



**SCIENTIFIC COMMITTEE
TWENTY-SECOND REGULAR SESSION**

11-19 August 2026
Apia, Samoa (Hybrid)

**TRENDS IN BIGEYE TUNA MEAN WEIGHT BY SPACE, TIME, AND FLEET IN THE
WCPO LONGLINE FISHERY**

WCPFC-SC22-2026-SA-IP13

11 July 2026

Submitted by C. Castillo-Jordán¹, J. Potts¹, T. Peatman², P. Hamer¹

¹ Oceanic Fisheries Programme, Pacific Community, Noumea, New Caledonia

² Shearwater Analytics Ltd., Frome, United Kingdom

Table of contents

1	Introduction	6
2	Methods	7
2.1	Data	7
2.2	Region-wide multi-fleet model	8
2.2.1	Exploratory data analysis	8
2.2.2	Model specification and selection	11
2.3	Single-fleet historical analyses	12
3	Results	14
3.1	Region-wide multi-fleet results	14
3.2	JPDW historical analysis (1957–2018)	16
3.3	USHW historical analysis (2000–2019)	23
3.4	Fleet comparison: JPDW versus USHW	25
4	Discussion	27
4.1	Comparison with Langley (2006)	27
4.2	Catching smaller fish, or fish getting smaller?	27
4.2.1	Population change vs sampling artefact	27
4.2.2	Fleet comparison: what it can and cannot show	28
4.2.3	Follow-up analyses to refine the decomposition	29
4.3	Considerations for the stock assessment	29
4.4	Methodological progress and limitations	30
4.4.1	A GAM rather than spatiotemporal models	31
4.4.2	Limitations	31
5	Conclusion	32
	Data and code availability	33
	Acknowledgements	34
	Declaration of generative AI and AI-assisted technologies in the SC paper preparation process	35
	References	36
A	Appendix: Glossary of terms, notation, and codes	38
A.1	Abbreviations and acronyms	38
A.2	Statistical notation	39
A.3	Flag-fleet codes	40
A.4	Key R functions and packages	41
A.5	Region-wide analysis: exploration of all data	42
A.6	Region-wide analysis: model diagnostics	44
A.7	Region-wide analysis: temporal autocorrelation (AR(1) robustness check)	49
A.8	Region-wide analysis: testing whether the recent pooled decline reflects the changing flag mix	50
A.9	Model diagnostics for the single-fleet historical GAMs	51
A.9.1	Summary statistics	51
A.9.2	Basis dimension checks (k.check)	52
A.9.3	Residual autocorrelation	53
A.9.4	Model appraise diagnostics	54
A.10	JPDW sensitivity analyses	56
A.11	JPDW spatial coverage relative to the study area	56

List of Figures

1	The five WCPFC stock-assessment regions used for the 2026 assessment used to filter the BET longline composition data. Note: longitudes follow the Pacific-centred 0–360°E convention throughout this paper.	8
2	Map of total sample size (top panel) and mean weight (bottom panel), averaged over all flags and years, for the region-wide analysis.	9
3	Map of sample size, summed across all years, for each flag retained in the region-wide analysis. Black outlines show the five assessment regions.	10
4	Number of weight samples per year, for each flag retained in the region-wide analysis.	11
5	Overall temporal trend in mean BET weight across the analysis period (top panel) and the seasonal effect (bottom panel), from <code>m_temp_flag</code> . Shading is the 95% CI.	14
6	Flag-specific temporal deviations from the global trend (log scale). Dashed line = zero (global trend); rug = observed quarters.	15
7	Flag-specific temporal trends in mean BET weight (kg), derived from the global temporal smooth plus each flag’s deviation smooth from <code>m_temp_flag</code> . Dashed black line = global trend; coloured lines = flag-specific trends; rug = observed quarter-years.	16
8	JPDW BET predicted mean weight (reported from 1960; see text), with 95% confidence band. Single-fleet GAM fitted to the full 1957–2018 record.	17
9	JPDW BET predicted spatial pattern of mean weight (kg), quarter 2. Each panel is the per-cell average of model predictions across the year window labelled (e.g. the 1970–1979 panel averages predictions for each year from 1970 to 1979). Each panel is masked to the 5° cells JPDW sampled within that window. The middle four panels each span exactly 10 years; the first (1957–1969, 13 years) and last (2010–2018, 9 years) panels extend to the start and end of JPDW BET sampling respectively, so the six panels together cover every year of the analysis. The progressive contraction of the coloured area shows the contraction of the JPDW fleet’s operational footprint over time.	19
10	JPDW BET quarterly mean weight binned by 5-year periods (1955–1959, 1960–1964, ...). Each box summarises the distribution of quarterly sample-weighted mean weights within the period; the connecting line traces the period medians.	20
11	JPDW BET quarterly mean weight binned by 10-year periods.	20
12	JPDW BET predicted spatial pattern of mean weight (kg) restricted to the 2010–2018 sampling footprint, replotted across the same six time windows as the by-decade spatial maps. Because every panel uses the same spatial mask (the cells the fleet sampled during 2010–2018), differences in colour between panels reflect temporal changes in predicted mean weight in the fixed area, not changes in the fleet’s operational footprint. This isolates the within-area temporal change from the spatial-contraction signal shown in the by-decade maps. The colour scale is shared with the by-decade maps to allow direct comparison.	21
13	Boxplots of JPDW BET predicted mean weight (kg) across all cells in the 2010–2018 sampling footprint, for each of the six time windows shown in the fixed-area spatial maps. Each box summarises the per-cell distribution of model predictions within the fixed spatial mask; the line connects period medians to trace the within-area temporal change. Because the spatial mask is held fixed across periods, any vertical shift between boxes is a within-area change in predicted mean weight, not an artefact of the fleet’s contracting operational footprint. The y-axis is cropped to the plausible range: a small number of cells in the sparsely-sampled early windows carry extreme extrapolated predictions from the per-window spatial smooth and fall off-scale; they do not affect the period medians shown.	22

14	Decomposition of the JPDW BET mean weight decline. Steelblue line: median predicted mean weight across cells in each period's own sampling footprint (unrestricted — spatial mask changes with fleet contraction). Orange line: median across the fixed 2010-2018 footprint (within-area — mask held constant). The shaded area between the two lines is the fleet-redistribution component: additional apparent decline arising from the fleet abandoning cells where fish were historically largest. Both lines converge in the final period (2010-2018) where the masks are identical. Time-window lengths are unequal, so the connecting lines should be read as period-to-period medians rather than a constant-rate trajectory.	23
15	USHW BET predicted mean weight, 2000–2019, with 95% confidence band.	24
16	USHW BET quarterly mean weight binned by 5-year periods. Each box summarises the distribution of quarterly sample-weighted mean weights within the period; the connecting line traces the period medians.	25
17	Comparison of single-fleet GAM predicted BET mean weight for JPDW (reported 1960-2018) and USHW (2000-2019). Each line is the temporal trend from the fleet's independent historical GAM at the fleet-mean spatial location (quarter 2); shaded bands are 95% confidence intervals. The grey band marks the 2000-2018 period over which both fleets were sampled. JPDW's large decline is concentrated before 2000, outside USHW's coverage; within the overlap period the two fleets fluctuate around broadly similar mean weights with overlapping confidence bands and neither shows a strong directional trend.	26
18	Number of weight samples by year and flag across the full source period (all years present in the raw data). Note the y-axis range changes by flag.	42
19	Sample size summed across all years in the source data, by flag.	43
20	Model appraisal diagnostics for <code>m_temp_flag</code> (<code>gratia::appraise</code>).	45
21	Estimated spatial effect on mean BET weight (log scale), controlling for temporal trend, seasonality, and flag.	46
22	Fitted vs Pearson residuals by flag.	47
23	Fitted vs Pearson residuals by year, coloured by flag.	48
24	Global temporal trend with and without the AR(1) correction. Shading = 95% CI.	49
25	Residual autocorrelation functions (ACF) for the two single-fleet historical GAMs. Dashed lines indicate the ± 0.1 threshold for reference. Both models show some residual autocorrelation at short lags, with JPDW displaying the most persistent multi-scale pattern.	53
26	Diagnostic plots (<code>gratia::appraise</code>) for <code>jpdw_mod</code>	54
27	Diagnostic plots (<code>gratia::appraise</code>) for <code>ushw_mod</code>	55

List of Tables

1	Total weight samples (individual fish) and row percentage by flag in each region.	11
2	Model identifiers used in the text, tables, and code	13
3	AIC comparison of candidate region-wide models.	14
4	JPDW BET predicted mean weight at milestone years, with 95% confidence intervals. Reported trend begins in 1960 (see text).	18
5	USHW BET predicted mean weight at milestone years (2000–2019), with 95% confidence intervals.	24
6	Abbreviations used in the paper	38
7	Mathematical notation	39
8	Flag-fleet codes referenced in the paper. The region-wide analysis retains the subset of these flags meeting the inclusion criteria (see Methods); AU, NZ, TWDW, TWOS, and VN are excluded from the region-wide analysis on data-sparsity or data-quality grounds.	40
9	Key R functions and packages	41
10	Overall fit statistics for the region-wide model (<code>m_temp_flag</code>).	44
11	Approximate significance of the smooth terms in <code>m_temp_flag</code> (<code>mgcv</code> summary). <code>edf</code> = effective degrees of freedom.	44
12	Basis-dimension adequacy check (<code>mgcv::k.check</code>) for <code>m_temp_flag</code> . A low p-value suggests the basis dimension <code>k</code> may be too small.	44

13	Pairwise concurvity (worst-case estimate) between the smooth terms of <code>m_temp_flag</code> . Values near 1 indicate near-complete confounding.	44
14	Lag-1 residual autocorrelation within the three densest JPDW cells (region-wide model). Cells are identified by the south-west corner of each 5° bin; <code>n</code> is the number of cell-quarter observations.	49
15	Summary statistics for the two single-fleet historical GAMs.	51
16	Basis dimension adequacy checks (<code>k.check</code>) for each model. A low <code>p</code> -value suggests the basis dimension <code>k</code> may be too small.	52
17	Sensitivity of the JPDW headline marginal decline (predicted mean weight at 1965 and 2015, and their difference) and of the observed within-area decline over the fixed 2010–2018 footprint, to the weight cap, error family, and temporal basis dimension. One setting is varied at a time from the base specification.	56
18	JPDW BET weight sampling by decade window, split by whether the 5-degree cell falls inside the five assessment regions (the study area used for the spatial maps) or outside them. Percentages are of the window’s total sampled fish. The share sampled outside the assessment regions is largest in the early windows and declines toward the 2010s, as the fleet’s footprint contracts into the WCPO.	56
19	JPDW BET weight records in the source database before 1957 — the years that fall outside the 1957–2018 single-fleet window and are therefore excluded from the JPDW analysis. Sampled fish per year (the totals are small: under 0.1% of all JPDW BET weight records combined).	57

Executive Summary

Variation in catch size composition is a key input to integrated stock assessment models, including the MULTIFAN-CL assessment of bigeye tuna (*Thunnus obesus*, BET) in the Western and Central Pacific Ocean (WCPO). This study re-examines size composition across the WCPO longline fishery using generalised additive models (GAMs) fitted via `mgcv`, extending Langley (2006).

A region-wide multi-fleet model fitted to 1990–2020 longline data (log mean weight, Gaussian family, sample-size weights; global temporal, seasonal, and spatial smooths, a flag random intercept, and flag-specific temporal deviations, structures compared by AIC) shows mean weight declining on average from approximately 47 kg in the early 1990s to around 31 kg by 2020, with substantial flag-to-flag heterogeneity around that global trend (e.g. CN persistently below trend; SB peaking around 2010–2015 then falling sharply).

Single-fleet analyses were run for Japan distant-water longline (JPDW, 1957–2018) and US Hawaii longline (USHW, 2000–2019). JPDW mean weight declines from approximately 46 kg in the mid-1960s to 35 kg by 2015, with the decline concentrated before 2000 and continuing at a reduced but still discernible rate thereafter; USHW shows no trend in mean weight over 2000–2019. A fixed-area decomposition holding the JPDW final-period footprint fixed indicates that within-area change (about 9 kg) accounts for the majority of the JPDW decline, with fleet redistribution responsible for a smaller secondary component.

Over their 2000–2018 overlap both fleets are similar and trendless, so neither shows evidence of a strong recent decline in mean weight within its sampled region. As USHW does not extend into the JPDW period of decline, the comparison cannot establish whether that decline was basin-wide in the north west Pacific; size signals are best treated as fleet-, region-, and time period-specific. A key outcome is that the declining JPDW weight trend was consistent even for the restricted area analysis. The trend may reflect a genuine decline in the underlying mean weight of the exploited BET population, and does not appear to be an artefact of fleet contraction. However, a change in JPDW selectivity over time cannot be ruled out. The observed trend may therefore reflect some combination of a real population-level decline and a selectivity shift, rather than either process alone.

Considerations for the BET stock assessment

- Treat the within-area decline as a candidate signal about BET size structure to be considered in the 2026 assessment; resolve its mechanism (truncation, growth-rate change, or recruitment) through the follow-up analyses identified in this paper.
- Recognise the fleet-redistribution component as a secondary bias under temporally-constant selectivity; its influence can be bounded with an in-assessment sensitivity run (time-block selectivity, or down-weighting the contracted-period composition).
- Fleets with stable spatial footprints (e.g. USHW) may provide useful complementary indicators of temporal size-structure change and should be evaluated alongside fleets with substantial footprint contraction, whose within-area size change is confounded with spatial redistribution.
- Standardising the composition data to remove the spatial/footprint effect at the data stage offers a more direct route to a population-level signal than relying on selectivity adjustments within the assessment alone.

1 Introduction

The Scientific Committee at its twenty-first regular session (SC21) recommended that the Stock Assessment and Modelling team (SAM) of the SPC Oceanic Fisheries Programme undertake a detailed analysis of bigeye tuna (BET) mean weight trends by space, time, and fleet over the past 10 to 20 years. The analysis is intended to improve understanding of these patterns and to inform the 2026 BET stock assessment. The present work responds to that recommendation.

The size composition of the catch is a key input to integrated stock assessment models, including the MULTIFAN-CL [MFCL; Fournier *et al.* (1998); Hampton *et al.* (2002)] assessment of bigeye tuna (*Thunnus obesus*, BET) in the Western and Central Pacific Ocean (WCPO). Mean weight — examined here by year, quarter, and spatial cell — is an integrated summary of the weight-composition signal that the assessment fits. Spatiotemporal heterogeneity in fish size composition has two implications for the assessment: it informs the regional and fishery structure assumed by the model, and it interacts with fleet-specific selectivity functions, which are assumed to be temporally constant in the assessment (Waterhouse *et al.*, 2014).

The foundational analysis of spatiotemporal variation in BET and yellowfin tuna size composition in the WCPO was conducted by Langley (2006) using descriptive median-by-block methods on Japanese distant-water longline data from 1948 to 2004. Langley’s analysis identified persistent spatial heterogeneity in size composition that informed the regional structure of the current MFCL assessment, and raised the methodological concern that spatial shifts in fishing effort can substantially alter the size composition of the catch even when the underlying population size structure is unchanged. Two decades later, around fourteen years of additional data and the opportunity to move beyond descriptive median-by-block methods to continuous, uncertainty-quantified smoothing motivate a re-examination.

The present analysis has three objectives:

1. Re-examine spatiotemporal variation in BET longline mean weight using flexible statistical methods (generalised additive models, joint estimation of fleet, temporal, and spatial effects, with confidence intervals on the fitted trends).
2. Extend the temporal coverage of Langley (2006)’s analysis by approximately 14 years for the Japanese distant-water longline fleet (JPDW), the longest-running BET weight sampling programmes available for the WCPO.
3. Provide complementary single-fleet analyses of the Japanese distant-water (JPDW) and US Hawaii (USHW) longline fleets, treated separately, and compare the two trajectories to assess whether size-structure signals are consistent across fleets — bearing directly on the constant fleet-selectivity assumption in the assessment.

The analysis is restricted to longline-caught BET. Purse-seine composition data are excluded because longline and purse seine target different size-classes of the BET population through fundamentally distinct selectivity processes; combining them in a single mean-weight model would conflate gear selectivity with biological size variation. Within the BET stock assessment, gear-specific composition data are integrated through gear-specific selectivity functions (Day *et al.*, 2023).

2 Methods

2.1 Data

Weight samples for BET were extracted from the WCPFC operational longline composition database held by the Pacific Community (SPC). The underlying samples derive from a range of port-sampling and auction programmes contributing to SPC’s data holdings; full details of these data sources and their relationship to the flag-fleet attribution are provided in a companion working and information paper (Hamer *et al.*, 2026; Peatman, 2026). We note that sampling sources for size data for individual flags vary across time, as we do not have specific details of these changes, we assume the sampling is consistent in its representation of catch size compositions.

The five assessment regions are defined as longitude–latitude boxes (shown in Figure 1; longitudes in the Pacific-centred $[0^\circ, 360^\circ]$ format), following the regional structure of the 2026 WCPO BET assessment (Kim *et al.*, 2026):

- **Region 1** (North): $\text{lon} \in [120^\circ, 210^\circ)$, $\text{lat} \in [10^\circ, 50^\circ)$
- **Region 2** (Northwest): $\text{lon} \in [110^\circ, 140^\circ)$, $\text{lat} \in [-10^\circ, 20^\circ)$
- **Region 3** (Central West): $\text{lon} \in [140^\circ, 185^\circ)$, $\text{lat} \in [-10^\circ, 10^\circ)$
- **Region 4** (Central East): $\text{lon} \in [185^\circ, 210^\circ)$, $\text{lat} \in [-10^\circ, 10^\circ)$
- **Region 5** (South): $\text{lon} \in [140^\circ, 210^\circ)$, $\text{lat} \in [-40^\circ, -10^\circ)$

Composition records are binned to 5° cells by their lower-left corner (longitude and latitude floored to the nearest 5°), so the lower bound is included and the upper excluded — a cell at 205°E spans $[205^\circ, 210^\circ)$, with the 210°E meridian forming the eastern edge. In code these bounds are applied with $\geq$ on the lower edge and $<$ on the upper.

Spatial coordinates were aggregated to $5^\circ \times 5^\circ$ grid cells. For each combination of fleet f , year-quarter t , and spatial cell, the response is the sample-weighted mean weight:

$$\bar{w}_{f,t,\text{cell}} = \frac{\sum_i w_i \cdot \text{freq}_i}{\sum_i \text{freq}_i}, \quad (1)$$

with sample size $n_{f,t,\text{cell}} = \sum_i \text{freq}_i$ used as the modelling weight.

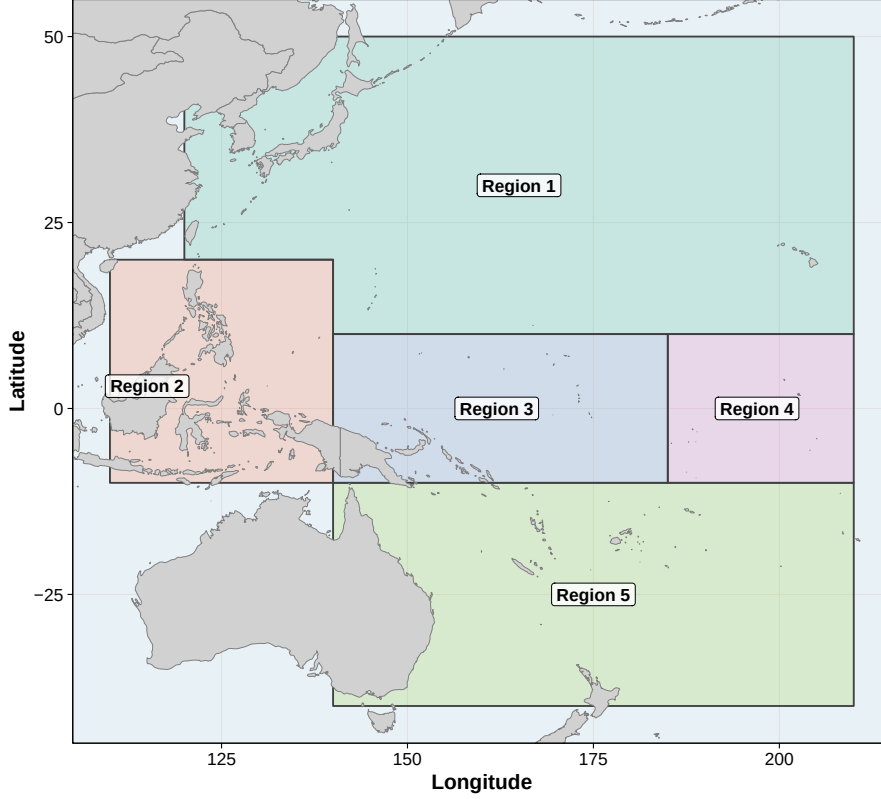


Figure 1: The five WCPFC stock-assessment regions used for the 2026 assessment used to filter the BET longline composition data. Note: longitudes follow the Pacific-centred 0–360°E convention throughout this paper.

