Supplemental Information: Full model tests and diagnostics

Tests for prior season catch as covariate

Table S1. Model selection tests of time-dependency the log catch during Jul-Sep monsoon period using F-tests of nested linear models. S_t is the catch during Jul-Sep and W_t is the catch during the post-monsoon (Oct-Mar). S_{t-1} and W_{t-1} are the catch during the prior season. S_{t-2} and W_{t-2} are the same for two seasons prior. Test A uses Jul-Sep catch as the explanatory variable. Test B uses Oct-Mar catch 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	Adj. R^2	F	p value	AICc	LOOCV RMSE
Naive Model 1983-2015 data						
$ln(S_t) = ln(S_{t-1}) + \epsilon_t$	33				126.63	1.596
Time dependency test A 1983-2015 data						
1. $ln(S_t) = \alpha + ln(S_{t-1}) + \epsilon_t$	32	-29.1			128.9	1.646
2. $ln(S_t) = \alpha + \beta ln(S_{t-1}) + \epsilon_t$	31	9.8	15.28	0	118.46	1.43
3. $ln(S_t) = \alpha + \beta_1 ln(S_{t-1}) + \beta_2 ln(S_{t-2}) + \epsilon_t$	30	12.7	2.05	0.163	118.88	1.418
Time dependency test B 1983-2015 data						
1. $ln(S_t) = \alpha + ln(W_{t-1}) + \epsilon_t$	32	11			116.64	1.367
2. $ln(S_t) = \alpha + \beta ln(W_{t-1}) + \epsilon_t$	31	20	4.48	0.043	114.48	1.319
3. $ln(S_t) = \alpha + \beta_1 ln(W_{t-1}) + \beta_2 ln(W_{t-2}) + \epsilon_t$	30	17.4	0.04	0.846	117.04	1.357
Time dependency test C 1983-2015 data						
1. $ln(S_t) = \alpha + ln(W_{t-1}) + \epsilon_t$	32	11			116.64	1.367
2. $ln(S_t) = \alpha + \beta ln(W_{t-1}) + \epsilon_t$	31	20	4.49	0.043	114.48	1.319
3a. $ln(S_t) = \alpha + \beta_1 ln(W_{t-1}) + \beta_2 ln(S_{t-1}) + \epsilon_t$	30	17.6	0.09	0.768	116.98	1.383
3b. $ln(S_t) = \alpha + \beta_1 ln(W_{t-1}) + \beta_2 ln(S_{t-2}) + \epsilon_t$		18.6	0.47	0.496	116.56	1.34

Table S2. Model selection tests of time-dependency the Jul-Sep catch 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	Adj. R^2	F	p value	AICc	LOOCV RMSE
Time dependency test A 1983-2015 data						
1. $ln(S_t) = \alpha + \beta ln(S_{t-1}) + \epsilon_t$	31	9.8			118.46	1.43
2. $ln(S_t) = \alpha + s(ln(S_{t-1})) + \epsilon_t$	29	19.9	2.73	0.085	116.75	1.363
3. $ln(S_t) = \alpha + s_1(ln(S_{t-1})) + s_2(ln(S_{t-2})) + \epsilon_t$	26.2	20.2	0.86	0.466	120.94	1.379
Time dependency test B 1983-2015 data						
1. $ln(S_t) = \alpha + \beta ln(W_{t-1}) + \epsilon_t$	31	20			114.48	1.319
2. $ln(S_t) = \alpha + s(ln(W_{t-1})) + \epsilon_t$	29.6	21.7	1.12	0.321	115.22	1.313
3. $ln(S_t) = \alpha + s_1(ln(W_{t-1})) + s_2(ln(W_{t-2})) + \epsilon_t$	27.3	18.9	0.36	0.732	119.59	1.352
Time dependency test C 1983-2015 data						
1. $ln(S_t) = \alpha + s(ln(W_{t-1})) + \epsilon_t$	29.6	21.7			115.22	1.313
2. $ln(S_t) = \alpha + s_1(ln(W_{t-1})) + s_2(ln(S_{t-1})) + \epsilon_t$	26.9	26.7	1.59	0.218	117.02	1.285
3. $ln(S_t) = \alpha + s_1(ln(W_{t-1})) + s_2(ln(S_{t-2})) + \epsilon_t$	26.8	23.4	1.06	0.381	118.61	1.312
Time varying test D 1983-2015 data						
1. $ln(S_t) = \alpha_t + \epsilon_t$	29				115.3	1.373
2. $ln(S_t) = \alpha_t + \beta_t t + \epsilon_t$	27				116.55	1.354
3a. $ln(S_t) = \alpha + \beta_t ln(S_{t-1}) + \epsilon_t$	28				117.14	1.49
3b. $ln(S_t) = \alpha + \beta_t ln(W_{t-1}) + \epsilon_t$					113.07	1.337

Table S3. Table S2 with 1956-1982 data instead of 1983 to 2015 data. See Table S1 for an explanation of the parameters and model set-up.

Model	Resid. df	Adj. R^2	F	p value	AICc	LOOCV RMSE
Time dependency test A 1956-1982 data						
1. $ln(S_t) = \alpha + \beta ln(S_{t-1}) + \epsilon_t$	23	1			62.58	0.809
2. $ln(S_t) = \alpha + s(ln(S_{t-1})) + \epsilon_t$	21.1	2.9	1	0.382	64.4	0.831
3. $ln(S_t) = \alpha + s_1(ln(S_{t-1})) + s_2(ln(S_{t-2})) + \epsilon_t$	18.9	8.3	1.42	0.267	67.06	1.023
Time dependency test B 1956-1982 data						
1. $ln(S_t) = \alpha + \beta ln(W_{t-1}) + \epsilon_t$	23	-3.7			63.76	0.814
2. $ln(S_t) = \alpha + s(ln(W_{t-1})) + \epsilon_t$	20.6	10.1	2.44	0.109	63.27	0.77
3. $ln(S_t) = \alpha + s_1(ln(W_{t-1})) + s_2(ln(W_{t-2})) + \epsilon_t$	17.4	18	1.44	0.264	67.62	0.783
Time dependency test C 1956-1982 data						
1. $ln(S_t) = \alpha + s(ln(W_{t-1})) + \epsilon_t$	21.5	5.7			65.74	0.79
2. $ln(S_t) = \alpha + s_1(ln(W_{t-1})) + s_2(ln(S_{t-1})) + \epsilon_t$	19.7	9.2	1.31	0.289	68.21	0.814
3. $ln(S_t) = \alpha + s_1(ln(W_{t-1})) + s_2(ln(S_{t-2})) + \epsilon_t$	18.6	15.5	1.49	0.252	65.73	0.972

Table S4. Model selection tests of time-dependency the post-monsoon Oct-Mar (W_t) catch using F-tests of nested models fit to 1983 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 W_{2014} includes quarters in 2014 and 2015. W_t is the Oct-Mar catch. S_{t-1} and W_{t-1} are the catch during the prior year. S_{t-2} and W_{t-2} are the same for two years prior. Test A uses the Jul-Sep (monsoon) catch as the explanatory variable. Test B uses Oct-Mar post-monsoon catch 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	Adj. R^2	F	p value	AICc	LOOCV RMSE
Naive Model 1983-2014 data						
$ln(W_t) = ln(W_{t-1}) + \epsilon_t$	32				92.86	0.999
Time dependency test A 1983-2014 data						
1. $ln(W_t) = \alpha + ln(S_{t-1}) + \epsilon$	31	-19.4			110.23	1.305
2. $ln(W_t) = \alpha + \beta ln(S_{t-1}) + \epsilon_t$	30	26.3	20.34	0	96.2	1.023
3. $ln(W_t) = \alpha + \beta_1 ln(S_{t-1}) + \beta_2 ln(S_{t-2}) + \epsilon_t$	29	26.8	1.21	0.281	97.52	1.048
Time dependency test B 1983-2014 data						
1. $ln(W_t) = \alpha + ln(W_{t-1}) + \epsilon_t$	31	25.5			95.14	1.031
2. $ln(W_t) = \alpha + \beta ln(W_{t-1}) + \epsilon_t$	30	37.5	6.74	0.015	90.91	1.048
3. $ln(W_t) = \alpha + \beta_1 ln(W_{t-1}) + \beta_2 ln(W_{t-2}) + \epsilon_t$	29	35.4	0.03	0.861	93.5	1.132
Time dependency test C 1983-2014 data						
1. $ln(W_t) = \alpha + \beta ln(W_{t-1}) + \epsilon_t$	30	37.5			90.91	1.048
2a. $ln(W_t) = \alpha + \beta_1 ln(W_{t-1}) + \beta_2 ln(S_{t-1}) + \epsilon_t$	29	35.6	0.11	0.746	93.41	1.078
2b. $ln(W_t) = \alpha + \beta_1 ln(W_{t-1}) + \beta_2 ln(S_{t-2}) + \epsilon_t$		35.4	0.01	0.923	93.52	1.191

