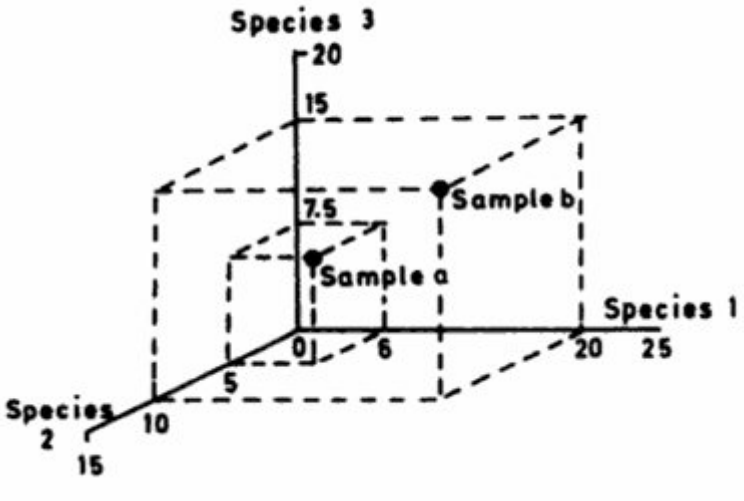
Ordination analysis

Theory

Ordination (from Latin *ordinatio*, putting things into order, or German *die Ordnung*, order) is a multivariate analysis, which arranges units (usually sites) in some order, and often includes plotting objects (points) along one or several axes representing this ordered relationship. Ordination stands in contrast to classification, which also deals with species composition data, but aims to chop the data into groups of sites with similar species composition (while groups are different from each other as much as possible). Ordination aims to find gradients or changes in species composition, handling data as if they are located continually along environmental gradients. Multivariate species composition data can be imagined as samples located in multidimensional hyperspace, where each dimension (axis) is represented by an abundance of one species (an example for two community samples with three species, projected onto 3D space defined by abundances of individual species, is on Fig. 1). Ordination can be applied also on other than species composition data, for example on the matrix of environmental variables (the dimensions of the space, in which the samples are located, is then defined by values of individual environmental variables, creating “environmental space”). The use of ordination in ecology was pioneered by English-born Australian botanist and ecologist David Goodall, whose first paper using an ordination analysis (PCA) was published in 1954 (Goodall 1954).

The main assumption of ordination is that analyzed data are redundant, i.e. they contain more variables (and dimensions) than is necessary to describe the information behind, and we can reduce the number of these dimensions without losing too much information. For example, in the case of species composition data, some of the species are usually ecologically similar (e.g. species that prefer to grow in wet instead of dry habitat), meaning that the dataset contains several redundant variables (species) telling the same story. Or, to explain the redundancy in another way, from occurrence (or absence) of one species we can often predict occurrence (or absence) of several other species (e.g. if the sample includes species of wet habitats, we may expect that species preferring dry habitats will not be present, while other wet-loving species may occur). In the case of ordination applied on the matrix of environmental variables, these are often correlated to each other (e.g. soil measurements of pH are often related to concentrations of Mg and Ca), also allowing for dimension reduction.

Figure 1: Multivariate species space defined by three species, with two samples located according to the abundance of three species. (From Zuur et al. 2007).



Since multidimensional space is not easy to display, describe or even just imagine, it is worth reducing it into a few main dimensions, while preserving maximum information. This also means that if the individual variables are completely independent of each other (e.g. each species have entirely different preferences), then ordination is not likely to find some reasonable reduction of the multidimensional space.

What ordination method does can be formulated in two alternative ways:

1. it searches for gradients in species composition (represented usually by ordination axes) and attempts to explain these gradients by environmental variables; or
2. it searches for the distribution of samples in reduced ordination space which maximally reflects the dissimilarity (= distance) between samples in terms of their species composition.

The first of these ways can be well represented by the algorithm of principal component analysis (PCA), which is searching for the directions in the multidimensional space (where dimensions are sample descriptors, e.g. species) that represent the most variation in the data (see details at the PCA & tb-PCA (linear unconstrained ordination)). These directions then become the ordination axes (individual ordination axes are, from the definition, always not correlated to each other, ie they are perpendicular). If the original data about species composition have high redundancy, most of the information can be represented by the position (*scores*) of samples along the first or first several ordination axes, which then represent the directions of the fastest change in species composition among individual samples (compositional turnover).

