Biodiversity Informatics, 15, 2020, pp. 1-10

CO-OCCURRENCE NETWORKS DO NOT SUPPORT IDENTIFICATION OF BIOTIC INTERACTIONS

A. Townsend Peterson[1], Jorge Soberón[1],
Janine M. Ramsey[2] and Luis Osorio-Olvera[1,3]
[1]Biodiveristy Institute, University of Kansas, Lawrence, Kansas 66045 USA;
[2]Instituto Nacional de Salud Pública, Tapachula, Chiapas, México;
[3]Centro de Cambio Global y Sustentabilidad, CONCAT, Villahermosa, Tabasco, México

Abstract. We assess a body of work that has attempted to use co-occurrence networks to infer the existence and type of biotic interactions between species. Although we see considerable promise in the approach as an exploratory tool for understanding patterns of co-occurrence of species, we note and describe numerous problems in the step of inferring biotic interactions from the co-occurrence patterns. These problems are both theoretical and empirical in nature, and limit confidence in inferences about interactions rather severely. We examine a series of examples that demonstrates striking discords between interactions inferred from co-occurrence patterns and previous experimental results and known life-history details.

Introduction

A series of contributions over the past decade has explored the application of networks of co-occurrence to identifying interactions among species (Sánchez-Cordero et al. 2008; Stephens et al. 2009; González-Salazar and Stephens 2012; Ibarra-Cerdeña et al. 2017; Stephens et al. 2019). That is, the authors purport to be able to use networks of co-occurrence or non-co-occurrence, based on spatial information in primary biodiversity databases, to infer biotic interactions (Sánchez-Cordero et al. 2008). Were such an inferential ability to prove feasible, it would represent an exciting new dimension in biodiversity informatics—for instance, with the emergence and existence of large-scale biodiversity information resources (Canhos et al. 2004; Stein and Wieczorek 2004), now beyond 10^9 records easily and readily available for analysis, species' interactions could be assessed and inferred on global scales, adding an important new dimension to what have been termed "essential biodiversity variables" (Pereira et al. 2013).

This paper, however, examines critically the basic proposition that spatial co-occurrence can be translated into a hypothesis and prediction of biotic interactions. Specifically, we assess (1) whether conceptually such a connection (co-occurrence biotic interactions) should be expected to exist, and (2) the degree to which practical considerations (e.g., sampling biases) may further cloud and confuse any such relationships. Finally, (3) we examine a series of examples of such analyses, and show that in many cases, by applying the proposed methodology, conclusions are reached that are even opposite to those of both common sense and actual experimental data.

Conceptual Bases

Should a co-occurrence biotic interaction connection exist?

Imagining a (wonderful) situation in which the occurrence data used for inferences are comprehensive, complete, and unbiased, one could then estimate *patterns* of co-occurrence with some confidence. However, whether spatial co-incidence can be a proxy for biotic interactions is a theme that has been debated for decades—the co-occurrence work makes brief reference to these debates (González-Salazar and Stephens 2012; Stephens et al. 2016; Stephens et al. 2017; Stephens et al. 2019), but generally fails to cite, mention, or assume their crucial elements (e.g., Connor and Simberloff 1979; Gilpin and Diamond 1982; Hubbell 2001; Peres-Neto et al. 2001; Ulrich 2004; Ulrich et al. 2017). Quite generally, spatial co-incidence of species' distributions may be a consequence of geographic constraint, history, shared climate or substrate preferences, migratory patterns, or many other factors (Morueta-Holme et al. 2016; Freilich et al. 2018). In tandem with co-occurrence patterns, since *non*-co-occurrence may derive similarly from causes not related to biotic interactions, patterns of co-incidence and non-co-incidence are no indication of the processes causing them (Bell 2005).

As such, the network patterns that are the focus of the inferences from the co-occurrence network analyses are simply numerical representations of patterns of co-occurrence in a particular database. In particular, their SPECIES tool (Stephens et al. 2019), which calculates indices of co-occurrence of species in the form of a statistic epsilon, is certainly useful as an exploratory tool focused on patterns of co-occurrence. However, going beyond the pattern to make inferences of *process*, is immediately suspect. the co-occurrence network group ostensibly tested the predictive capacity of their method empirically by correlating their interaction coefficient epsilon values with numbers of positive tests for the presence of a parasite in independent samples (González-Salazar and Stephens 2012). However, given almost 30 years of empirical and theoretical literature casting doubts on the robustness of such results, this claim cannot be accepted at face value and based on a single test, and rather needs to be examined critically, using a properly constructed battery of tests (Gotelli and Graves 1996; Gotelli 2000).

Practicalities

What other factors become important?

The co-occurrence network approach is based on databases of primary biodiversity occurrences, which is attractive in that it is *primary* (i.e., based on data deriving directly from individual occurrence records of each species), rather than secondary information (i.e., deriving from some interpretation or synthesis of primary data). However, such data are well-known to be massively influenced by biases related to sampling, in terms of the diverse logistic, practical, historical, and political factors that structure how biologists have been able to sample biodiversity on Earth, and report those data to the broader scientific community. These biases have been documented thoroughly in general (Yesson et al. 2007; Beck et al. 2013; Gaiji et al. 2013; Otegui et al. 2013a; Otegui et al. 2013b; Beck et al. 2014; Idohou et al. 2015; Anderson et al. 2016; Asase and Peterson 2016; Peterson and Soberón 2018), and specifically for Mexico (Bojórquez-Tapia et al. 1995; Peterson et al. 1998; Soberón et al. 2000; Soberón et al. 2007).

