

Review



**Cite this article:** Strydom T *et al.* 2021 A roadmap towards predicting species interaction networks (across space and time). *Phil. Trans. R. Soc. B* **376**: 20210063. <https://doi.org/10.1098/rstb.2021.0063>

Accepted: 29 July 2021

One contribution of 15 to a theme issue ‘Infectious disease macroecology: parasite diversity and dynamics across the globe’.

**Subject Areas:**

ecology, theoretical biology

**Keywords:**

ecological networks, machine learning, deep learning, ecological forecasting, biogeography

**Author for correspondence:**

Timothée Poisot  
e-mail: [timothee.poisot@umontreal.ca](mailto:timothee.poisot@umontreal.ca)

†These authors contributed equally to the work.

# A roadmap towards predicting species interaction networks (across space and time)

Tanya Strydom<sup>1,2,†</sup>, Michael D. Catchen<sup>2,3,†</sup>, Francis Banville<sup>1,2,4</sup>, Dominique Caron<sup>2,3</sup>, Gabriel Dansereau<sup>1,2</sup>, Philippe Desjardins-Proulx<sup>1,2</sup>, Norma R. Forero-Muñoz<sup>1,2</sup>, Gracielle Higino<sup>5</sup>, Benjamin Mercier<sup>2,4</sup>, Andrew Gonzalez<sup>2,3</sup>, Dominique Gravel<sup>2,4</sup>, Laura Pollock<sup>2,3</sup> and Timothée Poisot<sup>1,2</sup>

<sup>1</sup>Sciences Biologiques, Université de Montréal, Montréal, Canada H2V 0B3

<sup>2</sup>Québec Centre for Biodiversity Sciences, Montréal, Canada

<sup>3</sup>McGill University, Montréal, Canada

<sup>4</sup>Université de Sherbrooke, Sherbrooke, Canada

<sup>5</sup>Universidade Federal de Goiás, Goiânia, Brasil

TS, 0000-0001-6067-1349; DC, 0000-0002-2151-6693; GD, 0000-0002-2212-3584; TP, 0000-0002-0735-5184

Networks of species interactions underpin numerous ecosystem processes, but comprehensively sampling these interactions is difficult. Interactions intrinsically vary across space and time, and given the number of species that compose ecological communities, it can be tough to distinguish between a true negative (where two species never interact) from a false negative (where two species have not been observed interacting even though they actually do). Assessing the likelihood of interactions between species is an imperative for several fields of ecology. This means that to predict interactions between species—and to describe the structure, variation, and change of the ecological networks they form—we need to rely on modelling tools. Here, we provide a proof-of-concept, where we show how a simple neural network model makes accurate predictions about species interactions given limited data. We then assess the challenges and opportunities associated with improving interaction predictions, and provide a conceptual roadmap forward towards predictive models of ecological networks that is explicitly spatial and temporal. We conclude with a brief primer on the relevant methods and tools needed to start building these models, which we hope will guide this research programme forward.

This article is part of the theme issue ‘Infectious disease macroecology: parasite diversity and dynamics across the globe’.

## 1. Introduction

Ecosystems are, in large part, constructed by the interactions within them—organisms interact with one another and with their environment, either directly or indirectly. Interactions between individuals, populations, and species create networks of interactions that drive ecological and evolutionary dynamics and maintain the coexistence, diversity and functioning of ecosystems [1–3]. Species interaction networks underpin our understanding of numerous ecological processes [4,5]. Yet, even basic knowledge of species interactions (like being able to list them, or guess which ones may exist) remains one of the most severe biodiversity data shortfalls [6], in large part owing to the tedious, time-consuming, and expensive process of collecting species interaction data. Comprehensively sampling every possible interaction is not feasible given the sheer number of species on Earth, and the data we can collect about interactions tend to be biased and noisy [7]. This is then compounded as species interactions are typically measured as a binary variable (present or absent) even though it is evident interactions are not all-or-nothing. Empirically we know species

interactions occur probabilistically owing to variation in species abundances in space and time [8]. Different types of interactions vary in their intrinsic predictability (e.g. some fungal species engage in opportunistic saprotrophy [9], obligate parasites are more deterministic in their interactions than facultative parasites [10,11]). In addition to this variance in predictability, networks from different systems are structured by different mechanisms.

Still, like all of Earth's systems, species interaction networks have entered their 'long now' [12], where anthropogenic change will have long-term, low-predictability consequences [13] for our planet's ecology. Therefore, our field needs a roadmap towards models that enable prediction (for the present) and forecasting (for the future) of species interactions and the networks they form, and which accounts for their spatial and temporal variation [14,15]. As an example, in disease ecology, predicting potential hosts of novel disease (recently notably the search for wildlife hosts of betacoronaviruses; [16,17]) has received much attention. Network approaches have been used for the prediction of risk and dynamics of dengue [18], Chagas disease [19], Rickettsiosis [20], Leishmaniasis [21] and a myriad of infectious diseases in livestock and wildlife [22]. Additionally, prediction of interaction networks is a growing imperative for next-generation biodiversity monitoring, requiring a conceptual framework and a flexible set of tools to predict interactions that is explicitly spatial and temporal in perspective [23–25]. Developing better models for prediction of these interactions will rely on integration of data from many sources, and the sources for this data may differ depending on the type of interaction we wish to predict [26].

Interactions between species can be conceptualized in a multitude of ways (mutualistic versus antagonistic, strong versus weak, symmetric versus asymmetric, direct versus indirect) [27,28]. What is common among definitions of species interactions is that *at least* one of the species is affected by the presence of another [28]. Networks can be used to represent a variety of interaction types, including: *unipartite networks*: where each species can be linked to other species (often food webs), *bipartite networks*: where there are two pools of species and all interactions occur between species in each pool (typically used for pairwise interactions; e.g. hosts and parasites), and *k-partite networks*: which expand to more than two disjoint sets of interacting species (e.g. some parasitoid webs, seed dispersal networks and pollination networks [29]).

Methods for predicting interactions between species exist, but at the moment are difficult to generalize as they are typically based around a single mechanism at a single scale: position in the trophic niche [30,31], phylogenetic distance [32,33], functional trait matching [34], interaction frequency [35,36], or other network properties [37,38]. Species interaction networks, as we observe them on Earth today, are the product of ecological and evolutionary mechanisms interacting across spatial, temporal and organizational scales. The interwoven nature of these processes imposes structure on biodiversity data which is invisible when examined only through the lens of a single scale, however machine learning (ML) methods have enormous potential to find this structure in data [39], and have the potential to be used together with mechanistic models in order to make prediction of ecological dynamics more robust [40].

Here, we use a case study to show how ML models (specifically a deep neural network) can enable prediction

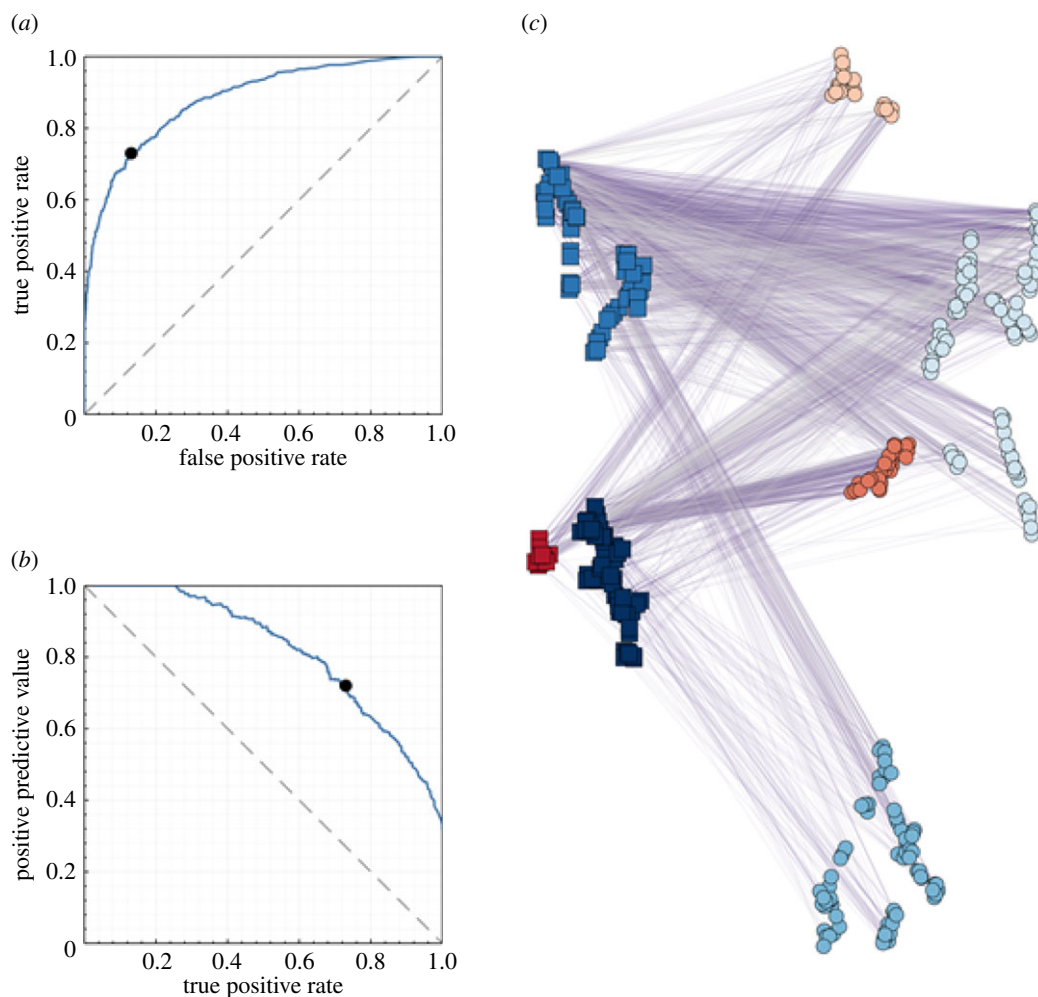
of species interactions: we construct a metaweb of host–parasite interactions across space, using predictors extracted from empirical data and accounting for the structure of co-occurrence between species. We use this case study to illustrate a roadmap for improving predictions using open data and ML methods; specifically, we focus on how emerging tools from ML can be used to deliver more accurate and more efficient predictions of ecological systems, and how the potential of these approaches will be magnified with increased data access. We then provide a non-exhaustive primer on the literature on interaction prediction, and identify the tools and methods most suited for the future of interaction network prediction models, covering the spatial, temporal and climatic dimensions of network prediction [41]. Both the case study and primer are largely geared towards binary (interactions are either present or absent) networks; there are limitations in data and tools that make it a more reasonable starting approach. First, most ecological networks do not have estimates of interaction strength, and particularly not estimates that are independent from relative abundances. Second, the methodological toolkit to analyse the structure of networks is far more developed for binary interactions [2], meaning that the predictions of binary interactions can be more readily interpreted.

We argue that adopting a more predictive approach to complex ecological systems (like networks) will establish a positive feedback loop with our understanding of these systems [42]: the tasks of understanding and predicting are neither separate nor opposed [43]. Instead, ML tools have the ability to capture a lot of our understanding into working assumptions, and comparing predictions to empirical data gives us better insights about how much we ignore about the systems we model (see [44], who provide an overview of deep learning techniques and concepts in ecology and evolution). Although data on species interaction networks are currently limited in the size of their spatial coverage, ML approaches have a demonstrated track record of revealing the 'unreasonable effectiveness' of data [45]; we argue that with a clear roadmap guiding the use of these methods, the task of predicting species interaction networks will become more attainable.

## 2. A case study: deep learning of spatially sparse host–parasite interactions

The premise of this manuscript is that we can predict interactions between species. In this section, we provide a proof-of-concept, where we use data from Hadfield *et al.* [46] describing 51 host–parasite networks sampled across space. In this data, as in most spatially distributed ecological networks, not all species co-occur across sites. As a direct consequence, there are pairs of species that may or may not be able to interact for which we have no data; furthermore, there are pairs of species that may interact, but have only been documented in a single location where the interaction was not detected. In short, there are ecological reasons to believe that a number of negative associations in the metaweb (*sensu* [47]) are false negatives.

Without any species-level information, we resort to using both co-occurrence and known interactions to predict novel interactions. To do this, we (i) extract features (equivalent to explanatory variables in a statistical model) for each species based on co-occurrence, (ii) use these features to train a



**Figure 1.** Proof-of-concept: an empirical metaweb (from Hadfield *et al.* [46]), i.e. a list of co-occurrences within a species pool, is converted into latent features using probabilistic PCA, then used to train a deep neural network to predict species interactions. Panels (a) and (b) represent, respectively, the receiver-operating-characteristic curve and the precision-recall curve, with the best classifier (according to Youden's  $J$ ) represented by a black dot. The expected performance of a neutral 'random-guessing' classifier is shown with a dashed line. Panel (c) shows the imputed using  $t$ -distributed stochastic neighbour embedding (tSNE), and the colours of nodes are the cluster to which they are assigned based on a  $k$ -means clustering of the tSNE output. Empirical interactions are shown in purple, and imputed interactions in grey.

