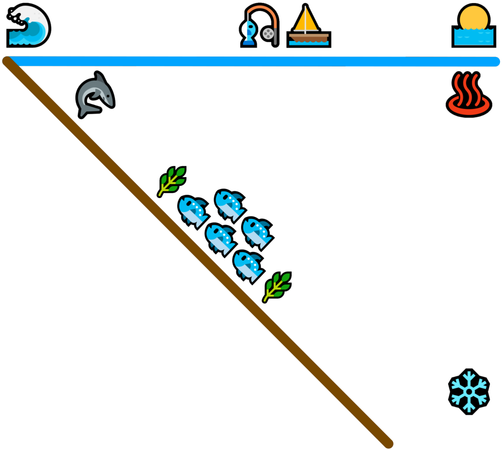
Modelling species' responses to climate change using a flexible non-linear Bayesian approach

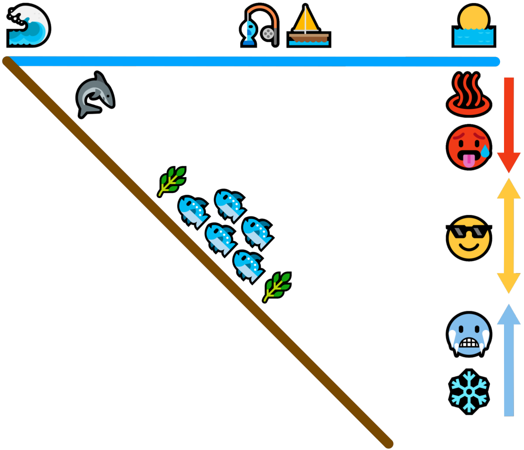
Hayden Rabel - Massey University & PRIMER-e

June 27, 2022 - ISBA World Meeting Montréal, Canada

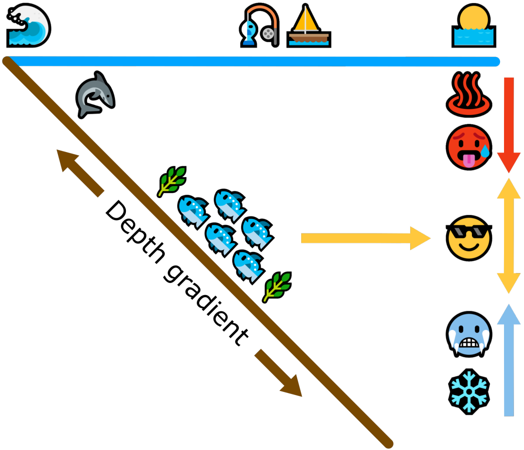
Species' responses



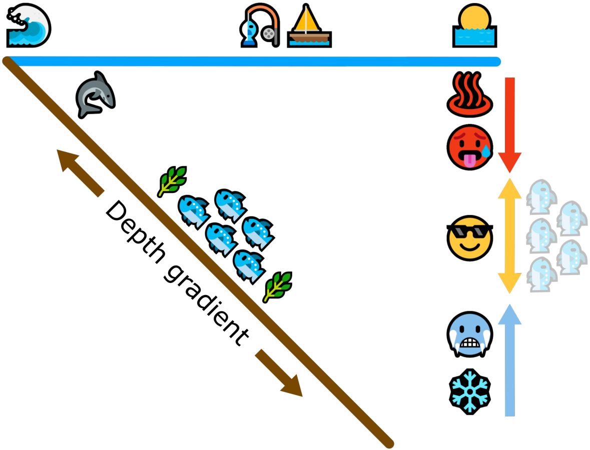
Species' responses



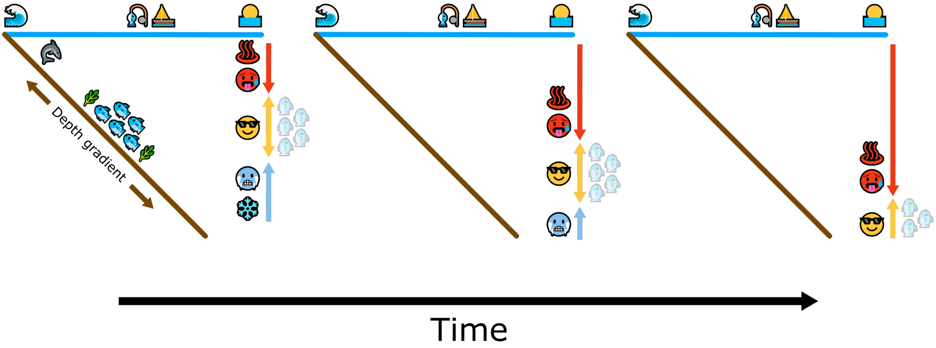
Species' responses



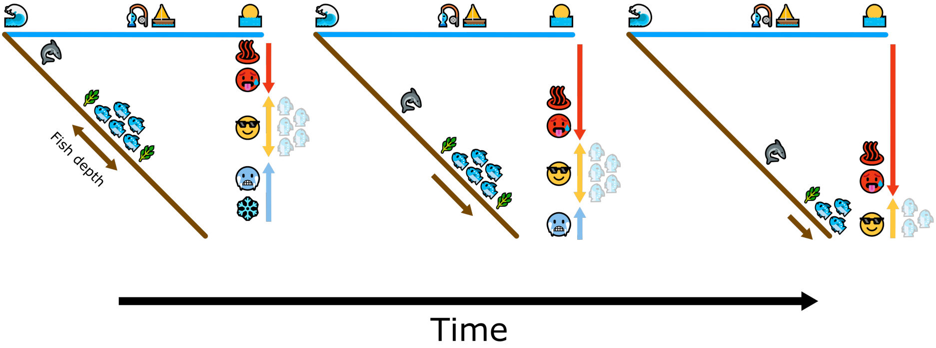
Species' responses



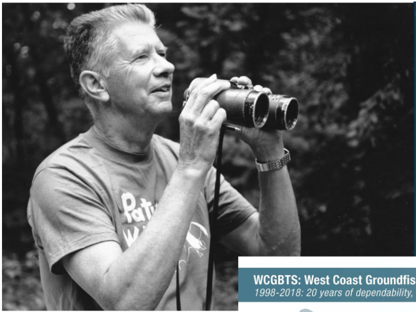
Species' responses

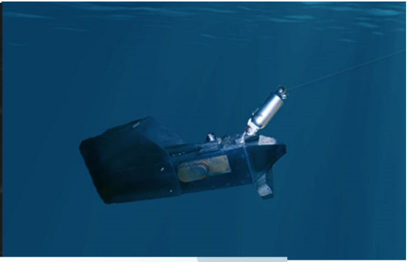


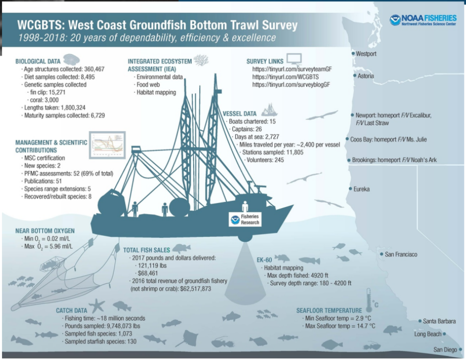
Species' responses



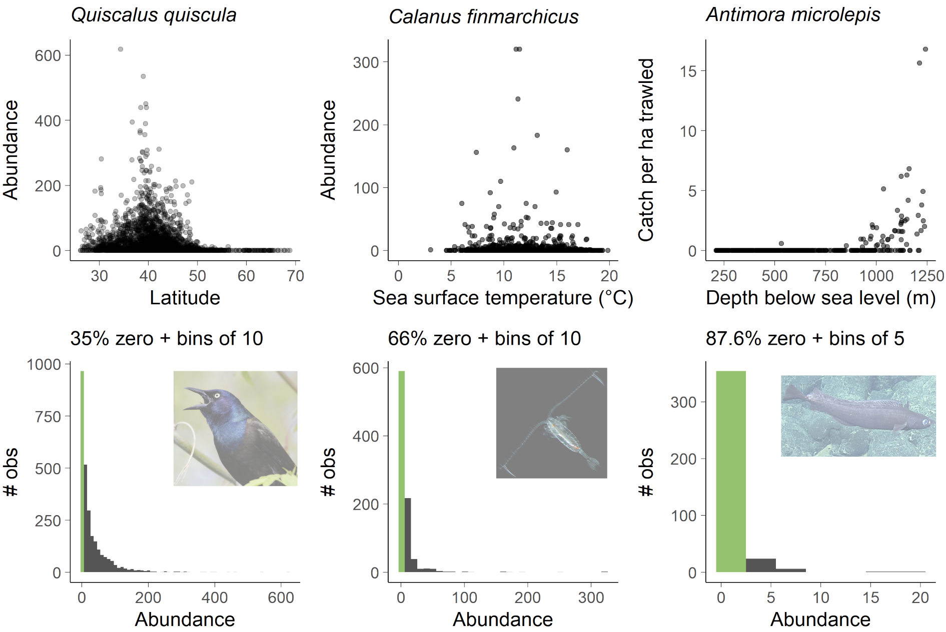
We need broad-scale data



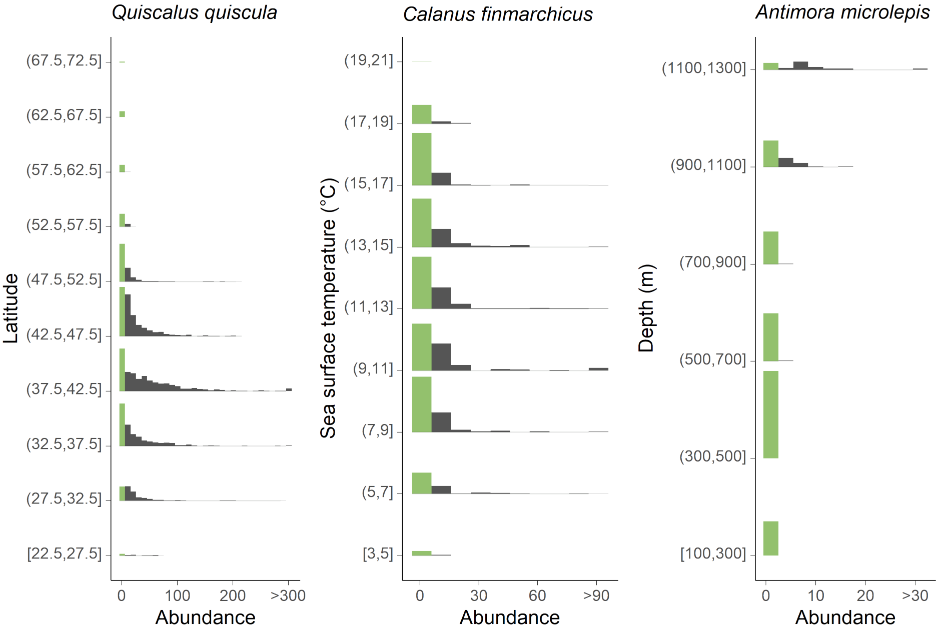




Species distributions along gradients



Species distributions along gradients



Some distributions for count data

Variance = mean:

$$y_i \sim \mathsf{Poisson}\left(\eta_i = \lambda_i, \theta = \{\}\right) \qquad \mathsf{E}\left[y_i\right] = \eta_i$$

Some distributions for count data

Variance $=$ mean:

$$y_i \sim \mathsf{Poisson}\left(\eta_i = \lambda_i, \theta = \{\}\right) \qquad \mathsf{E}\left[y_i\right] = \eta_i$$