To claim to estimate co-occurrence rates (let alone biotic interactions) robustly on the basis of such incomplete, sparse, uneven, and biased sampling is rather doubtful. For instance, in various descriptions of the co-occurrence network method (e.g., Stephens et al. 2019), the "probability" of species *C* being present in a random cell is stated as N_C/N, where N_C is the number of counts of presences of *C,* in a total of *N* cells. However, given incomplete, uneven, and biased sampling, N_C invariably underestimates the actual numbers of counts that would be obtained from a systematic documentation of the range of species C, so N_C/N will be, in the vast majority of cases, an underestimate of the true probability. The same problem occurs when estimating co-incidences of pairs or sets of species, which has been termed the "unseen shared species effect"—this problem is known to be serious, and detailed methods have been proposed to fix it (Colwell and Coddington 1994; Chao et al. 2000; Chao et al. 2005, 2006). However, these more robust estimation methods, which are clearly more appropriate than the simple, uncorrected N_C/N, are not used in the co-occurrence network methodology.

Details

The co-occurrence network work to date has used a mathematical notation that is extremely cumbersome to describe the methodology (Stephens et al. 2017; Stephens et al. 2019). For instance, X_a represents a cell, but since only the sub-index matters… why use X_a at all? We have also noted changes of meaning—for instance, in Stephens et al. (2017), X_a is used to denote a cell in one place, but elsewhere to denote a variable. The co-occurrence network methodology also falls into the trap of unnecessary introduction of symbols—in Stephens et al. (2009), the symbol for variable *i* in cell X_i is $B_i(X_i)$, but later these same variables are denoted *C* and X_i, which is confusing. Similarly, the definition of the epsilon index is presented in two different symbologies in the same paper (Stephens et al. 2009).

Much more appropriate would be to stick to the simpler and transparent symbols that are used in the software SPECIES. That is, n_j should be used to denote the number of incidences of species *j*. Then, $n_{j,k}$ can be used to denote the number of co-occurrences of species *j* and *k*; *N* is used to denote the total number of cells in the grid. Everything in the methodology (including the cardinality of complement sets) can be calculated from these three quantities. For instance, epsilon is:

$$\varepsilon(B_i \mid I_k) = \varepsilon(i,k) = \frac{n_k(n_{i,k} / n_k - n_i / N)}{\sqrt{n_k n_i (1 - n_i / N) / N}}$$

The above formula immediately raises questions. First, it is asymmetric: that is, $\varepsilon(i,k) \neq \varepsilon(k,i)$, and exploration of an example with Mexican felids (see below) suggests that the asymmetry can be by as much as 20%. What is the meaning of these two values for the co-occurrence of the two species?

Second, epsilon is dependent on the size of the grid, N, even though the magnitude and direction of the interaction of the two species should not depend on such contextual information. Third, in this formula, it is possible to have divisions by zero, which would produce undefined epsilon values. This undesirable situation will happen when $n_i = N$, which is relatively likely for coarse-grained grids. What is the meaning of these undefined values? Fourth, the authors claim that associations are significant when associated with epsilon values >1.96 (i.e., equivalent to two standard deviations of a normal distribution). Still, epsilon(Bi|lik) is most certainly not normally distributed, as any exploration of the SPECIES software will show, and therefore the use of a symmetric, normal distribution is not appropriate.

Finally, the gaps and biases in the *Sistema Nacional de Información sobre Biodiversidad* (SNIB) make the entire probabilistic argument behind the epsilon index very doubtful. The numbers in that formula, in general, cannot be regarded as probabilities, but as proportions of observations, for a given species, or proportions of observed co-occurrences, for pairs of species, in a particular database. Since the quantities N_{Xi}, $P(C \mid X_i)$, and $P(C)$ would mostly be underestimated relative to some hypothetical "true" value, often extremely so, the index epsilon will present serious problems. For example, imagine a grid of 32 km grid cells, it will have 6736 cells covering Mexico. Consider some species j with a true incidence of 60%, and some species k with a true incidence of 40%. If the true co-incidence of the two species is 20%, epsilon (j, k) is 3.79. However, if species j is sampled in only 10% of its localities, the epsilon value shifts to -0.75!

Worked examples

The co-occurrence network-based methodology has been implemented elegantly on a platform called SPECIES, which has been connected to the *Sistema Nacional para Información de la Biodiversidad* (SNIB), maintained by the *Comisión Nacional para el Uso y Conocimiento de la Biodiversidad* (CONABIO), of the Mexican government (see the "SPECIES" site[1]). This facility offers the opportunity to explore the methodology in relation to real data, and in diverse contexts. The outcome, however, is rather damning for the conclusions that one might wish to make from the co-occurrence network analyses.

For example, we explored the "interactions" between two taxa—the family Trogonidae (Aves) and the family Scarabeidae (Insecta) across Mexico, and found a complex set of attractions and interdependencies (Figure 1). That is, we noted some scarab species that were tightly and significantly associated with particular trogon or quetzal species, and others that were not closely associated at all. The interesting feature, however, is that trogons are arboreal and frugivorous, in largest part, and have tiny feet that would not permit any terrestrial activity (Forshaw 2009), whereas scarabs are terrestrial. We see no direct or indirect scenario that would lead to what could be termed biotic interactions between these two taxa, yet the epsilon index misinterprets distributional coincidence or distributional non-coincidence as positive or negative interactions.

