FINAL REPORT

for the project „Broad-scale functional patterns of Pannonian grasslands" funded by the National Research, Development and Innovation Office (ID 123997).

Principal Investigator: dr. Attila Lengyel

Project participant: dr. Zoltán Botta-Dukát

Host Institute: Centre for Ecological Research

Project period: 1/Oct/2017 – 30/Sept/2021

The main aim of the project was to contribute to the research of between-community variation following the trait-based approach. The study object was the Pannonian grasslands, with a focus on grasslands in the Kiskunság region; however, methodological aspects were also emphatic. The project consisted of four subtopics.

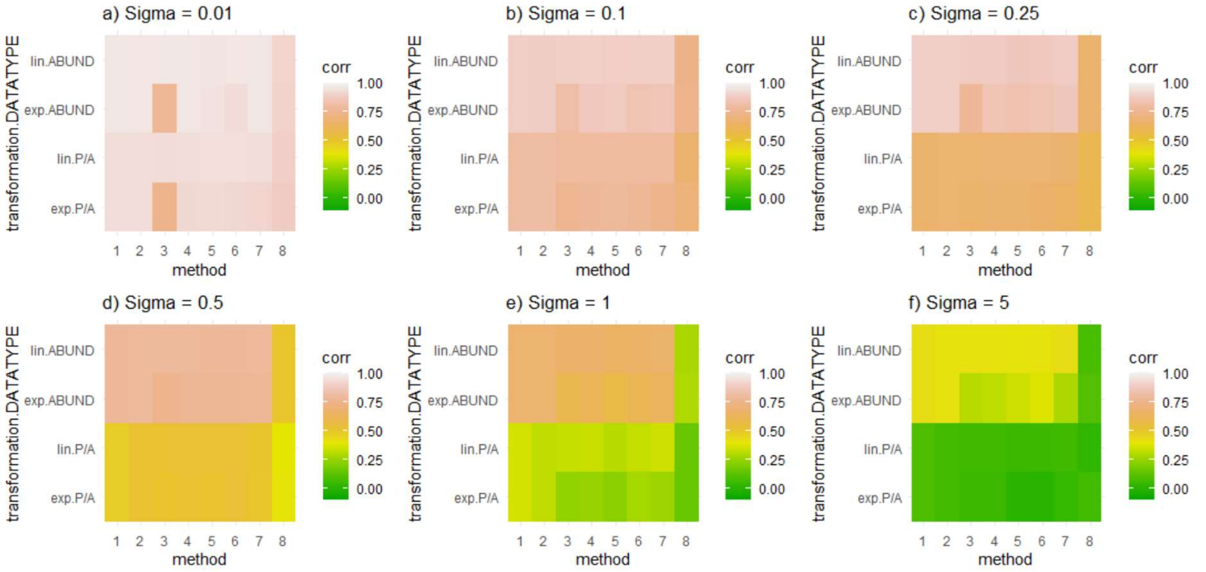
Subtopic 1 - Testing methods for plot-based multivariate analysis of functional community patterns

In the recent years, a great variety of indices has been proposed to quantify functional dissimilarity between pairs of communities. These indices were designed to follow the form of indices widely used for calculating species-based dissimilarity but they implement the trait information according to different concepts. Until now, no synthetic review and comparative test of performance was conducted on functional dissimilarity indices. We aimed at filling this gap using a simulation study.

