Comparing incomparable:

Variation in species composition explained by environment is not comparable among datasets with different levels of undersampling

David Zelený, Ching-Feng Li, Dominik Zukal,
Markéta Chudomelová, Helena Prokešová & Hana Sekerková

Masaryk University, Brno, Czech Republic

VEGETATION SCIENCE GROUP
MASARYK UNIVERSITY BRNO

Undersampling

Undersampling

Undersampling

community richness (synonym to *species pool*)

species richness (synonym to *alpha diversity*)

Disentangling the Drivers of β Diversity Along Latitudinal and Elevational Gradients

Nathan J. B. Kraft,[1,2]* Liza S. Comita,[3,4] Jonathan M. Chase,[5] Nathan J. Sanders,[6,7] Nathan G. Swenson,[8] Thomas O. Crist,[9] James C. Stegen,[10,11] Mark Vellend,[1,12] Brad Boyle,[13] Marti J. Anderson,[14] Howard V. Cornell,[15] Kendi F. Davies,[16] Amy L. Freestone,[17] Brian D. Inouye,[18] Susan P. Harrison,[15] Jonathan A. Myers[5]

Understanding spatial variation in biodiversity along environmental gradients is a central theme in ecology. Differences in species compositional turnover among sites (β diversity) occurring along gradients are often used to infer variation in the processes structuring communities. Here, we show that sampling alone predicts changes in β diversity caused simply by changes in the sizes of species pools. For example, forest inventories sampled along latitudinal and elevational gradients show the well-documented pattern that β diversity is higher in the tropics and at low elevations. However, after correcting for variation in pooled species richness (γ diversity), these differences in β diversity disappear. Therefore, there is no need to invoke differences in the mechanisms of community assembly in temperate versus tropical systems to explain these global-scale patterns of β diversity.

Some of the most striking and frequently documented patterns in ecology are that species richness in local communities generally declines with increasing latitude and elevation, such that the diversity of many clades peaks in lowland, tropical areas (*1*, *2*). The mechanisms underlying these gradients are often difficult to distinguish because multiple processes operating at multiple scales may govern geographic variation in diversity (*3*). For example, declines in diversity with elevation and latitude could result from deterministic community

emag.org **SCIENCE** VOL 333 23 SEPTEMBER 2011

Undersampling

Disentangling the Drivers of β Diversity Along Latitudinal and Elevational G

Nathan J. B. Kraft,[1,2]* Liza S. C
Nathan G. Swenson,[8] Thomas O
Marti J. Anderson,[14] Howard V.
Brian D. Inouye,[18] Susan P. Har

Understanding spatial variation in ecology. Differences in species co gradients are often used to infer v that sampling alone predicts chang pools. For example, forest invento well-documented pattern that β d after correcting for variation in po disappear. Therefore, there is no assembly in temperate versus trop

Some of the most striking a documented patterns in ec species richness in local com erally declines with increasing lat vation, such that the diversity of peaks in lowland, tropical areas (*1*,

Comment on "Disentangling the Drivers of β Diversity Along Latitudinal and Elevational Gradients"

Hanna Tuomisto* and Kalle Ruokolainen

Kraft *et al.* (Report, 23 September 2011, p. 1755) argued that the latitudinal trend in β diversity is spurious and just reflects a trend in γ diversity. Their results depend on the idiosyncrasies of their data, especially the latitudinally varying degree of undersampling and a local sampling setup that is not suitable for analyzing drivers of β diversity.

Kraft *et al.* (*1*) started with the premise that trends in β diversity are indicative of

dients a
data set
fine-gra
then cau
diversity
sity...an
iation in
whether
data set

First, the value of β diversity in a data set can be indicative of species responses to environ-

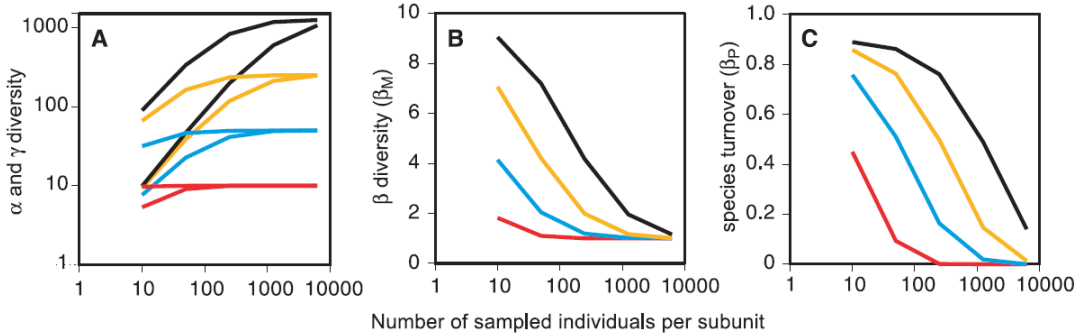
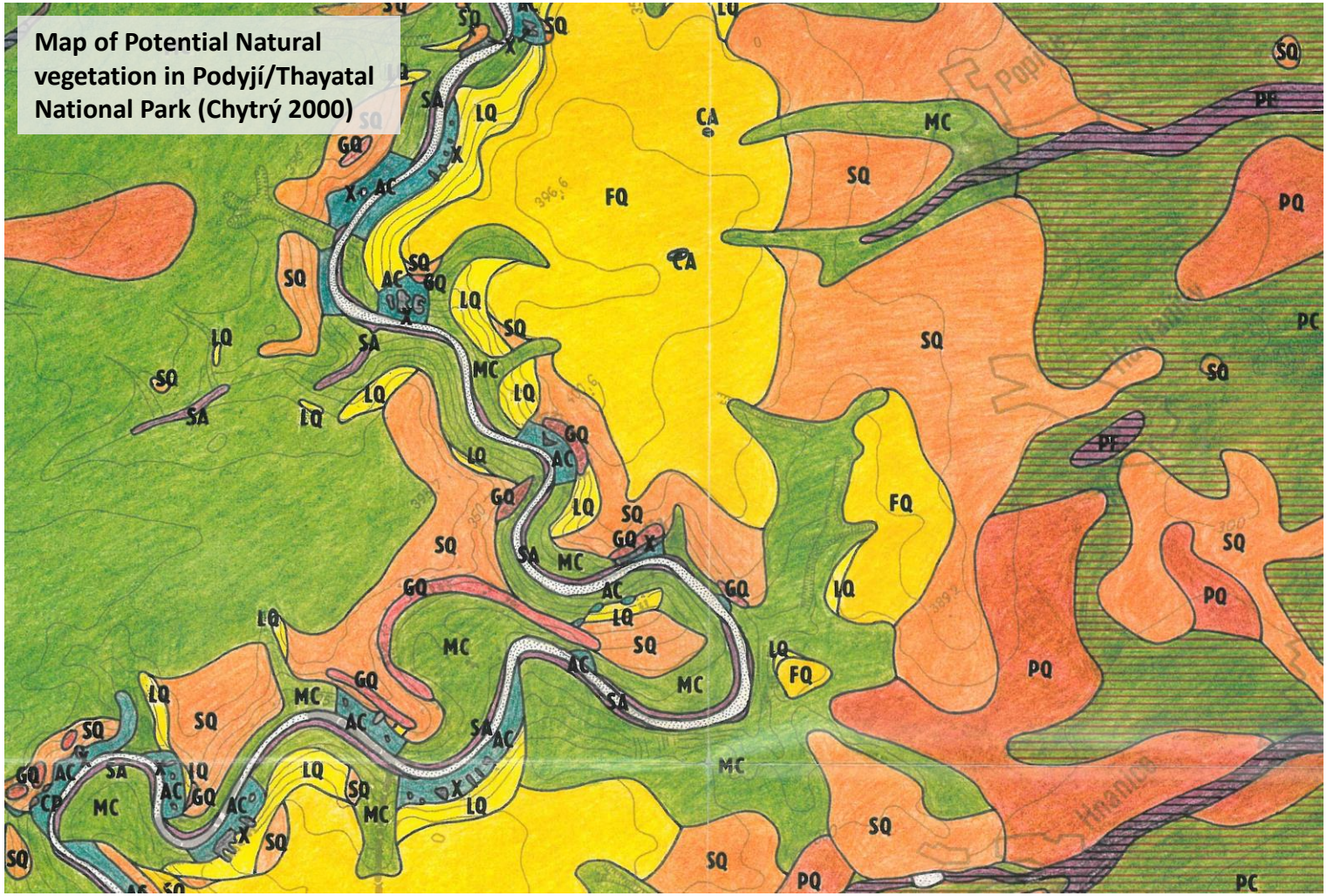


