



METHODS

Structural forecasting of species persistence under changing environments

Serguei Saavedra, 
 Lucas P. Medeiros  and
 Mohammad AlAdwani

Department of Civil and Environmental Engineering, MIT, 77 Massachusetts Av, 02139, Cambridge, MA, USA

*Correspondence: E-mail: sersaa@mit.edu

Abstract

The persistence of a species in a given place not only depends on its intrinsic capacity to consume and transform resources into offspring, but also on how changing environmental conditions affect its growth rate. However, the complexity of factors has typically taken us to choose between understanding and predicting the persistence of species. To tackle this limitation, we propose a probabilistic approach rooted on the statistical concepts of ensemble theory applied to statistical mechanics and on the mathematical concepts of structural stability applied to population dynamics models – what we call *structural forecasting*. We show how this new approach allows us to estimate a probability of persistence for single species in local communities; to understand and interpret this probability conditional on the information we have concerning a system; and to provide out-of-sample predictions of species persistence as good as the best experimental approaches without the need of extensive amounts of data.

Keywords

ecological communities, ensemble theory, experimental tests, nonlinear population dynamics, out-of-sample predictions, probability, statistical mechanics, structural stability.

Ecology Letters (2020)

INTRODUCTION

One of the long-standing questions in ecology is how can we know whether an observed species (e.g. bacteria, plant, insect or mammal species) in a given place (e.g. human host, natural habitat) will persist across a period of time (Pimm, 1991; Vellend, 2016). Even more simply, we can ask: what is the chance that a given species will persist? The difficulty in answering this question resides in knowing the exact equations governing the dynamics of ecological systems, together with the high uncertainty regarding the initial conditions, parameter values, intrinsic randomness, and more importantly, how the changing external conditions (such as biotic and abiotic factors) will affect the dynamics (Levins, 1968; Sugihara, 1994; Fukami, 2015; Boettiger, 2018; Cenci and Saavedra, 2018b). This complexity of multidimensional and changing factors has typically taken both theoretical and empirical studies to choose between understanding and predicting species persistence (Petchey *et al.*, 2015; Clark *et al.*, 2020).

Indeed, apart from collecting an enormous amount of data (which may not be possible to obtain under time and resource constraints) and building sophisticated learning algorithms (which may not generalise well to unseen data nor provide ecological understanding), how to understand and predict the persistence of species subject to changing environments remains an open question (Sugihara *et al.*, 2012; Harfoot *et al.*, 2014; Dietze, 2017; Cenci and Saavedra, 2019). Yet, having a framework that could unify the problems of understanding and predicting can help us to answer not only the question of what is the probability of persistence of a species, but also the question of why is this probability small or large. Notably, answering *what* and *why* questions are fundamental pillars of a science that is both predictive and descriptive

(Poincaré, 1905; Pearl, 2009). Moreover, this unification is of paramount importance in the face of rapidly changing environmental conditions, where understanding and predicting the presence (or absence) of species in ecological communities can help us to establish sustainable strategies essential to the maintenance of biodiversity and human well-being (Stenseth *et al.*, 2002; Walther, 2010; Dirzo *et al.*, 2014; Lu *et al.*, 2016; Rohr *et al.*, 2020).

On the one hand, theoretical and empirical work in ecology has focused on understanding the conditions under which a species can persist given an environmental context (Yodzis and Innes, 1992; Case, 2000; Turchin, 2003). It is worth noting that despite the fact that several ecological studies use the word prediction (Valdovinos, 2019), these studies are explanatory in nature as they do not address out-of-sample problems explicitly (Shmueli, 2010). Importantly, a large body of work has been built on either phenomenological or mechanistic models, where a context is defined by a set of interacting species (i.e. biotic factors) affecting each others' growth rate as well as by a specific set of abiotic factors shaping the parameter values of such models (Case, 2000). These studies have provided key insights regarding the mechanisms, as well as the necessary and sufficient conditions for species persistence (Svirezhev and Logofet, 1983; Hofbauer and Sigmund, 1998). However, it is still unclear how to use this knowledge to anticipate unknown observations (Fukami, 2015). For example, the well-known invasibility criterion, which states that a species persists if its invasion growth rate is positive, operates only if it is known a priori that all the other species in a community can also persist when removed and reintroduced – which is seldom the case (Saavedra *et al.*, 2017b; Grainger and Gilbert, 2019). In fact, using simulations, it has been shown that empirically parameterised models at the pairwise

level fail to predict the dynamics of larger experimental communities (Friedman *et al.*, 2017; Tuck *et al.*, 2018).

On the other hand, biodiversity forecasting is a well established and required area in ecological research (Clark *et al.*, 2001; Dietze, 2017). This field is providing the opportunity to anticipate rather than simply explain biodiversity changes in ecological communities contingent on explicit scenarios for climate change, land-use and species re-distributions (Dietze, 2017). Importantly, biodiversity forecasting spans and integrates many model-driven (parametric) and data-driven (non-parametric) methodologies, such as uncertainty propagation, statistics, informatics, Bayesian approaches, machine learning, Markov chain approaches, empirical dynamic modelling (Sugihara *et al.*, 2012; Harfoot *et al.*, 2014; Cazelles *et al.*, 2016; Dietze, 2017; Cenci and Saavedra, 2019; Adams *et al.*, 2020; Maynard *et al.*, 2020), as well as parameterising complex mechanistic models using either demographic, eco-physiological or allometric information (Preston, 1962; Pacala *et al.*, 1996; Dietze, 2017). However, the majority of these methodologies demands extensive amounts of data, their explanatory power has been contested, and their generalisation has not always been validated with experimental work (Dietze, 2017; Clark *et al.*, 2020).

As a response to the above limitations, recent experimental work has derived heuristic rules to predict species persistence (Friedman *et al.*, 2017). While these rules have been highly successful in making out-of-sample predictions (e.g. the persistence of species in microbial communities), their limitation resides on how to generalise these rules under the uncertainty derived from the unknown changing environmental conditions (Levins, 1968). For example, some of these rules completely eliminate the possibility of forming a community from a species pool when smaller subsets of species cannot persist in isolation (Friedman *et al.*, 2017). However, this is precisely the dynamics observed during different successional stages and when intransitive competition (e.g. rock–paper–scissors dynamics) operates in natural and experimental communities (Odum, 1969; Fukami, 2015; Levine *et al.*, 2017; Saavedra *et al.*, 2017a; Song *et al.*, 2018a). Moreover, the application of these rules can be experimentally extensive or unfeasible, calling for new descriptive and predictive approaches that can take into account the effects of changing environments using minimal information.

Although these theoretical and empirical studies are pushing community ecology into a more descriptive and predictive science, it has been emphasised that conceptual risks need to be taken in order to establish new approaches that can address the current limitations in understanding and predicting species persistence under changing conditions (Dietze, 2017). Indeed, finding a compromise between tractability (simplicity) and realism (complexity) has always been at the core of ecological modelling (May, 1976, 2004). This trade-off typically divides the capacity of understanding from predicting ecological dynamics (Sugihara, 1994). Importantly, while the trade-off between predicting and understanding nonlinear dynamical systems under changing conditions may be impossible to fully eliminate (Shmueli, 2010), research into the irreversibility of physical and biological systems has suggested that probabilistic approaches can be used to unify the

deterministic nature of their dynamics and their intrinsic variability due to changing environmental conditions and instabilities (Prigogine, 1962; Thom, 1972).

In statistical mechanics, given the difficulties in knowing even the initial conditions of a system (i.e. knowing the exact state of a system is an idealisation), we may represent the expected behaviour of a system by the *ensemble average* – the average of states (i.e. the possible copies of a system) consistent with given restrictions (e.g. same energy) but independent of initial conditions (Einstein, 1902; Gibbs, 1902). Each ensemble is then associated with the probability of a given qualitative or quantitative state. Conceptually, this framework moves the interpretation of an ensemble of equally likely states from an average behaviour to a potential behaviour. Focusing on the qualitative states of a system (e.g. a system with S different kinds of particles), the theory of ensembles can be integrated with the framework of structural stability (Smale, 1967; Arnold, 1988). Formally, a dynamical system is said to be structurally stable if the topology of the phase portrait is preserved under smooth changes of the vector field. In other words, structural stability can measure the level of environmental change that a system can sustain before undergoing some transition to a different qualitative state (e.g. from S to $S - 1$ different kinds of particles). Moreover, because the environmental conditions under which systems evolve can never be exactly duplicated, structural stability is a necessary condition for the observability of a system. Statistical mechanics and structural stability have allowed an integration of determinism and randomness in nonlinear dynamical systems (Prigogine *et al.*, 1973), opening new research areas and applications such as the successful research programs of quantum mechanics and morphogenesis (Planck, 1925; Thom, 1972; Prigogine and Stengers, 1984; Alberch, 1989).

