Appendices: Use of satellite data for understanding and predicting oil sardine (*Sardinella longiceps*) catch variability along the southwest coast of India

05 February, 2020

Appendix A: Tests for prior season catch as covariate

Table A1. Model selection tests of time-dependency the log catch during spawning months using F-tests of nested linear models. S_t is the catch during the spawning period (Jul-Sep). N_t is the catch during the non-spawning period (Oct-Jun). S_{t-1} and N_{t-1} are the catch during the prior season during and after the spawning period respectively. S_{t-2} and N_{t-2} are the same for two seasons prior. Test A uses catch during the spawning period as the explanatory variable. Test B uses catch during the non-spawning period as the explanatory variable. The numbers in front of the model equation indicate the level of nestedness. For Test C, there are two nested model sets, each with a different model 3. The Naive model is a model that uses the previous data point in the time series as the prediction; thus the Naive model has no estimated parameters.

Model	Residual df	MASE	Adj. R2	F	p value	AIC
Naive Model 1984-2015 data						
$ln(S_t) = ln(S_{t-1}) + \epsilon_t$	32	1				122.85
Time dependency test A 1984-2015 data						
1. $ln(S_t) = \alpha + ln(S_{t-1}) + \epsilon_t$	31	0.992	-29			124.83
2. $ln(S_t) = \alpha + \beta ln(S_{t-1}) + \epsilon_t$	30	0.814	10.3	15.14	0.001	114.14
3. $ln(S_t) = \alpha + \beta_1 ln(S_{t-1}) + \beta_2 ln(S_{t-2}) + \epsilon_t$	29	0.803	13.6	2.13	0.155	113.88
Time dependency test B 1984-2015 data						
1. $ln(S_t) = \alpha + ln(N_{t-1}) + \epsilon_t$	31	0.856	14.2			111.78
2. $ln(S_t) = \alpha + \beta ln(N_{t-1}) + \epsilon_t$	30	0.794	22.2	4.06	0.053	109.59
3. $ln(S_t) = \alpha + \beta_1 ln(N_{t-1}) + \beta_2 ln(N_{t-2}) + \epsilon_t$	29	0.797	19.6	0.01	0.919	111.57
Time dependency test C 1984-2015 data						
1. $ln(S_t) = \alpha + ln(N_{t-1}) + \epsilon_t$	31	0.856	14.2			111.78
2. $ln(S_t) = \alpha + \beta ln(N_{t-1}) + \epsilon_t$	30	0.794	22.2	4.08	0.053	109.59
3a. $ln(S_t) = \alpha + \beta_1 ln(N_{t-1}) + \beta_2 ln(S_{t-1}) + \epsilon_t$	29	0.804	20	0.16	0.688	111.4
3b. $ln(S_t) = \alpha + \beta_1 ln(N_{t-1}) + \beta_2 ln(S_{t-2}) + \epsilon_t$	29	0.778	20.8	0.45	0.508	111.09

Table A2. Model selection tests of time-dependency the catch during spawning months using non-linear responses instead of linear responses as in Table A1. See Table A1 for an explanation of the parameters and model set-up.

Model	Residual df	MASE	Adj. R2	F	p value	AIC
Time dependency test A 1984-2015 data						
1. $ln(S_t) = \alpha + \beta ln(S_{t-1}) + \epsilon_t$	30	0.814	10.3			114.14
2. $ln(S_t) = \alpha + s(ln(S_{t-1})) + \epsilon_t$	28.2	0.798	19.6	2.74	0.089	111.79
3. $ln(S_t) = \alpha + s_1(ln(S_{t-1})) + s_2(ln(S_{t-2})) + \epsilon_t$	25.5	0.77	20.7	0.97	0.416	113.23
Time dependency test B 1984-2015 data						
1. $ln(S_t) = \alpha + \beta ln(N_{t-1}) + \epsilon_t$	30	0.794	22.2			109.59
2. $ln(S_t) = \alpha + s(ln(N_{t-1})) + \epsilon_t$	28.6	0.761	24.4	1.26	0.287	109.52
3. $ln(S_t) = \alpha + s_1(ln(N_{t-1})) + s_2(ln(N_{t-2})) + \epsilon_t$	26.4	0.761	21.2	0.28	0.785	112.42
Time dependency test C 1984-2015 data						
1. $ln(S_t) = \alpha + s(ln(N_{t-1})) + \epsilon_t$	28.6	0.761	24.4			109.52
2. $ln(S_t) = \alpha + s_1(ln(N_{t-1})) + s_2(ln(S_{t-1})) + \epsilon_t$	26.1	0.698	28.5	1.49	0.242	109.55
3. $ln(S_t) = \alpha + s_1(ln(N_{t-1})) + s_2(ln(S_{t-2})) + \epsilon_t$	25.9	0.724	26.3	1.09	0.367	110.63

Table A3. Table A2 with 1956-1983 data instead of 1984 to 2015 data. See Table A1 for an explanation of the parameters and model set-up.

Model	Residual df	MASE	Adj. R2	F	p value	AIC
Time dependency test A 1956-1983 data						
1. $ln(S_t) = \alpha + \beta ln(S_{t-1}) + \epsilon_t$	24	0.633	-0.7			64.69
2. $ln(S_t) = \alpha + s(ln(S_{t-1})) + \epsilon_t$	22.1	0.614	-0.2	0.78	0.464	65.71
3. $ln(S_t) = \alpha + s_1(ln(S_{t-1})) + s_2(ln(S_{t-2})) + \epsilon_t$	19.9	0.58	3.1	1.19	0.329	66.35
Time dependency test B 1956-1983 data						
1. $ln(S_t) = \alpha + \beta ln(N_{t-1}) + \epsilon_t$	24	0.634	-3.8			65.48
2. $ln(S_t) = \alpha + s(ln(N_{t-1})) + \epsilon_t$	21.6	0.584	8.2	2.24	0.127	63.8
3. $ln(S_t) = \alpha + s_1(ln(N_{t-1})) + s_2(ln(N_{t-2})) + \epsilon_t$	18.5	0.495	16.9	1.56	0.231	63.13
Time dependency test C 1956-1983 data						
1. $ln(S_t) = \alpha + s(ln(N_{t-1})) + \epsilon_t$	22.5	0.586	4.3			66.2
2. $ln(S_t) = \alpha + s_1(ln(N_{t-1})) + s_2(ln(S_{t-1})) + \epsilon_t$	20.7	0.556	4.8	0.91	0.41	67.3
3. $ln(S_t) = \alpha + s_1(ln(N_{t-1})) + s_2(ln(S_{t-2})) + \epsilon_t$	19.5	0.55	12.9	1.42	0.266	63.79

Table A4. Model selection tests of time-dependency the N_t model using F-tests of nested models fit to 1984 to 2014 log landings data. The years are determined by the covariate data availability and end in 2014 since the landings data go through 2015 and N_{2014} includes quarters in 2014 and 2015. N_t is the catch during the non-spawning period (Qtrs 4 and 1: Oct-Mar) of season t (Jul-Jun). S_{t-1} and N_{t-1} are the catch during the prior sardine season during and after the spawning period respectively. S_{t-2} and N_{t-2} are the same for two seasons prior. Test A uses catch during the spawning period as the explanatory variable. Test B uses catch during the non-spawning period as the explanatory variable. Test C uses both. The numbers next to the model equations indicate the level of nestedness. The Naive model is a model that uses the previous data point in the time series as the prediction; thus the Naive model has no estimated parameters.

Model	Residual df	MASE	Adj. R2	F	p value	AIC
Naive Model 1984-2014 data						
$ln(N_t) = ln(N_{t-1}) + \epsilon_t$	31	1				90.87
Time dependency test A 1984-2014 data						
1. $ln(N_t) = \alpha + ln(S_{t-1}) + \epsilon$	30	1.363	-20.3			107.36
2. $ln(N_t) = \alpha + \beta ln(S_{t-1}) + \epsilon_t$	29	1.018	26.2	19.99	0	93.17
3. $ln(N_t) = \alpha + \beta_1 ln(S_{t-1}) + \beta_2 ln(S_{t-2}) + \epsilon_t$	28	1.009	26.6	1.15	0.292	93.92
Time dependency test B 1984-2014 data						
1. $ln(N_t) = \alpha + ln(N_{t-1}) + \epsilon_t$	30	0.999	24.7			92.87
2. $ln(N_t) = \alpha + \beta ln(N_{t-1}) + \epsilon_t$	29	0.978	37	6.63	0.016	88.28
3. $ln(N_t) = \alpha + \beta_1 ln(N_{t-1}) + \beta_2 ln(N_{t-2}) + \epsilon_t$	28	0.97	34.8	0.04	0.843	90.24
Time dependency test C 1984-2014 data						
1. $ln(N_t) = \alpha + \beta ln(N_{t-1}) + \epsilon_t$	29	0.978	37			88.28
2a. $ln(N_t) = \alpha + \beta_1 ln(N_{t-1}) + \beta_2 ln(S_{t-1}) + \epsilon_t$	28	0.964	35	0.12	0.729	90.15
2b. $ln(N_t) = \alpha + \beta_1 ln(N_{t-1}) + \beta_2 ln(S_{t-2}) + \epsilon_t$	28	0.978	34.7	0.01	0.919	90.27

Table A5. Model selection tests of time-dependency the N_t model using non-linear responses instead of linear responses as in Table A4 See Table A4 for an explanation of the parameters and model set-up.

