

Supplemental Information

S1. Supplemental Methods

S1.1. Mixed Effects Models to Explain Range_{CT}

We used mixed effect generalized linear models (GLMMs) to explore which of our 31 explanatory variables best explain observed trends in range_{CT}. In these models, species identity is included as a random effect. By treating species as a random effect, we assume that each species has unknown facets (e.g., biological, physiological, ecological) that might impact its particular range_{CT} in a manner not linked to our explanatory variables. In other words, we recognize that some of the variability observed in our range_{CT} values is likely due to the fact that we include multiple data points per species – and we assume that a given species is more likely to be similar in range_{CT} to itself than it is to another species.

Once we established a mixed effects modeling framework using species identity as a random effect, we tested two plausible model formats. One is a standard linear model, in which we assume that the response variable is a linear function of the explanatory variable and that the residuals follow a Gaussian distribution.

For species i and range_{CT} j , we use linear regression to model the expected value ($E[Y_{ij}]$) as:

$$\begin{aligned} E[Y_{ij}] &= \mu_{ij} = \beta_0 + \beta_1 X_{ij} + u_i \\ \text{var}(Y_{ij}) &\sim N(0, \sigma^2) \\ u_i &\sim N(0, \sigma_u^2) \end{aligned}$$

The second model structure we tested uses a Tweedie distribution with a log link function. The Tweedie distribution is used for data that are continuous and positive but also have exact zeros (unlike the gamma distribution, which disallows zero values). Here, we model the log of the expected value as:

$$\begin{aligned} \log(E[Y_{ij}]) &= \log(\mu_{ij}) = X_{ij}^T b + u_i \\ \text{var}(Y_{ij}) &= \phi \mu_{ij}^2 \\ u_i &\sim N(0, \sigma_u^2) \end{aligned}$$

Parameters b and ϕ are identified via maximum likelihood estimation.

We determined that the Tweedie distribution fit our data better than the Gaussian distribution in a number of ways. First, visual examination of expected versus observed values, as well as the

residuals, showed better alignment with a Tweedie distribution (Figure S5). Second, we compared average AIC and log likelihood values between models fit with Gaussian and Tweedie distributions. AIC and log likelihood were both lower for the Tweedie distribution, with significance determined by Wilcoxon rank-sum tests ($p < 0.001$ in both cases).

We undertook a similar mixed effects modeling approach to understanding the regionally-specific relationships between mean_{CT} and range_{CT} . Again, we treated species as a random effect and tested whether adding mean_{CT} as a fixed effect improved model performance. Model improvements were examined via a likelihood ratio test, which is used to test nested models (i.e., when one model is a subset of the other, as is the case here). We used the `lrtest` function in the `lmerTest` R package.

S1.2. Understanding Phenological Differences Between Provinces

We wanted to understand how larval fish phenology varied across the provinces in our dataset. We used range_{CT} to categorize the larval fish populations based on their phenological variability, hypothesizing that this grouping could help illuminate some general trends. The four categories we used were (1) highest range_{CT} , (2) high range_{CT} , (3) low range_{CT} , and (4) lowest range_{CT} , with each category representing 25% of the dataset. Given that our data were now categorical, we built a contingency table to examine a possible relationship between our range_{CT} categories and the provinces. We considered using a chi-squared test to judge whether our observations differed from a null expectation, but determined that a Fisher's exact test was more appropriate for our relatively small sample size per province.

We also wanted to examine whether there were any differences in mean_{CT} between provinces. Here, we were interested in mean_{CT} as a continuous variable. We considered testing for differences in mean_{CT} between groups (i.e. provinces) using a standard ANOVA. However, we determined through a Shapiro-Wilk test that the residuals in our data were not normally distributed ($p < 0.05$). Therefore, we used the non-parametric Kruskal-Wallis test, which functions similarly to an ANOVA but does not assume normality.

S2. Supplemental Tables and Figures

Table S1. Sensitivity tests of binning window. To test the impact of changing our binning choice on results, we moved the start year up incrementally, generating a total of 5 possible windows. We then reran our bootstrapping routine for each window (see Methods for details) and recalculated range_{CT} based on these new windows. Kruskal-Wallis tests were performed to test for differences in range_{CT} between binning windows.