First, we classified the indices into approaches based on algorithmic properties, and into families based on how trait information is taken into account. We recognized three approaches of functional dissimilarity indices. The *summary-based approach* includes indices that calculate some kind of summary statistics for each community and then calculate the distance function between them. Following this approach are the *typical value family*, where the mean or the median trait value represents a community, and the *distribution-based family*, where some trait distribution function (e.g. the probability density function) is applied instead of a single value. These methods require raw trait values of species for the communities under comparison. The *direct dissimilarity approach* can be applied if a species-by-species dissimilarity matrix is provided next to species abundances. One of the families following this approach is the *probabilistic family,* which can be deduced to quantify the expected similarity of randomly chosen individuals of the two communities. In the ordinariness-based family, the key concept is how ordinary (or, conversely, unique) the species in the two communities are. Similar species present in both communities mean high ordinariness and low uniqueness. The indices in the *diversity partitioning family* calculate gamma and alpha, and then multiplicative beta diversity for the two communities, and rescale its value to conform to algebraic requirements. Finally, indices in the *nearest neighbour family* rely on the most similar pair of species present in the pair of communities. The third approach is named the *classification-based approach*. Indices following this need an unambiguous assignment of species into functional groups, either in a hierarchical or in a non-hierarchical classification. Many of these indices are also used for calculating phylogenetic dissimilarity. The drawback of this approach with trait-based data is that the classification of species often requires further subjective decisions from the researcher regarding the choice of the classification method and number of functional groups. As these factors have no standards, classification-based indices cannot be evaluated without screening a range of possible decisions regarding how species are grouped. To keep our study focused, we decided not to provide a more in-depth review and performance test for them.

For the performance test, we used the simulation approach by Botta-Dukát & Czúcz (2016) to generate an artificial meta-community system. In this system, species have an environmental trait that responds to an environmental gradient. Trait information thus can be implemented as raw trait values or as the Gower distances between species. The environmental gradient determines species co-occurrence in local communities through a filtering process, while competitive exclusion has negligible role. We changed the data type (presence/absence vs. abundance), the niche width (i.e. strength of trait filtering), and the transformation function for between-species distances (linear vs. exponential). We tested how strongly trait-based dissimilarity correlates with environmental distance between communities. We built random forest models and variable importance scores to evaluate how data type, filtering, transformation function, as well as the choice of functional dissimilarity index determines correlation with the environmental distances.

Figure 1. Heat maps showing the interactive effects of niche width (sigma), transformation of between-species dissimilarities (lin = linear, exp = exponential), data type (ABUND = abundance, P/A = presence/absence), and dissimilarity index (1 – *CWMdis*, 2 – *CDFdis*, 3 – D_Q, 4 – *dsimcom*/Sørensen, 5 – *dissABC*/Sørensen, 6 – *generalized_Tradidiss*/generalized Canberra, uneven weighting, 7 – $\beta_{turnover}$, 8 – D_{CW}) on the correlation with environmental distance



The results showed that the most important variable in the random forest model is the niche width. In case of narrow niches (i.e. strong filtering), all dissimilarity indices correlated highly with environmental distance. With gradually weaker filtering effect, all indices decreased correlation; although, some of them more steeply than others. However, there were several indices which performed similarly well, e.g. the CWMdis index, dsimcom and generalized_Tradidiss obtained the highest correlations. Even the data type was more important than the choice of functional dissimilarity index – abundance data resulting in higher correlation.