Table S5. Model selection tests of time-dependency the W_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	Adj. R^2	F	p value	AICc	LOOCV RMSE
Time dependency test A 1983-2014 data						
1. $ln(W_t) = \alpha + \beta ln(S_{t-1}) + \epsilon_t$	30	26.3			96.2	1.023
2. $ln(W_t) = \alpha + s(ln(S_{t-1})) + \epsilon_t$	28.1	30	1.72	0.2	96.72	1.007
3. $ln(W_t) = \alpha + s_1(ln(S_{t-1})) + s_2(ln(S_{t-2})) + \epsilon_t$	25.1	36.6	1.84	0.167	98.16	1.004
Time dependency test B 1983-2014 data						
1. $ln(W_t) = \alpha + \beta ln(W_{t-1}) + \epsilon_t$	30	37.5			90.91	1.048
2. $ln(W_t) = \alpha + s(ln(W_{t-1})) + \epsilon_t$	28.6	45.9	4.06	0.042	87.74	0.955
3. $ln(W_t) = \alpha + s_1(ln(W_{t-1})) + s_2(ln(W_{t-2})) + \epsilon_t$	26.4	46.1	0.87	0.443	90.78	1.013
Time dependency test C 1983-2014 data						
1. $ln(W_t) = \alpha + s(ln(W_{t-1})) + \epsilon_t$	28.6	45.9			87.74	0.955
2. $ln(W_t) = \alpha + s_1(ln(W_{t-1})) + s_2(ln(S_{t-1})) + \epsilon_t$	26	44.6	0.58	0.615	92.46	1.057
3. $ln(W_t) = \alpha + s_1(ln(W_{t-1})) + s_2(ln(S_{t-2})) + \epsilon_t$	25.6	57.3	3.43	0.032	84.58	1.055
Time varying test D 1983-2014 data						
1. $ln(W_t) = \alpha_t + \epsilon_t$	28				93.87	1.043
2. $ln(W_t) = \alpha_t + \beta_t t + \epsilon_t$	26				99.36	1.045
3a. $ln(W_t) = \alpha + \beta_t ln(S_{t-1}) + \epsilon_t$	27				95.5	0.923
3b. $ln(W_t) = \alpha + \beta_t ln(W_{t-1}) + \epsilon_t$					91.82	1.031

Table S6. Table S5 with 1956-1982 data instead of 1983 to 2014 data. The years used in the fit start in 1958 since $t-2$ (which is 1956 for the 1958 data point) is used in the covariates. See Table S4 for an explanation of the parameters and model set-up.

Model	Residual df	Adj. R^2	F	p value	AICc	LOOCV RMSE
Time dependency test A 1958-1983 data						
1. $ln(W_t) = \alpha + \beta ln(S_{t-1}) + \epsilon_t$	24	-1.7			46.07	0.574
2. $ln(W_t) = \alpha + s(ln(S_{t-1})) + \epsilon_t$	22.1	16.2	3.53	0.052	43.29	0.542
3. $ln(W_t) = \alpha + s_1(ln(S_{t-1})) + s_2(ln(S_{t-2})) + \epsilon_t$	19.9	18.1	1.09	0.362	46.67	0.615
Time dependency test B 1958-1983 data						
1. $ln(W_t) = \alpha + \beta ln(W_{t-1}) + \epsilon_t$	24	-4.2			46.7	0.575
2. $ln(W_t) = \alpha + s(ln(W_{t-1})) + \epsilon_t$	21.6	29.1	5.69	0.009	39.78	0.468
3. $ln(W_t) = \alpha + s_1(ln(W_{t-1})) + s_2(ln(W_{t-2})) + \epsilon_t$	18.5	32.2	1.14	0.36	44.58	0.506
Time dependency test C 1958-1983 data						
1. $ln(W_t) = \alpha + s(ln(W_{t-1})) + \epsilon_t$	21.6	29.1			39.78	0.468
2a. $ln(W_t) = \alpha + s_1(ln(W_{t-1})) + s_2(ln(S_{t-1})) + \epsilon_t$	19	34.4	1.49	0.251	42.64	0.498
2b. $ln(W_t) = \alpha + s_1(ln(W_{t-1})) + s_2(ln(S_{t-2})) + \epsilon_t$	19.5	33.4	1.54	0.24	42.03	0.538

Tests for environmental variables as covariates

Table S7. Covariate tests for the Oct-Mar catch (W_t) using the more complex model. M is the base model with prior season Oct-Mar catch (W_{t-1}) and Jul-Sep catch two seasons prior (S_{t-2}) as the covariates. To the base model, the environmental covariates are added. Nearshore is 0-80km and regional is 0-160km. The SST data are from AVHRR. The models are nested sets, e.g. 1, 2a, 3a and 1, 2b, 3b.

Model	Resid. df	Adj. R^2	RMSE	AICc	LOOCV RMSE	LOOCV MdAE
catch only models 1983-2014 data						
null model: $ln(W_t) = ln(W_{t-1}) + \epsilon_t$	32		0.999	92.9	0.999	0.256
base model (M): 1. $ln(W_t) = \alpha + s(ln(W_{t-1})) + s(ln(S_{t-2})) + \epsilon_t$	26.6	57.3	0.7	84.6	1.055	0.345
Precipitation						
V_t = Jun-Jul Precipitation - ocean						
2a. $ln(W_t)$ = M + βV_t	25.7	57.6	0.685	86.5	1.083	0.365
3a. $ln(W_t)$ = M + $s(V_t)$	24.6	56.4	0.681	89.9	1.141	0.367
2b. $ln(W_t)$ = M + βV_{t-1}	25.6	55.7	0.7	87.9	1.066	0.375
3b. $ln(W_t)$ = M + $s(V_{t-1})$	24.5	57.7	0.669	89.1	1.058	0.347
V_t = Jun-Jul Precipitation - land						
2a. $ln(W_t)$ = M + βV_t	25.7	63.2	0.638	81.9†	1.071	0.376
3a. $ln(W_t)$ = M + $s(V_t)$	24.6	70.5	0.56	77.5††	0.965‡	0.292‡‡
2b. $ln(W_t)$ = M + βV_{t-1}	25.7	55.7	0.701	87.8	1.081	0.346
3b. $ln(W_t)$ = M + $s(V_{t-1})$	24.7	55.3	0.691	90.5	1.088	0.331
V_t = Apr-May Precipitation - ocean						
2a. $ln(W_t)$ = M + βV_t	25.6	55.7	0.7	87.9	1.071	0.372
3a. $ln(W_t)$ = M + $s(V_t)$	24.4	54.2	0.694	92.1	1.098	0.477
2b. $ln(W_t)$ = M + βV_{t-1}	25.6	56.8	0.692	87.2	1.041	0.36
3b. $ln(W_t)$ = M + $s(V_{t-1})$	24.4	56.4	0.677	90.5	1.049	0.357
V_t = Apr-May Precipitation - land						
2a. $ln(W_t)$ = M + βV_t	25.7	58.6	0.677	85.8	0.994‡	0.362
3a. $ln(W_t)$ = M + $s(V_t)$	23.8	57.6	0.66	91.2	0.998‡	0.475
2b. $ln(W_t)$ = M + βV_{t-1}	25.6	55.7	0.7	87.9	1.071	0.346
3b. $ln(W_t)$ = M + $s(V_{t-1})$	23.8	52.6	0.698	94.8	1.1	0.356
Sea surface temperature						
V_t = Mar-May SST - regional						
2a. $ln(W_t)$ = M + βV_t	25.7	58	0.682	86.2	1.06	0.428
3a. $ln(W_t)$ = M + $s(V_t)$	23.7	59.9	0.641	89.7	1.036	0.408
2b. $ln(W_t)$ = M + βV_{t-1}	25.6	58.8	0.676	85.6	1.036	0.509
3b. $ln(W_t)$ = M + $s(V_{t-1})$	24	57.1	0.667	91	1.048	0.532
V_t = Oct-Dec SST - nearshore						
2a. $ln(W_t)$ = M + βV_t	25.7	57.9	0.683	86.2	1.135	0.363
3a. $ln(W_t)$ = M + $s(V_t)$	24.8	56.7	0.68	89.3	1.176	0.382
2b. $ln(W_t)$ = M + βV_{t-1}	25.7	56.3	0.696	87.4	1.035	0.348
3b. $ln(W_t)$ = M + $s(V_{t-1})$	24.7	55.4	0.689	90.5	1.106	0.348