Model	Residual df	MASE	Adj. R2	F	p value	AIC
Time dependency test A 1984-2014 data						
1. $ln(N_t) = \alpha + \beta ln(S_{t-1}) + \epsilon_t$	29	1.018	26.2			93.17
2. $ln(N_t) = \alpha + s(ln(S_{t-1})) + \epsilon_t$	27.3	0.992	30.2	1.83	0.185	92.61
3. $ln(N_t) = \alpha + s_1(ln(S_{t-1})) + s_2(ln(S_{t-2})) + \epsilon_t$	24.4	0.94	36.4	1.79	0.177	91.62
Time dependency test B 1984-2014 data						
1. $ln(N_t) = \alpha + \beta ln(N_{t-1}) + \epsilon_t$	29	0.978	37			88.28
2. $ln(N_t) = \alpha + s(ln(N_{t-1})) + \epsilon_t$	27.6	0.874	45.3	3.88	0.047	84.75
3. $ln(N_t) = \alpha + s_1(ln(N_{t-1})) + s_2(ln(N_{t-2})) + \epsilon_t$	25.4	0.805	45.6	0.87	0.441	86.11
Time dependency test C 1984-2014 data						
1. $ln(N_t) = \alpha + s(ln(N_{t-1})) + \epsilon_t$	27.6	0.874	45.3			84.75
2. $ln(N_t) = \alpha + s_1(ln(N_{t-1})) + s_2(ln(S_{t-1})) + \epsilon_t$	25.1	0.856	43.8	0.53	0.634	87.37
3. $ln(N_t) = \alpha + s_1(ln(N_{t-1})) + s_2(ln(S_{t-2})) + \epsilon_t$	24.8	0.743	56.6	3.39	0.036	79.53

Table A6. Table A5 with 1956-1983 data instead of 1984 to 2014 data. See Table A4 for an explanation of the parameters and model set-up.

Model	Residual df	MASE	Adj. R2	F	p value	AIC
Time dependency test A 1956-1983 data						
1. $ln(N_t) = \alpha + \beta ln(S_{t-1}) + \epsilon_t$	24	0.641	-1.7			44.98
2. $ln(N_t) = \alpha + s(ln(S_{t-1})) + \epsilon_t$	22.1	0.534	16.2	3.53	0.052	41.11
3. $ln(N_t) = \alpha + s_1(ln(S_{t-1})) + s_2(ln(S_{t-2})) + \epsilon_t$	19.9	0.502	18.1	1.09	0.362	42
Time dependency test B 1956-1983 data						
1. $ln(N_t) = \alpha + \beta ln(N_{t-1}) + \epsilon_t$	24	0.681	-4.2			45.61
2. $ln(N_t) = \alpha + s(ln(N_{t-1})) + \epsilon_t$	21.6	0.507	29.1	5.69	0.009	37.12
3. $ln(N_t) = \alpha + s_1(ln(N_{t-1})) + s_2(ln(N_{t-2})) + \epsilon_t$	18.5	0.471	32.2	1.14	0.36	37.87
Time dependency test C 1956-1983 data						
1. $ln(N_t) = \alpha + s(ln(N_{t-1})) + \epsilon_t$	21.6	0.507	29.1			37.12
2a. $ln(N_t) = \alpha + s_1(ln(N_{t-1})) + s_2(ln(S_{t-1})) + \epsilon_t$	19	0.45	34.4	1.49	0.251	36.74
2b. $ln(N_t) = \alpha + s_1(ln(N_{t-1})) + s_2(ln(S_{t-2})) + \epsilon_t$	19.5	0.465	33.4	1.54	0.24	36.84

Appendix B: Tests for environmental variables as covariates

Table B1. Model selection tests of GPCP precipitation as an explanatory variable for the catch S_t during monsoon months (Jul-Sep) using 1984 to 2015 data. The data range is determined by the years for which SST was available in order to use a consistent dataset across covariate tests. The base model (M) with prior catch dependency was selected independently (Appendix A). To the base model, covariates are added. V_t is the covariate in same calendar year as the Jul-Sep catch. The specific hypothesis (Table 1) being tested is noted in parentheses. The models are tested as nested sets. Thus 1, 2a, 3a is a set and 1, 2b, 3b is another set. MASE is the mean absolute square error (residuals).

Model	Residual df	MASE	Adj. R2	F	p value	AIC
base model (M) 1984-2015 data						
1. $ln(S_t) = \alpha + s(ln(N_{t-1})) + \epsilon_t$	28.6	0.761	24.4			109.52
V_t = Jun-Jul Precipitation (S1)						
2. $ln(S_t)$ = M + βV_t	27.6	0.743	23.5	0.67	0.42	110.78
3. $ln(S_t)$ = M + $s(V_t)$	26	0.734	27	1.51	0.241	110.28
V_t = Apr-May Precipitation (S2)						
2. $ln(S_t)$ = M + βV_t	27.6	0.756	23.8	0.72	0.403	110.65
3. $ln(S_t)$ = M + $s(V_t)$	25.6	0.748	21.1	0.24	0.792	112.98

Table B2. Model selection tests of sea surface temperature off the Kerala coast (up to 80km offshore in boxes 2-5 in Figure 1), and upwelling indices as the explanatory variables (V_t) for the catch during monsoon months (Jul-Sep) using 1984 to 2015 data. The hypothesis tested (Table 1) is noted in parentheses. Three upwelling indices were tested. The nearshore-offshore temperature diferential (UPW), which is the offshore (box 13) minus nearshore (box 4) SST, the average nearshore SST along the Kerala coast (boxes 2-5), and the Bakun upwelling index based on wind stress. See Table B1 for an explanation of the models.

Model	Residual df	MASE	Adj. R2	F	p value	AIC
base model (M) 1984-2015 data						
1. $ln(S_t) = \alpha + s(ln(N_{t-1})) + \epsilon_t$	28.6	0.761	24.4			109.52
V_t = Ave Mar-May SST (S4)						
2a. $ln(S_t)$ = M + βV_t	27.6	0.772	23.3	0.59	0.449	110.87
3a. $ln(S_t)$ = M + $s(V_t)$	25.5	0.753	26.4	1.26	0.303	110.8
2b. $ln(S_t)$ = M + βV_{t-1}	27.6	0.754	22.8	0.39	0.533	111.07
3b. $ln(S_t)$ = M + $s(V_{t-1})$	25.7	0.723	26.1	1.35	0.275	110.74
V_t = Ave Oct-Dec SST (L1)						
2. $ln(S_t)$ = M + βV_{t-1}	27.6	0.759	21.8	0	0.952	111.5
$\Rightarrow$ 3. $ln(S_t)$ = M + $s(V_{t-1})$	26.1	0.768	21.8	0.66	0.482	112.32
V_t = Ave. Jun-Sep UPW (S4 and L2)						
2a. $ln(S_t)$ = M + βV_t	27.6	0.706	33.5	5.01	0.034	106.32
3a. $ln(S_t)$ = M + $s(V_t)$	25.5	0.682	34.1	0.83	0.455	107.24
2b. $ln(S_t)$ = M + βV_{t-1}	27.6	0.748	22.6	0.33	0.568	111.15
3b. $ln(S_t)$ = M + $s(V_{t-1})$	25.1	0.724	26.2	1.28	0.3	111.16
V_t = Ave. Jun-Sep SST (S4 and L2)						
2a. $ln(S_t)$ = M + βV_t	27.6	0.745	33.3	5.53	0.027	106.4
$\Rightarrow$ 3a. $ln(S_t)$ = M + $s(V_t)$	25.9	0.683	41	2.85	0.084	103.43
2b. $ln(S_t)$ = M + βV_{t-1}	27.6	0.742	23.2	0.54	0.468	110.89
3b. $ln(S_t)$ = M + $s(V_{t-1})$	25.5	0.715	22.2	0.54	0.599	112.57
V_t = Ave. Jun-Sep Bakun-UPW (L2)						
2a. $ln(S_t)$ = M + βV_t	27.6	0.776	28.4	3.62	0.069	108.66
$\Rightarrow$ 3a. $ln(S_t)$ = M + $s(V_t)$	25.5	0.633	47.8	5.66	0.009	99.8
2b. $ln(S_t)$ = M + βV_{t-1}	27.7	0.744	26.1	1.67	0.207	109.65
3b. $ln(S_t)$ = M + $s(V_{t-1})$	25.6	0.728	24.9	0.48	0.628	111.36

Table B2. Model selection tests of the multi-year average nearshore sea surface temperature and ENSO indices as the explanatory variables (V) for the catch during summer months (Jul-Sep) using 1984 to 2015 data. The hypothesis tested (Table 1) is noted in parentheses. The ENSO indices were the ONI index averaged over all months in the calendar year and the DMI index for Sep-Nov. The 2.5-year average SST is the average for Jan-Jun in the current calendar year and the prior 2 calendar years (30 months total). Thus the average does not include any months during the Jul-Sep catch. See Table B1 for an explanation of the models.

Model	Residual df	MASE	Adj. R2	F	p value	AIC
base model (M) 1984-2015 data						
1. $ln(S_t) = \alpha + s(ln(N_{t-1})) + \epsilon_t$	28.6	0.761	24.4			109.52
V_t = 2.5-year average SST (A1)						
2. $ln(S_t) =$ M + βV_t	27.6	0.723	33.2	5.52	0.027	106.43
3. $ln(S_t) =$ M + $s(V_t)$	26.2	0.653	41	3.22	0.07	103.26
V_t = ONI (A2)						
2. $ln(S_t) =$ M + βV_{t-1}	27.6	0.758	22	0.08	0.77	111.4
3. $ln(S_t) =$ M + $s(V_{t-1})$	26.6	0.733	23.6	1.16	0.294	111.28
V_t = DMI (A3)						
2. $ln(S_t) =$ M + βV_{t-1}	27.6	0.756	21.9	0.03	0.869	111.42
3. $ln(S_t) =$ M + $s(V_{t-1})$	24.7	0.761	19	0.41	0.744	114.41

Table B3. Model selection tests of GPCP precipitation as an explanatory variable for the catch (N_t) during post-monsoon months (Oct-May) using 1984 to 2014 data. The data range is determined by the years for which SST was available in order to use a consistent dataset across covariate tests. The base model (M) with prior catch dependency was selected independently (Appendix A). N_{t-1} is the post-monsoon catch in prior season, and S_{t-2} is the catch during Jul-Sep two seasons prior. To the base model, covariates are added. V_t is the covariate in the calendar year, and V_{t-1} is the covariate in the prior calendar year. The specific hypothesis (Table 1) being tested is noted in parentheses. The models are tested as nested sets. Thus 1, 2a, 3a is a set and 1, 2b, 3b is another set.