Variance $>$ mean (overdispersion):

$$y_i \sim \mathsf{NegativeBinomial2}\left(\eta_i = \mu_i, \theta = \{\phi\}\right) \qquad \mathsf{E}\left[y_i\right] = \eta_i$$

Some distributions for count data

Variance = mean:

$$y_i \sim \mathsf{Poisson}\left(\eta_i = \lambda_i, \theta = \{\}\right) \qquad \mathsf{E}\left[y_i\right] = \eta_i$$

Variance > mean (overdispersion):

$$y_i \sim \mathsf{NegativeBinomial2}\left(\eta_i = \mu_i, \theta = \{\phi\}\right) \qquad \mathsf{E}\left[y_i\right] = \eta_i$$

Zero Inflation (constant prob. excess zeros):

$$y_i \sim \mathsf{ZID}\left(\eta_i, \theta, \pi\right) \qquad \mathsf{E}\left[y_i\right] = (1-\pi)\eta_i$$

$$= \begin{cases} \pi + (1-\pi) \times \mathsf{D}\left(0, \theta\right), & \text{if } y_i = 0, \\ (1-\pi) \times \mathsf{D}\left(\eta_i, \theta\right), & \text{if } y_n > 0 \end{cases}$$

Some distributions for count data

Variance = mean:

$$y_i \sim \text{Poisson}\left(\eta_i = \lambda_i, \theta = \{\}\right) \qquad \text{E}\left[y_i\right] = \eta_i$$

Variance > mean (overdispersion):

$$y_i \sim \text{NegativeBinomial2}\left(\eta_i = \mu_i, \theta = \{\phi\}\right) \qquad \text{E}\left[y_i\right] = \eta_i$$

<u>Z</u>ero <u>I</u>nflation (constant prob. excess zeros):

$$y_i \sim \text{ZID}\left(\eta_i, \theta, \pi\right) \qquad \text{E}\left[y_i\right] = (1-\pi)\eta_i$$

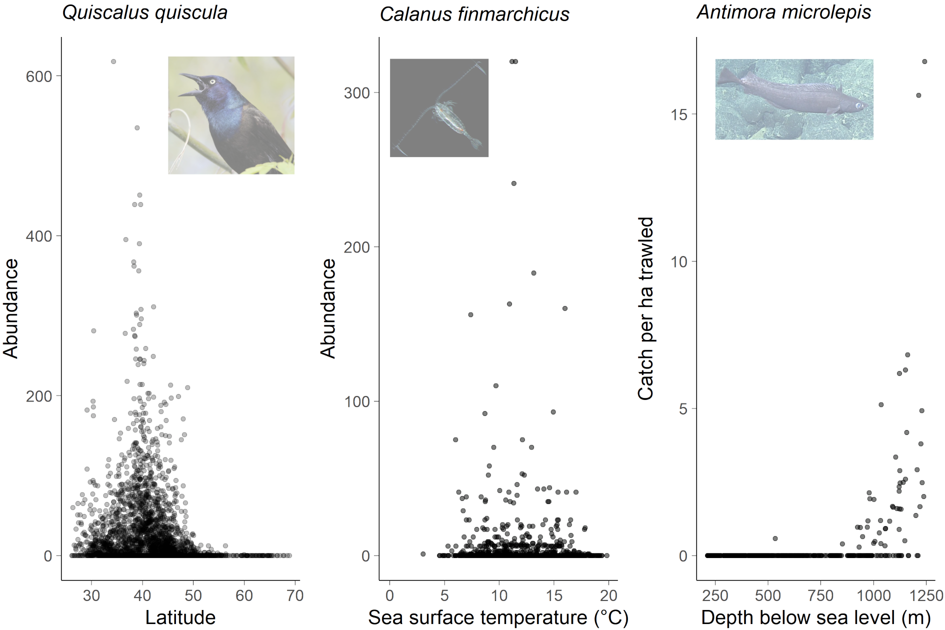
$$= \begin{cases} \pi + (1-\pi) \times \text{D}\left(0, \theta\right), & \text{if } y_i = 0, \\ (1-\pi) \times \text{D}\left(\eta_i, \theta\right), & \text{if } y_n > 0 \end{cases}$$

<u>ZI</u> (prob. excess zeros <u>L</u>inked to η:)

$$y_i \sim \text{ZIDL}\left(\eta_i, \theta, \pi_i, \gamma_0, \gamma_1\right) \qquad \text{E}\left[y_i\right] = (1-\pi)\eta_i$$

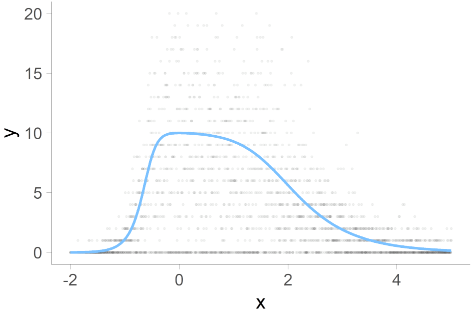
$$\text{logit } \pi_i = \gamma_0 - \gamma_1 \log\left(\eta_i\right)$$

Species-environment relationships aren't quite linear



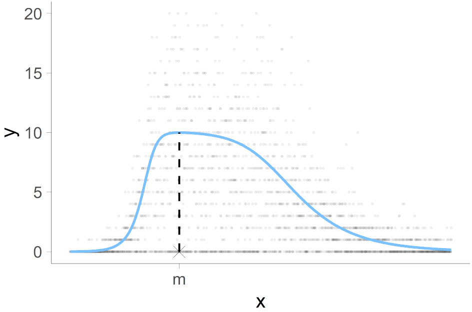
A non-linear predictor

- $\eta_i = f(x_i)$

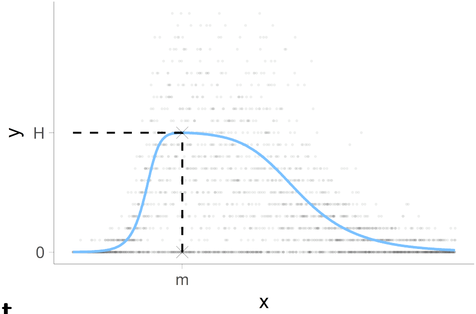


A non-linear predictor

- $\eta_i = f(x_i, m)$
- $m \in \mathbb{R}$: **mode**, optimum x

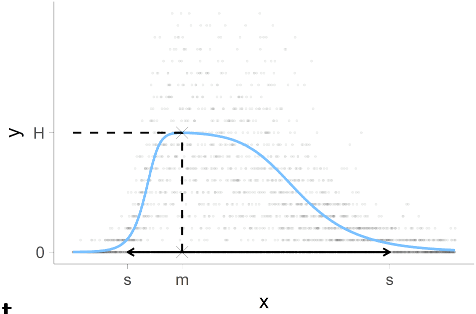


A non-linear predictor



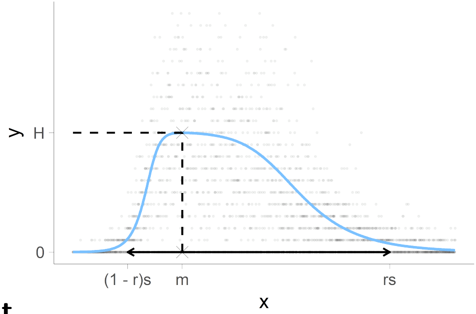
- $\eta_i = f(x_i, H, m)$
- $m \in \mathbb{R}$: **mode**, optimum x
- $H \in \mathbb{R}_{>0}$: maximum **Height**

A non-linear predictor



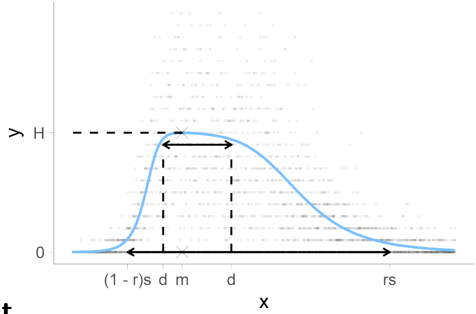
- $\eta_i = f(x_i, H, m, s)$
- $m \in \mathbb{R}$: **mode**, optimum x
- $H \in \mathbb{R}_{>0}$: maximum **Height**
- $s \in \mathbb{R}_{>0}$: **spread**, tolerable range

A non-linear predictor



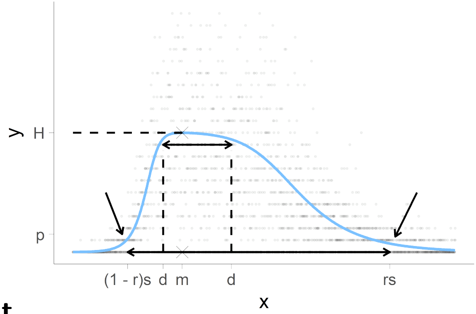
- $\eta_i = f(x_i, H, m, s, r)$
- $m \in \mathbb{R}$: **mode**, optimum x
- $H \in \mathbb{R}_{>0}$: maximum **Height**
- $s \in \mathbb{R}_{>0}$: **spread**, tolerable range
- $r \in (0, 1)$: **asymmetry**

A non-linear predictor



- $\eta_i = f(x_i, H, m, s, r, d)$
- $m \in \mathbb{R}$: **mode**, optimum x
- $H \in \mathbb{R}_{>0}$: maximum **Height**
- $s \in \mathbb{R}_{>0}$: **spread**, tolerable range
- $r \in (0, 1)$: **asymmetry**
- $d \in \mathbb{R}$: **broadness** or peakedness around the optimum

A non-linear predictor



- $\eta_i = f(x_i, H, m, s, r, d, p)$
- $m \in \mathbb{R}$: **mode**, optimum x
- $H \in \mathbb{R}_{>0}$: maximum **Height**
- $s \in \mathbb{R}_{>0}$: **spread**, tolerable range
- $r \in (0, 1)$: **asymmetry**
- $d \in \mathbb{R}$: **broadness** or peakedness around the optimum
- $p \in \mathbb{R}_{>0}$: **exagerration** or pinching of tails

The **modskurt** family $\eta_i = f(x_i)$

$$f(x_i|\psi) = H\left(re^{\frac{1}{p}\left(\frac{z_i}{r}-d\right)} + (1-r)e^{-\frac{1}{p}\left(\frac{z_i}{1-r}+d\right)} - e^{-\frac{d}{p}} + 1\right)^{-p}$$

$$z_i = \frac{x_i - m}{s}$$

$$\psi = \{H, m, s, r, d, p\}$$

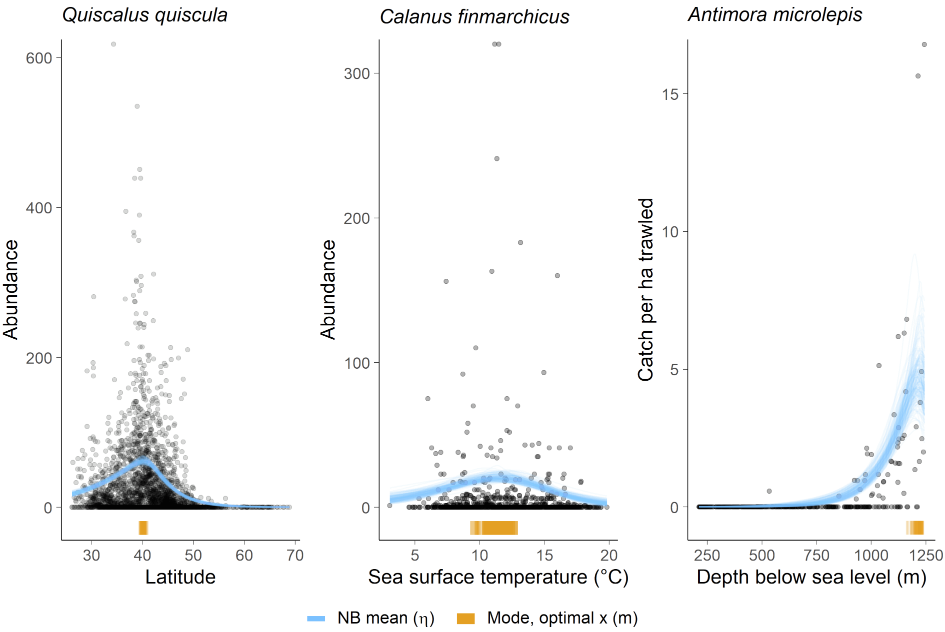
- All parameters for complex species-env relationships
- $p \to 1$ removes tail exagerration
- $d \to 0$ removes peak flatness
- $r \to 0.5$ removes asymmetry
- primer-e.github.io/senlm/articles/modskurt.html