The second way can be well understood on the algorithm of non-metric multidimensional scaling (NMDS), which is iteratively searching for the way how to represent the distribution of samples in low dimensional space (usually two dimensions) while preserving the distances between the samples similar to their original compositional distance (see details at PCoA & NMDS (distance-based unconstrained ordination)).

Types of ordination methods

Ordination methods discussed at this website are summarised in Tab. 1. They can be divided according to two criteria: **whether their algorithm includes also environmental variables** along to the species composition data (unconstrained ordination methods do not, while constrained do), and **what type of species composition data is used for analysis** (either raw data (sample-species matrix of species composition), pre-transformed raw data (e.g. using Hellinger transformation), or distance matrix (sample-sample symmetric matrix of distances between samples)).

	(a) Raw-data-based (classical approach)		**(b) Transformation--based**	**(c) Distance-based**
	Linear	**Unimodal**		
(1) Unconstrained (indirect) **Shiny app (how to use, blog)**	PCA Principal Component Analysis	CA & DCA Correspondence Analysis & Detrended Correspondence Analysis	tb-PCA Transformation-based Principal Component Analysis	PCoA, NMDS Principal Correspondence Analysis, Non-metric Multidimensional Scaling
(2) Constrained (direct, canonical) **Shiny app**	RDA Redundancy Analysis	CCA Canonical Correspondence Analysis	tb-RDA Transformation-based Redundancy Analysis	db-RDA Distance-based Redundancy Analysis

Table 1: Summary of ordination methods. Based on Lepš & Šmilauer (2003), extended for other methods from Legendre & Legendre (2012).

Does the ordination algorithm include also environmental variables?

(1) Unconstrained ordination (indirect gradient analysis)

Ordination axes are not constrained by environmental factors. The method aims to uncover the main gradients (directions of changes) in species composition data, and it returns unconstrained ordination axes, which correspond to the directions of greatest variability within the dataset. Optionally, these gradients can be *post hoc* (after the analysis) interpreted by environmental variables (if these are available). Environmental variables do not enter the ordination algorithm. Unconstrained ordination is primarily an exploratory analytical method, used to explore the pattern in multivariate data; it generates hypotheses, but does not test them.

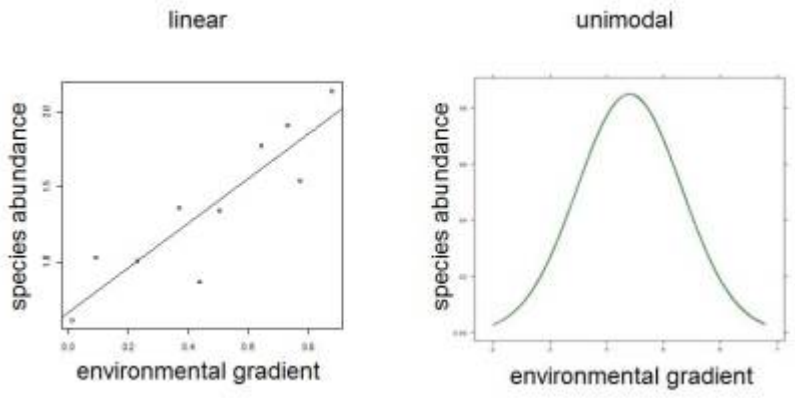
(2) Constrained ordination (direct gradient analysis, canonical ordination)

Ordination axes are constrained by environmental factors. It relates the species composition directly to the environmental variables and extracts the variance in species composition which is directly related to the environment. Environmental variables directly enter the algorithm, and the constrained ordination axes correspond to the directions of the variability in data which is explained by these environmental variables. The method is usually used as confirmatory analysis, i.e. it is able to test the hypotheses about the relationship between environmental factors on species composition (unlike unconstrained ordination, which is exploratory). It decomposes the total variance in species composition data into a fraction explained by environmental variables (related to constrained ordination axes) and not explained by environmental variables (related to unconstrained ordination axes). It offers several interesting opportunities when it comes to explanatory variables: forward selection (the selection of important environmental variables by excluding those which are not relevant for species composition), Monte Carlo permutation test (a test of significance of the variance explained by environmental factors) and variance partitioning (partitioning of the variance explained by different groups of environmental variables).

