North Pacific Fisheries Commission
NPFC-2026-TWG CMSA12-WP02

Standardized CPUE of chub mackerel (*Scomber japonicus*) from the RV Songhang trawl survey in the Northwest Pacific Ocean, 2021–2026

Yi Shang, Zhengyan Jiang, Qiuyun Ma, Yufei Zhou, Bilin Liu

College of Marine Living Resource Sciences and Management, Shanghai Ocean University, China

(Corresponding author: Qiuyun Ma, qyma@shou.edu.cn)

Working document submitted to the NPFC Technical Working Group on Chub Mackerel Stock Assessment (TWG-CMSA). All results in this document were produced by the R code described in Section 2 and are reproducible from the survey database.

Summary

The nominal CPUE of Chub mackerel (*Scomber japonicus*) was standardized using fishery-independent trawl survey data collected by the RV Songhang in the Northwest Pacific Ocean between 2021 and 2026. Fifty-four candidate generalized linear mixed models were fitted with the sdmTMB package, comprising all combinations of three response distributions (Tweedie, delta-gamma and delta-lognormal), two spatial random-effect settings and nine nested covariate sets, and were compared using the Akaike information criterion (AIC), convergence and Hessian diagnostics, and five-fold spatial block cross-validation. Among the 20 models that passed all diagnostics, a delta-lognormal model with no spatial random field and with sea surface temperature and chlorophyll-a concentration as covariates was selected. The standardized abundance index was highest in 2021 and 2023, declined in 2024, recovered slightly in 2025 and reached its six-year minimum in 2026. Coefficients of variation ranged from 0.466 to 1.320, peaking in 2026. We recommend that this standardized index could be used as one of abundance indices for the Chub mackerel stock assessment.

1. Introduction

Chub mackerel (*Scomber japonicus*) is an economically important pelagic species and a principal target of China's distant-water fishery in the Northwest Pacific Ocean. It is taken mainly by purse seiners, with some catch from midwater trawls. Because the stock fluctuates markedly and its distribution and migration are strongly driven by oceanographic conditions, catch per unit effort (CPUE) must be corrected for environmental and sampling biases before it can serve as an abundance index in stock assessment.

Several recent studies describe the biology and habitat of chub mackerel in this region. Growth is heterogeneous among cohorts, and dynamic energy budget models have been used to represent the complete life cycle. Habitat studies based on linear mixed models, ensemble habitat-suitability models under future climate scenarios, trawl-based habitat associations and comparisons with Blue

mackerel consistently identify temperature and productivity as the main environmental correlates of distribution, while the Kuroshio large meander has been linked to growth, distribution and recruitment and the spatio-temporal distribution and population dynamics of the stock have been described by Zhao et al. (2025). These studies indicate which environmental variables are the natural candidates for a CPUE standardization, but they rest on commercial catch data or on habitat models rather than on a fishery-independent survey.

Since 2021, the R/V Songhang, Shanghai Ocean University's distant-water fishery research vessel, has conducted systematic integrated surveys in the Northwest Pacific Ocean, with chub mackerel as a key target species; these surveys have accumulated extensive biological, fishing-ground and environmental data. Because the catches come from a research vessel sampling designed survey stations rather than a commercial fleet, the resulting index is fishery-independent and unaffected by changes in fleet behaviour, gear or market conditions. Therefore, this study has conducted CPUE standardization to estimate abundance index for Chub mackerel in the convention area.

2. Methods

2.1 The data

The data come from a fishery-independent survey rather than a commercial fishery: catches were taken by the RV Songhang during trawl surveys in the Northwest Pacific Ocean in 2021–2026. Location, time, catch, swept area and environmental variables are recorded for each station. Stations were planned on a 1° × 1° grid, so the nominal spatial resolution is approximately 1°; the realised trawl positions deviate slightly from the nominal grid points. The surveyed area covered 147–175°E and 34–48°N; in 2026 it extended further east and north, with about a third of the stations outside the area surveyed in 2021–2025. The surveys were conducted from June to August. Table 1 summarizes data representativeness. The survey took 1.83 t in 2021 but only 1.64 kg in 2026, a very large contrast between the two ends of the series.

Eight environmental variables were used: sea surface temperature (SST), salinity (SAL), chlorophyll-a concentration (CHL), mixed layer depth (MLD), sea surface height (SSH), dissolved oxygen (DO), eastward current velocity (ECV) and northward current velocity (NCV). SST, SAL and CHL were measured in situ at each station with a shipboard CTD profiler; MLD, SSH, DO, ECV and NCV came from the Copernicus Marine Environment Monitoring Service (https://marine.copernicus.eu/), matched to the position and date of each haul. Table 3 gives the ranges, sources, resolutions and inclusion rationale of the candidate explanatory variables.

The original dataset consisted of 473 observations; no record was missing spatial or environmental information, and one haul lost to gear failure in 2024 was removed, leaving 472 samples for analysis. Of these, 293 had a non-zero catch (62.08%).

Fig. 1 shows the spatial distribution of the survey stations in each year, and Figs 2 and 3 the corresponding nominal CPUE and catch of chub mackerel. Fig. 4 gives box plots of observed CPUE

by year and scatter plots against each candidate covariate; the monthly values are shown in Fig. 9 and Table 13. Catches were extremely skewed: the largest monthly catch was 1750.07 kg in July 2021, while the proportion of zero-catch hauls was 12.0% in 2021, 27.2% in 2022, 22.4% in 2023, 66.7% in 2024, 73.2% in 2025 and 90.4% in 2026. Swept area per haul increased in later years, dispersing effort more widely over a survey area that had itself expanded east and north.

2.2 Correlation analysis of variables

The candidate explanatory variables comprised a temporal variable (year) and the environmental variables SST, CHL, MLD, SSH, SAL, DO, ECV and NCV (Table 3). Because spatial and spatiotemporal random effects are included in the model (Section 2.3), longitude and latitude were not considered as explanatory variables. Screening proceeded in three steps: pairwise Pearson correlations were computed first to identify highly correlated pairs ($|r| \geq 0.7$), with the correlation matrix shown in Fig. 5; the variance inflation factor (VIF) was then calculated to assess multicollinearity (Table 9); and the covariates were finally compared through the nested model sets described in Section 2.3. All eight environmental covariates had VIF < 5 (maximum 3.72 for SST) and were retained as candidates. Among them, SST, SSH, SAL and DO were strongly correlated with one another ($|r| = 0.52–0.71$); the only pair above the correlation threshold was SST–SSH ($r = 0.71$), and because both had VIF < 5 they were both retained, with the question of whether they should appear together settled by the nested model comparison (Section 2.3). Chlorophyll-a concentration was much less strongly correlated with the other covariates ($|r| \leq 0.46$).

2.3 Model construction and selection

Generalized linear mixed models (GLMMs) were fitted using the sdmTMB package (version 1.0.0) in R 4.5. Because the observed catches contained many zeros and low values with a few extremely large catches, Tweedie, delta-gamma and delta-lognormal distributions were evaluated as alternative response distributions for standardization.

The full model included year, sea surface temperature (SST), chlorophyll-a concentration (CHL), mixed layer depth (MLD), sea surface height (SSH), salinity (SAL), dissolved oxygen (DO), eastward current velocity (ECV) and northward current velocity (NCV) as fixed effects. Year was treated as a categorical variable to estimate annual abundance indices, whereas the remaining variables were included as continuous covariates. The general form of the fitted model was:

$$\eta i = Xi\beta + \omega(si) + \delta(si, t)$$

where ηi is the linear predictor, $Xi\beta$ represents the fixed effects, $\omega(si)$ is the spatial random effect accounting for spatial autocorrelation, and $\delta(si, t)$ is the spatiotemporal random effect accounting for spatiotemporal variation, which can be modelled as an independent and identically distributed (IID) process (spatiotemporal = "iid", time = Year). Candidate models either included or excluded the spatial random effect ω, represented by a Gaussian random field estimated with the stochastic partial differential equation (SPDE) approach; a mesh built from the survey station coordinates with a 200 km minimum vertex distance gave 73 nodes.

The environmental covariates were added one at a time in the order SST, CHL, MLD, SSH, SAL, DO, ECV and NCV, giving nine nested sets from year alone to the full model; combining these with the three response distributions and the two spatial settings gave 3 × 2 × 9 = 54 candidate models (fit1–fit54; Table 8).

Model performance was evaluated using Akaike's Information Criterion (AIC), convergence diagnostics and Hessian matrix evaluation, together with ΔAIC, the coefficient of determination on the log scale (R^2) and the root-mean-square error relative to the mean observed CPUE (NRMSE) (Table 5). Each model was assigned a diagnostic status: √ (converged, positive-definite Hessian, no collapsed random-effect variance), △ (converged with a positive-definite Hessian but a random-effect standard deviation collapsed to zero, sigma_E < 0.001) or × (convergence failure or non-positive-definite Hessian). Models with status × or △ were excluded.

