A process-based framework for quantifying sampling completeness of species interaction networks

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

Corresponding Author: michael.catchen@umontreal.ca

[1] McGill University, Department of Biology, Québec, Canada

[2] Université de Montréal, Québec, Canada

[3] Québec Centre for Biodiversity Science, Québec, Canada

Abstract

Species interaction networks underpin ecosystem function and persistence, but their study is hindered by lack of empirical knowledge about interactions. Many interactions occur in nature that have not been observed or recorded. We develop a process-based framework for simulating species interaction accumulation curves that accounts for spatiotemporal variation species interaction networks. We do this by explicitly linking interaction realization and detection rates to species abundance. We implement this framework in a software package, SpeciesInteractionSamplers.jl, which enables researchers to evaluate monitoring strategies and assess the completeness of empirical interaction datasets under biologically realistic assumptions. Together, this framework and software provide a foundation for quantifying the Eltonian shortfall, and for improving the design of interaction monitoring networks.

1 Introduction

Species interactions — like predation, pollination, and parasitism — govern the flow of biomass and energy through ecosystems, ultimately enabling persistence of biodiversity and ecosystem functioning. Despite their importance for ecosystem processes, species interactions are a major knowledge gap in knowledge of life on Earth; the so-called "Eltonian shortfall" (Hortal et al. 2015). Better monitoring of species interactions is an imperative to understand and predict the impacts of anthropogenic change on biodiversity (Makiola et al. 2020), ecosystem function (Bartley et al. 2019), and to anticipate disease spillover to prevent future pandemics (Carlson et al. 2022, Becker et al.

2022). We lack sufficient data to meet these challenges (Poisot et al. 2021) because species interactions are difficult to sample (Jordano 2016a; b). Many, if not most, species interactions that occur in nature have not yet been detected and recorded (Cirtwill et al. 2019). We can be confident two species can feasibly interact if we have a record of it (assuming each species was correctly identified), but unsure that two species could not feasibly interact if we have never observed it, let alone if we have few (or zero) records of these species co-occurring. Consequently, empirically constructed interaction networks contain many false negatives, where two species interact, but we lack any record of it.

Recent work has aimed to address this data gap by using species-level information (e.g. traits, abundance, range, phylogenies, etc.) to predict interaction feasibility (Terry and Lewis 2020, Caron et al. 2022), and for spatially and temporally explicit prediction of species interaction networks (Strydom et al. 2021). One challenge here is that we don't have a good idea how many feasible interactions are "missing" from these networks, as species interactions vary in how often they are realized and their intrinsic detectability (Cirtwill et al. 2019). Here, we develop a spatiotemporally explicit process-based framework for modeling stochastic interaction realization and imperfect detection, enabling estimation of the sampling completeness of interaction networks grounded in ecological mechanisms.

Imperfect detection is a pervasive feature of ecological sampling and has motivated a long history of statistical approaches to account for it, e.g. rarefaction for species richness estimation (Sanders 1968, Simberloff 1972). Explicitly accounting for imperfect detection has become common in modeling occupancy (MacKenzie et al. 2002, Royle et al. 2005, Royle 2006), abundance (Royle 2004), ranges (Guillera-Arroita 2017), and species richness (Dorazio and Royle 2005) with extensions to joint modeling of co-occurrence (Pollock et al. 2014, Devarajan et al. 2020). Similarly, there has been considerable work on sampling properties of interaction networks, primarily food-webs (Martinez et al. 1999, McLeod et al. 2021). This work has motivated the use of rarefaction methods from species richness estimation (Chacoff et al. 2012) to evaluate the completeness of network sampling, quantifying the effect of underlying network topology on measurement bias (Aguiar et al. 2019), consideration of the impacts of spatial and taxonomic biases (Poisot et al. 2021), and by excluding biologically infeasible interactions from consideration (Jordano 2016a). Cirtwill et al. (2019) introduced a framework describing the construction of empirical interaction networks accounting for variability in realization and detection rates.

We extend the framework from Cirtwill et al. (2019) by developing a process-based model of interaction realization and detection that explicitly accounts for spatiotemporal variation in species occurrence and abundance, which intrinsically structures interactions across space and time (Poisot et al. 2015). This framework links realization and detection rates to species abundance, moving beyond rarefaction-based estimates of sampling completeness by linking interaction accumulation to ecological mechanisms. We implement this framework in an open source software package,

SpeciesInteractionSamplers.jl. This approach allows researchers to both (1) estimate sampling completeness under biologically realistic assumptions, and (2) evaluate the sampling design of interaction monitoring programs.

2 A Process-Based Framework for Spatiotemporal Species Interaction Realization and Detection

Quantifying the sampling completeness of species interaction networks is essential to assess the limitations of inference made from interaction data, and to make grounded predictions of interactions. To enable robust estimates of sampling completeness, we need to root this methodology in the ecological processes that enable species interactions in order to ground these estimates in ecological reality. We introduce a process-based model of the realization and detection of species interactions that explicitly accounts for spatiotemporal variation in species presence and abundance. Cirtwill et al. (2019) models how empirical species interaction networks are constructed by having all pairwise species combinations move through a sequence of filters from all pairs of species to those that are detected during data collection: (1) feasibility (an interaction is biologically possible based on the traits of each species (Banville et al. 2025)), (2) realization (a feasible interaction actually happens), and (3) detection (the interaction is observed and recorded by an observer).

