Seaweeds in Two Oceans: Beta-diversity (Appendices)

AJ Smit   *University of the Western Cape*

JJ Bolton   *University of Cape Town*

RJ Anderson   *Department of Agriculture, Forestry and Fisheries and University of Cape Town*

This file was generated in R using Rmarkdown, with a bit of LaTeX thrown in:

```
sessionInfo()
```

```
## R version 3.4.1 (2017-06-30)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Sierra 10.12.6
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.4/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.4/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## loaded via a namespace (and not attached):
##  [1] compiler_3.4.1  backports_1.1.1 magrittr_1.5    rprojroot_1.2
##  [5] tools_3.4.1     htmltools_0.3.6 yaml_2.1.14     Rcpp_0.12.13
##  [9] stringi_1.1.5   rmarkdown_1.6   knitr_1.17      stringr_1.2.0
## [13] digest_0.6.12   evaluate_0.10.1
```

Appendix A

Table 1: The 58 × 50 km sections of the South African coastline, with approximate GPS coordinates, delineation of sections, and some well-known sites in each section. Taken with permission from Bolton and Stegenga (2002).

Site	Lon	Lat	Limits	Including
1	16.72	-28.98	Orange River to just south of Holgats River	
2	16.94	-29.38	To just south of Wedge Point	Port Nolloth
3	17.08	-29.83	To just south of Melkbos Point	Kleinzee
4	17.26	-30.26	To Swartlintjies River	Skulpfontein Point, Swartlintjies
5	17.48	-30.68	To 10 km north of Groen River	River Hondeklip Bay, Spoeg River
6	17.72	-31.09	To just north of Brak River	Groen River, Island Point, Blougat
7	18.00	-31.46	To just north of Duiwegat	Voëlklip, Sout River, Blinkwater Bay
8	18.25	-31.85	To just south of Doring Bay	Olifants River, Strandfontein, Doring Bay
9	18.34	-32.30	To just north of Elands Bay	Lambert's Bay, Lang River
10	18.20	-32.72	To just north of Laaiplek	Elands Bay, Die Vlei, Dwarskersbos
11	17.85	-32.83	To just south of Cape Columbine	Laaiplek, St Helena Bay, Paternoster
12	18.03	-33.03	To just east of Saldahna	
13	18.01	-33.15	To Postberg	Langebaan Lagoon
14	18.32	-33.50	To just south of Modder River	Yzerfontein, Dassen Is., Grotto Bay
15	18.47	-33.91	To Sea Point	Melkbosstrand, Table Bay, Green Point
16	18.37	-34.21	To just north of Scarborough	Camps Bay, Hout Bay, Kommetjie
17	18.47	-34.11	To just east of Kalk Bay	Scarborough, Cape Point, Fishoek
18	18.82	-34.19	To just south of Gordons Bay	Muizenburg, Strandfontein, Strand
19	19.07	-34.35	To just east of Kleinmond	Rooi Els, Hangklip, Betty's Bay
20	19.34	-34.59	To just south of Danger Point	Bot River, Sand Bay, Hermanus, Die Kelders
21	19.66	-34.79	To just east of Quoin Point	Danger Point, Pearly Beach, Dyer Island
22	20.07	-34.75	To just east of Struis Bay	Die Mond, Cape Agulhas
23	20.48	-34.49	To just east of Skipskop	Struis Bay, Arniston
24	20.87	-34.39	To just east of Cape Infanta	Koppie Alleen, Cape Infanta, Bree River, Witsand
25	21.36	-34.42	To just east of Grootjongensfontein	Puntjie, Skurwe Bay
26	21.83	-34.38	To just west of Gouritzmond	Stil Bay, Bloukrans, Bull Point
27	22.12	-34.16	To just north of Mossel Bay	Gouritzmond, Vlees Bay, Pinnacle Rock
28	22.54	-34.01	To just west of Victoria Bay	Hartenbos, Klein and Groot Brak rivers, Herolds Bay, Skuinsbank
29	23.02	-34.08	To just west of The Heads, Knysna	Victoria Bay, Wilderness, Platbank, Oesterbank, Walker Bay
30	23.36	-34.10	To Jack's Point, south of Plettenberg Bay	The Heads, Neusgate
31	23.78	-34.01	To Elandbos River	Plettenberg Bay, Arch Rock, Die Punt, Blousloep
32	24.27	-34.08	To Skuinsklip	Storms River, Voëlkrans, Skietgate
33	24.74	-34.19	To Thys Point	Aasvoëlklip, Tsitsikamma River, Klipdrif River
34	25.04	-33.97	To just west of Gamtoos River	Cape St Francis, Krom River, Seekoei River, Jeffreys Bay
35	25.52	-34.04	To just east of Sardinia Bay	Van Stadens River, Claasen Point
36	25.70	-33.79	To just east of St George's Beach	Chelsea Point, Port Elizabeth, Bluewater Bay
37	26.18	-33.72	To just west of Woody Cape	St Croix Is., Sundays River
38	26.65	-33.70	To just west of Kenton-on-Sea	Seal Is., Bird Is., Cape Padrone, Cannon Rocks, Boknes
39	27.10	-33.52	To just east of Kleinemonde	Kasouga, Port Alfred
40	27.52	-33.27	To just east of Keiskamma River	Great Fish River, Madagascar Reef
41	27.93	-33.01	To just east of East London	Kayser's Beach, Kidd's Beach, Cove Rock
42	28.30	-32.73	To Haga-Haga	Gonubie, Cintsa River
43	28.68	-32.44	To Qora River	Morgans Bay, Kei Mouth, Nxaxo River, Mazeppa Bay
44	29.05	-32.11	To just east of Xora River	Dwesa, The Haven
45	29.37	-31.76	To Sharks Point	Mncwasa River, Coffee Bay, Hluleka
46	29.74	-31.46	To Mkozi River	Boulder Bay, Port St Johns, Montshe, Ntsubane
47	30.12	-31.18	To Mnyameni River	Cathedral Rock, Lambasi Bay, Wild Coast
48	30.41	-30.81	To just north of St Michaels-on-Sea	Mzamba, Port Edward, Southbroom, Margate
49	30.68	-30.41	To just south of Pennington	Port Shepstone, Mzumbe, Sezela
50	30.93	-30.01	To just south of Isipingo Beach	Scottburgh, Park Rynie, Umkomaas, Illovo, Amanzimtoti
51	31.15	-29.62	To Desainagar	Durban, Umhlanga Rocks, Umdloti Beach
52	31.46	-29.26	To just north of Zinkwazi Beach	Westbrook, Ballito, Blythdale Beach
53	31.82	-28.94	To just east of Mtunzini	Tugela River, Dunn's Reserve
54	32.21	-28.70	To Mbonambi Beach	Richards Bay
55	32.46	-28.32	To just north of First Rocks	Dawson's Rocks, Cape St Lucia, St Lucia
56	32.59	-27.87	To Bhukwini	Mission Rocks, Cape Vidal, Leven Point
57	32.72	-27.42	To just north of Gobey's Point	Liefeldts Rocks, Sodwana Bay
58	32.87	-26.97	To Kosimeer	Hulley Point, Black Rock

Appendix B

Spatial analysis background and code

The intention of this section is to show the approach and **R** scripts used to pull apart the spatial scales at which seaweed assemblages are structured around the coast of South Africa. Specifically, I wish to determine if these scales match those expressed by the coastal thermal provinces and the ocean regime underpinned by the Agulhas and Benguela Currents.