Following the premises above, here we advocate that understanding and predicting the persistence of single species under changing environmental conditions can also be accomplished by using a probabilistic approach (Lewontin, 1969; Cazelles *et al.*, 2016; Dietze, 2017; Song *et al.*, 2020). This new program can be developed by integrating the statistical concepts from the theory of ensembles applied to statistical mechanics (Einstein, 1902; Gibbs, 1902; Prigogine, 1962) with the mathematical concepts of structural stability applied to nonlinear population dynamics (Lewontin, 1969; May, 1976; Solé and Valls, 1992; Bastolla *et al.*, 2005; Rohr *et al.*, 2014; Saavedra *et al.*, 2017b) – what we call *structural forecasting*.

In this new program, the grand-canonical ensemble (per its name in the theory of ensembles when the system's states differ qualitatively) is formed by the different community configurations consistent with given restrictions (e.g. same governing laws and species interactions) formed by the various species that can potentially be present in a given place and time. Each of these community configurations can then be considered a micro-canonical ensemble (or a qualitative state of the grand ensemble), whose own states only differ quantitatively (i.e. same kinds of species but with different species abundances). In turn, individual species can represent subsystems of the grand ensemble formed by the micro ensembles containing such species. Thus, instead of aiming to study the future behaviour of a system (the realised micro ensemble)

or subsystem (the presence of a species) by inferring the main conditions acting upon it (which may be impossible to do), this program estimates the probability of a future behaviour based on the fraction of possible conditions compatible with such behaviour. In the remainder of this manuscript, we develop the conceptual framework of structural forecasting. Then, we provide a proof of concept using experimental trials of interacting microbial species. We show that structural forecasting can allow us to estimate the probability of persistence of single species in local communities, to understand and interpret this probability conditional on the information we have concerning a system, and to provide out-of-sample predictions of species persistence as good as the best experimental approaches without the need of obtaining extensive amounts of data.

UNDERSTANDING SPECIES PERSISTENCE

To understand the persistence of a species in an ecological community, it is important (although not necessary, see Sugi-hara, 1994; Cenci and Saavedra, 2019) to have tractable (mechanistic or phenomenological) population dynamics models upon which one can study cause–effect relationships between model parameters and model outputs (Case, 2000; Strogatz, 2014). The stochastic nature of ecological dynamics can then be incorporated through either random noise or systematic changes of parameter values (Turchin, 2003; Schreiber *et al.*, 2019; Yang *et al.*, 2019). Note that this stochasticity is independent from the intrinsic variability of the dynamics (e.g. bifurcations or chaotic behaviour), although a strict separation between stochastic and chaotic dynamics in ecological systems is unnecessary and potentially misleading (Ellner and Turchin, 1995; Boettiger, 2018). Moreover, this separation becomes unnecessary due to the virtually impossible measurement of the exact initial conditions and knowledge about the exact equations governing the dynamics of natural systems. Yet, both the observability of these systems and the fact that external conditions are seldom the same between any two points in time suggests that species persistence can be understood by studying the probabilistic nature of ecological communities through the lenses of the theory of ensembles and structural stability (Kerner, 1962; Lewontin, 1969; Thom, 1972).

As mentioned before, a grand ensemble can be defined as the copies of a system (community), whose states differ qualitatively under given restrictions (e.g. same interaction matrix). These qualitative states then can be defined as micro ensembles that belong to a region in the phase space (i.e. the space where all the possible quantitative states of a dynamical system are represented), where each of its own quantitative states (whose probability falls within certain small range) is equally likely inside the micro ensemble and all other quantitative states outside the micro ensemble have probability zero. Hence, a grand ensemble (local community of interacting species) can have different micro ensembles (subsets of species from the full community) each with different quantitative states (distribution of species abundances) and the density of each micro ensemble represents the probability of finding the system in such qualitative state. Thus, the probability of

persistence of a species (a subsystem within the grand ensemble) can be computed by the union of all the micro ensembles representing the qualitative states compatible with the persistence of this species (i.e. sharing that kind of species or subsystem). Moreover, it is possible to extend these concepts to the parameter space, where a point in this space corresponds to a vector of model parameters. That is, we can focus on regions in the parameter space compatible with a given state and the density of each region of which corresponds to the probability of finding the system in this region – assuming that each point in the parameter space is equally likely to be observed. This extension has the advantage of linking model parameters and model outputs, an essential property for understanding ecological dynamics.

Formally, the ensemble approach allows us to define and understand the probability $P(i)$ of persistence of a species i as.

$$P(i) = \sum_{j=1}^{n_i} P(E_j^{(i)}), \quad (1)$$

where $P(E_j^{(i)})$ corresponds to the probability of observing the j th micro ensemble compatible with the persistence of species i and n_i is the total number of micro ensembles that contain species i . Note that the micro ensembles $E_j^{(i)}$ are mutually exclusive. Recall that these potential micro ensembles are defined by the different combinations of kinds of species that can be observed in a community (the grand ensemble). For example, considering three species, the set of potential micro ensembles for species 1 is $E^{(1)} = \{(1), (1,2), (1,3), (1,2,3)\}$. Note that we are not focusing on the specific abundance (or biomass) of species, but simply on whether they are present (i.e. the qualitative state).

Thus, for a system with three species, we can compute the following probabilities:

$$P(1) = P(1, \bar{2}, \bar{3}) + P(1, 2, \bar{3}) + P(1, \bar{2}, 3) + P(1, 2, 3) \quad (2)$$

$$P(2) = P(\bar{1}, 2, \bar{3}) + P(1, 2, \bar{3}) + P(\bar{1}, 2, 3) + P(1, 2, 3) \quad (3)$$

$$P(3) = P(\bar{1}, \bar{2}, 3) + P(1, \bar{2}, 3) + P(\bar{1}, 2, 3) + P(1, 2, 3), \quad (4)$$

where $P(X, \bar{Y})$ denotes the joint probability of X persisting and Y not persisting. Applying the multiplication rule, Bayes theorem and marginalisation, Eqns 2–4 can be further expanded to obtain single species probabilities given by

$$P(1) = P(1|\bar{2}, \bar{3})[P(1, \bar{2}|\bar{3}) + P(\bar{1}, \bar{2}|\bar{3})][1 - P(3)] + P(1, 2|\bar{3})[1 - P(3)] + P(1, 3|\bar{2})[1 - P(2)] + P(1, 2, 3) \quad (5)$$

$$P(2) = P(2|\bar{1}, \bar{3})[P(2, \bar{1}|\bar{3}) + P(\bar{2}, \bar{1}|\bar{3})][1 - P(3)] + P(1, 2|\bar{3})[1 - P(3)] + P(2, 3|\bar{1})[1 - P(1)] + P(1, 2, 3) \quad (6)$$

$$P(3) = P(3|\bar{1}, \bar{2})[P(3, \bar{2}|\bar{1}) + P(\bar{3}, \bar{2}|\bar{1})][1 - P(1)] + P(1, 3|\bar{2})[1 - P(2)] + P(2, 3|\bar{1})[1 - P(1)] + P(1, 2, 3), \quad (7)$$

where $P(X|\bar{Y})$ denotes the conditional probability of X persisting given that Y does not persist. The equations above

become a linear system with three equations and three unknowns ($P(1)$, $P(2)$ and $P(3)$) as long as we can compute the rest of the probabilities from the system under consideration. If the matrix given by the probabilities involving more than one species has full rank (i.e. the matrix of coefficients has three independent rows), then the system of eqns 5–7 has a unique solution for $P(1)$, $P(2)$ and $P(3)$.

To calculate the probabilities above, we propose to integrate the ensemble approach with concepts on structural stability applied to population dynamics models. For this purpose, we need to estimate the fraction of the whole parameter space that is compatible with each micro ensemble. According to the theory of ensembles, all possible states within the micro ensemble must appear with an equal probability. This implies that the parameter space must be sampled uniformly. Note that this principle is the basis of ergodicity (i.e. time average equals the ensemble average) and the independence from initial conditions (Gibbs, 1902). Ecologically, using a uniform distribution for parameter values means that we assume no information about how environmental changes (i.e. changes in model parameters) will affect the growth rate of species. Additionally, these different states need to be consistent with given restrictions. Thus, we establish that the probability of a micro ensemble corresponds to the density of its own quantitative states governed by the same population dynamics and interaction matrix \mathbf{A} (i.e. the matrix whose elements correspond to the per capita effect of one species on the per capita growth rate of another species) of the grand ensemble.