Model	Resid. df	Adj. R^2	RMSE	AICc	LOOCV RMSE	LOOCV MdAE
Upwelling						
V_t = Jun-Sep nearshore-offshore SST differential						
2a. $ln(W_t)$ = M + βV_t	25.6	56.6	0.693	87.3	1.107	0.338
3a. $ln(W_t)$ = M + $s(V_t)$	24.4	55.7	0.683	91.1	1.13	0.396
2b. $ln(W_t)$ = M + βV_{t-1}	25.7	55.9	0.699	87.7	1.057	0.372
3b. $ln(W_t)$ = M + $s(V_{t-1})$	24.4	56.5	0.677	90.3	1.072	0.378
V_t = Jun-Sep SST - nearshore						
2a. $ln(W_t)$ = M + βV_t	25.6	59.8	0.667	84.8	1.128	0.408
3a. $ln(W_t)$ = M + $s(V_t)$	24		0.645	89.1	1.137	0.46
2b. $ln(W_t)$ = M + βV_{t-1}	25.7	56.7	0.693	87.2	1.041	0.409
3b. $ln(W_t)$ = M + $s(V_{t-1})$	24.1	55.7	0.679	91.8	1.091	0.442
V_t = Jun-Sep Ekman Mass Transport - nearshore						
2a. $ln(W_t)$ = M + βV_t	25.7	55.9	0.699	87.7	1.076	0.377
3a. $ln(W_t)$ = M + $s(V_t)$	24.2	55.2	0.684	91.9	1.101	0.366
2b. $ln(W_t)$ = M + βV_{t-1}	25.7	58.2	0.681	86	1.005	0.432
3b. $ln(W_t)$ = M + $s(V_{t-1})$	24.4	57.3	0.67	89.8	1.034	0.447
V_t = Apr-May Ekman Mass Transport - nearshore						
2a. $ln(W_t)$ = M + βV_t	25.7	55.8	0.701	87.7	1.098	0.348
3a. $ln(W_t)$ = M + $s(V_t)$	24.6	58.3	0.665	88.5	1.123	0.343
2b. $ln(W_t)$ = M + βV_{t-1}	25.6	57.7	0.685	86.5	1.067	0.394
3b. $ln(W_t)$ = M + $s(V_{t-1})$	24.5	60	0.65	87.6	1.051	0.407
V_t = Jun-Sep Ekman Pumping - nearshore						
2a. $ln(W_t)$ = M + βV_t	25.7	62.2	0.647	82.8	1.067	0.437
3a. $ln(W_t)$ = M + $s(V_t)$	24.4	61.9	0.633	86.2	1.127	0.437
2b. $ln(W_t)$ = M + βV_{t-1}	25.7	55.7	0.701	87.9	1.094	0.361
3b. $ln(W_t)$ = M + $s(V_{t-1})$	24.4	61.8	0.635	86.2	0.918‡‡	0.392
V_t = Jun-Sep Ekman Pumping - tip of India						
2a. $ln(W_t)$ = M + βV_t	25.7	58.8	0.676	85.5	1.086	0.403
3a. $ln(W_t)$ = M + $s(V_t)$	24.4	58	0.665	89.2	1.124	0.471
2b. $ln(W_t)$ = M + βV_{t-1}	25.7	55.6	0.702	87.9	1.085	0.364
3b. $ln(W_t)$ = M + $s(V_{t-1})$	24.5	60	0.651	87.5	1.004	0.507
V_t = Jan-Feb Ekman Pumping - tip of India						
2a. $ln(W_t)$ = M + βV_t	25.7	57.8	0.684	86.3	1.094	0.37
3a. $ln(W_t)$ = M + $s(V_t)$	24.9	56.8	0.681	89	1.118	0.376
2b. $ln(W_t)$ = M + βV_{t-1}	25.6	55.8	0.7	87.9	1.069	0.346
3b. $ln(W_t)$ = M + $s(V_{t-1})$	24.8	55	0.694	90.5	1.092	0.376
Ocean climate						
V_t = 2.5-year average SST - regional						
2a. $ln(W_t)$ = M + βV_t	25.6	67.6	0.599	77.9††	0.862‡‡	0.406

Model	Resid. df	Adj. R^2	RMSE	AICc	LOOCV RMSE	LOOCV MdAE
3a. $ln(W_t)$ = M + $s(V_t)$	24.8	69.5	0.571	78.1††	0.825‡‡‡	0.346
V_t = ONI Jul-Jun average						
2a. $ln(W_t)$ = M + βV_t	25.6	55.7	0.7	87.9	1.086	0.37
3a. $ln(W_t)$ = M + $s(V_t)$	24.7	57.4	0.674	88.9	1.078	0.397
V_t = PDO Jul-Jun average						
2a. $ln(W_t)$ = M + βV_t	25.7	56.2	0.696	87.5	1.043	0.34
3a. $ln(W_t)$ = M + $s(V_t)$	24.2	55.9	0.678	91.5	1.043	0.422
V_t = AMO Jul-Jun average						
2a. $ln(W_t)$ = M + βV_t	25.7	64.8	0.625	80.5†	0.914‡‡	0.393
3a. $ln(W_t)$ = M + $s(V_t)$	24.3	65.6	0.601	83.1	0.91‡‡	0.291‡‡
V_t = Sep-Nov DMI						
2a. $ln(W_t)$ = M + βV_t	25.7	56.4	0.696	87.2	1.089	0.33
3a. $ln(W_t)$ = M + $s(V_t)$	23.6	57.9	0.655	91.6	1.198	0.371
2b. $ln(W_t)$ = M + βV_{t-1}	25.7	55.6	0.702	87.9	1.076	0.331
3b. $ln(W_t)$ = M + $s(V_{t-1})$	23.8	69.1	0.564	81.1†	0.872‡‡	0.343
catch only models 1998-2014 data						
null model: $ln(W_t) = ln(W_{t-1}) + \epsilon_t$	17		0.432	22	0.432	0.133
base model (M): 1. $ln(W_t) = \alpha + p(ln(W_{t-1})) + p(ln(S_{t-2})) + \epsilon_t$	12	15.9	0.331	31	0.628	0.382
Chlorophyll						
V_t = Jul-Sep CHL - nearshore						
2a. $ln(W_t)$ = M + βV_t	11	11.4	0.325	36.5	0.738	0.361‡
3a. $ln(W_t)$ = M + $p(V_t)$	10	3.1	0.324	43.9	0.799	0.398
2b. $ln(W_t)$ = M + βV_{t-1}	11	18.7	0.311	35	0.614	0.378
3b. $ln(W_t)$ = M + $p(V_{t-1})$	10	10.8	0.311	42.5	1.619	0.525
V_t = Oct-Dec CHL - nearshore						
2a. $ln(W_t)$ = M + βV_t	11	11.6	0.325	36.4	0.652	0.337‡‡
3a. $ln(W_t)$ = M + $p(V_t)$	10	25.8	0.284	39.4	0.571‡	0.266‡‡‡
2b. $ln(W_t)$ = M + βV_{t-1}	11	41.1	0.265	29.5	0.497‡‡‡	0.307‡‡
3b. $ln(W_t)$ = M + $p(V_{t-1})$	10	35.7	0.264	37	0.6	0.329‡‡

Notes: LOOCV = Leave one out cross-validation. RMSE = root mean square error. MdAE = median absolute error. AICc = Akaike Information Criterion corrected for small sample size. † and †† = AICc greater than 2 and greater than 5 below model M (base catch model). ‡, ‡‡, and ‡‡‡ = LOOCV RMSE 5%, 10% and 20% below model M, respectively. t indicates current year and $t-1$ is the prior year. W_t spans two calendar years (Oct-Mar); t is the year in Oct. Thus if $t = 2014$, W_t is Oct 2014 to Mar 2015 and W_{t-1} is Oct 2013 to Mar 2014. For covariates that are multiyear, such as the multiyear average SST, t is the calendar year at the end of the multiyear span; thus the 2.5 year average SST for 2014 is Jan 2012 to Jun 2014.

Tests with upwelling and precipitation interactions

Table S8. Effect of interaction between upwelling and precipitation for the monsoon (Jul-Sep) and post-monsoon (Oct-Mar) catch (S_t and W_t) models. The models with upwelling-precipitation interaction are compared to the model with the 2.5 year average regional SST as a covariate (the model with AMO is similar). The upwelling index used is the SST nearshore-offshore differential (the Ekman Mass Transport index performed much more poorly). LI = Linear interaction. NLI = Non-linear interaction. ti() is a tensor (non-linear) interaction.

Model	Resid. df	Adj. R^2	RMSE	AICc	LOOCV RMSE	LOOCV MdAE
Jul-Sep catch only models 1983-2015 data						
null model: $ln(S_t) = ln(S_{t-1}) + \epsilon_t$	33		1.596	126.6	1.596	0.559
base model (M): $ln(W_t) = \alpha + s(ln(W_{t-1}) + \epsilon_t$	30	21.7	1.204	115.2	1.313	0.692
$ln(W_t) = M + s(SST)$	28.1	47	0.958	105.7††	1.375	0.558‡‡
$ln(S_t) = M + s(UPW)$	27.7	29.6	1.097	115.8	1.314	0.56‡‡
$ln(S_t) = M + s(Pr)$	28	29.9	1.1	115.3	1.327	0.62‡‡
LI: $ln(S_t) = M + \beta_1 UPW + \beta_2 Pr + \beta_3(UPW \times Pr)$	27.1	29.8	1.084	117	1.461	0.529‡‡‡
LI: $ln(S_t) = M + \beta(UPW \times Pr)$	29.1	20.6	1.193	117.2	1.357	0.63‡
NLI: $ln(S_t) = M + s(UPW \times Pr)$	27	30.1	1.078	117.2	1.791	0.653‡
NLI: $ln(S_t) = M + ti(Pr) + ti(UPW, Pr)$	21.9	38.7	0.909	127.5	1.565	0.624‡
Oct-Mar catch only models 1983-2014 data						
null model: $ln(W_t) = ln(W_{t-1}) + \epsilon_t$	32		0.999	92.9	0.999	0.256
base model (M): $ln(W_t) = \alpha + s(ln(W_{t-1}) + \epsilon_t$	29.1	45.9	0.824	87.7	0.955	0.323
$ln(W_t) = M + s(SST)$	27.1	65.9	0.632	76.4††	0.765‡‡	0.411
$ln(W_t) = M + s(UPW)$	26.8	44.8	0.798	92.5	1.01	0.411
$ln(W_t) = M + s(Pr)$	26.9	59.6	0.685	82.1††	0.906‡	0.246‡‡‡
LI: $ln(W_t) = M + \beta_1 UPW + \beta_2 Pr + \beta_3(UPW \times Pr)$	26.1	52.5	0.732	89	1.095	0.404
LI: $ln(W_t) = M + \beta(UPW \times Pr)$	28.1	44.4	0.822	90.2	1.002	0.43
NLI: $ln(W_t) = M + s(UPW \times Pr)$	26.4	48	0.769	91.5	0.983	0.288‡‡
NLI: $ln(W_t) = M + ti(Pr) + ti(UPW, Pr)$	20.7	68.3	0.532	92.4	0.968	0.316

Validation of catch base models

Test set-up

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 W_{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.