The region-wide multi-fleet analysis applies the following filters:

- Weight samples restricted to the five assessment regions defined above.
- Records with missing coordinates excluded; coordinates binned to $5^\circ \times 5^\circ$ cells.
- Individual weights ≤ 200 kg (larger fish removed as outliers).
- Restricted to 1990–2020 to ensure model convergence over a period of stable multi-fleet coverage.
- Flags with fewer than 15,000 individual fish weight samples removed as too data-sparse for reliable modelling; several further flags (AU, TWDW, TWOS, VN, NZ) removed on data-quality concerns and/or data-sparsity and restricted spatial coverage grounds (Section A.5). The flag codes used throughout (e.g. JPDW, USHW, CN, SB) denote national longline fleets and are defined in full in Table 8.

The full source dataset ranged from 1952 to 2024 and included weight and frequency data from over 13 million fish across 29 flags. The spatial and temporal coverage of several flags was too sparse for inclusion (Section A.5, Figure 18, Figure 19). After filtering, the region-wide analysis covers 1990–2020 and contains 8,455 cell-quarter observations from 8 flags: CN, FM, JPDW, MH, PW, SB, TO, USHW.

2.2 Region-wide multi-fleet model

2.2.1 Exploratory data analysis

A map of all sampling effort, across all time periods and flags, is given in Figure 2. For flag-specific effort, see Figure 3. The temporal coverage of sampling varies by flag (Figure 4).

A key structural feature of the data is that most flags fish almost exclusively within a single model region, making flag and region effects near-inseparable in any model (Table 1). With the exceptions of JPDW (which

spans Regions 1 and 3) and CN (which spans Regions 2 and 3), each flag contributes more than 90% of its weight samples from a single region. Flag-level and region-level temporal effects therefore cannot be independently estimated, and a model should include one or the other, not both.

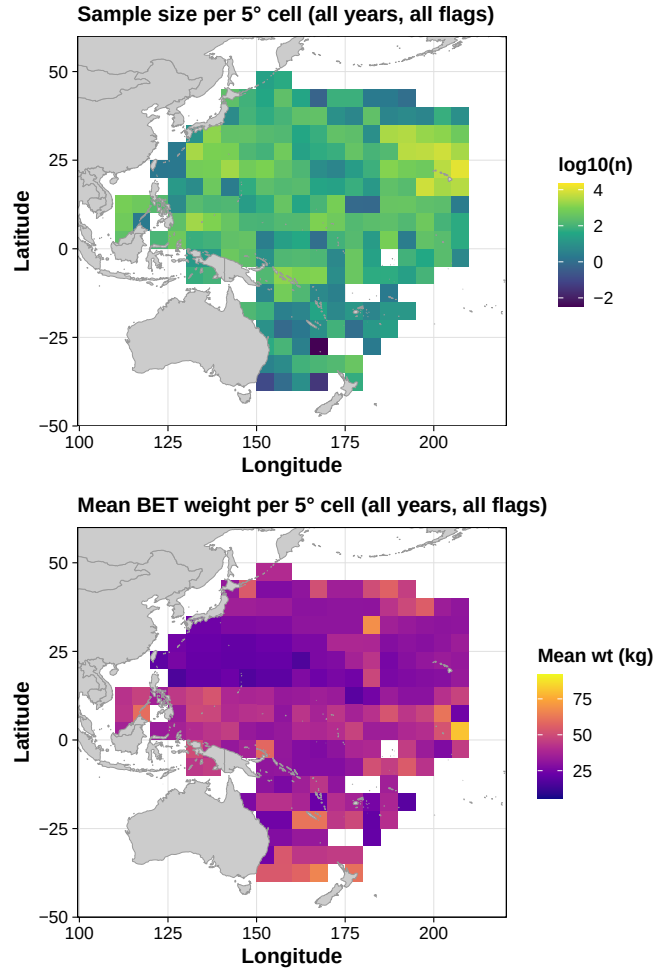


Figure 2: Map of total sample size (top panel) and mean weight (bottom panel), averaged over all flags and years, for the region-wide analysis.

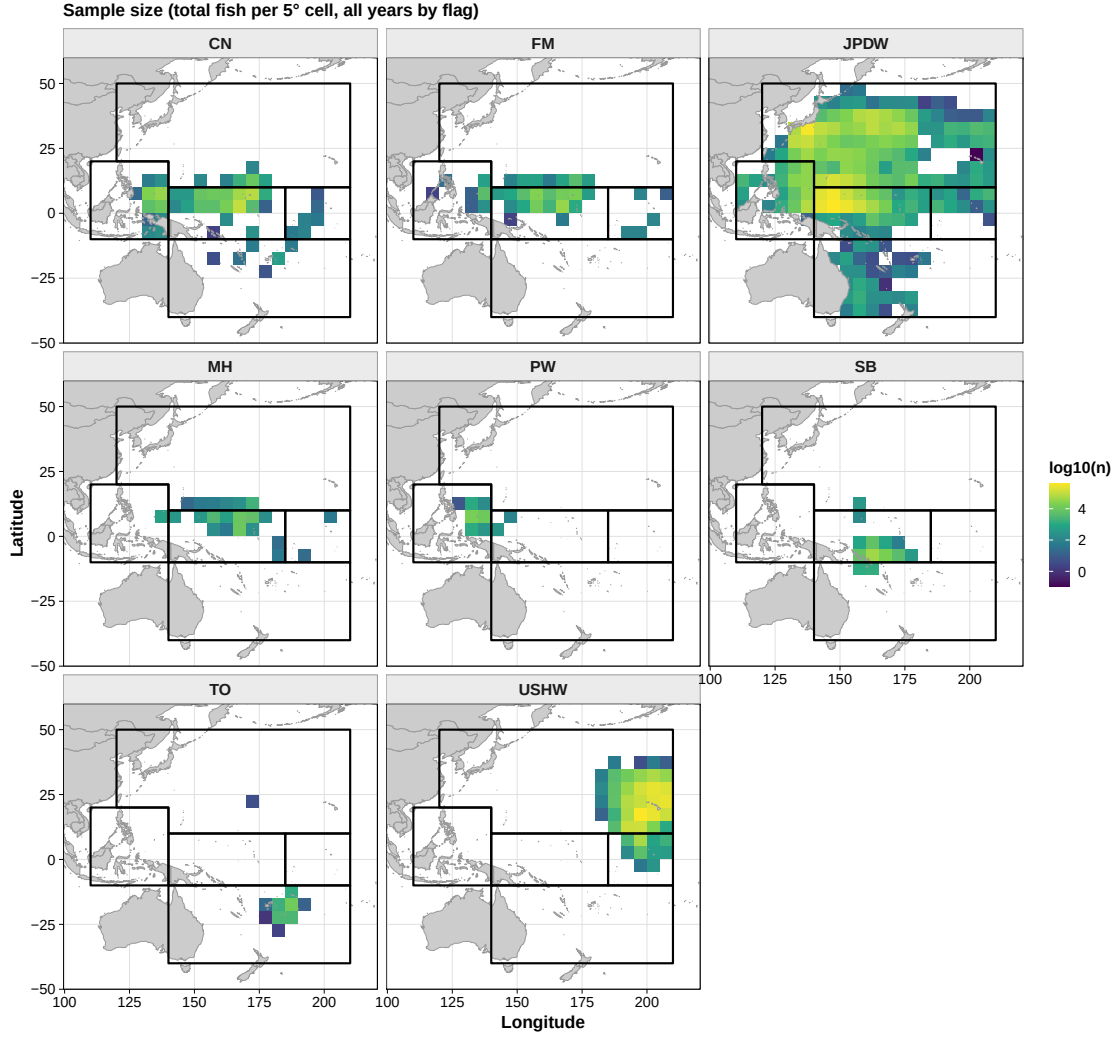


Figure 3: Map of sample size, summed across all years, for each flag retained in the region-wide analysis. Black outlines show the five assessment regions.

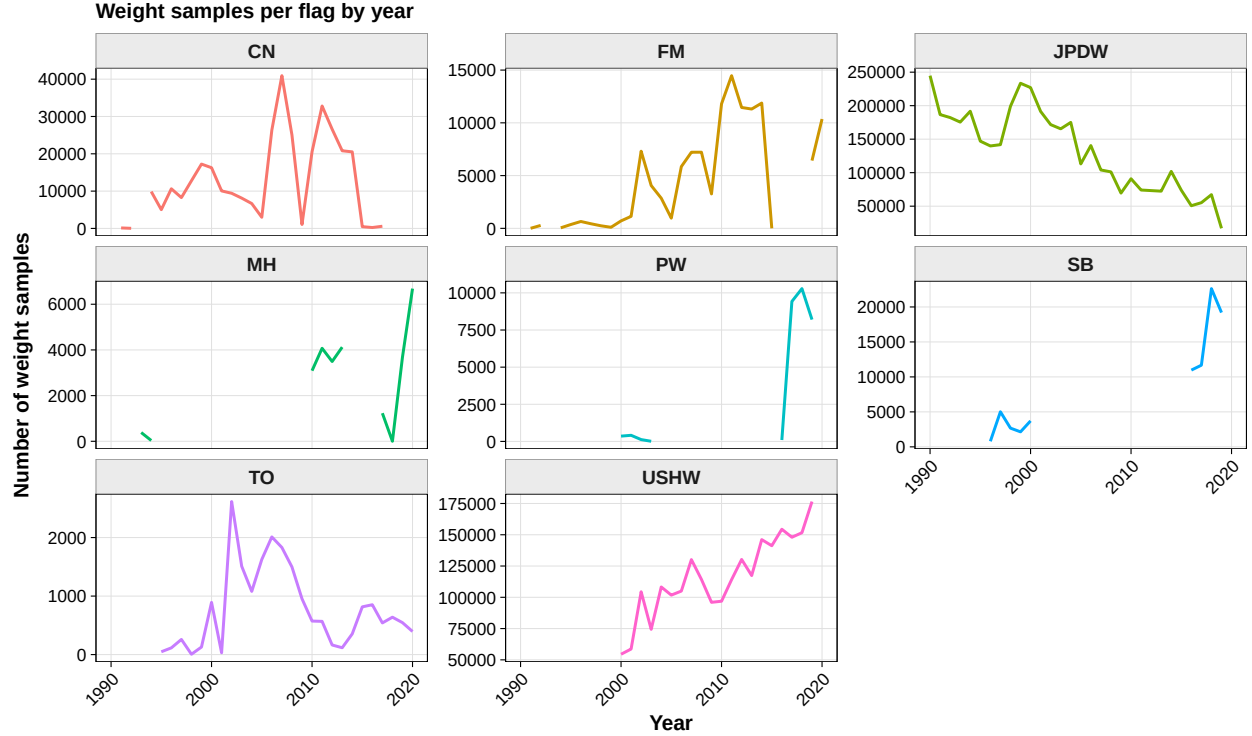


Figure 4: Number of weight samples per year, for each flag retained in the region-wide analysis.

Table 1: Total weight samples (individual fish) and row percentage by flag in each region.

Flag	Region 1	Region 2	Region 3	Region 4	Region 5
CN	10,869 (3.3%)	66,492 (19.9%)	256,049 (76.6%)	294 (0.1%)	615 (0.2%)
FM	6,224 (5.2%)	5,138 (4.3%)	108,902 (90.2%)	420 (0.3%)	0 (0%)
JPDW	2,032,998 (51.1%)	204,907 (5.2%)	1,724,505 (43.4%)	7,540 (0.2%)	7,583 (0.2%)
MH	2,262 (7.2%)	545 (1.7%)	28,334 (90.8%)	70 (0.2%)	0 (0%)
PW	0 (0%)	28,505 (98.6%)	404 (1.4%)	0 (0%)	0 (0%)
SB	348 (0.4%)	0 (0%)	76,043 (96.5%)	0 (0%)	2,407 (3.1%)
TO	4 (0%)	0 (0%)	0 (0%)	0 (0%)	20,165 (100%)
USHW	2,263,561 (97.4%)	0 (0%)	0 (0%)	60,342 (2.6%)	0 (0%)

2.2.2 Model specification and selection

The response variable is `log(mean_wt)` — the log of the weighted mean weight per cell-quarter — fitted with `family = gaussian()`. This is justified on two grounds. First, each cell-quarter response is a mean over many individual fish; although individual weights — like most organism size and weight data — are approximately log-normal (Limpert *et al.*, 2001), by the Central Limit Theorem (CLT) their mean is approximately normal, so the cell-quarter means are well described by a Gaussian. Second, residual diagnostics on a model fitted to untransformed `mean_wt` showed clear heteroscedasticity (variance increasing with the mean), which the log transform resolves (Box and Cox, 1964). The `total_n` column (sum of `freq`) is used as a precision weight in all models, since cells with more fish have more reliable mean estimates.

Three models are compared, all sharing a base structure of a global temporal smooth, a seasonal smooth, a spatial smooth, and a flag random intercept. The most general of the three adds flag-specific temporal deviations:

$$\mathbb{E}[\log \bar{w}_{f,t,\text{cell}}] = \beta_0 + s(t) + s(\text{qtr}) + s(\text{lon}, \text{lat}) + b_f + s(t, f), \quad (2)$$

where $s(t)$ is the global temporal smooth (year-quarter), $s(\text{qtr})$ a cyclic seasonal smooth, $s(\text{lon}, \text{lat})$ a global thin-plate spatial smooth, b_f a flag random intercept, and $s(t, f)$ the flag-specific temporal deviations (a factor-smooth, `bs = "fs"`) that distinguish `m_temp_flag` from the base model (`m_base`); a third candidate (`m_spat_flag`) instead allows the spatial smooth to vary by flag. All were fitted with `bam()` in the `mgcv` package (Wood, 2017), which implements large-data GAMs; estimation used `fREML` with `discrete = TRUE` for computational efficiency. Because all three candidates share the same response, family, weights, and fixed-effect structure, AIC is directly comparable and was used for model selection; the AIC reported by `mgcv` for penalised GAMs incorporates a correction for smoothing-parameter uncertainty (Wood *et al.*, 2016) rather than the nominal parameter count. The models are fitted without an autocorrelation term; an `AR(1)` refit examines the effect of residual temporal autocorrelation on the results (Section A.7). For this purpose each flag $\times$ 5° cell combination is treated as an independent series ordered by year-quarter, so lag-1 autocorrelation refers to residuals in successive quarters within the same cell and flag; distinct cells and flags are treated as independent.

2.3 Single-fleet historical analyses

Two single-fleet extensions complement the multi-fleet analysis, applying the same modelling approach to one fleet at a time. Because only a single flag is present, the two flag terms of the region-wide model (Equation 2) — the flag random intercept and the flag-specific temporal-deviation smooth — are dropped, leaving a base model with a global temporal smooth, a cyclic seasonal term, and a spatial smooth, fitted to the same sample-size-weighted 5° cell-quarter mean weights as the region-wide model (Section 2.1; here filtered to the single fleet):

$$\mathbb{E}[\log \bar{w}_{c,t}] = \beta_0 + s(t) + s(\text{qtr}) + s(\text{lon}, \text{lat}), \quad (3)$$

with t the year-quarter (as in Equation 1), $s(\text{qtr})$ a cyclic seasonal smooth, and $s(\text{lon}, \text{lat})$ a thin-plate spatial smooth, fitted by `bam()` with `weights = total_n`, a Gaussian family on the log scale, and fast restricted maximum likelihood (`fREML`, discretised).

JPDW (1957–2018): the longest available series; basis dimensions $K_T = 50$ (temporal) and $K_S = 50$ (spatial) reflect roughly six decades of quarterly data. K_T is kept moderate rather than maximal so that the temporal smooth is not over-flexible at the data-sparse start of the series, where a very high basis dimension produces an exaggerated boundary peak in the late 1950s. The window terminates at 2018 because JPDW weight contributions to the WCPFC database dropped sharply afterward (2019 contained approximately 2,000 records versus 28,000 in 2018, and 2020 onward were absent).

USHW (2000–2019): the US Hawaii longline fleet’s BET weight data spans approximately 2000–2019. The shorter extent and smaller, stable Hawaii-centred footprint motivate more conservative bases ($K_T = 25$, $K_S = 30$). USHW’s stable footprint contrasts with JPDW’s broad and shifting one, making it an informative independent comparison: if both fleets sampled the same declining sub-population, they should show concordant trends.

Reporting reference. Single-fleet trends are reported as the *marginal* temporal smooth $\exp(\beta_0 + s(t))$ — the model’s average-location trend, consistent with how the region-wide headline trend is reported. This is a geometric, average-location reference; the alternative footprint-averaged prediction (model predictions evaluated at the cells the fleet actually fished each year, averaged with sample-size weights) reads a few kilograms lower but shares the same shape. The fixed-area decomposition below works on the footprint-restricted predictions directly, so the within-area magnitudes are reported on that footprint basis.

Evolving spatial pattern. Because the base model’s $s(\text{lon}, \text{lat})$ is time-invariant, a supplementary model with a separate spatial surface per decade window — $s(\text{lon}, \text{lat}, \text{by} = \text{window})$ — is fitted for JPDW to visualise how the *predicted* spatial pattern of mean weight changes through time. Spatial predictions are

restricted to data-supported cells, so maps reflect sampled areas rather than spline extrapolation beyond the data's support.

Table 2 summarises the model identifiers used throughout the paper, across the region-wide and single-fleet analyses.

Table 2: Model identifiers used in the text, tables, and code

Model	Description
m_base	Region-wide base model: global temporal + seasonal + spatial smooths + flag random intercept
m_temp_flag	Most general region-wide model: m_base plus flag-specific temporal deviations (bs="fs")
m_spat_flag	Comparison model: flag-specific (by-flag) spatial smooths instead of flag temporal deviations
m_temp_flag_ar1	AR(1) robustness refit of m_temp_flag (Section A.7)
m_ju	Region-wide base structure refitted on JPDW + USHW only; compositional check (Section A.8)
jpdw_mod	Single-fleet JPDW historical model (1957–2018)
ushw_mod	Single-fleet USHW historical model (2000–2019)

3 Results

3.1 Region-wide multi-fleet results

Based on AIC model selection, `m_temp_flag` is the best-fitting of the three candidates (Table 3). Adding flag-specific temporal deviations (via `bs = "fs"`, a random smooth) improves fit by $\Delta\text{AIC} \approx 242$ over `m_base`. By contrast, allowing spatial patterns to differ by flag (`m_spat_flag`) is worse than `m_base`, indicating that a global spatial smooth plus a flag random intercept is sufficient to describe how flags differ in the sizes of fish they catch. A region-based temporal model is near-collinear with the flag structure given the $\text{flag} \times \text{region}$ confounding (Table 1), so was not explored.

Table 3: AIC comparison of candidate region-wide models.

model	df	AIC	delta_AIC
m_temp_flag	114.9	3565.7	0.0
m_base	84.6	3807.9	242.2
m_spat_flag	105.7	4214.7	649.0

Model diagnostics for `m_temp_flag` indicate the assumptions are reasonably met (Section A.6). A small number of observations have large absolute deviance residuals; all come from JPDW and USHW — the two largest flags by sample size. The most notable cluster is JPDW at $\text{lat} = 30^\circ$, $\text{lon} = 135^\circ$ (Kuroshio region), which repeatedly produces very low mean weights (as low as 11.8 kg) in Q2 (April–June) across many years.

The global temporal trend peaks at approximately 47 kg in the early 1990s and declines to around 31 kg by 2020 (Figure 5). The confidence interval widens at both ends of the series, where data are sparser and spline boundary effects apply. Mean weight also shows a clear seasonal effect (Figure 5).

