

Self-organization of ecosystems to exclude half of all potential invaders

Cillian Cockrell¹, Jacob D. O'Sullivan², J. Christopher D. Terry^{2,3}, Emmanuel C. Nwankwo², Kostya Trachenko⁴, and Axel G. Rossberg²

¹Department of Materials, Imperial College London, Exhibition Road, London, SW7 2AZ, UK

²School of Biological and Behavioural Sciences, Queen Mary University of London, Mile End Road, London, E1 4NS, UK

³Department of Biology, University of Oxford, 11a Mansfield Road, Oxford OX1 3SZ, UK

⁴School of Physical and Chemical Sciences, Queen Mary University of London, Mile End Road, London, E1 4NS, UK



(Received 6 October 2022; accepted 6 October 2023; published 25 January 2024)

Species-rich Lotka-Volterra competition models of ecosystem dynamics transition with increasing species pool size from a phase with well-defined stable equilibrium to a dynamic phase that remains incompletely understood. We analytically describe the statistical mechanics of the steady state deep inside this dynamic phase, characterized by incessant turnover in species composition, and extract the distribution of invasion fitness of random invaders. We find that steady state invasion probability universally equals 1/2. This striking result agrees well with observations in plants and animals.

DOI: [10.1103/PhysRevResearch.6.013093](https://doi.org/10.1103/PhysRevResearch.6.013093)

I. INTRODUCTION

Biodiverse ecosystems on large spatial scales exhibit such a wide range of reproducible, emergent high-level properties [1] that we might consider them a distinct state of matter, accessible to methods from statistical mechanics. Entropy maximization [2] and field-theoretical methods [3] have been invoked to describe such systems and explain, e.g., widely observed noninteger power laws relating study area to species counts (species richness) [3,4].

Such spatiotemporal structures formed by interacting species populations are reproduced in simulations of spatially coupled systems of Lotka-Volterra competition models [5,6] of the form

$$\frac{db_i}{dt} = \left[s_i - \sum_{j=1}^{S_{\text{pool}}} A_{ij} b_j \right] b_i + \iota_i \quad (1 \leq i \leq S_{\text{pool}}). \quad (1)$$

Here t is time, $b_i \geq 0$ are the time-dependent population biomasses of S_{pool} species i competing in the region, s_i are population growth rates in absence of competition, and the elements A_{ij} of matrix \mathbf{A} quantify the strengths of intraspecific ($i = j$) and interspecific ($i \neq j$) competition. The term in brackets in Eq. (1) approximates the momentary population growth rates of interacting species as a linear function of population biomasses. Rescaling biomass variables, one can set $A_{ii} = 1$ ($1 \leq i \leq S$). The ι_i represent spatial coupling (immigration) to neighboring locations. With $\iota_i = 0$, we can distinguish the S species extant ($b_i > 0$) in an equilibrium and all others ($b_i = 0$). Equation (1) then has several equilibria,

called ecological communities, distinguished by different (index) sets of extant species \mathcal{E} , called community compositions.

We consider a weaker interspecific ($i \neq j$) than intraspecific competition ($0 \leq A_{ij} < 1$) and absent or very weak spatial coupling ($\iota_i = 0$). Under weak coupling, extant and extinct species can be distinguished using heuristic criteria [5]. Substantial progress was recently made in understanding Eq. (1) when interspecific A_{ij} are assigned random values (fixed through time) and species richness $S = |\mathcal{E}|$ is large. We focus on the empirically well-supported case [7] where the A_{ij} are independently and identically distributed (i.i.d.) with mean μ and variance σ^2 (though correlated A_{ij} and A_{ji} are also often considered [8–10]). Higher moments play only minor roles in the limit $S \rightarrow \infty$ [11].

Random matrix theory predicts that for high species richness Eq. (1) becomes an ill-defined problem, a phenomenon called ecological structural instability ([12], Eq. (18.1)): when

$$S > S_{\text{ESI}} \stackrel{\text{def}}{=} \frac{(1 - \mu)^2}{\sigma^2}, \quad (2)$$

the area in the complex plane covered by the eigenvalues of the matrix $\{A_{ij}\}_{i,j \in \mathcal{E}}$ includes zero [13]. Population abundances are then highly sensitive to perturbations such as changes in the s_i or addition or removal of species. This implies more species than S_{ESI} are unlikely to coexist [5]. For $S_{\text{pool}} < 2S_{\text{ESI}}$, the “unique fixed point” phase, populations reach an equilibrium where $S < S_{\text{ESI}}$ species coexist, with the rest extinct [8]. At $S_{\text{pool}} = 2S_{\text{ESI}}$, where $S = S_{\text{ESI}}$, the system exhibits a phase transition [8]. For $S_{\text{pool}} > 2S_{\text{ESI}}$, in the so called “multiple attractors” (MA) phase, system dynamics is only partially understood. There is, with all $\iota_i = 0$, generally no stable equilibrium in this phase, because for each equilibrium there is at least one locally extinct species that could invade (i.e., its linear population growth rate is positive). These invasions are represented by heteroclinic orbits in phase space that each lead from one equilibrium to another [14]. The

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

resulting heteroclinic orbit network contains (as a directed graph) at least one strongly connected component [15] that is a stable attractor of system dynamics. On this attractor, the community transitions between different community compositions at an approximately constant rate on a logarithmic time scale [6]. For fixed small but positive ι_i , this network is perturbed, leading to slow switching between different community compositions on a linear time scale.

In the explicitly spatially extended case, so-called meta-community models [1], the ι_i are dynamically determined by neighboring communities, quantifying the coupling between the local Lotka-Volterra equations (1). For these models a similar phase transition is observed in simulations as the number of interacting communities increases [6,16].

In this paper, we make advances characterizing the statistical mechanics of the MA phase for $S_{\text{pool}} \gg S_{\text{ESI}} \gg 1$. This is achieved by approximating Eq. (1) by a community assembly model [17]. In this model we construct an ecological community by iterated invasions of random species. Extant species go extinct as a result of competition and a steady state is reached where invasions and extinctions are balanced. We assume, supported by simulations over a wide parameter range, that extinction occurs only to low-biomass species and only after the invasion of a new species. In other words, the changes in community composition and biomasses as a result of a single invasion are small—there are no mass extinctions or sudden changes. The community assembly process, particularly the probability of successful invasion p_{inv} , have been the subject of many studies [17–21]. It is a quantity of practical relevance in conservation ecology, where alien species invasions are considered a problem, and an important characteristic of community structure and dynamics. Our objective is to calculate p_{inv} in a steady-state community experiencing continuous species turnover through invasions and extinctions.