Fig. 1. The dependency of α and γ diversity (**A**), β diversity (**B**) and proportional species turnover (**C**) on species pool size and within-subunit number of individuals in 10-subunit data sets. Each color corresponds to a different species pool size: red, 10 species; blue, 50 species; orange, 250 species; black, 1250 species. Each line shows the mean of 1000 replicates where 10 subunits of either 10, 50, 250, 1250, or 6250 individuals were randomly drawn from a log-normal species-abundance distribution. In (A), the lower line for a given species pool size shows α diversity and the upper line, γ diversity. With a sufficiently large number of individuals, α and γ converge; all subunits were drawn from the same species pool, so any compositional differences among them are caused by undersampling. In (B), the factor by which γ exceeds α ($\beta_M = \gamma/\alpha$) decreases with the increasing number of individuals sampled. In (C), the proportion of species in the entire set of 10 subunits that does not fit within a single subunit ($\beta_P = 1 - \alpha/\gamma$) decreases with increasing number of individuals sampled.

Comparison of terminology

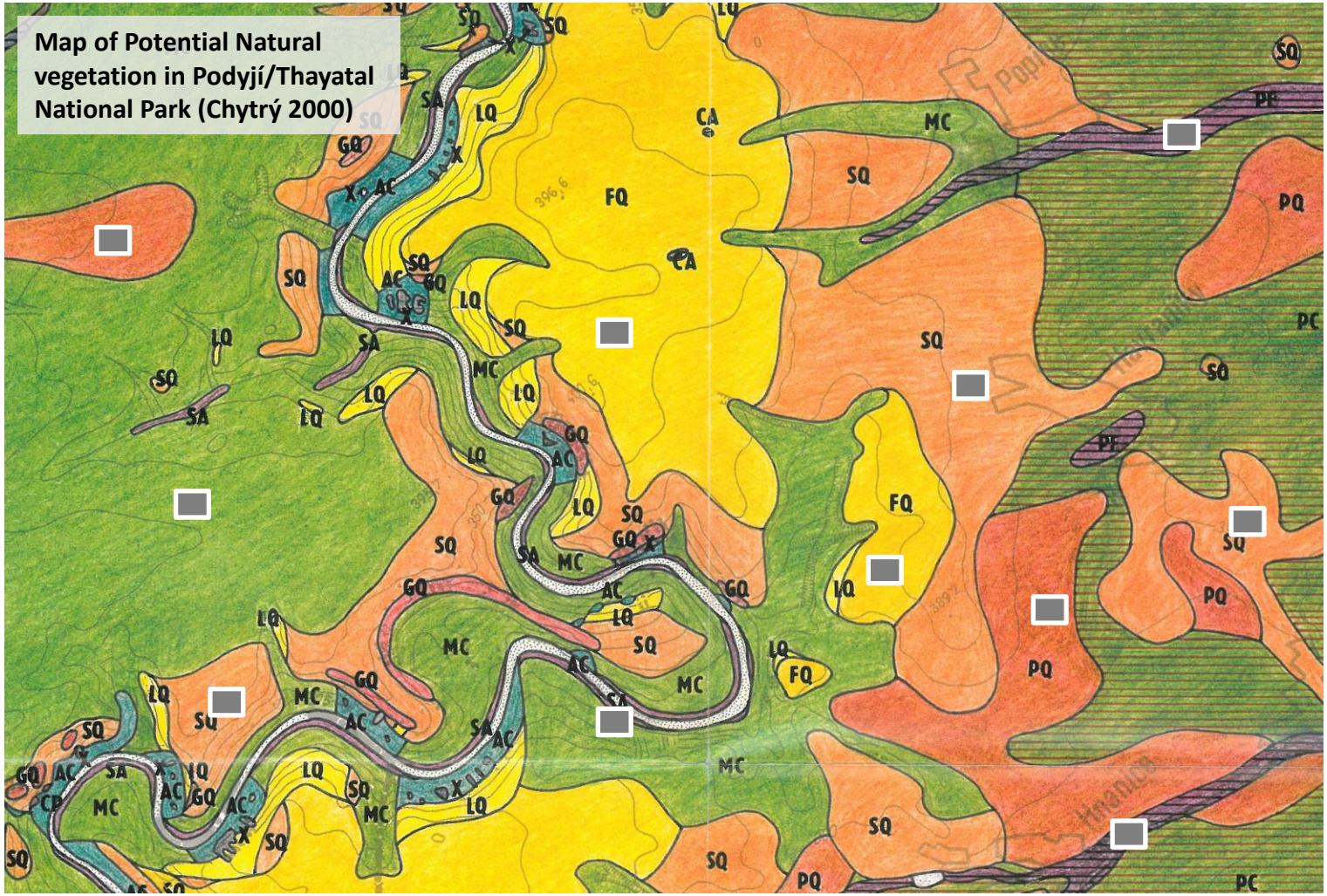
Richness/unseen species concept (Gotelli, Colwell, Chao etc.)	**Species pool/dark diversity concept** (Zobel, Pärtel etc.)
individual based	species based
community richness	species pool
species richness (of sample)	alpha diversity
unseen species	dark diversity
some species are more abundant than others -> species-abundance distribution	?