Model	Residual df	MASE	Adj. R2	F	p value	AIC
base model (M) 1984-2014 data						
1. $ln(N_t) = \alpha + s(ln(N_{t-1})) + s(ln(S_{t-2})) + \epsilon_t$	24.8	0.743	56.6			79.53
V_t = Jun-Jul Precipitation (S1)						
2a. $ln(N_t) = \text{M} + \beta V_t$	23.8	0.755	56.7	1.03	0.318	80.23
3a. $ln(N_t) = \text{M} + s(V_t)$	22.3	0.75	55.3	0.19	0.767	82.02
2b. $ln(N_t) = \text{M} + \beta V_{t-1}$	23.8	0.744	54.9	NA	NA	81.5
3b. $ln(N_t) = \text{M} + s(V_{t-1})$	22.3	0.701	56.4	1.32	0.28	81.18
V_t = Apr-May Precipitation (S2)						
2a. $ln(N_t) = \text{M} + \beta V_t$	23.8	0.742	55.1	0.11	0.735	81.34
3a. $ln(N_t) = \text{M} + s(V_t)$	21.7	0.73	53.7	0.36	0.707	83.39
2b. $ln(N_t) = \text{M} + \beta V_{t-1}$	23.8	0.723	56.2	0.74	0.397	80.6
3b. $ln(N_t) = \text{M} + s(V_{t-1})$	22	0.692	55.6	0.5	0.587	81.87

Table B4. Model selection tests of sea surface temperature off the Kerala coast (up to 80km offshore in boxes 2-5 in Figure 1), and upwelling indices as the explanatory variables (V) for the catch during post-monsoon months (Oct-May) using 1984 to 2014 data. The hypothesis tested (Table 1) is noted in parentheses. Three upwelling indices were tested. The nearshore-offshore temperature diferential (UPW), which is the offshore (box 13) minus nearshore (box 4) SST, the average nearshore SST along the Kerala coast (boxes 2-5), and the Bakun upwelling index based on wind stress. See Table B3 for an explanation of the models.

Model	Residual df	MASE	Adj. R2	F	p value	AIC
base model (M) 1984-2014 data						
1. $ln(N_t) = \alpha + s(ln(N_{t-1})) + s(ln(S_{t-2})) + \epsilon_t$	24.8	0.743	56.6			79.53
V_t = Ave Mar-May SST (S5)						
2a. $ln(N_t)$ = M + βV_t	23.8	0.701	59	2.84	0.107	78.53
3a. $ln(N_t)$ = M + $s(V_t)$	22	0.682	63.2	2.29	0.13	76.01
2b. $ln(N_t)$ = M + βV_{t-1}	23.8	0.762	57.1	1.33	0.26	79.93
3b. $ln(N_t)$ = M + $s(V_{t-1})$	22	0.747	57.4	0.79	0.455	80.61
V_t = Ave Oct-Dec SST (L1)						
2. $ln(N_t)$ = M + βV_{t-1}	23.8	0.748	54.9	NA	NA	81.5
3. $ln(N_t)$ = M + $s(V_{t-1})$	22.5	0.736	56	1.13	0.318	81.37
V_t = Ave. Jun-Sep SST-UPW (L2)						
$\Rightarrow$ 2a. $ln(N_t)$ = M + βV_t	23.8	0.759	62.2	4.91	0.038	76
3a. $ln(N_t)$ = M + $s(V_t)$	21.4	0.733	62.3	0.74	0.513	77.2
2b. $ln(N_t)$ = M + βV_{t-1}	23.8	0.742	54.9	0	0.979	81.49
3b. $ln(N_t)$ = M + $s(V_{t-1})$	21.4	0.711	56.5	1.12	0.351	81.6
V_t = Ave. Jun-Sep SST (L2)						
$\Rightarrow$ 2a. $ln(N_t)$ = M + βV_t	23.8	0.717	62.7	5.27	0.033	75.57
3a. $ln(N_t)$ = M + $s(V_t)$	21.9	0.714	61.8	0.39	0.67	77.33
2b. $ln(N_t)$ = M + βV_{t-1}	23.8	0.744	55.3	0.23	0.626	81.18
3b. $ln(N_t)$ = M + $s(V_{t-1})$	21.8	0.76	54.6	0.49	0.616	82.72
V_t = Ave. Jun-Sep Bakun-UPW (L2)						
2a. $ln(N_t)$ = M + βV_t	23.8	0.758	57.4	1.58	0.221	79.75
3a. $ln(N_t)$ = M + $s(V_t)$	21.8	0.672	60.3	1.58	0.228	78.55
2b. $ln(N_t)$ = M + βV_{t-1}	23.8	0.765	56.8	1.17	0.287	80.13
3b. $ln(N_t)$ = M + $s(V_{t-1})$	22	0.74	57.9	1.08	0.349	80.24

Table B5. Model selection tests of the multi-year average nearshore sea surface temperature and ENSO indices as the explanatory variables (V) for the catch during post-monsoon months (Oct-May) using 1984 to 2014 data. The hypothesis tested (Table 1) is noted in parentheses. The ENSO indices were the ONI index averaged over all months in the calendar year and the DMI index for Sep-Nov. The 2.5-year average SST is the average for Jan-Jun in the current calendar year and the prior 2 calendar years (30 months total). Thus the average does not include any months during the Oct-Mar catch. See Table B3 for an explanation of the models.

Model	Residual df	MASE	Adj. R2	F	p value	AIC
base model (M) 1984-2014 data						
1. $ln(N_t) = \alpha + s(ln(N_{t-1})) + s(ln(S_{t-2})) + \epsilon_t$	24.8	0.743	56.6			79.53
V_t = 2.5-year average SST (A1)						
2. $ln(N_t)$ = M + βV_t	23.8	0.667	64.7	7.68	0.012	73.9
$\Rightarrow$ 3. $ln(N_t)$ = M + $s(V_t)$	22.7	0.594	67.5	2.58	0.12	71.88
V_t = ONI (A2)						
2a. $ln(N_t)$ = M + βV_t	23.9	0.794	56.4	0.86	0.351	80.41
3a. $ln(N_t)$ = M + $s(V_t)$	22.8	0.79	56.9	0.93	0.351	80.5
2b. $ln(N_t)$ = M + βV_{t-1}	23.8	0.744	54.9	NA	NA	81.46
3b. $ln(N_t)$ = M + $s(V_{t-1})$	23	0.748	55.5	0.99	0.313	81.46
V_t = DMI (A3)						
2a. $ln(N_t)$ = M + βV_t	23.9	0.791	55.7	0.42	0.498	80.92
3a. $ln(N_t)$ = M + $s(V_t)$	21.2	0.77	56.8	1	0.404	81.54
2b. $ln(N_t)$ = M + βV_{t-1}	23.8	0.746	54.8	NA	NA	81.52
$\Rightarrow$ 3b. $ln(N_t)$ = M + $s(V_{t-1})$	21.1	0.678	67.5	4.34	0.018	72.69

Appendix C: Tests for Chlorophyll-a as a covariate

Table C1. Model selection tests of Chlorophyll-a (CHL) as an explanatory variable for the Jul-Sep catch (S_t) using 1998 to 2014 data. The data range is determined by the years for which CHL was available. V_t is CHL in the current season which spans two calendar years from July to June in the next year. V_{t-1} is CHL in the prior Jul-Jun season. Only CHL in Oct-Dec and Jan-Mar in the prior season is used since for the current season, these months are after the Jul-Sep catch being modeled. Non-linearity is modeled as a 2nd-order polynomial due to data constraints and appears as $p()$ in the model equations. The Jul-Sep catch is modeled as a function of Oct-Jun catch in the prior year only, without Jul-Sep catch 2-years prior as in the other covariate analyses (Appendix B). This is done due to data constraints. The models are nested; the Roman numeral indicates the level of nestedness. Models at levels II and higher are shown with the component that is added to the base level model (M) at top.

Model	Residual df	MASE	Adj. R2	F	p value	AIC
base model (M) 1998-2014 data						
1. $ln(S_t) = \alpha + p(ln(N_{t-1})) + \epsilon_t$	14	0.516	25.3			18.29
V_t = Jul-Sep Chlorophyll						
2. $ln(S_t) = M + \beta V_t$	13	0.503	24.6	0.69	0.427	19.2
3. $ln(S_t) = M + p(V_t)$	12	0.48	19.5	0.16	0.699	20.94
4. $ln(S_t) = M + p(V_t) + \beta V_{t-1}$	11	0.5	13.7	0.17	0.688	22.65
5. $ln(S_t) = M + p(V_t) + p(V_{t-1})$	10	0.497	5.1	0.01	0.935	24.64
V_t = Oct-Dec Chlorophyll						
2. $ln(S_t) = M + \beta V_{t-1}$	13	0.516	19.6	0	0.99	20.29
3. $ln(S_t) = M + p(V_{t-1})$	12	0.456	21.5	1.33	0.272	20.51
V_t = Jan-Mar Chlorophyll						
2. $ln(S_t) = M + \beta V_{t-1}$	13	0.522	20.6	0.16	0.697	20.08
3. $ln(S_t) = M + p(V_{t-1})$	12	0.526	16.7	0.4	0.541	21.52

Table C2. Model selection tests of Chlorophyll-a (CHL) as an explanatory variable for Oct-Jun catch (N_t) using 1998 to 2014 data. The data range is determined by the years for which CHL was available. V_t is CHL in the current season which spans two calendar years from July to June in the next year. V_{t-1} is CHL in the prior Jul-Jun season. Non-linearity is modeled as a 2nd-order polynomial due to data constraints and appears as $p()$ in the model equations. The Oct-Jun catch is modeled as a function of Oct-Jun catch in the prior year only, without Jul-Sep catch 2-years prior as in the other covariate analyses (Appendix B). This was done due to data constraints. The models are nested; the numeral indicates the level of nestedness. Models at levels 2 and higher are shown with the component that is added to the base level model (M) at top.