What we have so far

For obs $i \in 1, \ldots, N$

$$
\begin{aligned}
y_i &\sim \text{ZIDL}\left(\eta_i, \theta = \{\phi, \pi, \gamma_0, \gamma_1\}\right) \\
\eta_i &= f(x_i, H, m, s, r, d, p) \\
H &\sim \text{Beta}(2,4) \cdot \max y & H &\in (0, \max y) \\
m &\sim \text{Beta}(1,1) \cdot (u - l) + l & &\approx \text{Unif}\,[l = \min x, u = \max x] \\
s &\sim \text{HalfNormal}(1, \sigma_s) \cdot \bar{\sigma}_x & &\text{standardise over x} \\
r &\sim \text{Beta}(2,2) \\
d &\sim \text{HalfNormal}(0, \sigma_d) \\
p &\sim \text{HalfNormal}(1, \sigma_p) \\
\theta &\sim P(\mu_\theta, \sigma_\theta) & &\text{indep. dist priors}
\end{aligned}
$$

Posterior predictive checks



Responses need time

For year $t \in 1, \ldots, T$, each with observations $i \in 1, \ldots, N_t$

$$y_{t[i]} \sim \text{ZIDL}\left(\eta_{t[i]}, \theta = \{\phi, \pi, \gamma_0, \gamma_1\}\right)$$

Responses need time

For year $t \in 1, \ldots, T$, each with observations $i \in 1, \ldots, N_t$

$$y_{t[i]} \sim \mathsf{ZIDL}\left(\eta_{t[i]}, \theta = \{\phi, \pi, \gamma_0, \gamma_1\}\right)$$

Parameters for the error distribution consistent across years

$$\theta \sim P(\mu_\theta, \sigma_\theta) \quad \leftarrow \text{indep. dist priors}$$

Responses need time

For year $t \in 1, \ldots, T$, each with observations $i \in 1, \ldots, N_t$

$$y_{t[i]} \sim \text{ZIDL}\left(\eta_{t[i]}, \theta = \{\phi, \pi, \gamma_0, \gamma_1\}\right)$$

Parameters for the error distribution consistent across years

$$\theta \sim P(\mu_\theta, \sigma_\theta) \quad \leftarrow \text{indep. dist priors}$$

Different mean curves fit for each year

$$\eta_{t[i]} = f(x_{t[i]}, H_t, m_t, s_t, r_t, d_t, p_t)$$

Responses need time

For year $t \in 1, \ldots, T$, each with observations $i \in 1, \ldots, N_t$

$$y_{t[i]} \sim \text{ZIDL}\left(\eta_{t[i]}, \theta = \{\phi, \pi, \gamma_0, \gamma_1\}\right)$$

Parameters for the error distribution consistent across years

$$\theta \sim P(\mu_\theta, \sigma_\theta) \quad \leftarrow \text{indep. dist priors}$$

Different mean curves fit for each year

$$\eta_{t[i]} = f(x_{t[i]}, H_t, m_t, s_t, r_t, d_t, p_t)$$

Modal position as a function of time

$$m_t = g(t) + \epsilon_t$$

Responses need time

For year $t \in 1, \ldots, T$, each with observations $i \in 1, \ldots, N_t$

$$y_{t[i]} \sim \mathsf{ZIDL}\left(\eta_{t[i]}, \theta = \{\phi, \pi, \gamma_0, \gamma_1\}\right)$$

Parameters for the error distribution consistent across years

$$\theta \sim P(\mu_\theta, \sigma_\theta) \quad \leftarrow \text{indep. dist priors}$$

Different mean curves fit for each year

$$\eta_{t[i]} = f(x_{t[i]}, H_t, m_t, s_t, r_t, d_t, p_t)$$

Modal position as a function of time

$$m_t = g(t) + \epsilon_t$$

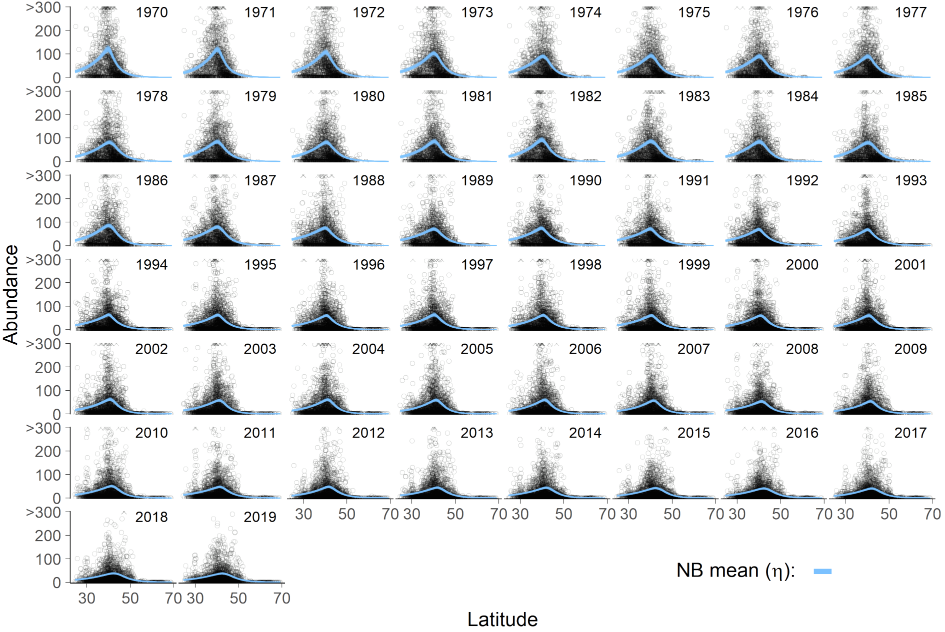
Remaining modskurt parameters are partially pooled across years

$$\begin{aligned} \psi &= \{H, s, r, d, p\} \\ \bar{\psi}_j &\sim P(\mu_\psi, \sigma_{\bar{\psi}}) \quad \leftarrow \text{average priors the same as earlier} \\ \psi_t &\sim \mathsf{Normal}(\bar{\psi}, \sigma_\psi) \quad \leftarrow \text{varying each year} \\ \sigma_\psi &\sim \mathsf{HalfNormal}(0, \tau_\psi) \quad \leftarrow \text{with some regularisation} \end{aligned}$$

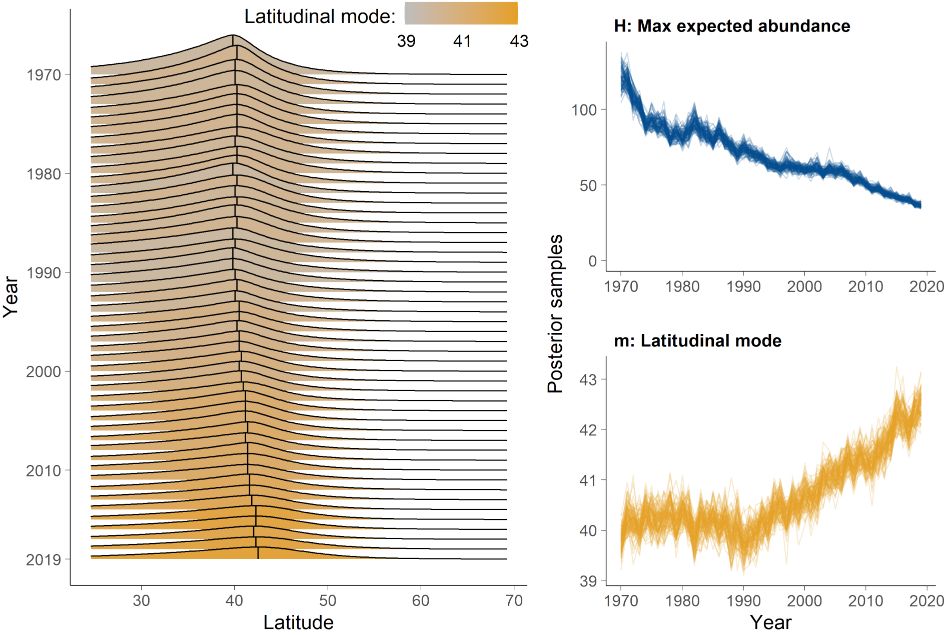
Technical issues

```
functions {
  vector zinbl(vector pooled, vector pars_t, array[] real x_t, array[] int y_t) {
    real H = pars_t[1]; real m = pars_t[2]; real s = pars_t[3];
    real r = pars_t[4]; real d = pars_t[5]; real p = pars_t[6];

    real phi = pooled[1]; real gamma_0 =  pooled[2]; real gamma_1 =  pooled[3];

    int N_t = y_t[1]; int Nz_t = y_t[2]; int Npos = N_t - Nz_t;

    vector[N_t] x = to_vector(x_t[3:(N_t + 2)]);
    array[Npos] int y = y_t[(Nz_t + 3):(N_t + 2)];

    vector[N_t] log_eta = log_modskurt(x, H, m, s, r, d, p);
    vector[N_t] logit_pi = fma(- gamma_1, log_eta, gamma_0);

    real lp = 0;
    if (Nz_t > 0) {
      lp += log_sum_exp(bernoulli_logit_lpmf(1 | logit_pi[1:Nz_t]),
                        bernoulli_logit_lpmf(0 | logit_pi[1:Nz_t]) + neg_binomial_2_log_lpmf(0 | log_eta[1:Nz_t], phi));
    }
    lp += bernoulli_logit_lpmf(0 | logit_pi[(Nz_t + 1):N_t]) + neg_binomial_2_log_lpmf(y | log_eta[(Nz_t + 1):N_t], phi);
    return [lp]';
  }
}
model {
  target += sum(map_rect(zinbl, pars_pooled, pars_t, x, y));
}
```

