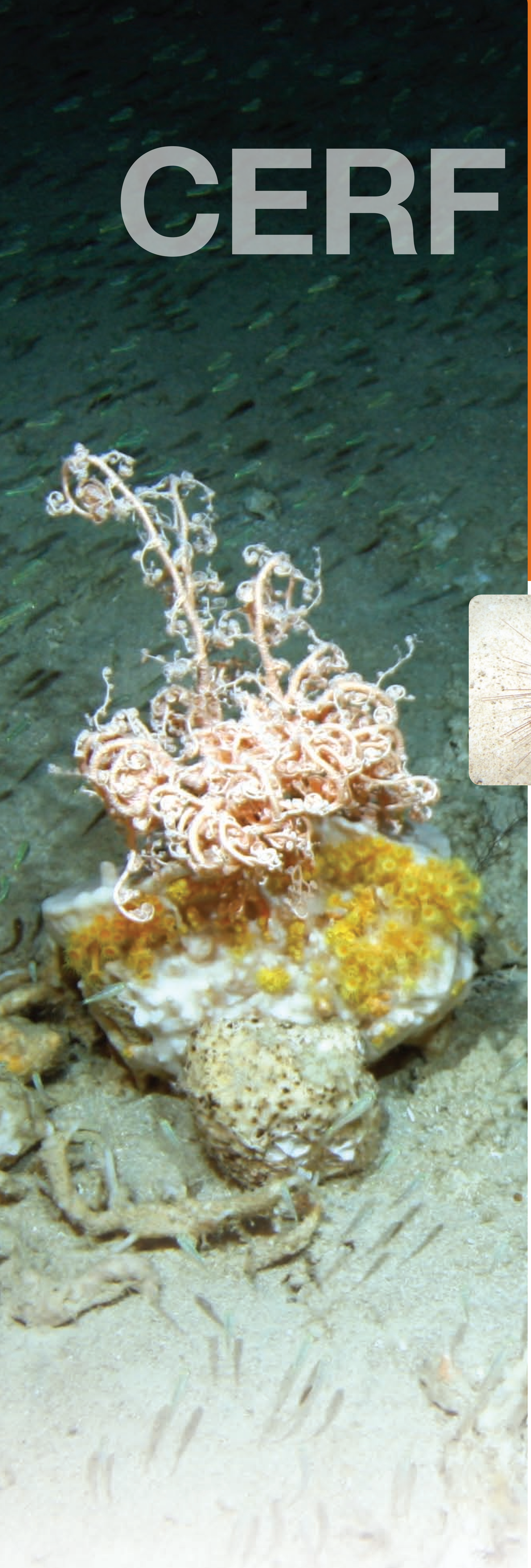
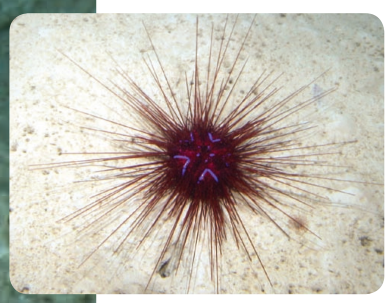
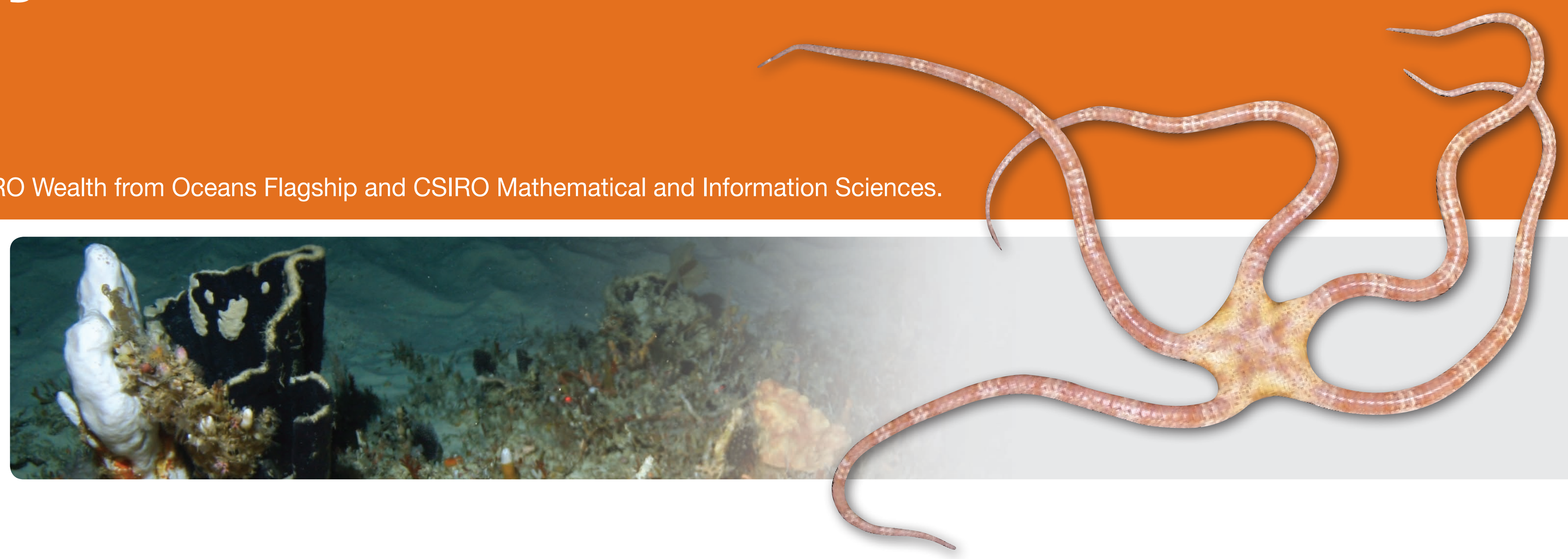


