Supplementary material 3

The document provides R code to handle the trait data in Supplementary material 2. From data loading and multivariate ordination, two main procedures are detailed: functional group extraction and functional niche breadth computation. We also provide a facility for empirical case studies.

1 Data loading

Load the data sheets included in Supplementary material 3.

```
data <- read.csv("TraitData.csv", h = T)
lab <- read.csv("TraitLabels.csv", h = T)
```

The object `data` is a species × traits data frame with taxonomy from the phylum to the species level recently updated (WoRMS Editorial Board, 2023) as the six first columns. The object `lab` is a data frame containing trait and trait modality labels, and enabling the manipulation of traits as column blocks; there are 15 traits identified in the column `lab$Column.name` and coded in `lab$Trait.code`. Isolate the data from taxonomy in the object `tab`.

```
tab <- data[7:ncol(data)]
rownames(tab) <- data$Species
```

Load the ade4 package. This package offers a large diversity of methods for one-table (Chessel et al., 2004) and multi-table ordinations (Dray et al., 2007).

```
require(ade4)
```

2 Multiple Factor Analysis

Multiple Factor Analysis (MFA; Escofier and Pagès, 1994; Abdi et al., 2013) is a K-table ordination technique (Thioulouse et al., 2018). A K-table is a series of matrices that match each other through rows or columns. Then, a specific ordination technique ("multi-table ordination") coordinates the K multivariate ordinations of the K matrices within a same system of multivariate axes. Firstly, create a list of 15 trait analyses (each of the 15 traits as a separate matrix); then, each trait matrix is processed by Fuzzy Correspondence Analysis (Chevénet et al., 1994).

```
k <- list()
for(i in 1:15){
  #Each trait is fuzzy coded
  fuz <- prep.fuzzy.var(tab[lab$Trait.code == i], table(lab$Trait.code)[i])
  #Fuzzy Correspondence Analysis
  k[[i]] <- dudi.fca(fuz, scan = F)
}
k <- ktab.list.dudi(k, tabnames = as.character(unique(lab$Trait)))
```

Execute the MFA. The option `"lambda1"` indicates that each matrix is weighted by the inverse of the first eigenvalue of its analysis. This operation provides an optimal solution to correlate column blocks (here, traits) and to balance the roles of blocks with different numbers of columns (Escofier and Pagès, 1983).

```
mfa1 <- mfa(k, option = "lambda1", scan = F, nf = 7)
```

`mfa1` is a list of different objects. Among them, `mfa1$li` contains species axis scores. 7 axes are kept for subsequent steps according to the eigenvalue diagram (Beauchard et al., 2023).

```
> head(round(mfa1$li, 2))
                      Axis1 Axis2 Axis3 Axis4 Axis5 Axis6 Axis7
Abludomelita obtusata  1.21 -0.42 -2.09 -0.18  0.04 -0.54 -0.12
Abra alba             -0.56  0.34  0.58 -3.33 -0.23 -2.06 -0.19
Abra nitida           -0.47  0.55  0.49 -3.41 -0.43 -2.04 -0.56
Abra prismatica       -0.47  0.55  0.49 -3.41 -0.43 -2.04 -0.56
Abra segmentum         0.69 -0.20  0.57 -3.20 -0.13 -1.31 -0.72
Abra tenuis           -0.47  0.55  0.49 -3.41 -0.43 -2.04 -0.56
```

3 Functional groups

Functional groups are derived from species axis scores. Create a Euclidean distance matrix and perform a clustering based on Ward's aggregation criterion (Murtagh and Legendre, 2014).

```
d <- dist(mfa1$li)
h <- hclust(d, "ward.D2")
grp <- cutree(h, 15)
```