Communities vs samples (e.g. vegetation plots)

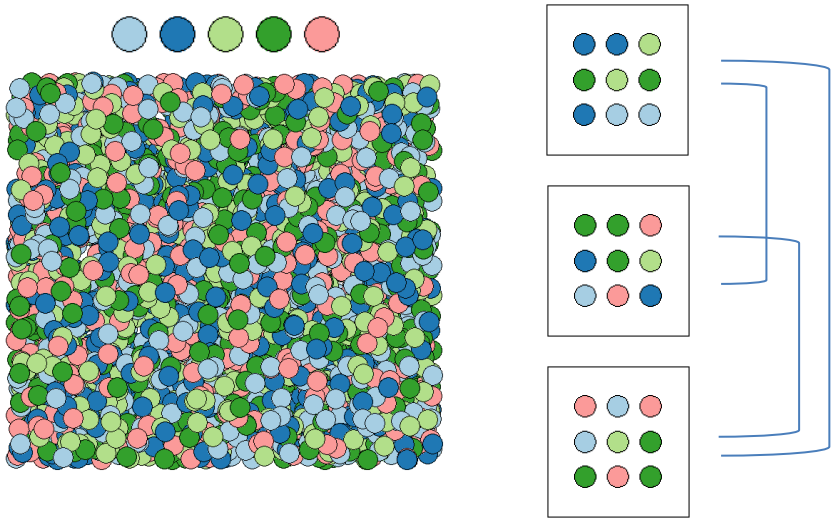


Communities vs samples (e.g. vegetation plots)

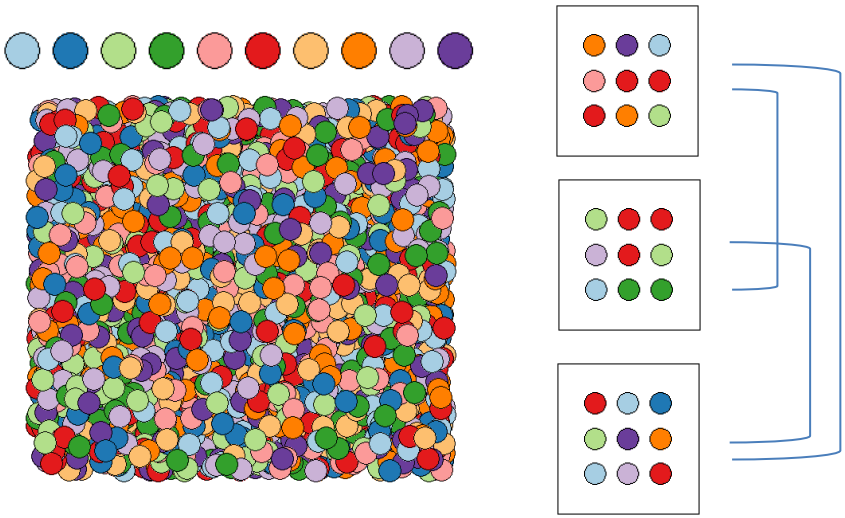
Communities vs samples (e.g. vegetation plots)



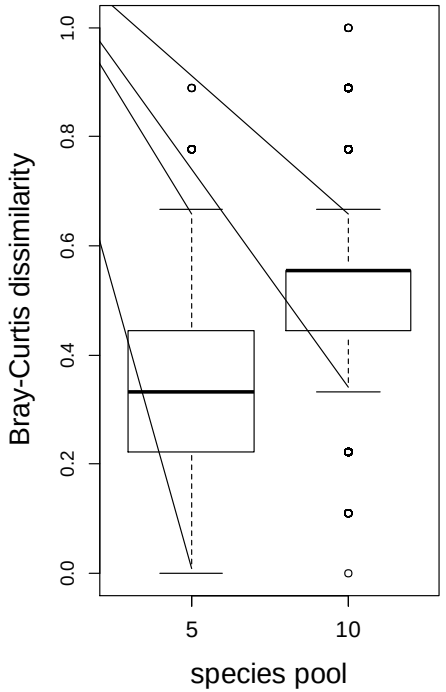
community richness = 5 species



community richness = 10 species



Effect of undersampling on beta diversity



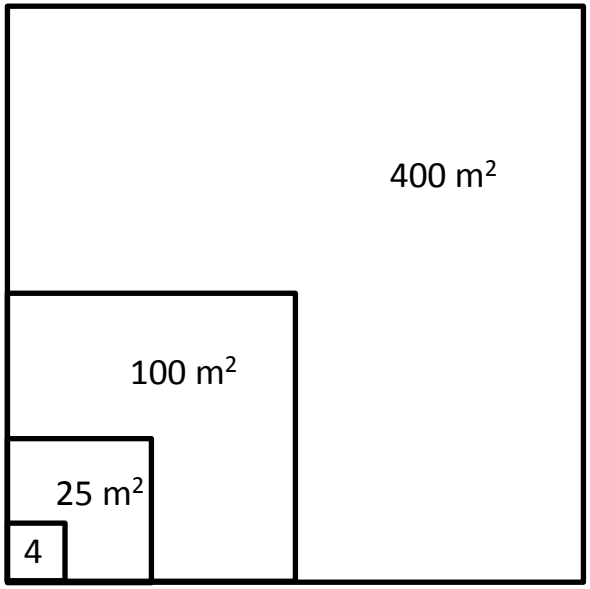
Studied vegetation

I II III IV V VI

I **Mesic beech forest** (*Galio odorati-Fagetum sylvaticae* Sougnez et Thill 1959)
II **Mountain spruce forest** (*Calamagrostio villosae-Piceetum abietis* Schlüter 1966)
III **Lowland floodplain forest** (*Fraxino pannonicae-Ulmetum glabrae* Aszód 1935 corr. Soó 1963)
IV **Acidophilous thermophilous oak forest** (*Sorbo torminalis-Quercetum* Svoboda ex Blažková 1962)
V **Basiphilous thermophilous oak forest** (*Euphorbio-Quercetum* Knapp ex Hübl 1959)
VI **Oak-hornbeam forest** (*Primulo veris-Carpinetum betuli* Neuhäusl et Neuhäuslová in Neuhäuslová-Novotná 1964)

Case study data:

- 6 localities
- 12 plots per locality
- 4 nested subplots per plot
- detailed measurements of environment



Studied vegetation

I II III IV V VI

I **Mesic beech forest** (*Galio odorati-Fagetum sylvaticae* Sougnez et Thill 1959)