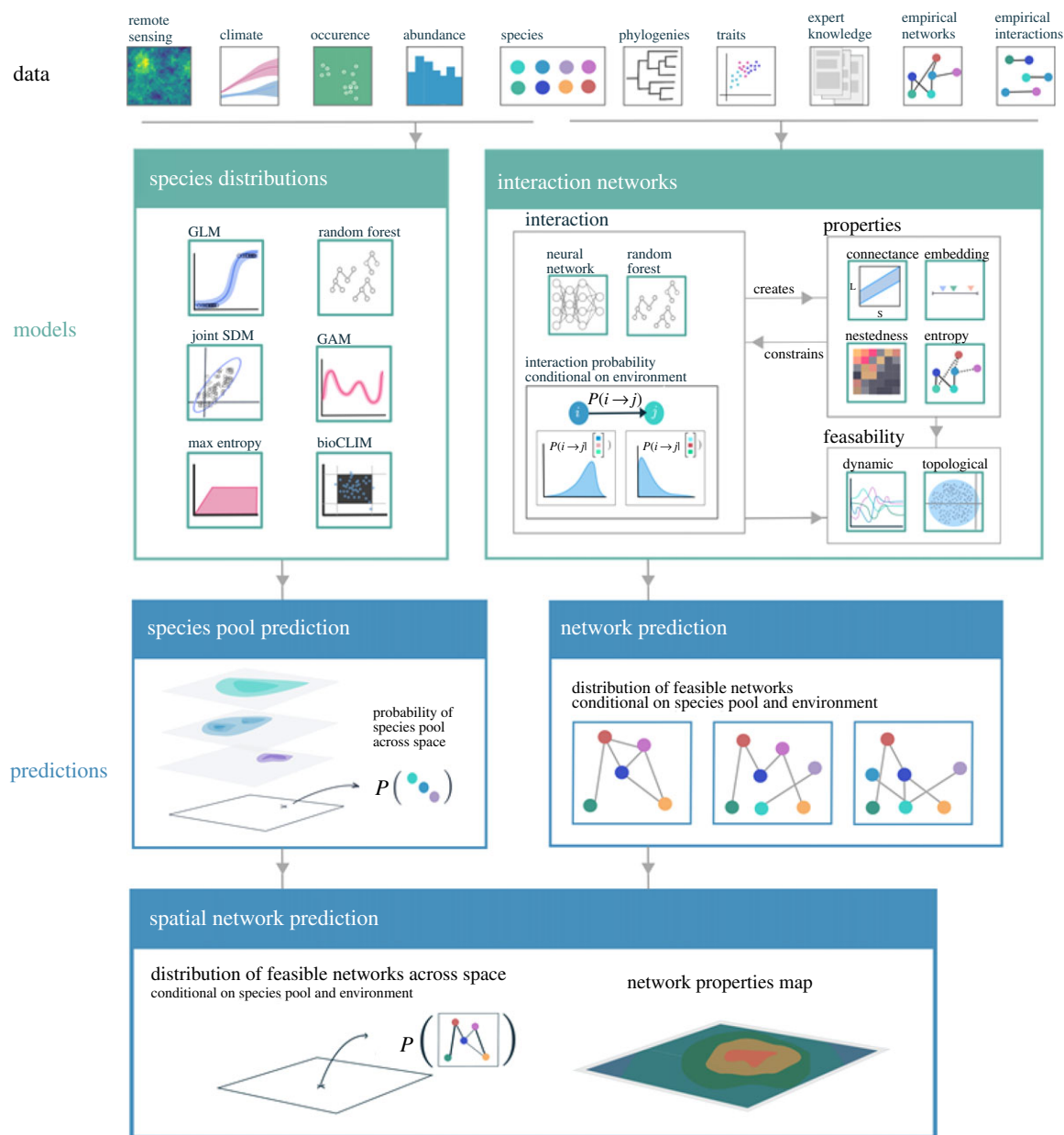
neural network to predict interactions, and (iii) apply this classifier (an algorithm that assigns a categorical output based on input features) to the original features to predict potential interactions across the entire species pool. ML relies on a lexicon that shares some terms with statistics, albeit with different meaning; we expand on the precise meanings in the 'How to validate a predictive model' section below. The outputs of the analysis are presented in figure 1, and the code to reproduce it is available at <https://osf.io/6jp4b/>; the entire example was carried out in *Julia 1.6.2* [48], using the *Flux* machine learning framework [49].

We first aggregate all species into a co-occurrence matrix  $A$  which represents whether a given pair of species ( $i, j$ ) was observed coexisting across any location. We then transform this co-occurrence matrix  $A$  via probabilistic principal component analysis (PCA) [50] and use the first 15 values from this PCA space as the feature vector for each species  $i$ . For each pair of (host, parasite) species ( $i, j$ ), we then feed the feature vectors ( $v_i, v_j$ ) into a neural network. The neural network uses four feed-forward layers (the weights for each layer is independent from the one before and after); the first layer uses the RELU activation function (which ignores input below a threshold), the rest use a  $\sigma$  function (which transforms linear activation energies into logistic responses). All layers

have appropriate dropout rates (in order to avoid over-fitting, only a fraction of the network is updated on each iteration:  $1 - 0.8$  for the first layer,  $1 - 0.6$  for the subsequent ones). This produces an output layer with a single node, which is the probability-score for interaction between species  $i$  and  $j$ .

We then train (equivalent to *fit*) this neural network by dividing the original dataset into testing and training sets (split 80-20 for training and testing, respectively). During the training of this neural network (using the ADAM optimizer learning rate), the  $5 \times 10^4$  batches of 64 items used for training were constrained to have at least 25% of positive interactions, as Poisot *et al.* [51] show slightly inflating the dataset with positive interactions enables us to counterbalance sampling biases. Furthermore, setting a minimum threshold of response balance is an established approach for datasets with strong biases [52]. Validating this model on the test data shows our model provides highly effective prediction of interactions between pairs of species not present in the training data (figure 1). The behaviour of the model was, in addition, checked by measuring the training and testing loss (difference between the actual value and the prediction, here using mean-squared error) and stopping well before they diverged (to avoid overfitting).

This case study shows that a simple neural network can be very effective in predicting species interactions even



**Figure 2.** A conceptual roadmap highlighting key areas for the prediction of ecological networks. Starting with the input of data from multiple sources, followed by a modelling framework for ecological networks and the landscape, which are then ultimately combined to allow for the prediction of spatially explicit networks.

without additional species-level data. Applying this model to the entire dataset (including species pairs never observed to co-occur) identified 1546 new possible interactions—746 (48%) of which were between pairs of species for which no co-occurrence was observed in the original dataset. This model reaches similar levels of predictive efficacy as previous studies that use far more species-level data and mechanistic assumptions [30], which serves to highlight the potential for including external sources of data for *improving* our prediction of interaction networks even further. For example, Krasnov *et al.* [53] collected traits data for this system that could be added to the model, in addition or in substitution to latent variables derived from observed interactions.

### 3. Predicting species interaction networks across space: challenges and opportunities

Here, we present a conceptual roadmap (figure 2) which shows a conceptual path from data to prediction of species

interaction networks, incorporating several modelling frameworks. We envisage this roadmap to be one conceptual path towards incorporating space into our prediction of interaction networks, and developing spatially explicit models of networks and their properties. In the following sections we discuss the challenges and opportunities for this path forward, and highlight two specific areas where it can have a strong impact: the temporal forecasting of species interaction networks structure, and the use of predicted networks for applied ecology and conservation biology.

#### (a) Challenges: constraints on predictions

##### (i) Ecological network data are scarce and hard to obtain

At the moment, prediction of species interactions is made difficult by the limited availability of data. Although we have seen a growth in species occurrence data, this growth is much slower for ecological interactions because species interactions are challenging to sample comprehensively [54,55] and sampling methodology has strong effects on the

resulting data [7]. In turn, the difficulty of sampling interactions can lead to biases in our understanding of network structure [7]. This knowledge gap has motivated a variety of approaches to deal with interactions in ecological research based on assumptions that do not always hold, such as the assumption that co-occurrence is equivalent to meaningful interaction strength [56]. Spatial biases in data coverage are prevalent at the global scale (with South America, Africa and Asia being under-represented) and different interaction types show biases towards different biomes [57]. These ‘spatial gaps’ serve as a limitation to our ability to confidently make predictions when accounting for real-world environmental conditions, especially in environments for which there are no analogous data.

Furthermore, empirical estimation of interaction *strength* is highly prone to bias as existing data are usually summarized at the taxonomic scale of the species or higher, thereby losing information that differentiates the strength in per-individual interactions from the strength of a whole species interaction [58]. Empirical estimations of interaction strength are still crucial [59], but are a hard task to quantify in natural communities [60–62], especially as the number of species composing communities increases, compounded by the possibility of higher-order interactions or non-linear responses in interactions [62]. Furthermore, interaction strength is often variable and context dependent and can be influenced by density-dependence and spatio-temporal variation in community composition [62].

### (ii) Powerful predictive tools work better on large data volumes

This scarcity of data limits the range of computational tools that can be used by network ecologists. Most deep learning methods, for instance, are very data expensive. The paucity of data is compounded by a collection of biases in existing datasets. Species interaction data are typically dominated by food webs, pollination and host–parasite networks [63,64]. This could prove to be a limiting factor when trying to understand or predict networks of under-represented interaction types or when trying to integrate networks of different types [65], especially given their inherent structural variation [66]. This stresses the need for an integrated, flexible and data-efficient set of computational tools which will allow us to predict ecological networks accurately from existing and imperfect datasets, but also enable us to perform model validation and comparison with more flexibility than existing tools. We argue that figure 1 is an example of the promise of these tools *even* when facing datasets of small size. The ability to extract and engineer features also serves to bolster our predictive power. Although it may be tempting to rely on approaches like bootstrapping to estimate the consistency of the predictions, we are confronted with the issues of low data volume and data bias—that we are more likely to observe interactions between some pairs of species (i.e. those that co-occur often, e.g. [67], and those with higher relative abundance, e.g. [68]). This introduces risk in training models on pseudo-replicated data. In short, the current lack of massive datasets must not be an obstacle to prediction; it is an ideal testing ground to understand how little data is sufficient to obtain actionable predictions, and how much we can rely on data inflation procedures to reach this minimal amount.

### (iii) Scaling-up predictions requires scaled-up data

We are also currently limited by the level of biological organization at which we can describe ecological networks. For instance, our understanding of individual-based networks (e.g. [69,70]) is still in its infancy [71] and acts as a resolution-limit. Similarly, the resolution of environmental (or landscape) data also limits our ability to predict networks at small scales, although current trends in remote sensing would suggest that this will become less of a hindrance with time [72]. Ecosystems are a quintessential complex-adaptive-system [73] with a myriad of processes at different spatial, temporal, and organizational scales that influence and respond to one another. Understanding how the product of these different processes drive the properties of ecosystems across different scales remains a central challenge of ecological research, and we should strive to work on methods that will integrate different empirical ‘snapshots’ of this larger system.

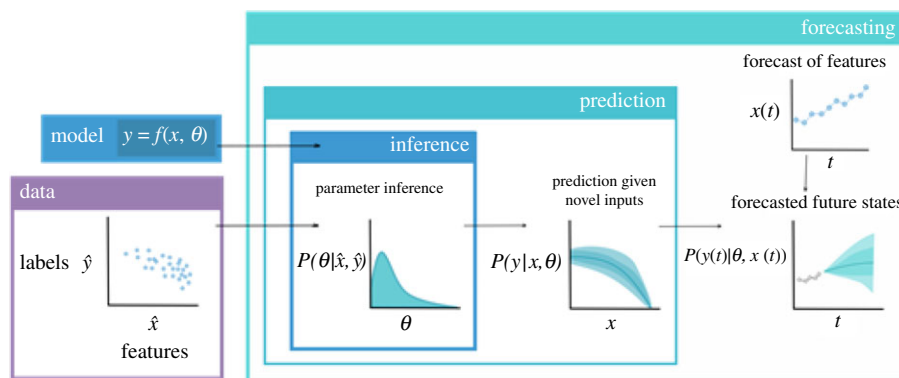
## (b) Opportunities: an emerging ecosystem of open tools and data

### (i) Data are becoming more interoperable

The acquisition of biodiversity and environmental data has tremendously increased over the past decades thanks to the rise of citizen science [74] and of novel technology [75], including wireless sensors [76], next-generation DNA sequencing [77], and remote sensing [78,79]. Open access databases, such as GBIF (<https://www.gbif.org/>) (for biodiversity data), NCBI (<https://www.ncbi.nlm.nih.gov/>) (for taxonomic and genomics data), TreeBASE (<https://www.treebase.org/treebase-web/home.html>) (for phylogenetics data), CESTE (<https://icestes.github.io/>) [80] (for metacommunity ecology and species traits data) and WorldClim (<https://www.worldclim.org/data/bioclim.html>) (for bioclimatic data) contain millions of data points that can be integrated to monitor and model biodiversity at the global scale. For species interactions data, at the moment Mangal (<https://mangal.io/#/>) is the most comprehensive open database of published ecological networks [81], and GloBI (<https://www.globalbioticinteractions.org/about>) is an extensive database of realized and potential species interactions [82]. Developing standard practices in data integration and quality control [83] and in next-generation biomonitoring [72] would improve our ability to make reliable predictions of ecosystem properties on increasing spatial and temporal scales. The advancement of prediction techniques coupled with a movement towards standardizing data collection protocols (e.g. [84] for plant functional traits) and metadata (e.g. DarwinCore)—which facilitates interoperability and integration of datasets—as well as a growing interest at the government level [85]—paints a positive picture for data access and usability in the coming years.

### (ii) Machine learning tools are becoming more accessible

This effort is also supported by a thriving ecosystem of data sources and novel tools. ML methods can often be more flexible and perform better than classical statistical methods, and can achieve a very high level of accuracy in many predictive and classification tasks in a relatively short amount of time (e.g. [86,87]). Increasing computing power combined with recent advances in ML techniques and applications shows



**Figure 3.** The nested nature of developing predictive and forecasting models, showcases the *forward problem* and how this relies on a hierarchical structure of the modelling process. The choice of a specific modelling technique and framework, as well as the data retained to be part of this model, proceeds directly from our assumptions about which ecological mechanisms are important in shaping both extant and future data.

promise in ecology and environmental science (see [88] for an overview). Moreover, ongoing developments in deep learning are aimed at improvement in low-data regimes and with unbalanced datasets [89,90]. Considering the current biases in network ecology [57] and the scarcity of data of species interactions, the prediction of ecological networks will undoubtedly benefit from these improvements. ML methods are emerging as the new standard in computational ecology in general [88,91], and in network ecology in particular [92], as long as sufficient, relevant data are available. Many studies have used ML models specifically with ecological interactions. Relevant examples include species traits used to predict interactions and infer trait-matching rules [93,94], automated discovery of food webs [95], reconstruction of ecological networks using next-generation sequencing data [92], and network inference from presence–absence data [96]. As many ecological and evolutionary processes underlie species interactions and the structure of their ecological networks (e.g. [68,97]), it can be difficult to choose relevant variables and model species interaction networks explicitly. A promising application of ML in natural sciences is scientific-ML, a framework that combines machine learning with mechanistic models [40,98].