We extend this framework in several ways: (1) By introducing a new filter, potential, which disentangles co-occurrence from realization to account for spatiotemporal variation in abundance and occurrence. Explicitly accounting for this variation is essential to account for the demographic and biogeographic processes that structure interactions across space. The species interactions that are realized at different locations and times is inherently structured by the distribution and abundance of species across space, and separating co-occurrence from realization allows us to model these processes independently, (2) by modeling the realization of interactions as a stochastic process dependent on species abundances and/or traits, and (3) by explicitly modeling imperfect detection of interactions by an observer.

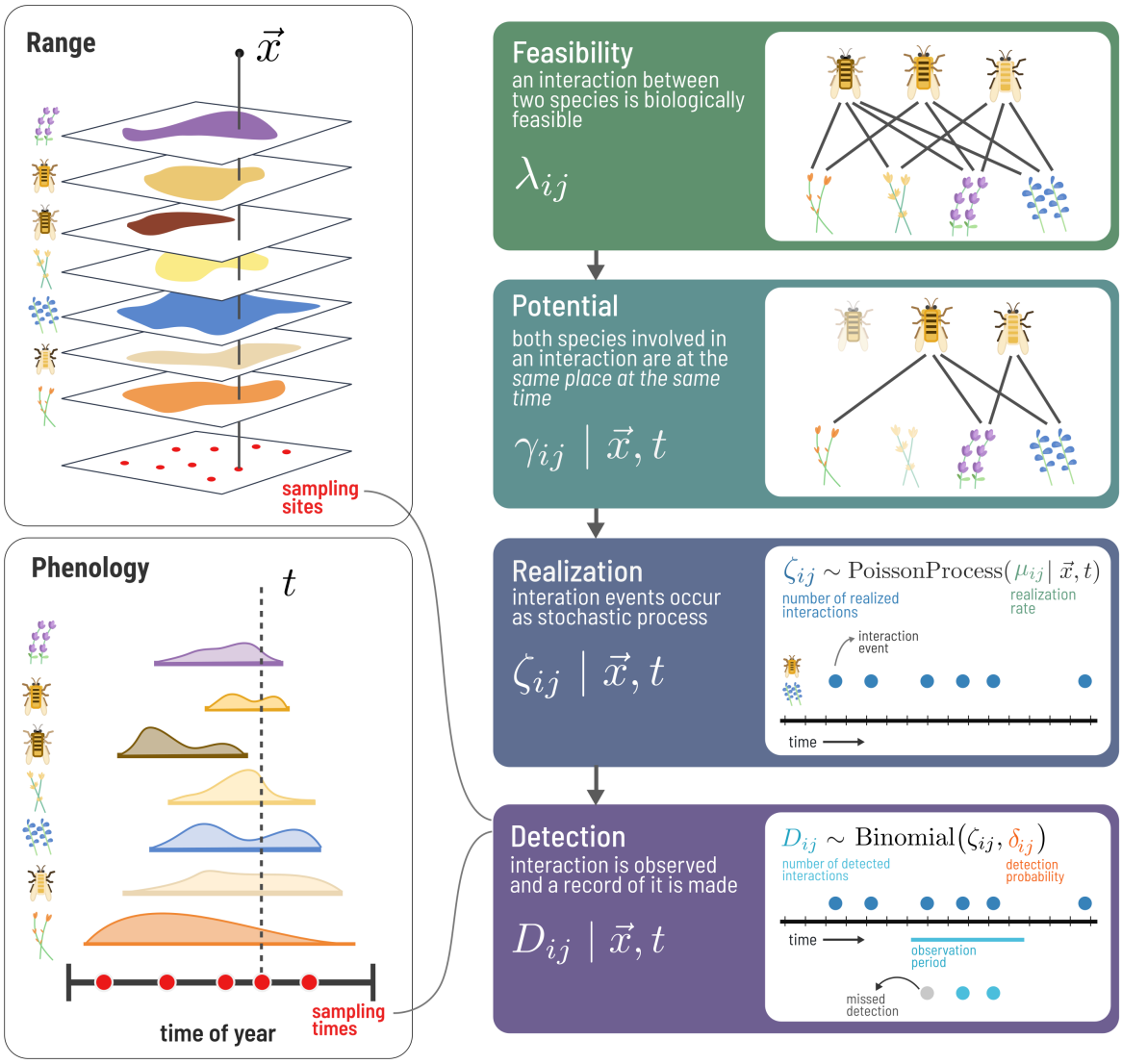

Figure 1: A spatiotemporal framework for interaction detection. Building upon the feasibility-realization-detection framework from Cirtwill et al. (2019), we introduce a new filter that feasible interactions go through before they are detected — potential — which explicitly accounts for species co-occurrence.

2.1 Interaction Feasibility

The first filter on interactions across all pairs of species to those that are detected is interaction feasibility, where an interaction is feasible if both species involved have the biological capability to interact Cirtwill et al. (2019). We denote the feasibility of an interaction between species i and j as $\lambda_{ij} \in \{0, 1\}$

The feasibility of a given interaction is governed by numerous factors depending on the particular system and taxa of interest. These can be morphological traits (e.g. body-size in food webs (Cohen et al. 2003), bill length in hummingbird pollinated systems (Vizentin-Bugoni et al. 2014), or based in genetics (e.g. the protein structure of receptors in a virus and host (Diamond 2009)).

Although an interaction between two species may be feasible ($\lambda_{ij} = 1$), this does not necessarily imply all individuals of each species can feasibly interact. Within-species variation in feasibility can vary, for example due to phenotypic plasticity enabling an interaction between some members of species and not others, or due to differences in morphology across life-stages.

