The missing link: discerning true from false negatives when sampling species interaction networks

Michael D. Catchen [1,2] Timothée Poisot [3,2] Laura J. Pollock [1,2] Andrew Gonzalez [1,4]

[1] McGill University [2] Québec Centre for Biodiversity Sciences [3] Université de Montréal [4] Québec Centre for Biodiversity Science

Correspondance to:

Michael D. Catchen — `michael.catchen@mcgill.ca`

Last revision: *January 14, 2023*

Abstract: Ecosystems are composed of networks of interacting species. These interactions allow communities of species to persist through time through both neutral and adaptive processes. Despite their importance, a robust understanding of (and ability to predict and forecast) interactions among species remains elusive. This knowledge-gap is largely driven by a shortfall of data—although species occurrence data has rapidly increased in the last decade, species interaction data has not kept pace, largely due to the intrinsic difficulty and effort required to sample interactions. This means there are many interactions between species that occur in nature, but we do not think this interaction occurs because we have no record of it. These so-called "false-negatives" bias data and hinder inference about the structure and dynamics of interaction networks. Here, we demonstrate the realized rate of false-negatives in data can be quite high, even in thoroughly sampled systems, due to the intrinsic variation in abundances across species in a community. We illustrate how a null model of occurrence detection can be used to estimate the false-negative rate in a given dataset. We also show how to directly incorporate uncertainty due to observation error into model-based predictions of interaction probabilities between species. One hypothesis is that interactions between "rare" species are themselves rare because these species are less likely to encounter one-another than species of higher relative abundance, and that this can (in part) explain the common pattern of nestedness in bipartite interaction networks. However, we demonstrate that across several datasets of spatial or temporally replicated networks, there are positive associations between species co-occurrence and interactions, which suggests these interactions among "rare" species actually exist but simply are not observed. Finally, we assess how false negatives influence various models of network prediction, and recommend directly accounting for observation error in predictive models. We conclude by discussing how the understanding of false-negatives can inform how we design monitoring schemes for species interactions.

Introduction

Species interactions drive many processes in evolution and ecology. A better understanding of species interactions is an imperative to understand the evolution of life on Earth, to mitigate the impacts of anthropogenic change on biodiversity (Makiola *et al.* 2020), and for predicting zoonotic spillover of disease to prevent future pandemics (Becker *et al.* 2021). At the moment we lack sufficient data to meet these challenges (Poisot *et al.* 2021), largely because species interactions are hard to sample (Jordano 2016). Over the past few decades biodiversity data has become increasingly available through remotely collected data and adoption of open data practices (Kenall *et al.* 2014; Stephenson 2020). Still, interaction data remains relatively scarce because sampling typically requires human observation. This induces a constraint on the amount, spatial scale, and temporal frequency of resulting data that it is feasible to collect by humans. Many crowdsourced methods for biodiversity data aggregation (e.g. GBIF, eBird) still rely on automated identification of species, which does not easily generalize to interaction sampling. There is interest in using remote methods for interaction sampling, which primarily detect co-occurrence and derive properties like species avoidance from this data (Niedballa *et al.* 2019). However, co-occurrence itself is not necessarily indicative of an interaction (Blanchet *et al.* 2020). This is an example of semantic confusion around the word "interaction"—for example one might consider competition a type of species interaction, even though it is marked by a lack of co-occurrence between species, unlike other types of interactions, like predation or parasitism, which require both species to be together at the same place and time. Here we consider interaction in the latter sense, where two species have fitness consequences on one-another if (and only if) they are in the sample place at the same time. In addition, here we only consider direct (not higher-order) interactions.

We cannot feasibly observe all (or even most) of the interactions that occur in an ecosystem. This means we can be confident two species actually interact if we have a record of it (assuming they are correctly identified), but not at all confident that a pair of species *do not* interact if we have *no record* of those species observed together. In other words, it is difficult to distinguish *true-negatives* (two species never interact) from *false-negatives* (two species interact sometimes, but we do not have a record of this interaction). For a concrete example of a false-negative in a food web, see fig. 1. Because even the most highly sampled systems will still contain false-negatives, there is increasing interest in combining species-level data (e.g. traits, abundance, range, phylogenetic relatedness, etc.) to build models to predict interactions between species we haven't observed together before (Strydom *et al.* 2021). However, the noise of false-negatives could impact the efficacy of our predictive models and have practical consequences for answering questions about interactions (de Aguiar *et al.* 2019). This data constraint is amplified as the interaction data we have is geographically biased toward the usual suspects (Poisot *et al.* 2021). We therefore need a statistical approach to assessing these biases in the observation process and their consequences for our understanding of interaction networks.

The importance of *sampling effort* and its impact on resulting ecological data has produced a rich body of literature.The recorded number of species in a dataset or sample depends on the total number of observations (Walther *et al.* 1995; Willott 2001), as do estimates of population abundance (Griffiths 1998). This relationship between sampling effort and spatial coverage and species detectability has motivated more quantitatively robust approaches to account for error in sampling data in many contexts: to determine if a given species is extinct (Boakes *et al.* 2015), to determine sampling design (Moore & McCarthy 2016), and to measure species richness across large scales (Carlson *et al.* 2020). In the context of interactions, an initial concern was the compounding effects of limited sampling effort combined with the amalgamation of data (across both study sites, time of year, and taxonomic scales) could lead any empirical set of observations to inadequately reflect the reality of how species interact (Paine 1988) or the structure of the network as a whole (Martinez *et al.* 1999; McLeod *et al.* 2021). Martinez *et al.* (1999) showed that in a plant-endophyte trophic network, network connectance is robust to sampling effort, but this was done in the context of a system for which observation of 62,000 total interactions derived from 164,000 plant-stems was feasible. In some systems (e.g. megafauna food-webs) this many observations is either impractical or infeasible due to the absolute abundance of the species in question.

The intrinsic properties of ecological communities create several challenges for sampling: first, species are not observed with equal probability—we are much more likely to observe a species of high abundance than one of very low abundance (Poisot *et al.* 2015). Canard *et al.* (2012) presents a null model of food-web structure where species encounter one-another in proportion to each species' relative-abundance. This assumes that there are no associations in species co-occurrence due to an interaction (perhaps because this interaction is "important" for both species; Cazelles *et al.* (2016)), but in this paper we later show increasing strength of associations leads to increasing probability of false-negatives in interaction data, and that these positive associations are common in existing network data. Second, observed co-occurrence is often equated with meaningful interaction strength, but this is not necessarily the case (Blanchet *et al.*

2020)—a true "non-interaction" would require that neither of two species, regardless of whether they co-occur, ever exhibit any meaningful effect on the fitness of the other. So, although co-occurrence is not directly indicative of an interaction, it *is* a precondition for an interaction.

Here, we illustrate how our confidence that a pair of species never interacts highly depends on sampling effort. We suggest that surveys of species interactions can benefit from simulation modeling of the sampling process. We demonstrate how the realized false-negative-rate of interactions is related to the relative abundance of the species pool, and introduce a method to produce a null estimate of the false-negative-rate given total sampling effort (the total count of all interactions seen among all species-pairs) and a method for including uncertainty into model predictions of interaction probabilities to account for observation error. We then show that positive associations in co-occurrence data can increase the realized number of false-negatives, and demonstrate these positive associations are rampant in network datasets, and conclude by recommending that the simulation of sampling effort and species occurrence can and should be used to help design surveys of species interaction diversity (Moore & McCarthy 2016), and by advocating use of null models like those presented here as a tool for both guiding design of surveys of species interactions and for modeling detection error in predictive models.

