INTERFACE

royalsocietypublishing.org/journal/rsif

Research

Cite this article: AlAdwani M, Saavedra S. 2020 Ecological models: higher complexity in, higher feasibility out. *J. R. Soc. Interface* **17**: 20200607.
http://dx.doi.org/10.1098/rsif.2020.0607

Received: 27 July 2020
Accepted: 28 October 2020

Subject Category:
Life Sciences–Mathematics interface

Subject Areas:
biocomplexity, biomathematics, computational biology

Keywords:
ecological systems, probability of feasibility, free-equilibrium points, functional responses, higher-order interactions, nonlinear dynamics

Author for correspondence:
Serguei Saavedra
e-mail: sersaa@mit.edu

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5202165.

THE ROYAL SOCIETY PUBLISHING

Ecological models: higher complexity in, higher feasibility out

Mohammad AlAdwani and Serguei Saavedra

Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

SS, 0000-0003-1768-363X

Finding a compromise between tractability and realism has always been at the core of ecological modelling. The introduction of nonlinear functional responses in two-species models has reconciled part of this compromise. However, it remains unclear whether this compromise can be extended to multispecies models. Yet, answering this question is necessary in order to differentiate whether the explanatory power of a model comes from the general form of its polynomial or from a more realistic description of multispecies systems. Here, we study the probability of feasibility (the existence of at least one positive real equilibrium) in complex models by adding higher-order interactions and nonlinear functional responses to the linear Lotka–Volterra model. We characterize complexity by the number of free-equilibrium points generated by a model, which is a function of the polynomial degree and system's dimension. We show that the probability of generating a feasible system in a model is an increasing function of its complexity, regardless of the specific mechanism invoked. Furthermore, we find that the probability of feasibility in a model will exceed that of the linear Lotka–Volterra model when a minimum level of complexity is reached. Importantly, this minimum level is modulated by parameter restrictions, but can always be exceeded via increasing the polynomial degree or system's dimension. Our results reveal that conclusions regarding the relevance of mechanisms embedded in complex models must be evaluated in relation to the expected explanatory power of their polynomial forms.

1. Introduction

Understanding and predicting the behaviour of ecological systems has been one of the greatest challenges in ecological research [1–4]. One promising route to accomplish this challenge has been based on the mathematical modelling of species abundances over time by assuming different functions of species interactions, growth and decline rates [5]. However, these terms are not uniquely represented: they are either arbitrarily or specifically chosen to provide tractability (such as the ability to analytically understand the effect of a change in a parameter) and preserve realism (such as mimicking as much as possible ecological mechanisms) [6]. Indeed, in principle, a tractable, realistic, mathematical model of a system can allow us to apply conventional methods to deduce and have a mechanistic knowledge about the behaviour of real-world systems [7,8]. Yet, finding a compromise between tractability and realism has not been easy [9–11].

Importantly, it has already been shown that in order to explain complex dynamics, it is not always necessary to have complex models [12]. For example, complex behaviour, such as transitions from point attractors to chaotic behaviour, can already emerge from population dynamics models with low-order polynomials (e.g. the one-dimensional deterministic logistic model) [13]. In fact, one of the best examples of simple tractable models in ecology is the well-known linear Lotka–Volterra (LV) model [14,15]. Yet, this model must be understood just as a first-order approximation to how complex ecological systems behave [8]. As a consequence, many modifications have been done to the linear LV model in the hope of adding realism and increasing their explanatory power [5]. In general, these modifications yield models of the form

$\mathrm{d}N_i/\mathrm{d}t = N_i f_i(\boldsymbol{N})/q_i(\boldsymbol{N})$ for $i = 1, 2, \ldots, n$, where the f's and q's are multivariate polynomials (in general with higher-order terms) in species abundances $\boldsymbol{N} = (N_1, N_2, \ldots, N_n)^{\mathrm{T}}$ [12].

A clear example of complexity added to the linear LV model is the incorporation of higher-order interactions (HOIs) that involve more than two variables [16]. The introduction of these higher-order terms has been justified in order to account for the possibility that the effect of a species i on the *per capita* growth rate of a species j might itself depend on the abundance of a third species k due to either compensatory effects or supra-additivity [5,17]. The addition of HOIs has been shown to stabilize dynamics in competition systems [18], promote diversity in ecological communities [19] and capture unexplained dynamics of linear LV models [20]. However, it has been debatable whether these terms are derived from fundamental principles [21], whether mathematically there is anything to prevent their inclusion into ecological models [22], or whether they are indeed useful to explain observed ecological dynamics [23].

Another example of complexity is the addition of functional responses, which have been one of the most studied and ecologically motivated polynomial fraction forms added to linear LV models [5,6,23,24]. Typically, functional responses correspond to the mechanistic (or phenomenological) description of how predators (consumers) search, attack and handle their prey (resources). Although the name of functional response was first introduced by Solomon [25], functional responses were broadly adopted after Holling [24] identified three types of responses: linear (Type I; linear LV model), hyperbolic (Type II) and sigmoid (Type III). For instance, the Beddington–DeAngelis functional response [26,27], which is a variation of Type II, has been one of the most widely used responses for modelling food webs [28]. Importantly, the introduction of functional responses has appeared to reconcile part of the compromise between tractability and realism across a variety of ecological models [5]. Yet, most of the analytical (tractable) work incorporating nonlinear functional responses (Types II and III) has been limited to two-species systems [5,6,24], remaining unclear whether this compromise can be extended to larger multispecies cases [22,29].

In general, one of the big questions derived from the addition of complexity (e.g. either HOIs or functional responses) is whether the explanatory power of a modified model comes from the general form of its polynomial or from a more realistic description of multispecies systems. To answer this question, we study the probability of feasibility in complex models (i.e. modifications to the linear LV model using multivariate polynomials) under an arbitrary choice of parameter values. Note that the observability or adaptability of an ecological system is associated with how much its structure can change while retaining its feasibility [30,31]. Thus, it is important to distinguish the minimum amount of information necessary in a model to explain such observability. Specifically, we study the probability of feasibility as a function of three key properties of these complex models: their polynomial degree (interaction order), dimension (number of species) and parameter restrictions (sign restrictions). We define the probability of feasibility as the frequency of finding in a model at least one feasible solution (i.e. a feasible free-equilibrium point where all its coordinates are real and positive) by randomly choosing parameter values under a given distribution [32–34]. Note that the existence of feasible equilibrium solutions is a crucial condition in the context of species coexistence, i.e. a necessary condition for persistence, permanence and the existence of bounded orbits in the feasibility domain [35].

We start illustrating our study using a one-dimensional toy model and demonstrating that its probability of feasibility increases as a function of its polynomial degree (and consequently its number of parameters) when parameter values are arbitrarily chosen from a given probability distribution. Next, we extend the toy example into a multidimensional case to show that the probability of generating a feasible multispecies system is an increasing function of its complexity. Specifically, we characterize complexity by the number of free-equilibrium points generated by a model, which is a function of the polynomial degree and system's dimension. Then, to illustrate the expected behaviour of complex models across different dimensions and parameter restrictions, we study modifications to the linear LV model using HOIs and functional responses. Finally, we discuss the implications of our results for the explanatory contribution to feasibility of complex ecological models.

2. Univariate complex models: conceptual illustration

To investigate the probability of feasibility of ecological systems using complex models, we start illustrating our methodology in one-dimensional (univariate) systems. For this purpose, let us consider the following one-dimensional dynamical system characterized by the state variable N as shown below:

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \frac{Nf(N)}{q(N)}, \tag{2.1}$$

where $f(N) = a_m N^m + a_{m-1} N^{m-1} + \cdots + a_1 N + a_0$ is a polynomial of degree m and $q(N)$ can be any other polynomial that shares no common factor with $Nf(N)$. Note that in the case when $f(N)$ is linear and $q(N) = 1$, we recover the one-dimensional version of the linear LV model (i.e. logistic growth model when $a_1 < 0$ [5]).

