ISC/25/ALBWG-01/01

An update of the standardized abundance index for the US surface fleet in the North Pacific

Peter T. Kuriyama[1,2] and Steven L. H. Teo[1]

[1]Southwest Fisheries Science Center, NOAA NMFS, La Jolla, CA, USA

[2]Email: peter.kuriyama@noaa.gov

This working paper was submitted to the ISC Albacore Working Group Intercessional Workshop, 10-14 March 2025, held at the National Taiwan University, Institute of Oceanography in Taipei, Taiwan.

ABSTRACT

This study updated the standardized index of abundance for the US surface fleet logbook data. Similar to previous efforts, the data were grouped into three time periods and generalized linear models were applied to the data. The full model that estimated fixed effects for year, area, and season were selected for each of the time periods. The estimates from 1999-2024 were similar to those estimated for the 2023 benchmark assessment. The recent CPUE estimates were relatively high in 2022, low for 2023, then relatively high in 2024. Spatiotemporal models with sdmTMB were also explored. The full GLM estimated in sdmTMB had identical coefficient estimates, and preliminary results that estimated a spatial random field were described. A more thorough investigation of spatiotemporal models will be completed prior to the data preparation workshop scheduled for the end of 2025.

INTRODUCTION

This working group paper describes preliminary results from reanalysis of the US surface fishery logbook data for albacore tuna in the north Pacific Ocean updated through 2024. The US surface fleet includes troll and pole-and-line gear, and has maintained logbook records dating back to the early 1960s. In 1961, the Southwest Fisheries Science Center developed and implemented a logbook program: participation was voluntary at first and became mandatory in 2005. As a result, logbook coverage of fishing activity has been variable, ranging from 7-33% from 1966-2004 but was estimated to be nearly 75% after 2005 (McDaniel 2006). While the albacore stock assessment model does not fit to the US surface fleet index of abundance (i.e., 0 weighting in the model) in the base model, it is evaluated in sensitivity runs. The model does not fit to the US surface fleet index because the data (and subsequent standardized indices) are not considered to be representative of the entire albacore stock in the north Pacific.

The treatment of the data and method of analysis have developed through time but have largely included estimating fixed effects for year, season, and area in a generalized linear model (GLM). In 2006, the data were standardized with a GLM and separated into two areas: Inshore (east of 130°W) and Offshore (west of 130°W). The 2013 index divided the data into eight fishing areas to account for fishing within 200 nm of the coast (Xu *et al.* 2013). The 2013 index also fit three GLMs, one for each time period, which were separated to account for differences in general fishing effort patterns. Fishing was largely concentrated close to shore from 1966-1978, expanded further offshore in 1979-1988, then concentrated close to shore in the northeast Pacific from 1999-2024 (Figure 1; Xu *et al.* 2013). Most recently, Teo (2022) evaluated six Bayesian generalized linear mixed models, some of which included random effects for vessel and space (utilizing the software INLA; Rue *et al.* 2009). The selected model was a negative binomial generalized linear mixed effect model with random vessel effects. However, the overall trends were similar to those from the GLM-based index so GLM results were included in the 2023 benchmark assessment sensitivities. The US surface fleet indices of abundance input to the 2023 assessment files were from a GLM with year, season, and area effects for data from 1999-2021.

This study updates the GLM-based index through 2024 and explores spatiotemporal modeling of the US surface fleet logbook data. There are a number of potential models to evaluate in preparation for the 2026 albacore benchmark assessment, and some of these possible directions are outlined here.

MATERIALS AND METHODS

Data

The SWFSC logbook data were aggregated according to previous methods (e.g., Xu *et al.* 2013, Teo 2022). These steps involved calculating nominal catch-per-unit-effort (CPUE) for each: 1° x 1° cell, month, and year. Nominal CPUE was the total catch divided by the total numbers of vessel-days. The aggregated data were also assigned a season and an area based on location. The seasons were January-March (1), April-June (2), July-September (3), and October-December (4). The area definitions were those described in Xu *et al.* (2013) and shown in Figure 2:

- Area 1- 200nm or less north of 48°N
- Area 2, 200nm or less, 40-48°N
- Area 3, 200nm or less, south of 40°N
- Area 4, more than 200nm, east of 140°W
- Area 5, north of 40°N, 140-160°W
- Area 6, south of 40°N, 140-160°W
- Area 7, north of 40°N, 160°E-160°W
- Area 8, south of 40°N, 160°E-160°W.