Model	Residual df	MASE	Adj. R2	F	p value	AIC
base model (M) 1998-2014 data						
1-M. $ln(N_t) = \alpha + p(ln(N_{t-1})) + \epsilon_t$	14	0.875	26.5			18.94
V_t = Jul-Sep Chlorophyll						
2. $ln(N_t) = M + \beta V_t$	13	0.893	23.1	0.32	0.587	20.45
3. $ln(N_t) = M + p(V_t)$	12	0.874	17.9	0.15	0.709	22.21
2. $ln(N_t) = M + \beta V_{t-1}$	13	0.86	25	0.69	0.422	20.03
3. $ln(N_t) = M + p(V_{t-1})$	11.7	0.839	21.7	0.27	0.677	21.36
V_t = Oct-Dec Chlorophyll						
2. $ln(N_t) = M + \beta V_t$	13	0.883	23.9	0.59	0.458	20.29
3. $ln(N_t) = M + p(V_t)$	12	0.744	29.5	2.22	0.167	19.62
4. $ln(N_t) = M + p(V_t) + \beta V_{t-1}$	11	0.679	40.8	2.99	0.114	17.16
5. $ln(N_t) = M + p(V_t) + p(V_{t-1})$	10	0.68	34.9	0	0.976	19.16
2. $ln(N_t) = M + \beta V_{t-1}$	13	0.764	39.4	3.87	0.074	16.41
3. $ln(N_t) = M + p(V_{t-1})$	11.3	0.728	37.7	0.49	0.595	17.62
V_t = Jan-Mar Chlorophyll						
2. $ln(N_t) = M + \beta V_t$	13	0.901	23.6	0.4	0.541	20.34
3. $ln(N_t) = M + p(V_t)$	12	0.829	23.9	0.89	0.367	20.92
2. $ln(N_t) = M + \beta V_{t-1}$	13	0.866	21.2	0.05	0.829	20.88
3. $ln(N_t) = M + p(V_{t-1})$	11.1	0.873	15.2	0.23	0.791	22.97

Table C3. Model selection tests of Chlorophyll-a as an explanatory variable for the catch during the non-spawning months (Oct-Jun) using box 5.

Model	Residual df	MASE	Adj. R2	F	p value	AIC
base model (M) 1998-2014 data						
1. $ln(N_t) = \alpha + p(ln(N_{t-1})) + \epsilon_t$	14	0.875	26.5			18.94
V_t = Jul-Sep Chlorophyll						
2. $ln(N_t) = M + \beta V_t$	13	0.865	22.5	0.24	0.635	20.6
3. $ln(N_t) = M + p(V_t)$	12	0.904	20.4	0.61	0.451	21.69
2. $ln(N_t) = M + \beta V_{t-1}$	13	0.839	28.5	1.33	0.271	19.22
3. $ln(N_t) = M + p(V_{t-1})$	12	0.837	25.2	0.07	0.789	20.42
V_t = Oct-Dec Chlorophyll						
2. $ln(N_t) = M + \beta V_t$	13	0.864	28.4	1.4	0.265	19.25
3. $ln(N_t) = M + p(V_t)$	12	0.844	24	0.26	0.62	20.91
4. $ln(N_t) = M + p(V_t) + \beta V_{t-1}$	11	0.666	35.6	2.9	0.119	18.62
5. $ln(N_t) = M + p(V_t) + p(V_{t-1})$	10	0.649	29.9	0.11	0.743	20.42
2. $ln(N_t) = M + \beta V_{t-1}$	13	0.739	35.5	2.88	0.116	17.48
3. $ln(N_t) = M + p(V_{t-1})$	11.7	0.732	34.2	0.52	0.534	18.39
V_t = Jan-Mar Chlorophyll						
2. $ln(N_t) = M + \beta V_t$	13	0.847	29.5	1.56	0.24	18.98
3. $ln(N_t) = M + p(V_t)$	12	0.804	31.6	1.33	0.276	19.11
2. $ln(N_t) = M + \beta V_{t-1}$	13	0.89	21.4	0.09	0.769	20.84
3. $ln(N_t) = M + p(V_{t-1})$	8.9	0.682	27.9	1.07	0.427	20.97

Appendix D: Covariates along the SE India coast

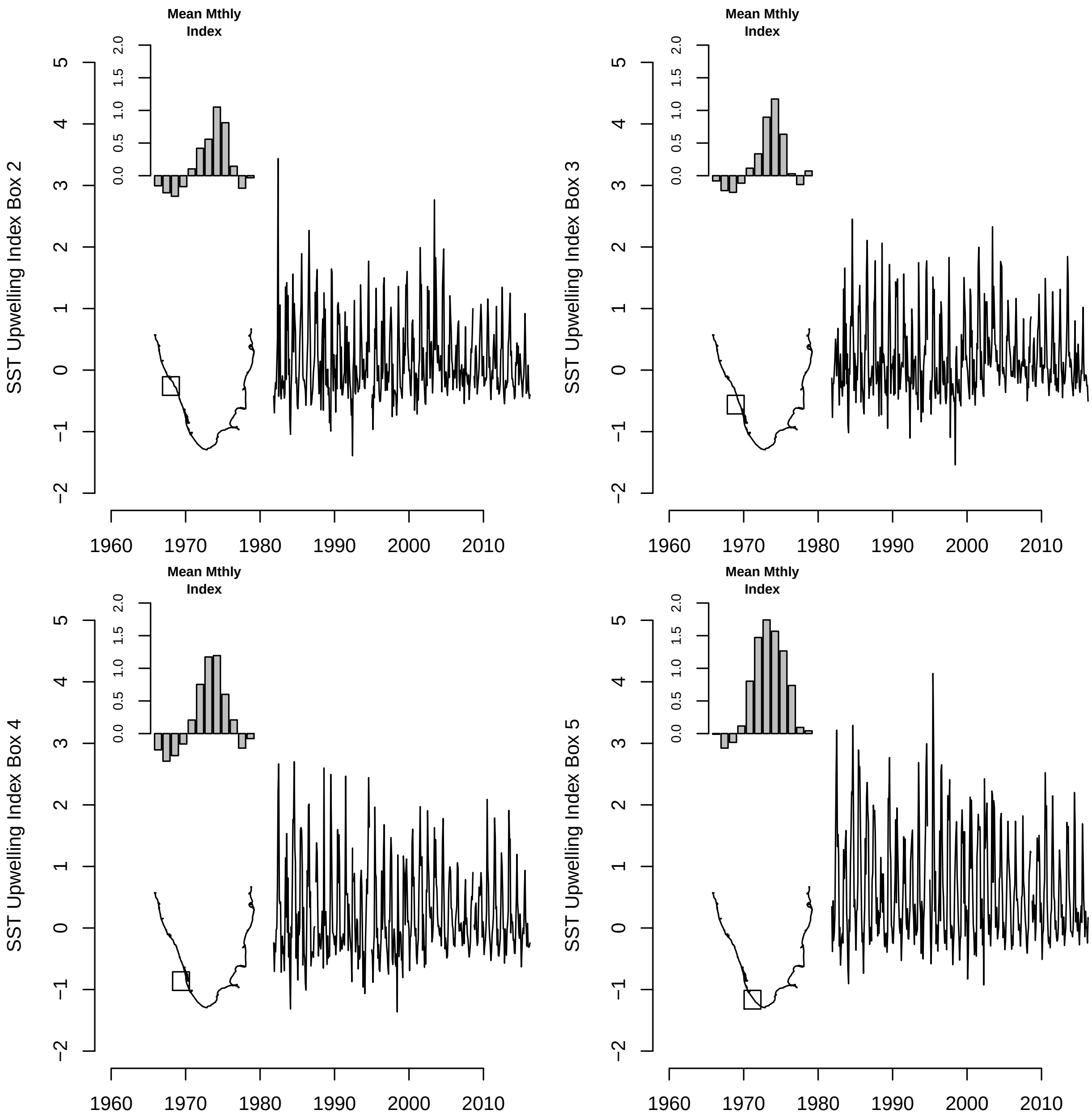

Figure D1. Upwelling index.