The data

I use two data sets. The first, Y, comprises distribution records of 846 macroalgal species within each of 58 × 50 km-long sections (Appendix A) of the South African coast (updated from Bolton and Stegenga, 2002). This represents *ca.* 90% of the known seaweed flora of South Africa, but excludes some very small and/or very rare species for which data are insufficient. The data are from verifiable literature sources and John Bolton and Rob Anderson's own collections, assembled from information collected by teams of phycologists over three decades (Bolton, 1986; Bolton and Stegenga, 2002; De Clerck et al., 2005; Stegenga et al., 1997). The second, E, is a dataset of *in situ* coastal seawater temperatures (Smit et al., 2013) derived from daily measurements over up to 40 years.

A third data set of explanatory variables — the spatial variables (S) — is constructed as per the instructions in section *Preparation of spatial variables*, later on.

Setting up the analysis environment

This is **R**, so first I need to find, install and load various packages. Some of the packages will be available on CRAN and can be accessed and installed in the usual way, but others will have to be downloaded from R Forge.

```
library(betapart)
library(vegan)
library(gridExtra)
library(packfor)
library(grid)
library(gridBase)
library(tidyr)
library(spdep) # for dnearneigh() in PCNM.R
library(AEM) # for moran.I.multi() in PCNM.R
# install.packages("devtools")
# install.packages("AEM", repos = "http://R-Forge.R-project.org")
# install.packages("packfor", repos = "http://R-Forge.R-project.org")
source("functions/pcoa_all.R")
source("functions/PCNM.R")
source("functions/spatial_MEM.R")
```

Now I get to the data. The first step involves the species table (Y). First I compute the Sørensen dissimilarity and then I decompose the dissimilarity into the 'turnover' (β_{sim}) and 'nestedness-resultant' (β_{sne}) components (Baselga, 2010; Baselga et al., 2013) using the `betapart.core()` and `betapart.pair()` functions of the **betapart** package (Baselga et al., 2013). These are placed into the matrices $Y1$ and $Y2$. Optionally, I can apply a prinipal components analysis (PCA) on Y to find the major patterns in the community data. In **vegan** this is done using the `rda()` function and not supplying the constraints (*i.e.* the environment table, E, or the spatial table, S). The formal analysis will use the species data in distance-based redundancy analyses (db-RDA as per **vegan**'s `capscale()` function) by coupling them with E and S.

```
# Read in the species data (note: on GitHub only the distance
# matrices obtained via 'beta.part' and 'beta.pair' (below)
# will be provided -- they are read in as 'Y1.Rdata' and 'Y2.Rdata';
# the raw data cannot be shared at this stage, but the distance matrix is provided):
# spp <- read.csv('../stats/seaweeds.csv')
# spp <- dplyr::select(spp, -1)
```

```
# Decompose total Sørensen dissimilarity into turnover and
# nestedness-resultant components:
# Y.core <- betapart.core(spp)
# Y.pair <- beta.pair(Y.core, index.family = "sor")

# Let Y1 be the turnover component (beta-sim):
# Y1 <- as.matrix(Y.pair$beta.sim)
# save(Y1, file = "data/Y1.Rdata")
load("data/Y1.Rdata")

# Let Y2 be the nestedness-resultant component (beta-sne):
# Y2 <- as.matrix(Y.pair$beta.sne)
# save(Y2, file = "data/Y2.Rdata")
load("data/Y2.Rdata")
```

It is now necessary to load the environmental data and some setup files that partition the 58 coastal sections (and the species and environmental data that fall within these sections) into bioregions.

The thermal (environmental) data contain various variables, but in the analysis I use only some of them. These data were obtained from many sites along the South African coast, but using interpolation (not included here) I calculated the thermal properties for each of the coastal sections for which seaweed data are available. Consequently we have a data frame with 58 rows and a column for each of the thermal metrics. Before use, I apply **vegan**'s decostand() function to scale the data to zero mean and unit variance.

Four bioregions are recognised for South Africa (Bolton and Anderson, 2004), namely the Benguela Marine Province (BMP; coastal sections 1–17), the Benguela-Agulhas Transition Zone (B-ATZ; 18–22), the Agulhas Marine Province (AMP; 19–43/44) and the East Coast Transition Zone (ECTZ; 44/45–58). My plotting functions partition the data into the bioregions and colour code the figures accordingly so I can see regional patterns in β-diversity emerging.

```
# Now comes in the in situ temperatures for the 58 coastal sections (interpolated temperaures as per version 2 of the South A
load('data/E.RData')
env <- as.data.frame(interpOut)

# I select only some of the thermal vars; the rest
# are collinear with some of the ones I import:
E1 <- dplyr::select(env, febMean, febRange, febSD, augMean,
                    augRange, augSD, annMean, annRange, annSD)

# Calculate z-scores:
E1 <- decostand(E1, method = "standardize")

# Load the coordinates of the coastal sections:
sites <- read.csv("data/sites.csv")
sites <- sites[, c(2, 1)]

# Load the bioregion definition:
bioreg <- read.csv('data/bioregions.csv', header = TRUE)
```

Preparation of spatial variables

I test the niche difference mechanism as the primary species compositional assembly process operating along South African shores. I suggest that the thermal gradient along the coast provides a suite of abiotic (thermal) conditions from which species can select based on their physiological tolerances, and hence this will structure β-diversity. For this mechanism to function one would assume that all species have equal access to all sections along this stretch of coast, thus following Beijerinck's 'Law' that everything is everywhere but the environment selects (Sauer, 1988) (but see main text!).

The basic approach to a spatial analysis structured around a biological response (*e.g.* community structure and composition; Y), environmental variables (E) and their spatial representation (S) involves an analysis of Moran's

eigenvector maps (MEM), followed by db-RDA and variance partitioning. Various literature sources discuss principle behind Moran's eigenvector maps (Dray et al., 2006, 2012). Worked examples are also presented in the excellent book *Numerical Ecology with R* (Borcard et al., 2011) in Section 7.4. The method followed here has been adapted from these and other sources.

Obtaining the MEMs to use in the analysis is based on the procedure introduced by Borcard and Legendre (2002), which was later modified by Dray et al. (2006). The basic approach involves:

1. Set up a geographic or Euclidian distance matrix representing the pairwise distances between the n sites ($D = [d_{ij}]$). I already did this when I applied the `decostand` function earlier.

2. Construct a truncated distance matrix by calculating a Minimum Spanning Tree ($S^\star$) and noting the following rules:
$$S^\star = \left\{ \begin{array}{cl} 0 & \text{if } i = j \\ d_{ij} & \text{if } d_{ij} \leq t \\ 4t & \text{if } d_{ij} > t \end{array} \right.$$

Weighting may be applied if desired, resulting in $S^\star_w$. It is not done here.

3. Do a Principal Coordinates Analysis (PCoA) of the truncated distance matrix $S^\star$.

The spatial properties imprinted on the species and their environment can be specified using a matrix of Euclidian or geographic distances. These coordinates are 'truncated' into a square (section × section) matrix containing non-negative values ($S^\star$). By convention the diagonal values are set to zero. A very basic spatial matrix is binary, where 1 codes for pairs of neigbouring sites while 0 denotes non-connected sites according to the chosen network topology. Such matrices are called 'binary connectivity matrices' and relate to graphs made using distance criteria derived from graph theory.