Additionally, the data were grouped into three time periods, based on the previously described shifts in fishing locations. Patterns of total effort and total catch are shown in Figure 3. Both effort and catch have declined since 2022 in the US surface fleet. Data from Season 1 (January-March) and cells with fewer than 3 vessel-days were dropped from the analysis. Nominal CPUE values from 2019-2024 are shown in Figure 4. There were logbook records for June 30th through September 21st in 2024.

Models

Three GLMs were fit to the 1961-2024 data, separated into three time periods. The full model was:

$$\ln\left(CPUE_{ijk} + 1\right) = X + Y_i + S_j + A_k + \epsilon_{ijk}$$

where $CPUE_{ijk}$ was the CPUE (fish per boat day) in year *i*, quarter *j*, and area *k*. *X* was the intercept representing the reference block, *Y* was the year factor, *S* was the season, and *A* was the area. The other models evaluated had only *Y*, or *Y* and *S*. Models were compared with AIC for each of the three periods. A bootstrap analysis was not conducted to calculate coefficients of variation. This will be done in advance of the data-focused 2025 ALBWG meeting.

The capabilities of the R package *sdmTMB* were explored for this working paper (Anderson *et al.* 2024). The goal of the package is to fit spatial and spatiotemporal Generalized Linear Mixed Effects Models in a user-friendly framework that utilizes syntax similar to the ‘glm’ command in R. As a first step, the GLM model with year, season, and area effects was replicated in *sdmTMB*. The results are preliminary and will be more thoroughly investigated in response to comments at the upcoming ALBWG meeting.

Additional model complexity was explored for the 1999-2024 data with a response variable of ln(CPUE+1) and *sdmTMB*. One model evaluated was one that estimated fixed effects for year and season and estimated a spatial Gaussian random field (in lieu of a fixed effect for area). Estimation of spatial fields might be most applicable for species with strong habitat preferences such as fish that prefer rocky reefs. This setting is not necessarily an ecological match to albacore dynamics and, as noted before, these examinations of *sdmTMB* are very preliminary.

RESULTS

The full models with year, season, and area effects were selected for each of the three time periods (Table 1). Overall, the trends for each of the models were very similar (Figure 5). The 1999-2024 standardized CPUE values were also very similar to those input into the 2023 benchmark assessment (Figure 6). The 2023 CPUE value was relatively low compared to those from 2022 and 2024 (estimates for 1999-2024 shown in Table 2). Significant coefficients for the model run are shown in Table 3.

Running the GLM with year, season, and area in *sdmTMB* resulted in identical coefficient estimates (Table 4). The annual indices were calculated differently than those shown in Table 2, and this calculation will be further investigated.

The model that estimated fixed effects for year and season and a spatial Gaussian random field had good convergence and a 1.5 second estimation time. The Matern range (which is a measure of the distance between independent samples in space) was 404 km. The difference between the coefficients is shown in Table 4. The spatial field estimated is shown in Figure 7 and the pattern seems to roughly match expectations given spatial patterns in the nominal landings data.

DISCUSSION

The GLM updated with data through 2024 had similar trends to estimates evaluated in the 2023 benchmark assessment model sensitivities. At minimum, the GLM results will likely be suitable for use in the upcoming benchmark assessment sensitivity runs.

Standardized CPUE values from the two *sdmTMB* models differ from those in the GLM. Currently, the specific *sdmTMB* calculation is not known but will be investigated further. The years 2022-2024 in the GLM are high, low, then high, and in the *sdmTMB* models the 2022 value is relatively high then declines through 2024.

