



HAL
open science

Equilibrium and surviving species in a large Lotka-Volterra system of differential equations

Maxime Clenet, François Massol, Jamal Najim

► **To cite this version:**

Maxime Clenet, François Massol, Jamal Najim. Equilibrium and surviving species in a large Lotka-Volterra system of differential equations. 2022. <hal-03655449>

HAL Id: hal-03655449

<https://hal.science/hal-03655449v1>

Preprint submitted on 29 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Equilibrium and surviving species in a large Lotka-Volterra system of differential equations

Maxime Clenet⁽¹⁾, François Massol⁽²⁾, Jamal Najim^{(1)*}

(1) CNRS and Université Gustave Eiffel, France

(2) CNRS, Université de Lille, INSERM, CHU, Institut Pasteur
Lille

April 29, 2022

Abstract

Lotka-Volterra (LV) equations play a key role in the mathematical modeling of various ecological, biological and chemical systems. When the number of species (or, depending on the viewpoint, chemical components) becomes large, basic but fundamental questions such as computing the number of surviving species still lack theoretical answers. In this paper, we consider a large system of LV equations where the interactions between the various species are a realization of a random matrix. We provide conditions to have a unique equilibrium and present a heuristics to compute the number of surviving species. This heuristics combines arguments from Random Matrix Theory, mathematical optimization (LCP), and standard extreme value theory. Numerical simulations, together with an empirical study where the strength of interactions evolves with time, illustrate the accuracy and scope of the results.

1 Introduction

Since May's seminal work [May72] and for the past decades, many theoretical studies addressed the issue of the coexistence of species in ecosystems.

Introduced in the 1920s by Lotka [Lot25] and Volterra [Vol26], the Lotka-Volterra (LV) model is a well-known classic in theoretical ecology and mathematics. It represents a first step in our understanding of ecosystems through the variety of its dynamical behaviours (single or multiple equilibria, cycles, chaos), its flexibility (many models can be approximated in the form of a LV model) and its mathematical calculability.

*Supported by CNRS Project 80 Prime — KARATE.

In this article, we consider large LV models with random parameters. Leveraging on the asymptotic understanding of large random matrices which naturally appear enables us to provide insights on equilibria and species coexistence for such models.

Model and assumptions.

Large Lotka-Volterra systems of differential equations arise in various scientific fields such as biology, ecology, chemistry, etc. Although our results are generic in nature and not specific to a given field, we will rely on the ecological terminology in the sequel.

A large system of Lotka-Volterra equations is a system of coupled ordinary differential equations (ODE) that write:

$$\frac{dx_k(t)}{dt} = x_k(t) \left(r_k - \theta x_k(t) + \sum_{\ell \in [n]} B_{k\ell} x_\ell(t) \right), \quad (1)$$

where $k \in [n] = \{1, \dots, n\}$.

Here, n represents the number of species in a food web or community, the unknown vector $\mathbf{x} = (x_k)_{k \in [n]}$ is the vector of abundances of the various species and evolves with time $t > 0$ according to the dynamics (1). Parameter r_k represents the intrinsic growth rate of species k , θ is an intraspecific feedback coefficient (most often positive due to competition), and $B_{k\ell}$ is the per capita effect of species ℓ on species k .

Remark 1. Notice that without interactions, i.e. $B = (B_{k\ell})_{k, \ell \in [n]} = 0$, system (1) is simply a system of uncoupled logistic differential equations.

We shall focus on the model where $r_k = \theta = 1$:

$$\frac{dx_k}{dt} = x_k (1 - x_k + (B\mathbf{x})_k), \quad k \in [n] \quad (2)$$

with matrix B admitting the following representation:

$$B = \frac{A}{\alpha\sqrt{n}} + \frac{\mu}{n} \mathbf{1}\mathbf{1}^T,$$

where $A = (A_{ij})$ is a matrix with random standardized ($\mathbb{E}A_{ij} = 0$ and $\text{var}(A_{ij}) = 1$) independent and identically distributed (i.i.d.) entries with finite fourth moment, $\alpha > 0$ is an extra parameter reflecting the interaction strength, and $\mu \in \mathbb{R}$ represents an arbitrary trend of the interactions. The $n \times 1$ vector $\mathbf{1}$ is a vector of ones.

