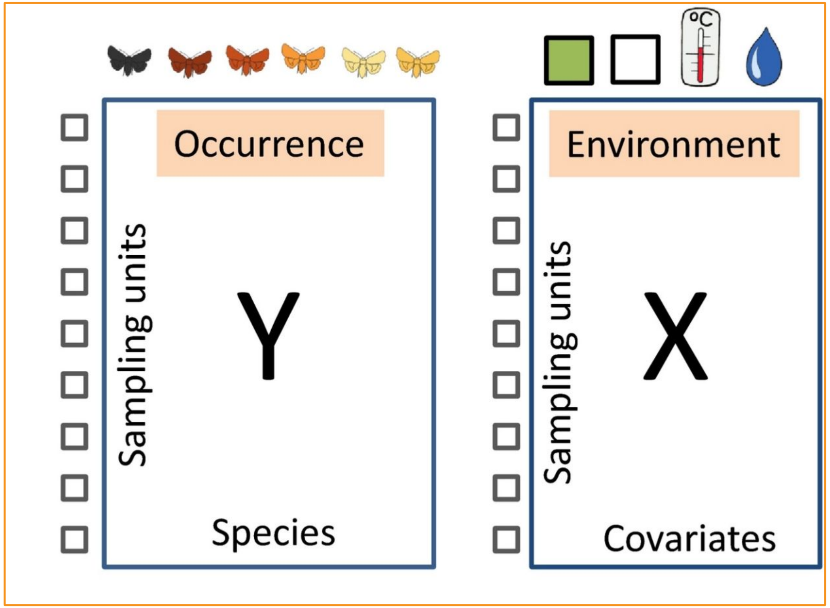
Scalable finite mixture of regression models for community ecology

Francis K.C. Hui Australian National University
Patricia Menendez University of Melbourne
Scott Foster CSIRO Data61
Skipton Woolley CSIRO Data61

Some images courtesy of SlidesCarnival and Google images
IBS-AR Conference 2025

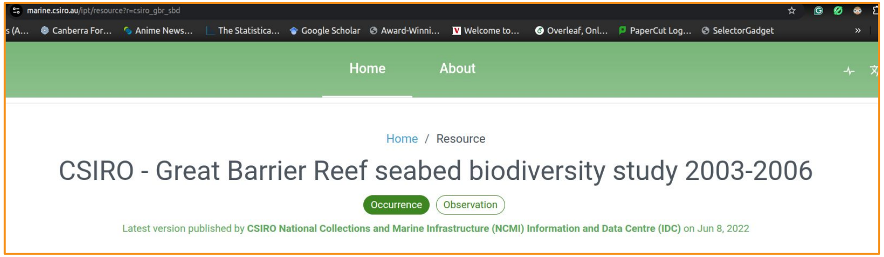
Community ecology data

- Multiple responses (potentially high-dimensional)
- Non-continuous responses with evident mean-variance relationship
- May have other data structures, but we will not worry about that in this presentation



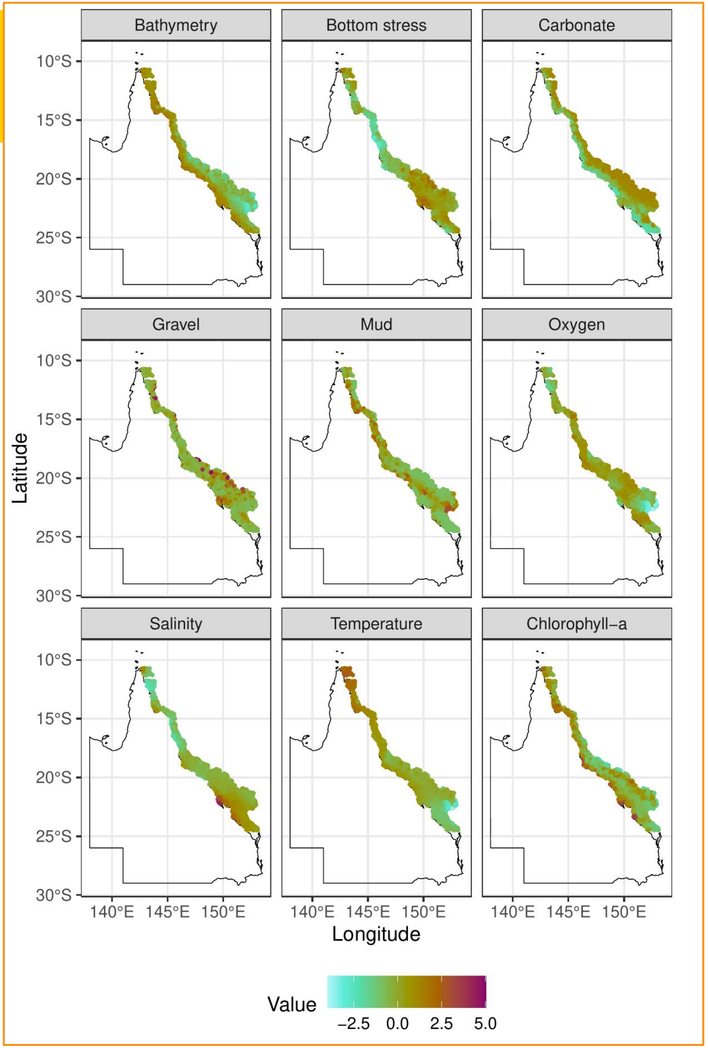
Community ecology data

- Great Barrier Reef Seabed biodiversity project



- For the purposes of this talk, we have:
  - Presence–absence (binary) responses
  - N = 1146 sites sampled
  - J = 235 species (median recorded prevalence = 31)

- For the purposes of this talk, we have:
  - Nine continuous environmental covariates
  - Standardized all to have mean zero and variance one



Species archetype models (SAMs)

- Aim: To understand how the assemblage distribution varies as a function of environment

- A starting point is to fit a stacked model e.g., separate binary logistic regression models to each species

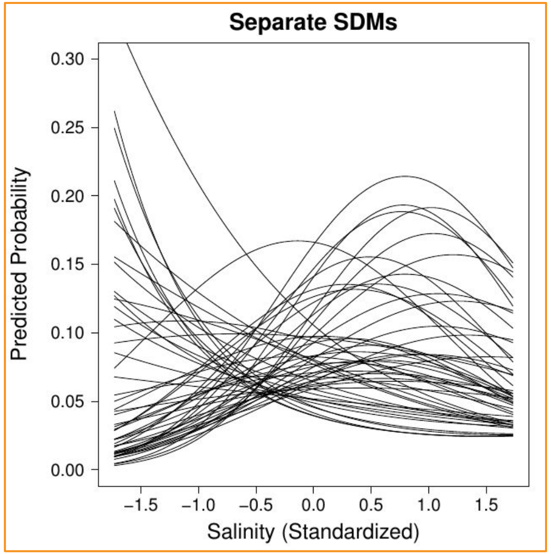
Consider a set of species $j = 1, \ldots, J$ recorded at a set of observational units $i = 1, \ldots, N$, along with measured covariates $\boldsymbol{x}_i$. Then a stacked model is characterized by

$$
\begin{aligned}
g(\mu_{ij}) &= \eta_{ij} = \boldsymbol{x}_i^\top \boldsymbol{\beta}_j \\
[y_{ij}] &= \text{Exp-Fam}(\mu_{ij}, \boldsymbol{\phi}_j) \\
\ell(\boldsymbol{\Psi}) &= \sum_{j=1}^{J} \left(\sum_{i=1}^{N} \log f(y_{ij} | \mu_{ij}, \boldsymbol{\phi}_j) \right)
\end{aligned}
$$