II. METHODS AND RESULTS

The state of our model, approximating Eq. (1), is characterized by the community composition \mathcal{E} . It changes by invasion of species $k \notin \mathcal{E}$, sampled at random conditional to k being able to invade (explained below). Since we are considering the case $S_{\text{pool}} \gg S_{\text{ESI}} = \mathcal{O}(S)$, k 's interactions A_{jk} , A_{kj} ($j \in \mathcal{E}$) are effectively i.i.d., sampled from a distribution with support $0 \leq A_{jk} < 1$ as above. The immigration terms ι_i in Eq. (1) are disregarded in this approximation. Denoting $\mathcal{E}^+ = \{k\} \cup \mathcal{E}$, population dynamics of the assembly model between invasions is given by

$$\frac{db_i}{dt} = \left[s_i - \sum_{j \in \mathcal{E}^+} A_{ij} b_j \right] b_i \quad (i \in \mathcal{E}^+), \quad (3)$$

with initial conditions such that the species in \mathcal{E} are in equilibrium and k has small, positive abundance ($0 < b_k \ll 1/|\mathcal{E}|$). From there, the system typically reaches a new unique stable equilibrium [found, e.g., by simulating Eq. (3)]. After this, the set \mathcal{E} of extant species is updated accordingly and the process is repeated, leading to a stochastic steady state where invasion and extinctions balance each other on average.

Our calculation makes use of a close relation between two quantities. First, the harvesting resistance of extant species i ,

defined as the inverse susceptibility $r_i = (d \ln(b_i)/ds_i)^{-1}$ [12], Sec. 14.5. It can be computed as

$$r_i = s_i - \sum_{j \in \mathcal{E} \setminus \{i\}} A_{ij} b_j, \quad (4)$$

with the b_j chosen to satisfy $s_i - \sum_{j \in \mathcal{E} \setminus \{i\}} A_{ij} b_j = 0$ for all $i \in \mathcal{E} \setminus \{i\}$ (see details in the Appendix) [12], Sec. 14.5. Second, invasion fitness, the linear growth rate of new invaders $k \notin \mathcal{E}$, which, by comparison of Eqs. (3) and (4), is given by Eq. (4) with $i = k$. As a general rule, $r_i > 0$ for all extant species, and k can invade if and only if $r_k > 0$. Invasion probability p_{inv} is the probability that $r_k > 0$ for a randomly sampled invader k .

For simplicity, we assume the same value s for all s_i . In this case, r_i and b_i are approximately proportional to each other, such that ([12], Eq. (17.23))

$$\frac{r_i}{b_i} \approx \text{const.} \approx (1 - \mu). \quad (5)$$

For new, random invaders $i = k$, the A_{kj} in Eq. (4) are random. Assuming a sparse matrix A_{ij} such that the extant biomasses are weakly correlated, the invasion fitness r_k , Eq. (4), is therefore for large S normally distributed. Since community turnover leads to slow random changes in invasion fitness, we therefore assume (and verify in the Appendix), that, for a randomly chosen potential invader k that never actually invades, r_k follows an approximate Ornstein-Uhlenbeck (OU) process ([12], Sec. 14.6) with mean \bar{r} and variance varr . We introduce a new, slow time scale T that measures the number of successful invasions in the assembly model and denote the OU relaxation rate on this time scale by ρ .

Now consider the alternative case where i is one of many ($S \gg 1$) extant species. Because r_i is given by the same formula as r_k , Eq. (4), we assume that between successive invasions the dynamics of r_i follows the same kind of random walk as r_k . At the moment when i invades, r_i follows a normal distribution with mean \bar{r} and variance varr , conditional to $r_i > 0$. Its extinction occurs when r_i reaches zero.

The relaxation rate ρ is therefore determined in two different ways: (i) as the relaxation rate ρ of the underlying OU process and (ii) as the relaxation rate of the sum in Eq. (4), of which each term performs, by Eq. (5), a random walk with the same relaxation rate parameter, while invasions and extinctions lead to removal and addition of terms. This gives a self-consistency condition that we can evaluate to determine p_{inv} .

To obtain an explicit expression for the relaxation rate given by (ii), we first make an additional simplifying assumption, later relaxed: with two parameters $0 < I, C < 1$, the i.i.d. distribution of A_{ij} for $i \neq j$ is such that $A_{ij} = I$ with probability C and $A_{ij} = 0$ otherwise. We denote by \mathcal{A}_i the subset of indices j in \mathcal{E} for which $A_{ij} = I$. We require that C is sufficiently small such that $|\mathcal{A}_i|$ follows an approximate Poisson distribution.

We standardize the variance of the OU process driving harvesting resistance by introducing scaled variables $y_i = r_i/(\text{varr})^{1/2}$ and defining $\bar{y} = \bar{r}/(\text{varr})^{1/2}$. Hence, $p_{\text{inv}} = \Phi(\bar{y})$, with Φ denoting the cumulative normal distribution. By Eq. (5), the scaled harvesting resistances y_j differ from the population biomasses b_j only by a constant factor. The

relaxation rate of the sum in Eq. (4) therefore equals the relaxation rate of

$$R_i = \sum_{j \in \mathcal{A}_i} y_j, \quad (6)$$

where the OU process followed by each y_j from the time it invades until it becomes extinct can be expressed through a Langevin equation

$$\frac{dy_j(T)}{dT} = -\rho(y_j(T) - \bar{y}) + \sqrt{2\rho}\xi_j(T), \quad (7)$$

with $\xi_j(T)$ denoting uncorrelated unit white noise [$\text{cor}(\xi_i(T), \xi_j(T')) = \delta_{i,j}\delta(T - T')$]. The coefficient of the noise term follows because the unconstrained OU process has unit variance by construction. The noise is uncorrelated due to our assumption of a sparse matrix A_{ij} and weak interspecific interactions. In addition, R_i changes through invasions and extinctions of species j . When a species j invades, its scaled harvesting resistance y_j follows a normal distribution conditional to $y_j > 0$, with density

$$p^{\text{inv}}(y) = \frac{1}{\sqrt{2\pi}\Phi(\bar{y})} \exp\left(-\frac{(y - \bar{y})^2}{2}\right). \quad (8)$$

Invasions of species j with $A_{ij} > 0$ occur at a rate C , adding the index j to \mathcal{A}_i and a term y_j to the sum defining R_i . Extinction of a species $j \in \mathcal{A}_i$ occurs when y_j reaches zero. The index j is then removed from \mathcal{A}_i and the process y_j from R_i .