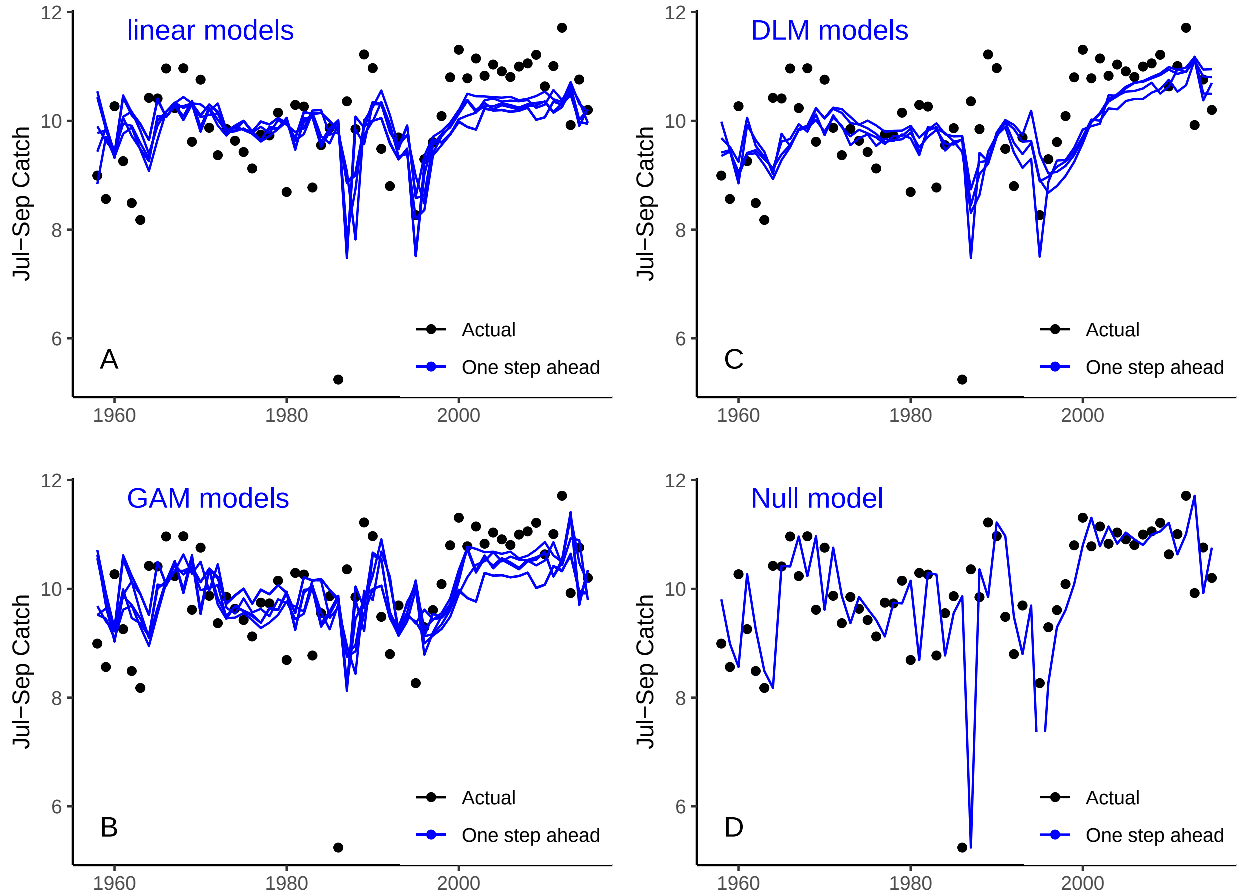


Figure S2. Leave one out (LOO) one step ahead predictions for the linear, GAM, and DLM models of Jul-Sep landings. The data point at year t on the x-axis is predicted from the data up to year t-1.

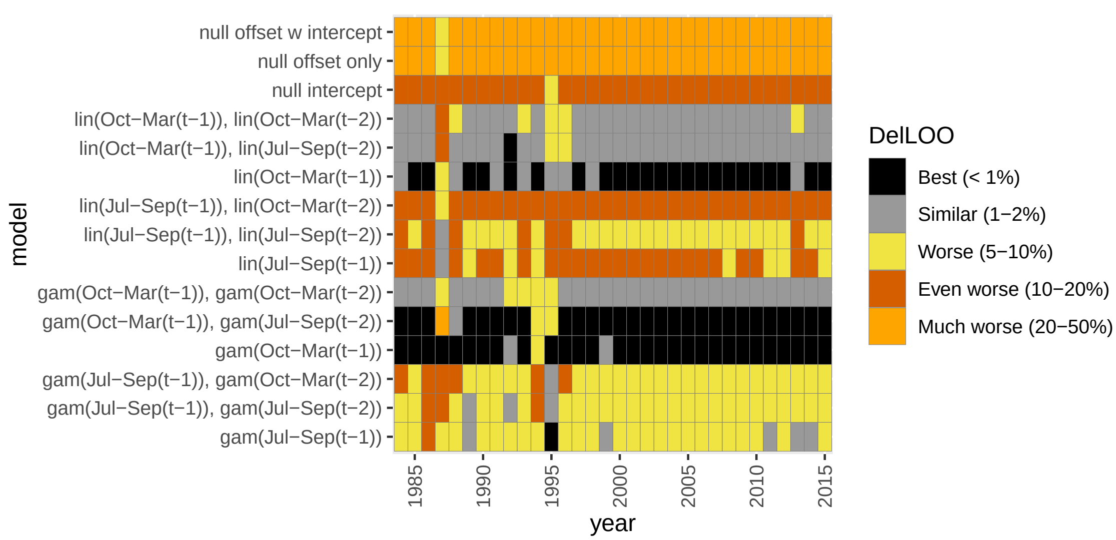


Figure S3. Leave-one-out predictive performance (leave out a year, fit, predict that year) for the Jul-Sep landings base models. The performance (DelLOO) is the RSME (root mean square error) between prediction and observed.

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).

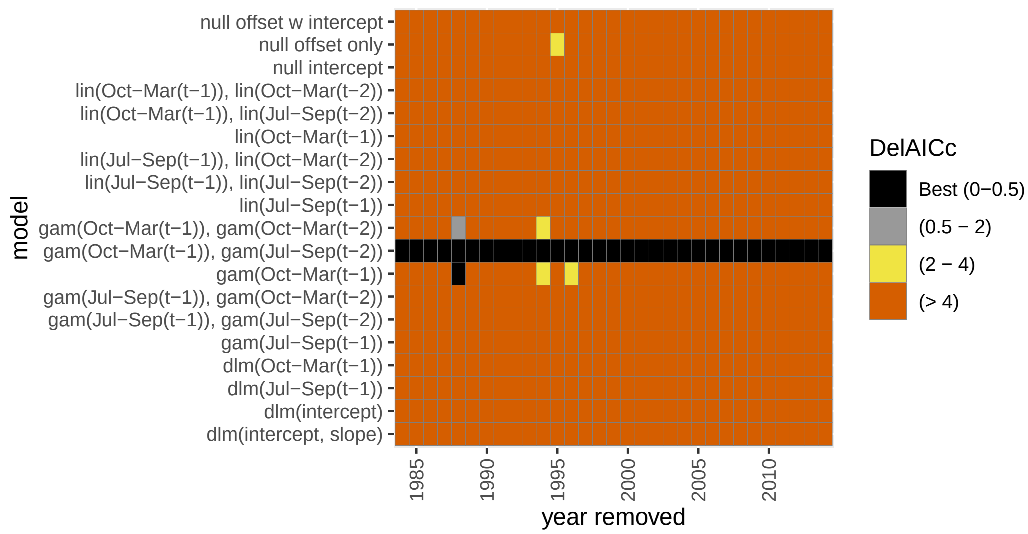


Figure S4. ΔAICc for the Oct-Mar landings base models with one year deleted using only the landings data that overlap with the environmental data 1984-2015. See Figure S1 for an explantion of the figure.

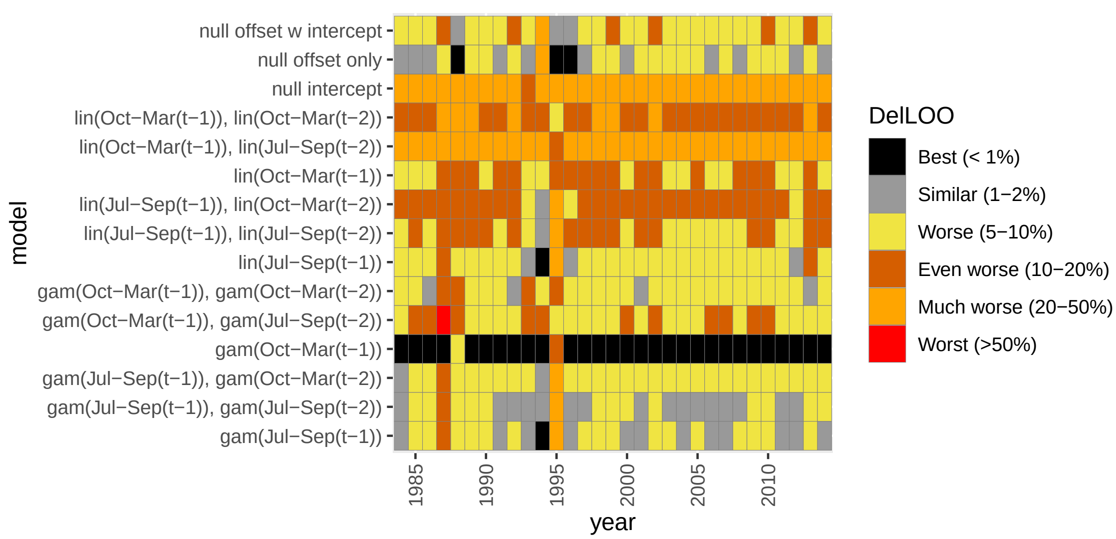


Figure S5. Leave-one-out predictive performance (leave out a year, fit, predict that year) for the Oct-Mar landings base models. The performance (DelLOO) is the RSME (root mean square error).

