Supplemental Information: Improving landings forecasts using environmental covariates: a case study on the Indian oil sardine (*Sardinella longiceps*)

EE Holmes, S BR, K Nimit, S Maity, DM Checkley, Jr., ML Wells, VL Trainer

Contents

Supplement 1: Full model tests and diagnostics

data version SardineForecast 1.11 (accepted paper used 1.11)

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. $ln(S_t) = \alpha + \beta_1 ln(W_{t-1}) + \beta_2 ln(S_{t-2}) + \epsilon_t$	30	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. $ln(S_t) = \alpha + \beta_t ln(W_{t-1}) + \epsilon_t$	28				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. $ln(W_t) = \alpha + \beta_1 ln(W_{t-1}) + \beta_2 ln(S_{t-2}) + \epsilon_t$	29	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. $ln(W_t) = \alpha + \beta_t ln(W_{t-1}) + \epsilon_t$	27				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	59.8	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.

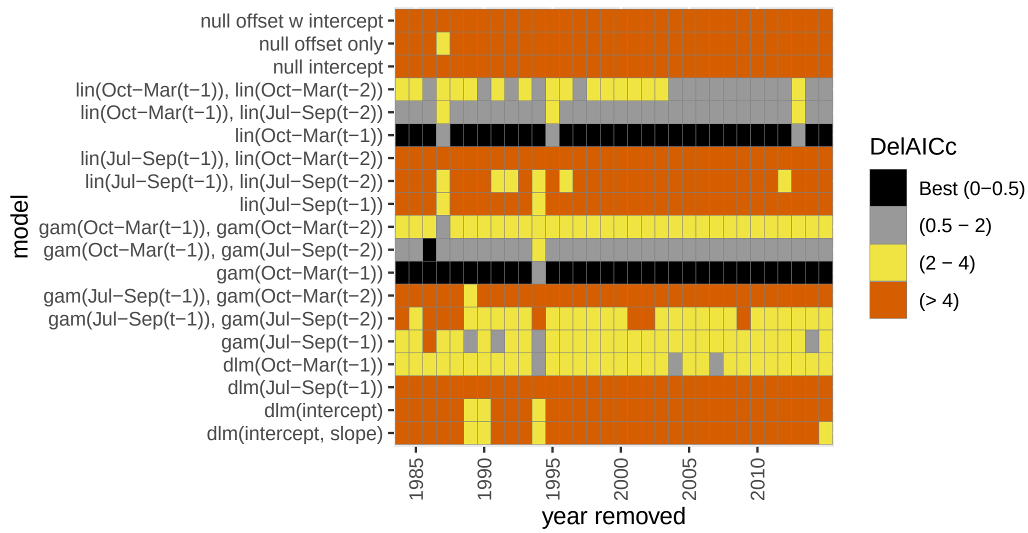


Figure S1. ΔAICc for the Jul-Sep landings base models with one year deleted using only the landings data that overlap with the environmental data 1984-2015.

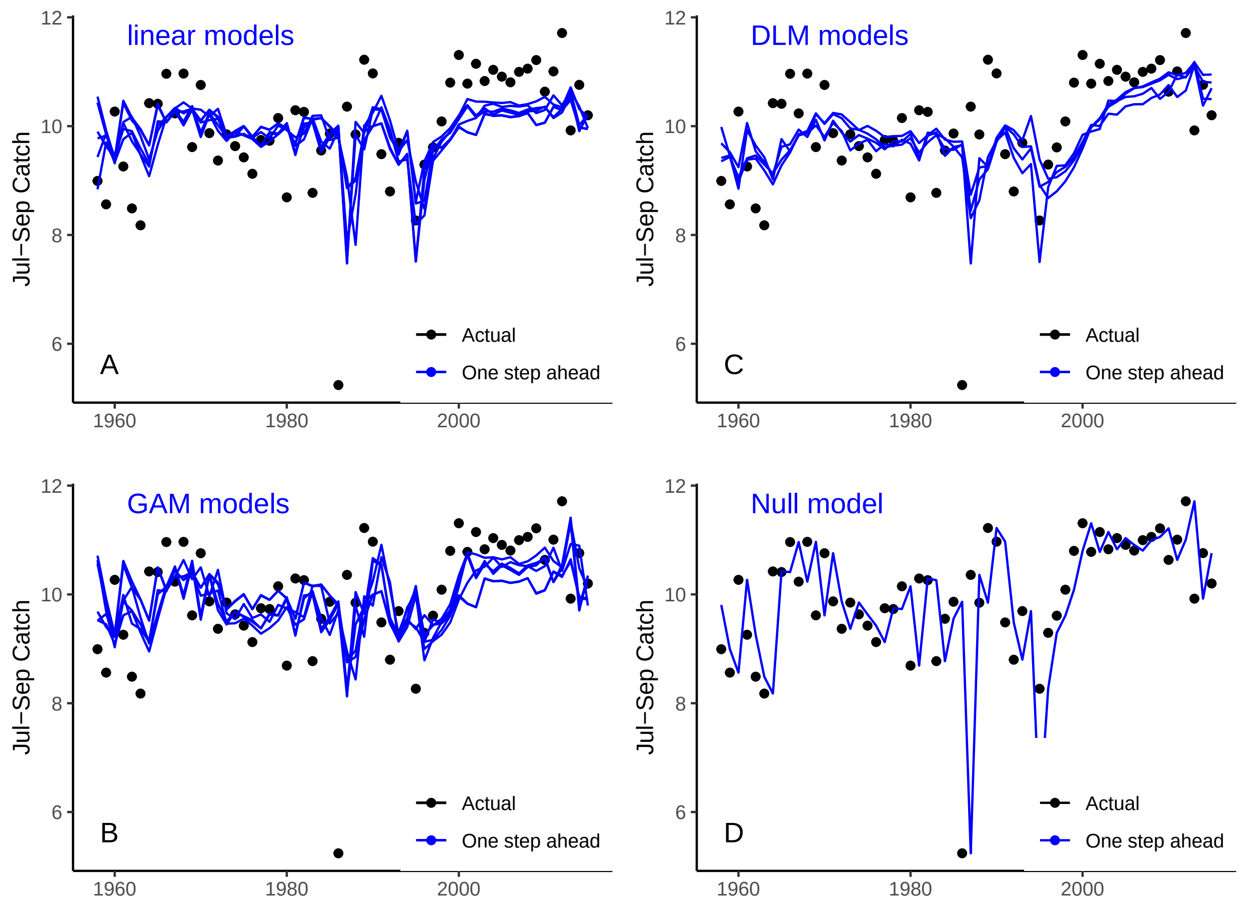


Figure S2. Leave one out (LOO) one step ahead predictions for the linear, GAM, and DLM models of Jul-Sep landings. The data point at year t on the x-axis is predicted from the data up to year t-1.

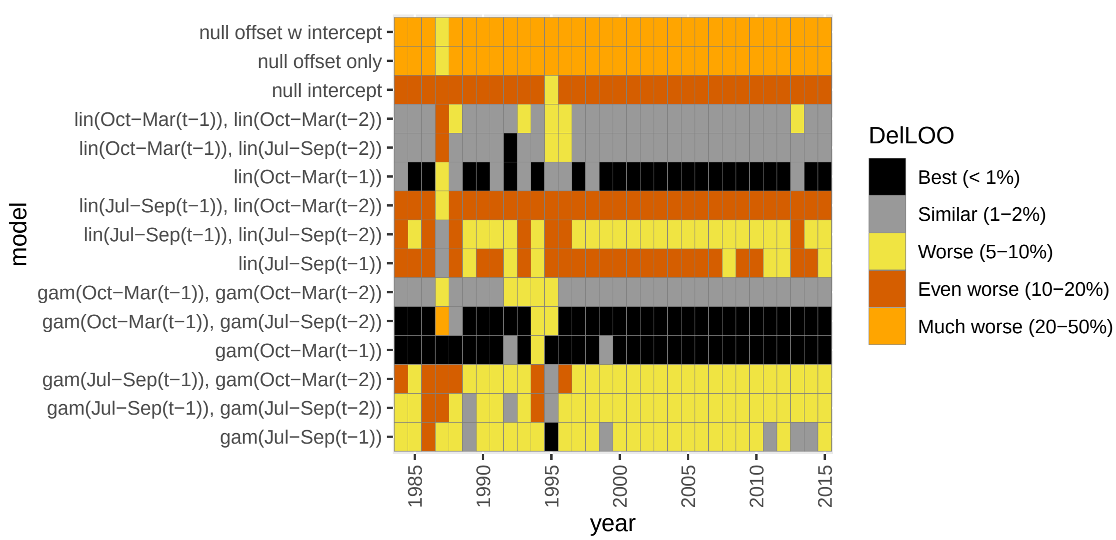


Figure S3. Leave-one-out predictive performance (leave out a year, fit, predict that year) for the Jul-Sep landings base models. The performance (DelLOO) is the RSME (root mean square error) between prediction and observed.

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).

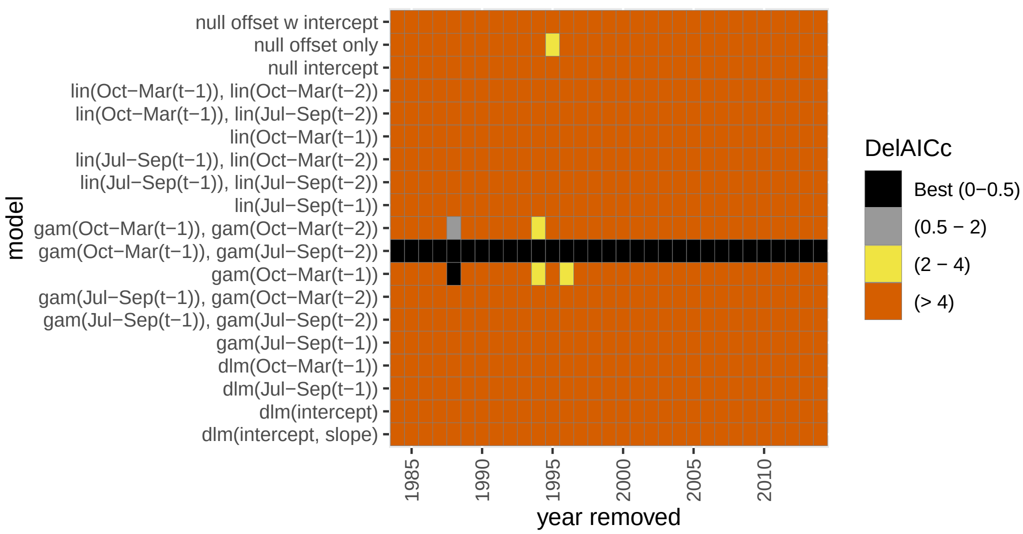


Figure S4. ∆AICc for the Oct-Mar landings base models with one year deleted using only the landings data that overlap with the environmental data 1984-2015. See Figure S1 for an explantion of the figure.

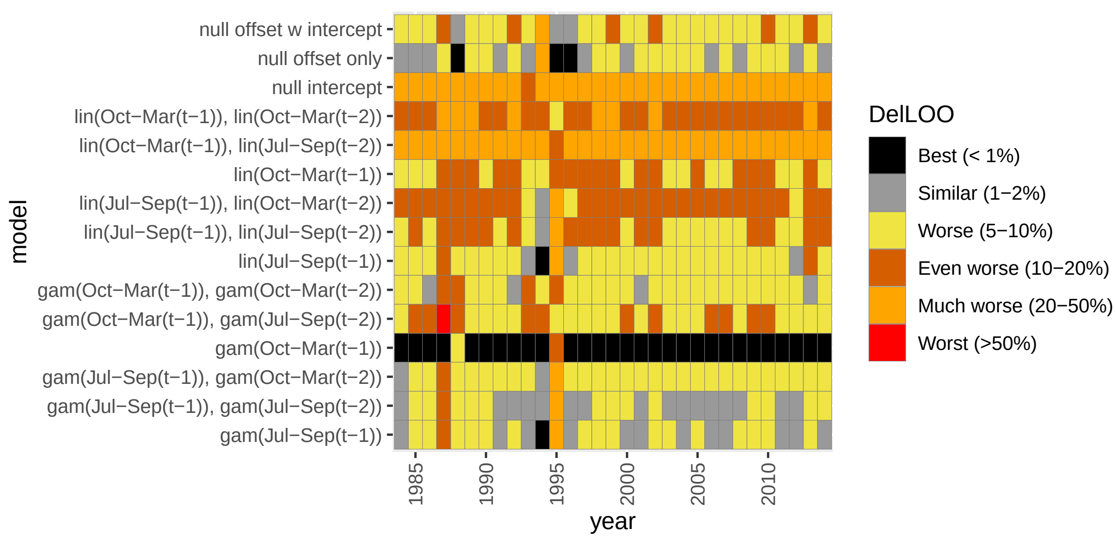


Figure S5. Leave-one-out predictive performance (leave out a year, fit, predict that year) for the Oct-Mar landings base models. The performance (DelLOO) is the RSME (root mean square error).

Comparison of land and oceanic rainfall measurements

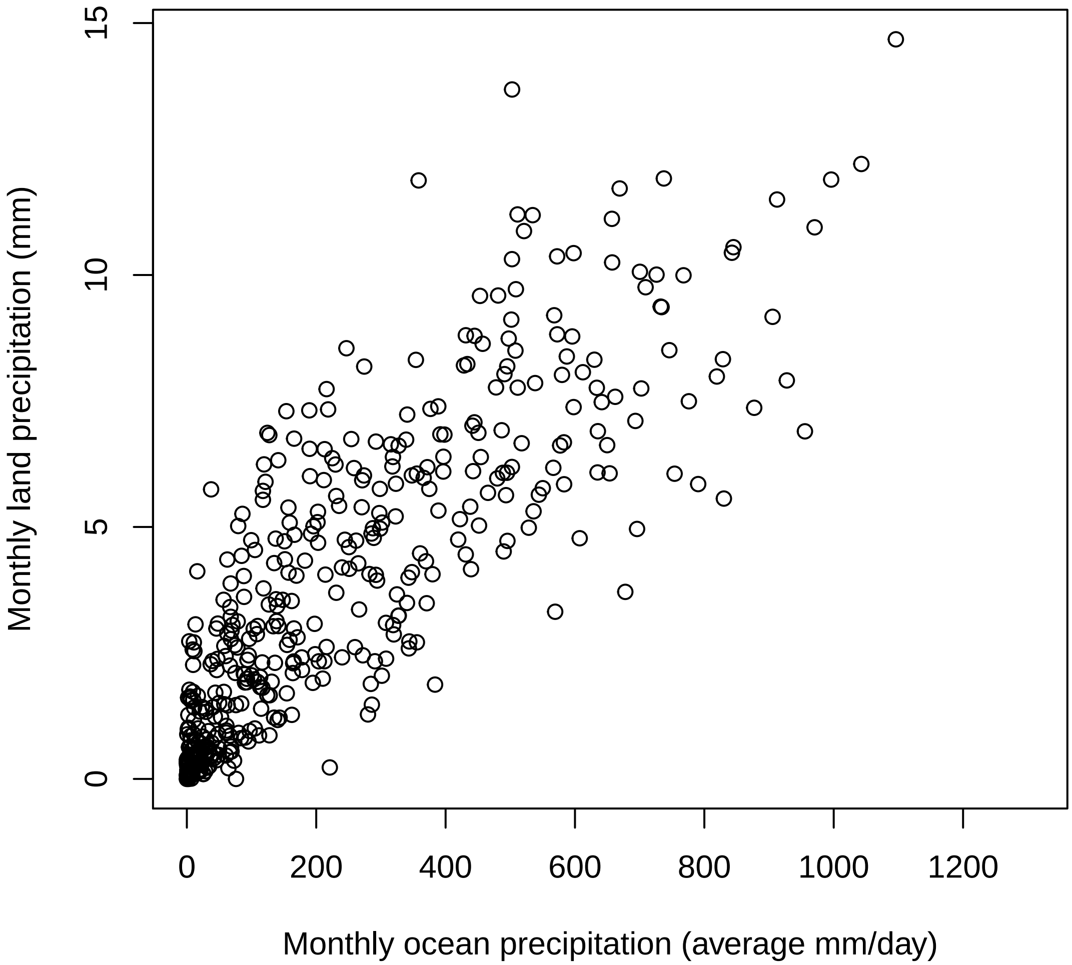


Figure S6. Monthly precipitation measured over land via land gauges versus the precipitation measured via remote sensing over the ocean.