We now evaluate the relaxation rate of R_i using moment equations, denoting expectation values by brackets $\langle \cdot \rangle$. With our assumption of negligible correlations, suppressing the index i and abbreviating $Z = |\mathcal{A}|$, we can write $\langle R \rangle = \langle Z \rangle \langle y \rangle$, where the second factor is the expectation of y_i for a randomly chosen extant species. According to above considerations, $\langle R \rangle$ changes over a small time interval $\delta T \geq 0$ to

$$\begin{aligned} \langle R(T + \delta T) \rangle &= \left\langle \sum_{j \in \mathcal{A}} y_j - \delta T \rho(y_j - \bar{y}) + \delta T C y^{\text{inv}} \right\rangle \\ &= \langle Z \rangle \langle y \rangle + [-\rho \langle Z \rangle (\langle y \rangle - \bar{y}) + C y^{\text{inv}}] \delta T, \end{aligned} \quad (9)$$

where y^{inv} denotes the mean associated with $P^{\text{inv}}(y)$. The sum above is over the extant species at time T . Extinctions of species can be disregarded at lowest order, because shortly before extinction their contribution to R is small. In the steady state the time-dependent term in Eq. (9) vanishes:

$$-\rho \langle y \rangle \langle Z \rangle + C y^{\text{inv}} = 0. \quad (10)$$

Making use of Eq. (10), we similarly derive in the Appendix the second moment:

$$\begin{aligned} \langle R(T)R(T + \delta T) \rangle &= \langle Z \rangle^2 \langle y \rangle^2 + \langle Z \rangle \langle y^2 \rangle + \langle Z \rangle \rho (\langle y \rangle \bar{y} - \langle y^2 \rangle) \delta T. \end{aligned} \quad (11)$$

Combining the moment equations, we first evaluate

$$\text{var} R = \langle R^2 \rangle - \langle R \rangle^2 = \langle Z \rangle \langle y^2 \rangle. \quad (12)$$

Then we calculate the short-term autocorrelation function

$$\text{cor}[R(T), R(T + \delta T)] = \frac{\text{cov}[R(T), R(T + \delta T)]}{\text{var} R} \quad (13)$$

and from this, considering that for an Ornstein Uhlenbeck process $\text{cor}[R(T), R(T + \delta T)] = \exp(-\tilde{\rho}|\delta T|)$, the relaxation rate of R as

$$\begin{aligned} \tilde{\rho} &= -\lim_{\delta T \rightarrow 0^+} \frac{d\text{cor}[R(T), R(T + \delta T)]}{d\delta T} \\ &= -\frac{\langle Z \rangle \rho (\langle y \rangle \bar{y} - \langle y^2 \rangle)}{\langle Z \rangle \langle y^2 \rangle} = -\frac{\rho (\langle y \rangle \bar{y} - \langle y^2 \rangle)}{\langle y^2 \rangle}. \end{aligned} \quad (14)$$

Equating $\tilde{\rho}$ and ρ in Eq. (14) yields our self-consistency condition. It simplifies to

$$\bar{y} = 0, \quad \text{and so} \quad p_{\text{inv}} = \Phi(\bar{y}) = \frac{1}{2}, \quad (15)$$

our main result. Our model ecosystems self-organise to exclude half of potential invaders.

Using the fact that dependence on $\langle Z \rangle$ and so C drops out of Eq. (14), we can now address the case of general distributions of the off-diagonal coefficients A_{ij} . Assume first that A_{ij} can attain at random not one but two values other than zero. In this case the dynamics of the sum in Eq. (4) through T is given by a linear combination of the dynamics of two sums of the form of Eq. (6). Since both have the same relaxation rate, Eq. (14), so has Eq. (4). One easily extends this argument to arbitrary discrete distributions for A_{ij} and then to the limit of continuous distributions.

Following [12], Sec. 14.6, we compute further properties of the system steady state. Applying the classical mean first passage time formula [22] to the process, Eq. (7), taking into account that the harvesting resistance of invaders is distributed as in Eq. (8), the mean time between invasion and extinction of a species is $\rho^{-1} \ln(2)$. Since species invade the community at a rate one by construction, the steady-state mean number of species in the community is $\langle S \rangle = \rho^{-1} \ln(2)$, so that

$$\rho = \frac{\ln(2)}{\langle S \rangle} \approx \frac{0.693}{\langle S \rangle}. \quad (16)$$

A previous attempt at deriving p_{inv} [12], Sec. 14.6 had incorrectly assumed $\rho = \langle S \rangle^{-1}$ on heuristic grounds.

To compute the distribution of y in a steady-state community, we construct the steady-state Fokker-Planck equation corresponding to Eq. (7) for $\bar{y} = 0$:

$$0 = \rho \frac{\partial [y P(y)]}{\partial y} + \rho \frac{\partial^2 P(y)}{\partial y^2} + \sqrt{\frac{2}{\pi}} \exp\left(-\frac{y^2}{2}\right). \quad (17)$$

The addition of $P^{\text{inv}}(y)$ in the last term of Eq. (17) describes invasion at rate one of new species with random harvesting resistances. These are balanced by extinctions, described the absorbing boundary conditions:

$$P(0) = 0, \quad \left. \frac{\partial P(y)}{\partial y} \right|_{y=0} = \rho^{-1}. \quad (18)$$

The solution to this equation is the distribution of y amongst extant species in the community, subject to Lotka-Volterra competition and a continuous turnover of interacting species through extinction and invasions. The distribution can be expressed using a hypergeometric function and is shown in Fig. 1. It has the moments $\langle y \rangle = 2^{1/2} \pi^{-1/2} \ln(2)^{-1} = 1.1511$ and $\langle y^2 \rangle = 1 + \ln(4)^{-1} = 1.7213$, which can be used to evaluate other characteristics of the steady state. From Eq. (17.25)