There are a wide range of potential applications of *sdmTMB* models, but there are a number of model settings that may affect the results and will be explored more fully. Specifically, spatial and spatiotemporal models are sensitive to the number of nodes used in the spatial meshes. A high number of nodes increases the spatial resolution of the model but increases run times. Models that were fit to data aggregated to 1 by 1 cells did not tend to converge when estimating a spatiotemporal process (as opposed to a spatial relationship shown in the results). The timestep of the models also might affect convergence. Spatiotemporal models did not seem to converge with annual time steps. One solution might be to add month-year time steps or perhaps even Julian dates to increase the temporal resolution. Finally, models estimated to unaggregated spatial data (that is, fit to individual logbook observations) seemed result in robust estimates of spatiotemporal processes. If these issues can be resolved, models that include environmental covariates like sea surface temperature or models that are fit to data from multiple countries may be utilized in the future.

REFERENCES

Anderson SC, Ward EJ, English PA, Barnett LAK, Thorson JT (2024). "*sdmTMB*: an R package for fast, flexible, and user-friendly generalized linear mixed effects models with spatial and spatiotemporal random fields." bioRxiv, 2022.03.24.485545. doi:10.1101/2022.03.24.485545.

McDaniel, J. D., P. R. Crone, and E. Dorval. 2006. Critical evaluation of important time series associated with albacore fisheries (United States, Canada, and Mexico) of the Eastern North Pacific Ocean. Page Report of the ISC Albacore Working Group Workshop, 28 November - 5 December, 2006. ISC/06/ALBWG/09, Shimizu, Shizuoka, Japan.

Teo, S. L. H. 2017. Relative abundance indices of juvenile albacore tuna for the US surface fishery in the north Pacific Ocean. ISC/17/ALBWG/09. Working document submitted to the ISC Albacore Working Group Meeting, 11-19 April 2017, Southwest Fisheries Science Center, La Jolla, California, USA.

Teo, S. L. H., J. Holmes, and S. Kohin. 2010. Joint standardized abundance index of US and Canada albacore troll fisheries in the North Pacific. ISC/10-3/ALBWG/01. Working document submitted to the ISC Albacore Working Group Meeting, 12-19 October 2010, Southwest Fisheries Science Center, NOAA, La Jolla, California, USA.

Xu, Y., S. L. H. Teo, and J. Holmes. 2013. An update of standardized abundance index of US and Canada albacore troll fisheries in the North Pacific (1966-2012). ISC/13/ALBWG-03/06. Working paper submitted to the ISC Albacore Working Group Workshop, 5-12 November, 2013.

TABLES

Table 1: AIC values associated with each of the GLMs for each of the time periods. The full model was selected for each time period.

Time periods	Model	AIC	deltaAIC
1966-1978	year	6823	423
	year + season	6696	296
	year + season + area	6400	0
1979-1998	year	21123	926
	year + season	21045	848
	year + season + area	20197	0
1999-2024	year	16737	259
	year + season	16635	156
	year + season + area	16479	0

Table 2: Standardized CPUE estimates for 1999-2024 from the GLM with year, season, and area effects.

Year	CPUE
1999	56.192
2000	53.778
2001	93.999
2002	93.076
2003	89.934
2004	106.39
2005	81.482
2006	145.108
2007	89.17
2008	93.072
2009	102.679
2010	81.521
2011	75.077
2013	99.673
2014	83.983
2015	98.675
2016	89.045
2017	59.887
2018	73.667
2019	105.094
2020	105.73
2021	97.382
2022	111.345
2023	63.191
2024	114.578

Table 3: Summary of coefficient significance from the GLM for ???