Following the assumptions above, we apply the framework of structural stability in ecology (Saavedra *et al.*, 2017b; Song *et al.*, 2018b) to estimate the probability of persistence of a micro ensemble $E_j^{(i)}$ conditional on a given interaction matrix \mathbf{A} as

$$P(E_j^{(i)}|\mathbf{A}) = \frac{\text{vol}(E_j^{(i)} \cap \mathbb{B}^S)}{\text{vol}(\mathbb{B}^S)}, \quad (8)$$

where S is the number of species, $\text{vol}(\mathbb{B}^S)$ is the volume of the S -dimensional unit ball (i.e. the full parameter space) and $\text{vol}(E_j^{(i)} \cap \mathbb{B}^S)$ corresponds to the volume of the intersection of the domain of the micro ensemble with the unit ball. That is, the probability $P(E_j^{(i)}|\mathbf{A})$ is given by the fraction of parameter values compatible with the states of the micro ensemble. In our example with three species (i.e. $S=3$), if $E_j^{(i)} = \{1, 2, 3\}$, then the probability $P(1, 2, 3|\mathbf{A})$ reduces to finding the fraction of the parameter space compatible with the persistence of all three species together. Similarly, if the micro ensemble is $E_j^{(i)} = \{1, 2\}$, then the probability $P(1, 2, \bar{3}|\mathbf{A})$ consists on finding $P(1, 2|\bar{3}, \mathbf{A})$ (recall that $[1 - P(3|\mathbf{A})]$ will be solved as an unknown variable in Eqns 5–7). It is also worth noting that this approach is formulated independently from the precise specification of initial conditions – following the premise of the theory of ensembles.

ILLUSTRATIVE EXAMPLE

To illustrate our approach, we calculated the probabilities of persistence for three-competing species. We assumed that the competition system can be described by a model topologically equivalent to the classic Lotka–Volterra (LV) model (Case,

2000) – that is, the unstable and stable fixed points in the classic LV model must be mapped into a pair of unstable and stable fixed points in the modified model (Cenci and Saavedra, 2018a). The classic LV competition model written in the r-formalism is given by:

$$\frac{dN_i}{dt} = N_i \left(r_i - \sum_{j=1}^S a_{ij} N_j \right) \quad (9)$$

$$\frac{d\mathbf{N}}{dt} = \text{diag}(\mathbf{N})(\mathbf{r} - \mathbf{A}\mathbf{N}), \quad (10)$$

where N_i is the abundance (or biomass) of species i , r_i is the intrinsic growth rate of species i , S is the number of species and a_{ij} is an element of the interaction matrix \mathbf{A} representing the per capita competitive effect of species j on species i . Note that under competition dynamics, it is expected to assume $r_i > 0$ and $a_{ij} > 0$. Bold characters denote matrix notation, for example, $\mathbf{r} = [r_1, \dots, r_S]^T$ (the superindex^T denotes the transpose operator). The classic LV model has been derived from thermodynamics principles (Michaelian, 2005), principles of conservation of mass and energy (Svirezhev and Logofet, 1983), from chemical kinetics in large populations (Täuber, 2011), and can exhibit a rich behaviour such as chaotic dynamics and limit cycles (Vano *et al.*, 2006).

Importantly, by assuming this classic LV model, it is possible to simplify the parameter space to a single phenomenological parameter (e.g. the intrinsic growth rates r_i), representing how environmental conditions (biotic and abiotic factors) affect the balance between mortality and resource intake. Recall that the interaction matrix \mathbf{A} is the restriction we are imposing on the micro ensembles (can be thought of as analogous to temperature in thermodynamics). Otherwise, if we treated \mathbf{A} also as a random variable, we would end up with no information about the system other than what one can obtain from standard random matrix theory (Serván *et al.*, 2018). One can simply apply small perturbations to \mathbf{A} ; however, this would lead us to sensitivity analysis (small perturbations around an expected value) rather than structural stability analysis (Meszéna *et al.*, 2006), making results dependent on the type of perturbation (Saavedra *et al.*, 2014).

As a starting point, we propose to estimate the probabilities of species persistence by focusing on the probability of feasibility of micro ensembles. We define the probability of feasibility as the probability of finding a positive solution (i.e. $N_i^* > 0$ for all species i under equilibrium $dN_i/dt = 0$) in the LV model. Note that feasibility provides the necessary condition for persistence, permanence and the existence of bounded orbits (Hofbauer and Sigmund, 1998). Thus, for a given interaction matrix \mathbf{A} , feasibility in LV models will be satisfied as long as the direction of the \mathbf{r} -vector (under the r-formalism) falls inside the feasibility domain $D_F(\mathbf{A}) = \{\mathbf{r} = N_1^* \mathbf{v}_1 + \dots + N_S^* \mathbf{v}_S, \text{ with } N_1^*, \dots, N_S^* > 0\}$, where \mathbf{v}_i is the i th column vector of \mathbf{A} and N_i^* is the feasible (i.e. positive) abundance of species i at equilibrium: $\mathbf{N}^* = \mathbf{A}^{-1} \mathbf{r}$. Because it is only necessary to know the different directions of \mathbf{r} -vectors, their magnitude can be normalised by any norm (Rohr *et al.*, 2016; later we will be considering the ℓ_1 norm). For example, for the micro ensemble $\{1, 2, 3\}$, the probability of feasibility can be calculated as.

$$P(1,2,3|\mathbf{A}) = \frac{2^S \text{vol}(D_F(\mathbf{A}) \cap \mathbb{B}^S)}{\text{vol}(\mathbb{B}^S)}, \quad (11)$$

where 2^S normalises the unit ball \mathbb{B}^S of parameter space (i.e. $\|\mathbf{r}\|_1 \leq 1$) to the positive orthant (i.e. $\mathbb{B}^S \cap \mathbb{R}_{\geq 0}^S$) and $\text{vol}(D_F(\mathbf{A}) \cap \mathbb{B}^S)$ corresponds to the volume of the intersection of the feasibility domain with the unit ball.

For micro ensembles with only two species (e.g. $\{1,2\}$), the probability of feasibility can also be calculated using eqn 11, but the interaction matrix \mathbf{A} needs to be modified. The modification can be done by either reducing the system by one dimension (e.g. removing the columns and rows related to species 3) or by making one species independent from the rest (e.g. setting to zero the interspecific interactions of species 3). This procedure will provide a conditional probability $P(i,j|\mathbf{A}_{-k})$, where i and j are the two observed species, \mathbf{A}_{-k} is the modified interaction matrix without species k . In the case of ensembles with a single species (e.g. $\{1\}$), it is also possible to use Equation 11 and calculate $P(i,j|\mathbf{A}_{-k,j})$, where $\mathbf{A}_{-k,j}$ is the modified interaction matrix \mathbf{A}_{-k} with the addition of moving the column vector of species j to its own axis, such that the new feasibility domain covers both the region where species i and j persist and the region where only species i persists

(Saavedra *et al.*, 2017b). Following this methodology, we can calculate all the probabilities required in eqns 5–7 and solve for $P(1|\mathbf{A})$, $P(2|\mathbf{A})$ and $P(3|\mathbf{A})$ – the probabilities of persistence of single species.

Figures 1a and b shows the regions of the parameter space of intrinsic growth rates \mathbf{r} (2000 directions of \mathbf{r} -vectors sampled uniformly inside the full parameter space (Song and Saavedra, 2018)) compatible with the feasibility of each micro ensemble derived from two illustrative 3-species competition systems under classic LV dynamics. Note that, for competition systems, the normalised parameter space of \mathbf{r} (the intersection between the positive orthant and the unit ball) can be visualised as the $S-1$ dimensional unit simplex (i.e. $\sum_{i=1}^S r_i = 1$, the ℓ_1 norm) (Saavedra *et al.*, 2017b). Each simplex (Figs 1a and b) corresponds to one interaction matrix \mathbf{A} , and each point inside the simplex represents a fixed point (a solution) of eqn 9 obtained by simulating the classic LV (Type I) model with a single initial condition under a given direction of the \mathbf{r} -vector. The closer a point is to a vertex i (representing the parameter r_i), the larger the fraction of resources that such species i uses from the environment. The first matrix (Fig. 1a) corresponds to a case where all species have similar competition effects on each other. The second matrix (Fig. 1b)