Comparison of land and oceanic rainfall measurements

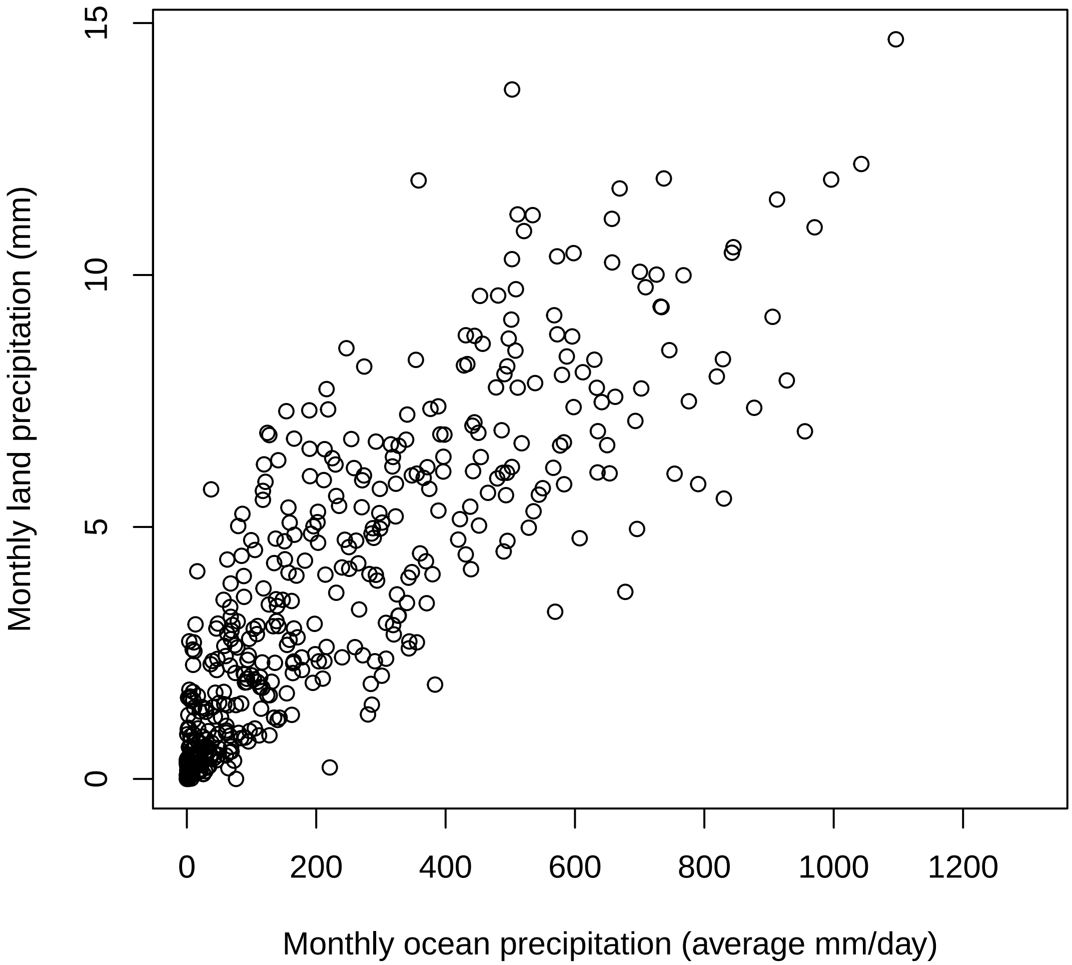


Figure S6. Monthly precipitation measured over land via land gauges versus the precipitation measured via remote sensing over the ocean.

Comparison of multiyear average regional SST from AVHRR and ICOADS

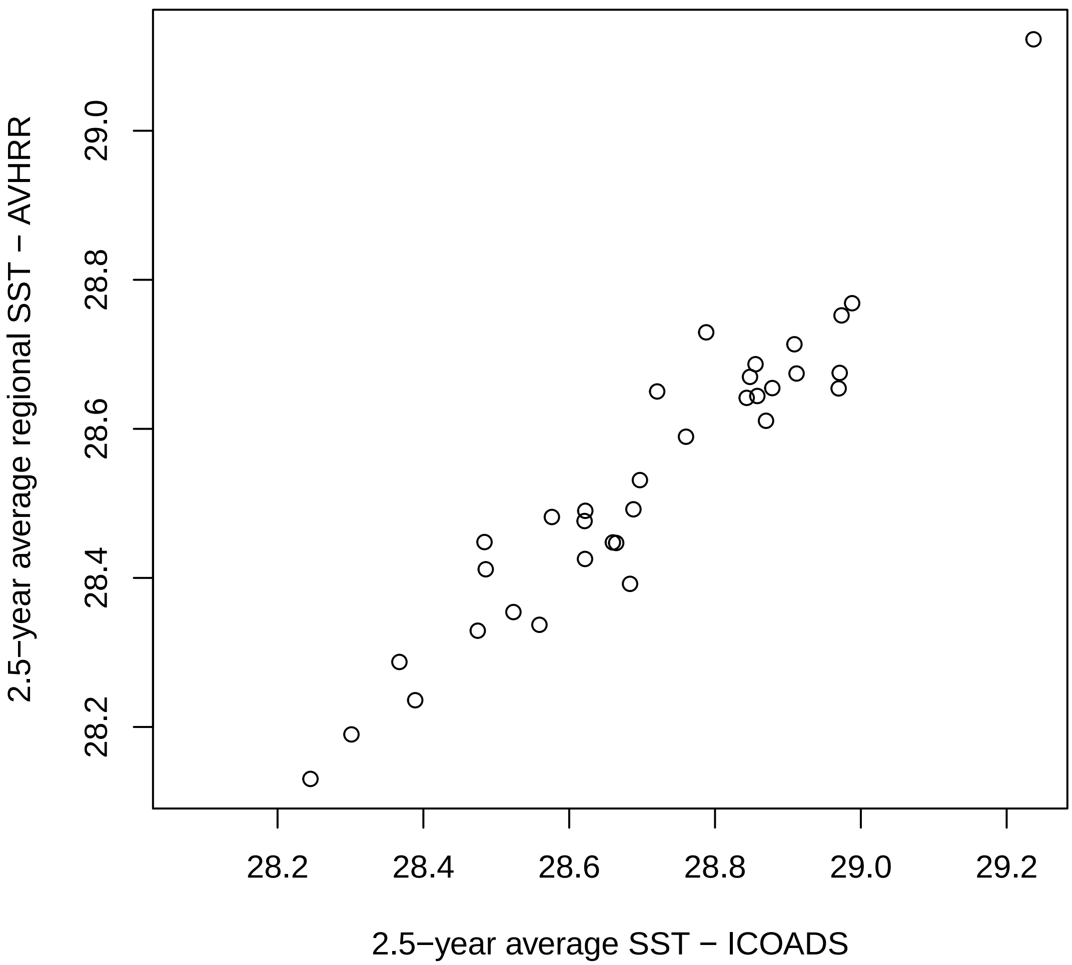


Figure S7. Multiyear average regional SST from the AVHRR versus ICOADS data sets.

SST Product Comparisons

Load needed R libraries

```
library(ggplot2)
library(tidyr)
library(dplyr)
library(raster)
library(rasterVis)
```

Get the data

Get SST from the Daily Optimum Interpolation (OI), AVHRR Only, Version 2. This is on a 0.25 degree grid. The data are from https://coastwatch.pfeg.noaa.gov/erddap/info/ncdcOisst2Agg/index.html.

```
dates <- c("2010-01-01", "2011-01-01")
lats <- c(7, 15); lons <- c(70,78)
sst1 <- getdata("ncdcOisst2Agg", date=dates, lat=lats, lon=lons, pars="sst",
                altitude=0, alt.name="zlev")
```

```
## data read from ncdcOisst2Agg-7-15-70-78-2010-01-01-2011-01-01.csv
## data ncdcOisst2Agg date 2010-01-01-2011-01-01, latitude 7-15, longitude 70-78
```

```
sst1$date <- format(sst1$time, "%Y-%m-%d")
sst1$month <- format(sst1$time, "%m")
sst1$year <- format(sst1$time, "%Y")
sst1.mon <- sst1 %>% group_by(year, month, latitude, longitude) %>%
  summarize(sst = mean(sst, na.rm=TRUE))
sst1.mon$date <- paste0(sst1.mon$year, "-", sst1.mon$month, "-", "01")
```