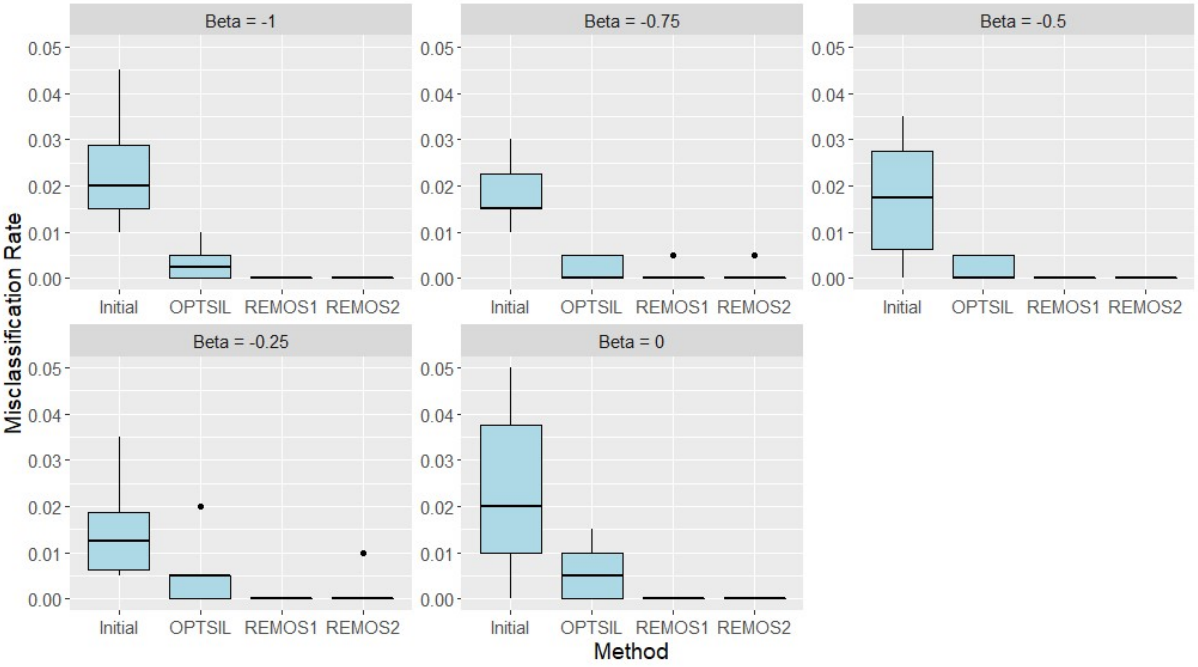
We concluded that the difference between the behaviour of many available indices of functional dissimilarity is surprisingly small. The strength of filtering process (i.e. the length of the gradient) and the data type influence the correlation with environmental distance more strongly than the choice of the dissimilarity index. These results are uploaded to the bioRxiv preprint server (see Lengyel et al. 2021a) and is ahead of resubmission to Ecography.

Subtopic 2 – Functional classification of grasslands in the Kiskunság Sand Ridge

As a part of an earlier project, we showed an example of trait-based vegetation classification on phytosociological relevés of wet and mesic grasslands of Poland (Lengyel et al. 2020). In this paper, we argued that species with highly varying cover percentage and unique combination of trait values have especially high influence on the classification. Such species often occur as constant and dominant elements of clusters. On the other hand, we concluded that functional vegetation classifications heavily depend on which traits are included in the analysis; therefore, no universal functional classification can be expected. We compiled a data set of Kiskunság grasslands and ran pilot classifications; they supported the above statements. However, we did not see it likely to publish a Q1 paper due to the lack of novel and general findings. Hence, we shifted our scope to a methodological challenge instead, and developed a new non-hierarchical classification algorithm called REMOS (RE-allocation of Misclassified Obsects based on Silhouette width).

REMOS starts with an initial, existing classification of objects into discrete groups. First, it calculates the silhouette width index that takes a negative value for objects which seem more similar to objects in a different group than those in the same group. Such objects can be deemed misclassified. In the next step, the misclassified object(s) are moved to the cluster whose objects are the most similar. Then, the silhouette index is updated, and if further misclassified objects exist, a new re-allocation step is done. The algorithm is, therefore, iterative repeats the calculation of silhouette widths and re-allocation of misclassified objects. We introduced two versions of REMOS. REMOS1 re-allocates only a single object at once, which is the object with the most negative silhouette width. In contrast, REMOS2 re-allocates all misclassified objects in each round. Using artificial point patterns and real data sets, we compared REMOS1 &2 with a similar silhouette-based optimization method, OPTSIL. The initial classification was a beta-flexible solution with beta values changed between -1 and 0. With most data sets, OPTSIL achieved higher mean silhouette width; however, REMOS1 & 2 outperformed OPTSIL in the number of correctly classified objects (that is, they provided less misclassified objects). There was no clear difference between the two methods in terms of diagnostic species. However, REMOS ran magnitudes faster than OPTSIL.

Figure 2. Differences in misclassification rate between the initial classification (without optimization), REMOS1, REMOS2, and OPTSIL across different beta values of the flexible-beta classification.



We concluded that if the aim of the classification is to produce as few misclassifications as possible, and/or runtime is a limiting factors, then REMOS is a straightforward choice.

REMOS is published in Journal of Vegetation Science (Lengyel et al. 2021b).

