Appendix A: Environmental indices of petrale recruitment, and estimates of the abundance spatial distribution of juveniles

Nick Tolimieri

Prior ROMS based indicators of petrale recruitment

Haltuch et al. (2020) examined the relationship between recruitment deviations from the 2019 petrale assessment (Wetzel 2019) and oceanographic drivers based on model output from a Regional Ocean Modeling System (ROMS) model for the California Current Ecosystem (Neveu et al. 2016). Potential drivers were selected based a conceptual life history model, which was used to generate life-stage-specific and spatio-temporally-specific mechanistic hypotheses regarding oceanographic variables that might likely influence survival at each life stage. The study area encompassed the region from 40 to 48°N in the California Current Ecosystem with individual predictors limited by depth or distance from shore (Table A1). Model selection resulted in a single model in which four oceanographic variables explained 73% of the variation in in the recruitment deviations. Recruitment deviations were:

- (1) positively correlated with degree days during the female precondition period (DDpre),
- (2) positively correlated with mixed-layer depth during the egg stage (MLDegg),
- (3) negatively correlated with cross-shelf transport during the larval stage (CSTlarv), and
- (4) negatively correlated with cross-shelf transport during the benthic juvenile stage (CSTbjuv).

These results suggested that ROMS output might be useful as the basis for an environmental index of recruitment for petrale to allow for better model precision and near-term forecasting. However, while the ROMS model used by Haltuch et al. (2020) was consistent in structure and inputs for 1980-2010, the ROMS model was updated beginning in 2011 producing potential discontinuities in the ROMS predictions and the identified drivers from the earlier 1980-2010 time period.

Here, outputs from the updated ROMS model for 2011-2022 were compared to the 1980-2010 mode looking for discontinuities in ROMS time series used in Haltuch et al. (2020) with a focus on the the four predictors identified as important in that work. The conceptual life-history model and abbreviation for terms are shown in Table 1.

Life stage	Year	Stage duration	Stage depth	Abbrv.	Hypothesis	ROMS variable
Preconditio ning	Year 0 May– October		Bottom depths of 50–200 m	DDpre	Higher bottom water temperatures increases food demand resulting in lower egg production, egg quality, or probability of spawning and lowers recruitment (likely a bell-shaped relationship)	Mean bottom water temperature (°C, 4 days)
Spawning	Year 1 November– March		Bottom depths of 250–500 m	Tpre.a	Bottom water temperature acts as a spawning cue with fish less likely to spawn at high temperature resulting in lower recruitment	Mean bottom water temperature (°C, 4 days)
				Tpre.b	Water column temperature acts as a spawning cue with fish less likely to spawn at high temperature resulting in lower recruitment	Mean water column temperature (°C, 4 days)
Egg, surface	Year 1 November– April	6–14 days	Water column from surface to MLD	MLDegg	Eggs are buoyant so mixed-layer depth limits how far they rise in the water column affecting later transport	Mean mixed-layer depth (m)
				LSTegg1	Transport in the water column above the MLD to settlement habitat affects recruitment	Mean long-shore transport between the surface and MLD (m/s, 4 days cumulative)
				CSTegg	Transport in the water column above the MLD to settlement habitat affects recruitment (aka Advection reduces recruitment while retention enhances recruitment)	Mean cross-shelf transport between the surface and MLD (m/s, 4 days cumulative)
				DDegg1	Growth/Predation hypothesis: growth rate is faster in warm water leading to reduced time vulnerable to predators	Degree days in the water column calculated from mean water column temperature between the surface and MLD (days,
Egg, sinking	Year 1 November– April	6–14 days	Water column from MLD to 400 m	LSTegg2	Transport in the water column to settlement habitat affects recruitment (aka Advection reduces recruitment while retention enhances recruitment)	Mean long-shore transport from the MLD to 400 m (m/s, 4 days cumulative)
				CSTegg2	Transport in the water column to settlement habitat affects recruitment	Mean cross-shelf transport from the MLD to 400 m (m/s, 4 days cumulative)
				DDegg2	Growth/Predation hypothesis: growth rate is faster in warm water leading to reduced time vulnerable to predators	Degree days in the water column calculated from mean water column temperature between the MLD and 400 m (days, 4 days)
Larvae (both yolk sack and feeding)	Year 1 December - May	~5 months for all pelagic stages	Water column from 0–50 m	LSTlarv	North to south transport in the water column brings northern zooplankton and	Mean long-shore transport in the water column at 50– 150 km offshore (m/s, 4 days cumulative)

Table 1: Conceptual life-history model for petrale, including abbreviation of ROMS term. Reproduced from Haltuch et al. (2020).

Life stage	Year	Stage duration	Stage depth	Abbrv.	Hypothesis	ROMS variable
				CSTlarv	Transport in the water column to settlement habitat affects recruitment (aka Advection reduces recruitment while retention enhances recruitment)	Mean cross-shelf transport in the water column at 50– 150 km offshore (m/s, 4 days cumulative)
				DDlarv	Growth/Predation hypothesis: growth rate is faster in warm water leading to reduced time vulnerable to predators	Degree days in the water column calculated from mean water column temperature at 50 – 150 km offshore (days, 4 days)
Pelagic juveniles (feeding pelagics)	Year 1 March– June	~5 months for all pelagic stages	Water column from 0 to 150 m	LSTpjuv	North to south transport brings northern zooplankton and leads to higher survival and recruitment, Transport to settlement habitat affects recruitment	Mean long-shore transport in the water column at 80– 120 km offshore (m/s, 4 days cumulative)
				CSTpjuv	Transport in the water column to settlement habitat affects recruitment (aka Advection reduces recruitment while retention enhances recruitment)	Mean cross-shelf transport in the water column at 80– 120 km offshore (m/s, 4 days cumulative)
				DDpjuv	Growth/Predation hypothesis: growth rate is faster in warm water leading to reduced time vulnerable to predators	Degree days in the water column calculated from mean water column temperature at 80– 120 km offshore (days, 4 days)
Benthic Juvenile (Age-0)	Year 1 April– October		Bottom depths from 50–150 m AND 150–500 m	LSTbjuv	Bottom water transport to settlement habitat affects recruitment (Advection reduces recruitment while retention enhances recruitment)	Mean long-shore transport at bottom depths of 50–150 m and 150 – 500 m (m/s, 4 days cumulative); Two depth ranges are considered due to uncertainty regarding juvenile distributions
				CSTbjuv	Bottom water transport to settlement habitat affects recruitment (aka Advection reduces recruitment while retention enhances recruitment)	Mean cross-shelf transport at bottom depths of 50–150 m and 150–500 m (m/s, 4 days cu- mulative); Two depth ranges are considered due to uncertainty regarding juvenile distributions

