MARINE BIODIVERSITY RESEARCH

Prediction and Management of Australia's Marine Biodiversity

Model based grouping of species across environmental gradients

Ross Darnell (CSIRO Mathematics, Informatics and Statistics)
Piers Dunstan (CSIRO Marine and Atmospheric Research)
Scott Foster (CSIRO Mathematics, Informatics and Statistics)

CERF Biodiversity Hub Prediction Project

Motivation

Predict *"bioversity"* using physical variables having measured individual species ($S \approx 200$) biomass at many sites ($N \approx 800$) on the Great Barrier Reef, along with many physical measurements ($P \approx 20$) taken at each site.

- **Frequently used methods to avoid**
  - aggregate species, then model or
  - model individual species and then aggregate.
- **Avoid taxanomic aggregation.**
- **Ignore sampling issues (don't miss Hideyasu's talk)**
- **Longitude and latitude NOT to be used as predictors**

Possible Solution

- **Define *"biodiversity"* as species presence/absence.**
- **Model presence using a reasonably sufficient function of covariates.**
- **Maximum level of heterogeneity among regression coefficients supported by data.**

The Model

The probability that species j is observed is

$$\Pr(Y_j = 1|g) \sim \text{Bernoulli}(\boldsymbol{\mu_g})$$

where $g = 1, \ldots, G$ and $\boldsymbol{\mu_g} = (\mu_{1g}, \ldots, \mu_{Ng})^T$ being the vector of probabilities at each site for the g-th species group, G being the number of groups.

A mixture of GLMs. (Wedel & DeSarbo, 1995)

The mixture terms

"g" is not known but

$$y_j = z_1\boldsymbol{\mu}_1 + z_2\boldsymbol{\mu}_2 + \ldots + z_G\boldsymbol{\mu}_g$$

where $\boldsymbol{z} = (z_1, \ldots, z_G)^T$ being an observation from a multinomial with mean probability vector $\boldsymbol{\pi}$.

The likelihood contribution for the jth species is

$$L_j(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_G, \boldsymbol{\pi}) = \sum_{g=1}^{G} \pi_g \prod_{i=1}^{N} \mathsf{Pr}(y_{ij}|p_g = 1)$$

$\mathsf{Pr}(y_{ij}|p_g = 1)$ is the probability of observing species jth at the ith site.

$\widehat{G}$?

Our “suggestions” (others welcome).

- **BIC (**$-2\ell(G) + p \log S$**)**
- Δ**BIC**
- **min** π

Estimated membership

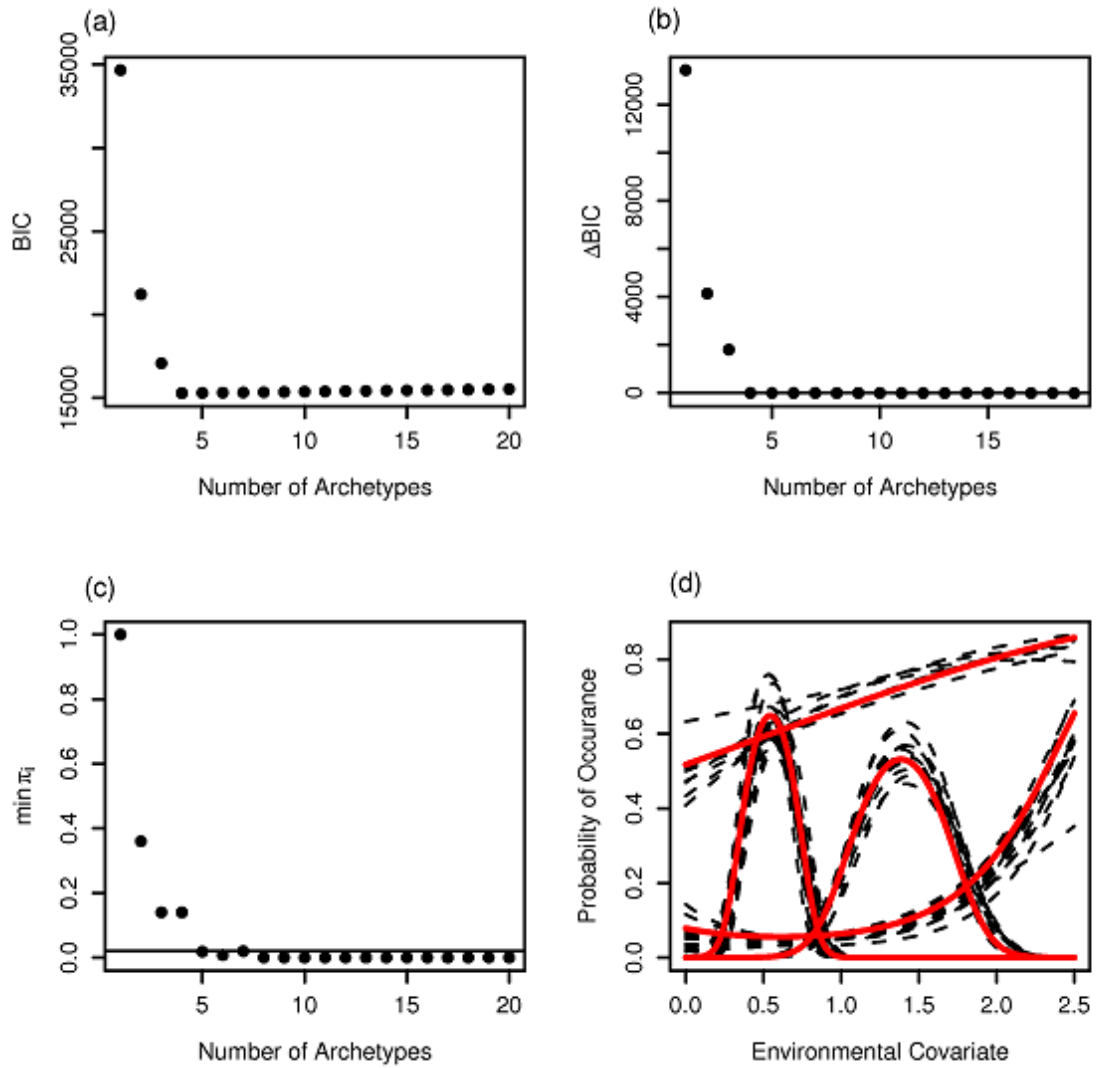
Posterior probability for j th species in the gth species-archetype