Remark 2. Although matrix B is a complex random object, a result by Tao [Tao13, Theorem 1.7] fully describes its asymptotic spectrum: Assume that $|\mu| > 1/\alpha$, then for any fixed $\varepsilon > 0$, almost surely eventually all the eigenvalues of B but one are in the disk $\{z \in \mathbb{C} : |z| \leq 1/\alpha + \varepsilon\}$ while one extra eigenvalue takes the value $\mu + o(1)$.

Remark 2 is illustrated in Fig. 1.

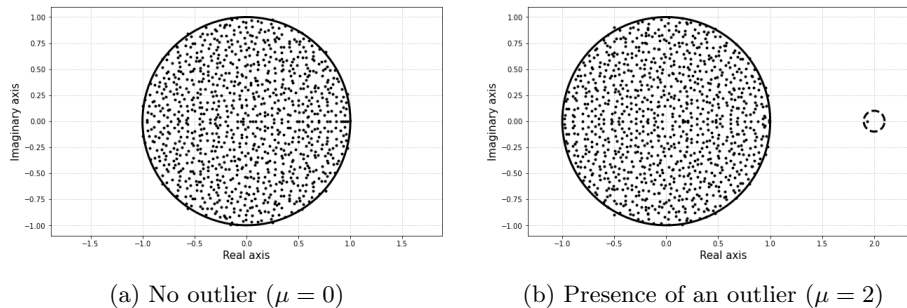


Figure 1: Spectrum of non-Hermitian matrix B in the complex plan ($n = 1000$, $\alpha = 1$). In Fig. 1a, $\mu = 0$ and the solid line circle represents the boundary of the circular law. In Fig. 1b, $\mu = 2$ and there is an eigenvalue in the small dashed circle centered at 2, as predicted by [Tao13, Th. 1.7] - see also Remark 2.

Presentation of the main results.

Unique equilibrium. In the study of the behaviour of $\mathbf{x}(t)$ as $t \rightarrow \infty$ the existence of an equilibrium \mathbf{x}^* to Eq. (2) is an important prior to any stability property of $\mathbf{x}(t)$. By equilibrium, we mean the existence of a vector $\mathbf{x}^* = (x_k^*)$ satisfying

$$x_k^*(1 - x_k^* + (B\mathbf{x}^*)_k) = 0 \quad \text{for } k \in [n].$$

General results on LV systems state that $\mathbf{x}(t) > 0$ (componentwise) as long as $\mathbf{x}(0) > 0$ [HS98]. However, a possible equilibrium \mathbf{x}^* will only verify $\mathbf{x}^* \geq 0$, i.e. some components x_k^* will take the value zero.

In Theorem 2, we provide sufficient conditions on the parameters α and μ to ensure the existence of a unique equilibrium. These conditions rely on the “typical” behaviour of the random matrix B in large dimension $n \rightarrow \infty$.

Evaluating the number of surviving species. Given a unique equilibrium \mathbf{x}^* , an important question is to describe the set of surviving/vanishing species. In this perspective, we introduce the set

$$\mathcal{S} = \{i \in [n], x_i^* > 0\} \tag{3}$$

of surviving species. In Section 3, we provide a heuristics to compute asymptotically the ratio $\frac{|\mathcal{S}|}{n}$ and understand via a system of equations the dependence between parameters α and μ and the number of surviving species. A complementary result addressing the elliptic random matrix model by means of theoretical physics methods can be found in [Bun17] (dynamical cavity method) and in [Gal18] (generating functional techniques).

Notice that in [BN21], Bizeul and Najim have studied a different normalization for α in the case $\mu = 0$, namely $\alpha \sim \sqrt{2 \log(n)}$, to guarantee the survival of every species (feasibility of the equilibrium). Indeed, a consequence of Dougoud

et al.'s results [DVR⁺18] is that some species will go to extinction if $\alpha > 0$ is fixed (i.e. does not increase sufficiently with n).

An empirical study of LV systems with changing interaction strengths.

Equipped with results on the existence of a unique equilibrium, one pending question is to understand what happens when the coefficient α varies for the same matrix A and the same parameter μ . In particular, when the value of α increases above a certain critical value, all species will coexist [BN21]; conversely, for sufficiently low values of α , the existence of a feasible equilibrium is not warranted anymore, however a unique and stable equilibrium may exist. How species equilibrium abundances change between these two states and how $|S|$ varies will be the focus of Section 4.

