MARINE BIODIVERSITY RESEARCH

Prediction and Management of Australia's Marine Biodiversity

RAD Biodiversity: Modelling many species' counts together

March 2009, Keio University

Scott Foster (CSIRO Mathematical and Information Science)

Piers Dunstan (CSIRO Marine and Atmospheric Research)

Outline

Voyage of Discovery

The Question

Our Approach
- RAD Description
- Modelling RADs
- Application to WA data
- Current Limitations

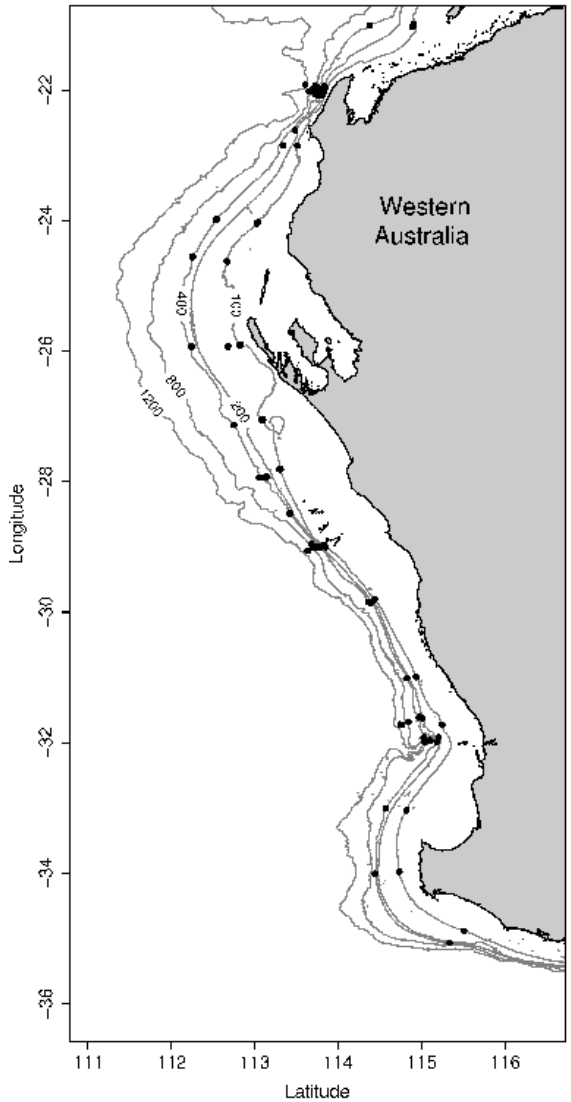
Conclusions

Voyage of Discovery

- Aim was to categorise biodiversity on lower shelf and upper slopes
- Study area was south west Australia
- 120 benthic samples taken
- 6 phyla used in this study
- Species counts generated by museum experts

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

Voyage of Discovery – sample locations



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

Voyage of Discovery – biology data

- **1548 species encountered**
  - 55% found at only one site
  - 89.7% found at 5 or less sites
  - Most abundant species found at only 25 sites
- **Very little information on each species!**
  - Almost excludes species based analysis

Why Look at These Data?

Why look at these data?

- **To investigate biodiversity – of course!**

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

Why look at these data?

- **To investigate biodiversity – of course!**
  - Very vague
  - Many possible definitions of biodiversity
  - Many aspects to biodiversity

Our Question

How does biodiversity change with the environment?