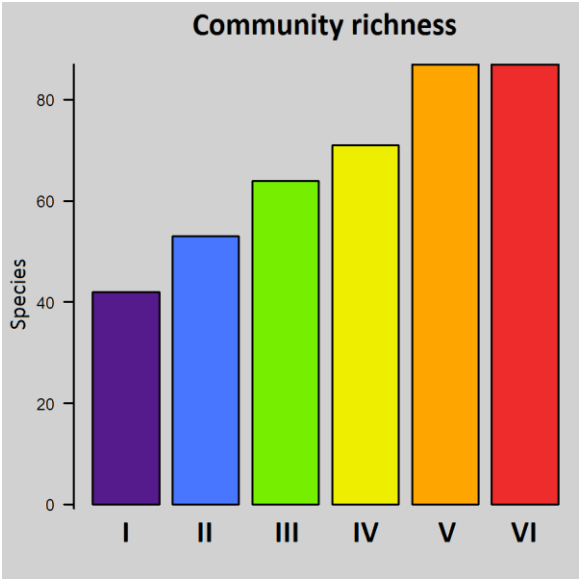
II **Mountain spruce forest** (*Calamagrostio villosae-Piceetum abietis* Schlüter 1966)

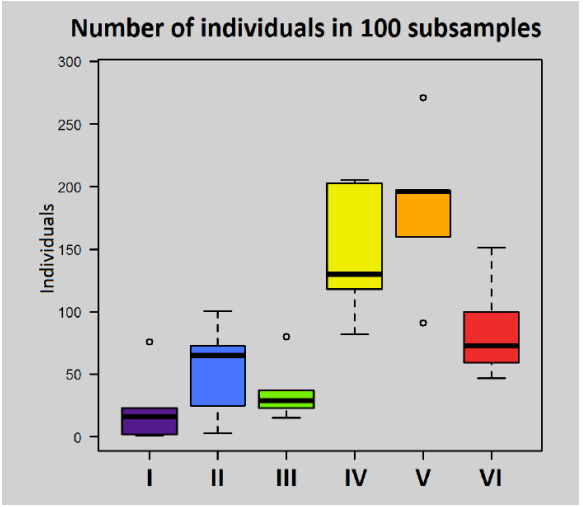
III **Lowland floodplain forest** (*Fraxino pannonicae-Ulmetum glabrae* Aszód 1935 corr. Soó 1963)

IV **Acidophilous thermophilous oak forest** (*Sorbo torminalis-Quercetum* Svoboda ex Blažková 1962)

V **Basiphilous thermophilous oak forest** (*Euphorbio-Quercetum* Knapp ex Hübl 1959)

VI **Oak-hornbeam forest** (*Primulo veris-Carpinetum betuli* Neuhäusl et Neuhäuslová in Neuhäuslová-Novotná 1964)





Studied vegetation

I II III IV V VI

I **Mesic beech forest** (*Galio odorati-Fagetum sylvaticae* Sougnez et Thill 1959)

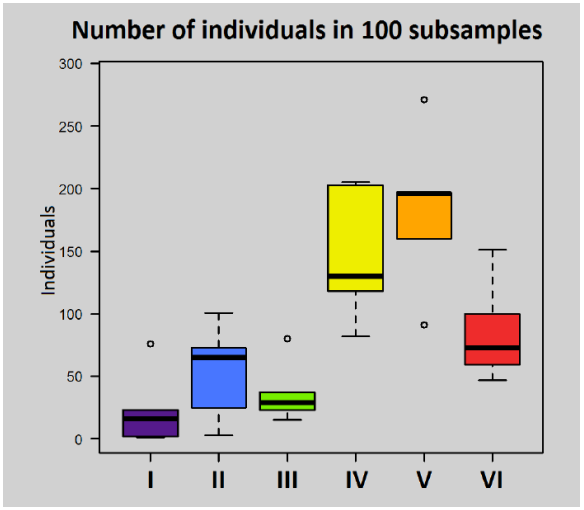
II **Mountain spruce forest** (*Calamagrostio villosae-Piceetum abietis* Schlüter 1966)

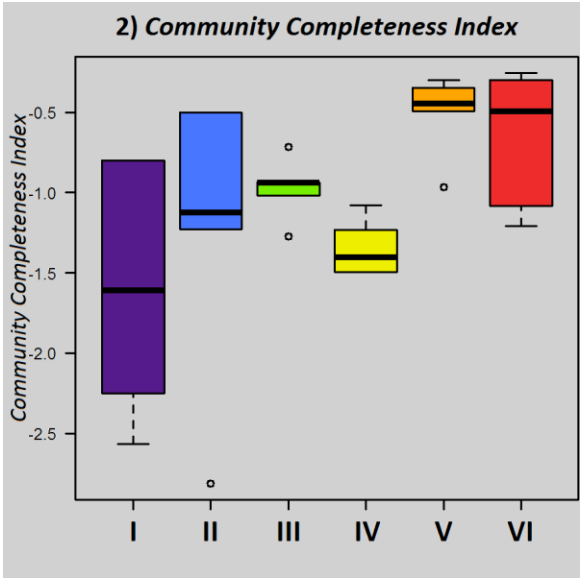
III **Lowland floodplain forest** (*Fraxino pannonicae-Ulmetum glabrae* Aszód 1935 corr. Soó 1963)

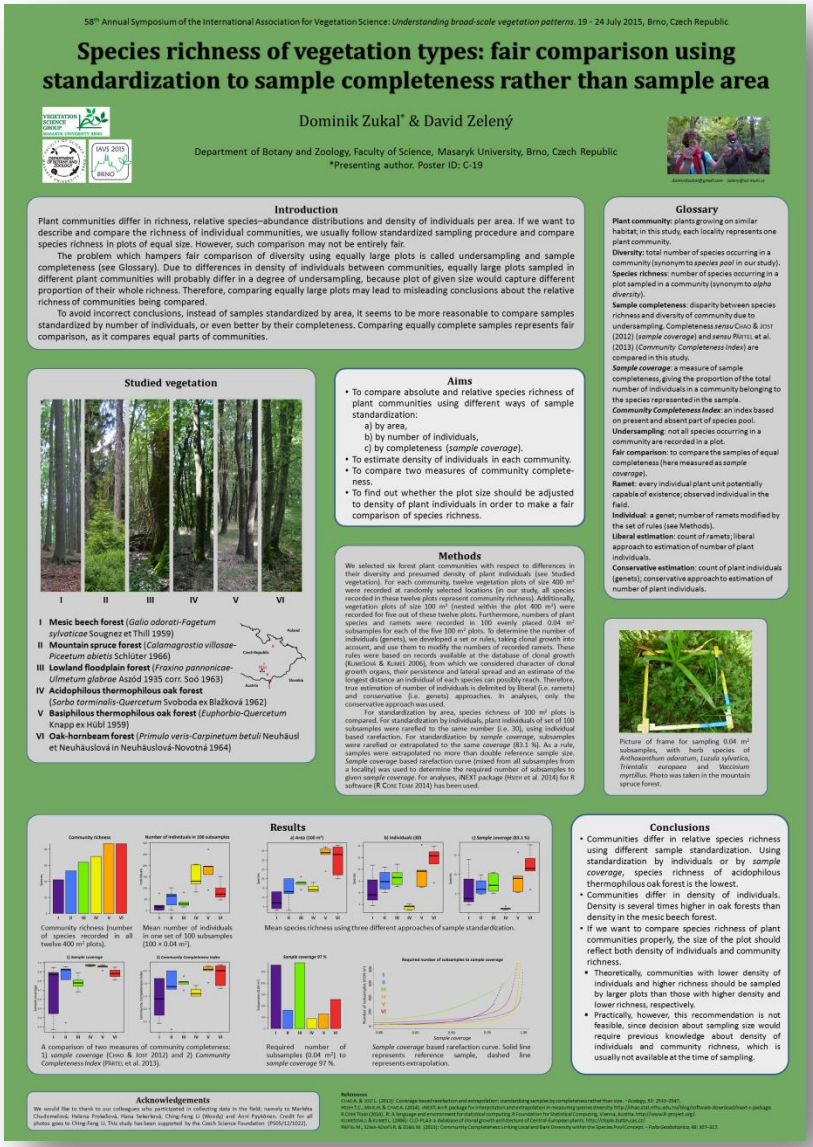
IV **Acidophilous thermophilous oak forest** (*Sorbo torminalis-Quercetum* Svoboda ex Blažková 1962)