Accounting for false-negatives in species interactions

In this section, we demonstate how difference in relative-abundance can lead to many false-negatives in interaction data. We also introduce a method for producing a null estimate of the false-negative-rate in datasets via simulation, and a method for incorporating uncertainty directly into predictions of species interactions to account for observation error.

How many observations of a non-interaction do we need to be confident it's a true negative?

We start with a naive model of interaction detection: we assume that every interacting pair of species is incorrectly observed as not-interacting with an independent and fixed probability, which we denote p_{fn} and subsequently refer to as the False-Negative-Rate (FNR). If we observe the same species not-interacting N times, then the probability of a true-negative (denoted p_{tn}) is given by $p_{tn} = 1 - (p_{fn})^N$. This relation

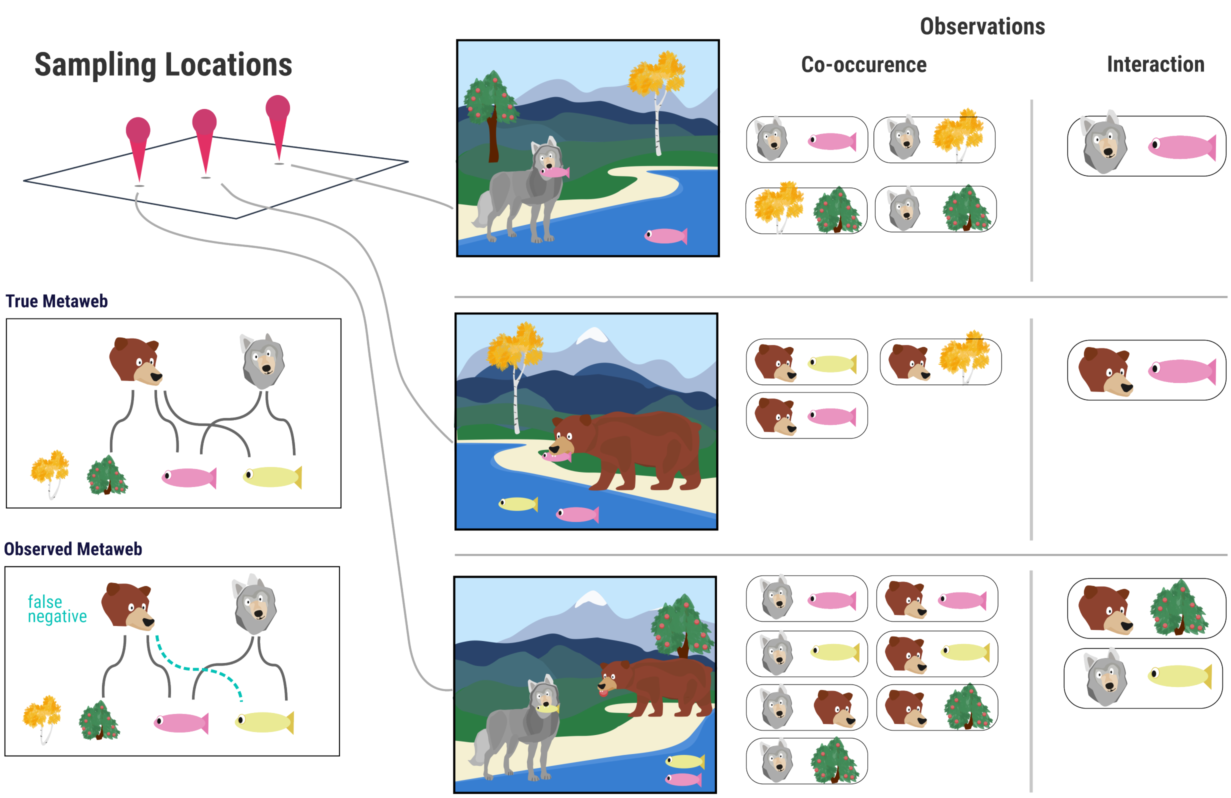

Figure 1 This conceptual example considers sample of the trophic community of bears, wolves, salmon (pink fish), pike (yellow fish), berry trees, and aspen trees. The true metaweb (all realized interactions across the entire spatial extent) is shown on the left. In the center is what a hypothetical ecologist samples at each site. Notice that although bears are observed co-occurring with both salmon and pike, there was never a direct observation of bears eating pike, even though they actually do. Therefore, this interaction between bears and pike is

(the probability-mass-function of geometric distribution, a special case of the negative-binomial distribution) is shown in fig. 2(A) for varying values of p_{fn} and illustrates a fundamental link between our ability to reliably say an interaction doesn't exist—p_{tn}—and the number of times N we have observed a given species. In addition, note that there is no non-zero p_{fn} for which we can ever *prove* that an interaction does not exist—no matter how many observations of non-interactions N we have, $p_{tn} < 1$.

From fig. 2(A) it is clear that the more often we see two species co-occurring, but *not interacting*, the more likely the interaction is a true-negative. This has several practical consequences: first it means negatives taken outside the overlap of the range of each species aren't informative because co-occurrence was not possible, and therefore neither was an interaction. Second, we can use this relation to compute the expected number of total observations needed to obtain a "goal" number of observations of a particular pair of species (fig. 2(B)). As an example, if we hypothesize that A and B do not interact, and we want to see species A and B both co-occurring and *not interacting* 10 times to be confident this is a true negative, then we need an expected 1000 observations of all species if the relative abundances of A and B are both 0.1.

Because the true FNR is latent, we can never actually be sure what the actual number of false-negatives in our data—however, we can use simulation to estimate the FNR for datasets of a given size using neutral models of observation. If some of the "worst-case" FNRs presented in fig. 2(A) seem unrealistically high, considering that species are observed in proportion to their relative abundance. In the next section we demonstrate that the distribution of abundance in ecosystems can lead to very high realized values of FNR (p_{fn}) simply as an artifact of sampling effort.

False-negatives as a product of relative abundance

We now show that the realized FNR changes drastically with sampling effort due to the intrinsic variation of the abundance of individuals of each species within a community. We do this by simulating the process of observation of species interactions, applied both to 243 empirical food webs from the Mangal database (Banville *et al.* 2021) and random food-webs generated using the niche model, a simple generative model of food-web structure that accounts for allometric scaling (Williams & Martinez 2000). Our neutral model of observation assumes each observed species is drawn in proportion to each species' abundance at that place and time. The abundance distribution of a community can be reasonably-well described by a log-normal distribution (Volkov *et al.* 2003). In addition to the log-normal distribution, we also tested the