As mentioned before, we study the feasibility of a system as defined by its capacity to have at least one feasible equilibrium point (the equilibrium point is both real and positive) under an arbitrary choice of parameter values. This implies that the feasibility problem of Model (2.1) is identical to the feasibility problem of the system defined by the model $\mathrm{d}N/\mathrm{d}t = Nf(N)$, as they both involve analysing the real and positive roots of the polynomial $f(N)$. Therefore, we can think of the feasibility problem in Model (2.1) as the same as the feasibility problem of the modified one-dimensional linear LV model with higher-order terms. Note that the dynamical stability criterion can be relaxed in this case, as it is linked to the feasibility problem [36]. That is, when Model (2.1) has k positive equilibrium points (without multiple or complex roots), a stable feasible free-equilibrium point is followed by an unstable one, making the number of positive stable equilibrium points to be either floor or ceil $k/2$ for $k = 0, 1, \ldots, m$ [22]. This implies that one can derive the stability problem from the feasibility one.

It is well known that the feasibility of any system depends on the specifics given by the model parametrization and constraints [37]. However, in the absence of information about the exact parameter values, as in most of the ecological research, these values are randomly chosen from a probability distribution [11,38]. This parameter uncertainty transfers the feasibility problem to the probability of having at least one feasible equilibrium point by randomly choosing parameter values under given conditions [32,39]. For illustration purposes, let us consider the case when the a's are all Gaussians i.i.d. centred on zero (mean zero), and let us denote $p_G(m)$ the probability that at least one root of $f(N)$ is positive (i.e. feasible). Note that $p_G(m)$ is independent of the distribution's variance simply because $f(N)$ and $cf(N)$ have identical roots for any constant $c \neq 0$. Under this Gaussian case, it has been demonstrated [40] that the expected number of positive real roots $E(m)$ as $m \to \infty$ is given by

$$E(m) = \frac{R(m)}{2}, \tag{2.2}$$

where

$$R(m) = \frac{2}{\pi}\log(m) + 0.6257358072\ldots + O\left(\frac{1}{m}\right). \tag{2.3}$$

Note that $R(m)$ corresponds to the expected number of real roots, while $E(m)$ assumes no *a priori* tendency for positive or negative roots in $R(m)$. That is, the density of real zeros is an even function [40].

Next, let us try to find how equation (2.2) can be inserted into the expression of the probability of feasibility $p_G(m)$. To provide a numerical approximation, let us assume that the location of the m roots of $f(N)$ are independent of each other and positive with probability p_i for $i = 1, 2, \ldots, m$. Therefore, the probability of feasibility (to have at least one positive and real root) becomes $p_G(m) = 1 - (1 - p_1)(1 - p_2) \ldots (1 - p_m)$. By applying Jensen's inequality (i.e. $f(p_1) + f(p_2) + \cdots + f(p_m) \geq mf((p_1 + p_2 + \cdots + p_m)/m)$) to the convex function $f(x) = -\log(1 - x)$, we obtain $p_G(m) \geq 1 - (1 - E(m)/m)^m$, where the expected number of positive roots $E(m) = p_1 + p_2 + \cdots + p_m$. From the formula of $p_G(m)$, one can derive the upper bound $p_G(m) \leq 1 - (1 - \max(p_1, p_2, \ldots, p_m))^m$. Assuming that $p_G(m)$ is continuous for any $m \geq 1$ implies the existence of $\hat{E}(m)$ such that the probability of feasibility can be written as

$$p_G(m) = 1 - \left(1 - \frac{\hat{E}(m)}{m}\right)^m, \tag{2.4}$$

where $\hat{E}(m)$ is an overestimate of the expected number of positive roots (that is, $E(m) \leq \hat{E}(m) \leq m \max(p_1, p_2, \ldots, p_m)$). This allows us to infer the mathematical form of $E(m)$ by finding $\hat{E}(m)$ such that $p_G(m)$ is the best fit of $1 - (1 - \hat{E}(m)/m)^m$.

Figure 1 provides a numerical confirmation of the positive relationship between the probability of feasibility $p_G(m)$ and the degree m of the polynomial $f(N)$ under an arbitrary choice of parameter values (no parameter restrictions). The probability is calculated numerically over 10^4 simulations using i.i.d. parameters from a Gaussian distribution with mean zero and standard deviation one. The figure shows the best fit to the data using equation (2.4), where $\hat{E}(m) = a\log(m) + b + c/m$, $a = 0.391$, $b = 0.356$ and $c = 0.141$. Note that these values are close (still an overestimate) to $1/\pi$ and to the constant term in equation (2.2). Also, notice the sharp increase in the probability of feasibility for small m's. That is, when the polynomial degree m is relatively small, only a few extra terms are needed to add a noticeable increase in the probability of feasibility. Once m is large enough, the rate of increase in the probability of feasibility diminishes sharply even if considerably more extra terms are added.

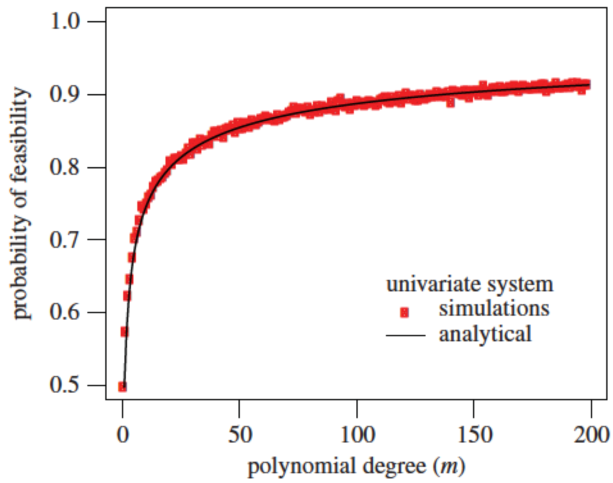

Figure 1. Probability of feasibility increases as a function of the polynomial degree in one-dimensional systems. The figure shows the probability that at least one root is feasible $p_G(m)$ in Model (2.1) as a function of the degree $m \geq 1$ of the polynomial $f(N)$ (using 10^4 trial points for each polynomial degree m). Note that parameters are all i.i.d. Gaussian with mean zero and standard deviation one. The probability $p_G(m)$ is independent of the distribution's variance. By plotting (solid line) $p_G(m)$ and fitting it with $1 - (1 - \hat{E}(m)/m)^m$ (where $\hat{E}(m) = a\log(m) + b + c/m$), we find that the best-fit parameters ($R^2 = 0.9966$) are $a = 0.391 \pm 0.004$ (an overestimate value that is close to $1/\pi$ in the expression of $E(m)$ in equations (2.2) and (2.3)), $b = 0.356 \pm 0.013$ (an overestimate value that is close to the constant term in the expression of $E(m)$ in equations (2.2) and (2.3)) and $c = 0.141 \pm 0.016$. The term c/m, which is present in both $E(m)$ (as an order quantity $O(1/m)$) and $\hat{E}(m)$, is not significant for large polynomial degrees m (as it is small compared to either the log or the constant term in both $\hat{E}(m)$ and $E(m)$). However, the best fit of c takes care of fitting the probability of feasibility with low polynomial degrees without altering the fact that the values of a and b in $\hat{E}(m)$ are overestimate values and are close to the ones in the expression of $E(m)$.

