COMMUNITY ECOLOGY 18(1): 117-119, 2017
1585-8553 © AKADÉMIAI KIADÓ, BUDAPEST
DOI: 10.1556/168.2017.18.1.13

On the operative use of community overlap in analyzing incidence data

D. Schmera

MTA Centre for Ecological Research, Balaton Limnological Institute and MTA Centre for Ecological Research, GINOP Sustainable Ecosystem Group, Klebelsberg K. u. 3, H-8392 Tihany, Hungary.
E-mail: schmera.denes@okologia.mta.hu

Keywords: Co-occurrence, Overlapping species, Similarity, Total overlap size.

Abstract: Although overlap of communities is a key issue in studies ranging from community ecology to biogeography, a clear definition of community overlap and related terms hinder the development of the field. The absence of a unified terminology is remarkable even when the overlap of a pair or multiple communities is characterized. As a remedy, I suggest a definition of community overlap and two measures of it (number of overlapping species and total overlap size). Although both measures quantify different aspects of community overlap, in studying pairs of communities they yield identical results. The present findings demonstrate the need for a unified terminology in research on community overlap as well as for pairwise and multiple measures for quantifying the phenomenon.

Introduction

The analysis of incidence data has a long history in studies ranging from community ecology to biogeography (Veech 2014). Species incidence among sites can be interpreted in two complementary ways (Arita 2015): in analyses by species (Pielou 1977) and in analyses by sites (Koleff et al. 2003). Here, I focus on analyses by sites and thus, when studying overlap, I concentrate on the overlap of communities or, in other words, on community overlap. Although community overlap is a key concept in studying community patterns and therefore frequently used for quantifying compositional similarity (Jost et al. 2011), the lack of clearly defined terms and measures still hinders the development of the field. For instance, Arita (2015) has claimed fairly recently that [community] overlap "corresponds to the number of species that are shared between sites" and defined [community] overlap as "the number of sets [sites] sharing a given species". Although both definitions are related to community overlap, I feel that some clarification is in order. I argue that community overlap is a multifaceted phenomenon (a pattern) that can be quantified in different ways. Obviously, the operative use of community overlap requires the separation of the phenomenon (i.e., the pattern) from its measures (i.e., a quantitative property of the pattern), as well as the use of distinct and self-explanatory terms for both the phenomenon and its measures. In my view, Arita (2015) violates these criteria because he understands the term overlap as a phenomenon (the pattern), as well as two distinct measures (e.g., the number of species and the number of sites). As a remedy, I suggest a *definition* of community overlap and propose *measures* quantifying different properties of the pattern. I hope that the new terminology allows an operative use of community overlap in analyzing incidence data.

Definition of community overlap and related terms

I suggest that, in analyzing species incidence data by sites, *community overlap represents the intersection in the composition of communities*. This definition means that *community overlap is a phenomenon* which exists when species occur in more than one community. In other words, overlap among communities exists when the set of communities contains at least one species present in two or more of them and does not exist when all species are present only in a single community. It follows that community overlap is manifested through *overlapping species, i.e., through species with at least two occurrences in the set of communities*.

Measures of community overlap

Two-community situation

Figure 1 shows a hypothetical set of two communities. Species 4 and 5 are overlapping species while species 1, 2, 3 and 6 are not. A straightforward way of quantifying overlap is via the number of overlapping species. Although Arita (2015, in his Table 1) calls this number as the number of overlaps, I disagree with this terminology, because this would confuse two phenomena (community overlap and overlapping species) and a phenomenon with a measure (community overlap and number of overlapping species). I suggest that the measure counting overlapping species should be termed as the *number of overlapping species*.

It follows that widely used similarity indices express community overlap in a relativized form. In a more formal way, similarity indices are commonly expressed in terms of a 2 × 2 contingency table in which *a* refers to the number of species present in both sites being compared (number of shared species, or the number of overlapping species), *b* to the number