Truncation produced by Minimum Spanning Trees (MST) focuses on the binary relationships between neighbouring sites, discarding any other connections (*i.e.* some sites are considered to be neighbours, while for others the relationships are null). One could also choose a Gabriel graph or another kind of network topology. Such matrix representations show section-to-section connectivities. In the case of South Africa's coastline data, the MST causes sections to be connected only to other sections adjacent to two sides of it: for example, Section **4** is directly connected to *only* Sections **3** and **5**; sections at the termini of the coastal 'string' of sections are each connected to only one other section. The binary connectivity matrices, also called *topology-based connectivity matrices*, can be produced from Euclidian or geographic coordinates using functions in at least two **R** packages (I start with geographic coordinates). One option is to use the **spdep** package's `mst.nb()` function to calculate a MST, but there are also options in the **vegan** package and elsewhere. The neighbours list arrived at from the MST represents the spatial component, $S^\star$. The MST results in small connectivity artefacts in the Saldanha Bay region where the closest sections are not necessarily the ones adjacent one-another following along the path around the coast, because sections at opposite sides of the bay may in fact be closer together. This topological inconsistency does not affect the spatial analysis in any way.

Once the truncated distance matrix has been prepared, it is subjected to a PCoA and I keep the eigenvectors that represent positive spatial correlation (positive Moran's I). For the MEM analysis I use the function `PCNM()` that resides in the `functions` folder in the file `PCNM.R` (see notes inside about authorship). PCNM stands for Principal Coordinates Analysis of Neighbourhood Matrices (the neighbourhood matrix in this instance being the MST). This method automatically constructs the spatial variables and calculates the Moran's I for each. The MEMs are completely orthogonal and represent the spatial structures over the full range of scales from 50 to 2,700 km. The large eigenvectors represent broad spatial scales while smaller ones cover finer features. The *spatial data* will be used as a set of explanatory variables in the multiple regression type analyses applied to a species dissimilarity matrix (*i.e.* the db-RDA; Dray et al., 2012)

The code below reproduces the spatial analysis in the paper. Due to the length of the output I have prevented the script from returning any output here; rather, if the reader is for some odd reason interested in repeating this analysis, s/he may find the data and scripts in my GitHub repository, and the full code can be run in its entirety. Well, I hope this will work, but if it doesn't (probably very likely) then write to me at ajsmit@uwc.ac.za and I shall assist — this may depend on if your email has a catchy title that will make it stand out from among all the other emails which other people think are equally important.

```
## Auto PCNM:
S.auto <- PCNM(dist(sites), silent = TRUE)
# summary(S.auto)


# The truncation distance:
S.dmin <- S.auto$thresh


# The number of eigenvalues:
S.len <- length(S.auto$values)


# Expected value of I, no spatial correlation:
S.auto$expected_Moran


# Select eigenfunction with positive spatial correlation:
S.sel <- which(S.auto$Moran_I$Positive == TRUE)
# length(S.sel)
# there are 27 MEMs, i.e. 27 of the PCNM variables (eigenvalues) relate
# significantly to Moran's I


# Extract the eigenvectors associated with those MEMs:
S.pos <- as.data.frame(S.auto$vectors)[, S.sel]
```

The code below lets us visualise the configuration of the 58 coastal sections as represented by the minimum spanning tree. Because the sites are constrained by the coast the MST network topology results in a string of coastal sections arranged along the shore between Section 1 and Section 58. This spatial network therefore also captures the spatial connectivity in the seaweed's dispersal ability along the shore, although no directionality is associated with dispersal. In the paper I discuss the possible influence of ocean currents (*e.g.* Wernberg et al., 2013) and I pointed out that it is tempting to assume that seaweeds would disperse in the direction the major ocean currents. These kinds of networks could conceivably be configured to model dispersal due to currents, but here it is simply used for representing the spatial scale of the study region.

```
# The spatial netwwork topology of the coastal sections can be seen by:
plot(S.auto$spanning, sites)
```

db-RDA on the MEMs

The next step of the spatial analysis is to apply a db-RDA with the seaweed data ($Y1$ and $Y2$) coupled with the MEMs. I now run a full (global) db-RDA on the significant, positive MEMs selected above, and I then perform a permutation test to see if the fit is significant.

```
# Run the db-RDA on the Y1 data:
S.Y1.cs <- capscale(Y1 ~., S.pos)


# Permutation test to test for the significance of the global fit:
anova(S.Y1.cs, parallel = 4) # ... yes, significant!


# The global adjusted R2 --- the variance explained by the constrained axes:
S.Y1.cs.R2 <- RsquareAdj(S.Y1.cs)$adj.r.squared


# Variance explained by full model:
sum(S.Y1.cs$CCA$eig) / S.Y1.cs$tot.chi * 100


# And on the Y2 data (uncommented, but same as above):
S.Y2.cs <- capscale(Y2 ~., S.pos)
S.Y2.cs.R2 <- RsquareAdj(S.Y2.cs)$adj.r.squared
sum(S.Y2.cs$CCA$eig) / S.Y2.cs$tot.chi * 100
```

Since the analysis is significant, I compute the adjusted R^2 and run forward selection of the MEMs. The forward selection procedure of Blanchet et al. (2008) is implemented in the **packfor** package for R, and I use it to reduce the number of MEM variables and retain only those that best fit the biotic data. Forward selection prevents the inflation of the overall type I error and reduces the number of explanatory variables used in the final model, which improves parsimony. I then run a new db-RDA analysis on the 'best' (reduced) set of MEM variables that was selected.

```r
# Forward selection on Y1:
S.Y1.fwd <- forward.sel(Y1, as.matrix(S.pos), adjR2thresh = S.Y1.cs.R2)


# Forward selection on Y2:
S.Y2.fwd <- forward.sel(Y2, as.matrix(S.pos), adjR2thresh = S.Y2.cs.R2)
```

```r
# Write the significant MEMs to a new object:
S.Y1.no.sig <- nrow(S.Y1.fwd)
S.Y1.sign <- sort(S.Y1.fwd[, 2])
S.Y1.red <- S.pos[, c(S.Y1.sign)]
colnames(S.Y1.red) <- paste(rep("MEM", S.Y1.no.sig),
                            as.character(S.Y1.sign), sep = "")

# Identity of significant MEMs:
colnames(S.Y1.red)

# Run a new db-RDA on the best MEM variables:
S.Y1.s2 <- capscale(Y1 ~., data = S.Y1.red)
# no need to check these for collinearity as the
# MEMs are completely orthogonal..

# Permutation test to test for significance:
anova(S.Y1.s2, parallel = 4)

# Test by axis:
anova(S.Y1.s2, by = "axis", parallel = 4)

# The significant axes:
S.Y1.axis.test <- anova(S.Y1.s2, by = "terms", parallel = 4)
S.Y1.ax <- which(S.Y1.axis.test[, 4] < 0.05)
S.Y1.sign.ax <- colnames(S.Y1.red[,S.Y1.ax])

# Test by terms:
anova(S.Y1.s2, by = "terms", parallel = 4)

# The adjusted R2 --- the variance explained by the constrained axes:
S.Y1.s2.R2 <- RsquareAdj(S.Y1.s2)$adj.r.squared

# Variance explained by reduced model:
sum(S.Y1.s2$CCA$eig) / S.Y1.s2$tot.chi * 100

# Show only the first 6 rows:
scores(S.Y1.s2, display = "bp", choices = c(1:4))[1:6, ]
```

```r
# As above, but now with Y2:
S.Y2.no.sig <- nrow(S.Y2.fwd)
S.Y2.sign <- sort(S.Y2.fwd[, 2])
S.Y2.red <- S.pos[, c(S.Y2.sign)]
colnames(S.Y2.red) <- paste(rep("MEM", S.Y2.no.sig),
                            as.character(S.Y2.sign), sep = "")
colnames(S.Y2.red)
```

```
S.Y2.s2 <- capscale(Y2 ~., data = S.Y2.red)


anova(S.Y2.s2, parallel = 4) # ... yes, significant!


anova(S.Y2.s2, by = "axis", parallel = 4)


S.Y2.axis.test <- anova(S.Y2.s2, by = "terms", parallel = 4)
S.Y2.ax <- which(S.Y2.axis.test[, 4] < 0.05)
S.Y2.sign.ax <- colnames(S.Y2.red[,S.Y2.ax])


S.Y2.s2.R2 <- RsquareAdj(S.Y2.s2)$adj.r.squared


sum(S.Y2.s2$CCA$eig) / S.Y2.s2$tot.chi * 100


scores(S.Y2.s2, display = "bp", choices = c(1:4))
```