2.2 Interaction Potential

Within Cirtwill et al. (2019)'s framework, both species co-occurrence and interaction are bundled into the realization filter that describes if interactions are realized during a given sampling event or not. We disentangle these two processes to make this framework spatiotemporally explicit by introducing a new filter, interaction potential, between feasibility and realization (Figure 1), which is dependent on the co-occurrence of both species involved in an interaction at a particular place x and time t.

This separates the co-occurrence component of interactions from their stochastic realization, enabling explicit modeling of the spatiotemporal heterogeneity in interaction networks due to variation in species abundance and occurrence, which can then influence the rate of interaction realization. We denote the occurrence of species i at location x and time t as $\gamma_i^{(x,t)}$, which is either a binary variable indicating presence (i.e. $\gamma_i^{(x,t)} \in \{0, 1\}$), or a measure of relative species density (i.e. $\gamma_i^{(x,t)}$ is between 0 and 1).

2.3 Interaction Realization

We introduce a family of process-based models of interaction realization. In the real world, species interactions are not regular and deterministic, but sporadic and stochastic. Therefore, we model the interaction realization as a Poisson process, where the number of events per unit of time is described by a rate parameter μ, and during a single unit of time, the number of events that occur is distributed as a Poisson(μ) distribution. We are then left with two separate challenges when choosing how to define interaction rates: (1) defining the total number of expected number of interactions across all interacting species per unit of time and (2) defining the rate of each particular interaction.

To govern the expected total number of interactions, we introduce a "realization energy" parameter ε, which governs the total number of expected interactions across all pairs of potentially interacting species at a given site and/or time. We then define a set of models for determining per-interaction rates that are **(1)** Abundance-based, **(2)** Trait-Matching based, and **(3)** Both Abundance and Trait-Matching based.

Abundance Based Realization

The core idea of abundance-based realization is that more abundant species are more likely to encounter each other in nature, and therefore the rate of interaction realization between two species i and j is a function of each species' abundance (the mass action hypothesis; Staniczenko et al. (2013)). We use this idea to describe a family of interaction realization models.

We start by defining a global distribution relative abundance $\mathcal{R}$ that is pooled across all locations and times. We denote the relative abundance of species i as R_i, and by definition

$$\sum_i R_i = 1$$

We then define the rate interaction $\mu_{ij}^{(x,t)}$ between species i and j at location x and time t as

$$\mu_{ij}^{(x,t)} = \frac{\varepsilon \cdot \lambda_{ij} \cdot f\left(r_i \gamma_i^{(x,t)}, r_j \gamma_j^{(x,t)}\right)}{\sum_{mn} \lambda_{mn} f\left(r_m \gamma_m^{(x,t)}, r_n \gamma_n^{(x,t)}\right)}$$

where $\gamma_i^{(x,t)}$ is the presence/abundance of species i at location x and time t, and f is a function for combining abundances, e.g. arithmetic or geometric mean of abundances, the product of abundances, or minimum/maximum abundance.

Note that the reason for normalization (the denominator in the above equation) is to ensure $\sum_{ij} \mu_{ij}^{(x,t)} = 1$, meaning the expected number of realized interactions is controlled only by the energy parameter, thereby decoupling parameters that control species-level occurrence and density from the overall number of realized interactions per site/time.

Trait-Matching Based Realization

A second model of interaction rates relates interaction realization rates to species' traits. This enables incorporating species "preferences" among potential interactions into interaction rates.

We consider each species having an arbitrary dimensional set of traits $z_i \in \mathbb{R}^N$, and we compute species traits matching using the Euclidean distance $|z_i - z_j|$, and compute the contribution to interaction rate using a kernel function k (e.g. Gaussian or exponential distance). The rate interaction $\mu_{ij}^{(x,t)}$ between species i and j at location x and time t is then computed in a similar way as before:

$$\mu_{ij}^{(x,t)} = \frac{\varepsilon \cdot \lambda_{ij} \cdot k\left(z_i, z_j\right)}{\sum_{mn} \lambda_{mn} k(z_m, z_n)}$$

Abundance and Trait-Based Realization

We can also combine abundance and trait-based realization by incorporating both species abundances and the trait-kernel into the computation of $\mu_{ij}^{(x,t)}$ as

$$\mu_{ij}^{(x,t)} = \frac{\varepsilon \cdot \lambda_{ij} \cdot f\left(r_i\gamma_i^{(x,t)}, r_j\gamma_j^{(x,t)}\right) \cdot k(z_i, z_j)}{\sum_{mn} \lambda_{mn} f\left(r_m\gamma_m^{(x,t)}, r_n\gamma_n^{(x,t)}\right) \cdot k(z_m, z_n)}$$

2.4 Interaction Detection

Finally, to model detection, we introduce an abundance-mediated model of detection probability. We denote the probability of successful detection of an interaction between species i and j as δ_{ij}. We model species-level detection probability as $\delta_i = 1 - (1 - r_i)^{\alpha}$ where α is a scaling parameter that controls detectability, and as $\alpha \to 0$, the probability of detecting rare species approaches 0. We then treat the probability of detecting an interaction as function of each species' detection probabilities, i.e. $\delta_{ij} = f(\delta_i, \delta_j)$, where f is e.g. the arithmetic or geometric mean product or minimum/maximum abundance, mirroring how the abundance can be combined to compute realization rates.