Importantly, the example above illustrating a monotonic and saturating behaviour of the probability of feasibility as a function of the polynomial degree is robust to the choice of the probability distribution (see the electronic supplementary material). This is true as long as the addition of parameters does not decrease the probability of obtaining an odd sign sequence in the coefficients of $f(N)$ (i.e. the number of consecutive sign changes in $a_m, a_{m-1}, \ldots, a_1, a_0$ is an odd integer—see electronic supplementary material).

3. Complex multispecies models

To investigate whether the probability of feasibility in complex multispecies models has similar patterns to those shown in one-dimensional models, we focus on the multivariate generalization of Model (2.1):

$$\frac{dN_i}{dt} = \frac{N_i f_i(N)}{q_i(N)}, \quad i = 1, \ldots, n, \tag{3.1}$$

where $\mathbf{N}$ is a vector of species abundances. Model (3.1) can be characterized by two quantities: its number of free-equilibrium points and the joint distribution of its parameters.

Equilibrium points (known as N_i^*) are the solutions to all N_i in equation (3.1) when the left-hand side of the equation is equal to zero. These equilibrium points can be classified as free or rigid [22]. Free-equilibrium points have non-zero components (can be complex) and can move freely within the state space as a function of parameter values, while rigid-equilibrium points are restricted in space such that they contain at least one zero (i.e. $N_i^* = 0$). That is, rigid-equilibrium points are restricted to particular regions of the state space regardless of the values that model parameters can take and contain at least one zero coordinate (i.e. boundary-equilibrium points). Instead, the locations of free-equilibrium points are not restricted in space and are completely dependent on model parameters. This implies that only free-equilibrium points can lead to feasible systems (i.e. $N_i^* > 0$ for $i = 1, \ldots, n$). These definitions further reveal that the number of free-equilibrium points (Θ) is, in fact, the multivariate generalization of the polynomial degree m seen in the one-dimensional case.

Following the definitions above, we use Θ as the measure of complexity of a model. When parameters of $f_i(N)$ are independent for all i, Θ can be analytically obtained by computing the number of complex roots of equation (3.1) [22,41]. However, when parameters are not independent, that Θ becomes an upper bound and the exact value can be computed using the PHClab software package [42]. Note that other measures of complexity have been used in the literature [6,43]. However, these other measures are either at the level of system complexity (such as dimensionality or connectivity) or at the level of assumed mechanisms in a model (e.g. Type I versus Type II functional responses). Instead our measure of complexity makes no prior assumption about the complexity of a model, but integrates all this information to provide a measure of the enrichment in dynamics that can be derived from a model [22].

Assuming equation (2.4) as the expression for the probability of feasibility in multidimensional systems where m is replaced by Θ, $p(\Theta)$ can be further simplified under two key observations. (i) The number of free-equilibrium points Θ is expected to be large in multidimensional systems. This observation has been shown for LV models with HOIs under an arbitrary choice of parameter values, where Θ increases exponentially with the dimension of the system [22]. (ii) The overestimate $\hat{E}(\Theta)$ is very small compared to Θ (i.e. $\hat{E}(\Theta)/\Theta \ll 1$). This second observation has been shown for one-dimensional systems with standard Gaussian distributions (see previous section). Specifically, $E(m) \approx \log(m)/\pi$, which is much smaller than m for large m. Building on these two observations, we can rewrite the probability of feasibility in multidimensional systems as

$$p(\Theta) \approx 1 - \exp(-\hat{E}(\Theta)). \tag{3.2}$$

The goodness in the approximation of equation (3.2) to equation (2.4) can be shown by sampling over the Θ–$\hat{E}$ space. For example, setting $\Theta = 1000$ and $\hat{E} = 1$, the evaluated expression $1 - (1 - \hat{E}(\Theta)/\Theta)^{\Theta} = 0.632305$ is close to the evaluated expression $1 - \exp(-\hat{E}(\Theta)) = 0.63212$. Importantly, the joint distribution of parameters in Model (3.1) affects how $\hat{E}$ is related to Θ. As we have shown, in one-dimensional systems with i.i.d. probability distributions, $\hat{E}$ increases with the polynomial degree m. Thus, following equation (3.2), the assumption $\hat{E}(\Theta_1) > \hat{E}(\Theta_2)$ when $\Theta_1 > \Theta_2$ implies that $p(\Theta_1) > p(\Theta_2)$. That is, we expect that $p(\Theta)$ increases with Θ in multidimensional systems as well. However, as in the univariate case, there are also exceptions to this pattern. Specifically, when $\Theta > 1$ is small, the relationship $\hat{E}(\Theta) > \hat{E}(1)$ can be violated. This is because models with $\Theta = 1$ and $\Theta > 1$ are fundamentally different. Using arbitrary model parameters with the LV model (i.e. $\Theta = 1$), the solo free-equilibrium point must have real coordinates. Instead, for complex models (i.e. $\Theta > 1$) free-equilibrium points are generally complex [22]. Thus, when Θ is small but $\Theta > 1$, the comparison between $\hat{E}(\Theta)$ and $\hat{E}(1)$ is unclear and becomes dependent on the distribution. However, $\hat{E}(\Theta) > \hat{E}(1)$ is expected to hold with the increase of free-equilibrium points. These results show that the complexity of a model can be characterized by its number of free-equilibrium points (Θ), which are a function of the polynomial degree and system's dimension.

As in the univariate case, stability in the multivariate case is related to the feasibility problem [36]. In multidimensional systems, we do not necessarily need an asymptotically stable free equilibrium point for the existence of species coexistence [35,44]. For example, species coexistence can be possible when there exists both a trajectory from the initial condition towards any of the feasible free equilibria and if after some sufficiently large time, the maximum distance between the feasible equilibrium and the trajectory is bounded [45]. That is, in the presence of an attracting direction in any of the feasible free-equilibrium points, species coexistence is possible given an appropriate initial condition [22]. Thus, as we have discussed for the one-dimensional case (where a stable equilibrium is typically followed by an unstable one), increasing the number of feasible free-equilibrium points alone increases on average the probability of the existence of at least one trajectory compatible with such points. For example, let us assume a scenario where we have all repelling feasible free-equilibrium points (i.e. given an unstable system), then adding an extra feasible free-equilibrium point increases the probability of the existence of a non-repeling direction and consequently attaining coexistence. In the next sections, we test our hypotheses above by illustrating the expected behaviour of specific multidimensional models using modifications to the linear LV model with HOIs and nonlinear functional responses.

3.1. Higher-order interactions

The multidimensional model with HOIs can be generally written as [23]

$$\left.\begin{aligned} \frac{dN_i}{dt} &= N_i\Big(r_i + \sum_{l=1}^{m'-1} \sum_{1\le j_1 \le j_2 \le \ldots \le j_l \le n} a_{ij_1 j_2,\ldots,j_l} N_{j_1} N_{j_2} \ldots N_{j_l}\Big), \\ i &= 1, \ldots, n, \end{aligned}\right\} \tag{3.3}$$

where the r's represent species growth rates, m' is the interaction order (with $m' = 2$ we recover the linear LV model), double indexed a's represent pairwise species interaction coefficients, and the remaining a's correspond to HOIs. Note that the feasibility problem in Model (3.3) is identical to the feasibility problem of $dN_i/dt = N_i f_i(N)$—which is the same as the feasibility problem studied in Model (3.1). That is, in both cases, the feasibility problem involves solving the multivariate polynomial system $f_i(N^*) = 0$ for $i = 1, 2, \ldots, n$. Hence, without loss of generalization, we can think of the feasibility problem of a general fractional polynomial system as that of an LV model with HOIs.

