Supplemental Information: Full model tests and diagnostics

```
## File .here already exists in /Users/eli.holmes/Documents/GitHub/SardinePaper2
```

Table S1. 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	LOOCV
Naive Model 1984-2015 data							
$ln(S_t) = ln(S_{t-1}) + \epsilon_t$	32	1				122.85	1.599
Time dependency test A 1984-2015 data							
1. $ln(S_t) = \alpha + ln(S_{t-1}) + \epsilon_t$	31	0.992	-29			124.83	1.65
2. $ln(S_t) = \alpha + \beta ln(S_{t-1}) + \epsilon_t$	30	0.814	10.3	15.14	0.001	114.14	1.43
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	1.414
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	1.346
2. $ln(S_t) = \alpha + \beta ln(N_{t-1}) + \epsilon_t$	30	0.794	22.2	4.06	0.053	109.59	1.308
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	1.346
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	1.346
2. $ln(S_t) = \alpha + \beta ln(N_{t-1}) + \epsilon_t$	30	0.794	22.2	4.08	0.053	109.59	1.308
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	1.37
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	1.331

Table S2. Model selection tests of time-dependency the catch during spawning months using non-linear or time-varying linear responses instead of time-constant linear responses as in Table S1. See Table S1 for an explanation of the parameters and model set-up.

Model	Residual df	MASE	Adj. R2	F	p value	AIC	LOOCV
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	1.43
2. $ln(S_t) = \alpha + s(ln(S_{t-1})) + \epsilon_t$	28.2	0.798	19.6	2.74	0.089	111.79	1.371
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	1.382
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	1.308
2. $ln(S_t) = \alpha + s(ln(N_{t-1})) + \epsilon_t$	28.6	0.761	24.4	1.26	0.287	109.52	1.299
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	1.342
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	1.299
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	1.273
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	1.295
Time varying test D 1984-2015 data							
1. $ln(S_t) = \alpha_t + \epsilon_t$	29	0.658				114.45	1.373
2. $ln(S_t) = \alpha_t + \beta_t t + \epsilon_t$	27	0.85				114.24	1.354
3a. $ln(S_t) = \alpha + \beta_t ln(S_{t-1}) + \epsilon_t$	28	0.723				115.66	1.49
3b. $ln(S_t) = \alpha + \beta_t ln(N_{t-1}) + \epsilon_t$	28	0.794				111.59	1.337

Table S3. Table S2 with 1956-1983 data instead of 1984 to 2015 data. See Table S1 for an explanation of the parameters and model set-up.

Model	Residual df	MASE	Adj. R2	F	p value	AIC	LOOCV
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	0.821
2. $ln(S_t) = \alpha + s(ln(S_{t-1})) + \epsilon_t$	22.1	0.614	-0.2	0.78	0.464	65.71	0.844
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	1.053
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	0.821
2. $ln(S_t) = \alpha + s(ln(N_{t-1})) + \epsilon_t$	21.6	0.584	8.2	2.24	0.127	63.8	0.783
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	0.785
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	0.8
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	0.829
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	0.967

Table S4. 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	LOOCV
Naive Model 1984-2014 data							
$ln(N_t) = ln(N_{t-1}) + \epsilon_t$	31	1				90.87	1.015
Time dependency test A 1984-2014 data							
1. $ln(N_t) = \alpha + ln(S_{t-1}) + \epsilon$	30	1.363	-20.3			107.36	1.324
2. $ln(N_t) = \alpha + \beta ln(S_{t-1}) + \epsilon_t$	29	1.018	26.2	19.99	0	93.17	1.035
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	1.062
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	1.048
2. $ln(N_t) = \alpha + \beta ln(N_{t-1}) + \epsilon_t$	29	0.978	37	6.63	0.016	88.28	1.062
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	1.148
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	1.062
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	1.093
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	1.208

Table S5. Model selection tests of time-dependency the N_t model using non-linear or time-varying linear responses instead of time-constant linear responses as in Table S4 See Table S4 for an explanation of the parameters and model set-up.