Comparison of multiyear average regional SST from AVHRR and ICOADS

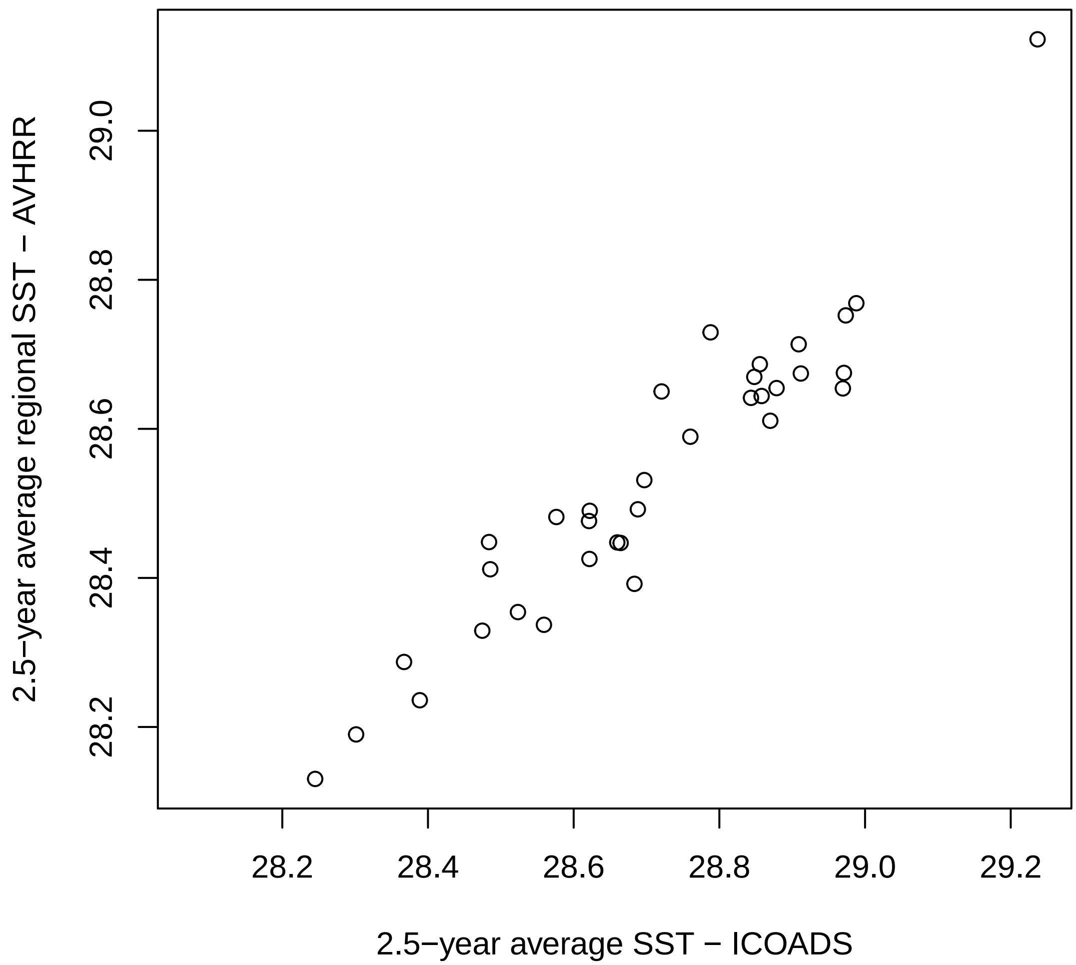


Figure S7. Multiyear average regional SST from the AVHRR versus ICOADS data sets.

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
}
```

```r
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

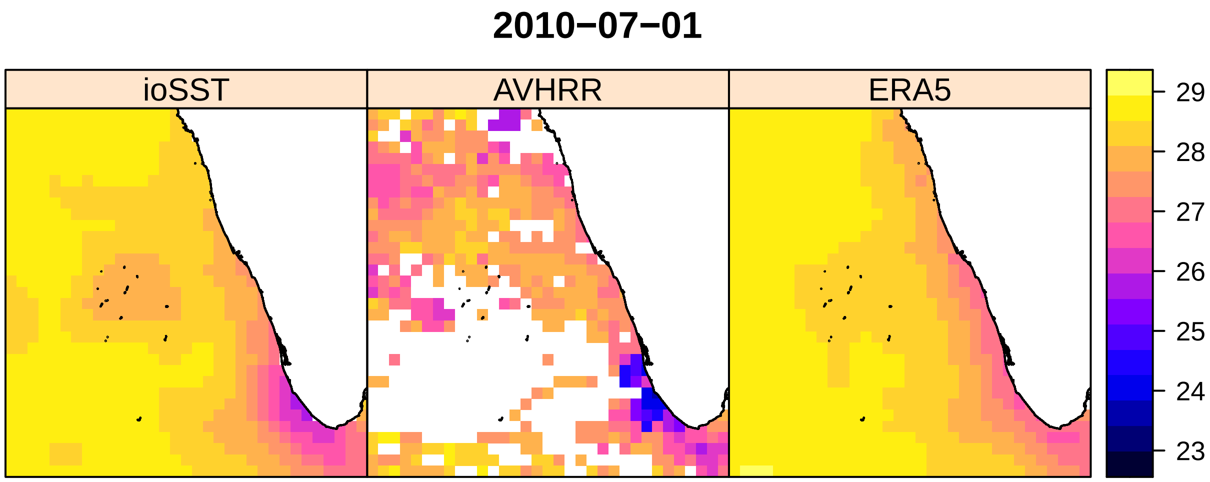
```r
sst3 <- getdata("hawaii_soest_d124_2bb9_c935", date=dates, lat=lats, lon=lons,
                    pars="sst", eserver="http://apdrc.soest.hawaii.edu/erddap")
```

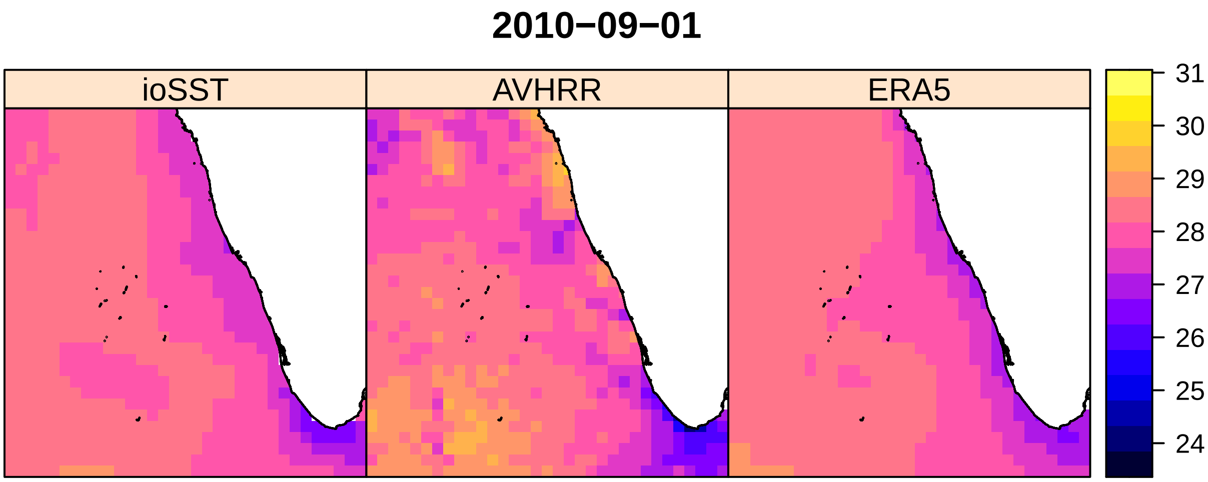
```
## data read from hawaii_soest_d124_2bb9_c935-7-15-70-78-2010-01-01-2011-01-01.csv
## data hawaii_soest_d124_2bb9_c935 date 2010-01-01-2011-01-01, latitude 7-15, longitude 70-78
```

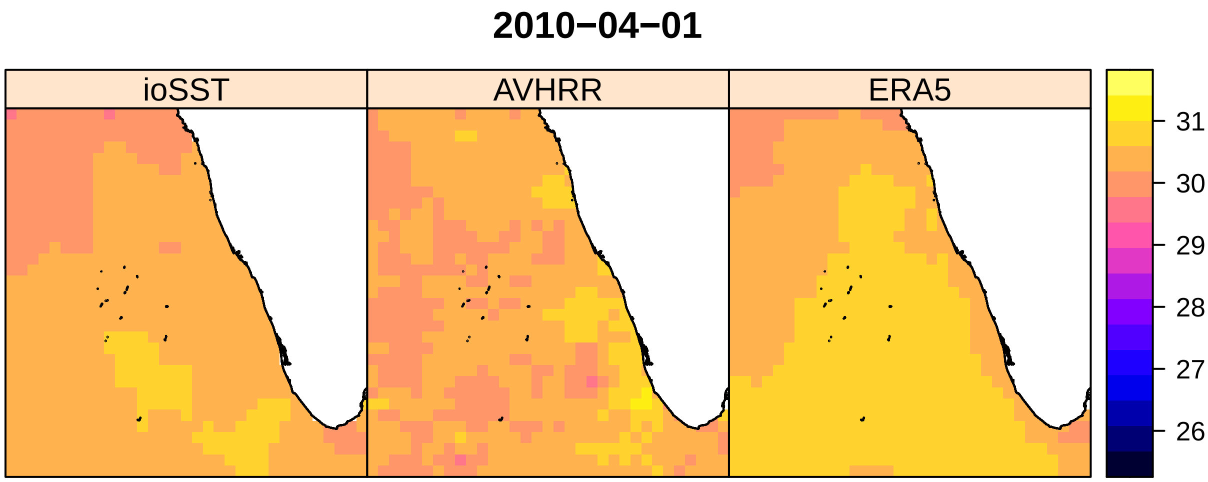
```r
sst3$date <- format(sst3$time, "%Y-%m-%d")
sst3$month <- format(sst3$time, "%m")
sst3$year <- format(sst3$time, "%Y")
sst3$sst <- sst3$sst-273.15
sst3$latitude <- sst3$latitude - 0.125
sst3$longitude <- sst3$longitude - 0.125
sst3.mon <- sst3 %>% group_by(year, month, latitude, longitude) %>%
  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



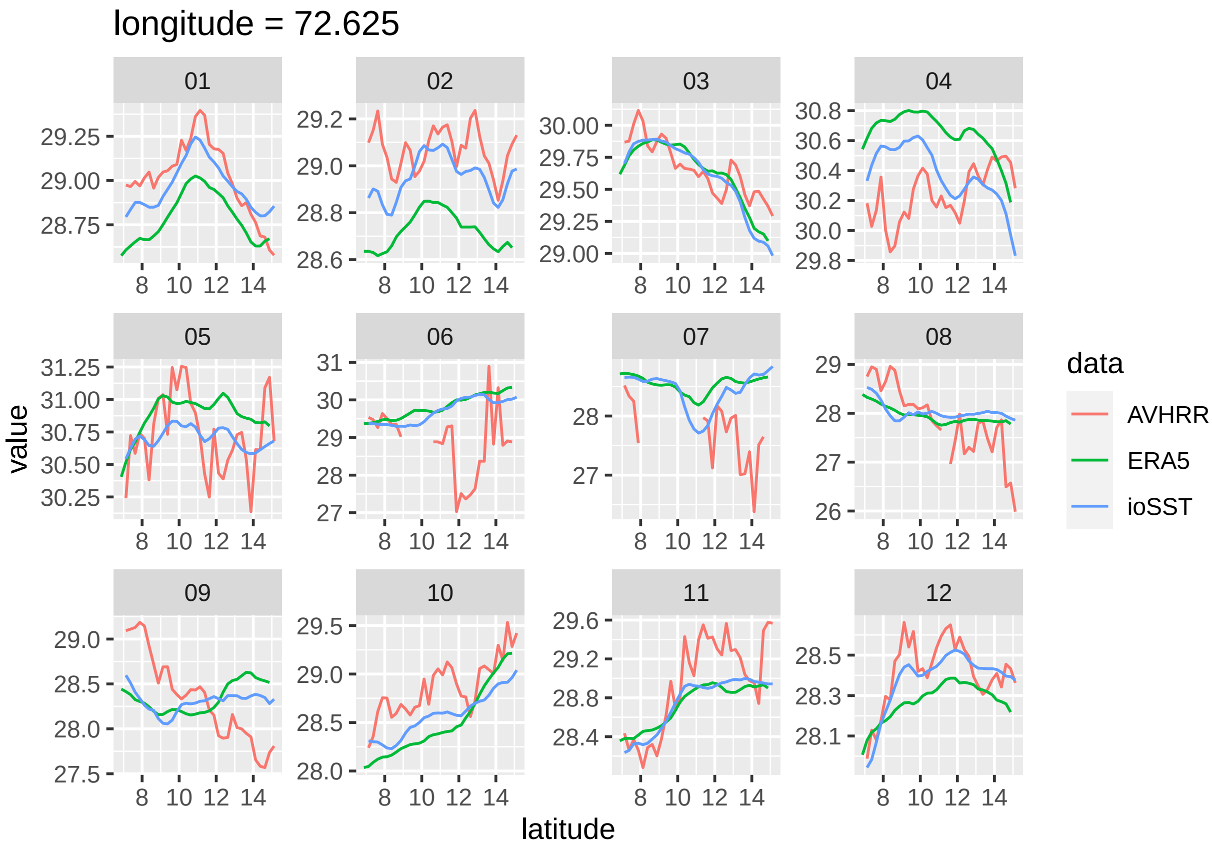
SST along one longitude line

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")
```

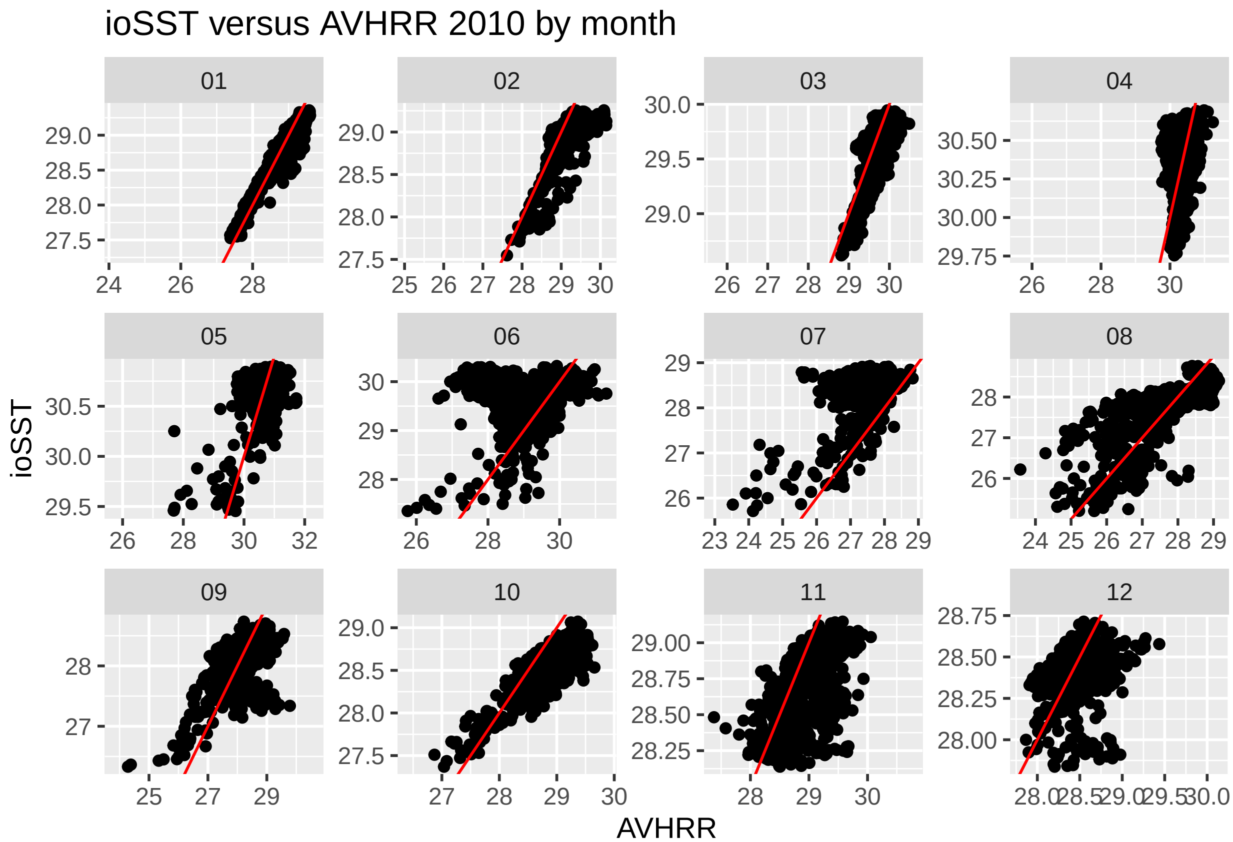
The SST from AVHRR can be quite different from ioSST and ERA5, e.g. Sept 2010. Here the SST along one longitude line is shown for a specific day.

```
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]))
```