DD = degree days; T = temperature; MLD = mixed-layer depth, LST = longshore transport, CST = crossshelf transport; pre = female preconditioning stage; egg = egg stage, larv = larval stage, pjuv = pelagic juveniles, bjuv = benthic juveniles.

ROMS outputs

Visual comparison of the 1980-2010 and the 2011-2022 ROMS outputs show distinct discontinuities between the two time periods in multiple time series (Fig. A1). In particular, DDpre and MLDegg (from the original recruitment analysis) both show distinction changes in scale and trend across the 2010/2011 boundary. The cross-shelf transport time series from the original model (CSTlarv and CSTbjuv.a) are not as distinct but do suggest some discontinuity between ROMS models. Other parameters also show rapid changes, such as the longshore transport terms for the egg stage (LSTegg and LSTegg2) as well as the temperature terms. In some cases, the 2014-2016 marine heatwave may have influenced the 2011+ outputs (and thus be real), but this effect is unlikely the case for many of the terms that jump sharply from 2010 to 2011.



Figure 1: ROMS predictors used in the original 1980-2010 analysis but updated to include 2011-2022 outputs. DD = degree days, CST = cross-shelf transport, LST = long-shore transport, MLD = mixed layer depth, pre = prespawning, egg = egg stage, larv = larval stage, pjuv = pelagic juveniles, bjuv = benthic juveniles. 'a' and 'b' suffixes indicate drivers with the same time period but different depth ranges due to differences in the literature. Dotted red line indicates the 2010/2011 switch in model outputs.

There are multiple possibilities for how an individual ROMS time series may change over the 2010/2011 boundary and how those changes might impact the ROMS-recruitment model from Haltuch et al. (2020). There might be no change in the extracted ROMS variable such that the relationship between the ROMS predictor and petrale recruitment deviations remains the same. That is, there is no change in the intercept, slope, or variance (Fig. A2c). This outcome would be ideal because the current ROMS-recruitment model could be used unchanged with updated ROMS time series. The overall relationship might stay the same but variability in the ROMS parameter mighty change due to new observed model inputs leading to more correct estimates (Fig. A2b). Essentially, the variance increase (or decreases) from 2011+ due to change in the ROMS model structure, including new or different time series of observed data, but the slope and intercept remain the same. This outcome could be modeled by allowing the variance to change across the 2010/2011 boundary.



Figure 2: Hypothetical relationships between ROMS predictors and petrale recruitment deviations across the 2010/,2011 boundary. a) the relationship is the same, b) a change in variance of the ROMS predictor, c) a change in the intercept, and d) a change in the slope. Colors represent different time periods.

Additionally, the absolute value of the ROMS variable might change (e.g., water column temperature is warmer than previously modeled), but the overall relationship (slope) remains the same but with a different y-intercept (Fig. A2c). Here, one could add a time-period term to the model to account for different intercepts in the two blocks of time.

Finally, the relationship between the ROMS predictor and recruitment deviations might actually change (Fig. A2d). While one could include an interaction term in the model, this

case represents the least desired outcome, as it is difficult to interpret the meaning of the interaction. Relationships between environmental variables and biological outcomes are often non-linear, with performance decreasing both above and below a species optimum (e.g., growth and temperature relationships for fish). Obviously, there can also be some combination of b-d, or the relationship can just breakdown completely.

Refitting the Haltuch et al. (2020) model

To examine the impact of updating the ROMS model on the ROMS-recruitment relationship, the ROMS time series were updated (following Haltuch et al. (2020)) to include the new 2011+ ROMS outputs. The full ROMS time series were then fit against recruitment deviations from the 2019 petrale stock assessment (Wetzel 2019). The best-fit model from Haltuch et al. (2020) was used as the base model:

Recruitment Deviations ~ DDpre + MLDegg + CSTlarv + CSTbjuv.a

Recruitment deviations ran from 1981-2018 because some parameters were averaged over the winter and included data from two calendar years (i.e., Dec 2010 in the 2011 time series), and the first year with all predictors is 1981. The year 2011 was excluded from the analysis because some of the averaged ROMS variables would include data from both ROMS time periods and there were obvious step changes in some variables.

In addition to the base model, the following model structures were also investigated. These structures included the base model, plus:

- (1) different intercepts by period (1981-2010 vs, 2012-2018)
- (2) different variance by period
- (3) interactions between Period and the four terms in the original best-fit model (un
- (4) the model with interactions back-fit to remove non-significant interaction terms.
- (5) a separate four-factor generalized additive model (GAM) was also included to better visualize non-linear relationships between the ROMS predictors and petrale recruitment deviations.

For the linear models, the model including the interaction terms but with some interactions removed had the lowest AIC and fewest coefficients (Table 2) and was chosen as the most parsimonious model (Burnham & Anderson 1998).

