

**Electronic Supplement S1: Selected Similarity and Distance Measures Used in Vegetation Ecology**
*Principal Component Analysis (PCA)*

PCA frequently bases on Pearson's correlation coefficient (intrinsically standardizing species vectors), which is nothing else but a vector product (Wildi 2017),

$$s_{j,k} = \sum_{h=1}^p a_{hj} a_{hk}, \quad (1)$$

where  $s_{j,k}$  is a similarity between vectors  $j$  and  $k$ . If vector  $a$  is centred only then a variance-covariance matrix  $S$  results. When  $a$  is standardized then  $S$  is a correlation matrix with range  $-1 \leq s_{j,k} \leq +1$ .

*Principal Coordinates Analysis (PCOA) and Nonmetric Multivariate Scaling (NMDS)*

PCOA and NMDS require distance matrices of multivariate plot data. To achieve the effect of variable standardisation a distance measure can be derived from Pearson's correlation coefficient among plots by taking the one complement and adjusting the range to  $0 \leq d_{j,k} \leq +1$ ,

$$dc = (1 - d_{hk})/2. \quad (2)$$

In the present paper this is called *correlation as distance*,  $dc$ .

Any intrinsic transformation is avoided when choosing Euclidean distance,  $de$ . For sampling units 1, 2 this is

$$de_{1,2} = \sqrt{\sum_{j=1}^p (x_{1j} - x_{2j})^2} \quad (3)$$

where  $p$  are the variables, the species vectors in this case. Within an ordination this is the length of the straight line between two data points. Euclidean distance has no upper bound and it will increase with the number of variables addressed. To limit the range from 0 to the square root of two, the sampling units are normalized and the coefficient thereafter called *chord distance*,  $dc$ , implemented as:

$$dc_{1,2} = \sqrt{2 \left( 1 - \frac{\sum_{j=1}^p x_{1j} x_{2j}}{\sum_{j=1}^p x_{1j}^2 \sum_{j=1}^p x_{2j}^2} \right)} \quad (4)$$

An alternative to Euclidean distance is Manhattan distance,  $dm$ . For any two sampling units it is

$$dm_{1,2} = \sum_{j=1}^p |x_{1j} - x_{2j}| \quad (5)$$

This is the sum of differences occurring on all variables  $p$ .

Another practice is using distance measures adjusting for species sum of scores, correcting for unbalanced weights of species. The Bray-Curtis distance,  $dbc$ , is among the most popular in vegetation science. For sampling units 1, 2 this is

$$dbc_{1,2} = \frac{\sum_{j=1}^p |x_{1j} - x_{2j}|}{\sum_{j=1}^p (x_{1j} + x_{2j})} \quad (6)$$

It follows the principle of Manhattan distance but the resulting distance is weighted by the inverse of total scores occurring in anyone species involved. Bray-Curtis distance is a semi-metric and when used in PCOA the non-metric proportion will result in negative Eigenvalues.

*Canberra distance, dca* bases on the same idea, but it applies weighting to each species individually:

$$dca_{1,2} = \frac{1}{pp} \sum_{j=1}^p \left( \frac{|x_{1j} - x_{2j}|}{(x_{1j} + x_{2j})} \right) \quad (7)$$

In this  $pp$  is the total number of variables (species). Bray-Curtis and Canberra distance have in common that they have range 0 to 1.

## Electronic Supplement S2: Using R Function `ordpanel {dave}`

*Purpose:*

Printing a panel of ordinations of a data frame of vegetation data. Measuring the predictive power of all ordinations based on a measured numerical site factor.

*Call:*

```
o.ordpanel<- ordpanel(vegdata,sitefactor,sf.name,exp.vec,methlist,cxt,dotsz=0.4,lwdd=0.1,
PSURF,mymodel=1)
```

`vegdata:` a data frame of vegetation data, use `vegdata=sveg` to get Figure 1

`sitefactor` numeric site factor, use `sitefactor=ssit$pH.peat` for Figures 2 and 3

`sf.name` name of site factor, use `sf.name="pH.peat"` for Figures 2 and 3

`exp.vec` a vector containing all exponents included in the ordination panel.

For Figure 1 this is `exp.vec <- c(2.50,2.00,1.50,1.25,1.00,0.75,0.50,0.25,0.01)`

This vector can have any length.

`methlist` The method numbers in the panel of ordinations.

Numbers from 1 to 12 are allowed. In Figure 1 this is `methlist = c(1:12)`.

`cxt` Text size for method names in Figure 1, e.g., `cxt = 0.8`

`dotsz` Maximum dot size for scores of 3rd axis, e.g., `dotsz = 1.0`

`lwdd` Line width of dots in the ordinations, e.g., `lwdd = 0.2`

`PSURF` If `PSURF=TRUE` trend surfaces are added to all ordinations in Figure 1 based on parameter `sitefactor`.

`mymodel` A model for analysis (1–4, see methods section)

*Display of Results:*

```
summary(o.ordpanel,...)
```

`o.ordpanel` An object of class "ordpanel". This will do all the evaluations, including the ordinations of ordinations based on the `sitefactor` as specified in function `ordpanel` and `mymodel` chosen therein.