Model	Residual df	MASE	Adj. R2	F	p value	AIC	LOOCV
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	1.035
2. $ln(N_t) = \alpha + s(ln(S_{t-1})) + \epsilon_t$	27.3	0.992	30.2	1.83	0.185	92.61	1.016
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	1.012
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	1.062
2. $ln(N_t) = \alpha + s(ln(N_{t-1})) + \epsilon_t$	27.6	0.874	45.3	3.88	0.047	84.75	0.966
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	1.02
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	0.966
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	1.081
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	1.062
Time varying test D 1984-2014 data							
1. $ln(N_t) = \alpha_t + \epsilon_t$	28	0.443				92.98	1.043
2. $ln(N_t) = \alpha_t + \beta_t t + \epsilon_t$	26	0.455				96.96	1.045
3a. $ln(N_t) = \alpha + \beta_t ln(S_{t-1}) + \epsilon_t$	27	0.789				93.96	0.923
3b. $ln(N_t) = \alpha + \beta_t ln(N_{t-1}) + \epsilon_t$	27	0.978				90.28	1.031

Table S6. Table S5 with 1956-1983 data instead of 1984 to 2014 data. See Table S4 for an explanation of the parameters and model set-up.

Model	Residual df	MASE	Adj. R2	F	p value	AIC	LOOCV
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	0.574
2. $ln(N_t) = \alpha + s(ln(S_{t-1})) + \epsilon_t$	22.1	0.534	16.2	3.53	0.052	41.11	0.542
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	0.615
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	0.575
2. $ln(N_t) = \alpha + s(ln(N_{t-1})) + \epsilon_t$	21.6	0.507	29.1	5.69	0.009	37.12	0.468
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	0.506
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	0.468
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	0.498
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	0.538

Tests for environmental variables as covariates

Table S7. 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 $+ \beta V_t$	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 $+ \beta V_t$	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 S8. 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 minus nearshore SST, the average nearshore SST along the Kerala coast (boxes 2-5), and the Bakun upwelling index based on wind stress. See Table S7 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 = Mar-May r-SST (S5)						
2a. $ln(S_t)$ = M + βV_t	27.6	0.772	22.9	0.45	0.509	111.03
3a. $ln(S_t)$ = M + $s(V_t)$	25.6	0.748	26.5	1.37	0.271	110.68
2b. $ln(S_t)$ = M + βV_{t-1}	27.6	0.751	23.3	0.57	0.454	110.88
3b. $ln(S_t)$ = M + $s(V_{t-1})$	25.7	0.722	25.8	1.19	0.32	110.91
V_t = Oct-Dec ns-SST (L1)						
2. $ln(S_t)$ = M + βV_{t-1}	27.6	0.759	21.8	0	0.952	111.5
3. $ln(S_t)$ = M + $s(V_{t-1})$	26.1	0.768	21.8	0.66	0.482	112.32
V_t = Jun-Sep UPW (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 = Jun-Sep ns-SST (L2)						
⇒ 2a. $ln(S_t)$ = M + βV_t	27.6	0.745	33.3	5.53	0.027	106.4
⇒ 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 = Jun-Sep Bakun-UPW (L2)						
2a. $ln(S_t)$ = M + βV_t	27.6	0.776	28.4	3.62	0.069	108.66
⇒ 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 S9. 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 nearshore 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 S7 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 ave ns-SST (A1)						
$\Rightarrow$ 2. $ln(S_t)$ = M + βV_t	27.6	0.723	33.2	5.52	0.027	106.43
$\Rightarrow$ 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 = Sep-Nov 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 S10. 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. $ln(N_t) =$ M + βV_t	23.8	0.755	56.7	1.03	0.318	80.23
3a. $ln(N_t) =$ M + $s(V_t)$	22.3	0.75	55.3	0.19	0.767	82.02
2b. $ln(N_t) =$ M + βV_{t-1}	23.8	0.744	54.9	0	1	81.5
3b. $ln(N_t) =$ M + $s(V_{t-1})$	22.3	0.701	56.4	1.32	0.28	81.18
V_t = Apr-May Precipitation (S2)						
2a. $ln(N_t) =$ M + βV_t	23.8	0.742	55.1	0.11	0.735	81.34
3a. $ln(N_t) =$ M + $s(V_t)$	21.7	0.73	53.7	0.36	0.707	83.39
2b. $ln(N_t) =$ M + βV_{t-1}	23.8	0.723	56.2	0.74	0.397	80.6
3b. $ln(N_t) =$ M + $s(V_{t-1})$	22	0.692	55.6	0.5	0.587	81.87