Model	AIC	Parameters	Delta AIC
All interactions	-1.882	10	0.000
GAM	-0.979	37	0.903
Back-fit with interactions	-0.514	8	1.368
Period	19.980	6	21.862
Orginal best-fit	22.843	5	24.725
Different variances	30.276	8	32.158

Table 2: Model fitting criteria for the ROMS-recruitment model fits. Bold indicates the moswl chosen as the bestfit model based on delta AIC < 2.0 and the fewest parameters.

For 1981-2018, the best-fit model included Period and two interactions (Back-fit with interactions) because it had a Δ AIC < 2.0 and the fewest parameters:

Recruitment Deviations ~ *Period* + *DDpre* + *MLDegg* + *CSTlarv* + *CSTbjuv.a* + *DDprePeriod* + *MLDeggPeriod*

This model explained slightly less variance ($r^2 = 0.71$) than the original 1980-2010 model ($r^2 = 0.73$), but had 7 parameters (excluding the intercept) and 37 data points (1981:2010, 2012-2018), or just over five points per predictor term. Examination of the model coefficients (Table 3) indicates that the relationships between recruitment deviations and DDpre and MLDegg changed across the 2010/2011 boundary from positive to negative.

Overall the model fit to the data was good and captured a decline in recruitment later in the time series (Fig. A3). Forecasting through 2022 (latest ROMS output availability), suggested high recruitment in 2022. This model also over-predicted recruitment in 2015 during the 2014-2016 marine heatwave.

Model	Estimate	SE	t-value	p-value
Intercept	3.862	1.984	1.946	0.061
Period (before)	-7.605	2.073	-3.669	0.001
DDpre	-0.005	0.004	-1.308	0.201
MLDegg	-0.067	0.023	-2.874	0.008
CSTlarv	-38.082	7.108	-5.358	0.000
CSTbjuv	-38.795	11.695	-3.317	0.002
DDpre x Period	0.016	0.004	3.637	0.001
MLDegg x Period	0.112	0.025	4.505	0.000

Table 3: Parameter estimates from the best-fit model including combined ROMS time series from 1981-2018. Model $r^2 = 0.71$, p < 0.001.



Figure 3: Model fit for ROMS predictors to the recruitment deviations for 1981-2018. Black line = model prediction including forecast to 2019-2022. Points are the recruitment deviations from the 2019 petrale stock assessment. Error envelopes indicate 1.0 s.e.

As noted earlier, the interaction makes it somewhat difficult to draw conclusions on the credibility of the model - especially given the large discontinuities in the ROMS outputs and the change in slope of some of the relationships. Therefore, the four-factor GAM output was examined to better understand the relationships between the ROMS predictors and recruitment deviations. The four-factor GAM explained about 65% of the variance in

petrale recruitment deviations (Fig. A4, $r^2 = 0.67$). While the fit is slightly lower than the linear model, the GAM did not over-predict the more or less average 2018 recruitment deviation, while slightly over-predicting several lower recruitment deviations earlier in the time series. In particular, the GAM model caught the low recruitment in 2015 but under predicted recruitment in 2016.



Figure 4: GAM model fit to the recruitment deviations for 1981-2018. Black line = model prediction including forecast to 2019-2022. Points are the recruitment deviations from the 2019 petrale stock assessment. Error envelopes indicate 1.0 s.e.

For CSTlarv the new data fall within the same range as the 1980-2010 models, although there may be a slight change in the slope (Fig. A5). CSTbjuv appearsnon-linear but largely due to one data point (Fig. A5). Nevertheless, the newer values are all lower than the earlier period with the exception of one point.

Likewise, for MLD, the newer ROMS model predicts a shallower MLD and there appears to be a shift in the slope of the relationship (Fig. A5). With a continuous time series, one might reasonably interpret this non-linear relationship as reasonable, but given that the shift is associated with two different models, it seems more likely that the pattern is an artifact of the model discontinuity.

Other models

For DDpre, the 2011+ data are much "warmer" and in a different part of the graph, with the exception of one earlier period data point for DDpre (Fig. A5). For DDpre, the convex shape of the relationship would make some sense if there were no obvious shifts in temperature between time periods. Temperature typically impacts growth and survival for many fishes with an optimal temperature and declining performance in either warmer or colder waters. However, given the discontinuity between with the 2011+ data being warmer (Fig. A1), it seem more likely that this result is an artifact of the different ROMS models and not a biological one.



Figure 5: Smooths from the four-factor GAM. Black points are 1981-2010. Red points are 2012-2018. Numbers in parentheses on the y-axis are the estimated degrees of freedom where edf = 1.0 suggests a linear relationship.

Replacing ROMS predictors with observed variables

Because of the difficulties with the updated ROMS outputs, we also examined replacing specific ROMS predictors with observed variables or variables derived from observations, like the cumulative upwelling index.

West Coast Groundfish Bottom Trawl Survey (WCGBTS) bottom temperature data

In addition to bottom trawls for groundfish biomass and biological data, the WCGBTS (Keller et al. 2017) collects environmental data like bottom temperature. The survey time series begins in 2003, covers approximately 50 - 1200 m, and is conducted from May to October. This time period overlaps with the preconditioning period for petrale sole (Haltuch et al. 2019) suggesting that directly observed bottom temperature might be used to replace the DDpre term from the ROMS model.

Bottom temperature (btemp) from the WCGBTS was averaged for each year for May -October and then lagged one year to match the pre-conditioning year (btemp_pre). This lagging sets the available observed time series at 2004-2022. Degree days was also calculated as bottom temperature minus the reference temperature of 3.5 °C and averaged for each year. Normally one would sum degree days, but it was averaged here due varying samples sizes each year and multiple observations each year. Estimated this way degree days (btemp_DDpre) is essentially the same as btemp_pre but 3.5 degrees °C lower. The resulting time series are shown in Figure 6 along with DDpre and Tpre.a from the ROMS output for comparison. Interestingly, there were no significant correlations between the ROMS predictors and observed bottom temperature data whether compared across 2004-2022 or when separated by time period (2004-2010, 2011-2022) (Fig. A6).