Notations

Denote by $\rho(C)$ the spectral radius of matrix C , by $\|C\|$ the spectral norm of matrix C , and by $\|\mathbf{u}\|$ the euclidean norm of vector \mathbf{u} . We represent by δ_x the Dirac measure at x :

$$\delta_x(E) = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{else} \end{cases}.$$

We denote by $\xrightarrow{a.s.}$ the almost sure convergence of random quantities and by \xrightarrow{weak} the weak convergence of measures. Given a set S , we denote by $|S|$ its cardinality.

2 Equilibrium and stability results

A primer on Random Matrix Theory. We first recall some results on Random Matrix Theory (RMT), which provides a number of valuable insights to understand the asymptotic behaviour of A . We begin by the almost sure (a.s.) convergence of the spectral radius and the spectral norm:

$$\rho(A/\sqrt{n}) \xrightarrow[n \rightarrow \infty]{a.s.} 1 \quad \text{and} \quad \|A/\sqrt{n}\| \xrightarrow[n \rightarrow \infty]{a.s.} 2.$$

We also have the a.s. weak convergence of the spectral measure of A/\sqrt{n} to the circular law (see for instance [BC12]):

$$(a.s.) \quad \frac{1}{n} \sum_{k \in [n]} \delta_{\lambda_k(A/\sqrt{n})} \xrightarrow[n \rightarrow \infty]{weak} \frac{\mathbf{1}_{\{x^2+y^2 \leq 1\}}}{\pi} dx dy,$$

where $(\lambda_k(A/\sqrt{n}); k \in [n])$ is the spectrum of A/\sqrt{n} . This convergence is illustrated in Fig. 1a.

The description of the spectral norm of the deterministic part of matrix B is more straightforward:

$$\left\| \mu \frac{\mathbf{1}\mathbf{1}^T}{n} \right\| = |\mu|.$$

Notice that both the random and deterministic parts of matrix B do not vanish asymptotically and thus have a macroscopic effect on the dynamics of system (1), as recalled in Remark 2 where the asymptotic spectrum of B is described.

The non-invadability condition. A key element to understand the dynamics of the LV system (1) is the existence of an equilibrium $\mathbf{x}^* = (x_k^*)_{k \in [n]}$ such that

$$\begin{cases} x_k^* (1 - x_k^* + (B\mathbf{x}^*)_k) = 0, & \forall k \in [n], \\ x_k^* \geq 0. \end{cases} \quad (4)$$

and the study of its stability, that is the convergence of a solution \mathbf{x} to the equilibrium \mathbf{x}^* : $\mathbf{x}(t) \xrightarrow[t \rightarrow \infty]{} \mathbf{x}^*$ if $\mathbf{x}(0)$ is sufficiently close to \mathbf{x}^* .

It is well known that for LV equations, the fact that $\mathbf{x}(0) > 0$ (component-wise) implies that $\mathbf{x}(t) > 0$ for every $t > 0$, but one can have some components $x_k(t)$ of $\mathbf{x}(t)$ vanishing to zero. As a consequence, we will only consider non negative equilibria $\mathbf{x}^* \geq 0$ with possibly vanishing components.

Notice that the situation substantially differs whether $\mathbf{x}^* > 0$ or \mathbf{x}^* has vanishing component. In the former case, the equilibrium set of equations becomes a linear equation:

$$\mathbf{x}^* = \mathbf{1} + B\mathbf{x}^* .$$

In the latter case, the equilibrium equations are no longer linear.

In the centered case $\mu = 0$, the existence of a positive solution has been studied in [BN21] and requires $\alpha \gg \sqrt{2 \log(n)}$ (while we consider α fixed here).

A naive and systematic way to solve (4) is to choose a priori a subset $\mathcal{I} \subset [n]$, to set the corresponding components $\mathbf{x}_{\mathcal{I}} = (x_i^*)_{i \in \mathcal{I}}$ to zero, and to solve the remaining linear system:

$$\mathbf{x}_{\mathcal{I}^c} = \mathbf{1}_{|\mathcal{I}^c|} + B_{\mathcal{I}^c} \mathbf{x}_{\mathcal{I}^c} .$$

If there exists $\mathbf{x}_{\mathcal{I}^c} \geq 0$ that solves the previous equation, then $\mathbf{x} = \begin{pmatrix} \mathbf{x}_{\mathcal{I}} \\ \mathbf{x}_{\mathcal{I}^c} \end{pmatrix}$ satisfies (4) and is a potential equilibrium. The number of subcases $\mathcal{I} \subset [n]$ is 2^n and in particular grows exponentially as $n \rightarrow \infty$.