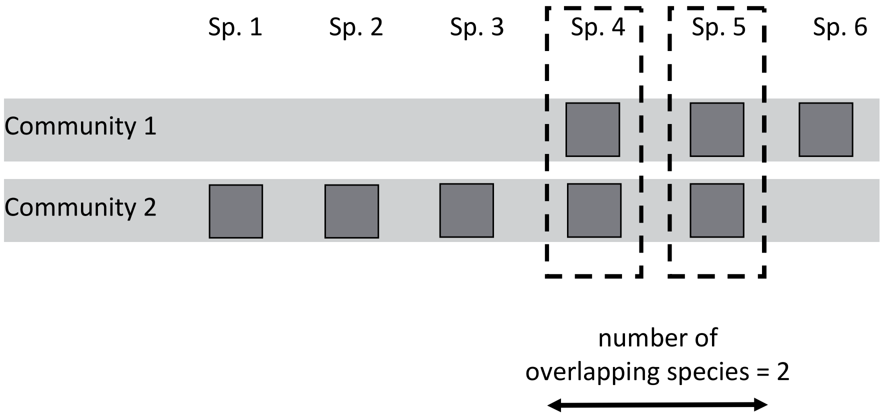

Figure 1. Schematic representation of the intersection of two communities. Communities are in rows (Communities 1 and 2), species are in columns (Sp. 1 -6). Species presence is displayed with a square. Overlapping species are highlighted by dashed border.

of species present only in the first and *c* to the number of species in the second. The Simpson similarity index (Simpson 1943) quantifies the number of overlapping species in relation to the number of species in the poorer community:

$$\frac{a}{a+\min(b,c)},$$

the Jaccard similarity (Jaccard 1912) expresses the number of overlapping species divided by the number of species present in the two communities being compared:

$$\frac{a}{a+b+c},$$

while the Sørensen similarity (Sørensen 1948) is obtained as the number of overlapping species divided by the average number of species in the two communities:

$$\frac{2a}{2a+b+c}=\frac{a}{\frac{1}{2}(\ a+b)+(a+c)}.$$

Multi-community situation

Figure 2 shows a hypothetical set of three communities. Species 9, 10, and 11 are overlapping species. In examining overlapping species, one should recognize that they show quantitative differences: species 9 and 11 occur in 2 sites, while species 10 in 3 sites. Arita (2015, p. 9 of the online document) recognized this and argued that, "in multisite analyses, a distinction has to be made between general overlap (the number of sets sharing a given species) and the number of species that are shared by each pair of sites". This wording is misleading, because the phenomenon is confounded with its measurement.

To avoid confusion, I suggest to characterize overlapping species by a quantitative property called *overlap size*, and – in agreement with the intention of Arita (2015) – I also suggest to quantify it as the occurrence frequency of the species minus one (note that overlap size can also be applied to singleton species but it is equal to 0). It is important to note that overlap size is a measure related to overlapping species, but not to community overlap. To characterize community overlap with the overlap sizes, I propose the sum overlap sizes of species in the set of communities, which may be called as the *total overlap size*. In a formal way, if n_i is the range size (number of sites occupied) by species *i* and *S* is the total number of species in the set of communities, then total overlap size equals to

$$\sum_{i=1}^{S} n_i - S .$$

Remember that if pairs of communities are studied, then total overlap size equals to the number of overlapping species. This suggests that community overlap is a complex phenomenon and its multifaceted nature might be hidden when only pairs of communities are studied.

A careful reader should realize at this point that the general overlap indices of Arita (2015) use total overlap size in a relativized form. If s_j denotes the species richness of community *j*, then the Simpson general overlap index (see Table 1 in Arita 2015) expresses the total overlap size in relation to the maximum of total overlap size if the communities show nested design:

$$\frac{\sum_{i=1}^{S} n_i - S}{\sum_{i=1}^{S} n_i - \max(s_i)},$$

while the Jaccard general overlap index (Koch 1957) is the total overlap size in relation to the maximum of total overlap size with *N* communities and *S* species (see Table 1 in Arita 2015):