What type of species composition data is used for analysis?

(a) Raw-data-based methods (classical approach)

Figure 2: An assumption of linear (left) vs unimodal (right) response of species fitness/abundance along environmental distinguishes the use of linear vs unimodal ordination method.



Methods based on analysis of raw sample-species matrices with abundance or presence/absence data. Within these methods, two categories are traditionally recognized, differing by an assumption of species response along the environmental gradient:

- **linear** (Fig. 2, left panel) – assume that species response linearly along environmental gradient, which could be true for rather homogeneous ecological data, where ecological gradients are rather short (Fig. 4 left);
- **unimodal** (Fig. 2, right panel) – species response unimodally along the gradient, having its optima at a certain gradient position; this model is more close to the reality of ecological data and is more suitable for heterogeneous datasets (structured by a strong or long ecological gradient, with high species turnover and many zeroes in the species matrix) with rather long environmental gradients (Fig. 4 right).

(b) Transformation-based methods (tb-PCA and tb-RDA)

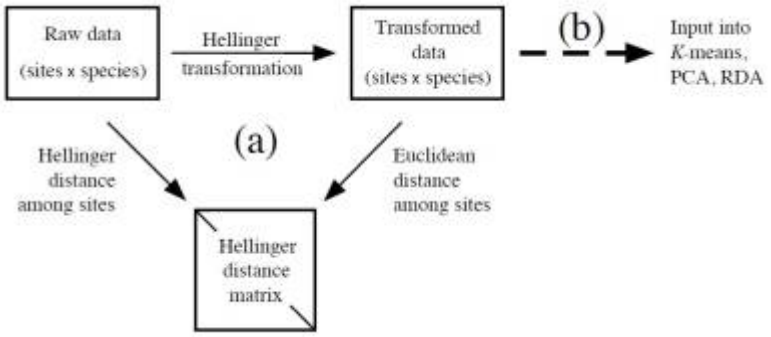


Figure 3: (a) Hellinger distance can be calculated either directly from raw species composition data, or by first transforming species composition data by Hellinger transformation (standardization) and then applying Euclidean distance on these transformed data. (b) Hellinger transformed data could be used in ordination methods using Euclidean distance (e.g. PCA, RDA, K-means clustering), in which case these methods use Hellinger distances (Hellinger transformed data + Euclidean distance = Hellinger distance). From Legendre & Legendre (2012), modified from Legendre & Gallagher (2001).

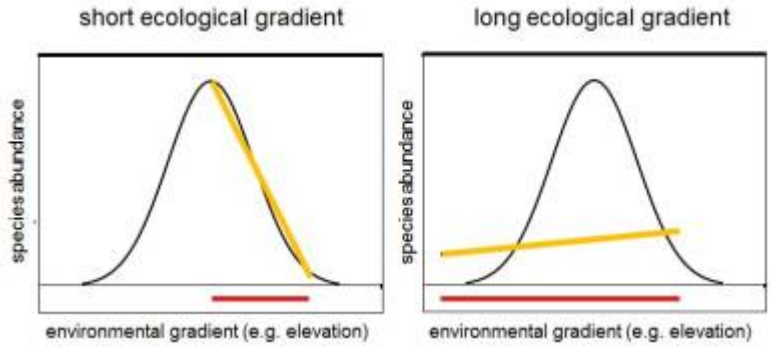
This category includes linear raw-data-based ordination methods (PCA, RDA), applied on sample×species data transformed by Hellinger (or one of several other) transformations. The Euclidean distance (implicit for PCA/RDA (Fig. 3) when applied on Hellinger-transformed species composition data results in Hellinger distance, which is more suitable for ecological data because (contrary to Euclidean distance) it is asymmetrical (ignores double zeros). Legendre & Gallagher (2001) consider this as a preferable way how to analyse heterogeneous data (otherwise not suitable for linear methods) using linear ordinations[1)]. Additionally to Hellinger transformation, the other suitable transformation is chord transformation, and other possible (but less suitable) transformations are species profile transformation, chi-square distance and chi-square metric transformations.