Because the leading models differed by less than the 2-unit AIC threshold used to distinguish them, five-fold spatial block cross-validation was used as an independent, out-of-sample criterion. Folds were defined by k-means clustering of the station coordinates into five spatially contiguous blocks, so that neighbouring stations were always held out together: with random fold assignment the training set would contain stations adjacent to almost every validation station, which inflates the scores of models able to fit local spatial pattern and would bias the comparison between the spatial settings themselves. Withholding whole blocks instead reflects the applied task of predicting abundance in areas not yet surveyed. Block sizes were 113, 99, 112, 31 and 117 stations. Each model was refitted on the remaining folds, recording the expected log pointwise predictive density (elpd_cv), the cross-validation root-mean-square error (cv_RMSE) and the number of folds with a positive-definite Hessian (Table 10).

The selection rule was fixed in advance: (1) discard models with status × or △; (2) identify the minimum AIC among the remainder; (3) within ΔAIC < 2 of that minimum, select the model with the best cross-validated predictive density and the most stable Hessian across folds.

2.4 Extraction of annual trends

Annual standardized CPUE indices were extracted with predict() and get_index() from sdmTMB. Predictions were generated for each year on the same footprint — all 472 station positions, with the environmental covariates held at their observed values — and the expected catches were summed with get_index() to give the annual index. Because the footprint includes stations that were surveyed only in 2026, the earlier years partly extrapolate into those areas (Section 4). Holding the covariates fixed removes confounding by year-to-year differences in the environmental conditions encountered. Because the expected catch of a haul under the delta-lognormal model is the product of its two components, the index combines the probability of encountering chub mackerel with the catch rate when they are present. Both indices were normalized by their means over the survey period for comparison:

$$RI_t = I_t / \text{mean}(I)$$

The nominal index was calculated as the ratio of the total catch to the total swept area over the hauls that carry an effort record in each survey year (all 472 hauls; Table 13). Uncertainty in the standardized index was quantified by the standard error returned by get_index(), which is on the log scale, and the coefficient of variation was calculated as:

$$\mathrm{CV} = \mathrm{sqrt}(\exp(\mathrm{SE}^2) - 1)$$

Confidence intervals were taken from the limits returned by get_index(), computed as exp(log(index) ± 1.96 × SE). Index sensitivity to mesh resolution was tested by refitting the selected model at 150, 200, 300 and 400 km (Table 11).

3. Result

3.1 Model selection and diagnostics

A total of 54 candidate generalized linear mixed models were fitted using combinations of Tweedie, delta-gamma and delta-lognormal distributions with different spatial and spatiotemporal structures (Table 5). The observed catches were highly skewed, with a large proportion of zero or low catches and several extremely high catch observations (Fig. 4). Model performance was evaluated based on the Akaike information criterion (AIC), convergence diagnostics, and Hessian matrix evaluation. The lowest AIC among all models with results was obtained by the delta-lognormal model fit51 (AIC = 2024.6), but this model was rejected because its spatiotemporal random-effect standard deviation collapsed to zero (sigma_E = 0.0002). Among the models that passed all diagnostics, fit41 and fit39 had the lowest AIC values (both 2036) and differed by only 0.07, which does not distinguish them, and cross-validation gave fit39 both the better predictive density and the more stable Hessian across folds (Table 10). fit39 was therefore selected as the final model for CPUE standardization.

Parameter estimates of the selected model are presented in Table 6. Because fit39 is a delta-lognormal model it has two components. In the occurrence component, the year effects for 2023–2026 were all significantly negative and became progressively stronger, and chlorophyll-a concentration was significantly negative, whereas sea surface temperature was not significant. In the positive-catch component, only chlorophyll-a concentration was significant; neither sea surface temperature nor any year effect was significant. All estimated coefficients were biologically reasonable and were associated with acceptable standard errors, indicating stable parameter estimation and successful model convergence.

Model diagnostics suggested that the selected model adequately described the observed data (Fig. 6). The residual histogram and Q–Q plot indicated that the residuals approximately followed a normal distribution, while the residual-versus-fitted plot showed no obvious systematic pattern. Correlations between the residuals and all explanatory variables were very small ($|r| \leq 0.066$; Table 12), indicating that no strong systematic structure remained. The explanatory power of the model is nevertheless low ($R^2 = 0.179$ on the log scale and 0.0024 on the raw scale; NRMSE = 7.41), which

reflects the extreme skew of the catch data rather than a misspecification of the environmental relationships; the index should therefore be treated as a relative indicator only.

3.2 Effects of explanatory variables

The estimated partial effects of the covariates retained in the selected model on the expected CPUE are presented in Fig. 7. The expected CPUE increased with chlorophyll-a concentration (CHL), rising steeply and non-linearly on the response scale, from about 2 kg at the lowest observed CHL (0.07 mg/m³) to about 140 kg at the 99th percentile of observed CHL (3.4 mg/m³), which is consistent with the planktivorous feeding ecology of chub mackerel. The effect of sea surface temperature (SST) was slightly negative but essentially flat over the observed range (8–28 °C). Because the selected model is a delta model, CHL enters the two components with opposite signs: it is negative in the occurrence component and positive in the positive-catch component. The index is the product of the two components, so the positive effect on the catch rate dominates and the net partial effect on the expected CPUE is strongly positive, as Fig. 7 shows. Such sign reversal is not uncommon when covariates are spatially structured, and the partial effect on the expected value is the quantity that should be interpreted.

Sea surface temperature was not significant in either component, and its coefficient is difficult to interpret on its own: SST is strongly correlated with latitude ($r = -0.80$), so its independent effect cannot be separated from the latitudinal gradient. The remaining candidate covariates (MLD, SSH, SAL, DO, ECV and NCV) were not retained in the selected model. Once SST and CHL were included, adding them did not lower the AIC by more than the 2-unit threshold used to distinguish models (Table 5), although the four-covariate model fit41 was nominally 0.07 AIC units better than fit39 and was carried into the cross-validation comparison on that basis.

3.3 Standardized abundance index

The standardized abundance index and nominal index values from 2021 to 2026 are presented in Fig. 6, and the corresponding estimates are summarized in Table 7. Compared with the nominal index, the standardized abundance index represents a relative abundance estimate after accounting for the effects of environmental covariates, year effects, and spatiotemporal variability included in the model.

The standardized and nominal indices did not follow the same trajectory. The nominal index rose from 2021, reached its maximum in 2024 and fell sharply thereafter, whereas the standardized index was highest in 2021, declined in 2022, recovered in 2023 and then fell to its lowest values in 2025–2026. The standardized index was therefore higher than the nominal index in 2021, 2022, 2023, 2025 and 2026, but markedly lower in 2024, when the nominal index reached its peak. The largest discrepancy occurred in 2024: the nominal index ranked that year first of the six, whereas the standardized index ranked it fifth, below 2021, 2022, 2023 and 2025.

The coefficients of variation of the standardized abundance index ranged from 0.466 to 1.320, the highest values occurring in the years of lowest abundance.

The sensitivity of the index to the resolution of the spatial mesh is summarized in Table 11.

The 200 km row reproduces the relative index of Table 7 exactly, and across 150, 200 and 300 km the relative index varies by at most 6% and the ranking of the years is unchanged. The 400 km refit is reported for completeness only: it did not reach a positive-definite Hessian and is therefore excluded under the same diagnostic rule that was applied to the candidate models in Section 2.3. The index is thus insensitive to the mesh resolution over the range in which the fitted model is numerically sound, which is expected given that the selected model does not retain a persistent spatial field.

4. Discussion

The two indices agree that 2026 was the year of lowest abundance, but they rank the intervening years very differently: the nominal index peaks in 2024, whereas the standardized index peaks in 2021 and places 2024 fifth of the six years. This divergence shows that the raw ratio of total catch to total swept area does not control for the conditions under which the hauls were made, whereas the standardized model estimates a common year effect while holding the environmental covariates at their observed values. The contrast is a concrete illustration of why nominal survey CPUE should not be used directly as an abundance index. By incorporating environmental covariates together with the year effect and the spatiotemporal random effect, the selected delta-lognormal model produced an index less affected by sampling conditions.