V **Basiphilous thermophilous oak forest** (*Euphorbio-Quercetum* Knapp ex Hübl 1959)

VI **Oak-hornbeam forest** (*Primulo veris-Carpinetum betuli* Neuhäusl et Neuhäuslová in Neuhäuslová-Novotná 1964)







58th Annual Symposium of the International Association for Vegetation Science: *Understanding broad-scale vegetation patterns*. 19 - 24 July 2015, Brno, Czech Republic

Species richness of vegetation types: fair comparison using standardization to sample completeness rather than sample area

Dominik Zukal* & David Zelený

Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech Republic
*Presenting author. Poster ID: C-19

Introduction

Plant communities differ in richness, relative species–abundance distributions and density of individuals per area. If we want to describe and compare the richness of individual communities, we usually follow standardized sampling procedure and compare species richness in plots of equal size. However, such comparison may not be entirely fair.

The problem which hampers fair comparison of diversity using equally large plots is called undersampling and sample completeness (see Glossary). Due to differences in density of individuals between communities, equally large plots sampled in different plant communities will probably differ in a degree of undersampling, because plot of given size would capture different proportion of their whole richness. Therefore, comparing equally large plots may lead to misleading conclusions about the relative richness of communities being compared.

To avoid incorrect conclusions, instead of samples standardized by area, it seems to be more reasonable to compare samples standardized by number of individuals, or even better by their completeness. Comparing equally complete samples represents fair comparison, as it compares equal parts of communities.

Glossary

Plant community: plants growing on similar habitat; in this study, each locality represents one plant community.
Diversity: total number of species occurring in a community (synonym to species *pool* in our study).
Species richness: number of species occurring in a plot sampled in a community (synonym to *alpha diversity*).
Sample completeness: disparity between species richness and diversity of community due to undersampling. Completeness *sensu* Chao & Jost (2012) (*sample coverage*) and *sensu* Pärtel et al. (2013) (*Community Completeness Index*) are compared in this study.
Sample coverage: a measure of sample completeness, giving the proportion of the total number of individuals in a community belonging to the species represented in the sample.
Community Completeness Index: an index based on present and absent part of species pool.
Undersampling: not all species occurring in a community are recorded in a plot.
Fair comparison: to compare the samples of equal completeness (here measured as *sample coverage*).
Ramet: every individual plant unit potentially capable of existence; observed individual in the field.
Individual: a genet; number of ramets modified by the set of rules (see Methods).
Liberal estimation: count of ramets; liberal approach to estimation of number of plant individuals.
Conservative estimation: count of plant individuals (genets); conservative approach to estimation of number of plant individuals.

Studied vegetation

I II III IV V VI

I **Mesic beech forest** (*Galio odorati-Fagetum sylvaticae* Sougnez et Thill 1959)
II **Mountain spruce forest** (*Calamagrostio villosae-Piceetum abietis* Schlüter 1966)
III **Lowland floodplain forest** (*Fraxino pannonicae-Ulmetum glabrae* Aszód 1935 corr. Soó 1963)
IV **Acidophilous thermophilous oak forest** (*Sorbo torminalis-Quercetum* Svoboda ex Blažková 1962)
V **Basiphilous thermophilous oak forest** (*Euphorbio-Quercetum* Knapp ex Hübl 1959)
VI **Oak-hornbeam forest** (*Primulo veris-Carpinetum betuli* Neuhäusl et Neuhäuslová in Neuhäuslová-Novotná 1964)

Aims

- To compare absolute and relative species richness of plant communities using different ways of sample standardization:
  a) by area,
  b) by number of individuals,
  c) by completeness (*sample coverage*).
- To estimate density of individuals in each community.
- To compare two measures of community completeness.
- To find out whether the plot size should be adjusted to density of plant individuals in order to make a fair comparison of species richness.

Methods

We selected six forest plant communities with respect to differences in their diversity and presumed density of plant individuals (see Studied vegetation). For each community, twelve vegetation plots of size 400 m² were recorded at randomly selected locations (in our study, all species recorded in these twelve plots represent community richness). Additionally, vegetation plots of size 100 m² (nested within the plot 400 m²) were recorded for five out of these twelve plots. Furthermore, numbers of plant species and ramets were recorded in 100 evenly placed 0.04 m² subsamples for each of the five 100 m² plots. To determine the number of individuals (genets), we developed a set or rules, taking clonal growth into account, and use them to modify the numbers of recorded ramets. These rules were based on records available at the database of clonal growth (Klimešová & Klimeš 2006), from which we considered character of clonal growth organs, their persistence and lateral spread and an estimate of the longest distance an individual of each species can possibly reach. Therefore, true estimation of number of individuals is delimited by liberal (i.e. ramets) and conservative (i.e. genets) approaches. In analyses, only the conservative approach was used.

For standardization by area, species richness of 100 m² plots is compared. For standardization by individuals, plant individuals of set of 100 subsamples were rarefied to the same number (i.e. 30), using individual based rarefaction. For standardization by *sample coverage*, subsamples were rarefied or extrapolated to the same coverage (83.1 %). As a rule, samples were extrapolated no more than double reference sample size. *Sample coverage* based rarefaction curve (mixed from all subsamples from a locality) was used to determine the required number of subsamples to *given sample coverage*. For analyses, iNEXT package (Hsieh et al. 2014) for R software (R Core Team 2014) has been used.