Edit the vector `grp` and attribute the functional group labels of Beauchard et al. (2023) in the corresponding order from group 1 to group 15.

```
lab.grp <- c("SurfDiff",
             "ShalShel",
             "MinAbr",
             "SesBiot",
             "EpiLarge",
             "Deep3D",
             "EpiErect",
             "MajBiot",
             "SmallTub",
             "DeepTub",
             "Fouler",
             "MinBiot",
             "EpiSmal",
             "MajAbr",
             "Borer")
lab.grp <- rep(lab.grp, table(grp))
names(lab.grp) <- data$Species[order(grp)]
lab.grp <- lab.grp[order(names(lab.grp))]
```

```
> data.frame(lab.grp)[1:30, drop = F,]
                             lab.grp
Abludomelita obtusata       SurfDiff
Abra alba                   ShalShel
Abra nitida                 ShalShel
Abra prismatica             ShalShel
Abra segmentum              ShalShel
Abra tenuis                 ShalShel
Acanthocardia echinata      ShalShel
Acanthocardia tuberculata   ShalShel
Acanthochitona crinita        MinAbr
Acanthochitona fascicularis   MinAbr
Acanthodoris pilosa         SurfDiff
Acanthostepheia malmgreni   SurfDiff
Achelia echinata            SurfDiff
```

```
Acrocnida brachiata          SesBiot
Acteon tornatilis            SesBiot
Actinia equina              EpiLarge
Adalaria proxima            SurfDiff
Aeolidia papillosa          SurfDiff
Aequipecten opercularis     ShalShel
Afruca tangeri                Deep3D
Agelas oroides              EpiLarge
Alcyonium acaule            EpiLarge
Alcyonium coralloides       EpiErect
Alcyonium digitatum         EpiLarge
Alcyonium glomeratum        EpiLarge
Alcyonium hibernicum        EpiLarge
Alcyonium palmatum          EpiLarge
Alderia modesta             SurfDiff
Alitta succinea              MajBiot
Alitta virens                MajBiot
```

4 Functional niche breadth

The object `mfa1$li` is a weighted average score of the object `mfa1$lisup` which returns the 15 species axis scores along each MFA axis. Here, the first 5 species for the first trait:

```
> head(round(mfa1$lisup, 2))
```

```
                                                Fac1  Fac2  Fac3  Fac4 Fac5  Fac6  Fac7
Abludomelita obtusata.Substratum depth distribution -0.15  0.14 -0.13 -0.21 0.05 -0.08 -0.13
Abra alba.Substratum depth distribution             -1.06 -0.45 -0.11 -0.51 0.55 -0.35  0.21
Abra nitida.Substratum depth distribution           -0.86  0.01 -0.30 -0.64 0.28 -0.32 -0.24
Abra prismatica.Substratum depth distribution       -0.86  0.01 -0.30 -0.64 0.28 -0.32 -0.24
Abra segmentum.Substratum depth distribution        -0.86  0.01 -0.30 -0.64 0.28 -0.32 -0.24
```

The 15 scores of the first species:

```
> round(mfa1$lisup[mfa1$TL$L == "Abludomelita obtusata",], 2)
```

```
                                                Fac1  Fac2  Fac3  Fac4  Fac5  Fac6  Fac7
Abludomelita obtusata.Substratum depth distribution -0.15  0.14 -0.13 -0.21  0.05 -0.08 -0.13
Abludomelita obtusata.Biodiffusion                  -0.05 -0.09 -0.44 -0.50  0.10 -0.80  0.03
Abludomelita obtusata.Downward conveying             0.21 -0.09  0.05 -0.01 -0.02  0.00 -0.02
Abludomelita obtusata.Upward conveying               0.11  0.09 -0.05 -0.05  0.06  0.00  0.07
Abludomelita obtusata.Regeneration                   0.02  0.12 -0.01 -0.06  0.07  0.01  0.07
Abludomelita obtusata.Bioerosion                     0.00 -0.07 -0.11  0.12  0.09 -0.23  0.01
Abludomelita obtusata.Biodeposition                  0.06 -0.36 -0.51 -0.08 -0.09  0.16 -0.16
Abludomelita obtusata.Biostabilisation               0.18 -0.13  0.02 -0.03  0.04  0.02  0.00
Abludomelita obtusata.Ventilation/Pumping            0.45  0.14  0.01  0.00 -0.09  0.04 -0.07
Abludomelita obtusata.Endo-bioconstruction type      0.48  0.13 -0.01  0.00 -0.12 -0.14 -0.02
Abludomelita obtusata.Endo-bioconstruction depth     0.48  0.13 -0.01  0.00 -0.12 -0.09 -0.02
Abludomelita obtusata.Endo-bioconstruction width     0.48  0.13  0.00  0.01 -0.12 -0.08 -0.02
Abludomelita obtusata.Epi-bioconstruction type       0.24 -0.35 -0.96  0.17  0.06  0.17  0.03
Abludomelita obtusata.Epi-bioconstruction extension  0.24 -0.35 -0.96  0.17  0.06  0.17  0.03
Abludomelita obtusata.Epi-bioconstruction size       0.24 -0.35 -0.96  0.17  0.06  0.17  0.04
```