Species archetype models (SAMs)

- Aim: To understand how the assemblage distribution varies as a function of environment

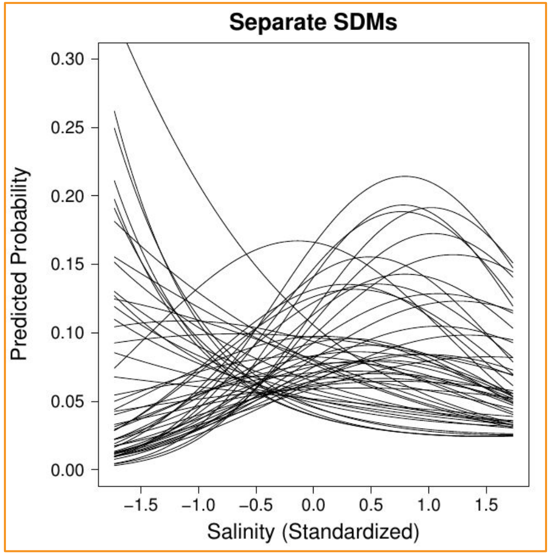
- A starting point is to fit a stacked model e.g., separate binary logistic regression models to each species

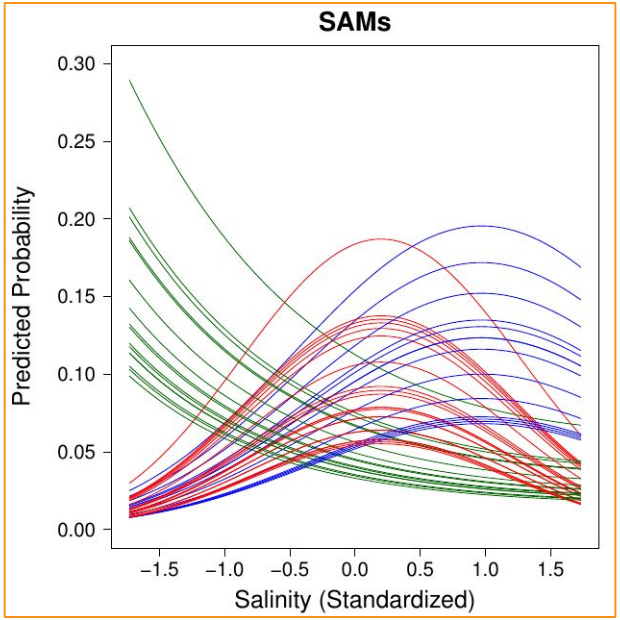


Species archetype models (SAMs)

- Aim: To understand how the assemblage distribution varies as a function of environment

- A starting point is to fit a stacked model e.g., separate binary logistic regression models to each species





Species archetype models (SAMs)

- Aim: To understand how the assemblage distribution varies as a function of environment
- Cluster species with similar environmental responses into so-called archetypal responses

Consider a set of species $j = 1, \ldots, J$ recorded at a set of observational units $i = 1, \ldots, N$, along with covariates $\boldsymbol{x}_i$. Let $z_{jk} = 1$ if species $j = 1, \ldots, J$ belongs to archetype $k = 1, \ldots, K$ and zero otherwise. If we denote $\mu_{ijk} = \mathsf{E}(y_{ij}|\boldsymbol{x}_i, z_{jk} = 1)$, then a species archetypal model or SAM is characterized by

$$
\begin{aligned}
g(\mu_{ijk}) &= \alpha_j + \boldsymbol{x}_i^\top \boldsymbol{\beta}_k \\
[y_{ij}|z_{jk} = 1] &= \text{Exp-Fam}(\mu_{ijk}, \boldsymbol{\phi}_j) \\
\ell(\boldsymbol{\Psi}) &= \sum_{j=1}^{J} \log \left\{ \sum_{k=1}^{K} \omega_k \prod_{i=1}^{N} f(y_{ij}|\mu_{ijk}, \boldsymbol{\phi}_j) \right\}
\end{aligned}
$$

Species archetype models (SAMs)

- Aim: To understand how the assemblage distribution varies as a function of environment

- Cluster species with similar environmental responses into so-called archetypal responses
    - A "partial" finite mixture of regression models

Consider a set of species $j = 1, \ldots, J$ recorded at a set of observational units $i = 1, \ldots, N$, along with covariates $\boldsymbol{x}_i$. Let $z_{jk} = 1$ if species $j = 1, \ldots, J$ belongs to archetype $k = 1, \ldots, K$ and zero otherwise. If we denote $\mu_{ijk} = \mathsf{E}(y_{ij}|\boldsymbol{x}_i, z_{jk} = 1)$, then a species archetypal model or SAM is characterized by

$$
\begin{aligned}
g(\mu_{ijk}) &= \alpha_j + \boldsymbol{x}_i^\top \boldsymbol{\beta}_k \\
[y_{ij}|z_{jk} = 1] &= \text{Exp-Fam}(\mu_{ijk}, \boldsymbol{\phi}_j) \\
\ell(\boldsymbol{\Psi}) &= \sum_{j=1}^{J} \log \left\{ \sum_{k=1}^{K} \omega_k \prod_{i=1}^{N} f(y_{ij}|\mu_{ijk}, \boldsymbol{\phi}_j) \right\}
\end{aligned}
$$

Species archetype models (SAMs)

- Unlike standard mixture models, partial finite mixture of regression models are more computationally burdensome to fit
  - Species-specific intercepts/dispersion parameters often done separately in a conditional maximization step (ECM)
  - Or update all parameters in a single M-step. This requires a single, large memory GLM which may not scale well with N and J (and K)

Consider a set of species $j = 1, \ldots, J$ recorded at a set of observational units $i = 1, \ldots, N$, along with covariates $\boldsymbol{x}_i$. Let $z_{jk} = 1$ if species $j = 1, \ldots, J$ belongs to archetype $k = 1, \ldots, K$ and zero otherwise. If we denote $\mu_{ijk} = \mathsf{E}(y_{ij} | \boldsymbol{x}_i, z_{jk} = 1)$, then a species archetypal model or SAM is characterized by

$$
\begin{aligned}
g(\mu_{ijk}) &= \alpha_j + \boldsymbol{x}_i^\top \boldsymbol{\beta}_k \\
[y_{ij} | z_{jk} = 1] &= \text{Exp-Fam}(\mu_{ijk}, \boldsymbol{\phi}_j) \\
\ell(\boldsymbol{\Psi}) &= \sum_{j=1}^{J} \log \left\{ \sum_{k=1}^{K} \omega_k \prod_{i=1}^{N} f(y_{ij} | \mu_{ijk}, \boldsymbol{\phi}_j) \right\}
\end{aligned}
$$

Approximate and scalable SAMs (asSAMs)

- Approximate each species-archetype contribution with a quadratic/normal approximation
  - Note the maximizer is species-specific!