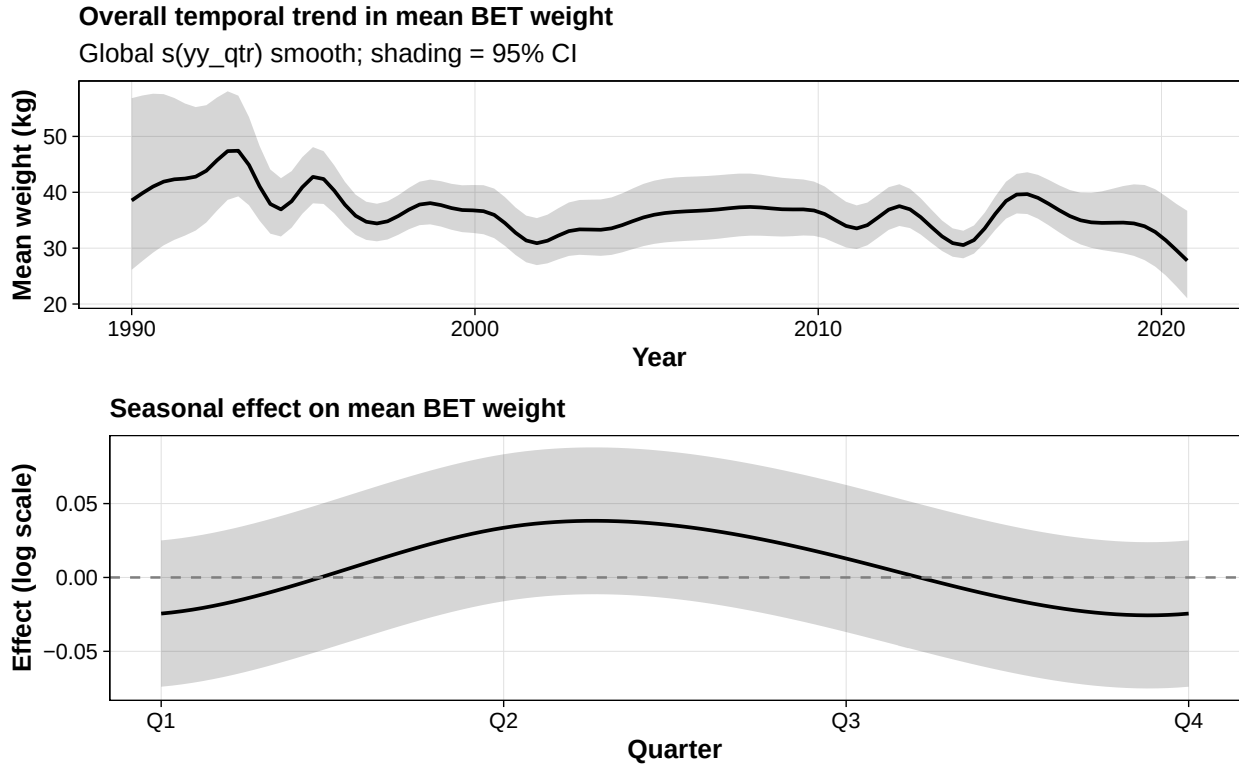


Figure 5: Overall temporal trend in mean BET weight across the analysis period (top panel) and the seasonal effect (bottom panel), from `m_temp_flag`. Shading is the 95% CI.

The `s(yy_qtr, flag.fleet, bs = "fs")` term fits flag-specific deviations from the global trend (Figure 6). These deviations are penalised toward zero, so data-sparse flags are appropriately shrunk toward the global trend.

On the kilogram scale (Figure 7), CN has been consistently below the global trend since the early 2000s; SB shows a pronounced hump — above trend around 2010–2015 followed by a sharp decline to well below 30 kg by 2020; JPDW dipped below the global trend in the 1990s and has tracked at or above it in recent years. Flags with sparse coverage (MH, PW, TO) have wide confidence intervals and their trend lines are largely a product of the penalisation toward zero.

It is important to note that these are deviations *relative to* the declining region-wide average over 1990–2020: a flag tracking “at or above” the global trend in recent years is still consistent with a falling absolute mean weight, because the reference trend is itself declining. The JPDW deviation here therefore does not conflict with the absolute long-term JPDW decline documented in the single-fleet analysis (Figure 8; Table 4), which spans a longer window (1957–2018) and is measured on the kilogram scale rather than against a pooled multi-fleet baseline.

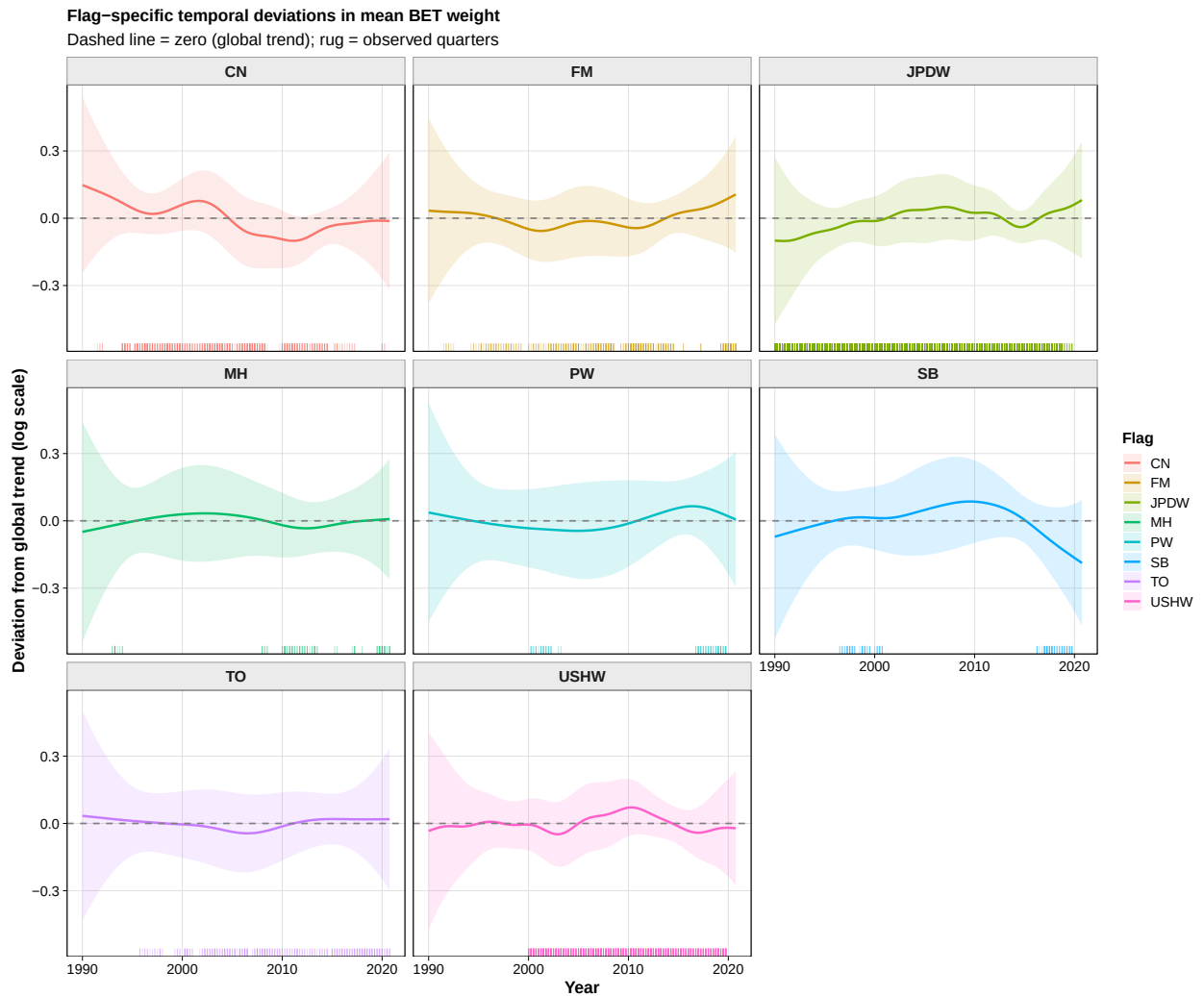


Figure 6: Flag-specific temporal deviations from the global trend (log scale). Dashed line = zero (global trend); rug = observed quarters.

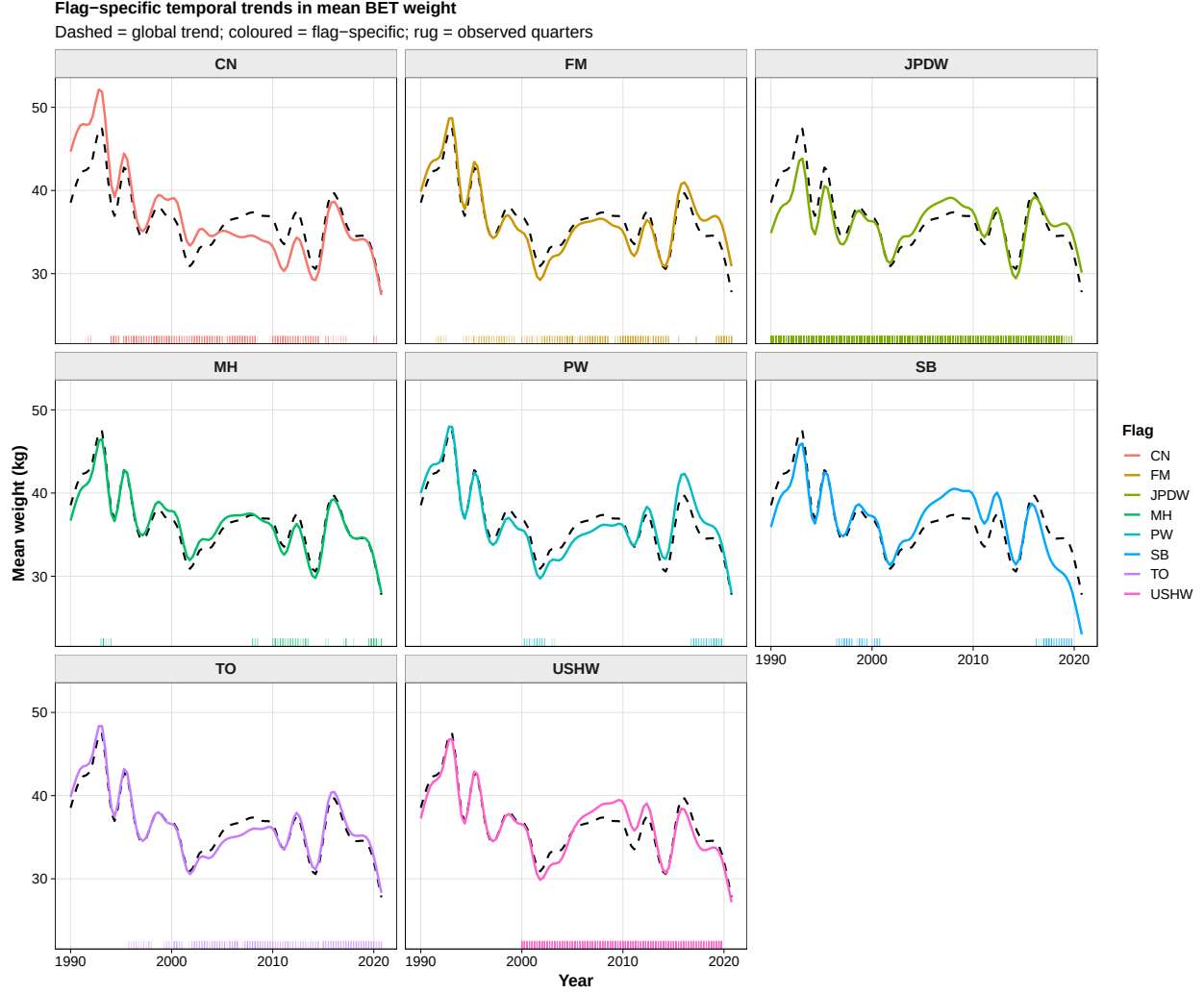


Figure 7: Flag-specific temporal trends in mean BET weight (kg), derived from the global temporal smooth plus each flag's deviation smooth from `m_temp_flag`. Dashed black line = global trend; coloured lines = flag-specific trends; rug = observed quarter-years.

3.2 JPDW historical analysis (1957–2018)

The JPDW fleet provides the longest available time series (1957–2018) and therefore offers the clearest opportunity to examine long-term changes in BET mean weight. The following analyses evaluate temporal trends, residual structure, spatial shifts, and the relative contributions of spatial redistribution and within-area changes.

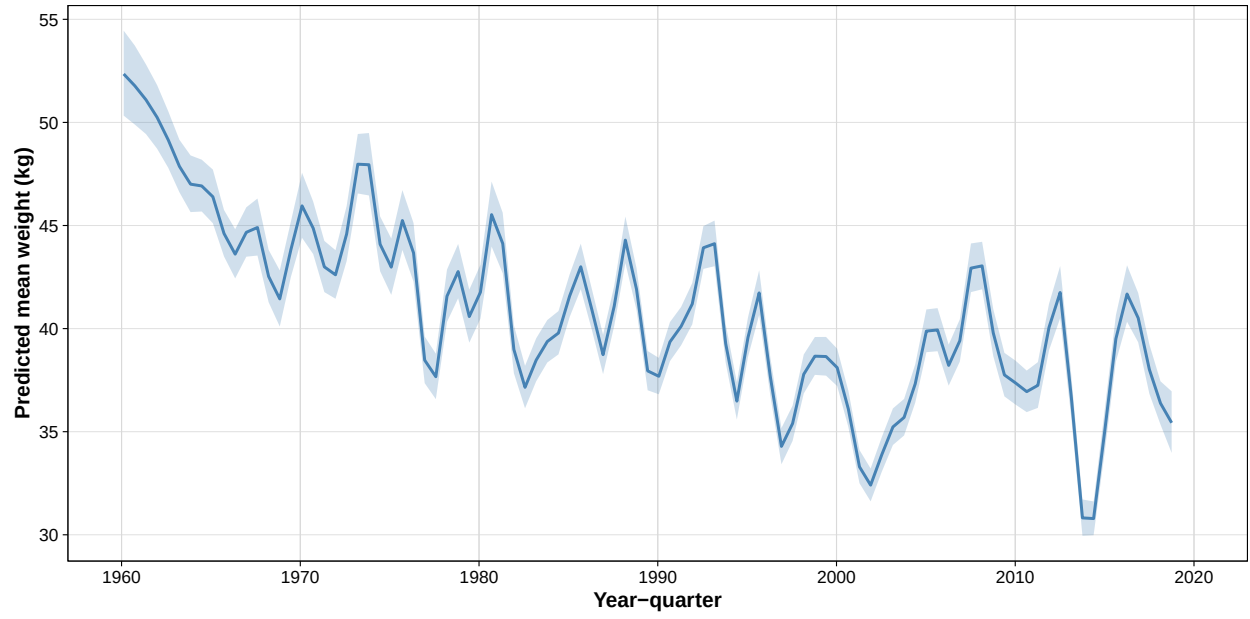


Figure 8: JPDW BET predicted mean weight (reported from 1960; see text), with 95% confidence band. Single-fleet GAM fitted to the full 1957–2018 record.

Table 4: JPDW BET predicted mean weight at milestone years, with 95% confidence intervals. Reported trend begins in 1960 (see text).

Year	Mean wt (kg)	Lower 95% CI	Upper 95% CI
1965	46.4	45.1	47.7
1985	41.6	40.6	42.6
2000	38.1	37.2	39.0
2015	35.0	34.0	36.1

The JPDW analysis (Figure 8; Table 4) shows a clear long-term decline in mean weight. The temporal trend is reported from 1960 onward: the late 1950s sit at the boundary of the smooth, where sparse, spatially limited early sampling leaves it weakly constrained and prone to an exaggerated edge value, so those years are set aside (the model fit and the spatial decomposition below both retain the full record from 1957). From roughly 46 kg in the mid-1960s the trend declines to 35 kg by 2015, a fall of about 11 kg (robust to the weight cap, error family, and temporal basis dimension; Section A.10). The mid-1960s and 2015 confidence intervals do not overlap (45–48 kg vs 34–36 kg), establishing the long-term decline as statistically robust. The intermediate milestones (Table 4) trace a steady decline from the mid-1960s to around 2000, after which the decline continues at a reduced but still discernible rate — the underlying quarterly series fluctuates from year to year (Figure 8), and the slight upturn at the very end sits within a wider, sparser-data boundary interval.

Residual diagnostics for the JPDW model (Section A.9, Figure 25) show residual autocorrelation at short lags: the multi-scale temporal structure typical of size-composition series, also present in the region-wide model (Section A.7). In both cases the structure is documented as a known limitation rather than corrected, since a single AR(1) term does not fully capture it; the long-term trend is the quantity of interest and is robust to this structure.

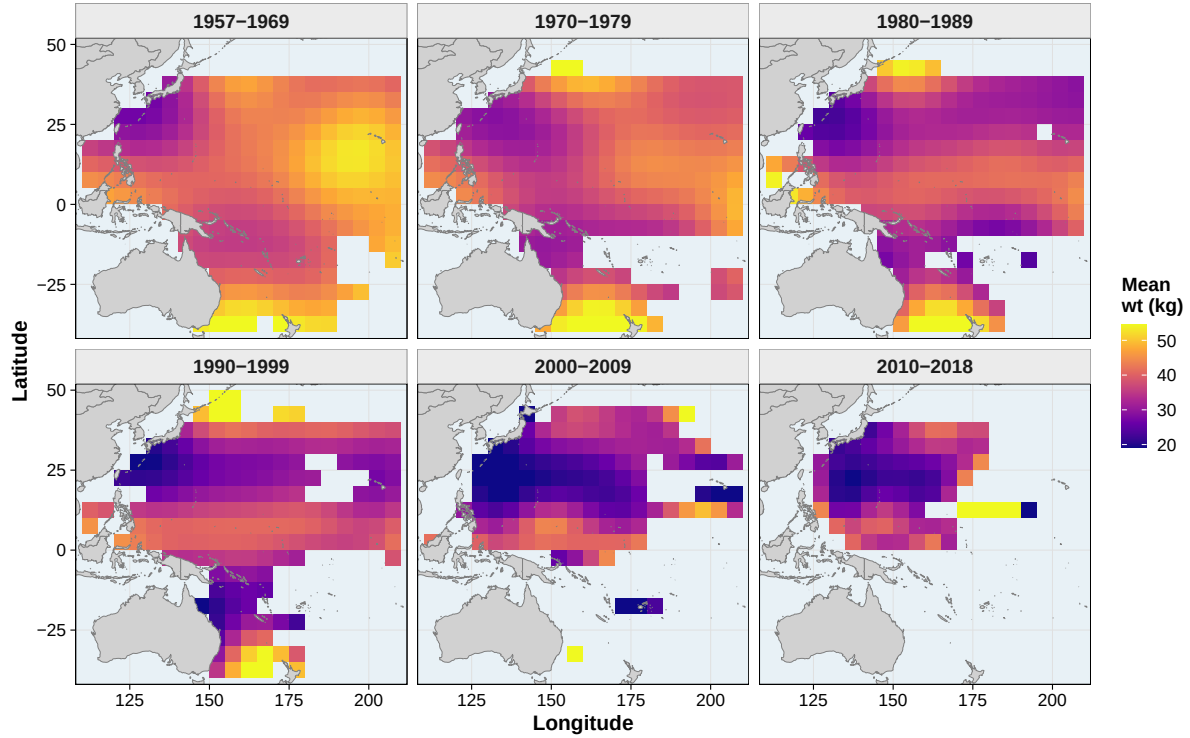


Figure 9: JPDW BET predicted spatial pattern of mean weight (kg), quarter 2. Each panel is the per-cell average of model predictions across the year window labelled (e.g. the 1970-1979 panel averages predictions for each year from 1970 to 1979). Each panel is masked to the 5° cells JPDW sampled within that window. The middle four panels each span exactly 10 years; the first (1957-1969, 13 years) and last (2010-2018, 9 years) panels extend to the start and end of JPDW BET sampling respectively, so the six panels together cover every year of the analysis. The progressive contraction of the coloured area shows the contraction of the JPDW fleet’s operational footprint over time.

The decadal spatial sequence (Figure 9) shows the spatial contraction of the JPDW fleet’s operational footprint. In the 1960s and 1970s the fleet sampled broadly across the western and central Pacific — including sampling beyond the five assessment regions (Section A.11) — whereas by the 2010s it was largely restricted to a narrower band within the WCPO. This spatial contraction is essential context for interpreting the temporal decline in mean weight. The decline is further quantified by boxplots at 5-year (Figure 10) and 10-year (Figure 11) resolution, showing the progressive downward shift in quarterly sample-weighted mean weight.

