

# Hierarchical Dirichlet Regression Model for Benthic Cover in the Abrolhos Bank

Pamela M. Chiroque-Solano

Sciences Department at University of Lisboa, CE3C, Portugal  
Institute of Biology and SAGE-COPPE, Federal University of Rio de Janeiro, Brazil.

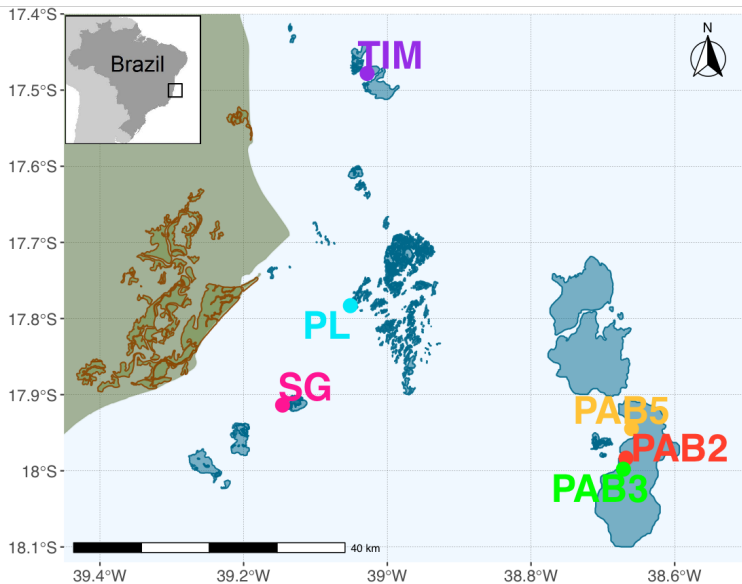
This work was done in collaboration with Mariana S. Sá and Larissa M. Martins.

11st of August, 2021

# Framework

- Multivariate regression with constraint response.
- Challenge:
  - ▶ Unbalanced;
  - ▶ Lot of missing data;
  - ▶ Identificability issues
- Problem: To understand the variability by localization

# Objective



# Model

## Maier (2014) and Holger (2018)

### Alternative parametrization

- $\mathbf{Y}_l \sim D(\mu_l, \phi_l)$  with parameter  $\alpha_{cl} = \mu_{cl}\phi_l$
- $\mu_{cl}$  : level term
- $\phi_l$  : precision term

### Reference component: $c^*$

$c^*$  should be chosen

### Sharing information equation

$$\begin{aligned}\beta_{cl} &= \beta_c + \epsilon_{\beta_l}, & \epsilon_{\beta_l} &\sim \mathcal{N}(0, V_\beta) \\ \theta_l &= \theta + \epsilon_{\theta_l}, & \epsilon_{\theta_l} &\sim \mathcal{N}(0, V_\theta)\end{aligned}$$

# Inference procedure

Let  $\Theta = (\beta, \phi)$  be the vector of parameters

Proper independent prior distribution for the parametric vector  $\Theta$  are Normal with zero mean and precision  $1/K$  (one over K) for all effects of the model.

The joint posterior distribution does not have a known closed form

$$\pi(\Theta \mid \mathbf{y}) \propto L(\Theta \mid \mathbf{y}) \prod_l^L \pi(\phi_l) \prod_c^C \pi(\beta_{cl}) \quad (1)$$

Sampling from the posterior distribution

by Markov chain Monte Carlo (MCMC) via the Stan software.

# Results and Conclusions

The results validate the original hypotheses

Sites near the coast (inshore) are more variable than the offshore sites.

## Main conclusions

- The proposed model quantifies the heteroscedasticity through precision effects via hierarchical structures by site;
- The method is flexible;
- The reference component has been chosen using objective criteria;
- The proposal allows to obtain adequate predictions.
- This work contributes to the United Nations's Sustainable Development Goal 14 - "Life Under Water".

## Sampling from the posterior distribution

by Markov chain Monte Carlo (MCMC) via the Stan software.

# Referencias

- Gelman, Andrew, and Jennifer Hill. 2006. Data Analysis Using Regression and Multilevel/Hierarchical Models. Analytical Methods for Social Research. Cambridge University Press. <https://doi.org/10.1017/CBO9780511790942>.
- Holger, and Sennhenn-Reulen. 2018. “Bayesian Regression for a Dirichlet Distributed Response Using Stan.” ArXiv.org. <https://arxiv.org/abs/1808.06399>.
- Maier, Marco J. 2014. “DirichletReg: Dirichlet Regression for Compositional Data in R.” Research Report Series/Department of Statistics and Mathematics 125. Vienna: WU Vienna University of Economics and Business. <http://epub.wu.ac.at/4077/>.
- Wang, K., G. Tian, and M Tang. 2011. Dirichlet and Related Distributions: Theory, Methods and Applications. Wiley Series in Probability; Statistics.

Thank you  
[pamela@dme.ufrj.br](mailto:pamela@dme.ufrj.br)