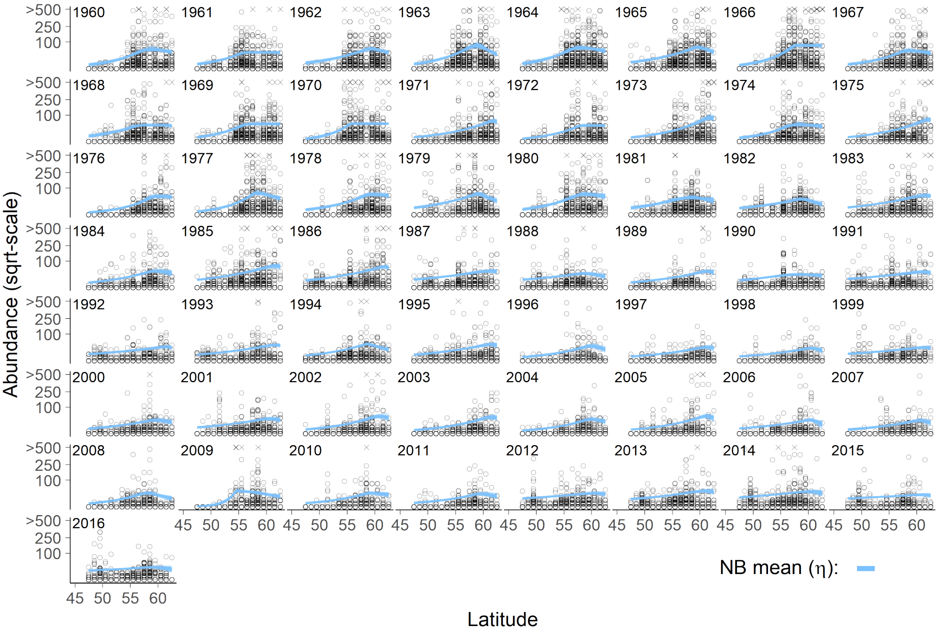
Quiscalus quiscula (common grackle) + lat



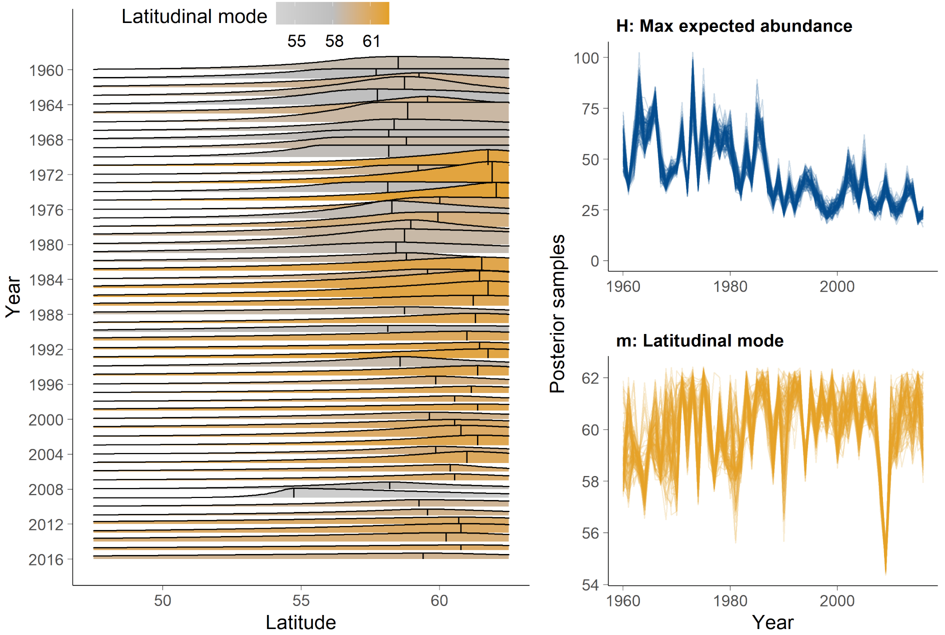
Quiscalus quiscula (common grackle) + lat



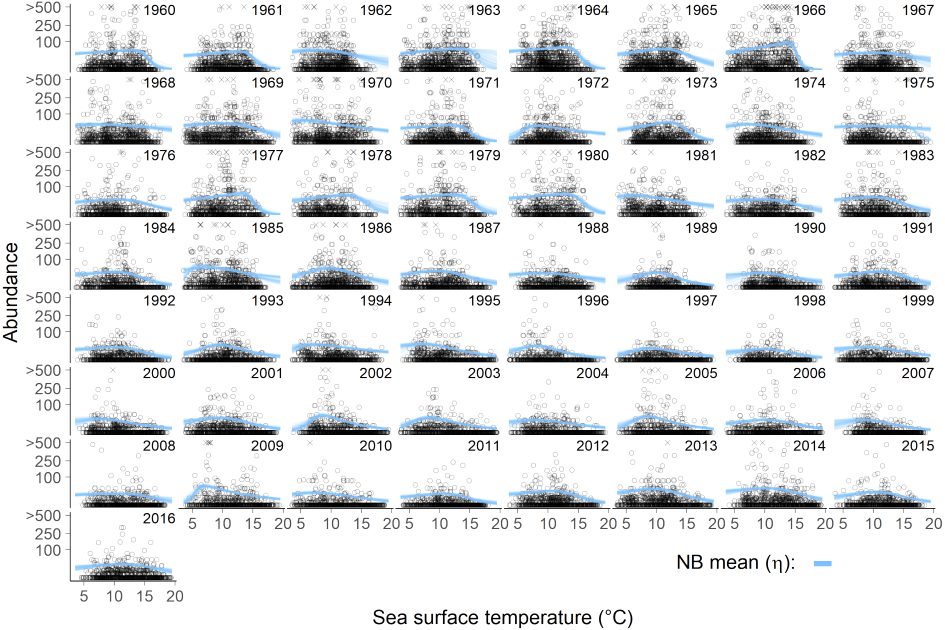
Calanus finmarchicus (cold water copepod) + lat



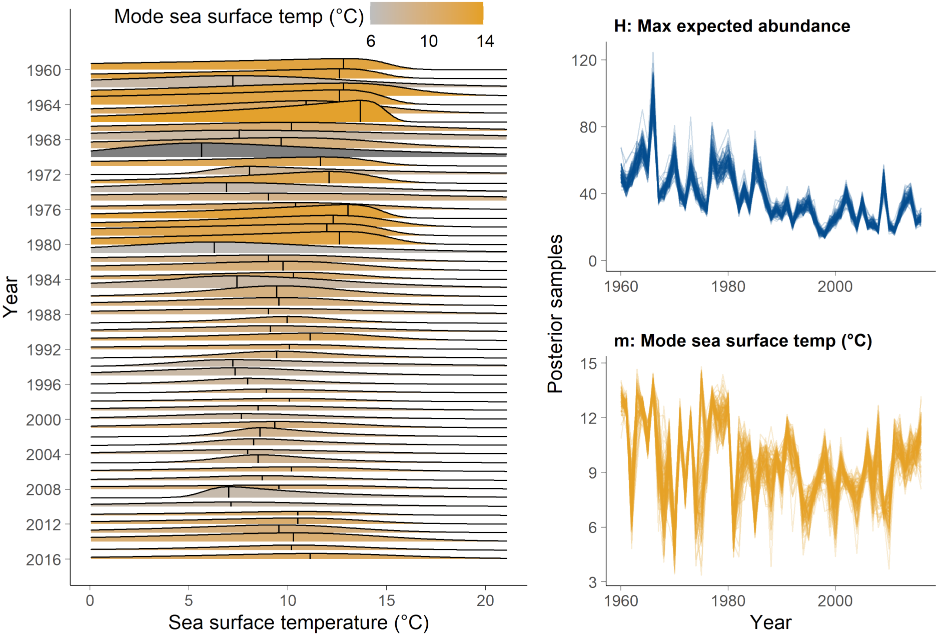
Calanus finmarchicus (cold water copepod) + lat



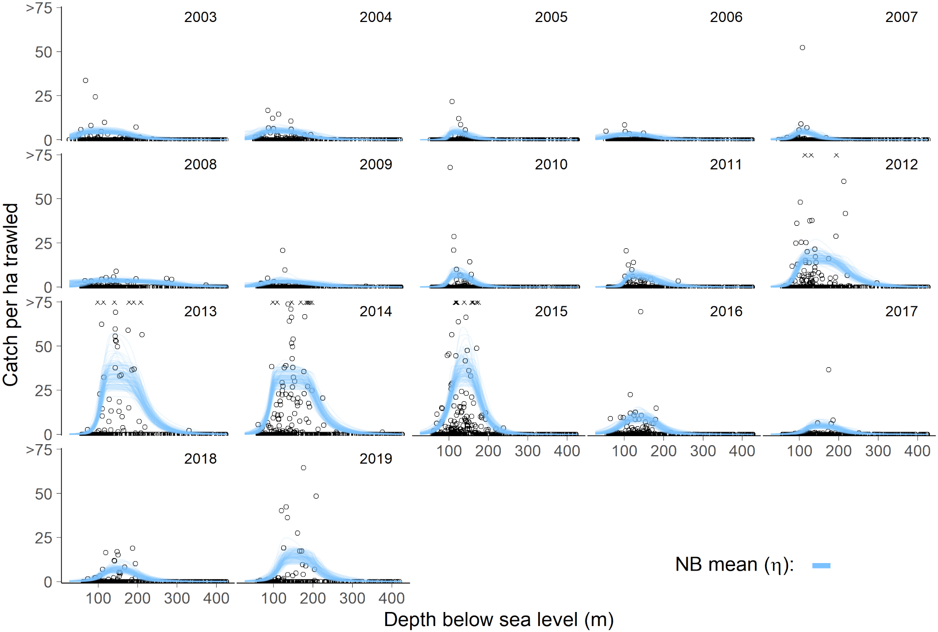
Calanus finmarchicus (cold water copepod) + temp



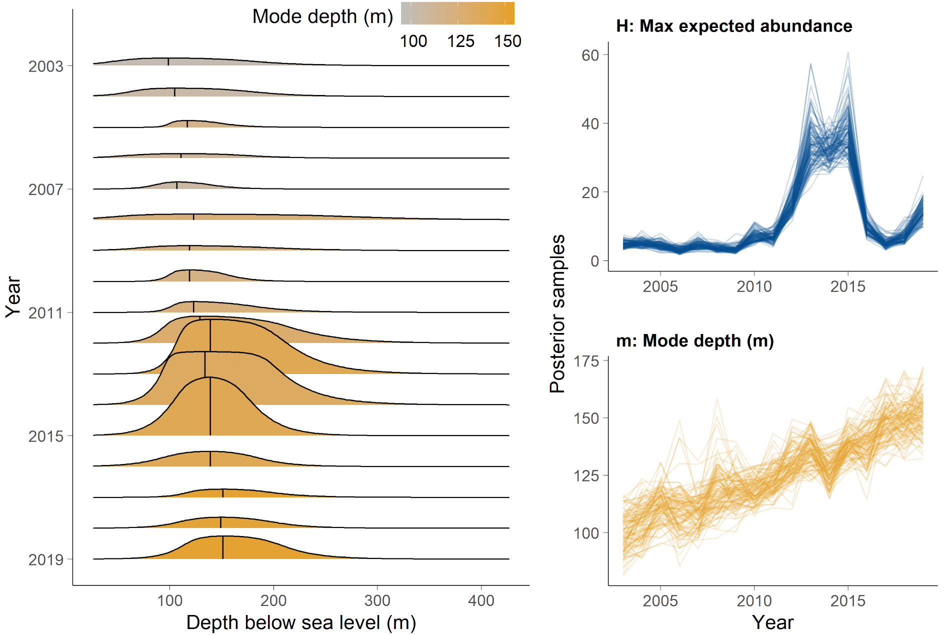
Calanus finmarchicus (cold water copepod) + temp