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2.82763	0.08757	32.291	< 2e-16	***
yearf2000	-0.04391	0.07873	-0.558	0.577029	
yearf2001	0.51451	0.07925	6.492	9.26e-11	***
yearf2002	0.50464	0.09082	5.557	2.89e-08	***
yearf2003	0.47029	0.09863	4.768	1.91e-06	***
yearf2004	0.63833	0.09940	6.422	1.47e-10	***
yearf2005	0.37160	0.10066	3.692	0.000225	***
yearf2006	0.94870	0.11113	8.536	< 2e-16	***
yearf2007	0.46176	0.11216	4.117	3.90e-05	***
yearf2008	0.50460	0.11255	4.483	7.50e-06	***
yearf2009	0.60283	0.10139	5.946	2.94e-09	***
yearf2010	0.37208	0.09370	3.971	7.25e-05	***
yearf2011	0.28973	0.09669	2.997	0.002743	**
yearf2013	0.57312	0.10333	5.546	3.06e-08	***
yearf2014	0.40183	0.12513	3.211	0.001330	**
yearf2015	0.56305	0.12085	4.659	3.26e-06	***
yearf2016	0.46036	0.12426	3.705	0.000214	***
yearf2017	0.06369	0.10181	0.626	0.531639	
yearf2018	0.27078	0.11870	2.281	0.022579	*
yearf2019	0.62608	0.11250	5.565	2.75e-08	***
yearf2020	0.63211	0.11161	5.664	1.56e-08	***
yearf2021	0.54987	0.11016	4.991	6.19e-07	***
yearf2022	0.68386	0.11468	5.963	2.64e-09	***
yearf2023	0.11738	0.10829	1.084	0.278418	
yearf2024	0.71248	0.20897	3.410	0.000656	***
seasonf3	0.41764	0.05892	7.088	1.55e-12	***
seasonf4	0.16696	0.07266	2.298	0.021607	*
areaf2	0.58173	0.06254	9.302	< 2e-16	***
areaf3	0.46749	0.08983	5.204	2.02e-07	***
areaf4	0.05136	0.07396	0.694	0.487419	
areaf5	0.45362	0.08170	5.552	2.96e-08	***
areaf6	0.54338	0.12209	4.451	8.74e-06	***
areaf7	0.23939	0.09016	2.655	0.007956	**
areaf8	0.14446	0.09413	1.535	0.124933	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 4: Comparison of coefficients estimated with the glm command, *sdmTMB* and *sdmTMB* with a spatial field.

	glm	*sdmTMB*	*sdmTMB*_spatial
Intercept	2.828	2.828	3.311
yearf2000	-0.044	-0.044	-0.072
yearf2001	0.515	0.515	0.473
yearf2002	0.505	0.505	0.45
yearf2003	0.47	0.47	0.433
yearf2004	0.638	0.638	0.623
yearf2005	0.372	0.372	0.321
yearf2006	0.949	0.949	0.894
yearf2007	0.462	0.462	0.501
yearf2008	0.505	0.505	0.452
yearf2009	0.603	0.603	0.51
yearf2010	0.372	0.372	0.235
yearf2011	0.29	0.29	0.227
yearf2013	0.573	0.573	0.522
yearf2014	0.402	0.402	0.392
yearf2015	0.563	0.563	0.502
yearf2016	0.46	0.46	0.477
yearf2017	0.064	0.064	0.047
yearf2018	0.271	0.271	0.252
yearf2019	0.626	0.626	0.583
yearf2020	0.632	0.632	0.585
yearf2021	0.55	0.55	0.482
yearf2022	0.684	0.684	0.673
yearf2023	0.117	0.117	-0.056
yearf2024	0.712	0.712	0.855
seasonf3	0.418	0.418	0.301
seasonf4	0.167	0.167	0.117
areaf2	0.582	0.582	NA
areaf3	0.467	0.467	NA
areaf4	0.051	0.051	NA
areaf5	0.454	0.454	NA
areaf6	0.543	0.543	NA
areaf7	0.239	0.239	NA
areaf8	0.144	0.144	NA