The uncertainty of the standardized index is high, with coefficients of variation between 0.466 and 1.320, and it is higher in the later years of the series. Several factors contribute: the time series covers only six survey years; only 293 of the 472 station records had a non-zero catch, and chub mackerel showed pronounced spatial clustering; and the explanatory power of the model is low, so the index depends heavily on the estimated year effects and random effects. The 2026 value requires particular care, because the surveyed area extended markedly further east and north than in the preceding five years (to 174.02°E and 47.98°N), with 19 of the 52 stations lying outside the area surveyed during 2021–2025; the earlier indices therefore incorporate extrapolation into these newly added areas. Only 5 of the 52 stations in 2026 yielded chub mackerel catch, so the 2026 index is estimated from very little information and its confidence interval is correspondingly wide. For these reasons the index should be used to describe relative interannual change and not as an absolute measure of biomass.

The monitor programme reported here follows a systematic design and is fishery-independent, and it is among the more extensive data sets currently available for chub mackerel on the high seas of the Northwest Pacific Ocean, where comparable fishery-independent information remains limited. The index was derived with a pre-specified model-selection rule and is supported by convergence, Hessian and cross-validation diagnostics, which indicate that it is suitable for use as a relative measure of interannual change in abundance. We recommend that this standardized index could be used as one of the input abundance indices for the Chub mackerel stock assessment., and that it be updated as further survey years become available.

Acknowledgements

This work was supported by the National Key R&D Program of China, the National Natural Science Foundation of China , and the High Seas Fishery Resources Survey Programme. We thank the entire survey team of the R/V Songhang for their dedicated work in carrying out the surveys and collecting the data used in this study.

References

Akaike H. Factor analysis and AIC[J]. Psychometrika, 1987, 52(3): 317-332.

Anderson S C, Ward E J, English P A, et al. sdmTMB: an R package for fast, flexible, and user-friendly generalized linear mixed effects models with spatial and spatiotemporal random fields[J]. Journal of Statistical Software, 2025, 115(2): 1-46.

Cai K, Kindong R, Ma Q, et al. Growth heterogeneity of chub mackerel (*Scomber japonicus*) in the Northwest Pacific Ocean[J]. Journal of Marine Science and Engineering, 2022, 10(2): 301.

Dormann C F, Elith J, Bacher S, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance[J]. Ecography, 2013, 36(1): 27-46.

He W, Guan W, Cao R. Dynamic energy budget model for the complete life cycle of chub mackerel in the Northwest Pacific[J]. Fisheries Research, 2024, 270: 106902.

Jiang Z, Zhou Y, Ma Q, et al. Preliminary standardized fishery-independent index of Japanese sardine (*Sardinops melanostictus*) based on the Chinese research vessel “Song hang” from 2021 to 2025[R]. NPFC-2026-SSC JS01-WP05, 2026.

Kindong R, Sarr O, Wang J Q, et al. Size distribution patterns of silky shark *Carcharhinus falciformis* shaped by environmental factors in the Pacific Ocean[J]. Science of the Total Environment, 2022, 850: 157927.

Kristensen K, Nielsen A, Berg C W, et al. TMB: automatic differentiation and Laplace approximation[J]. Journal of Statistical Software, 2016, 70(5): 1-21.

Li J, Tang F, Wu Y, et al. A study on the impact of environmental factors on chub mackerel *Scomber japonicus* fishing grounds based on a linear mixed model[J]. Fishes, 2024, 9(8): 323.

Lindgren F, Rue H, Lindstrom J. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach[J]. Journal of the Royal Statistical Society, 2011, 73(4): 423-498.

Lo N C H, Jacobson L D, Squire J L. Indices of relative abundance from fish spotter data based on delta-lognormal models[J]. Canadian Journal of Fisheries and Aquatic Sciences, 1992, 49(12): 2515-2526.

Maunder M N, Punt A E. Standardizing catch and effort data: a review of recent approaches[J].

Fisheries Research, 2004, 70(2-3): 141-159.

North Pacific Fisheries Commission. Conservation and Management Measure for Chub Mackerel (CMM 2025-07)[R]. Tokyo: NPFC, 2025.

North Pacific Fisheries Commission. Stock assessment report for chub mackerel – 2025[R]. Tokyo: NPFC, 2026.

Shono H. Application of the Tweedie distribution to zero-catch data in CPUE analysis[J]. Fisheries Research, 2008, 93(1-2): 154-162.

Stefánsson G. Analysis of groundfish survey abundance data: combining the GLM and delta approaches[J]. ICES Journal of Marine Science, 1996, 53(3): 577-588.

Sun Y, Zhang H, Jiang K, et al. Simulating the changes of the habitats suitability of chub mackerel (*Scomber japonicus*) in the high seas of the North Pacific Ocean using ensemble models under medium to long-term future climate scenarios[J]. Marine Pollution Bulletin, 2024, 207: 116873.

Thorson J T. Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments[J]. Fisheries Research, 2019, 210: 143-161.

Thorson J T, Shelton A O, Ward E J, et al. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes[J]. ICES Journal of Marine Science, 2015, 72(5): 1297-1310.

Wang Z, Ito S. Impacts of Kuroshio large meander on growth, distribution, and recruitment of chub mackerel (*Scomber japonicus*) during 2018-2024 in the Northwestern Pacific[J]. Progress in Oceanography, 2026, 247: 103784.

Zhang H, Zhu H, Zhang F, et al. Analysis of relative abundance distribution and environmental differences for blue mackerel (*Scomber australasicus*) and chub mackerel (*Scomber japonicus*) on the high seas of the North Pacific Ocean[J]. Animals, 2025, 15(19): 2822.

Zhao G, Zhang H, Tang F. The spatio-temporal distribution and population dynamics of chub mackerel (*Scomber japonicus*) in the high seas of the Northwest Pacific Ocean[J]. Animals, 2025, 15(8): 1135.

Zhao J, Ye X C, Liu Z L, et al. Comparison of sampling designs for fishery resource surveys in the Northwest Pacific Ocean[J]. Journal of Shanghai Ocean University, 2024, 33(4).

Zou X, Zhou J, Xu H, et al. Association between the habitat of Pacific chub mackerel in the northwest Pacific Ocean and environmental factors based on trawl fisheries data[J]. Marine and Coastal Fisheries, 2026, 18(5): vtag012.

Tables

Table 1. Catch and effort information for the R/V Songhang trawl survey (CPUE fleet), 2021–2026

Year	No. of stations	Stations with positive catch	Total catch of CPUE fleet (MT)	Total effort of CPUE fleet (swept area, km²)	Nominal index (kg/km²)
2021	142	125	1.830	45.119	40.56
2022	92	67	0.215	17.971	11.97
2023	85	66	0.720	31.280	23.00
2024	45	15	0.701	14.829	47.25
2025	56	15	0.089	13.481	6.58
2026	52	5	0.002	16.002	0.10

Table 2. Filter "rules" used on the data for CPUE standardization and the effect on the overall sample size

Filter applied	Number of records remaining	Number removed	Number of records with chub mackerel catch > 0
Initial dataset	473	—	293
Remove records with missing spatial or environmental variables	473	0	293
Remove the haul with a gear failure	472	1	293
Final dataset	472	—	293

Table 3. Summary of the explanatory variables considered in the CPUE standardization

Variable	Abbrev.	Range	Unit	Data source and resolution	Rationale for inclusion
Year	Year	2021–2026	—	Survey record; one value per survey year	Temporal variable; required to estimate the annual abundance index
Sea surface temperature	SST	7.99–28.10	°C	Shipboard CTD, in situ; haul level	Influences the distribution and migration of chub mackerel and the location of fishing grounds
Chlorophyll-a concentration	CHL	0.07–5.11	mg/m³	Shipboard CTD, in situ; haul level	Proxy for primary productivity and food availability for a planktivorous species
Mixed layer depth	MLD	10.38–37.31	m	CMEMS reanalysis; matched to haul position and date	Affects vertical mixing and the vertical distribution of prey
Sea surface	SSH	−0.29–	m	CMEMS reanalysis;	Proxy for mesoscale eddies and frontal

height		1.28		matched to haul position and date	structures that aggregate pelagic fish
Salinity	SAL	32.21–34.82	PSU	Shipboard CTD, in situ; haul level	Describes water-mass structure and habitat suitability
Dissolved oxygen	DO	206.99–339.69	µmol/L	CMEMS reanalysis; matched to haul position and date	Affects habitat suitability and vertical distribution
Eastward current velocity	ECV	−0.86–1.78	m/s	CMEMS reanalysis; matched to haul position and date	Affects transport, retention and aggregation of pelagic fish
Northward current velocity	NCV	−1.39–1.15	m/s	CMEMS reanalysis; matched to haul position and date	Affects transport, retention and aggregation of pelagic fish