- **Do not want biodiversity based on species**
- **Do want biodiversity based on species observations**
- **Biodiversity indices are an option**
  - Do not seem to vary with covariates appropriately

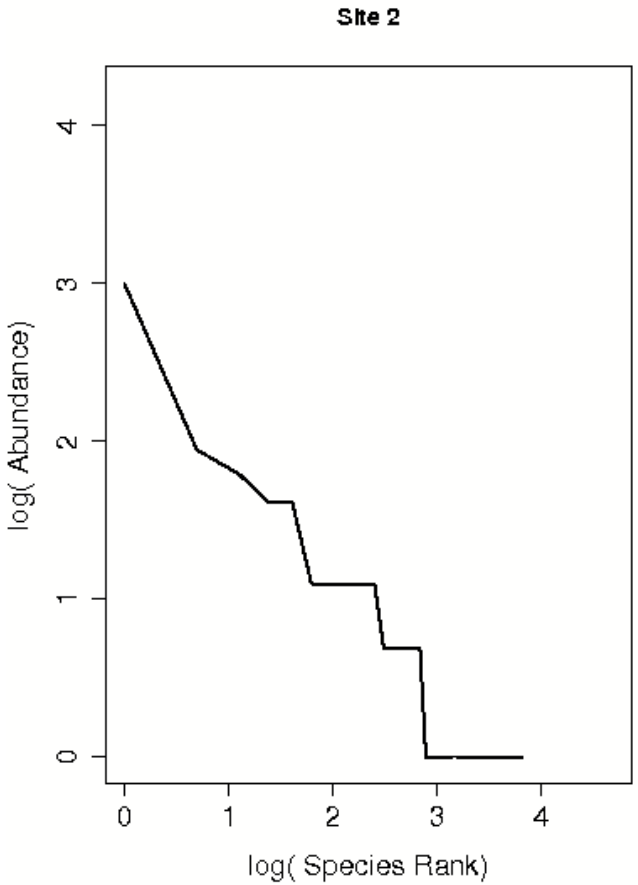
Our Approach

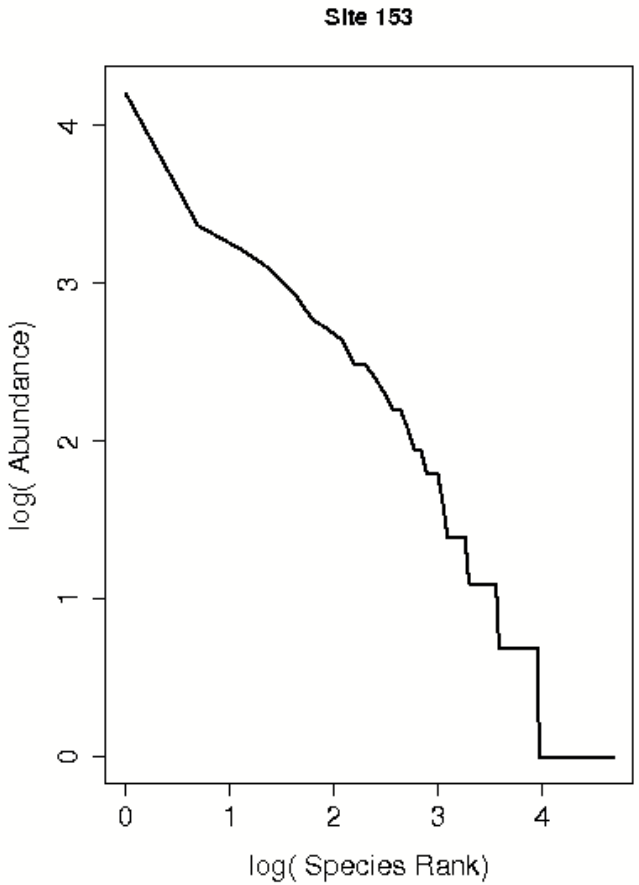
- **Model rank abundance distributions (RADs) of observed species counts**
  - Multivariate outcomes
- **Not species based**
- **Species counts preserved**
- **Allow RADs to change with the environment**

What is a RAD?