Region	Type	n	statistic	df	p
GB	BANK	156	0.461	5	0.993
GB	SB	48	0.217	5	0.999
GOM	BASIN	18	5.466	5	0.362
GOM	IS	72	1.708	5	0.888
GOM	MS	84	0.549	5	0.99
MAB	IS	60	0.702	5	0.983
MAB	MS	144	3.068	5	0.689
MAB	SB	72	0.408	5	0.995
SNE	C	66	0.074	5	1
SNE	IS	162	0.346	5	0.997
SNE	MS	174	0.587	5	0.989
SNE	SB	66	1.215	5	0.943

Table S2. Populations excluded from analysis because their larvae occur in winter. Species abbreviations as in Table 1, region and province abbreviations as in Figure 1. Six species were excluded entirely from our analyses due to winter occurrence: Atlantic menhaden (*Brevoortia tyrannus*; bretyr), Atlantic herring (*Clupea harengus*; cluhar), Atlantic cod (*Gadus morhua*; gadmor), spot (*Leiostomus xanthurus*; leixan), Atlantic croaker (*Micropogonias undulatus*; micund), and summer flounder (*Paralichthys dentatus*; parden).

Abbreviation	Region_Province
ammssp	MAB_MS
benspp	GB_SB, SNE_SB
botssp	GB_SB, MAB_MS, MAB_SB

bretyr	GB_BANK, GB_SB, MAB_IS, MAB_MS, MAB_SB, SNE_C, SNE_IS, SNE_MS
centstr	MAB_MS
cermad	MAB_MS, MAB_SB
citarc	MAB_MS, SNE_MS
cluhar	GB_BANK, GB_SB, GOM_BASIN, GOM_IS, GOM_MS, MAB_IS, MAB_MS, MAB_SB, SNE_C, SNE_IS, SNE_MS
cycspp	GB_BANK, GOM_MS, MAB_SB
diaspp	GB_SB, MAB_MS
enccim	MAB_IS
etrsp	MAB_MS
gadmor	GB_BANK, GOM_IS, GOM_MS, SNE_C, SNE_IS, SNE_MS, SNE_SB
leixan	MAB_IS, MAB_MS, MAB_SB
melaeg	GOM_BASIN, MAB_MS
meralb	SNE_MS
merbil	MAB_MS, MAB_SB, SNE_IS, SNE_MS
micund	MAB_IS, MAB_MS
parden	MAB_IS, MAB_MS, MAB_SB, SNE_IS, SNE_MS
pepspp	MAB_MS
polvir	GB_BANK, GB_SB, GOM_IS, GOM_MS, SNE_C, SNE_IS, SNE_SB
scoaqu	MAB_IS, MAB_MS
sebspp	GB_BANK
urospp	GB_BANK, GB_SB, MAB_IS, MAB_MS, MAB_SB

Table S3. List of all variables tested with generalized linear mixed models. Listed in order of descending AIC. Abbreviations: bt = bottom temperature; sst = sea surface temperature; se = standard error; ssal = surface salinity; bsal = bottom salinity; smax = salinity maximum intrusions.

Variable	Log likelihood	AIC	Slope	Sign of slope
Smax CT	-91.37	190.74	14.52	positive
Smax SE	-93.86	195.72	0.74	positive
Smax Mean	-95.05	198.11	0.71	positive
Ssal SE	-112.17	232.34	14.43	positive
Bsal SE	-114.33	236.66	15.17	positive
Mean CT	-114.83	237.66	0.28	positive
SST SE	-115.09	238.18	2.60	positive
Region	-114.53	241.05	0.56	NA
Bsal CT SE	-116.69	241.39	-1.11	negative
Ssal CT SE	-116.71	241.42	-0.97	negative
BT CT SE	-116.82	241.64	0.08	positive
SST CT SE	-116.85	241.70	0.10	positive
BT Range	-116.98	241.96	-0.04	negative
BT SE	-117.07	242.15	0.29	positive
BT CT SE	-117.54	243.09	-0.38	negative
Bsal Range	-117.89	243.78	-0.17	negative
SST CT	-117.92	243.84	-0.44	negative
Ssal CT	-118.21	244.42	0.27	positive
Bsal CT	-118.22	244.45	0.25	positive
Habitat (Walsh et al. 2015)	-118.34	244.67	0.28	positive

