_k}(\hat{r}_{i,k} - r_{i,k})^2} \quad (9)$$

$$\mathrm{Bias}_{i,k} = \hat{r}_{i,k} - r_{i,k} \quad (10)$$

where RMSE(t) determines accuracy of the estimated abundance trend $\hat{r}_{i,k}$ compared to the true abundance trend $r_{i,k}$ for species i and simulation run k and n_k is the number of simulations. The scenarios H2.S4.E1 and H4.S10.E1 act as controls, for which $\hat{r}_i$ estimated from the nominal CPUE indices is expected to be unbiased.

To evaluate how accurately the nominal and standardized CPUE followed the simulated yearly biomass $B_{i,y}$, the indices were rescaled through normalization by the geometric mean and compared based on the root-mean-squared error of the form(Thorson et al., 2012b):

$$\mathrm{RMSE(y)}_i = \sqrt{\frac{1}{n_k n_y}\sum_{k=1}^{n_k}\sum_{y=1}^{n_y}(\hat{I}_{i,y,k} - I_{i,y,j})^2} \quad (11)$$

where $\mathrm{RMSE(y)}_{i,}$ determines accuracy of the estimated year effect for species i, n_k is the number of simulation runs, n_y is the number of years, $\hat{I}_{i,y,k}$ is the normalized abundance index for species i, simulation run k and year y based on nominal or standardized CPUE and $I_{i,y,k}$ is the normalized true abundance index for species i, simulation run k and year y based on the simulated biomass.

Comparisons of accuracy of the estimated abundance trends and year effects among species were achieved by calculating the difference in $RMSE(t)_i$ and $RMSE(y)_i$ between the nominal CPUE and DPC procedures with lowest estimation errors. The purpose of the comparisons was to examine the influence of species distribution among habitat on the relative performance of each procedure.

Revised results

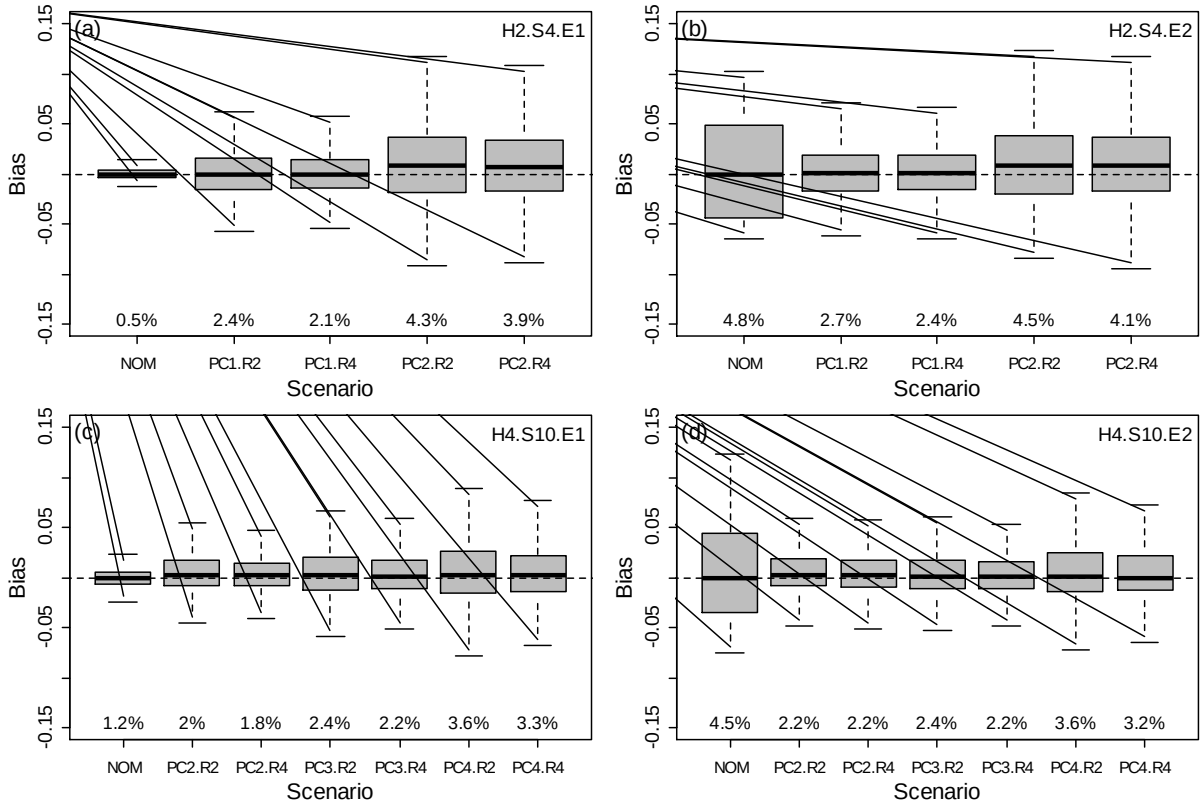

Fig. 1. Boxplots showing the difference between the true and estimated log-linear trend in abundance, averaged over all species, for each scenario and standardization model, with NOM denoting the nominal CPUE. The corresponding mean-root-squared errors (%) for the estimated abundance trend (RMSE(t)) are displayed below each boxplot.

Table 1. Root-mean-squared errors (%) of the estimated year effect (RMSE(y)), averaged over all species, for each scenario and standardization model

Scenario	H2.S4.E1	H2.S4.E2	H4.S10.E1	H4.S10.E2
Nominal	17.3%	43.5%	30.9%	42.4%
PC1.R2	16.3%	16.7%	-	-
PC2.R2	24.9%	26.2%	16.9%	17.9%
PC3.R2	-	-	17.7%	17.7%
PC4.R2	-	-	24.3%	24.1%
PC1.R4	**14.8%**	**16.7%**	-	-
PC2.R4	22.5%	23.7%	16.9%	18.6%
PC3.R4	-	-	**16.7%**	**16.8%**
PC4.R4	-	-	22.3%	22.1%

Table 2. Comparisons of differences in RMSE(t) and RMSE(y) mean by species between the nominal CPUE and the best-performing DPC model (PC3.R4) for control and test simulations of the four-habitat scenario (H4.S10). Positive values indicate a better performance of the DPC model.

Scenario	Species	ΔRSME(t) (Nom - DPC)		ΔRSME(y) (Nom - DPC)	
		Control	Test	Control	Test
H4.S10	KOB	-0.001	0.008	0.23	0.37
	GLBK	-0.001	0.007	0.19	0.31
.	HAKE	0.000	0.008	0.24	0.35
	CRPN	-0.004	0.004	0.10	0.23
	PANG	-0.002	0.006	0.18	0.30
	SNTR	-0.003	0.005	0.16	0.29
	ROMN	-0.002	0.005	0.11	0.23
	DGRD	-0.002	0.004	0.10	0.21
	RSTM	-0.002	0.005	0.10	0.22
	SHRK	-0.005	-0.001	-0.03	0.03