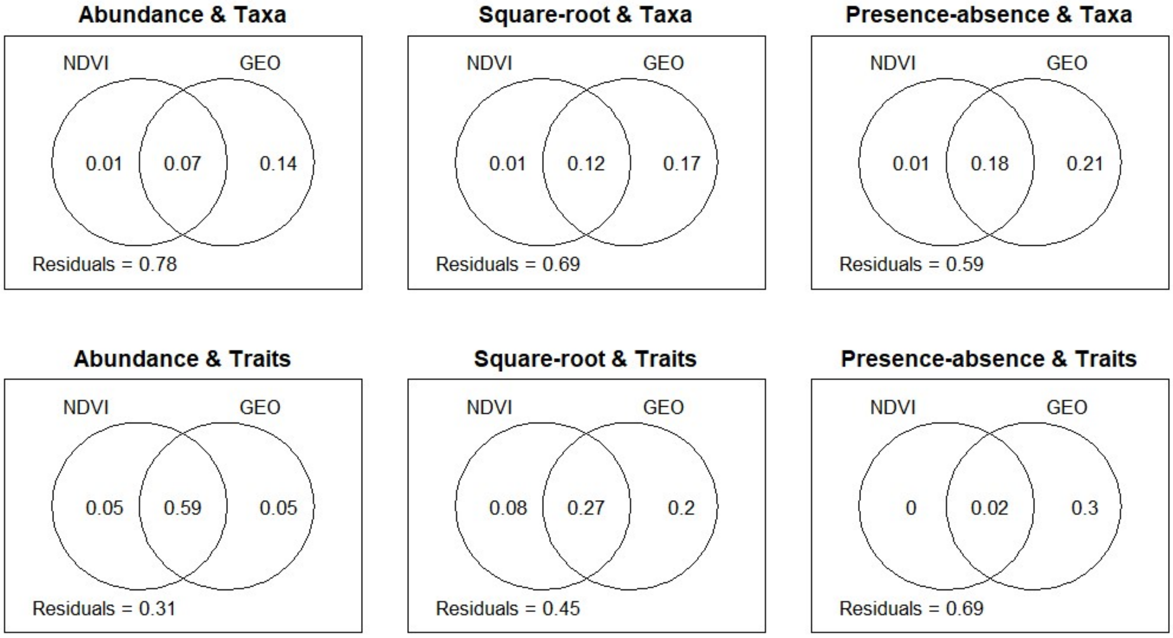
Subtopic 3 – Functional redundancy and vulnerability in Kiskunság sand grasslands

We conducted two studies on different aspects of diversity of Kiskunság grasslands. For both studies we used the data set of the Coarse-scale Vegetation Ecology Group of the host institute collected on three sites in the same region. In each site, vegetation was sampled using phytosociological relevés in a design to represent a long productivity (wetness) gradient. The total number of plots is 364. For each plot, normalized vegetation difference index (NDVI) was measured in the field as a proxy for productivity. Trait data was gathered for the species from public databases, literature and own measurements, of which here we used canopy height, specific leaf area and seed mass.

In the first study, we examined how between-community dissimilarity, calculated with and without considering traits, is explained by NDVI measurements and spatial location. Additionally, we tested how the data type (raw abundances, squared-rooted abundances, presence/absence data) influences this relationship.

We found that taking account of trait improved the amount of between-community variation explained by NDVI only if abundance data was also considered. With presence/absence data, taxon-based models explained higher amount of total variation than trait-based models in two out of three sites. Spatial location showed stronger relationship with presence/absence data than with abundances in cases of both taxa and traits.

Figure 3. Variation fractions of taxonomical and functional dissimilarity matrices explained by NDVI (as a proxy for productivity) and GEO (spatial distance) using different treatments of abundance data from the Ásotthalom site.