Table 4. Result of the model selection: full, null and selected models

Role	Model	Distribution family / spatial	Variables	AIC	ΔAIC	R^2 (log)	NRMSE
Null (year effect only)	fit10	Tweedie / on	Year	2429	393	0.068	7.406
Full (year + all 8 covariates)	fit18	Tweedie / on	Year + SST + CHL + MLD + SSH + SAL + DO + ECV + NCV	2401	365	0.121	7.400
Minimum AIC among models passing all diagnostics	fit41	delta-lognormal / off	Year + SST + CHL + MLD + SSH	2036	0	0.181	7.402
Selected model	fit39	delta-lognormal / off	Year + SST + CHL	2036	0.07	0.178	7.407

Table 5. Comparison of the 54 candidate models based on AIC, convergence diagnostics and model performance

Model	Diagnostics	AIC	ΔAIC	R^2 (log)	NRMSE
fit41	√	2036	0	0.181	7.402
fit39	√	2036	0	0.178	7.407
fit16	√	2399	363	0.120	7.400
fit12	√	2400	364	0.155	7.401
fit17	√	2401	365	0.120	7.400
fit7	√	2401	365	0.142	7.395
fit15	√	2401	365	0.167	7.396

fit3	√	2401	365	0.160	7.397
fit13	√	2401	365	0.154	7.401
fit6	√	2401	365	0.170	7.392
fit18	√	2401	365	0.121	7.400
fit4	√	2402	366	0.159	7.397
fit14	√	2402	366	0.161	7.399
fit5	√	2402	366	0.164	7.396
fit8	√	2402	367	0.141	7.396
fit9	√	2403	367	0.141	7.395
fit10	√	2429	393	0.068	7.406
fit11	√	2431	395	0.093	7.404
fit1	√	2431	395	0.067	7.404
fit2	√	2432	396	0.099	7.399

Table 6. Estimated coefficients and uncertainty in the selected model

Component	Variable	Estimate	SE	95% CI	Significance
1: Occurrence probability (logit)	(Intercept)	3.3669	0.9096	(1.584, 5.150)	***
1: Occurrence probability (logit)	fYear2022	-0.9160	0.4943	(-1.885, 0.053)	ns
1: Occurrence probability (logit)	fYear2023	-1.1573	0.5143	(-2.165, -0.149)	***
1: Occurrence probability (logit)	fYear2024	-2.9396	0.5521	(-4.022, -1.857)	***
1: Occurrence probability (logit)	fYear2025	-3.1908	0.5280	(-4.226, -2.156)	***
1: Occurrence probability (logit)	fYear2026	-4.8736	0.6692	(-6.185, -3.562)	***
1: Occurrence probability (logit)	SST	-0.0340	0.0388	(-0.110, 0.042)	ns
1: Occurrence probability (logit)	CHL	-0.7593	0.2005	(-1.152, -0.366)	***
2: Positive catch (log)	(Intercept)	1.1376	0.8428	(-0.514, 2.789)	ns
2: Positive catch (log)	fYear2022	-0.3602	0.4354	(-1.214, 0.493)	ns
2: Positive catch (log)	fYear2023	0.4774	0.4325	(-0.370, 1.325)	ns
2: Positive catch (log)	fYear2024	0.1019	0.6062	(-1.086, 1.290)	ns
2: Positive catch (log)	fYear2025	0.5697	0.5853	(-0.577, 1.717)	ns
2: Positive catch (log)	fYear2026	-0.8456	0.8725	(-2.556, 0.864)	ns
2: Positive catch (log)	SST	-0.0086	0.0370	(-0.081, 0.064)	ns
2: Positive catch (log)	CHL	1.5171	0.1611	(1.201, 1.833)	***
Random effect (component	range	450.9304	337.8557	(103.837,	—

1)				1958.236)	
Random effect (component 1)	sigma_E	0.5938	0.3335	(0.198, 1.785)	—
Random effect (component 2)	range	470.1958	271.7946	(151.442, 1459.863)	—
Random effect (component 2)	phi	1.7159	0.0803	(1.566, 1.881)	—
Random effect (component 2)	sigma_E	0.6076	0.2189	(0.300, 1.231)	—

Table 7. Nominal and standardized CPUE of the R/V Songhang trawl survey, 2021–2026

Year	Nominal CPUE (kg/km²)	Standardized CPUE index	SE (log scale)	CV	95% CI lower	95% CI upper	Relative index
2021	40.56	17057.58	0.462	0.487	6902.82	42151.06	2.356
2022	11.97	5287.42	0.443	0.466	2217.29	12608.54	0.730
2023	23.00	14891.07	0.587	0.641	4715.11	47028.36	2.057
2024	47.25	2624.29	0.614	0.677	787.84	8741.48	0.362
2025	6.58	3398.11	0.608	0.668	1032.69	11181.66	0.469
2026	0.10	181.52	1.004	1.320	25.35	1299.60	0.025

Table 8. Specifications of the 54 candidate sdmTMB models used for CPUE standardization

Model	Distribution family	spatial	spatiotemporal	Variables
fit1	Tweedie	off	iid	factor(Year)
fit2	Tweedie	off	iid	factor(Year)+SST
fit3	Tweedie	off	iid	factor(Year)+SST+CHL
fit4	Tweedie	off	iid	factor(Year)+SST+CHL+MLD
fit5	Tweedie	off	iid	factor(Year)+SST+CHL+MLD+SSH
fit6	Tweedie	off	iid	factor(Year)+SST+CHL+MLD+SSH+SAL
fit7	Tweedie	off	iid	factor(Year)+SST+CHL+MLD+SSH+SAL+DO
fit8	Tweedie	off	iid	factor(Year)+SST+CHL+MLD+SSH+SAL+DO+ECV
fit9	Tweedie	off	iid	factor(Year)+SST+CHL+MLD+SSH+SAL+DO+ECV+NCV
fit10	Tweedie	on	iid	factor(Year)
fit11	Tweedie	on	iid	factor(Year)+SST
fit12	Tweedie	on	iid	factor(Year)+SST+CHL
fit13	Tweedie	on	iid	factor(Year)+SST+CHL+MLD
fit14	Tweedie	on	iid	factor(Year)+SST+CHL+MLD+SSH
fit15	Tweedie	on	iid	factor(Year)+SST+CHL+MLD+SSH+SAL
fit16	Tweedie	on	iid	factor(Year)+SST+CHL+MLD+SSH+SAL+DO
fit17	Tweedie	on	iid	factor(Year)+SST+CHL+MLD+SSH+SAL+DO+ECV
fit18	Tweedie	on	iid	factor(Year)+SST+CHL+MLD+SSH+SAL+DO+ECV+NCV
fit19	delta_gamma	off	iid	factor(Year)
fit20	delta_gamma	off	iid	factor(Year)+SST
fit21	delta_gamma	off	iid	factor(Year)+SST+CHL
fit22	delta_gamma	off	iid	factor(Year)+SST+CHL+MLD
fit23	delta_gamma	off	iid	factor(Year)+SST+CHL+MLD+SSH
fit24	delta_gamma	off	iid	factor(Year)+SST+CHL+MLD+SSH+SAL
fit25	delta_gamma	off	iid	factor(Year)+SST+CHL+MLD+SSH+SAL+DO
fit26	delta_gamma	off	iid	factor(Year)+SST+CHL+MLD+SSH+SAL+DO+ECV
fit27	delta_gamma	off	iid	factor(Year)+SST+CHL+MLD+SSH+SAL+DO+ECV+NCV
fit28	delta_gamma	on	iid	factor(Year)
fit29	delta_gamma	on	iid	factor(Year)+SST
fit30	delta_gamma	on	iid	factor(Year)+SST+CHL
fit31	delta_gamma	on	iid	factor(Year)+SST+CHL+MLD
fit32	delta_gamma	on	iid	factor(Year)+SST+CHL+MLD+SSH
fit33	delta_gamma	on	iid	factor(Year)+SST+CHL+MLD+SSH+SAL
fit34	delta_gamma	on	iid	factor(Year)+SST+CHL+MLD+SSH+SAL+DO