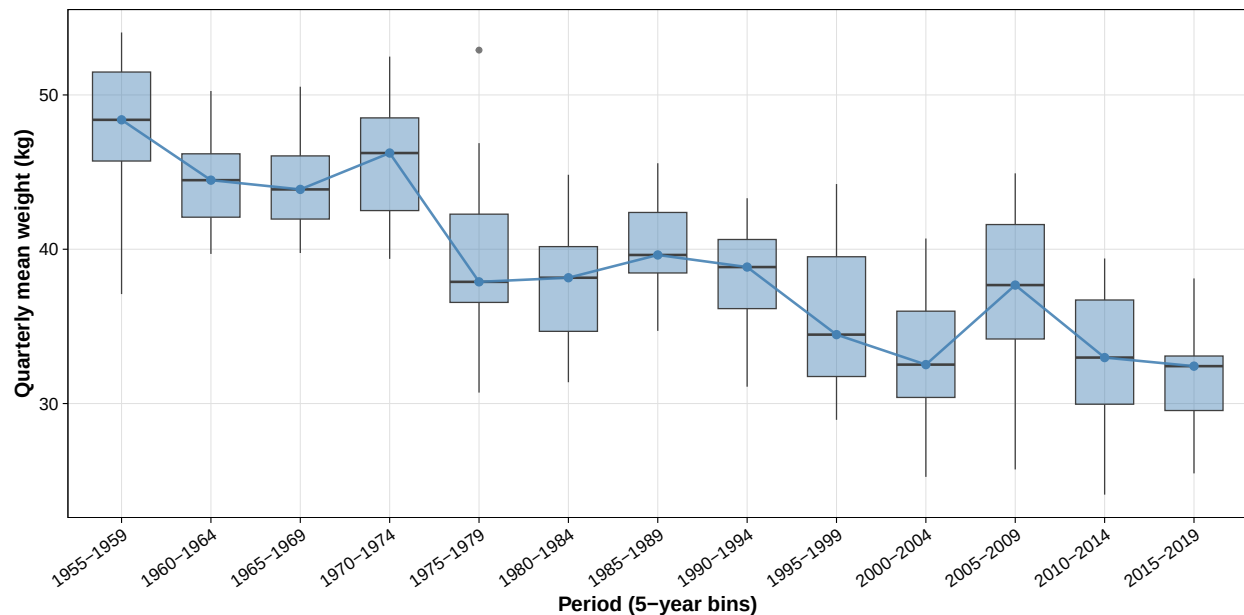


Figure 10: JPDW BET quarterly mean weight binned by 5-year periods (1955–1959, 1960–1964, ...). Each box summarises the distribution of quarterly sample-weighted mean weights within the period; the connecting line traces the period medians.

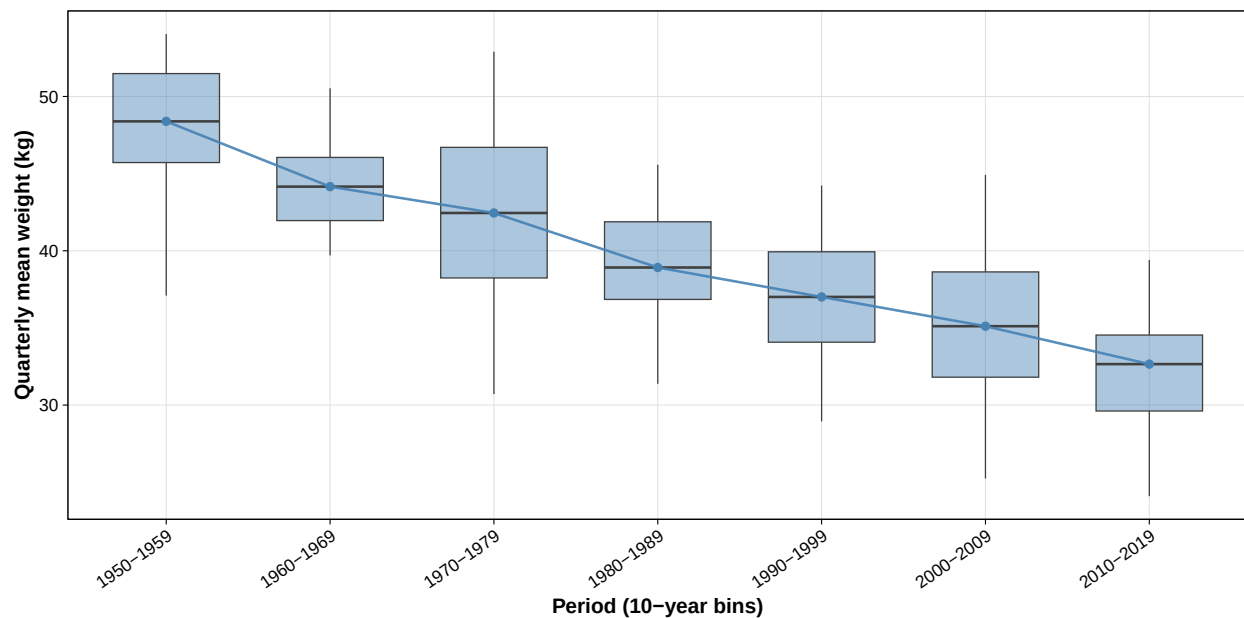


Figure 11: JPDW BET quarterly mean weight binned by 10-year periods.

To isolate genuine temporal change from this fleet-redistribution confounding, Figure 12 replots the same six time windows but restricts every panel to the **2010–2018 sampling footprint**. Because the spatial mask is held fixed across panels, any colour difference between panels reflects a temporal change in predicted mean weight *in the same area*, not a change in which areas were sampled. Comparing the 1957–1969 panel with the 2010–2018 panel in this figure asks: in cells the fleet fishes today, has the BET mean weight there changed over time?

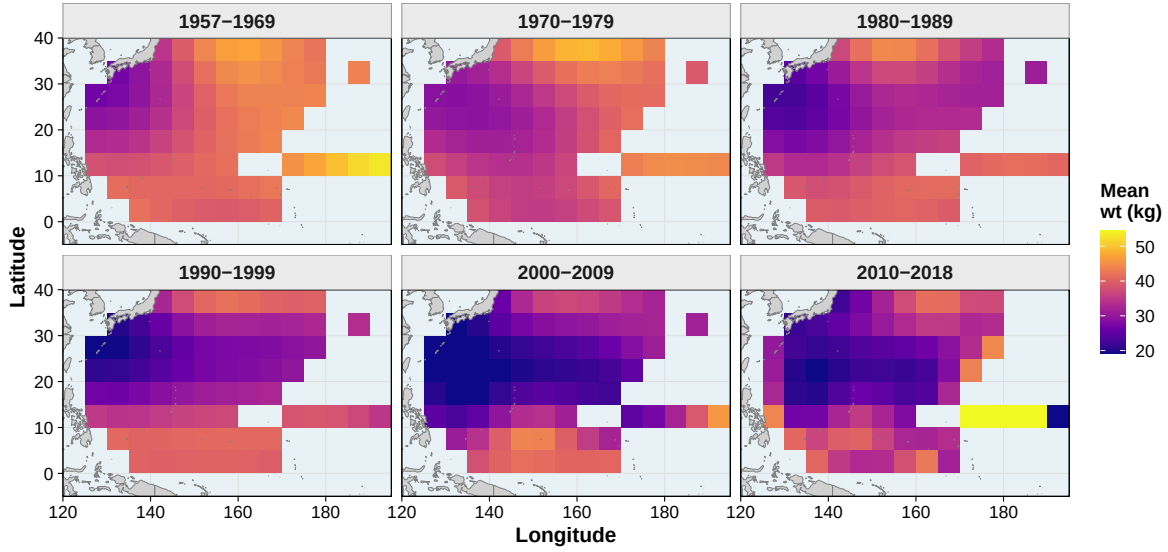


Figure 12: JPDW BET predicted spatial pattern of mean weight (kg) restricted to the 2010-2018 sampling footprint, replotted across the same six time windows as the by-decade spatial maps. Because every panel uses the same spatial mask (the cells the fleet sampled during 2010-2018), differences in colour between panels reflect temporal changes in predicted mean weight in the fixed area, not changes in the fleet’s operational footprint. This isolates the within-area temporal change from the spatial-contraction signal shown in the by-decade maps. The colour scale is shared with the by-decade maps to allow direct comparison.

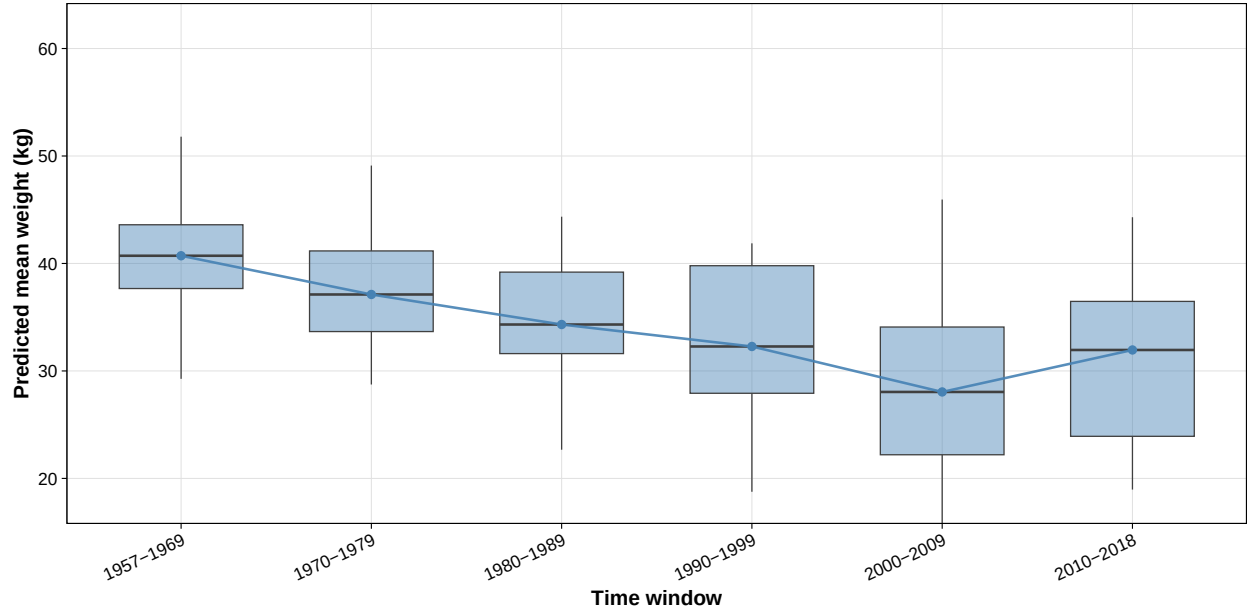


Figure 13: Boxplots of JPDW BET predicted mean weight (kg) across all cells in the 2010-2018 sampling footprint, for each of the six time windows shown in the fixed-area spatial maps. Each box summarises the per-cell distribution of model predictions within the fixed spatial mask; the line connects period medians to trace the within-area temporal change. Because the spatial mask is held fixed across periods, any vertical shift between boxes is a within-area change in predicted mean weight, not an artefact of the fleet’s contracting operational footprint. The y-axis is cropped to the plausible range: a small number of cells in the sparsely-sampled early windows carry extreme extrapolated predictions from the per-window spatial smooth and fall off-scale; they do not affect the period medians shown.

As shown in Figure 14, the two median trajectories compare the unrestricted per-cell median (mask tracks each period’s footprint) against the fixed-area median (mask held at 2010–2018). The vertical gap between them is the fleet-redistribution contribution. Summed across the full comparison, redistribution accounts for approximately 2 kg of the total 11 kg decline, against 9 kg attributable to the within-area component; because the two trajectories are identical by construction in 2010–2018, this redistribution contribution is necessarily concentrated in the earliest, broadest-footprint periods.

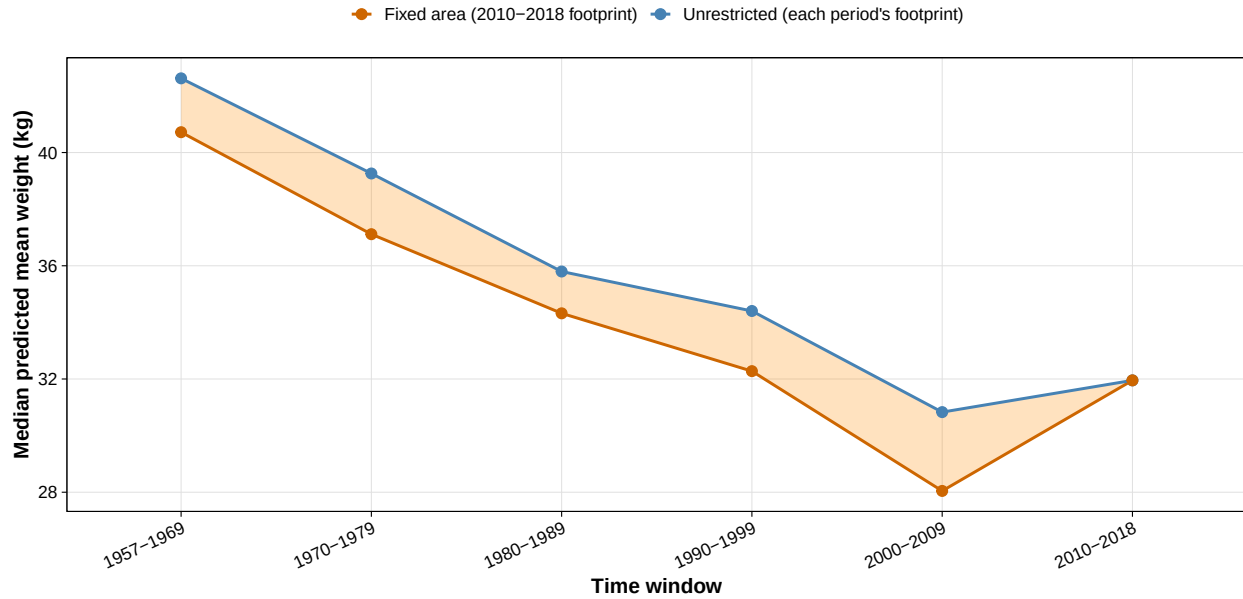


Figure 14: Decomposition of the JPDW BET mean weight decline. Steelblue line: median predicted mean weight across cells in each period’s own sampling footprint (unrestricted — spatial mask changes with fleet contraction). Orange line: median across the fixed 2010-2018 footprint (within-area — mask held constant). The shaded area between the two lines is the fleet-redistribution component: additional apparent decline arising from the fleet abandoning cells where fish were historically largest. Both lines converge in the final period (2010-2018) where the masks are identical. Time-window lengths are unequal, so the connecting lines should be read as period-to-period medians rather than a constant-rate trajectory.

3.3 USHW historical analysis (2000–2019)

The USHW fleet provides an independent, more recent Hawaii-centred perspective spanning approximately two decades. Unlike JPDW, its spatial footprint remained relatively stable through time, allowing evaluation of temporal change with minimal confounding from large-scale redistribution.

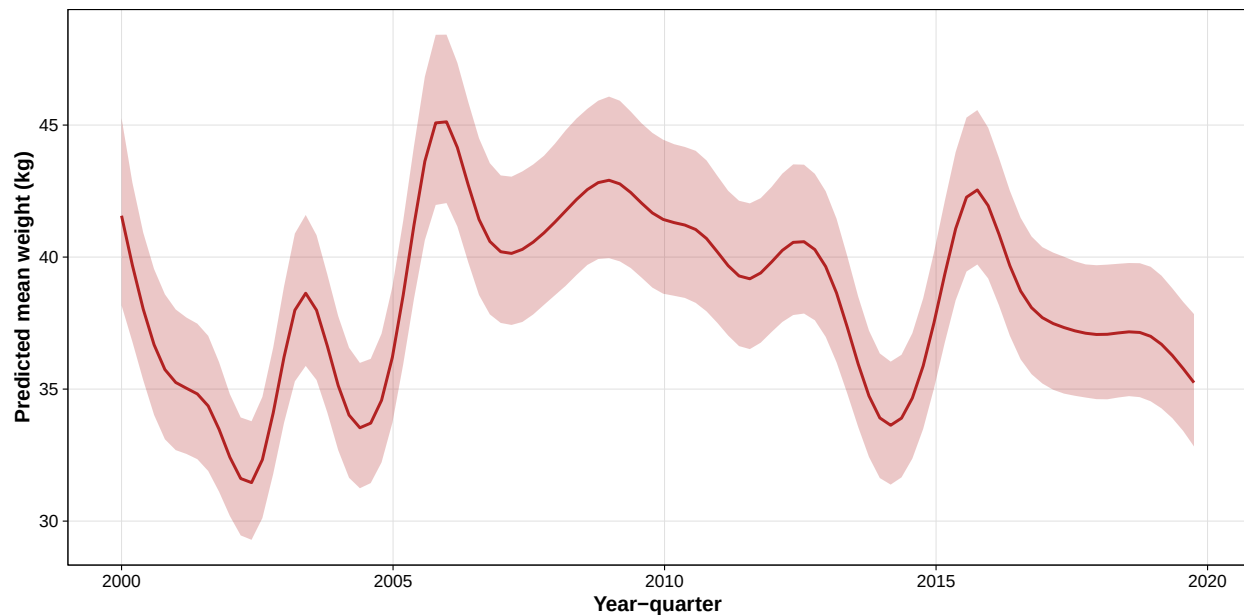


Figure 15: USHW BET predicted mean weight, 2000–2019, with 95% confidence band.

Table 5: USHW BET predicted mean weight at milestone years (2000–2019), with 95% confidence intervals.

Year	Mean wt (kg)	Lower 95% CI	Upper 95% CI
2000	41.6	38.2	45.3
2006	45.1	42.0	48.4
2010	41.4	38.6	44.4
2019	37.0	34.5	39.6

The USHW analysis (Figure 15; Table 5) shows mean weights that fluctuate within a narrow range (approximately 31–45 kg) across the two-decade window without a clear directional trend. The 5-year period boxplot (Figure 16) is consistent with this pattern, with medians remaining within a narrow range across all bins.

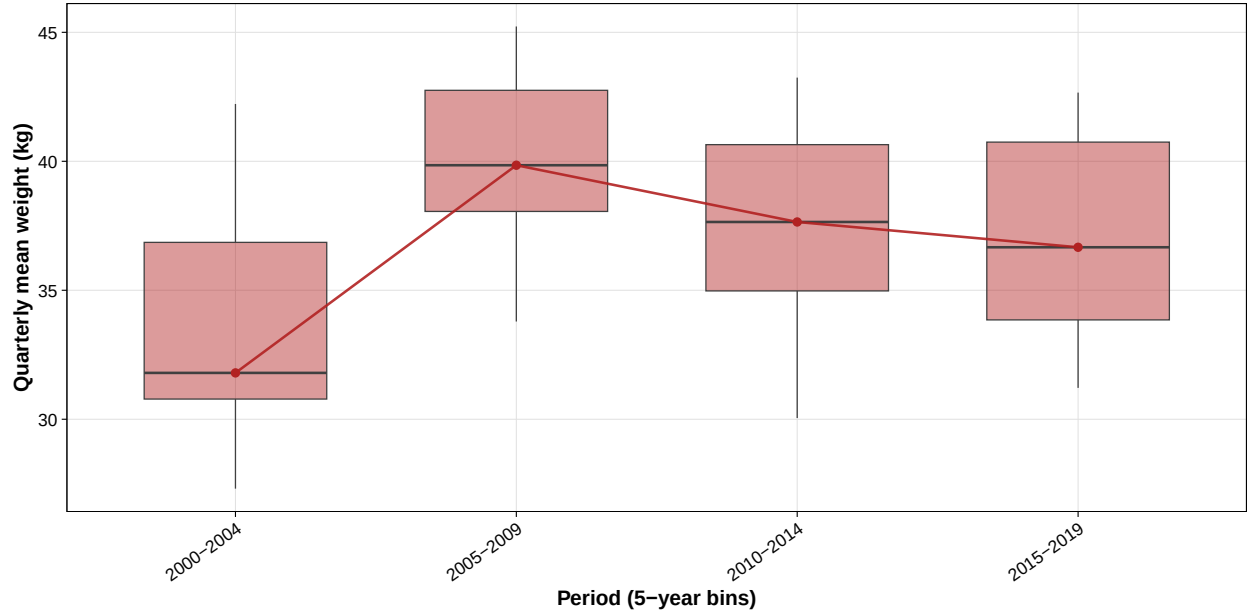


Figure 16: USHW BET quarterly mean weight binned by 5-year periods. Each box summarises the distribution of quarterly sample-weighted mean weights within the period; the connecting line traces the period medians.

Because USHW shows no directional trend and its spatial footprint is stable across the 2000–2019 window, the fixed-area decomposition was not applied to USHW — there is no decline to decompose.

3.4 Fleet comparison: JPDW versus USHW

Comparing the two single-fleet trajectories on a common temporal scale provides a direct test of whether the long-term JPDW decline is mirrored by a second, independently-sampled fleet. The key feature to note is *when* the JPDW decline occurs: it is concentrated in the early decades (pre-2000; Table 4), well before USHW weight sampling begins. Over the 2000–2018 window in which both fleets are observed, JPDW continues to decline, but much more slowly than in the early decades. Because this overlap window falls entirely after JPDW’s early-decades decline, however, the comparison does not provide comparative information on that earlier decline. As such, we cannot conclude whether the decline seen in JPDW’s weights is indicative of a basin-wide decline in bigeye mean weight or one confined to the north west Pacific region.

Figure 17 overlays the predicted mean weight trends from the two single-fleet historical analyses on a common time axis.