- **Rank Abundance Distribution**
- **Listing of the observed species counts from most abundant to least abundant**
- **Species labels are then discarded**
- **Not dependent on species identity – can compare different types of communities**
- **Multivariate observation**

RAD examples





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

Information in a RAD

- **RADs capture information on:**
  - The number of individuals (N)
  - The number of species (S), and
  - The relative abundance of those species ($\boldsymbol{n}$)
- **These are three attributes of a community that biodiversity if often defined by.**

Modelling goals

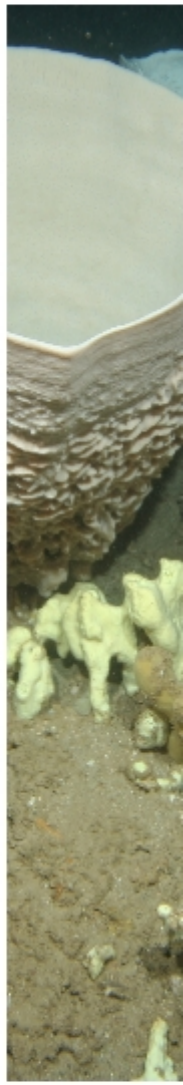
- **Model multivariate RAD observation as a function of environmental gradients**
- **Predict RADs and/or relevant measures of biodiversity with sensible measures of uncertainty**

The modelling approach – Broadly

- **The RAD is completely defined by $(S, \boldsymbol{n})$**
  - Equivalent to $(S, N, \boldsymbol{n})$ with sum constraint
  - Still a multivariate response
- **Model these aspects through**

$$\Pr(S, N, \boldsymbol{n}) = \Pr(N) \Pr(S|N) \Pr(\boldsymbol{n}|N, S)$$

- **Modelling task now decomposed into 3 separate tasks**
  - A model for abundance (univariate)
  - A model for conditional richness (univariate), and
  - A model for relative abundance (multivariate).



A Model for Total Abundance

- **Generalized linear models and similar**
- **Select covariates**
- **Diagnostics using randomised quantile residuals, *sans* randomisation**

A Model for Conditional Species Richness

- **Species richness is commonly analysed *marginally* to abundance via GLM etc**
- ***Conditionally* the statistical model should reflect the identity $S \leq N$**
- **We use a truncated Poisson or Negative Binomial model with log-likelihood**

$$\ell_S\left(\boldsymbol{\tau}_S, \theta_S; \boldsymbol{S} | \boldsymbol{N}\right) = \sum_{i=1}^{T} \left[\log\left\{\Pr\left(Y = S_i\right)\right\} - \log\left\{\Pr\left(Y \leq N_i\right)\right\}\right]$$

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

A Model for Conditional Species Richness

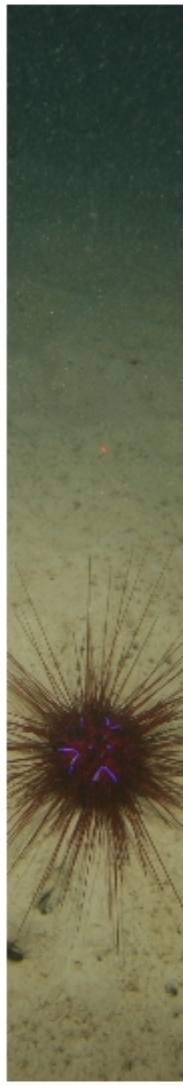
- **Estimation via maximisation of log-likelihood using numerical methods**
- **Model selection**
- **Diagnostics using quantile residuals**
- **Expectations available using brute-force**
- ***Marginal* species richness predictions via parametric bootstrap ('integrating' out total abundance)**

A Model for Conditional Relative Abundance

- **Condition N individuals into S categories (ranks)**
- **Could use multinomial framework but need to specify mean probabilities**
  - Decreasing function (due to ranking)
  - Many (many!) possible from theoretical ecology
  - We use a relation of the broken stick (niche pre-emption) as this empirically agreed with data