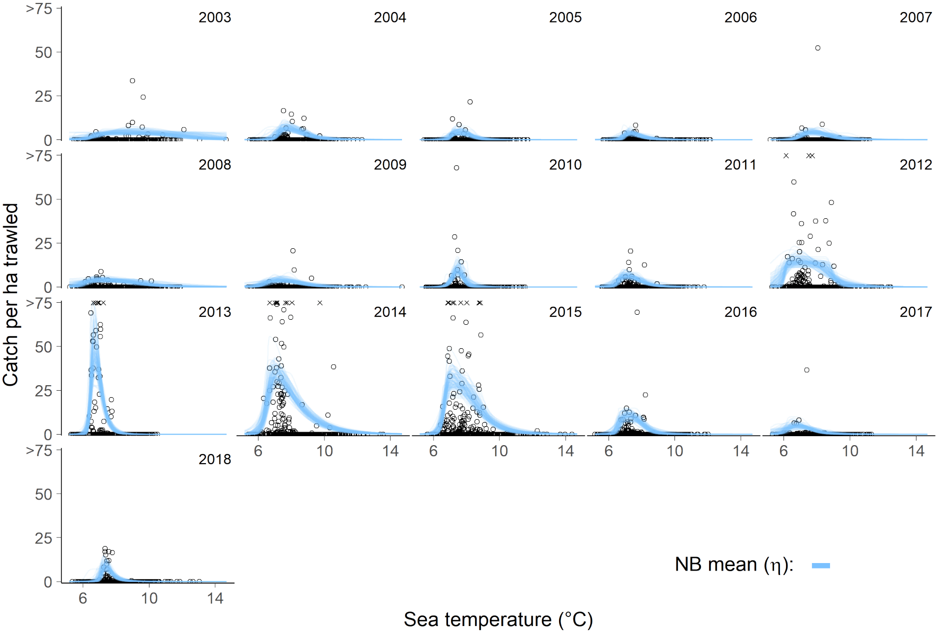
Thaleichthys pacificus (eulachon) + depth



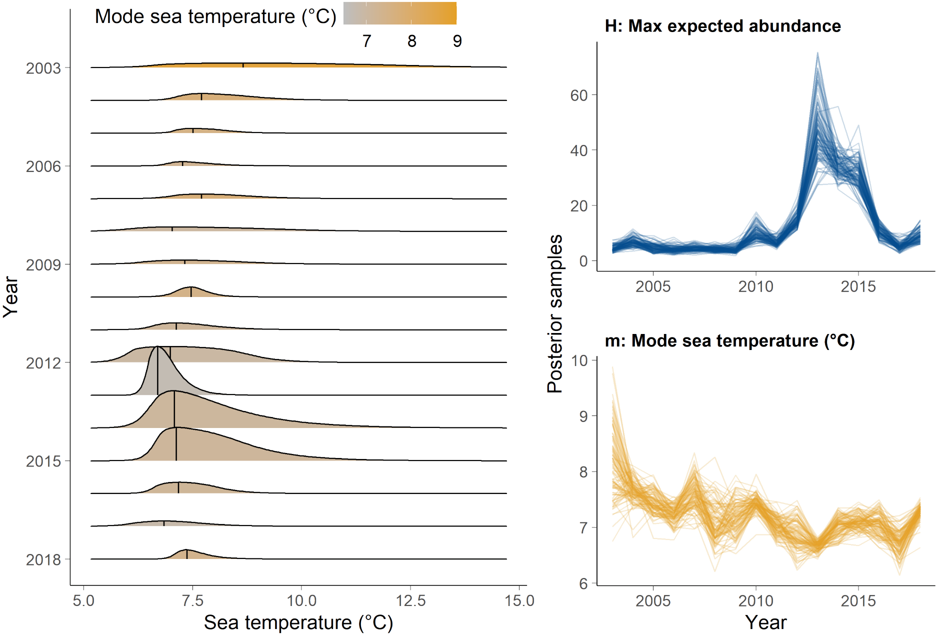
Thaleichthys pacificus (eulachon) + depth



Thaleichthys pacificus (eulachon) + temp



Thaleichthys pacificus (eulachon) + temp



In summary

- Flexible and interpretable model for exploring species-environment data over time (some things to iron out)
- Bayesian approach propagates uncertainty across multiple levels, and benefits from regularising prior model
- Building from principles may offer inferential advantages over more predictive approaches

Next steps

- Do ironing
- Formal models for causality
- Publish R package to extend `brms`

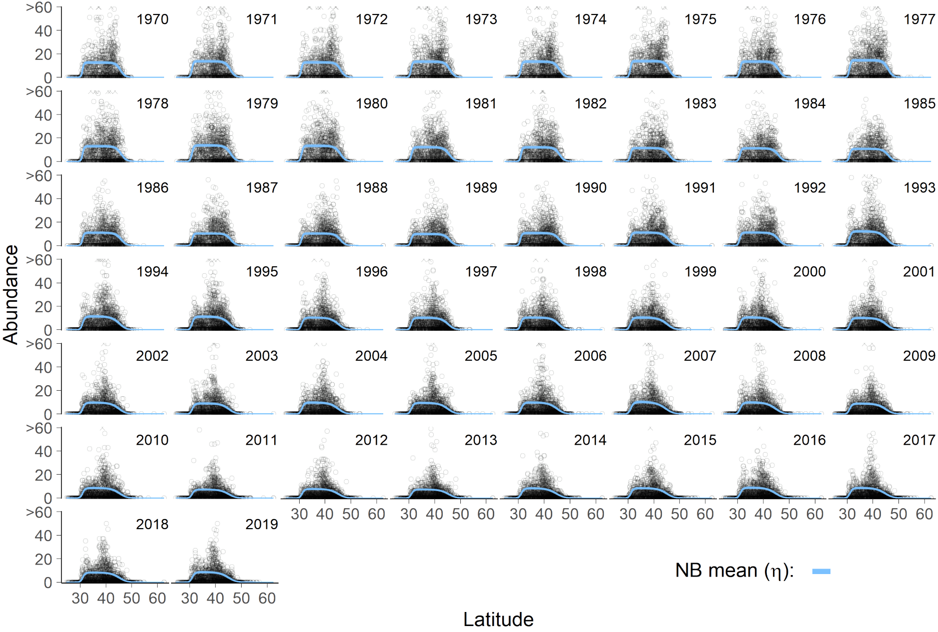
Acknowledgements

- My supervisors and team: Marti Anderson, Adam Smith, Winston Sweatman, Andrew Punnet, Paulo Mateus Martins
- Massey University Masters Research Scholarship
- PRIMER-e Inaugural Masterate Scholarship
- New Zealand Institute for Advanced Study - Center for Theoretical Chemistry and Physics HPC cluster.
- Royal Society of New Zealand Marsden Grant (19-466 MAU-145)
- Canadian Wildlife Service & United States Geological Survey, UK Marine Biological Association, National Oceanic and Atmospheric Administration Fisheries

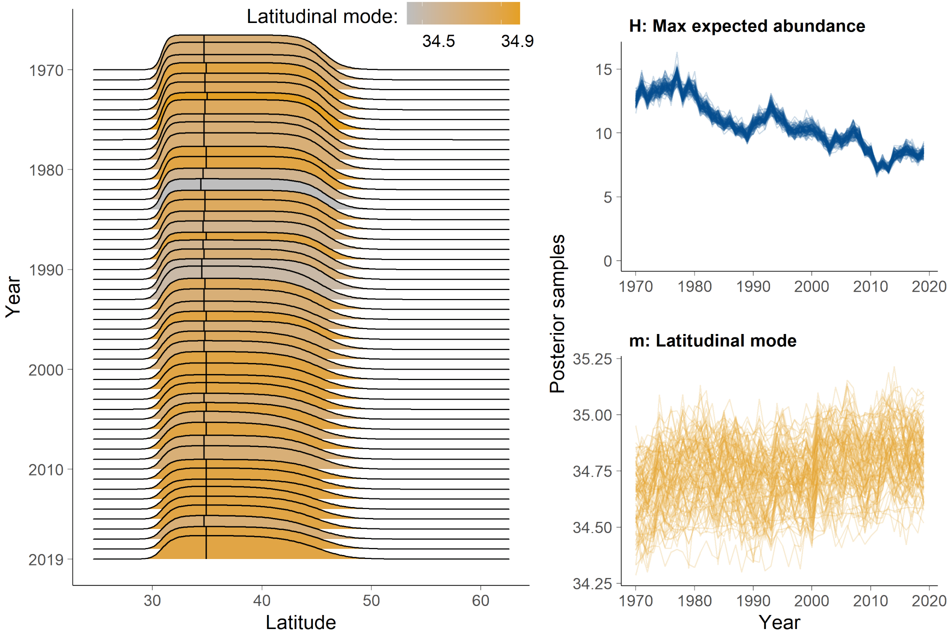
Extras

…

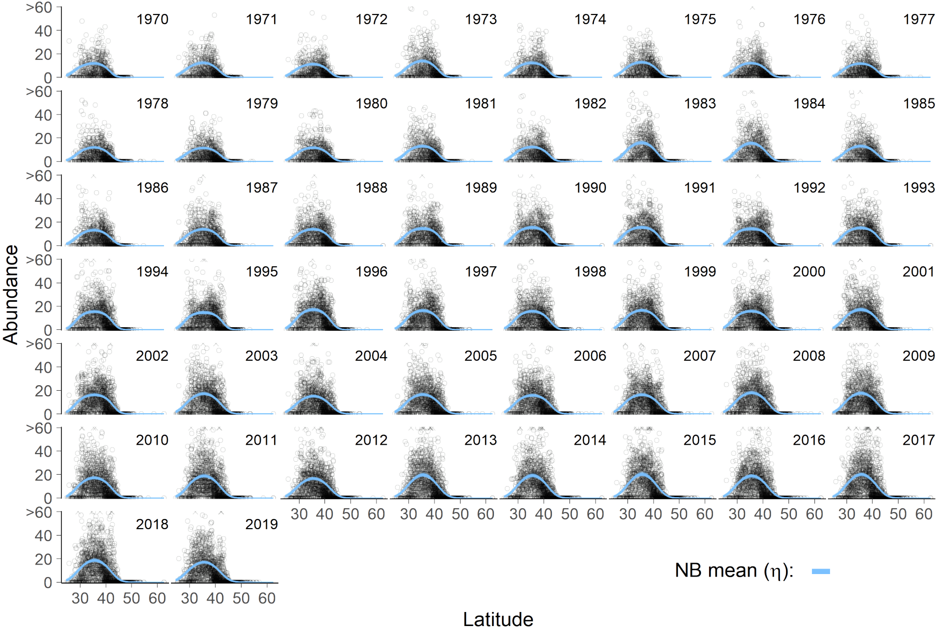
Hylocichla mustelina (wood thrush) + lat



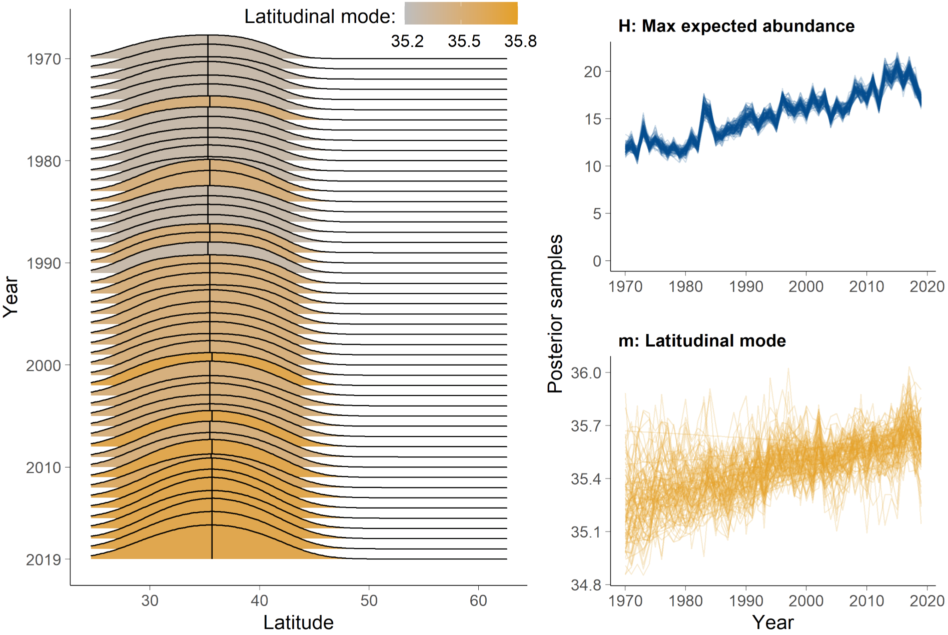
Hylocichla mustelina (wood thrush) + lat