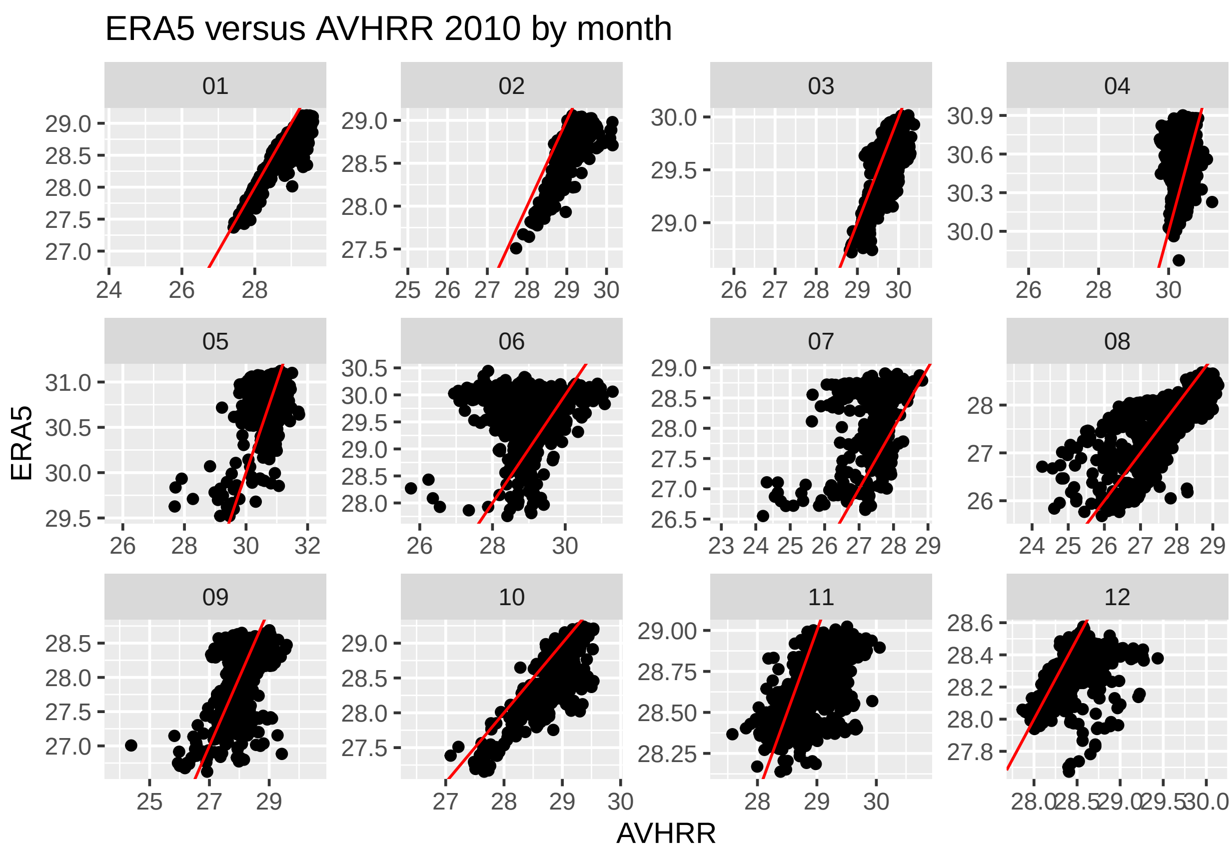


Pairwise comparison by month

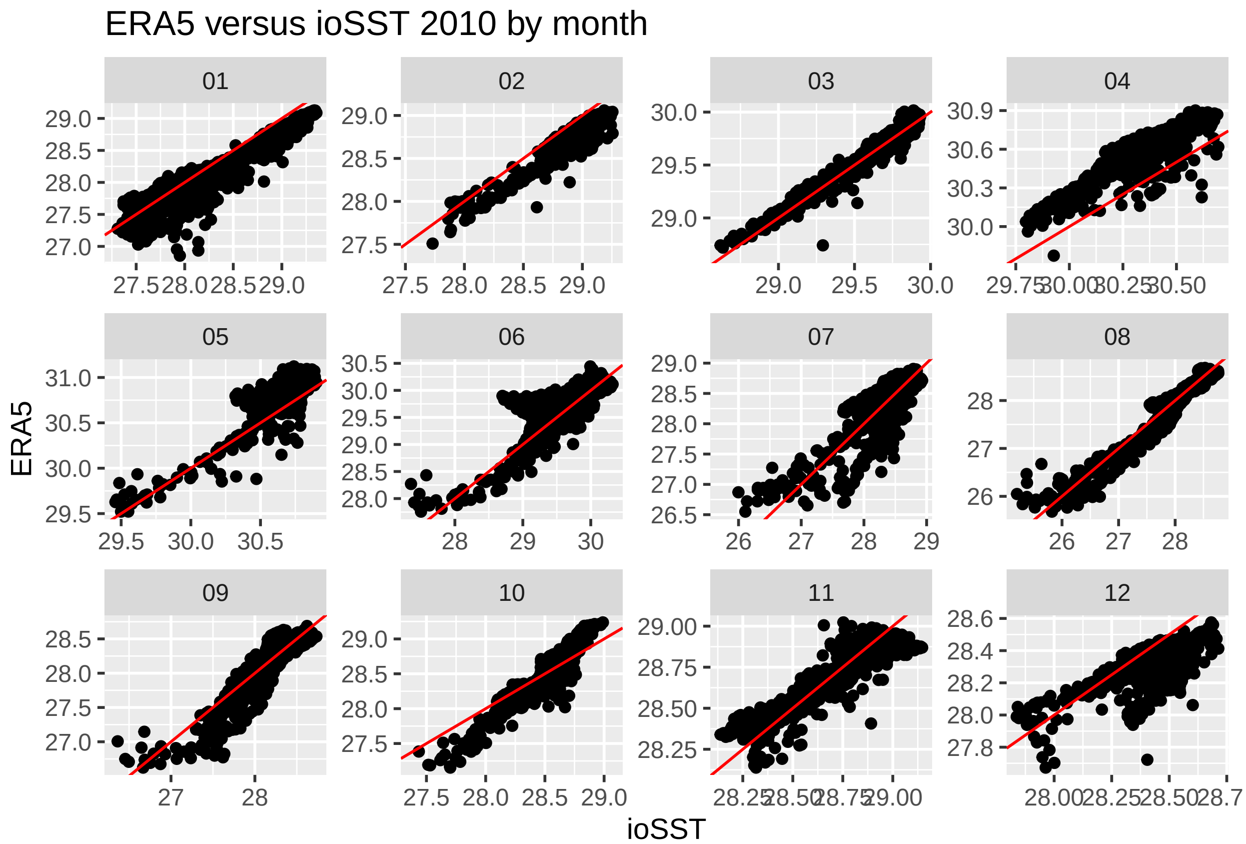
```
dfl2 <- dfl %>% pivot_wider(names_from=data, values_from="value")
dfl2$month <- format(as.Date(dfl2$date), "%m")
ggplot(dfl2, aes(x=AVHRR, y=ioSST)) + geom_point() + geom_abline(col="red") +
  facet_wrap(~month, scales="free") + ggtitle("ioSST versus AVHRR 2010 by month")
```




```
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")
```




```r
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

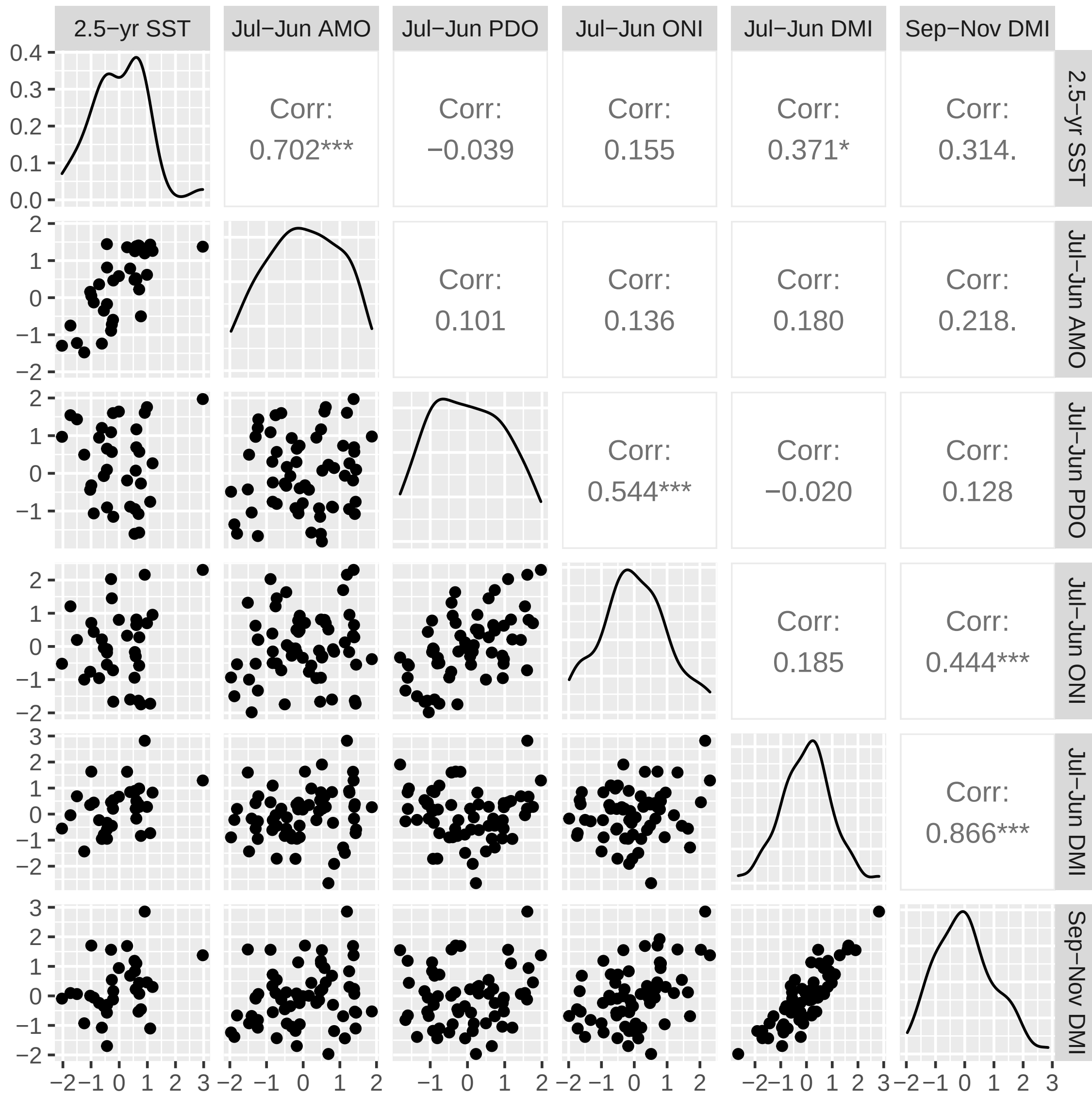


Figure S8. Multiyear average regional SST from AVHRR versus ocean climate indices (AMO, ONI, PDO and DMI) 1983-2016. The corvariates are those used in the analyses. Multi-year SST is the 2.5-year average of the regional (0-160km from coast) SST, so January $t-2$ to July t. The climate indices are 12-month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

Supplement 2: Data sources and raw data

data version SardineForecast 1.11 (accepted paper used 1.11)

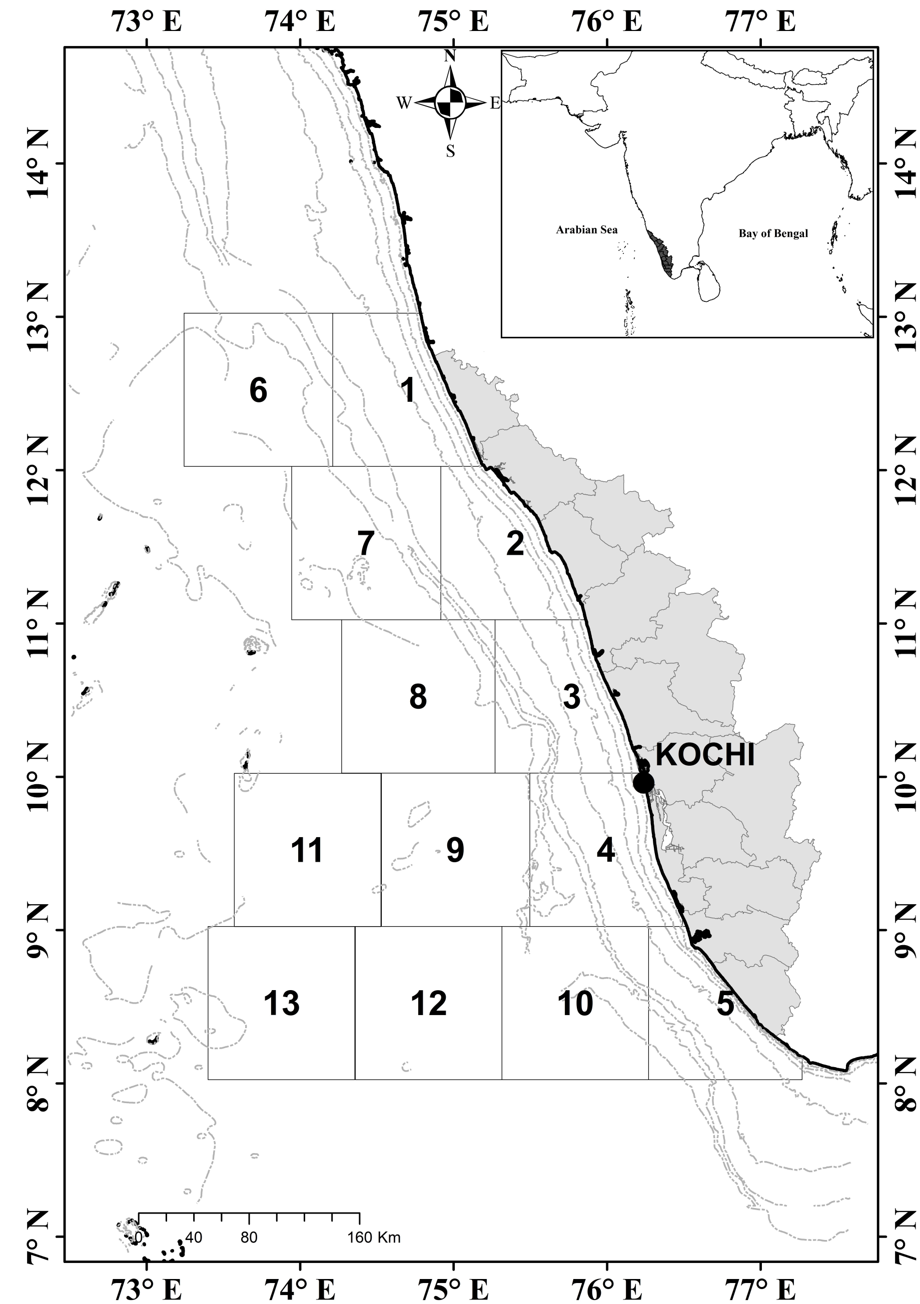


Figure S9. Study area with the boxes used for remote sensing variables, unless noted otherwise in the details. Specifically, the upwelling indices using winds (EMT and W_e) were based on a 0.25 degree grid and all values within 2 degrees latitude of the coast and between 8 and 13 degrees longitude N were averaged.

Landings Data

We used landings (in metric tons) of oil sardines in Kerala State 1956-2015. The data were collected and processed into a total landings estimate based on a stratified sampling of landing sites along the southwest coast of India throughout the year. The program is run by the Central Marine Fisheries Research Institute (CMFRI) in Cochin, India. We obtained the data from reports published by CMFRI; see references.

References

CMFRI reports were downloaded from the CMFRI Publication repository http://www.cmfri.org.in.

1956-1968 Antony Raja BT (1969). "Indian oil sardine." CMFRI Bulletin, 16, 1-142.

1968-1978 Pillai VN (1982). Physical characteristics of the coastal waters off the south-west coast of India with an attempt to study the possible relationship with sardine, mackerel and anchovy fisheries. Thesis, University of Cochin.

1975-1984 Kerala Jacob T, Rajendran V, Pillai PKM, Andrews J, Satyavan UK (1987). "An appraisal of the marine fisheries in Kerala." Central Marine Fisheries Research Institute. Report.

1975-1984 Kernataka Kurup KN, Nair GKK, Annam VP, Kant A, Beena MR, Kambadkar L (1987). "An appraisal of the marine fisheries of Karnataka and Goa." Central Marine Fisheries Research Institute. Report.

1985-2015 Provided by CMFRI directly via a data request.

SST Data

We used two primary reanalysis SST data sets: ioSST and ICOADS SST. We also compared to a third non-reanalysis data set. Reanalysis means that the data set uses multiple data sources and may use interpolation to fill in missing values. See Supplement 1 for comparisons of the data sources on a grid.

AVHRR Data For the paper, we used the Daily Optimum Interpolation (OI), AVHRR Only, Version 2.1 data set (oiSST) by the Group for High Resolution Sea Surface Temperature (GHRSST). "AVHRR Only" is in contrast to the "AMSR+AVHRR Product (AMSR = Advanced Microwave Scanning Radiometer).

https://www.ncei.noaa.gov/metadata/geoportal/rest/metadata/item/gov.noaa.nodc:GHRSST-AVHRR_OI-NCEI-L4-GLOB/html

and downloaded from

https://coastwatch.pfeg.noaa.gov/erddap/griddap/ncdcOisst21Agg.html

This data set is on a 0.25 degree and provides complete ocean temperature fields constructed by combining bias-adjusted observations from different platforms (AVHRR satellite instruments, ships, buoys) on a regular global grid, with gaps filled in by interpolation.

The ioSST data were compared to non-reanalysis AVHRR data. For 1981 to 2003, we used the Pathfinder Version 5.2 (L3C) monthly day and night product on a 0.0417 degree grid. These SST data use the Advanced Very-High Resolution Radiometer (AVHRR) instrument on the Pathfinder satellites and were provided by the Group for High Resolution Sea Surface Temperature (GHRSST) and the US National Oceanographic Data Center. This project was supported in part by a grant from the NOAA Climate Data Record (CDR) Program for satellites. For 2004 to 2016, we used the NOAA CoastWatch sea surface temperature (SST) products derived from NOAA's Polar Operational Environmental Satellites (POES). The SST estimates use the Advanced Very-High Resolution Radiometer (AVHRR) instruments on the POES satellites and are on a 0.1 degree grid.

ICOADS The International Comprehensive Ocean-Atmosphere Data Set (ICOADS) is a collection of surface marine data. SST data from 1960 onward were used, which are on a 1 degree grid. The nearshore data (boxes 1 to 5) are not as accurate as the AVHRR-based SST data. The ICOADS data were only used for the regional SST measurements not for nearshore or for the SST-differential based upwelling estimate.

These last two SST data sets were downloaded from the NOAA ERDDAP server:

https://coastwatch.pfeg.noaa.gov/erddap/info/ncdcOisst21Agg/index.html

https://coastwatch.pfeg.noaa.gov/erddap/info/erdAGsstamday/index.html

https://coastwatch.pfeg.noaa.gov/erddap/info/erdPH2sstamday/index.html.

https://coastwatch.pfeg.noaa.gov/erddap/info/esrlIcoads1ge/index.html

The SST values were averaged across the thirteen 1 degree by 1 degree boxes which parallel the bathymetry (Figure S1).

References

The AVHRR data were provided by GHRSST and the US National Oceanographic Data Center. The data were downloaded from NOAA CoastWatch-West Coast Regional Node and Southwest Fisheries Science Center's Environmental Research Division.

The ICOADS data were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their web site at http://www.esrl.noaa.gov/psd/

Casey KS, Brandon TB, Cornillon P, Evans R (2010). "The past, present, and future of the AVHRR Pathfinder SST program." In Oceanography from space, 273–287. Springer.

Simons RA. 2019. ERDDAP. https://coastwatch.pfeg.noaa.gov/erddap . Monterey, CA: NOAA/NMFS/SWFSC/ERD.

Walton C, Pichel W, Sapper J, May D (1998). "The development and operational application of nonlinear algorithms for the measurement of sea surface temperatures with the NOAA polar-orbiting environmental satellites." Journal of Geophysical Research: Oceans, 103(C12), 27999–28012.

Huang, Boyin; Liu, Chunying; Banzon, Viva F.; Freeman, Eric; Graham, Garrett; Hankins, Bill; Smith, Thomas M.; Zhang, Huai-Min. (2020): NOAA 0.25-degree Daily Optimum Interpolation Sea Surface Temperature (OISST), Version 2.1. [indicate subset used]. NOAA National Centers for Environmental Information. https://doi.org/10.25921/RE9P-PT57. Accessed November 10, 2020.

Upwelling Data

Four upwelling indices were used: a SST nearshore offshore differential (degree Celcius), the nearshore SST (degree Celcius), the Ekman Mass Transport perpendicular to the coast (kg m^{-1} s^{-1}), and Ekman Pumping at the tip of India (m s^{-1}). The The SST data were downloaded from the NOAA CoastWatch ERDDAP server. See the SST data description above. The Ekman Mass Transport and Ekman Pumping were computed from the meridonal and zonal winds from the Reanalysis Data ERA5 monthly Wind velocities downloaded from the Asia Pacific Data-Research Center ERRDAP server. ERA5 uses both the QSAT and ASCAT scatterometer measurements. See Supplement 3 for comparisons of the wind products.

The SST differential upwelling indices (degree Celcius) were computed from the ioSST data set (see above in SST section) which is based on AVHRR and thus accurate close to the coast.The SST-based UPW index is the difference between the neashore box (1 to 5) and a box 3 degrees longitude offshore at the same latitude.

The Ekman Mass Transport (EMT) index of upwelling is based upon Ekman's theory of mass transport due to wind stress. The index is computed from the `ektrx` and `ektry`, which are the x- and y- components of Ekman Transport (kg m^{-1} s^{-1}) computed from the longitudinal and latitudinal wind speeds. The functions for computing EMT are below; we used a coast angle of 158 degrees for the India west coast near Kochi (74.5E 11.5N). The `ekrtrx` and `ekrtry` were computed for latitude 8 to 13 and up to 2 degrees longitude from the coast. The values were then averaged to give a coastal average.

We used monthly winds from the ERA5 Reanalysis product from the European Centre for Medium-Range Weather Forecasts (ECMWF). Using monthly winds (as opposed to daily winds) in the EMT calculation does lead to bias, however monthly winds have been used in many recent papers (see citations in the main text) on upwelling in this region, and our paper sought to test indices that have been previously used. We used the winds closest to the sea surface. See Supplement 3 for a full discussion of the EMT (and Ekman Pumping) calculations along with comparisons of different wind products.

The Reanalysis Data ERA5 monthly 3d Wind velocities were downloaded from

http://apdrc.soest.hawaii.edu/erddap/griddap/hawaii_soest_66d3_10d8_0f3c.html. Our function to compute the EMT perpendicular to the coast was:

```
getEMT <- function(u, v){
  dat <- list()
  pa <- 1.22 # kg/m3 air pressure
  omega <- 7.272205e-05 #(rad/s)
  f <- 2*omega*sin(pi*dat$latitude/180) #Coriolis parameter
  # Compute tau; get taux and tauy from u and v
  uv_mag <- sqrt(dat$u^2+dat$v^2) # wind speed
  tau  <- pa*Cd(uv_mag)*uv_mag*uv_mag # wind stress
  tauy <- pa*Cd(uv_mag)*uv_mag*dat$v
  taux <- pa*Cd(uv_mag)*uv_mag*dat$u
  EMT <- tau/f
  # MASS TRANSPORT IS PERPENDICULAR TO WIND
  # u/taux positive is west to east wind so into india w coast
  # wind blowing west to east means EMT toward equator (negative)
  EMTy <- -1*taux/f
  # v/tauy negative is wind blowing south toward equator
  # Negative v (south) means EMT to west (off-shore)
  # EMTx is the one that drives upwelling since coast is mostly n-s
  # Negative EMTx = directed offshore = positive upwelling
  EMTx <- tauy/f
  dat$uv_mag <- uv_mag
  dat$tau <- tau
  dat$taux <- taux
  dat$tauy <- tauy
  dat$EMT <- EMT
  dat$ektrx <- EMTx
  dat$ektry <- EMTy
  return(dat)
}

EMTperp <- function(ektrx, ektry, coast_angle) {
  pi <- 3.1415927
  degtorad <- pi/180.
  alpha <- (180 - coast_angle) * degtorad
  s1 <- cos(alpha)
  t1 <- sin(alpha)
  s2 <- -1 * t1
  t2 <- s1
  perp <- (s1 * ektrx) + (t1 * ektry)
  para <- (s2 * ektrx) + (t2 * ektry)
  return(list(perp=perp, para=para))
}

getEkmanPump <- function(dat, res){
  # dat is the ektrx and ektry for a grid
  # dat has latitude, longitude, ektrx and ektry
```

```
  # res is the grid resolution; 0.25 deg for the ERA5 wind data
  dat$Wey <- NA; dat$Wex <- NA; dat$We <- NA
  lons <- sort(unique(dat$longitude))
  lats <- sort(unique(dat$latitude))

  # constants
  # h is 1 degree lat or lon in meters
  h.meter <- function(lat){
    m_per_deg_lat <- 111132.954 - 559.822 * cos( 2 * pi * lat/180 ) +
                     1.175 * cos( 4 * pi * lat / 180)
   m_per_deg_lon <- 111132.954 * cos ( pi*lat/180 )
   return(list(lat=m_per_deg_lat, lon=m_per_deg_lon))
  }
  psw <- 1023.6 #kg/m3 density of sea water

  if(length(lats)>2)
    for(lat in (min(lats)+res):(max(lats)-res)){
      dEMTy.lat <- (dat$ektry[dat$latitude==(lat+res)] -
         dat$ektry[dat$latitude==(lat-res)])/
         (2*res*h.meter(lat)$lat)
      Wey <- dEMTy.lat/psw
      dat$Wey[dat$latitude==lat] <- Wey
    }
  if(length(lons)>2)
    for(lon in (min(lons)+res):(max(lons)-res)){
      dEMTx.lon <- (dat$ektrx[dat$longitude==(lon+res)] -
      dat$ektrx[dat$longitude==(lon-res)])/
         (2*res*h.meter(dat$latitude[dat$longitude==(lon+res)])$lon)
      Wex <- dEMTx.lon/psw
      dat$Wex[dat$longitude==lon] <- Wex
    }

  dat$We <- dat$Wex+dat$Wey

  return(dat)
}
```

References

SST data: These data were provided by GHRSST and the US National Oceanographic Data Center. This project was supported in part by a grant from the NOAA Climate Data Record (CDR) Program for satellites. The data were downloaded from NOAA CoastWatch-West Coast Regional Node and Southwest Fisheries Science Center's Environmental Research Division.

ERA5: Hersbach, H., Bell, B., Berrisford, P., et al. The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society. 2020; 146: 1999–2049. https://doi.org/10.1002/qj.3803

ERA5 website: https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5

Bakun A (1973). "Coastal upwelling indices, west coast of North America." US Department of Commerce, NOAA Technical Report NMFS SSRF-671.

Chamberlain S (2019). rerddap: General Purpose Client for 'ERDDAP' Servers. R package version 0.6.5, https://CRAN.R-project.org/package=rerddap.

Mendelssohn R (2020). rerddapXtracto: Extracts Environmental Data from 'ERDDAP' Web Services. R package version 0.4.7, https://CRAN.R-project.org/package=rerddapXtracto.

Precipitation Data

We used three precipitation data sets off the southwest coast of India. Two are satellite derived and one is based on land gauges.

The National Climatic Data Center provides basic information on the Global Precipitation Climatology Project (GPCP) Precipitation data set. The data set consists of monthly precipitation estimates (average mm/day) from January 1979 to the present. The precipitation estimates merge several satellite and in situ sources into a final product. Data are provided on a 2.5 degree grid. The GPCP Precipitation data are provided by the NOAA/NCEI Global Precipitation Climatology Project and were downloaded from https://www.ncei.noaa.gov/data/global-precipitation-climatology-project-gpcp-monthly. Two boxes were defined, one off the Kerala coast and one off the Karnataka coast, and the average values of all grid points within these boxes were used. The boxes are Kerala Lat(8.75, 11.25), Lon(73.25, 75.75) Karnataka Lat(13.75, 16.25), Lon(71.25, 73.75)

The land gauge data set is a monthly rainfall (in mm) area weighted average for each state in India starting from 1901 onwards based on rain gauges. The data are provided by the India Meteorological Department (Ministry of Earth Sciences). The 1901 to 2014 data were downloaded from the Open Government Data Platform India https://data.gov.in. The 2015 and 2016 data were extracted from the yearly Rainfall Statistics reports (see references).

NASA's Tropical Rainfall Measuring Mission (TRMM) website provides background on the TRMM precipitation data (https://pmm.nasa.gov/). 1997 to 2015 monthly precipitation estimates on a 0.25 degree grid were downloaded from the Tropical Rainfall Measuring Mission (TRMM) website. The data were averaged in the 2.5 x 2.5 degree boxes 1 to 13 used for the other satellite data.

References

Adler R, Huffman G, Chang A, Ferraro R, Xie P, Janowiak J, Rudolf B, Schneider U, Curtis S, Bolvin D, Gruber A, Susskind J, Arkin P (2003). "The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-Present)." Journal of Hydrometeorology, 4, 1147-1167.

Adler R, Wang J, Sapiano M, Huffman G, Chiu L, Xie PP, Ferraro R, Schneider U, Becker A, Bolvin D, Nelkin E, Gu G, Program NC (2016). "Global Precipitation Climatology Project (GPCP) Climate Data Record (CDR), Version 2.3 (Monthly)." National Centers for Environmental Information. doi: 10.7289/V56971M6.

Purohit MK, Kaur S (2016). "Rainfall Statistics of India - 2016." India Meterorological Department (Ministry of Earth Sciences). http://hydro.imd.gov.in/hydrometweb/.

Kothawale DR, Rajeevan M (2017). Monthly, seasonal and annual rainfall time series for all-India, homogeneous regions and meteorological subdivisions: 1871-2016. Report, Indian Institute of Tropical Meteorology..

NCEI (2017). Global Precipitation Climatology Project Monthly Product Version 2.3. Retrieved from National Centers for Environmental Information website:
https://www.ncei.noaa.gov/data/global-precipitation-climatology-project-gpcp-monthly/access/.

Chlorophyll Data

We used Chlorophyll-a products developed by the Ocean Biology Processing Group in the Ocean Ecology Laboratory at the NASA Goddard Space Flight Center.

For 1997 to 2002, we used the Chlorophyll-a 2014.0 Reprocessing (R2014.0) product from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) on the Orbview-2 satellite. These data are on a 0.1 degree grid with units of mg m^{-3}. See reference below.

For 2003 to 2017, we used the MODIS-Aqua product on a 4km grid with units of mg m^{-3}. These CHL data are taken from measurements gathered by the Moderate Resolution Imaging Spectroradiometer (MODIS) on NASA's Aqua Spacecraft. See reference below.

Both CHL data sets were downloaded from the NOAA ERDDAP server:

https://coastwatch.pfeg.noaa.gov/erddap/info/erdSW1chlamday/index.html

https://coastwatch.pfeg.noaa.gov/erddap/info/erdMH1chlamday/index.html.

The CHL values were averaged across the thirteen 1 degree by 1 degree boxes which parallel the bathymetry (Figure S1).

References

NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group; (2014): SeaWiFS Ocean Color Data; NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group. https://dx.doi.org/10.5067/ORBVIEW-2/SEAWIFS_OC.2014.0

NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group. Moderate-resolution Imaging Spectroradiometer (MODIS) Aqua Chlorophyll Data; 2014 Reprocessing. NASA OB.DAAC, Greenbelt, MD, USA. https://dx.doi.org/10.5067/AQUA/MODIS/L3M/CHL/2014

Hu C, Lee Z, Franz B (2012). "Chlorophyll aalgorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference." Journal of Geophysical Research: Oceans, 117(C1).

Ocean Climate Indices

We used the following ocean climate indices: Oceanic Nino Index, Dipole Mode Index, Pacific Decadal Oscillation index and Atlantic Multidecadal Oscillation index.

The Oceanic Nino Index (ONI) is one of the primary indices used to monitor the El Nino-Southern Oscillation (ENSO). The ONI index is 3 month running mean of ERSST.v5 SST anomalies in the Niño 3.4 region (5°N-5°S, 120°-170°W)], based on centered 30-year base periods updated every 5 years.
The ONI was downloaded from the National Weather Service Climate Prediction Center

http://www.cpc.ncep.noaa.gov/data/indices/oni.ascii.txt

The DMI is the monthly Dipole Mode Index. The DMI is defined by the SSTA (SST anomaly) difference between the western Indian Ocean (10°S–10°N, 50°E–70°E) and the southeastern Indian Ocean (10°S–0°, 90°E–110°E). The data were downloaded from

https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Data/dmi.long.data The original data source is the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) via the page

http://www.jamstec.go.jp/frcgc/research/d1/iod/e/iod/about_iod.html.

The PDO is the Pacific Decadal Oscillation. The PDO index is defined by the sea surface temperature anomaly over the North Pacific Ocean (north of 20°N). The data were downloaded from the NOAA Physical Sciences Laboratory:

https://psl.noaa.gov/tmp/gcos_wgsp/data.143.131.2.6.325.11.4.55

The AMO is the Atlantic Multidecadal Oscillation. The AMO index is based on sea-surface temperature anomalies in the North Atlantic Ocean. The data were downloaded from the NOAA Physical Sciences Laboratory:

https://psl.noaa.gov/tmp/gcos_wgsp/data.143.131.2.6.325.11.25.33

References

Saji NH, Yamagata T (2003). "Possible impacts of Indian Ocean Dipole mode events on global climate." Climate Research, 25(2), 151-169.

Climate Prediction Center. El Nino-Southern Oscillation.

https://www.cpc.ncep.noaa.gov/products/precip/CWlink/MJO/enso.shtml

NOAA Physical Sciences Laboratory. Pacific Decadal Oscillation.

https://psl.noaa.gov/pdo/

van den Dool HM, Saha S, Johansson Å (2000). “Empirical orthogonal teleconnections.” Journal of Climate, 13, 1421–1435.

Mantua NJ, Hare SR, Zhang Y, Wallace JM, Francis RC (1997). “A Pacific interdecadal climate oscillation with impacts on salmon production.” Bulletin of the American Meteorology Society, 78, 1069-1079.

Enfield DB, Mestas-NuÃ±ez AM, Trimble PJ (2001). “The Atlantic Multidecadal Oscillation and its relation to rainfall and river flows in the continental U.S.” Geophysical Research Letters, 28, 2077–2080.

Raw Data

The data used in the paper were prepared from monthly data in the SardineForecast package. csv of those data can be found in the data folder in the repository. The file `setup.Rmd` has the code to make the data in this table.

Table S9. Landings data and environmental covariates, part 1. W=Oct-Mar Kerala landings. S=Jul-Sep Kerala landings. Values have been rounded to 2 digits. ns- = nearshore (0-80km) boxes 2 to 5 in Figure 1 (main text). r- = regional (0-160km) boxes 1 to 5 and 7 to 10. o-Precip is precipitation over the ocean; l-Precip is precipitation over land.

Year	W	S	Jun-Jul o-Precip	Apr-May o-Precip	Jun-Jul l-Precip	Apr-May l-Precip	Jun-Sep Bakun UPW	Jun-Sep coastal EMT	Jun-Sep We tip	Jun-Sep SST-diff UPW
1956	9.76	5.19	NA	NA	611.10	251.45	NA	NA	NA	NA
1957	12.27	9.81	NA	NA	853.65	225.70	NA	NA	NA	NA
1958	11.02	8.99	NA	NA	668.00	244.30	NA	NA	NA	NA
1959	10.15	8.56	NA	NA	1014.25	248.95	NA	NA	NA	NA
1960	12.27	10.27	NA	NA	615.60	373.30	NA	NA	NA	NA
1961	11.58	9.26	NA	NA	1075.85	297.30	NA	NA	NA	NA
1962	11.19	8.49	NA	NA	598.00	283.75	NA	NA	NA	NA
1963	10.66	8.18	NA	NA	556.75	126.70	NA	NA	NA	NA
1964	12.33	10.42	NA	NA	566.80	89.05	NA	NA	NA	NA
1965	11.61	10.41	NA	NA	531.40	162.15	NA	NA	NA	NA
1966	12.01	10.96	NA	NA	549.05	131.30	NA	NA	NA	NA
1967	12.06	10.23	NA	NA	641.55	157.50	931.11	NA	NA	NA
1968	12.06	10.97	NA	NA	1002.65	111.65	2216.34	NA	NA	NA
1969	11.66	9.62	NA	NA	684.65	172.30	1027.24	NA	NA	NA
1970	12.06	10.76	NA	NA	546.70	210.90	584.32	NA	NA	NA
1971	11.72	9.87	NA	NA	769.10	215.25	815.40	NA	NA	NA
1972	10.97	9.37	NA	NA	558.10	261.75	794.72	NA	NA	NA
1973	11.21	9.85	NA	NA	600.25	125.70	1067.60	NA	NA	NA
1974	11.40	9.63	NA	NA	635.55	174.75	998.95	NA	NA	NA
1975	10.94	9.43	NA	NA	697.85	143.00	959.12	NA	NA	NA
1976	11.38	9.12	NA	NA	419.15	105.15	1054.11	NA	NA	NA
1977	11.26	9.75	NA	NA	676.45	204.35	1270.40	NA	NA	NA
1978	11.54	9.73	NA	NA	722.40	235.35	1639.26	NA	NA	NA
1979	11.25	10.15	7.13	0.64	622.55	84.85	1682.78	-1087.51	9.83	NA
1980	10.72	8.69	7.28	2.97	749.95	110.05	1250.51	-1189.04	8.65	NA
1981	11.80	10.29	8.01	3.16	701.10	121.10	564.19	-1149.29	8.83	0.97
1982	11.63	10.26	7.92	2.01	561.85	104.30	262.33	-1133.71	9.92	1.03
1983	11.71	8.77	5.53	0.33	453.00	44.55	949.40	-1174.91	8.60	0.73
1984	11.39	9.55	8.25	3.98	748.10	123.35	1186.72	-1602.71	10.92	1.42
1985	10.76	9.86	6.83	4.50	608.80	160.40	1306.33	-1621.21	11.14	1.13
1986	9.19	5.24	7.05	2.46	461.35	94.90	1365.76	-1682.67	11.42	1.23
1987	9.43	10.36	4.53	2.30	396.80	82.75	1075.50	-1377.96	10.50	0.62
1988	11.24	9.85	12.44	3.25	507.05	167.40	453.01	-1257.44	10.71	0.65
1989	11.46	11.22	8.56	3.53	554.10	155.45	579.77	-1694.33	11.21	1.04
1990	11.50	10.97	5.94	3.29	582.00	265.15	653.88	-1783.02	10.51	0.89
1991	10.42	9.49	11.92	1.49	1000.80	105.20	574.93	-1491.14	9.70	0.57
1992	9.88	8.80	8.99	4.07	793.55	130.70	676.15	-1310.65	12.21	0.80
1993	9.80	9.69	9.30	2.97	716.60	112.75	637.71	-1488.76	11.14	0.76
1994	6.57	5.91	8.72	4.01	900.25	147.90	575.00	-1266.01	6.94	0.89
1995	9.67	8.27	6.69	1.96	597.95	245.25	701.69	-1366.14	11.85	0.90

Year	W	S	Jun-Jul o-Precip	Apr-May o-Precip	Jun-Jul l-Precip	Apr-May l-Precip	Jun-Sep Bakun UPW	Jun-Sep coastal EMT	Jun-Sep We tip	Jun-Sep SST-diff UPW
1996	10.64	9.29	7.67	2.16	634.20	99.30	621.31	-1641.69	12.44	0.70
1997	10.57	9.61	8.29	1.97	757.35	97.10	691.06	-946.29	7.02	0.54
1998	10.93	10.09	8.42	2.66	686.95	106.35	439.74	-1407.18	14.11	0.51
1999	11.45	10.80	7.42	5.22	653.85	282.40	453.43	-1271.77	9.17	0.94
2000	11.20	11.31	5.18	4.82	488.50	110.40	710.33	-1579.93	12.01	0.96
2001	11.11	10.78	9.07	7.39	648.30	238.25	661.91	-1608.59	9.98	0.95
2002	11.98	11.15	5.54	4.52	404.95	228.80	600.95	-1732.19	11.00	1.02
2003	11.88	10.83	10.15	1.87	518.10	108.55	627.16	-1350.33	9.19	0.97
2004	11.90	11.04	6.97	4.65	545.80	353.60	921.00	-1464.15	10.71	1.02
2005	11.82	10.91	6.72	4.73	756.00	170.10	956.08	-1496.50	10.98	1.29
2006	11.87	10.81	4.97	5.62	573.25	297.10	1029.44	-1120.20	7.43	0.74
2007	12.01	11.00	11.90	3.26	866.75	173.05	851.51	-1137.68	7.28	0.72
2008	11.51	11.06	8.74	2.55	493.35	91.15	685.70	-1076.27	7.83	0.45
2009	11.58	11.21	8.07	4.61	680.55	129.10	905.05	-1194.75	10.80	0.97
2010	12.27	10.63	10.02	3.86	649.55	165.05	373.33	-1340.18	10.30	1.02
2011	12.18	11.01	6.85	4.11	664.45	143.40	519.07	-1229.90	7.19	0.99
2012	12.44	11.71	5.78	1.94	402.60	140.65	666.39	-1290.93	8.62	0.91
2013	11.40	9.92	8.88	3.36	936.45	84.30	762.91	-1602.39	9.00	1.21
2014	11.30	10.76	5.05	3.31	566.10	173.35	864.24	-1414.09	10.39	0.90
2015	NA	10.20	8.32	3.62	687.95	174.65	581.45	-1067.67	9.41	0.71

Table S10. Landings data and environmental covariates, part 2.

	Year	Apr-May r-SST	Oct-Dec ns-SST	Jun-Sep ns-SST	Jul-Jun ONI	Jul-Jun PDO	Jul-Jun AMO	Sep-Nov DMI	2.5 yr ave r-SST ICOADS	2.5 yr ave r-SST AVHRR
26	1981	29.67	28.13	27.29	-0.15	0.87	-0.09	-0.43	28.68	NA
27	1982	29.19	29.02	27.20	0.08	0.19	-0.11	0.58	28.62	NA
28	1983	29.18	28.31	27.66	1.54	1.00	-0.20	-0.10	28.62	28.48
29	1984	29.34	27.87	26.24	-0.48	1.45	-0.14	-0.43	28.62	28.49
30	1985	29.29	28.15	26.49	-0.68	0.48	-0.31	-0.09	28.37	28.29
31	1986	29.18	27.88	26.55	-0.33	0.90	-0.27	-0.01	28.25	28.13
32	1987	29.56	29.18	27.97	0.94	1.40	-0.17	0.33	28.30	28.19
33	1988	29.73	28.63	27.49	0.64	1.49	0.08	-0.24	28.70	28.53
34	1989	29.36	28.32	26.74	-1.23	-0.20	-0.12	-0.20	28.86	28.69
35	1990	29.17	28.56	26.95	0.02	-0.02	-0.10	-0.08	28.62	28.43
36	1991	29.70	28.41	27.70	0.38	-0.90	-0.05	0.17	28.52	28.35
37	1992	29.46	28.15	26.94	1.12	0.54	-0.17	-0.49	28.58	28.48
38	1993	29.34	28.27	27.16	0.21	1.10	-0.26	-0.03	28.49	28.41
39	1994	29.18	28.37	27.02	0.20	1.30	-0.26	0.64	28.39	28.24
40	1995	29.54	28.44	27.22	0.57	-0.24	-0.02	-0.15	28.56	28.34
41	1996	29.25	28.08	27.13	-0.65	0.88	0.04	-0.74	28.68	28.39
42	1997	29.68	29.21	28.02	-0.08	0.62	-0.06	1.10	28.66	28.45
43	1998	30.30	28.27	27.78	1.64	1.46	0.19	-0.50	28.91	28.71
44	1999	28.89	28.03	27.18	-1.21	-0.63	0.24	0.01	28.97	28.75
45	2000	29.07	28.59	27.09	-1.17	-0.98	0.06	-0.05	28.69	28.49
46	2001	29.57	28.54	27.09	-0.51	-0.35	0.00	-0.18	28.47	28.33
47	2002	29.68	28.86	27.39	-0.01	-0.76	0.12	0.39	28.48	28.45
48	2003	29.93	28.49	27.74	0.65	1.07	0.06	-0.03	28.88	28.65
49	2004	29.27	28.71	26.94	0.26	0.54	0.23	0.04	28.97	28.68
50	2005	29.64	28.20	27.21	0.53	0.65	0.23	-0.28	28.97	28.65
51	2006	29.01	28.78	27.65	-0.35	0.12	0.24	0.63	28.66	28.45
52	2007	29.76	28.44	27.60	0.29	-0.13	0.22	0.22	28.76	28.59
53	2008	29.33	28.94	27.70	-1.12	-0.74	0.12	0.12	28.87	28.61
54	2009	29.69	28.68	27.11	-0.37	-1.35	0.01	0.07	28.91	28.67
55	2010	30.28	28.43	27.37	0.75	0.28	0.21	-0.27	28.99	28.77
56	2011	29.40	28.99	27.27	-1.15	-0.91	0.23	0.43	28.85	28.67
57	2012	29.10	29.15	27.59	-0.64	-1.38	0.06	0.28	28.84	28.64
58	2013	29.65	28.54	26.70	-0.07	-0.80	0.20	0.04	28.86	28.64
59	2014	29.94	28.76	27.60	-0.17	0.10	0.07	0.13	28.72	28.65
60	2015	29.97	29.32	28.33	0.57	1.59	0.09	0.50	28.79	28.73

Table S10. Landings data and environmental covariates, part 3.

	Year	Jul-Sep log CHL	Oct-Dec log CHL
42	1997	0.03	-0.85
43	1998	1.38	0.01
44	1999	1.41	0.08
45	2000	1.67	-0.49
46	2001	1.51	-0.06
47	2002	1.84	-0.27
48	2003	2.12	0.22
49	2004	2.49	-0.01
50	2005	1.75	0.70
51	2006	2.07	-0.14
52	2007	1.69	-0.28
53	2008	1.57	-0.43
54	2009	1.90	1.04
55	2010	1.92	0.38
56	2011	1.84	0.25
57	2012	1.52	-0.14
58	2013	1.71	0.04
59	2014	1.73	-0.54
60	2015	0.94	-0.63

Supplement 3: Ekman Mass Transport and Pumping Calculations

This report shows the Ekman Mass Transport (EMT) upwelling calculations and the Ekman Pumping calculations. The calculations are compared to values provided by the Environmental Research Division (ERD) at the Southwest Fisheries Science Center. Although we do not use the ERD upwelling indices in the paper, they are shown here to illustrate why we are not using them and to confirm that our EMT calculations are correct.

The second part of the report compares four different wind-products. The ERD products use pressure to estimate winds and this is known to be inaccurate near the equator. Instead, researchers studying winds off the coast of India use scatterometer instruments (ASCAT and QSCAT). In our paper, we use winds from the ERA5 Reanalysis product, which uses ASCAT and QSCAT data. This allows us to use one product for the entire time-range of our catch data. The wind and derived Ekman Mass Transport and Pumping values are compared using ASCAT data directly versus the ERA5 product.

Hersbach, H, Bell, B, Berrisford, P, et al. The ERA5 global reanalysis. Q J R Meteorol Soc. 2020; 146: 1999–2049. https://doi.org/10.1002/qj.3803

Load R libraries

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

Get the data

Get pressure, wind, and Ekman Transport from ERDDAP servers. This gets the data for the current month and has all the needed parameters. Current FNMOC grid resolution is 1 degree while older resolution was 2.5 degree and later interpolated to 1 degree.

Get data for latitude degrees 7 to 15N and longitude degrees 70 to 78E from https://coastwatch.pfeg.noaa.gov/erddap/griddap/erdlasFnTransMon. The `getdata()` function is at the end of this report. The downloaded data file `erdlasFnTransMon-7-15-70-78-2020-11-01-2020-11-02.csv` is also at the end of the report.