A few visualisations

Now I make a visualisation to reveal the spatial arrangement of the MEMs used in the final db-RDA involving the spatial variables (*i.e.* S.Y1.s2 and S.Y2.s2). The spatial configuration relates to broad scales as seen in Fig. 3 in the paper. Here are plots of the site scores for the MEMs and $Y1$ and $Y2$ (a few panels belonging with Fig. 3):

```
# Plot the first canonical axis of the db-RDA with the significant MEMs for Y1;
# (see Fig. 3):
S.Y1.axes <- scores(S.Y1.s2, choices = c(1:3), display = "lc", scaling = 1)
S.Y1.plt.axis1 <- ggmap() +
  geom_point(data = sites, aes(x = Longitude, y = Latitude,
                               size = abs(S.Y1.axes[, 1]),
                               col = ifelse(S.Y1.axes[, 1] < 0, "a", "b")), shape = 1) +
  scale_size_continuous(guide = FALSE) +
  scale_colour_manual(guide = FALSE, values = c("black", "grey60")) +
  ggtitle(expression(paste("CAP1 of spatial variables, ", beta[sim])))

# And the same for Y2 (see Fig. 3):
S.Y2.axes <- scores(S.Y2.s2, choices = c(1:3), display = "lc", scaling = 1)
S.Y2.plt.axis1 <- ggmap() +
  geom_point(data = sites, aes(x = Longitude, y = Latitude,
                               size = abs(S.Y2.axes[, 1]),
                               col = ifelse(S.Y2.axes[, 1] < 0, "a", "b")), shape = 1) +
  scale_size_continuous(guide = FALSE) +
  scale_colour_manual(guide = FALSE, values = c("black", "grey60")) +
  ggtitle(expression(paste("CAP1 of spatial variables, ", beta[sne])))
```

Now that I know that spatial structures are present in the seaweed data I check how these significant spatial patterns (two significant canonical axes, CAP1 and CAP2) are related to the environmental variables using linear regression. Checks for normality are also done but none of the output is printed here.

Next I want to show the ordination biplots of the MEM variables with respect to the sites using scaling = 2 (species) and showing the LC scores. Now I can see the major directions of influence of the spatial variables with respect to the sites. The code below produces a few panels of Fig. 2:

```
# A few of the panels that go with Fig. 2;
# first for Y1...:
S.Y1.scrs <- scores(S.Y1.s2, display = c("sp","wa","lc","bp","cn"))
S.Y1.df_sites <- data.frame(S.Y1.scrs$constraints)
S.Y1.df_sites$bioreg <- bioreg$bolton
S.Y1.df_sites$section <- seq(1:58)
colnames(S.Y1.df_sites) <- c("x", "y", "Bioregion", "Section")
```

```r
multiplier <- ordiArrowMul(S.Y1.scrs$biplot)
S.Y1.bp <- S.Y1.scrs$biplot * multiplier
S.Y1.bp <- as.data.frame(S.Y1.bp)
S.Y1.bp$labels <- rownames(S.Y1.bp)
colnames(S.Y1.bp) <- c("x", "y", "labels")
S.Y1.bp.sign <- S.Y1.bp[S.Y1.bp$labels %in% S.Y1.sign.ax,]

# A modification of the vegan ordiArrowTextXY() function to prevent the
# "plot.new has not been called yet" from occuring
source("functions/text_mult.R")

S.Y1.text <- text.mult(S.Y1.scrs$biplot)
S.Y1.text <- as.data.frame(S.Y1.text)
S.Y1.text$labels <- rownames(S.Y1.text)
colnames(S.Y1.text) <- c("x", "y", "labels")
S.Y1.text.sign <- S.Y1.text[S.Y1.text$labels %in% S.Y1.sign.ax,]

S.Y1.p <- ggplot(data = S.Y1.df_sites, aes(x, y, colour = Bioregion)) +
  geom_point(size = 4.0) +
  geom_text(aes(label = Section), size = 3.0, col = "white") +
  geom_segment(data = S.Y1.bp,
               aes(x = 0, y = 0, xend = x, yend = y),
               arrow = arrow(length = unit(0.2, "cm")),
               color = "red", alpha = 1, size = 0.7) +
  geom_text(data = as.data.frame(S.Y1.text),
            aes(x, y, label = rownames(S.Y1.text)),
            color = "black") +
  xlab("CAP1") + ylab("CAP2") +
  ggtitle(expression(paste("Spatial variables and ", beta[sim]))) +
  theme_grey() +
  theme(panel.grid.minor = element_blank(),
        legend.position = "none",
        aspect.ratio = 0.8)

# ...then for Y2:
S.Y2.scrs <- scores(S.Y2.s2, display = c("sp","wa","lc","bp","cn"))
S.Y2.df_sites <- data.frame(S.Y2.scrs$constraints)
S.Y2.df_sites$bioreg <- bioreg$bolton
S.Y2.df_sites$section <- seq(1:58)
colnames(S.Y2.df_sites) <- c("x", "y", "Bioregion", "Section")

multiplier <- ordiArrowMul(S.Y2.scrs$biplot, fill = 0.25)
S.Y2.bp <- S.Y2.scrs$biplot * multiplier
S.Y2.bp <- as.data.frame(S.Y2.bp)
S.Y2.bp$labels <- rownames(S.Y2.bp)
colnames(S.Y2.bp) <- c("x", "y", "labels")
S.Y2.bp.sign <- S.Y2.bp[S.Y2.bp$labels %in% S.Y2.sign.ax,]

S.Y2.text <- text.mult(S.Y2.scrs$biplot, fill = 0.25)
S.Y2.text <- as.data.frame(S.Y2.text)
S.Y2.text$labels <- rownames(S.Y2.text)
colnames(S.Y2.text) <- c("x", "y", "labels")
S.Y2.text.sign <- S.Y2.text[S.Y2.text$labels %in% S.Y2.sign.ax,]

S.Y2.p <- ggplot(data = S.Y2.df_sites, aes(x, y, colour = Bioregion)) +
  geom_point(size = 4.0) +
  geom_text(aes(label = Section), size = 3.0, col = "white") +
```

```
    geom_segment(data = S.Y2.bp.sign,
                aes(x = 0, y = 0, xend = x, yend = y),
                arrow = arrow(length = unit(0.2, "cm")),
                color = "red", alpha = 1, size = 0.7) +
    geom_text(data = as.data.frame(S.Y2.text.sign),
                aes(x, y, label = rownames(S.Y2.text.sign)),
                color = "black") +
    xlab("CAP1") + ylab("CAP2") +
    ggtitle(expression(paste("Spatial variables and ", beta[sne]))) +
    theme_grey() +
    theme(panel.grid.minor = element_blank(),
            legend.position = "none",
            aspect.ratio = 0.8)
```

Analysis of the thermal variables

As before with the spatial variable, I now do a db-RDA involving all the thermal variables (E) followed by forward selection. There is less explanation provided here as the reader should now be familiar with db-RDA — the procedure is the same as with the MEMs, just different explanatory variables are supplied. Another difference is that the thermal variables are not necessarily orthogonal, so I check for collinearity using variance inflation factors (VIF).