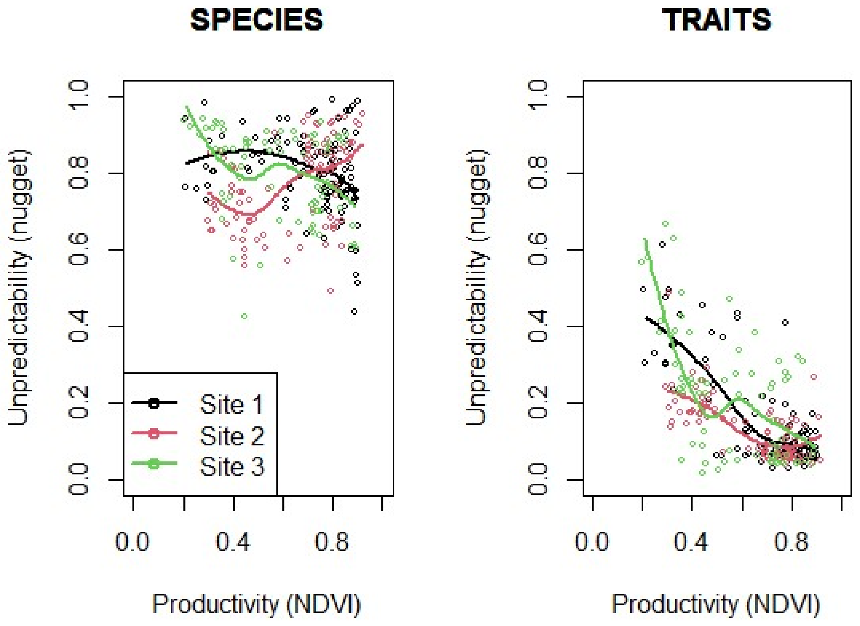


With these results we argue that the involvement of trait information improves the environmental matching of community data only if abundances are also considered. This finding add an important note to the fundamental of trait-based ecology stating that the application of traits instead of taxa improves our understanding of ecosystem properties.

A manuscript of this study was uploaded to the bioRxiv preprint server (Lengyel et al. 2021c) and is currently submitted to journal Oikos.

In the second study, we investigated the role of chance in the spatial pattern in relation with the productivity gradient. Improving on the methodology by Brownstein et al. (2012), we introduced a method with which it is possible to give local estimates of unpredictability on single vegetation plots with respect to species or trait composition. We made such estimates for each plot in each site, using a species-based and a trait-based dissimilarity index (that is, reflecting species and trait composition). We assessed how unpredictability changes along the productivity gradient.

Figure 4. Estimates of plot-level unpredictability along the productivity gradient based on species (left) and traits (right).



Taxonomical unpredictability was relatively high and showed no clear trend along productivity (wetness). In contrast, trait-based unpredictability was lower, and decreased towards more productive plots. In other words, species composition is constantly hard to predict across the productivity gradient but trait composition is becoming easier to predict towards wetter grasslands. There are several, non-exclusive explanations for this pattern. We think that the most parsimonious explanations are that (1) the strength of the filtering processes change along the gradient, or (2) the relative importance of the selected traits in the community assembly mechanisms change (increase).

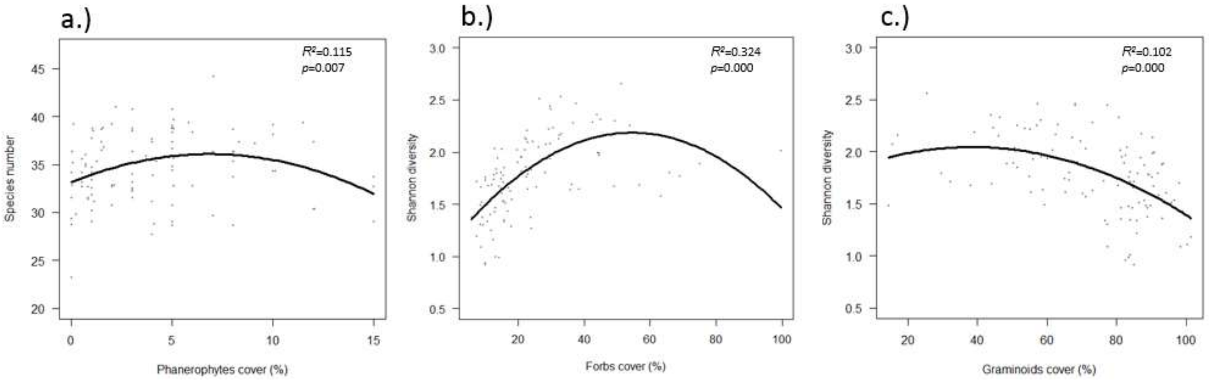
These results were only presented on conferences yet.