The number of detected interactions between species i and j is then drawn from a binomial distribution as $D_{ij}^{(x,t)} \sim \text{Binomial}\left(\zeta_{ij}^{(x,t)}, \delta_{ij}\right)$, where $\zeta_{ij}^{(x,t)}$ is the number of realized interactions drawn from $\zeta_{ij}^{(x,t)} \sim \text{Poisson}\left(\mu_{ij}^{(x,t)}\right)$.

3 Case Studies

Here we demonstrate how this framework can be used to quantify the sampling completeness of an interaction network as a function of how many total (non-unique) interactions have been detected in a given dataset. We do this using a software package we developed, SpeciesInteractionSamplers.jl, in the Julia language, which implements this framework. SIS contains numerous methods for generating feasible networks, ranges, and phenologies, and a variety of methods for interaction realization and detection, including the abundance-based models described above. The full set of features is described in the documentation here.

In these case studies, we explore how this relationship changes as a function of realization mode, detection, and spatial sampling effort. We measure sampling completeness using interaction accumulation curves, which measure the proportion of feasible interactions that have been detected as a function of the total number of (non-unique) interactions that have been detected.

3.1 Realization Rate

We first consider how the number of species, the shape of the species abundance distribution, and the relationship between abundance and realization rate affects interaction accumulation curves in networks without any spatial or temporal structure. All below results use the product of species abundances to parameterize realization rate, and the mean of species-level detectability ($\delta_i = 1 -$

$(1 - r_i)^\alpha$ with $\alpha = 20$) to parameterize the probability of interaction detection, unless one of these factors is the variable of interest being varied.

Because the total number of detected interactions is not an input parameter to the framework, we use an adaptive sampling scheme to generate interaction accumulation curves. Realization energy is the parameter we have control over, but the distribution of total detected interactions for a given energy value is a function of the models of realization and detection and their parameters. To get an even sample across the number of total detected interactions, we first do a small "pilot" run across a range of energy values with a small number of samples, and use linear regression to estimate the relationship between realization energy and total detected interactions. This enables simulation of interaction accumulation curves that are evenly spread across the number of total detected interactions. Here, each pilot run consists of 25 replicates of generated interaction networks across thirty energy values spread evenly between 0 and 50,000 to estimate the relationship between energy and number of detected interactions. Then, bins of total detected interactions of size 20, ranging from 0 to 10,000, are filled until 100 replicates are reached per bin.

Figure 2 contains simulations of food-web interaction accumulation curves under different numbers of species, abundance distributions, and realization modes. All feasible interaction networks are generated using the niche model of food webs (Williams and Martinez 2000) with connectance of 0.2. In Figure 2(A) we consider different species pool sizes. Unsurprisingly, networks composed of more species require more detected interactions to research the same level of completeness. In Figure 2(B), we look at how the shape of the species abundance distribution (SAD) impacts interaction accumulation. We model the SAD as a $\text{LogNormal}(\sigma)$ distribution (Preston 1948) with σ controlling the width of the distribution. Low values of σ correspond to nearly uniform SADs, and high σ corresponds to uneven distributions, with many individuals of common species and few individuals of rare species. We find that the more even the distribution of abundances is, the more quickly the sampling completeness of the network increases.

In Figure 2(C), we examine the impact of how abundances are combined to create realization rate in networks with even ($\sigma = 0.2$) and uneven ($\sigma = 2.0$) log-normal abundance distributions. The impact of the method for combining abundances used to compute realization rate is context-dependent: when abundances are long-tailed, the way abundances are combined to compute realization rate produces slower accumulation of unique detected interactions, as interactions involving very rare species have far lower realization rate when the rate is scaled by multiplying relative abundances than taking the mean. This extends to using the minimum of species abundance, which is slightly slower to accumulate detected interactions than the mean. In contrast, when abundances across species are even, there is little difference between product/mean/minimum modes.

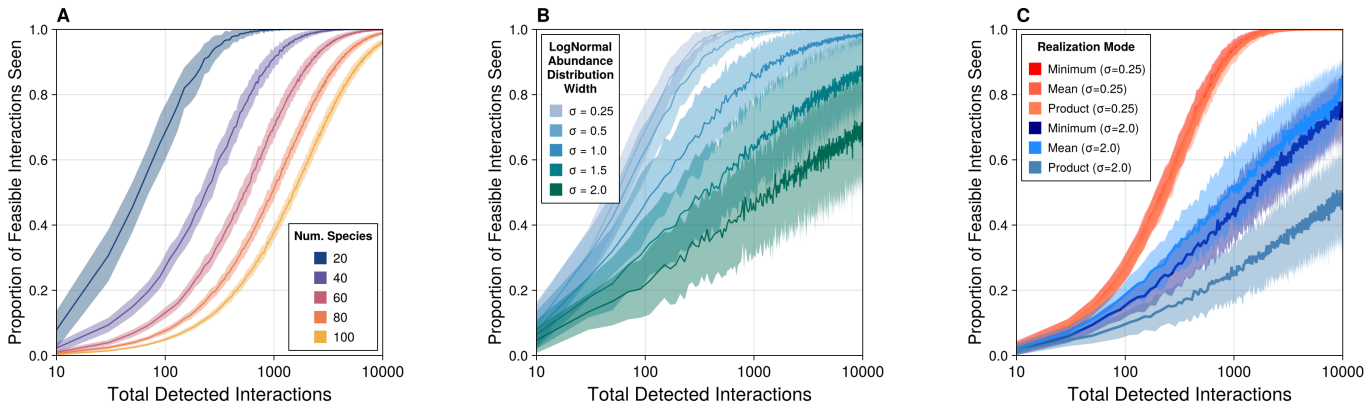

Figure 2: (A) The interaction accumulation curve for different sized species pools (colors). (B) The interaction accumulation curve across log-normal abundance distributions with different variances (colors). (C) The interaction accumulation curve under even (red) and uneven (blue) species abundance distributions. Different shades indicate different methods of combining abundance when computing realization rate (mean, minimum, product).