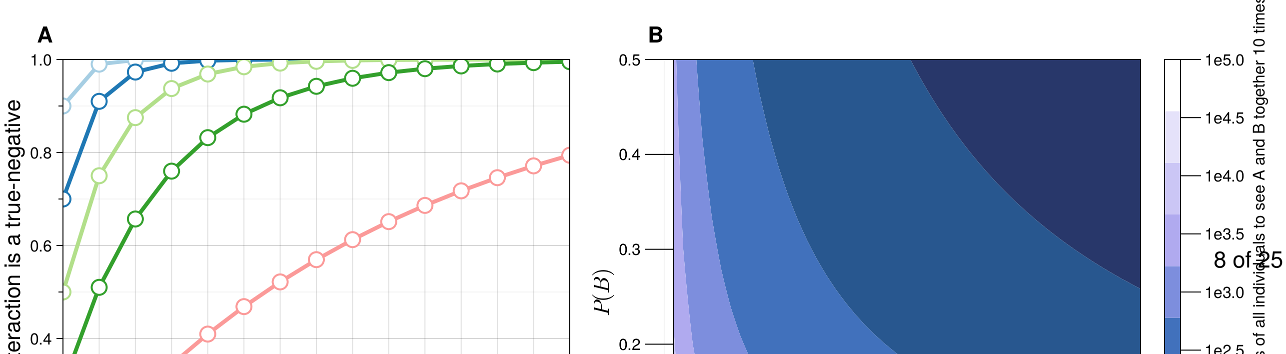

Figure 2 **(A)** Th probability that a observed interaction is a tru negative (y-axis) given hov many time it has bee sampled as a nor interaction (x-axis). Each colo reflects different value of p_{f} the false negative- rate (FNR)—thi is effectivel the cdf of th geometric distributior **(B)** Th expected number of tota observation needed (colors) to observ 10 co occurrences between a specie with relativ abundance $P(A)$ (x-axis and a secon species wit relative abundance $P(Y)$. **(C)** false- negative- rate (y-axis as a functio of tota sampling

case where the abundance distribution is derived from power-law scaling $Z^{(log(T_i)-1)}$ where T_i is the trophic level of species i and Z is a scaling coefficient (Savage *et al.* 2004), which yields the same qualitative behavior. The practical consequence of abundance distributions spanning many orders of magnitude of abundance is that observing two "rare" species interacting requires two low probability events: observing two rare species *at the same time*.

To simulate the process of observation, for an ecological network M with S species, we sample abundances for each species from a standard-log-normal distribution. For each true interaction in the adjacency matrix M (i.e. $M_{ij} = 1$) we estimate the probability of observing both species i and j at a given place and time by simulating n observations of all individuals of any a species, where the species of the individual observed at the $\{1, 2, \dots, n\}$-th observation is drawn from the generated log-normal distribution of abundances. For each pair of species (i, j), if both i and j are observed within the n-observations, the interaction is tallied as a true positive if $M_{ij} = 1$. If only one of i or j are observed—but not both—in these n observations, but $M_{ij} = 1$, this is counted as a false-negative, and a true-negative otherwise. For each pair of species (i, j), if both i and j are observed within the n-observations, the interaction is tallied as a true positive if $M_{ij} = 1$. If only one of i or j are observed—but not both—in these n observations, but $M_{ij} = 1$, this is counted as a false-negative, and a true-negative otherwise ($M_{ij} = 0$). This process is illustrated conceptually in fig. 3(A).

In fig. 2(C) we see this model of observation applied to niche model networks across varying levels of species richness, and in fig. 2(D) the observation model applied to Mangal food webs. For all niche model simulations in this manuscript, for a given number of species S the number of interactions is drawn from the flexible-links model fit to Mangal data (MacDonald *et al.* 2020), effectively drawing the number of interactions L for a random niche model food-web as

$$L \sim \text{BetaBinomial}(S^2 - S + 1, \mu\phi, 1 - \mu\phi)$$

where the maximum *a posteriori* (MAP) estimate of (μ, ϕ) applied to Mangal data from (MacDonald *et al.* 2020) is $(\mu = 0.086, \phi = 24.3)$. All simulations were done with 500 independent replicates of unique niche model networks per unique number of observations n. All analyses presented here are done in Julia v1.8 (Bezanson *et al.* 2015) using both EcologicalNetworks.jl v0.5 and Mangal.jl v0.4 (Banville *et al.* 2021) and are hosted on Github). Note that the empirical data, for the reasons described above, very likely already contains many false-negatives, we'll revisit this issue in the final section.

From fig. 2(C) it is evident that the number of species considered in a study is inseparable from the false-negative-rate in that study, and this effect should be taken into account when designing samples of ecological networks in the future. We see a similar qualitative pattern in fig. 2(D) where the FNR drops off quickly as a function of observation effort, mediated by total richness. The practical consequence of the bottom row of fig. 2 is whether the total number of observations of all species (the x-axis) for the threshold FNR we deem acceptable (the y-axis) is feasible. This raises two points: first, empirical data on interactions are subject to the practical limitations of funding and human-work hours, and therefore existing data tend to fall on the order of hundreds or thousands observations of individuals per site. Clear aggregation of data on sampling effort has proven difficult to find and a meta-analysis of network data and sampling effort seems both pertinent and necessary, in addition to the effects of aggregation of interactions across taxonomic scales (Gauzens *et al.* 2013; Giacomuzzo & Jordán 2021). This inherent limitation on in-situ sampling means we should optimize where we sample across space so that for a given number of samples, we obtain the maximum information possible. Second, what is meant by "acceptable" FNR? This raises the question: does a shifting FNR lead to rapid transitions in our ability inference and predictions about the structure and dynamics of networks, or does it produce a roughly linear decay in model efficacy? We explore this in the next section.

We conclude this section by advocating for the use of neutral models similar to above to generate expectations about the number of false-negatives in a data set of a given size. This could prove fruitful both for designing surveys of interactions but also because we may want to incorporate models of imperfect detection error into predictive interactions models, as Joseph (2020) does for species occurrence modeling. Additionally, we emphasize that one must consider the context for sampling—is the goal to detect a particular species (as in fig. 2(C)), or to get a representative sample of interactions across the species pool? These arguments are well-considered when sampling individual species (Willott 2001), but have not yet been adopted for designing samples of communities.

Including observation error in interaction predictions

Here we show how to incorporate uncertainty into model predictions of interaction probability to account for imperfect observation (both false-negatives and false-positives). Models for interaction prediction typically yield a probability of interaction between each pair of species, p_{ij}. When these are considered with uncertainty, it is usually model-uncertainty, e.g. the variance in the interaction probability prediction across several cross-validation folds, where the data is split into training and test sets several times. The method we introduce adjusts the value of a model's predictions to produce a distribution of interaction probabilities, which are adjusted by a given false-negative-rate p_{fn} and false-positive-rate p_{fp} (outlined in figure fig. 3). We describe first how to sample from this distribution of adjusted interaction probabilities via simulation, and show that this distribution can be well-approximated analytically.

We then consider the output prediction from an arbitrary prediction model, which is the probability p_{ij} that two species i and j interact. To get an estimate of p_{ij} that accounts for observation error, we resample the probability of each interaction p_{ij} by simulating a set of several 'particles,' where each particle is a realization of an interaction occurring (either true or false with probabilities p_{ij} and $1 - p_{ij}$ respectively) and then being correctly observed with probabilities given by p_{fp} and p_{fn} to yield a single boolean outcome for each particle. ("Resampling" within fig. 3 (B)). Over many samples of particles, the resulting frequency of 'true' outcomes is a single resample of the interaction probability p^*_{ij}. Across several samples each of several particles, this forms a distribution of probabilities which are adjusted by the true and false negative rates.