For species $j = 1, \dots, J$, write $\boldsymbol{\theta}_{jk} = (\alpha_j^\top, \boldsymbol{\beta}_k^\top, \boldsymbol{\phi}_j^\top)^\top$, and let $L_{jk}(\boldsymbol{\theta}_{jk}) = \prod_{i=1}^N f(y_{ij}|\mu_{ijk}, \boldsymbol{\phi}_j)$.

Define $\boxed{\tilde{\boldsymbol{\theta}}_j = (\tilde{\boldsymbol{\alpha}}_j, \tilde{\boldsymbol{\beta}}_j, \tilde{\boldsymbol{\phi}}_j)} = \arg\max_{\boldsymbol{\alpha}_j, \boldsymbol{\beta}_k, \boldsymbol{\phi}_j} \log\{L_{jk}(\boldsymbol{\theta}_{jk})\}$, and $\boldsymbol{I}(\tilde{\boldsymbol{\theta}}_j) = -\nabla^2 \log\{L_{jk}(\tilde{\boldsymbol{\theta}}_j)\}$.

Approximate and scalable SAMs (asSAMs)

- Approximate each species-archetype contribution with a quadratic/normal approximation
  - Note the maximizer is species-specific!

For species $j = 1, \ldots, J$, write $\boldsymbol{\theta}_{jk} = (\alpha_j^\top, \boldsymbol{\beta}_k^\top, \boldsymbol{\phi}_j^\top)^\top$, and let $L_{jk}(\boldsymbol{\theta}_{jk}) = \prod_{i=1}^N f(y_{ij}|\mu_{ijk}, \boldsymbol{\phi}_j)$.

Define $\tilde{\boldsymbol{\theta}}_j = (\tilde{\boldsymbol{\alpha}}_j, \tilde{\boldsymbol{\beta}}_j, \tilde{\boldsymbol{\phi}}_j) = \arg\max_{\boldsymbol{\alpha}_j, \boldsymbol{\beta}_k, \boldsymbol{\phi}_j} \log\{L_{jk}(\boldsymbol{\theta}_{jk})\}$, and $\boldsymbol{I}(\tilde{\boldsymbol{\theta}}_j) = -\nabla^2 \log\{L_{jk}(\tilde{\boldsymbol{\theta}}_j)\}$. Then consider the quadratic approximation:

$$\begin{aligned}
\log\{L_{jk}(\boldsymbol{\theta}_{jk})\} &= \sum_{i=1}^N \log\{f(y_{ij}|\mu_{ijk}, \boldsymbol{\phi}_j)\} \\
&\approx \sum_{i=1}^N \log\{f(y_{ij}|\tilde{\mu}_{ij}, \tilde{\boldsymbol{\phi}}_j)\} - \frac{1}{2}\left(\boldsymbol{\theta}_{jk} - \tilde{\boldsymbol{\theta}}_j\right)^\top \boldsymbol{I}(\tilde{\boldsymbol{\theta}}_j)\left(\boldsymbol{\theta}_{jk} - \tilde{\boldsymbol{\theta}}_j\right),
\end{aligned}$$

where $\tilde{\mu}_{ij} = g^{-1}(\boldsymbol{u}_i^\top \tilde{\boldsymbol{\alpha}}_j + \boldsymbol{x}_i^\top \tilde{\boldsymbol{\beta}}_j)$.

Approximate and scalable SAMs (asSAMs)

- Approximate each species-archetype contribution with a quadratic/normal approximation
  - Note the maximizer is species-specific!

- After some algebraic manipulation and collecting terms that constant wrt parameters

$$
\begin{aligned}
\ell(\boldsymbol{\Psi}, \boldsymbol{\omega}) &= \sum_{j=1}^{J} \log \left(\sum_{k=1}^{K} \omega_k \exp \left[\log\{ L_{jk}(\boldsymbol{\theta}_{jk}) \} \right] \right) \\
&\approx C_0 + \sum_{j=1}^{J} \log \left[\sum_{k=1}^{K} \omega_k \exp \left\{ -\frac{1}{2} \left(\tilde{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_{jk} \right)^\top \boldsymbol{I}(\tilde{\boldsymbol{\theta}}_j) \left(\tilde{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_{jk} \right) \right\} \right] \\
&= C_1 + \sum_{j=1}^{J} \log \left[\sum_{k=1}^{K} \omega_k \mathcal{N} \left\{ \tilde{\boldsymbol{\theta}}_j | \boldsymbol{\theta}_{jk}, \boldsymbol{I}(\tilde{\boldsymbol{\theta}}_j)^{-1} \right\} \right] \triangleq \ell_{\text{assam}}(\boldsymbol{\Psi}, \boldsymbol{\omega}), \qquad (1)
\end{aligned}
$$

Approximate and scalable SAMs (asSAMs)

- asSAM = finite mixture of multivariate normals with known covariance matrices

- asSAMs is *very* amenable to using EM-algorithm
  - M-step updates are all closed-form (details in manuscript)
  - Need a pre-step to form the quadratic/normal approximation, but we know how to do this!