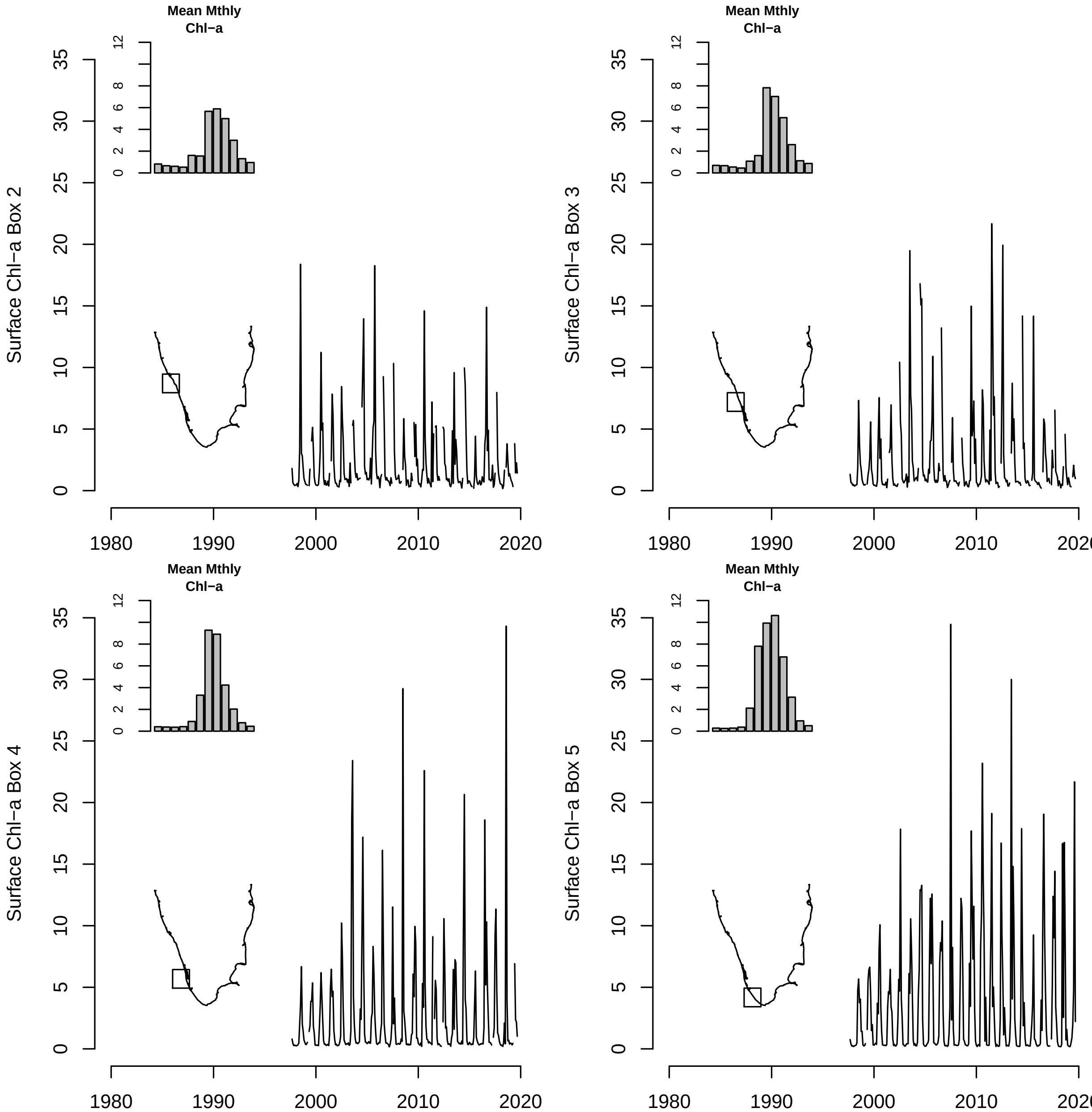

Figure D2. Chlorophyll-a.

Appendix E: Comparison of land and oceanic rainfall measurements

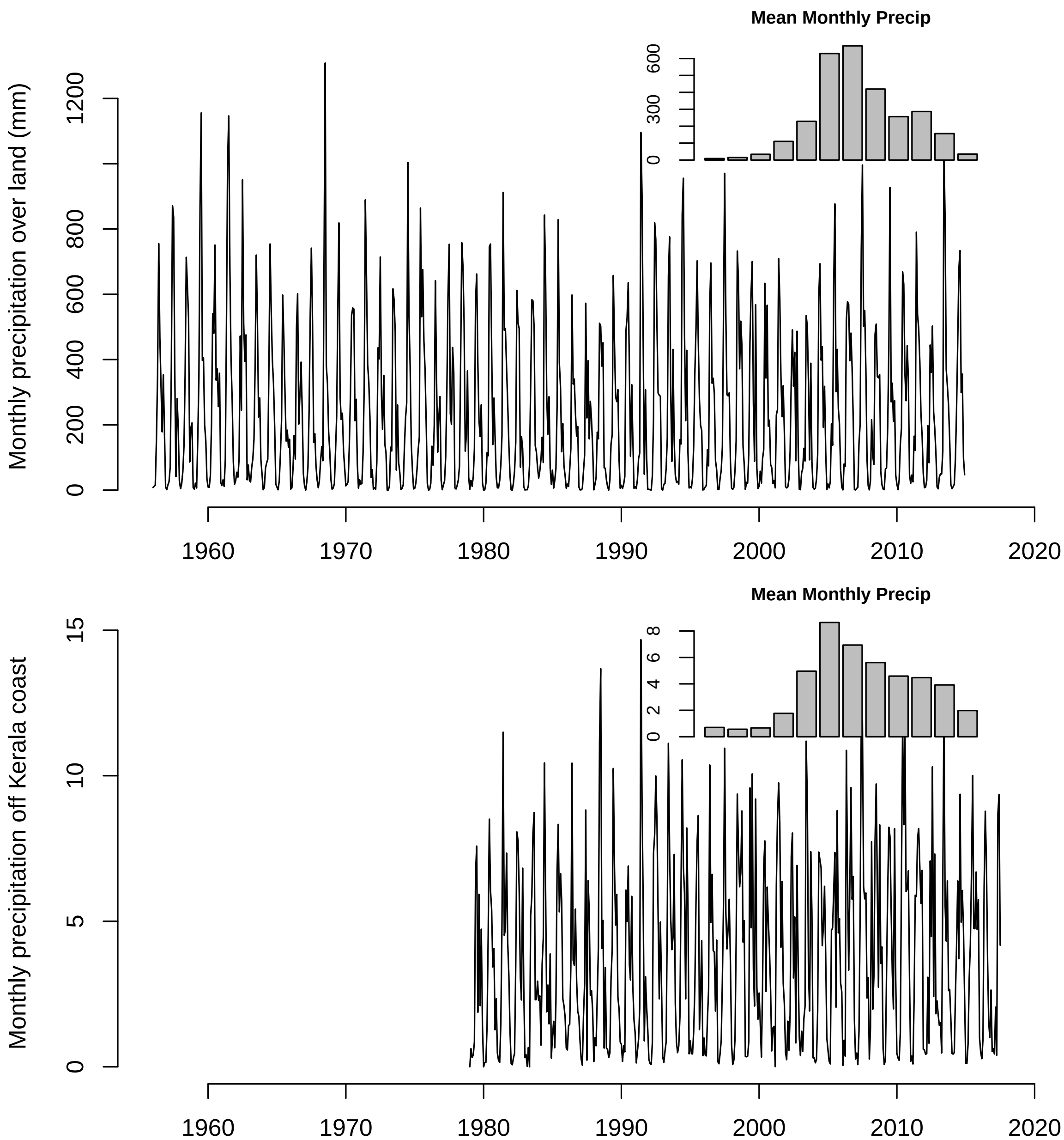

Figure E1

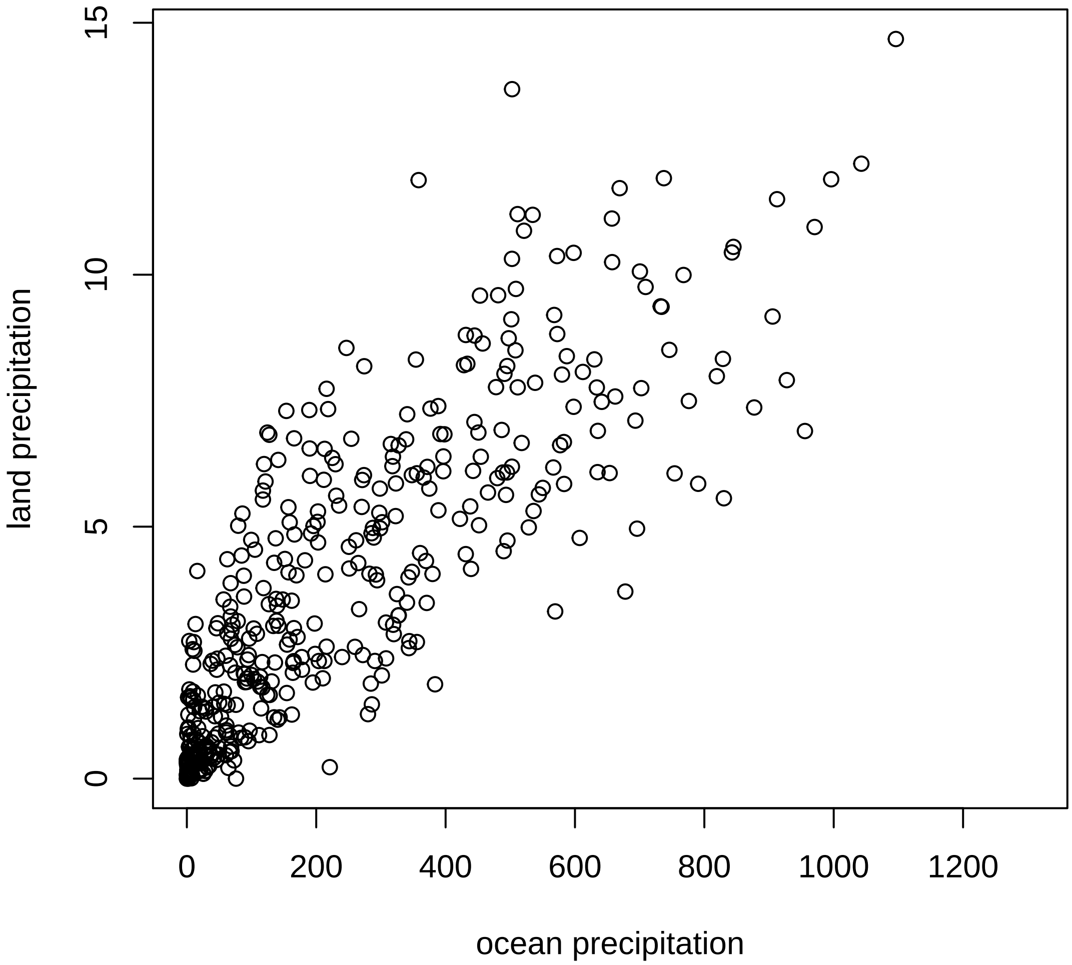

Figure E2. Monthly precipitation measured over land via land gauges versus the precipitation measured via remote sensing over the ocean.