There is also an analytic way to approximate this distribution using the normal approximation to binomial. As a reminder, as the total number of samples n from a binomial distribution with success probability p from approaches infinity, the sum of total successes across all samples approaches a normal distribution with mean np and variance $np(1-p)$. We can use this to correct the estimate p_{ij} based on the expected false-negative-rate p_{fn} and false-positive rate p_{fp} to obtain the limiting distribution as the number of resamples approaches infinity for the resampled p^*_{ij} for a given number of particles n_p. We do this by first adjusting for the rates of observation error to get the mean resampled probability, $\mathbb{E}[p^*_{ij}]$, as

$$\mathbb{E}[p^*_{ij}] = p_{ij}(1 - p_{fp}) + (1 - p_{ij})p_{fn}$$

which yields the normal approximation

$$\sum_{i=1}^{n_p} p^*_{ij} \sim \mathcal{N}\left(n_p \cdot \mathbb{E}[p^*_{ij}], \sqrt{n_p \mathbb{E}[p^*_{ij}](1 - \mathbb{E}[p^*_{ij}])}\right)$$

which then can be converted back to a distribution of frequency of successes to yield the final approximation

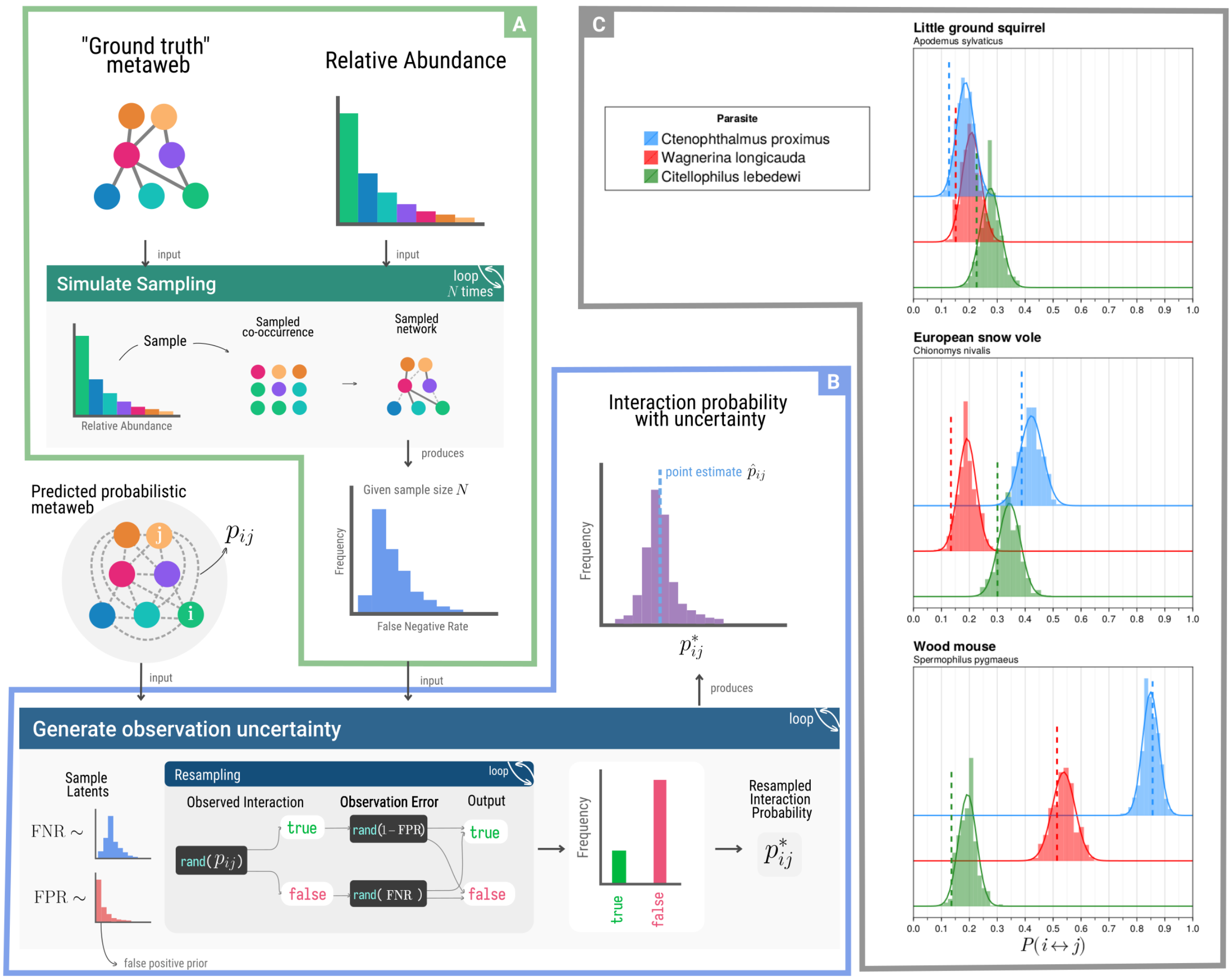

Figure 3. (A) The process for estimating the false negative-rate (FNR) for an interaction dataset consisting of N total observed interactions. (B) The method for resampling interaction probability based on estimates of false negative and false positive rates. (C) The method for interaction probability resampling applied to three mammals and three parasites from the Hadfield *et al.* (2014) dataset.

$$p_{ij}^* \sim \mathcal{N}\left(\mathbb{E}[p_{ij}^*], \sqrt{\frac{\mathbb{E}[p_{ij}^*](1-\mathbb{E}[p_{ij}^*])}{n_p}}\right) \quad (1)$$

We can then further truncate to remain on the interval $(0, 1)$ (as the output is a probability, although in practice often the probability mass outside $(0, 1)$ is exteremly low. As an example case study, we use a boosted-regression-tree to predict interactions in a host-parasite network (Hadfield *et al.* 2014) (with features derived in the same manner as Strydom *et al.* (2021) derives features on this data) to produce a set of interaction predictions. We then applied this method to a set of a few resampled interaction probabilities between mammals and parasite species shown in figure fig. 3(C).

Why is this useful? For one, this analytic method avoids the extra computation required by simulating samples from this distribution directly. Further, it enables continuous exanination of the number of particles n_p as a uncertainty width. The natural analogue for the number of particles sampled is the number of observations of co-occurrence for a given pair of species—the fewer the particles, the higher the variance of the resulting approximation. The normal approximation is undefined for 0 particles (i.e. 0 observations co-occurrence), although as n_p approaches 0 the approxated normal (once truncated) approaches a uniform distribution on the interval $(0, 1)$, the maximum entropy distribution where we have no information about the possibility of an interaction.

This also has implications for what we mean by 'uncertainty' in interaction predictions. A model's prediction can be 'uncertain' in two different ways: (1) the model's predictions may have high variance, or (2) the model's predictions may be centered around a probability of interaction of 0.5, where we are the most unsure about whether this interaction exists. Improving the incorporation of different forms of uncertainty in probabilistic interaction predictions seems a necessary next step toward understanding what pairs of species we know the least about, in order to prioritize sampling to provide the most new information possible.

