

On the inference of positive and negative species interactions and their relation to abundance

Andrew J. Rominger

Why do rare species persist in ecosystems? Rare species seem to be at a disadvantage by pure probabilistic odds¹ and perhaps also from poorly adapted species-environment and species-species interactions², though negative density-dependence may help rare species persist^{3,4}. The question of rarity and persistence thus remains unresolved. In a recent paper, Calatayuda et al.⁵ (CEA) have contributed toward helping resolve this question. They compiled an impressive collection of datasets, across many taxa and environments, capturing spatially replicated species abundance measures. With these data they inferred species-species interaction networks. CEA found that rare species were statistically associated with positive interactions whereas common species were associated with negative interactions, indicating that positive interactions, such as facilitation, may help rare species persist⁵. However, the use of abundance data to infer species interactions is difficult and often inaccurate⁶⁻⁸. Here, I show that the finding of an association between abundance and interaction type as reported by CEA can be explained by statistical artifacts in the inference of species interactions from abundance data. These artifacts arise because of spatial clustering in species abundances. It would therefore not be supported to assign biological interpretations to associations between interaction types and abundances until more data can be brought to bear on the subject of interaction types and the persistence of rare species.

When interaction networks are inferred from spatially replicated abundance data, species-species co-occurrences are quantified by a metric (e.g., CEA use Schoener similarity⁹) and then a null model is used to assess whether these co-occurrence metrics deviate substantially enough from null expectations to suggest a non-random interaction, either in the positive direction (suggesting positive interactions) or the negative direction (suggesting negative interactions). However, if abundances are driven by processes, both probabilistic or deterministic, other than species-species interactions, these null models may not reveal true interactions, but rather artifacts of other processes. Species abundances are not evenly distributed within species nor across space (often referred to as spatial clustering)¹⁰⁻¹⁴. This ubiquitous pattern can be accounted for by purely probabilistic processes from neutral birth-death-immigration¹⁵ to mechanistically agnostic statistical-mechanical properties of large assemblages¹³. Thus, the simple observation of uneven abundances does by itself indicate the influence of deterministic species interactions. The data compiled by CEA⁵ indeed confirm the ubiquity of uneven species abundances (Supplementary Figs. 2 and 3).

In Figure 1 I first reproduce key results from CEA's Figure 2(B-C). Then to evaluate whether these results can be produced simply from spatial clustering alone I simulate purely random data that match the unevenness of abundance found in the observed data but contain absolutely no species interactions. These random data are simulated as follows:

- 1) The number of species S , number of sites M , and shape of the best fitting species abundance distribution (SAD) are sampled (with replacement) from the observed data
- 2) S species abundances $x_i \dots x_S$ are sampled from the SAD
- 3) For each x_i , within-species counts are distributed across the M sites according to a spatial species abundance distribution (SSAD) that is either negative binomial (in the case of spatial clustering) or Poisson (in the case of spatial evenness)
- 4) The resulting simulated site by species matrix is fed through the same analytically pipeline (described in CEA) as the observed data to infer positive and negative interactions.

All analyses are carried out in R¹⁷ and can be fully reproduced by installing the R package accompanying this paper, as detailed in the supplement.

In the case of a Poisson SSAD the one parameter (the mean) is fully specified by the average site-level abundance of a given species. In the case of a negative binomial SSAD, the mean parameter is again specified by the site-level average, but the size or clustering parameter k is not fully specified. To capture the rough