Table S11. 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 minus nearshore SST, the average nearshore SST along the Kerala coast (boxes 2-5), and the Bakun upwelling index based on wind stress. See Table S10 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 = Mar-May r-SST (S5)						
2a. $ln(N_t)$ = M + βV_t	23.8	0.707	59.1	2.95	0.101	78.46
⇒ 3a. $ln(N_t)$ = M + $s(V_t)$	22	0.686	63.7	2.39	0.119	75.68
2b. $ln(N_t)$ = M + βV_{t-1}	23.8	0.765	58.4	2.13	0.159	78.96
3b. $ln(N_t)$ = M + $s(V_{t-1})$	21.9	0.748	57.6	0.45	0.633	80.52
V_t = Oct-Dec ns-SST (L1)						
2. $ln(N_t)$ = M + βV_{t-1}	23.8	0.748	54.9	0	1	81.5
3. $ln(N_t)$ = M + $s(V_{t-1})$	22.5	0.736	56	1.13	0.318	81.37
V_t = Jun-Sep UPW (L2)						
⇒ 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 = Jun-Sep ns-SST (L2)						
⇒ 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 = 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 S12. 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 nearshore 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 S10 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 ave ns-SST (A1)						
$\Rightarrow$ 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	0	1	81.46
3b. $ln(N_t)$ = M + $s(V_{t-1})$	23	0.748	55.5	0.99	0.313	81.46
V_t = Sep-Nov 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	0	1	81.52
$\Rightarrow$ 3b. $ln(N_t)$ = M + $s(V_{t-1})$	21.1	0.678	67.5	4.34	0.018	72.69

Table S13. 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 and the base model without S_{t-2}. 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. $ln(N_t) = \alpha + s(ln(N_{t-1})) + \epsilon_t$	27.6	0.874	45.3			84.75
V_t = Jun-Jul Precipitation (S1)						
2a. $ln(N_t) = \mathrm{M} + \beta V_t$	26.6	0.873	43.3	0	1	86.74
3a. $ln(N_t) = \mathrm{M} + s(V_t)$	24.9	0.849	45.3	1.29	0.288	86.66
2b. $ln(N_t) = \mathrm{M} + \beta V_{t-1}$	26.6	0.872	44	0.35	0.556	86.33
3b. $ln(N_t) = \mathrm{M} + s(V_{t-1})$	24.9	0.867	43.4	0.53	0.566	87.69
V_t = Apr-May Precipitation (S2)						
2a. $ln(N_t) = \mathrm{M} + \beta V_t$	26.6	0.872	43.5	0.1	0.742	86.59
3a. $ln(N_t) = \mathrm{M} + s(V_t)$	24.5	0.863	41.3	0.24	0.802	89.08
2b. $ln(N_t) = \mathrm{M} + \beta V_{t-1}$	26.6	0.86	44	0.35	0.553	86.32
3b. $ln(N_t) = \mathrm{M} + s(V_{t-1})$	24.8	0.833	44.7	0.85	0.431	87.01

Table S14. 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 and base model without S_{t-2}. The hypothesis tested (Table 1) is noted in parentheses. Three upwelling indices were tested. The nearshore-offshore temperature diferential (UPW), which is the offshore minus nearshore SST, the average nearshore SST along the Kerala coast (boxes 2-5), and the Bakun upwelling index based on wind stress. See Table S10 for an explanation of the models.

Model	Residual df	MASE	Adj. R2	F	p value	AIC
base model (M) 1984-2014 data						
1. $ln(N_t) = \alpha + s(ln(N_{t-1})) + \epsilon_t$	27.6	0.874	45.3			84.75
V_t = Mar-May r-SST (S5)						
2a. $ln(N_t) = \text{M} + \beta V_t$	26.6	0.854	46.1	1.44	0.242	85.16
3a. $ln(N_t) = \text{M} + s(V_t)$	24.7	0.841	46.6	0.82	0.447	85.97
2b. $ln(N_t) = \text{M} + \beta V_{t-1}$	26.6	0.81	49.7	3.6	0.07	82.99
3b. $ln(N_t) = \text{M} + s(V_{t-1})$	24.6	0.777	50.9	1.02	0.379	83.45
V_t = Oct-Dec ns-SST (L1)						
2. $ln(N_t) = \text{M} + \beta V_{t-1}$	26.6	0.828	45.6	1.12	0.299	85.47
3. $ln(N_t) = \text{M} + s(V_{t-1})$	25.2	0.839	44.5	0.28	0.684	86.89
V_t = Jun-Sep UPW (L2)						
$\Rightarrow$ 2a. $ln(N_t) = \text{M} + \beta V_t$	26.6	0.863	54	6.63	0.017	80.25
3a. $ln(N_t) = \text{M} + s(V_t)$	24.1	0.834	56	1.22	0.319	80.36
2b. $ln(N_t) = \text{M} + \beta V_{t-1}$	26.6	0.876	43.3	0	1	86.74
3b. $ln(N_t) = \text{M} + s(V_{t-1})$	24	0.804	46.9	1.46	0.253	86.25
V_t = Jun-Sep ns-SST (L2)						
$\Rightarrow$ 2a. $ln(N_t) = \text{M} + \beta V_t$	26.6	0.857	51.6	4.71	0.04	81.79
3a. $ln(N_t) = \text{M} + s(V_t)$	24.6	0.86	51.3	0.63	0.545	83.19
2b. $ln(N_t) = \text{M} + \beta V_{t-1}$	26.6	0.819	45.2	0.92	0.345	85.67
3b. $ln(N_t) = \text{M} + s(V_{t-1})$	24.5	0.801	43.6	0.36	0.711	87.78
V_t = Jun-Sep Bakun-UPW (L2)						
2a. $ln(N_t) = \text{M} + \beta V_t$	26.6	0.825	46.2	1.5	0.231	85.1
3a. $ln(N_t) = \text{M} + s(V_t)$	24.5	0.746	47	0.89	0.428	85.85
2b. $ln(N_t) = \text{M} + \beta V_{t-1}$	26.6	0.825	44.2	0.44	0.507	86.22
3b. $ln(N_t) = \text{M} + s(V_{t-1})$	24.7	0.807	44.7	0.83	0.446	87.07