Positive associations in co-occurrence increase the false-negative-rate

The model above doesn't consider the possibility that there are positive or negative associations which shift the probability of species cooccurrence away from what is expected based on their relative abundances due to their interaction (Cazelles *et al.* 2016). However, here we demonstrate that the probability of having a false-negative can be higher if there is some positive association in the occurrence of species A and B. If we denote the probability that we observe the co-occurrence of two species A and B as $P(AB)$ and if there is no association between the marginal probabilities of observing A and observing B, denoted $P(A)$ and $P(B)$ respectively, then the probability of observing their co-occurrence is the product of the marginal probabilities for each species, $P(AB) = P(A)P(B)$. In the other case where there is some positive strength of association between observing both A and B because this interaction is "important" for each species, then the probability of observation both A and B, $P(AB)$, is greater than $P(A)P(B)$ as $P(A)$ and $P(B)$ are not independent and instead are positively correlated, i.e. $P(AB) > P(A)P(B)$. In this case, the probability of observing a single false-negative in our naive model from fig. 2(A) is $p_{fn} = 1 - P(AB)$, which due to the above inequality implies $p_{fn} > 1 - P(A)P(B)$. This indicates an increasingly greater probability of a false negative as the strength of association gets stronger, $P(AB) \rightarrow P(AB) \gg P(A)P(B)$. However, this still does not consider variation in species abundance in space and time (Poisot *et al.* 2015). If positive or negative associations between species structure variation in the distribution of $P(AB)$ across space/time, then the spatial/temporal biases induced by data collection would further impact the realized false-negative-rate, as the probability of false negative would not be constant for each pair of species across sites.

To test for these positive associations in data we scoured Mangal for datasets with many spatial or temporal replicates of the same system, which led the the resulting seven datasets set in figure fig. 4. For each dataset, we compute the marginal probability $P(A)$ of occurrence of each species A across all networks in the dataset. For each pair of interacting species A and B, we then compute and compare the probability of co-occurrence if each species occurs independently, $P(A)P(B)$, to the empirical joint probability of co-occurrence, $P(AB)$. Following our analysis above, if $P(AB)$ is greater than $P(A)P(B)$, then we expect our neutral estimates of the FNR above to underestimate the realized FNR. In fig. 4, we see the difference between $P(AB)$ and $P(A)P(B)$ for the seven suitable datasets with enough spatio-temporal replicates and a shared taxonomic backbone (meaning all individual networks use common species identifiers) found on Mangal to perform this analysis. Further details about each dataset are reported in tbl. 1.

In each of these datasets, the joint probability of co-occurrence $P(AB)$ is decisively greater than our expectation if species co-occur in proportion to their relative abundance $P(A)P(B)$. This suggests that there may not be as many "neutrally forbidden links" (Canard *et al.* 2012) as we might think, and that the reason we do not have records of interactions between rare species is probably due to observation error. This has serious ramifications for the widely observed property of nestedness seen in bipartite networks

(Bascompte & Jordano 2007)—perhaps the reason we have lots of observations between generalists is because they are more abundant, and this is particularly relevant as we have strong evidence that generalism drives abundance (Song *et al.* 2022a), not vice-versa.

Table 1: The datasets used in the above analysis (Fig 2). The table reports the type of each dataset, the total number of networks in each dataset (N), the total species richness in each dataset (S), the connectance of each metaweb (all interactions across the entire spatial-temporal extent) (C), the mean species richness across each local network $\bar{S}$, the mean connectance of each local network $\bar{C}$, the mean β-diversity among overlapping species across all pairs of network species ($\bar{\beta}_{OS}$), and the mean β-diversity among all species in the metaweb ($\bar{\beta}_{WN}$). Both metrics are computed using KGL β-diversity (Koleff *et al.* 2003)

Network	Type	N	S	C	$\bar{S}$	$\bar{C}$	$\bar{\beta}_{OS}$	$\bar{\beta}_{WN}$
Kopelke *et al.* (2017)	Food Web	100	98	0.037	7.87	0.142	1.383	1.972
Thompson & Townsend (2000)	Food Web	18	566	0.014	80.67	0.049	1.617	1.594
Havens (1992)	Food Web	50	188	0.065	33.58	0.099	1.468	1.881
Ponisio *et al.* (2017)	Pollinator	100	226	0.079	23.0	0.056	1.436	1.870
Hadfield *et al.* (2014)	Host-Parasite	51	327	0.085	32.71	0.337	1.477	1.952
Closs & Lake (1994)	Food Web	12	61	0.14	29.09	0.080	1.736	1.864
CaraDonna *et al.* (2017)	Pollinator	86	122	0.18	21.42	0.312	1.527	1.907

The impact of false-negatives on network properties and prediction

Here, we assess the effect of false-negatives on our ability to make predictions about interactions, as well as their effect on network structure. The prevalence of false-negatives in data is the catalyst for interaction prediction in the first place, and as a result methods have been proposed to counteract this bias (Stock *et al.* 2017; Poisot *et al.* 2022). However, it is feasible that the FNR in a given dataset is so high that it could induce too much noise for an interaction prediction model to detect the signal of possible interaction between species.

To test this we use the dataset from Hadfield *et al.* (2014) that describes host-parasite interaction networks sampled across 51 sites, and the same method as Strydom *et al.* (2021) to extract latent features for each species in this dataset based on applying PCA to the co-occurrence matrix. We then predict a metaweb (equivalent to predicting true or false for an interaction between each species pair, effectively a binary classification problem) from these species-level features using four candidate models for binary

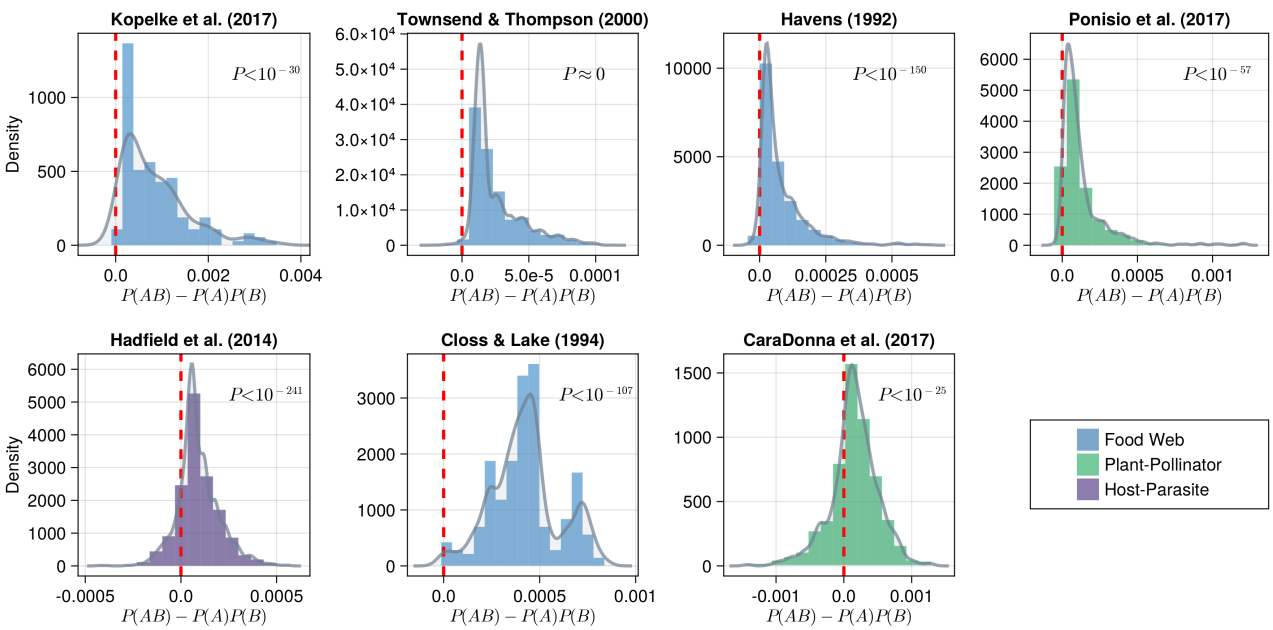

Figure 4 The difference between joint-probability of co-occurrence ($P(AB)$) and expected probability of co-occurrence under independence ($P(A)P(B)$) for interacting species for each dataset. The red dashed line indicates 0 (no association). Each histogram represents a density, meaning the area of the entire curve sums to 1. The continuous density estimate (computed using local smoothing) is shown in grey. The p-value of each plot is the result of a one-sided t-test comparing the mean of each distribution to 0.