I start with the full model and then run forward selection and repeat the db-RDA on the reduced set. Analyses shown for $Y1$ and $Y2$:

```
# First Y1:
E.Y1.cs <- capscale(Y1 ~., E1)

# Is the fit significant?
anova(E.Y1.cs, parallel = 4) # ... yes!

# The adjusted R2 --- the variance explained by the constrained axes:
E.Y1.R2a <- RsquareAdj(E.Y1.cs)$adj.r.squared

# Variance explained by full model:
sum(E.Y1.cs$CCA$eig) / E.Y1.cs$tot.chi * 100
```

```
# ...and now Y2:
E.Y2.cs <- capscale(Y2 ~., E1)
anova(E.Y2.cs, parallel = 4) # ... yes!
E.Y2.R2a <- RsquareAdj(E.Y2.cs)$adj.r.squared
sum(E.Y2.cs$CCA$eig) / E.Y2.cs$tot.chi * 100
```

```
# Forward selection on Y1:
E.Y1.fwd <- forward.sel(Y1, E1, adjR2thresh = E.Y1.R2a, nperm = 999)

# Forward selection on Y1:
E.Y2.fwd <- forward.sel(Y2, E1, adjR2thresh = E.Y2.R2a, nperm = 999)
```

```
# Write the significant envs to a new object, and
# identity of significant envs in increasing order;
# first Y1:
E.Y1.sign <- E.Y1.fwd %>%
  dplyr::select(variables) %>%
  as.vector()

E.Y1.red <- E1[, E.Y1.sign[,1]]
```

```r
# Run a new env analysis on the best env variables:
E.Y1.cs2 <- capscale(Y1 ~., E.Y1.red)

# Check for collinearity:
vif.cca(E.Y1.cs2)

# If there are significant collinearity the collinear variables can be removed:
# E.red <- dplyr::select(E.red, -augMean)
# E.cs2 <- capscale(Y1 ~ ., E.red)
#
# check for collinearity again:
# vif.cca(E.cs2) # much better

# Test for significance:
anova(E.Y1.cs2, parallel = 4) # ... yes!

# Which axes are significant?
anova(E.Y1.cs2, by = "axis", parallel = 4) # ... yes!

# The significant axes:
E.Y1.axis.test <- anova(E.Y1.cs2, by = "terms", parallel = 4)
E.Y1.ax <- which(E.Y1.axis.test[, 4] < 0.05)
E.Y1.sign.ax <- colnames(E.Y1.red[,E.Y1.ax])

# The adjusted R2 --- the variance explained by the constrained axes:
E.Y1.cs2.R2 <- RsquareAdj(E.Y1.cs2)$adj.r.squared

# Variance explained by reduced (final) model:
sum(E.Y1.cs2$CCA$eig) / E.Y1.cs2$tot.chi * 100

# The biplot scores for constraining variables:
scores(E.Y1.cs2, display = "bp", choices = c(1:2))
```

```r
# ...then Y2
E.Y2.sign <- E.Y2.fwd %>%
  dplyr::select(variables) %>%
  as.vector()

E.Y2.red <- E1[, E.Y2.sign[,1]]

E.Y2.sign <- sort(E.Y2.fwd[, 1])
E.Y2.red <- data.frame(E1[, c(E.Y2.sign)])
colnames(E.Y2.red) <- E.Y2.sign

E.Y2.cs2 <- capscale(Y2 ~., E.Y2.red)

vif.cca(E.Y2.cs2)
# E.red <- dplyr::select(E.red, -augMean)
# E.cs2 <- capscale(Y2 ~ ., E.red)

# vif.cca(E.cs2) # much better

anova(E.Y2.cs2, parallel = 4) # ... yes!

E.Y2.axis.test <- anova(E.Y2.cs2, by = "terms", parallel = 4)
# E.Y2.ax <- which(E.Y2.axis.test[, 4] < 0.05) # doesn't work...
# E.Y2.sign.ax <- colnames(E.Y2.red[,E.Y2.ax])
```

```
E.Y2.sign.ax <- ”annMean” # a manual cheat

anova(E.Y2.cs2, by = ”terms”, parallel = 4) # ... yes!

E.Y2.cs2.R2 <- RsquareAdj(E.Y2.cs2)$adj.r.squared

sum(E.Y2.cs2$CCA$eig) / E.Y2.cs2$tot.chi * 100

scores(E.Y2.cs2, display = ”bp”, choices = c(1:2))
```

Now I make the remaining panels of Fig. 3, these showing the spatial arrangement associated with the site scores of the environmental variables for $Y1$ and $Y2$:

```
# Plot the two significant canonical axes of the
# db-RDA with the significant MEMs. This part of Fig. 3:
E.Y1.axes <- scores(E.Y1.cs2, choices = c(1:2),
                    display = ”lc”, scaling = 1)
E.Y1.plt.axis1 <- ggmap() +
  geom_point(data = sites, aes(x = Longitude, y = Latitude, size = E.Y1.axes[, 1]),
             col = ”black”, shape = 1) +
  scale_size_continuous(guide = FALSE) +
  ggtitle(expression(paste(”CAP1 of thermal variables, ”, beta[sim])))

E.Y1.plt.axis2 <- ggmap() +
  geom_point(data = sites, aes(x = Longitude, y = Latitude, size = E.Y1.axes[, 2]),
             col = ”black”, shape = 1) +
  scale_size_continuous(guide = FALSE) +
  ggtitle(expression(paste(”CAP2 of thermal variables, ”, beta[sim])))

E.Y2.axes <- scores(E.Y2.cs2, choices = c(1:3),
                    display = ”lc”, scaling = 1)

E.Y2.plt.axis1 <- ggmap() +
  geom_point(data = sites, aes(x = Longitude, y = Latitude, size = E.Y2.axes[, 1]),
             col = ”black”, shape = 1) +
  scale_size_continuous(guide = FALSE) +
  ggtitle(expression(paste(”CAP1 of thermal variables, ”, beta[sne])))
```