3.2 Detection

We now shift toward focusing on the impact of the interaction detection model on the interaction accumulation curve by adjusting the parameters of the abundance-dependent detection model. We adjust the adaptive sampling scheme used above to fit a unique linear model for each unique detectability scaling parameter to predict the average detected interactions per site as a function of the input energy. In practice, we found fitting a log-log regression, i.e.

$$\log(\text{energy}) \sim \log(\text{total detected})$$

produced a more even number of samples per bin.

In Figure 3(A), we see the species-level detection probability across different values of the scaling parameter α. The interaction detection rate, when being combined as the product of species-level detection, is shown for each combination of species relative abundance in Figure 3(B). Finally, the proportion of feasible interactions seen is shown as a function of both the total number of detected interactions and a gradient of the detectability scaling parameter in Figure 3(C). This emphasizes the importance of the detectability of interactions between rare species on the rate of interaction accumulation.

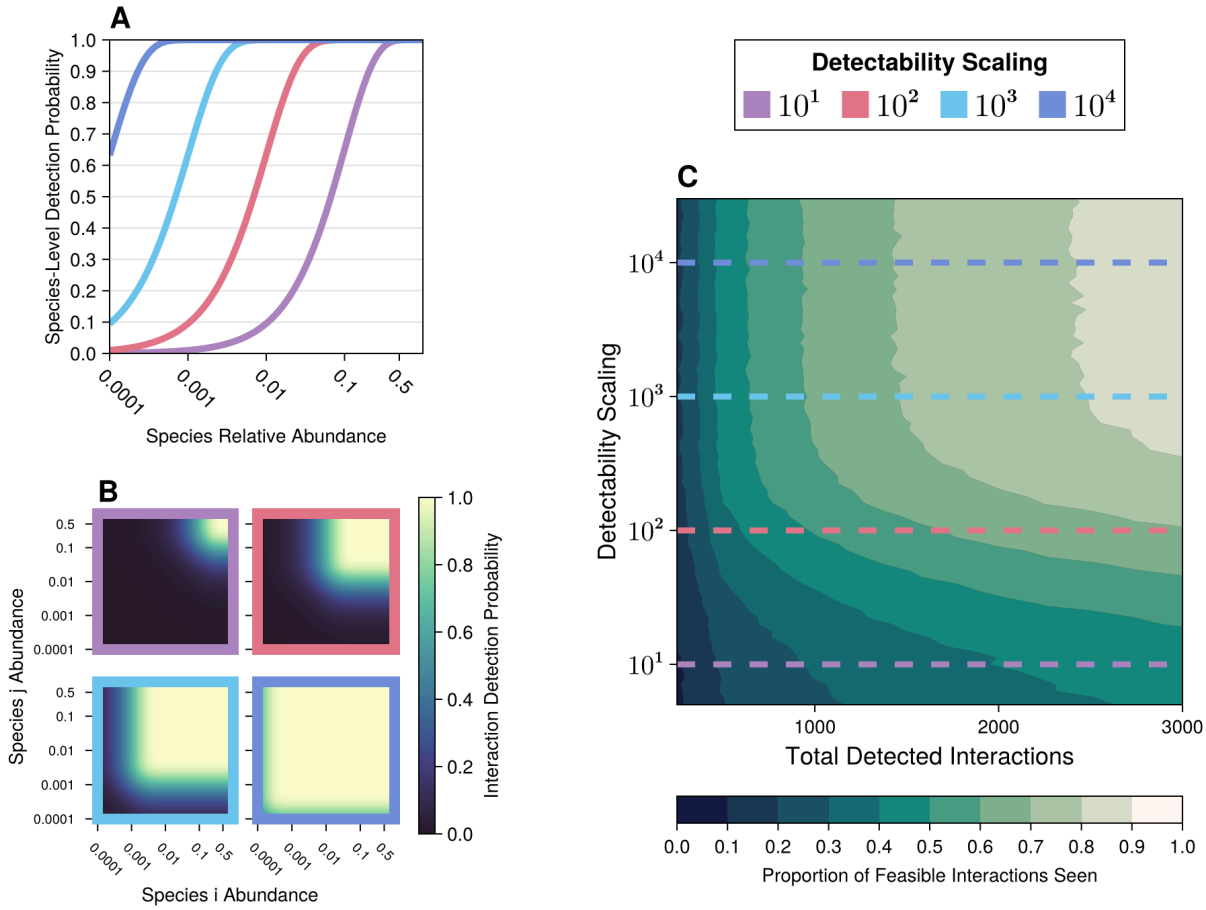

Figure 3: (**A**) The species-level detectability as a function of relative abundance (x-axis) under different detectability scaling parameters (colors). (**B**) The interaction detection probability across the relative-abundance of each involved species using the product mode of combining species-level scores. The color of each box corresponds to the detectability scaling parameter. (**C**) The proportion of feasible interactions that have been detected (colors) across a continuous gradient of the detectability scaling parameter (y-axis) and the total number of non-unique detected interactions (x-axis).