classification—three often used machine-learning (ML) methods (Boosted Regression Tree (BRT), Random Forest (RF), Decision Tree (DT)), and one naive model from classic statistics (Logistic Regression (LR)). Each of the ML models are bootstrap aggregated (or bagged) with 100 replicates each. We partition the data into 80-20 training-test split, and then seed the training data with false negatives at varying rates, but crucially do nothing to the test data. We fit all of these models using MLJ.jl, a high-level Julia framework for a wide-variety of ML models (Blaom *et al.* 2020). We evaluate the efficacy of these models using two common measures of binary classifier performance: the area under the receiver-operator curve (ROC-AUC) and the area under the precision-recall curve (PR-AUC), for more details see Poisot (2022). Here, PR-AUC is slightly more relevant as it is a better indicator of prediction of false-negatives. The results of these simulations are shown in fig. 5(A&B).

One interesting result seen in fig. 5(A&B) is that the ROC-AUC value does not approach random in the same way the PR-AUC curve does as we increase the added FNR. The reason for this is that ROC-AUC is fundamentally not as useful a metric in assessing predictive capacity as PR-AUC. As we keep adding more false-negatives, the network eventually becomes a zeros matrix, and these models can still learn to predict "no-interaction" for all possible species pairs, which does far better than random guessing (ROC-AUC = 0.5) in terms of the false positive rate (one of the components of ROC-AUC). This highlights a more broad issue of label class imbalance, meaning there are far more non-interactions than interactions in data. A full treatment of the importance of class-balance is outside the scope of this paper, but is explored in-depth in Poisot (2022).

Although these ML models are surprisingly performant at link prediction given their simplicity, there have been several major developments in applying deep-learning methods to many tasks in network inference and prediction—namely graph-representation learning (GRL, Khoshraftar & An (2022)) and graph convolutional networks (Zhang *et al.* 2019). At this time, these advances can not yet be applied to ecological networks because they require far more data than we currently have. We already have lots of features that could be used as inputs into these models (i.e. species level data about occurrence, genomes, abundance, etc.), but our network datasets barely get into the hundreds of local networks sampled across space and time (tbl. 1). Once we start to get into the thousands, these models will become more useful, but this can only be done with systematic monitoring of interactions. This again highlights the need to optimize our sampling effort to maximize the amount of information contained in our data given the expense of sampling interactions.

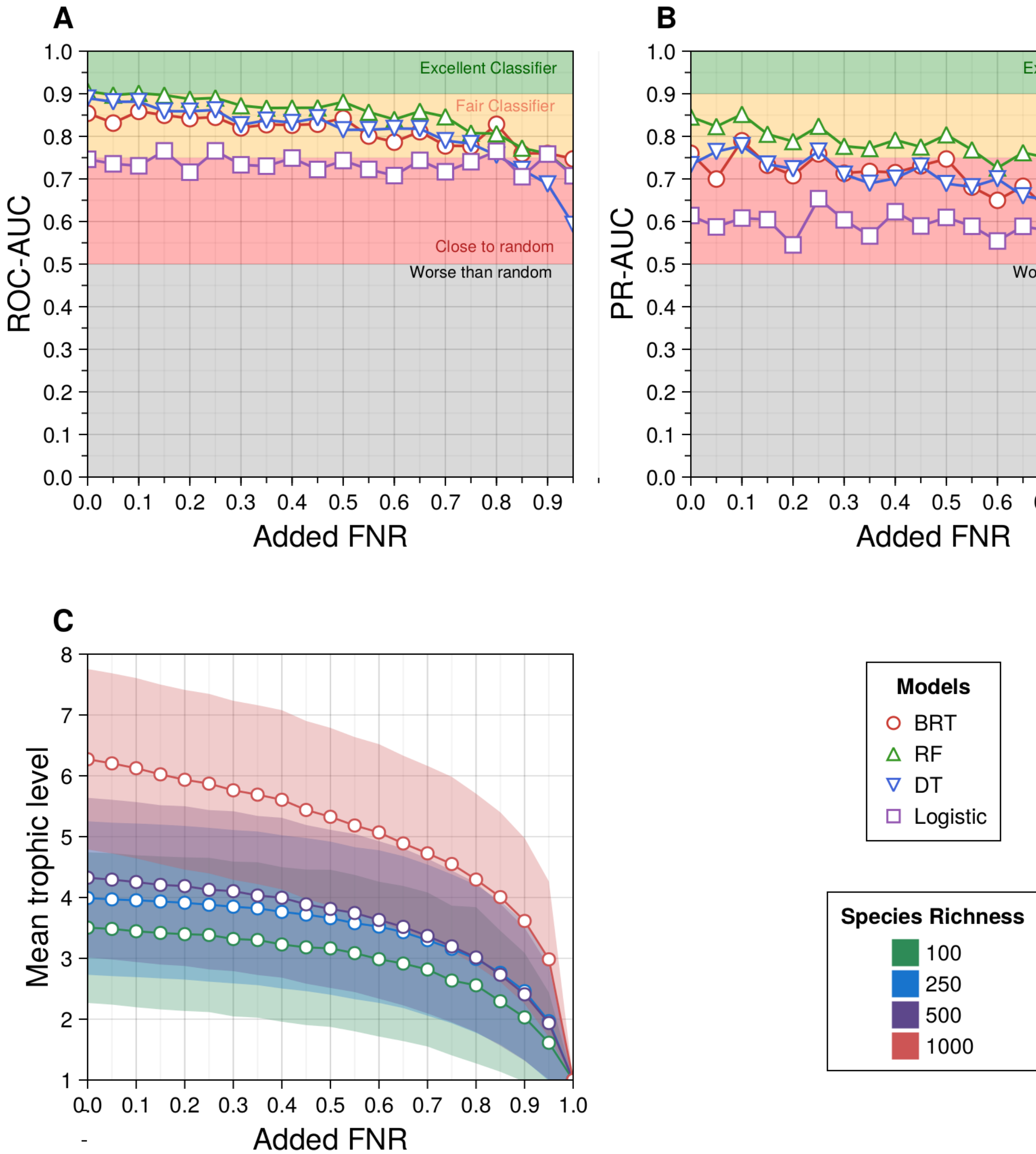

Figure 5 **(A)** Th area-under the receiver operator curve (ROC-AUC and **(B** The area under th precision-recall curv (PR-AUC; right) for eac different predictive model (colors/sha across spectrum of th proportion of adde false-negatives (x-axis). **(C** The mea trophic-leve of all specie in a networ generated with th niche mode across different species richnesses (colors). Fo each valu of the FNF the mea trophic level wa computed across 5 replicates. The shade region for eac line is on standard-deviation across thos

We also consider how the FNR affects network properties. In fig. 5(C) we see the mean trophic level across networks simulated using the niche model (as above), across a spectrum of FNR values. In addition to the clear dependence on richness, we see that mean trophic level, despite varying widely between niche model simulations, tends to be relatively robust to false-negatives and does not deviate widely from the true value until very large FNRs, i.e. $p_{fn} > 0.7$. This is not entirely unsurprising. Removing links randomly from a food-web is effectively the inverse problem of the emergence of a giant component (more than half of the nodes are in a connected network) in random graphs (see Li *et al.* (2021) for a thorough review). The primary difference being that we are removing edges, not adding them, and thus we are witnessing the dissolution of a giant component, rather than the emergence of one. Further applications of percolation theory (Li *et al.* 2021) to the topology of sampled ecological networks could improve our understanding of how false-negatives impact the inferences about the structure and dynamics on these networks.