$$p_{ij} = \frac{1}{K} \exp\left(-\beta_i \log j\right)$$

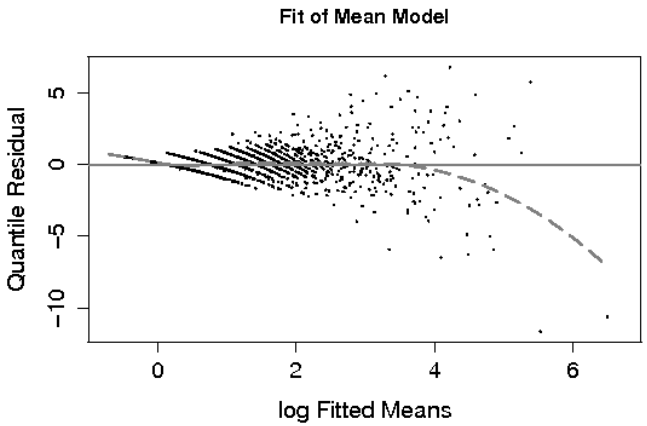
- **Defines all S_i probabilities with single parameter β_i**
- **Model β_i as a linear combination of environmental gradients**

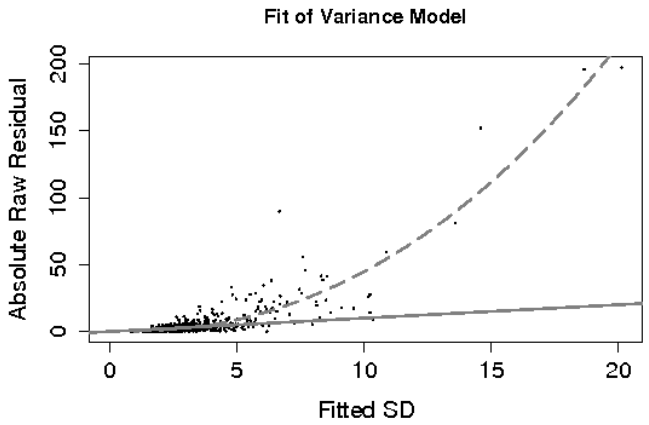


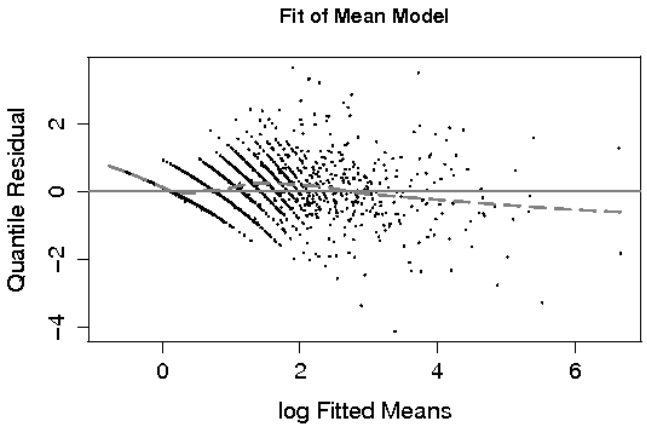
A Model for Conditional Relative Abundance

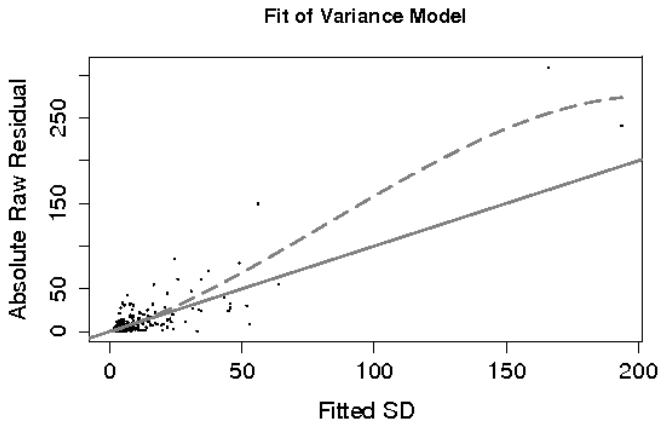
- **Model selection**
- **Quantile residuals used to inspect mean model**
- **Raw residuals used to inspect variance model (not standardised)**
- **Residuals not great for multinomial model (next slide)**
- ***Marginal* predictions available via parametric bootstrap**