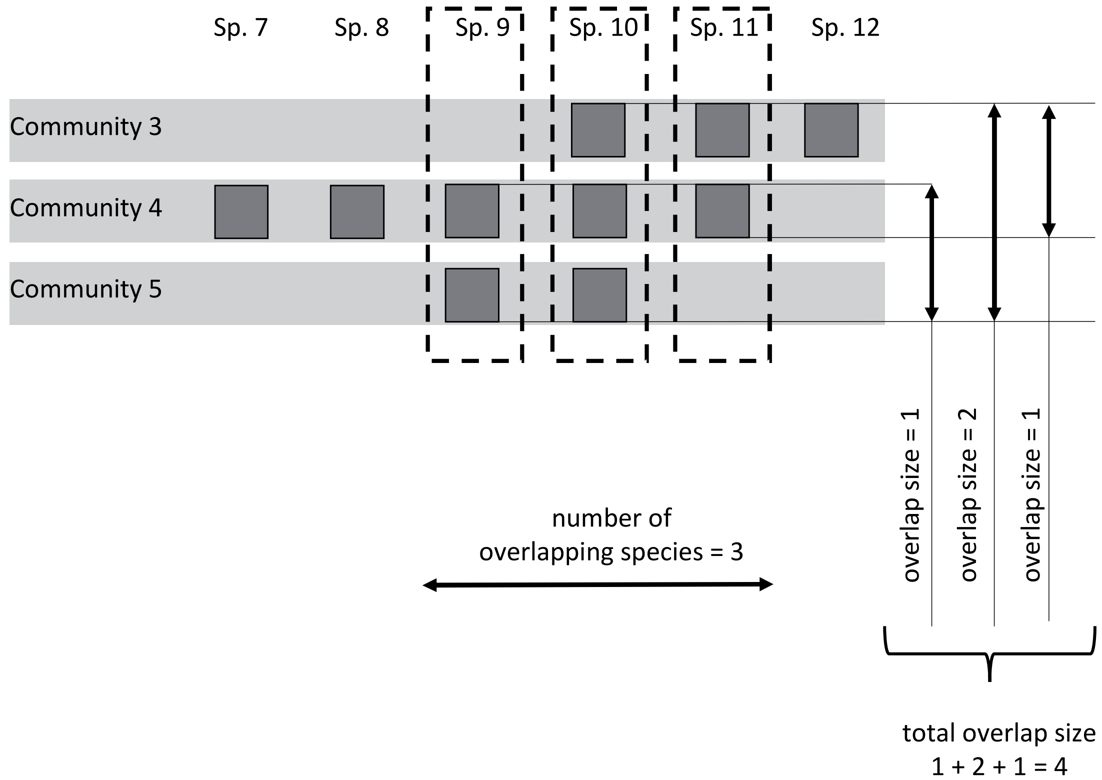

Figure 2. Schematic representation of the intersection of 3 communities. Communities are in rows (Communities 3, 4 and 5), species are in columns (Sp. 7 -12). Species presence is displayed with a square. Overlapping species are highlighted by dashed border.

$$\frac{\sum_{i=1}^{S} n_i - S}{S(N-1)}.$$

That is, the total overlap size is already part of the toolkit of numerical ecology in a relativized form, called as general overlap measures (Arita 2015).

Conclusions

Although community overlap is a key issue in studies ranging from community ecology to biogeography, I found that the phenomenon and its measurement are not clearly separated and that the same term is used with different meanings. To avoid these shortcomings, I provided a definition of community overlap and a related term (overlapping species). I suggested two measures of community overlap: the number of overlapping species and the total overlap size. Both measures can be applied and interpreted when pairs or multiple communities are studied. I argue that the new definitions and measures allow the operative study of community overlap, and thus contribute to the proper interpretation of compositional similarity.

Acknowledgement. This research was supported by GINOP-2.3.2-15-2016-00019 project.

References

Arita, H.T. 2015. Multisite and multispecies measures of overlap, co-occurrence, and co-diversity. *Ecography* doi: 10.1111/ecog.01942

Jaccard, P. 1912. The distribution of the flora in the alpine zone. *New Phytol.* 11: 37–50.

Jost, L., A. Chao and R.R. Chazdon. 2011. Compositional similarity and β (beta) diversity. In A.E. Magurran and B.J. McGill (eds), *Biological Diversity. Frontiers in Measurement and Assessment.* Oxford University Press, Oxford. pp. 66–84.

Koch, L.F. 1957. Index of biotic dispersity. *Ecology* 38, 145–148.

Koleff, P., K.J. Gaston and J.L. Lennon. 2003. Measuring beta diversity for presence-absence data. *J. Animal Ecol.* 72: 367–382.

Pielou, E.C. 1977. The latitudinal spans of seaweed species and their patterns of overlap. *J. Biogeogr.* 4: 299–311.

Simpson, G.G. 1943. Mammals and the nature of continents. *Amer. J. Sci.* 241: 1–31.

Sørensen, T.A. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content, and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Viden- skabernes Selskabs Biologiske Skrifter* 5: 1–34.

Veech J.A. 2014. The pairwise approach to analysing species co-occurrence. *J. Biogeogr.* 41: 1029–1035.

Received April 15, 2017
Revised May 4, 2017
Accepted May 18, 2017