fit35	delta_gamma	on	iid	factor(Year)+SST+CHL+MLD+SSH+SAL+DO+ECV
fit36	delta_gamma	on	iid	factor(Year)+SST+CHL+MLD+SSH+SAL+DO+ECV+NCV
fit37	delta_lognormal	off	iid	factor(Year)
fit38	delta_lognormal	off	iid	factor(Year)+SST
fit39	delta_lognormal	off	iid	factor(Year)+SST+CHL
fit40	delta_lognormal	off	iid	factor(Year)+SST+CHL+MLD
fit41	delta_lognormal	off	iid	factor(Year)+SST+CHL+MLD+SSH
fit42	delta_lognormal	off	iid	factor(Year)+SST+CHL+MLD+SSH+SAL
fit43	delta_lognormal	off	iid	factor(Year)+SST+CHL+MLD+SSH+SAL+DO
fit44	delta_lognormal	off	iid	factor(Year)+SST+CHL+MLD+SSH+SAL+DO+ECV
fit45	delta_lognormal	off	iid	factor(Year)+SST+CHL+MLD+SSH+SAL+DO+ECV+NCV
fit46	delta_lognormal	on	iid	factor(Year)
fit47	delta_lognormal	on	iid	factor(Year)+SST
fit48	delta_lognormal	on	iid	factor(Year)+SST+CHL
fit49	delta_lognormal	on	iid	factor(Year)+SST+CHL+MLD
fit50	delta_lognormal	on	iid	factor(Year)+SST+CHL+MLD+SSH
fit51	delta_lognormal	on	iid	factor(Year)+SST+CHL+MLD+SSH+SAL
fit52	delta_lognormal	on	iid	factor(Year)+SST+CHL+MLD+SSH+SAL+DO
fit53	delta_lognormal	on	iid	factor(Year)+SST+CHL+MLD+SSH+SAL+DO+ECV
fit54	delta_lognormal	on	iid	factor(Year)+SST+CHL+MLD+SSH+SAL+DO+ECV+NCV

Table 9. Variance inflation factors (VIF) of the explanatory variables

Variable	VIF	Variable	VIF
SSH	2.20	SAL	1.83
SST	3.72	DO	2.16
CHL	1.36	ECV	1.05
MLD	1.09	NCV	1.02

Table 10. Five-fold spatial block cross-validation of the leading candidate models

Model	Distribution family / spatial	Variables	elpd_cv	cv_RMSE	Folds with positive-definite Hessian
fit39	delta_lognormal / off	factor(Year)+SST+CHL	-1058.37	208.053	2/5
fit41	delta_lognormal / off	factor(Year)+SST+CHL+MLD+SSH	-1067.85	209.821	2/5
fit40	delta_lognormal / off	factor(Year)+SST+CHL+MLD	-1068.07	207.798	3/5
fit49	delta_lognormal / on	factor(Year)+SST+CHL+MLD	-1050.00	205.050	1/5
fit3	tweedie / off	factor(Year)+SST+CHL	-5717.20	202.898	5/5
fit7	tweedie / off	factor(Year)+SST+CHL+MLD+SSH+SAL+DO	-6118.05	206.032	5/5
fit12	tweedie / on	factor(Year)+SST+CHL	-7188.77	202.826	5/5
fit15	tweedie / on	factor(Year)+SST+CHL+MLD+SSH+SAL	-7457.48	202.949	5/5

Table 11. Sensitivity of the standardized index to the resolution of the spatial mesh (relative index)

Mesh spacing (km)	No. of mesh nodes	2021	2022	2023	2024	2025	2026
150	91	2.372	0.712	2.056	0.360	0.475	0.025
200	73	2.356	0.730	2.057	0.362	0.469	0.025
300	45	2.414	0.688	2.055	0.361	0.457	0.024
400	37	2.134	0.954	1.941	0.401	0.540	0.030

Table 12. Correlation coefficients between the model residuals and the explanatory variables

Variable	SST	CHL	MLD	SSH	SAL	DO	ECV	NCV
Pearson r	-0.058	0.032	-0.063	-0.066	0.030	-0.000	0.007	0.019

Table 13. Monthly catch, effort and nominal CPUE of the R/V Songhang trawl survey

Year	Month	No. of stations	Catch (kg)	Records with effort	Effort (km²)	Nominal CPUE (kg/km²)
2021	6	36	78.12	36	11.315	6.90
2021	7	90	1750.07	90	29.336	59.66
2021	8	16	1.79	16	4.467	0.40
2022	6	42	34.86	42	8.460	4.12
2022	7	50	180.34	50	9.511	18.96
2023	6	38	82.32	38	14.003	5.88
2023	7	45	636.75	45	16.562	38.45
2023	8	2	0.52	2	0.715	0.72
2024	6	15	137.22	15	5.130	26.75
2024	7	30	563.39	30	9.699	58.09
2025	6	17	20.05	17	4.282	4.68
2025	7	26	50.83	26	6.233	8.15
2025	8	13	17.81	13	2.967	6.00
2026	6	5	0.35	5	1.152	0.30
2026	7	29	1.28	29	8.866	0.14
2026	8	18	0.01	18	5.985	0.00

Table 14. Fork length composition (%) of chub mackerel (*Scomber japonicus*) caught by R/V Songhang in the Northwest Pacific Ocean, 2021–2026

Fork length class (mm)	2021	2022	2023	2024	2025	2026
0–39	—	—	—	—	—	1.08
40–79	—	—	0.31	0.26	28.83	31.18
80–119	14.17	—	8.15	2.07	1.80	36.56
120–159	64.53	22.10	27.43	23.51	23.60	29.03
160–199	8.89	45.54	27.43	46.51	30.00	2.15
200–239	8.89	10.71	26.65	10.85	15.77	—
240–279	3.03	19.42	7.37	8.01	—	—
280–319	0.35	2.23	2.51	8.53	—	—
320–359	0.07	—	0.16	0.26	—	—
360–399	0.07	—	—	—	—	—
Total	100.00	100.00	100.00	100.00	100.00	100.00
Number	1418	448	638	387	1110	93

Figures

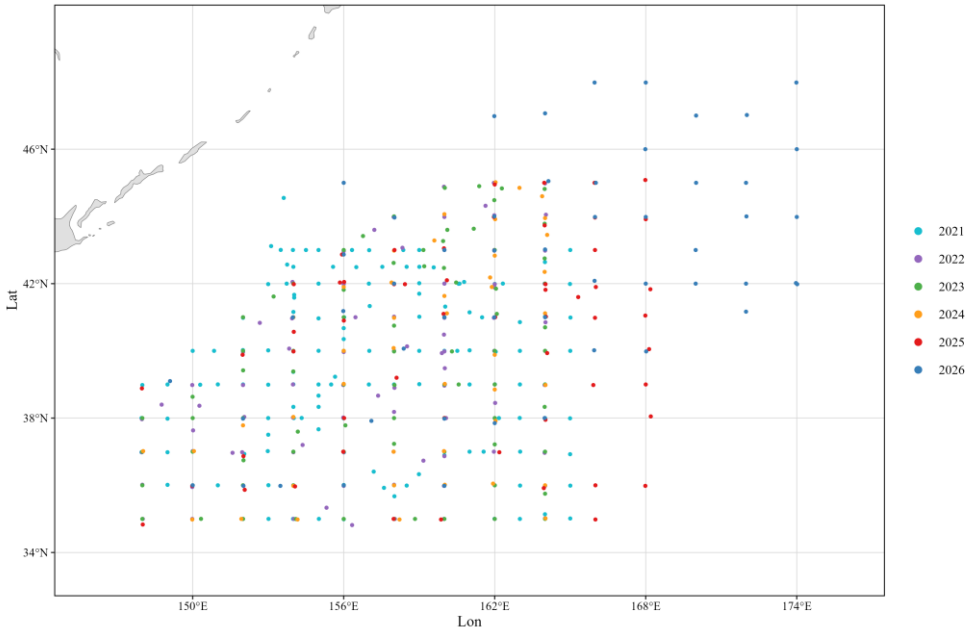

Fig. 1. Spatial distribution of trawl survey stations in the Northwest Pacific Ocean, 2021–2026.

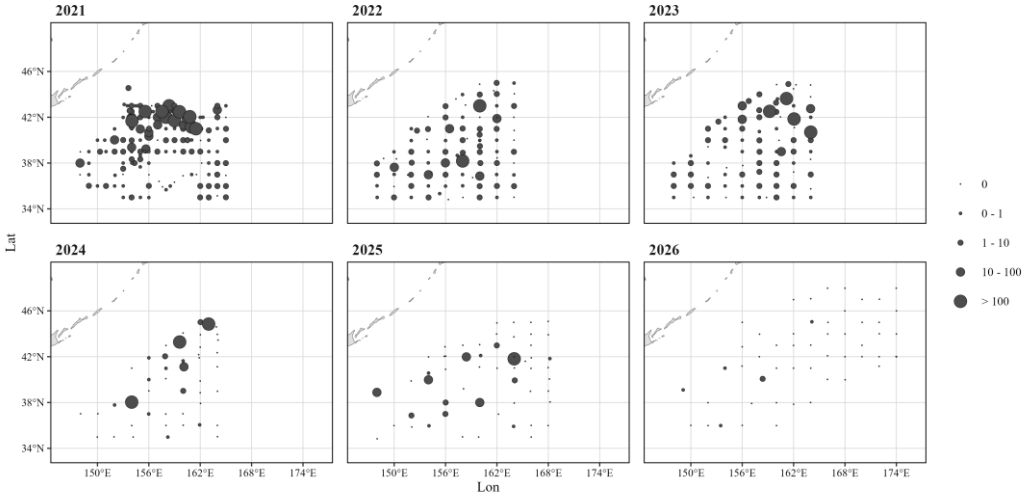

Fig. 2. Spatial distribution of the nominal CPUE (kg/km²) of chub mackerel by year, 2021–2026.