(c) Distance-based methods

Methods using the matrix of distances between samples measured by distance coefficients, and projecting these distances into two- or more-dimensional ordination diagrams. **Distance-based RDA (db-RDA)** is the combination of PCoA, applied on raw data using selected distance measure, and RDA applied on eigenvectors resulting from PCoA. It offers an alternative to RDA (based on Euclidean distances) and tb-RDA (based on Hellinger distances if transformed by Hellinger transformation), with a freedom to choose distance measure suitable for investigated data[2)].

Linear or unimodal ordination method?

Figure 4: Hypothetical species response along environmental gradient. Redline indicates the segment of the gradient actually sampled, and the yellow line indicates how would the species response looks like if fitted by a linear model. According to Lepš & Šmilauer (2003).



In the case that we sampled rather a short fraction of the environmental gradient (short red line segment at the left figure of Fig. 4), we may assume that species response (although fundamentally unimodal) can be modelled as linear (yellow line segment). In the case of the long gradient (figure at right), to model species response as linear would be wrong (right figure of Fig. 4).

To decide whether to apply **linear or unimodal ordination method** on your data, you can use the rule of thumb introduced by Lepš & Šmilauer (2003): first, calculate DCA (detrended by segments) on your data, and

check the length of the *first* DCA axis (which is scaled in units of standard deviation, S.D.). The length of first DCA axis > 4 S.D. indicates a heterogeneous dataset on which unimodal methods should be used, while the length < 3 S.D. indicates a homogeneous dataset for which linear methods are suitable (see Fig. 5. In the grey zone between 3 and 4 S.D., both linear and unimodal methods are OK. Note that while linear methods should not be used for heterogeneous data, unimodal methods can be used for homogeneous data, but linear methods, in this case, are more powerful and should be preferred. Alternatively, if your data are heterogeneous, but you still want to use linear ordination methods (PCA, RDA), apply them on Hellinger transformed species composition data to calculate ordination based on Hellinger distances (as recommended e.g. by Legendre & Gallagher (2001)).