```
dates <- c("2020-11-01T00:00:00Z", "2020-11-02T06:00:00Z")
dat <- getdata("erdlasFnTransMon", date=dates)
```

```
## data read from erdlasFnTransMon-7-15-70-78-2020-11-01-2020-11-02.csv
## data erdlasFnTransMon date 2020-11-01T00:00:00Z-2020-11-02T06:00:00Z, latitude 7-15, longitude 70-78
```

```
cat(colnames(dat))
```

```
## time latitude longitude P_msl u v uv_mag taux tauy curl ektrx ektry
```

Calculations of wind from FNMOC pressure grid

Schwing, F. B., O'Farrell, M., Steger, J., and Baltz, K., 1996: Coastal Upwelling Indices, West Coast of North America 1946 - 1995, NOAA Technical Memorandum NMFS-SWFSC-231

Schwing et al 1996 gives the calculations for computing the wind vector from the pressures reported by FNMOC. First the pressure differential in the y (north-south) and x (east-west) directions is computed. y is the latitude center of the box. x is the longitude center of the box. P is in Pascals. Note the the pressure reported by FNMOC must be multiplied by 100 to get Pascals. h is the resolution of the grid in radians ($= \pi \times 1$ degree / 180 degree). Equation 1 in Schwing et al 1996:

$$\frac{\Delta P}{\Delta x} = (P_{x+h,y} - P_{x-h,y})/2h \qquad \frac{\Delta P}{\Delta y} = (P_{x,y+h} - P_{x,y-h})/2h \tag{1}$$

The east ($\vec{u}_g$) and north ($\vec{v}_g$) components of the geostrophic wind are (Equation 2 in Schwing et al 1996):

$$\vec{u}_g = -\frac{1}{f\,p_a\,R}\frac{\Delta P}{\Delta y} \tag{2}$$

$$\vec{v}_g = \frac{1}{f\,p_a\,R\,cos(\pi y/180)}\frac{\Delta P}{\Delta x} \tag{3}$$

In the equation, $p_a = 1.22\,kg/m^3$ and R is the radius of the earth in meters $= 6371000\,m$. f is the Coriolis parameters and is $2\Omega sin(\pi y/180)$, where $\Omega - 7.272205 \times 10^{-5}$ and y is the latitude of the center of the grid cell in degrees (and $\pi y/180$ is the latitude in radians). Positive $\vec{u}$ is wind blowing west to east. Positive $\vec{v}$ is wind blowing south to north (in the northern hemisphere). Notice the the east-west geostrophic wind is associated with the north-south pressure differential while the north-south wind is associated with the east-west differential.

On the last line of page 7 in Schwing et al 1996, they state "To approximate frictional effects, the geostrophic wind at the sea surface is estimated by rotating the geostrophic wind 15 deg to the left and reducing its magnitude by 30%". The actual angle that the wind is rotated in the numbers provided by ERD however appears to be 30 degrees. The following equation rotates and reduces the magnitude. α is the angle in radians ($\pi \times 30/180$).

$$\vec{u} = 0.7(cos(\alpha)\,u_g - sin(\alpha)\,v_g) \tag{4}$$

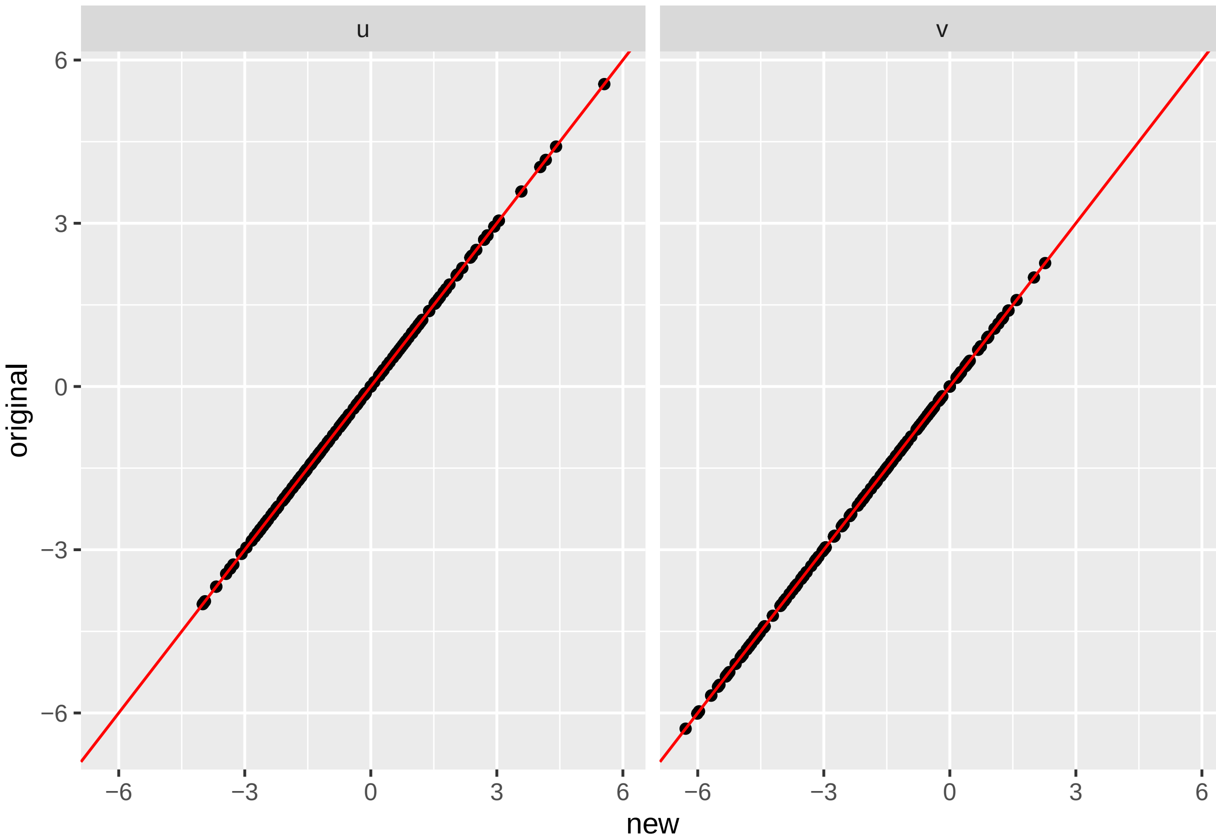
$$\vec{v} = 0.7(sin(\alpha)\,u_g + cos(\alpha)\,v_g) \tag{5}$$

The `getwindfromP()` function at the end of this report shows the R code for these calculations.