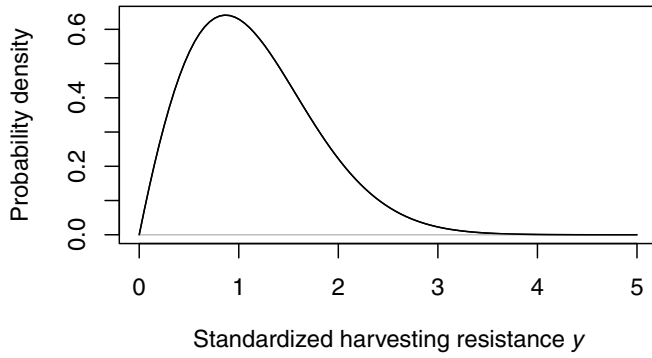


FIG. 1. The parameter-free distribution of standardized harvesting resistance in the steady state of our model.

of Ref. [12], Sec. 14.6, for example, we obtain the approximate mean steady-state species richness

$$\langle S \rangle = \frac{S_{\text{ESI}}}{\langle y^2 \rangle}, \quad (19)$$

and from Eq. (17.26) the approximate standard deviation of the fitness of new random invaders k ,

$$\text{SD}_r = (\text{var} r_k)^{1/2} = \frac{\langle y^2 \rangle}{\langle y \rangle} \cdot \frac{\sigma^2}{\mu(1-\mu)} s. \quad (20)$$

Numerical simulations of the assembly model (detailed in the Appendix) reveal the main effects limiting the accuracy of our calculations. First, r_i as given by Eq. (4) cannot exceed $s_i = s$, which constrains the degree to which the r_i and y can be described by an OU process or Eqs. (17) and (18). This becomes a problem when SD_r , given by Eq. (20), is of the magnitude of s . The smaller SD_r is compared to s ; the better simulations agree with the analytic predictions (Fig. 2). Second, the actual proportion of nonzero off-diagonal entries in $\{A_{ij}\}_{i,j \in \mathcal{E}}$ is smaller than C , possibly a reflection of the constraint that $\{A_{ij}\}_{i,j \in \mathcal{E}}$ must permit coexistence [7,23]. This effect disappears when $\text{SD}_r \ll s$ [Fig. 2(a)]. We therefore extrapolated all our numerical results to $\text{SD}_r \rightarrow 0$ [Figs. 2(b)–2(e)]. In this limit, we verified our main result, Eq. (15),

for invasion probability to within 1.8% relative error; p_{inv} is particularly robust to parameter changes in simulations [Fig. 2(b)]. We further verified Eq. (16) for the relaxation rate of invasion fitness to within 0.6% [Fig. 2(c)], Eq. (19) for expected species richness $\langle S \rangle$ to within 1.0% [Fig. 2(d)], and Eq. (20) for SD_r itself to within 1.9% [Fig. 2(e)].

III. DISCUSSION AND CONCLUSIONS

Structure and dynamics of ecosystems result from processes far from thermodynamic equilibrium, and fundamental physical laws help little in illuminating laws of ecology. Yet, just as quantum physics and relativity explain why some particles have spin 1/2, basic ecological principles imply that $p_{\text{inv}} = 1/2$. Potentially, this is not the only constant of nature emerging from ecological principles. Other candidates are the diet partitioning exponent of predators [24], which empirically appears narrowly constrained around 0.54(2), or the ratio by which species richness declines from one level to the next, observed to be close to 1/3 in both models [12,25] and data [12,26].

Except for substantial work seeking universal exponents linking individual body masses to physiological and ecological rates [27,28], researchers do not generally consider that universal ecological constants might exist. Most ecological studies are designed to identify dependencies between quantities rather than testing constancy under well defined conditions. Nevertheless, there is striking empirical evidence consistent with our prediction that 50% of attempted invasions succeed.

Invasion ecology distinguishes two processes: “establishment,” the successful invasion of a local ecological community by an introduced species, and “spread,” the subsequent successful invasion into a metacommunity. A metaanalysis of the invasion success of vertebrates transported between Europe and North America [20] found that the mean establishment success for species introduced from Europe to North America was $59.6 \pm 11.6\%$ (standard error) and $52.4 \pm 11.9\%$ in the other direction. For spread, the corresponding figures were $65.0 \pm 16.5\%$ and $54.4 \pm 17.4\%$, respectively. A

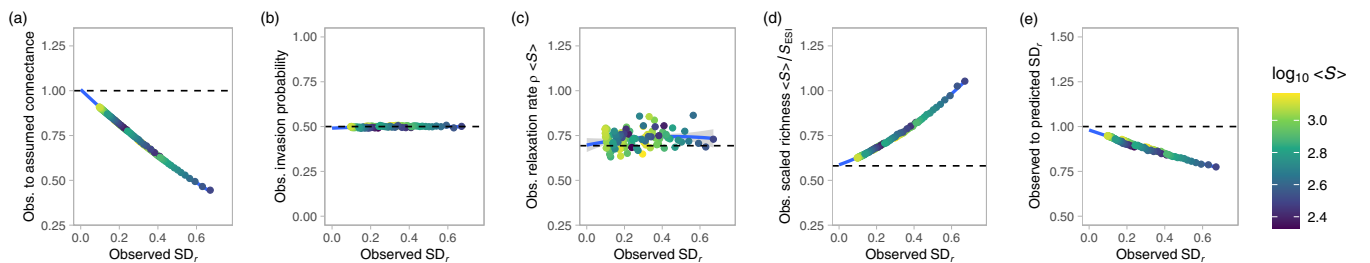


FIG. 2. Numerical test of theory. Panels compare five predictions and assumptions of our theory with steady-state outcomes of 105 simulations. Simulation parameters were chosen as $C = 10^{-2}, 10^{-2.5}, 10^{-3}$, and I such that S_{ESI} as given by Eq. (2) equals 300, 350, 400, ..., 2000 [N.B.: $\mu = IC$, $\sigma^2 = I^2 C(1 - C)$; see Appendix for details]. The abscissa in each graph represents the standard deviation SD_r of the fitness r of random invaders, as observed in simulations. Solid lines and 95% confidence ranges (gray) correspond to the quadratic regressions used to extrapolate to $\text{SD}_r = 0$ in the large-system limit. Horizontal dashed lines represent our analytic predictions. Panel (a) compares the sampled proportion of nonzero off-diagonal entries in $\{A_{ij}\}_{i,j \in \mathcal{E}}$ (“connectance”) with the assumed value C . (b) tests our prediction for invasion probability, Eq. (15), (c) our prediction for the relaxation rate ρ , Eq. (16). Panel (d) compares observed species richness with that predicted by Eq. (19), with S_{ESI} given by model parameters. (e) compares the observed standard deviation of invasion fitness, computed as in Eq. (4), with that predicted by Eq. (20) (using the realized connectance for C rather than the model parameter).