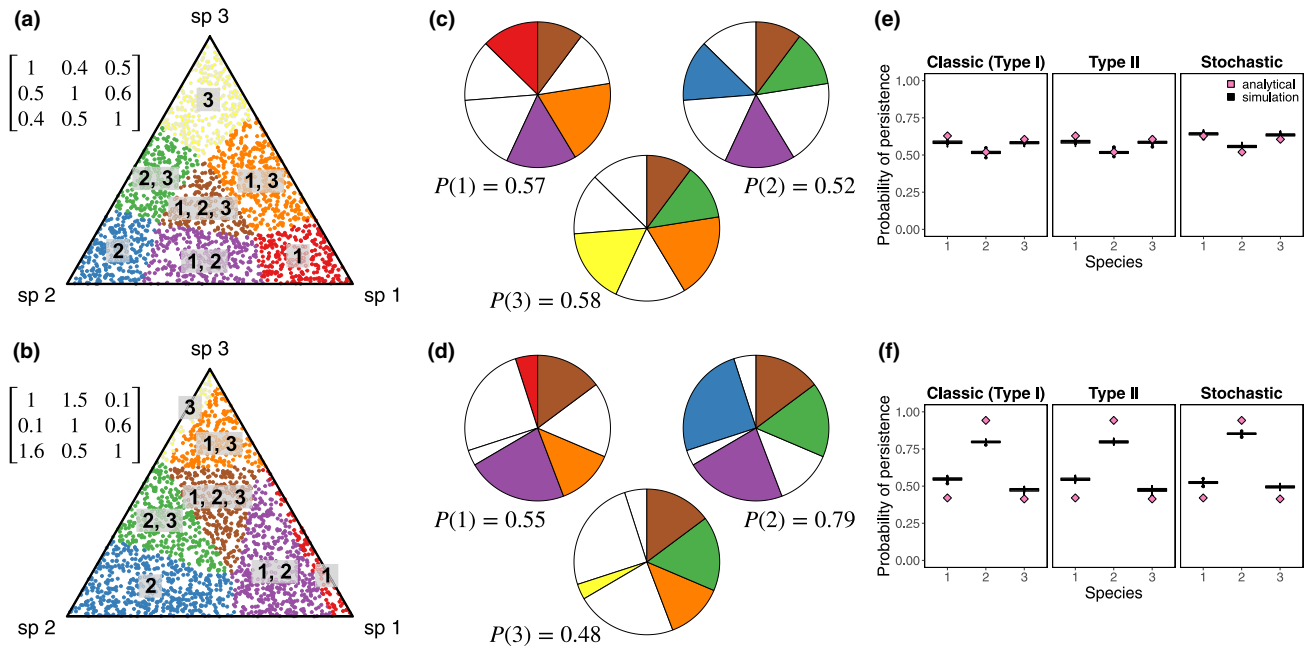


Figure 1 Probabilistic approach to understand species persistence under changing environments. The figure represents two illustrative communities of three-competing species. Panels (a and b) present the association between the parameter space of \mathbf{r} -vectors and the equilibrium state \mathbf{N}^* each interaction matrix \mathbf{A} (see text for details). The parameter space is represented as a two-dimensional unit simplex, meaning that each point is a vector of intrinsic growth rates that sums one and simplex vertices indicate the species that has $r_i = 1$. Each color inside the simplex corresponds to a different micro ensemble of species that resulted from simulating Lotka–Volterra (LV) dynamics using the interaction matrix \mathbf{A} shown and a given \mathbf{r} -vector. For example, the micro ensemble $\{1,2,3\}$ (brown points) implies that all the three species had positive solutions, whereas micro ensemble $\{1,2\}$ (purple points) implies that species 3 went extinct (zero solution). This simplex is constructed by simulating LV dynamics with 2000 different directions of \mathbf{r} -vectors sampled uniformly over the unit simplex under the same initial condition \mathbf{N} . Panels (c and d) represent the micro ensembles from the simplex that contains species 1, 2 or 3 and their respective frequency of occurrence under the simulations. The sum of the frequency of occurrence of these micro ensembles in the pie chart corresponds to a numerical approximation to the probability of persistence of a given species ($P(1|\mathbf{A})$, $P(2|\mathbf{A})$ and $P(3|\mathbf{A})$). Panels (e and f) represent the distribution (box plots) of $P(1|\mathbf{A})$, $P(2|\mathbf{A})$ and $P(3|\mathbf{A})$ generated by simulating LV dynamics over 100 different initial conditions. Left, middle and right panels correspond to the simulations of the classic (Type I), Type II and stochastic LV dynamics (see text for details). The pink diamonds correspond to the analytical probabilities $P(1|\mathbf{A})$, $P(2|\mathbf{A})$ and $P(3|\mathbf{A})$ calculated for each individual species following Equations 5–7 in the text

corresponds to transitive dynamics (i.e. the matrix allows for a rock–paper–scissors competition loop (Saavedra *et al.*, 2017b)), where species 2 is the strongest competitor.

Figures 1c and d shows the corresponding probabilities of persistence of single species ($P(i|\mathbf{A}), i=1,2,3$) using the same initial conditions as in Figures 1a and b. Each pie chart represents the probability associated with all the micro ensembles (see Figs 1a and b) to which a species can belong. Thus, the sum of the probabilities of each micro ensemble (i.e. fraction of simulations in which that micro ensemble is feasible) gives the probability of persistence of single species under changing environments. Recall that we are sampling uniformly the directions within the parameter space that represents the external conditions and, consequently, all states inside a micro ensemble have equal probability – following the principle of the theory of ensembles.

Figures 1e and f shows the distribution of the probability of persistence of single species ($P(i|\mathbf{A}), i=1,2,3$) using the classic LV model across 100 initial conditions as black box plots (200 000 simulations: 100 initial conditions and 2000 directions of \mathbf{r} -vectors for each initial condition) and the analytical approximations computed by solving eqns 5–7 as pink symbols. As expected, these probabilities confirm that in the first matrix, the three species have a similar probability of persistence, while in the second matrix species 2 has the largest probability. Importantly, the analytical probabilities were close to the numerical ones, indicating that the solutions are only weakly influenced by the initial conditions – as the theory of ensembles requires in order to be implemented. Note that differences between the analytical and numerical probabilities can be attributed to the number of samples (here 200 000), size and number of integration steps (here 10^{-2} over 200 time steps), species extinction thresholds (here 10^{-4}) and dynamical instabilities (leading to different boundary equilibria) considered for the simulations.

Furthermore, Figures 1e and f shows that the estimated probabilities of persistence of single species are also consistent with widely used modified LV models, as expected from previous work (Cenci and Saavedra, 2018a). Specifically, we repeated the simulations above, but instead of using the classic LV model (eqn 9), we used a Type II LV model (middle panel) and a stochastic LV model (right panel). The Type II LV model is given by multiplying \mathbf{A} in Eqn. 9 by the diagonal matrix \mathbf{M} with elements $m_{ii} = 1/(1 + N_i)$ (Hastings and Powell, 1991). The stochastic LV model is given by adding the term $v(t)/\sqrt{S}$ to eqn (9), where $v(t)$ is the Gaussian white noise with zero mean and correlations given by

$$\langle v_i(t)v_j(t') \rangle = \beta_{ij}\delta(t-t') \text{ using } \beta_{ij} = N_i \left(r_i - \sum_{j=1}^S a_{ij}N_j \right) \forall i=j \text{ and}$$

zero otherwise, while δ characterises the white noise (McKane *et al.*, 2014). Overall, Figure 1 shows that our proposed approach can be used to estimate and understand either analytically or numerically the probability of persistence of single species under changing environments.

PREDICTING SPECIES PERSISTENCE

While probability is not intrinsically intended to establish categorical predictions – a probabilistic thinking is rooted on the idea of (conditional) expected values (de Finetti, 2017) – the

value of a theory can only be assessed based on its capacity to explain and predict the natural world (Poincaré, 1905). Following this premise, the probabilities that can be derived from our proposed approach correspond to the probability of persistence of single species assuming that all parameter values are equally likely (i.e. the prior is a uniform distribution of intrinsic growth rate directions). Ecologically, this can be interpreted as either the context under which conditions change arbitrarily, or the context under which species did not have time to adapt to the current conditions and, as a consequence, parameter values can take any potential value (Levins, 1968; Song *et al.*, 2017, 2018a; Cenci *et al.*, 2018). Then, predictions can be made by establishing a threshold (λ) against which probabilities can be compared. For example, if we choose $\lambda=0.5$, then we predict that species i will persist if this species has a probability of persistence $P(i)>0.5$. Alternatively, one can establish as a criterion that only the species belonging to the micro ensemble (out of all possible ones) with the largest probability of feasibility will persist, or we can also introduce standard statistical approaches for prediction analyses (Rohr *et al.*, 2010).