Figure 6: Relationships between observed temperature from the trawl survey (WCGBTS) and modeled ROMS variables. a) time series of observed bottom temperature during the preconditioning period (btemp_pre), observed degree days during the preconditioning period (btemp_DDpre), DDpre, and Tpre.a, b) correlation between Tre.a and trawl bottom temperature, and c) correlation between DDpre and trawl bottom temperature. There were no significant correlations between observed bottom temperature data and modeled ROMS variables whether tested across both time periods (2004-2022, extent of the trawl data) or separated by ROMS time periods (2004-2010 and 2011-2022, p > 0.05 in all cases).

Cumulative upwelling index (CUTI)

The negative correlation with cross-shelf transport during the larval and benthic juvenile stages seen by Haltuch et al. (2020) indicates that recruitment is better when there is offshore transport in surface waters. This condition implies that better recruitment may correlate with upwelling either due to some transport-related interaction or better productivity during upwelling conditions leading to better feeding conditions. Therefore, CUTI was examined as a potential replacement for CSTlarv and CSTbjuv.a by averaging daily values for December-May and April-October, respectively (Fig. A7).



Figure 7: Cumulative upwelling indices (CUTI) calculated for the larval and benthic juvenile stages. Error envelopes indicate 1.0 s.d.

Cross-shelf transport and CUTI were correlated during the larval stage (Fig. A8a, r = -0.68, p < 0.001), but not during the benthic juvenile stage Fig. A8b, r = 0.01, p = 0.972).



Figure 8: Relationships between CUTI and ROMS variables.

Fitting new models

Although correlations between the ROMS variables in the best-fit model from Haltuch et al. (2020) and observed ocean temperature and the CUTI time series were weak, these new variables were used to fit the following model:

Recruitment deviations = btemp_pre + cuti_larv + cuti_juv

This model explained only 12% of the variance and was non-significant (p = 0.7). The model did capture some trends in recruitment deviations (Fig. A9) but missed both high and low recruitment events, suggesting that while it might be improved, but the relationship does not seem valuable at present.



Figure 9: Model fit for the bottom-temperature and CUTI predictors. Points are the recruitment devations from the 2019 stock assessment; line is the predicted fit with standard error. Note that there is no prediction for 2021 because there was no bottom trawl temperature data in 2020, which was necessary for calculating btemp_pre.

Basin scale predictors

The Ecosystem Status Report from the California Current Integrated Ecosystem Assessment reports several basin-scale indicators as indicative of environmental conditions along the west coast of the USA (Harvey 2022). These indicators include the Ocean Niño Index (ONI), Pacific Decadal Oscillation (PDO) index, and the North Pacific Gyre Oscillation (NPGO) index. These indicators were tested here to determine whether they could be used as the basis of an environmental index of petrale recruitment. The Coastal Upwelling Transport Index (CUTI) was also included as a measure of large scale cross-shelf transport for the basin. Indices were averaged for the spring (April-June) and summer (July-Sept). The ONI, PDO, and NPGO were also lagged to the pre-conditioning year for females because prior analyses (Haltuch et al. 2019) suggested that climate in the form of degree days was important during the pre-conditioning period (May-October).

Model selection (delta AIC < 2.0 and fewest parameters, Burnham & Anderson (1998)) was then used to select the best-fit model with the restriction that no model could have more than five predictor variables. Predictors that were correlated (r > 0.75) were excluded from the same model. For example, spring and summer NPGO indices were highly correlated and

excluded from the same model (Fig. A10). See Tolimieri et al. (2018) and Haltuch et al. (2020) for more detail on overall methodology.



Figure 10: Correlations among basin-scale predictors.

Model selection produced a large number of similarly weighted models. There were 22 models with an AICc < 2.0 (Table 4). AICc weights for these models were low and relatively similar. Most models included the NPGO during the pre-conditioning spring (npgo_spr_pre), NPGO during the summer of recruitment (npgo_sum), and PDO in the spring of the recruitment year (pdo_spring), while other terms occurred in some model but not all. Because it had the fewest terms and these terms were consistent across all models, the following model was chosen as the best-fit for the basin-scale parameters:

Recruitment deviations ~ npgo_spr_pre + npgo_sum + pdo_spr

Intercept	cuti_larv	npgo_aut	npgo_spr_pre	mus_opdn	npgo_sum_pre	oni_aut	oni_spr_pre	pdo_aut	pdo_spr	pdo_spr_pre	pdo_sum_pre	pdo_win	R2	AICc	delta AICc	Weight
0.16	-	-	-0.28	0.15	-	-	-0.29	-	-0.24		0.14	-	0.60	12.87	0.00	0.08
0.09			-0.25	0.17			-0.21		-0.20	0.14			0.60	13.05	0.18	0.08
0.09			-0.42	0.15	0.26				-0.20	0.13			0.60	13.39	0.52	0.06
0.08			-0.19	0.17					-0.21	0.12			0.55	13.69	0.82	0.06
0.07			-0.24	0.18				-0.12				-0.13	0.54	13.82	0.94	0.05
0.02			-0.17	0.17				-0.10		0.11		-0.15	0.59	13.87	1.00	0.05
-0.01	0.08		-0.28	0.19				-0.12				-0.11	0.59	13.96	1.09	0.05
-0.04	0.10		-0.50	0.17	0.28			-0.14					0.59	14.00	1.12	0.05
0.13			-0.27	0.18					-0.20				0.49	14.15	1.27	0.04
0.06			-0.19	0.16				-0.11			0.10	-0.19	0.59	14.20	1.32	0.04
0.01		0.14	-0.14					-0.14		0.14		-0.15	0.59	14.21	1.33	0.04
0.09			-0.24	0.16				-0.10	-0.15				0.54	14.25	1.38	0.04
0.07			-0.20	0.18					-0.14	0.13		-0.11	0.58	14.40	1.53	0.04
0.17			-0.22			-0.10	-0.41		-0.29		0.18		0.58	14.51	1.64	0.04
0.03			-0.17	0.21						0.14		-0.17	0.53	14.57	1.70	0.04
-0.04	0.10	0.14	-0.53		0.33			-0.18					0.58	14.59	1.71	0.04
0.15			-0.48	0.17	0.24				-0.19				0.53	14.60	1.72	0.04
-0.06	0.09		-0.24	0.18				-0.14					0.53	14.60	1.73	0.04
0.06			-0.27	0.22		0.08		-0.16				-0.11	0.58	14.64	1.77	0.03
0.11			-0.30	0.16			-0.18	-0.10	-0.14				0.58	14.66	1.79	0.03
0.13			-0.22	0.17					-0.15		0.13	-0.15	0.58	14.73	1.86	0.03
0.15			-0.33	0.18			-0.17		-0.19				0.53	14.84	1.97	0.03