- Model selection is easy/scalable
  - Choose K using BIC or some variation thereof
  - Archetypal (mixture) coefficients: Deploy sparse linear modelling ideas e.g., LASSO, SCAD, BAR etc...

$$\ell_{\text{assam}}(\boldsymbol{\Psi}, \boldsymbol{\omega}) = \sum_{j=1}^{J} \log \left[\sum_{k=1}^{K} \omega_k \mathcal{N} \left\{ \tilde{\boldsymbol{\theta}}_j | \boldsymbol{\theta}_{jk}, \boldsymbol{I}(\tilde{\boldsymbol{\theta}}_j)^{-1} \right\} \right]$$

Application to Great Barrier Reef dataset

- Great Barrier Reef Seabed biodiversity project

- Recall:
    - Presence-absence (binary) responses
    - N = 1146 sites sampled
    - J = 235 species (median recorded prevalence = 31)

    - Nine continuous environmental covariates
    - Standardized all to have mean zero and variance one

- asSAMs application
    - All covariates included as linear terms only
    - Bernoulli distribution with logit link
    - BIC to choose K; BAR penalty to perform selection on archetypal coefficients (penalized asSAMs or pasSAMS)

ELSEVIER

Journal of Multivariate Analysis

Volume 168, November 2018, Pages 334-351

Broken adaptive ridge regression and its asymptotic properties

Linlin Dai [a], Kani Chen [b], Zhihua Sun [c], Zhenqiu Liu [d], Gang Li [e]

Show more

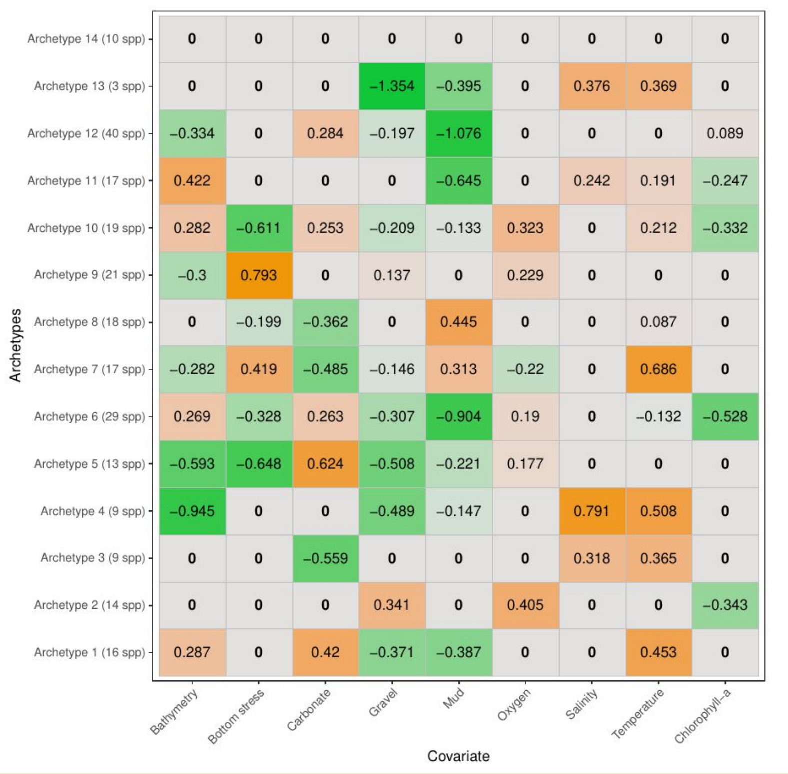
Add to Mendeley Share Cite

https://doi.org/10.1016/j.jmva.2018.08.007 Get rights and content

Under an Elsevier user license Open archive

Application to Great Barrier Reef dataset

- Great Barrier Reef Seabed biodiversity project
- asSAMs application
  - K = 14 species archetypes chosen
  - All covariates important
  - Environment-agnostic archetype
  - Most species classified with high probability; this is typical of SAMs

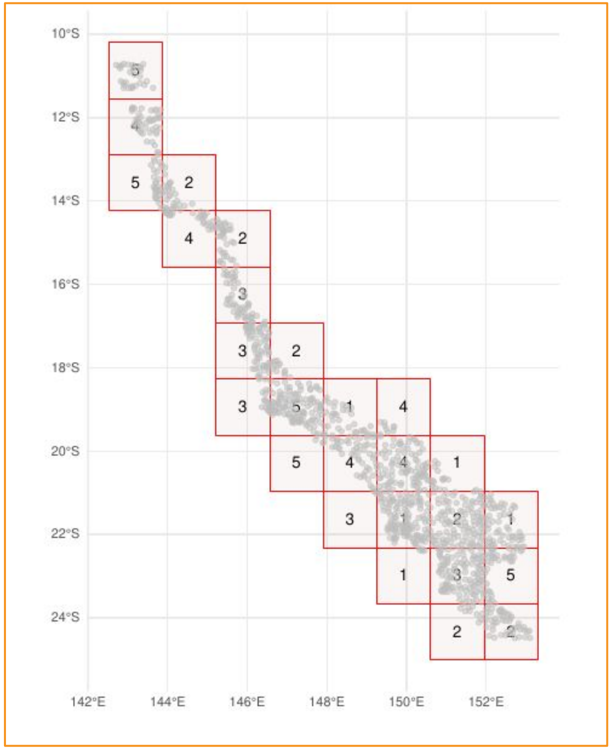


Application to Great Barrier Reef dataset

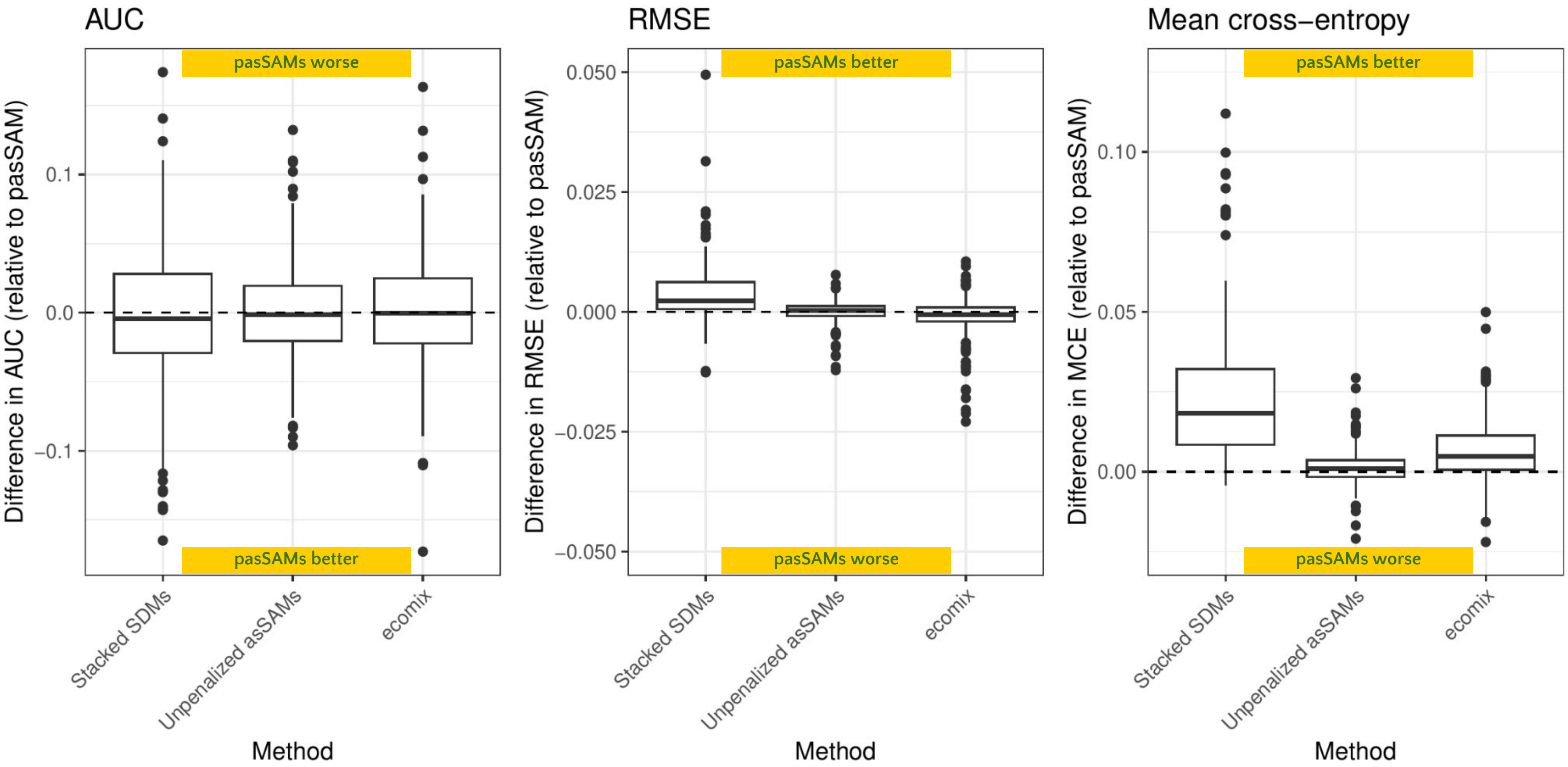
- Great Barrier Reef Seabed biodiversity project