more detailed metaanalysis confirmed that for birds invasion probability is $50 \pm 2.6\%$, but reports $79 \pm 1.7\%$ for mammals [21]. These probabilities are largely unchanged over a factor 10^6 in invaded land mass area [21]. One study [19] estimating the invasion probability for 15 plant species into 0.5×2 m plots in an old field reports probabilities around 50% for most species (for the equilibrium richness of ≈ 13 species per plot, we extract a median of 51% and an average of $56 \pm 5\%$).

Considering the higher value for mammals, we note that our theory assumes that $\tau = \text{cor}(A_{ij}, A_{ji})$ is zero, while interspecific resource competition can lead to $\tau > 0$ and predator-prey interactions to $\tau < 0$. It is known that $\tau > 0$ decreases p_{inv} ([12], Sec. 17.3). Predator-prey interactions between mammals might have the opposite effect, increasing p_{inv} .

Other studies, especially with plants, report lower invasion probabilities. Only 10% of 27 009 plant species introduced into Australia have formed self-sustaining wild populations [29]. However, this figure is dominated by 25 360 ornamental gardening plants [29], for many of which we can assume that the climate and soil of Australia are unsuitable, violating our assumption that all s_i in Eq. (4) are equal. Our theory must break down when variability in s_i (called environmental filtering [30]) is comparable or larger in magnitude than variability in the sum in Eq. (4), given by Eq. (20) (called biotic filtering [31]). Including environmental filtering into the theory is an important next step.

Regarding the solutions of Eq. (1), here we have obtained a comprehensive description of the situation when $S_{\text{pool}} \gg S_{\text{ESI}}$, deep inside the MA phase. The equilibrium species richness $\langle S \rangle$ predicted by Eq. (19) is by a factor $\langle y^2 \rangle$ smaller than the richness S_{ESI} at the phase transition. This is because steady states in species turnover arise already when the system is sufficiently sensitive that the perturbation induced by a new invader leads, on average, to extinction of a rare species, while $S = S_{\text{ESI}}$ corresponds to infinite sensitivity [[12], Eq. (17.24)]. The precise nature of the transition from $S_{\text{pool}} = 2S_{\text{ESI}}$ to $S_{\text{pool}} \gg 2S_{\text{ESI}}$ remains unclear. Simulations indicate the existence of an intermediate parameter range where the community dynamically switches between a few recurring configurations (Clementsian temporal turnover, [6]). Whether this constitutes a distinct phase in the macroecological limit $S_{\text{ESI}} \rightarrow \infty$ remains unclear.

ACKNOWLEDGMENTS

This work forms part of the project ‘‘Mechanisms and prediction of large-scale ecological responses to environmental change’’ funded by the Natural Environment Research Council (NE/T003510/1).

APPENDIX

1. Derivation of Eq. (11)

We derive Eq. (11), making first use of Eq. (10) and then of the fact that $\text{var}Z = \langle Z^2 \rangle - \langle Z \rangle^2 = \langle Z \rangle$ on account of the Poisson distribution of Z :

$$\begin{aligned} & (R(T)R(T + \delta T)) \\ &= \left\langle \sum_{j,l \in \mathcal{A}} y_j(T) y_l(T + \delta T) + \sum_{j \in \mathcal{A}} y_j C y^{\text{inv}} \delta T \right\rangle \end{aligned}$$

$$\begin{aligned} &= \left\langle \sum_{j,l \in \mathcal{A}} y_j [y_l - \rho(y_l - \bar{y}) \delta T] \right\rangle + \langle y \rangle \langle Z \rangle C y^{\text{inv}} \delta T \\ &= \left\langle \sum_{\substack{j,l \in \mathcal{A} \\ j \neq l}} y_j [y_l - \rho(y_l - \bar{y}) \delta T] \right\rangle \\ &\quad + \left\langle \sum_{j \in \mathcal{A}} y_j [y_j - \rho(y_j - \bar{y}) \delta T] \right\rangle + \langle y \rangle \langle Z \rangle C y^{\text{inv}} \delta T \\ &= \langle Z(Z-1) \rangle [\langle y \rangle^2 - \rho(\langle y \rangle^2 - \langle y \rangle \bar{y}) \delta T] \\ &\quad + \langle Z \rangle [\langle y^2 \rangle - \rho(\langle y^2 \rangle - \langle y \rangle \bar{y}) \delta T] \\ &\quad + \langle y \rangle \langle Z \rangle \rho(\langle y \rangle - \bar{y}) \langle Z \rangle \delta T \\ &= \langle Z \rangle^2 \langle y \rangle^2 + \langle Z \rangle \langle y^2 \rangle \\ &\quad + (\langle Z \rangle^2 - \langle Z^2 \rangle) \rho(\langle y \rangle^2 - \langle y \rangle \bar{y}) \delta T \\ &\quad + \langle Z \rangle \rho(\langle y \rangle^2 - \langle y^2 \rangle) \delta T \\ &= \langle Z \rangle^2 \langle y \rangle^2 + \langle Z \rangle \langle y^2 \rangle - \langle Z \rangle \rho(\langle y \rangle^2 - \langle y \rangle \bar{y}) \delta T \\ &\quad + \langle Z \rangle \rho(\langle y \rangle^2 - \langle y^2 \rangle) \delta T \\ &= \langle Z \rangle^2 \langle y \rangle^2 + \langle Z \rangle \langle y^2 \rangle + \langle Z \rangle \rho(\langle y \rangle \bar{y} - \langle y^2 \rangle) \delta T. \end{aligned}$$

2. Computation of harvesting resistance

Without loss of generality, we show that Eq. (4) evaluates to harvesting resistance $r_1 = (d \ln(b_1)/ds_1)^{-1}$ if species $i = 1$ is extant, assuming the other $S - 1$ extant species have indices $2, \dots, S$.