However, in many occasions, the set of species under investigation may have already evolved towards specific environmental conditions or may have been observed under current or controlled conditions (Frederickson, 2017; Song *et al.*, 2020). For example, in laboratory experiments, typically the species under investigation are well adapted to the experimental conditions when in isolation (Hoek *et al.*, 2016). This selection process can bias the range of parameter values (e.g. intrinsic growth rates) under which a given system is observed, potentially allowing the existence of highly unlikely micro ensembles. To empirically explain these cases, we used publicly available data from a very detailed and controlled study performing persistence experiments by co-inoculating different combinations of heterotrophic soil-dwelling bacterial species at varying initial fractions, and propagating them through five growth–dilution cycles. To illustrate our approach numerically and graphically over the simplex of all possible intrinsic growth rates (as in Fig. 1), we selected eight systems (see Fig. 2 for details) of three-competing species from the experimental data (i.e. the systems that have only competitive interactions among the species).

Figure 2 shows the location of the direction of the \mathbf{r} -vector inside the unit simplex for each of the eight experimental competition systems. These \mathbf{r} -vectors were inferred by fitting via least-squares the classic LV competition model (eqn 9) to the observed abundance time series of species monocultures (Friedman *et al.*, 2017). Additionally, the figure shows the size of the different potential micro ensembles for each studied system by performing numerical simulations (using Eqn 9) using the inferred \mathbf{A} matrices and 2000 directions of \mathbf{r} -vectors sampled uniformly inside the full parameter space (Song and Saavedra, 2018). Each interaction matrix \mathbf{A} was inferred through pairwise tournaments by fitting via least-squares Eqn. 9 to the observed time series of species abundances (Friedman *et al.*, 2017). Note that different methods to infer the interaction matrix have been introduced in the literature involving regression methods, press or pulse perturbations, Bayesian approaches and machine learning algorithms (Laska and Wootton, 1998; Cao *et al.*, 2016; Dietze, 2017; Martin *et al.*,

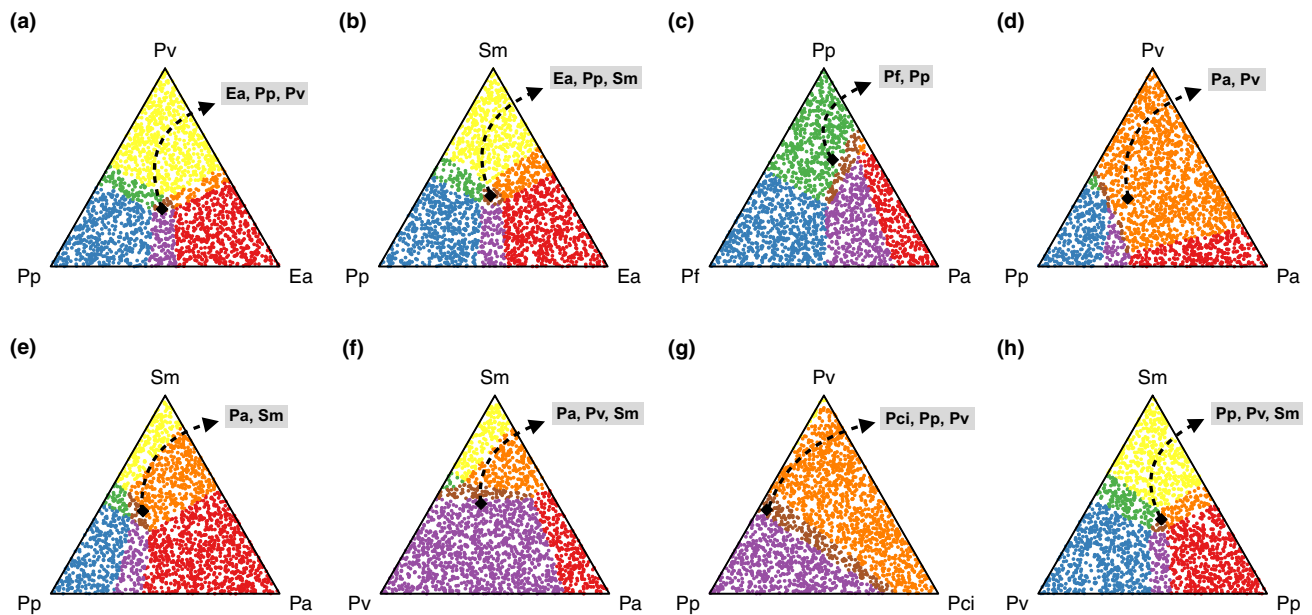


Figure 2 Micro ensembles in experimental data of microbial communities. The figure shows eight different microbial competition systems (a–h) of three species extracted from Friedman *et al.* (2017). Each two-dimensional simplex was formed by running Lotka–Volterra (LV) dynamics using the inferred competition matrix \mathbf{A} from a given system, 2000 different directions of \mathbf{r} -vectors sampled uniformly over the unit simplex and a single initial condition \mathbf{N} . Each point in a given simplex corresponds to the simulation outcome using one \mathbf{r} -vector and each colour represents a given micro ensemble (see Fig. 1). The black diamond on each simplex corresponds to the inferred \mathbf{r} -vector reported in the experimental study. The label associated with the inferred \mathbf{r} -vector corresponds to the ensemble of species that was observed in the majority of LV simulation outcomes (more than 70 out of 100 initial conditions) using all the inferred parameters. Species names: *Enterobacter aerogenes* (Ea), *Pseudomonas aurantiaca* (Pa), *Pseudomonas citronellolis* (Pci), *Pseudomonas fluorescens* (Pf), *Pseudomonas putida* (Pp), *Pseudomonas veronii* (Pv) and *Serratia marcescens* (Sm)

2018; Cenci *et al.*, 2019; Maynard *et al.*, 2020). Interestingly, the figure shows that the majority (five out of eight) of inferred \mathbf{r} -vectors are located within the least probable (smallest) micro ensemble – the one containing all three species (brown points), revealing how highly tuned these experimental systems can be to their environments.

The examples above further imply that predictions on these systems may be done more accurately by comparing the estimated probabilities of persistence $P(i|\mathbf{A})$ (derived from the grand ensemble average) against the threshold λ given by the micro ensemble average formed by all species together. That is, λ can be defined by the expected probability of species persistence $\omega(\mathbf{A})$ within the micro ensemble formed by $E = \{1, \dots, S\}$. This probability can be calculated as $\omega(\mathbf{A}) = P(1, \dots, S|\mathbf{A})^{\frac{1}{S}}$, where $P(1, \dots, S|\mathbf{A})$ corresponds to the probability of feasibility of the full micro ensemble described by the interaction matrix \mathbf{A} . Mathematically, $\omega(\mathbf{A})$ represents the probability of S independent Bernoulli trials (a species is present or not), whose product corresponds to the probability of the full micro ensemble. That is, differently from the probability of persistence of a single species $P(i|\mathbf{A})$, $\omega(\mathbf{A})$ is the expected probability of persistence for any randomly chosen species within the community (Song *et al.*, 2020). Ecologically, this probability can also be interpreted as the expected fraction of species that can persist in a system described by \mathbf{A} , assuming that parameter values (and therefore quantitative states) are all equally likely. This is also analogous to an isomorphic system, where its elements can be treated as independent and as having equal likelihood (Prigogine and Stengers,

1984). Therefore, the threshold $\lambda = \omega(\mathbf{A})$ can be interpreted as the minimum potential (probability) that a randomly chosen species needs to have in order to persist under changing conditions consistent with given restrictions \mathbf{A} .

PROOF OF CONCEPT

We provide a proof of concept to illustrate the usefulness of structural forecasting to predict species persistence in field or experimental settings. Specifically, we follow the methodology described above to make out-of-sample predictions of the competition outcomes of the eight laboratory experiments involving three-competing species (24 individual species outcomes in total; Friedman *et al.*, 2017). We predict that species i persists if $P(i|\mathbf{A}) > \lambda = \omega(\mathbf{A})$. While the values of $P(i|\mathbf{A})$ are analytically calculated, similar results can be obtained by computing probabilities numerically using an extensive sample of different directions of \mathbf{r} -vectors and initial conditions. To illustrate the performance of our predictions, we built a confusion matrix displaying the percentages of true positives, false positives, false negatives and true negatives (Fig. 3a). Figure 3b shows that our predictions match the experimental outcomes (i.e. sum of true positives and true negatives) in 71% of the cases (a one-sided Binomial test $B(24, 0.5)$ for this prediction accuracy gives $P = 0.03$).