## 4. A primer on predicting ecological networks

Within the constraints outlined in the previous section, we now provide a primer on the background concepts necessary to build predictive models of species interaction networks, with a focus on using ML approaches in the modelling process. As figure 2 illustrates, this involves a variety of numerical and computational approaches; therefore, rather than an exhaustive summary, we aim to convey a high-level understanding that translates the core concepts into their application to ecological networks.

### (a) Models

#### (i) What is a predictive model?

Models are used for many purposes, and the term ‘model’ itself embodies a wide variety of meanings in scientific discourse. All models can be thought of as a function,  $f$ , that takes a set of inputs  $x$  (also called features, descriptors or independent variables) and parameters  $\theta$  (called weights in the contents of neural networks), and maps them to predicted

output states  $y$  (also called label, response or dependent variable) based on the input to the model:  $y = f(x, \theta)$ .

A given model  $f$  can be used for either descriptive or predictive purposes. Many forms of scientific inquiry are based around using models *descriptively*, a practice also called inference, the inverse problem, fitting a model, or training a model [99]. In this context, the goal of using a model is to estimate the parameters,  $\theta$ , that best explain a set of empirical observations,  $\{\hat{x}, \hat{y}\}$ . In some cases, these parameter values are themselves of interest (e.g. the strength of selection, intrinsic growth rate, dispersal distance), but in others cases, the goal is to compare a set of competing models  $f_1, f_2, \dots$  to determine which provides the most parsimonious explanation for a dataset. The quantitative representation of ‘effects’ in these models—the influence of each input on the output—is often assumed to be linear, and within the frequentist worldview, the goal is often to determine if the coefficient corresponding with an input is non-zero to determine its ‘significance’ (often different from its ecological relevance; [100]) in influencing the outcome.

Models designed for inference have use—descriptive models of networks can reveal underlying mechanisms that structure ecological communities, given a proper null model [101]. However, in order for ecology to develop as a predictive science [102], interest has grown in developing models that are used not just for description of data, but also for prediction. Predictive models are based in the *forward problem*, where the aim is to predict new values of the output  $y$  given an input  $x$  and our estimate value of  $\theta$  [99]. Because the forward problem relies on an estimate of  $\theta$ , then, the problem of inference is nested within the forward problem (figure 3): working towards a predictive view of ecological networks will give us the needed tools to further our understanding of them.

#### (ii) What do you need to build a predictive model?

To build a predictive model, one needs the following: first, **data**, split into features  $\hat{x}$  and labels  $\hat{y}$  (figure 3). Second, a **model**  $f$ , which maps features  $x$  to labels  $y$  as a function of parameters  $\theta$ , i.e.  $y = f(x, \theta)$ . Third, a **loss function**  $L(\hat{y}, y)$ , which describes how far a model’s prediction  $y$  is from an empirical value  $\hat{y}$ . Lastly, **priors** on parameters,  $P(\theta)$ , which describe the modeller’s *a priori* belief about the value of the parameters; rather than making an analysis implicit, specifying priors has the merit of making the modeller’s

assumptions explicit, which is a most desirable feature when communicating predictions to stakeholders [103]. Often an important step before fitting a model is feature engineering: adjusting and reworking the features to better uncover feature-label relationships [104]. This can include projecting the features into a lower dimensional space, as we did through a probabilistic PCA in the case study, or removing the covariance structure using a Whitening approach. Then, when a model is fitted (synonymous with parameter inference or the inverse problem, see figure 3), a fitting algorithm attempts to estimate the values of  $\theta$  that minimizes the mean value of loss function  $L(\hat{y}, y)$  for all labels  $\hat{y}$  in the provided data  $Y$ . In a Bayesian approach, this typically relies on drawing candidate parameter values from priors and applying some form of sampling to generate a *posterior* estimate of parameters,  $P(\theta|\hat{x}, \hat{y})$ . In the training of neural networks, this usually involves some form of error back-propagation across the edges in order to tune their weights, and the biases of each nodes.

### (iii) How do we validate a predictive model?

After we fit a model, we inevitably want to see how ‘good’ (meaning, ‘fit for purpose’) it is. This process can be divided into two parts: (i) model selection, where the modeller chooses from a set of possible models, and (ii) model assessment, where the modeller determines the performance characteristics of the chosen model [105].

In the context of *model selection*, a naive initial approach is to simply compute the average error between the model’s prediction and the true data we have, and choose the model with the smallest error—however, this approach inevitably results in *overfitting*. One approach to avoid overfitting is using information criteria (e.g. Akaike information criteria, Bayesian information criteria, minimum description length) based around the heuristic that good models maximize the ratio of information provided by the model to the number of parameters it has. However, when the intended use-case of a model is prediction the relevant form of validation is *predictive accuracy*, which should be tested with *cross-validation*. Cross-validation methods divide the original dataset into two—one which is used to fit the model (called the *training* set) and one used to validate its predictive accuracy on the data that it hasn’t ‘seen’ yet (called the *test* set) [106]. This procedure is often repeated across different test and training subdivisions of the dataset (either picked randomly or stratified by some criteria, like balance between positive and negative interactions in the case study) to determine the uncertainty associated with our measurement owing to our choice of test and training sets [107], in the same conceptual vein as data bootstrapping: the mean value of the validation metric gives an overall estimate of its performance, and the variance around this mean represents the effect of using different data for training and testing. In a robust model/dataset combination, we expect this variance to be low, although there are no prescriptive guidelines as to how little variance is acceptable; the choice of whether to use a model is often left to the best judgement of the modeller.

We still have to define what *predictive accuracy* means in the context of interaction network prediction. In the proof-of-concept, we used a neural-network to perform binary classification by predicting the presence/absence of an interaction between any two species. There are two ways

for the model to be right: the model predicts an interaction and there is one (a *true positive* (TP)), or the model predicts no interaction and there isn’t one (a *true negative* (TN)). Similarly, there are two ways for the model to be wrong: the model predicts an interaction which does not exist (a *false positive* (FP)), or the model predicts no interaction and it does exist (a *false negative* (FN)).

A naive initial approach to measure how well a model does is *accuracy*, i.e. the proportion of values it got correct. However, consider what we know about interaction networks: they are often very sparse, with connectance usually below a third [108]. If we build a model that always guesses there will be no interaction between two species, it will be correct in the majority of cases because the majority of potential interactions in a network typically do not exist. Therefore this ‘empty-matrix’ model would always have an *accuracy* of  $1 - C$ , where  $C$  is the observed connectance, which would almost always be greater than 50%. Understanding model performance within sensitivity-specificity space may be more informative, where sensitivity evaluates how good the model is at predicting true interactions (TP rate) and specificity refers to the prediction of true ‘non-interactions’ (TN rate). It must be noted that in ecological networks, there is no guarantee that the ‘non-interactions’ (assumed TNs) in the original dataset are indeed TNs [27,55]. This can result in the positive/negative values, and the false omission/discovery being artificially worse, and specifically decrease our confidence in predicted interactions.

In response to the general problem of biases in classifiers, many metrics have been proposed to measure binary-classifiers [109,110] and are indicative of how well the model performs with regards to some aspect of accuracy, sensitivity, specificity and/or precision (table 1). Ultimately the choice of metric will depend on the intended use of the model: there is not a single definition of ‘success,’ but rather different interpretation of what sources of error are acceptable for a given application.

In the ML literature, a common way of visualizing this extensive list of possible metrics is through the use of receiver-operating-characteristic (ROC; FP rate on the  $x$ -axis, and TP rate on the  $y$ -axis) and precision-recall (PR; TP-rate on the  $x$ -axis, positive-predictive-value on the  $y$ -axis) curves (figure 1). These curves are generated by considering a continuum of thresholds of classifier acceptance, and computing the values of ROC/PR metrics for each value of the threshold. The area-under-the-curve (AUC) is then used as a validation metric and are typically called area-under-the-curve receiver-operator-curve (AUC-ROC) and area-under-the-curve precision-recall (AUC-PR) (e.g. ROC-AUC in table 1). These measures have the unstated assumption that the training and testing set are ‘correct,’ or at least correct enough that the number of true/false positive/negatives are meaningful; although should this assumption be true, there would be no need for any predictive approach—but it is a well established fact that ML systems are resilient—even relatively high uncertainties in the data [45].

## (b) Networks and interactions as predictable objects

### (i) Why predict networks and interactions at the same time?

Ecological networks are quite sparse, and larger networks tend to get sparser [111]; in other words, although networks are composed of a set of interactions between species pairs, they also form a much larger set of species pairs that do not interact. If we aim to predict the structure of networks

**Table 1.** Overview of the validation statistics applied to the case study, alongside the criteria indicating a successful classifier and a guide to interpretation of the values. (Taken together, these validation measures indicate that the model performs well, especially considering that it is trained from a small volume of data.)

| name                      | value | success    | description   |
|---------------------------|-------|------------|---|
| random accuracy           | 0.56  |            | fraction of correct predictions if the classifier is random   |
| accuracy                  | 0.81  | →1         | observed fraction of correct predictions                      |
| balanced accuracy         | 0.80  | →1         | average fraction of correct positive and negative predictions |
| true positive rate        | 0.77  | →1         | fraction of interactions predicted                            |
| true negative rate        | 0.83  | →1         | fraction of non-interactions predicted                        |
| false positive rate       | 0.16  | →0         | fraction of non-interactions predicted as interactions        |
| false negative rate       | 0.22  | →0         | fraction of interactions predicted as non-interactions        |
| ROC – AUC                 | 0.86  | →1         | proximity to a perfect prediction (ROC – AUC = 1)             |
| Youden's J                | 0.60  | →1         | informedness of predictions (trust in individual prediction)  |
| Cohen's $\kappa$          | 0.58  | $\geq 0.5$ |   |
| positive predictive value | 0.66  | →1         | confidence in predicted interactions                          |
| negative predictive value | 0.89  | →1         | confidence in predicted non-interactions                      |
| false omission rate       | 0.10  | →0         | expected proportion of missed interactions                    |
| false discovery rate      | 0.33  | →0         | expected proportion of wrongly imputed interactions           |

from the 'bottom-up'—by considering each pairwise combination of  $S$  different species—we are left with  $(S-1)^2$  interaction values to estimate, a majority of which will be 0. Instead, we can use our existing understanding of the mechanisms that structure ecological networks to whittle down the set of feasible adjacency matrices, thereby reducing the amount of information we must predict, and making the problem of predicting interactions less daunting. The processes that structure ecological networks do not only occur at the scale of interactions—there are also processes at the network level which limit what interactions (or how many) are realistic. The realized structure of a network is the synthesis of the interactions forming the basis for network structure, and the network structure refining the possible interactions—'Part makes whole, and whole makes part' [112].

Another argument for the joint prediction of networks and interactions is to reduce circularity and biases in the predictions. As an example, models like linear filtering [37] generate probabilities of non-observed interactions existing, but do so based on measured network properties. Some recent models make interaction-level predictions (e.g. [113]); these are not unlike stacked species distribution models, which are individually fit, but collectively outperformed by joint models or rule-based models [114]. By relying on adequate testing of model performance of biases (i.e. optimizing not only accuracy, but paying attention to measures like false discovery and false omission rates), and developing models around a feedback loop between network and interaction prediction, it is likely that the quality of the predicted networks will be greatly improved compared to current models.

## (ii) What network properties should we use to inform our predictions of interactions?

There are many dimensions of network structure [2], yet there are two arguments to support basing network prediction around a single property: *connectance* (the ratio of actual edges to possible edges in the network). First, connectance

is ecologically informative—it relates to resilience to invasion [115,116], can increase robustness to extinction in food webs [117], while decreasing it in mutualistic networks [118], and connectance relates to network stability [3]. Second, most (if not all) ecological network properties covary with connectance [117,119].

Within the network science literature, there are numerous methods for predicting edges based on network properties (e.g. block models [120] based on modularity, hierarchical models [121] based on embedding, etc.). However, in the context of species interaction networks, these properties often covary with connectance. As a result we suggest that using connectance as the primary property of interest is most likely to be practical to formulate at the moment. We have models to estimate species richness over space [122], and because we can predict connectance from species richness alone [111], we can then derive distributions of network properties from richness estimates, that can serve to penalize further models that formulate their predictions at the scale of each possible interaction.

## (iii) How do we predict how species that we have never observed together will interact?

A neutral approach to ecological interactions would assume the probability of an interaction mirrors the relative abundance of both species, and would be unaffected by trait variation [8,94]; more accurately, a neutral assumption states that the relative abundances are sufficient to predict the structure of networks, and this view is rather well supported in empirical and theoretical systems [123,124]. However, functional-trait based proxies could enable better predictions of ecological interactions [34,125–127]. Selection on functional traits could cause interactions to be conserved at some evolutionary scales, and therefore predictions of interaction could be informed by phylogenetic analyses [32,128,129]. Phylogenetic matching in bipartite networks is



consistent across scales [130], even in the absence of strong selective pressure [131].