Let \mathbf{A} and \mathbf{s} be, respectively, the interaction matrix and the vector of linear growth rates s_i entering Eq. (3), but restricted to the S extant species. Define $\mathbf{M} = \mathbf{A}^{-1}$. We break both \mathbf{A} and \mathbf{M} into four blocks to separate the entries for the species with index 1 from all other species, assigned a mnemonic index ‘‘o’’

$$\mathbf{A} = \begin{pmatrix} A_{11} & \mathbf{A}_{1o} \\ \mathbf{A}_{o1} & \mathbf{A}_{oo} \end{pmatrix}, \quad (\text{A1})$$

and, using a well-known block-wise inversion formula [32],

$$\begin{aligned} \mathbf{M} &= \begin{pmatrix} M_{11} & \mathbf{M}_{1o} \\ \mathbf{M}_{o1} & \mathbf{M}_{oo} \end{pmatrix} \\ &= \begin{pmatrix} M_{11} & -M_{11} \mathbf{A}_{1o} \mathbf{A}_{oo}^{-1} \\ -\mathbf{A}_{oo}^{-1} \mathbf{A}_{o1} M_{11} & \mathbf{A}_{oo}^{-1} + \mathbf{A}_{oo}^{-1} \mathbf{A}_{o1} M_{11} \mathbf{A}_{1o} \mathbf{A}_{oo}^{-1} \end{pmatrix}, \end{aligned} \quad (\text{A2})$$

where $M_{11} = (A_{11} - \mathbf{A}_{1o} \mathbf{A}_{oo}^{-1} \mathbf{A}_{o1})^{-1}$. Similarly, we break up

$$\mathbf{s} = \begin{pmatrix} s_1 \\ \mathbf{s}_o \end{pmatrix}. \quad (\text{A4})$$

By Eq. (3) of the main text, $\mathbf{b} = \mathbf{A}^{-1} \mathbf{s} = \mathbf{M} \mathbf{s}$ is the vector of the equilibrium biomasses b_i of the extant species. In particular, $b_1 = M_{11} s_1 + \mathbf{M}_{1o} \mathbf{s}_o$. Hence, by the definition of harvesting resistance,

$$r_1 = \left(\frac{d \ln b_1}{ds_1} \right)^{-1} = \frac{b_1}{db_1/ds_1} = \frac{M_{11} s_1 + \mathbf{M}_{1o} \mathbf{s}_o}{M_{11}}. \quad (\text{A5})$$

By Eq. (A2), $\mathbf{M}_{10} = -M_{11}\mathbf{A}_{10}\mathbf{A}_{00}^{-1}$. Inserting this above, M_{11} cancels out and we obtain

$$r_1 = s_1 - \mathbf{A}_{10}\mathbf{A}_{00}^{-1}s_0. \quad (\text{A6})$$

Defining $\mathbf{b}'_0 = \mathbf{A}_{00}^{-1}s_0$, this becomes

$$r_1 = s_1 - \mathbf{A}_{10}\mathbf{b}'_0, \quad (\text{A7})$$

which is Eq. (4) of the main text for $i = 1$, with \mathbf{b}'_0 defined as prescribed below this equation.

Hence, we have shown that Eq. (4) of the main text provides the value of harvesting resistance in line with its definition. The fact that some entries of the abbreviation $\mathbf{b}'_0 = \mathbf{A}_{00}^{-1}s_0$ entering this equation may be negative does not invalidate this derivation and does not imply that negative population biomasses actually arise in our calculation.

3. General simulation method

With a direct simulation of the model, computation time increases with system size approximately as S_{ESI}^4 : by a factor S_{ESI}^2 because of the vector-matrix operation in Eq. (3), by a factor S_{ESI} because the time to reach equilibrium grows approximately as S_{ESI} [33], and by another factor S_{ESI} because turning over an existing community of size $O(S_{\text{ESI}})$, so to generate statistically independent data, requires $O(S_{\text{ESI}})$ invasions.

To allow evaluation of the model for values of S_{ESI} up to 2000, we therefore computed the new equilibrium of Eq. (3) after invasion of a species by directly solving the linear system of equilibrium equations,

$$0 = s_i - \sum_{j \in \mathcal{E}^+} A_{ij}b_j \quad (i \in \mathcal{E}^+), \quad (\text{A8})$$

rather than through numerical integration of the model. When one or more species in the solution of Eq. (A8) had negative population biomass b_j , we considered the one with lowest b_j extinct, removed it from \mathcal{E} , and solved Eq. (A8) once more, repeating until all populations biomasses were nonnegative. In the simulation, we counted as iterations of the model both cases where a randomly sampled species could invade and where it could not.

The above procedure is plausible when only a single species attains negative biomass after an invasion. When there are several candidates for extinction, one might be concerned that the order of species removal affects the outcome. We note, however, that (i) in the model steady state only a single species goes extinct after each successful invasion on average and fluctuations around this mean are small; (ii) indirect interactions between species are of similar magnitude as direct interactions [12,34], implying that extinct species are not usually directly interacting with the added species or with each other; (iii) when a species i goes extinct, $|b_i|$ is generally small both before and after the species addition step (especially for large S_{ESI}), implying that the effect of its removal on other species in \mathcal{E}^+ remains small. Because of this, concerns about the order of species removal have little bearing in practice.

To achieve the desired efficiency gain from directly solving Eq. (A8), one must keep in mind that the direct numerical solution of this equation requires time of order $\mathcal{O}(S_{\text{ESI}}^3)$,

implying that this would not fundamentally improve the algorithm compared to direct simulation. Instead, we therefore kept track of both the matrix $\{A_{ij}\}_{i,j \in \mathcal{E}}$ and its inverse between evaluations of Eq. (A8), using again the block-wise inversion formula [32] to account for addition or removal of a single species in \mathcal{E} .