For each species, calculation of the 7 axis score variances.

```
niche <- apply(mfa1$lisup, 2, function(x) tapply(x, mfa1$TL$L, var))
> head(round(niche, 2))
```

```
                     Fac1 Fac2 Fac3 Fac4 Fac5 Fac6 Fac7
Abludomelita obtusata 0.04 0.04 0.15 0.03 0.01 0.06 0.00
Abra alba             0.38 0.10 0.19 0.28 0.08 0.06 0.05
Abra nitida           0.35 0.08 0.20 0.29 0.06 0.06 0.05
Abra prismatica       0.35 0.08 0.20 0.29 0.06 0.06 0.05
Abra segmentum        0.11 0.02 0.18 0.30 0.02 0.06 0.01
```

```
Abra tenuis           0.35 0.08 0.20 0.29 0.06 0.06 0.05
```

Functional niche breadth as the row sum:

```
niche <- apply(niche, 1, sum)
> data.frame(niche)[1:20, drop = F,]
                                 niche
Abludomelita obtusata       0.3359305
Abra alba                   1.1367663
Abra nitida                 1.0855359
Abra prismatica             1.0855359
Abra segmentum              0.7019672
Abra tenuis                 1.0855359
Acanthocardia echinata      0.7368649
Acanthocardia tuberculata   0.7368649
Acanthochitona crinita      1.1024077
Acanthochitona fascicularis 1.0698649
Acanthodoris pilosa         0.3394978
Acanthostepheia malmgreni   0.4648499
Achelia echinata            0.3485928
Acrocnida brachiata         2.5011771
Acteon tornatilis           1.7326316
Actinia equina              0.6469424
Adalaria proxima            0.3394978
Aeolidia papillosa          0.3229695
Aequipecten opercularis     0.6569418
Afruca tangeri              4.0007271
```