$$\tau_{jg} = \frac{\pi_g f(y_j; \beta_g)}{\sum_{k=1}^{G} \pi_k f(y_j; \beta_k)}$$

A toy example

- **Generate datasets where $S = 50$, $G = 4$, and $N = 100$.**
- **Single covariate.**
- **Fit intercept, linear and quadratic term.**

Mixture of GLMs



MARINE BIODIVERSITY RESEARCH
Prediction and Management of Australia's Marine Biodiversity

Coefficients for toy data ($G = 4\ p = 3$)

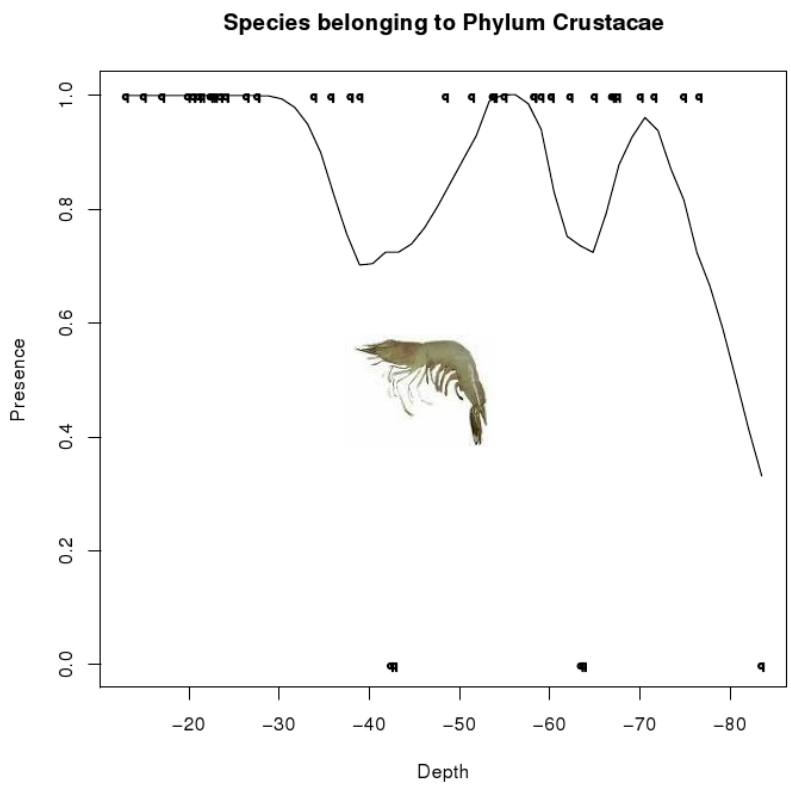
group	covariate	original value	estimated value	se
β_1	intercept	-9	-10.494	0.246
	covariate	35	39.415	0.370
	covariate^2	-32	-35.267	0.128
β_2	intercept	0	0.019	1.157
	covariate	0.7	0.677	1.660
	covariate^2	0	0.009	0.576
β_3	intercept	-16	-16.912	1.069
	covariate	23	24.196	4.094
	covariate^2	-8.2	-8.608	3.683
β_4	intercept	-3	-2.896	0.139
	covariate	-0.6	-1.021	0.267
	covariate^2	0.8	0.967	0.107