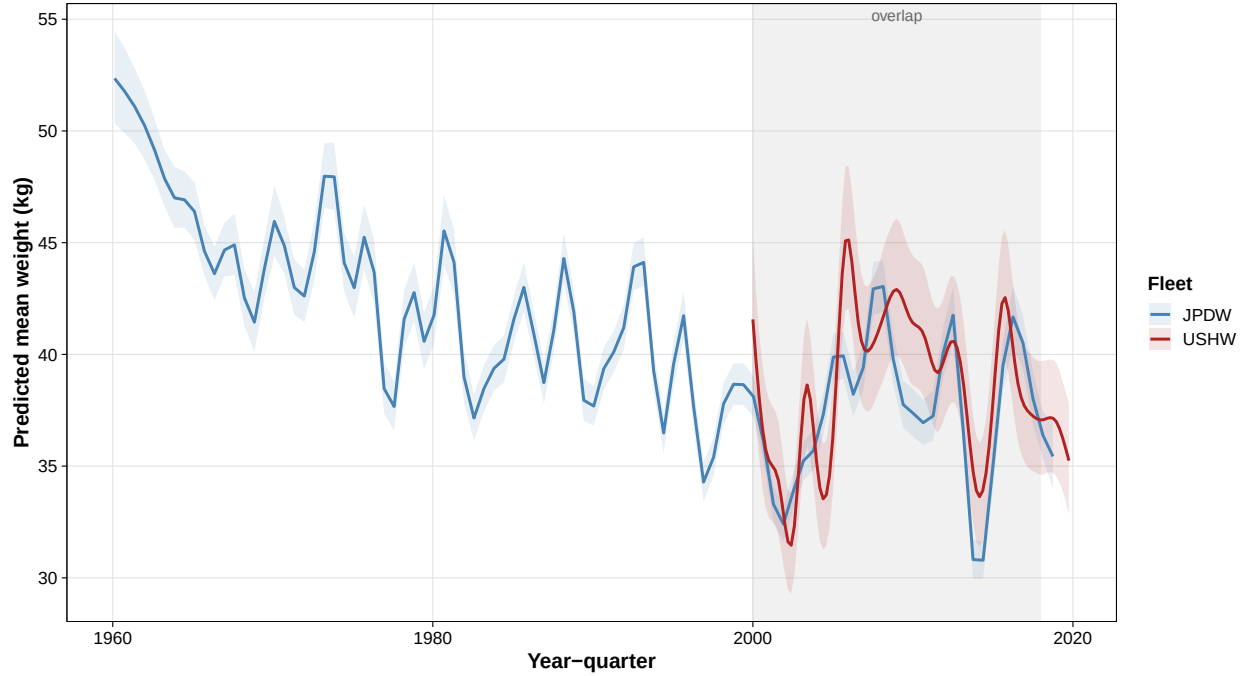


Figure 17: Comparison of single-fleet GAM predicted BET mean weight for JPDW (reported 1960-2018) and USHW (2000-2019). Each line is the temporal trend from the fleet’s independent historical GAM at the fleet-mean spatial location (quarter 2); shaded bands are 95% confidence intervals. The grey band marks the 2000-2018 period over which both fleets were sampled. JPDW’s large decline is concentrated before 2000, outside USHW’s coverage; within the overlap period the two fleets fluctuate around broadly similar mean weights with overlapping confidence bands and neither shows a strong directional trend.

The fleet comparison (Figure 17) places the two trajectories on a common axis. The dominant feature is that JPDW’s decline is largely historical: by 2000 — when USHW sampling begins — most of the JPDW decline has already occurred. Over the 2000–2018 overlap the two fleets fluctuate around broadly similar mean weights, their confidence bands overlap, and neither shows a strong directional trend. The comparison therefore does *not* demonstrate a contemporaneous divergence; rather, it shows that the recent period is stable and the short-term dynamics are similar in both sampled regions, and that the large JPDW decline predates the only window in which both fleets are sampled together.

4 Discussion

4.1 Comparison with Langley (2006)

The spatial heterogeneity Langley documented (smaller fish in the western north Pacific/PNG waters; larger fish in equatorial regions) is preserved here in the region-wide spatial smooth (Figure 21) and in the JPDW spatial maps (Figure 9). The JPDW temporal decline Langley identified through 2004 has continued through 2018, with a non-monotonic rate (steep early, slowing through the 1990s, and continuing at a reduced but still discernible rate from the early 2000s). Most importantly, Langley (2006) cautioned that fleet redistribution can mimic population change — this analysis directly tests and quantifies that concern through the fixed-area decomposition of the JPDW decline (Section 4.2).

4.2 Catching smaller fish, or fish getting smaller?

The JPDW long-term decline, set against the fleet’s contracting operational footprint (Figure 9), raises a fundamental interpretive question: do the data indicate that *fish are getting smaller* (a population-level shift in BET size structure within the area still fished), or that *fleets are catching smaller fish* (a change in what is sampled, because the fleet has redistributed into areas where fish were always smaller)?

Schematically, observed mean weight reflects three influences:

$$\bar{w}_{\text{observed}} \propto \bar{w}_{\text{population}} \times (\text{fleet selectivity}) \times (\text{spatial availability}),$$

where $\bar{w}_{\text{population}}$ is the true mean weight of the sub-population in the area fished (the *within-area* signal the fixed-area analysis below isolates), fleet selectivity is the size-dependent capture probability of the gear (assumed constant in the assessment, so it does not drive temporal change here), and spatial availability is the effect of *which* cells the fleet samples — the *fleet-redistribution* component addressed below. This is a conceptual schematic, not a fitted decomposition; the fixed-area analysis provides the empirical separation.

The fixed-area comparison (Figure 12; Figure 13) holds the spatial mask constant at the 2010–2018 JPDW footprint and asks what the model predicts for earlier periods *over the same cells*. Because the mask is fixed, any period-to-period change isolates the within-area component from the fleet-redistribution component.

4.2.1 Population change vs sampling artefact

Within the 2010–2018 fleet footprint, the median predicted mean weight across cells declines from approximately **41 kg in 1957–1969 to 32 kg in 2010–2018, a drop of about 9 kg** (22%; Figure 13), proceeding gradually rather than through a single abrupt step.

Figure 14 compares the unrestricted per-cell median trajectory (mask tracking each period’s actual footprint) with the fixed-area median trajectory (mask held constant at 2010–2018). The two converge in the final period (identical mask); in earlier periods the unrestricted median is higher because the fleet’s broader footprint included higher-weight cells later abandoned. The vertical gap between the lines is the fleet-redistribution contribution to the apparent decline.

This decomposition is descriptive. Each period is summarised by the median rather than the mean to reduce sensitivity to sparsely sampled cells and local smooth extrapolation, and the within-area and redistribution magnitudes are point estimates — we have not yet attached formal uncertainty to the *partitioning* between them, so the component sizes should be read as indicative. Critically, the decomposition establishes *where* the decline occurs — within the retained footprint rather than between abandoned cells — not *why*: it does not by itself distinguish a biological cause (truncation, growth, recruitment) from a within-cell behavioural or oceanographic one.

On this basis, Figure 14 indicates that the majority of the apparent JPDW decline is a within-area change in BET mean weight, with a smaller secondary component from the fleet ceasing to sample the southern equatorial cells that historically held the largest fish — the shaded gap, most pronounced in the earliest, broadest-footprint periods.

Whether the within-area change is a population signal or a residual sampling artefact is the question that matters for interpretation, since only a within-area change carries information about the BET population. Fleet redistribution is a well-documented confounding factor in large-scale longline fisheries. Hoyle *et al.* (2021) showed for Indian Ocean BET that mean-size trends from logbook data are sensitive to fleet-distribution shifts and must be disentangled from population signals before use in assessment, and Wu *et al.* (2026) note the analogous difficulty of separating projected BET habitat contraction from historical effort reallocation. The fixed-area decomposition is the size-composition analogue: holding the footprint constant isolates the within-area signal from the wider between-cell confounding. Three points argue that the within-area decline is not a between-cell sampling artefact:

1. **The fixed-area decomposition removes the availability confounding by construction.** By holding the spatial mask constant at the 2010–2018 footprint and predicting backward over *the same cells*, Figure 13 and Figure 14 show that the within-area trajectory cannot be produced by the fleet moving among cells — the cell set is fixed. The substantial within-area decline (~9 kg) that survives this control is therefore not attributable to between-cell spatial redistribution. Spatial redistribution explains only the additional gap between the unrestricted and fixed-area trajectories, which is the smaller component.
2. **The interquartile range narrows over time within the fixed area** (from ~35–45 kg to ~27–35 kg). A pure between-cell sampling effect along a stable spatial gradient would not necessarily contract the spread; the narrowing is more consistent with a within-area size-structure change than with a between-cell availability artefact.
3. **A pure-availability mechanism would require an implausibly steep, stable spatial gradient.** For redistribution alone to generate a ~9 kg within-area shift, the abandoned cells would have to be both systematically larger-fish cells *and* unchanging in their own size structure over six decades — yet the cells the fleet retained also decline, which the availability hypothesis cannot explain.

These points rule out between-cell redistribution as the dominant driver, but not the mechanism of the within-area change itself, which the mean-weight data cannot resolve: fishing truncation (Barnett *et al.*, 2017), growth, or recruitment change, or a non-fishing driver such as warming (Lin and Ito, 2024) all remain open, and motivate the environmental covariates and follow-up analyses below. For the assessment, the distinction is what matters: a constant-selectivity model may interpret the decline in large fish as a population signal and explain it through fishing mortality or other population processes, even if part of the observed change arises from unmodelled changes in availability or selectivity rather than from the stock itself (Section 4.3).

4.2.2 Fleet comparison: what it can and cannot show

Comparing the two single-fleet trajectories on a common axis is less conclusive than the within-area decomposition, because the fleets barely overlap in the period that matters. JPDW’s decline is concentrated before 2000 — most of the fall from about 46 kg in the mid-1960s to 38 kg occurs over 1965–2000 (Table 4) — whereas USHW weight sampling only begins around 2000. Over the 2000–2018 window in which both fleets are observed, JPDW continues to decline, but much more slowly than in the early decades, sits at mean weights similar to USHW, and the two confidence bands overlap (Figure 17). The comparison therefore does *not* show a contemporaneous divergence between the fleets.

What it does show is more limited but still useful:

- **USHW** (stable, Hawaii-centred footprint) shows no decline over its 2000–2019 window: there is no evidence of an ongoing, recent reduction in BET mean weight in its region.
- **JPDW** (broad, contracting footprint) shows a large decline that is largely historical; over the recent overlap period it continues to decline, but much more slowly than in the early decades. Its within-area decline (the dominant component after the decomposition) is consistent with a persistent within-area size-structure change in the region it fishes, compounded, to a much lesser degree, by redistribution out of historically large-fish cells.

Two consequences follow. First, because USHW does not extend back into the decades when the JPDW decline occurred, the comparison **cannot establish whether that historical decline was basin-wide or confined to JPDW’s region** — the fleets are not contemporaneous where it would count. Second, the recent period (post-2000) appears stable in both sampled regions, so the assessment-relevant size signal in JPDW composition data is dominated by the historical decline that the fixed-area decomposition isolates, rather than by an ongoing recent trend.

The two largest long-running fleets show no *large* recent decline individually, yet the region-wide *pooled* trend continues to fall to 2020 (from approximately 47 kg in the early 1990s to around 31 kg; Figure 5). The two observations are not in conflict, and the reconciliation is itself part of this paper’s argument. The region-wide decline is front-loaded — a sharp drop across the early-to-mid 1990s followed by a long plateau, with a further dip at the very end of the series — and part of its late-period decline is driven by particular flags rather than by JPDW or USHW: SB rises above trend around 2010–2015 and then falls sharply, and CN sits persistently below trend (Figure 6; Figure 7). As flags enter, leave, and change their share of the samples through time, the pooled mean can drift downward independently of any single fleet’s own trend. A direct check quantifies how much. Refitting the region-wide base structure on JPDW and USHW alone, over 2000–2018 (avoiding the sparse final two years) the full eight-flag global trend falls by about 2.1 kg while the two-fleet trend falls by about 1.5 kg; extending to 2020 gives 5.4 and 2.7 kg (Section A.8). Roughly 29% of the 2000–2018 pooled decline is therefore associated with the changing flag mix (the gap between the two trends), the remainder a decline present in both of the two main fleets. The recent pooled decline is thus only *partly* compositional — substantial enough to matter, but not so dominant that the decline can be dismissed as a pure averaging artefact.

The practical message for the assessment is therefore carefully bounded: size-structure signals from longline composition data are **fleet-, region-, and time period-specific**, and pooling fleets of different times and footprints under constant selectivity risks conflating a historical, region-specific decline with current conditions. This need for spatially- and temporally-resolved treatment is recognised in the broader stock assessment literature (Thorson and Haltuch, 2019; Maunder *et al.*, 2020; Hoyle *et al.*, 2021).

4.2.3 Follow-up analyses to refine the decomposition

The decomposition is robust but the mechanism remains underdetermined. Distinguishing between population truncation (smaller upper tail), growth-rate change (whole distribution shift), and recruitment change requires additional information: (1) CPUE analysis by size class; (2) multi-fleet spatially-aligned comparison (consistency of selectivity); (3) length-frequency distribution shape analysis; (4) age composition from otolith ageing (Farley *et al.*, 2020); (5) environmental and climate covariates (Pacific Decadal Oscillation, El Niño–Southern Oscillation [ENSO], Warm Pool position) which may help explain the regional pattern of the historical decline (Lehodey *et al.*, 2003; Senina *et al.*, 2020).

4.3 Considerations for the stock assessment

The decomposition has direct implications for the MFCL workflow. Within MFCL, gear-specific composition data are integrated through fleet-specific selectivity functions that are assumed constant over time (Fournier *et al.*, 1998; Day *et al.*, 2023). The assumption of temporally constant selectivity is a widely noted limitation in tuna stock assessments: when fishing fleets redistribute spatially over time, the realised (population-level) selectivity changes even if contact selectivity remains fixed (Sampson and Scott, 2012; Waterhouse *et al.*, 2014; Maunder *et al.*, 2020; Hoyle *et al.*, 2021). If left unaccounted for, such selectivity changes can bias key assessment quantities, including estimates of fishing mortality and recruitment (Ono *et al.*, 2015; Stewart and Monnahan, 2017). Figure 14 identifies two components of the JPDW decline, each mapping to a different concern for the assessment:

- **The within-area component** (the larger contributor) appears to carry information about BET size structure that the assessment should consider. Under the current constant-selectivity assumption, this signal will be absorbed into estimates of recruitment, growth, or fishing mortality. Whether that absorption is biologically appropriate depends on which mechanism drives the within-area decline (fishing

truncation vs growth vs recruitment) — a question the follow-up analyses listed above are needed to resolve.

- **The fleet-redistribution component** (the secondary contributor) is a selectivity artefact: it reflects the fleet sampling different parts of the BET population over time, not a change in the population itself or in the fleet’s intrinsic gear selectivity. Under the current constant-selectivity assumption, this component will be misattributed to population dynamics, introducing a modest systematic bias.

Possible avenues, for the assessment team to consider, include:

1. **The fleet-redistribution bias can be bounded directly with an in-assessment sensitivity run.** Rather than resolve the selectivity question in advance, it can be tested inside the 2026 assessment: fit the base model alongside a sensitivity that either (i) excludes JPDW composition data from the most contracted period (the recent years when its operational footprint was smallest — the ~2010–2018 footprint used as the fixed mask above), or (ii) allows JPDW selectivity a single time-block at the footprint-contraction point, then compare terminal depletion ($SB_{\text{recent}}/SB_{F=0}$), the recruitment deviates, and the growth/mortality estimates between the two fits. If the estimates barely move, the contracted-period JPDW size signal is not materially influencing the assessment and no further treatment is needed; if they move, the size of the movement quantifies how much the within-area and redistribution signals matter and which estimated quantity (most plausibly recruitment, under constant selectivity) is absorbing them. The 2026 assessment already stratifies the fishery-dependent composition data by region, which accounts for part of the spatial structure in fleet sampling and is an improvement in the right direction (Teears, 2026). Should the sensitivity nonetheless indicate a material effect, natural avenues for future assessments include time-block selectivity for the affected fleets — or, more generally, state-space formulations that estimate the degree of temporal change in selectivity directly from the data rather than pre-specifying block boundaries (Nielsen and Berg, 2014) — and ways of down-weighting the composition data from specific periods (e.g. separate early- and late-period fleets), so that intervals of atypical spatial coverage contribute less to the fit. These treatments alone, however, would not guarantee resolution of the bias, which remains confounded with other sources of uncertainty in the assessment.
2. **Greater weight on fleets with stable spatial footprints (e.g. USHW) as size-structure indices for the recent period.** USHW’s stability over 2000–2019 indicates no recent decline in the Hawaii-centred sub-population it samples, in contrast to the long-term (largely pre-2000) decline in the broader WCPO area JPDW historically sampled, though the shorter USHW series warrants caution (below). What makes such a fleet informative is the stability of its coverage, not the size of the area: a stable footprint gives an interpretable temporal signal, whereas a shifting one confounds real change with changes in sampling — reliably separable only when the coverage shift is not too extreme (Ducharme-Barth *et al.*, 2022).
3. **Follow-up analyses to identify the mechanism of the within-area decline** (detailed above): these are needed to determine whether the within-area component reflects population truncation, growth-rate change, or recruitment dynamics — distinctions that affect how MFCL should represent the signal.

4.4 Methodological progress and limitations

The present analysis improves on Langley (2006) in three respects:

1. **Continuous spatiotemporal smoothing** rather than discrete median-by-block analysis. The GAM framework resolves finer-scale patterns and provides confidence intervals on the fitted trends.
2. **Joint multi-fleet estimation** with shared spatiotemporal structure but fleet-specific temporal smooths via the `bs = "fs"` factor-smooth construction (Pedersen *et al.*, 2019).
3. **Transparent treatment of residual autocorrelation.** The JPDW single-fleet analysis indicates that long fish-size composition time series show multi-scale temporal structure (residual ACF persisting across multiple short lags rather than decaying as a single AR(1) process would; Figure 25) that AR(1) cannot fully accommodate. The long single-fleet models therefore use the temporal smooth alone and document this as a known limitation rather than concealing it through misspecified inference. In the

region-wide multi-fleet model, residual autocorrelation over the shorter 1990–2020 span is moderate and predominantly short-lag; an AR(1) refit leaves the trend essentially unchanged, so the uncorrected model is used with that robustness check documented (Section A.7).

4.4.1 A GAM rather than spatiotemporal models

A GAM framework was chosen over purpose-built spatiotemporal models (sdmTMB, VAST, INLA/SPDE (Lindgren *et al.*, 2011)) for three practical and analytical reasons.

First, the analytical goal is descriptive — characterising smooth trends in mean weight by fleet, time, and space and decomposing an apparent decline — rather than producing a spatially-explicit abundance index, which is where Gaussian-random-field models excel (Thorson, 2019). The smooth and factor-smooth constructions in `mgcv` (Wood, 2017) capture the required spatial-by-quarter (via continuous-by scaling) and fleet-specific temporal structure directly and transparently.

Second, the fixed-area decomposition central to this paper requires straightforward control over predictions along arbitrary fixed spatial masks across time. In GAM, this is direct: predictions can be generated over any spatial subset and propagated backward through the fitted smooth. In random-field frameworks where the latent field is tied to a mesh, it is more complex.

Third, `mgcv::bam()` (Wood *et al.*, 2015) scales efficiently to the large composition dataset with discrete covariates and parallel fREML. Fitting a full spatiotemporal random-field model at this volume is considerably more computationally demanding.

A follow-up using spatiotemporal random-field models — the SPDE approach (Lindgren *et al.*, 2011), implemented in packages such as `sdmTMB` (Anderson *et al.*, 2025), `VAST` (Thorson, 2019) and `inlabru` (Bachl *et al.*, 2019) — would bring real advantages: a principled treatment of space-time correlation that propagates into the uncertainty, and geostatistically based prediction into unsampled cells rather than spline extrapolation. Additive latent-Gaussian formulations of this kind can also combine temporal components at different scales — a smooth long-term trend, a short-lag autoregressive term, and a seasonal term — offering a more natural representation of the multi-scale residual autocorrelation noted above than the single AR(1) term assessed here. This is identified as a worthwhile methodological extension for future work.

4.4.2 Limitations

Three key limitations apply:

- The USHW comparison is limited by shorter duration (2000–2019) and narrower spatial footprint; extending to additional fleets (TWOS, CN, AU) would strengthen inference.
- The decomposition partitioning is currently presented as a point estimate without propagated uncertainty, and the relative contributions of within-area change and spatial redistribution should therefore be interpreted as approximate.
- Mean weight is an integrated indicator that cannot uniquely identify underlying size or age structure (Audzijonyte *et al.*, 2025); resolving the mechanism (truncation, growth-rate change, or recruitment) behind the observed decline requires the follow-up analyses identified in Section 4.2.3.