3.3 Spatial Coverage

Finally, we consider spatial variability in species occurrence. A practical challenge for sampling species interactions given the spatial variation in species ranges is determining how many sampling locations are needed to get a complete picture of the regional metaweb.

We generate 30 species food-webs using the niche model (Williams and Martinez 2000), with a connectance of 0.2. Ranges were generated using the Midpoint-Displacement algorithm with an autocorrelation parameter of 0.85, where the prevalence of each species is drawn from a $\text{Beta}(10, 10)$ distribution (meaning range sizes are highly concentrated around 50% of the total extent). We simulate spatial networks with 64 replicates per unique number of sampling sites and energy values ranging for 100 to 1000 (in intervals of 100), governing the expected number of interactions realized per sampling site. For simplicity, we use a homogenous distribution of energy

across space, so the expected number of realized interactions is the same across all sites. Similar to the previous case study, we adjust the adaptive sampling scheme by fitting a unique linear model for each number of sampling sites to predict the average detected interactions per site as a function of the input energy.

In Figure 4, we show the proportion of total feasible interactions in the metaweb detected across the number of sampled sites and the average number of (non-unique) detected interactions per sampling site. This shows that beyond a given level of sampling effort at each site (an average of 20 interactions detected per site), the primary driver of interaction accumulation is the total number of sites, not the effort at each site.

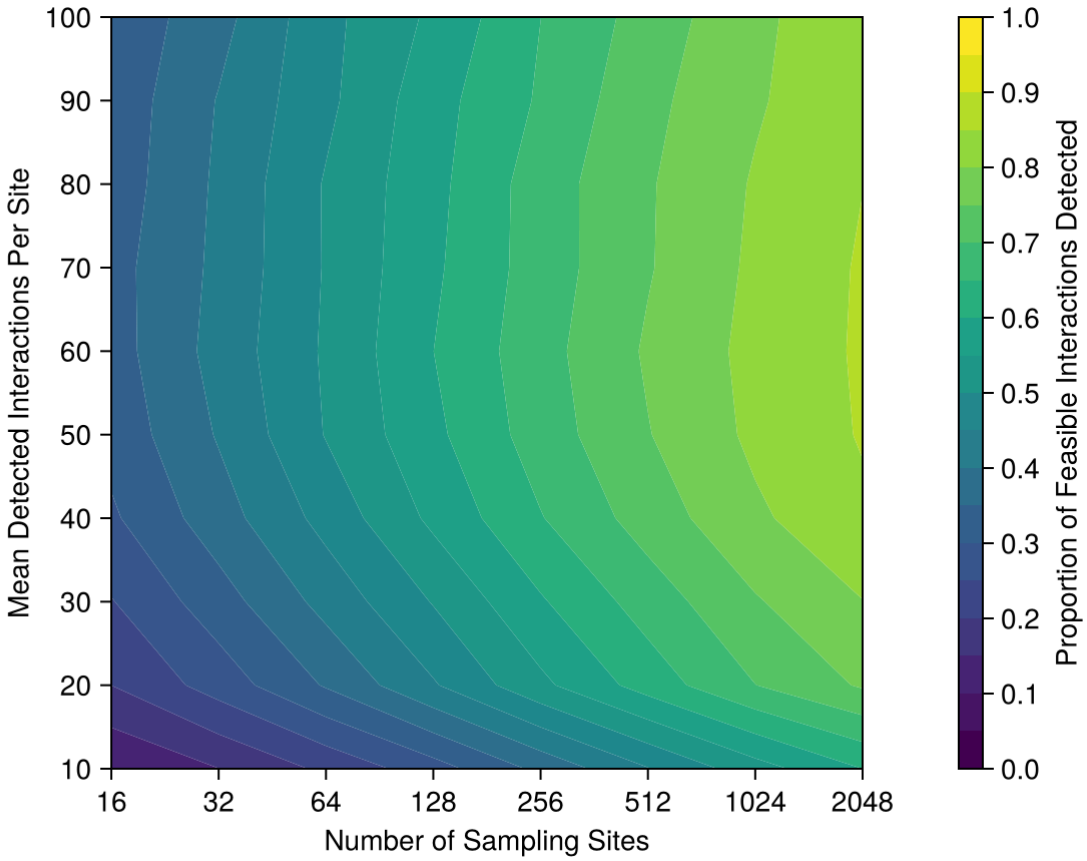

Figure 4: The proportion of feasible interactions that have been detected (colors) as a function of the number of sampling sites (x-axis; log scale) and the mean number of interactions detected at each site (y-axis).

4 Discussion

Species interactions underpin ecosystem function, but still remain a massive gap in our understanding of ecology (Hortal et al. 2015) due to the intrinsic difficulty imposed by interaction sampling (Jordano 2016a). Here, we have introduced a framework to calculate the size of this shortfall in empirical data by extending the framework for interaction detection from Cirtwill et al.

(2019) to a process-based model of interaction realization and detection. This enables simulation of interaction accumulation curves in spatiotemporally variable networks, where realization and detection is a function of species abundances, to estimate the sampling completeness of networks as a function of effort. We provide an associated software package, SpeciesInteractionSamplers.jl, which enables generation of feasible interaction networks, species ranges, phenologies, and abundances using a variety of models, and a variety of models of interaction realization and detection.