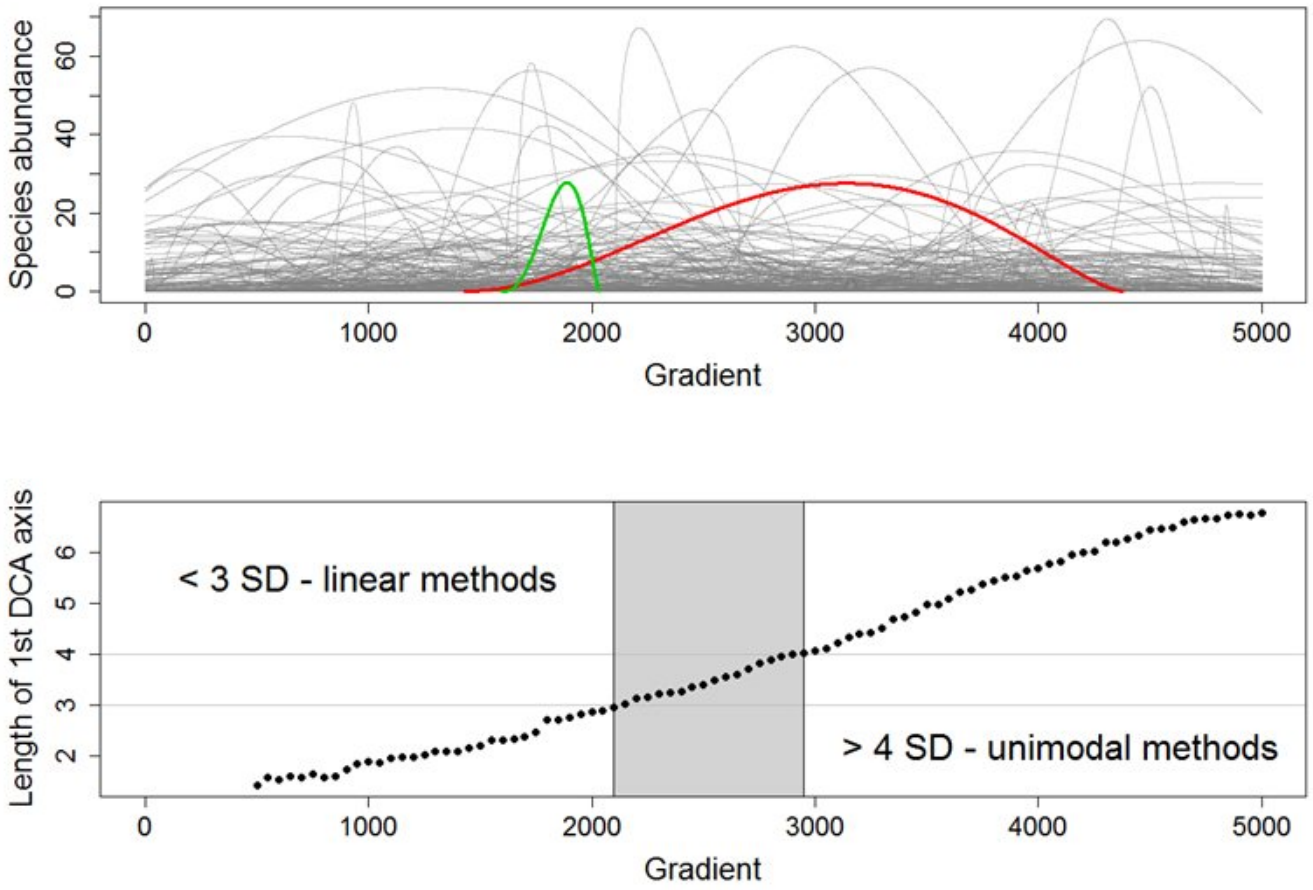


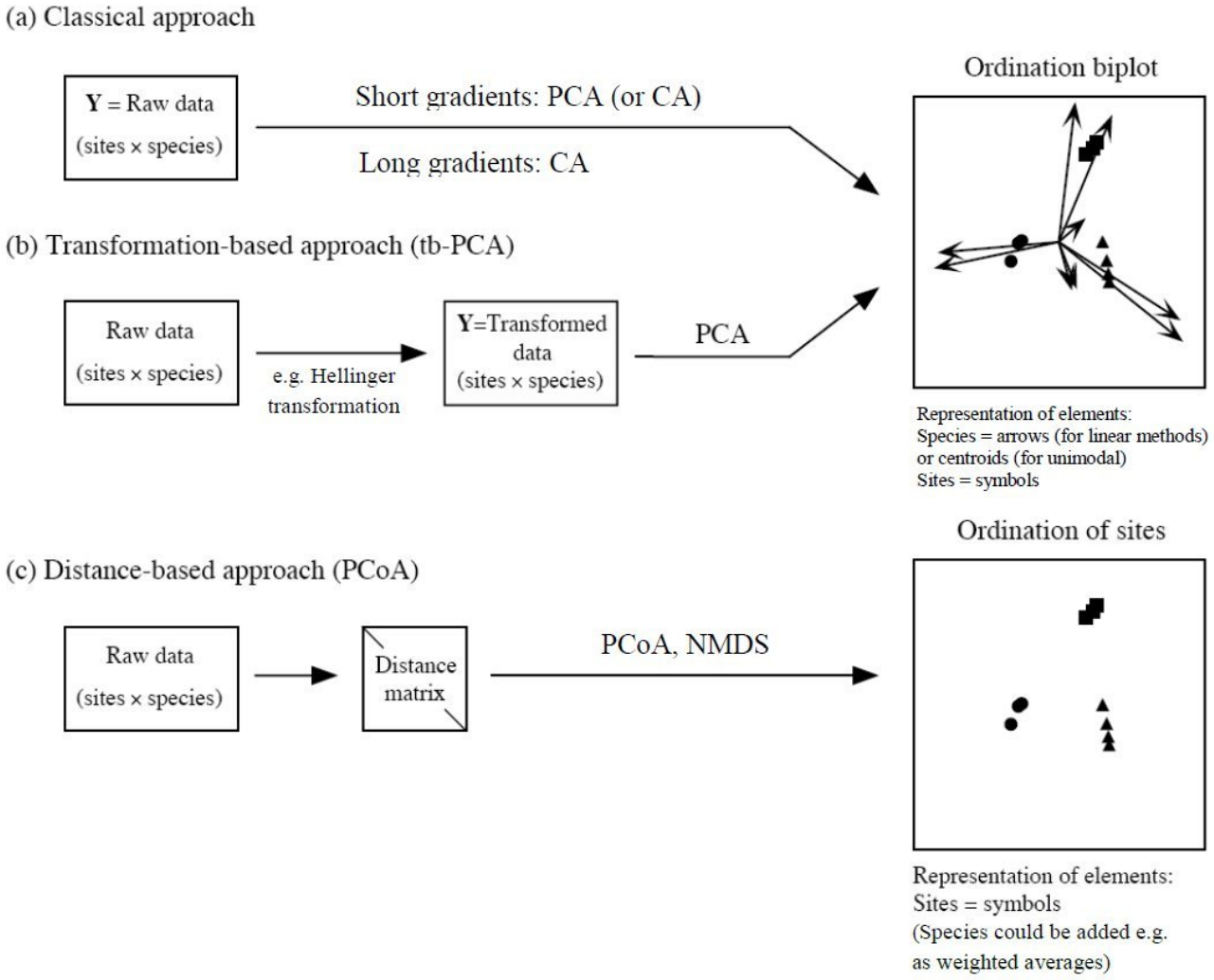
Figure 5: Illustration of the rule how to select whether to use linear ordination methods (like PCA or RDA) or unimodal (CA, DCA or CCA) on the data. The upper diagram shows a simulated community structured by a single environmental gradient, with a number of species response curves. The diagram below shows the relationship between the length of the gradient sampled in the simulated community on the x-axis (in arbitrary units), and the length of the first DCA ordination axis (in units of S.D.). The dataset which according to DCA is rather homogeneous (< 3 S.D.) has an environmental gradient up to 2000 units long; the longer gradient results into a heterogeneous dataset for which linear methods are not suitable.

Summary: Three alternative approaches for ordination

The schemas below show the three alternative approaches you can use for the ordination of community ecology data, for either unconstrained or constrained ordination (Fig. 6). It does not make much sense to combine individual approaches within the same type of ordination. For example, you can decide to analyze your data by either a) PCA/CA methods (depending on whether community composition data are homogeneous or heterogeneous), or b) transformation-based PCA (i.e. by pre-transforming your species composition data, e.g. by Hellinger standardization, and then using PCA; here doesn't matter whether community composition data are homogeneous or heterogeneous), or c) by distance-based approaches like PCoA or NMDS. But it often does not

make much sense to combine these approaches. For example, if you decide for approach a) above (PCA or CA), but in the case that you select PCA (since you concluded, e.g. by using preanalysis by DCA, that your data are reasonably homogeneous and you thus don't have to face double-zero problem), you also apply Hellinger standardization on your community composition data - in this case, you, in the end, opt for approach b) above, and you didn't need to check the heterogeneity of compositional data at all.