Following our analysis of one-dimensional systems, let us assume that in Model (3.3), the r's and the a's are all Gaussians i.i.d. with mean zero (the variance does not affect the probability of feasibility or the location of

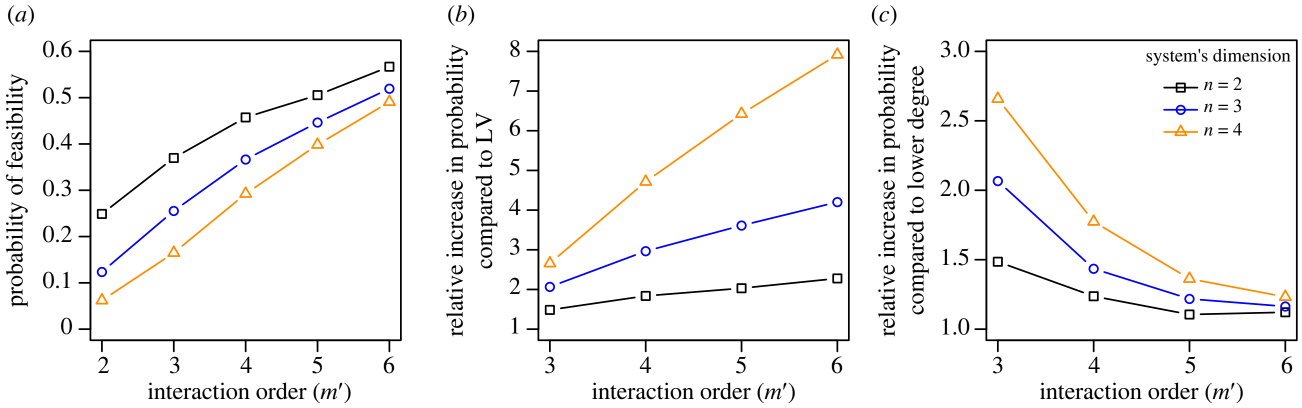

Figure 2. Probability of feasibility increases as a function of systems' dimension and polynomial degree. Panel (*a*) shows the probability of feasibility (i.e. $p_G(n, m')$) as a function of system size n and interaction order m' in the linear Lotka–Volterra (LV) model (i.e. $p_G(n, m' = 2)$) and modifications with higher-order interactions (i.e. $p_G(n, m' > 2)$). In general, probabilities decrease with system size, but increase as a function of interaction order. Panel (*b*) shows the relative increase in the probability of feasibility of modified models compared to the linear LV model (i.e. $p_G(n, m')/p_G(n, 2)$). The higher the system size and interaction order, the higher the relative increase. Panel (*c*) shows the relative increase in the probability of feasibility of modified models compared to lower degree models (i.e. $p_G(n, m')/p_G(n, m' - 1)$) for $n = 2, 3, 4$ species communities. The higher the interaction order, the higher the saturation and, in turn, the lower the relative increase. Note that parameters (r's and the a's) in equation (3.3) are all assumed to be standard Gaussians i.i.d. with zero mean and unit variance.

free-equilibrium points since for any constants $c \neq 0$, the roots of the multivariate polynomial system $cf_i(N^*) = 0$ for $i = 1, 2, \ldots, n$ do not change). For illustration purposes, let us consider multispecies systems of dimension two, three, and four (i.e. $n = 2, 3, 4$) with interaction order given by $m' = 2, 3, 4, 5, 6$. Then, we define $p_G(n, m')$ as the probability of feasibility with n species and interaction order m'. The probability of feasibility is calculated using the PHClab package [42], which numerically solves the polynomial system defined by equation (3.3) after setting $\mathrm{d}N_i/\mathrm{d}t = 0$ and deleting N_i from the right-hand side (r.h.s.). Under a generic choice of parameter values, it has been shown [22] that the number of free-equilibrium points is given by $\Theta = (m' - 1)^n$. It is also well known that in the case of the linear LV model (i.e. $m' = 2$, $\Theta = 1$) under an arbitrary choice of parameter values (distribution centred on zero), the probability of feasibility is given by $p_G(n, 2) = 1/2^n$ for all n [46,47]. However, as in the one-dimensional case, figure 2*a* shows that when HOIs are added, the probability of feasibility increases as a function of the polynomial degree m'. The figure also shows that if two multispecies models have the same polynomial degree m', the one with the lower dimension n exhibits a higher probability of feasibility (i.e. $p_G(n_1, m') > p_G(n_2, m')$ for all m' if and only if $n_1 < n_2$). This result can be expected from the fact that the probability of feasibility in a system decreases on average as the number of species increases [47,48]. Note that if two multispecies models have the same interaction order m', but different dimension n, they also differ in the number of model parameters and free-equilibrium points Θ.

Next, we use the results above to study how the explanatory power of feasibility with complex models changes relative to the linear LV model. Figure 2*b* shows that multispecies models with the same interaction order $m' > 2$ exhibit a relative increase in the probability of feasibility compared to the linear LV model (i.e. $p_G(n, m')/p_G(n, 2)$) as a function of their dimension n. For example, adding up to quadratic terms (i.e. $m' = 3$), the relative probability of feasibility increases by a factor of 1.5, 2 and 2.6 for 2, 3 and 4 species, respectively. Note that increasing the interaction order substantially increases the number of model parameters in a high-dimensional system, which turns into high amplifications in probability. Nevertheless, figure 2*c* shows that this relative increase in the probability of feasibility reduces as more parameters are added (i.e. $p_G(n, m')/p_G(n, m' - 1)$). That is, adding extra parameters to a multispecies model, that is already defined by a large number of parameters, increases the relative probability of feasibility less than in a multispecies model with fewer number of parameters. This implies that the largest relative increase in the probability of feasibility will happen when adding HOIs to the linear LV model with a large number of species.

3.2. Nonlinear functional responses and parameter restrictions

The cases above did not consider any sort of parameter restrictions, hence we now shift our focus to study how the probability of feasibility with complex multispecies models changes as a function of sign restrictions. Typically, these restrictions are imposed into models to specify particular structures and dynamics, such as antagonistic, competitive and mutualistic [5]. In particular, these dynamics are expressed and modified through a variety of nonlinear functional responses [6,24]. Hence, to explicitly incorporate functional responses into our general multidimensional model (equation (3.1)), we use the form

$$\frac{\mathrm{d}N_i}{\mathrm{d}t} = N_i\Big(r_i + a_{ii}N_i - \sum_{\substack{1 \le j \le n \\ j \neq i}} a_{ij}N_j\phi_{ij}\Big), \quad i = 1, \ldots, n, \tag{3.4}$$

where

$$\phi_{ij} = \begin{cases} 1, & m'' = 1 \\ \dfrac{N_{p(i,j)}^{m''-2}}{1 + \sum_{k \in I'_{q(i,j)}} h_{k,q(i,j)} N_k^{m''-1}}, & m'' = 2, 3, \ldots \end{cases} \tag{3.5}$$

Functional responses (i.e. ϕ's), which are quotients of two polynomials determined by the abundances of the prey, are dependent on the parameters: m'', p's, q's, I''s and h's. The parameter m'' in equation (3.5) indicates the type of the functional response (i.e. Type I, Type II, and Type III functional responses are represented by $m'' = 1$, $m'' = 2$, and $m'' = 3$, respectively). For any pair of species (i,j), the functions $p(i, j)$ and $q(i, j)$ represent the indices of the prey (or resource) and predator (or consumer), respectively (i.e. if i is the prey of j then $p(i, j) = i$ and $q(i, j) = j$). For each species k, we define I_k to be the set of indices that represent all their prey. By defining I_k for each species, all connections between species are known. To allow for different responses within the same species, for every predator $q(i, j)$, we define $I'_{q(i,j)}$ to be a subset of $I_{q(i,j)}$ which contains the index of the prey $p(i, j)$ (i.e. $p(i, j) \in I'_{q(i,j)} \subseteq I_{q(i,j)}$). The h's in equation (3.5) are constants and represent prey handling time in Type II functional responses. Note that the r's and the a's continue to represent species' growth rates and interaction coefficients, respectively.