Picture of frame for sampling 0.04 m² subsamples, with herb species of *Anthoxanthum odoratum*, *Luzula sylvatica*, *Trientalis europaea* and *Vaccinium myrtillus*. Photo was taken in the mountain spruce forest.

Results

Community richness (number of species recorded in all twelve 400 m² plots).

Mean number of individuals in one set of 100 subsamples (100 × 0.04 m²).

Mean species richness using three different approaches of sample standardization.

A comparison of two measures of community completeness: 1) *sample coverage* (Chao & Jost 2012) and 2) *Community Completeness Index* (Pärtel et al. 2013).

Required number of subsamples (0.04 m²) to *sample coverage* 97 %.

Sample coverage based rarefaction curve. Solid line represents reference sample, dashed line represents extrapolation.

Conclusions

- Communities differ in relative species richness using different sample standardization. Using standardization by individuals or by *sample coverage*, species richness of acidophilous thermophilous oak forest is the lowest.
- Communities differ in density of individuals. Density is several times higher in oak forests than density in the mesic beech forest.
- If we want to compare species richness of plant communities properly, the size of the plot should reflect both density of individuals and community richness.
  - Theoretically, communities with lower density of individuals and higher richness should be sampled by larger plots than those with higher density and lower richness, respectively.
  - Practically, however, this recommendation is not feasible, since decision about sampling size would require previous knowledge about density of individuals and community richness, which is usually not available at the time of sampling.

Acknowledgements

We would like to thank to our colleagues who participated in collecting data in the field; namely to Markéta Chudomelová, Helena Prokešová, Hana Sekerková, Ching-Feng Li (Wendy) and Anni Pyykönen. Credit for all photos goes to Ching-Feng Li. This study has been supported by the Czech Science Foundation (P505/12/1022).

References

Chao A. & Jost L. (2012): Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. – Ecology, 93: 2533–2547.

Hsieh T.C., Ma K.H. & Chao A. (2014): iNEXT: An R package for interpolation and extrapolation in measuring species diversity. http://chao.stat.nthu.edu.tw/blog/software-download/inext-r-package.

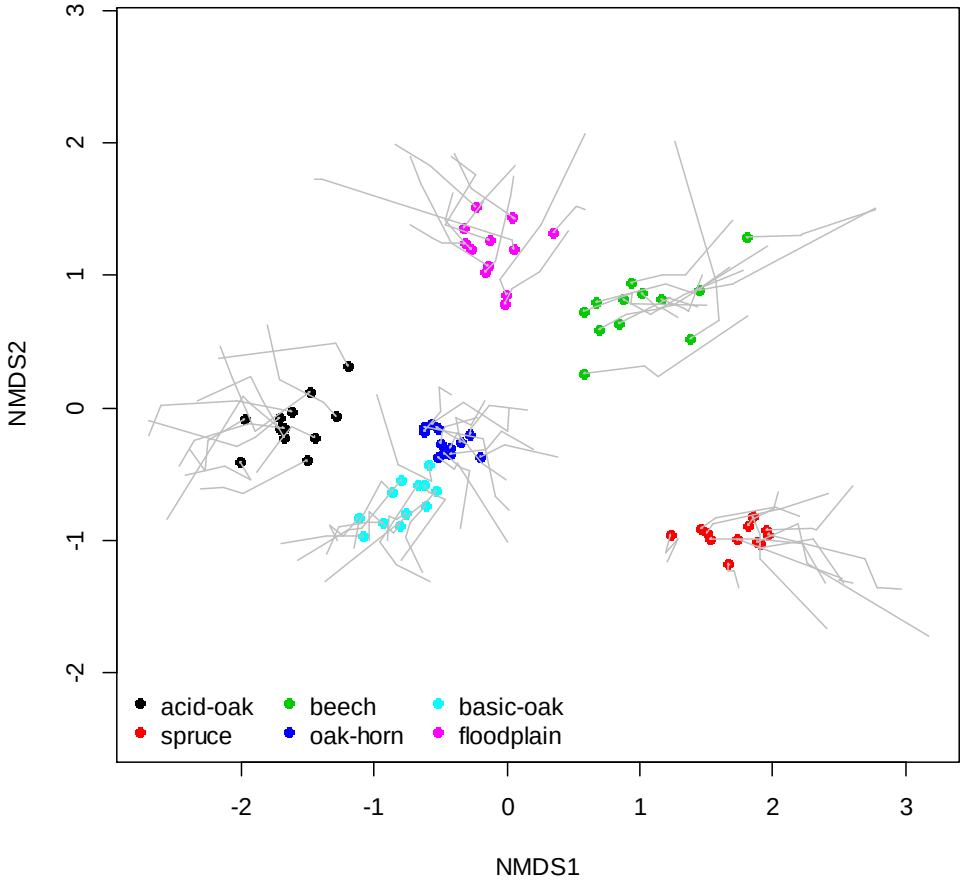
R Core Team (2014): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Klimešová J. & Klimeš L. (2006): CLO-PLA 3: a database of clonal growth architecture of Central-European plants. http://clopla.butbn.cas.cz/.

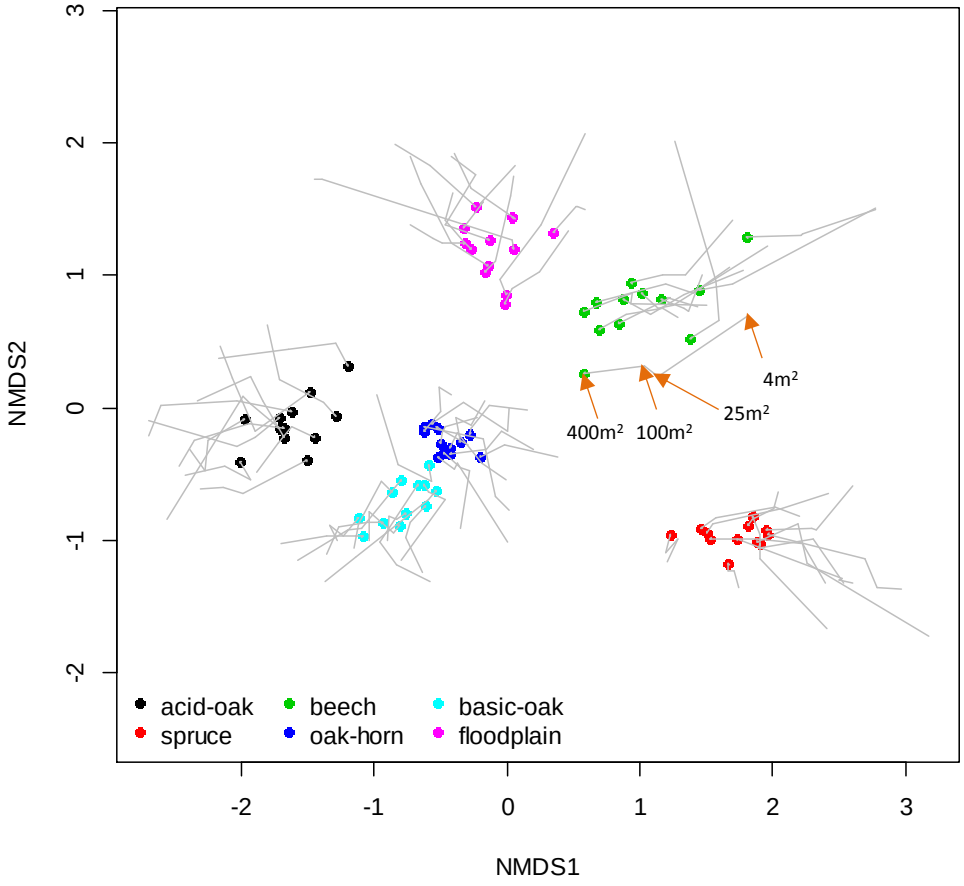
Pärtel M., Szava-Kovats R. & Zobel M. (2013): Community Completeness: Linking Local and Dark Diversity within the Species Pool Concept. – Folia Geobotanica, 48: 307–317.