Variable	Log likelihood	AIC	Slope	Sign of slope
Season (Walsh et al. 2015)	-117.51	245.03	0.46	NA
Bsal Mean	-119.48	246.97	0.08	positive
Ssal Range	-119.78	247.57	0.06	positive
BT Mean	-119.91	247.81	-0.05	negative
Ssal Mean	-120.10	248.20	-0.01	negative
Habitat (Froese and Pauly, 2000)	-119.24	248.48	-0.05	NA
SST Mean	-120.61	249.23	-0.03	negative
Season	-112.96	249.93	0.11	NA
SST Range	-121.08	250.16	-0.02	negative
Province	-118.91	253.82	-0.01	NA

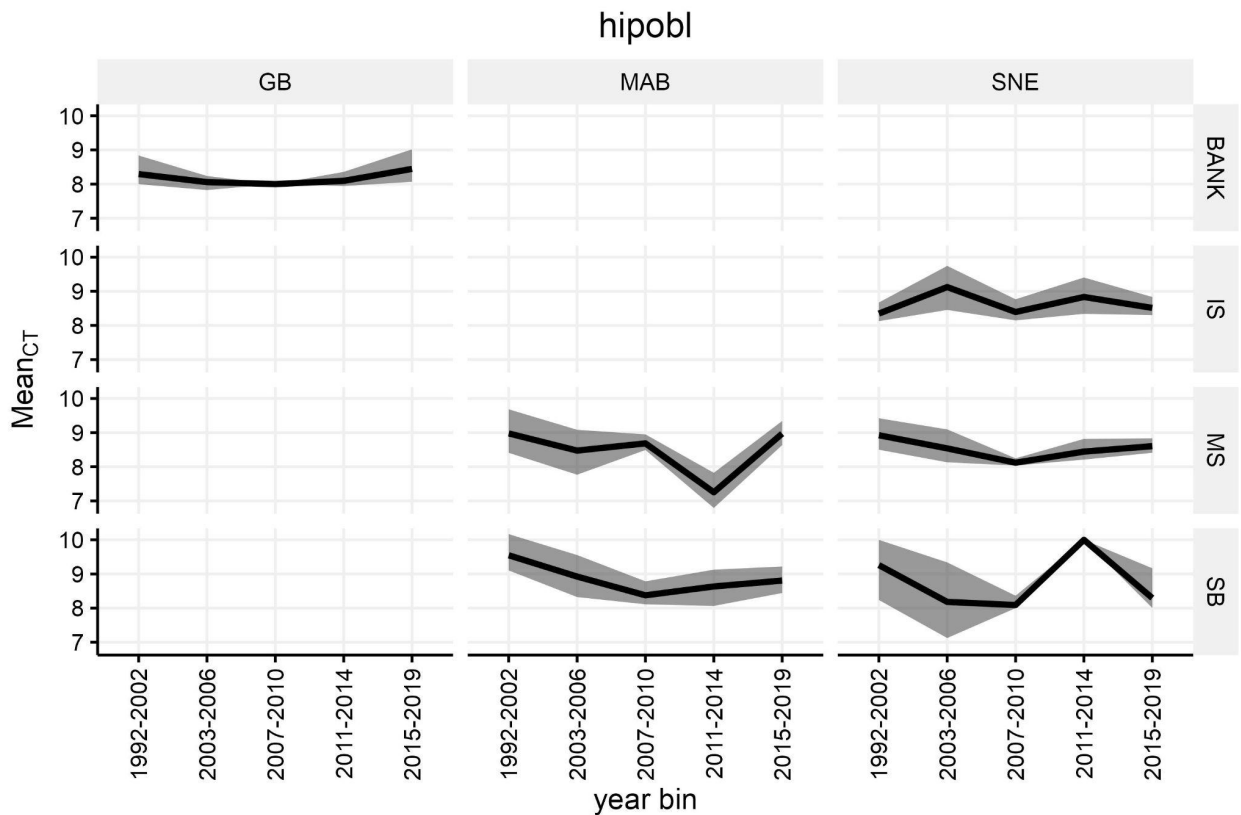
Table S4. Games Howell post-hoc test results comparing mean_{CT} across provinces.

