Pitcher, C.R., Lawton, P., Ellis, N., Smith, S.J., Incze, L.S., Wei, C-L., Greenlaw, M.E., Wolff, N.H., Sameoto, J.A. & Snelgrove, P.V.R. (2012) Exploring the role of environmental variables in shaping patterns of seabed biodiversity composition in regional-scale ecosystems. *Journal of Applied Ecology*

Appendix S4. Quantifying performance of biologically informed mapping.

The potential improvement of the "biologically informed" approach to mapping biodiversity, using Gradient Forest, over a purely physical approach to marine *bio*regional mapping was assessed. This assessment requires a "benchmark" measure of compositional differences among sampled survey sites. While there is no standardised measure of beta diversity patterns, and there are a variety of alternative metrics, the commonly used Bray-Curtis (BC) dissimilarity measure (Bray & Curtis 1957) was calculated for the GBR sled and trawl datasets. These were compared with the Euclidean distances (ED) between the same sites located in the biologically informed environment space, after transformation using the Gradient Forest cumulative importance distributions. A performance measure was then calculated, analogous to the stress diagnostic of the popular non-metric Multi-Dimensional Scaling (nMDS, Kruskal 1964), for the fit of a monotonic regression between the ED in biologically transformed environment space and the BC dissimilarities. We note that Gradient Forest does not directly fit a BC dissimilarity matrix, so this represents a severe test.

A similar procedure was followed for the "uninformed" environmental predictors. The BC dissimilarities were compared with the EDs between the same sites located in the physical environment space, after normalising the raw environmental predictors to unit variance (i.e. the variables have neutral weight and no re-scaling within their range).

The fits of the monotonic regressions and the stress diagnostics of the "biologically informed" approach were ~25% and ~50% better than those of the uninformed predictors, for the GBR sled and trawl datasets respectively (Table S4-1). This indicated that the biologically informed approach provided an improved representation of the patterns of biodiversity composition as present in the raw BC dissimilarities. Note that both of these approaches represent multi-dimensional composition constrained by relationships with the environment; when multi-dimensional representation of composition is not constrained by the environment (e.g. a conventional nMDS), the stress is lower (Table S4-1).

Table S4-1: Stress performance diagnostics for different methods of representing biological composition in multi-dimensional space.

Description of comparison	Stress	
	Sled	**Trawl**
3D nMDS ordination of BC dissimilarities of biological data matrix	0.214	0.165
ED matrix of RF transformed site data vs BC dissimilarity matrix	0.344	0.276
ED matrix of normalised site environmental data vs BC dissimilarity matrix	0.435	0.417

References

Bray, J.R. & Curtis, J.T. (1957) An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, **27**, 325–349.

Kruskal, J.B. (1964) Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, **29**, 115–129.