Poster of Dominik Zukal **C-19**

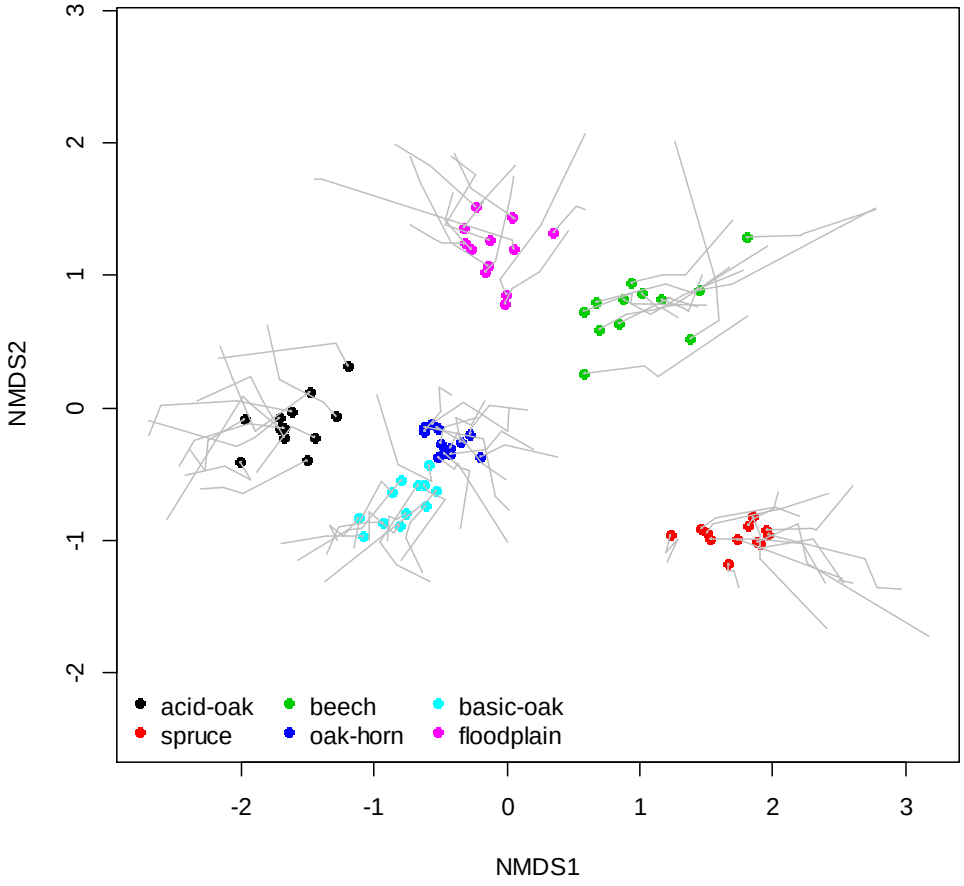
Differences in species composition among six communities



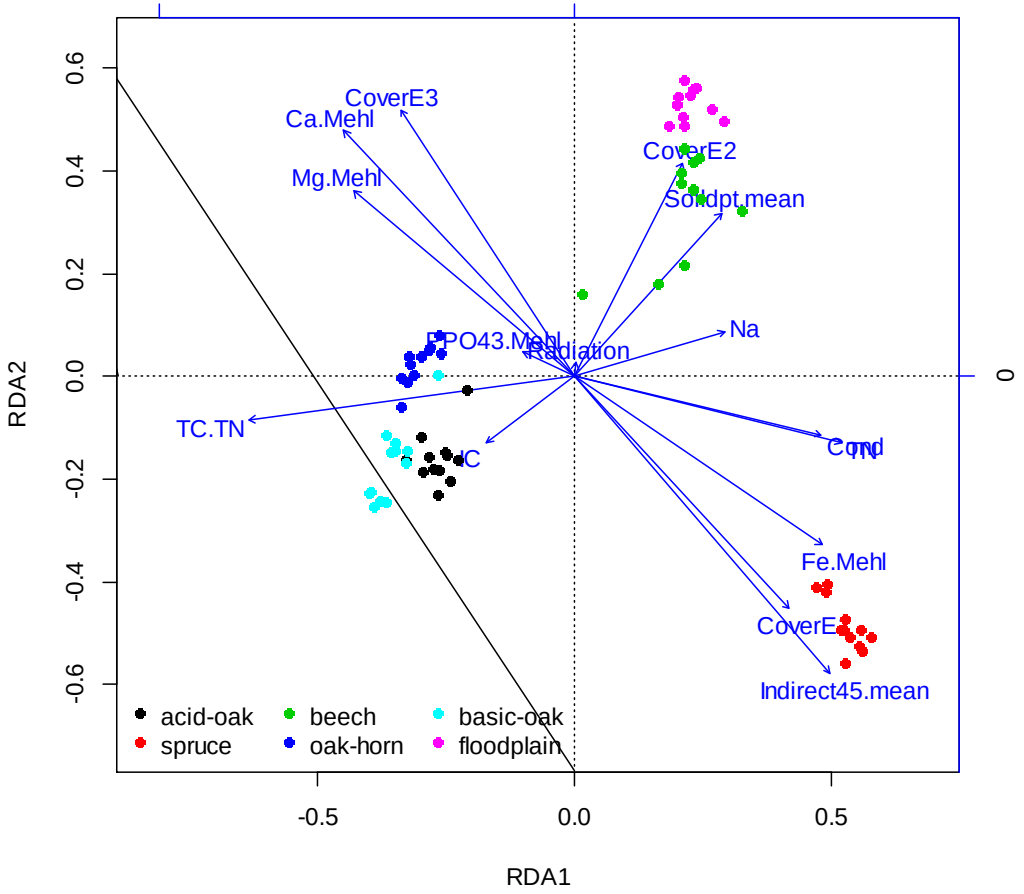
Differences in species composition among six communities