To provide a more concrete calibration of the co-occurrence network interaction coefficients, we explored situations in which actual field experiments have been conducted. That is, we explored the co-occurrence of the rodents *Dipodomys merriami* and *Perognathus longimembris* (based on data derived from the Global Biodiversity Information Facility; https://www.gbif.org/occurrence/download/0000059-190415153152247), which yielded a substantial positive epsilon value of 15.37, which is highly statistically significant using the SPECIES methodology. A positive epsilon should reflect positive interactions (i.e., mutualism, symbiosis), yet detailed field experiments (Lemen and Freeman 1983; Lemen and Freeman 1986) indicate that these species rather have a strong *negative* interaction, in which *Dipodomys* depresses populations of *Perognathus* dramatically. Similarly, we compared three *Dipodomys* spp. against a suite of six other rodent species (based on data derived from the Global Biodiversity Information Facility[2]) all of the pairwise epsilon values were >4.39, and all were statistically significant, yet Heske et al. (1994) documented diverse, strong negative interactions among these same species.

In a further example, we took the six cat species (family Felidae) occurring in Mexico, and calculated

---

[1] http://species.conabio.gob.mx

[2] https://www.gbif.org/occurrence/download/0000059-190415153152247)

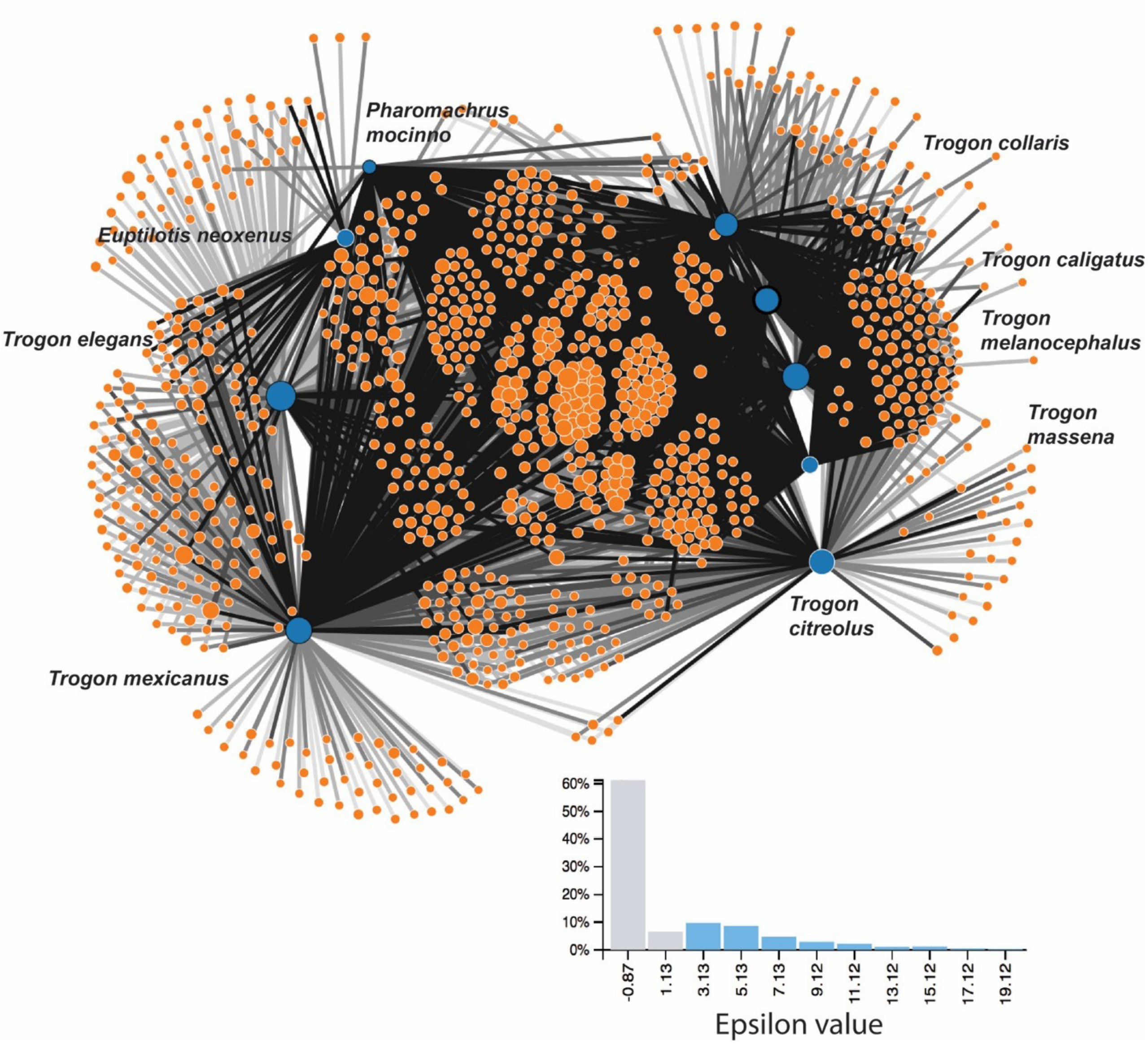

Figure 1. Summary and visualization of a co-occurrence network of trogons and quetzals (Trogonidae; blue circles) and scarab beetles (Scarabeidae, orange circles). The trogons and quetzals are labeled as to species, and highland species are aligned along the left, whereas the lowland species are aligned along the right. Scarab species are not labeled, as they are quite numerous. Inset: Distribution of epsilon values, with significant values shown in blue.