Multinomial Model Residuals









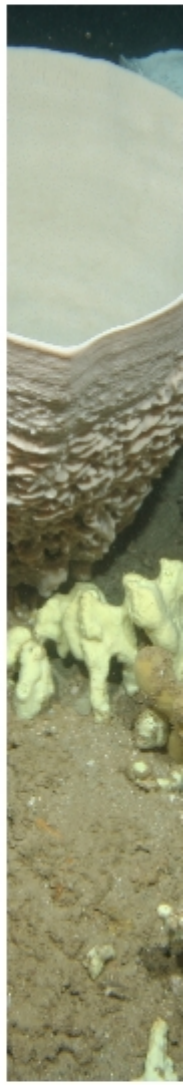
MARINE BIODIVERSITY RESEARCH
Prediction and Management of Australia's Marine Biodiversity

Extensions of Multinomial Model

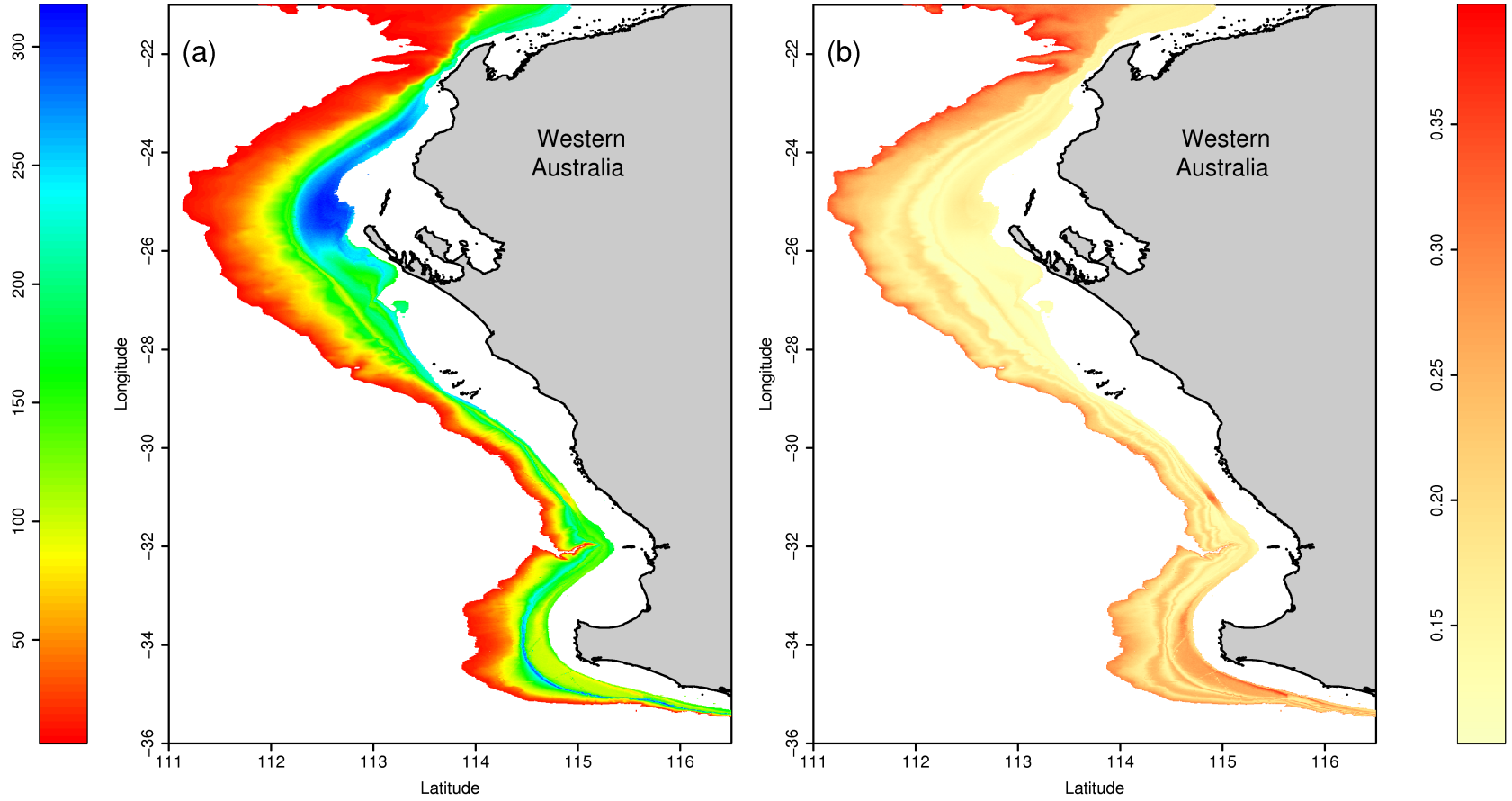
- **Dirichlet-multinomial (DM) provides constant over-dispersion**
  - Provides modest benefit to model fit
  - Not worth the effort
- **Modified DM (M-DM) obtained by modifying DM**
  - Alter the beta-binomial marginals of the DM so that over-dispersion is a decreasing function of rank
  - Some optimisation heart-ache but is now conquered
- **Residuals look much better (see previous slide)**