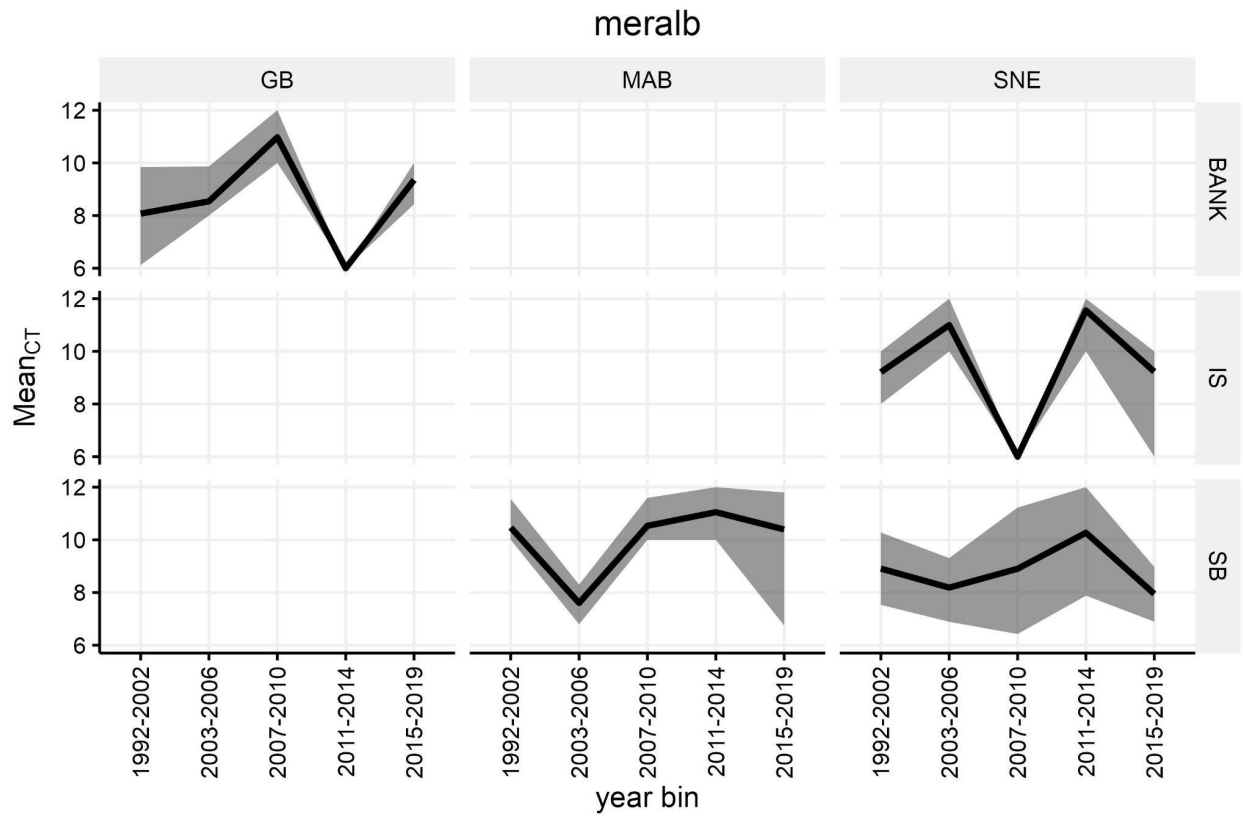
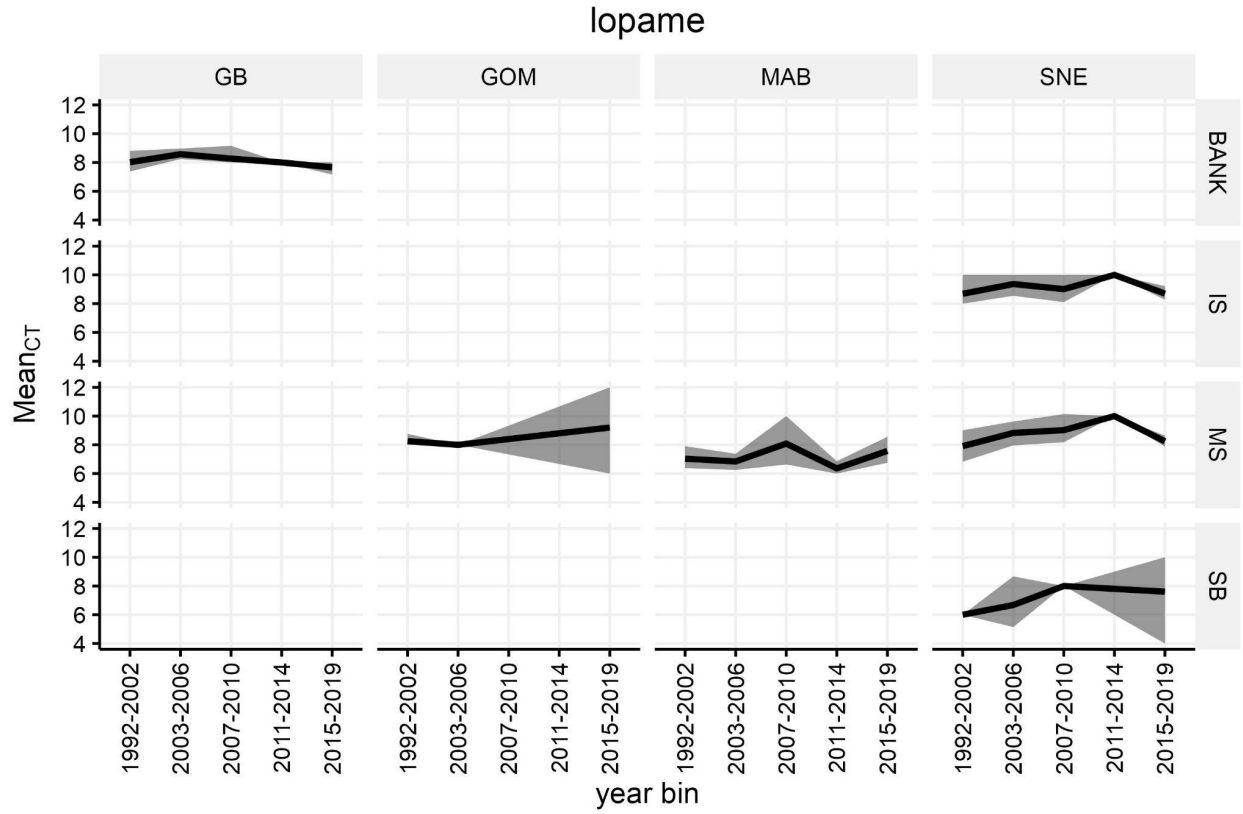
Group1	Group2	Statistic	DF	p-value
BANK	BASIN	0.42	6.90	1.00
BANK	C	1.55	23.89	0.64
BANK	IS	0.46	34.87	1.00
BANK	MS	0.57	34.05	0.99
BANK	SB	2.90	33.41	0.07
BASIN	C	1.92	7.72	0.46
BASIN	IS	0.06	3.58	1.00
BASIN	MS	0.05	3.45	1.00
BASIN	SB	2.32	3.59	0.37
C	IS	2.22	15.90	0.28
C	MS	2.34	15.48	0.24
C	SB	4.41	15.78	0.01
IS	MS	0.15	82.21	1.00
IS	SB	3.33	55.99	0.02
MS	SB	3.24	56.80	0.02