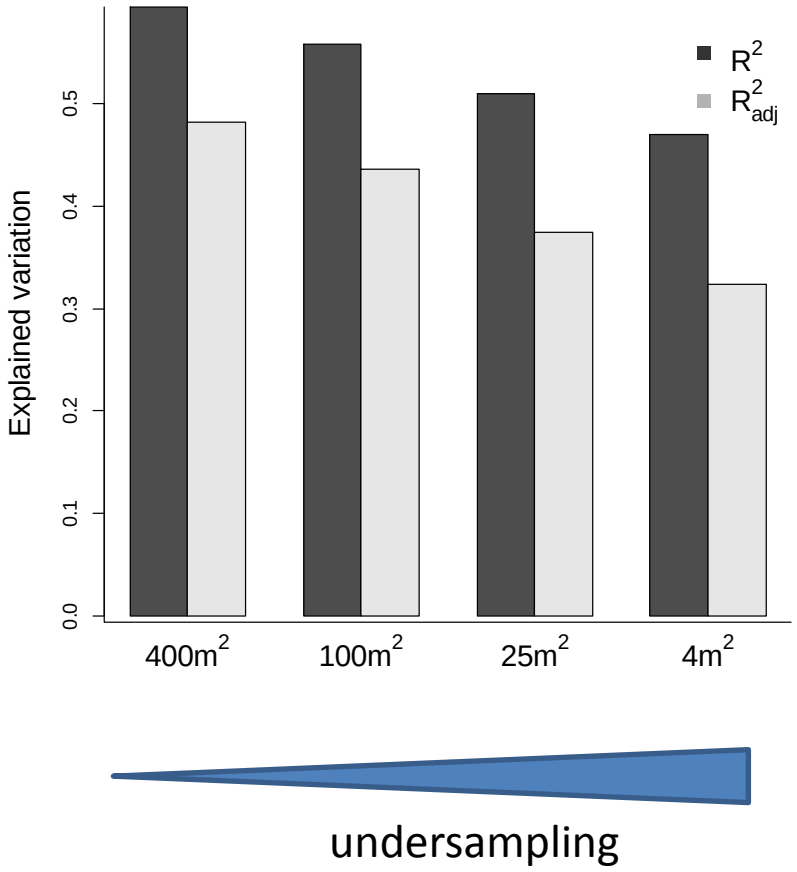
Differences in species composition among six communities



Explaining differences in species composition by environment (RDA)



Relationship between explained variation and plot size



Explained variation in beta diversity depends on:	The way to fix it:
number of samples and number of explanatory variables	use of adjusted R^2 (Peres-Neto et al. 2006)
quality of explanatory variables (not all relevant ones were measured)	e.g. use of spatial variables as surrogates for environmental factors (Dray et al. 2012)
level of undersampling (and community completeness)	by null models (e.g. Chase et al. 2011, Myers et al. 2012, De Caceres et al. 2012) *(but only for individual based data, and only unders assumption that undersampling doesn't differ within the dataset)*

Conclusions

Explained variation in beta diversity cannot be compared between different datasets, unless we are reasonably sure that the level of undersampling is on comparable level

Explained vari
be compared l
unless we are
of undersampl

Soinnen (2014) *Ecology*

Ecology, 95(12), 2014, pp. 3284–3292
© 2014 by the Ecological Society of America

A quantitative analysis of species sorting across organisms and ecosystems

JANNE SOININEN[1]

Department of Geosciences and Geography, University of Helsinki, P.O. Box 64, FIN-00014 Finland

Abstract. The estimation of the relative roles of environment, space, and their interactions in structuring community composition is one of the central topics of community ecology. I conducted a quantitative review to determine if the degree of species sorting (SS) by the abiotic environment varied predictably between organism types and ecosystems. SS was quantified as the relative fraction of community variation that is explained by environmental variables. I integrated data from 326 variation partition analyses in a generalized linear model, and found that a mean of 26.1% (minimum 0%, maximum 88%) of the variance in community composition was explained by environmental variables. I also found that organism body size and dispersal group were not related to the degree of SS. SS varied among trophic positions, being highest in autotrophs and omnivores and lowest in herbivores and decomposers. SS also varied among ecos
environments. Stuc
on presence–absen
of by polynomials
emerged for trophi
different dispersal
on how the degree
insights into the n
environment.

Key words: body scale; spatial variatio

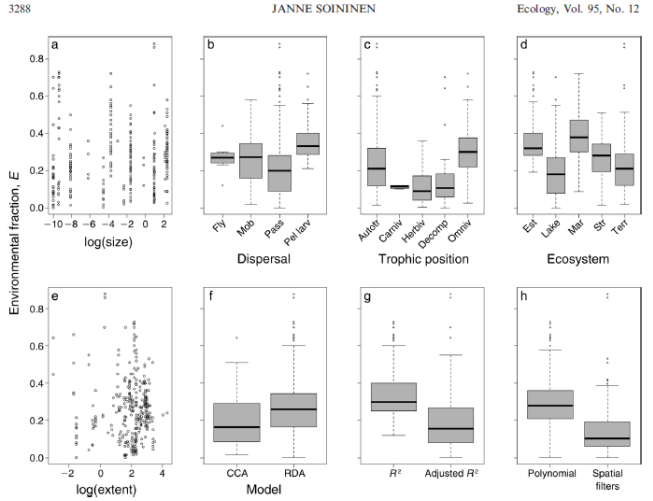


FIG. 1. Observed values for the whole data set ($n = 326$ SS values analyzed) of the environmental fraction (E, the relative amount of community variation that is explained by environmental variables) against (a) log-transformed body size (originally measured in g), (b) dispersal mode, (c) trophic position, (d) ecosystem type, (e) log-transformed study extent (originally measured in km), (f) model type (canonical correspondence analysis and redundancy analysis), (g) adjusting, and (h) spatial filters vs. polynomial. Lines in boxes are medians, box ends are quartiles, whiskers show 95th percentiles, and circles show outliers. Abbreviations are: Fly, flying; Mob, mobile; Pass, passive; Pel larv, pelagic larvae; Autotr, autotroph; Carniv, carnivore; Herbiv, herbivore; Decomp, decomposer; Omniv, omnivore; Est, estuary; Mar, marine; Str, stream; Terr, terrestrial.

Conclusions

Explained variation in beta diversity cannot be compared between different datasets, unless we are reasonably sure that the level of undersampling is on comparable level

Thank you for your attention!