Biodiversity Measures from RAD Model

- **Many (all?) ecologists would consider the following base information:**
  - the amount of life
  - the variety of life (richness)
  - the manner in which communities are structures (e.g. evenness)
- **These are available through this model as**
  - Total abundance
  - *Marginal* Species richness
  - Derivative of probability function for *marginal* evenness
  - All other indices used can be derived from the information provided from the predicted RAD

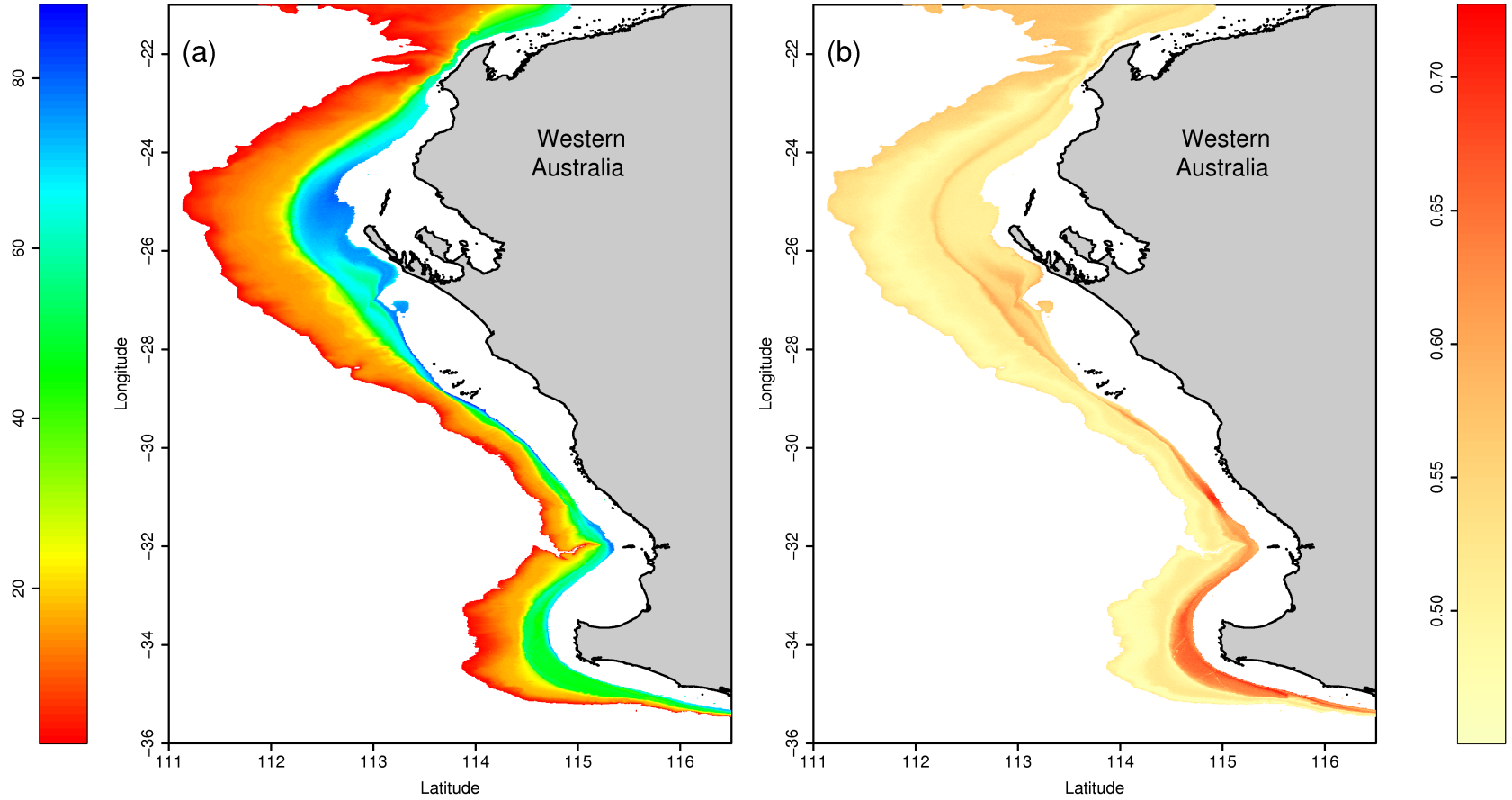


Voyage of Discovery Data – abundance



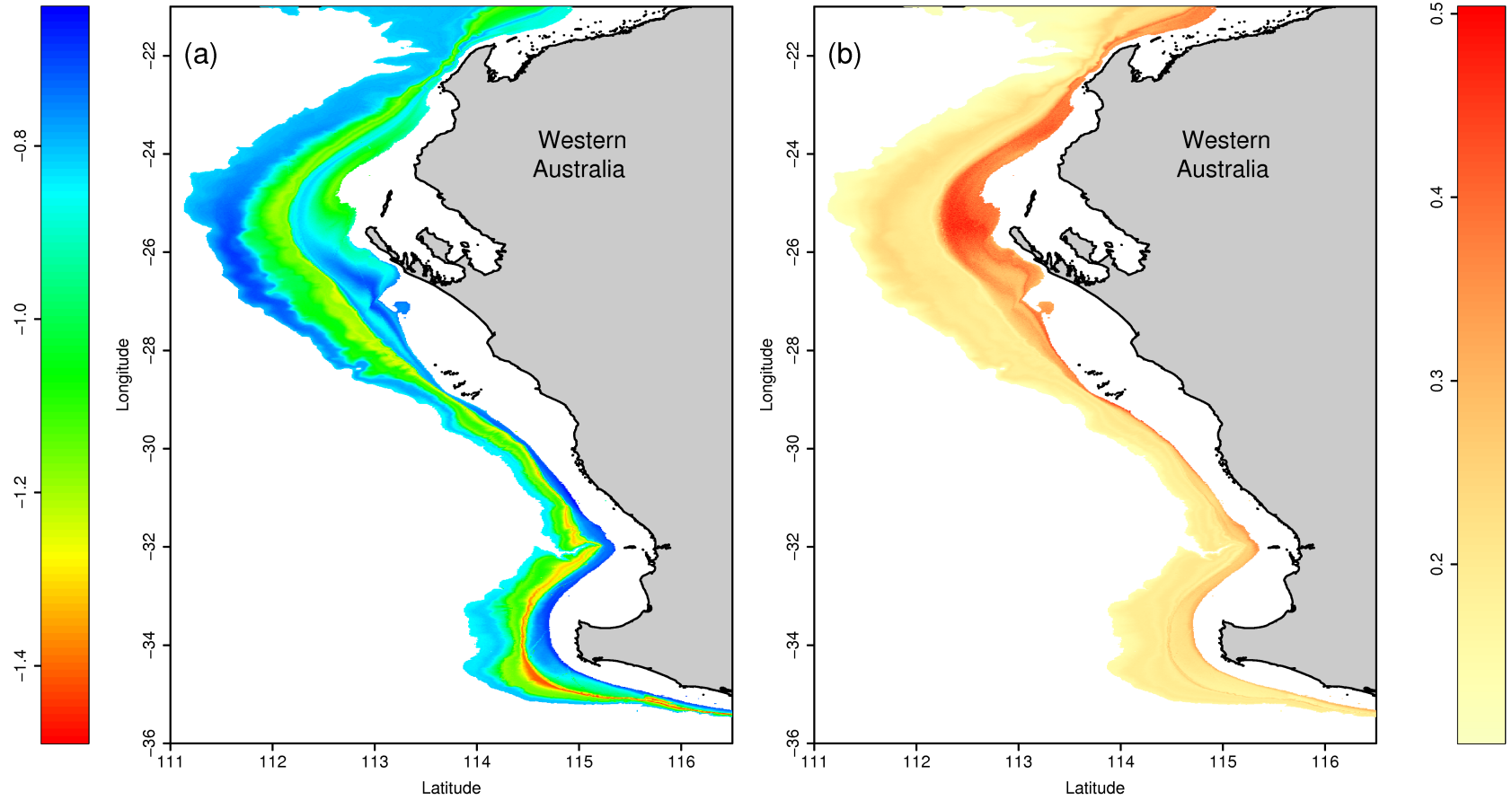
MARINE BIODIVERSITY RESEARCH
Prediction and Management of Australia's Marine Biodiversity