These tools provide a method to estimate the number of "missing" interactions in sampled networks as a function of sampling effort, and provide a sandbox for testing different strategies for designing monitoring schemes to most effectively sample interactions in spatiotemporally variable networks. This work has practical implications for the design of surveys of species interactions. First, the structure of species co-occurrence must be considered when sampling interactions. This is impacted by the goal of sampling: if the aim is targeting interactions whose feasibility we are most uncertain about, sampling design should select locations that maximize the chance of observing species co-occurrence. If instead the goal is broad-scale characterization of the ecological network across a large region, the goal should be to target sampling to be representative of the co-occurrence structure across space. This framework and software can then be used to test the efficacy of different monitoring design programs under known "ground-truth" scenarios of interaction network structure and spatiotemporal distribution.

Second, given the increased interest in using models to predict feasible but unobserved interactions between species based on traits (Gravel et al. 2013, Strydom et al. 2021, Caron et al. 2022), these methods provide a useful tool for estimating the number of interactions "missing" from the sampled data as a function of sampling effort and the spatiotemporal distribution of species. Having even a rough estimate of true class balance (the ratio of how many positives to. negatives are expected in the true data) can be useful in training models to predict interactions (Poisot 2022).

Further, explicitly accounting for imperfect detection in occurrence and abundance modeling is increasingly common (Royle et al. 2005). Although the use of predictive models has seen increased use for imputing feasible interactions (Stock et al. 2017, Young et al. 2021), explicitly accounting for detection error is not yet commonly adopted in these methods. The process-based models introduced here provide a basis for incorporating the impact of abundance on interaction realization and detection, and could be introduced into Bayesian models of interaction count data like presented in Becker et al. (2022).

Finally, the consequences of sampling completeness of species interaction data can impact our understanding of the impact of species interaction network structure on dynamics. For example, inferring the dynamic stability of a network could be prone to error if the observed network is not sampled "enough". Song et al. (2022) show that the stability of sampled subnetworks can be used to

infer the stability of the overall metaweb beyond a threshold of samples. Future work could focus on how robust this result is when species interaction coefficients are inferred from networks sampled with realization and detection rates influenced by species abundance. Further, predictions about the consequences of network rewiring (Thompson and 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.

A better understanding of sampling effects on the construction of species interaction networks is a necessity to have a statistically robust estimate of the current state and change over time of species interactions. This work provides a process-based framework for simulating the rate of interaction accumulation when sampling, while accounting for spatiotemporal variation in interaction networks (Poisot et al. 2015) and the effects of abundance on interaction realization and detection. This provides a necessary step forward for developing a theory of interaction monitoring to improve our understanding of species interaction networks and how they are changing.

Code and Data Availability: The code for SpeciesInteractionSamplers.jl is available on Github here.

Acknowledgements: MDC is funded by an IVADO Postdoctoral Fellowship. TP is funded by an NSERC Discovery grant, a Discovery Acceleration Supplement grant, and a Wellcome Trust grant (223764/Z/21/Z). AG and MDC were supported by the Liber Ero Chair in Conservation.

References

Aguiar MAM de, Newman EA, Pires MM, et al (2019) Revealing Biases in the Sampling of Ecological Interaction Networks. PeerJ 7:e7566. https://doi.org/10.7717/peerj.7566

Banville F, Strydom T, Blyth PSA, et al (2025) Deciphering Probabilistic Species Interaction Networks. Ecology Letters 28:e70161. https://doi.org/10.1111/ele.70161

Bartley TJ, McCann KS, Bieg C, et al (2019) Food Web Rewiring in a Changing World. Nature Ecology & Evolution 3:345–354. https://doi.org/10.1038/s41559-018-0772-3

Becker DJ, Albery GF, Sjodin AR, et al (2022) Optimising Predictive Models to Prioritise Viral Discovery in Zoonotic Reservoirs. The Lancet Microbe 3:e625–e637. https://doi.org/10.1016/S2666-5247(21)00245-7

Carlson CJ, Albery GF, Merow C, et al (2022) Climate Change Increases Cross-Species Viral Transmission Risk. Nature 607:555–562. https://doi.org/10.1038/s41586-022-04788-w

Caron D, Maiorano L, Thuiller W, Pollock LJ (2022) Addressing the Eltonian Shortfall with Trait-Based Interaction Models. Ecology Letters 25:889–899. https://doi.org/10.1111/ele.13966

Chacoff NP, Vázquez DP, Lomáscolo SB, et al (2012) Evaluating Sampling Completeness in a Desert Plant–Pollinator Network. Journal of Animal Ecology 81:190–200. https://doi.org/10.1111/j.1365-2656.2011.01883.x

Cirtwill AR, Eklöf A, Roslin T, et al (2019) A Quantitative Framework for Investigating the Reliability of Empirical Network Construction. Methods in Ecology and Evolution 10:902–911. https://doi.org/10.1111/2041-210X.13180

Cohen JE, Jonsson T, Carpenter SR (2003) Ecological Community Description Using the Food Web, Species Abundance, and Body Size. Proceedings of the National Academy of Sciences 100:1781–1786. https://doi.org/10.1073/pnas.232715699

Devarajan K, Morelli TL, Tenan S (2020) Multi-Species Occupancy Models: Review, Roadmap, and Recommendations. Ecography 43:1612–1624. https://doi.org/10.1111/ecog.04957

Diamond MS (2009) Virus and Host Determinants of West Nile Virus Pathogenesis. PLOS Pathogens 5:e1000452. https://doi.org/10.1371/journal.ppat.1000452

Dorazio RM, Royle JA (2005) Estimating Size and Composition of Biological Communities by Modeling the Occurrence of Species. Journal of the American Statistical Association 100:389–398. https://doi.org/10.1198/016214505000000015