Now get monthly SST from AVHRR. This is the night and day monthly averages on a 0.0417° grid. https://coastwatch.pfeg.noaa.gov/erddap/info/erdPH2sstamday/index.html.

```
sst2 <- getdata("erdPH2sstamday", date=dates, lat=rev(lats), lon=lons, pars="sea_surface_temperature")
```

```
## data read from erdPH2sstamday-15-7-70-78-2010-01-01-2011-01-01.csv
## data erdPH2sstamday date 2010-01-01-2011-01-01, latitude 15-7, longitude 70-78
```

```
sst2$date <- format(sst2$time, "%Y-%m-%d")
sst2$month <- format(sst2$time, "%m")
sst2$year <- format(sst2$time, "%Y")
sst2$lon1 <- NA
sst2$lat1 <- NA
lats1 <- sort(unique(sst1$latitude))
lons1 <- sort(unique(sst1$longitude))
for(i in lats1){
  sst2$lat1[sst2$latitude > i-0.25 & sst2$latitude < i+0.25] <- i
}
for(i in lons1){
  sst2$lon1[sst2$longitude > i-0.25 & sst2$longitude < i+0.25] <- i
}
```

```
sst2.mon <- sst2 %>% group_by(year, month, lat1, lon1) %>%
  summarize(sst = mean(sea_surface_temperature, na.rm=TRUE))
sst2.mon$date <- paste0(sst2.mon$year, "-", sst2.mon$month, "-", "01")
colnames(sst2.mon) <- colnames(sst1.mon)
```

Now get the Reanalysis Data ERA5 monthly sst from http://apdrc.soest.hawaii.edu/erddap/info/hawaii_soest_d124_2bb9_c935/index.html

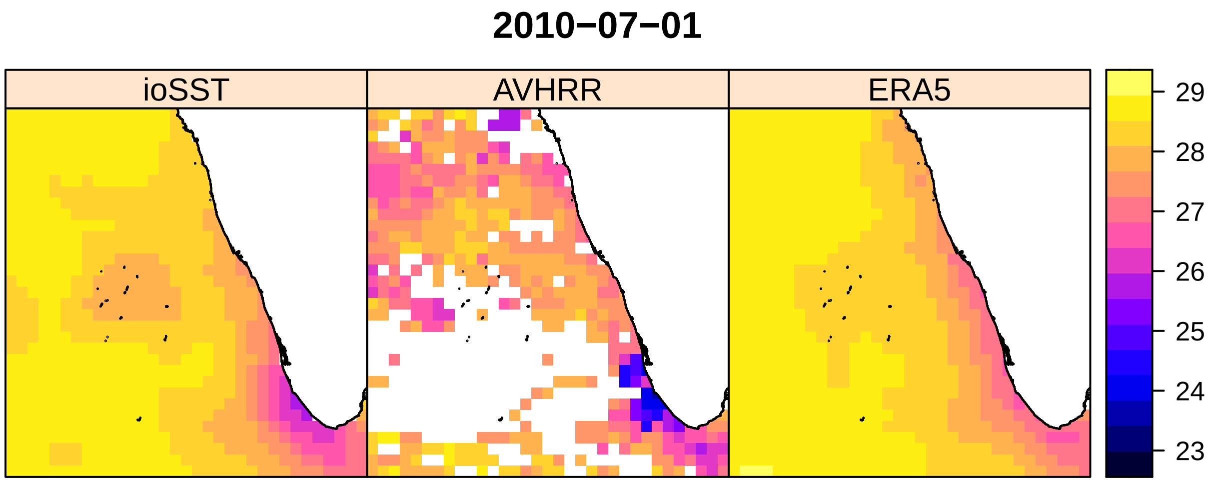
```
sst3 <- getdata("hawaii_soest_d124_2bb9_c935", date=dates, lat=lats, lon=lons,
                    pars="sst", eserver="http://apdrc.soest.hawaii.edu/erddap")
```

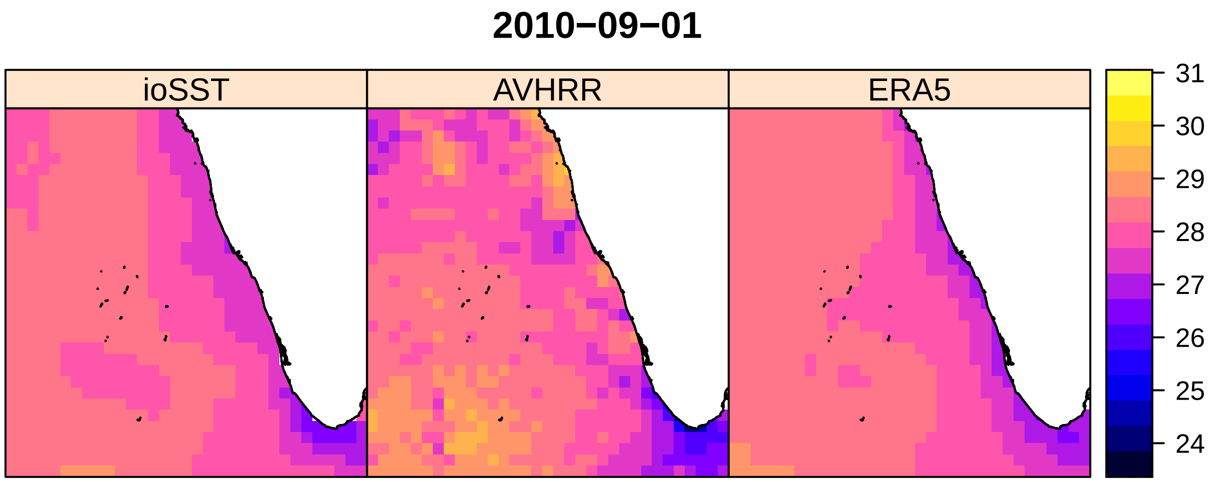
```
## data read from hawaii_soest_d124_2bb9_c935-7-15-70-78-2010-01-01-2011-01-01.csv
## data hawaii_soest_d124_2bb9_c935 date 2010-01-01-2011-01-01, latitude 7-15, longitude 70-78
```

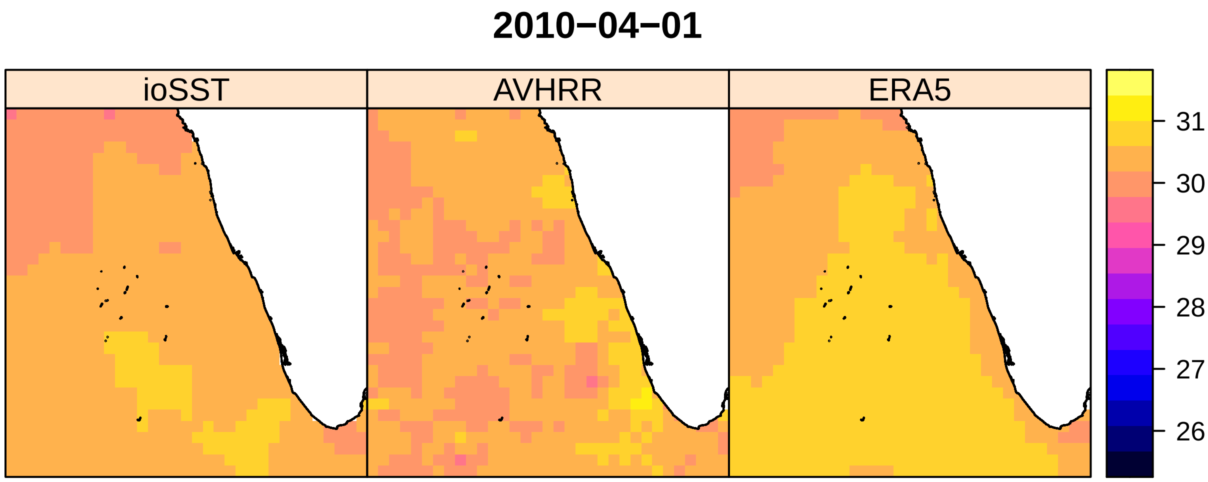
```
sst3$date <- format(sst3$time, "%Y-%m-%d")
sst3$month <- format(sst3$time, "%m")
sst3$year <- format(sst3$time, "%Y")
sst3$sst <- sst3$sst-273.15
sst3$latitude <- sst3$latitude - 0.125
sst3$longitude <- sst3$longitude - 0.125
sst3.mon <- sst3 %>% group_by(year, month, latitude, longitude) %>%
  summarize(sst = mean(sst, na.rm=TRUE))
sst3.mon$date <- paste0(sst3.mon$year, "-", sst3.mon$month, "-", "01")
```

Compare SST on specific dates

We can see this especially for certain months such as April 2010.