Discussion

Species interactions enable the persistence and functioning of ecosystems, but our understanding of interactions is limited due to the intrinsic difficulty of sampling them. Here we have provided a null model for the expected number of false-negatives in an interaction dataset. We demonstrated that we expect many false-negatives in species interaction datasets purely due to the intrinsic variation of abundances within a community. We also, for the first time to our knowledge, measured the strength of association between co-occurrence and interactions (Cazelles *et al.* 2016) across many empirical systems, and found that these positive associations are both very common, and showed algebraically that they increase the realized FNR. We have also shown that false-negatives could further impact our ability to both predict interactions and infer properties of the networks, which highlights the need for further research into methods for correcting this bias in existing data.

A better understanding of how false-negatives impact species interaction data is a practical necessity—both for inference of network structure and dynamics, but also for prediction of interactions by using species level information. False-negatives could pose a problem for many forms of inference in network ecology. For example, inferring the dynamic stability of a network could be prone to error if the observed network is not sampled “enough.” What exactly “enough” means is then specific to the application, and should be assessed via methods like those here when designing samples. Further, predictions about network rewiring (Thompson & Gonzalez 2017) due to range shifts in response to climate change could be error-prone without accounting for interactions that have not been observed but that still may become climatically infeasible. As is evident from fig. 2(A), we can never guarantee there are no false-negatives in data. In recent years, there has been interest toward explicitly accounting for false-negatives in models (Stock *et al.* 2017; Young *et al.* 2021), and a predictive approach to networks—rather than expecting our samples to fully capture all interactions (Strydom *et al.* 2021). As a result, better models for predicting interactions are needed for interaction networks. This includes explicitly accounting for observation error (Johnson & Larremore 2021)—certain classes of models have been used to reflect hidden states which account for detection error in occupancy modeling (Joseph 2020), and could be integrated in the predictive models of interactions in the future.

This work has several practical consequences for the design of surveys for species' interactions. Simulating the process of observation could be a powerful tool for estimating the sampling effort required by a study that takes relative abundance into account, and provides a null baseline for expected FNR. It is necessary to take the size of the species pool into account when deciding how many total samples is sufficient for an "acceptable" FNR (fig. 2(C & D)). Further the spatial and temporal turnover of interactions means any approach to sampling prioritization must be spatiotemporal. We demonstrated earlier that observed negatives outside of the range of both species aren't informative, and therefore using species distribution models could aid in this spatial prioritization of sampling sites.

We also should address the impact of false-negatives on the inference of process and causality in community ecology. We demonstrated that in model food webs, false-negatives do not impact the measure of total trophic levels until very high FNR (figure fig. 5(C)), although we cannot generalize this further to other properties. This has immediate practical concern for how we design what taxa to sample—does it matter if the sampled network is fully connected? It has been shown that the stability of subnetworks can be used to infer the stability of the metaweb paper beyond a threshold of samples (Song *et al.* 2022b). But does this extend to other network properties? And how can we be sure we are at the threshold at which we can be confident our sample characterizes the whole system? We suggest that modeling observation error as we have done here can address these questions and aid in the design of samples of species interactions. To try to survey to avoid all false-negatives is a fool's errand. Species ranges overlap to form mosaics, which themselves are often changing in time. Communities and networks don't end in space, and the interactions that connect species on the 'periphery' of a given network to species outside the spatial extent of a given sample will inevitably appear as false-negatives in practical samples. The goal should instead be to sample a system enough to have a statistically robust estimate of the current state and empirical change over time of an ecological community at a given spatial extent and temporal resolution, and to determine what the sampling effort required should be prior to sampling.

Our work highlights the need for a quantitatively robust approach to sampling design, both for interactions (Jordano 2016) and all other aspects of biodiversity (Carlson *et al.* 2020). As anthropogenic forces create rapid shifts in our planet's climate and biosphere, this is an imperative to maximize the amount of ecological information we get in our finite samples, and make our inferences and decisions based on this data as robust as possible. Where we choose to sample, and how often we choose to sample there, has strong impacts on the inferences we make from data. Incorporating a better understanding of sampling effort and bias to the design of biodiversity monitoring systems, and the inference and predictive models we apply to this data, is imperative in understanding how biodiversity is changing, and making forecasts that can guide conservation action.

Acknowledgements

AG & MDC acknowledge the support of the Liber Ero Chair for Biodiversity conservation and NSERC.

References

Banville, F., Vissault, S. & Poisot, T. (2021). Mangal.jl and EcologicalNetworks.jl: Two complementary packages for analyzing ecological networks in Julia. *Journal of Open Source Software*, 6, 2721.

Bascompte, J. & Jordano, P. (2007). Plant-Animal Mutualistic Networks: The Architecture of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 38, 567–593.

Becker, D.J., Albery, G.F., Sjodin, A.R., Poisot, T., Bergner, L.M., Dallas, T.A., *et al.* (2021). Optimizing predictive models to prioritize viral discovery in zoonotic reservoirs.

Bezanson, J., Edelman, A., Karpinski, S. & Shah, V.B. (2015). Julia: A Fresh Approach to Numerical Computing.

Blanchet, F.G., Cazelles, K. & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions. *Ecology Letters*, 23, 1050–1063.

Blaom, A.D., Kiraly, F., Lienart, T., Simillides, Y., Arenas, D. & Vollmer, S.J. (2020). MLJ: A Julia package for composable machine learning. *Journal of Open Source Software*, 5, 2704.

Boakes, E.H., Rout, T.M. & Collen, B. (2015). Inferring species extinction: The use of sighting records. *Methods in Ecology and Evolution*, 6, 678–687.

Canard, E., Mouquet, N., Marescot, L., Gaston, K.J., Gravel, D. & Mouillot, D. (2012). Emergence of Structural Patterns in Neutral Trophic Networks. *PLOS ONE*, 7, e38295.

CaraDonna, P.J., Petry, W.K., Brennan, R.M., Cunningham, J.L., Bronstein, J.L., Waser, N.M., *et al.* (2017). Interaction rewiring and the rapid turnover of plantpollinator networks. *Ecology Letters*, 20, 385–394.

Carlson, C.J., Dallas, T.A., Alexander, L.W., Phelan, A.L. & Phillips, A.J. (2020). What would it take to describe the global diversity of parasites? *Proceedings of the Royal Society B: Biological Sciences*, 287, 20201841.

Cazelles, K., Araújo, M.B., Mouquet, N. & Gravel, D. (2016). A theory for species co-occurrence in interaction networks. *Theoretical Ecology*, 9, 39–48.