And now I make the remaining panels of Fig. 2 for Y1 and Y2 and the environmental constraining vectors:

```
# The ordiplots in Fig. 2:
E.Y1.scrs <- scores(E.Y1.cs2, display = c(”sp”,”wa”,”lc”,”bp”,”cn”))
E.Y1.df_sites <- data.frame(E.Y1.scrs$constraints)
E.Y1.df_sites$bioreg <- bioreg$bolton
E.Y1.df_sites$section <- seq(1:58)
colnames(E.Y1.df_sites) <- c(”x”, ”y”, ”Bioregion”, ”Section”)

multiplier <- ordiArrowMul(E.Y1.scrs$biplot)
E.Y1.bp <- E.Y1.scrs$biplot * multiplier
E.Y1.bp <- as.data.frame(E.Y1.bp)
E.Y1.bp$labels <- rownames(E.Y1.bp)
colnames(E.Y1.bp) <- c(”x”, ”y”, ”labels”)
E.Y1.bp.sign <- E.Y1.bp[E.Y1.bp$labels %in% E.Y1.sign.ax,]

E.Y1.text <- text.mult(E.Y1.scrs$biplot)
E.Y1.text <- as.data.frame(E.Y1.text)
E.Y1.text$labels <- rownames(E.Y1.text)
```

```r
colnames(E.Y1.text) <- c("x", "y", "labels")
E.Y1.text.sign <- E.Y1.text[E.Y1.text$labels %in% E.Y1.sign.ax,]


E.Y1.p <- ggplot(data = E.Y1.df_sites, aes(x, y, colour = Bioregion)) +
  geom_point(size = 4.0) +
  geom_text(aes(label = Section), size = 3.0, col = "white") +
  geom_segment(data = E.Y1.bp.sign,
               aes(x = 0, y = 0, xend = x, yend = y),
               arrow = arrow(length = unit(0.2, "cm")),
               color = "red", alpha = 1, size = 0.7) +
  geom_text(data = as.data.frame(E.Y1.text.sign),
            aes(x, y, label = rownames(E.Y1.text.sign)),
            color = "black") +
  xlab("CAP1") + ylab("CAP2") +
  ggtitle(expression(paste("Thermal variables and ", beta[sim]))) +
  theme_grey() +
  theme(panel.grid.minor = element_blank(),
        legend.position = "none",
        aspect.ratio = 0.8)


E.Y2.scrs <- scores(E.Y2.cs2, display = c("sp","wa","lc","bp","cn"))
E.Y2.df_sites <- data.frame(E.Y2.scrs$constraints)
E.Y2.df_sites$bioreg <- bioreg$bolton
E.Y2.df_sites$section <- seq(1:58)
colnames(E.Y2.df_sites) <- c("x", "y", "Bioregion", "Section")


multiplier <- ordiArrowMul(E.Y2.scrs$biplot, fill = 0.45)
E.Y2.bp <- E.Y2.scrs$biplot * multiplier
E.Y2.bp <- as.data.frame(E.Y2.bp)
E.Y2.bp$labels <- rownames(E.Y2.bp)
colnames(E.Y2.bp) <- c("x", "y", "labels")
E.Y2.bp.sign <- E.Y2.bp[E.Y2.bp$labels %in% E.Y2.sign.ax,]


E.Y2.text <- text.mult(E.Y2.scrs$biplot, fill = 0.45)
E.Y2.text <- as.data.frame(E.Y2.text)
E.Y2.text$labels <- rownames(E.Y2.text)
colnames(E.Y2.text) <- c("x", "y", "labels")
E.Y2.text.sign <- E.Y2.text[E.Y2.text$labels %in% E.Y2.sign.ax,]


E.Y2.p <- ggplot(data = E.Y2.df_sites, aes(x, y, colour = Bioregion)) +
  geom_point(size = 4.0) +
  geom_text(aes(label = Section), size = 3.0, col = "white") +
  geom_segment(data = E.Y2.bp.sign,
               aes(x = 0, y = 0, xend = x, yend = y),
               arrow = arrow(length = unit(0.2, "cm")),
               color = "red", alpha = 1, size = 0.7) +
  geom_text(data = as.data.frame(E.Y2.text.sign),
            aes(x, y, label = rownames(E.Y2.text.sign)),
            color = "black") +
  xlab("CAP1") + ylab("CAP2") +
  ggtitle(expression(paste("Thermal variables and ", beta[sne]))) +
  theme_grey() +
  theme(panel.grid.minor = element_blank(),
        legend.position = c(.80, .75),
        aspect.ratio = 0.8)
```

Here I now assemble the various panels into what we see produced in Fig. 2 in the paper:

```
pdf("Fig2.pdf", width = 9, height = 8)
grid::grid.newpage()
grid::pushViewport(grid::viewport(layout = grid::grid.layout(2,2)))
vplayout <- function(x, y) grid::viewport(layout.pos.row = x, layout.pos.col = y)
print(E.Y1.p, vp = vplayout(1,1))
print(E.Y2.p, vp = vplayout(1,2))
print(S.Y1.p, vp = vplayout(2,1))
print(S.Y2.p, vp = vplayout(2,2))
dev.off()
```

And I do the same with assembling the panels that form Fig. 3 in the paper:

```
pdf("Fig3.pdf", width = 9, height = 7)
grid::grid.newpage()
grid::pushViewport(grid::viewport(layout = grid::grid.layout(3,2)))
vplayout <- function(x, y) grid::viewport(layout.pos.row = x, layout.pos.col = y)
print(E.Y1.plt.axis1, vp = vplayout(1,1))
print(E.Y1.plt.axis2, vp = vplayout(1,2))
print(E.Y2.plt.axis1, vp = vplayout(2,1))
print(S.Y1.plt.axis1, vp = vplayout(3,1))
print(S.Y2.plt.axis1, vp = vplayout(3,2))
dev.off()
```

Partitioning of variance

Lastly, using **vegan**'s varpart() function, I partition the variance between the MEM variables and the thermal variables (Peres-Neto and Legendre, 2010; Peres-Neto et al., 2006).

```
# These lines of code produce a few figures to visually understand
# the variance partitioning on Y1:
vp2.Y1 <- varpart(Y1, E.Y1.red, S.Y1.red)
par(mfrow = c(1, 2))
showvarparts(2, c("Environ-\nment","","Spatial",""))
plot(vp2.Y1, digits = 2)
par(mfrow = c(1, 1))

# Now I test the significant fractions [a], [b] and [c]...
ES.Y1.red <- cbind(E.Y1.red, S.Y1.red)

# Fraction E | S; pure environmental, i.e. [a]:
anova.cca(capscale(Y1 ~ augMean + febRange + febSD + augSD +
                Condition(MEM1 + MEM2 + MEM3 + MEM4 + MEM5 +
                          MEM6 + MEM7 + MEM8 + MEM9 + MEM10 +
                          MEM13 + MEM15 + MEM16 +
                          MEM18 + MEM19 + MEM20),
            data = ES.Y1.red), parallel = 4, step = 1000)

# Fraction S | E; pure spatial, i.e. [c]:
anova.cca(capscale(Y1 ~ MEM1 + MEM2 + MEM3 + MEM4 + MEM5 +
                          MEM6 + MEM7 + MEM8 + MEM9 + MEM10 +
                          MEM13 + MEM15 + MEM16 +
                          MEM18 + MEM19 + MEM20 +
                Condition(augMean + febRange + febSD + augSD),
            data = ES.Y1.red), parallel = 4, step = 1000)

# Fraction E; environmental, i.e. [a] + [b]:
anova.cca(capscale(Y1 ~., E.Y1.red), parallel = 4, step = 1000)
```

```r
# Fractions S; spatial, i.e. [b] + [c]:
anova.cca(capscale(Y1 ~., S.Y1.red), parallel = 4, step = 1000)

# Fractions E + S; spatial and environmental, i.e. [a] + [b] + [c]:
anova.cca(capscale(Y1 ~., cbind(E.Y1.red, S.Y1.red)), parallel = 4, step = 1000)

# And now the partitioning of the variance in Y2:
(vp2.Y2 <- varpart(Y2, E.Y2.red, S.Y2.red))
par(mfrow = c(1, 2))
showvarparts(2, c("Environ-\nment","","Spatial",""))
plot(vp2.Y2, digits = 2)
par(mfrow = c(1, 1))

# Tests the significant fractions [a], [b] and [c]...
ES.Y2.red <- cbind(E.Y2.red, S.Y2.red)

# Fraction E | S; pure environmental, i.e. [a]:
anova.cca(capscale(Y2 ~ annMean +
                Condition(MEM1 + MEM2 + MEM3 + MEM5),
             data = ES.Y2.red), parallel = 4, step = 1000)

# Fraction S | E; pure spatial, i.e. [c]:
anova.cca(capscale(Y2 ~ MEM1 + MEM2 + MEM3 + MEM5 +
                Condition(annMean),
             data = ES.Y2.red), parallel = 4, step = 1000)

# Fraction E; environmental, i.e. [a] + [b]:
anova.cca(capscale(Y2 ~., E.Y2.red), parallel = 4, step = 1000)

# Fractions S; spatial, i.e. [b] + [c]:
anova.cca(capscale(Y2 ~., S.Y2.red), parallel = 4, step = 1000)

# Fractions E + S; spatial and environmental, i.e. [a] + [b] + [c]:
anova.cca(capscale(Y2 ~., cbind(E.Y2.red, S.Y2.red)), parallel = 4, step = 1000)
```