To further analyse the performance of our predictions, we used as a benchmark the predictions derived from the best experimental approach (called the assembly rule) proposed by Friedman *et al.*, 2017. This approach establishes that if a

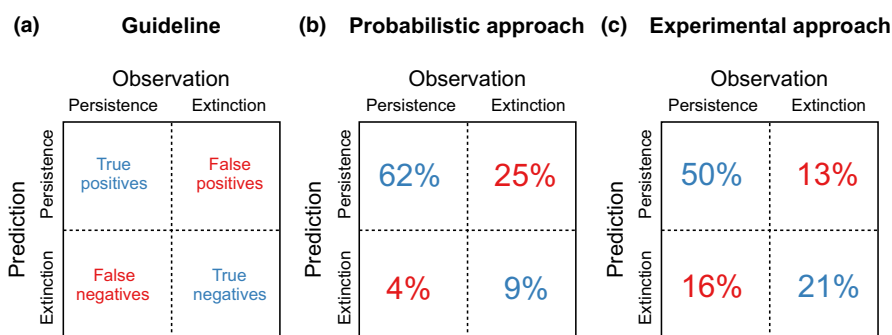


Figure 3 Structural forecasting of species persistence of eight experimental systems of three-competing microbial species. Panel (a) shows how our predictions can be displayed as a confusion matrix containing the fraction of true positives, false positives, false negatives and true negatives. Correct predictions are depicted in blue and wrong predictions in red. We define positives and negatives as cases of species persistence and extinction respectively. The data are composed of eight competition systems of three species (24 individual species outcomes in total) reported in Friedman *et al.* (2017). Panel (b) shows the confusion matrix for our predictions using structural forecasting and the outcomes of the experiment. For comparison purposes, Panel (c) shows the confusion matrix for the predictions using the best experimental approach (the assembly rule) and the outcomes of the experiment (Friedman *et al.*, 2017). Note that the accuracy of predictions can be estimated by the sum of correct predictions (i.e. true positives plus true negatives)

species is out-competed in a pairwise tournament by any of the species forming the potential three-species micro ensemble, then this species will not persist in the three-species tournament. To perform these predictions, we used the experimental pairwise tournaments reported in Friedman *et al.*, 2017 for our studied subset of competing species. Similar to our predictions, Figure 3c shows that these heuristic predictions match

the experiments in 71% of the cases. Interestingly, this experimental approach yields higher percentages for true negatives than our predictions, indicating that this method is more efficient in detecting extinctions. However, the experimental approach performs worse than our probabilistic approach for true positives, indicating that our method has more power to detect surviving species.

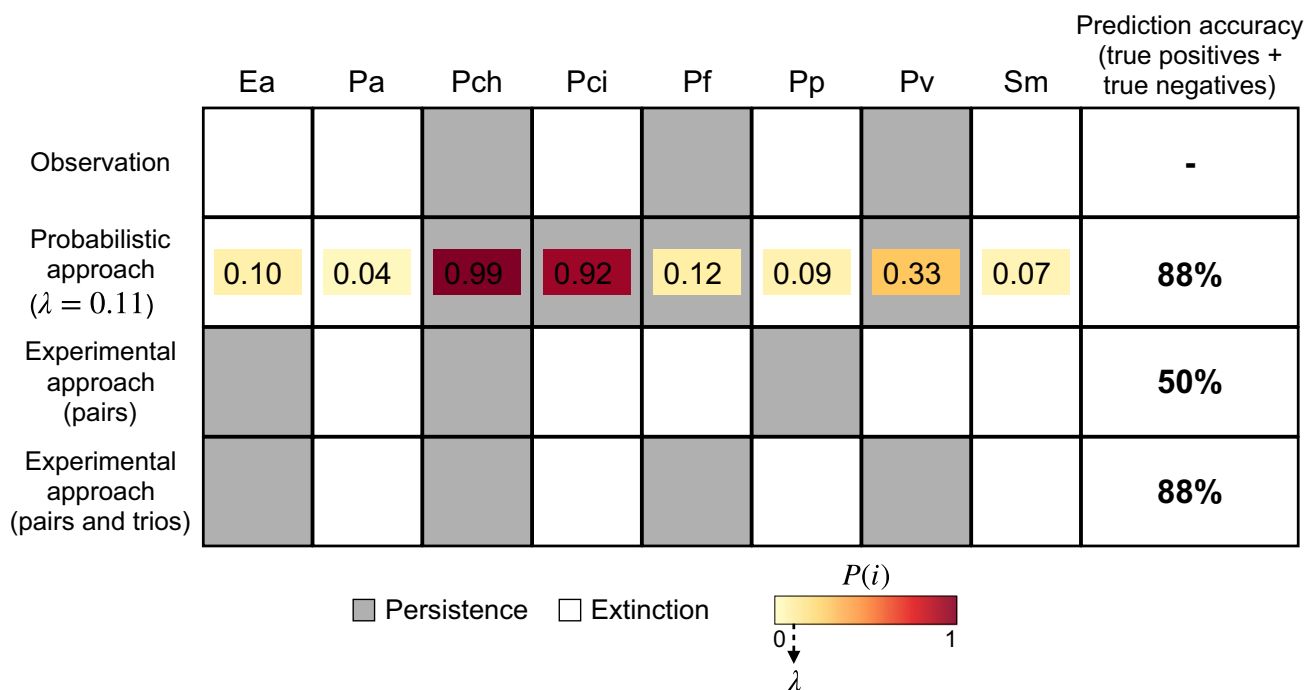


Figure 4 Structural forecasting of species persistence of an experimental system of eight microbial species involving competition and facilitation interactions. The first row (observation) shows the outcome of the experiment as reported in Friedman *et al.* (2017). Grey and white squares correspond to species that persisted or went extinct respectively. The second row shows the predictions from our structural forecasting together with the estimated probability for each species ($P(i)$, the darker the color the higher). We used the interaction matrix inferred from pairwise tournaments reported in Friedman *et al.* (2017). The threshold for extinction was computed as $\lambda = \omega(\mathbf{A}) = 0.11$ (see text for details). The third row shows the predictions from the experimental approach using only pairwise tournaments, while the fourth row shows the predictions from the experimental approach using pairwise and trio tournaments (assembly rules reported in Friedman *et al.* (2017)). The last column shows the overall prediction accuracy (i.e. the percentage of correct predictions) for each approach. Species names: *Enterobacter aerogenes* (Ea), *Pseudomonas aurantiaca* (Pa), *Pseudomonas chlororaphis* (Pch), *Pseudomonas citronellolis* (Pci), *Pseudomonas fluorescens* (Pf), *Pseudomonas putida* (Pp), *Pseudomonas veronii* (Pv) and *Serratia marcescens* (Sm)

Additionally, we tested our probabilistic approach beyond the case of three-competing species. In particular, we tested the outcome of combining eight microbial species involving both competition and facilitation in the pairwise interaction matrix \mathbf{A} as reported in Friedman *et al.*, 2017 (Fig. 4). Note that this interaction matrix was parameterised using pairwise tournaments only. The dimensionality and nature of this eight-species community demanded us to estimate the probabilities ($P(i|\mathbf{A})$) computationally, only the threshold ($\lambda = \omega(\mathbf{A})$) was computed analytically. That is, the probabilities $P(i|\mathbf{A})$ were estimated by running classic LV dynamics with 100 000 directions of \mathbf{r} -vectors sampled uniformly on the positive orthant of the unit ball and random initial conditions. Figure 4 shows that our probabilistic approach correctly predicted seven out of eight cases (recall that we predict that a species persists if $P(i|\mathbf{A}) > \lambda = \omega(\mathbf{A})$).

Finally, we compared our predictions against the predictions from the experimental approach of Friedman *et al.* (2017). Figure 4 shows that using all the pairwise tournaments, the experimental approach correctly predicted four out of eight cases. However, adding the 56 results (plus replicates) from trio tournaments, the experimental approach correctly predicted seven out of eight cases (Fig. 4), revealing the importance of adding information about changing conditions (Friedman *et al.*, 2017). Overall, these findings illustrate that structural forecasting can provide accurate predictions and be as good as the best experimental approaches without the need of extensive or potentially unfeasible experimental work.

FINAL THOUGHTS

Probability is defined as the (conditional) belief about the occurrence of a given event (de Finetti, 2017). This belief is updated according to new available information, providing a measure of the uncertainty we have concerning an event at any given point in time. Similarly, our knowledge about the outcomes (e.g. species extinctions) of ecological communities is uncertain given the several unknown factors affecting ecological and evolutionary dynamics (e.g. governing equations, intrinsic randomness, initial conditions, parameter values and external perturbations), but it has been shown that these outcomes can be approximated following different assumptions, simplifications and inferences (Levins, 1968; Margalef, 1968; Vandermeer, 1970; May, 1976; Sviridov and Logofet, 1983; Sugihara, 1994; Case, 2000; Turchin, 2003). Hence, these uncertainties reveal that we ought to move away from making categorical analyses of ecological communities and embrace probability as a natural approach to understand and predict ecological dynamics (Lewontin, 1969; Cazelles *et al.*, 2016; Dietze, 2017; Song *et al.*, 2020).