Table S15. 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 and base model without S_{t-2}. 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 nearshore 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 S10 for an explanation of the models.

Model	Residual df	MASE	Adj. R2	F	p value	AIC
base model (M) 1984-2014 data						
1. $ln(N_t) = \alpha + s(ln(N_{t-1})) + \epsilon_t$	27.6	0.874	45.3			84.75
V_t = 2.5-year ave ns-SST (A1)						
⇒ 2. $ln(N_t)$ = M + βV_t	26.6	0.736	55	8.08	0.009	79.58
⇒ 3. $ln(N_t)$ = M + $s(V_t)$	25.3	0.664	60.3	3.48	0.064	76.34
V_t = ONI (A2)						
2a. $ln(N_t)$ = M + βV_t	26.6	0.883	48.4	2.66	0.115	83.81
3a. $ln(N_t)$ = M + $s(V_t)$	25.5	0.877	47.9	0.37	0.567	84.69
2b. $ln(N_t)$ = M + βV_{t-1}	26.6	0.875	43.8	0.25	0.614	86.45
3b. $ln(N_t)$ = M + $s(V_{t-1})$	25.6	0.844	46.3	1.86	0.185	85.53
V_t = Sep-Nov DMI (A3)						
2a. $ln(N_t)$ = M + βV_t	26.6	0.904	48.1	2.52	0.126	83.99
3a. $ln(N_t)$ = M + $s(V_t)$	23.6	0.882	48	0.73	0.548	85.86
2b. $ln(N_t)$ = M + βV_{t-1}	26.6	0.855	43.9	0.26	0.606	86.38
3b. $ln(N_t)$ = M + $s(V_{t-1})$	23.7	0.853	43.4	0.68	0.571	88.43

Tests for Chlorophyll-a as a covariate

Table S16. Model selection tests of nearshore (0-80km boxes 2 to 5) chlorophyll-a (ns-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. Only CHL in Oct-Dec and Jan-Mar in the current season (which runs Jul-Jun) is not used since 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. 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 ns-CHL						
2. $ln(S_t) = M + \beta V_t$	13	0.515	25.2	0.85	0.377	19.05
3. $ln(S_t) = M + p(V_t)$	12	0.489	27.8	1.22	0.295	19.1
4. $ln(S_t) = M + p(V_t) + \beta V_{t-1}$	11	0.484	21.3	0.02	0.898	21.07
5. $ln(S_t) = M + p(V_t) + p(V_{t-1})$	10	0.479	13.8	0.04	0.851	23.01
V_t = Oct-Dec ns-CHL						
2. $ln(S_t) = M + \beta V_{t-1}$	13	0.517	20	0.08	0.787	20.2
3. $ln(S_t) = M + p(V_{t-1})$	12	0.447	27	2.25	0.16	19.29
V_t = Jan-Mar ns-CHL						
2. $ln(S_t) = M + \beta V_{t-1}$	13	0.515	19.6	0.01	0.934	20.28
3. $ln(S_t) = M + p(V_{t-1})$	12	0.519	13.9	0.13	0.721	22.09