- asSAMs application
  - K = 14 species archetypes chosen
  - All covariates important
  - Environment-agnostic archetype
  - Most species classified with high probability; this is typical of SAMs

- 5-fold spatial cross-validation to compare predictions
  - Separate logistic regression models
  - Penalized asSAMs or pasSAMs
  - Unpenalized asSAMs i.e., no selection on archetypal coefficients
  - SAMs fitted `ecomix` (no approximations)



Application to Great Barrier Reef dataset



Application to Great Barrier Reef dataset

- Great Barrier Reef Seabed biodiversity project

- asSAMs application
  - K = 14 species archetypes chosen
  - All covariates important
  - Environment-agnostic archetype
  - Most species classified with relatively high probability

- 5-fold spatial cross-validation to compare predictions
  - Separate logistic regression models
  - Penalized asSAMs or pasSAMs
    - ~5.7 mins per fold
  - Unpenalized asSAMs i.e., no selection on archetypal coefficients
    - ~42 seconds per fold
  - SAMs fitted `ecomix` (no approximations)
    - ~56 mins per fold

Concluding remarks

- Manuscript in review; https://github.com/fhui28/assam

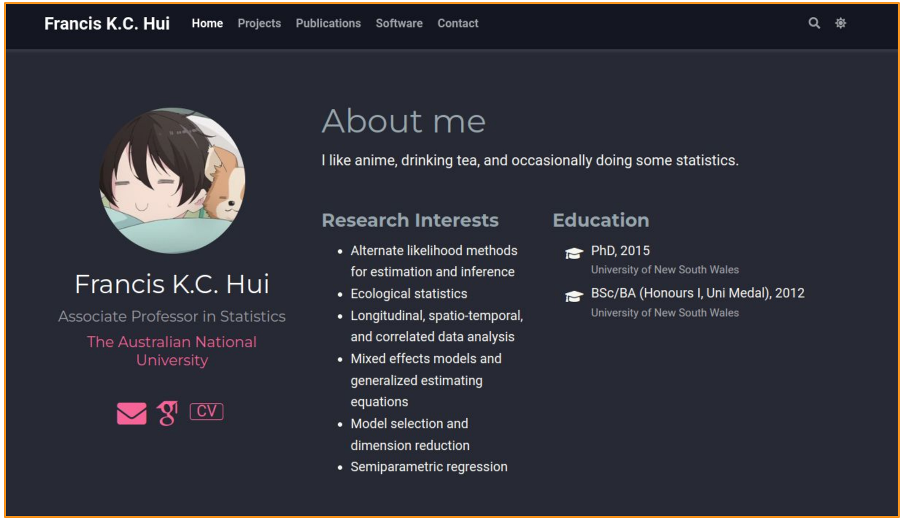
- The package allows:
  - A number of response types
  - Fast approximate bootstrap for uncertainty quantification
  - Specific-specific effects besides intercepts e.g., sampling effort, survey effect
  - Specific-specific spatial fields

- Future extensions to semi-parametric/ML-based archetypal responses

Thanks for listening!

Questions?

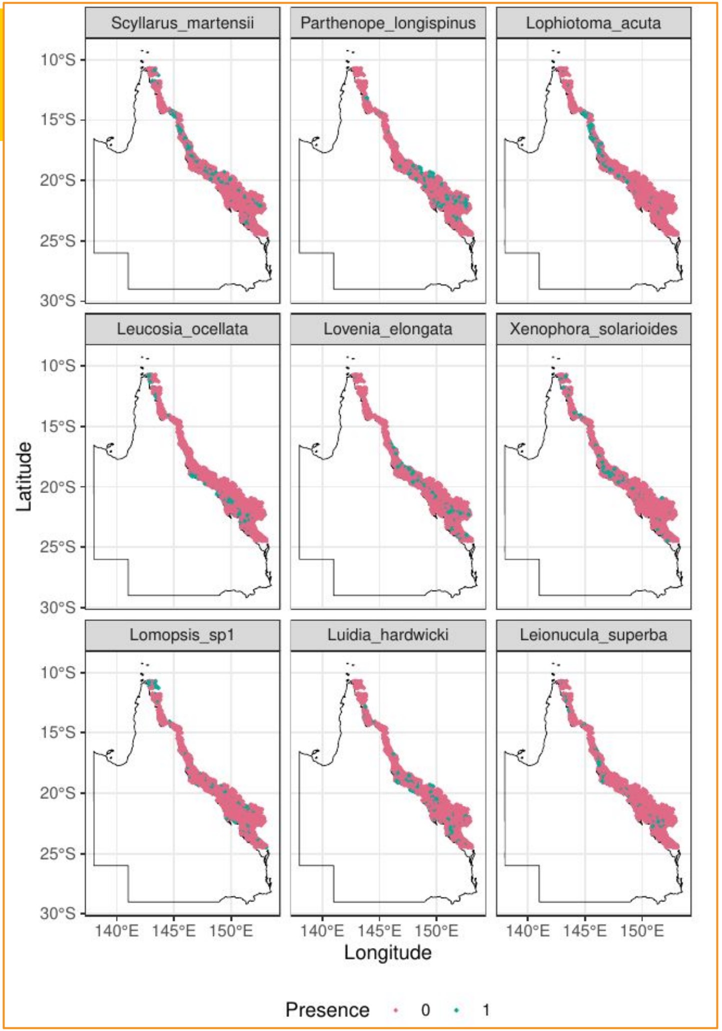
- francis.hui@anu.edu.au
- https://francishui.netlify.app/