Here, we advocate for a probabilistic thinking to understand and predict the persistence of single species embedded in ecological communities under changing environmental conditions. Borrowing concepts from the theory of ensembles applied to statistical mechanics and the mathematics of structural stability applied to population dynamics models, we have introduced a probabilistic approach called structural forecasting and have used numerical and analytical tools to illustrate its power and applicability. While we have

just centred on a small fraction of the possible applications and study systems (i.e. using feasibility in LV competition systems), we hope that future work can adopt probabilistic approaches for understanding and performing out-of-sample predictions of biodiversity changes in different ecological communities. This can involve multitrophic systems, different conditions for persistence, different prediction thresholds and other types of phenomenological or mechanistic models. We believe that structural forecasting can be used to better understand the environmental factors acting on ecological systems, and to be able to predict whether a species invades or remains in a community, a community evolves towards a particular trajectory or a community recovers after a perturbation. In the face of larger and more frequent climatic variations, building probabilistic and testable methodologies to understand and predict the persistence of species under changing conditions is of central importance for developing successful strategies for sustaining entire ecosystems and human health (Stenseth *et al.*, 2002; Walther, 2010; Dirzo *et al.*, 2014; Smith *et al.*, 2014; Lu *et al.*, 2016; Vellend, 2018; Rohr *et al.*, 2020).

ACKNOWLEDGEMENTS

SS acknowledges funding NSF grant No. DEB-2024349. We thank Rudolf P. Rohr, Chuliang Song and Pengjuan Zu for their insightful comments on a previous version.

AUTHOR CONTRIBUTIONS

SS designed the study. All the authors performed the analyses. SS wrote a first version of the manuscript and all the authors contributed with revisions.

DATA ACCESSIBILITY STATEMENTS

The R code and data supporting the results are archived on GitHub: https://github.com/MITEcology/ELE_Saavedra_etal_2020.

REFERENCES

- Adams, M.P., Sisson, S.A., Helmstedt, K.J., Baker, C.M., Holden, M.H., Plein, M. *et al.* (2020). Informing management decisions for ecological networks, using dynamic models calibrated to noisy time-series data. *Ecol. Lett.*, 23, 607–619.
- Alberch, P. (1989). The logic of monsters: evidence for internal constraint in development and evolution. *Geobios*, 22, 21–57.
- Arnold, V.I. (1988). *Geometrical Methods in the Theory of Ordinary Differential Equations 2nd*. Springer, New York.
- Bastolla, U., Lässig, M., Manrubia, S.C. & Valleriani, A., (2005). Biodiversity in model ecosystems, i: coexistence conditions for competing species. *J. Theor. Biol.*, 235, 521–530.
- Boettiger, C. (2018). From noise to knowledge: how randomness generates novel phenomena and reveals information. *Ecol. Lett.*, 21, 1255–1267.
- Cao, H.T., Gibson, T.E., Bashan, A. & Liu, Y.Y. (2016). Inferring human microbial dynamics from temporal metagenomics data: Pitfalls and lessons. *BioEssays*, 39, 1600188.
- Case, T.J. (2000). *An Illustrated Guide to Theoretical Ecology*. Oxford Univ. Press, Oxford.

- Cazelles, K., Mouquet, N., Mouillot, D. & Gravel, D. (2016). On the integration of biotic interaction and environmental constraints at the biogeographical scale. *Ecography*, 39, 921–931.
- Cenci, S., Montero-Castaño, A. & Saavedra, S. (2018). Estimating the effect of the reorganization of interactions on the adaptability of species to changing environments. *J. of Theor. Bio.*, 437, 115–125.
- Cenci, S. & Saavedra, S. (2018a). Structural stability of nonlinear population dynamics. *Phys. Rev. E*, 97, 012401.
- Cenci, S. & Saavedra, S. (2018b). Uncertainty quantification of the effects of biotic interactions on community dynamics from nonlinear time-series data. *J. R. Soc. Interface*, 15, 20180695.
- Cenci, S. & Saavedra, S. (2019). Non-parametric estimation of the structural stability of non-equilibrium community dynamics. *Nature Ecology & Evolution*, 3, 912.
- Cenci, S., Sugihara, G. & Saavedra, S. (2019). Regularized s-map for inference and forecasting with noisy ecological time series. *Methods Ecol. Evol.*, 10, 650–660.
- Clark, A.T., Turnbull, L.A., Tredennick, A., Allan, E., Harpole, W.S., Mayfield, M.M. et al. (2020). Predicting species abundances in a grassland biodiversity experiment: trade-offs between model complexity and generality. *J. Ecol.*, 108, 774–787.
- Clark, J.S., Carpenter, S.R., Barber, M., Collins, S., Dobson, A., Foley, J.A. et al. (2001). Ecological forecasts: An emerging imperative. *Science*, 293, 657–660.
- de Finetti, B. (2017). *Theory of probability: a critical introductory treatment*. Wiley, New York.
- Dietze, M.C. (2017). *Ecological Forecasting*. Princeton Univ, Press.
- Dirzo, R., Young, H.S., Galetti, M., Ceballos, G., Isaac, N.J.B. & Collen, B. (2014). Defaunation in the anthropocene. *Science*, 345, 401–406.
- Einstein, A. (1902). Kinetische theorie der warmegleichgewichtes und des zweiten hauptsatzes der thermodynamik. *Ann. Phys.*, 9, 417–433.
- Ellner, S.P. & Turchin, P. (1995). Chaos in a noisy world: new methods and evidence from time-series analysis. *Am. Nat.*, 145, 343–375.
- Frederickson, M.E. (2017). Mutualisms are not on the verge of breakdown. *Trends Ecol. Evol.*, 32, 727–734.
- Friedman, J., Higgins, L.M. & Gore, J. (2017). Community structure follows simple assembly rules in microbial microcosms. *Nature Ecology & Evolution*, 1, 0109.
- Fukami, T. (2015). Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annu. Rev. Ecol. Evol. Syst.*, 46, 1–23.
- Gibbs, J.W. (1902). *Elementary Principles in Statistical Mechanics*. Charles Scribner's Sons, New York.
- Grainger, T.N. & Gilbert, J.M.L.B. (2019). The invasion criterion: A common currency for ecological research. *Trends Ecol. Evol.*, 34, 925–935.
- Harfoot, M.B.J., Newbold, T., Tittensor, D.P., Emmott, S., Hutton, J., Lyutsarev, V. et al. (2014). Emergent global patterns of ecosystem structure and function from a mechanistic general ecosystem model. *PLoS Biol.*, 12, e1001841.
- Hastings, A. & Powell, T. (1991). Chaos in a three-species food chain. *Ecology*, 72, 896–903.
- Hoek, T.A., Axelrod, K., Biancalani, T., Yurtsev, E.A., Liu, J. & Gore, J. (2016). Resource availability modulates the cooperative and competitive nature of a microbial cross-feeding mutualism. *PLoS Biol.*, 14, e1002540.
- Hofbauer, J. & Sigmund, K. (1998). *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge.
- Kerner, E.H. (1962). Gibbs ensemble and biological ensemble. *Ann. N. Y. Acad. Sci.*, 96, 975–984.
- Laska, M.S. & Wootton, J.T. (1998). Theoretical concepts and empirical approaches to measuring interaction strength. *Ecology*, 79, 461–476.
- Levine, J.M., Bascompte, J., Adler, P.B. & Allesina, S. (2017). Beyond pairwise mechanisms of species coexistence in complex communities. *Nature*, 546, 56.
- Levins, R. (1968). *Evolution in Changing Environments: Some Theoretical Explorations*. Princeton University Press, NJ.
- Lewontin, R.C. (1969). The meaning of stability. *Brookhaven Symp. Biol.*, 22, 13–24.
- Lu, X., Gray, C., Brown, L.E., Ledger, M.E., Milner, A.M., Mondragon, R.J. et al. (2016). Drought rewires the cores of food webs. *Nature Climate Change* page, 6(9), 875–878. <https://doi.org/10.1038/nclimate3002>.
- Margalef, R. (1968). *Perspectives in Ecological Theory*. University of Chicago Press, Chicago.
- Martin, B.T., Munch, S.B. & Hein, A.M. (2018). Reverse-engineering ecological theory from data. *Proc. R. Soc. B*, 285, 0180422.
- May, R.M. (1976). Simple mathematical models with very complicated dynamics. *Nature*, 261, 459–467.
- May, R.M. (2004). Uses and abuses of mathematics in biology. *Science*, 303, 790–793.
- Maynard, D.S., Miller, Z.R. & Allesina, S. (2020). Predicting coexistence in experimental ecological communities. *Nature Ecology & Evolution*, 4, 91–100.
- McKane, A.J., Biancalani, T. & Rogers, T. (2014). Stochastic pattern formation and spontaneous polarisation: The linear noise approximation and beyond. *Bull. Math. Biol.*, 76, 895–921.
- Meszéna, G., Gyllenberg, M., Pásztor, L. & Metz, J.A.J. (2006). Competitive exclusion and limiting similarity: a unified theory. *Theor. Popul. Biol.*, 69, 68–87.
- Michaelian, K. (2005). Thermodynamic stability of ecosystems. *J. Theor. Biol.*, 237, 323–335.
- Odum, E.P. (1969). The strategy of ecosystem development. *Science*, 164, 262–270.
- Pacala, S.W., Canham, C.D., Saponara, J., Silander, J.A., Kobe, R.K. & Ribbens, E. (1996). Forest models defined by field measurements: Estimation, error analysis and dynamics. *Ecol. Monogr.*, 66, 1–43.
- Pearl, J. (2009). *Causality*. Cambridge University Press, Cambridge.
- Petchey, O.L., Pontarp, M., Massie, T.M., Kéfi, S., Ozhul, A., Weilenmann, M. et al. (2015). The ecological forecast horizon, and examples of its uses and determinants. *Ecol. Lett.*, 18, 597–611.
- Pimm, S.L. (1991). *The balance of nature. Ecological Issues in the conservation of species and communities*. University of Chicago, Chicago.
- Planck, M. (1925). *A Survey of Physical Theory*. Dover Publications, New York.
- Poincaré, H. (1905). *Science and Hypothesis*. The Walter Scott Publishing, New York.
- Preston, F.W. (1962). The canonical distribution of commonness and rarity: Part i. *Ecol. Monogr.*, 43, 185–215.
- Prigogine, I. (1962). *Introduction to Nonequilibrium Statistical Mechanics*. Interscience, New York.
- Prigogine, I., George, C., Henin, F. & Rosenfeld, L. (1973). A unified formulation of dynamics and thermodynamics. *Chemica Scripta*, 4, 5–32.
- Prigogine, I. & Stengers, I. (1984). *Order out of Chaos: Man's new dialogue with nature*. Bantam, New York.
- Rohr, J.R., Civitello, D.J., Halliday, F.W., Hudson, P.J., Lafferty, K.D., Wood, C.L. & et al. (2020). Towards common ground in the biodiversity – disease debate. *Nature Ecology & Evolution*, 4, 24–33.
- Rohr, R.P., Saavedra, S. & Bascompte, J. (2014). On the structural stability of mutualistic systems. *Science*, 345, 1253497.
- Rohr, R.P., Saavedra, S., Peralta, G., Frost, C.M., Bersier, L.-F., Bascompte, J. et al. (2016). Persist or produce: A community trade-off tuned by species evenness. *Am. Nat.*, 188, 411–422.
- Rohr, R.P., Scherer, H., Kehrl, P., Mazza, C. & Bersier, L.F. (2010). Modeling food webs: Exploring unexplained structure with latent traits. *Am. Nat.*, 176, 170–177.
- Saavedra, S., Cenci, S., del Val, E., Boege, K. & Rohr, R.P. (2017a). Reorganization of interaction networks modulates the persistence of species in late successional stages. *J. of Animal Ecology*, 86, 1136–1146.
- Saavedra, S., Rohr, R.P., Bascompte, J., Godoy, O., Kraft, N.J. & Levine, J.M. (2017b). A structural approach for understanding multispecies coexistence. *Ecol. Monogr.*, 87, 470–486.