Table S17. Model selection tests of nearshore (0-80km boxes 2 to 5) chlorophyll-a (ns-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-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. 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 ns-CHL						
2. $ln(N_t) = M + \beta V_t$	13	0.879	24	0.5	0.496	20.26
3. $ln(N_t) = M + p(V_t)$	12	0.869	18.6	0.13	0.727	22.07
2. $ln(N_t) = M + \beta V_{t-1}$	13	0.832	31.5	2.15	0.17	18.5
3. $ln(N_t) = M + p(V_{t-1})$	11.4	0.748	35.8	1.29	0.302	18.12
V_t = Oct-Dec ns-CHL						
2. $ln(N_t) = M + \beta V_t$	13	0.889	24	0.69	0.425	20.26
3. $ln(N_t) = M + p(V_t)$	12	0.721	35.1	3.57	0.088	18.2
2. $ln(N_t) = M + \beta V_{t-1}$	13	0.718	45.3	5.23	0.043	14.66
3. $ln(N_t) = M + p(V_{t-1})$	10.8	0.711	39.2	0.11	0.907	17.41
V_t = Jan-Mar ns-CHL						
2. $ln(N_t) = M + \beta V_t$	13	0.848	32.6	1.86	0.203	18.21
3. $ln(N_t) = M + p(V_t)$	12	0.846	27	0	0.967	20.21
2. $ln(N_t) = M + \beta V_{t-1}$	13	0.86	21.3	0.07	0.796	20.85
3. $ln(N_t) = M + p(V_{t-1})$	10.2	0.812	18.2	0.55	0.647	22.7

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:

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

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 LOOCV prediction. The LOOCV 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 S1 shows 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.

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 S2), 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 leave one out predictive error (the root mean squared error which penalizes large predictive errors) is lowest for the models with Oct-Mar in the prior season (Figure S3).

It should be noted that none of the Jul-Sep 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.

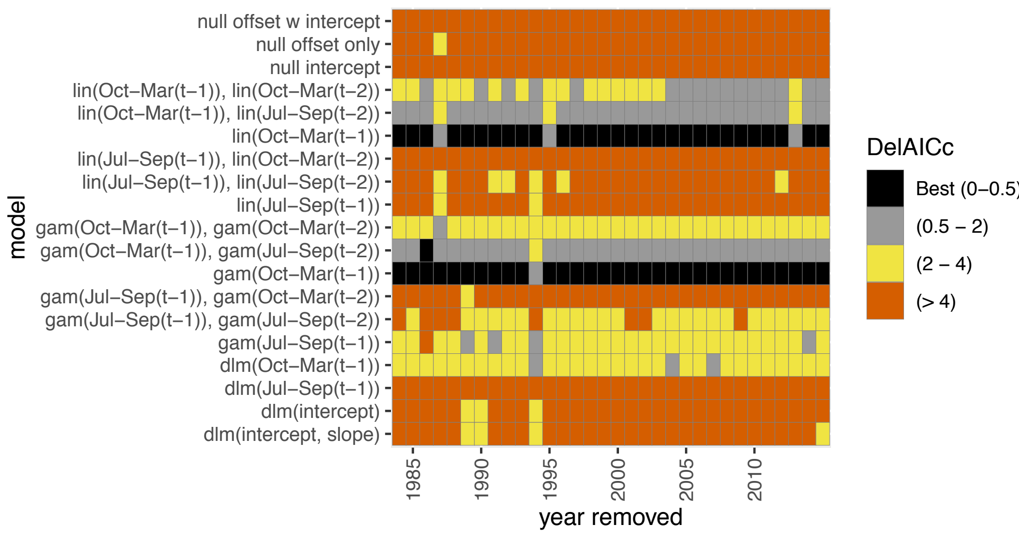

Figure S1. ΔAICc for the Jul-Sep landings base models with one year deleted using only the landings data that overlap with the environmental data 1984-2015.