```
wind1 <- getwindfromP(dat)
```

Compare to wind numbers reported by ERD

Show that this is the same as what is reported by SWFSC-ERD.



Calculations of Ekman Mass Transport and Pumping

The following equations are in

Schwing, F. B., O'Farrell, M., Steger, J., and Baltz, K., 1996: Coastal Upwelling Indices, West Coast of North America 1946 - 1995, NOAA Technical Memorandum NMFS-SWFSC-231

The same equations and the non-linear C_d discussed below are also given in

Shafeeque et al. 2019: Effect of precipitation on chlorophyll-a in an upwelling dominated region along the West coast of India. Journal of Coastal Research, 86, 218-224. https://doi.org/10.2112/SI86-032.1

Wind stress is computed as

$$\vec{\tau} = p_a\, C_d\, |\vec{w}|\, \vec{w} \tag{6}$$

$$\tau_x = p_a\, C_d\, |\vec{w}|\, \vec{u} \qquad \tau_y = p_a\, C_d\, w\, \vec{v} \tag{7}$$

where C_d is the coefficient of drag and $|\vec{w}|$ is the wind speed (in m/s) and $\vec{w}$ is the wind vector. C_d is the empirical drag coefficient. It is a non-linear drag coefficient based on Large and Pond (1981) and modified for low wind speeds as in Trenberth et al. (1990), per discussion on the ERD upwelling page. This is

$$C_d = \begin{cases} 2.18 \times 10^-3, & \text{if } |\vec{w}| < 1 \\ (0.62 + \frac{1.56}{|\vec{w}|}) \times 10^-3 & \text{if } 1 < |\vec{w}| < 3 \\ 1.14 \times 10^-3 & \text{if } 3 \geq |\vec{w}| < 10 \\ (0.49 + 0.065|\vec{w}|) \times 10^-3 & \text{if } 10 \geq |\vec{w}| \end{cases} \tag{8}$$

Ekman Mass Transport (EMT kg/m s) is perpendicular (rotated 90 degrees clockwise) to wind stress. The EMT in the x and y directions is (Equation 4 in Schwing et al 1996):

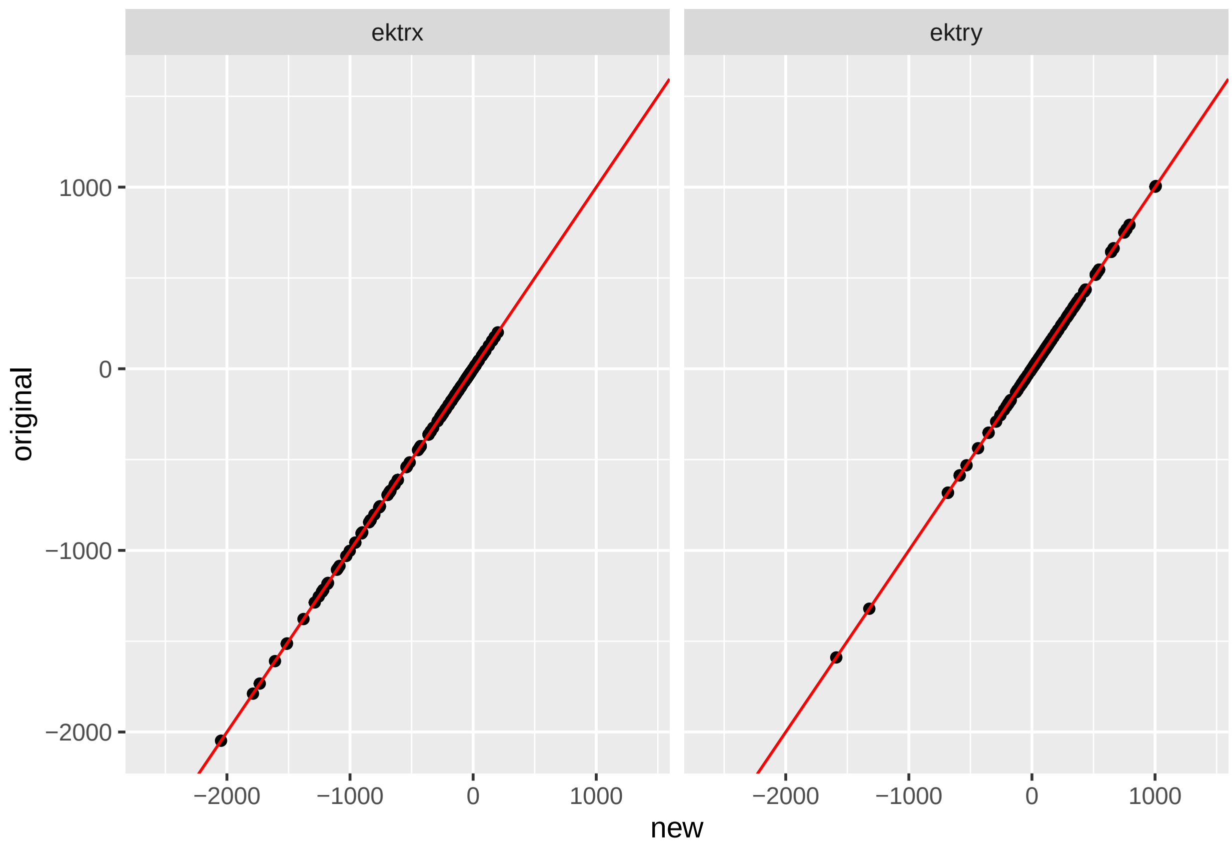
$$EMT_y = -\tau_x/f \qquad EMT_x = \tau_y/f \tag{9}$$

The `getEMT()` and `Cd()` functions shows how to compute these. `coast_angle` is added here for the upwelling index discussed below.

Compare EMT computed from wind to that reported by ERD

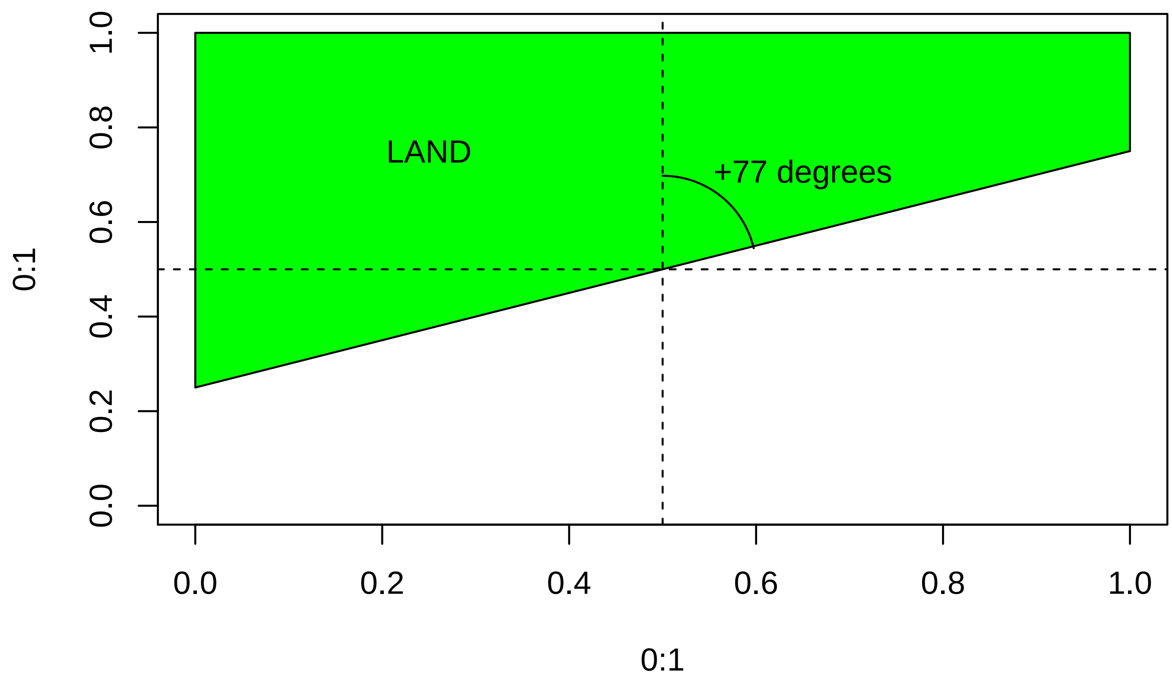
The values are the same, confirming that the equations are correct.

```
wind1 <- getEMT(wind1, coast_angle=158)
```



Upwelling index

ERD's Bakun upwelling index is derived from the EMT that is perpendicular and away from the coast. Their upwelling index is computed by finding the EMT vector that is perpendicular to the coast and defining a positive vector as away from the coast and a negative vector as toward the coast. ERD then divides this value by 10. The coast angle is defined as the degrees rotation clockwise away from north-south with land to the west (see figure). The coast angle for the SW coast of India is 158 degrees.



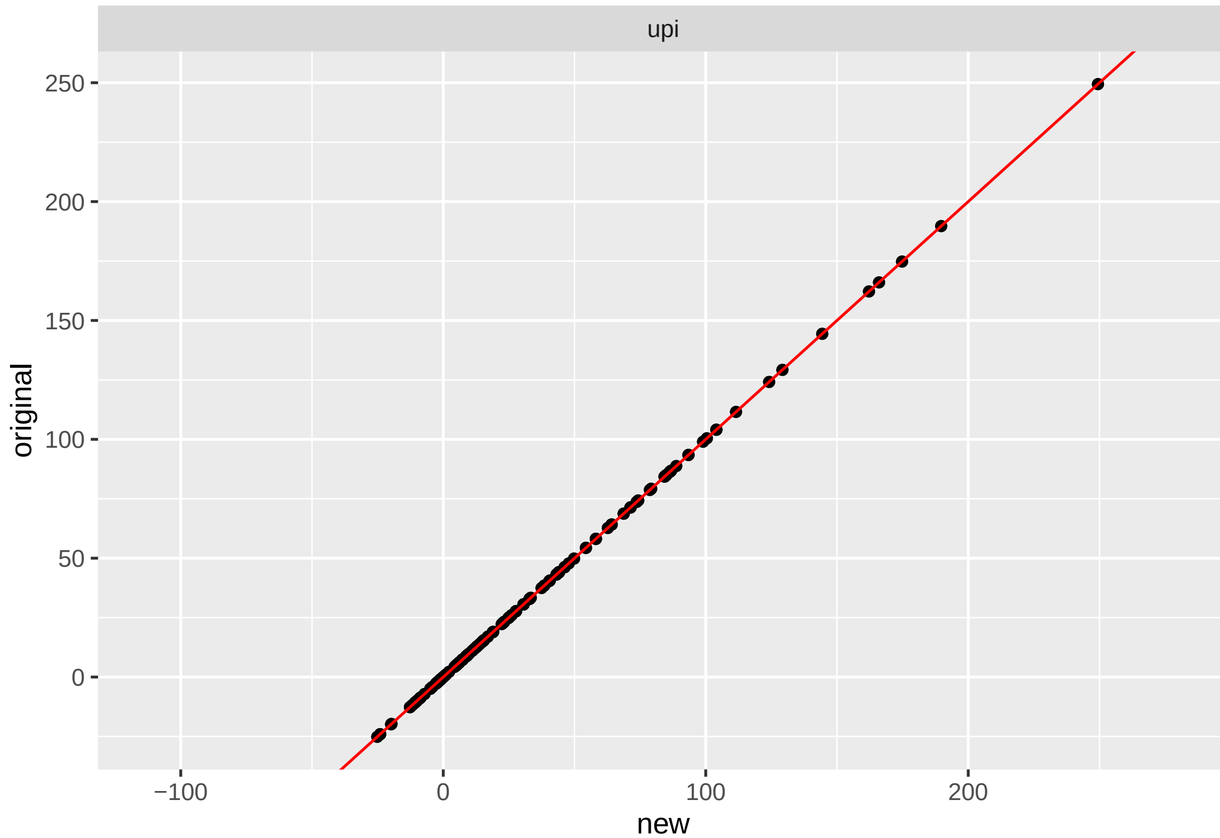
The equation to rotate the EMT vector is the following with the coast angle θ in degrees.

$$\alpha = (360 - \theta)\pi/180 \tag{10}$$

$$upi = \frac{EMT_x cos(\alpha) + EMT_y sin(\alpha)}{10} \tag{11}$$

The `upwell()` function shows how to compute this from the coast angle and the EMT in the x and y directions. Here the upwelling index using the data from the ERD EMT values is added to the data frame.

```
wind1$upi_orig <- upwell(wind1$ektrx_orig, wind1$ektry_orig, 158)
```



Ekman Pumping

$$\begin{aligned}\frac{\Delta EMT_x}{\Delta x} &= ((EMT_x)_{x+h,y} - (EMT_x)_{x-h,y})/2h_{m,lon} \\ \frac{\Delta EMT_y}{\Delta y} &= ((EMT_y)_{x,y+h} - (EMT_y)_{x,y-h})/2h_{m,lat}\end{aligned} \tag{12}$$

h_m is the resolution of the grid in meters (in the lon east-west and lat north-south directions). The resolution in meters varies by latitude. The following function returns the resolution in meters in the north-south and east-west directions as a function of latitude.

```
h.meter <- function(lat){
    m_per_deg_lat <- 111132.954 - 559.822 * cos( 2 * pi * lat/180 ) +
                     1.175 * cos( 4 * pi * lat / 180)
   m_per_deg_lon <- 111132.954 * cos ( pi*lat/180 )
   return(list(lat=m_per_deg_lat, lon=m_per_deg_lon))
  }
```

Ekman Pumping (m/s) is the sum of these divided by the density of sea water.

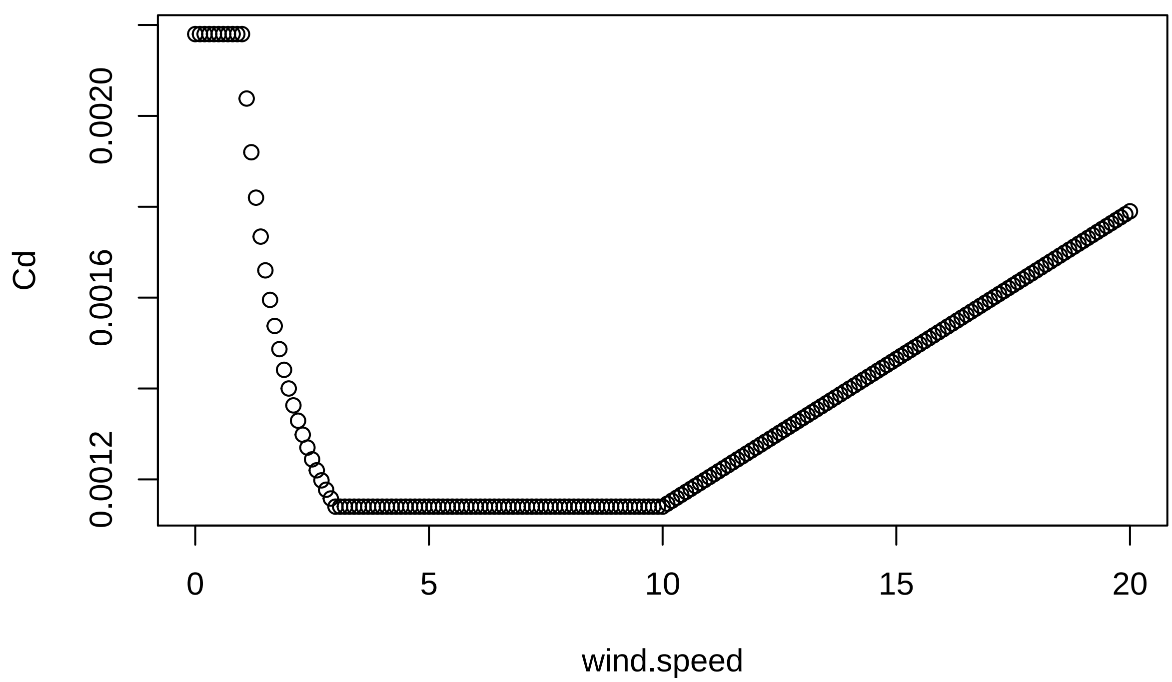
$$W_e = \frac{1}{p_{sw}}(\frac{\Delta EMT_x}{\Delta x} + \frac{\Delta EMT_y}{\Delta y}) \tag{13}$$

where p_{sw} is the density of sea water (assumed to be $1023.6kgm^{-3}$).

Comparison of wind stress from ASCAT winds versus ERD product

In the coastal upwelling products (EMT based) provided by NOAA ERD, a non-linear coefficient of drag is used:

```
plot(seq(0,20,0.1), Cd(seq(0,20,0.1)), xlab="wind.speed", ylab="Cd")
```



Comparing the τ computed from the equations above and that provided by ERD in their ASCAT wind stress product, it is clear that this non-linear C_d is not used. The wind stress products from ERD were not used for this reason.

Get data for a few days in 2010. The wind stress and Ekman upwelling (Pumping) computed by ERD is downloaded from https://coastwatch.pfeg.noaa.gov/erddap/info/erdQAstress1day/index.html and the winds are from https://coastwatch.pfeg.noaa.gov/erddap/info/erdQAwind1day/index.html.

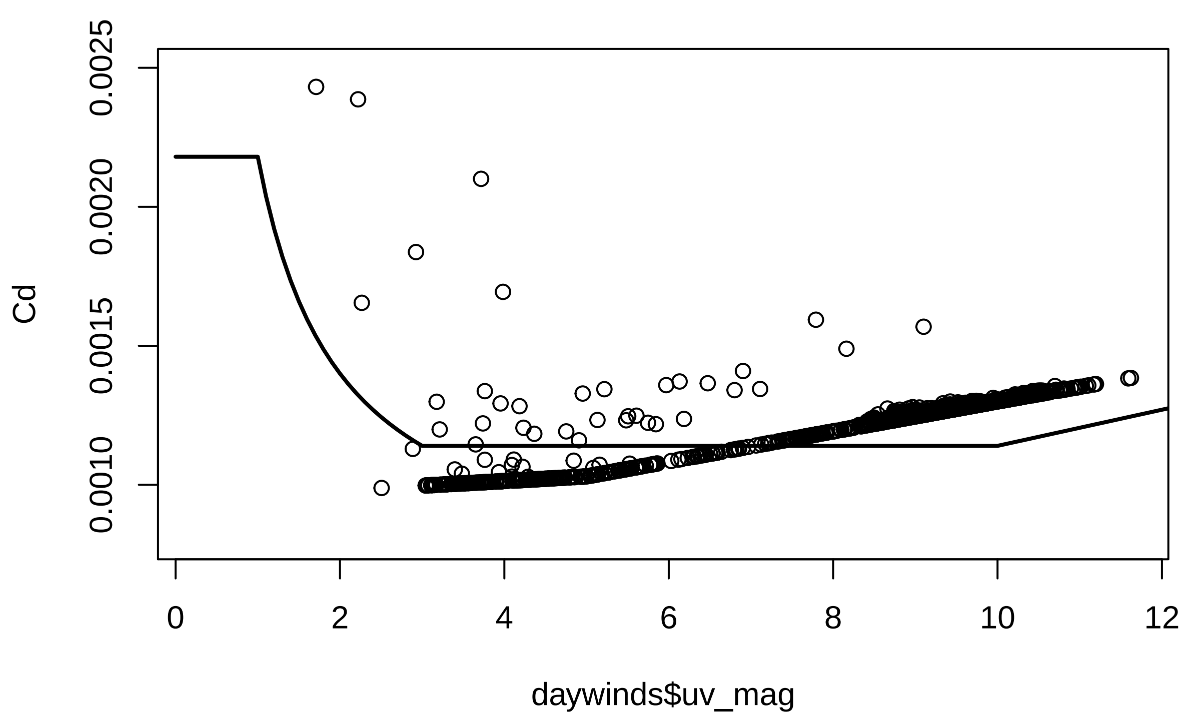
```
#10m winds
lats <- c(7, 15); lons <- c(70,78)
daywinds <- data.frame()
for(dates in c("2010-07-01", "2010-09-01", "2010-12-01")){
tmp <- getdata("erdQAwind1day", date=dates, lat=lats, lon=lons, altitude=10)
windmon5 <- getdata("erdQAstress1day", date=dates, lat=lats, lon=lons, altitude=0)
windmon5$u <- tmp$x_wind
windmon5$v <- tmp$y_wind
windmon5$uv_mag <- sqrt(tmp$y_wind^2 + tmp$x_wind^2)
windmon5$date <- format(windmon5$time, "%Y-%m-01")
windmon5$month <- format(as.Date(windmon5$date), "%m")
```