5 Conclusion

A flexible GAM-based re-analysis extends the Langley (2006) analysis using continuous spatiotemporal smoothing, joint multi-fleet estimation with fleet-specific temporal smooths, and AIC-based model selection. The region-wide model (`m_temp_flag`) is a Gaussian GAM on log mean weight, combining a global temporal smooth, a seasonal smooth, a global spatial smooth, a flag random intercept, and flag-specific temporal deviations. It shows the region-wide *average* mean weight declining substantially since the early 1990s, from a peak of approximately 47 kg to around 31 kg by 2020 — a decline robust across model specifications and to an AR(1) error structure.

This pooled average, however, masks substantial heterogeneity between flags rather than describing a uniform basin-wide trend: CN has caught below-average-weight fish for most of the series, while SB shows a pronounced hump around 2010–2015 followed by a sharp recent decline. Because most flags fish almost exclusively within a single region (JPDW being the main flag with meaningful cross-regional coverage), apparent spatial heterogeneity in the trends cannot be cleanly separated from flag-level effects — a structural feature of the data rather than a resolvable modelling choice. Against this region-wide backdrop, two single-fleet historical analyses examine the longest-running fleets — JPDW and USHW — in detail.

Single-fleet analyses show: JPDW declines from ~46 kg (mid-1960s) to ~35 kg (2015), with the decline concentrated between the mid-1960s and 2000 and continuing at a reduced but still discernible rate thereafter; USHW shows no trend over 2000–2019. **The central single-fleet result is a fixed-area decomposition** of the JPDW decline into (1) **within-area decline (~9 kg, dominant)** — isolated by holding the 2010–2018 spatial footprint fixed and consistent with a within-area size-structure change in the area still fished, and (2) **fleet-redistribution component (secondary)** — the fleet’s contraction out of historically large-fish cells, a selectivity artefact. The two fleets barely overlap where it counts: JPDW’s decline predates USHW’s 2000-onward window, and over the period both are sampled their mean weights are broadly similar and trendless. The comparison therefore shows no ongoing recent decline in either sampled region, but it cannot establish whether the historical JPDW decline was basin-wide.

Considerations for stock assessment:

1. **The within-area decline appears to be the larger component of the apparent JPDW trend and a candidate signal about BET size structure** that should be considered in the assessment. Whether it reflects population truncation, growth-rate change, or recruitment dynamics requires the follow-up analyses set out in Section 4.2.3 before the next assessment cycle.
2. **The fleet-redistribution component is a secondary bias under constant selectivity.** Its influence can be bounded directly within the 2026 assessment — for example a sensitivity that either excludes the contracted-period JPDW composition or grants JPDW selectivity a single time-block at the footprint-contraction point (see Section 4.3). Such treatments alone, however, would not guarantee resolution of the bias, which remains confounded with other estimated quantities in the assessment.
3. **Fleets with stable spatial footprints (e.g. USHW) may provide useful complementary indicators of temporal size-structure change and should be evaluated alongside fleets showing substantial footprint contraction.** USHW shows no recent decline in its region, while the JPDW signal is strongly influenced by a historical decline coincident with footprint contraction; treating the two as interchangeable under a single constant-selectivity assumption risks conflating temporal trends with fleet/area differences.
4. **Standardising the composition data to remove the spatial/footprint effect at the data stage offers a more direct route to a population-level signal than relying on selectivity adjustments within the assessment alone.** A within-assessment time-block still leaves the assessment structure to absorb the redistribution artefact and any real population signal together, since it cannot separate them; a spatiotemporal standardisation of the composition data — the random-field approach — would separate the two at the data stage instead, yielding a spatially-standardised mean-weight series that could serve as a direct, decomposition-free input to the assessment, consistent with spatiotemporal approaches already advocated for tuna composition data specifically (Maunder *et al.*, 2020).

Data and code availability

The weight composition data analysed here are held in the WCPFC operational longline composition database and are subject to the WCPFC Rules and Procedures for the Protection, Access and Dissemination of Data. The analysis was implemented as a reproducible R and Quarto workflow using the `mgcv` package; the model-fitting and reporting code is available from the authors on request.

Acknowledgements

This work was supported by funding from the Western and Central Pacific Fisheries Commission under the regular ‘scientific services’ and Project 127/127a: Review and reconciliation of size data collected in the WCPFC-CA for stock assessment purposes.

This analysis was conducted at the Pacific Community (SPC) Oceanic Fisheries Programme. The authors thank the WCPFC CMMs noted in this paper for their collection and provision of weight data to the WCPFC, and acknowledge discussions with colleagues in the Stock Assessment and Modelling team. We also thank Simon Hoyle for helpful comments during the early stages of this work.

Declaration of generative AI and AI-assisted technologies in the SC paper preparation process

During the preparation of this work the authors used Claude (Anthropic) to assist with drafting and editing text, and with refining the visual presentation of figures. All analytical work, model specifications, diagnostic interpretations, and scientific conclusions are those of the authors, who reviewed and verified all AI-assisted output and take full responsibility for the content of the paper.

References

- Anderson, S. C., Ward, E. J., English, P. A., Barnett, L. A. K., and Thorson, J. T. 2025. [sdmTMB: An R package for fast, flexible, and user-friendly generalized linear mixed effects models with spatial and spatiotemporal random fields](#). *Journal of Statistical Software*, 115: 1–46.
- Audzijonyte, A., Andersen, K. H., Atkinson, D., Bigman, J. S., Blanchard, J. L., Coghlan, A. R., Heather, F., *et al.* 2025. [Which body size metrics should be used for assessing temperature impacts on fish growth and size?](#) *Global Change Biology*, 31: e70296.
- Bachl, F. E., Lindgren, F., Borchers, D. L., and Illian, J. B. 2019. [inlabru: An R package for Bayesian spatial modelling from ecological survey data](#). *Methods in Ecology and Evolution*, 10: 760–766.
- Barnett, L. A. K., Branch, T. A., Ranasinghe, R. A., and Essington, T. E. 2017. [Old-growth fishes become scarce under fishing](#). *Current Biology*, 27: 2843–2848.
- Box, G. E. P., and Cox, D. R. 1964. [An analysis of transformations](#). *Journal of the Royal Statistical Society: Series B (Methodological)*, 26: 211–243.
- Day, J., Magnusson, A., Tears, T., Hampton, J., Davies, N., Castillo Jordán, C., Peatman, T., *et al.* 2023. Stock assessment of bigeye tuna in the western and central Pacific Ocean: 2023. WCPFC-SC19-2023/SA-WP-05. Western; Central Pacific Fisheries Commission.
- Ducharme-Barth, N. D., Grüss, A., Vincent, M. T., Kiyofuji, H., Aoki, Y., Pilling, G., Hampton, J., *et al.* 2022. [Impacts of fisheries-dependent spatial sampling patterns on catch-per-unit-effort standardization: A simulation study and fishery application](#). *Fisheries Research*, 246: 106169.
- Farley, J., Krusic-Golub, K., Eveson, P., Clear, N., Roupsard, F., Sanchez, C., Nicol, S., *et al.* 2020. Age and growth of yellowfin and bigeye tuna in the western and central Pacific Ocean from otoliths. WCPFC-SC16-2020/SA-WP-02. Western; Central Pacific Fisheries Commission.
- Fournier, D. A., Hampton, J., and Sibert, J. R. 1998. MULTIFAN-CL: A length-based, age-structured model for fisheries stock assessment, with application to south pacific albacore (*Thunnus alalunga*). *Canadian Journal of Fisheries and Aquatic Sciences*, 55: 2105–2116.
- Hamer, P., Castillo-Jordán, C., Panizza, A., and Peatman, T. 2026. Review and reconciliation of size data collected in the WCPFC-CA for stock assessment purposes: Phase 2 progress report. WCPFC-SC22-2026/ST-WP-02. Western; Central Pacific Fisheries Commission.
- Hampton, J., Kleiber, P., Takeuchi, Y., Kurota, H., and Maunder, M. N. 2002. Stock assessment of bigeye tuna in the western and central Pacific Ocean. WCPFC Working Paper, SCTB15 BET-1. Western; Central Pacific Fisheries Commission.
- Hoyle, S. D., Chang, S.-T., Fu, D., Itoh, T., Lee, S. I., Lucas, J., Matsumoto, T., *et al.* 2021. Review of size data from Indian Ocean longline fleets, and its utility for stock assessment. Working Paper, IOTC-2021-WPTT23-07. Indian Ocean Tuna Commission, Online.
- Kim, K. *et al.* 2026. Stock assessment of bigeye tuna in the western and central Pacific Ocean: 2026. WCPFC-SC22-2026/SA-WP-04. Western; Central Pacific Fisheries Commission.
- Langley, A. D. 2006. Spatial and temporal variation in the size composition of the yellowfin and bigeye longline catch in the WCPO. WCPFC Working Paper, SC2-2006/ME IP-2. Western; Central Pacific Fisheries Commission, Manila, Philippines.
- Lehodey, P., Chai, F., and Hampton, J. 2003. [Modelling climate-related variability of tuna populations from a coupled ocean–biogeochemical-populations dynamics model](#). *Fisheries Oceanography*, 12: 483–494.
- Limpert, E., Stahel, W. A., and Abbt, M. 2001. [Log-normal distributions across the sciences: Keys and clues](#). *BioScience*, 51: 341–352.
- Lin, Z., and Ito, S. 2024. [Fish weight reduction in response to intra- and interspecies competition under climate change](#). *Fish and Fisheries*, 25: 455–470.
- Lindgren, F., Rue, H., and Lindström, J. 2011. [An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach](#). *Journal of the Royal Statistical Society: Series B*, 73: 423–498.
- Maunder, M. N., Thorson, J. T., Xu, H., Oliveros-Ramos, R., Hoyle, S. D., Tremblay-Boyer, L., Lee, H.-H., *et al.* 2020. [The need for spatio-temporal modeling to determine catch-per-unit effort based indices of abundance and associated composition data for inclusion in stock assessment models](#). *Fisheries Research*, 229: 105594.

- Nielsen, A., and Berg, C. W. 2014. [Estimation of time-varying selectivity in stock assessments using state-space models](#). Fisheries Research, 158: 96–101.
- Ono, K., Licandeo, R., Muradian, M. L., Cunningham, C. J., Anderson, S. C., Hurtado-Ferro, F., Johnson, K. F., *et al.* 2015. [The importance of length and age composition data in statistical age-structured models for marine species](#). ICES Journal of Marine Science, 72: 31–43.
- Peatman, T. 2026. Analysis of size frequency data for the 2026 bigeye and yellowfin assessments. WCPFC-SC22-2026/SA-IP-06. Western; Central Pacific Fisheries Commission.
- Pedersen, E. J., Miller, D. L., Simpson, G. L., and Ross, N. 2019. [Hierarchical generalized additive models in ecology: An introduction with mgcv](#). PeerJ, 7: e6876.
- Sampson, D. B., and Scott, R. D. 2012. An exploration of the shapes and stability of population-selection curves. Fish and Fisheries, 13: 89–104.
- Senina, I., Lehodey, P., Sibert, J., and Hampton, J. 2020. [Integrating tagging and fisheries data into a spatial population dynamics model to improve its predictive skills](#). Canadian Journal of Fisheries and Aquatic Sciences, 77: 576–593.
- Simpson, G. L. 2024. gratia: Graceful ggplot-based graphics and other functions for GAMs fitted using mgcv. <https://gavinsimpson.github.io/gratia/>.
- Stewart, I. J., and Monnahan, C. C. 2017. Implications of process error in selectivity for approaches to weighting compositional data in fisheries stock assessments. Fisheries Research, 192: 126–134.
- Teears, T. 2026. Background analyses and data inputs for the 2026 WCPO yellowfin and bigeye tuna stock assessment. WCPFC-SC22-2026/SA-IP-03. Western; Central Pacific Fisheries Commission.
- Thorson, J. T. 2019. [Guidance for decisions using the Vector Autoregressive Spatio-Temporal \(VAST\) package in stock, ecosystem, habitat and climate assessments](#). Fisheries Research, 210: 143–161.
- Thorson, J. T., and Haltuch, M. A. 2019. [Spatiotemporal analysis of compositional data: Increased precision and improved workflow using model-based inputs to stock assessment](#). Canadian Journal of Fisheries and Aquatic Sciences, 76: 401–414.
- Waterhouse, L., Sampson, D. B., Maunder, M., and Semmens, B. X. 2014. [Using areas-as-fleets selectivity to model spatial fishing: Asymptotic curves are unlikely under equilibrium conditions](#). Fisheries Research, 158: 15–25.
- Wood, S. N., Goude, Y., and Shaw, S. 2015. Generalized additive models for large data sets. Journal of the Royal Statistical Society: Series C (Applied Statistics), 64: 139–155.
- Wood, S. N., Pya, N., and Säfken, B. 2016. [Smoothing parameter and model selection for general smooth models](#). Journal of the American Statistical Association, 111: 1548–1563.
- Wood, S. N. 2017. Generalized additive models: An introduction with R. Chapman; Hall/CRC.
- Wu, Y.-L., Liang, T.-Y., Hsu, W.-P., Shimada, T., Wirasatriya, A., Wetchayont, P., Qiu, C., *et al.* 2026. [Projected impacts of climate change on the spatial distribution and habitat preference of tropical tuna in the Pacific Ocean](#). Marine Environmental Research, 216: 107868.

A Appendix: Glossary of terms, notation, and codes

This appendix collects the abbreviations (Table 6), statistical notation (Table 7), fleet codes (Table 8), and key R functions and packages (Table 9) used throughout the paper, for ease of reference. The model identifiers are summarised in the Methods (Table 2).

A.1 Abbreviations and acronyms

Table 6: Abbreviations used in the paper

Term	Definition
BET	Bigeye tuna (<i>Thunnus obesus</i>)
YFT	Yellowfin tuna (<i>Thunnus albacares</i>)
WCPO	Western and Central Pacific Ocean
WCPFC	Western and Central Pacific Fisheries Commission
SPC	Pacific Community (Secretariat of the Pacific Community)
OFP	Oceanic Fisheries Programme (of SPC)
SAM	Stock Assessment and Modelling team (of OFP)
SC21 / SC22	Scientific Committee, twenty-first / twenty-second regular session
PAW	Pre-Assessment Workshop
MFCL	MULTIFAN-CL, the length-based age-structured assessment platform (Fournier <i>et al.</i> , 1998)
LL	Longline (gear type)
GAM	Generalised additive model
AR(1)	First-order autoregressive (error) process
ACF	Autocorrelation function
CPUE	Catch per unit effort
ENSO	El Niño–Southern Oscillation
AIC	Akaike information criterion
EDF	Effective degrees of freedom (of a fitted smooth)
fREML	Fast restricted maximum likelihood (the fitting criterion used by <code>mgcv::bam</code>)
VAST	Vector Autoregressive Spatio-Temporal (spatiotemporal modelling package) (Thorson, 2019)
INLA	Integrated Nested Laplace Approximation
SPDE	Stochastic partial differential equation (approach to spatial random fields)
CI	Confidence interval

A.2 Statistical notation

Table 7: Mathematical notation

Symbol	Meaning
$\mathbb{E}[\cdot]$	Expectation
$\bar{w}_{f,t,\text{cell}}$	Sample-weighted mean weight (kg) for fleet f , year-quarter t , spatial cell (Equation 1)
w_i	Individual fish whole weight (kg)
freq_i	Number of fish represented by record i
n	Sample size (number of fish) in a cell $\times$ quarter; used as the model weight
$\hat{w}_i$	Model-fitted mean weight on the kg scale, $\exp(\hat{\eta}_i)$
η_i	Linear predictor, i.e. fitted $\mathbb{E}[\log \bar{w}]$ (log-kg scale)
β_0	Model intercept
b_f	Flag random intercept (region-wide model)
$s(\cdot)$	A penalised smooth function
t	Time, expressed as year-quarter (<code>yy_qtr</code> , e.g. 2005.25 = Q2 2005)
$\hat{\rho}$	Pooled lag-1 residual autocorrelation (Section A.7)
k	Basis dimension of a smooth (maximum complexity)
N	Number of fleet $\times$ year-quarter $\times$ cell observations

A.3 Flag-fleet codes

Table 8: Flag-fleet codes referenced in the paper. The region-wide analysis retains the subset of these flags meeting the inclusion criteria (see Methods); AU, NZ, TWDW, TWOS, and VN are excluded from the region-wide analysis on data-sparsity or data-quality grounds.

Code	Fleet
AU	Australia longline
CN	China longline
FM	Federated States of Micronesia longline
JPDW	Japan distant-water longline
MH	Marshall Islands longline
NZ	New Zealand longline
PW	Palau longline
SB	Solomon Islands longline
TO	Tonga longline
TWDW	Taiwanese distant-water longline
TWOS	Taiwanese offshore longline
USHW	United States (Hawaii) longline
VN	Vietnam longline

A.4 Key R functions and packages

Table 9: Key R functions and packages

Function / package	Role
<code>mgcv::bam</code>	Fits GAMs to large datasets; supports <code>discrete=TRUE</code> , AR(1) via <code>rho/AR.start</code> (Wood <i>et al.</i> , 2015; Wood, 2017)
<code>bs = "fs"</code>	Factor-smooth basis: one shared smoothing parameter, flag-specific curves (Pedersen <i>et al.</i> , 2019)
<code>bs = "tp"</code>	Thin-plate regression spline basis (spatial smooth)
<code>bs = "cr"</code>	Cubic regression spline basis (temporal smooth)
<code>bs = "cc"</code>	Cyclic cubic spline basis (seasonal smooth)
<code>bs = "re"</code>	Random-effect basis (flag random intercept)
<code>gaussian()</code>	Gaussian family fitted to <code>log(mean_wt)</code> (response logged prior to fitting)
<code>gratia::appraise</code>	Visual model diagnostics (Simpson, 2024)
<code>mgcv::k.check</code>	Basis-dimension adequacy check

A.5 Region-wide analysis: exploration of all data

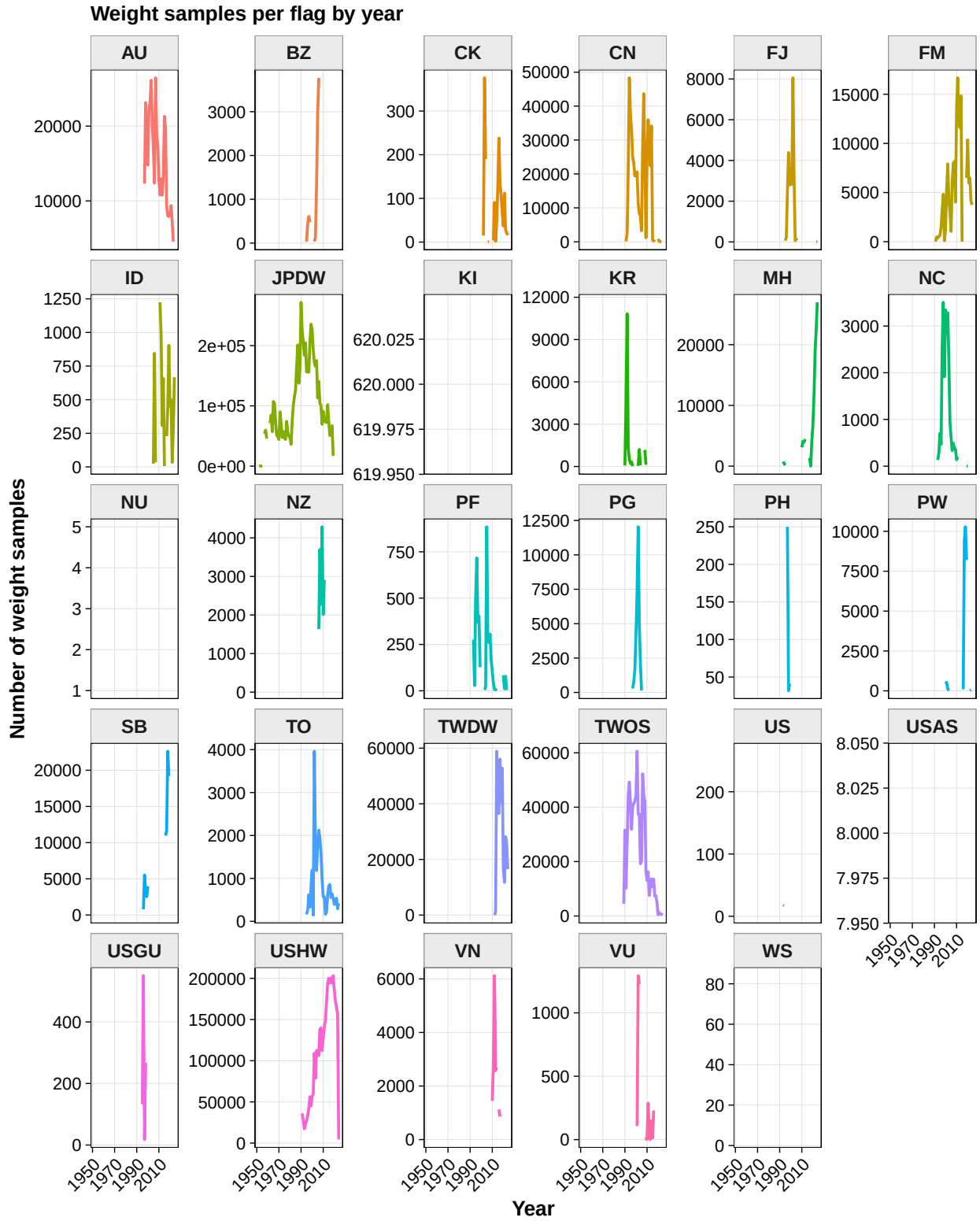


Figure 18: Number of weight samples by year and flag across the full source period (all years present in the raw data). Note the y-axis range changes by flag.