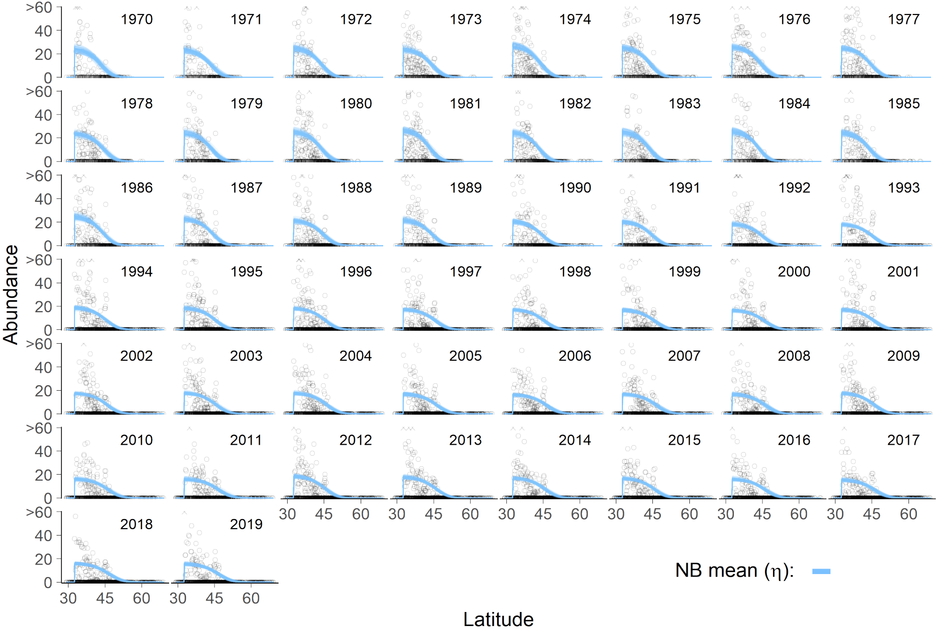
Baeolophus bicolor (tufted titmouse) + lat



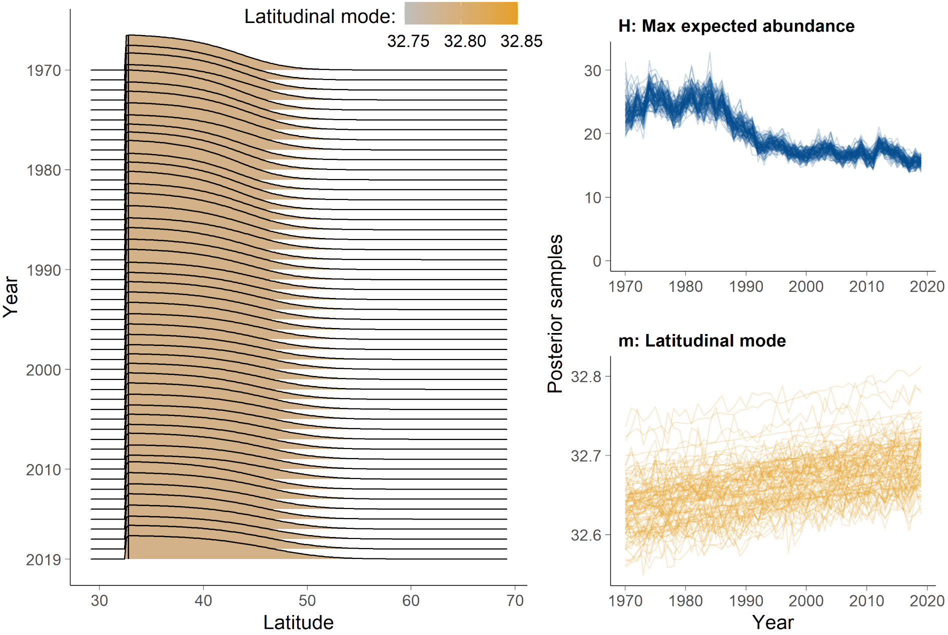
Baeolophus bicolor (tufted titmouse) + lat



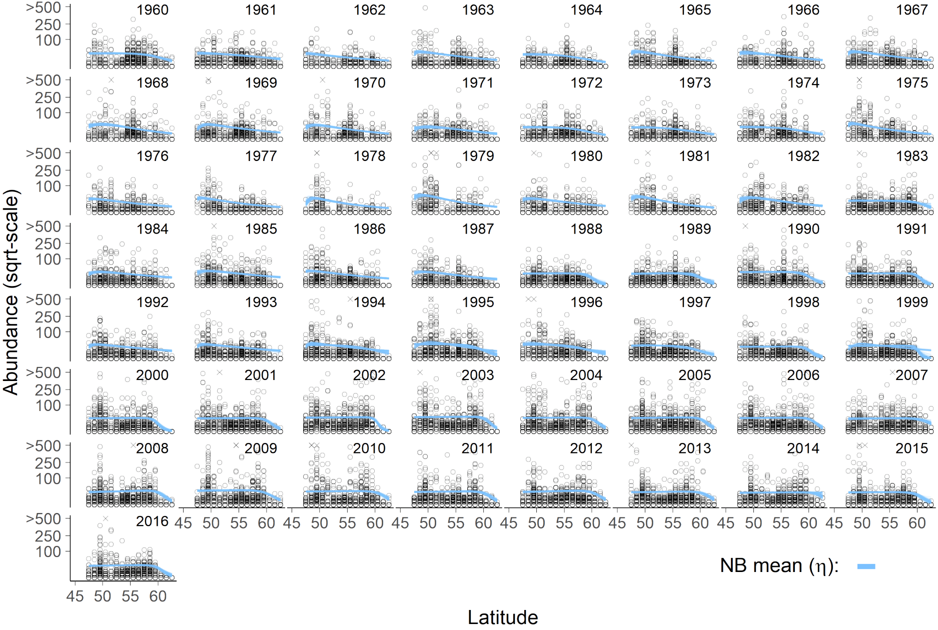
Aphelocoma californica (Cali scrub jay) + lat



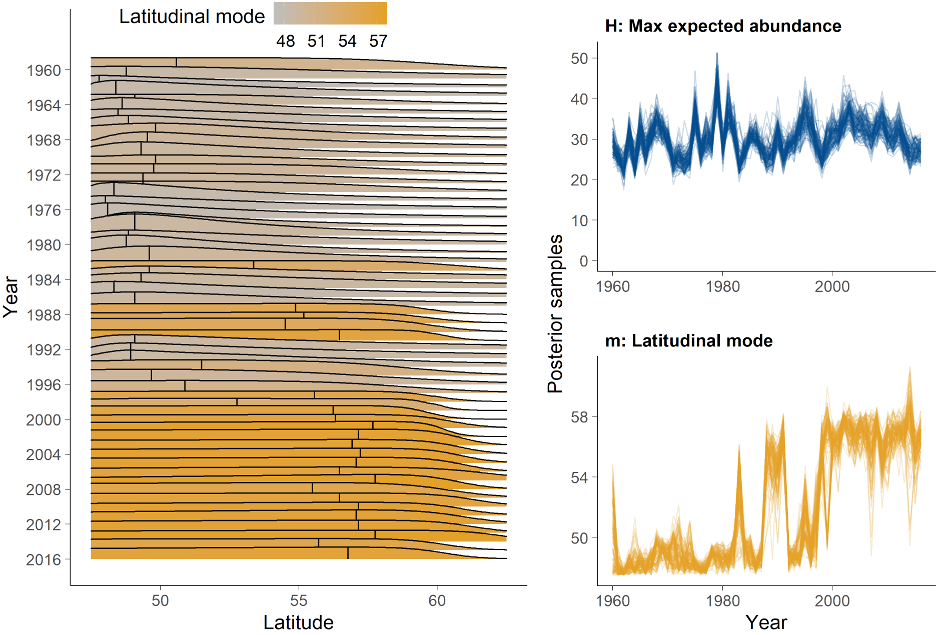
Aphelocoma californica (Cali scrub jay) + lat



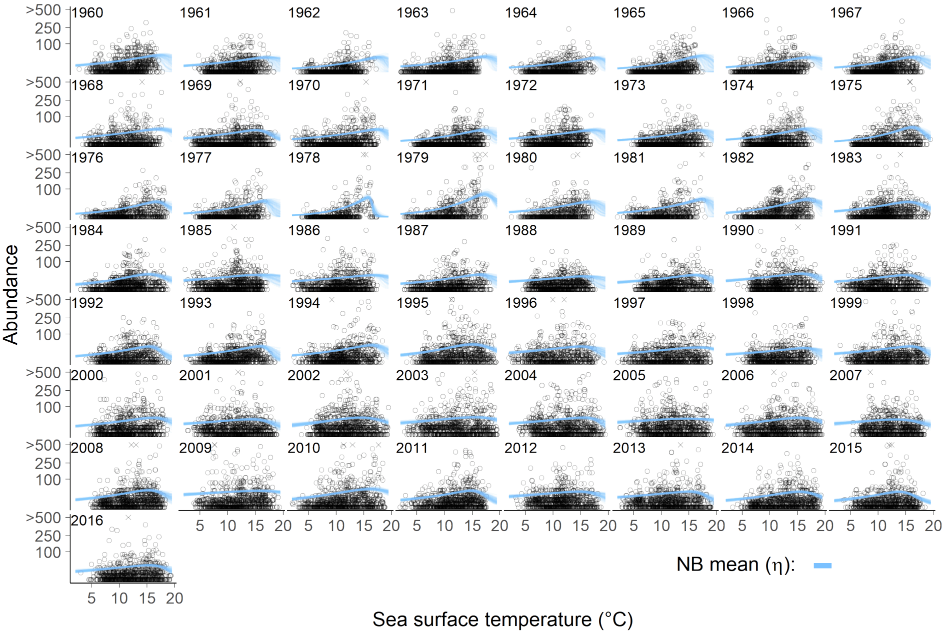
Calanus helgolandicus (warm water copepod) + lat