```
res <- attr(windmon5, "resolution")
windmon5 <- getEMT(windmon5, coast_angle=158)
daywinds <- rbind(daywinds, windmon5)
}
```

```
## data read from erdQAwind1day-7-15-70-78-2010-07-01-2010-07-01.csv
## data erdQAwind1day date 2010-07-01-2010-07-01, latitude 7-15, longitude 70-78
## data read from erdQAstress1day-7-15-70-78-2010-07-01-2010-07-01.csv
## data erdQAstress1day date 2010-07-01-2010-07-01, latitude 7-15, longitude 70-78
## data read from erdQAwind1day-7-15-70-78-2010-09-01-2010-09-01.csv
## data erdQAwind1day date 2010-09-01-2010-09-01, latitude 7-15, longitude 70-78
## data read from erdQAstress1day-7-15-70-78-2010-09-01-2010-09-01.csv
## data erdQAstress1day date 2010-09-01-2010-09-01, latitude 7-15, longitude 70-78
## data read from erdQAwind1day-7-15-70-78-2010-12-01-2010-12-01.csv
## data erdQAwind1day date 2010-12-01-2010-12-01, latitude 7-15, longitude 70-78
## data read from erdQAstress1day-7-15-70-78-2010-12-01-2010-12-01.csv
## data erdQAstress1day date 2010-12-01-2010-12-01, latitude 7-15, longitude 70-78
```

Dividing τ_x computed by ERD by $p_a = 1.22$, wind speed, and zonal wind speed should return C_d. Plotting this we see that the C_d function used is very different. The dots would fall on the solid black line if there were the same.

```
daywinds$a=daywinds$taux_orig/(1.22*daywinds$uv_mag*daywinds$u)
plot(daywinds$uv_mag, daywinds$a, ylim=c(0.0008, 0.0025), ylab="Cd")
lines(seq(0,20,0.1), Cd(seq(0,20,0.1)), lwd=2)
```



Comparison of monthly wind products

Get the FNMOC monthly winds. This is on a 1 degree grid. The data are from https://coastwatch.pfeg.noaa.gov/erddap/info/erdlasFnWPr/index.html.

```
dates <- c("2010-01-01", "2019-12-16")
windmon1 <- getdata("erdlasFnWPr", date=dates, lat=lats, lon=lons)
```

```
## data read from erdlasFnWPr-7-15-70-78-2010-01-01-2019-12-16.csv
## data erdlasFnWPr date 2010-01-01-2019-12-16, latitude 7-15, longitude 70-78
```

```
windmon1$date <- format(windmon1$time, "%Y-%m-01")
colnames(windmon1) <- stringr::str_replace(colnames(windmon1), "_mean", "")
windmon1 <- getEMT(windmon1, coast_angle=158)
```

Now get ASCAT monthly 10m winds https://coastwatch.pfeg.noaa.gov/erddap/info/erdQAwindmday/index.html. . This is on a 0.25 degree grid.

```
#10m winds
windmon2 <- getdata("erdQAwindmday", date=dates, lat=lats, lon=lons, altitude=10)
```

```
## data read from erdQAwindmday-7-15-70-78-2010-01-01-2019-12-16.csv
## data erdQAwindmday date 2010-01-01-2019-12-16, latitude 7-15, longitude 70-78
```

```
windmon2$u <- windmon2$x_wind
windmon2$v <- windmon2$y_wind
windmon2$date <- format(windmon2$time, "%Y-%m-01")
windmon2 <- getEMT(windmon2, coast_angle=158)
```

Now get the Reanalysis Data ERA5 monthly 10m winds from http://apdrc.soest.hawaii.edu/erddap/info/hawaii_soest_66d3_10d8_0f3c/index.html The altitude specification is in millibar. Air pressure at 10m is ca 1000 millibar, thus altitude is set to 1000.

```
windmon4 <- getdata("hawaii_soest_66d3_10d8_0f3c", date=dates, lat=lats, lon=lons,
                    altitude=1000, eserver="http://apdrc.soest.hawaii.edu/erddap",
                    alt.name="LEV")
```

```
## data read from hawaii_soest_66d3_10d8_0f3c-7-15-70-78-2010-01-01-2019-12-16.csv
## data hawaii_soest_66d3_10d8_0f3c date 2010-01-01-2019-12-16, latitude 7-15, longitude 70-78
```

```
windmon4$date <- format(windmon4$time, "%Y-%m-01")
windmon4 <- getEMT(windmon4, coast_angle=158)
```

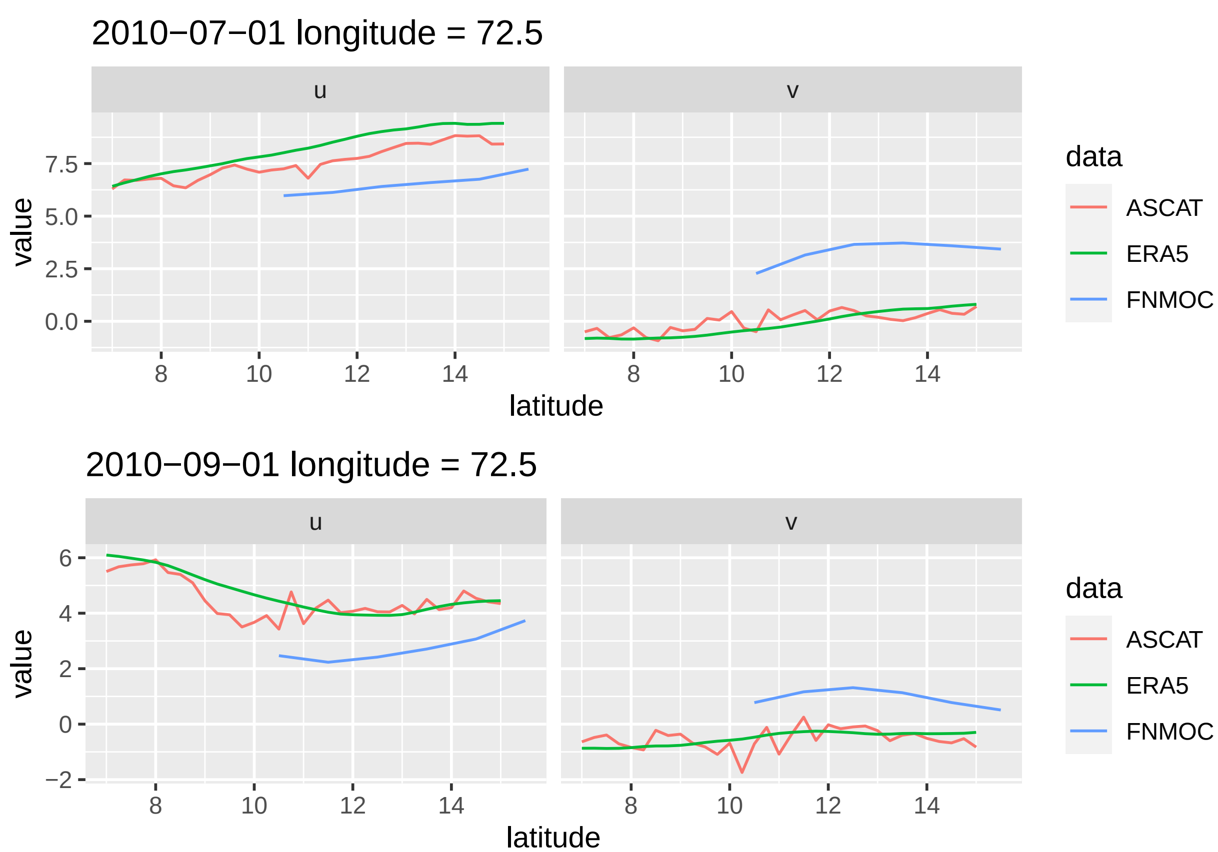
Create the data frame.

```
cols <- c("date", "latitude", "longitude", "u", "v", "EMTperp", "We")
df <- rbind(data.frame(windmon1[,cols], data="FNMOC"),
            data.frame(windmon2[,cols], data="ASCAT"),
            data.frame(windmon4[,cols], data="ERA5"))
dfl <- df %>%
  pivot_longer(!date & !latitude & !longitude & !data,
               names_to = "name",
               values_to = "value")
```

Winds on specific dates

FNMOC winds are quite different. ERA5 are much smoother than ASCAT winds.

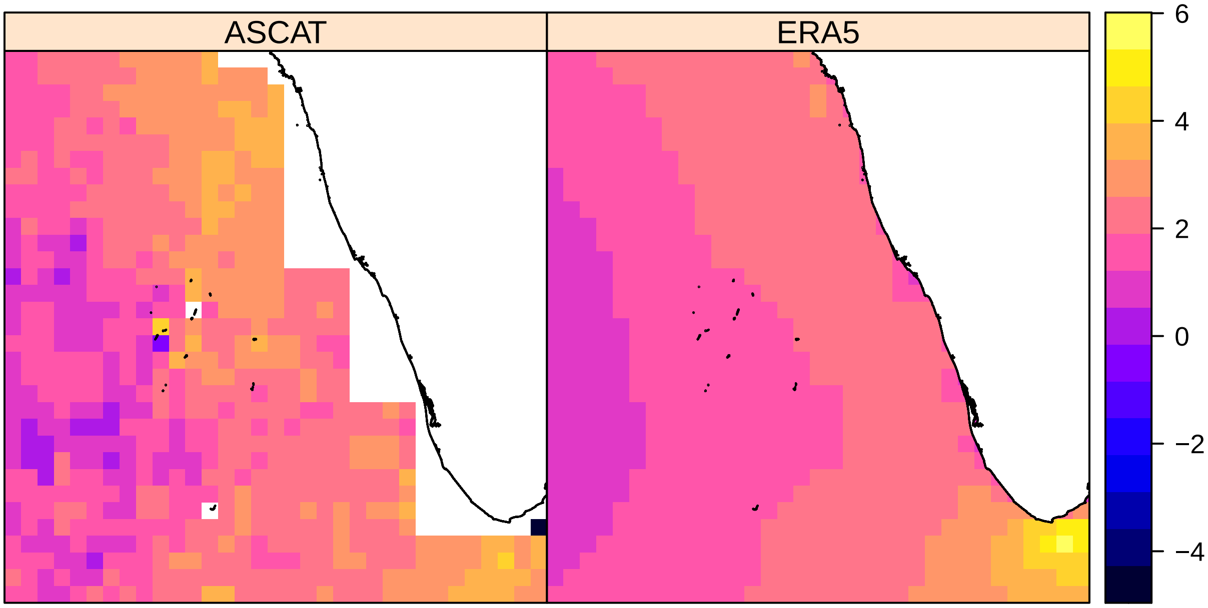
```
thedate <- "2010-07-01"
plotlons <- c(72.5)
pars <- c("u", "v")
p1 <- ggplot(subset(dfl, longitude==plotlons[1] & date==thedate & name%in%pars),
             aes(x=latitude, y=value, color=data)) + geom_line() +
  facet_wrap(~name) + ggtitle(paste(thedate, "longitude =", plotlons[1]))
thedate <- "2010-09-01"
p2 <- ggplot(subset(dfl, longitude==plotlons[1] & date==thedate & name%in%pars),
             aes(x=latitude, y=value, color=data)) + geom_line() +
  facet_wrap(~name) + ggtitle(paste(thedate, "longitude =", plotlons[1]))
gridExtra::grid.arrange(p1,p2)
```



We can see the ASCAT versus ERA5 difference especially for certain months such as April 2010. Here `indiashp2` is a shapefile for the coast of India.

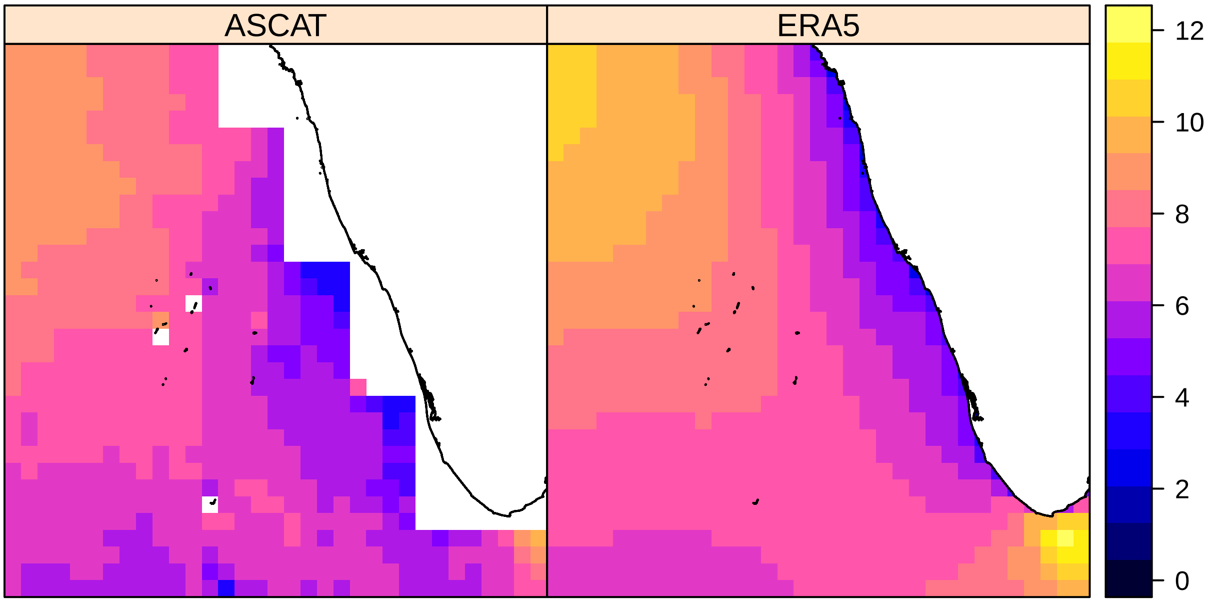
```
thedate <- "2010-04-01"
rdf <- subset(windmon2, date==thedate)
rdf$x <- rdf$longitude
rdf$y <- rdf$latitude
rdf <- rdf[, c("x", "y", "u")]
rst1 <- rasterFromXYZ(rdf); names(rst1) <- "ASCAT"
rdf <- subset(windmon4, date==thedate)
```