5 Averaging trait scores in empirical case studies

In our 13 case studies, we related habitat descriptors and trawling intensity (matrix R) to trait data (matrix Q) through RLQ analysis. Matrix L (samples × taxa) is the link in this procedure. However, our trait data are defined at the species level whereas taxa are not always identified at this level in field data. Therefore, we provide here a command that automatises trait data averaging for taxa identified at levels higher than the species level. Load the command.

```
source("…/avg_trait.R")
> avg.trait
avg.trait <- function(taxocenosis, taxonomy, traits, labels,
                      which.traits = unique(labels$Trait.code),
                      find.upper = FALSE, check.block = TRUE,
                      ori.col.names = FALSE){
  tab.avg <- traits
  for(i in 1:5){
    fac <- factor(taxonomy[,i])
    w <- apply(traits, 2, function(x) tapply(x, fac, mean))
    w <- data.frame(w)
    tab.avg <- rbind(tab.avg, w)
  }
  tab.res <- tab.avg[rownames(tab.avg) %in% taxocenosis, drop = F,]
  tab.tax <- data.frame()
  for(i in 6:1){
    w <- taxonomy[taxonomy[,i] %in% taxocenosis,]
    w$tax.lev <- rep(7-i, nrow(w))
    if(i < 6 & length(w[,i]) > 0)
      w[(i+1):6] <- ""
    w$Species <- w[,i]
    tab.tax <- unique(rbind(tab.tax, w))
  }
  not.found <- taxocenosis[taxocenosis %in% rownames(tab.res) == F]
  if(length(not.found) > 0 & find.upper == T){
    for(i in 1:length(not.found)){
      w <- strsplit(not.found[i], " ")[[1]]
      if(length(w) > 1){
        genus <- w[[1]][1]
```

```
        w <- taxonomy[taxonomy[,5] %in% genus,]
        if(nrow(w) > 0){
          w <- w[1,]
          w$Species <- not.found[i]
          w$tax.lev <- 2
          tab.tax <- unique(rbind(tab.tax, w))
          w <- tab.avg[rownames(tab.avg) == genus, drop = F,]
          rownames(w) <- not.found[i]
          tab.res <- rbind(tab.res, w)
        }
      }
    }
  }
  colnames(tab.tax)[6] <- "Taxon"
  tab.tax <- tab.tax[order(tab.tax$Taxon),]
  rownames(tab.tax) <- 1:nrow(tab.tax)
  tab.res <- tab.res[order(rownames(tab.res)), drop = F,]
  tab.res <- tab.res[labels$Trait.code %in% which.traits]
  tab.res <- tab.res[apply(tab.res, 2, sum) > 0]
  if(check.block == TRUE){
    newlabels <- labels[labels$Column.name %in% colnames(tab.res),]
    w <- table(newlabels$Trait.code)
    w <- as.numeric(names(w)[which(w > 1)])
    newlabels <- newlabels[newlabels$Trait.code %in% w,]
    tab.res <- tab.res[colnames(tab.res) %in% newlabels$Column.name]
    labels <- labels[labels$Trait.code %in% which.traits,]
    w <- labels$Trait.code[labels$Trait.code %in% newlabels$Trait.code == F]
    if(length(w) > 0){
      warning <- unique(labels$Trait[labels$Trait.code %in% w])
      warning <- paste(warning, collapse = ", ")
      warning <- paste("Warning: no variation in ", warning, sep = "")
      cat(warning, "\n")
      cat("Data discarded", "\n")
      if(length(which.traits) == 1) stop("No output", call. = F)
    }
  }
  if(ori.col.names == FALSE){
    lab <- labels[labels$Column.name %in% colnames(tab.res),]
    lab$Trait.code <- rep(1:length(unique(lab$Trait.code)), table(lab$Trait.code))
    w <- numeric()
    for(i in 1:max(lab$Trait.code)){
      w <- c(w, 1:table(lab$Trait.code)[i])
    }
    lab$Modality.code <- w
    w <- paste(rep("T", ncol(tab.res)), lab$Trait.code, sep = "")
    w <- paste(w, paste(rep("M", ncol(tab.res)), lab$Modality.code, sep = ""), sep = ".")
    colnames(tab.res) <- w
    lab$Column.name <- w
  }else{
    lab <- labels[labels$Column.name %in% colnames(tab.res),]
  }
  not.found <- not.found[not.found %in% tab.tax$Taxon == F]
  list(tab = tab.res, labels = lab, tax = tab.tax, not.found = not.found)
}
```

This command calculates mean trait modality scores for all taxa present in the entire effect trait data set (object `tab` as previously created), from the phylum to the genus level. The argument `taxocenosis` (character vector) is a set of taxa from different clades at different taxonomic levels from phylum to species, as typically encountered in benthic community data (i.e., column labels of matrix L). Then, the command extracts the raw trait data at the species level or averaged at higher levels for the corresponding elements of the taxocenosis.

The object `taxonomy` is the data frame containing the 6 first columns of the `data` object previously created.

```
> head(data[1:6])
      Phylum        Class     Order    Family        Genus               Species
1 Arthropoda Malacostraca Amphipoda Melitidae Abludomelita Abludomelita obtusata
2   Mollusca     Bivalvia  Cardiida Semelidae         Abra             Abra alba
3   Mollusca     Bivalvia  Cardiida Semelidae         Abra           Abra nitida
4   Mollusca     Bivalvia  Cardiida Semelidae         Abra       Abra prismatica
```

```
5   Mollusca     Bivalvia  Cardiida Semelidae         Abra        Abra segmentum
6   Mollusca     Bivalvia  Cardiida Semelidae         Abra           Abra tenuis
```

```
taxon <- head(data[1:6])
```

The argument `traits` is the data frame `tab` and the argument `labels` the data frame `lab` as initially loaded. `which.traits` is a numeric vector identifying traits desired for analysis in the column `lab$Trait.code`. When `find.upper = TRUE`, the command calculates average trait scores at the genus level for taxa present in `taxocenosis` at the species level but only at the genus level in `traits`.