In order to decrease the number of potential solutions to (4), we first notice that relying on standard properties of dynamical systems, see for instance [Tak96, Theorem 3.2.5], a necessary condition for the equilibrium \mathbf{x}^* to be stable is that

$$1 - x_k^* + (B\mathbf{x}^*)_k \leq 0 . \quad (5)$$

The condition (5) is better known in ecology as the non-invadability condition [LM96]. In reference to the ODE (2), the requirement for a given species $k \in [n]$ to be non-invasive is equivalent to:

$$\left(\frac{1}{x_k} \frac{dx_k}{dt} \right)_{x_k \rightarrow 0^+} \leq 0 . \quad (6)$$

The main interpretation is as follows: if one adds species k with a very low abundance in the system, it will not be able to invade the system as a result of condition (6).

As a consequence, we will now focus on the following set of conditions:

$$\begin{cases} x_k^* (1 - x_k^* + (B\mathbf{x}^*)_k) = 0 & \text{for } k \in [n], \\ 1 - x_k^* + (B\mathbf{x}^*)_k \leq 0 & \text{for } k \in [n], \\ \mathbf{x}^* \geq 0 & \text{componentwise.} \end{cases} \quad (7)$$

This casts the problem of finding a non negative equilibrium into the class of Linear Complementarity Problems (LCP), which we describe hereafter.

Linear Complementarity Problem (LCP). LCP is a class of problems from mathematical optimization which in particular encompasses linear and quadratic programs; standard references are [MY88, CPS09]. Given a $n \times n$ matrix M and a $n \times 1$ vector \mathbf{q} , the associated LCP denoted by $LCP(M, \mathbf{q})$ consists in finding two $n \times 1$ vectors \mathbf{z}, \mathbf{w} satisfying the following set of constraints:

$$\begin{cases} \mathbf{z} \geq 0, \\ \mathbf{w} = M\mathbf{z} + \mathbf{q} \geq 0, \\ \mathbf{w}^T \mathbf{z} = 0 \end{cases} \Leftrightarrow w_k z_k = 0 \quad \text{for all } k \in [n]. \quad (8)$$

Since \mathbf{w} can be inferred from \mathbf{z} , we denote $\mathbf{z} \in LCP(M, \mathbf{q})$ if (\mathbf{w}, \mathbf{z}) is a solution of (8).

A theorem by Murty [Mur72] states that the $LCP(M, \mathbf{q})$ has a unique solution (\mathbf{w}, \mathbf{z}) iff M is a P -matrix, that is:

$$\det(M_{\mathcal{I}}) > 0, \quad \forall \mathcal{I} \subset [n], \quad M_{\mathcal{I}} = (M_{k\ell})_{k,\ell \in \mathcal{I}}.$$

In view of (7), we look for $\mathbf{x}^* \in LCP(I - B, -\mathbf{1})$.

The equilibrium \mathbf{x}^* and its stability. For a generic LV system

$$\frac{dy_k(t)}{dt} = y_k(r_k + (C\mathbf{y})_k), \quad k \in [n], \quad (9)$$

Takeuchi and Adachi (see for instance [Tak96, Th. 3.2.1]) provide a criterion for the existence of a unique equilibrium \mathbf{y}^* and the global stability of the LV system.

Theorem 1 (Takeuchi and Adachi [TA80]). *If there exists a positive diagonal matrix Δ such that $\Delta C + C^T \Delta$ is negative definite, then $LCP(-C, \mathbf{r})$ admits a unique solution. In particular, for every $\mathbf{r} \in \mathbb{R}^n$, there is a unique equilibrium \mathbf{y}^* to (9), which is globally stable in the sense that for every $\mathbf{y}_0 > 0$, the solution to (9) which starts at $\mathbf{y}(0) = \mathbf{y}_0$ satisfies*

$$\mathbf{y}(t) \xrightarrow[t \rightarrow \infty]{} \mathbf{y}^*.$$

Combining this result (setting $C = -(I - B)$) with results from RMT, we can guarantee the existence of a globally stable equilibrium \mathbf{x}^* of (1) for a wide range of the set (α, μ) . Denote by

$$\mathcal{A} = \left\{ (a, m) \in \mathbb{R}_+^* \times \mathbb{R} : a > \sqrt{2}, m < \frac{1}{2} + \frac{1}{2} \sqrt{1 - \frac{2}{a^2}} \right\} \quad (10)$$

the set of *admissible parameters*.