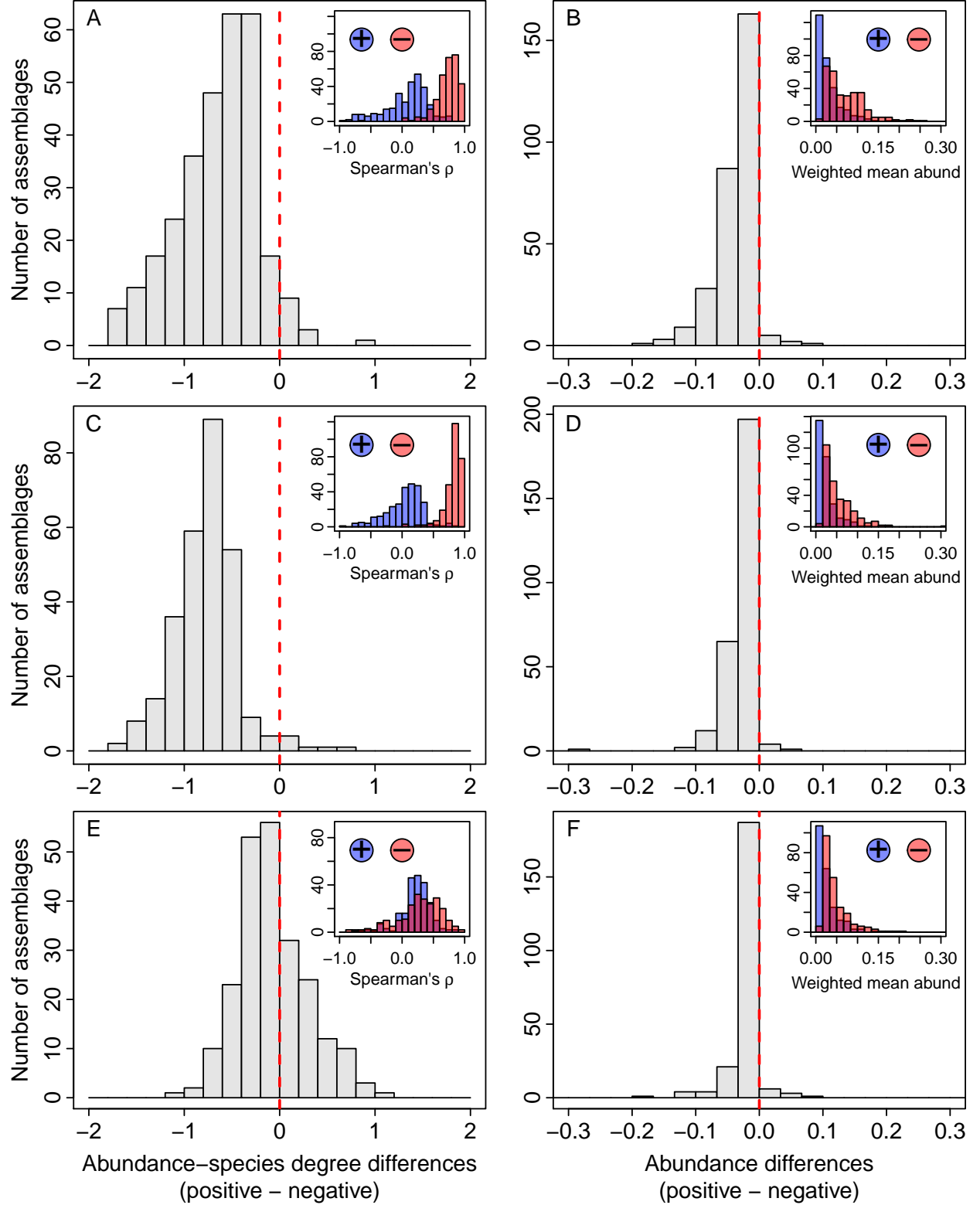


Figure 1: Distributions of correlations between network centrality (i.e. species degree) and abundance (left panels) and distributions of weighted mean abundances (right panels). The main figures show the differences between positive and negative interaction networks, while the inset figures show the separate distributions for each network. The results of CEA Figure 2(B-C) are reproduced here in panels A-B; panels C-D show data simulated with a negative binomial SSAD and no species interactions; panels E-F show data simulated with a Poisson SSAD and no species interactions.