Specifically, if \mathbf{A}' is the matrix obtained from \mathbf{A} by the addition of row \mathbf{v} and column \mathbf{u} , i.e.,

$$\mathbf{A}' = \begin{pmatrix} \mathbf{A} & \mathbf{u} \\ \mathbf{v}^T & 1 \end{pmatrix}, \quad (\text{A9})$$

then

$$\mathbf{A}'^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + c^{-1}(\mathbf{A}^{-1}\mathbf{u})(\mathbf{v}^T\mathbf{A}^{-1}) & -c^{-1}\mathbf{A}^{-1}\mathbf{u} \\ -c^{-1}\mathbf{v}^T\mathbf{A}^{-1} & c^{-1} \end{pmatrix} \quad (\text{A10})$$

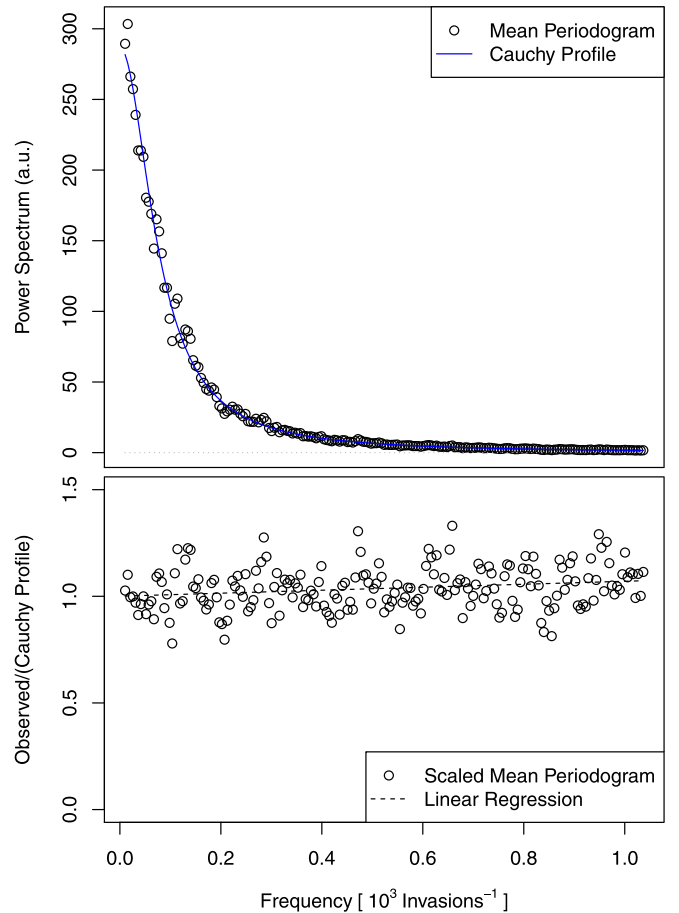


FIG. 3. Test for consistency of invasion fitness (or suppression) with Ornstein-Uhlenbeck process. Top: average of 100 periodograms (circles) compared with a fitted Cauchy profile for a community with $C = 10^{-2}$ and $S_{\text{ESI}} = 2000$, illustrating the good fit. Bottom: ratio of averaged periodogram to Cauchy profile. In principle, the ratio should be one at each frequency, with residual variance close to $1/100 = 0.01$ (since periodograms have a coefficient of variation of one). We find that there is only a small increase with frequency—over the frequency range in which the Cauchy profile declines by a factor 180, the ratio increases by just 7%—and residual variance is 0.0097, demonstrating the good fit.

with $c = 1 - \mathbf{v}^T \mathbf{A}^{-1} \mathbf{u}$, which allowed us to compute $(\mathbf{A}')^{-1}$ from \mathbf{A}^{-1} and vice versa with $\mathcal{O}(S_{\text{ESI}}^2)$ operations. To ensure against loss of accuracy, we computed the inverse of $\{A_{ij}\}_{i,j \in \mathcal{E}}$ directly every 1000 iterations and when $|c| < 10^{-7}$ in a species removal step.

4. Simulation parameters

We evaluated the above assembly model for communities with values of connectance $C = 10^{-2}$, $10^{-2.5}$, 10^{-3} and interaction strength

$$I = [S_{\text{ESI}} C (1 - C)]^{-1/2} \quad (\text{A11})$$

for $S_{\text{ESI}} = 300, 350, 400, \dots, 2000$, running the model over $200 \times S_{\text{ESI}}$ iterations, starting with a single species.

5. Ornstein-Uhlenbeck test and determination of relaxation rate ρ

Of the quantities displayed in Fig. 2 of the main text, only the relaxation rate ρ is difficult to extract from simulations. To obtain ρ , we simulated the invasion fitness of 200 test species (which never actually invade) throughout a simulation.

To avoid artifacts due to small numbers of interaction partners or due to the fact that invasion fitness cannot exceed $s_i (= 1)$ in our model, which might arise for low C , we artificially fixed the connectance of test species (i.e., the probability for these species to be suppressed by a given resident species) at $40/S_{\text{ESI}}$.

For each test species, we sampled the sum of the biomasses b_i of the suppressing resident species (hereafter called “suppression”) after every $S_{\text{ESI}}/10$ iterations and discarded the first 100 samples as burn in. To circumvent the problem of determining the expected mean suppression, we paired the first 100 test species up with the other 100 test species and computed time series for the difference of suppression of each pair, as this difference has mean zero and the same autocorrelation function as both components.

To these 100 time series, we then applied a cosine-bell taper and computed their periodograms. We then averaged the 100 periodograms to estimate the power spectrum. To the lowest 200 nonzero frequencies f of this power spectrum we obtained the least-square fit to a Cauchy profile (the Fourier transform of the exponential autocovariance function of an Ornstein-Uhlenbeck process),

$$\frac{a}{(\rho')^2 + (2\pi f)^2}, \quad (\text{A12})$$

with fitting parameters $a, \rho' > 0$.

As a test for our assumption that invasion fitness (and so suppression) follows an Ornstein-Uhlenbeck process, we verified that the Cauchy profile, Eq. (A12), described the observed power spectra well (see example in Fig. 3). Remaining small deviations from the Cauchy profile are likely attributable to numerical limitations, including the finite duration of the simulated time series.

From the fitted ρ' we computed ρ by taking account of the subsampling of the time series and, dividing by the observed value of p_{inv} , of the fact that T in the main text counts successful invasions only.