Theorem 2. *Let $(\alpha, \mu) \in \mathcal{A}$, then a.s. matrix $(I - B) + (I - B)^T$ is eventually positive definite: with probability one, for a given realization ω , there exists $N(\omega)$ such that for $n \geq N(\omega)$, $(I - B^\omega) + (I - B^\omega)^T$ is positive definite. In particular, there exists a unique (random) globally stable equilibrium $\mathbf{x}^* \in LCP(I - B^\omega, -\mathbf{1})$ to (7).*

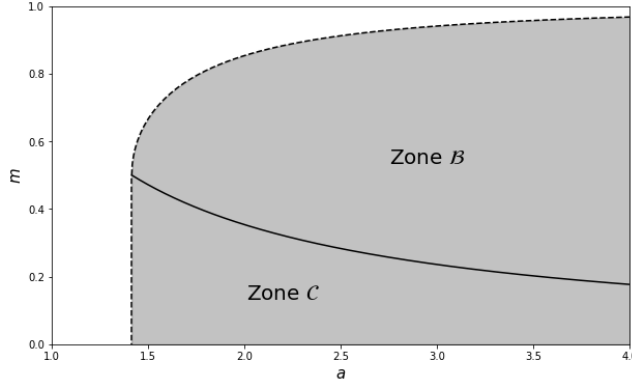


Figure 2: The shaded area represents the set \mathcal{A} given by (10) yielding the existence of a unique (random) globally stable equilibrium \mathbf{x}^* . Area \mathcal{A} is divided into two zones \mathcal{B} and \mathcal{C} . Both zones correspond to parameters (α, μ) for which matrix $2I - (B + B^T)$ is definite positive, as stated in Theorem 2. In zone \mathcal{B} , $\lambda_{\max}(B + B^T)$ corresponds to a spiked eigenvalue (μ above the critical threshold $(\alpha\sqrt{2})^{-1}$). In zone \mathcal{C} , $\lambda_{\max}(B + B^T)$ corresponds to the right edge of the semi-circle law. Notice that zone \mathcal{C} extends to negative values along the y -axis.

Proof. We have

$$I - B + I - B^T = 2I - (B + B^T) = 2I - \left(\frac{A + A^T}{\alpha\sqrt{n}} + \frac{2\mu}{n} \mathbf{1}\mathbf{1}^* \right).$$

Notice that $2I - (B + B^T)$ is positive definite iff the top eigenvalue of $B + B^T$ is lower than 2:

$$\lambda_{\max}(B + B^T) < 2 \quad (11)$$

We first focus on the random part $(A + A^T)/\alpha$ which is a symmetric matrix with independent $\mathcal{N}(0, 2/\alpha^2)$ entries above the diagonal (note that the distribution of the diagonal entries is different from the off-diagonal entries, with no asymptotic effect). In this case, it is well known that the largest eigenvalue of the normalized matrix (or equivalently its spectral norm since the matrix is symmetric) a.s. converges to the right edge of the support of the semi-circle law (see [BS10, Th. 5.2]):

$$\lambda_{\max} \left(\frac{A + A^T}{\alpha\sqrt{n}} \right) \xrightarrow[n \rightarrow \infty]{a.s.} \frac{2\sqrt{2}}{\alpha}. \quad (12)$$

In the centered case ($\mu = 0$), condition (11) occurs if $\alpha > \sqrt{2}$.

We now consider the general case where $\mu \neq 0$. Notice that the rank-one perturbation matrix $P = \frac{2\mu}{n} \mathbf{1}\mathbf{1}^*$ admits a unique non zero eigenvalue 2μ . Denote by $\check{A} = \frac{A + A^T}{\alpha\sqrt{n}}$. We are concerned with the top eigenvalue of the symmetric matrix $\check{A} + P$. Based on a result by Capitaine et al. [CDMF09, Th. 2.1], we have:

$$\lambda_{\max}(\check{A} + P) \xrightarrow[n \rightarrow \infty]{a.s.} \begin{cases} 2\mu + \frac{1}{\alpha^2\mu} & \text{if } \mu > \frac{1}{\sqrt{2}\alpha}, \\ \frac{2\sqrt{2}}{\alpha} & \text{else.} \end{cases}$$

This result is illustrated in Figure 3.

Assume first that $\mu \leq \frac{1}{\alpha\sqrt{2}}$ (corresponding to zone \mathcal{C} in Fig. 2), then $\lambda_{\max}(\check{A} + P) \xrightarrow[n \rightarrow \infty]{a.s.} \frac{2\sqrt{2}}{\alpha}$, which is strictly lower than 2 (cf. condition (11)) if $\alpha > \sqrt{2}$. Hence $\lambda_{\max}(\check{A} + P)$ is eventually strictly lower than 2 under this condition.