A separate family of methods are based on network embedding (as in the proof-of-concept). A network embedding projects each node of the network into a lower-dimensional latent space. Previous explorations of the dimensionality of food webs have revealed that a reduced number of dimensions (7) was sufficient to capture most of their structure [132]; however, recent quantifications of the complexity of the embedding space of bipartite ecological networks found a consistent high complexity [133], suggesting that the precise depth of embedding required may vary considerably across systems. Embedding enables us to represent the structure of a network, which previously required the  $S^2$  dimensions of an adjacency matrix, with a smaller number of dimensions. The position of each node in this lower dimensional space is then treated as a latent measurement corresponding to the role of that species in the network (e.g. [51], where a network of about 1500 species was most accurately described using 12 dimensions). Species close together in the latent space should interact with a similar set of species [134,135]. However, these models are sensitive to sampling biases as they are limited to species for which there is already interaction data, and as a result a methodological breakthrough is needed to extend these models to species for which there is little or no interaction data.

#### (iv) How do we quantify interaction strength?

Species interaction networks can also be used as a means to quantify and understand *interaction strength*. Interaction strength, unlike the qualitative presence or absence of an interaction, is a continuous measurement which attempts to quantify the effect of one species on another. This results in weighted networks representing different patterns of ‘flows’ between nodes—which can be modelled in a variety of ways [136]. Interaction strength can generally be divided into two main categories (as suggested by [137]): (i) the strength of an interaction between individuals of each species, or (ii) the effect that changes in one species population has on the dynamics of the other species. It can be measured as the effect over a period of time (in the units of biomass or energy flux [137,138]) or the relative importance of one species on another [4,62,139]. One recurring observation is that networks are often composed of many weak interactions and few strong interactions [137]. The distribution of interaction strength within a network effects its stability [140,141] and functioning [142,143], and serves to benefit multi-species models [62]. Alternatively, understanding flow in modules within networks can aid in understanding the organization of networks [144,145] or the cascading effects of perturbations [146].

In some systems, quantifying interaction strength is relatively straightforward; this includes a lot of host–parasite systems. For example, freshwater cyprinid fishes can be divided in micro-habitats (fins, skin, digestive system, gill subsections) and the parasites counted in each of these micro-habitats, giving within-host resolution [147]; marine sparids and labrids have similarly been studied this way, see notably [148–150]. In some cases, within-host assessments of interaction strengths can reveal macro-ecological events, like in the conservatism of micro-habitat use in amphibian

hosts by helminths [151]. Even ectoparasites can provide reliable assessments of interaction strength; for example, when rodent hosts are minimally disturbed during capture, fine combing of their fur will result in exhaustive ectoparasites inventories [46,152–155]. Parasites have the desirable property of usually remaining intact within their host during the interaction, as opposed to prey items as can be recovered through e.g. gut content analysis or stable isotopes [156,157]. As network ecology is starting to explore the use of predictive models, leading up to forecasting, we argue that host–parasite systems can provide data that are reliable and trustworthy enough that they can become the foundations for methodological development and benchmark studies, thereby providing more information about host–parasite systems and supporting the technical development of the field.

However, in most situations, much like quantifying the occurrence of an interaction, quantifying interaction *strength* in the field is challenging in the majority of systems, and one must often rely on proxies. In some contexts, interaction strength can be estimated via functional foraging [158], where the primary basis for inferring interaction is foraging behaviour like searching, capture and handling times. In food-webs, metabolic-based models use body mass, metabolic demands, and energy loss to infer energy fluxes between organisms [159,160]. In addition, food-web energetics models can be incorporated at various resolutions for a specific network, ranging from individual-based data to more lumped data at the species level or trophic group, depending on data availability [138,159]. Taken together, these considerations impose too many constraints on predicting continuous interaction strength at the moment, resulting in our primary focus in binary present/absent interactions within this manuscript.

#### (v) How do we determine what interaction networks are feasible?

For several decades, ecologists have aimed to understand how networks of many interacting species persist through time. The diversity–stability paradox, first explored by May [161], shows that under a neutral set of assumptions ecological networks should become decreasingly stable as the number of species increases. Yet, in the natural world we observe networks of interactions that consist of far more species than May’s model predicts [162]. As a result, understanding what aspects of the neutral assumptions of May’s model are incorrect has branched many investigations into the relationship between ecological network structure and persistence [163]. These assumptions can be split into dynamical assumptions and topological assumptions. Topologically, we know that ecological networks are not structured randomly. Some properties, like the aforementioned connectance, are highly predictable [111]. Generative models of food-webs (based on network embeddings) fit empirical networks more effectively than random models [164]. These models have long used allometry as a single-dimensional niche space—naturally we want to extend this to traits in general. The second approach to stability is through *dynamics*. Early models of community dynamics rely on the assumption of linear interaction effects, but in recent years models of bioenergetic community dynamics have shown promise in basing our understanding of energy flow in food-webs in the understood relationship between allometry and metabolism [165]. An additional consideration is

the multidimensional nature of ‘stability’ and ‘feasibility’ (e.g. resilience to environmental change versus extinctions) [166] and how different disturbances propagate across levels of biological organization [167,168]. Recent approaches such as structural stability [169,170] allow us to think of network feasibility in rigorous mathematical terms, which may end up as usable parameters to penalize network predictions.

#### (vi) What taxonomic scales are suitable for the prediction of species interactions?

If we use different trait-based proxies to predict potential interactions between species, the choice of such proxies should be theoretically linked to the taxonomic and spatial scale we are using in our prediction [171]. At some scales, we can use morphological traits of co-occurring species to assess the probability of interaction between them [34]. On broader taxonomic scales, we can infer interaction probability through the phylogenetic distance, assuming that functional traits themselves are conserved [129]. In this case, we can think of the probability that one species will interact with another as the distance between them in niche-space [93], and this can be modelled by simulating neutral expectations of trait variation on phylogenetic trees [128]. At the narrowest scales, we may be interested in predicting behavioural traits like foraging behaviour [34], and at this scale we may need to consider abundance’s effect on the probability of an encounter [58].

#### (vii) What about indirect and higher-order interactions?

Although network ecology often assumes that interactions go strictly from one node to the other, the web of life is made up of a variety of interactions. Indirect interactions—either higher-order interactions between species, or interaction strengths that themselves interact—have gained interest in recent years [172,173]. One mathematical tool to describe these situations is hypergraphs: hypergraphs are the generalization of a graph, allowing a broad yet manageable approach to complex interactions [174], by allowing for particular interactions to occur beyond a pair of nodes. An additional degree of complexity is introduced by multi-layer networks [175]. Multi-layer networks include edges across ‘variants’ of the networks (timepoints, locations, or environments). These can be particularly useful to account for metacommunity structure [176], or to understand how dispersal can inform conservation action [177]. Ecological networks are intrinsically multi-layered [178]. However, *prima facie*, increasing the dimensionality of the object we need to predict (the multiple layers rather than a single network) makes the problem more complicated. Yet, multi-layer approaches improve prediction in social networks [179–181], and they may prove useful in network ecology going forward.

#### (c) Space

Although networks were initially used to describe the interactions *within* a community, interest in the last decade has shifted towards understanding their structure and variation over space [182,183], and has established network ecology as an important emerging component of biogeography and macroecology.

#### (i) How much do networks vary over space?

Networks can vary across space either in their structural properties (e.g. connectance or degree distribution) or in their composition (identity of nodes and edges). Interestingly, variation in the structural properties of ecological networks primarily responds to changes in the size of the network. The number of links in ecological networks scales with the number of species [111,184], and connectance and size drive the rest of network structure [117,119,185]. Species turnover in space results in changes in the composition of ecological networks. However, this is not the only reason network composition varies [8]. Intraspecific variation can result in interaction turnovers without changes in species composition [186]. Similarly, changes in species abundances can lead to variation in interaction strengths [123,187]. Variation in the abiotic environment and indirect interactions [173] could modify the occurrence and strength of individual interactions. Despite this, empirical networks tend to share a common backbone [188] and functional composition [189] across space.

#### (ii) How do we predict what the species pool at a particular location is?

As the species pool forms the basis for network structure, predicting which species are present at a particular location is essential to predict networks across space. Species distribution models (SDMs) are increasingly ubiquitous in macroecology—these models predict the range of a species based on known occurrences and environmental conditions, such as climate and land cover [190,191]. Including interactions or co-occurrences in SDMs generally improves predictive performance [192]. Several approaches exist to combine multiple SDMs: community assemblage at a particular site can be predicted either by combining independent single-species SDMs (stacked-SDMs, SSDMs) or by directly modelling the entire species assemblage and multiple species at the same time (joint SDMs; JSDMs) [193]. Building on the JSMD framework, hierarchical modelling of species communities [194] has the advantage of capturing processes that structure communities. Spatially explicit species assemblage modelling constrains SDM predictions using macroecological models [195]—for example, variation in species richness across space can constrain assemblage predictions [196].

The next step is to constrain distribution predictions using network properties. This builds on previous calls to adopt a probabilistic view: a probabilistic species pool [197], and probabilistic interactions through Bayesian networks [198]. Blanchet *et al.* [56] argue that the probabilistic view avoids confusion between interactions and co-occurrences, but that it requires prior knowledge of interactions. This could potentially be solved through our framework of predicting networks first, interactions next, and finally the realized species pool.

#### (iii) How do we combine spatial and network predictions?

In order to predict networks across space, we need to combine multiple models—one which predicts what the species pool will be at a given location, and one to predict what interaction networks composed from this species pool are likely to be (figure 2). Both of these models contain uncertainty, and when we combine them the uncertainty from each model should be propagated into the combined model. The

Bayesian paradigm provides a convenient solution to this—if we have a chain of models where each model feeds into the next, we can sample from the posterior of the input models. A different approach is *ensemble modelling* which combines the predictions made by several models, where each model is predicting the same thing [199]. Error propagation, an important step in building any ecological model, describes the effect of the uncertainty of input variables on the uncertainty of output variables [200,201]. Benke *et al.* [202] identifies two broad approaches to model error propagation: analytically using differential equations or stochastically using Monte-Carlo simulation methods. Errors induced by the spatial or temporal extrapolation of data also need to be taken into account when estimating the uncertainty of a model's output [203].

## (d) Time

### (i) Why should we forecast species interaction networks?

Forecasting species interactions are critical for informing ecosystem management [204] and systematic conservation prioritization [205], and for anticipating extinctions and their consequences [206,207]. Ecological interactions shape species distributions at both local and broad spatial scales, and including interactions in SDM models typically improves predictive performance [192,208,209]. However, these tend to rely on approaches involving estimating pairwise dependencies based on co-occurrence, using surrogates for biotic-interaction gradients, and hybridizing SDMs with dynamic models [192]. Most existing models to predict the future distribution of species ignore interactions [210]. Changes in species ranges and phenology will inevitably create spatio-temporal mismatches and affect encounter rates between species [211], which will further shift the distribution of species across space. New interactions will also appear between species that are not currently co-occurring [211]. Only by forecasting how species will interact can we hope to have an accurate portrait of how biodiversity will be distributed under the future climate.

Forecasting how climate change will alter biodiversity is also crucial for maximizing conservation outcomes. Improving SDMs through interactions is crucial for conservation, as nearly 30% of models in SDM studies are used to assess population declines or landscape ability to support populations [212]. Reliable predictions about how ecological networks will change over time will give us critical information that could be communicated to decision-makers and the scientific community about what future environmental risks we are awaiting and how to mitigate them [213]. Not only this, but how biodiversity is structured influences the functioning of the whole ecosystem, community stability and persistence [214,215]. Will climate change impact the distribution of network properties (e.g. connectance)? If so, which regions or species groups need special conservation efforts? These overarching questions are yet to be answered (but see [216–218]). We believe that the path towards forecasting ecological networks provides useful guidelines to ultimately better predict how climate change will affect the different dimensions of biodiversity and ecosystem functioning.

### (ii) How do we turn a predictive model into a forecasting model?

On some scales, empirical time-series encode enough information about ecological processes for ML approaches to

make accurate forecasts. However, there is an intrinsic limit to the predictability of ecological time series [219]. A forecast inherently has a *resolution limit* in space, time and organization. For example, one could never hope to predict the precise abundance of every species on Earth on every day hundreds of years into the future. There is often a trade-off between the resolution and horizon of forecast, e.g. a lower resolution forecast, like primary production will be at a maximum in the summer, is likely to be true much further into the future than a higher resolution forecast. If we want to forecast the structure of ecological networks beyond the forecasting horizon of time-series-based methods, we need forecasts of our predictive model's inputs—a forecast of the distribution of both environmental conditions and the potential species pool across space (figure 3).

### (iii) How can we validate a forecasting model?

Often the purpose of building a forecasting model is to inform *present* action [220]. Yet, the nature of forecasting—trying to predict the future—is that you can only know if a forecast is 'right' once it is too late to change it. If we want to maximize the chance that reality falls within a forecasting model's predictions, there are two directions to approach this problem: the first is to extend model validation techniques to a forecasting context, and the second is to attempt to maximize the amount of uncertainty in the forecast without compromising its resolution. Cross-validation (see *How do we validate a predictive model?*) can be used to test the efficacy of a forecasting model. Given a time-series of  $N$  observations, a model can iteratively be trained on the first  $n$  time-points of data, and the forecasting model's accuracy can be evaluated on the remaining time-points it has not 'seen' [106]. This enables us to understand both how much temporal data is required for a model to be robust, and also enables us to explore the *forecasting horizon* of a process. Furthermore, this approach can also be applied in the opposite temporal direction—if we have reliable data from the past, 'hindcasting' can also be used to test a forecast's robustness.

However, these methods inevitably bump into a hard-limitation on what is feasible for a forecasting model. The future is uncertain. Any empirical time-series we use to validate a model was collected in past conditions that may not persist into the future. Any system we wish to forecast will undergo only one of many possible scenarios, yet we can only observe the realized outcome of the system under the scenario that actually unfolds. It is therefore impossible to assess the quality of a forecasting model in scenarios that remain hypothetical. If the goal is to maximise the probability that reality will fall within the forecast's estimates, forecasts should incorporate as much uncertainty about the future scenario as possible—one way to do this is ensemble modelling [199]. However, as we increase the amount of uncertainty we incorporate into a forecasting model, the resolution of the forecast's predictions could shrink [221], and therefore the modeller should be mindful of the trade-off between resolution and accuracy when developing any forecast. Finally, ensemble models are not guaranteed to give more accurate results: for example, Becker *et al.* [16] noted that the ensemble model outperforms the best-in-class models, which should be taken as an indication that careful model building and selection is of the utmost importance when dealing with a problem as complex as the prediction of species interactions.