Table 4: Model selection parameters and coefficients for the models with delta AIC < 2.0. NPGO/npgo = Northern Pacific Gyre Oscillation; ONI/oni = Ocean Nino Index, PDO/pdo = Pacific decadal oscillation; spr = spring; sum = summer; pre = female precoditioning year. The best-fit model is shown in bold.

Petrale recruitment deviations were negatively correlated with the NPGO in the precondition spring and with the PDO in the spring of the age-0 year. Recruitment was positively correlated with the NPGO in the summer of the recruitment year (Table 5).

Model	Estimate	SE	t-value	p-value
Intercept	0.134	0.058	2.332	0.027
npgo_spr_pre	-0.268	0.065	-4.150	0.000
npgo_sum	0.182	0.059	3.107	0.004
pdo_spr	-0.201	0.070	-2.859	0.008

Table 5: Parameter estimates from the best-fit model using basin scale predictors: ONI, PDO, NGPO, and CUTI. Model $r^2 = 0.49$, p < 0.001.

NPGO/npgo = Nothern Pacific Gyre Ossilation; ONI/oni = Ocean Nino Index, PDO/pdo = Pacific decadal oscillation; spr = spring; sum = summer; pre = female precoditioning year.

This basin-scale predictors model explained 49% of the variation in recruitment deviations ($r^2 = 0.49$, p < 0.001, Fig. A11). The model under-estimated highs and lows in petrale recruitment and predicted increases in recruitment later in the time series. While basin-scale predictors lack direct mechanistic explanations (compared to the ROMS predictors) and there are always concerns about the non-stationarity of relationships between the basin-scale indicators and physical processes, this model does track recruitment well.

For comparison, the predictions and fit for the model with the highest r^2 (and delta AIC < 2.0) are also shown in Figure A11. This model explained about 11% more variance (Table 4) and did a better job of matching the higher and lower recruitment (as one would expect from a model with more terms). These results suggest that the three-term model is generally good at capturing variation in petrale recruitment but that other drivers impact the more extreme recruitment that the base mode either over or under predicts.



Figure 11: Model fits for different models using the basin-scale predictors. NPGO = North Pacific Gyre Oscillation; PDO = Pacific Decadal Oscillation; spr = April=June; sum = July-September; pre = pre-conditioning year. Line is the predicted fit with standard error; Points are the recruitment deviations; dotted line indicates the zero recruitment deviation for reference.

From a stock assessment context, the inclusion of the npgo_sum term (Table 5) is a problem since the index would be available until the end of September at the earliest, causing the index to be a year behind the assessment. Removing the npgo_sum (and npgo_spr) term explained less variance but the model was also significant ($r^2 = 0.31$, p = 0.006)(Fig. A11). However, the spring and summer NPGO are highly correlated (r = 0.97, Fig. A10). Therefore, the best-fit model was re-run replacing npgo_sum with npgo_spr. This model explained slightly less variation in recruitment than the best-fit model but performed well overall ($r^2 = 0.42$, p = 0.002) (Fig. A11).

Copernicus Marine Environment Monitoring Service (CMEMS) Oceanographic Products

Given the difficulties with the combined ROMS analyses across 2010/2011, we also investigated alternative oceanographic model products produced by Copernicus Marine

Environment Monitoring Service (CMEMS) (<u>https://marine.copernicus.eu/</u>) and Mercator Ocean International (MOI) (<u>https://www.mercator-ocean.eu/</u>) to test if this modelling framework could be used to produce an environmental index of petrale recruitment.

We combined two CMEMS products: the Global Ocean Reanalysis and Simulation (GLORYS12V1: GLOBAL_MULTIYEAR_PHY_001_030, <u>https://doi.org/10.48670/moi-000211</u>) (Fernandez & Lellouche 2018, Jean-Michel et al. 2021, Drevillon et al. 2022) and the Copernicus Marine global analysis and forecast (CMGAF,

GLOBAL_ANALYSISFORECAST_PHY_001_024; <u>https://doi.org/10.48670/moi-00016</u>) (Le Galloudec et al. 2022). The data are served by the Copernicus Marine Service (<u>https://marine.copernicus.eu/</u>). When downloaded the data covered: GLORYS: 1993-01-01 to 2020-10-31 and CMGAF: 2020-11-01 to 2023-06-01. Note both the reanalysis and the analysis and forecast walk forward in time. For the CMGAF, time series are updated at regular intervals beginning with a daily forecast and hindcast simulation, and a weekly 'hindcast-best analysis' with data assimilation through -15 days (Le Galloudec et al. 2022).