Closs, G.P. & Lake, P.S. (1994). Spatial and Temporal Variation in the Structure of an Intermittent-Stream Food Web. *Ecological Monographs*, 64, 1–21.

de Aguiar, M.A.M., Newman, E.A., Pires, M.M., Yeakel, J.D., Boettiger, C., Burkle, L.A., *et al.* (2019). Revealing biases in the sampling of ecological interaction networks. *PeerJ*, 7, e7566.

Gauzens, B., Legendre, S., Lazzaro, X. & Lacroix, G. (2013). Food-web aggregation, methodological and functional issues. *Oikos*, 122, 1606–1615.

Giacomuzzo, E. & Jordán, F. (2021). Food web aggregation: Effects on key positions. *Oikos*, 130, 2170–2181.

Griffiths, D. (1998). Sampling effort, regression method, and the shape and slope of sizeabundance relations. *Journal of Animal Ecology*, 67, 795–804.

Hadfield, J.D., Krasnov, B.R., Poulin, R. & Nakagawa, S. (2014). A Tale of Two Phylogenies: Comparative Analyses of Ecological Interactions. *The American Naturalist*, 183, 174–187.

Havens, K. (1992). Scale and Structure in Natural Food Webs. *Science*, 257, 1107–1109.

Johnson, E.K. & Larremore, D.B. (2021). Bayesian estimation of population size and overlap from random subsamples.

Jordano, P. (2016). Sampling networks of ecological interactions. *Functional Ecology*, 30, 1883–1893.

Joseph, M.B. (2020). Neural hierarchical models of ecological populations. *Ecology Letters*, 23, 734–747.

Kenall, A., Harold, S. & Foote, C. (2014). An open future for ecological and evolutionary data? *BMC Evolutionary Biology*, 14, 66.

Khoshraftar, S. & An, A. (2022). A Survey on Graph Representation Learning Methods.

Koleff, P., Gaston, K.J. & Lennon, J.J. (2003). Measuring beta diversity for presenceabsence data. *Journal of Animal Ecology*, 72, 367–382.

Kopelke, J.-P., Nyman, T., Cazelles, K., Gravel, D., Vissault, S. & Roslin, T. (2017). Food-web structure of willow-galling sawflies and their natural enemies across Europe. *Ecology*, 98, 1730–1730.

Li, M., Liu, R.-R., Lü, L., Hu, M.-B., Xu, S. & Zhang, Y.-C. (2021). Percolation on complex networks: Theory and application. *Physics Reports*, Percolation on complex networks: Theory and application, 907, 1–68.

MacDonald, A.A.M., Banville, F. & Poisot, T. (2020). Revisiting the Links-Species Scaling Relationship in Food Webs. *Patterns*, 1.

Makiola, A., Compson, Z.G., Baird, D.J., Barnes, M.A., Boerlijst, S.P., Bouchez, A., *et al.* (2020). Key Questions for Next-Generation Biomonitoring. *Frontiers in Environmental Science*, 7.

Martinez, N.D., Hawkins, B.A., Dawah, H.A. & Feifarek, B.P. (1999). Effects of Sampling Effort on Characterization of Food-Web Structure. *Ecology*, 80, 1044–1055.

McLeod, A., Leroux, S.J., Gravel, D., Chu, C., Cirtwill, A.R., Fortin, M.-J., *et al.* (2021). Sampling and asymptotic network properties of spatial multi-trophic networks. *Oikos*, 130, 2250–2259.

Moore, A.L. & McCarthy, M.A. (2016). Optimizing ecological survey effort over space and time. *Methods in Ecology and Evolution*, 7, 891–899.

Niedballa, J., Wilting, A., Sollmann, R., Hofer, H. & Courtiol, A. (2019). Assessing analytical methods for detecting spatiotemporal interactions between species from camera trapping data. *Remote Sensing in*

Ecology and Conservation, 5, 272–285.

Paine, R.T. (1988). Road Maps of Interactions or Grist for Theoretical Development? *Ecology*, 69, 1648–1654.

Poisot, T. (2022). Guidelines for the prediction of species interactions through binary classification.

Poisot, T., Bergeron, G., Cazelles, K., Dallas, T., Gravel, D., MacDonald, A., *et al.* (2021). Global knowledge gaps in species interaction networks data. *Journal of Biogeography*, 48, 1552–1563.

Poisot, T., Ouellet, M.-A., Mollentze, N., Farrell, M.J., Becker, D.J., Brierly, L., *et al.* (2022). Network embedding unveils the hidden interactions in the mammalian virome.

Poisot, T., Stouffer, D.B. & Gravel, D. (2015). Beyond species: Why ecological interaction networks vary through space and time. *Oikos*, 124, 243–251.

Ponisio, L.C., Gaiarsa, M.P. & Kremen, C. (2017). Opportunistic attachment assembles plantpollinator networks. *Ecology Letters*, 20, 1261–1272.

Savage, V.M., Gillooly, J.F., Brown, J.H., West, G.B. & Charnov, E.L. (2004). Effects of Body Size and Temperature on Population Growth. *The American Naturalist*, 163, 429–441.

Song, C., Simmons, B.I., Fortin, M.-J. & Gonzalez, A. (2022a). Generalism drives abundance: A computational causal discovery approach. *PLOS Computational Biology*, 18, e1010302.

Song, C., Simmons, B.I., Fortin, M.-J., Gonzalez, A., Kaiser-Bunbury, C.N. & Saavedra, S. (2022b). Rapid monitoring for ecological persistence.

Stephenson, P. (2020). Technological advances in biodiversity monitoring: Applicability, opportunities and challenges. *Current Opinion in Environmental Sustainability*, Open issue 2020 part A: Technology Innovations and Environmental Sustainability in the Anthropocene, 45, 36–41.

Stock, M., Poisot, T., Waegeman, W. & De Baets, B. (2017). Linear filtering reveals false negatives in species interaction data. *Scientific Reports*, 7, 45908.

Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., *et al.* (2021). A roadmap towards predicting species interaction networks (across space and time). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376, 20210063.

Thompson, P.L. & Gonzalez, A. (2017). Dispersal governs the reorganization of ecological networks under environmental change. *Nature Ecology & Evolution*, 1, 1–8.

Thompson, R.M. & Townsend, C.R. (2000). Is resolution the solution?: The effect of taxonomic resolution on the calculated properties of three stream food webs. *Freshwater Biology*, 44, 413–422.

Volkov, I., Banavar, J.R., Hubbell, S.P. & Maritan, A. (2003). Neutral theory and relative species abundance in ecology. *Nature*, 424, 1035–1037.

Walther, B.A., Cotgreave, P., Price, R.D., Gregory, R.D. & Clayton, D.H. (1995). Sampling Effort and Parasite Species Richness. *Parasitology Today*, 11, 306–310.

Williams, R.J. & Martinez, N.D. (2000). Simple rules yield complex food webs. *Nature*, 404, 180–183.

Willott, S.j. (2001). Species accumulation curves and the measure of sampling effort. *Journal of Applied Ecology*, 38, 484–486.

Young, J.-G., Valdovinos, F.S. & Newman, M.E.J. (2021). Reconstruction of plantpollinator networks from observational data. *Nature Communications*, 12, 3911.

Zhang, S., Tong, H., Xu, J. & Maciejewski, R. (2019). Graph convolutional networks: A comprehensive review. *Computational Social Networks*, 6, 11.