## 5. Conclusion: why should we predict species interaction networks?

Because we almost can, and because we definitely should.

A better understanding of species interactions, and the networks they form, would help unify the fields of community, network, and spatial ecology; improve the quantification of the functional relationships between species [222,223]; re-evaluate metacommunities in light of network structure [224]; and enable a new line of research into the biogeography of species interactions [225,226] which incorporates a synthesis of both Eltonian and Grinnellian niche [113]. Furthermore, the ability to reliably predict and forecast species interactions would inform conservation efforts for protecting species, communities, and ecosystems. Integration of species interactions into the assessment of vulnerability to climate change is a needed methodological advancement [227]. International panels draw on models to establish scientific consensus [212], and they can be improved through more effective prediction of species distributions and interactions [228]. Furthermore, recent studies argue for a shift in focus from species to interaction networks for biodiversity conservation to better understand ecosystem processes [204].

We should invest in network prediction because the right conditions to do so reliably and rapidly are beginning to emerge. Given the possible benefits to a variety of ecological disciplines that would result from an increased ability to predict networks, we feel strongly that the research agenda we outline here should be picked up by the community. Although novel technologies are bringing massive amounts of data to some parts of ecology (primarily environmental DNA and remote sensing, but now more commonly image analysis and bioacoustics), it is even more important to be intentional about *reconciling* data. This involves not only the work of understanding the processes encoded within data, but also the groundwork of developing pipelines to bridge the ever-expanding gap between 'high-throughput' and 'low-throughput' sampling methods. An overall increase in the volume of data will not result in an increase of our predictive capacity as long as this data increase is limited to specific aspects of the problem. In the areas, we highlight in figure 2, many data steps are still limiting: documenting empirical interactions is natural history work that does not

lend itself to systematic automation; expert knowledge is by design a social process that may be slightly accelerated by text mining and natural language processing (but is not yet, or not routinely, or at scale). These limitations are affecting our ability to reconstruct networks.

But the tools to which we feed these data, incomplete as they may be, are gradually getting better; that is, they can do predictions faster, they handle uncertainty and propagate it well, and they can accommodate data volumes that are lower than we may expect [94]. It is clear attempting to predict the structure of ecological networks at any scale is a methodological and ecological challenge; yet it will result in qualitative changes in our understanding of complex adaptive systems, as well as changes to our ability to use information about network structure for conservation decisions. It is perhaps even more important to forecast the structure of ecological networks because it is commonly neglected as a facet of biodiversity that can (and should) be managed. In fact, none of the Aichi targets mention biostructure or its protection, despite this being recognized as an important task [14], either implicitly or explicitly. Being able to generate reliable datasets on networks in space or time will make this information more actionable.

**Data accessibility.** This article has no additional data.

**Authors' contributions.** All authors contributed to the drafting, writing and editing of the manuscript.

**Competing interests.** We declare we have no competing interests.

**Funding.** T.S., N.R.F.-M. and T.P. are funded by a donation from the Courtois Foundation; F.B., N.R.F.-M. and T.P. are funded by IVADO; B.M. is funded by the NSERC Alexander Graham Bell Canada Graduate Scholarship and the FRQNT master's scholarship; F.B., G.D., N.R.F.-M. and G.H. are funded by the NSERC BIOS<sup>2</sup> CREATE program; G.D. is funded by the FRQNT doctoral scholarship; D.C., T.S., L.P. and T.P. are funded by the Canadian Institute of Ecology & Evolution; this research was enabled in part by support provided by Calcul Québec ([www.calculquebec.ca](http://www.calculquebec.ca)) and Compute Canada ([www.computeCanada.ca](http://www.computeCanada.ca)). This work was supported by funding to the Viral Emergence Research Initiative (VERENA) consortium including NSF BII 2021909. A.G. and M.D.C. are supported in part by the Liber Ero Chair.

**Acknowledgements.** We acknowledge that this study was conducted on land within the traditional unceded territory of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat and Omàmiwininiwak nations.

## References

- Albrecht J *et al.* 2018 Plant and animal functional diversity drive mutualistic network assembly across an elevational gradient. *Nat. Commun.* **9**, 3177. (doi:10.1038/s41467-018-05610-w)
- Delmas E *et al.* 2018 Analysing ecological networks of species interactions. *Biol. Rev.* **94**, 112540. (doi:10.1111/brv.12433)
- Landi P, Minoarivelo HO, Brönström Å, Hui C, Dieckmann U. 2018 Complexity and stability of ecological networks: a review of the theory. *Popul. Ecol.* **60**, 319–345. (doi:10.1007/s10144-018-0628-3)
- Heleno R *et al.* 2014 Ecological networks: delving into the architecture of biodiversity. *Biol. Lett.* **10**, 20131000. (doi:10.1098/rsbl.2013.1000)
- Pascual M, Dunne JA 2006 *Ecological networks: linking structure to dynamics in food webs*. Oxford, UK: Oxford University Press.
- Hortal J, de Bello F, Diniz-Filho JA, Lewinsohn TM, Lobo JM, Ladle RJ. 2015 Seven shortfalls that beset large-scale knowledge of biodiversity. *Annu. Rev. Ecol. Evol. Syst.* **46**, 523–549. (doi:10.1146/annurev-ecolsys-112414-054400)
- de Aguiar MA *et al.* 2019 Revealing biases in the sampling of ecological interaction networks. *PeerJ* **7**, e7566. (doi:10.7717/peerj.7566)
- Poisot T, Stouffer DB, Gravel D. 2015 Beyond species: why ecological interaction networks vary through space and time. *Oikos* **124**, 243–251. (doi:10.1111/oik.01719)
- Smith GR, Finlay RD, Stenlid J, Vasaitis R, Menkis A. 2017 Growing evidence for facultative biotrophy in saprotrophic fungi: data from microcosm tests with 201 species of wood-decay basidiomycetes. *New Phytol.* **215**, 747–755. (doi:10.1111/nph.14551)
- Luong LT, Mathot KJ. 2019 Facultative parasites as evolutionary stepping-stones towards parasitic lifestyles. *Biol. Lett.* **15**, 20190058. (doi:10.1098/rsbl.2019.0058)
- Poisot T, Stanko M, Miklisová D, Morand S. 2013 Facultative and obligate parasite communities exhibit different network properties. *Parasitology* **140**, 1340–1345. (doi:10.1017/S0031182013000851)
- Carpenter SR. 2002 Ecological futures: building an ecology of the long now. *Ecology* **83**, 2069–2083.

13. Burkle LA, Marlin JC, Knight TM. 2013 Plant-pollinator interactions over 120 years: loss of species, co-occurrence, and function. *Science* **339**, 1611–1615. (doi:10.1126/science.1232728)
14. McCann K. 2007 Protecting biostructure. *Nature* **446**, 29. (doi:10.1038/446029a)
15. Seibold S, Cadotte MW, MacIvor JS, Thorn S, Müller J. 2018 The necessity of multitrophic approaches in community ecology. *Trends Ecol. Evol.* **33**, 754–764. (doi:10.1016/j.tree.2018.07.001)
16. Becker DJ *et al.* 2020 Predicting wildlife hosts of betacoronaviruses for SARS-COV-2 sampling prioritization: a modeling study. *bioRxiv* (doi:10.1101/2020.05.22.111344)
17. Wardeh M, Baylis M, Blagrove MSC. 2021 Predicting mammalian hosts in which novel coronaviruses can be generated. *Nat. Commun.* **12**, 780. (doi:10.1038/s41467-021-21034-5)
18. Zhao N *et al.* 2020 Machine learning and dengue forecasting: comparing random forests and artificial neural networks for predicting dengue burden at national and sub-national scales in Colombia. Edited by Marc Choisy. *PLoS Negl. Trop. Dis.* **14**, e0008056. (doi:10.1371/journal.pntd.0008056)
19. Rengifo-Correa L, Stephens CR, Morrone JJ, Téllez-Rendón JL, Gonzalez-Salazar C. 2017 Understanding transmissibility patterns of chagas disease through complex vector-host networks. *Parasitology* **144**, 760. (doi:10.1017/S0031182016002468)
20. Morand S, Chairisi K, Kritiyakan A, Kumlert R. 2020 Disease ecology of rickettsial species: a data science approach. *Trop. Med. Infect. Dis.* **5**, 64. (doi:10.3390/tropicalmed5020064)
21. Stephens CR. 2009 Using biotic interaction networks for prediction in biodiversity and emerging diseases. *PLoS ONE* **4**, e5725. (doi:10.1371/journal.pone.0005725)
22. Craft ME. 2015 Infectious disease transmission and contact networks in wildlife and livestock. *Phil. Trans. R. Soc. B* **370**, 20140107. (doi:10.1098/rstb.2014.0107)
23. Edwards FA, Edwards DP, Hamer KC, Fayle TM. 2021 Tropical land-use change alters trait-based community assembly rules for dung beetles and birds. *Oecologia* **195**, 705–717. (doi:10.1007/s00442-020-04829-z)
24. Magioli M, de Barros Ferraz KMPM. 2021 Deforestation leads to prey shrinkage for an apex predator in a biodiversity hotspot. *Mammal Res.* **66**, 245–255. (doi:10.1007/s13364-021-00556-9)
25. Zhang M, He F. 2021 Plant breeding systems influence the seasonal dynamics of plant-pollinator networks in a subtropical forest. *Oecologia* **195**, 751–758. (doi:10.1007/s00442-021-04863-5)
26. Gibb R *et al.* 2021 Data proliferation, reconciliation, and synthesis in viral ecology. *bioRxiv* (doi:10.1101/2021.01.14.426572)
27. Jordano P. 2016 Chasing ecological interactions. *PLoS Biol.* **14**, e1002559. (doi:10.1371/journal.pbio.1002559)
28. Morales-Castilla I, Matias MG, Gravel D, Araújo MB. 2015 Inferring biotic interactions from proxies. *Trends Ecol. Evol.* **30**, 347–356. (doi:10.1016/j.tree.2015.03.014)
29. Pocock MJO, Evans DM, Memmott J. 2012 The robustness and restoration of a network of ecological networks. *Science* **335**, 973–977. (doi:10.1126/science.1214915)
30. Gravel D, Poisot T, Albouy C, Velez L, Mouillot D. 2013 Inferring food web structure from predator-prey body size relationships. *Methods Ecol. Evol.* **4**, 1083–1090. (doi:10.1111/2041-210X.12103)
31. Petchev OL, Beckerman AP, Riede JO, Warren PH. 2008 Size, foraging, and food web structure. *Proc. Natl Acad. Sci. USA* **105**, 4191–4196. (doi:10.1073/pnas.0710672105)
32. Elmasri M, Farrell MJ, Davies TJ, Stephens DA. 2020 A hierarchical Bayesian model for predicting ecological interactions using scaled evolutionary relationships. *Ann. Appl. Stat.* **14**, 221–240. (doi:10.1214/19-AOAS1296)
33. Pomeranz JPF, Thompson RM, Poisot T, Harding JS. 2018 Inferring predator-prey interactions in food webs. *Methods Ecol. Evol.* **10**, 356–367. (doi:10.1111/2041-210X.13125)
34. Bartomeus I, Gravel D, Tylianakis JM, Aizen MA, Dickie IA, Bernard-Verdier M. 2016 A common framework for identifying linkage rules across different types of interactions. *Funct. Ecol.* **30**, 1894–1903. (doi:10.1111/1365-2435.12666)
35. Vázquez DP, Morris WF, Jordano P. 2005 Interaction frequency as a surrogate for the total effect of animal mutualists on plants. *Ecol. Lett.* **8**, 1088–1094. (doi:10.1111/j.1461-0248.2005.00810.x)
36. Weinstein BG, Graham CH. 2017 On comparing traits and abundance for predicting species interactions with imperfect detection. *Food Webs* **11**, 17–25. (doi:10.1016/j.fooweb.2017.05.002)
37. Stock M, Poisot T, Waegeman W, De Baets B. 2017 Linear filtering reveals false negatives in species interaction data. *Sci. Rep.* **7**, 45908. (doi:10.1038/srep45908)
38. Terry JCD, Lewis OT. 2020 Finding missing links in interaction networks. *Ecology* **101**, e03047. (doi:10.1002/ecy.3047)
39. Desjardins-Proulx P, Poisot T, Gravel D. 2019 Artificial intelligence for ecological and evolutionary synthesis. *Front. Ecol. Evol.* **7**, 402. (doi:10.3389/fevo.2019.00402)
40. Rackauckas C, Ma Y, Martensen J, Warner C, Zubov K, Supekar R, Skinner D, Ramadhan A, Edelman A. 2020 Universal differential equations for scientific machine learning. *arXiv* 2001.04385.
41. Burkle LA, Alarcon R. 2011 The future of plant-pollinator diversity: understanding interaction networks across time, space, and global change. *Am. J. Bot.* **98**, 528–538. (doi:10.3732/ajb.1000391)
42. Houlahan JE, McKinney ST, Anderson TM, McGill BJ. 2017 The priority of prediction in ecological understanding. *Oikos* **126**, 1–7. (doi:10.1111/oik.03726)
43. Maris V, Huneman P, Coreau A, Kéfi S, Pradel R, Devictor V. 2017 Prediction in ecology: promises, obstacles and clarifications. *Oikos* **127**, 171–183. (doi:10.1111/oik.04655)
44. Borowiec ML, Frandsen P, Dikow R, McKeeken A, Valentini G, White AE. 2021 Deep learning as a tool for ecology and evolution. (doi:10.32942/osf.io/n3as)
45. Halevy A, Norvig P, Pereira F. 2009 The unreasonable effectiveness of data. *IEEE Intell. Syst.* **24**, 8–12. (doi:10.1109/MIS.2009.36)
46. Hadfield JD, Krasnov BR, Poulin R, Nakagawa S. 2014 A tale of two phylogenies: comparative analyses of ecological interactions. *Am. Nat.* **183**, 174–187. (doi:10.1086/674445)
47. Dunne J. 2006 The network structure of food webs. In *Ecological networks: linking structure to dynamics in food webs* (eds M Pascual, JA Dunne), pp. 27–86. Oxford, UK: Oxford University Press.
48. Bezanson J, Edelman A, Karpinski S, Shah V. 2017 Julia: a fresh approach to numerical computing. *SIAM Rev.* **59**, 65–98. (doi:10.1137/141000671)
49. Innes M. 2018 Flux: elegant machine learning with Julia. *J. Open Source Softw.* **3**, 602. (doi:10.21105/joss.00602)
50. Tipping ME, Bishop CM. 1999 Probabilistic principal component analysis. *J. R. Stat. Soc. B (Stat. Methodol.)* **61**, 611–622. (doi:10.1111/1467-9868.00196)
51. Poisot T *et al.* 2021 Imputing the mammalian virome with linear filtering and singular value decomposition. *arXiv* 2105.14973 (<http://arxiv.org/abs/2105.14973>)
52. Lemaître G, Nogueira F, Aridas CK. 2017 Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. *J. Mach. Learn. Res.* **18**, 1–5.
53. Krasnov BR, Shenbrot GI, Khokhlova IS, Degen AA. 2016 Trait-based and phylogenetic associations between parasites and their hosts: a case study with small mammals and fleas in the palearctic. *Oikos* **125**, 29–38. (doi:10.1111/oik.02178)
54. Bennett AE, Evans DM, Powell JR. 2019 Potentials and pitfalls in the analysis of bipartite networks to understand plant-microbe interactions in changing environments. *Funct. Ecol.* **33**, 107–117. (doi:10.1111/1365-2435.13223)
55. Jordano P. 2016 Sampling networks of ecological interactions. *Funct. Ecol.* **30**, 1883–1893. (doi:10.1111/1365-2435.12763)
56. Blanchet FG, Cazelles K, Gravel D. 2020 Co-occurrence is not evidence of ecological interactions. *Ecol. Lett.* **23**, 1050–1063. (doi:10.1111/ele.13525)
57. Poisot T, Bergeron G, Cazelles K, Dallas T, Gravel D, MacDonald A, Mercier B, Violet C, Vissault S. 2021 Global knowledge gaps in species interaction networks data. *J. Biogeogr.* **48**, jbi.14127. (doi:10.1111/jbi.14127)
58. Wells K, O'Hara RB. 2013 Species interactions: estimating per-individual interaction strength and covariates before simplifying data into per-species ecological networks. *Methods Ecol. Evol.* **4**, 1–8. (doi:10.1111/j.2041-210x.2012.00249.x)
59. Novak M, Wootton JT. 2008 Estimating nonlinear interaction strengths: an observation-based method