FIGURES

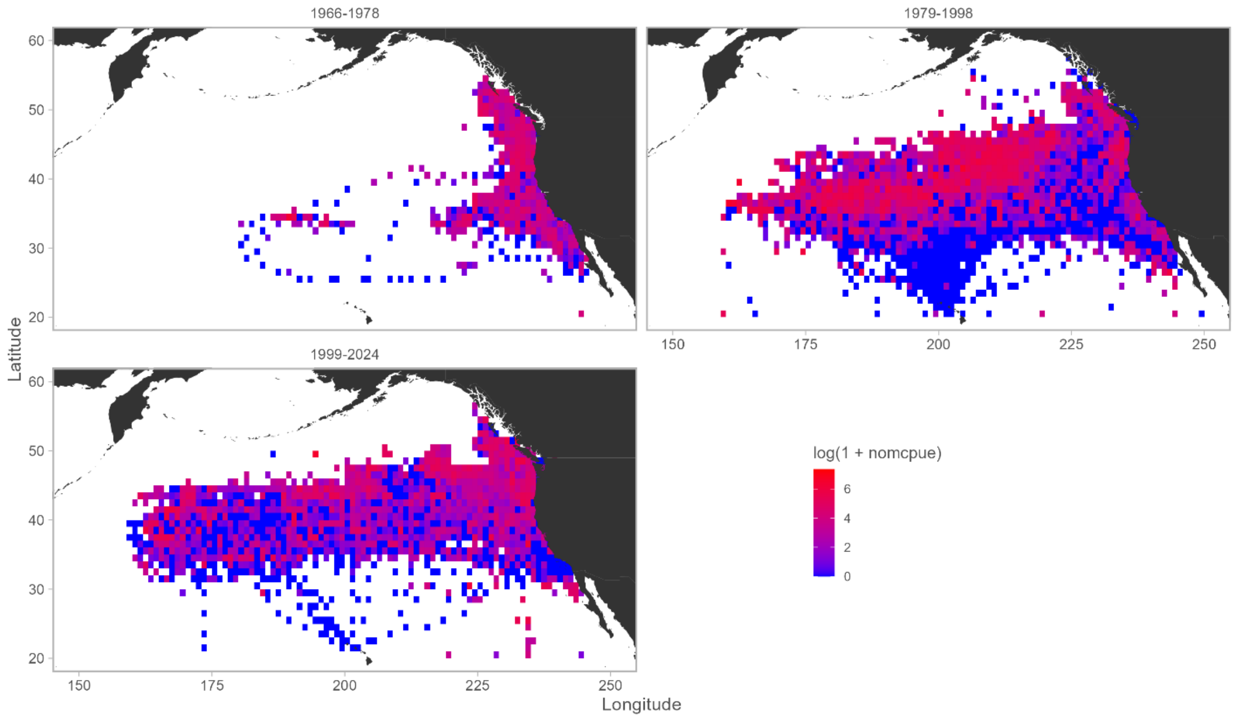

Figure 1: Nominal CPUE (total fish / number of vessel-days; plotted on log scale) values for 1x1 degree cells calculated across three time periods (1966-1978; 1979-1998; 1999-2024). Nominal CPUE was concentrated close to the coast for 1966-1978, shifted further offshore for 1979-1998, and concentrated inshore and north of 40degrees N for 1999-2024.