It has been common to use forms of functional responses where the ϕ's are functions of the abundance of a single prey for which $p(i, j)$ is the only element in $I'_{q(i,j)}$ [6]. Nevertheless, the ϕ's can also be functions of the abundances of all prey, making $I'_{q(i,j)} = I_{q(i,j)}$. Here, we consider four commonly used cases: Type I, Type II with $I'_{q(i,j)} = I_{q(i,j)}$ (i.e. Beddington–DeAngelis functional response), Type II with $I'_{q(i,j)} = \{p(i, j)\}$ and Type III with $I'_{q(i,j)} = \{p(i, j)\}$. We denote these responses by T_1, T_{2m}, T_{2s} and T_{3s}, respectively. Note that in equation (3.5) $\phi_{ij} = \phi_{ji}$; however, this symmetry can be broken to allow for more generalized types of functional responses by replacing the double subscript constant $h_{k,q(i,j)}$ with a triple subscript constant $h_{k,i,j}$ in equation (3.5).

In Model (3.4), the a's are not necessarily restricted to any particular value or sign. However, in predator–prey models, a_{ij} and a_{ji} have opposite signs for every $i \neq j$. Moreover, the ratio $|a_{q(i,j),p(i,j)}|/|a_{p(i,j),q(i,j)}|$, which is denoted by ϵ_{ij}, is usually a constant between 0 and 1, and reflects the fraction of prey that is converted into a predator's abundance. This implies that the probability of feasibility with functional responses (or higher-order terms in general) can be different when adding or not parameter restrictions (e.g. sign restrictions defining who eats whom). Thus, to study the effect of sign restrictions in the coefficients of a's, we rewrite Model (3.4) as

$$\left.\begin{aligned} \frac{\mathrm{d}N_i}{\mathrm{d}t} &= N_i\Big(r_i + a_{ii}N_i - \sum_{j\in S_i\setminus I_i} a_{ij}N_j\phi_{ij} + \sum_{j\in I_i} \epsilon_{ji}a_{ji}N_j\phi_{ij}\Big),\\ di &= 1, \ldots, n, \end{aligned}\right\} \quad (3.6)$$

where $S_i = \{1, 2, \ldots, n\}\backslash\{i\}$ (backslash symbol means set difference) and all a_{ij}'s are non-negative except when $i = j$ (unrestricted in sign).

Additionally, it is worth noticing that the feasibility in Model (3.4) is dependent on the common numerator of its r.h.s., and the solution becomes similar to the previous case of an LV model with HOIs, where parameters are linked. That is, the higher the diversity and order of functional responses added into a model, the higher the order of terms added to the numerator of the r.h.s. of equation (3.4). To show this, let us write all ϕ_{ij}'s as quotients of two polynomials $\phi_{ij} = \phi^U_{ij}/\phi^D_{ij}$. Thus, equation (3.4) has a common denominator given by $\Phi_i = \Pi_{k\in S_i}\phi^D_{ik}$, whose number of terms and leading order depend on the specified functional responses. Then, let us define $\Phi_{ij} = \Phi_i\phi^U_{ij}/\phi^D_{ij}$, where Φ_{ij} is the same as Φ_i but with the term ϕ^D_{ij} replaced by ϕ^U_{ij}—which also depends on the specified functional responses. This process implies that the common numerator of the r.h.s. of equation (3.4) (after deleting N_i outside the bracket) is a multivariate polynomial expressed in terms of species abundances given by the following expression:

$$r_i\Phi_i + a_{ii}N_i\Phi_i - \sum_{\substack{1\le j\le n\\ j\neq i}} a_{ij}N_j\Phi_{ij}, \quad i = 1, 2, \ldots, n. \quad (3.7)$$

Therefore, the roots of equation (3.7) determine the free-equilibrium points of equation (3.4), allowing the common numerator of the r.h.s. of equation (3.6) to be written in a form similar to equation (3.7). Note that when moving from Type T_1 to T_{2s} and then to T_{3s}, the order of added terms increases. Moreover, when T_{2m} functional responses are used, which are functions of all prey abundances of a specific predator, there will be fewer distinct denominators in each line of equation (3.4). This is because all prey of a specific predator will have the identical denominator of equation (3.5), and, in return, it will have fewer higher-order terms added to the numerator in equation (3.7) than what would be added by T_{2s}.

To numerically compute the probability of feasibility for different types of functional responses (i.e. T_1, T_{2m}, T_{2s} and T_{3s}), we consider the following models and parameter distributions. Unrestricted Model (M_U): Model (3.4) is used to represent species dynamics (the parameters are a's, h's and r's) and the distribution of parameters is given by (i) all a's are uniform in $[-\sqrt{3}, \sqrt{3}]$, (ii) all r's are uniform in $[-\sqrt{3}, \sqrt{3}]$ except for r_1 (uniform in $[0, 2\sqrt{3}]$) and r_n (uniform in $[-2\sqrt{3}, 0]$, and (iii) all h's are uniform in $[0, 2\sqrt{3}]$ (notice that all these parameters have a unit variance). Restricted Model (M_R): Model (3.6) is used to represent species dynamics (the parameters are a's, h's, r's and the ϵ's) and the distribution of parameters is given by (i) all a_{ij}'s are uniform in $[0, 2\sqrt{3}]$ except when $i = j$ (a_{ii}'s are uniform between $[-\sqrt{3}, \sqrt{3}]$), (ii) all r's are uniform in $[-\sqrt{3}, \sqrt{3}]$ except for r_1 (uniform in $[0, 2\sqrt{3}]$) and r_n (uniform in $[-2\sqrt{3}, 0]$), (iii) all h's are uniform in $[0, 2\sqrt{3}]$, and (iv) all ϵ's are uniform in [0, 1]. In addition, we assume an interaction network defined by $p(i, j) = \min(i, j)$ for every $1 \le i, j \le n$, where $i \neq j$ (i.e. all species are connected to each other). This assumption requires that $I_k = \{1, 2, \ldots, k-1\}$ for all k where I_1 is an empty set (i.e. every species is a predator of all lower indexed species and is a prey to all higher indexed species). Next, we compute the probability of feasibility $p(n$, type, model) for n-species systems ($n = 2, 3, 4$). Note that $p(2, T_{2m}, M_i) = p(2, T_{2s}, M_i)$ for $i = \{U, R\}$ as the solo predator in the network has a single prey. These parameter values are chosen to simply illustrate the effect of sign restrictions and are not intended to reflect any specific ecological process.

In general, as the complexity of a model increases (either with dimension or with adding extra processes), its number of roots (free-equilibrium points) also increases, which leads to an increase in the computational time needed to solve multivariate polynomials for a single trial. Thus, any exponential increase in the number of free-equilibrium points makes computing the probability of feasibility a hard task. Thus, to reasonably compute these probabilities, we use 25 000 trials for each combination $p(n$, type, $M_i)$. The results are presented in table 1. Note that the probability of feasibility agrees with the theoretical value $p(n, T_1, M_U) = 1/2^n$ up to 2 digits, which is a good indicator for comparison purposes.

Table 1 shows that the number of free-equilibrium points (Θ) of both models (M_U and M_R) increases as a function of the polynomial degree, i.e. functional responses T_1, T_{2m}, T_{2s}

Table 1. Probability of feasibility and number of free-equilibrium points as a function of system's size and polynomial degree (functional response). For $n = 2$, 3, 4 species systems (columns) and four different types of functional responses T_1, T_{2m}, T_{2s}, T_{3s} (rows), the table shows the probability of feasibility for sign-unrestricted models ($p_U \equiv p(n,$ type, $M_U)$) and sign-restricted models ($p_R \equiv p(n,$ type, $M_R)$). For each of these combinations, the table also shows the average (the median is almost identical) number of free-equilibrium points (Θ) computed using the solver PHClab (with default settings). Note that Θ can numerically fluctuate either because parameters yield fewer roots or because the solver eliminates leading terms with small coefficients (which do not affect the existence of a feasible root). Probabilities decrease with system size. Both probabilities and free-equilibrium points tend to increase with polynomial degree. For each system size, p_U increases with the value of Θ. Similarly, for $\Theta > 1$, p_R increases with the value of Θ.