Overall the CMEMS analysis followed Tolimieri et al. (2018) and Haltuch et al. (2020). More specifically, data for water column temperature, bottom temperature, and mixed-layer depth were downloaded as daily values for 40-48 °N and processed as follows for each life-history-stage predictor:

- 1. Subsetted data by bottom depth, mixed-layer depth, and distance from shore as relevant (see Table A1)
- 2. Calculated the daily average
- 3. Subsetted #2 by the relevant time periods (months in Table A1)
- 4. Calculated the annual average (or sum for degree days) for 1993-2023 for that potential predictor

For transport variables, monthly means from the CMEMS models were used to reduce processing time but followed the same overall model selection process as as above. Overall, the combined CMEMS time series did not show obvious break points from 2020 to 2021 (Fig. A12).

Model selection followed Tolimieri et al. (2018) and Haltuch et al. (2020). Briefly, CMEMS predictors were pre-screened for correlations among variables and non-linear relationships with petrale recruitment deviations. Correlated predictors ($r \ge 0.75$, Fig. A@ref(fig: glorys-correlations)) were excluded from the same model. Non-linearity for individual terms was evaluated by comparing the linear model to a model including both the linear and quadratic forms of the predictor. If the AICc of the quadratic form was lower, we included the quadratic form as a potential predictor as well, but required that the linear term appear in models that included the quadratic from.

Model selection was carried out using the 'dredge' function in the MuMIn package in R (R Core Team (2023), Bartoń (2023)). Candidate models were evaluated based on their delta AIC and number of predictors.



Figure 12: Time series of processed CMEMS indicators used in the model fitting.



Figure 13: Correlations between CMEMS time series. DD = degree days, T = temperature, MLD = mixed-layer depth, LST = longshore transport, CST = cross-shelf transport, pre = female precondition period prior to spawning, egg = egg stage, larv = larval stage, pjuv = pelagic juveniles, bjuv = benthic juveniles.

Results

Only two candidate models had delta AICc values of 4.0 or less; only one model had a delta AICc of less than 2.0. (Table A(tab:mtable)). The best-fit model included degree days during the pelagic juvenile period (DDpjuv) in both its linear and quadratic form, and long-shore transport (LSTlarv) in the larval stage (Table A(tab:mtable)).

Table 6: Terms and coefficients for the candidate models with a delta AIC value of 4.0 or less.

Int	DDpjuv	DDpjuv ²	LSTlarv	LSTIarv ²	LSTpjuv	LSTpjuv ²	R ²	AICc	delta	weight
24.859	-6.683	0.446	15.776	474.695			0.683	6.014	0.000	0.870
26.754	-7.212	0.480			5.958	471.358	0.633	9.820	3.805	0.130

DD = degree days; LST = longshore transport, larv = larval stage, pjuv = pelagic juveniles.

This best-fit model explained 68% of the variance in the petrale recruitment deviations from 1993-2018 (Fig A14) and generally captured the peaks and lows in the recruitment deviations well. Predictions of potential petrale recruitment through 2023 are included in Figure A14. However, the model did over-predict recruitment in 2015 during the large marine heatwave off the us west coast. When predicted through 2023, this index suggests strong incoming recruitment equivalent to historical highs during the 2006-2008 period.



Figure 14: Relationship between recruitment deviations from the 2019 assessment and the MOI index for 1993-2018 and predictions for 2019-2023. Points are the recruitment deviations from the assessment; solid like is the MOI environemental index of recruitment, and grey envelope represents 95 confidence intervals.

Model diagnostics and testing

Model diagnostics and testing followed Tolimieri et al. (2018) and Haltuch et al. (2020). Model testing was carried out to determine how stable the best-fit model was to both individual years and the precision of the estimates of recruitment deviations. Tests included:

- 1. Boot-strap analysis on the best-fit model was used to estimate bias
- 2. Jackknife analysis on the best-fit model was used to determine the impact of individual years on the model fit.
- 3. Refit the best-fit model using data for 1993-2013 and then predict 2014-2018.
- 4. Recruitment deviations were re-sampled 1000 times from a log-normal distribution to evaluate the impact of the precision of these estimates on the fit of the best-fit model.

- 5. Individual years were jackknifed and then the entire model selection process was rerun to determine the impact of individual years on the selection of the predictors in the best-fit model.
- 6. Recruitment deviations were re-sampled 100 times and the entire model selection process rerun to evaluate the impact of the precision of these estimates on the predictors included in the best-fit model.

Diagnostics for the best-fit model

There was no obvious autocorrelation in the residuals (Fig. A15). Diagnostic plots did suggest some under-prediction when recruitment deviations were high (Fig. A16.



Model 1 Autocorrelation plot

Figure 15: Autocorrelation plot for the best-fit model



Figure 16: Diagnostic plots for the best-fit model.

Model testing

Bias (estimated from bootstrap analysis with 1000 bootstraps) was generally low (Table A7) at around 5-15% of the scaled coefficient with the exception of LSTlarv, which had a small impact overall on the model relationship (standardized coefficient of -0.028.

Predictor	Coefficient	Bias	SE	Std Coef	Std Bias	Std SE
Intercept	24.859	-1.930	11.363	-0.288	0.016	0.074
DDpjuv	-6.683	0.506	3.134	-0.096	0.015	0.060
LSTlarv	15.776	-0.448	4.352	-0.028	0.011	0.051
I.DDpjuv.2.	0.446	-0.033	0.215	0.163	-0.014	0.076
I.LSTlarv.2.	474.695	-27.744	124.598	0.167	-0.010	0.052

Table 7: Standardized coefficients and estimates of bias from 1000 bootstraps.

DD = degree days; LST = longshore transport, larv = larval stage, pjuv = pelagic juveniles.