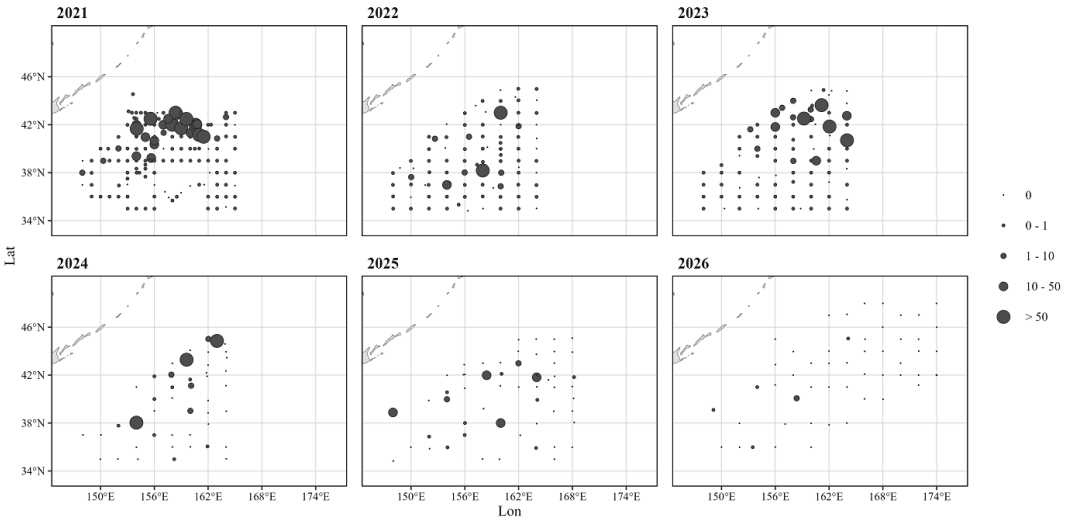

Fig. 3. Spatial distribution of the chub mackerel catch (kg) by year, 2021–2026.

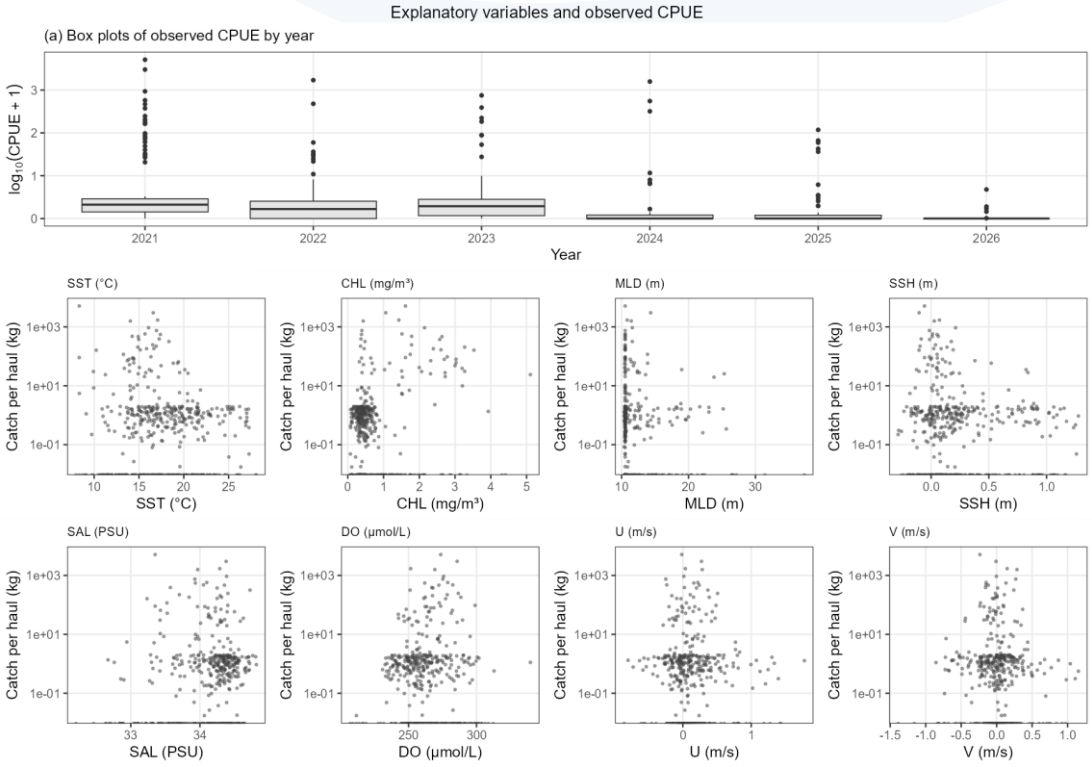

Fig. 4. (a) Annual distribution of observed chub mackerel CPUE; (b) catch per haul against each candidate explanatory variable.

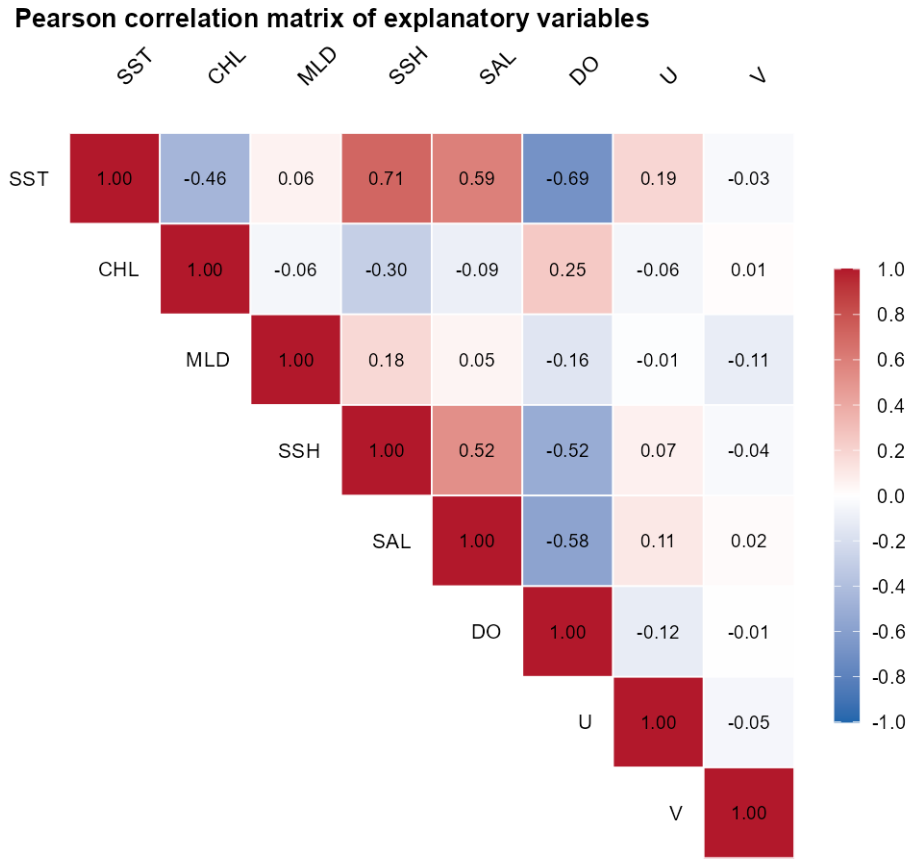

Fig. 5. Pearson correlation matrix of the eight environmental covariates considered for CPUE standardization.