- for species-rich food webs. *Ecology* **89**, 2083–89. (doi:10.1890/08-0033.1)
60. Sala E, Graham MH. 2002 Community-wide distribution of predator-prey interaction strength in kelp forests. *Proc. Natl Acad. Sci. USA* **99**, 3678–3683. (doi:10.1073/pnas.052028499)
61. Wootton JT. 1997 Estimates and tests of per capita interaction strength: diet, abundance, and impact of intertidally foraging birds. *Ecol. Monogr.* **67**, 45–64. (doi:10.1890/0012-9615(1997)067%5B0045:EAOPC%5D2.0.CO;2)
62. Wootton JT, Emmerson M. 2005 Measurement of interaction strength in nature. *Annu. Rev. Ecol. Evol. Syst.* **36**, 419–444. (doi:10.1146/annurev.ecolsys.36.091704.175535)
63. Ings TC *et al.* 2009 Ecological networks—beyond food webs. *J. Anim. Ecol.* **78**, 253–269. (doi:10.1111/j.1365-2656.2008.01460.x)
64. Poisot T, Bergeron G, Cazelles K, Dallas T, Gravel D, Macdonald A, Mercier B, Violet C, Vissault S. 2020 Environmental biases in the study of ecological networks at the planetary scale. *bioRxiv*. (doi:10.1101/2020.01.27.921429)
65. Fontaine C, Guimarães PR, Kéfi S, Loeuille N, Memmott J, van der Putten WH, van Veen FJF, Thébaud E. 2011 The ecological and evolutionary implications of merging different types of networks. *Ecol. Lett.* **14**, 1170–1181. (doi:10.1111/j.1461-0248.2011.01688.x)
66. Michalska-Smith MJ, Allesina S. 2019 Telling ecological networks apart by their structure: a computational challenge. *PLoS Comput. Biol.* **15**, e1007076. (doi:10.1371/journal.pcbi.1007076)
67. Cazelles K, Araújo MB, Mouquet N, Gravel D. 2015 A theory for species co-occurrence in interaction networks. *Theor. Ecol.* **9**, 39–48. (doi:10.1007/s12080-015-0281-9)
68. Vazquez DP, Bluethgen N, Cagnolo L, Chacoff NP. 2009 Uniting pattern and process in plant-animal mutualistic networks: a review. *Ann. Bot.* **103**, 1445–1457. (doi:10.1093/aob/mcp057)
69. Araújo MS, Guimarães Jr PR, Svanbäck R, Pinheiro A, Guimarães P, Reis SF, Bolnick DI. 2008 Network analysis reveals contrasting effects of intraspecific competition on individual vs. population diets. *Ecology* **89**, 1981–1993. (doi:10.1890/07-0630.1)
70. Tinker M, Guimarães Jr PR, Novak M, Marquitti FM, Bodkin JL, Staedler M, Bentall G, Estes JA. 2012 Structure and mechanism of diet specialisation: testing models of individual variation in resource use with sea otters. *Ecol. Lett.* **15**, 475–483. (doi:10.1111/j.1461-0248.2012.01760.x)
71. Guimarães PR. 2020 The structure of ecological networks across levels of organization. *Annu. Rev. Ecol. Evol. Syst.* **51**, 433–460. (doi:10.1146/annurev-ecolsys-012220-120819)
72. Makiola A *et al.* 2020 Key questions for next-generation biomonitoring. *Front. Environ. Sci.* **7**, 197. (doi:10.3389/fenvs.2019.00197)
73. Levin SA. 1998 Ecosystems and the biosphere as complex adaptive systems. *Ecosystems* **1**, 431–436. (doi:10.1007/s100219900037)
74. Dickinson JL, Zuckerman B, Bonter DN. 2010 Citizen science as an ecological research tool: challenges and benefits. *Annu. Rev. Ecol. Evol. Syst.* **41**, 149–172. (doi:10.1146/annurev-ecolsys-102209-144636)
75. Stephenson PJ. 2020 Technological advances in biodiversity monitoring: applicability, opportunities and challenges. *Curr. Opin. Environ. Sustain.* **45**, 36–41. (doi:10.1016/j.cosust.2020.08.005)
76. Porter J *et al.* 2005 Wireless sensor networks for ecology. *BioScience* **55**, 561–572. (doi:10.1641/0006-3568(2005)055%5B0561:WSNFE%5D2.0.CO;2)
77. Creer S, Deiner K, Frey S, Porazinska D, Taberlet P, Thomas WK, Potter C, Bik HM. 2016 The ecologist's field guide to sequence-based identification of biodiversity. *Methods Ecol. Evol.* **7**, 1008–1018. (doi:10.1111/2041-210X.12574)
78. Lausch A *et al.* 2016 Linking earth observation and taxonomic, structural and functional biodiversity: local to ecosystem perspectives. *Ecol. Indic.* **70**, 317–339. (doi:10.1016/j.ecolind.2016.06.022)
79. Skidmore AK, Pettorelli N. 2015 Agree on biodiversity metrics to track from space: ecologists and space agencies must forge a global monitoring strategy. *Nature* **523**, 403–406. (doi:10.1038/523403a)
80. Jeliakova A *et al.* 2020 A global database for metacommunity ecology, integrating species, traits, environment and space. *Sci. Data* **7**, 6. (doi:10.1038/s41597-019-0344-7)
81. Poisot T *et al.* 2016 Mangal making ecological network analysis simple. *Ecography* **39**, 384–90. (doi:10.1111/ecog.00976)
82. Poelen JH, Simons JD, Mungall CJ. 2014 Global biotic interactions: an open infrastructure to share and analyze species-interaction datasets. *Ecol. Inf.* **24**, 148–159. (doi:10.1016/j.ecoinf.2014.08.005)
83. Kissling WD *et al.* 2018 Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. *Biol. Rev.* **93**, 600–625. (doi:10.1111/brv.12359)
84. Pérez-Harguindeguy N *et al.* 2013 New handbook for standardised measurement of plant functional traits worldwide. *Aust. J. Bot.* **61**, 167. (doi:10.1071/BT12225)
85. Scholes RJ *et al.* 2012 Building a global observing system for biodiversity. *Curr. Opin. Environ. Sustain.* **4**, 139–146. (doi:10.1016/j.cosust.2011.12.005)
86. Cutler DR, Edwards Jr TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ. 2007 Random forests for classification in ecology. *Ecology* **88**, 2783–2792. (doi:10.1890/07-0539.1)
87. Krizhevsky A, Sutskever I, Hinton GE. 2017 ImageNet classification with deep convolutional neural networks. *Commun. ACM* **60**, 84–90. (doi:10.1145/3065386)
88. Christin S. 2019 Applications for deep learning in ecology. *Methods Ecol. Evol.* **10**, 1632–1644. (doi:10.1111/2041-210X.13256)
89. Antoniou A, Storkey A, Edwards H. 2018 Data augmentation generative adversarial networks. (<http://arxiv.org/abs/1711.04340>) [cs, Stat], March. <http://arxiv.org/abs/1711.04340>.
90. Chawla NV. 2010 Data mining for imbalanced datasets: an overview. In *Data mining and knowledge discovery handbook* (eds O Maimon, L Rokach), pp. 875–86. Boston, MA: Springer. (doi:10.1007/978-0-387-09823-4\_45)
91. Olden JD, Lawler JJ, LeRoy Poff N. 2008 Machine learning methods without tears: a primer for ecologists. *Q. Rev. Biol.* **83**, 171–193. (doi:10.1086/587826)
92. Bohan DA, Vacher C, Tamaddon-Nezhad A, Raybould A, Dumbrell AJ, Woodward G. 2017 Next-generation global biomonitoring: large-scale, automated reconstruction of ecological networks. *Trends Ecol. Evol.* **32**, 477–487. (doi:10.1016/j.tree.2017.03.001)
93. Desjardins-Proulx P, Laigle I, Poisot T, Gravel D. 2017 Ecological interactions and the Netflix problem. *PeerJ* **5**, e3644. (doi:10.7717/peerj.3644)
94. Pichler M, Boreux V, Klein A-M, Schleuning M, Hartig F. 2020 Machine learning algorithms to infer trait-matching and predict species interactions in ecological networks. *Methods Ecol. Evol.* **11**, 281–293. (doi:10.1111/2041-210X.13329)
95. Bohan DA, Caron-Lormier G, Muggleton S, Raybould A, Tamaddon-Nezhad A. 2011 Automated discovery of food webs from ecological data using logic-based machine learning. *PLoS ONE* **6**, e29028. (doi:10.1371/journal.pone.0029028)
96. Sander EL, Wootton JT, Allesina S. 2017 Ecological network inference from long-term presence-absence data. *Sci. Rep.* **7**, 7154. (doi:10.1038/s41598-017-07009-x)
97. Segar ST, Fayle TM, Srivastava DS, Lewinsohn TM, Lewis OT, Novotny V, Kitching RL, Maunsell SC. 2020 The role of evolution in shaping ecological networks. *Trends Ecol. Evol.* **35**, 454–66. (doi:10.1016/j.tree.2020.01.004)
98. Chuang KV, Keiser MJ. 2018 Adversarial controls for scientific machine learning. *ACS Chem. Biol.* **13**, 2819–2821. (doi:10.1021/acscchembio.8b00881)
99. Stouffer DB. 2019 All ecological models are wrong, but some are useful. *J. Anim. Ecol.* **88**, 192–195. (doi:10.1111/1365-2656.12949)
100. Martínez-Abraín A. 2008 Statistical significance and biological relevance: a call for a more cautious interpretation of results in ecology. *Acta Oecologica* **34**, 9–11. (doi:10.1016/j.actao.2008.02.004)
101. Connor N, Barberán A, Clauset A. 2017 Using null models to infer microbial co-occurrence networks. *PLoS ONE* **12**, e0176751. (doi:10.1371/journal.pone.0176751)
102. Evans MR, Norris EJ, Benton TG. 2012 Predictive ecology: systems approaches. *Phil. Trans. R. Soc. B* **367**, 163–169. (doi:10.1098/rstb.2011.0191)
103. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. 2000 Bayesian methods in health technology assessment: a review. *Health Tech. Assess.* **4**, no. 38. York, UK: York Publishing Services.
104. Kuhn M, Johnson K. 2019 *Feature engineering and selection: a practical approach for predictive models*. London, UK: Chapman and Hall/CRC.