Appendix C

Network graph of β-diversity

I delved deeper into the patterns of β-diversity by examining the properties of the full dissimilarity matrix, which gives regional β-diversity mentioned above. This matrix describes all pairwise combinations of sections (58² − 1 = 3363), and as such gives us a regional perspective (Anderson et al., 2013). The usual visualisation approach is to plot the dissimilarity metric as a function of geographical distance along the gradient or with respect to the distance between corresponding pairs of sections (Davidar et al., 2007; *e.g.* Nekola et al., 1999); these visualisations are provided here. The plots of dissimilarities were colour-coded according to the bioregion to which the section pairs belong (the Benguela Marine Province (BMP; **1–17**), the Benguela-Agulhas Transition Zone (B-ATZ; **18–22**), the Agulhas Marine Province (AMP; **19–43/44** — the location of this transition is somewhat uncertain at this stage) and the East Coast Transition Zone (ECTZ; **44/45–58**) (*sensu* Bolton and Anderson, 2004) to distinguish bioregional properties of species distribution from the wider geographical scale structure along the whole coastline. In doing so, the change in β-diversity per unit of separating distance between sections (km^{-1}) could be calculated for each bioregion using linear regression. Since the connectivity between sections is constrained by their location along the shore, I calculated the distances between sections not as 'as the crow flies' distances (*e.g.* Section 1 is not connected in a straight line to Section 58 because of the intervening land in-between), but as the great circle geodesic distances between each pair of sections along a network of connected sections (vertices on a network graph). In other words, travelling from Section 1 to Section 58 requires travelling first along the coast through Section 2, then Section 3, and eventually all the way up to Section 58. The total distance between a pair of arbitrary sections is therefore the cumulative sum of the great circle distances between each consecutive pair of intervening sections along the 'route'. This information is encapsulated as a square geodesic distance matrix, and can supply the distance along the abscissa against which species dissimilarities are plotted along the ordinate. The plots showing the relationship between β-diversity with distance are limited because they do not provide a geographical context. To overcome this problem, I relied on a visualisation technique not commonly found in biogeographical studies to explicitly provide the geographical context. I structured the sections as vertices of a network graph and assigned to them their geographical coordinates to force a familiar layout of the graph — when plotted on geographic coordinates, the sections form a map of South Africa. The species dissimilarities were assigned as edge weights (the lines connecting the 58 coastal sections) between pairs of sections, and added to the map. The weights are directly proportional to the thickness of the edges, and colours assigned to vertices (points, or the 58 coastal sections) cluster the sections into their bioregions. Initially I used the **igraph** package that many people rave about, but I found it bothersome. So I devised a cunning way to create network graphs from scratch with some **dplyr** and **ggplot2** magick. I suppose that if I really wanted to I could have made neat functions here (and elsewhere) to reduce some of the repetitive nature of my code, but I really couldn't be bother doing that.

```r
# Visualise the pairwise dissimilarities as network graphs where the
# vertices are geographical coordinates and the edge lengths are the geodesic
# distances.
# These visualisations appear in the paper as Fig. 4.
colnames(sites) <- c("lat", "lon")
sites <- cbind(data.frame(site = seq(1:58)), sites)

Y1.sl <- as.data.frame(expand.grid(seq(1:58), seq(1:58)))
colnames(Y1.sl) <- c("to", "from")

Y2.sl <- Y1.sl

# First Y1:
Y1.sl$Y1 <- as.vector(Y1)

Y1.sl.BMP <- Y1.sl %>%
  dplyr::left_join(., sites, by = c("to" = "site")) %>%
  dplyr::left_join(., sites, by = c("from" = "site")) %>%
  dplyr::filter(Y1 <= 0.5 & Y1 != 0) %>%
  dplyr::filter(from != to & from <= 16)
```

```
Y1.sl.BATZ <- Y1.sl %>%
  dplyr::left_join(., sites, by = c("to" = "site")) %>%
  dplyr::left_join(., sites, by = c("from" = "site")) %>%
  dplyr::filter(Y1 <= 0.5 & Y1 != 0) %>%
  dplyr::filter(from != to & from > 16 & from <= 21)

Y1.sl.AMP <- Y1.sl %>%
  dplyr::left_join(., sites, by = c("to" = "site")) %>%
  dplyr::left_join(., sites, by = c("from" = "site")) %>%
  dplyr::filter(Y1 <= 0.5 & Y1 != 0) %>%
  dplyr::filter(from != to & from > 21 & from <= 41)

Y1.sl.ECTZ <- Y1.sl %>%
  dplyr::left_join(., sites, by = c("to" = "site")) %>%
  dplyr::left_join(., sites, by = c("from" = "site")) %>%
  dplyr::filter(Y1 <= 0.5 & Y1 != 0) %>%
  dplyr::filter(from != to & from > 41)

Y1.sl <- rbind(Y1.sl.BMP, Y1.sl.BATZ, Y1.sl.AMP, Y1.sl.ECTZ)

# and then Y2:
Y2.sl$Y2 <- as.vector(Y2)

Y2.sl.BMP <- Y2.sl %>%
  dplyr::left_join(., sites, by = c("to" = "site")) %>%
  dplyr::left_join(., sites, by = c("from" = "site")) %>%
  dplyr::filter(Y2 <= 0.5 & Y2 != 0) %>%
  dplyr::filter(from != to & from <= 16)

Y2.sl.BATZ <- Y2.sl %>%
  dplyr::left_join(., sites, by = c("to" = "site")) %>%
  dplyr::left_join(., sites, by = c("from" = "site")) %>%
  dplyr::filter(Y2 <= 0.5 & Y2 != 0) %>%
  dplyr::filter(from != to & from > 16 & from <= 21)

Y2.sl.AMP <- Y2.sl %>%
  dplyr::left_join(., sites, by = c("to" = "site")) %>%
  dplyr::left_join(., sites, by = c("from" = "site")) %>%
  dplyr::filter(Y2 <= 0.5 & Y2 != 0) %>%
  dplyr::filter(from != to & from > 21 & from <= 41)

Y2.sl.ECTZ <- Y2.sl %>%
  dplyr::left_join(., sites, by = c("to" = "site")) %>%
  dplyr::left_join(., sites, by = c("from" = "site")) %>%
  dplyr::filter(Y2 <= 0.5 & Y2 != 0) %>%
  dplyr::filter(from != to & from > 41)

# Load coastline
load("data/coast.RData")

sa_lats <- c(-38, -26); sa_lons <- c(14, 34)

net.plot.Y1 <- function(dissim = NULL, title = NULL, col.seq = NULL) {
  ggplot(dissim, aes(lon.x, lat.x)) +
    geom_polygon(data = south_africa_coast,
                 aes(x = long, y = lat, group = group),
                 show.legend = FALSE, fill = "#F9FAEC") +
```

```
    geom_curve(aes(xend = lon.y, yend = lat.y, col = Y1, alpha = (1-Y1)-0.4),
              curvature = 0.3) +
    geom_point(data = sites, aes(x = lon, y = lat, fill = bioreg$bolton),
              col = "black", shape = 21) +
    scale_fill_manual(breaks = c("AMP", "B-ATZ", "BMP", "ECTZ"),
                      values = col.seq, name = "bioregion", guide = FALSE) +
    scale_colour_gradient(name = expression(paste(beta[sim])),
                          low = "black", high = "red") +
    coord_fixed(ratio = 1, expand = TRUE) +
    scale_x_continuous(breaks = seq(15, 35, 5),
                       labels = scales::unit_format("°E", sep = "")) +
    scale_y_continuous(breaks = seq(-35, -25, 5),
                       labels = c("35°S", "30°S", "25°S")) +
    scale_alpha_continuous(guide = FALSE) +
    theme_grey() + xlab(NULL) + ylab(NULL) +
    theme(panel.grid.minor = element_blank()) +
    ggtitle(title)}

a <- net.plot.Y1(Y1.sl.BMP, "Benguela Marine Province",
                col.seq = c("black", "black", "white", "black")) + # alphabetical
  theme(legend.direction = "horizontal",
        legend.position = c(x = 0.5, y = 0.8),
        legend.key.height = unit(0.3, "cm"))
b <- net.plot.Y1(Y1.sl.BATZ, "Benguela-Agulhas Transition Zone",
                col.seq = c("black", "white", "black", "black")) +
  theme(legend.position = "none")
c <- net.plot.Y1(Y1.sl.AMP, "Agulhas Marine Province",
                col.seq = c("white", "black", "black", "black")) +
  theme(legend.position = "none")
d <- net.plot.Y1(Y1.sl.ECTZ, "East Coast Transition Zone",
                col.seq = c("black", "black", "black", "white")) +
  theme(legend.position = "none")

pdf("Fig4.pdf", width = 7, height = 3.8)
grid::grid.newpage()
grid::pushViewport(grid::viewport(layout = grid::grid.layout(2,2)))
vplayout <- function(x, y) grid::viewport(layout.pos.row = x, layout.pos.col = y)
print(a, vp = vplayout(1,1))
print(b, vp = vplayout(1,2))
print(c, vp = vplayout(2,1))
print(d, vp = vplayout(2,2))
dev.off()
```