Community ecology data

- Great Barrier Reef Seabed biodiversity project

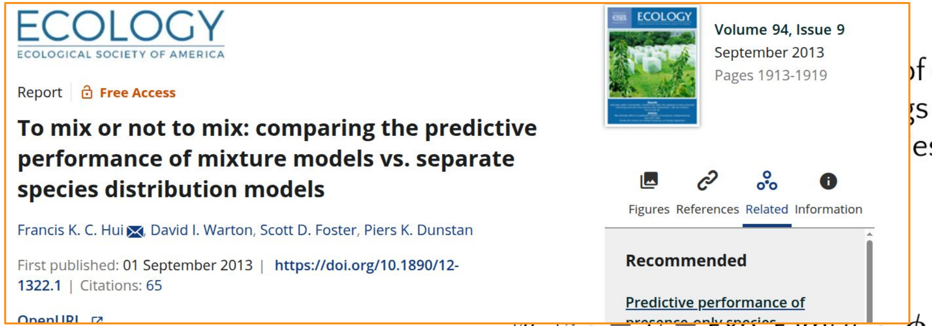
- For the purposes of this talk, we have:
    - Presence-absence (binary) responses
    - N = 1146 sites sampled
    - J = 235 species (median recorded prevalence = 31)



Species archetype models (SAMs)

- Aim: To understand how each species' distribution varies as a function of environment

- Cluster species with similar environmental responses into so-called archetypal responses
  - Simpler interpretation and easier to deploy for ecologist/policy makers
  - Borrow strength across species


ECOLOGY
ECOLOGICAL SOCIETY OF AMERICA

Volume 94, Issue 9
September 2013
Pages 1913-1919

Report | Free Access

To mix or not to mix: comparing the predictive performance of mixture models vs. separate species distribution models

Francis K. C. Hui, David I. Warton, Scott D. Foster, Piers K. Dunstan

First published: 01 September 2013 | https://doi.org/10.1890/12-1322.1 | Citations: 65

Figures References Related Information

Recommended

Predictive performance of


Finite Mixture of Regression Modeling for High-Dimensional Count and Biomass Data in Ecology

Published: 16 May 2013

Volume 18, pages 357–375, (2013) Cite this article

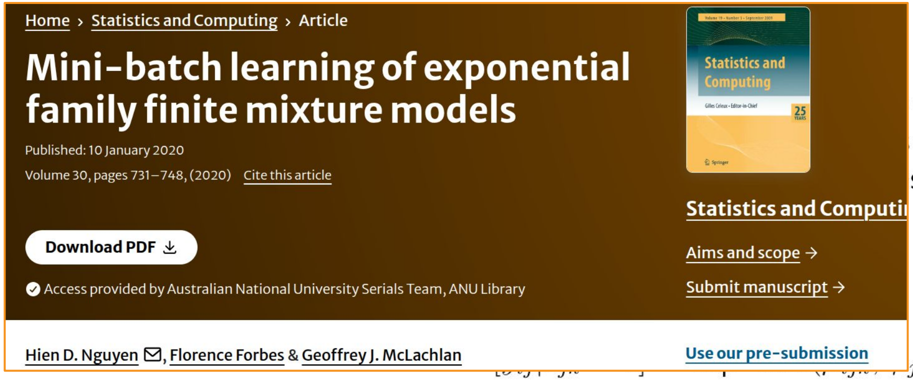
Journal of Agricultural, Biological, and Environ

$$[y_{ij}|z_{jk}=1] = \text{Exp-Fam}(\mu_{ijk}, \phi_j)$$

$$\ell(\Psi) = \sum_{j=1}^{J} \log\left\{\sum_{k=1}^{K} \omega_k \prod_{i=1}^{N} f(y_{ij}|\mu_{ijk}, \phi_j)\right\}$$

Species archetype models (SAMs)

- Unlike standard mixture models, partial finite mixture of regression models are more computationally burdensome to fit
  - Some ways to get around this, but not easy to generalize if we have random effects, smooth covariate terms etc…



$$\ell(\Psi) = \sum_{j=1}^{J} \log\left\{\sum_{k=1}^{K} \omega_k \prod_{i=1}^{N} f(y_{ij}|\mu_{ijk}, \phi_j)\right\}$$

Species archetype models (SAMs)

- Unlike standard mixture models, partial finite mixture of regression models are more computationally burdensome to fit
  - Some ways to get around this, but not easy to generalize if we have random effects, smooth covariate terms etc...

- Unlike (partial) finite mixture of regressions model in other settings, we have multiple observations per "object" we wish to cluster (N sites within each species)

Consider a set of species $j = 1, \dots, J$ recorded at a set of observational units $i = 1, \dots, N$, along with covariates $\boldsymbol{x}_i$. Let $z_{jk} = 1$ if species $j = 1, \dots, J$ belongs to archetype $k = 1, \dots, K$ and zero otherwise. If we denote $\mu_{ijk} = \mathsf{E}(y_{ij}|\boldsymbol{x}_i, z_{jk} = 1)$, then a species archetypal model or SAM is characterized by

$$
\begin{aligned}
g(\mu_{ijk}) &= \alpha_j + \boldsymbol{x}_i^\top \boldsymbol{\beta}_k \\
[y_{ij}|z_{jk} = 1] &= \text{Exp-Fam}(\mu_{ijk}, \boldsymbol{\phi}_j) \\
\ell(\boldsymbol{\Psi}) &= \sum_{j=1}^{J} \log \left\{ \sum_{k=1}^{K} \omega_k \boxed{\prod_{i=1}^{N} f(y_{ij}|\mu_{ijk}, \boldsymbol{\phi}_j)} \right\}
\end{aligned}
$$

Species archetype models (SAMs)

- Aim: To understand how each species' distribution varies as a function of environment

- Cluster species with similar environmental responses into so-called archetypal responses
  - Simpler interpretation and easier to deploy for ecologist/policy makers
  - Borrow strength across species
  - A "partial" finite mixture of regression models