	$n = 2$			$n = 3$			$n = 4$		
type/size	Θ	p_U	p_R	Θ	p_U	p_R	Θ	p_U	p_R
T_1	1	0.2543	0.5378	1	0.1256	0.2299	1	0.0662	0.0923
T_{2m}	3	0.2661	0.3248	10	0.1391	0.1817	23	0.0760	0.0957
T_{2s}	same as T_{2m}			12	0.1457	0.2031	62	0.0920	0.1226
T_{3s}	5	0.2842	0.3486	33	0.1807	0.2802	289	0.1347	0.2422

and T_{3s} in that order. Focusing on the unrestricted model (M_U) and controlling for the number of species n (for any n, columns in table), table 1 shows that the increase in the probability of feasibility is consistent with the increase in Θ (i.e. complexity). That is, $p(n, T_1, M_U) < p(n, T_{2m}, M_U) < p(n, T_{2s}, M_U) < p(n, T_{3s}, M_U)$. These inequalities are also present in the restricted model (M_R), except for $n = 2$ and $n = 3$, where $p(n, T_1, M_R) > p(n, T_{2m}, M_R)$. However, at $n = 4$ (under a higher Θ when T_{2m} is used) the inequality is recovered again. It is worth mentioning that unlike the case of the LV model with HOIs, where both the number of free-equilibrium points and the number of parameters increase as a function of the polynomial degree, in the case of functional responses T_{2m}, T_{2s} and T_{3s}, the number of parameters is constant. Nevertheless, the increase in the probability of feasibility, while not as high as in the LV model with HOIs, is still observed despite fixing the number of parameters and whether the a's are restricted in sign or not.

The analysis above allows us to make a distinction between complex models and the linear LV model. Table 1 reveals that the probability of feasibility is a monotonic and saturating function of complexity when $\Theta > 1$ (e.g. moving from type T_{2m} to T_{2s} to T_{3s}). However, when we compare cases with $\Theta = 1$ against cases with $\Theta > 1$, the probability of feasibility of a linear LV model (i.e. $\Theta = 1$) will be exceeded only as soon as a minimum level of complexity (Θ^*) is reached. For example, under the restricted model M_R, the probability of feasibility will exceed that of the linear LV model when $\Theta \geq 23$ for $n = 2$, 3, 4. This level of complexity (Θ) differs for each distribution. By contrast, under the unrestricted model M_U, this level decreases to $\Theta \geq 3$ for all n's regardless of functional response.

The previous results can be explained by noticing the fundamental difference between complex models and the linear LV model. Under arbitrary model parameters, the solo free-equilibrium point in LV model must be real (i.e. all its coordinates are real). Instead, in complex models free-equilibrium points are generally complex [22]. Thus, the initial entrance to the complex domain represents a handicap for complex models, yet this is quickly recovered by the increase of free-equilibrium points. These concepts can be verified analytically: defining the probability of feasibility $p(\Theta)$ by the form $1 - (1 - \hat{E}(\Theta)/\Theta)^{\Theta}$, let us assume that $p(\Theta) < p(1)$, leading to $\hat{E}(\Theta) \leq \Theta(1 - (1 - \hat{E}(1))^{1/\Theta}$. If Θ increases, the r.h.s. of the inequality will approach $-\ln(1 - \hat{E}(1))$ independently of Θ. However, we know already that $\hat{E}(\Theta)$ increases with Θ. This implies the existence of a minimum Θ^* for which $\hat{E}(\Theta^*) > -\ln(1 - \hat{E}(1))$ and subsequently $p(\Theta^*) > p(1)$. Because of the expected drastic increase in Θ as a function of the dimension of the system [22], it can be proved that no matter the parameter restrictions imposed in a model, Θ^* can always be exceeded by increasing either dimensionality or polynomial degree. This increase will yield a higher probability of feasibility than in the linear LV model regardless of the specific mechanisms invoked.

Indeed, table 1 shows that the relative increase in the probability of feasibility compared to the linear LV model (i.e. $p(n$, type, $M_i)/p(n, T_1, M_i)$, for $i = \{U, R\}$) is a function of n for any given functional response. For instance, when T_{3s} is used as functional response, the probability for models M_U and M_R increases by a factor of 1.1, 1.4, 2 and 0.6, 1.2, 2.6 for $n = 2$, 3, 4 species, respectively. This increasing pattern is also consistent using T_{2m} and T_{2s}. Therefore, adding nonlinear processes to a linear LV model (T_1) increases the number of free-equilibrium points; which, in turn, contributes to the increase in the probability of feasibility—this is further magnified as the number of species increases. Additionally, note that the number of parameters is fixed for the functional responses T_{2m}, T_{2s} and T_{3s}, which differ only by a few terms (the h's) compared to T_1. Therefore, unlike the case of HOIs, the relative increase in the probability of feasibility is not necessarily larger when moving from Type I to Type II (either T_{2m} or T_{2s}) than the increase observed when moving from Type II to Type III. For example, focusing on the model M_R with $n = 4$ species and moving from T_1 to T_{2s} and then from T_{2s} to T_{3s}, the relative increase is 1.3 and 2, respectively.

It is also worth noticing that while functional responses T_{2m}, T_{2s} and T_{3s} are widely used in the literature, the functional response T_{2m} can be considered a more realistic type [6]. Interestingly, table 1 shows that models with T_{2m} have fewer free-equilibrium points (less complexity) than the other types, and their probability of feasibility is the closest to the linear LV model. This can suggest that realistic models should deviate the least from the probability of feasibility of the linear LV model. Nevertheless, the difference in its probability of feasibility compared to the linear LV model will increase with dimensionality. This is evident by the number of free-equilibrium points, which exceeds that obtained from the LV model with HOIs at interaction order $m' = 3$ (i.e. adding up to quadratic terms to the linear LV model results in 2^n free equilibrium points) at $n \geq 3$—implying at least an exponential increase in Θ with dimensionality. Focusing on T_{2m}, if the symmetry is broken in equation (3.5) by replacing a few $h_{k,q(i,j)}$ with $h_{k,i,j}$ (which can differ only slightly from $h_{k,q(i,j)}$), the number of free-equilibrium points can go beyond that of T_{3s}, increasing the probability of feasibility significantly. Furthermore, the probability of feasibility with functional responses (under parameter restrictions) is smaller than in the linear LV model for $n = 2$, 3 species except when $n = 3$ and the functional response T_{3s} is used. However, this pattern already disappears with four species, confirming that one cannot directly extrapolate our understanding of ecological dynamics from low to high dimensions [39,49,50].

4. Discussion

One of the main goals in ecological research is to understand the main factors that contribute to the persistence of multispecies systems [5,51]. While simple ecological models (such as the linear LV model) are typically modified for the purpose of adding realism and dynamical richness, tractability is usually compromised [6,12]. For example, it is well known that in the linear LV model (Type I), the number of feasible equilibrium solutions (a crucial condition for the persistence of ecological systems) is always one regardless of the dimension of the system, making this a limited but tractable model [35,52]. By contrast, the addition of higher-order terms (specifically, polynomial fractions such as nonlinear functional responses and higher-order interactions) invariantly increases the number of free-equilibrium solutions, making these rich but untractable models [22]. This reveals that without knowing the exact parameter values in a model, it is necessary to study from a probabilistic point of view the contribution of ecological processes (both mechanistic and phenomenological) to explaining the dynamics of multispecies systems.