Table S5. Games Howell post-hoc test results comparing mean_{CT} across regions.

Group1	Group2	Statistic	DF	p-value
GB	GOM	0.87	38.81	0.82
GB	MAB	2.44	36.95	0.09
GB	SNE	0.14	42.42	1.00
GOM	MAB	4.75	49.14	0.00
GOM	SNE	1.40	71.53	0.50
MAB	SNE	3.23	81.15	0.01

Figure S1. Some examples of species-specific results for temporal trends of mean_{CT} (units: months). Mean_{CT} (black line) and 95% confidence interval (grey ribbon) for each year bin calculated from bootstrapped dataset. Region and province abbreviations explained in the text; species abbreviations as in Table 1.





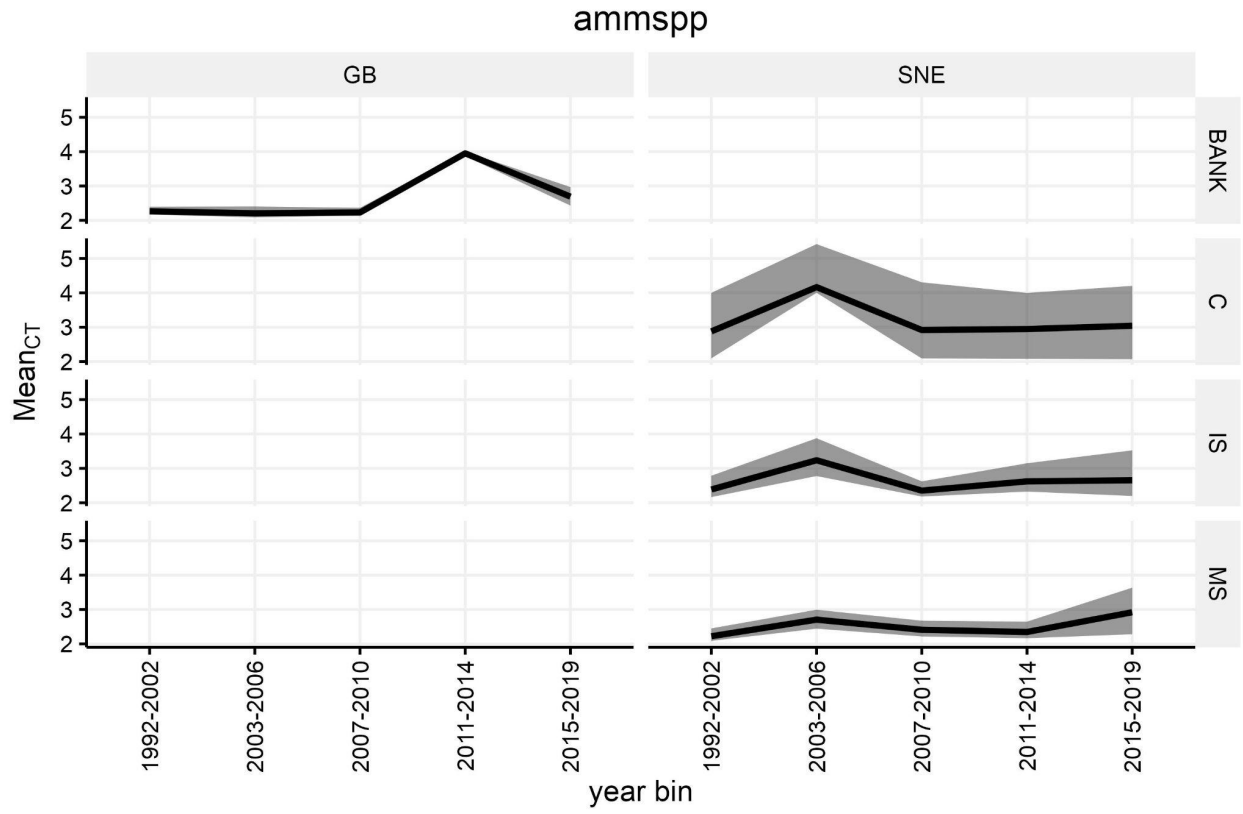


Figure S2. Temporal gaps in EcoMon data necessitate binning. Left, number of years that each region-province combination was sampled in each bimonthly season (full coverage = 21 years). Right, number of year bins in which each region-province was sampled in each bimonthly season (full coverage = 5 year bins).