And that's it, folks. You'll notice that I haven't reproduced Fig. 5 here. I'll leave that up to you… or ask me and I'll send the code. All of the matrices have (mostly) been calculated above and they can be used together with some **dplyr** and **ggplot2** know-how in the creation of that graph.

Legalise seaweed!

References

Anderson, M. J., Tolimieri, N., and Millar, R. B. (2013). Beta diversity of demersal fish assemblages in the North-Eastern Pacific: interactions of latitude and depth. *PLOS ONE* 8, e57918.

Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography* 19, 134–143.

Baselga, A., Orme, D., Villeger, S., Bortoli, J. D., and Leprieur, F. (2013). *betapart: Partitioning beta diversity into*

turnover and nestedness components. Available at: http://CRAN.R-project.org/package=betapart.

Blanchet, F. G., Legendre, P., and Borcard, D. (2008). Forward selection of explanatory variables. *Ecology* 89, 2623–2632.

Bolton, J. J. (1986). Marine phytogeography of the Benguela upwelling region on the west coast of southern Africa: A temperature dependent approach. *Botanica Marina* 29, 251–256.

Bolton, J. J., and Anderson, R. J. (2004). "Marine Vegetation," in *Vegetation of southern africa*, eds. R. M. Cowling, D. M. Richardson, and S. M. Pierce (Cambridge University Press), 348–370.

Bolton, J. J., and Stegenga, H. (2002). Seaweed species diversity in South Africa. *South African Journal of Marine Science* 24, 9–18.

Borcard, D., Gillet, F., and Legendre, P. (2011). *Numerical Ecology with R*. Springer New York Available at: https://books.google.co.za/books?id=dtQNxsH4Y2wC.

Davidar, P., Rajagopal, B., Mohandass, D., Puyravaud, J.-P., Condit, R., Wright, S., et al. (2007). The effect of climatic gradients, topographic variation and species traits on the beta diversity of rain forest trees. *Global Ecology and Biogeography* 16, 510–518.

De Clerck, O., Bolton, J. J., Anderson, R. J., and Coppejans, E. (2005). Guide to the seaweeds of KwaZulu-Natal. *Scripta Botanica Belgica* 33, 294 pp.

Dray, S., Legendre, P., and Peres-Neto, P. R. (2006). Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling* 196, 483–493.

Dray, S., Pélissier, R., Couteron, P., Fortin, M. J., Legendre, P., Peres-Neto, P. R., et al. (2012). Community ecology in the age of multivariate multiscale spatial analysis. *Ecological Monographs* 82, 257–275.

Nekola, J. C., White, P. S., Carolina, N., Nekola, C., and Curriculum, P. S. W. (1999). The distance decay of similarity in biogeography and ecology. *Journal of Biogeography* 26, 867–878.

Peres-Neto, P. R., and Legendre, P. (2010). Estimating and controlling for spatial structure in the study of ecological communities. *Global Ecology and Biogeography* 19, 174–184.

Peres-Neto, P. R., Legendre, P., Dray, S., and Borcard, D. (2006). Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology* 87, 2614–2625.

Sauer, J. D. (1988). *Plant migration: The dynamics of geographic patterning in seed plant species*. University of California Press.

Smit, A. J., Roberts, M., Anderson, R. J., Dufois, F., Dudley, S. F. J., Bornman, T. G., et al. (2013). A coastal seawater temperature dataset for biogeographical studies: large biases between *in situ* and remotely-sensed data sets around the coast of South Africa. *PLOS ONE* 8, e81944.

Stegenga, H., Bolton, J. J., and Anderson, R. J. (1997). Seaweeds of the South African west coast. *Contributions of the Bolus Herbarium* 18, 3–637.

Wernberg, T., Thomsen, M. S., Connell, S. D., Russell, B. D., Waters, J. M., Zuccarello, G. C., et al. (2013). The footprint of continental-scale ocean currents on the biogeography of seaweeds. *PLOS ONE* 8, e80168.