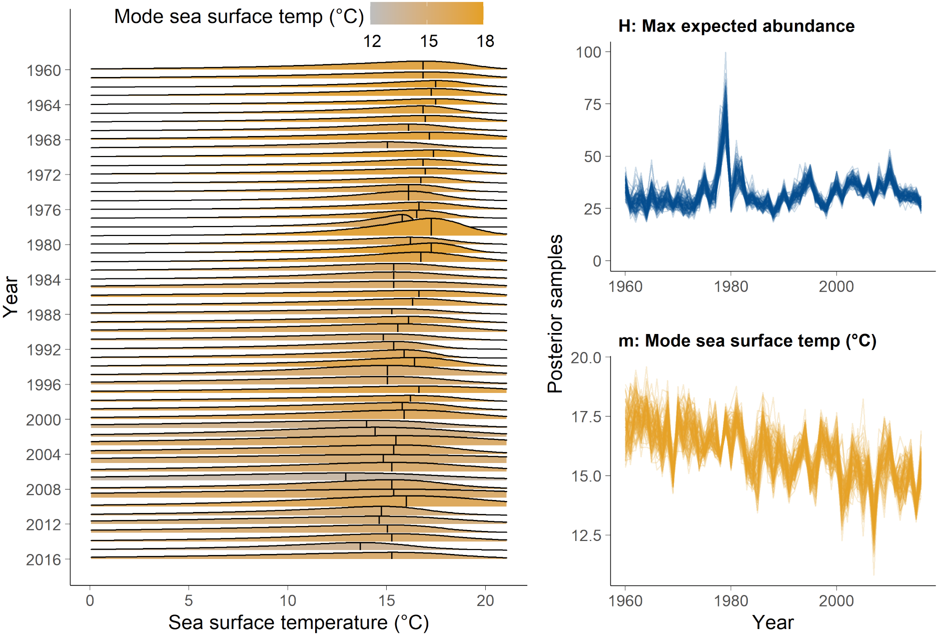
Calanus helgolandicus (warm water copepod) + lat



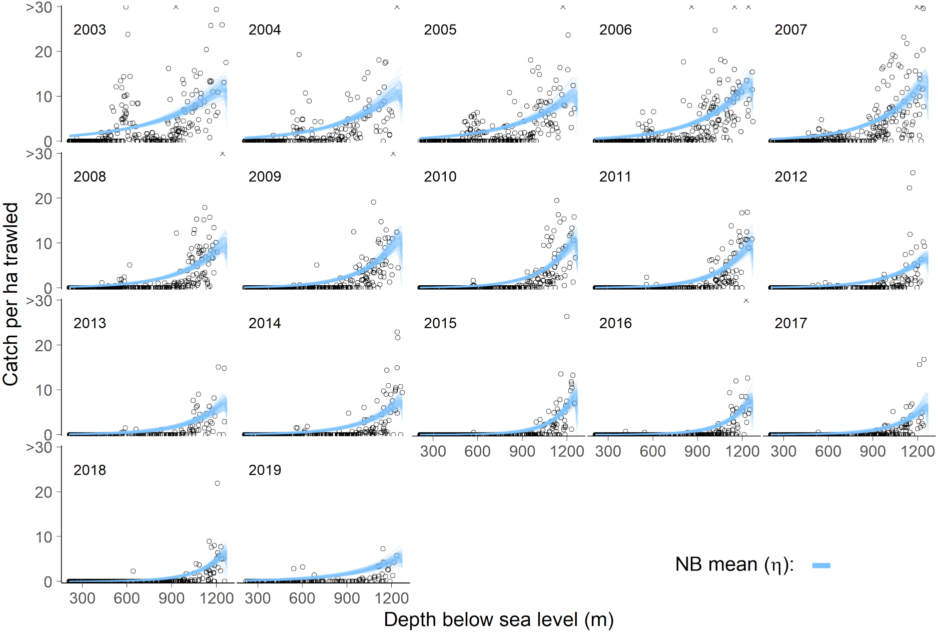
Calanus helgolandicus (warm water copepod) + temp



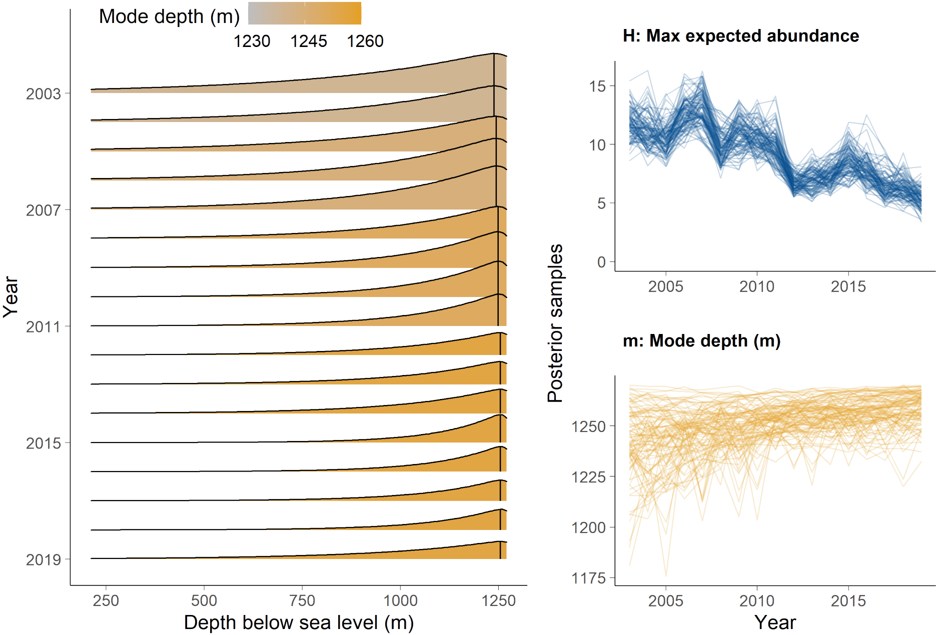
Calanus helgolandicus (warm water copepod) + temp



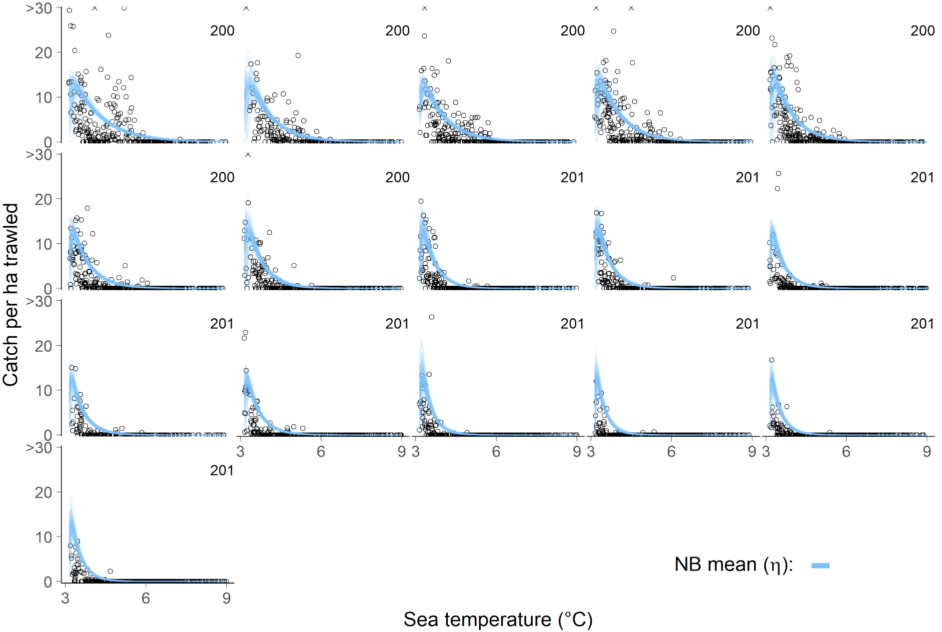
Antimora microlepis (pacific flatnose) + depth



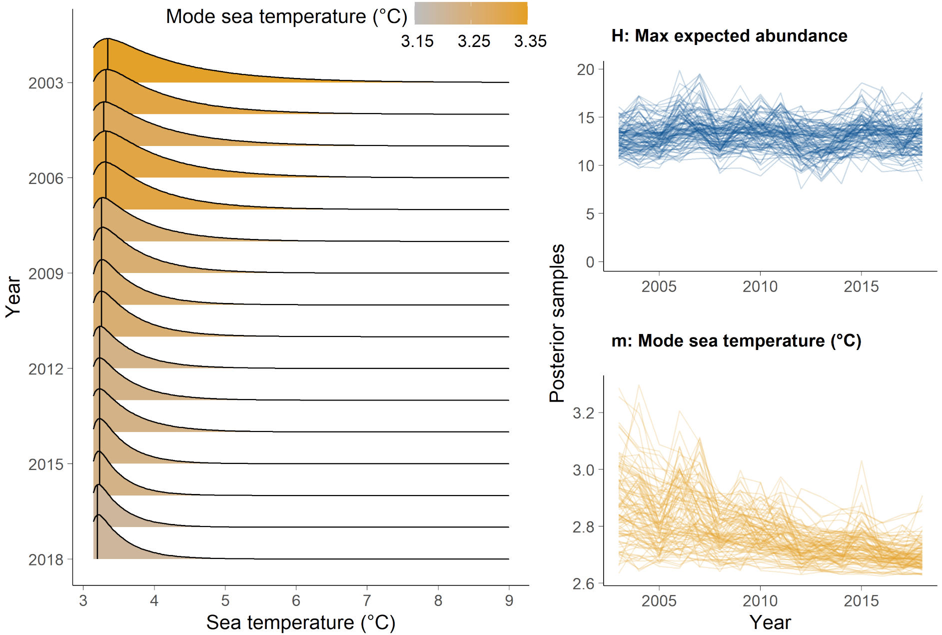
Antimora microlepis (pacific flatnose) + depth



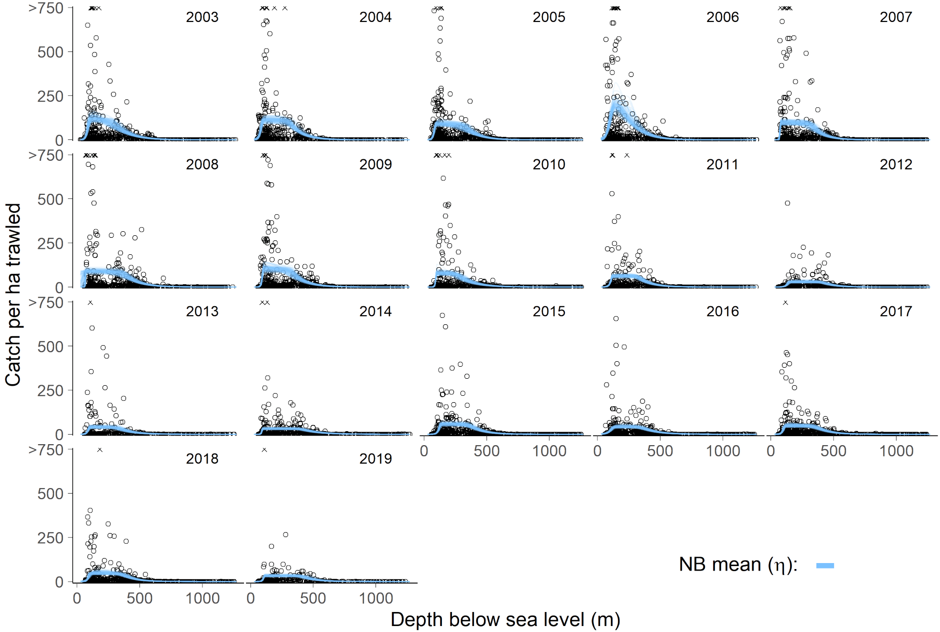
Antimora microlepis (pacific flatnose) + temp