Some issues in modelling biodiversity using spatially modelled covariates

Hideyasu Shimadzu[1], Scott D. Foster[2]

[1] Marine and Coastal Environment Group, Geoscience Australia. [2] CSIRO Wealth from Oceans Flagship and CSIRO Mathematical and Information Sciences.





Statistical models have enhanced the understanding of the relationship between biodiversity and the environment. Typically, some sort of regression analysis is performed where physical variables are covariates. It is frequently the situation that the covariates are not observed; they are spatial predictions. This study indicates that this process may bias the statistical distribution and the resulting parameter estimates if the variance of the predictions is ignored.

Great Barrier Reef data

Data comprises of 1189 sites, where biological and physical variables were measured (Pitcher et al 2007, see Figure 1). The biological outcome variables were presence/absence of a particular species, and species richness (number of species) in a benthic sled sample. The physical variables used in this study were: depth, %carbonate and %mud. We randomly selected 200 sites from this complete set to mimic performing a biological survey (without measuring physical data).

> **Figure 1:** 200 study sites (orange) randomly selected from 1189 sled survey sites (black).

Model and approximate bias

Let Y_i be a biological response such as species presence/absence or richness, and let $\boldsymbol{x}_i$ be the vector of physical covariates at the survey site $\boldsymbol{i}$. A generalised linear model is often used with

$$\mathrm{E}\left[Y_i|\boldsymbol{x}_i\right] = h\left(\eta_i\right) = h\left(\boldsymbol{x}_i^\top \boldsymbol{\tau}\right),$$

where h is an inverse link function and $\boldsymbol{\tau}$ is a p x 1 vector of unknown parameters. However, the physical covariates are commonly not observed but are estimated by spatial predictions, $\tilde{\boldsymbol{x}}_i$, based on observations at other sites. Allowing for variance from the spatially predicted covariate, the approximate mean and covariance of the outcomes are

$$\mathrm{E}\left[\mathrm{E}\left[Y_i|\tilde{\boldsymbol{x}}_i\right]|\left\{\boldsymbol{X}_o\right\}\right] = h\left(\tilde{\boldsymbol{x}}_i^\top \boldsymbol{\tau}\right) + \frac{1}{2}\frac{d^2 h}{d\eta_i^2}\boldsymbol{\tau}^\top \boldsymbol{\Sigma}_{ii}\boldsymbol{\tau},$$

and

$$\mathrm{Cov}\left[Y_i, Y_j|\left\{\boldsymbol{X}_o\right\}\right] = \mathrm{E}\left[\mathrm{Cov}\left[Y_i, Y_j|\left\{\boldsymbol{X}_o\right\}\right]\right] + \frac{dh}{d\eta_i}\frac{dh}{d\eta_j}\boldsymbol{\tau}^\top \boldsymbol{\Sigma}_{ij}\boldsymbol{\tau},$$

where $\boldsymbol{\Sigma}_{ij}$ is the cross-covariance of prediction for the ith and jth physical covariates and $\{\boldsymbol{X}_o\}$ are the observed physical covariates.