105. Hastie T, Tibshirani R, Friedman JH 2009 *The elements of statistical learning: data mining, inference, and prediction*, 2nd edn. Springer Series in Statistics. New York, NY: Springer.
106. Bishop CM 2006 *Pattern recognition and machine learning. Information Science and Statistics*. New York, NY: Springer.
107. Arlot S, Celisse A. 2010 A survey of cross-validation procedures for model selection. *Stat. Surv.* **4**, 40–79. (doi:10.1214/09-SS054)
108. Cohen JE, Briand F, Newman C 1990 *Community food webs: data and theory. Biomathematics*. Berlin, Heidelberg, Germany: Springer.
109. Drummond C, Holte RC. 2006 Cost curves: an improved method for visualizing classifier performance. *Mach. Learn.* **65**, 95–130. (doi:10.1007/s10994-006-8199-5)
110. Gu Q, Zhu L, Cai Z. 2009 Evaluation measures of the classification performance of imbalanced data sets. In *Computational intelligence and intelligent systems* (eds Zhihua Cai, Zhenhua Li, Zhuo Kang, Yong Liu), vol. 51, pp. 461–471. Berlin, Heidelberg, Germany: Springer. (doi:10.1007/978-3-642-04962-0\_53)
111. MacDonald AAM, Banville F, Poisot T. 2020 Revisiting the links-species scaling relationship in food webs. *Patterns* **1**, 100079. (doi:10.1016/j.patter.2020.100079)
112. Levins R, Lewontin RC 1987 *The dialectical biologist*. Cambridge, MA: Harvard University Press.
113. Gravel D *et al.* 2019 Bringing Elton and Grinnell together: a quantitative framework to represent the biogeography of ecological interaction networks. *Ecography* **42**, 401–415. (doi:10.1111/ecog.04006)
114. Zurell D, Zimmermann NE, Gross H, Baltensweiler A, Sattler T, Wüest RO. 2020 Testing species assemblage predictions from stacked and joint species distribution models. *J. Biogeogr.* **47**, 101–113. (doi:10.1111/jbi.13608)
115. Baiser B, Russell GJ, Lockwood JL. 2010 Connectance determines invasion success via trophic interactions in model food webs. *Oikos* **119**, 1970–1976. (doi:10.1111/j.1600-0706.2010.18557.x)
116. Smith-Ramesh LM, Moore AC, Schmitz OJ. 2016 Global synthesis suggests that food web connectance correlates to invasion resistance. *Glob. Change Biol.* **23**, 465–473. (doi:10.1111/gcb.13460)
117. Dunne JA, Williams RJ, Martinez ND. 2002 Food-web structure and network theory: the role of connectance and size. *Proc. Natl Acad. Sci. USA* **99**, 12 917–12 922. (doi:10.1073/pnas.192407699)
118. Vieira MC, Almeida-Neto M. 2015 A simple stochastic model for complex coextinctions in mutualistic networks: robustness decreases with connectance. *Ecol. Lett.* **18**, 144–152. (doi:10.1111/ele.12394)
119. Poisot T, Gravel D. 2014 When is an ecological network complex? Connectance drives degree distribution and emerging network properties. *PeerJ* **2**, e251. (doi:10.7717/peerj.251)
120. Yen T-C, Larremore DB. 2020 Community detection in bipartite networks with stochastic blockmodels. *Phys. Rev. E* **102**, 032309. (doi:10.1103/PhysRevE.102.032309)
121. Kawakatsu M, Chodrow PS, Eikmeier N, Larremore DB. 2021 Emergence of hierarchy in networked endorsement dynamics. *arXiv* 2007.04448 (<http://arxiv.org/abs/2007.04448>)
122. Jenkins CN, Pimm SL, Joppa LN. 2013 Global patterns of terrestrial vertebrate diversity and conservation. *Proc. Natl Acad. Sci. USA* **110**, E2602–E2610. (doi:10.1073/pnas.1302251110)
123. Canard EF, Mouquet N, Mouillot D, Stanko M, Miklisova D, Gravel D. 2014 Empirical evaluation of neutral interactions in host-parasite networks. *Am. Nat.* **183**, 468–479. (doi:10.1086/675363)
124. Canard E, Mouquet N, Marescot L, Gaston KJ, Gravel D, Mouillot D. 2012 Emergence of structural patterns in neutral trophic networks. *PLoS ONE* **7**, e38295. (doi:10.1371/journal.pone.0038295)
125. Bartomeus I. 2013 Understanding linkage rules in plant-pollinator networks by using hierarchical models that incorporate pollinator detectability and plant traits. *PLoS ONE* **8**, e69200. (doi:10.1371/journal.pone.0069200)
126. Cirtwill AR, Ekf A, Roslin T, Wootton K, Gravel D. 2019 A quantitative framework for investigating the reliability of empirical network construction. *Methods Ecol. Evol.* **10**, 902–911. (doi:10.1111/2041-210X.13180)
127. Cirtwill AR, Eklöf A. 2018 Feeding environment and other traits shape species' roles in marine food webs. *Ecol. Lett.* **21**, 875–884. (doi:10.1111/ele.12955)
128. Davies TJ. 2021 Ecophylogenetics redux. *Ecol. Lett.* **24**, 1073–1088. (doi:10.1111/ele.13682)
129. Gómez JM, Verdú M, Perfectti F. 2010 Ecological interactions are evolutionarily conserved across the entire tree of life. *Nature* **465**, 918–921. (doi:10.1038/nature09113)
130. Poisot T, Stouffer DB. 2018 Interactions retain the co-phylogenetic matching that communities lost. *Oikos* **127**, 230–238. (doi:10.1111/oik.03788)
131. Coelho MT, Rodrigues JF, Rangel TF. 2017 Neutral biogeography of phylogenetically structured interaction networks. *Ecography* **40**, 1467–1474. (doi:10.1111/ecog.02780)
132. Eklöf A *et al.* 2013 The dimensionality of ecological networks. *Ecol. Lett.* **16**, 577–583. (doi:10.1111/ele.12081)
133. Strydom T, Dalla Riva GV, Poisot T. 2021 SVD entropy reveals the high complexity of ecological networks. *Front. Ecol. Evol.* **9**, 308. (doi:10.3389/fevo.2021.623141)
134. Rohr RP, Scherer H, Kehrl P, Mazza C, Bersier L-F. 2010 Modeling food webs: exploring unexplained structure using latent traits. *Am. Nat.* **176**, 170–177. (doi:10.1086/653667)
135. Rossberg AG, Matsuda H, Amemiya T, Itoh K. 2006 Food webs: experts consuming families of experts. *J. Theor. Biol.* **241**, 552–563. (doi:10.1016/j.jtbi.2005.12.021)
136. Borrett SR, Scharler UM. 2019 Walk partitions of flow in ecological network analysis: review and synthesis of methods and indicators. *Ecol. Indic.* **106**, 105451. (doi:10.1016/j.ecolind.2019.105451)
137. Berlow EL *et al.* 2004 Interaction strengths in food webs: issues and opportunities. *J. Anim. Ecol.* **73**, 585–598. (doi:10.1111/j.0021-8790.2004.00833.x)
138. Barnes AD, Jochum M, Lefcheck JS, Eisenhauer N, Scherber C, O'Connor MI, de Ruiter P, Brose U. 2018 Energy flux: the link between multitrophic biodiversity and ecosystem functioning. *Trends Ecol. Evol.* **33**, 186–197. (doi:10.1016/j.tree.2017.12.007)
139. Brown JH, Gillooly JF, Allen AP, Savage VM, West GB. 2004 Toward a metabolic theory of ecology. *Ecology* **85**, 1771–1789. (doi:10.1890/03-9000)
140. Neutel A-M. 2002 Stability in real food webs: weak links in long loops. *Science* **296**, 1120–1123. (doi:10.1126/science.1068326)
141. de Ruiter PC, Neutel AM, Moore JC. 1995 Energetics, patterns of interaction strengths, and stability in real ecosystems. *Science* **269**, 1257–1260. (doi:10.1126/science.269.5228.1257)
142. Duffy JE. 2002 Biodiversity and ecosystem function: the consumer connection. *Oikos* **99**, 201–219. (doi:10.1034/j.1600-0706.2002.990201.x)
143. Montoya JM, Rodríguez MA, Hawkins BA. 2003 Food web complexity and higher-level ecosystem services. *Ecol. Lett.* **6**, 587–593. (doi:10.1046/j.1461-0248.2003.00469.x)
144. Farage C, Edler D, Eklöf A, Rosvall M, Pilosof S. 2021 Identifying flow modules in ecological networks using Infomap. *Methods Ecol. Evol.* **12**, 778–786. (doi:10.1111/2041-210X.13569)
145. Montoya JM, Solé RV. 2002 Small world patterns in food webs. *J. Theor. Biol.* **214**, 405–412. (doi:10.1006/jtbi.2001.2460)
146. Gaiarsa MP, Guimarães PR. 2019 Interaction strength promotes robustness against cascading effects in mutualistic networks. *Sci. Rep.* **9**, 676. (doi:10.1038/s41598-018-35803-8)
147. Simková A, Kadlec D, Gelnar M, Morand S. 2002 Abundance-prevalence relationship of gill congeneric ectoparasites: testing the core satellite hypothesis and ecological specialisation. *Parasitol. Res.* **88**, 682–686. (doi:10.1007/s00436-002-0650-3)
148. Desdevises Y. 2006 Determinants of parasite species richness on small taxonomical and geographical scales: lamellodiscus monogeneans of northwestern Mediterranean sparid fish. *J. Helminthol.* **80**, 235–241.
149. Morand S, Simková A, Matejusová I, Plaisance L, Verneau O, Desdevises Y. 2002 Investigating patterns may reveal processes: evolutionary ecology of ectoparasitic monogeneans. *Int. J. Parasitol.* **32**, 111–119. (doi:10.1016/s0020-7519(01)00347-2)
150. Sasal P, Niquil N, Bartoli P. 1999 Community structure of digenean parasites of sparid and labrid fishes of the Mediterranean Sea: a new approach. *Parasitology* **119**, 635–648. (doi:10.1017/s0031182099005077)
151. Badets M *et al.* 2011 Correlating early evolution of parasitic plathyhelminths to Gondwana breakup. *Syst. Biol.* **60**, 762–781. (doi:10.1093/sysbio/syr078)
152. Dickinson ER, Millins C, Biek R. 2020 Sampling scale and season influence the observed relationship