2010−07−01
ioSST
AVHRR
ERA5
29
28
27
26
25
24
23


2010−09−01
ioSST
AVHRR
ERA5
31
30
29
28
27
26
25
24



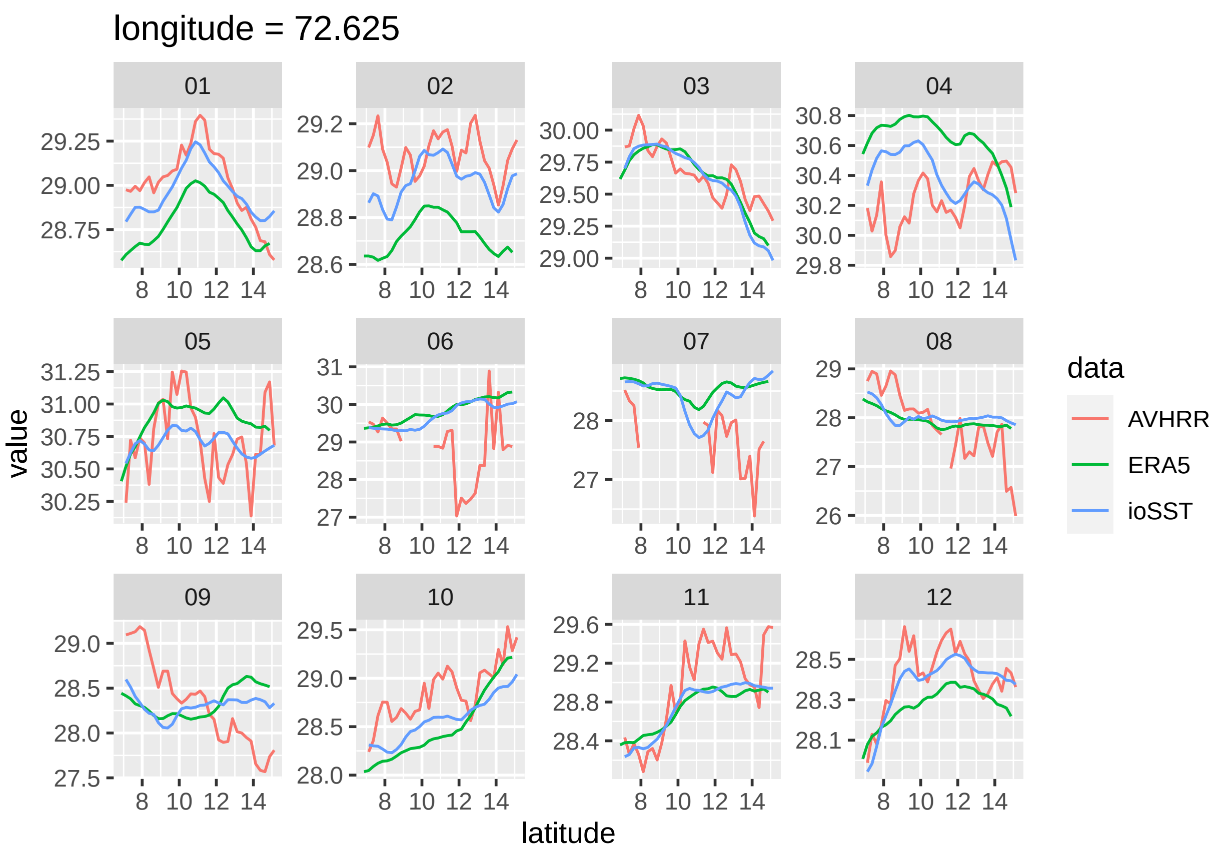
SST along one longitude line

Create the data frame.

```
cols <- c("date", "latitude", "longitude", "sst")
df <- rbind(data.frame(sst1.mon[,cols], data="ioSST"),
            data.frame(sst2.mon[,cols], data="AVHRR"),
            data.frame(sst3.mon[,cols], data="ERA5"))
dfl <- df %>%
  pivot_longer(!date & !latitude & !longitude & !data,
                      names_to = "name",
                      values_to = "value")
```

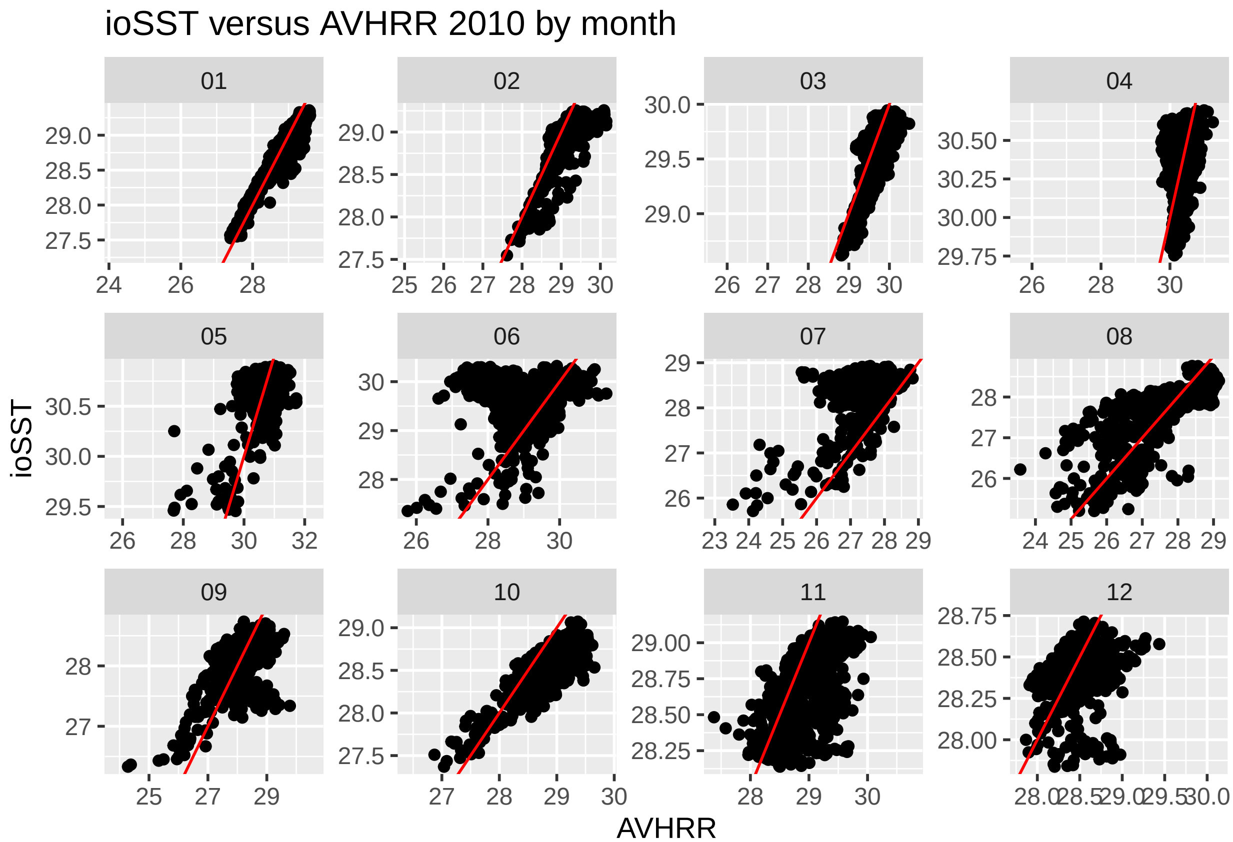
The SST from AVHRR can be quite different from ioSST and ERA5, e.g. Sept 2010. Here the SST along one longitude line is shown for a specific day.

```
thedate <- "2010-07-01"
plotlons <- c(72.625)
dfl$month=format(as.Date(dfl$date), "%m")
dfl$year=format(as.Date(dfl$date), "%Y")
pars <- c("sst")
ggplot(subset(dfl, longitude==plotlons[1] & name%in%pars & year==2010),
       aes(x=latitude, y=value, color=data)) + geom_line() +
  facet_wrap(~month, scales="free") +
  ggtitle(paste("longitude =", plotlons[1]))
```