Simulation studies

To assess the size of relative bias a simulation study was performed. We simulated biological data given the observed data and analysed them using the spatially predicted covariates ($\tilde{\boldsymbol{x}}_i$) and the observed covariates $\boldsymbol{x}_i$. A total of 1000 simulations were performed and are indexed by $\boldsymbol{k}$ in the equations below.

Simulation model 1

$$\mathrm{E}\left[Y_i^{(k)}\right] = h\left(\tilde{\boldsymbol{x}}_i^\top \boldsymbol{\tau}^{(k)}\right);$$

Simulation model 2

$$\mathrm{E}\left[Y_i^{(k)}\right] = h\left(\boldsymbol{x}_i^\top \boldsymbol{\tau}^{(k)}\right).$$

Results

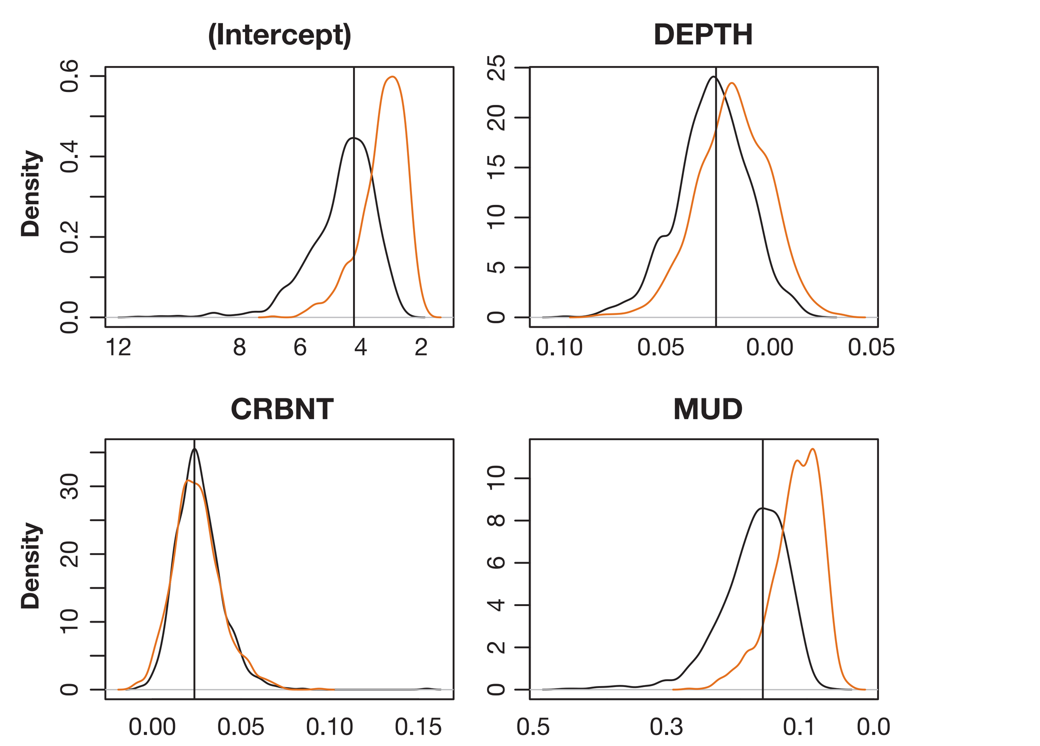
Relative size of bias

		Min.	Median	Max.
Presence/absence	bias/mean	-19.0%	100.1%	839400.0%
(*Cheilostomata Hippaliosina spp*)	bias/variance	0.0%	19.2%	1339000.0%
Richness	bias/mean	1.3%	3.5%	17.2%
	bias/variance	73.3%	361.6%	1200.0%