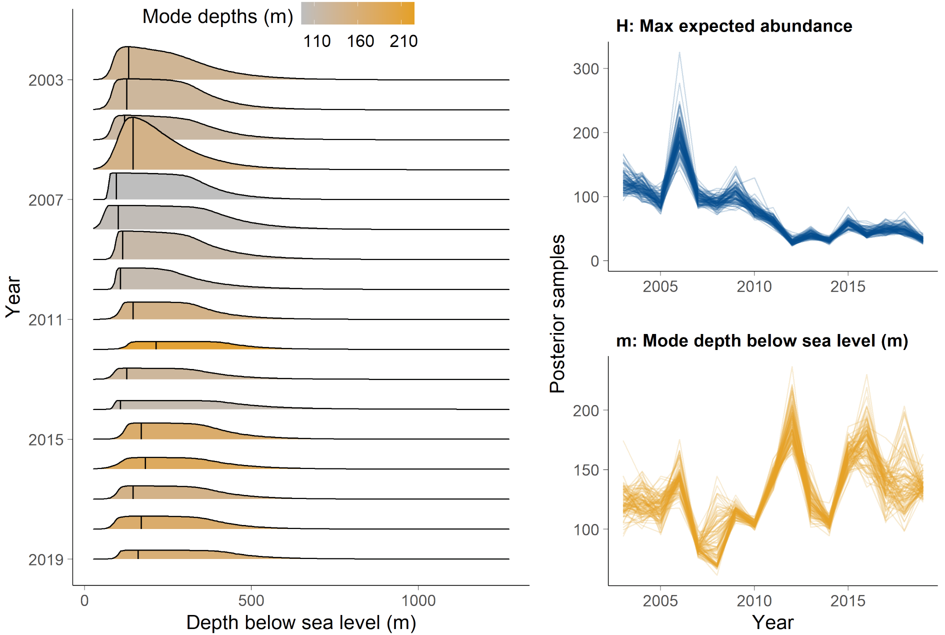
Antimora microlepis (pacific flatnose) + temp



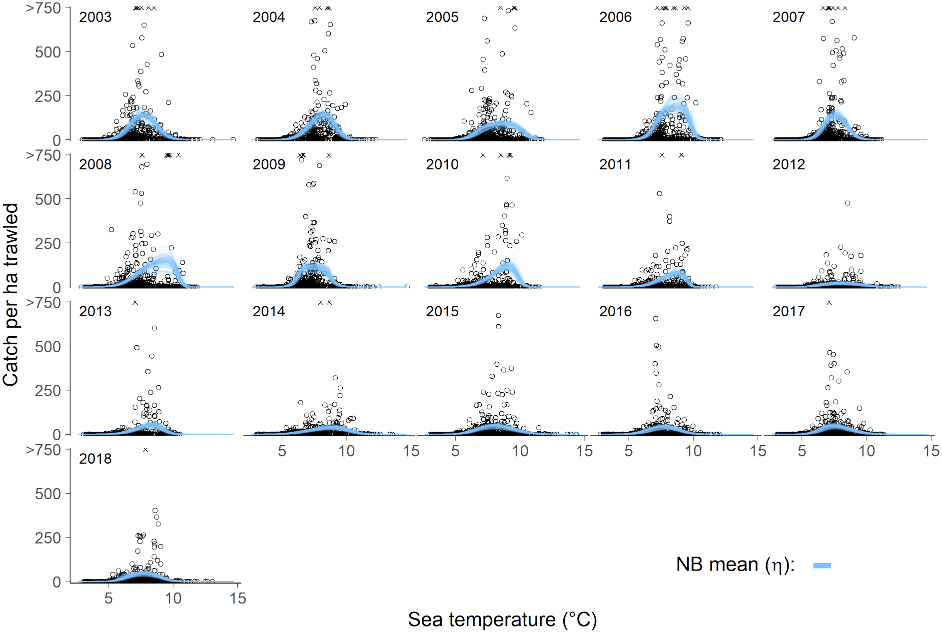
Merluccius productus (pacific hake) + lat



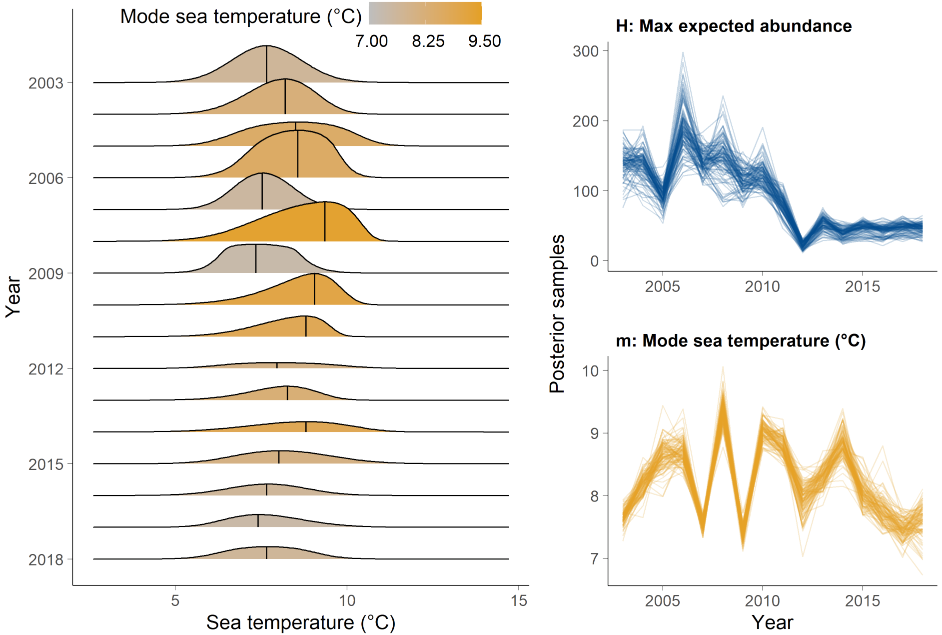
Merluccius productus (pacific hake) + lat



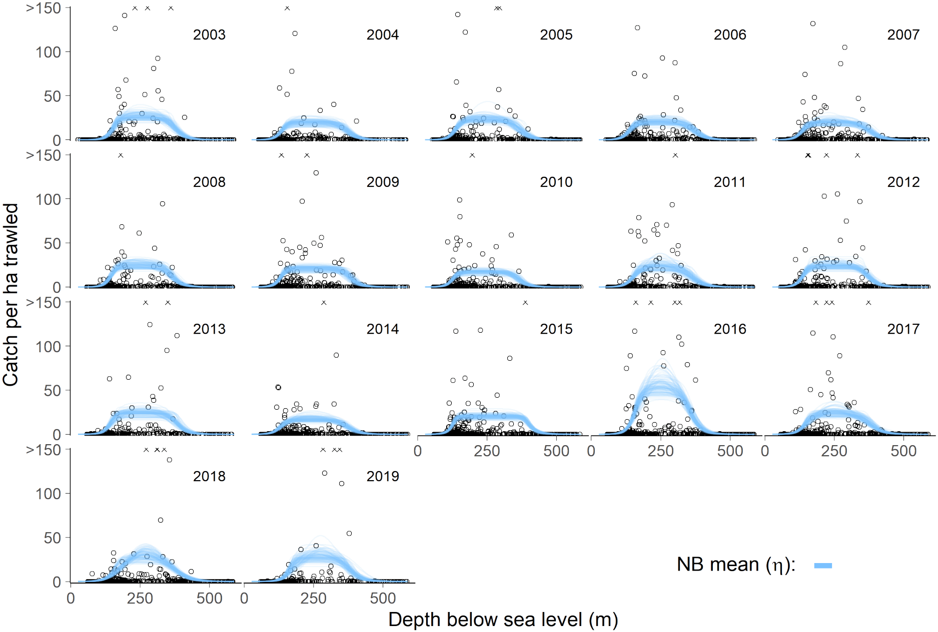
Merluccius productus (pacific hake) + temp



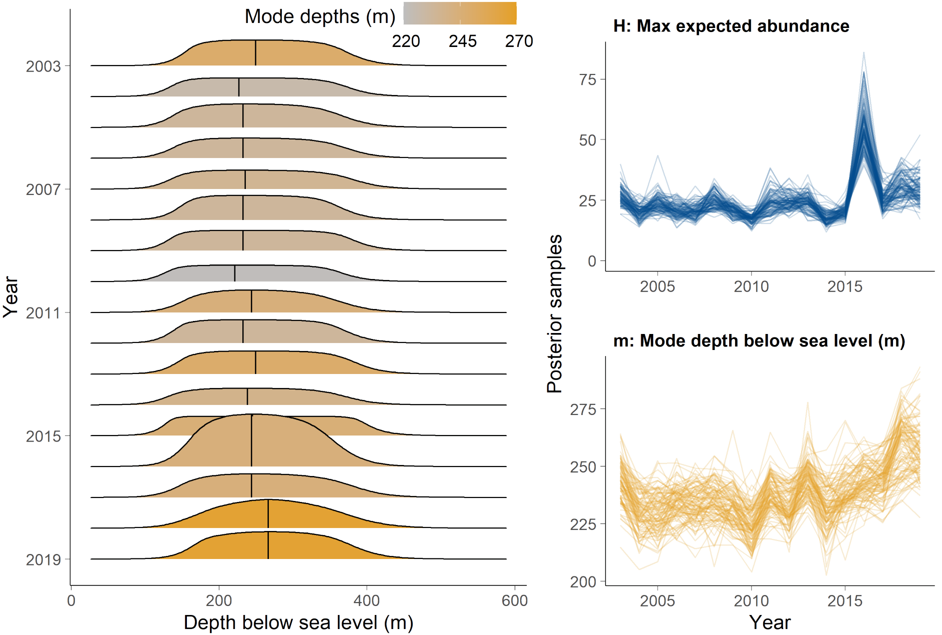
Merluccius productus (pacific hake) + temp



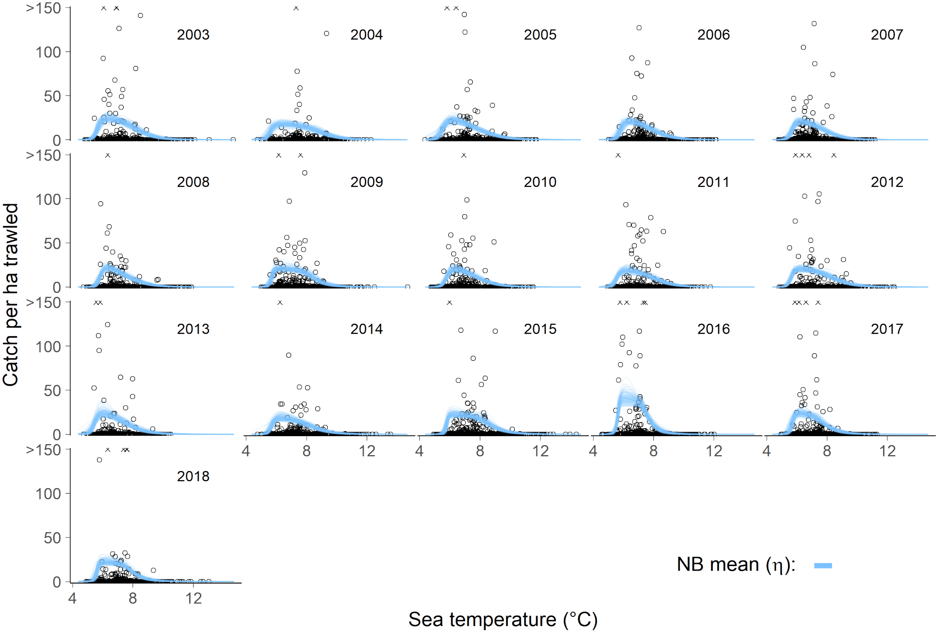
Sebastes crameri (darkblotched rockfish) + lat



Sebastes crameri (darkblotched rockfish) + lat



Sebastes crameri (darkblotched rockfish) + temp



Sebastes crameri (darkblotched rockfish) + temp