Let's consider a virtual fauna composed of 6 taxa of different levels as an example:

```
> taxoc <- c("Abra", "Callianassa subterranea",
             "Cnidaria", "Cumacea", "Geodiidae",
             "Golfingia elongata")
```

The 15 available traits and their respective code:

```
> unique(lab[c("Trait.code", "Trait")])
   Trait.code                          Trait
1           1 Substratum depth distribution
6           2                   Biodiffusion
9           3             Downward conveying
12          4               Upward conveying
15          5                   Regeneration
18          6                  Bioerosion
22          7                 Biodeposition
25          8               Biostabilisation
28          9           Ventilation/Pumping
31         10     Endo-bioconstruction type
39         11    Endo-bioconstruction depth
44         12    Endo-bioconstruction width
48         13      Epi-bioconstruction type
56         14 Epi-bioconstruction extension
62         15      Epi-bioconstruction size
```

Suppose that the traits of interest for analysis are "Substratum depth distribution" and "Biodiffusion".

```
w <- avg.trait(taxocenosis = taxoc, taxonomy = taxon,
               traits = tab, labels = lab, which.traits = c(1, 2),
               find.upper = FALSE, check.block = TRUE,
               ori.col.names = FALSE)
```

`w` is a list of 4 objects:

```
> names(w)
[1] "tab"       "labels"    "taxonomy"  "not.found"
```

`w$tab` is the taxa × traits data frame.

```
> round(w$tab, 2)
                        T1.M1 T1.M2 T1.M3 T1.M4 T1.M5 T2.M1 T2.M2 T2.M3
Abra                      0.0  3.00  0.60  0.00     0  0.00  3.00  0.00
Callianassa subterranea   0.0  1.00  2.00  3.00     3  0.00  0.00  3.00
Cnidaria                  2.9  0.19  0.03  0.02     0  2.81  0.19  0.00
Cumacea                   3.0  2.27  0.00  0.00     0  0.14  1.91  0.95
Geodiidae                 3.0  0.00  0.00  0.00     0  3.00  0.00  0.00
```

In column names, “T” means trait and “M” modality, followed by a numerical code. These labels are found in the object `w$lab` that provides explicit labels.

```
> w$lab
  Column.name Trait.code Modality.code                          Trait Modality
1       T1.M1          1             1 Substratum depth distribution        0
2       T1.M2          1             2 Substratum depth distribution      0-5
3       T1.M3          1             3 Substratum depth distribution     5-15
4       T1.M4          1             4 Substratum depth distribution    15-30
5       T1.M5          1             5 Substratum depth distribution      >30
6       T2.M1          2             1                   Biodiffusion     Null
7       T2.M2          2             2                   Biodiffusion      Low
8       T2.M3          2             3                   Biodiffusion     High
```

The object `w$taxonomy` is a data frame that can be used for some verifications.

```
> w$taxonomy
      Phylum        Class          Order         Family       Genus                  Taxon tax.lev
1   Mollusca     Bivalvia        Cardiida      Semelidae        Abra                   Abra       2
2 Arthropoda Malacostraca        Decapoda Callianassidae Callianassa Callianassa subterranea       1
3   Cnidaria                                                                      Cnidaria       6
4 Arthropoda Malacostraca         Cumacea                                        Cumacea       4
5   Porifera Demospongiae Tetractinellida      Geodiidae                        Geodiidae       3
```

The column `w$taxonomy$tax.lev` identifies taxonomic levels from phylum (6) to species (1). Here, averaging trait modality scores for high levels such as Cnidaria (phylum) and Cumacea (order) remains reliable. Most Cnidaria are epibenthic and non-bioturbative, and scores in Cumacea are relatively similar from one species to another. However, this example is deliberately simplistic, and user should be extremely careful when selecting more traits for real taxocenoses that often exceed several tens of taxa in marine benthic case studies. In our case studies, most of times we discarded taxa from the family to the phylum level prior to averaging.