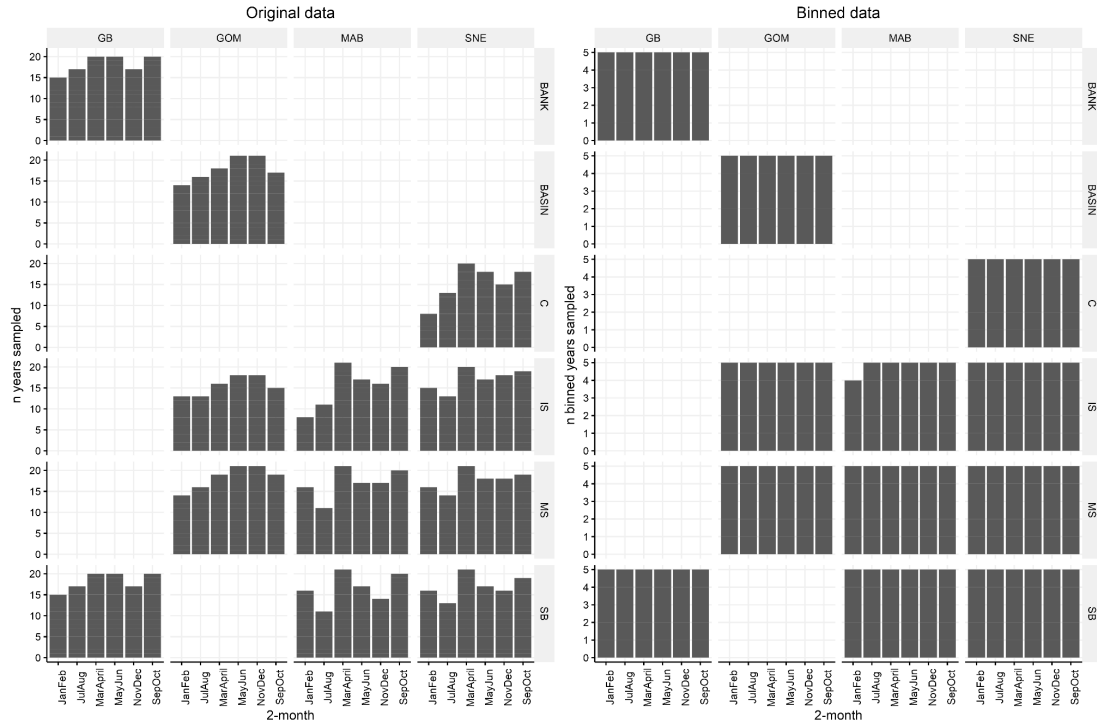


Figure S3. Comparison between advancing (trending earlier; negative; $n = 16$) and delayed (trending later, positive; $n = 6$) groups in terms of Mean_{CT} . Results do not support the hypothesis that spring-spawned larvae are advancing while fall-spawned larvae are delayed.