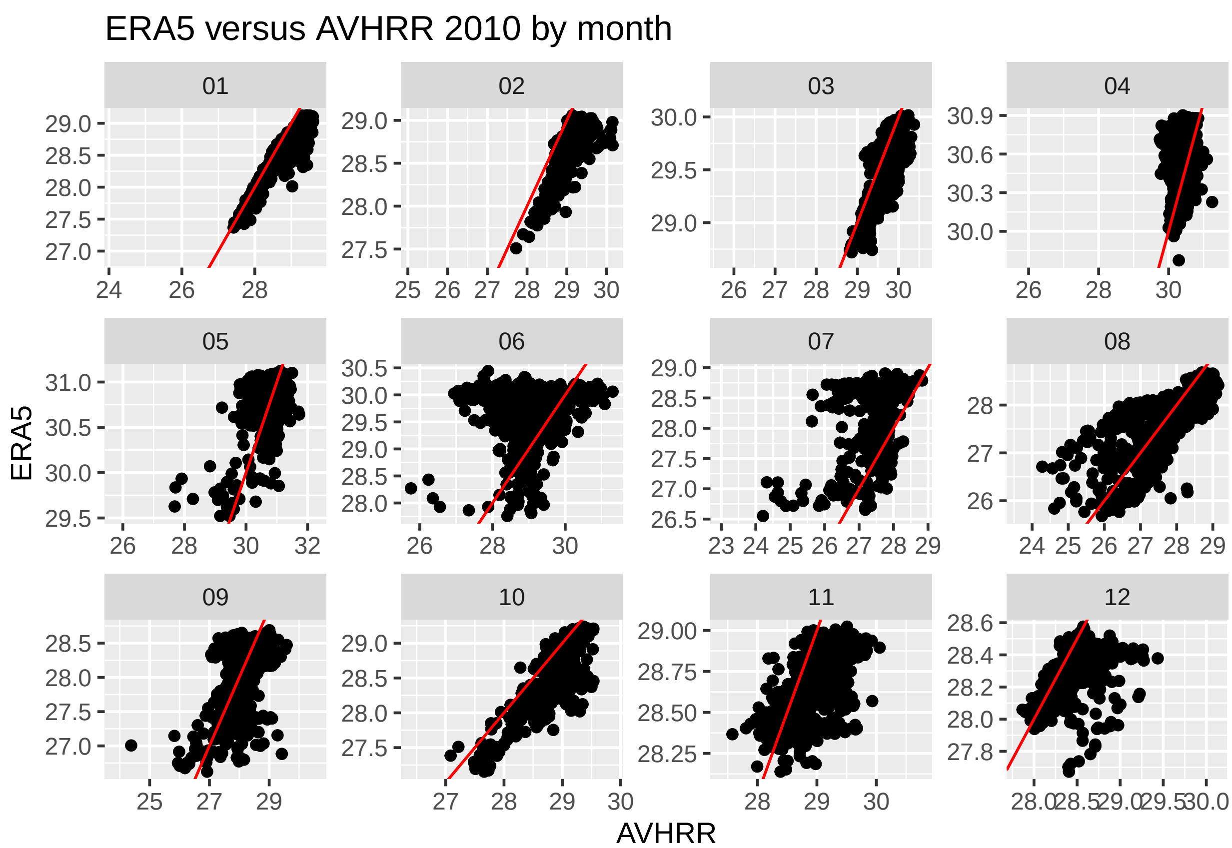


Pairwise comparison by month

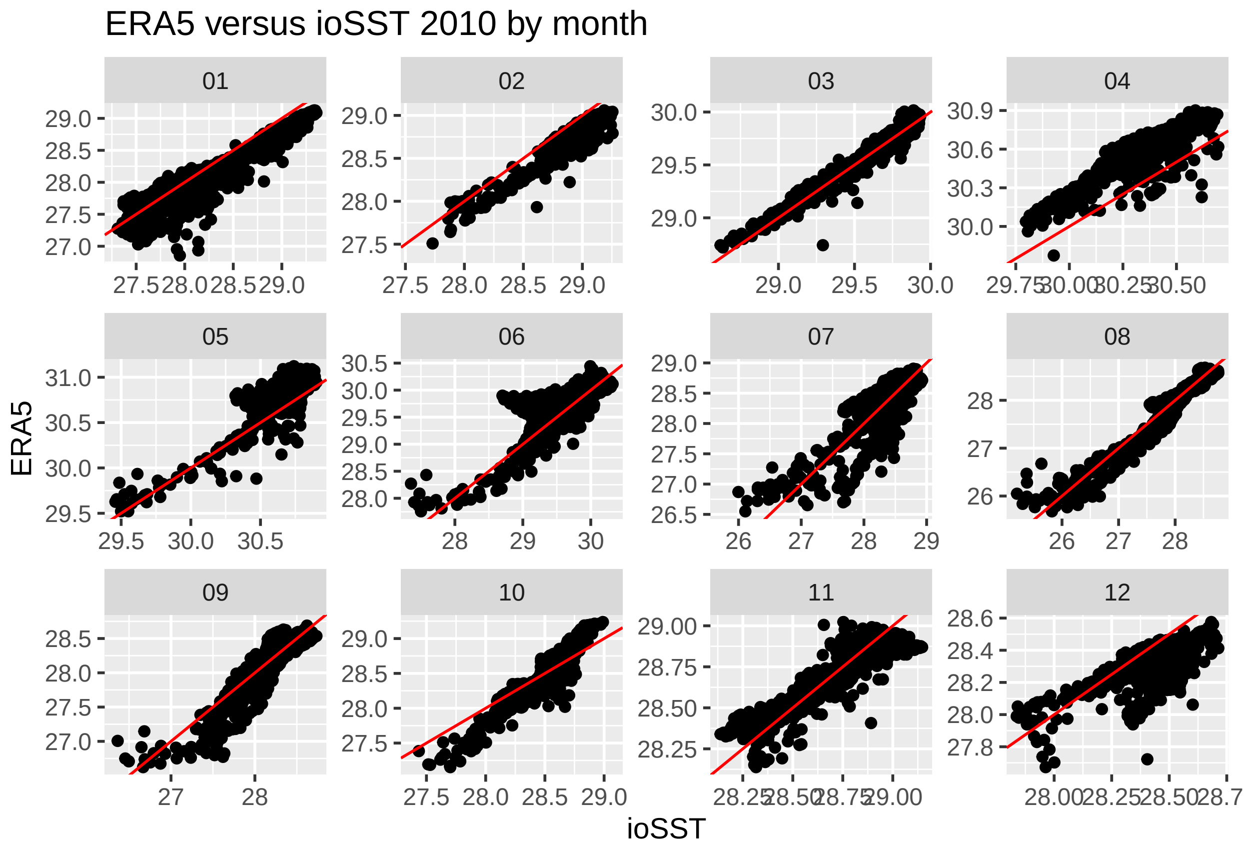
```
dfl2 <- dfl %>% pivot_wider(names_from=data, values_from="value")
dfl2$month <- format(as.Date(dfl2$date), "%m")
ggplot(dfl2, aes(x=AVHRR, y=ioSST)) + geom_point() + geom_abline(col="red") +
  facet_wrap(~month, scales="free") + ggtitle("ioSST versus AVHRR 2010 by month")
```




```
dfl2 <- dfl %>% pivot_wider(names_from=data, values_from="value")
dfl2$month <- format(as.Date(dfl2$date), "%m")
ggplot(dfl2, aes(x=AVHRR, y=ERA5)) + geom_point() + geom_abline(col="red") +
  facet_wrap(~month, scales="free") + ggtitle("ERA5 versus AVHRR 2010 by month")
```




```
dfl2 <- dfl %>% pivot_wider(names_from=data, values_from="value")
dfl2$month <- format(as.Date(dfl2$date), "%m")
ggplot(dfl2, aes(x=ioSST, y=ERA5)) + geom_point() + geom_abline(col="red") +
  facet_wrap(~month, scales="free") + ggtitle("ERA5 versus ioSST 2010 by month")
```



Conclusion

The ioSST and ERA5 products are similar (linear relationship) though biased for some months. The ioSST product was chosen since it uses only AVHRR data and would be similar to previous analyses of remote-sensing SST in the region.

Comparison of multiyear average regional SST and ocean climate indices

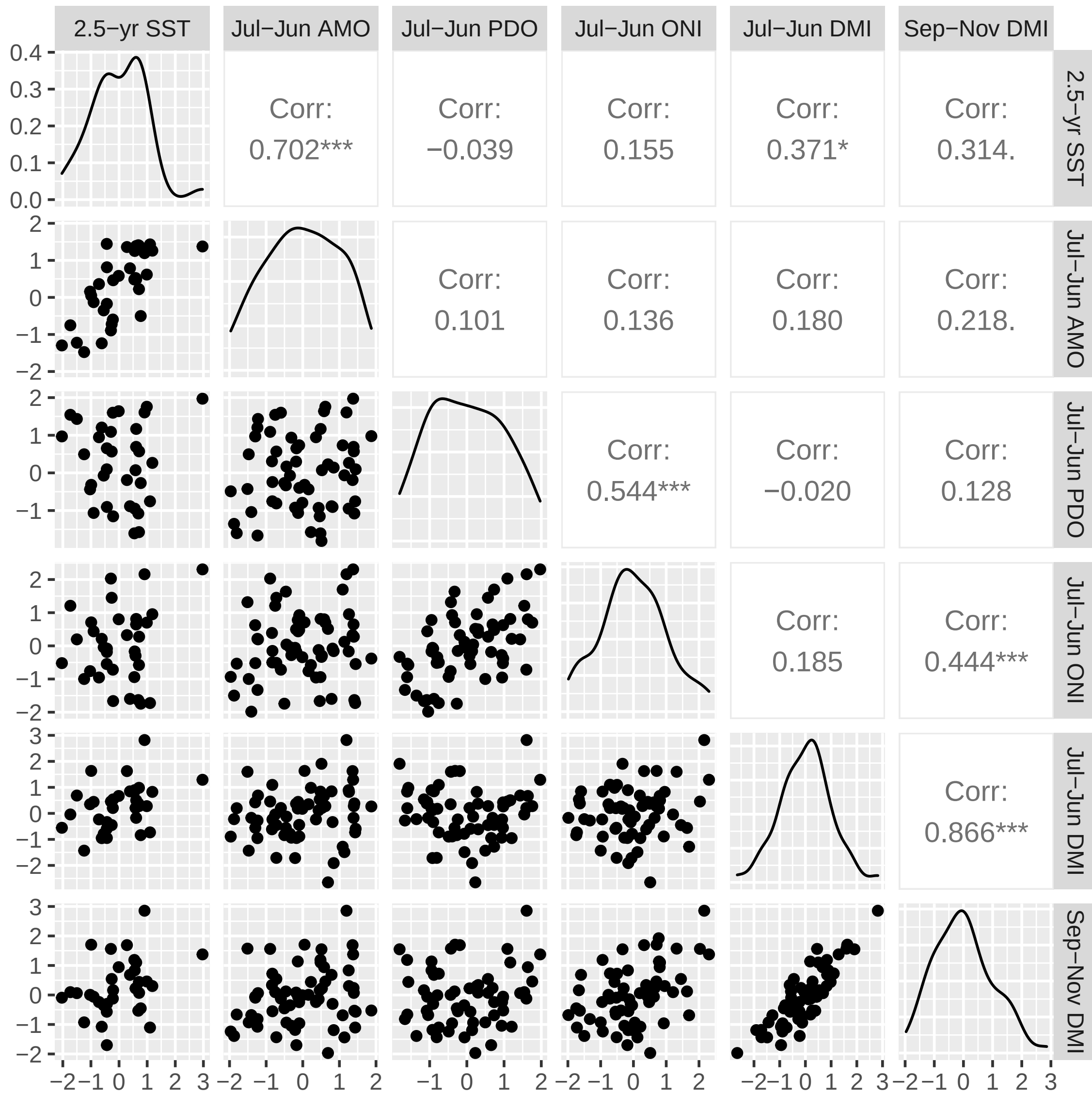


Figure S8. Multiyear average regional SST from AVHRR versus ocean climate indices (AMO, ONI, PDO and DMI) 1983-2016. The corvariates are those used in the analyses. Multi-year SST is the 2.5-year average of the regional (0-160km from coast) SST, so January $t-2$ to July t. The climate indices are 12-month average from July $t-1$ to June t. For the analyses, Sep-Nov average DMI in the prior year was used so that is also added.