epsilon values for each pair; however, we used two distinct databases to construct the presence-absence matrices (PAMs) from which epsilon is calculated: one was the 2018 SNIB database, and the other was for the same species, but with distributional data derived from the IUCN extent-of-occurrence datasets (IUCN 2016). Both PAMs have errors, of course, but they also have contrasting properties: the IUCN data generally overpredict alpha (i.e., single-site) diversity and under-predict beta (i.e., among-site) diversity, whereas the opposite is true for the SNIB database (Lira-Noriega et al. 2007). The outcomes were quite contrasting (Figure 2): epsilon values from the IUCN data were centered on zero but quite variable, whereas those from the SNIB data were generally positive but less variable. This result highlights that the epsilon index is database-dependent, such that inferences about true processes will remain doubtful, even if, as commented above, the database were the outcome of perfect and comprehensive sampling.

DISCUSSION

Importance in ecology

Finding a way by which to infer process from pattern has always been a Holy Grail in ecology. The specific challenge of inferring biotic interactions

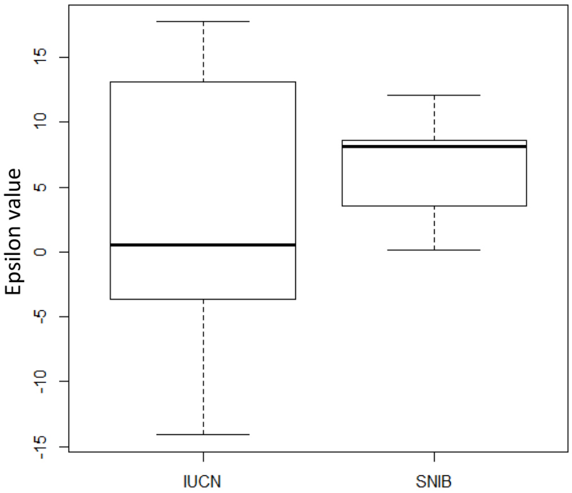

Figure 2. Boxplots of epsilon values obtained for pairwise comparisons among the six species of Felidae in Mexico, from the *Sistema Nacional para Información de la Biodiversidad* (SNIB) database, and from the International Union for the Conservation of Nature (IUCN) database. Both datasets were assessed at a spatial resolution of ½º, or about 55 km.

from data documenting only occurrences is complicated by the facts that interactions are scale-dependent, and that co-occurrences are determined by complex interactions among dispersal, physiology, behavior, habitat preferences, and evolutionary history. The bottom line is that epsilon, particularly when used with real-world data with the ever-present biases and gaps, produces highly doubtful results that are certainly not interpretable as indices of real co-incidence, much less of ecological interactions.

Now, fixes exist for this set of problems and challenges. (1) One can perform a completeness analysis for every grid cell, and use in analyses only those cells that have good completeness indices (Sousa-Baena et al. 2014). This approach will produce a database with a species x site matrix that has fewer rows (i.e., fewer sites), but those sites will have inventories that are more directly comparable. (2) One can use a species x site matrix created not from observations, but from expert data, such as IUCN's extent of occurrence maps (Hurlbert and Jetz 2007). This approach will probably overestimate incidences and co-incidences (n_j, n_k, and $n_{j,k}$), but the error would be much less marked than when using raw occurrence data, which will inevitably underestimate these three quantities. (3) One can develop detailed species distribution models, and use careful interpretations of the outputs of the models to create the species x site matrix (Rojas-Soto et al. 2003; Cooper and Soberón 2018). These three approaches allow researchers to deal with the serious problems involved in using a database of simple occurrence data; epsilon values deriving from such analyses will be a truer measure of co-incidence of species *j* and *k*.

If one desires to interpret patterns of co-occurrence as reflecting biotic interactions, approaches have been explored that would clarify and refine that interpretation. For instance, Morueta-Holme et al. (2016) refined simple interpretations of networks by correcting for indirect effects of other species, avoiding spurious associations driven by regional-scale distributions, and describing associations in multi-species contexts—these refinements permitted some degree of direct interpretation of co-occurrence networks in the context of interactions. Such refinements, however, appear not to have been contemplated—much less implemented—by the co-occurrence network group.

In sum, we perceive a set of points that would improve the ideas and methods of those who would wish to interpret co-occurrence networks as interactions. They should (1) refer to their epsilon values as simple measures of association, rather than making conclusions about interactions and other biological processes that may or may not be producing those associations. (2) They should make clear that the data in the SNIB are simply examples, and make explicit that unless the database is a "true," complete, and comprehensive presence-absence database, their epsilon calculations are not appropriate measures of co-incidence, much less indicators of ecological interactions. Finally (3), those using co-occurrence network properties should describe for what type of database and in what sense epsilon can be calculated and how this index can and should be interpreted. This latter point clearly affects much of the substance of the work, such that no doubt exists that those using these methods are seriously overinterpreting the meaning of the epsilon statistic.

Importance in epidemiology and public health

The arena in which co-occurrence networks have been used most intensively is that of detecting species interactions relevant to transmission of pathogens from reservoir or host species to humans via insect vectors (Stephens et al. 2009). The interaction networks of interest in this area are interactions between pathogens and vectors, pathogens and hosts, and vectors and hosts. In each case, knowledge of these biotic associations is too-often incomplete, fragmentary, and/or incorrect (Woolhouse and Gowtage-Sequeria 2005; Peterson et al. 2007; Estrada-Peña et al. 2015). In this sense, the potential inferences that these researchers explore are exciting, but only if they hold robust and well-founded promise of anticipating real associations.