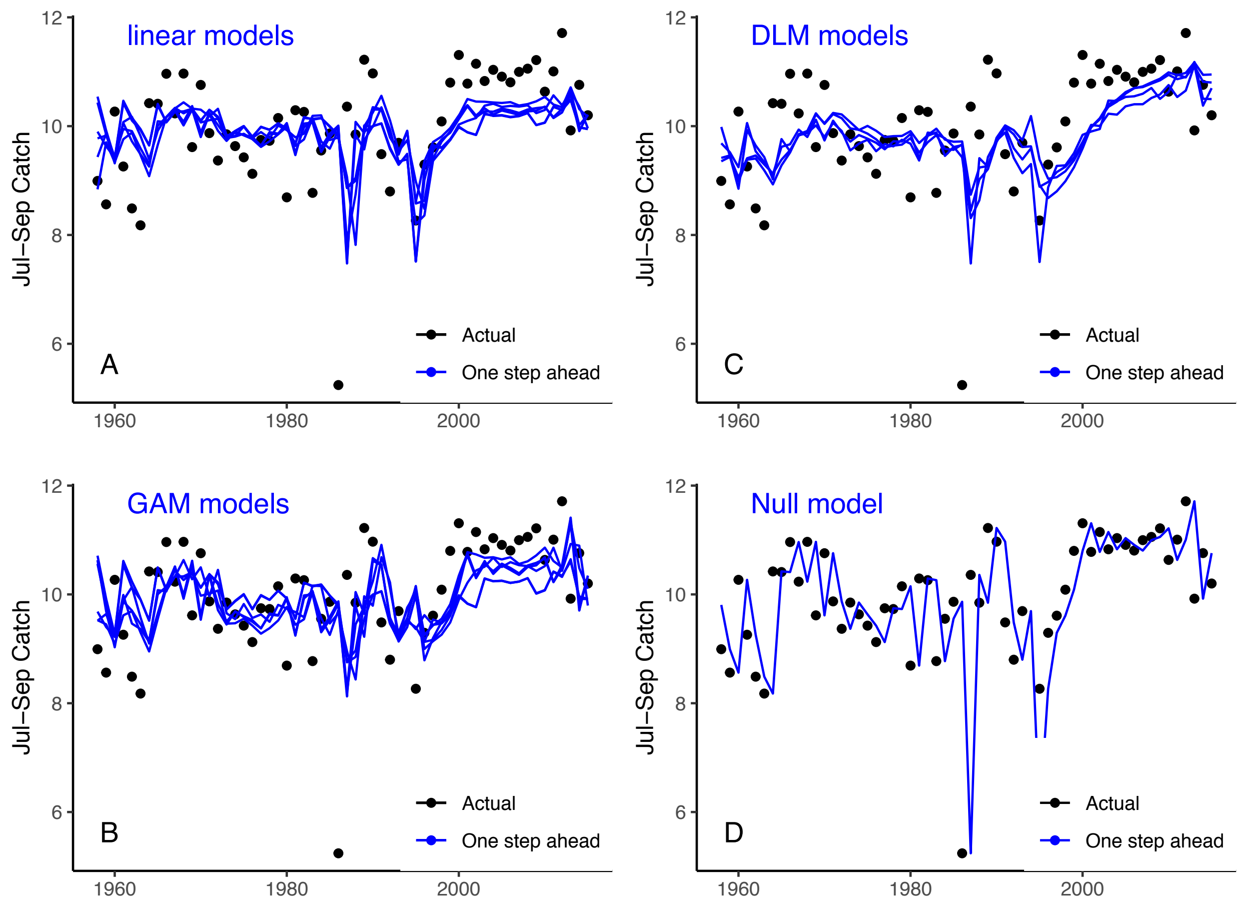

Figure S2. Leave one out (LOO) one step ahead predictions for the linear, GAM, and DLM models of Jul-Sep landings. The data point at year t on the x-axis is predicted from the data up to year t-1.