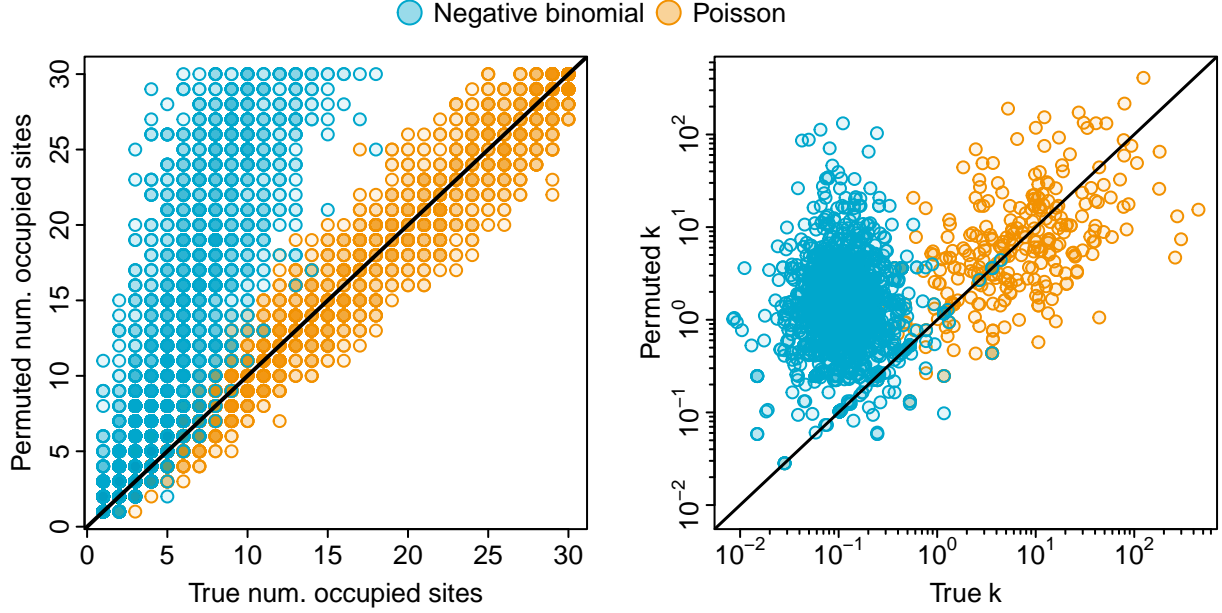


Figure 2: Comparison of SSAD statistics for true and permuted site by species matrices. Colors correspond to the true, un-permuted SSAD. Panel (A) shows how permutation affects number of occupied sites. Panel (B) shows how permutation affects maximum likelihood estimates of the clustering parameter k (B). Points are semi-transparent to help display density. Lines are 1:1 lines.

features of the data, I sample k from a linear relationship (with noise) between the maximum likelihood estimates of k and the relative abundance of each species (Supplementary Fig. 2).

Figure 1 A–D shows that with a negative binomial SSAD, simulated data closely match observed findings: the correlation between abundance and species’ network degree skews more negative in positive interaction networks (i.e. more rare species are more highly connected in positive interaction networks), and positive interaction networks tend to contain more rare species than negative networks. This correspondence between real and simulated patterns largely disappears when we instead use a Poisson SSAD, highlighting the importance of spatial aggregation in driving the spurious results.

My findings do not depend on simulating SAD and SSAD shapes from the data: in Supplementary Figure 4 I show that the spurious relationship between abundance and interaction type occurs even when simulating data from just one arbitrary SAD function with the one arbitrary spatially clustered SSAD for all species. In this simulation, again, replacing the spatially clustered SSAD with a Poisson SSAD breaks the spurious association as in Figure 1 (E–F).

Why do negative binomial SSADs reproduce the results while Poisson SSADs fail to? The null model algorithm used here and in CEA fixes row and column marginals, but within any given species, the way its total abundance is allocated across sites by the null model has a potentially large combinatorial space to explore. I compare known SSADs to their permuted counterpart in Figure 2 and find that the null model transforms negative binomial SSADs to a more Poisson shape, while leaving Poisson SSADs probabilistically unchanged. Specifically, when starting with a negative binomial SSAD, the null model inflates the number of sites individuals are allocated to (more similarly to a Poisson SSAD) and increases the inferred k parameter, indicating less spatial clustering in the permuted matrices compared to their non-permuted, negative binomial starting points.

At a mathematical level, clustered SSADs, compared to spatially even SSADs, actually increase the probability that rare species will appear aggregated with each other and common species will appear repelled. Consider for example two rare species, one with a single individual and the other with abundance 5, distributed across 5 sites. Their Schoener similarity is maximized when all individuals occur at the same site, such as this site

by species matrix

$$X_{rare} = \begin{bmatrix} 1 & 5 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

If we define $Q(x_i; \mu = 1)$ as the probability of observing x_i individuals in site i given an SSAD with mean parameter μ , then the probability of the above configuration is $P(X_{rare}) = Q(5; \mu = 1) (Q(0; \mu = 1)^4)$. Under a negative binomial SSAD with $k = 0.1$, $P(X_{rare}) = 4.58 \times 10^{-3}$ whereas under a Poisson SSAD $P(X_{rare}) = 5.61 \times 10^{-5}$.

Conversely, for two common species, say each with abundance 50, an example configuration that *minimizes* their Schoener similarity would be

$$Y_{min} = \begin{bmatrix} 50 & 0 \\ 0 & 50 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

We calculate the probability of any such scenario where no abundances overlap as $P(Y_{min}) = 4 \left((Q(50; \mu = 10) Q(0; \mu = 10)^4)^2 \right)$. With a negative binomial SSAD with $k = 0.1$, $P(Y_{min}) = 1.41 \times 10^{-7}$ whereas with a Poisson SSAD $P(Y_{min}) = 1.61 \times 10^{-72}$.