Gravel D, Poisot T, Albouy C, et al (2013) Inferring Food Web Structure from Predator–Prey Body Size Relationships. Methods in Ecology and Evolution 4:1083–1090. https://doi.org/10.1111/2041-210X.12103

Guillera-Arroita G (2017) Modelling of Species Distributions, Range Dynamics and Communities under Imperfect Detection: Advances, Challenges and Opportunities. Ecography 40:281–295. https://doi.org/10.1111/ecog.02445

Hortal J, De Bello F, Diniz-Filho JAF, et al (2015) Seven Shortfalls That Beset Large-Scale Knowledge of Biodiversity. Annual Review of Ecology, Evolution, and Systematics 46:523–549. https://doi.org/10.1146/annurev-ecolsys-112414-054400

Jordano P (2016b) Chasing Ecological Interactions. PLOS Biology 14:e1002559. https://doi.org/10.1371/journal.pbio.1002559

Jordano P (2016a) Sampling Networks of Ecological Interactions. Functional Ecology 30:1883–1893. https://doi.org/10.1111/1365-2435.12763

MacKenzie DI, Nichols JD, Lachman GB, et al (2002) Estimating Site Occupancy Rates When Detection Probabilities Are Less Than One. Ecology 83:2248–2255. https://doi.org/10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2

Makiola A, Compson ZG, Baird DJ, et al (2020) Key Questions for Next-Generation Biomonitoring. Frontiers in Environmental Science 7:

Martinez ND, Hawkins BA, Dawah HA, Feifarek BP (1999) Effects of Sampling Effort on Characterization of Food-Web Structure. Ecology 80:1044–1055. https://doi.org/10.1890/0012-9658(1999)080[1044:EOSEOC]2.0.CO;2

McLeod A, Leroux SJ, Gravel D, et al (2021) Sampling and Asymptotic Network Properties of Spatial Multi-Trophic Networks. Oikos 130:2250–2259. https://doi.org/10.1111/oik.08650

Poisot T (2022) Guidelines for the Prediction of Species Interactions through Binary Classification

Poisot T, Bergeron G, Cazelles K, et al (2021) Global Knowledge Gaps in Species Interaction Networks Data. Journal of Biogeography 48:1552–1563. https://doi.org/10.1111/jbi.14127

Poisot T, Stouffer DB, Gravel D (2015) Beyond Species: Why Ecological Interaction Networks Vary through Space and Time. Oikos 124:243–251. https://doi.org/10.1111/oik.01719

Pollock LJ, Tingley R, Morris WK, et al (2014) Understanding Co-occurrence by Modelling Species Simultaneously with a Joint Species Distribution Model (\textsc{JSDM}). Methods in Ecology and Evolution 5:397–406. https://doi.org/10.1111/2041-210X.12180

Preston FW (1948) The Commonness, And Rarity, of Species. Ecology 29:254–283. https://doi.org/10.2307/1930989

Royle JA (2006) Site Occupancy Models with Heterogeneous Detection Probabilities. Biometrics 62:97–102. https://doi.org/10.1111/j.1541-0420.2005.00439.x

Royle JA (2004) \emph{N} -Mixture Models for Estimating Population Size from Spatially Replicated Counts. Biometrics 60:108–115. https://doi.org/10.1111/j.0006-341X.2004.00142.x

Royle JA, Nichols JD, Kéry M (2005) Modelling Occurrence and Abundance of Species When Detection Is Imperfect. Oikos 110:353–359. https://doi.org/10.1111/j.0030-1299.2005.13534.x

Sanders HL (1968) Marine Benthic Diversity: A Comparative Study. The American Naturalist 102:243–282. https://doi.org/10.1086/282541

Simberloff D (1972) Properties of the Rarefaction Diversity Measurement. The American Naturalist 106:414–418

Song C, Simmons BI, Fortin M-J, et al (2022) Rapid Monitoring for Ecological Persistence. 2022.07.02.498308

Staniczenko PPA, Kopp JC, Allesina S (2013) The Ghost of Nestedness in Ecological Networks. Nature Communications 4:1391. https://doi.org/10.1038/ncomms2422

Stock M, Poisot T, Waegeman W, De Baets B (2017) Linear Filtering Reveals False Negatives in Species Interaction Data. Scientific Reports 7:45908. https://doi.org/10.1038/srep45908

Strydom T, Catchen MD, Banville F, 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. https://doi.org/10.1098/rstb.2021.0063

Terry JCD, Lewis OT (2020) Finding Missing Links in Interaction Networks. Ecology 101:e3047. https://doi.org/10.1002/ecy.3047

Thompson PL, Gonzalez A (2017) Dispersal Governs the Reorganization of Ecological Networks under Environmental Change. Nature Ecology & Evolution 1:1–8. https://doi.org/10.1038/s41559-017-0162

Vizentin-Bugoni J, Maruyama PK, Sazima M (2014) Processes Entangling Interactions in Communities: Forbidden Links Are More Important than Abundance in a Hummingbird–Plant Network. Proceedings of the Royal Society B: Biological Sciences 281:20132397. https://doi.org/10.1098/rspb.2013.2397

Williams RJ, Martinez ND (2000) Simple Rules Yield Complex Food Webs. Nature 404:180–183. https://doi.org/10.1038/35004572

Young J-G, Valdovinos FS, Newman MEJ (2021) Reconstruction of Plant–Pollinator Networks from Observational Data. Nature Communications 12:3911. https://doi.org/10.1038/s41467-021-24149-x