- Saavedra, S., Rohr, R.P., Gilarranz, L.J. & Bascompte, J. (2014). How structurally stable are global socioeconomic systems? *J. R. Soc. Interface*, 11, 20140693.
- Schreiber, S.J., Yamamichi, M. & Strauss, S.Y. (2019). When rarity has costs: coexistence under positive frequency-dependence and environmental stochasticity. *Ecology*, 100, e02664.
- Serván, C.A., Capitán, J.A., Grilli, J., Morrison, K.E. & Allesina, S. (2018). Coexistence of many species in random ecosystems. *Nature Ecology & Evolution*, 2, 1237.
- Shmueli, G. (2010). To explain or to predict? *Stat. Sci.*, 25, 289–310.
- Smale, S. (1967). Differentiable dynamical systems. *Bull. Amer. Math. Soc.*, 73, 747–817.
- Smith, M.B., Kelly, C. & Alm, E.J. (2014). How to regulate faecal transplants. *Nature*, 506, 290–291.
- Solé, R.V. & Valls, J. (1992). On structural stability and chaos in biological systems. *J. Theor. Biol.*, 155, 87–102.
- Song, C., Ahn, S.V., Rohr, R.P. & Saavedra, S. (2020). Towards a probabilistic understanding about the context-dependency of species interactions. *Trends Ecol. Evol.*, 35, 384–396.
- Song, C., Altermatt, F., Pearse, I. & Saavedra, S. (2018a). Structural changes within trophic levels are constrained by within-family assembly rules at lower trophic levels. *Ecol. Lett.*, 21, 1221–1228.
- Song, C., Rohr, R.P. & Saavedra, S. (2017). Why are some plant-pollinator networks more nested than others? *J. Anim. Ecol.*, 86, 1417–1424.
- Song, C., Rohr, R.P. & Saavedra, S. (2018). b. A guideline to study the feasibility domain of multi-trophic and changing ecological communities. *J. of Theoretical Biology*, 450, 30–36.
- Song, C. & Saavedra, S. (2018). Will a small randomly-assembled community be feasible and stable? *Ecology*, 99, 743–751.
- Stenseth, N.C., Mysterud, A., Ottersen, G., Hurrell, J.W., Chan, K.S. & Lima, M. (2002). Ecological effects of climate fluctuations. *Science*, 297, 1292–1296.
- Strogatz, S.H. (2014). *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Westview press, Boulder.
- Sugihara, G. (1994). Nonlinear forecasting for the classification of natural time series. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 348, 477–495.
- Sugihara, G., May, R., Ye, H., Hsieh, C.-H., Deyle, E., Fogarty, M. *et al.* (2012). Detecting causality in complex ecosystems. *Science*, 338, 496–500.
- Svirezhev, Y.M. & Logofet, D.O. (1983). *Stability of Biological Communities*. Mir Publishers, Moscow.
- Täuber, U.C. (2011). Stochastic population oscillations in spatial predator-prey models. *J. Phys: Conf. Ser.*, 319, 012019.
- Thom, R. (1972). *Stabilité structurelle et morphogénèse*. InterÉditions, Paris.
- Tuck, S.L., Porter, J., Rees, M. & Turnbull, L.A. (2018). Strong responses from weakly interacting species. *Ecol. Lett.*, 21, 1845–1852.
- Turchin, P. (2003). *Complex Population Dynamics: A Theoretical/Empirical Synthesis*. Princeton University Press, Princeton.
- Valdovinos, F.S. (2019). Mutualistic networks: moving closer to a predictive theory. *Ecol. Lett.*, 22, 1517–1534.
- Vandermeer, J.H. (1970). The community matrix and the number of species in a community. *Am. Nat.*, 104, 73–83.
- Vano, J.A., Wildenberg, J.C., Anderson, M.B., Noel, J.K. & Sprott, J.C. (2006). Chaos in low-dimensional lotka–volterra models of competition. *Nonlinearity*, 19, 2391.
- Vellend, M. (2016). *The Theory of Ecological Communities*. Princeton University Press, NJ.
- Vellend, M. (2018). Are local losses of biodiversity causing degraded ecosystem function? In: *Effective conservation science: data not dogma* (eds Kareiva, P., Marvier, M. & Silliman, B.). Oxford University Press, Oxford.
- Walther, G.R. (2010). Community and ecosystem responses to recent climate change. *Phil. Trans. R. Soc. B*, 365, 2019–2024.
- Yang, Q., Fowler, M.S., Jackson, A.L. & Donohue, I. (2019). The predictability of ecological stability in a noisy world. *Nature Ecology & Evolution*, 3, 51–259.
- Yodzis, P. & Innes, S. (1992). Body size and consumer-resource dynamics. *Am. Nat.*, 132, 1151–1175.

Editor, Carl Boettiger

Manuscript received 12 March 2020

First decision made 7 June 2020

Manuscript accepted 8 July 2020