(1) Unconstrained ordination analysis



(2) Constrained ordination analysis

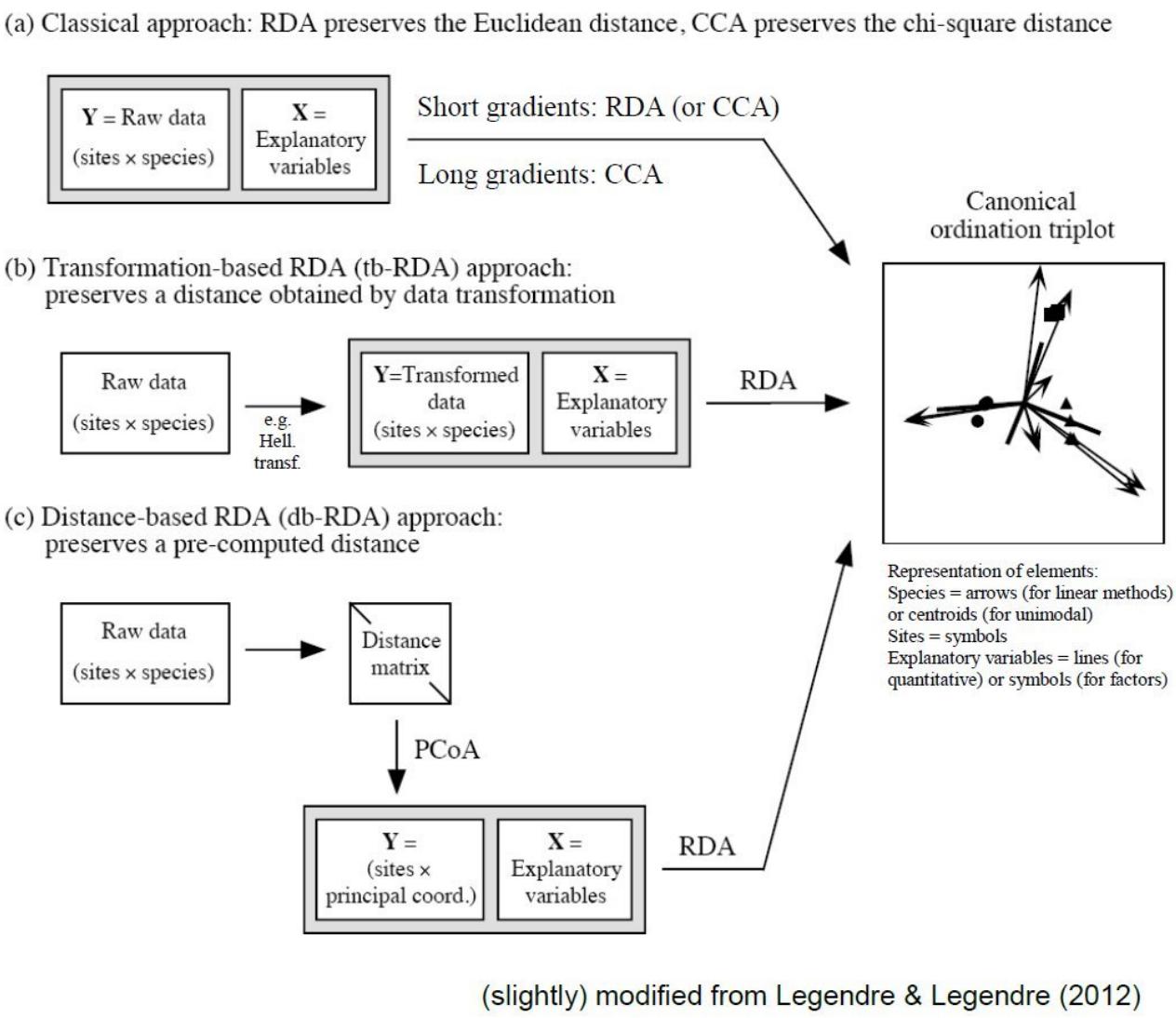


Figure 6: Three alternative approaches to unconstrained (above) and constrained (below) ordination analysis. From Legendre & Legendre (2012), slightly modified by D. Zelený.

Further reading

- Ordination methods - an overview (Mike Palmer)

1)
An alternative opinion, however, was presented by Minchin & Rennie at the ESA conference in 2010.
2)
Distance measure used in PCoA and db-RDA must be Euclidean, i.e. it must obey triangle inequality principle, otherwise it will produce negative eigenvalues, which in db-RDA may result in unrealistically high explained variation.

From:
https://anadat-r.davidzeleny.net/ - **Analysis of community ecology data in R**

Permanent link:
https://anadat-r.davidzeleny.net/doku.php/en:ordination

Last update: **2022/03/09 23:54**