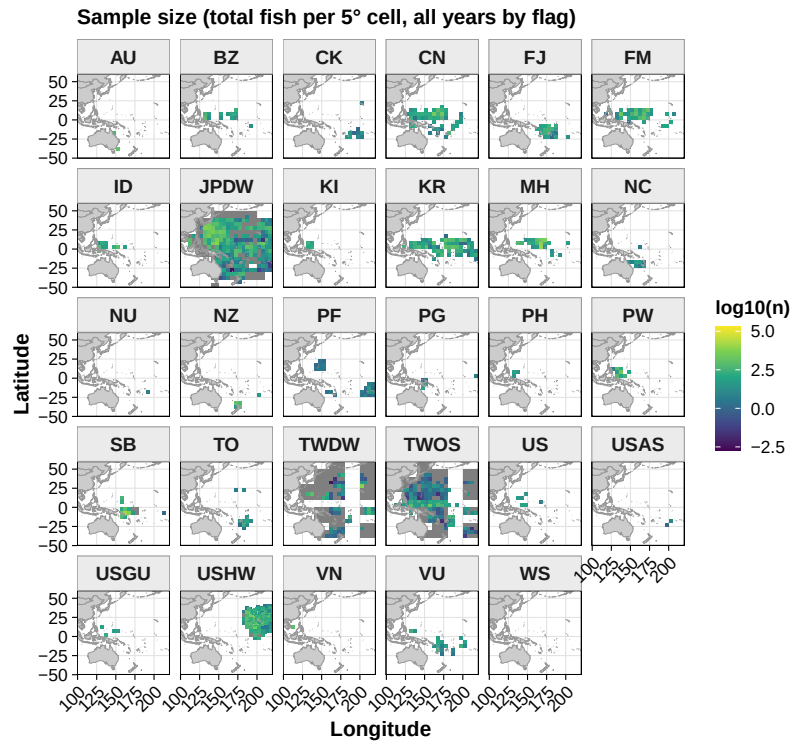


Figure 19: Sample size summed across all years in the source data, by flag.

A.6 Region-wide analysis: model diagnostics

The residual diagnostics for `m_temp_flag` are acceptable. Table 10 reports the overall fit statistics and Table 11 the approximate significance of each smooth term; Table 12 reports basis-dimension adequacy, and Table 13 the pairwise concurvity between model terms. The **appraise** diagnostic plots follow (Figure 20): the residuals-vs-linear-predictor panel shows no systematic fan pattern, indicating the log transform has resolved the heteroscedasticity seen in the untransformed model, while the QQ plot shows some deviation in the tails, driven by a number of outlying cell-quarter combinations (over-prediction of mean weight when small, under-prediction when large). Concurvity — the smooth analogue of multicollinearity — has values near 1 where terms are near-completely confounded.

Table 10: Overall fit statistics for the region-wide model (`m_temp_flag`).

Model	Family	Link	N (rows)	Dev. expl. (%)	Adj. R ²	Scale est.
<code>m_temp_flag</code>	gaussian	identity	8455	74	0.737	15.5

Table 11: Approximate significance of the smooth terms in `m_temp_flag` (`mgcv` summary). edf = effective degrees of freedom.

Smooth	edf	Ref.df	F	p-value
<code>s(yy_qtr)</code>	28.20	28.9	30.50	0.0000
<code>s(qtr)</code>	1.98	2.0	120.00	0.0000
<code>s(lat,lon)</code>	45.30	47.3	338.00	0.0000
<code>s(flag.fleet)</code>	5.40	7.0	16.60	0.0000
<code>s(yy_qtr,flag.fleet)</code>	29.00	113.0	2.73	0.0259

Table 12: Basis-dimension adequacy check (`mgcv::k.check`) for `m_temp_flag`. A low p-value suggests the basis dimension `k` may be too small.

Smooth	k'	edf	k-index	p-value
<code>s(yy_qtr)</code>	29	28.20	0.971	0.0125
<code>s(qtr)</code>	2	1.98	1.000	0.6780
<code>s(lat,lon)</code>	49	45.30	0.932	0.0000
<code>s(flag.fleet)</code>	8	5.40	NA	NA
<code>s(yy_qtr,flag.fleet)</code>	120	29.00	0.971	0.0300

Table 13: Pairwise concurvity (worst-case estimate) between the smooth terms of `m_temp_flag`. Values near 1 indicate near-complete confounding.

Term	para	<code>s(yy_qtr)</code>	<code>s(qtr)</code>	<code>s(lat,lon)</code>	<code>s(flag.fleet)</code>	<code>s(yy_qtr,flag.fleet)</code>
para	1.000	0.002	0.003	0.048	0.468	0.045
<code>s(yy_qtr)</code>	0.160	1.000	0.003	0.039	0.165	0.541
<code>s(qtr)</code>	0.007	0.000	1.000	0.004	0.006	0.001
<code>s(lat,lon)</code>	0.975	0.019	0.032	1.000	0.771	0.129
<code>s(flag.fleet)</code>	1.000	0.015	0.005	0.383	1.000	0.134
<code>s(yy_qtr,flag.fleet)</code>	1.000	0.618	0.026	0.406	1.000	1.000

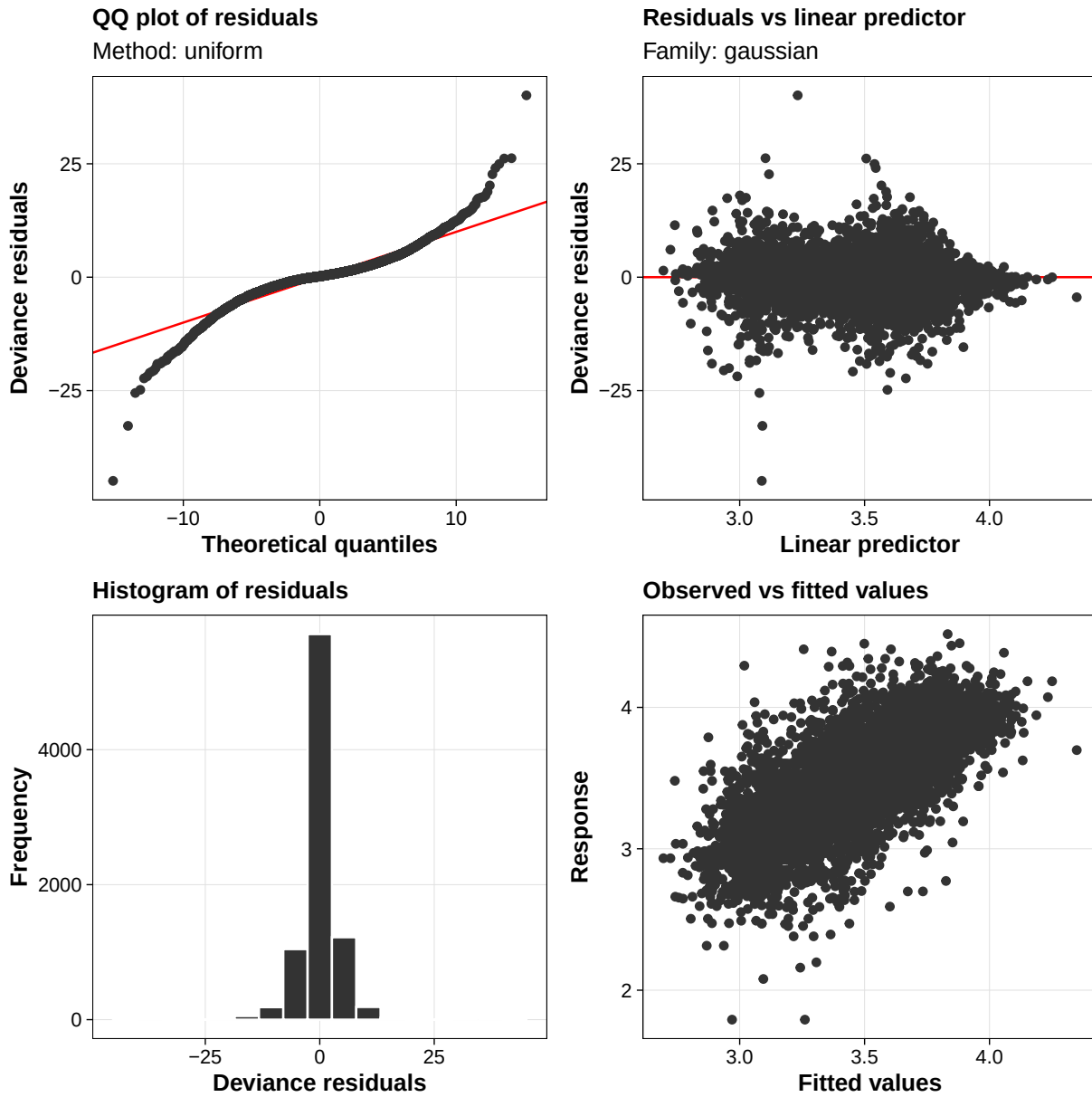


Figure 20: Model appraisal diagnostics for `m_temp_flag` (gratia::appraise).

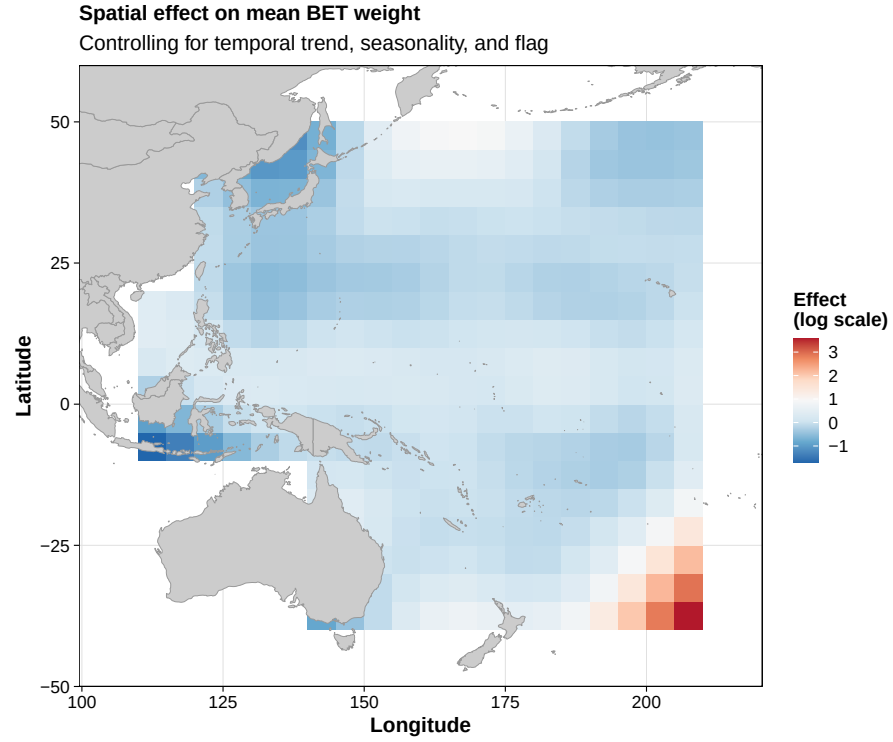


Figure 21: Estimated spatial effect on mean BET weight (log scale), controlling for temporal trend, seasonality, and flag.

Fitted-versus-Pearson-residual plots, disaggregated by flag (Figure 22) and by year (Figure 23), show no strong residual structure in either dimension.

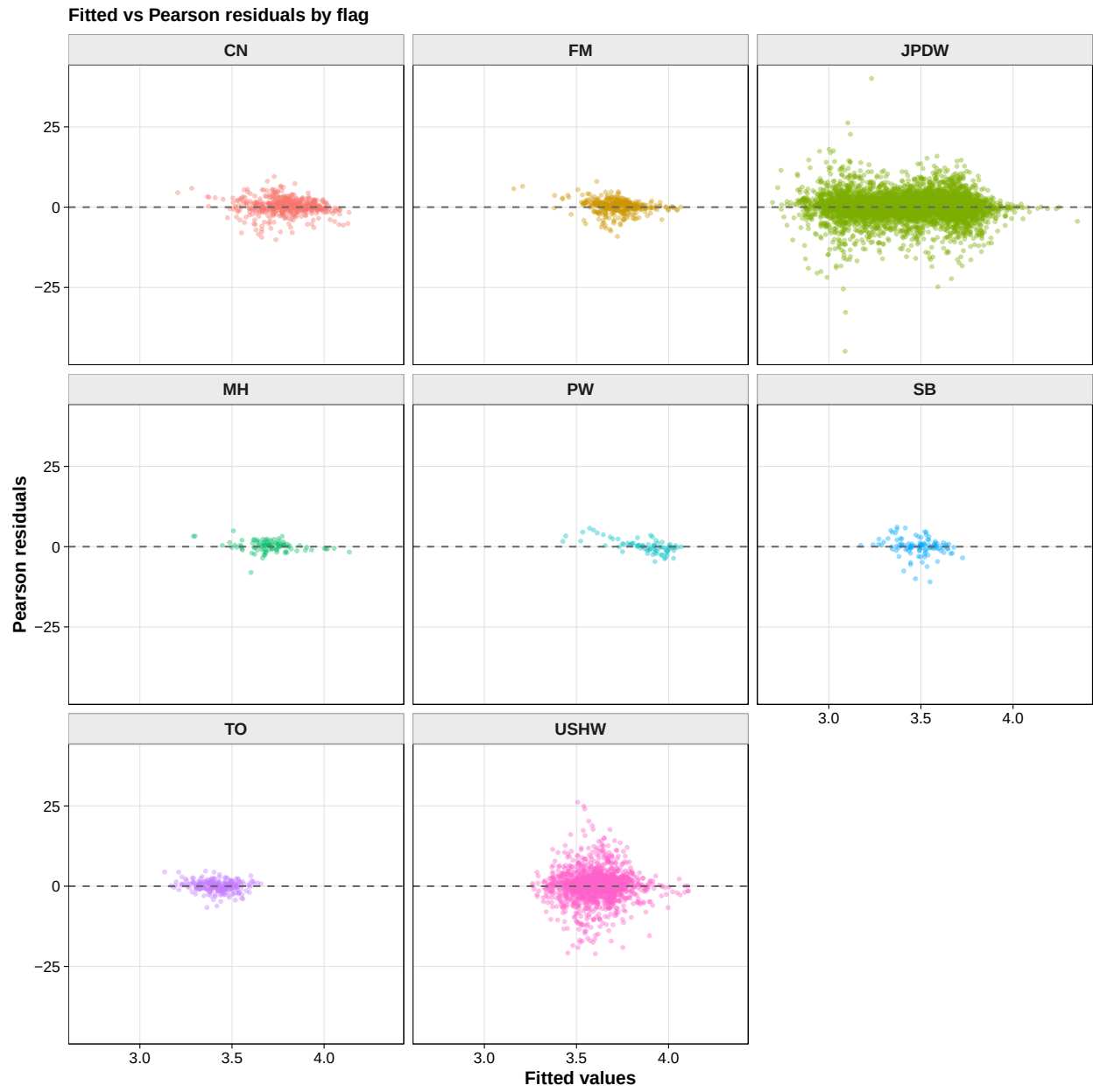


Figure 22: Fitted vs Pearson residuals by flag.

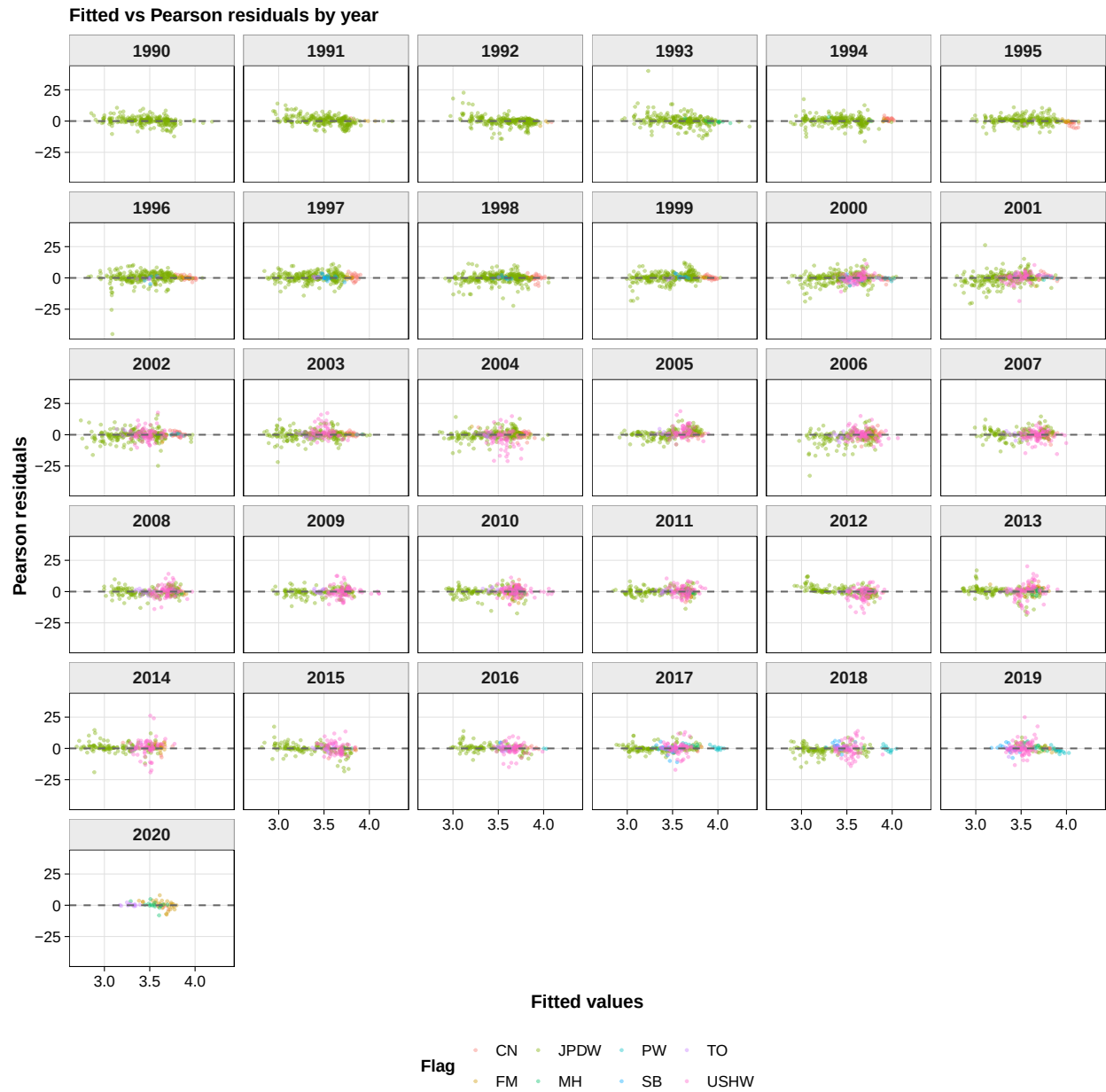


Figure 23: Fitted vs Pearson residuals by year, coloured by flag.

A.7 Region-wide analysis: temporal autocorrelation (AR(1) robustness check)

The pooled lag-1 residual autocorrelation (computed in the methods) was $\hat{\rho} = 0.31$. Lag-1 ACF on residuals within the three densest JPDW cells is shown in Table 14 for context.

Table 14: Lag-1 residual autocorrelation within the three densest JPDW cells (region-wide model). Cells are identified by the south-west corner of each 5° bin; n is the number of cell-quarter observations.