The co-occurrence network group has focused most on transmission of *Leishmania* spp. parasites from vertebrates via sandflies of the genus *Lutzomyia* to humans (e.g., González-Salazar and Stephens 2012; Stephens et al. 2016). The key assertion is that geographic co-occurrence implies biotic interactions such as vectoring and hosting pathogens—this proposition can be examined empirically for this same system. For instance, Pech-May et al. (2010), in detailed sampling at two sites in the Calakmul region of southeastern Mexico, found 7 species of *Lutzomyia* species present, with *Leishmania* infections common; however, those infections were common in only three species (*Lu. shannoni, Lu. cruciata, Lu. ylephiletor*), and rare or absent in the remaining species. Similar results obtained in a later study in the same region, with several species co-occurring with *Leishmania* infections, but apparently not interacting with the pathogen (Pech-May et al. 2016). Similarly, Sánchez-García et al. (2010) collected large samples of 10 sandfly species in the Chetumal region of southeastern Mexico, but found only three of them to carry *Leishmania* infections. These carefully designed studies demonstrate the differential vector capacity of phlebotomine species, despite co-occurrence, and the rash nature of conclusions based simply on distributional coincidence.

In terms of *Leishmania*-mammal interactions, Rodríguez-Rojas et al. (2017) sampled rodents and sandflies in northeastern Mexico, and tested them for infection with *Leishmania*. Four of 10 rodent species were infected with *Le. mexicana* at an overall infection rate of ~10%, yet 9 sandfly species—including two species (*Lu. cruciata, Lu. shannoni*) for which sampling was numerous enough that *Leishmania* should have been detected—revealed no positive samples. This comment is not to assert that such infections do not exist, but rather to emphasize the tenuous, assumption-laden, and uncertain nature of the inferential linkage between co-occurrence in space and participation in biotic interactions, and more importantly, linkages to human disease risk.

To consider a very different class of diseases, arboviruses are of increasing concern in many regions in light of the important disease burden that they create, particularly where they are invasive (Gubler 1998). Vectorial capacity and competence are key factors driving pathogen transmission. For instance, dengue virus is now well-established in the Americas via the invasive vector species *Aedes aegypti* and *A. albopictus*, and overlaps broadly in terms of geographic distribution with large numbers of mammal species, which those using co-occurrence networks would interpret as positive interactions (e.g., host-parasite interactions). Nonetheless, dengue has been identified only tentatively in several bat species in Mexico, which likely reflects incidental infection, since an elegant recent study demonstrates that those species are not competent reservoirs, and neither co-incidence, viremia or duration of the latter are sufficient for vector infection (Vicente-Santos et al. 2017)—bats are therefore dead-end hosts for the virus, despite broad co-occurrence.

A recent review of Zika virus (ZIKV) etiology (Gutiérrez-Bugallo et al. 2019) sets out current information regarding infection and competency of vector and host associations, focusing on regions where the virus circulates. Despite known ZIKV infections in numerous species of bats in Africa and Asia, they have not been found to be competent reservoirs, *contra* the predictions of González-Salazar et al. (2017). Experimentally infected rodents, particularly mice, are not susceptible to ZIKV infection and development of sustained viremia, thanks at least in part to differences in viral detection by the rodents' innate immune system (Ding et al. 2018). Indeed, even among invertebrates found infected with ZIKV, only a handful are competent vectors (Garcia-Rejon et al. 2010). These results contrast and highlight the incomplete and weak inferences that derive from assumptions that co-occurrence implies biotic interactions that may impact human disease risk.

General conclusions

A key point in this debate, quite clearly, is why species co-occur and why do other species not co-occur? The naive application of the competitive exclusion principle (Gause 1934) would suggest that two species with identical ecological niches will not be able to coexist. This idea immediately causes concern for the co-occurrence network methodology, which uses co-occurrence to infer interactions: competitors would never be found co-occurring. However, the complexities that enter into this debate are quite daunting. That is, many species pairs indeed interact strongly in a negative sense, but are quite able to exist at different times, or on spatial scales finer than those that are considered by the co-occurrence network methodology. That is, a substantial body of ecological theory treats how the competitive exclusion principle is necessarily modified by the existence of spatial heterogeneity in a landscape (Amarasekare 2003), or by the possibility of temporal partitioning to avoid strong, negative interactions between species (Chesson and Warner 1981). Indeed, the classic book, *Geographical Ecology* (MacArthur 1972) has entire chapters treating mechanisms of co-existence of species pairs that might otherwise compete.

A related question is that of why species do not co-occur. The co-occurrence network methodology interprets non-co-occurrences as evidence of negative interactions between species. However, many other factors may enter the picture. For instance, the species concerned may simply have distinct ecological niches—that is, they may belong to lineages that have adapted to different sets of conditions, and for that reason do not co-occur. They may also have different areas of origin, which is still reflected in distinct distributional areas. The important point is that such species may never have even been *close to* one another, much less interacted (D'Amen et al. 2018).