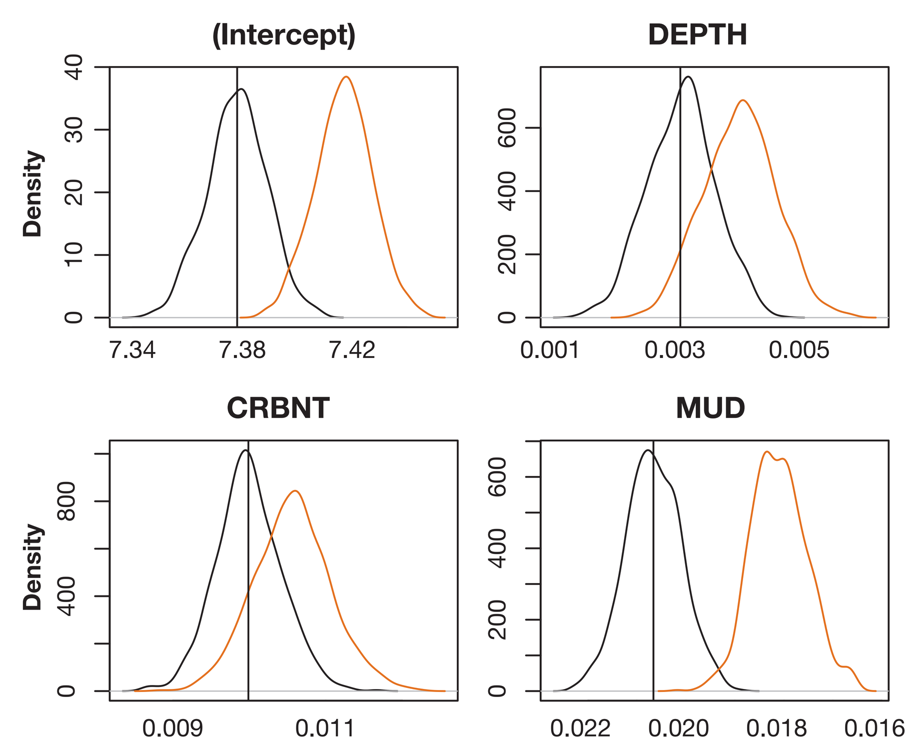
Summary statistics of the 200 sites

Conclusions

- Using spatially modelled covariates leads to bias in outcome distribution and parameter estimates
- Simulation shows this is a real problem not just a theoretical one
- Will need to account for uncertainty in future models



> **Figure 2:** Presence/absence – Empirical distribution of estimates from observed physical data (black) and predicted physical data (orange).



> **Figure 3:** Richness – Empirical distribution of estimates from observed physical data (black) and predicted physical data (orange).

Acknowledgements

We would like to thank everybody who contributed to the collection of the data. Specific thanks go to Roland Pitcher, Zhi Huang, Kathy Haskard, Mark Palmer, Ross Darnell, Piers Dunstan, Jin Li, and Bill Venables.

References

Pitcher R. *et al.* (2007) Seabed Biodiversity on the Continental Shelf of the Great Barrier Reef World Heritage Area. CSIRO Marine and Atmospheric Research.

MARINE BIODIVERSITY RESEARCH

Prediction and Management of Australia's Marine Biodiversity

Australian Government

Department of the Environment, Water, Heritage and the Arts

tafi Tasmanian Aquaculture and Fisheries Institute

MUSEUMVICTORIA

Australian Government

Geoscience Australia

National Research FLAGSHIPS

Wealth from Oceans

CSIRO

Australian Government

Australian Institute of Marine Science

The Marine Biodiversity Hub is funded through the Commonwealth Environment Research Facilities Program (CERF), administered through the Australian Government's Department of the Environment, Water, Heritage and the Arts. The key aim of CERF is to provide sound advice to inform environmental public policy objectives and to better the management of Australia's unique environment. (Our stakeholder partners are: AFMA, APPEA, CFA, DAFF, DEWHA, DAFF, the Tourism CRC, and WWF Australia)

Further information

Hideyasu Shimadzu

Phone: +61 2 6249 5842

Email: Hideyasu.Shimadzu@ga.gov.au

Marine and Coastal Environment Group, Petroleum and Marine Division,

Geoscience Australia
GPO Box 378 Canberra ACT 2601, Australia

www.marinehub.org