Jackknifing the best-fit model by year (removing individual years and refitting the best-fit model) did not show strong effects of individual years on the fit of the model (Fig. A17) with the majority of r^2 values falling between 0.65 and 0.70.



Figure 17: Results of jackknife re-fitting of the best-fit model to evaluate the impact of individual years on the model fit.

Predictions of individuals years from the jackknife analysis fell largely along the same trend as for the best-fit model and within the 95% confidence limits (Fig. A18). Prediction 2014 to 2018 based on refitting the best-fit model to data for 1993-2013 captured the last five years relatively well, although estimates for 2015 and 2016 were high (Fig. A18).



Figure 18: Fit of the best-fit model to the recruitment deviations from the 2019 petrale asessment. Solid line is the predicted fit and dashed lines are the 95% confidence limits. Open circles are the log recruitment deviations from the 2015 sablefish assessment. Stars are predicted values from jackknife analysis removing individual years one at a time. Red points are predictions from fitting best-fit model to 1993-2013 and then predicting 2014-2018.

To examine the impact of the precision of the estimate of recruitment deviations on the fit of the best-fit model, recruitment deviations were re-sampled from a log-normal distribution using the estimate and standard error for each year and the model refit. This procedure was completed 1000 times. The performance of the best-fit model was fairly stable (Table A8). The mean r^2 for the across all 1000 iterations was $r^2 = 0.567$ with a lower 95% quantile of $r^2 = -6.602$.

Table 8: Results of resampling recruitment deviations and refitting the best fit model 1000 times. Best-fit = best-fit model; Mean = mean values over the 1000 refits, median = median values over the 1000 refits; and 2.5 and 97.5% confidence intervals.

Term	R ²	Int	DDpjuv	DDpjuv ²	LSTlarv	LSTlarv ²	р
Best-fit	0.683	24.859	-6.683	0.446	15.776	474.695	0.000
Mean	0.567	24.572	-6.602	0.440	15.791	474.485	0.005
Median	0.573	24.554	-6.596	0.441	15.781	475.272	0.001
2.5%	0.385	11.931	-9.792	0.194	10.321	367.356	0.000
97.5%	0.737	35.892	-3.061	0.662	21.215	576.199	0.031

DD = degree days; LST = longshore transport, larv = larval stage, pjuv = pelagic juveniles.

Jackknifing years and re-running the entire model selection process also produced stable results (Table A9). DDpjuv and LSTlarv occurred in the 23 and 24 (0 respectively) of the top model across the 100 jackknifes. CSTegg1 and DDegg1 each occurred in three models, while MLDegg and its quadratic form occurred in two.

Predictor	No
CSTegg1	3
DDegg1	3
DDpjuv	23
DDpjuv ²	23
LSTlarv	24
LSTlarv ²	24
MLDegg	2
MLDegg ²	2

Table 9: Result of jackknifing individual years and re-running the model selection process. The results are the number of times a predictor appeared in the model with the lowest AICc in each iteration of the re-selection.

DD = degree days; T = temperature; MLD = mixed-layer depth, LST = longshore transport, CST = cross-shelf transport; egg = egg stage, larv = larval stage, pjuv = pelagic juveniles.

A similar analysis was conducted in which recruitment deviations were re-sample from a log-normal distribution and the entire model selection process was rerun to determine how the variability in recruitment estimates might impact the overall model selection process. LSTlarv and DDpjuv and their quadratic terms were included in at least 79% of the refits (Table A10). The two linear terms alone appeared in 90% (LSTlarv) and 85% (DDpjuv) of models, providing strong support for stability these terms in relation to the precision of estimates of recruitment deviations.

Predictor	No. models	No. jackknifes
CSTbjuv.a	6	4
CSTbjuv.b	1	1
CSTegg1	59	33
CSTegg2	6	5
CSTlarv	1	1
CSTpjuv	1	1
DDbjuv.a	2	2
DDbjuv.b	1	1
DDegg1	48	39
DDegg2	18	18
DDlarv	26	22
DDpjuv	85	61
DDpre	7	6
CSTegg2 ²	6	5
DDpjuv ²	79	56
LSTIarv ²	87	68
LSTpjuv ²	26	26
LSTpjuv ²	45	27
LSTbjuv.a	0	0
LSTbjuv.b	1	1
LSTegg1	1	1
LSTegg2	0	0
LSTIarv	90	69
LSTpjuv	31	30
MLDegg	45	27
Tpre.a	7	4
Tpre.b	8	5

Table 10: Result of resampling recruitment deviations and completing the model selection process 100 times. The results are the number of times a predictor appeared in the model with the lowest AICc from each re-dredging iteration.

DD = degree days; T = temperature; MLD = mixed-layer depth, LST = longshore transport, CST = cross-shelf transport; pre = female preconditioning stage; egg = egg stage, larv = larval stage, pjuv = pelagic juveniles, bjuv = benthic juveniles.

Finally, the data were subset to 1993-2013 and the entire model selection process was rerun. The resulting best-fit model included the same terms as for the full analysis. This 1993-2013 model was then used to predict petrale recruitment deviations through 2023. The overall fit of the model though 2013 was high ($r^2 = 0.78$) (Fig. A19). However, like the best-fit model, it over-predicted recruitment for 2015 and 2016 during the height of the marine heatwave, suggesting that some fundamental dynamics may have shifted during this period.



Figure 19: Fit of the best-fit model from the 1993-2013 analysis and predictions through 2023. Black line is the model prediction with 95% confidence limits. Black points are the recruitment deviations used in the model development; red points are future recruitments.

Distribution and abundance of juvenile petrale

Tolimieri et al. (2020) used spatio-temporal models to examine the distribution and abundance of 13 species of juvenile fishes along the West Coast. Those results are updated here following Tolimieri et al. (2020) but using the 'sdmTMB' package (Anderson et al. 2022) for R instead of the 'VAST' package (Thorson 2019).