We contrast this with a configuration that would maximize the Schoener similarity between these two common species:

$$Y_{max} = \begin{bmatrix} 10 & 10 \\ 10 & 10 \\ 10 & 10 \\ 10 & 10 \\ 10 & 10 \end{bmatrix}$$

The probability of this configuration is $P(Y_{max}) = Q(10; \mu = 10)^{10}$. For the same negative binomial $P(Y_{max}) = 5.76 \times 10^{-22}$, and for the Poisson $P(Y_{max}) = 9.40 \times 10^{-10}$.

Thus a spatially clustered SSAD, compared to a spatially even SSAD, gives more probability to configurations where rare species appear aggregated and common species appear over-dispersed. Because the null model algorithm permutes site by species matrices to resemble more spatially even SSADs this probabilistic difference between spatially clustered versus even SSADs accounts for the prevalence of rare species in positive interaction networks and common species in negative interaction networks.

Great caution should be used when inferring species interactions from abundance data. More fundamentally than the spurious association of abundance with interaction type, my analysis shows that statistically significant species interactions are inferred from data simulated without any interactions. In data simulated with a negative binomial SSAD, on average 75% of species were placed in positive interaction networks and 74% in negative interaction networks with a significance cutoff of $\alpha = 0.05$. With the Poisson SSAD these simulated numbers were 71% for positive networks and 25% for negative networks. For the observed data, on average 73% of species were placed in positive interaction networks and 60% in negative interaction networks. It is becoming increasingly appreciated that abundance data alone are not sufficient to distinguish between different ecological processes^{18,19}. The question of why rare species persist is fascinating, and CEA should be commended for making a concerted effort to illuminate possible mechanisms underlying the phenomenon; however, to reach robust conclusions, other types of data, such as actual experimental measurement of interaction strengths, are needed in addition to abundance data.

References

1. McGill, B. J., Hadly, E. A. & Maurer, B. A. Community inertia of quaternary small mammal assemblages in north america. *Proceedings of the National Academy of Sciences* **102**, 16701–16706 (2005).
2. Hutchinson, G. E. The paradox of the plankton. *The American Naturalist* **95**, 137–145 (1961).
3. Leigh Jr, E. G. *et al.* Why do some tropical forests have so many species of trees? *Biotropica* **36**, 447–473 (2004).
4. Yenni, G., Adler, P. B. & Ernest, S. M. Strong self-limitation promotes the persistence of rare species. *Ecology* **93**, 456–461 (2012).
5. Calatayud, J. *et al.* Positive associations among rare species and their persistence in ecological assemblages. *Nat Ecol Evol* (2019).
6. Freilich, M. A., Wieters, E., Broitman, B. R., Marquet, P. A. & Navarrete, S. A. Species co-occurrence networks: Can they reveal trophic and non-trophic interactions in ecological communities? *Ecology* **99**, 690–699 (2018).
7. Carr, A., Diener, C., Baliga, N. S. & Gibbons, S. M. Use and abuse of correlation analyses in microbial ecology. *The ISME journal* **13**, 2647–2655 (2019).
8. Rajala, T., Olhede, S. C. & Murrell, D. J. When do we have the power to detect biological interactions in spatial point patterns? *Journal of Ecology* **107**, 711–721 (2019).
9. Schoener, T. W. The anolis lizards of bimini: Resource partitioning in a complex fauna. *Ecology* **49**, 704–726 (1968).
10. McGill, B. & Collins, C. A unified theory for macroecology based on spatial patterns of abundance. *Evolutionary Ecology Research* **5**, 469–492 (2003).
11. Engen, S., Lande, R. & Sæther, B.-E. A general model for analyzing taylor’s spatial scaling laws. *Ecology* **89**, 2612–2622 (2008).
12. Zillio, T. & He, F. Modeling spatial aggregation of finite populations. *Ecology* **91**, 3698–3706 (2010).
13. Harte, J. *The maximum entropy theory of ecology*. (Oxford University Press, 2011).
14. Connolly, S. R., Hughes, T. P. & Bellwood, D. R. A unified model explains commonness and rarity on coral reefs. *Ecology letters* **20**, 477–486 (2017).
15. Kendall, D. G. Stochastic processes and population growth. *Journal of the Royal Statistical Society. Series B (Methodological)* **11**, 230–282 (1949).
16. Hubbell, S. P. *The unified neutral theory of biodiversity and biogeography*. (Princeton University Press, 2001).
17. R Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2018).
18. McGill, B. J. *et al.* Species abundance distributions: Moving beyond single prediction theories to integration within an ecological framework. *Ecology letters* **10**, 995–1015 (2007).
19. Morlon, H. *et al.* Taking species abundance distributions beyond individuals. *Ecology Letters* **12**, 488–501 (2009).