The object `w$not.found` is a vector listing the taxa for which no corresponding documentation was found.

```
> w$not.found
[1] "Golfingia elongata"
```

Although this species is not present in the data set, the sister species *Golfingia vulgaris* does. By assuming that these sister species are functionally similar, scores of *G. vulgaris* can be used to document *G. elongata*. When setting `find.upper = TRUE`, the command attributes averaged scores of species from the same genus present in `traits` and only at the genus level in `taxocenosis`.

```
w <- avg.trait(taxocenosis = taxoc, taxonomy = taxon,
               traits = tab, labels = lab, which.traits = c(1, 2),
               find.upper = TRUE, check.block = TRUE,
               ori.col.name = FALSE)
```

```
> round(w$tab, 2)
                        T1.M1 T1.M2 T1.M3 T1.M4 T1.M5 T2.M1 T2.M2 T2.M3
Abra                      0.0  3.00  0.60  0.00     0  0.00  3.00  0.00
Callianassa subterranea   0.0  1.00  2.00  3.00     3  0.00  0.00  3.00
Cnidaria                  2.9  0.19  0.03  0.02     0  2.81  0.19  0.00
Cumacea                   3.0  2.27  0.00  0.00     0  0.14  1.91  0.95
Geodiidae                 3.0  0.00  0.00  0.00     0  3.00  0.00  0.00
Golfingia elongata        0.0  1.00  3.00  3.00     3  0.00  0.00  3.00

> w$not.found
character(0)
```

If the fauna does not exhibit variation for a selected trait in `which.traits`, the trait is discarded. For instance, the previously illustrated fauna is not variable in bioerosion (trait 6), all the taxa having a unique score for the modality "None". Only trait 1 is kept in the output.

```
> w <- avg.trait(taxocenosis = taxoc, taxonomy = taxon,
                 traits = tab, labels = lab, which.traits = c(1, 6),
                 find.upper = TRUE, check.bock = TRUE,
                 ori.col.names = FALSE)
Warning: no variation in Bioerosion
Data discarded

> round(w$tab, 2)
                        T1.M1 T1.M2 T1.M3 T1.M4 T1.M5
Abra                     0.0  3.00  0.60  0.00     0
Callianassa subterranea  0.0  1.00  2.00  3.00     3
Cnidaria                 2.9  0.19  0.03  0.02     0
Cumacea                  3.0  2.27  0.00  0.00     0
Geodiidae                3.0  0.00  0.00  0.00     0
Golfingia elongata       0.0  1.00  3.00  3.00     3
```

Any 0 column is removed. In that case, the column names of `w$tab` and the content of `w$lab` are relabelled. When `ori.col.names = TRUE`, original labels are kept. In case the object `traits` contains only one trait as a single variable (e.g., species niche breadth), the argument `check.block` must be attributed `FALSE` (`TRUE` by default).

References

Abdi, H., Williams, L. J., and Valentin, D. (2013). Multiple factor analysis: principal component analysis for multitable and multiblock data sets. Wiley Interdiscip. Rev. Comput. Stat. 5, 149–179. doi:10.1002/wics.1246

Beauchard, O., Thompson, M. S. A., Ellingsen, K., Piet, G. J., Laffargue, P., and Soetaert, K. (2023). Assessing sea floor functional biodiversity and vulnerability. Mar. Ecol. Prog. Ser. 708, 21–43. doi: 10.3354/meps14270

Chessel, D., Dufour, A. B., and Thioulouse, J. (2004). The ade4 package-I – One-table methods. R News 4, 5–10. https://cran.r-project.org/doc/Rnews

Chevénet, F., Dolédec, S., and Chessel, D. (1994). A fuzzy coding approach for the analysis of long-term ecological data. Freshw. Biol. 31, 295−309. doi: 10.1111/j.1365-2427.1994.tb01742.x

Dray, S., Dufour, A. B., and Chessel, D. (2007). The ade4 Package – II: Two-table and K-table methods. R News 7, 47–54. https://cran.r-project.org/doc/Rnews

Escofier, B., and Pagès, J. (1983). Méthode pour l'analyse de plusieurs groupes de variables. Application à la caractérisation des vins rouges du Val de Loire. Revue de Statistique Appliquée 31, 43–59.

Escofier, B., and Pagès, J. (1994). Multiple factor analysis (AFMULT package). Comput. Stat. Data Anal. 18, 121–140. doi: 10.1016/0167-9473(94)90135-X

Murtagh F., and Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? J. Classif. 31, 274–295. doi: 10.1007/s00357-014-9161-z

Thioulouse, J., Dray, S., Dufour, A. B., Siberchicot, A., Jombart, T., and Pavoine, S. (2018). Multivariate analysis of ecological data with ade4. New York: Springer.

WoRMS Editorial Board (2023). World Register of Marine Species. https://www.marinespecies.org