Focusing on the feasibility of ecological systems (defined here as the probability of exhibiting at least one positive real root under an arbitrary choice of parameter values) in complex models (defined here as modifications to the LV model using multivariate polynomial fractions and with $\Theta > 1$), we have shown that the probability of feasibility is a monotonic and saturating function of its complexity, regardless of the specific mechanism invoked. We have characterized this complexity by the number of free-equilibrium points (Θ) generated by a model, which is a function of the model's polynomial degree and dimension. We have found that the probability of feasibility in a complex model ($\Theta > 1$) will exceed the one in a linear LV model ($\Theta = 1$) as soon as a minimum level of complexity (Θ^*) is reached. Importantly, this minimum level is modulated by parameter restrictions, but can always be exceeded via increasing the polynomial degree or system's dimension.

It is worth recalling that the number of free-equilibrium points in a model and its number of parameters are two different descriptors [22]. For example, the LV model with Type II functional responses has the same number of parameters as that of Type III, yet the number of free-equilibrium points is different in both models. This difference is important as we have shown that it is expected that the model with more free-equilibrium points will have a higher probability of feasibility. These findings could be perceived as a desirable advantage for complex models, as they can provide a higher probability of generating a feasible multispecies system (and richer dynamics). Unfortunately, this increase in probability happens no matter what type of specific mechanism is added, it all depends on its polynomial form—limiting the capacity to distinguish the actual contribution of a specific ecological process to the feasibility of a multispecies system.

For example, studies have investigated population dynamics resulting from mutualism by employing functional responses based on density-dependent benefits and costs [53–55], i.e. the ϕ's in equation (3.4) are replaced with $\phi - \phi^{\text{cost}}$ where functional responses are modified to add a cost term. However, as we have shown, adding cost terms to penalize for some benefits will not decrease the probability of feasibility, actually they will increase it. Similarly, in the study of food-web models [28], it is common practice to use multispecies functional responses (i.e. polynomial fractions of more than a single species) in order to include the effect of other predators or prey [56]. Note that the Types I, II, III functional responses are functions of the prey density only—a single species. But, as we have shown, any of these modifications can only increase the probability of feasibility. As a third example, the simple and ecologically motivated idea of introducing carrying capacities to limit the growth of species (i.e. the total growth rate G_i of species i is replaced by $G_i(1 - N_i/K_i)$, where K_i is its carrying capacity) [29] also increases the probability of feasibility. Thus, regardless of whether a higher-order term (nonlinear mechanism) is ecologically well motivated, expected to limit or enrich dynamics, or has absolutely no meaning, it is expected to increase the probability of feasibility in a multispecies model. This suggests that the explanatory contribution to feasibility of a proposed ecological mechanism must be evaluated by its deviation from the expected behaviour of its polynomial form.

The contribution of different ecological processes has been studied by showing how additional terms can help us to fit observed data [54]. Yet, fitting data has the same effect as introducing parameter restrictions [22]. Hence, under this fitting process, it is only expected that any additional process will increase on average the probability of explaining the dynamics of the feasible system. Furthermore, under certain cases, adding more process into a model can leave the probability the same as in the original restricted case (e.g. a linear LV mutualistic model with no self-regulation: adding negative-density dependence will increase the probability, while adding positive density dependence will leave the probability invariant; yet, this involves modifying an already restricted model rather than restricting a modified model). Thus, studies using fitting methodologies should contrast their results by using out-of-sample validations [33].

Our results motivate us to reconsider what constitutes a realistic model, or how much complexity can be appropriate to add into a model to mimic realistic ecological mechanisms. Do we need models to fit perfectly data? Or do we need models to explain and predict dynamics with minimal available information? Because it is virtually impossible to know the exact form of the equations governing the dynamics of multispecies systems, as well as the exact value of initial conditions, we believe that a first step towards answering these questions implies understanding the extent to which the complexity of a model provides an advantage over other models by virtue of their specific mechanisms invoked and not simply by their polynomial form. Otherwise, any mechanism can explain equally well any ecological dynamics, introducing the problem of model or structural unidentifiability [57]. Thus, in order to advance our causative knowledge of ecological dynamics, we need to understand the expected outcomes of our proposed models and their alternative hypotheses.

Data accessibility. The code supporting the results can be found at https://github.com/MITEcology/Interface_AlAdwani_2020.

Authors' contributions. S.S. and M.A. designed the study; M.A. performed the study; S.S. supervised the study; S.S. and M.A. analysed results and wrote the manuscript.

Competing interests. The authors declare no competing financial interests.

Funding. M.A. would like to thank Kuwait Foundation for the Advancement of Sciences (KFAS) and Kuwait Chamber of Commerce and Industry (KCCI) for their support. Also, M.A. would like to thank MathWorks Inc. for their generous fellowship. S.S. acknowledges funding by NSF grant no. DEB-2024349.

Acknowledgements. We would like to thank Alan Edelman whose work on random polynomials inspired part of this paper. We would also like to thank Abdulaziz Beljadid and Washington Taylor for valuable discussions, and Chuliang Song and Lucas P. Medeiros for their constructive comments on a previous draft of the manuscript.