Recap — Mixtures of GLMs

- **Model based grouping of species for prediction across environmental gradients**
- **Simultaneous estimation of effects and species groups.**
- **A model is fitted to each species archetype simultaneously.**
- **The correct number of archetypes can be determined by comparing BIC with models with different G.**
- **Prediction is possible for each archetype.**
- **Currently implemented for presence/absence only.**

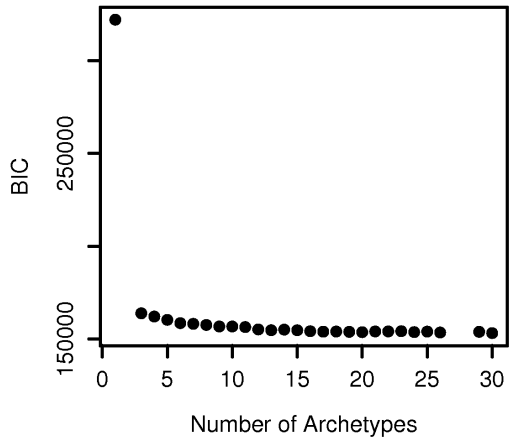
Archetypes on the GBR

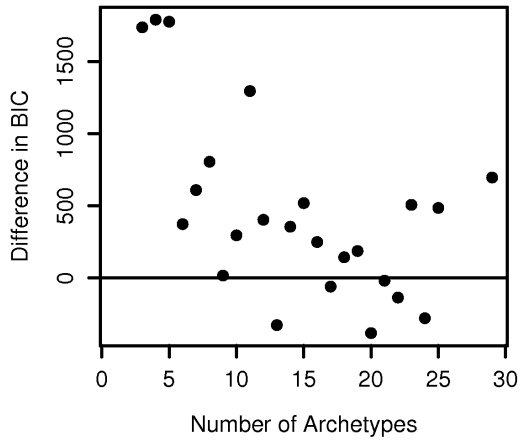
- **Apply the method to extensive data set from Great Barrier Reef (Pitcher et al.)**
- **Estimate number of archetypes**
- **Estimate model parameters for each archetype**
- **Predict probability of presence for archetypes across the GBR**
- **Each model is fitted with physical covariates**

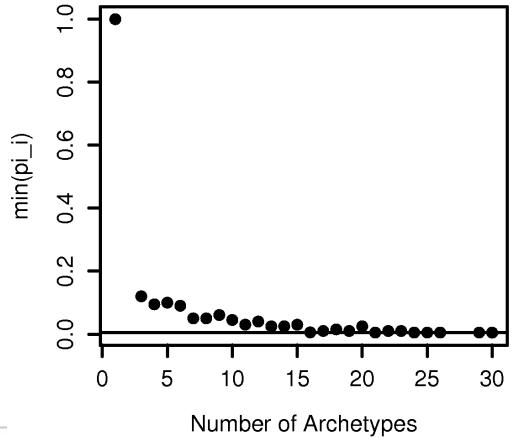


MARINE BIODIVERSITY RESEARCH
Prediction and Management of Australia's Marine Biodiversity