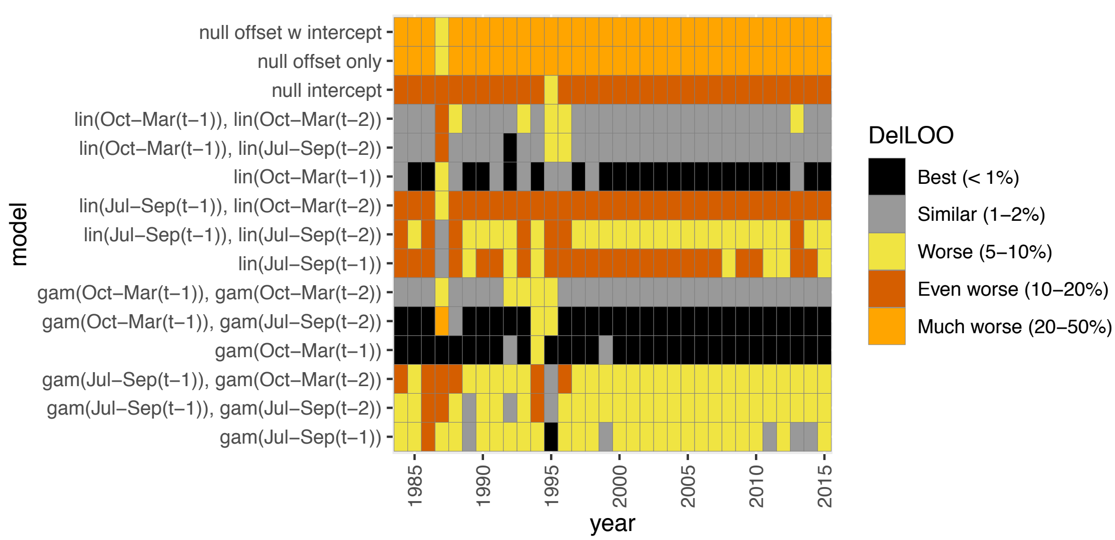

Figure S3. Leave-one-out predictive performance (leave out a year, fit, predict that year) for the Jul-Sep landings base models. The performance (DelLOO) is the RSME (root mean square error) between prediction and observed.

Validation of the Oct-Mar landings base models

Figure S4 shows that for Oct-Mar landings with the 1984 to 2015 data, the best model was always GAM with Oct-Mar in the prior season and Jul-Sep landings two seasons prior. For the one step ahead predictions, a simpler models had the lower prediction errors: GAM with Oct-Mar in the prior season as the only covariate (Figure S5).

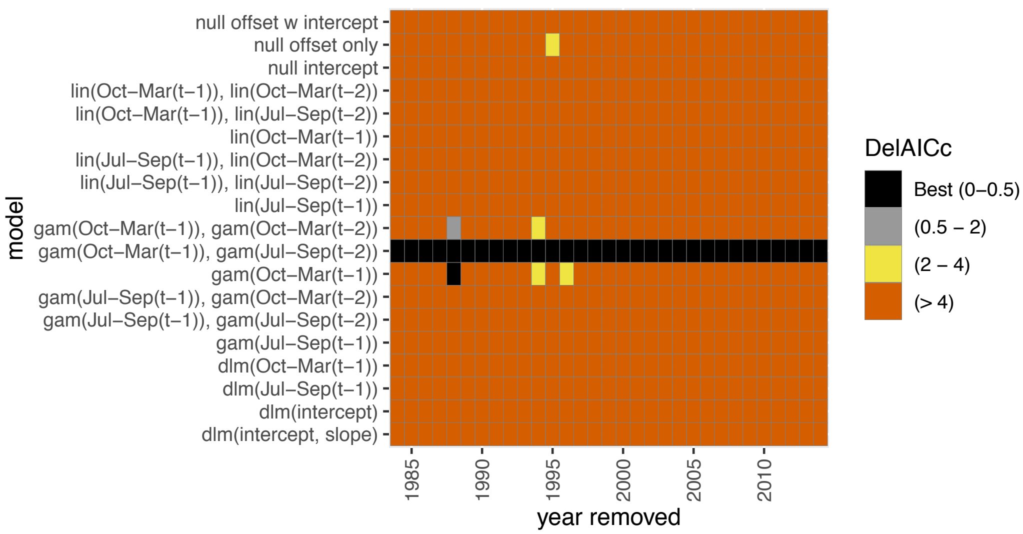

Figure S4. ΔAICc for the Oct-Mar landings base models with one year deleted using only the landings data that overlap with the environmental data 1984-2015. See Figure S1 for an explantion of the figure.