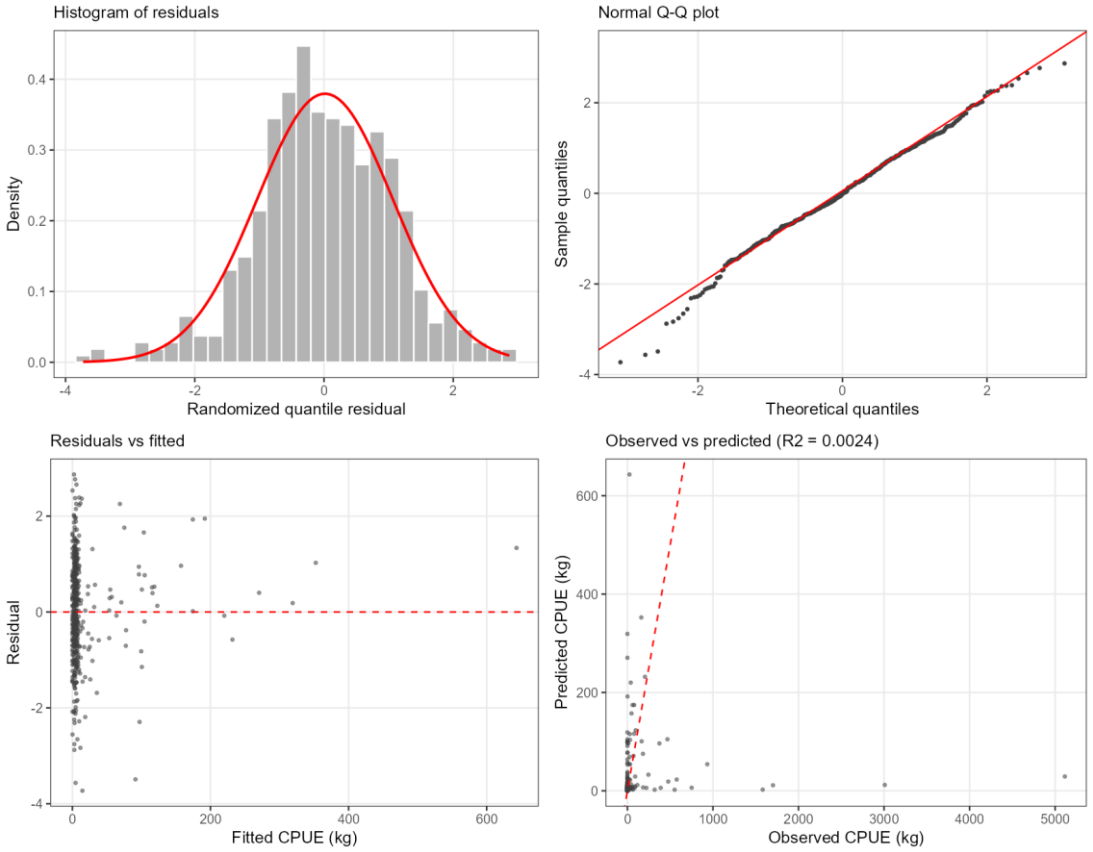

Fig. 6a. Diagnostic plots for the selected delta-lognormal model (fit39).

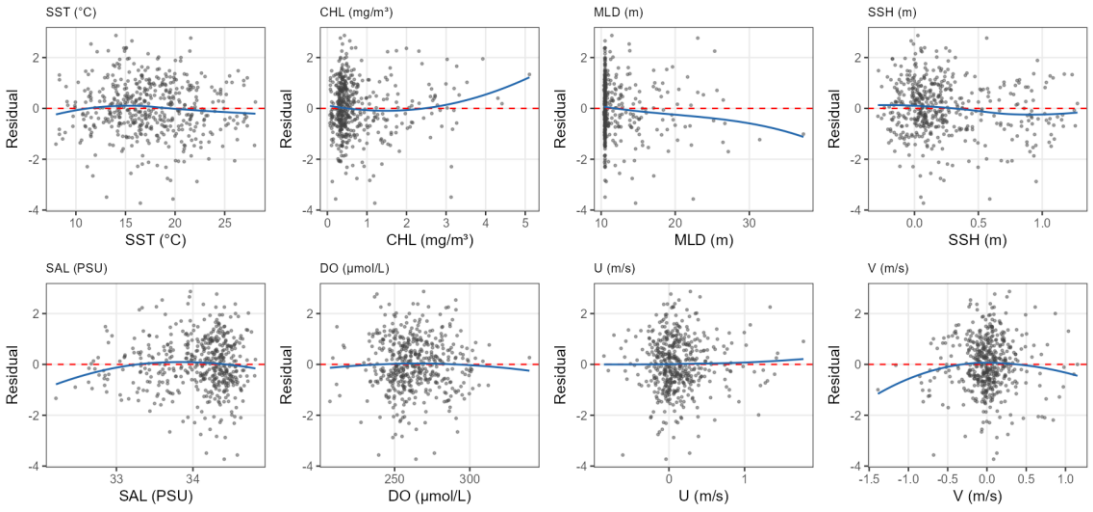

Fig. 6b. Residuals of the selected model against each candidate explanatory variable.

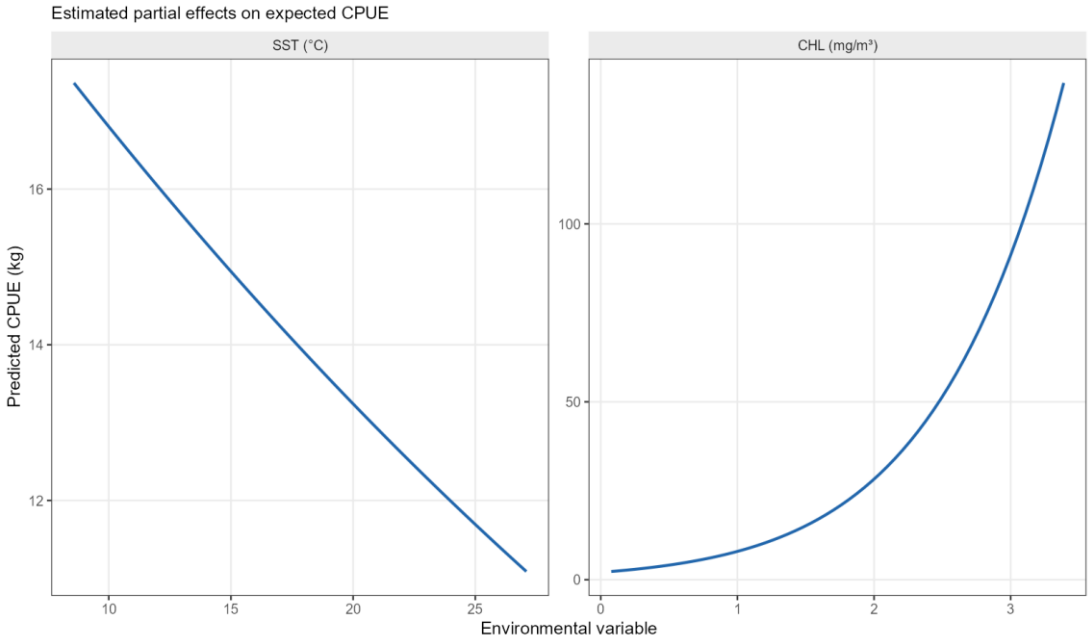

Fig. 7. Estimated partial effects of the covariates retained in the selected model (SST and CHL) on the expected CPUE (kg), averaged over the observed hauls.

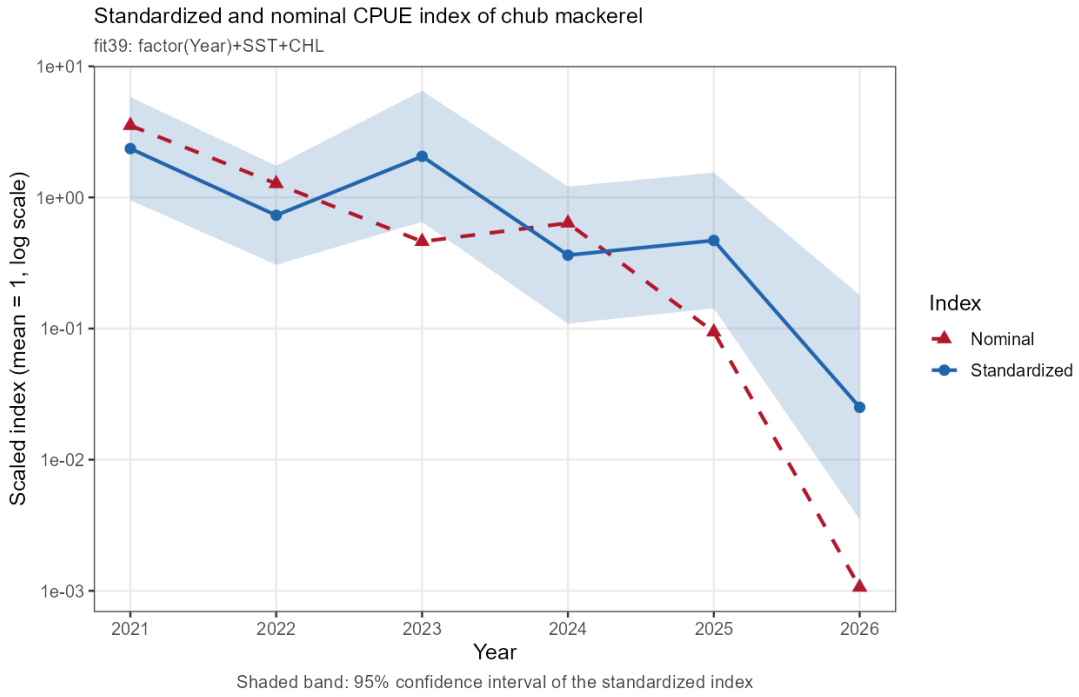

Fig. 8a. Scaled standardized and nominal CPUE indices of chub mackerel, 2021–2026