Selecting archetypes from the GBR

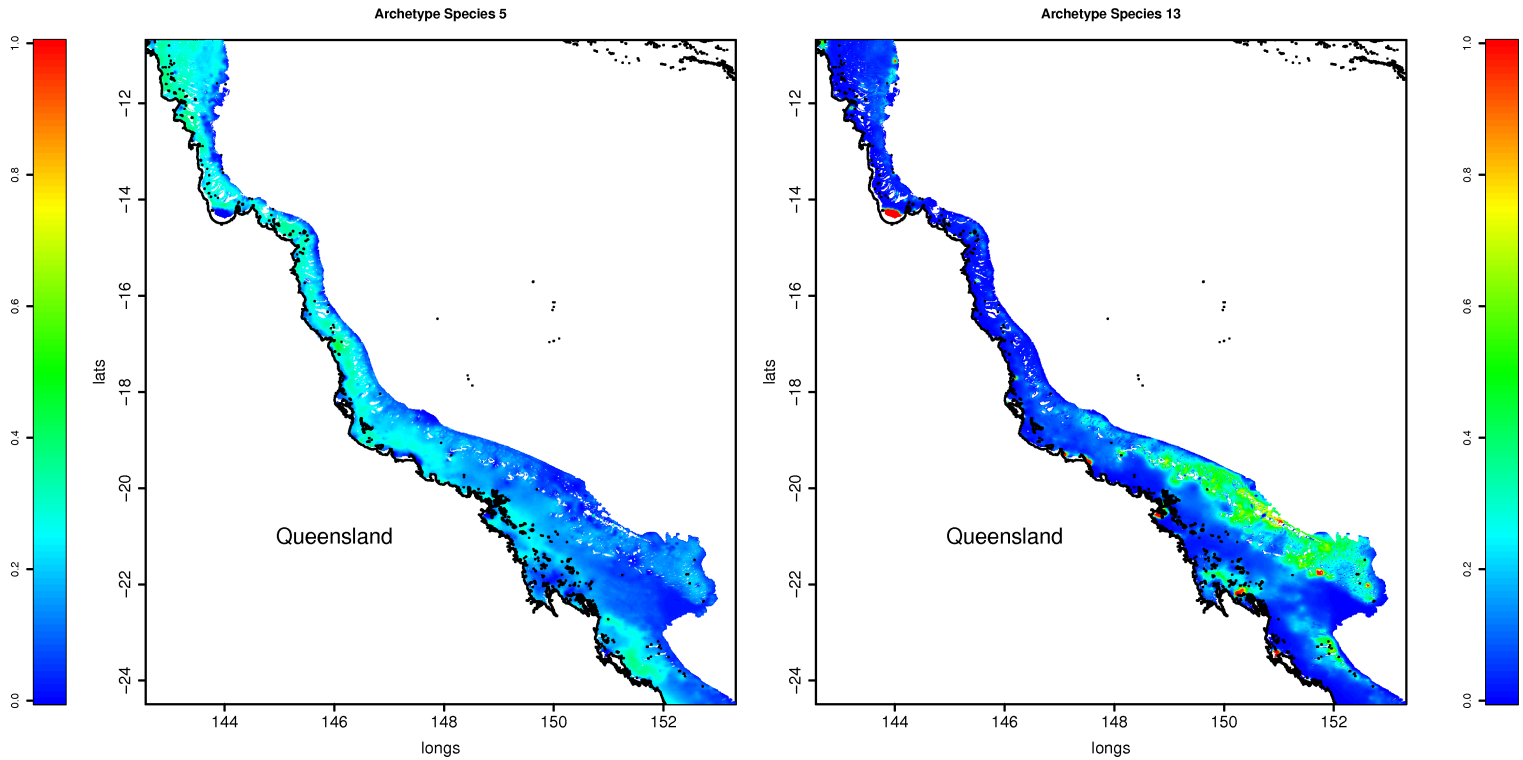






MARINE BIODIVERSITY RESEARCH
Prediction and Management of Australia's Marine Biodiversity

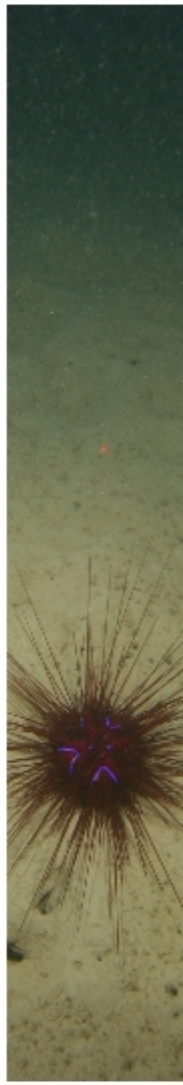
Archetypes 5 and 13



MARINE BIODIVERSITY RESEARCH
Prediction and Management of Australia's Marine Biodiversity

Archetype Species

- Each archetype represents many species
- Each archetype represents the response of its member species to the environment
- Prediction of archetype with error
- Captures both species with restricted distributions and ubiquitous species
- Each archetype can be used as a management unit



MARINE BIODIVERSITY RESEARCH
Prediction and Management of Australia's Marine Biodiversity

Funding and data were provided by the CERF Marine Biodiversity Hub. The ideas and methods described here were the result of discussions with other members of the CERF Prediction Project.

Australian Government

Geoscience Australia

Australian Government

AUSTRALIAN INSTITUTE OF MARINE SCIENCE

Australian Government

Department of the Environment, Water, Heritage and the Arts

MARINE BIODIVERSITY RESEARCH

Prediction and Management of Australia's Marine Biodiversity