Voyage of Discovery Data – richness



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

Voyage of Discovery Data – evenness



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

Limitations and Future Needs

- **Biomass data**
- **Model for relative abundance is not quite right**
  - Likelihood doesn't guarantee decreasing observations
  - Zeros
  - Still useful
- **More theoretical work on the modified Dirichlet-multinomial**
- **Computing speed**
  - Estimation (automatic differentiation?)
  - Bootstrap predictions

Conclusions

- **We feel the RAD approach is a useful first step**
  - Models ecologically meaningful quantity
  - Indices predicted as simple summaries of models
- **Still require polishing and generalisation**

Conclusions (cont.)

- **Plenty to do**
- **Plenty to consider**
- **A lot of options**
- **But very few simple solutions!**
- **Lots of interesting issues (statistically and ecologically)**

CERF Is

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

Thanks

- **Alan Williams, Franzis Althaus, Felicity McEnnulty, Rudy Kloser, Gordon Keith, and Gary Poore**
- **Captain and Officers of the RV Southern Surveyor**
- **Jeff Dunn and Mike Fuller**

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

Bibliography

Dunn, P. K. and Smyth, G. K. (1996).
Randomized Quantile Residuals.
Journal of Computational and Graphical Statistics **5,** 236-244.

Dunstan, P. K. and Foster, S. D. (2009).
RAD biodiversity: Prediction of rank abundance distributions from deep water benthic assemblages.
in review.

Foster, S. D. and Dunstan, P. K. (2009).
The Analysis of Biodiversity Using Rank Abundance Distributions.
Biometrics to appear.

McGill, B. J., Etienne, R. S., Gray, J. S., Alonso, D., Anderson, M. J., Benecha, H. K., Dornelas, M., Enquist, B. J., Green, J. L., He, F., Hurlbert, A. H., Magurran, A. E., Marquet, P. A., Maruer, B. A., Ostling, A., Soykan, C. U., Ugland, K. I., and White, E. P. (2007).
Species abundance distributions: Moving beyond single prediction theories to integration within an ecological framework.
Ecology Letters **10,** 995–1015.

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