Appendix F: Chlorophyll-a images in 2016

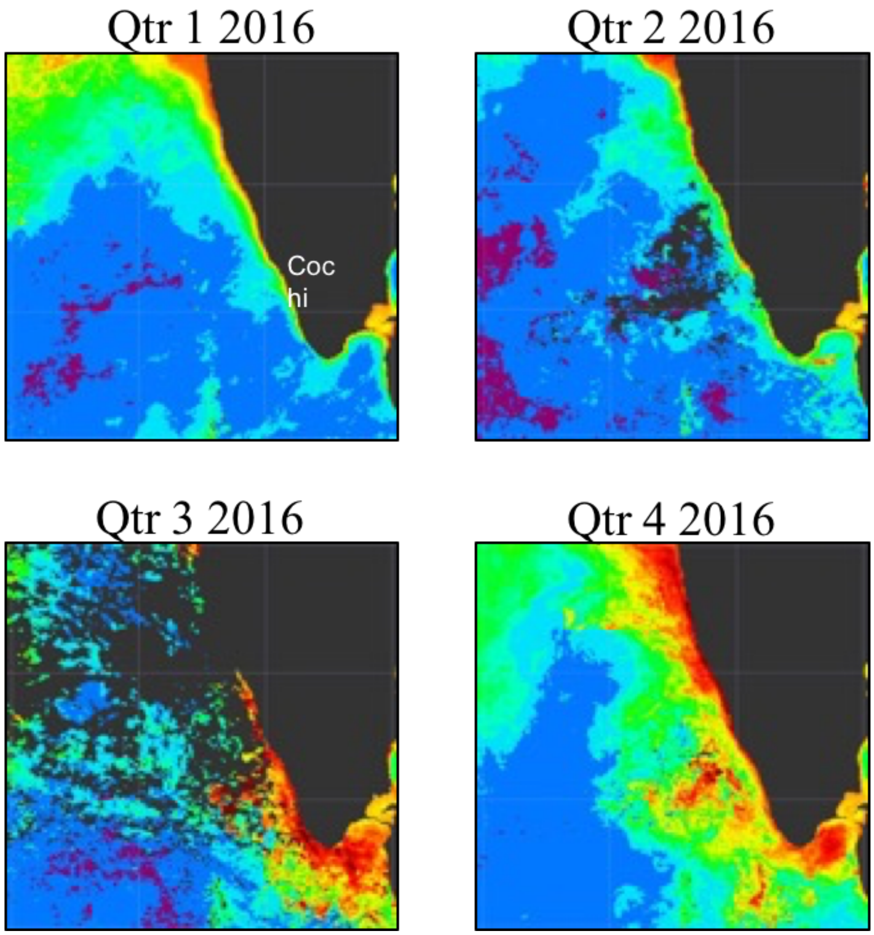

Figure 1

Appendix G: Influential years tests

Validation of the landings base model

This describes a variety of cross-validations used to select the base model for landing. The base model is the model with no environmental covariates only prior landings as covariates.

Three types of base models were fit. The first two were GAM and linear models with Jul-Sep and Oct-Mar in the prior season only or prior season and two seasons prior as covariates. c is the response variable: landings during the two seasons, either Jul-Sep or Oct-Mar.

$$
\begin{aligned}
\text{GAM t-1} &: X_t = \alpha + s(c_{t-1}) + e_t \\
\text{Linear t-1} &: X_t = \alpha + \beta c_{t-1} + e_t \\
\text{GAM t-1, t-2} &: X_t = \alpha + s(c_{t-1}) + s(d_{t-2}) + e_t \\
\text{Linear t-1, t-2} &: X_t = \alpha + \beta c_{t-1} + d_{t-2} + e_t
\end{aligned}
$$

where c_{t-1} was either S_{t-1} (Jul-Sep landings in prior season) or N_{t-1} (Oct-Mar landings in prior season) and d_{t-2} was the same but 2 seasons prior.

These types of models do not allow the model parameters (the intercept α and effect parameter β) to vary in time. The second type of models were dynamic linear models (DLMs). DLMs allow the parameters to evolve in time. Two types of DLMs were used, an intercept only model where the intercept α evolves and a linear model where the effect parameter β is allowed to evolve:

$$
\begin{aligned}
\text{DLM intercept only} &: X_t = \alpha_t + e_t \\
\text{DLM intercept and slope} &: X_t = \alpha_t + \beta_t t + e_t \\
\text{DLM intercept and effect} &: X_t = \alpha_t + \beta_t c_{t-1} + e_t
\end{aligned}
$$

In addition to the GAM, linear and DLM models, three null models were included in the tested model sets:

$$
\begin{aligned}
\text{intercept only} &: X_t = \alpha + e_t \\
\text{intercept and prior catch} &: X_t = \alpha_t + X_{t-1} + e_t \\
\text{prior catch only} &: X_t = X_{t-1} + e_t
\end{aligned}
$$

The 'intercept only' is a flat level model. The 'prior catch only' simply uses the prior value of the time series (in this case landings) as the prediction and is a standard null model for prediction. The 'intercept and prior catch' combines these two null models.

The models were fit to the 1956-2015 landings (full data) and 1984-2015 (data that overlap the environmental covariates).

The model performance was measured by AIC, AICc and LOO prediction. The LOO prediction error is the data point t minus the predited value for data point t. This is repeated for all data points t. The influence of single data points to on model performance was evaluated by leaving out one data point, fitting to the remaining data and computing the model performance (via AIC, AICc or LOO prediction error).

Results: Jul-Sep landings

The Figure 1D and 2D show the ΔAIC for the models: GAM, linear, and DLM. The figure shows that for the 1984-2015 data with any year left out, the set of models that has the lowest AIC was always the GAM or linear model with Oct-Mar in the prior season. There were cases where deleting a year removed one of these two from the 'best' category, but they were still in the 'competitive' category with a ΔAIC less than 2. With the full data set, 1956 to 2015, the best models were GAM with Jul-Sept in the prior season and two seasons prior. The DLM with Oct-Mar in prior season was also supported; this model allowed the effect of Oct-Mar catch to vary over time.

AIC gives us a measure of how well the models fit the data, with a penalty for the number of estimated parameters. We look at the one-step-ahead predictive performance (Figure 3D), we see that all the GAM, linear and DLM models have a hard time adjusting to shifts in the data (e.g. after 1998). The null models can adjust quickly but has large errors when there are rapid changes. The root mean squared error (which penalizes large predictive errors) is lowest for the models with Oct-Mar in the prior season for the more recent data (Figure 4D) and for the full data, many models have similar predictive performance.

It should be noted that none of the models has a particularly high adjusted R^2. The values are generally less than 0.3. The Jul-Sep landings tend to be highly variable and not related to the catch in prior years. Jul-Sep is during the monsoon during which fishing is not always possible due to sea-state and there is a 6-week fishing ban during this time.

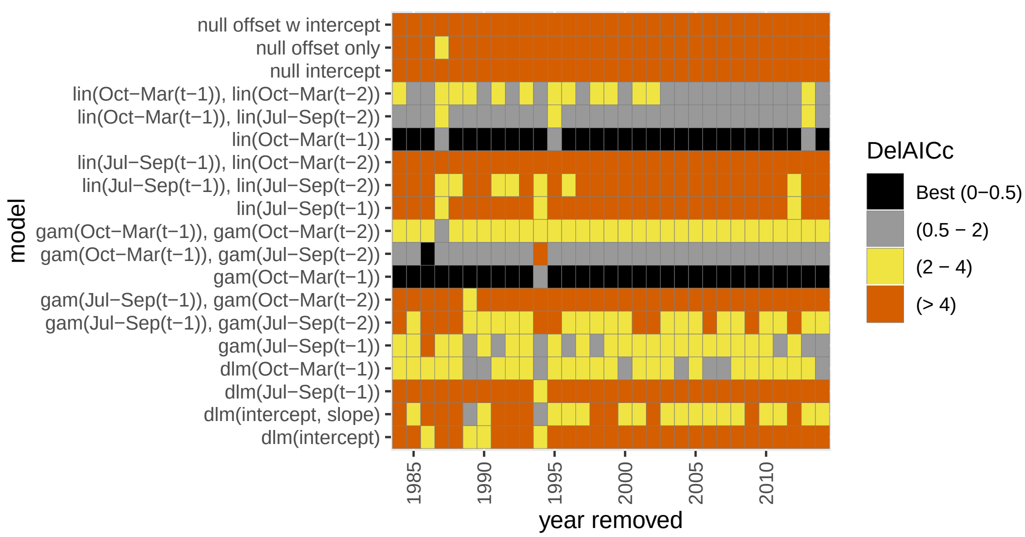

Figure 1D. ΔAICc for the Jul-Sep landings base models with one year deleted using only the landings data that overlap with the environmental data 1984-2015. See Figure 1D for an explantion of the figure.

Figure 2D. ΔAIC for the Jul-Sep landings base models with one year deleted using the full landings data set 1956-2015. ΔAIC is AIC of the model minus the AIC of the best (lowest AIC model) in the set. Black models were the best models in the set and within 0.5 AIC of each other. Grey are models within 2 of the best model, thus competitive to the best models. Deleted year is shown on the x-axis; the farthest right column has no year removed.