Consider a set of species $j = 1, \ldots, J$ recorded at a set of observational units $i = 1, \ldots, N$, along with covariates $\boldsymbol{x}_i$. Let $z_{jk} = 1$ if species $j = 1, \ldots, J$ belongs to archetype $k = 1, \ldots, K$ and zero otherwise. If we denote $\mu_{ijk} = \mathrm{E}(y_{ij}|\boldsymbol{x}_i, z_{jk} = 1)$, then a species archetypal model or SAM is characterized by

$$\begin{aligned} g(\mu_{ijk}) &= \alpha_j + \boldsymbol{x}_i^\top \boldsymbol{\beta}_k \\ [y_{ij}|z_{jk} = 1] &= \text{Exp-Fam}(\mu_{ijk}, \boldsymbol{\phi}_j) \\ \ell(\boldsymbol{\Psi}) &= \sum_{j=1}^{J} \log\left\{\sum_{k=1}^{K} \omega_k \prod_{i=1}^{N} f(y_{ij}|\mu_{ijk}, \boldsymbol{\phi}_j)\right\} \end{aligned}$$

Approximate and scalable SAMs (asSAMs)

- Inference via bootstrapping
  - Uncertainty due to making the quadratic/normal approximation
  - Uncertainty due to sampling variability given on the approximation

$$\ell_{\text{assam}}(\boldsymbol{\Psi}, \boldsymbol{\omega}) = \sum_{j=1}^{J} \log \left[\sum_{k=1}^{K} \omega_k \mathcal{N} \left\{ \tilde{\boldsymbol{\theta}}_j | \boldsymbol{\theta}_{jk}, \boldsymbol{I}(\tilde{\boldsymbol{\theta}}_j)^{-1} \right\} \right]$$

Approximate and scalable SAMs (asSAMs)

- Fast approximate bootstrap for asSAMs
  - ~~Uncertainty due to making the quadratic/normal approximation~~ (goes away with large N?)
  - Uncertainty due to sampling variability given the approximation (dominant source; goes away with large J and N?)

$$\ell_{\text{assam}}(\boldsymbol{\Psi}, \boldsymbol{\omega}) = \sum_{j=1}^{J} \log \left[\sum_{k=1}^{K} \omega_k \mathcal{N} \left\{ \tilde{\boldsymbol{\theta}}_j | \boldsymbol{\theta}_{jk}, \boldsymbol{I}(\tilde{\boldsymbol{\theta}}_j)^{-1} \right\} \right]$$

Approximate and scalable SAMs (asSAMs)

- Fast approximate bootstrap for asSAMs
  - Bootstrap confidence intervals for parameter estimates, fitted values, predictions follow

Given asSAM estimates $(\hat{\boldsymbol{\Psi}}^\top, \hat{\boldsymbol{\omega}}^\top)^\top$,

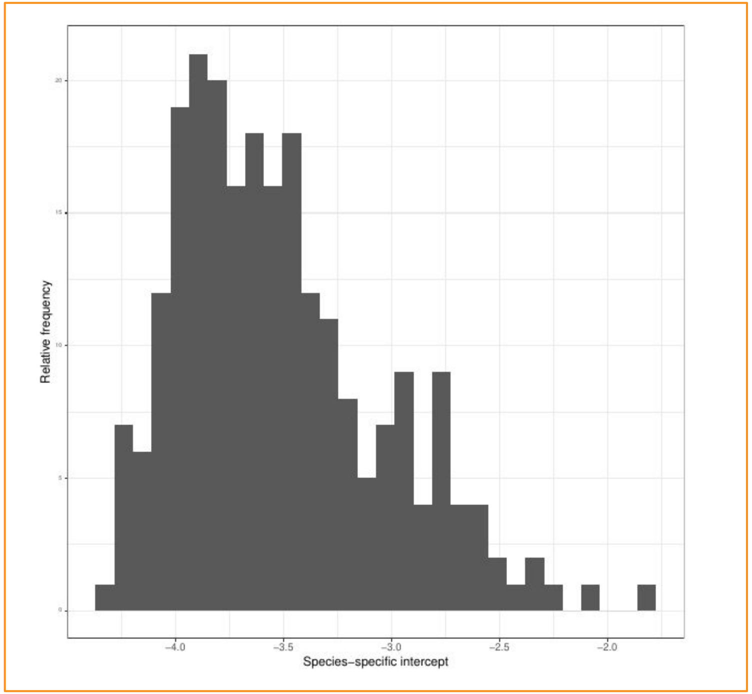
1. For species $j = 1, \ldots, J$, simulate component labels $\boldsymbol{z}_j^* = (z_{j1}^*, \ldots, z_{jK}^*)^\top$ from a multinomial distribution with trial size equal to one and probability vector $\hat{\boldsymbol{\omega}}$;
2. Conditional on $z_{jk}^* = 1$, simulate $\boldsymbol{\theta}_j^* = (\boldsymbol{\alpha}_j^{*\top}, \boldsymbol{\beta}_j^{*\top}, \boldsymbol{\phi}_j^{*\top})^\top$ from a multivariate normal distribution with mean vector $\hat{\boldsymbol{\theta}}_{jk} = (\hat{\boldsymbol{\alpha}}_j^\top, \hat{\boldsymbol{\beta}}_k^\top, \hat{\boldsymbol{\phi}}_j^\top)^\top$ and covariance matrix $\boldsymbol{I}(\tilde{\boldsymbol{\theta}}_j)^{-1}$;
3. Given bootstrap dataset $\{\boldsymbol{\theta}_j^*; j = 1, \ldots, J\}$, maximize $\ell_{\text{assam}}(\boldsymbol{\Psi}, \boldsymbol{\omega})$ and obtain bootstrap asSAM estimates $(\hat{\boldsymbol{\Psi}}_b^{*\top}, \hat{\boldsymbol{\omega}}_b^{*\top})^\top$.

$$\ell_{\text{assam}}(\boldsymbol{\Psi}, \boldsymbol{\omega}) = \sum_{j=1}^{J} \log \left[\sum_{k=1}^{K} \omega_k \mathcal{N} \left\{ \tilde{\boldsymbol{\theta}}_j | \boldsymbol{\theta}_{jk}, \boldsymbol{I}(\tilde{\boldsymbol{\theta}}_j)^{-1} \right\} \right]$$