- between the density of deer and questing *Ixodes ricinus* nymphs. *Parasit. Vectors* **13**, 493. (doi:10.1186/s13071-020-04369-8)
153. Karbowski G, Miklisová D, Stanko M, Werszko J, Hajdul-Marwicz M, Szewczyk T, Rychlik L. 2019 The competition between immatures of *Ixodes ricinus* and *Dermacentor reticulatus* (Ixodida: Ixodidae) ticks for rodent hosts. *J. Med. Entomol.* **56**, 448–452. (doi:10.1093/jme/tjy188)
154. Matthee S, Stekolnikov AA, van der Mescht L, Froeschke G, Morand S. 2020 The diversity and distribution of chigger mites associated with rodents in the South African savanna. *Parasitology* **147**, 1038–1047. (doi:10.1017/S00311820000748)
155. Sanchez S, Serrano E, Gómez MS, Felu C, Morand S. 2014 Positive co-occurrence of flea infestation at a low biological cost in two rodent hosts in the Canary archipelago. *Parasitology* **141**, 511–521. (doi:10.1017/S0031182013001753)
156. Macías-Hernández N, Athey K, Tonzo V, Wangenstein OS, Arnedo M, Harwood JD. 2018 Molecular gut content analysis of different spider body parts. *PLoS ONE* **13**, e0196589. (doi:10.1371/journal.pone.0196589)
157. Schmid-Araya JM, Schmid PE, Tod SP, Esteban GF. 2016 Trophic positioning of meiofauna revealed by stable isotopes and food web analyses. *Ecology* **97**, 3099–3109. (doi:10.1002/ecy.1553)
158. Portalier SMJ, Fussmann GF, Loreau M, Cherif M. 2019 The mechanics of predator-prey interactions: first principles of physics predict predator-prey size ratios. *Funct. Ecol.* **33**, 323–334. (doi:10.1111/1365-2435.13254)
159. Berlow EL, Dunne J, Martinez ND, Stark PB, Williams RJ, Brose U. 2009 Simple prediction of interaction strengths in complex food webs. *Proc. Natl Acad. Sci. USA* **106**, 187–191. (doi:10.1073/pnas.0806823106)
160. Yodzis P, Innes S. 1992 Body size and consumer-resource dynamics. *Am. Nat.* **139**, 1151–1175. (doi:10.1086/285380)
161. May RM 1974 *Stability and complexity in model ecosystems*, vol. 1. Princeton, NJ: Princeton University Press.
162. Albouy C *et al.* 2019 The marine fish food web is globally connected. *Nat. Ecol. Evol.* **3**, 1153–1161. (doi:10.1038/s41559-019-0950-y)
163. Allesina S, Tang S. 2012 Stability criteria for complex ecosystems. *Nature* **483**, 205–208. (doi:10.1038/nature10832)
164. Allesina S, Alonso D, Pascual M. 2008 A general model for food web structure. *Science* **320**, 658–661. (doi:10.1126/science.1156269)
165. Delmas E, Brose U, Gravel D, Stouffer DB, Poisot T. 2017 Simulations of biomass dynamics in community food webs. *Methods Ecol. Evol.* **8**, 881–886. (doi:10.1111/2041-210X.12713)
166. Domínguez-García V, Dakos V, Kéfi S. 2019 Unveiling dimensions of stability in complex ecological networks. *Proc. Natl Acad. Sci. USA* **116**, 25714–25720. (doi:10.1073/pnas.1904470116)
167. Gravel D, Massol F, Leibold MA. 2016 Stability and complexity in model meta-ecosystems. *Nat. Commun.* **7**, 12457. (doi:10.1038/ncomms12457)
168. Kéfi S, Domínguez-García V, Donohue I, Fontaine C, Thébault E, Dakos V. 2019 Advancing our understanding of ecological stability. *Ecol. Lett.* **22**, 1349–1356. (doi:10.1111/ele.13340)
169. Ferrera A, Pascual-García A, Bastolla U. 2016 Effective competition determines the global stability of model ecosystems. *Theor. Ecol.* **10**, 1–11. (doi:10.1007/s12080-016-0322-z)
170. Saavedra S, Rohr RP, Bascompte J, Godoy O, Kraft NJB, Levine JM. 2017 A structural approach for understanding multispecies coexistence. *Ecol. Monogr.* **87**, 470–86. (doi:10.1002/ecm.1263)
171. Wiens JA. 1989 Spatial scaling in ecology. *Funct. Ecol.* **3**, 385–397. (doi:10.2307/2389612)
172. Golubski AJ, Abrams PA. 2011 Modifying modifiers: what happens when interspecific interactions interact? *J. Anim. Ecol.* **80**, 1097–1108. (doi:10.1111/j.1365-2656.2011.01852.x)
173. Golubski AJ, Westlund EE, Vandermeer J, Pascual M. 2016 Ecological networks over the edge: hypergraph trait-mediated indirect interaction (TMII) structure. *Trends Ecol. Evol.* **31**, 344–354. (doi:10.1016/j.tree.2016.02.006)
174. Carletti T, Fanelli D, Nicoletti S. 2020 Dynamical systems on hypergraphs. *J. Phys.: Complex.* **1**, 035006. (doi:10.1088/2632-072X/aba8e1)
175. Hutchinson MC, Bramon Mora B, Pilofof S, Barner AK, Kéfi S, Thébault E, Jordano P, Stouffer DB. 2019 Seeing the forest for the trees: putting multilayer networks to work for community ecology. *Funct. Ecol.* **33**, 206–217. (doi:10.1111/1365-2435.13237)
176. Gross T, Allhoff KT, Blasius B, Brose U, Drossel B, Fahimipour AK, Guill C, Yeakel JD, Zeng F. 2020 Modern models of trophic meta-communities. *Phil. Trans. R. Soc. B* **375**, 20190455. (doi:10.1098/rstb.2019.0455)
177. Albert CH, Rayfield B, Dumitru M, Gonzalez A. 2017 Applying network theory to prioritize multispecies habitat networks that are robust to climate and land-use change: prioritizing a network for biodiversity. *Conserv. Biol.* **31**, 1383–1396. (doi:10.1111/cobi.12943)
178. Pilofof S, Porter MA, Pascual M, Kéfi S. 2017 The multilayer nature of ecological networks. *Nat. Ecol. Evol.* **1**, 101. (doi:10.1038/s41559-017-0101)
179. Jalili M, Orouskhani Y, Asgari M, Alipourfard N, Perc M. 2017 Link prediction in multiplex online social networks. *R. Soc. Open Sci.* **4**, 160863. (doi:10.1098/rsos.160863)
180. Najari S, Salehi M, Ranjbar V, Jalili M. 2019 Link prediction in multiplex networks based on interlayer similarity. *Physica A* **536**, 120978. (doi:10.1016/j.physa.2019.04.214)
181. Yasami Y, Safaei F. 2018 A novel multilayer model for missing link prediction and future link forecasting in dynamic complex networks. *Physica A* **492**, 2166–2197. (doi:10.1016/j.physa.2017.11.134)
182. Baiser B *et al.* 2019 Ecogeographical rules and the macroecology of food webs. *Global Ecol. Biogeogr.* **28**, 1204–1218. (doi:10.1111/geb.12925)
183. Trøjsgaard K, Olesen JM. 2016 Ecological networks in motion: micro- and macroscopic variability across scales. *Funct. Ecol.* **30**, 1926–1935. (doi:10.1111/1365-2435.12710)
184. Brose U, Ostling A, Harrison K, Martinez ND. 2004 Unified spatial scaling of species and their trophic interactions. *Nature* **428**, 167–171. (doi:10.1038/nature02297)
185. Riede JO, Rall BC, Banasek-Richter C, Navarrete SA, Wieters EA, Emmerson MC, Jacob U, Brose U. 2010 Scaling of food-web properties with diversity and complexity across ecosystems. *Adv. Ecol. Res.* **42**, 139–170. (doi:10.1016/B978-0-12-381363-3.00003-4)
186. Bolnick DI *et al.* 2011 Why intraspecific trait variation matters in community ecology. *Trends Ecol. Evol.* **26**, 183–192. (doi:10.1016/j.tree.2011.01.009)
187. Vázquez DP, Melián CJ, Williams NM, Blüthgen N, Krasnov BR, Poulin R. 2007 Species abundance and asymmetric interaction strength in ecological networks. *Oikos* **116**, 1120–1127. (doi:10.1111/j.0030-1299.2007.15828.x)
188. Mora BB, Gravel D, Gilarranz LJ, Poisot T, Stouffer DB. 2018 Identifying a common backbone of interactions underlying food webs from different ecosystems. *Nat. Commun.* **9**, 1–8. (doi:10.1038/s41467-018-05056-0)
189. Dehling DM *et al.* 2020 Similar composition of functional roles in Andean seed-dispersal networks, despite high species and interaction turnover. *Ecology* **101**, e03028. (doi:10.1002/ecy.3028)
190. Elith J *et al.* 2006 Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**, 129–151. (doi:10.1111/j.2006.0906-7590.04596.x)
191. Guisan A, Thuiller W. 2005 Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* **8**, 993–1009. (doi:10.1111/j.1461-0248.2005.00792.x)
192. Wisz MS *et al.* 2013 The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biol. Rev.* **88**, 15–30. (doi:10.1111/j.1469-185X.2012.00235.x)
193. Norberg A *et al.* 2019 A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecol. Monogr.* **89**, e01370. (doi:10.1002/ecm.1370)
194. Ovaskainen O, Tikhonov G, Norberg A, Guillaume Blanchet F, Duan L, Dunson D, Roslin T, Abrego N. 2017 How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecol. Lett.* **20**, 561–576. (doi:10.1111/ele.12757)
195. Guisan A, Rahbek C. 2011 SESAM a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *J. Biogeogr.* **38**, 1433–1444. (doi:10.1111/j.1365-2699.2011.02550.x)
196. D'Amen M, Dubuis A, Fernandes RF, Pottier J, Pellissier L, Guisan A. 2015 Using species richness and functional traits predictions to constrain assemblage predictions from stacked species



- distribution models. *J. Biogeogr.* **42**, 1255–1266. (doi:10.1111/jbi.12485)
197. Karger DN *et al.* 2016 Delineating probabilistic species pools in ecology and biogeography. *Global Ecol. Biogeogr.* **25**, 489–501. (doi:10.1111/geb.12422)
198. Staniczenko PPA, Sivasubramaniam P, Suttle KB, Pearson RG. 2017 Linking macroecology and community ecology: refining predictions of species distributions using biotic interaction networks. *Ecol. Lett.* **20**, 693–707. (doi:10.1111/ele.12770)
199. Parker WS. 2013 Ensemble modeling, uncertainty and robust predictions. *WIREs Clim. Change* **4**, 213–223. (doi:10.1002/wcc.220)
200. Draper D. 1995 Assessment and propagation of model uncertainty. *J. R. Stat. Soc. Ser. B* **57**, 45–97. (doi:10.1111/j.2517-6161.1995.tb02015.x)
201. Parysow P, Gertner G, Westervelt J. 2000 Efficient approximation for building error budgets for process models. *Ecol. Modell.* **135**, 111–125. (doi:10.1016/S0304-3800(00)00347-1)
202. Benke KK, Norng S, Robinson NJ, Benke LR, Peterson TJ. 2018 Error propagation in computer models: analytic approaches, advantages, disadvantages and constraints. *Stoch. Environ. Res. Risk Assess.* **32**, 2971–2985. (doi:10.1007/s00477-018-1555-8)
203. Peters DPC, Herrick JE. 2004 Strategies for ecological extrapolation. *Oikos* **106**, 627–636. (doi:10.1111/j.0030-1299.2004.12869.x)
204. Harvey E, Gounand I, Ward CL, Altermatt F. 2017 Bridging ecology and conservation: from ecological networks to ecosystem function. *J. Appl. Ecol.* **54**, 371–379. (doi:10.1111/1365-2664.12769)
205. Pollock LJ, O'Connor LMJ, Mokany K, Rosauer DF, Talluto MV, Thuiller W. 2020 Protecting biodiversity (in all its complexity): new models and methods. *Trends Ecol. Evol.* **35**, 1119–1128. (doi:10.1016/j.tree.2020.08.015)
206. McDonald-Madden E, Sabbadin R, Game ET, Baxter PW, Chades I, Possingham HP. 2016 Using food-web theory to conserve ecosystems. *Nat. Commun.* **7**, 10245. (doi:10.1038/ncomms10245)
207. McWilliams C, Lurgi M, Montoya JM, Sauve A, Montoya D. 2019 The stability of multitrophic communities under habitat loss. *Nat. Commun.* **10**, 2322. (doi:10.1038/s41467-019-10370-2)
208. Araújo MB, Luoto M. 2007 The importance of biotic interactions for modelling species distributions under climate change. *Global Ecol. Biogeogr.* **16**, 743–753. (doi:10.1111/j.1466-8238.2007.00359.x)
209. Pigot AL, Tobias JA. 2013 Species interactions constrain geographic range expansion over evolutionary time. *Ecol. Lett.* **16**, 330–338. (doi:10.1111/ele.12043)
210. Urban MC *et al.* 2016 Improving the forecast for biodiversity under climate change. *Science* **353**, aad8466. (doi:10.1126/science.aad8466)
211. Gilman SE, Urban MC, Tewsbury J, Gilchrist GW, Holt RD. 2010 A framework for community interactions under climate change. *Trends Ecol. Evol.* **25**, 325–331. (doi:10.1016/j.tree.2010.03.002)
212. Araújo MB *et al.* 2019 Standards for distribution models in biodiversity assessments. *Sci. Adv.* **5**, eaat4858. (doi:10.1126/sciadv.aat4858)
213. Kindsvater HK, Dulvy NK, Horswill C, Juan-Jordá M-J, Mangel M, Matthiopoulos J. 2018 Overcoming the data crisis in biodiversity conservation. *Trends Ecol. Evol.* **33**, 676–688. (doi:10.1016/j.tree.2018.06.004)
214. Stouffer DB, Bascompte J. 2010 Understanding food-web persistence from local to global scales. *Ecol. Lett.* **13**, 154–161. (doi:10.1111/j.1461-0248.2009.01407.x)
215. Thompson RM *et al.* 2012 Food webs: reconciling the structure and function of biodiversity. *Trends Ecol. Evol.* **27**, 689–97. (doi:10.1016/j.tree.2012.08.005)
216. Albouy C, Guilhaumon F, Leprieur F, Lasram FB, Somot S, Aznar R, Velez L, Le Loc'h F, Mouillot D. 2013 Projected climate change and the changing biogeography of coastal Mediterranean fishes. Edited by Peter Pearman. *J. Biogeogr.* **40**, 534–547. (doi:10.1111/jbi.12013)
217. Hattab T, Leprieur F, Lasram FBR, Gravel D, Loc'h FL, Albouy C. 2016 Forecasting fine-scale changes in the food-web structure of coastal marine communities under climate change. *Ecography* **39**, 1227–1237. (doi:10.1111/ecog.01937)
218. Kortsch S, Primicerio R, Fossheim M, Dolgov AV, Aschan M. 2015 Climate change alters the structure of arctic marine food webs due to poleward shifts of boreal generalists. *Proc. R. Soc. B* **282**, 20151546. (doi:10.1098/rspb.2015.1546)
219. Pennekamp F *et al.* 2019 The intrinsic predictability of ecological time series and its potential to guide forecasting. *Ecol. Monogr.* **89**, e01359. (doi:10.1002/ecm.1359)
220. Dietze MC *et al.* 2018 Iterative near-term ecological forecasting: needs, opportunities, and challenges. *Proc. Natl Acad. Sci. USA* **115**, 1424–1432. (doi:10.1073/pnas.1710231115)
221. Lei L, Whitaker JS. 2017 Evaluating the trade-offs between ensemble size and ensemble resolution in an ensemble-variational data assimilation system. *J. Adv. Model. Earth Syst.* **9**, 781–789. (doi:10.1002/2016MS000864)
222. Dehling DM, Stouffer DB. 2018 Bringing the Eltonian niche into functional diversity. *Oikos* **127**, 1711–1723. (doi:10.1111/oik.05415)
223. O'Connor LM, Pollock LJ, Braga J, Ficetola GF, Maiorano L, Martinez-Almoyna C, Montemaggiore A, Ohlmann M, Thuiller W. 2020 Unveiling the food webs of tetrapods across Europe through the prism of the Eltonian niche. *J. Biogeogr.* **47**, 181–192. (doi:10.1111/jbi.13773)
224. Guzman LM, Germain RM, Forbes C, Straus S, O'Connor MI, Gravel D, Srivastava DS, Thompson PL. 2019 Towards a multi-trophic extension of metacommunity ecology. *Ecol. Lett.* **22**, 19–33. (doi:10.1111/ele.13162)
225. Braga J *et al.* 2019 Spatial analyses of multi-trophic terrestrial vertebrate assemblages in Europe. *Global Ecol. Biogeogr.* **28**, 1636–1648. (doi:10.1111/geb.12981)
226. Massol F, Dubart M, Calcagno V, Cazelles K, Jacquet C, Kéfi S, Gravel D. 2017 Chapter four - island biogeography of food webs. In *Advances in ecological research*. Networks of invasion: a synthesis of concepts (eds DA Bohan, AJ Dumbrell, F Massol), vol. 56, pp. 183–262. New York, NY: Academic Press. (doi:10.1016/bs.aecr.2016.10.004)
227. Foden WB, Young BE. 2016 *IUCN SSC Guidelines for assessing species' vulnerability to climate change*. Cambridge, UK and Gland, Switzerland: IUCN.
228. Syfert MM, Joppa L, Smith MJ, Coomes DA, Bachman SP, Brummitt NA. 2014 Using species distribution models to inform IUCN red list assessments. *Biol. Conserv.* **177**, 174–184. (doi:10.1016/j.biocon.2014.06.012)