Overall, indeed, the SPECIES software is quite attractive in that it is fast and user-friendly, constituting an elegant exploratory tool for primary biodiversity occurrence datasets. The epsilon calculations are valid tools, but should be interpreted as indicating co-*occurrence* in a particular database *only*, which may have many meanings and interpretations, depending on the particular situation and conditions. We note that more sophisticated approaches to these questions of links between co-occurrence and interactions have been published (e.g., Morueta-Holme et al. 2016) that take into account ecological niche differences and other factors (see review in Dormann et al. 2018), and that simple, site- or pixel-based spatial coincidence could be refined via consideration of spatial topology or via fuzzy spatial matching (Visser and de Nijs 2006). As we have discussed and demonstrated above, the further inferential step of interpreting epsilon as summarizing the magnitude and direction of biotic interactions is quite inappropriate.

Acknowledgments

The presence-absence matrices were compiled by Raul Sierra, a collaborator of Stephens, as part of a collaboration to analyze the performance of SPECIES when applied to different types of databases. LOO thanks CONACyT for postdoctoral fellowship support (number 740751, CVU: 368747). This research was supported in part by a grant from the National Science Foundation (IIA-1920946).

Competing Interests

The authors have declared that no competing interests exist.

Literature Cited

Amarasekare, P. 2003. Competitive coexistence in spatially structured environments: A synthesis. Ecology Letters 6:1109-1122.

Anderson, R. P., M. Araújo, A. Guisan, J. M. Lobo, E. Martínez-Meyer, A. T. Peterson, and J. Soberón. 2016. Data fitness for use in distribution modelling: Are species occurrence data in global online repositories fit for modeling species distributions? The case of the Global Biodiversity Information Facility (GBIF). Global Biodiversity Information Facility, Copenhagen.

Asase, A., and A. T. Peterson. 2016. Completeness of digital accessible knowledge of the plants of Ghana. Biodiversity Informatics 11:1-11.

Beck, J., L. Ballesteros-Mejia, P. Nagel, and I. J. Kitching. 2013. Online solutions and the 'Wallacean shortfall': What does GBIF contribute to our knowledge of species' ranges? Diversity and Distributions 19:1043-1050.

Beck, J., M. Böller, A. Erhardt, and W. Schwanghart. 2014. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. Ecological Informatics 19:10-15.

Bell, G. 2005. The co-distribution of species in relation to the neutral theory of community ecology. Ecology 86:1757-1770.

Bojórquez-Tapia, L. A., I. Azuara, E. Ezcurra, and O. A. Flores V. 1995. Identifying conservation priorities in Mexico through geographic information systems and modeling. Ecological Applications 5:215-231.

Canhos, V. P., S. de Souza, R. de Giovanni, and D. A. L. Canhos. 2004. Global biodiversity informatics: Setting the scene for a "new world" of ecological forecasting. Biodiversity Informatics 1:1-13.

Chao, A., R. L. Chazdon, R. K. Colwell, and T. J. Shen. 2005. A new statistical approach for assessing similarity of species composition with incidence and abundance data. Ecology Letters 8:148-159.

Chao, A., R. L. Chazdon, R. K. Colwell, and T. J. Shen. 2006. Abundance-based similarity indices and their estimation when there are unseen species in samples. Biometrics 62:361-371.

Chao, A., W.-H. Hwang, Y.-C. Chen, and C.-Y. Kuo. 2000. Estimating the number of shared species in two communities. Statistica Sinica 10:227-246.

Chesson, P. L., and R. R. Warner. 1981. Environmental variability promotes coexistence in lottery competitive systems. American Naturalist 117:923-943.

Colwell, R. K., and J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. Philosophical Transactions of the Royal Society of London B 335:101-118.

Connor, E. F., and D. Simberloff. 1979. The assembly of species communities: Chance or competition? Ecology 60:1132-1140.

Cooper, J. C., and J. Soberón. 2018. Creating individual accessible area hypotheses improves stacked species distribution model performance. Glob. Ecol. Biogeogr. 27:156-165.

D'Amen, M., H. K. Mod, N. J. Gotelli, and A. Guisan. 2018. Disentangling biotic interactions, environmental filters, and dispersal limitation as drivers of species co-occurrence. Ecography 41:1233-1244.

Ding, Q., J. M. Gaska, F. Douam, L. Wei, D. Kim, M. Balev, B. Heller, and A. Ploss. 2018. Species-specific disruption of STING-dependent antiviral cellular defenses by the Zika virus NS2B3 protease. Proceedings of the National Academy of Sciences USA 115:E6310-E6318.

Dormann, C. F., M. Bobrowski, D. M. Dehling, D. J. Harris, F. Hartig, H. Lischke, M. D. Moretti, J. Pagel, S. Pinkert, and M. Schleuning. 2018. Biotic interactions in species distribution modelling: 10 questions to guide interpretation and avoid false conclusions. Global Ecology and Biogeography 27:1004-1016.

Estrada-Peña, A., J. de la Fuente, R. S. Ostfeld, and A. Cabezas-Cruz. 2015. Interactions between tick and transmitted pathogens evolved to minimise competition through nested and coherent networks. Scientific Reports 5:10361.

Forshaw, J. M. 2009. Trogons: A Natural History of the Trogonidae. Princeton University Press, Princeton.

Freilich, M. A., E. Wieters, B. R. Broitman, P. A. Marquet, and S. A. Navarrete. 2018. Species co-occurrence networks: Can they reveal trophic and non-trophic interactions in ecological communities? Ecology 99:690-699.

Gaiji, S., V. Chavan, A. H. Ariño, J. Otegui, D. Hobern, R. Sood, and E. Robles. 2013. Content assessment of the primary biodiversity data published through GBIF network: Status, challenges and potentials. Biodiversity Informatics 8:94-172.