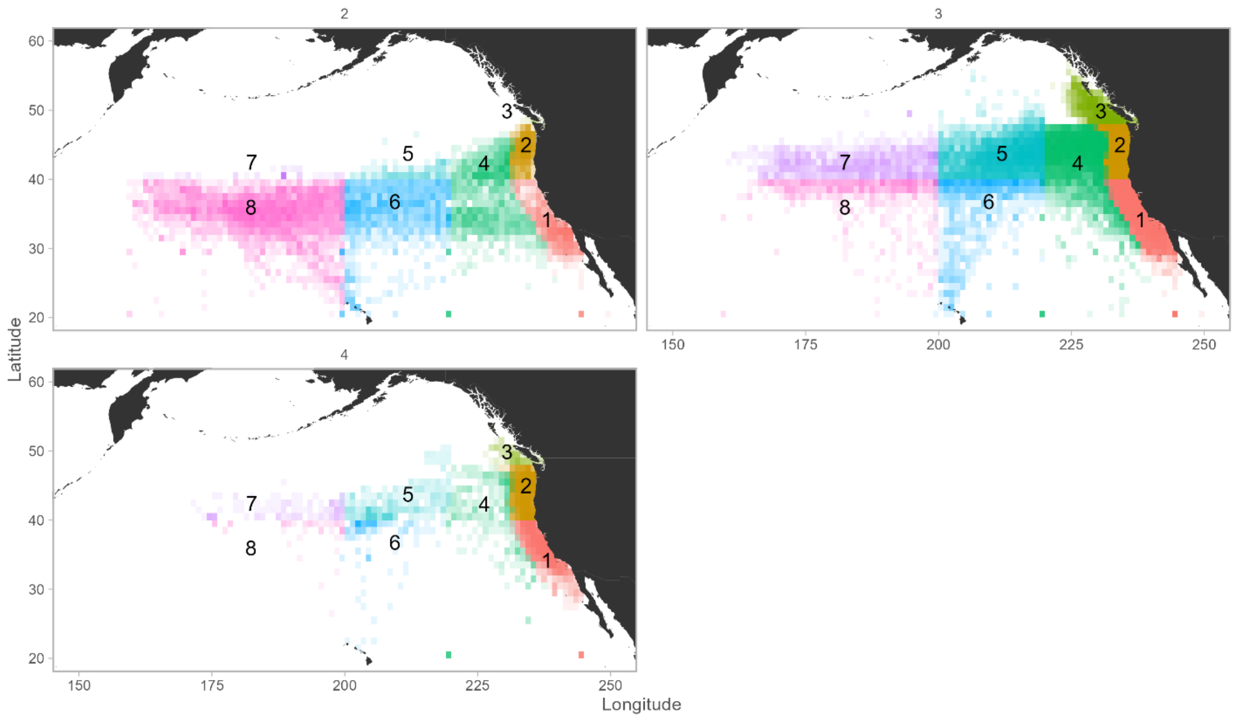

Figure 2: Nominal CPUE aggregated across all years shown for each season (2-top left; 3- top right; 4-bottom left). Colors and labels indicate areas.

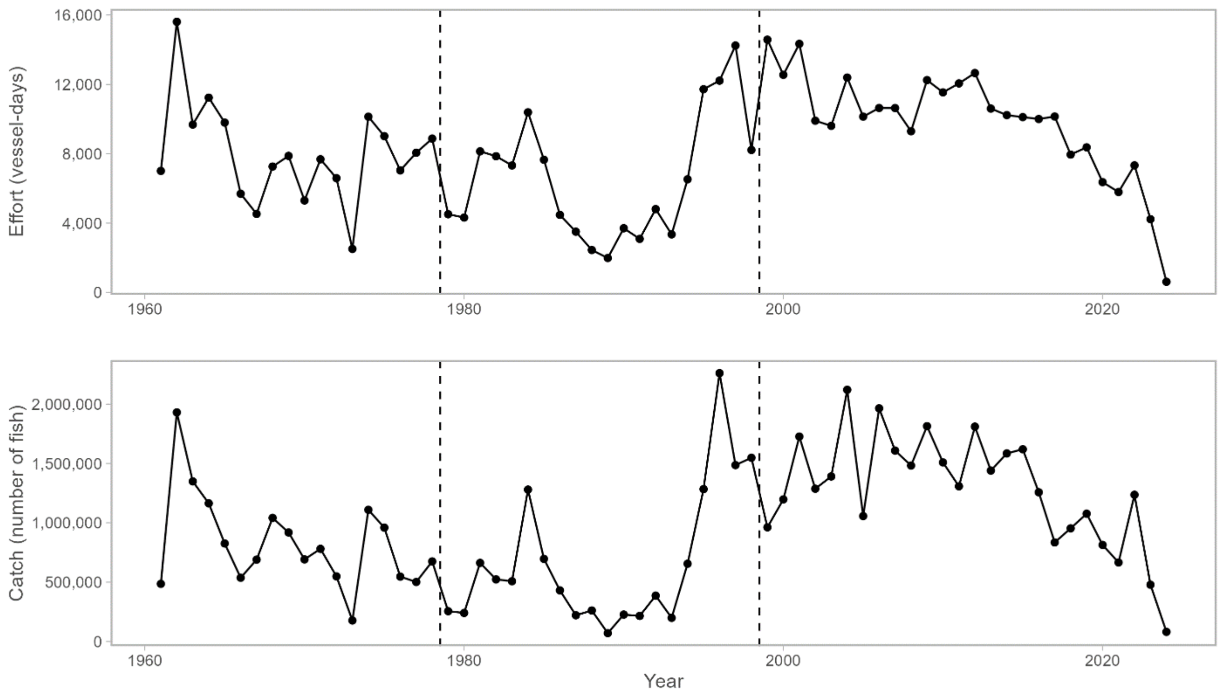

Figure 3: Annual effort (top; number of vessel-days) and catch (bottom; number of fish) by year. The values include all seasons and all areas. Vertical lines indicate the break points for the three periods (1961-1978, 1979-1998, 1999-2024).