```
rdf$x <- rdf$longitude
rdf$y <- rdf$latitude
rdf <- rdf[, c("x", "y", "u")]
rst2 <- rasterFromXYZ(rdf); names(rst2) <- "ERA5"
rst <- stack(rst1, rst2)
proj4string(rst) <- "+proj=longlat +datum=WGS84"
at=seq(-3000,0,50)
spplot(rst) + layer(sp.polygons(indiashp2, fill="white"))
```



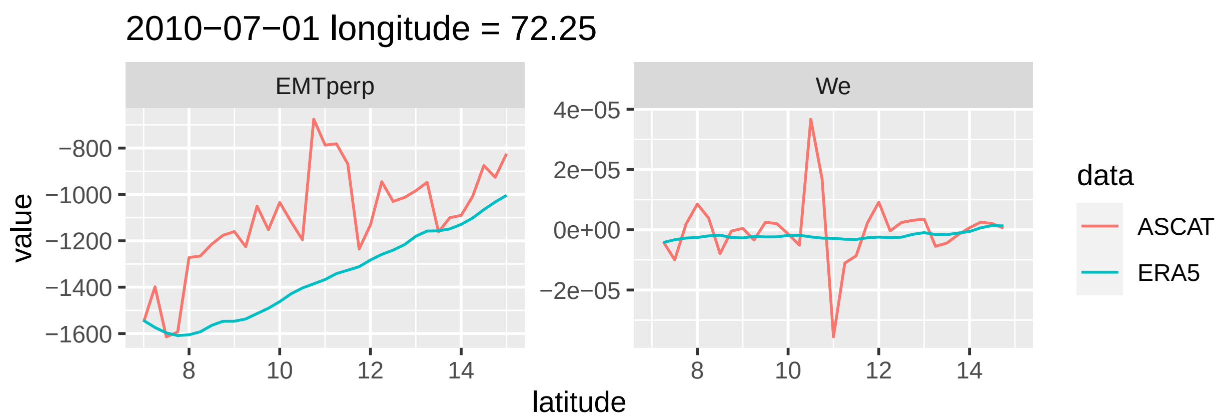
But less so for others, such as August 2010.

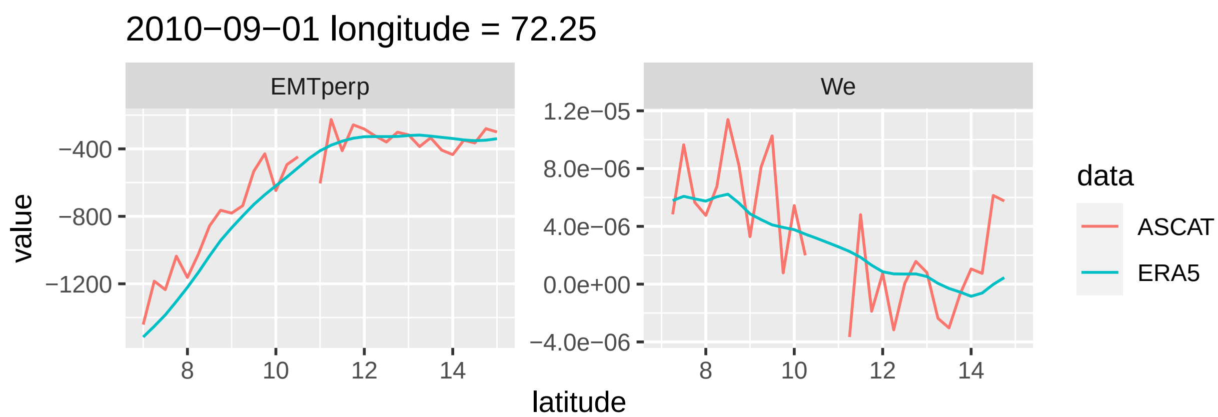
```
thedate <- "2010-08-01"
rdf <- subset(windmon2, date==thedate)
rdf$x <- rdf$longitude
rdf$y <- rdf$latitude
rdf <- rdf[, c("x", "y", "u")]
rst1 <- rasterFromXYZ(rdf); names(rst1) <- "ASCAT"
rdf <- subset(windmon4, date==thedate)
rdf$x <- rdf$longitude
rdf$y <- rdf$latitude
rdf <- rdf[, c("x", "y", "u")]
rst2 <- rasterFromXYZ(rdf); names(rst2) <- "ERA5"
rst <- stack(rst1, rst2)
proj4string(rst) <- "+proj=longlat +datum=WGS84"
at=seq(-3000,0,50)
spplot(rst) + layer(sp.polygons(indiashp2, fill="white"))
```



Ekman Mass Transport and Pumping

```
thedate <- "2010-07-01"
plotlons <- 72.25
pars <- c("EMTperp", "We")
p1 <- ggplot(subset(dfl, longitude==plotlons[1] & date==thedate & name%in%pars), aes(x=latitude, y=valu
  facet_wrap(~name, scales = "free_y") +
  ggtitle(paste(thedate, "longitude =", plotlons[1]))
thedate <- "2010-09-01"
p2 <- ggplot(subset(dfl, longitude==plotlons[1] & date==thedate & name%in%pars), aes(x=latitude, y=valu
  facet_wrap(~name, scales = "free_y") +
  ggtitle(paste(thedate, "longitude =", plotlons[1]))
gridExtra::grid.arrange(p1,p2)
```





Compare the coastal EMT (used in paper)

Define the coast latitude-longitude line.

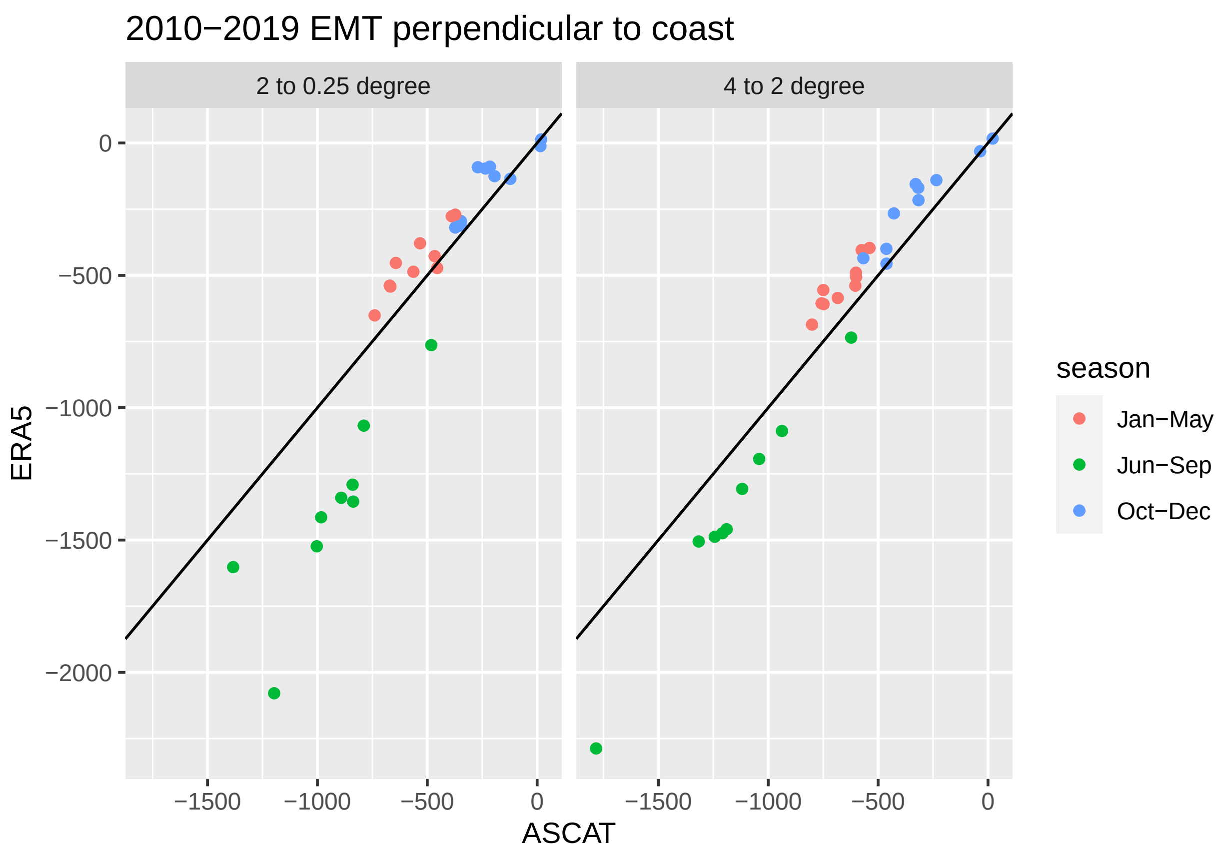
```
b=as(as(indiashp2, "SpatialLinesDataFrame"), "SpatialPointsDataFrame")
b=subset(b, Line.NR==129)
coastcoord=coordinates(b)
coastcoord <- coastcoord[coastcoord[,"x"]<77.5,]
coastlats = unique(windmon2$latitude)
coastlons = c()
for(i in coastlats){
  tmp <- coastcoord[,"y"]
  coastlons <- c(coastlons, min(coastcoord[,"x"][which(abs(tmp-i)==min(abs(tmp-i)))]))
}
```

Define the coast as 2 to 0.25 degrees offshore.

```
dfl2 <- dfl %>%
  pivot_wider(names_from = "data", values_from = "value")
dfl2.coast1 <- subset(dfl2, latitude >= 8 & latitude <= 13)
dfl2.coast2 <- subset(dfl2, latitude >= 8 & latitude <= 13)
for(i in 1:length(coastlats)){
  thelat <- coastlats[i]
  thelon <- coastlons[i]
  dfl2.coast1 <- dfl2.coast1[!(dfl2.coast1$latitude==thelat &
```

```r
    (dfl2.coast1$longitude<(thelon-2) | dfl2.coast1$longitude>=(thelon-0.25))),]
  dfl2.coast2 <- dfl2.coast2[!(dfl2.coast2$latitude==thelat &
    (dfl2.coast2$longitude<(thelon-4) | dfl2.coast2$longitude>=(thelon-2))),]
}
dfl2.coast1$coast <- "2 to 0.25 degree"
dfl2.coast2$coast <- "4 to 2 degree"
dfl2.coast <- rbind(dfl2.coast1, dfl2.coast2)
dfl2.coast$month <- format(as.Date(dfl2.coast$date), "%m")
dfl2.coast$year <- format(as.Date(dfl2.coast$date), "%Y")
dfl2.coast.mean <- dfl2.coast %>% group_by(year, month, name, coast) %>%
  summarize(FNMOC = mean(FNMOC, na.rm=TRUE),
            ASCAT = mean(ASCAT, na.rm=TRUE),
            ERA5 = mean(ERA5, na.rm=TRUE))
dfl2.coast.mean$season <- NA
dfl2.coast.mean$season[as.numeric(dfl2.coast.mean$month) %in% 6:9] <- "Jun-Sep"
dfl2.coast.mean$season[as.numeric(dfl2.coast.mean$month) %in% 10:12] <- "Oct-Dec"
dfl2.coast.mean$season[as.numeric(dfl2.coast.mean$month) %in% 1:5] <- "Jan-May"
```

```r
dfl2.coast.season <- dfl2.coast.mean %>% group_by(year, season, name, coast) %>%
  summarize(FNMOC = mean(FNMOC, na.rm=TRUE),
            ASCAT = mean(ASCAT, na.rm=TRUE),
            ERA5 = mean(ERA5, na.rm=TRUE))
ggplot(subset(dfl2.coast.season, name=="EMTperp" & ASCAT<1000),
       aes(x=ASCAT, y=ERA5, col=season)) + geom_point() + geom_abline() +
  facet_wrap(~coast) +
  ggtitle("2010-2019 EMT perpendicular to coast")
```



Conclusion

For EMT perpendicular to the coast, there is a linear relationship between the ERA5 product versus ASCAT. But the values from ERA5 are biased downward (more negative) in the summer monsoon (Jun-Sep) period. The bias is stronger when using values closer to the coast and suggests that it is due to the nearshore gaps in the ASCAT data seen in the raster plot. The missing raster points nearshore would be those with lower EMT values in the summer. However the generally linear relationship supports using the ERA5 product and the bias would not affect the analysis since the models include an intercept.

Functions in R

```
upwell <-  function (ektrx, ektry, coast_angle)
{
    pi <- 3.1415927
    degtorad <- pi/180
    alpha <- (360 - coast_angle) * degtorad
    s1 <- cos(alpha)
    t1 <- sin(alpha)
    s2 <- -1 * t1
    t2 <- s1
    perp <- (s1 * ektrx) + (t1 * ektry)
    para <- (s2 * ektrx) + (t2 * ektry)
    return(perp/10)
}
```

```
Cd <-  function (wind.speed)
{
    wind.orig <- wind.speed
    wind.speed[is.na(wind.speed)] <- -Inf
    Cd <- rep(NA, length(wind.orig))
    Cd[wind.speed <= 1] <- 2.18
    Cd[wind.speed < 3 & wind.speed > 1] <- 0.62 + 1.56/wind.speed[wind.speed <
        3 & wind.speed > 1]
    Cd[wind.speed >= 3 & wind.speed < 10] <- 1.14
    Cd[wind.speed >= 10] <- 0.49 + 0.065 * wind.speed[wind.speed >=
        10]
    Cd[is.na(wind.orig)] <- NA
    return(Cd * 0.001)
}

EMTperp <-  function (ektrx, ektry, coast_angle)
{
    pi <- 3.1415927
    degtorad <- pi/180
    alpha <- (180 - coast_angle) * degtorad
    s1 <- cos(alpha)
    t1 <- sin(alpha)
    s2 <- -1 * t1
    t2 <- s1
    perp <- (s1 * ektrx) + (t1 * ektry)
    para <- (s2 * ektrx) + (t2 * ektry)
    return(list(perp = perp, para = para))
}

getEMT <-  function (dat, coast_angle = NULL)
{
    if (!any(colnames(dat) == "u"))
        stop("need u in data")
    if (!any(colnames(dat) == "v"))
        stop("need v in data")
    for (i in c("uv_mag", "tau", "taux", "tauy", "EMT", "ektrx",
        "ektry", "upi")) if (any(colnames(dat) == i)) {
        warning(paste0(i, " is already in data. Changing to ",
            i, "_orig\n"))
        colnames(dat)[colnames(dat) == i] <- paste0(i, "_orig")
    }
    pa <- 1.22
    omega <- 7.272205e-05
    f <- 2 * omega * sin(pi * dat$latitude/180)
    uv_mag <- sqrt(dat$u^2 + dat$v^2)
    tau <- pa * Cd(uv_mag) * uv_mag * uv_mag
    tauy <- pa * Cd(uv_mag) * uv_mag * dat$v
    taux <- pa * Cd(uv_mag) * uv_mag * dat$u
    EMT <- tau/f
    EMTy <- -1 * taux/f
    EMTx <- tauy/f
    dat$uv_mag <- uv_mag
    dat$tau <- tau
    dat$taux <- taux
```

```
    dat$tauy <- tauy
    dat$EMT <- EMT
    dat$ektrx <- EMTx
    dat$ektry <- EMTy
    dat$We <- getEPump(dat)$We
    if (!missing(coast_angle)) {
        dat$EMTperp <- EMTperp(dat$ektrx, dat$ektry, coast_angle)$perp
        dat$upi <- upwell(dat$ektrx, dat$ektry, coast_angle)
    }
    return(dat)
}

getwindfromP <-  function (dat)
{
    if (is.null(attr(dat, "resolution")))
        stop("need resolution in the data.frame")
    if (!any(c("P_msl", "p_msl") %in% colnames(dat)))
        stop("need P_msl or p_msl (pressure) in data")
    for (i in c("u", "v", "P")) if (any(colnames(dat) == i)) {
        warning(paste0(i, " is already in data. Changing to ",
            i, "_orig", "\n"))
        colnames(dat)[colnames(dat) == i] <- paste0(i, "_orig")
    }
    res <- attr(dat, "resolution")
    lons <- sort(unique(dat$longitude))
    lats <- sort(unique(dat$latitude))
    dat$ug <- NA
    dat$vg <- NA
    dat$P <- 100 * dat[, tolower(colnames(dat)) == "p_msl"]
    h <- pi * res/180
    pa <- 1.22
    omega <- 7.272205e-05
    f <- 2 * omega * sin(pi * dat$latitude/180)
    R <- 6371 * 1000
    if (length(lats) > 2)
        for (lat in (min(lats) + res):(max(lats) - res)) {
            dPdlat <- (dat$P[dat$latitude == (lat + res)] - dat$P[dat$latitude ==
                (lat - res)])/(2 * h)
            constant <- (f * pa * R)[dat$latitude == lat]
            ug <- -1 * dPdlat/constant
            dat$ug[dat$latitude == lat] <- ug
        }
    if (length(lons) > 2)
        for (lon in (min(lons) + res):(max(lons) - res)) {
            dPdlon <- (dat$P[dat$longitude == (lon + res)] -
                dat$P[dat$longitude == (lon - res)])/(2 * h)
            constant <- (f * pa * R * cos(pi * dat$latitude/180))[dat$longitude ==
                lon]
            vg <- dPdlon/constant
            dat$vg[dat$longitude == lon] <- vg
        }
    ang <- pi * 30/360
    u <- 0.7 * (cos(ang) * dat$ug - sin(ang) * dat$vg)
    v <- 0.7 * (sin(ang) * dat$ug + cos(ang) * dat$vg)
```

```
    dat$u <- u
    dat$v <- v
    return(dat)
}

getdata <-  function (id, pars = NULL, lat = c(7, 15), lon = c(70, 78), date = NULL,
    altitude = 10, alt.name = "altitude", eserver = "https://coastwatch.pfeg.noaa.gov/erddap")
{
    url <- paste0(eserver, "/info/", id, "/index.csv")
    meta <- read.csv(url)
    if (!missing(date) && length(date) == 1)
        date <- c(date, date)
    if (missing(date))
        date <- c(meta$Value[meta$Attribute.Name == "time_coverage_start"],
            meta$Value[meta$Attribute.Name == "time_coverage_end"])
    lat1 <- lat[1]
    lat2 <- lat[2]
    lon1 <- lon[1]
    lon2 <- lon[2]
    fil <- paste0(id, "-", lat1, "-", lat2, "-", lon1, "-", lon2,
        "-", stringr::str_sub(date[1], 1, 10), "-", stringr::str_sub(date[2],
            1, 10), ".csv")
    dfil <- file.path(here::here(), "SupplementFiles", fil)
    if (!file.exists(dfil)) {
        if (missing(pars)) {
            pars <- unique(meta$Variable.Name)
            pars <- pars[!(pars %in% c("NC_GLOBAL", "time", "latitude",
                "longitude", alt.name))]
        }
        alttag <- ifelse(alt.name %in% meta$Variable.Name, paste0("[(",
            altitude, "):1:(", altitude, ")]"), "")
        val <- paste0("[(", date[1], "):1:(", date[2], ")]",
            alttag, "[(", lat1, "):1:(", lat2, ")][(", lon1,
            "):1:(", lon2, ")]")
        val2 <- paste0(pars, val, collapse = ",")
        url <- paste0(eserver, "/griddap/", id, ".csv?", val2)
        download.file(url, destfile = dfil)
        cat("data saved to", fil, "\n")
    }
    else {
        cat("data read from", fil, "\n")
    }
    dat <- read.csv(file = dfil, stringsAsFactors = FALSE)
    dat <- dat[-1, ]
    for (i in 2:ncol(dat)) dat[[i]] <- as.numeric(dat[[i]])
    dat$time <- as.POSIXlt(dat$time, format = "%Y-%m-%dT%H:%M:%OSZ",
        tz = "UTC")
    attr(dat, "resolution") <- min(abs(diff(dat$latitude))[abs(diff(dat$latitude)) !=
        0], na.rm = TRUE)
    cat(paste0("data ", id, " date ", date[1], "-", date[2],
        ", latitude ", lat1, "-", lat2, ", longitude ", lon1,
        "-", lon2), "\n")
    dat
}
```