References

1. Petchey OL, Brose U, Rall BC. 2010 Predicting the effects of temperature on food web connectance. *Phil. Trans. R. Soc. B* **365**, 2081–2091. (doi:10.1098/rstb.2010.0011)
2. Fukami T. 2015 Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annu. Rev. Ecol. Evol. Syst.* **46**, 1–23. (doi:10.1146/annurev-ecolsys-110411-160340)
3. Levins R. 1968 *Evolution in changing environments: some theoretical explorations*. Princeton, NJ: Princeton University Press.
4. Sugihara G. 1994 Nonlinear forecasting for the classification of natural time series. *Phil. Trans. R. Soc. Lond. A* **348**, 477–495. (doi:10.1098/rsta.1994.0106)
5. Case TJ. 2000 *An illustrated guide to theoretical ecology*. Oxford, UK: Oxford University Press.
6. Turchin P. 2003 *Complex population dynamics: a theoretical/empirical synthesis*. Princeton, NJ: Princeton University Press.
7. Svirezhev YM, Logofet DO. 1983 *Stability of biological communities*. Moscow, Russia: Mir Publishers.
8. Strogatz SH. 2015 *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Boulder, CO: Westview Press.
9. Kantz H, Schreiber T. 2004 *Nonlinear time series analysis*. Cambridge, UK: Cambridge University Press.
10. Shmueli G. 2010 To explain or to predict? *Stat. Sci.* **25**, 289–310. (doi:10.1214/10-STS330)
11. Dietze MC. 2017 *Ecological forecasting*. Princeton, NJ: Princeton University Press.
12. May RM. 2004 Uses and abuses of mathematics in biology. *Science* **303**, 790–793. (doi:10.1126/science.1094442)
13. May RM. 1976 Simple mathematical models with very complicated dynamics. *Nature* **261**, 459–467. (doi:10.1038/261459a0)
14. Lotka AJ. 1920 Analytical note on certain rhythmic relations in organic systems. *Proc. Natl Acad. Sci. USA* **6**, 410–415. (doi:10.1073/pnas.6.7.410)
15. Volterra V, Brelot M. 1931 *Leçons sur la théorie mathématique de la lutte pour la vie*. Gauthier-Villars: Cahiers scientifiques.
16. Levine JM, Bascompte J, Adler PB, Allesina S. 2017 Beyond pairwise mechanisms of species coexistence in complex communities. *Nature* **546**, 56–64. (doi:10.1038/nature22898)
17. Letten AD, Stouffer DB. 2019 The mechanistic basis for higher-order interactions and non-additivity in competitive communities. *Ecol. Lett.* **22**, 423–436. (doi:10.1111/ele.13211)
18. Grilli J, Barabas G, Michalska-Smith MJ, Allesina S. 2017 Higher-order interactions stabilize dynamics in competitive network models. *Nature* **58**, 210–213. (doi:10.1038/nature23273)
19. Bairey E, Kelsic ED, Kishony R. 2016 High-order species interactions shape ecosystem diversity. *Nat. Commun.* **7**, 12285. (doi:10.1038/ncomms12285)
20. Mayfield MM, Stouffer DB. 2017 Higher-order interactions capture unexplained complexity in diverse communities. *Nat. Ecol. Evol.* **1**, 0062. (doi:10.1038/s41559-016-0062)
21. Abrams PA. 1983 Arguments in favor of higher order interactions. *Am. Nat.* **121**, 887–891. (doi:10.1086/284111)
22. AlAdwani M, Saavedra S. 2019 Is the addition of higher-order interactions in ecological models increasing the understanding of ecological dynamics? *Math. Biosci.* **315**, 108222. (doi:10.1016/j.mbs.2019.108222)
23. Vandermeer JH. 1969 The competitive structure of communities: an experimental approach with protozoa. *Ecology* **50**, 362–371. (doi:10.2307/1933884)
24. Holling CS. 1959 Some characteristics of simple types of predation and parasitism. *Can. Entomol.* **91**, 385–398. (doi:10.4039/Ent91385-7)
25. Solomon M. 1949 The natural control of animal populations. *J. Anim. Ecol.* **18**, 1–35. (doi:10.2307/1578)
26. Beddington JR. 1975 Mutual interference between parasites or predator and its effect on searching efficiency. *J. Anim. Ecol.* **44**, 331–340. (doi:10.2307/3866)
27. DeAngelis DL, Goldstein RA, O'Neill RV. 1975 A model for trophic interaction. *Ecology* **56**, 881–892. (doi:10.2307/1936298)
28. Rossberg AG. 2013 *Food webs and biodiversity*. New York, NY: Wiley.
29. Ong TWY, Vandermeer JH. 2015 Coupling unstable agents in biological control. *Nat. Commun.* **6**, 5991. (doi:10.1038/ncomms6991)
30. Thom R. 1972 *Stabilité structurelle et morphogenèse*. Paris, France: InterÉditions.
31. Alberch P. 1991 From genes to phenotype: dynamical systems and evolvability. *Genetica* **84**, 5–11. (doi:10.1007/BF00123979)
32. Song C, Rohr RP, Saavedra S. 2018 A guideline to study the feasibility domain of multi-trophic and changing ecological communities. *J. Theor. Biol.* **450**, 30–36. (doi:10.1016/j.jtbi.2018.04.030)
33. Saavedra S, Medeiros LP, AlAdwani M. 2020 Structural forecasting of species persistence under changing environments. *Ecol. Lett.* **23**, 1511–1521. (doi:10.1111/ele.13582)
34. Medeiros LP, Boege K, Zaldivar-Riverón A, Saavedra S. In press. Observed ecological communities are formed by species combinations that are among the most likely to persist under changing environments. *Am. Nat.* (doi:10.1086/711663)
35. Hofbauer J, Sigmund K. 1998 *Evolutionary games and population dynamics*. Cambridge, UK: Cambridge University Press.
36. Song C, Saavedra S. 2017 Will a small randomly assembled community be feasible and stable? *Ecology* **99**, 743–751. (doi:10.1002/ecy.2125)
37. Rohr RP, Saavedra S, Bascompte J. 2014 On the structural stability of mutualistic systems. *Science* **345**, 1253497. (doi:10.1126/science.1253497)
38. Uricchio LH, Daws SC, Spear ER, Mordecai EA. 2019 Priority effects and nonhierarchical competition shape species composition in a complex grassland community. *Am. Nat.* **193**, 213–226. (doi:10.1086/701434)
39. Saavedra S, Rohr RP, Bascompte J, Godoy O, Kraft NJB, Levine JM. 2017 A structural approach for understanding multispecies coexistence. *Ecol. Monogr.* **87**, 470–486. (doi:10.1002/ecm.1263)
40. Edelman A, Kostlan E. 1995 How many zeros of a random polynomial are real? *Bull. Am. Math. Soc.* **32**, 1–37. (doi:10.1090/S0273-0979-1995-00571-9)
41. Sommese AJ, Wampler CW. 2005 *The numerical solution of systems of polynomials arising in engineering and science*. Singapore: World Scientific.
42. Guan Y, Verschelde J. 2008 PHClab: a MATLAB/Octave interface to PHCpack. In *Software for algebraic geometry* (eds M Stillman,

J Verschelde, N Takayama), pp. 15–32. New York, NY: Springer. (doi:10.1007/978-0-387-78133-4_2)

43. May RM. 1972 Will a large complex system be stable? *Nature* **238**, 413–414. (doi:10.1038/238413a0)
44. Schreiber SJ, Yamamichi M, Strauss SY. 2019 When rarity has costs: coexistence under positive frequency-dependence and environmental stochasticity. *Ecology* **100**, e02664. (doi:10.1101/161919)
45. Cenci S, Saavedra S. 2018 Structural stability of nonlinear population dynamics. *Phys. Rev. E* **97**, 012401. (doi:10.1103/PhysRevE.97.012401)
46. Serván CA, Capitán JA, Grilli J, Morrison KE, Allesina S. 2018 Coexistence of many species in random ecosystems. *Nat. Ecol. Evol.* **2**, 1237–1242. (doi:10.1038/s41559-018-0603-6)
47. Goh BS, Jennings LS. 1977 Feasibility and stability in randomly assembled Lotka-Volterra models. *Ecol. Modell*. **3**, 63–71. (doi:10.1016/0304-380 0(77)90024-2)
48. Grilli J, Adorisio M, Suweis S, Barabás G, Banavar JR, Allesina S, Maritan A. 2017 Feasibility and coexistence of large ecological communities. *Nat. Commun.* **8**, 14389. (doi:10.1038/ncomms14389)
49. Barabás G, D'Andrea R, Stump SM. 2018 Chesson's coexistence theory. *Ecol. Monogr.* **88**, 277–303. (doi:10.1002/ecm.1302)
50. Song C, Barabás G, Saavedra S. 2019 On the consequences of the interdependence of stabilizing and equalizing mechanisms. *Am. Nat.* **194**, 627–639. (doi:10.1086/705347)
51. Vellend M. 2016 *The theory of ecological communities (MPB-57)*. Princeton, NJ: Princeton University Press.
52. Takeuchi Y. 1996 *Global dynamical properties of Lotka–Volterra systems*. Singapore: World Scientific.
53. Holland JN, DeAngelis DL, Bronstein JL. 2002 Population dynamics and mutualism: functional responses of benefits and costs. *Am. Nat.* **159**, 231–244. (doi:10.1086/338510)
54. Valdovinos FS. 2019 Mutualistic networks: moving closer to a predictive theory. *Ecol. Lett.* **22**, 1517–1534. (doi:10.1111/ele.13279)
55. Peralta G, Stouffer DB, Bringa EM, Vázquez DP. 2020 No such thing as a free lunch: interaction costs and the structure and stability of mutualistic networks. *Oikos* **129**, 503–511. (doi:10.1111/oik.06503)
56. 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)
57. Angulo MT, Moreno JA, Lippner G, Barabási AL, Liu YY. 2017 Fundamental limitations of network reconstruction from temporal data. *J. R. Soc. Interface* **14**, 20160966. (doi:10.1098/rsif.2016.0966)