Ocean Time series Workshop 15-18 June 2021 - Q&A Questions

Day 1

1) 1-I´m working with a database (south Atlantic angler answer´s) that included several types of numerical predictor variables such as Age (continuous), Gender (binary), Willingness to comply with a rule (High=4, Good=3, Low=2, nothing=0), other categorical and ordinal variables but with different levels, some NA, etc. **What´s the correct transformation and distance/dissimilarity?** Is PCA a good ordination method for this type of data matrix?
2) 2-I´m working with a database (Sub-Antarctic nototheniods fish) that included several types of life history traits such as: length at first maturity, longevity, Growth, reproductive season (summer =1, etc), Gender (binary). **What´s the correct transformation and distance/dissimilarity?** Is NMDS a good ordination method for this type of data matrix?
3) 3- I cannot yet see **the logic behind the use of the PCA axis or PCOA axis as input in other multivariate methods such as DbRDA, NMDS**, etc. For example, if the first two PCA axis are the best maximizing the variance, What´s the new information that provided in a further analysis?
4) 4- **Can the Chord, Log-chord, Hellinger and chi-square transform data used in a PERMANOVA analysis?**
5) **In the section 6.1.1 partial RDA from practical exercises,** Where the formula comes from? What do the variables represent (lambda, centroid.sc1.sites, centroid.sc2.sites, sqrt(n-1)?
6) In the PCoA video, Pierre introduces several correction options (Lingoes, Cailliez and square root correction) to eliminate negative eigenvalues produced during a PCoA of a non-euclidean dissimilarity matrix. **How should one use which correction to apply to a dataset?**
7) 1) Where can we find **the spider data used in the lectures video**s? (in the Chesapeake Bay practical exercises folder)
8) 2) Which algorithms do **the packages {stats} and {vegan} use for the PCA calculation,** EDV or SDV?
9) 3) **In the PCoA lecture PDF, slide 34-36, it is not clear for me how is the abundance data represented in the biplot.** That means, there is an ordination of the distance matrix onto the PCoA space, is it possible to add abundance data in that space?
10) 4) And last, in the video-lecture of canonical analyses, the professor Legendre mentioned a couple of times a **previous lecture on multivariate lineal regression**, that I believed is not included in this worksop. Is there any possibility to have the link to that video?

Day 2

11) How **RDA, MRT and Hierarchical Agglomerative clustering handle NA's in the different matrices?** I haven't seen examples, and should be interesting to know to understand is possible impact once testing space-time interactions.
12) **How does one deal with temporal autocorrelation in partial RDA**, and in general the other models we are learning about?
13) **1) If the permutation test for homogeneity of multivariate dispersions returns a significant p-value (betadisper followed by permutest), i.e. we reject H0 of homogenous variance — what would be the most appropriate course of action? If standard transformations (sqrt, log-chord,...) do not help, would it be best** to identify and remove the replicates which are outliers and thus cause increased dispersion at individual sites.
14) 2) I usually always think in terms of cause and response, and thus the asymmetric methods are more intuitive for me. In the Canonical Analysis lecture, you provide a useful example of using CoIA to analyse water chemistry and physiography matrices. Just to confirm that I understand the potential application of symmetric methods: **would this for example also be useful to compare species matrices Y1 of infauna and Y2 of epifauna in a soft-bottom environment? Neither is explanatory, but correlations may still exist. Or would you say the application of symmetric methods is more common to environmental variables?**
15) 3) To have more freedom in terms of the distance measure used, **would it be sensible to use Canonical Analysis of Principal coordinates (a.k.a CAP; Anderson & Willis 2003) as a symmetric method for canonical analyses?** This is not covered in the lecture, but I believe it can be implemented using BiodiversityR::CAPdiscrim() in R
16) from the Test space-time interaction video. **For the K-means plot; 1. How were the 5 species of Trichoptera selected? 2. Is there a maximum number of species that could have been visualized together?**
17) Within the stimodels function, Ti represents the number of temporal sampling steps or a matrix of temporal coordinates. Lets say you collect data in multiple 3 months blocks, over several years, and you are interested to find out whether there is a space-time interaction over the entire data-set. Within blocks the time interval between sampling is equispaced, but between blocks there can be as little as 1 day to as much as several months. Do I understand it correctly that **you then provide a matrix in which the temporal coordinates reflect this sampling pattern?**
18) 1) As in other methods, **Should the explanatory matrix (e.g. environmental variables) be standarized?**
19) 2) In the example carried out by De'ath (2002), **the Chi-square transformation is used. Doesn't this transformation give too much importance to rare species?** Is correct for this type of euclidean method?
20) 1) In the section 4.2.3, – We test the conservative hypothesis that a polynomial trend through the multivariate data would represent the faunal variation better than a linear trend. If the data are represented better by a poly2 or more, is it still possible to do an RDA?
21) 2) What does this result (significant linear trend) involves for the interpretation of the previous space constrained hierarchical clustering?
22) Is linearity assumption necessary to test before applaying RDA?