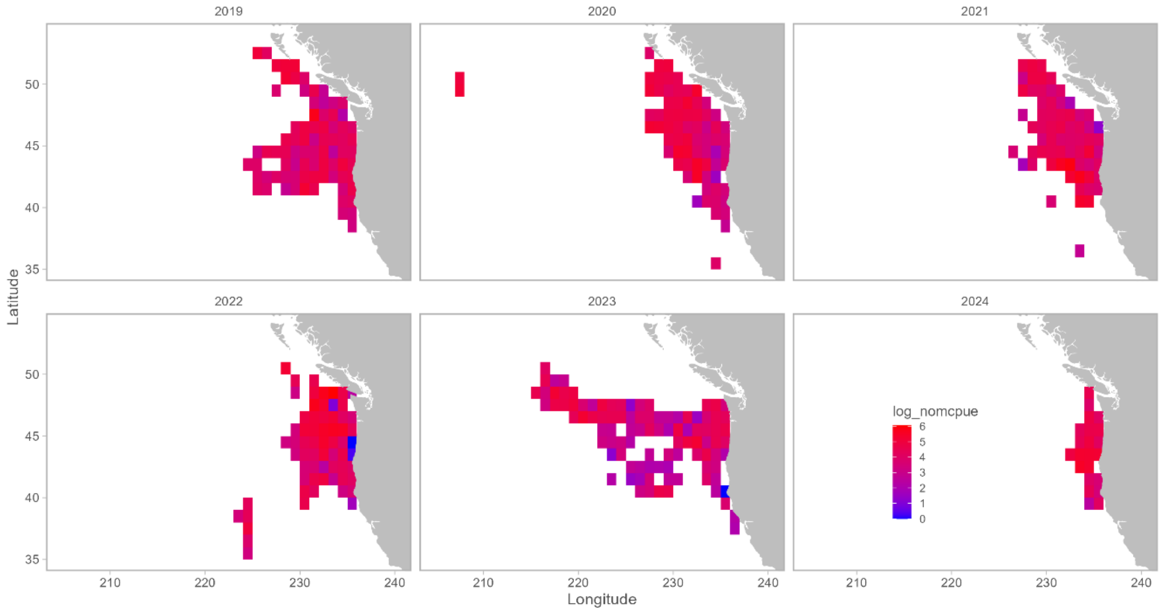

Figure 4: Nominal CPUE values (log scale) from 2019-2024 for 1x1 degree cells. Values from 2024 were from July through September.

Figure 5: Standardized CPUE values for the three time periods (rows) and three models (colors).