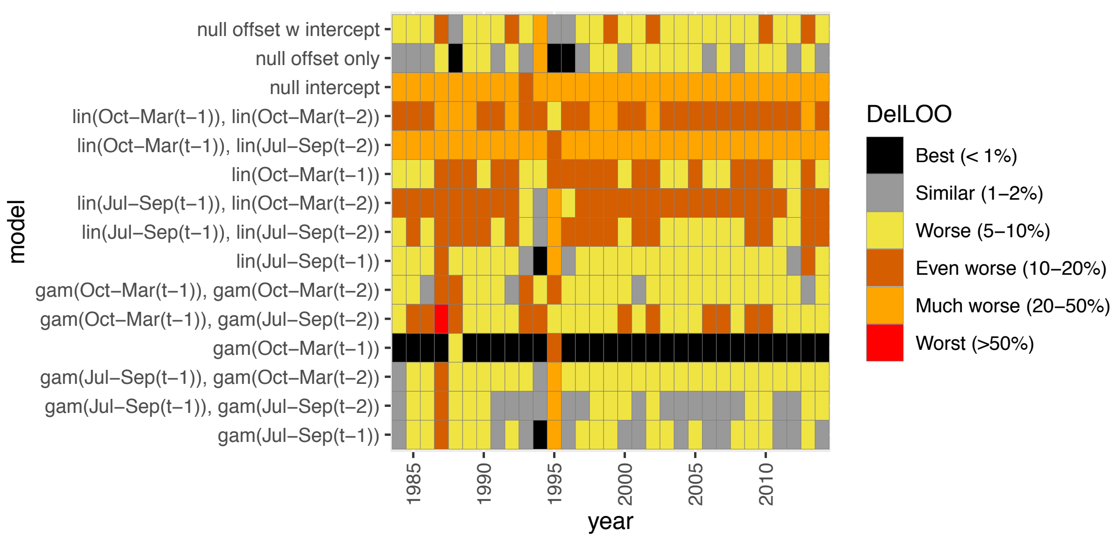

Figure S5. Leave-one-out predictive performance (leave out a year, fit, predict that year) for the Oct-Mar landings base models. The performance (DelLOO) is the RSME (root mean square error).

Comparison of land and oceanic rainfall measurements

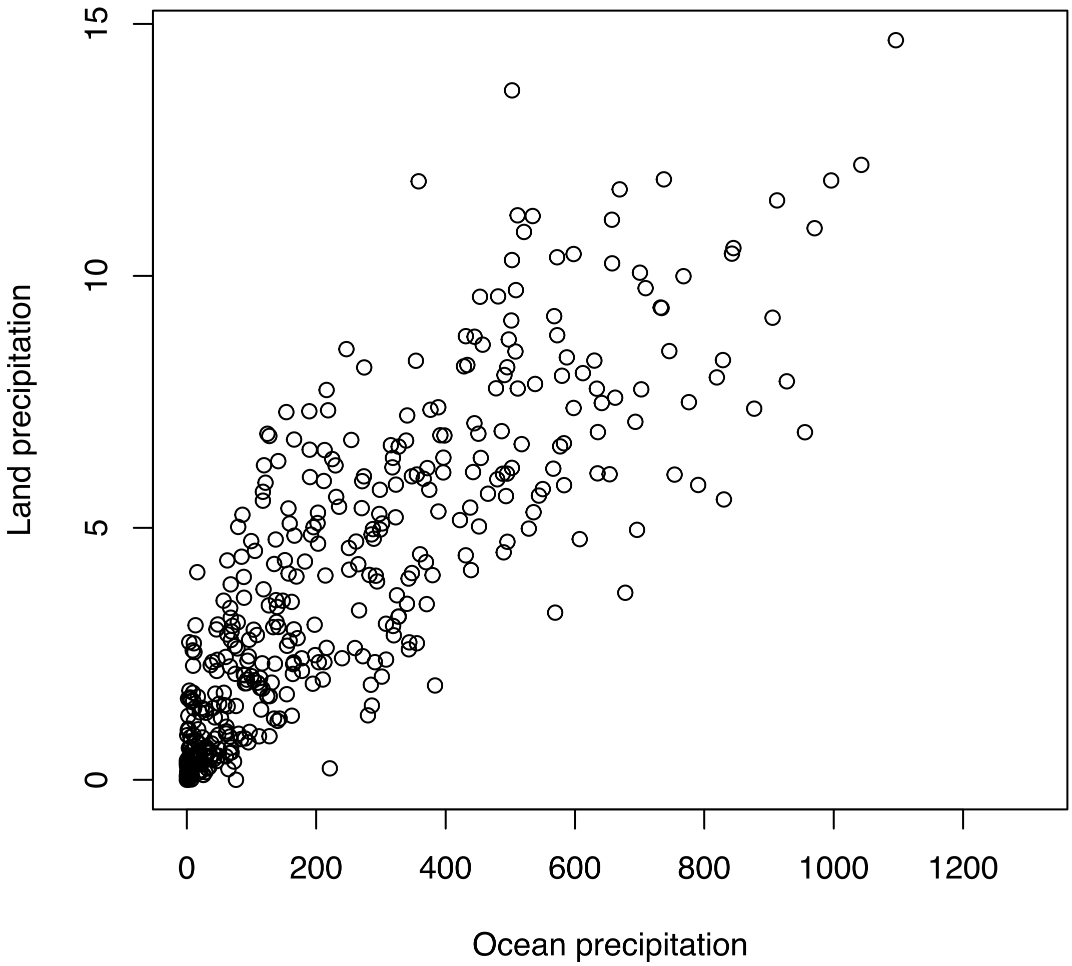

Figure S6. Monthly precipitation measured over land via land gauges versus the precipitation measured via remote sensing over the ocean.