Latitude	Longitude	n	Lag-1 ACF
5	140	110	0.283
5	145	119	0.223
10	145	113	0.250

There is some residual temporal autocorrelation, meaning standard errors from `m_temp_flag` are likely somewhat underestimated. A refitted model with an AR(1) error structure assesses the practical importance of this (Figure 24). Point estimates are virtually identical across the two models and the confidence intervals are broadly comparable through the data-dense period (mid-1990s to 2020). The substantive findings are robust to the AR(1) correction; `m_temp_flag` was therefore used for all interpretation, with the caveat that its confidence intervals may be modestly optimistic.

Global temporal trend: AR(1) vs no correction

Shading = 95% CI

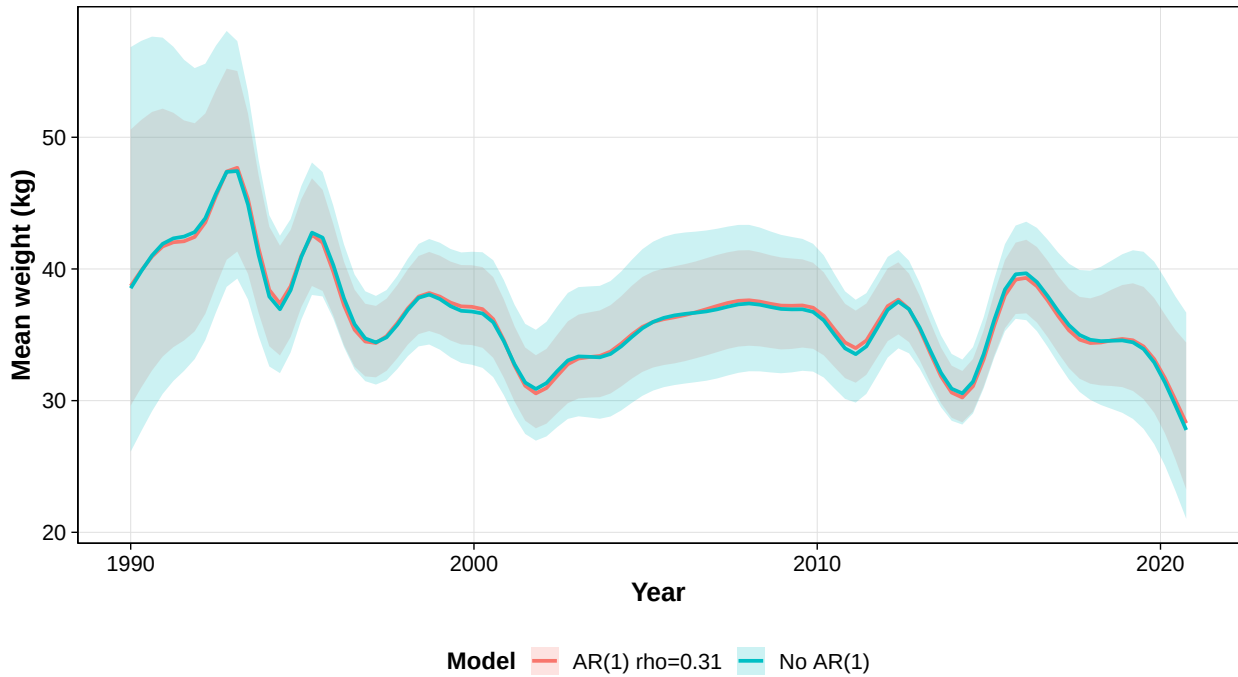


Figure 24: Global temporal trend with and without the AR(1) correction. Shading = 95% CI.

A.8 Region-wide analysis: testing whether the recent pooled decline reflects the changing flag mix

The pooled region-wide trend (Figure 5) is a sample-size-weighted average across all eight flags. Because the flags differ both in the mean size of fish they catch and in *when* they fish, a shift over time in which flags dominate the data can drag the pooled average down even if every individual fleet is flat — a composition effect, distinct from any real size-structure change. This is the concern here, because the recent part of the pooled decline coexists with two long-running fleets, JPDW and USHW, that are individually declining only marginally after 2000 (Figure 17).

To separate the two, the region-wide base structure (global temporal, seasonal, and spatial smooths plus a flag random intercept) was refitted on the JPDW + USHW subset alone, and its global trend compared with the full eight-flag trend. If the recent decline were purely compositional, dropping the other six flags would flatten it; if it reflects a shared decline, it should persist.

Over 2000–2018 — stopping before the sparse final two years, where the global smooth is least constrained — the eight-flag trend falls by about 2.1 kg while the two-fleet refit falls by about 1.5 kg (the 2000–2020 window gives 5.4 kg and 2.7 kg). The two-fleet trend therefore retains a sizeable part of the decline: roughly 29% of the 2000–2018 pooled decline (and 50% of the 2000–2020 decline) reflects the changing flag mix, with the remainder present in both of the two main fleets. The recent pooled decline is thus only partly compositional — not a pure averaging artefact — and this conclusion does not depend on the final two years.

A.9 Model diagnostics for the single-fleet historical GAMs

This section presents the diagnostic summaries for the two single-fleet historical GAMs (Equation 3). Both share the same structure: `bam()` with a Gaussian family on the log scale, `weights = total_n`, a temporal smooth `s(yy_qtr)`, a cyclic seasonal smooth `s(qtr)`, and a thin-plate spatial smooth `s(lon, lat)`. Table 15 provides summary statistics, Table 16 reports basis-dimension adequacy, Figure 25 compares residual autocorrelation across fleets, and the `appraise` diagnostic plots for each model follow (Figure 26, Figure 27).

A.9.1 Summary statistics

Table 15: Summary statistics for the two single-fleet historical GAMs.

Model	Fleet	Period	N (rows)	Dev. expl. (%)	Adj. R ²
jpdw_mod	JPDW	1957-2018	19636	71.0	0.707
ushw_mod	USHW	2000-2019	1846	60.1	0.589

A.9.2 Basis dimension checks (k.check)

Table 16: Basis dimension adequacy checks (k.check) for each model. A low p-value suggests the basis dimension k may be too small.

Model	Smooth	k'	edf	k-index	p-value
JPDW	s(yy_qtr)	49	48.155	0.980	0.058
JPDW	s(qtr)	2	1.954	0.997	0.460
JPDW	s(lat,lon)	49	46.468	0.901	0.000
USHW	s(yy_qtr)	24	22.440	0.579	0.000
USHW	s(qtr)	2	1.990	0.585	0.000
USHW	s(lat,lon)	29	22.607	0.937	0.018

A.9.3 Residual autocorrelation

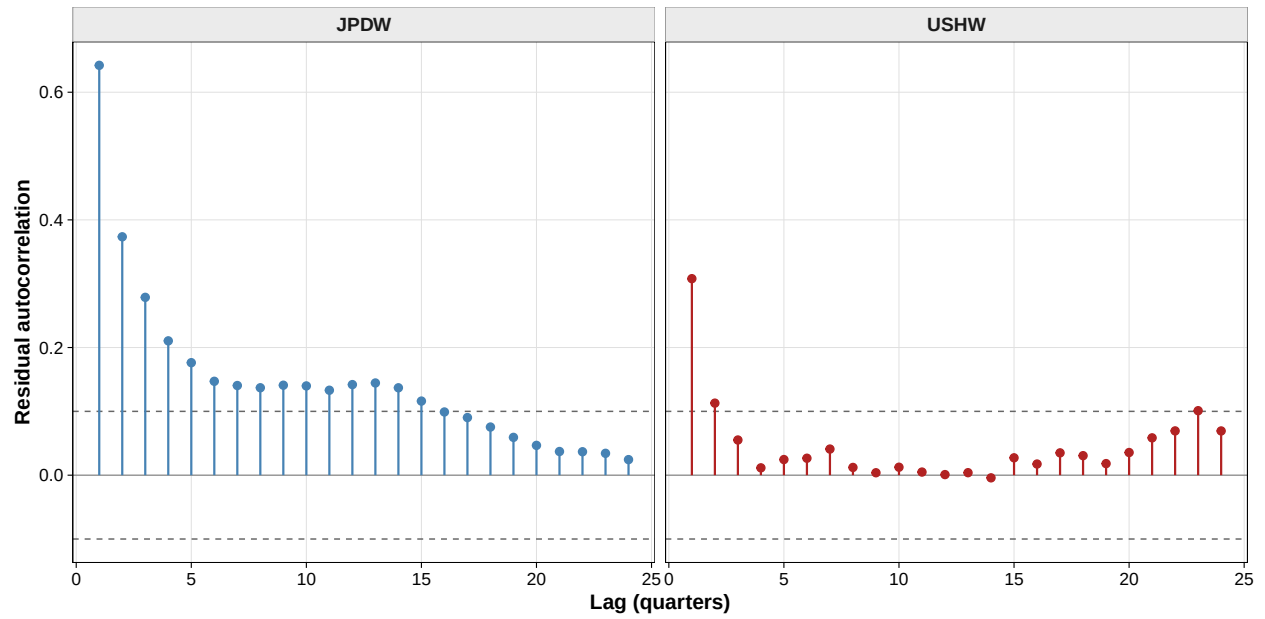


Figure 25: Residual autocorrelation functions (ACF) for the two single-fleet historical GAMs. Dashed lines indicate the ± 0.1 threshold for reference. Both models show some residual autocorrelation at short lags, with JPDW displaying the most persistent multi-scale pattern.

A.9.4 Model appraise diagnostics

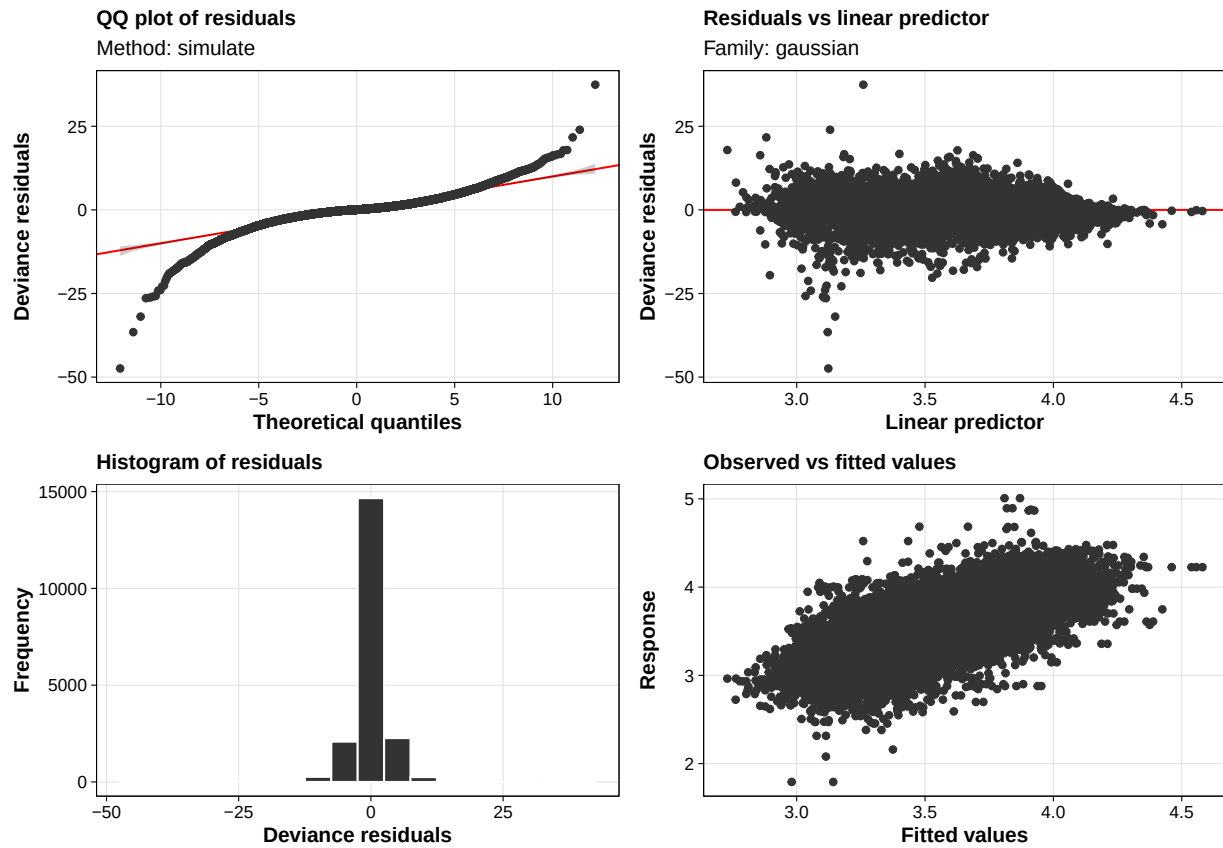


Figure 26: Diagnostic plots (`gratia::appraise`) for `jpdw_mod`.

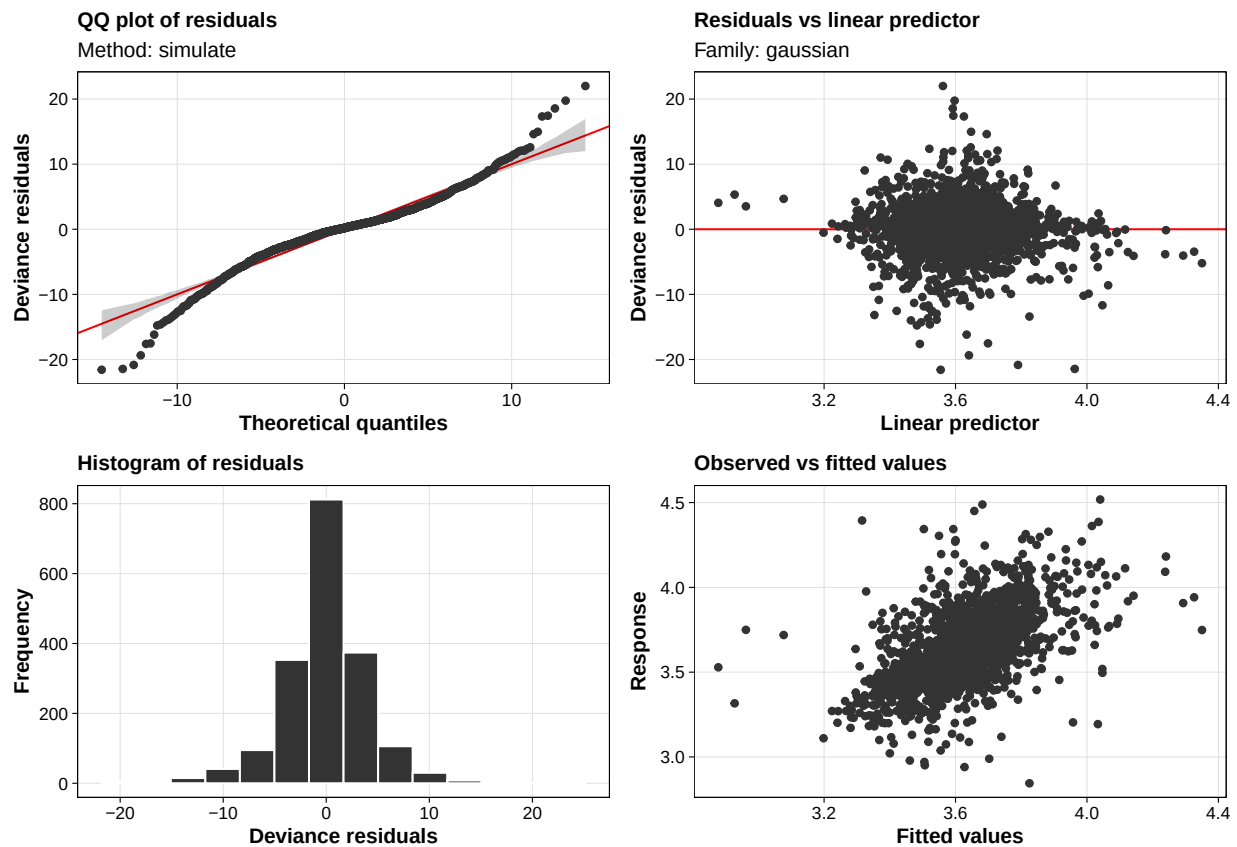


Figure 27: Diagnostic plots (`gratia::appraise`) for `ushw_mod`.

A.10 JPDW sensitivity analyses

The two headline JPDW results — the long-term marginal decline (mid-1960s to 2015) and the within-area drop from the fixed-area decomposition — are checked here against the main modelling choices: the 200 kg weight cap, the error family (Gaussian on the log scale versus Gamma with a log link), and the temporal basis dimension K_T (Table 17). The marginal decline is recomputed from each refit; the within-area column is the *observed* sample-size-weighted mean-weight decline over the fixed 2010–2018 footprint (model-free, so unaffected by the error family or K_T by construction — its constancy across those rows confirms the within-area signal is present in the raw data rather than induced by the smoother, and it responds only to the weight cap). Note this observed within-area figure is the descriptive analogue of the model-based ~9 kg decomposition value in the main text (which uses predicted cell medians), not the identical quantity.

Table 17: Sensitivity of the JPDW headline marginal decline (predicted mean weight at 1965 and 2015, and their difference) and of the observed within-area decline over the fixed 2010–2018 footprint, to the weight cap, error family, and temporal basis dimension. One setting is varied at a time from the base specification.

Variant	1965 (kg)	2015 (kg)	Decline 1965-2015 (kg)	Obs. within-area (kg)
Base (cap 200 kg, Gaussian-log, k=50)	46.4	35.0	11.4	11.5
No weight cap	46.4	35.0	11.4	11.5
Gamma (log link)	45.7	34.5	11.2	11.5
Lower basis (k=30)	45.8	32.9	13.0	11.5
Higher basis (k=80)	46.8	35.3	11.5	11.5

If the decline and the within-area drop are stable in sign and broadly stable in magnitude across these rows, the headline conclusions do not hinge on the specific modelling choices.

A.11 JPDW spatial coverage relative to the study area

The single-fleet JPDW model is fitted to all 5° cells the fleet sampled, including cells outside the five WCPFC assessment regions; the by-decade spatial maps (Figure 9), by contrast, are restricted to the study area. Because JPDW’s historical footprint extended well beyond the assessment regions, it is useful to quantify how much of the fleet’s weight sampling fell outside the study area, and how that share changed over time (Table 18).

Table 18: JPDW BET weight sampling by decade window, split by whether the 5-degree cell falls inside the five assessment regions (the study area used for the spatial maps) or outside them. Percentages are of the window’s total sampled fish. The share sampled outside the assessment regions is largest in the early windows and declines toward the 2010s, as the fleet’s footprint contracts into the WCPO.

Window	Total fish	Inside study area (%)	Outside study area (%)
1957-1969	756,000	84.7	15.3
1970-1979	580,933	79.9	20.1
1980-1989	1,167,310	76.9	23.1
1990-1999	1,972,709	93.4	6.6
2000-2009	1,458,911	100.0	0.0
2010-2018	659,589	100.0	0.0

In the earliest window (1957–1969), about 15.3% of JPDW’s sampled fish came from cells outside the five assessment regions, falling to about 0% by 2010–2018. The early out-of-region sampling reflects the fleet’s broad historical footprint (much of it east of the assessment area); the contraction toward almost exclusively

in-region sampling is the same operational-footprint contraction shown in Figure 9. The single-fleet GAM uses the full footprint, so the early trend is informed by these out-of-region cells; restricting the maps (and the fixed-area decomposition, which is anchored on the 2010–2018 in-region footprint) to the study area is therefore the conservative choice for interpretation. Temporally, the single-fleet window (1957–2018) also bounds the analysis. In the source database, JPDW BET weight sampling spans 1952–2019; the window excludes the 962 fish recorded before 1957 (about 0% of all JPDW BET weight records) and the 16,962 recorded after 2018 (about 0.3%), the latter reflecting the sharp post-2018 drop in JPDW contributions noted in the Methods. Together with the spatial restriction above, the 1957–2018 window defines the data actually entering the single-fleet analysis; the excluded records sit at the sparse temporal margins of the JPDW series and are unlikely to alter the long-term trend.

The pre-1957 JPDW BET records can be inspected directly. JPDW sampling is visible in the database well before the analysis start (Figure 18); Table 19 summarises the JPDW BET weight records before 1957, by year.

Table 19: JPDW BET weight records in the source database before 1957 — the years that fall outside the 1957–2018 single-fleet window and are therefore excluded from the JPDW analysis. Sampled fish per year (the totals are small: under 0.1% of all JPDW BET weight records combined).

Year	Sampled fish
1952	380
1953	353
1954	7
1955	221

These years are present in the raw data but sit outside the analysis window, so they do not contribute to the fitted JPDW trend, milestones, spatial maps, or decomposition. If Table 19 is empty, JPDW BET weight sampling in the database begins in 1957 and nothing is excluded at the early boundary.