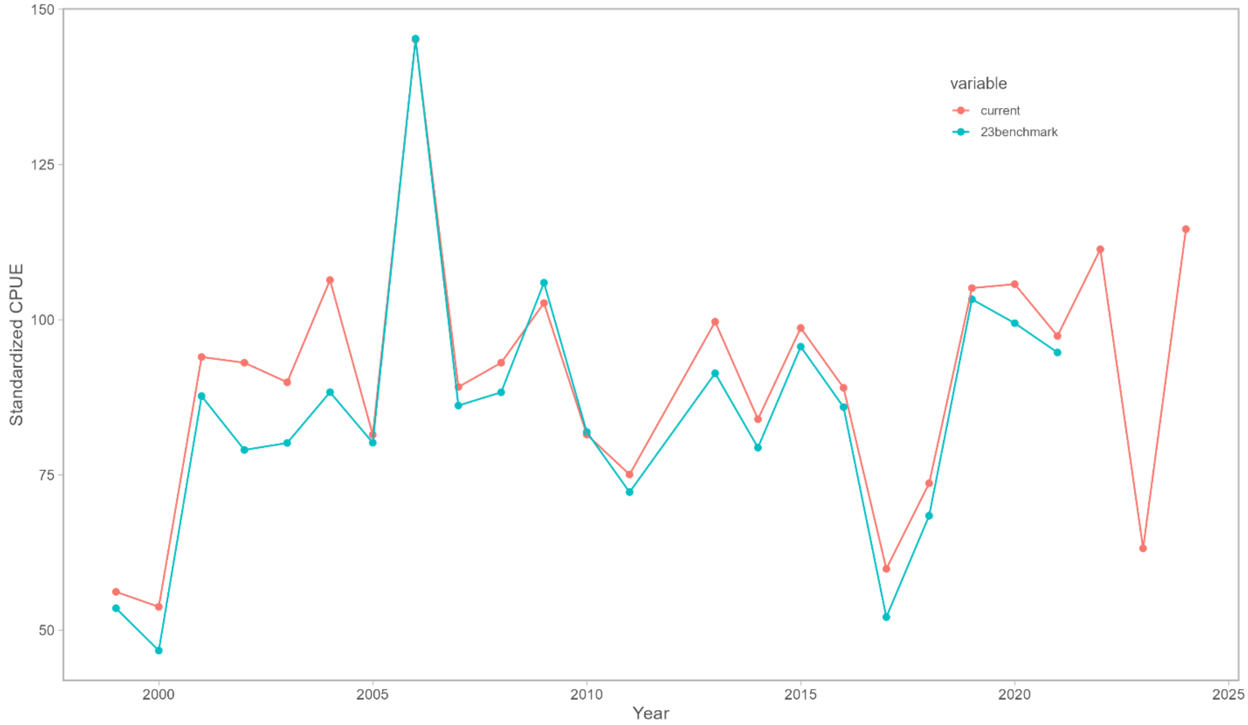

Figure 6: Standardized CPUE values estimated from the full GLM (year + season + area) input to the 2023 benchmark assessment files (blue) and the current data through 2024 (red).

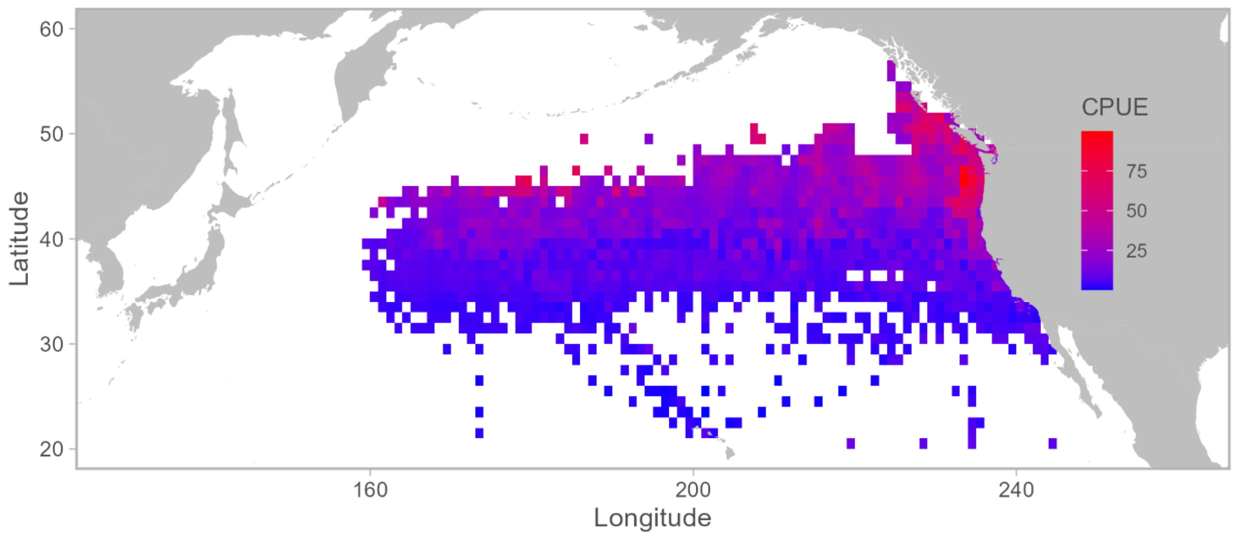

Figure 7: Spatial field estimated in *sdmTMB*. The model was fit to log(CPUE + 1) and the values shown here a transformed to linear space.