Subtopic 4 – Effects of management attributes on functional diversity, redundancy and vulnerability

The data set we previously planned to use seemed unsatisfactory; therefore, I had to join another group of researchers who was conducting research on a similar topic and carried out field sampling campaigns in 2017 and 2018. For this reason, the focus and methodology of the research shifted slightly and I was not able to contribute as the lead author.

Our aim was to evaluate the effects on vegetation diversity and composition of those management properties, e.g. intensity of herbage removal and temporal complexity, which are usually neglected in most studies. We sampled 117 plots on meadow steppe grasslands in the Kiskunság. These grasslands were managed by mowing, grazing by cattle or both combined. We differentiated grasslands according to herbage removal intensity: “low” meant <0.5 AU/ha or mowing once a year, “high” meant >0.5 AU/ha or mowing once a year with aftermath grazing. We also distinguished grasslands according to the temporal complexity of management: “high” for different management regime each year, “low” for repetitions in subsequent years. We modelled species-based diversity measures, contribution of functional groups and physiognomical variables as function of the type, intensity and temporal complexity of management using generalized mixed effect models. We also tested pairwise correlations between the summed covers of functional group and diversity indices.

Figure 5 Effect of plant functional type covers on diversity meaures. Fits of models are expressed by beta R^2-values derived from GLMER models.



We found that temporal complexity significantly increased diversity indices, and this effect was stronger than those of the other management properties. We showed that the higher species richness on grazed grasslands in comparison with mown ones is attributable to the nurse effect of small shrubs that provide microrefugia for species not tolerating biomass removal. We found that the summed cover of forbs positively, while graminoids negatively correlated with species richness.

Based on these results we hope that the maintenance of species-rich steppe grasslands can become more successful. We also discuss the possibilities and the challenges of including more detailed information on management into modelling and conservation planning.

The first manuscript was rejected in the Journal for Nature Conservation. Now we are ahead of the second submission to another journal.

References

Botta-Dukát, Z. & Czúcz, B. (2016). Testing the ability of functional diversity indices to detect trait convergence and divergence using individual-based simulation. Methods in Ecology and Evolution, 7, 114-126. https://doi.org/10.1111/2041-210X.12450

Brownstein, G., Steel, J.B., Porter, S., Gray, A., Wilson, C., Wilson, P.G. and Bastow Wilson, J. (2012). Chance in plant communities: a new approach to its measurement using the nugget from spatial autocorrelation. Journal of Ecology 100: 987-996. https://doi.org/10.1111/j.1365-2745.2012.01973.x

Lengyel, A, Roberts, DW, Botta-Dukát, Z. (2021b). Comparison of silhouette-based reallocation methods for vegetation classification. Journal of Vegetation Science 32: e12984. https://doi.org/10.1111/jvs.12984

Lengyel, A, Swacha, G, Botta-Dukát, Z, Kącki, Z. (2020). Trait-based numerical classification of mesic and wet grasslands in Poland. Journal of Vegetation Science 31: 319–330. https://doi.org/10.1111/jvs.12850

Lengyel, A. & Botta-Dukát, Z. (2021a). Review and performance evaluation of trait-based between-community dissimilarity measures. bioRxiv 2021.01.06.425560; doi: https://doi.org/10.1101/2021.01.06.425560

Lengyel, A., Barabás, S., Berki, B., Csecserits, A., Gyalus, A., Lhotsky, B., Ónodi, G., Rédei, T., Botta-Dukát, Z. (2021c). Using trait data improves correlation between environment and community data only if abundances are considered. bioRxiv 2021.09.27.461896; doi: https://doi.org/10.1101/2021.09.27.461896