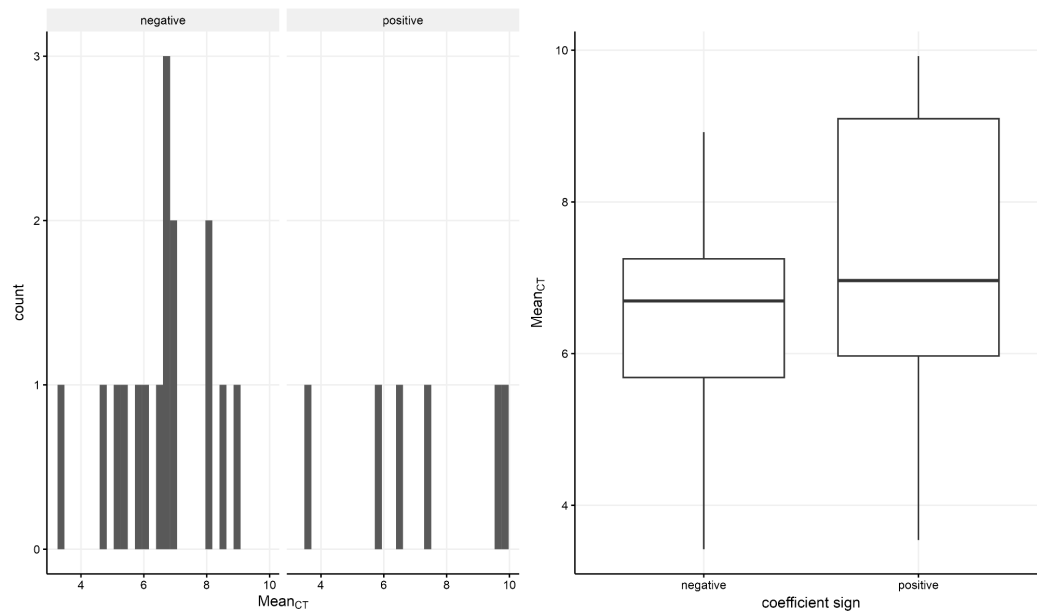


Figure S4. A) Fisher’s exact test for differences in range_{CT} among regions. Colors represent quartiles of range_{CT} (lowest, low, high, highest), and representation of each range_{CT} quartile within each region is noted in white. B) Kruskal-Wallis differences in mean_{CT} among regions. Overall Kruskal-Wallis p-value and post-hoc Games-Howell p-values noted above each significant comparison.

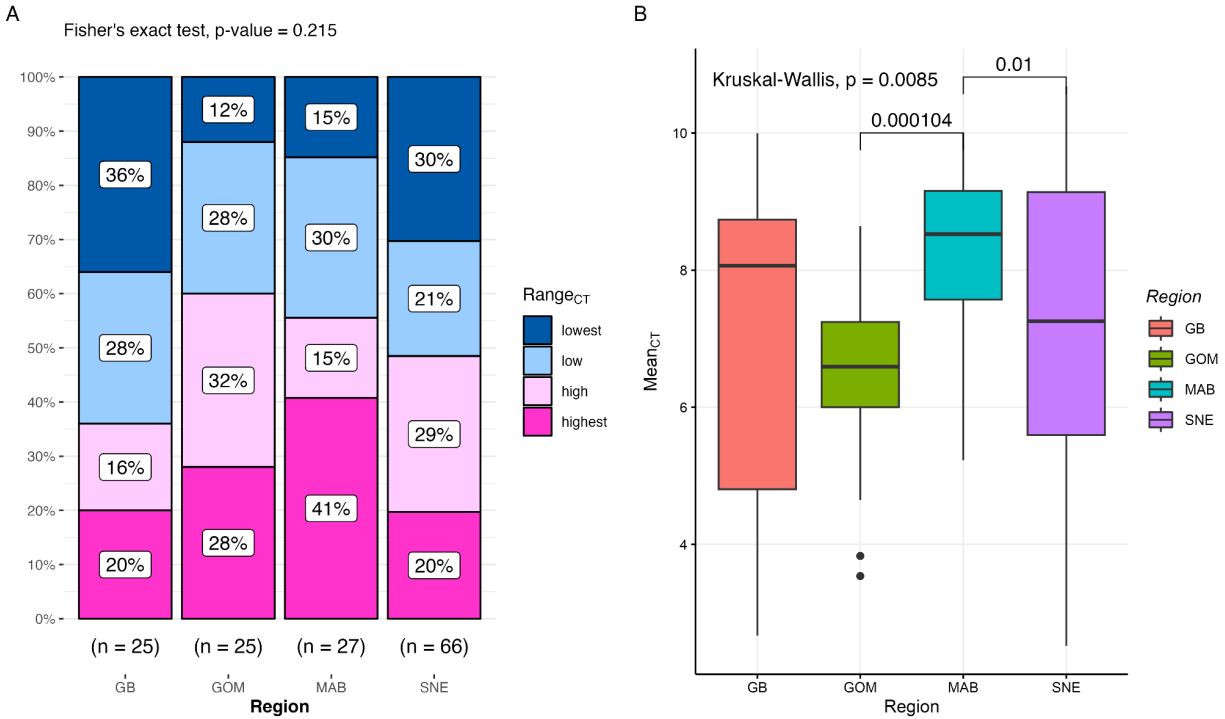


Figure S5. Tweedie distribution (A) fits data range_{CT} data better than Gaussian distribution (B), visualized using standard plots.