Application to Great Barrier Reef dataset

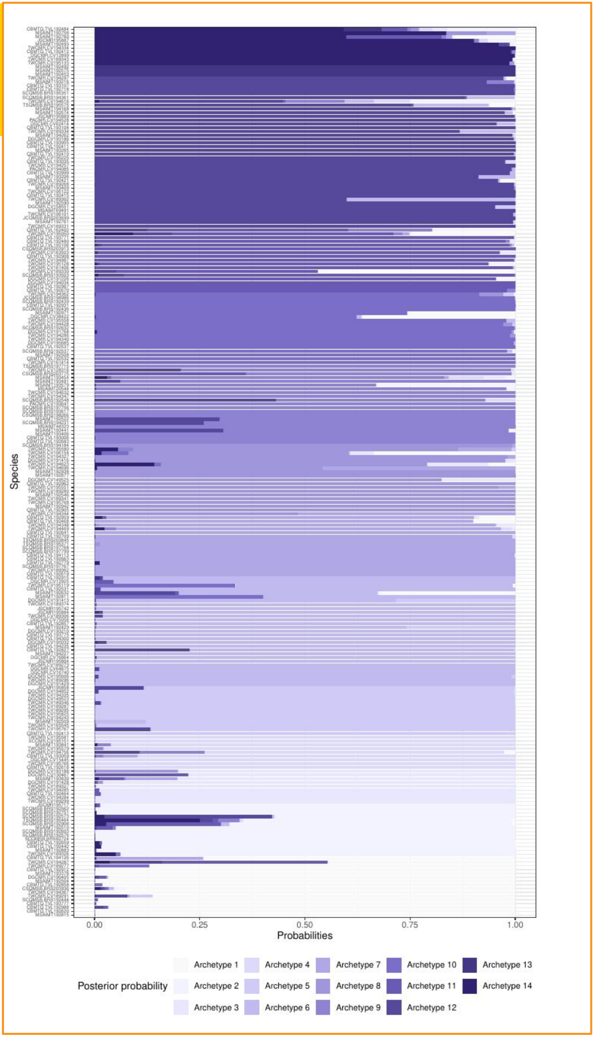
- Example: Great Barrier Seabed biodiversity project

- asSAMs application
  - K = 14 species archetypes chosen
  - All covariates important
  - Environment-agnostic archetype
  - Most species classified with relatively high probability; this is typical of (as)SAMs
  - The rarity of most species is clear!



Application to Great Barrier Reef dataset

- Great Barrier Reef Seabed biodiversity project
- asSAMs application
  - K = 14 species archetypes chosen
  - All covariates important
  - Environment-agnostic archetype
  - Most species classified with relatively high probability; this is typical of (as)SAMs



Concluding remarks

- Manuscript in preparation; https://github.com/fhui28/assam

- The package allows:
    - A number of response types
    - Specific-specific effects besides intercepts e.g., sampling effort, survey effect
    - Specific-specific spatial fields

- Extensions to semi-parametric/ML-based archetypa
    - Careful consideration of how to perform the quadratic
      issue here)
    - Spatially-varying effects/spatio-temporal asSAMs foll

- Hierarchical asSAMs?
    - Fit a finite mixture on the species-specific slopes rather than on the responses
    - Allows for heterogeneity within an archetype

Appendix

Algorithm 1 Computing asSAM estimates

Require: Multivariate abundance data $\{(y_{ij}, \boldsymbol{x}_i); i = 1, \ldots, N; j = 1, \ldots, J\}$; number of archetypes set to K; mapping matrix $\boldsymbol{M}$; tolerance value e.g., $\epsilon = 10^{-4}$.

1: Fit J separate generalized linear models, via parallel computing if possible. That is, for $j = 1, \ldots, J$, compute

$$\tilde{\boldsymbol{\theta}}_j = (\tilde{\boldsymbol{\alpha}}_j, \tilde{\boldsymbol{\beta}}_j, \tilde{\boldsymbol{\phi}}_j) = \arg \max_{\boldsymbol{\alpha}_j, \boldsymbol{\beta}_k, \boldsymbol{\phi}_j} \log\{L_{jk}(\boldsymbol{\theta}_{jk})\},$$

and also $\boldsymbol{I}(\tilde{\boldsymbol{\theta}}_j) = -\nabla^2 \log\{L_{jk}(\tilde{\boldsymbol{\theta}}_j)\}$.

2: Construct a set of initial values $(\hat{\boldsymbol{\Psi}}^{(0)\top}, \hat{\boldsymbol{\omega}}^{(0)\top})^\top$ from step 1 e.g., apply a K-medoids algorithm to the estimates $\{\tilde{\boldsymbol{\beta}}_j; j = 1, \ldots, J\}$ to obtain $(\hat{\boldsymbol{\beta}}_1^{(0)\top}, \ldots, \hat{\boldsymbol{\beta}}_K^{(0)\top})^\top$.

3: **for** $t = 1, 2 \ldots$ **do**

- *E-step:* Construct $\hat{\boldsymbol{\theta}}_{jk}^{(t)} = (\hat{\boldsymbol{\alpha}}_j^{(t)}, \hat{\boldsymbol{\beta}}_k^{(t)\top}, \hat{\boldsymbol{\phi}}_j^{(t)\top})^\top$, and compute the posterior probabilities

$$\hat{\tau}_{jk}^{(t+1)} = \frac{\hat{\omega}_k^{(t)} \mathcal{N}\{\tilde{\boldsymbol{\theta}}_j | \hat{\boldsymbol{\theta}}_{jk}^{(t)}, \boldsymbol{I}(\tilde{\boldsymbol{\theta}}_j)^{-1}\}}{\sum_{k'=1}^K \hat{\omega}_{k'}^{(t)} \mathcal{N}\{\tilde{\boldsymbol{\theta}}_j | \hat{\boldsymbol{\theta}}_{jk'}^{(t)}, \boldsymbol{I}(\tilde{\boldsymbol{\theta}}_j)^{-1}\}}; \; j = 1 \ldots, J, k = 1, \ldots, K$$

- *M-step:* Update the mixing proportions as $\hat{\omega}_k^{(t+1)} = J^{-1} \sum_{j=1}^J \hat{\tau}_{jk}^{(t+1)}$, and the remaining parameters as

$$\hat{\boldsymbol{\Psi}}^{(t+1)} = \left(\boldsymbol{M}^\top \boldsymbol{W}^{(t+1)} \boldsymbol{M}\right)^{-1} \boldsymbol{M}^\top \boldsymbol{W}^{(t+1)} \tilde{\boldsymbol{\Theta}},$$

where $\tilde{\boldsymbol{\Theta}} = (\boldsymbol{1}_K^\top \otimes \tilde{\boldsymbol{\theta}}_1^\top, \ldots, \boldsymbol{1}_K^\top \otimes \tilde{\boldsymbol{\theta}}_J^\top)^\top$ and $\boldsymbol{W}^{(t+1)}$ is a block-diagonal matrix where block $j = 1, \ldots, J$ equals $\text{Diag}(\hat{\boldsymbol{\tau}}_j^{(t+1)}) \otimes \boldsymbol{I}(\tilde{\boldsymbol{\theta}}_j)$.

until $|\ell_{\text{assam}}(\hat{\boldsymbol{\Psi}}^{(t+1)}, \hat{\boldsymbol{\omega}}^{(t+1)}) - \ell_{\text{assam}}(\hat{\boldsymbol{\Psi}}^{(t)}, \hat{\boldsymbol{\omega}}^{(t)})| < \epsilon$.

4: **end for**

5: **return** Estimates $(\hat{\boldsymbol{\Psi}}^\top, \hat{\boldsymbol{\omega}}^\top)^\top$ and posterior probabilities $\{\hat{\tau}_{jk}; j = 1, \ldots, j; k = 1, \ldots, K\}$.