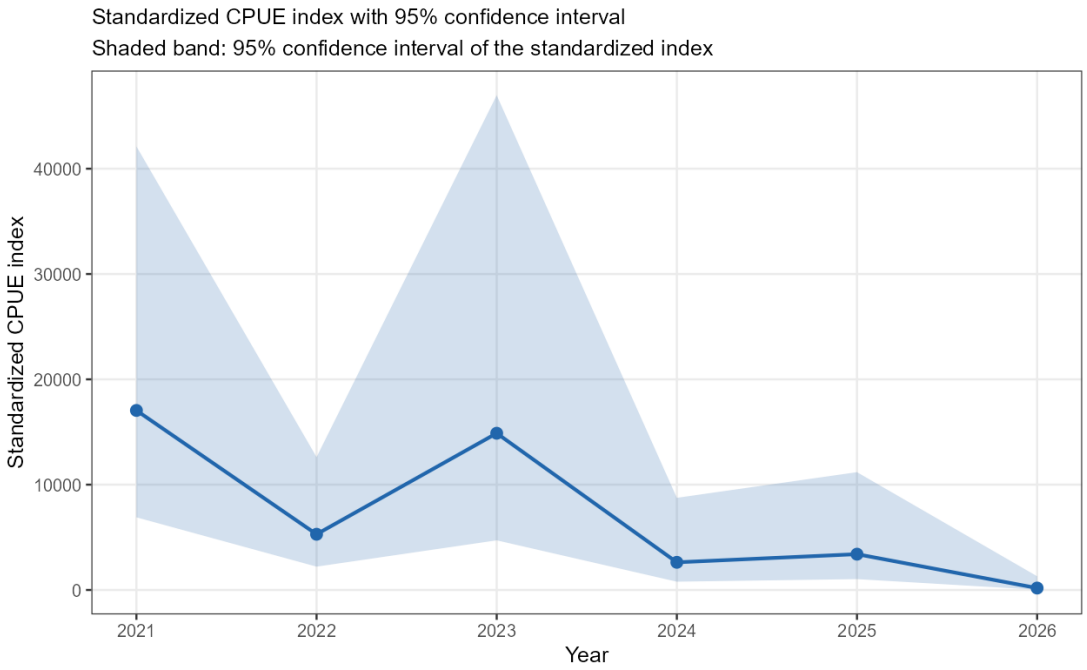

Fig. 8b. Standardized CPUE index of chub mackerel with its 95% confidence interval, 2021–2026.

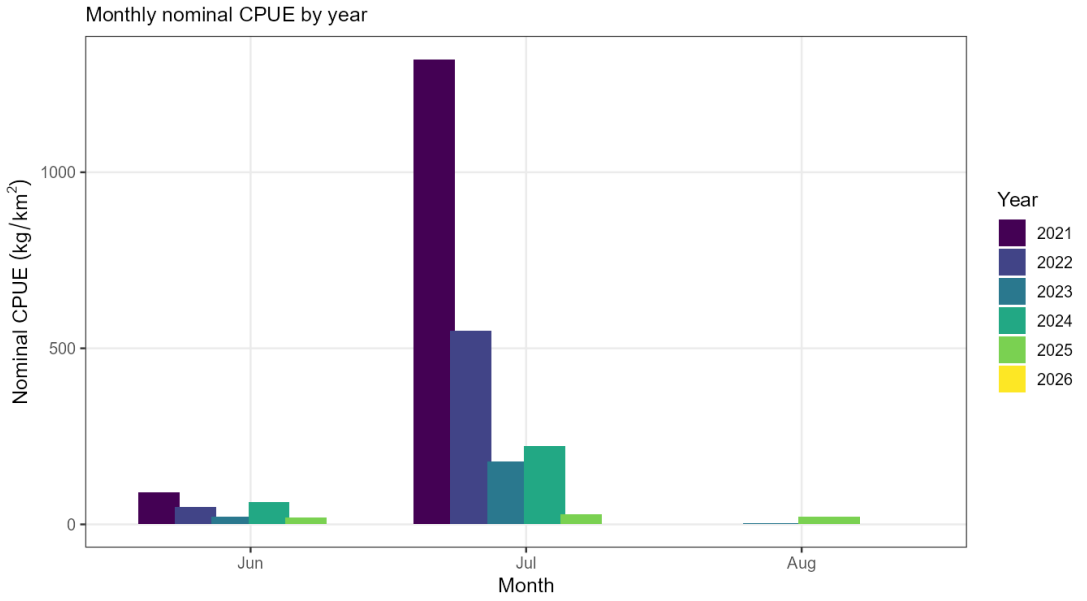

Fig. 9. Monthly nominal CPUE of the R/V Songhang trawl survey by year, 2021–2026.

APPENDICES

Appendix I. Checklist for the CPUE standardization protocol

No.	Step-by-step protocols	Yes/No	Note
1	Provide a description of the type of data (logbook, observer, survey, etc.) and the "resolution" of the data, including the representativeness of the data in two tables.	Yes	See Section 2.1 and Tables 1 and 2.
2	Conduct a thorough literature review to identify potential explanatory variables that may influence CPUE values.	Yes	See Section 1.
3	Plot annual/monthly spatial catch, effort and nominal CPUE distributions and determine temporal and spatial resolution for CPUE standardization.	Yes	See Figs 1–3, Fig. 9, and Table 13.
4	Make scatter plots (for continuous variables) and/or box plots (for categorical variables) and present a correlation matrix to evaluate correlations between each pair of those variables.	Yes	See Fig. 4 (box plot by year and scatter plots of catch against every candidate covariate) and Fig. 5 (correlation matrix).
5	Describe selected explanatory variables based on (2)–(4) to develop the full model for CPUE standardization.	Yes	See Sections 2.1 and 2.3, and Table 3.
6	Specify model type and software (packages) and fit the data to the assumed statistical models.	Yes	See Section 2.3: sdmTMB 1.0.0 in R 4.5; GLMM with Tweedie, delta-gamma and delta-lognormal distributions, 54 candidate models.
7	Evaluate and select the best model(s) using methods such as likelihood ratio test, information criteria, cross validation, etc.	Yes	See Section 2.3 (pre-specified selection rule), Section 3.1 and Tables 4, 5 and 10. A likelihood ratio test is not applicable because the models are not all nested and the random-effect variances lie on the boundary of the parameter space.
8	Provide diagnostic plots to support the chosen model is appropriate and the assumptions are met.	Yes	See Fig. 6a–b (histogram and Q–Q plot of randomized quantile residuals, residuals versus fitted, observed versus predicted, and residuals versus each covariate) and Table 12.
9	Present estimated values of parameters and uncertainty in the parameters in a table.	Yes	See Table 6 (both components of the delta model and the random-effect parameters, with standard errors and 95% confidence intervals).

10	Present the relationship between the dependent and independent variables and check whether it is interpretable.	Yes	See Fig. 7 and Section 3.2. The sign reversal of the CHL coefficient between the two components is discussed there.
11	Extract yearly standardized CPUE and standard error by a method that accounts for spatial heterogeneity of effort; provide details on how the index was extracted.	Yes	See Section 2.4. The index was extracted with predict() and get_index() on the same footprint of 472 station positions in every year, with a constant area weight of 1; area weighting is not required because the footprint is constant across years, although the earlier years partly extrapolate onto stations that were surveyed only in 2026.
12	Calculate uncertainty (SD, CV, CI) for the standardized CPUE for each year and explain how it was calculated.	Yes	See Section 2.4 and Table 7: the log-scale standard error from get_index(), CV = $\sqrt{\exp(SE^2) - 1}$, and 95% limits $\exp(\log(\text{index}) \pm 1.96\ SE)$.
13	Provide a table and a plot of nominal and standardized CPUEs over time, and explain the reasons when the trends differ largely.	Yes	See Table 7, Fig. 8a–b and Section 3.3; the divergence between the two indices in 2024 is discussed.