23) Except that one is standardized and the other not, I can't see the main difference between Procrustes and co-inertia analyses. For example, in what case should we prefered one to the other?

Day 3

24) How does taxonomic level influence **ordination**?
25) Much has been discussed about looking for interaction between factors, and you mentioned some considerations for the time scales of sampling. **Can you provide some direction on uncoupling aspects of a single factor and using them to relate to response data?** For example, environmental explanatory variables that may be a composite of signals at multiple time scales, such as current meter readings that have tidal signals with an hourly/daily component and a monthly component. **What approach do you suggest for decomposing such data, to then look for correlations in the response data** (e.g. community data) that may relate to one of these components? How would you deal with interaction between the decomposed factors?
26) 1. We can compute LCBDs both from a Y response matrix and D dissimilarity matrix. However, in the case we use D, we cannot compute SCBDs nor the p-values. Therefore, I would always use rather Y instead of D, except if I am interested in a particular dissimilarity index for a given research question. Therefore **I am confused about the actual computation process performed by beta.div(): how and when beta.div() uses Y or D** (e.g. method = "none" does not work in beta.div()).
27) 2. From a dissimilarity matrix, we can only retrieve p-values of LCBD but not the SCBDs. **However, let's imagine we transpose the Y response matrix to have species in rows and sites in columns.** Then we recompute the dissimilarity matrix and the LCBD which may now be interprated as SCBDs. Would that make sense?
28) **2.bis Why can't SCBD be tested for significance with beta.div()?** Is it useless?
29) 1. **Slide 16 - I am confused - which data are being presented here?** The fish data shown on slide 14 are only at 11 sites and do not include a site 17, so it can't be them, but the data on slide 15 only has 10 sites, so it can't be those either....
30) 2. **If LCBDs and SCBDs and BDtotal can be computed from either matrix Y or D, is there an advantage to doing it with one or the other?** I would be inclined to use Y, as then I would not have to worry about the type of dissimilarity I chose.
31) 3. **Can LCBDs and SCBDs be computed between two successive sites** to obtain the results of adjacent/directional beta diversity, which is introduced at the beginning of the lecture?
32) p-value adjustments for multiple testing in temporal beta-diversity analyses: In the BCI example of the lecture, the Holm correction was used. **How do you choose between the different methods (Hommel, Holm, Bonferroni) for p-value adjustments for multiple testing** in TBI analysis? Do you have a general recommendation/preference?
33) From Legendre and De Cáceres 2013: "Type III coefficients (percentage difference) the square root of the distances must be taken before they are used in PCoA. The matrix of principal coordinates can be used as the response data in RDA; this is the distance-based RDA method

proposed by Legendre & Anderson (1999)". I cannot yet see the logic behind the use of principal coordinates matrix in RDA.

34) Could you give examples of methods to compute permutation test that preserves spatial correlation in R? (LBCD lecture)

Day 4

35) Prior to looking at periodograms, you need to detrend the data. Are there any preferred ways to detrend data (e.g., depending on your type of data)?
36) You stated that all three components (autocorrelated signal, trend, noise) of a data series need to be seperated. So first you "detrend" the data, then you analyse the periodic autocorrelated signal using autocorrelograms or periodogramms. Does the second step allow to seperate the periodic signal from the noise? In the examples you showed we only used simmulated data without added noise. On slide 7 you mentioned the seperation of noise as a thrid step in the analysis.
37) You mentionned Dutilleul cannot deal with NAs. However, as it computes a R^2 with a fitted model, I don't understand why since we don't necessarily need equidistant points in time in a linear regression for example.
38) Once we identify the significant periods, how to generate multiple vectors from the univariate data (i.e. one vector for the component of the data related to each period), in order to compare these with community/multivariate data?
39) Do you have a smart solution for resolving directional data (specifically, current direction) in time series correlation analysis? Directional data is either measured in radians (0-2π) or degrees (0-360°), and when using this in its raw/untransformed format, this fails to highlight correlations in the norther sector (e.g. between NNW (355°) and NNE (5°)).
40) Some of the participants in Break Out Room 1 expressed a lot of interest in the topics covered in the time series analysis pre-recorded lecture and are keen to try the techniques discussed in the lecture on their datasets. I looked at the NEwR book but did not find any chapter covering periodograms. Do you know of any training script and/or dataset that you would recommend to the participants who would like to try out the techniques you discuss in the lecture?