Garcia-Rejon, J. E., B. J. Blitvich, J. A. Farfan-Ale, M. A. Loroño-Pino, W. A. C. Chim, L. F. Flores-Flores, E. Rosado-Paredes, C. Baak-Baak, J. Perez-Mutul, and V. Suarez-Solis. 2010. Host-feeding preference of the mosquito, *Culex quinquefasciatus*, in Yucatan state, Mexico. Journal of Insect Science 10:32.

Gause, G. F. 1934. Experimental analysis of Vito Volterra's mathematical theory of the struggle for existence. Science 79:16-17.

Gilpin, M. E., and J. M. Diamond. 1982. Factors contributing to non-randomness in species co-occurrences on islands. Oecologia 52:75-84.

González-Salazar, C., and C. R. Stephens. 2012. Constructing ecological networks: A tool to infer risk of transmission and dispersal of leishmaniasis. Zoonoses and Public Health 59:179-193.

González-Salazar, C., C. R. Stephens, and V. Sánchez-Cordero. 2017. Predicting the potential role of non-human hosts in Zika virus maintenance. EcoHealth 14:171-177.

Gotelli, N. J. 2000. Null model analysis of species co-occurrence patterns. Ecology 81:2606-2621.

Gotelli, N. J., and G. R. Graves. 1996. Null Models in Ecology. Smithsonian Institution, Washington, D.C.

Gubler, D. J. 1998. Resurgent vector-borne diseases as a global health problem. Emerging Infectious Diseases 4:442.

Gutiérrez-Bugallo, G., L. A. Piedra, M. Rodriguez, J. A. Bisset, R. Lourenço-de-Oliveira, S. C. Weaver, N. Vasilakis, and A. Vega-Rúa. 2019. Vector-borne transmission and evolution of Zika virus. Nature Ecology and Evolution 3:561-569.

Heske, E. J., J. H. Brown, and S. Mistry. 1994. Long-term experimental study of a Chihuahuan Desert rodent community: 13 years of competition. Ecology 75:438-445.

Hubbell, S. T. 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press, Princeton, N.J.

Hurlbert, A. H., and W. Jetz. 2007. Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. Proc. Natl. Acad. Sci. USA 104:13384-13389.

Ibarra-Cerdeña, C. N., L. Valiente-Banuet, V. Sánchez-Cordero, C. R. Stephens, and J. M. Ramsey. 2017. *Trypanosoma cruzi* reservoir—triatomine vector co-occurrence networks reveal meta-community effects by synanthropic mammals on geographic dispersal. PeerJ 5:e3152.

Idohou, R., A. Arino, A. Assogbadjo, R. G. Kakai, and B. Sinsin. 2015. Diversity of wild palms (Arecaceae) in the Republic of Benin: Finding the gaps in the national inventory combining field and digital accessible knowledge. Biodiversity Informatics 10:45-55.

IUCN. 2016. Spatial Data Download.[3] International Union for the Conservation of Nature, Gland.

Lemen, C., and P. W. Freeman. 1983. Quantification of competition among coexisting heteromyids in the Southwest. Southwestern Naturalist 28:41-46.

Lemen, C. A., and P. W. Freeman. 1986. Interference competition in a heteromyid community in the Great Basin of Nevada, USA. Oikos 46:390-396.

Lira-Noriega, A., J. Soberón, A. G. Navarro-Sigüenza, Y. Nakazawa, and A. T. Peterson. 2007. Scale-dependency of diversity components estimated from primary biodiversity data and distribution maps. Diversity and Distributions 13:185-195.

MacArthur, R. 1972. Geographical Ecology. Princeton University Press, Princeton, N.J.

Morueta-Holme, N., B. Blonder, B. Sandel, B. J. McGill, R. K. Peet, J. E. Ott, C. Violle, B. J. Enquist, P. M. Jørgensen, and J. C. Svenning. 2016. A network approach for inferring species associations from co-occurrence data. Ecography 39:1139-1150.

Otegui, J., A. H. Ariño, V. Chavan, and S. Gaiji. 2013a. On the dates of GBIF mobilised primary biodiversity records. Biodiversity Informatics 8:173-184.

Otegui, J., A. H. Ariño, M. A. Encinas, and F. Pando. 2013b. Assessing the primary data hosted by the Spanish node of the Global Biodiversity Information Facility (GBIF). PLoS ONE 8:e55144.

Pech-May, A., F. J. Escobedo-Ortegón, M. Berzunza-Cruz, and E. A. Rebollar-Téllez. 2010. Incrimination of four sandfly species previously unrecognized as vectors of *Leishmania* parasites in Mexico. Medical and Veterinary Entomology 24:150-161.

Pech-May, A., G. Peraza-Herrera, D. A. Moo-Llanes, J. Escobedo-Ortegón, M. Berzunza-Cruz, I. Becker-Fauser, A. C. Montes de Oca-Aguilar, and E. Rebollar-Tellez. 2016. Assessing the importance of four sandfly species (Diptera: Psychodidae) as vectors of *Leishmania mexicana* in Campeche, Mexico. Medical and Veterinary Entomology 30:310-320.

Pereira, H. M., S. Ferrier, M. Walters, G. N. Geller, R. H. G. Jongman, R. J. Scholes, M. W. Bruford, N. Brummitt, S. H. M. Butchart, and A. C. Cardoso. 2013. Essential biodiversity variables. Science 339:277-278.

Peres-Neto, P. R., J. D. Olden, and D. A. Jackson. 2001. Environmentally constrained null models: Site suitability as occupancy criterion. Oikos 93:110-120.