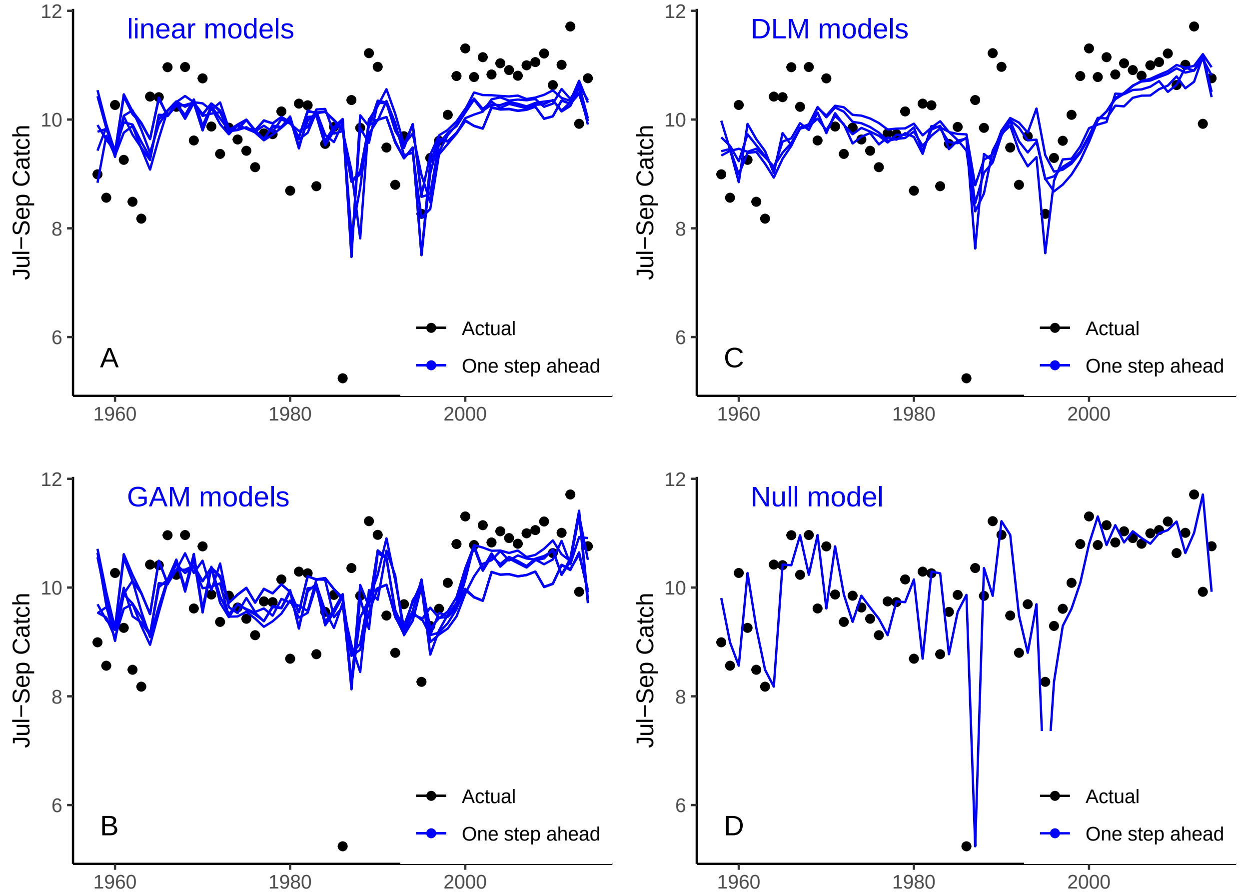

Figure 3D. Leave one out (LOO) one step ahead prediction errors for the linear, GAM, and DLM models of Jul-Sep landings. The data point at year t on the x-axis is predicted from the data up to year t-1.

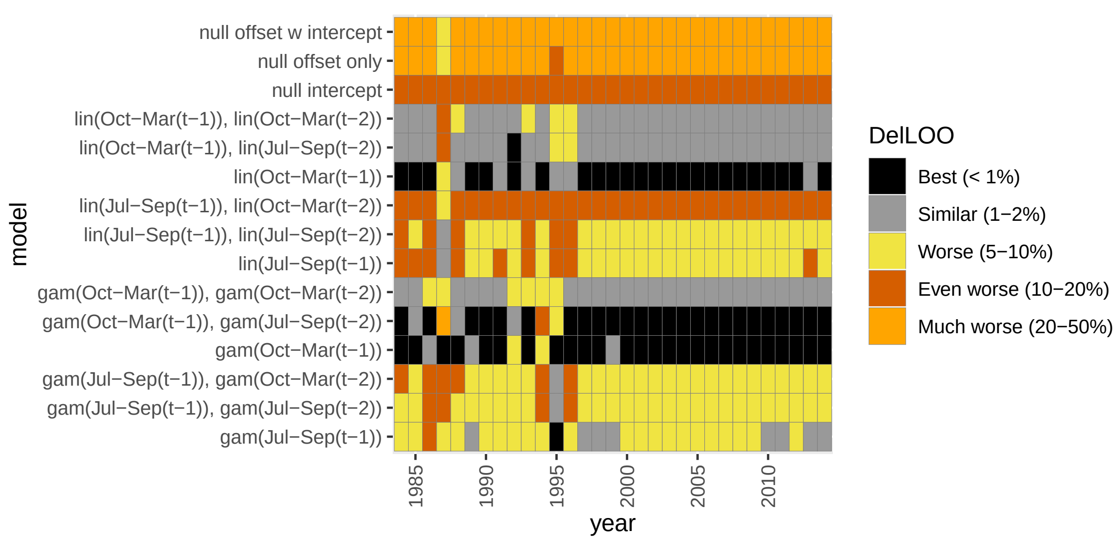

Figure 4D. Leave-one-out predictive performance (leave out a year, fit, predict that year) for the Jul-Sep landings base models. The performance (DelLOO) is the RSME (root mean square error).

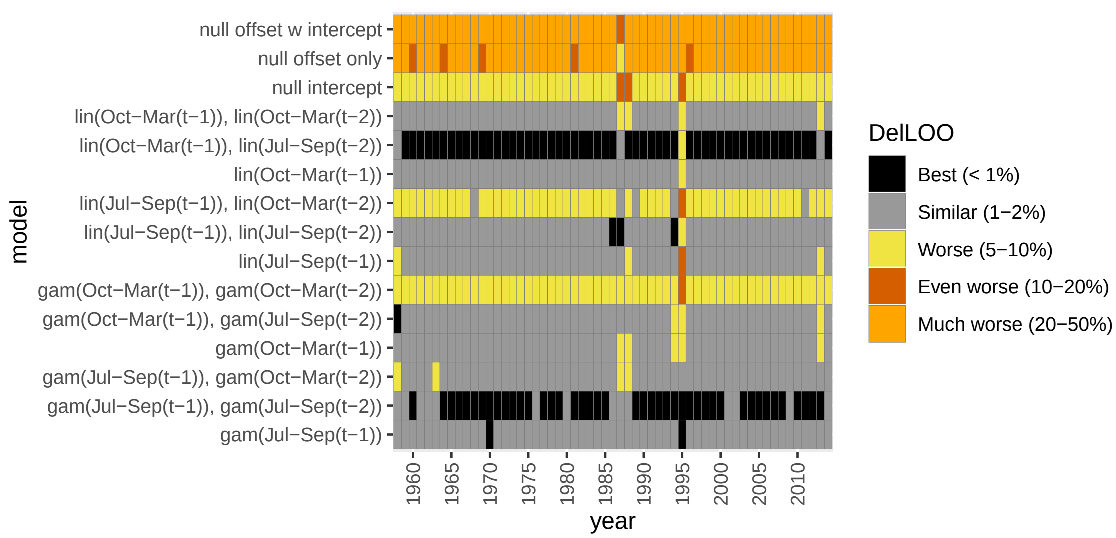

Figure 5D. Leave-one-out predictive performance (leave out a year, fit, predict that year) for the Jul-Sep landings base models. The performance (DelLOO) is the RSME (root mean square error).

Validation of the Oct-Mar landings base models

The Figure 6D shows that for Oct-Mar landings with the 1984 to 2014 data, the best model was always GAM with Oct-Mar in the prior season and Jul-Sep landings two seasons prior. For the full data set, the simpler GAM model with only Oct-Mar landings in the prior season was best. For the one step ahead predictions, simpler models had the lower prediction errors: GAM with Oct-Mar in the prior season for the recent data and linear with Oct-Mar in the prior season for the full data set.

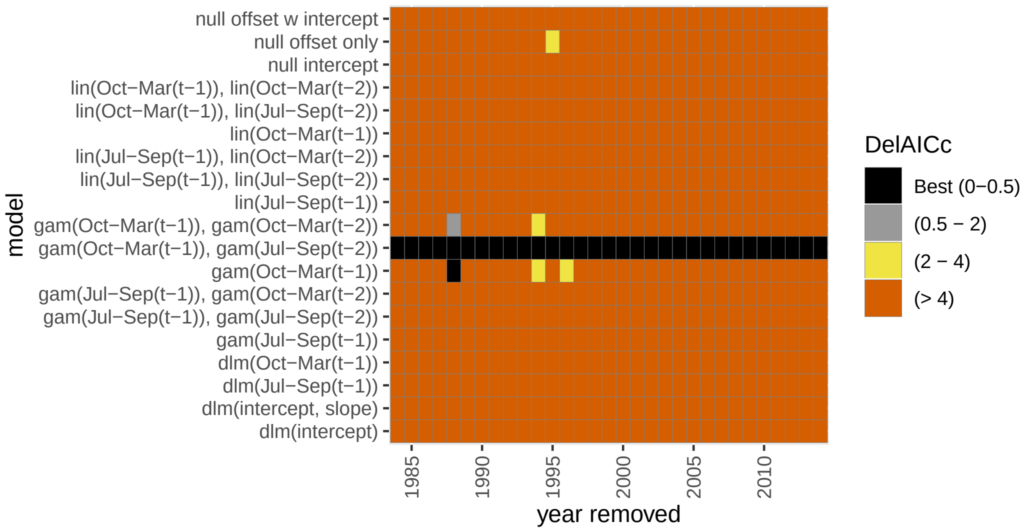

Figure 6D. ΔAICc for the Oct-Mar landings base models with one year deleted using only the landings data that overlap with the environmental data 1984-2015. See Figure 1D for an explantion of the figure.