-
- [1] M. A. Leibold and J. M. Chase, *Metacommunity Ecology* (Princeton University Press, Princeton, NJ, 2017).
 - [2] J. Harte, T. Zillio, E. Conlisk, and A. B. Smith, Maximum entropy and the state-variable approach to macroecology, *Ecology* **89**, 2700 (2008).
 - [3] J. P. O'Dwyer and J. L. Green, Field theory for biogeography: A spatially explicit model for predicting patterns of biodiversity, *Ecol. Lett.* **13**, 87 (2010).
 - [4] J. Harte, A. B. Smith, and D. Storch, Biodiversity scales from plots to biomes with a universal species–area curve, *Ecol. Lett.* **12**, 789 (2009).
 - [5] J. D. O'Sullivan, R. J. Knell, and A. G. Rossberg, Metacommunity-scale biodiversity regulation and the self-organised emergence of macroecological patterns, *Ecol. Lett.* **22**, 1428 (2019).
 - [6] J. D. O'Sullivan, J. C. D. Terry, and A. G. Rossberg, Intrinsic ecological dynamics drive biodiversity turnover in model metacommunities, *Nat. Commun.* **12**, 3627 (2021).
 - [7] M. Barbier, C. de Mazancourt, M. Loreau, and G. Bunin, Fingerprints of high-dimensional coexistence in complex ecosystems, *Phys. Rev. X* **11**, 011009 (2021).
 - [8] G. Bunin, Ecological communities with Lotka-Volterra dynamics, *Phys. Rev. E* **95**, 042414 (2017).
 - [9] T. Galla, Dynamically evolved community size and stability of random Lotka-Volterra ecosystems, *Europhys. Lett.* **123**, 48004 (2018).
 - [10] A. Altieri, F. Roy, C. Cammarota, and G. Biroli, Properties of equilibria and glassy phases of the random Lotka-Volterra model with demographic noise, *Phys. Rev. Lett.* **126**, 258301 (2021).
 - [11] M. Barbier, J.-F. Arnoldi, G. Bunin, and M. Loreau, Generic assembly patterns in complex ecological communities, *Proc. Natl. Acad. Sci. USA* **115**, 2156 (2018).
 - [12] A. G. Rossberg, *Food Webs and Biodiversity: Foundations, Models, Data* (Wiley, 2013).
 - [13] T. Tao, V. Vu, and M. Krishnapur, Random matrices: Universality of ESDs and the circular law, *Ann. Probab.* **38**, 2023 (2010).
 - [14] J. Hofbauer, Heteroclinic cycles in ecological differential equations, *Equadiff 8* (Mathematical Institute, Slovak Academy of Sciences, Bratislava, 1994), pp. 105–116.
 - [15] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Giant strongly connected component of directed networks, *Phys. Rev. E* **64**, 025101(R) (2001).
 - [16] F. Roy, M. Barbier, G. Biroli, and G. Bunin, Complex interactions can create persistent fluctuations in high-diversity ecosystems, *PLoS Comput. Biol.* **16**, e1007827 (2020).
 - [17] W. M. Post and S. L. Pimm, Community assembly and food web stability, *Math. Biosci.* **64**, 169 (1983).
 - [18] R. M. Arthur, Species packing, and what competition minimizes, *Proc. Natl. Acad. Sci. USA* **64**, 1369 (1969).

- [19] S. J. Meiners, M. L. Cadenasso, and S. T. A. Pickett, Beyond biodiversity: Individualistic controls of invasion in a self-assembled community, *Ecol. Lett.* **7**, 121 (2004).
- [20] J. M. Jeschke and D. L. Strayer, Invasion success of vertebrates in Europe and North America, *Proc. Natl. Acad. Sci. USA* **102**, 7198 (2005).
- [21] J. M. Jeschke, Across islands and continents, mammals are more successful invaders than birds, *Divers. Distrib.* **14**, 913 (2008).
- [22] C. W. Gardiner, *Handbook of Stochastic Methods*, 2nd ed. (Springer, Berlin, 1990).
- [23] G. Bunin, Interaction patterns and diversity in assembled ecological communities, [arXiv:1607.04734v1](https://arxiv.org/abs/1607.04734v1) [q-bio.PE].
- [24] A. G. Rossberg, K. D. Farnsworth, K. Satoh, and J. K. Pinnegar, Universal power-law diet partitioning by marine fish and squid with surprising stability-diversity implications, *Proc. R. Soc. B* **278**, 1617 (2011).
- [25] A. G. Rossberg, R. Ishii, T. Amemiya, and K. Itoh, The top-down mechanism for body-mass–abundance scaling, *Ecology* **89**, 567 (2008).
- [26] S. Jennings, K. J. Warr, and S. Mackinson, Use of size-based production and stable isotope analyses to predict trophic transfer efficiencies and predator-prey body mass ratios in food webs, *Mar. Ecol. Prog. Ser.* **240**, 11 (2002).
- [27] R. H. Peters, *The Ecological Implications of Body Size* (Cambridge University Press, Cambridge, 1983).
- [28] G. B. West, J. H. Brown, and B. J. Enquist, A general model for the origin of allometric scaling laws in biology, *Science* **276**, 122 (1997).
- [29] J. G. Virtue, S. J. Bennett, and R. P. Randall, Plant introductions in Australia: How can we resolve 'weedy' conflicts of interest? in *Weed Management: Balancing People, Planet, Profit. 14th Australian Weeds Conference, Wagga Wagga, New South Wales, Australia, 6-9 September 2004: Papers and Proceedings* (Weed Society of New South Wales, 2004), pp. 42–48.
- [30] N. J. B. Kraft, P. B. Adler, O. Godoy, E. C. James, S. Fuller, and J. M. Levine, Community assembly, coexistence and the environmental filtering metaphor, *Funct. Ecol.* **29**, 592 (2015).
- [31] O. Boet, X. Arnan, and J. Retana, The role of environmental vs. biotic filtering in the structure of European ant communities: A matter of trait type and spatial scale, *PLOS ONE* **15**, e0228625 (2020).
- [32] W. W. Hager, Updating the inverse of a matrix, *SIAM Rev.* **31**, 221 (1989).
- [33] A. G. Rossberg, A. Caskenette, and L.-F. Bersier, Structural instability of food webs and food-web models and their implications for management, in *Adaptive Food Webs: Stability and Transitions of Real and Model Ecosystems*, edited by J. C. Moore, P. C. de Ruiter, K. S. McCann, and W. V. (Cambridge University Press, Cambridge, 2017), pp. 373–383.
- [34] T. Fung, K. D. Farnsworth, D. G. Reid, and A. G. Rossberg, Impact of biodiversity loss on production in complex marine food webs mitigated by prey-release, *Nat. Commun.* **6**, 6657 (2015).