Data come from the West Coast Groundfish Bottom Trawl Survey (WCGBTS)(Keller et al. 2017) for 2003-2022. There are no data for 2020 as COVID restrictions prevented the survey from being completed. The individual biomass data contain estimates of age, length, and biomass for subsamples of each haul and occasionally data for the entire haul when catch is low. There is length information (cm total length) for all individuals in the subsample but many individual fishes lack weight or age data due to time-constraints in the field. To expand the subsample, the following procedure was followed:

- 1. Missing weights for individuals in the subsample were estimated by first estimating the length-weight relationship from existing data and using this relationship to estimate the missing weights.
- 2. Individual fish were allocated to age classes following Tolimieri et al. (2020) by using length-age relationships from the WCGBTS data to determine age-class maximum lengths. See Tolimieri et al. (2020) for more detail. The maximum lengths used here were taken from Tolimieri et al. (2020). Maximum length was 21 cm total length, and depth ranged from 50-200 m. Based on otolith analyses, these fish would

be age-1 and age-2 fishes. Thus, these results do not represent recruitment but juvenile abundance several years later.

3. The proportional biomass of juveniles in each subsamble was calculated and used to estimate the total biomass of juvenile fishes in the full trawl.

The juvenile biomass index was low in recent years (Fig. A20), which does not match up with the predictions of recruitment from the basin-scale models. Observed modeled recruitment deviations from the 2019 assessment and the observed juvenile biomass in the trawl survey do not align either. Juvenile abundance for petrale was highest in the 2008-2011 time period (Fig. A20) with high densities just south of 45°N and moderate abundance to the south until about 37 °N. This period of high abundance followed several strong recruitment events from 2006-2008 (Fig. A14) suggesting a 2-3 year delay between recruitment events and when they are observed in the trawl survey, which is consistent with the age of the fishes in the trawl survey. This delay would also function to reduce the correlation between the recruitment deviations from the assessment and observed abundance of agee-1 and age-2 petrale since the index in Figure A20 integrates several age classes.

Age-1 and age-2 petrale were widely distributed from 35 to 46 °N during the period of high abundance from 2008-2011 (Fig. A21). Abundance was highest around the bank system at approximately 44 °N, even in years of lower abundance.



Biomass index 1.000 Latitude 0.100 8.819 Year

Figure 20: Index of juvenile abudance from the species distribution modelling.

Figure 21: Distribution juvenile petrale sole along the West Coast from 2003-2022 from the species distribution modeling.

References

Anderson SC, Ward EJ, English PA, Barnett LA (2022) sdmTMB: An r Package for Fast, Flexible, and User-Friendly Generalized Linear Mixed Effects Models with Spatial and Spatiotemporal Random Fields. bioRxiv.

Bartoń K (2023) MuMIn: Multi-model inference.

- Burnham KP, Anderson DR (1998) Model selection and inference: A practical informationtheoretic approach. Springer-Verlag, New York, NY.
- Drevillon M, Fernandez E, Lellouche JM (2022) For the Global Ocean Physical Multi Year Product GLOBAL_MULTIYEAR_PHY_001_030. Copernicus Product User Manual 1:1– 25.
- Fernandez E, Lellouche JM (2018) Product User Manual for the Global Ocean Physical Reanalysis Product Glorys12v1. Copernicus Product User Manual 4:1–15.
- Haltuch MA, Tolimieri N, Lee Q, Jacox MG (2020) Oceanographic Drivers of Petrale Sole Recruitment in the California Current Ecosystem. Fisheries Oceanography 29:122– 136.
- Harvey A C. J. Leising (2022) 2022-2023 California Current Ecosystem Status Report a report of the NOAA California Current Integrated Ecosystem Assessment Team (CCIEA) to the Pacific Fishery Management Council, March 7, 2023.
- Jean-Michel L, Eric G, Romain B-B, Gilles G, Angélique M, Marie D, Clément B, Mathieu H, Olivier LG, Charly R (2021) The Copernicus Global 1/12 Oceanic and Sea Ice Glorys12 Reanalysis. Frontiers in Earth Science 9:698876.
- Keller AA, Wallace JR, Methot RD (2017) The Northwest Fisheries Science Center's West Coast Groundfish Bottom Trawl Survey: History design, and description. U.S. Department of Commerce.
- Le Galloudec O, Law Chune S, Nouel L, Fernandez E, Derval C, Tressol M, Dussurget R, Bairdeau A, Tonani M (2022) Product User Manual for Global Ocean Physical Analysis and Forecasting Product GLOBAL_ANALYSISFORECAST_PHY_001_024. Copernicus Product User Manual 1:1–41.
- Neveu E, Moore AM, Edwards CA, Fiechter J, Drake P, Crawford WJ, Jacox MG, Nuss E (2016) An Historical Analysis of the California Current Circulation Using ROMS 4d-Var: System Configuration and Diagnostics. Ocean Modelling 99:133–151.
- R Core Team (2023) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Thorson JT (2019) Guidance for Decisions Using the Vector Autoregressive Spatio-Temporal (VAST) Package in Stock, Ecosystem, Habitat and Climate Assessments. Fisheries Research 210:143–161.

- Tolimieri N, Haltuch MA, Lee Q, Jacox MG, Bograd SJ (2018) Oceanographic Drivers of Sablefish Recruitment in the California Current. Fisheries Oceanography 27:458– 474.
- Tolimieri N, Wallace J, Haltuch M (2020) Spatio-Temporal Patterns in Juvenile Habitat for 13 Groundfishes in the California Current Ecosystem. PLoS One 15.
- Wetzel CR (2019) Status of Petrale Sole (*Eopsetta Jordani*) Along the U.S. West Coast in 2019. Pacific Fishery Management Council, 7700 Ambassador Place NE, Suite 101, Portland, OR 97220.