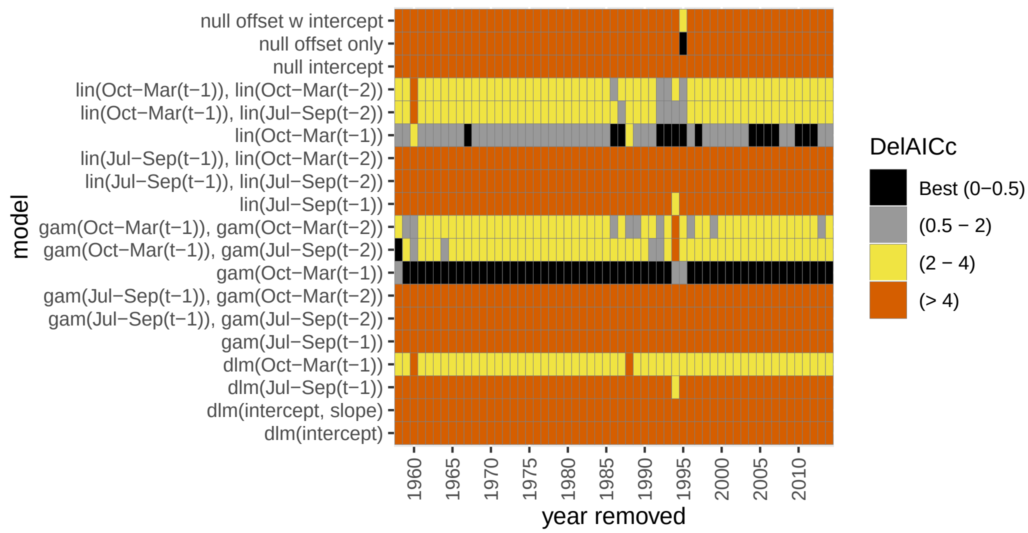

Figure 7D. ΔAIC for the Oct-Mar landings base models with one year deleted using the full landings data set 1956-2015. ΔAIC is AIC of the model minus the AIC of the best (lowest AIC model) in the set. Black models were the best models in the set and within 0.5 AIC of each other. Grey are models within 2 of the best model, thus competitive to the best models. Deleted year is shown on the x-axis; the farthest right column has no year removed.

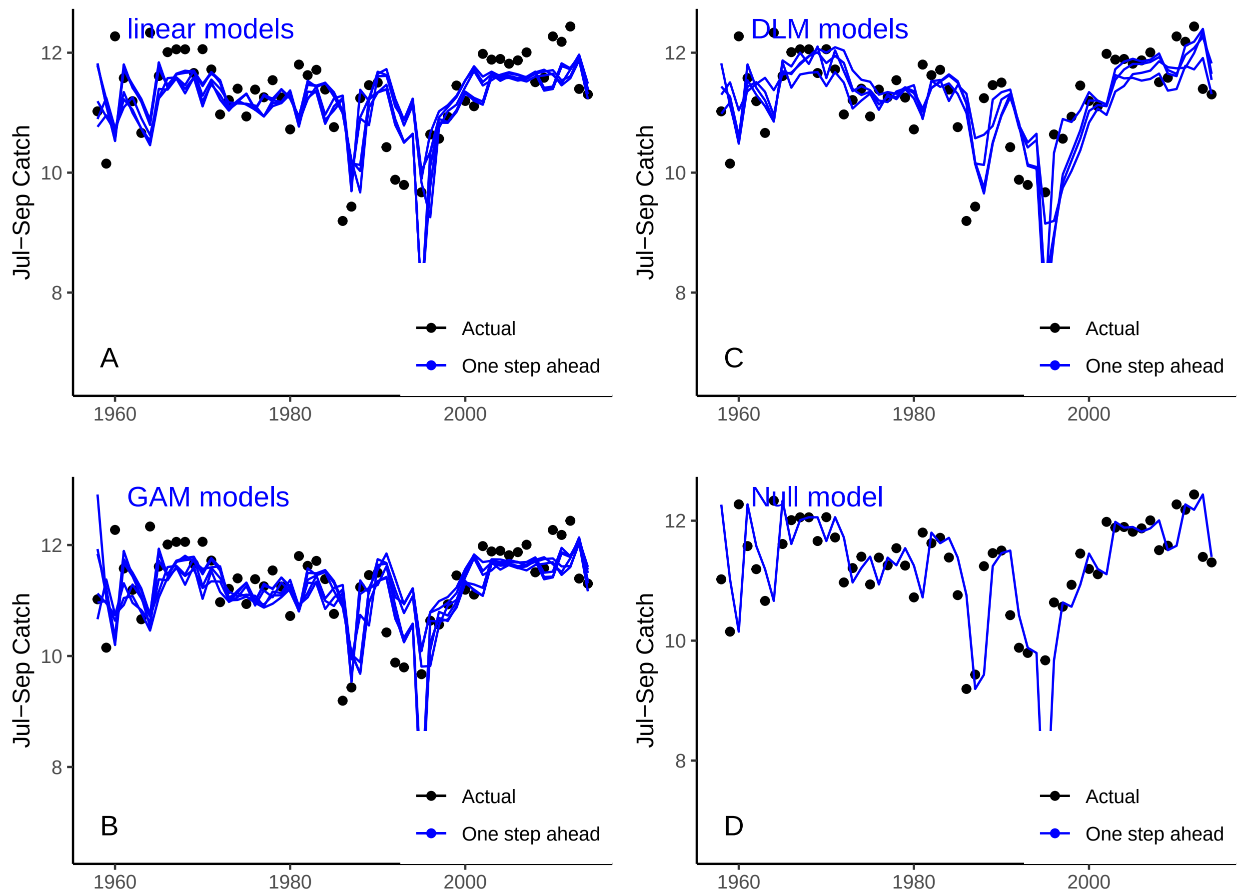

Figure 8D. Leave one out (LOO) one step ahead prediction errors for the linear, GAM, and DLM models of Oct-Mar landings. The data point at year t on the x-axis is predicted from the data up to year t-1.

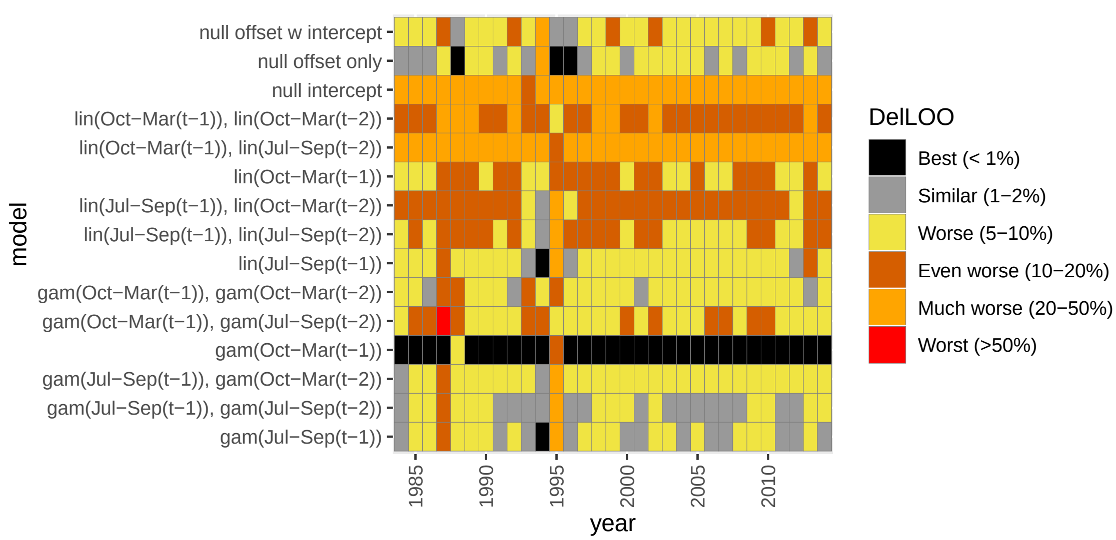

Figure 9D. Leave-one-out predictive performance (leave out a year, fit, predict that year) for the Oct-Mar landings base models. The performance (DelLOO) is the RSME (root mean square error).

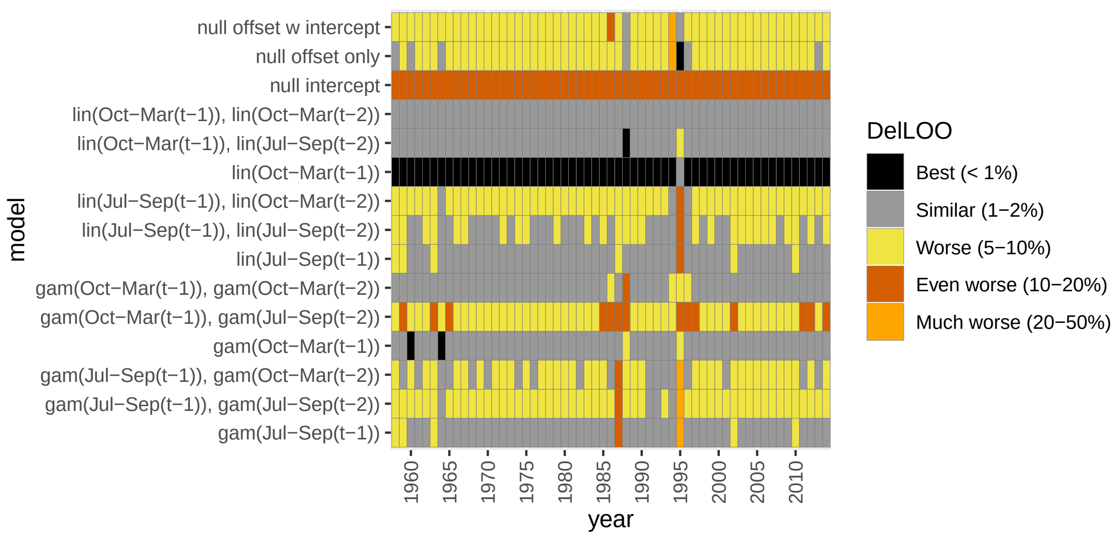

Figure 10D. Leave-one-out predictive performance (leave out a year, fit, predict that year) for the Oct-Mar landings base models. The performance (DelLOO) is the RSME (root mean square error).