Peterson, A. T., A. G. Navarro-Sigüenza, and H. Benítez-Díaz. 1998. The need for continued scientific collecting: A geographic analysis of Mexican bird specimens. Ibis 140:288-294.

Peterson, A. T., M. Papeş, D. S. Carroll, H. Leirs, and K. M. Johnson. 2007. Mammal taxa constituting potential coevolved reservoirs of filoviruses. Journal of Mammalogy 88:1544-1554.

Peterson, A. T., and J. Soberón. 2018. Essential biodiversity variables are not global. Biodiversity and Conservation 27:1277-1288.

Rodríguez-Rojas, J. J., Á. Rodríguez-Moreno, M. Berzunza-Cruz, G. Gutiérrez-Granados, I. Becker, V. Sánchez-Cordero, C. R. Stephens, I. Fernández-Salas, and E. A. Rebollar-Téllez. 2017. Ecology of phlebotomine sandflies and putative reservoir hosts of leishmaniasis in a border area in northeastern Mexico: Implications for the risk of transmission of *Leishmania mexicana* in Mexico and the USA. Parasite 24:33.

Rojas-Soto, O. R., O. Alcantara-Ayala, and A. G. Navarro. 2003. Regionalization of the avifauna of the Baja California Peninsula, Mexico: A parsimony analysis of endemicity and distributional modelling approach. Journal of Biogeography 30:449-461.

Sánchez-García, L., M. Berzunza-Cruz, I. Becker-Fauser, and E. A. Rebollar-Téllez. 2010. Sand flies naturally infected by *Leishmania* (*L.*) *mexicana* in the peri-urban area of Chetumal city, Quintana Roo, México. Transactions of the Royal Society of Tropical Medicine and Hygiene 104:406-411.

Sánchez-Cordero, V., D. Stockwell, S. Sarkar, H. Liu, C. R. Stephens, and J. Giménez. 2008. Competitive interactions between felid species may limit the southern distribution of bobcats *Lynx rufus*. Ecography 31:757-764.

Soberón, J., R. Jiménez, J. Golubov, and P. Koleff. 2007. Assessing completeness of biodiversity databases at different spatial scales. Ecography 30:152-160.

Soberón, J. M., J. B. Llorente, and L. Oñate. 2000. The use of specimen-label databases for conservation

---

[3] http://www.iucnredlist.org/technical-documents/spatial-data

purposes: An example using Mexican papilionid and pierid butterflies. Biodiversity and Conservation 9:1441-1466.

Sousa-Baena, M. S., L. C. Garcia, and A. T. Peterson. 2014. Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. Diversity Distrib. 20:369-381.

Stein, B. R., and J. Wieczorek. 2004. Mammals of the world: MaNIS as an example of data integration in a distributed network environment. Biodiversity Informatics 1:14-22.

Stephens, C., V. Sánchez-Cordero, and C. González Salazar. 2017. Bayesian inference of ecological interactions from spatial data. Entropy 19:547.

Stephens, C. R., J. Giménez-Heau, C. González, C. N. Ibarra-Cerdeña, V. Sánchez-Cordero, and C. González-Salazar. 2009. Using biotic interaction networks for prediction in biodiversity and emerging diseases. PLoS ONE 4:e5725.

Stephens, C. R., C. González-Salazar, V. Sánchez-Cordero, I. Becker, E. Rebollar-Tellez, Á. Rodríguez-Moreno, M. Berzunza-Cruz, C. D. Balcells, G. Gutiérrez-Granados, and M. Hidalgo-Mihart. 2016. Can you judge a disease host by the company it keeps? Predicting disease hosts and their relative importance: a case study for leishmaniasis. PLoS Neglected Tropical Diseases 10:e0005004.

Stephens, C. R., R. Sierra-Alcocer, C. González-Salazar, J. M. Barrios, J. C. Salazar Carrillo, E. Robredo Ezquivelzeta, and E. del Callejo Canal. 2019. SPECIES: A platform for the exploration of ecological data. Ecology and Evolution 9:1638-1653.

Ulrich, W. 2004. Species co-occurrences and neutral models: Reassessing JM Diamond's assembly rules. Oikos 107:603-609.

Ulrich, W., W. Kryszewski, P. Sewerniak, R. Puchałka, G. Strona, and N. J. Gotelli. 2017. A comprehensive framework for the study of species co-occurrences, nestedness and turnover. Oikos 126:1607-1616.

Vicente-Santos, A., A. Moreira-Soto, C. Soto-Garita, L. G. Chaverri, A. Chaves, J. F. Drexler, J. A. Morales, A. Alfaro-Alarcón, B. Rodríguez-Herrera, and E. Corrales-Aguilar. 2017. Neotropical bats that co-habit with humans function as dead-end hosts for dengue virus. PLoS Neglected Tropical Diseases 11:e0005537.

Visser, H., and T. de Nijs. 2006. The Map Comparison Kit. Environmental Modeling and Software 21:346-358.

Woolhouse, M. E. J., and S. Gowtage-Sequeria. 2005. Host range and emerging and reemerging pathogens. Emerging Infectious Diseases 11:1842.

Yesson, C., P. W. Brewer, T. Sutton, N. Caithness, J. S. Pahwa, M. Burgess, W. A. Gray, R. J. White, A. C. Jones, F. A. Bisby, and A. Culham. 2007. How global Is the Global Biodiversity Information Facility? PLoS ONE 2:e1124.