Assume now that $\mu > \frac{1}{\alpha\sqrt{2}}$ (corresponding to zone \mathcal{B} in Fig. 2), then

$$\lambda_{\max}(\check{A} + P) \xrightarrow[n \rightarrow \infty]{a.s.} 2\mu + \frac{1}{\alpha^2\mu}.$$

We are interested in the conditions for which $2\mu + \frac{1}{\alpha^2\mu} < 2$ or equivalently

$$2\alpha^2\mu^2 - 2\alpha^2\mu + 1 < 0. \quad (13)$$

An elementary study of the polynomial $\xi(X) = 2\alpha^2X^2 - 2\alpha^2X + 1$ yields that ξ 's discriminant is positive if $\alpha > \sqrt{2}$,

$$\xi(\mu^\pm) = 0 \quad \Leftrightarrow \quad \mu^\pm = \frac{1}{2} \pm \frac{1}{2} \sqrt{1 - \frac{2}{\alpha^2}},$$

and $\xi\left(\frac{1}{\alpha\sqrt{2}}\right) < 0$, so that $\frac{1}{\alpha\sqrt{2}} \in (\mu^-, \mu^+)$. In particular condition (13) is fulfilled if

$$\mu \in \left(\frac{1}{\alpha\sqrt{2}}, \frac{1}{2} + \frac{1}{2} \sqrt{1 - \frac{2}{\alpha^2}} \right).$$

Under this condition, (13) is fulfilled and a.s. $\limsup_{n \rightarrow \infty} \lambda_{\max}(\check{A} + P) < 2$, which completes the proof: we can then rely on Theorem 1 to conclude. \square

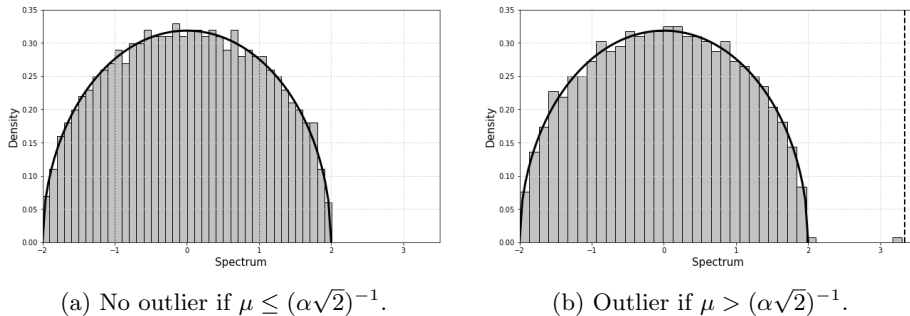


Figure 3: Spectrum (histogram) of the Hermitian random matrix $B + B^T$ ($n = 1000$, $\alpha = \sqrt{2}$). The solid line represents the semi-circular law. In Fig. 3a, $\mu = 0$ and there is no outlier. In Fig. 3b, $\mu = 1.5$ and one can notice the presence of an eigenvalue outside the bulk of the circular law. The dashed line indicates its theoretical value.

3 A heuristic approach to the proportion and distribution of the surviving species

3.1 Proportion of surviving species

In Section 2, we have presented conditions on parameters α, μ for the existence of a globally stable equilibrium \mathbf{x}^* to (1) under the non-invadability condition. As \mathbf{x}^* depends on the realization of matrix B , it is a random vector. Moreover since $\alpha > 0$ is fixed and does not depend on n , the equilibrium \mathbf{x}^* will feature vanishing components (see the original argument in [DVR⁺18] and the discussion in [BN21]). In an ecological context, we shall refer to these non-vanishing components as the surviving species, the vanishing components corresponding to the species going to extinction with $x_k^* = 0$ and $x_k(t) \xrightarrow[t \rightarrow \infty]{} 0$.

In this section, we assume that the A_{ij} 's are $\mathcal{N}(0, 1)$ -distributed and describe the proportion of non-vanishing components of the equilibrium \mathbf{x}^* ; we also describe the distribution of the surviving species $x_i^* > 0$ which turns out to be a truncated Gaussian.

Remark 3. *The Gaussianity assumption facilitates the explanation of the heuristics but does not seem necessary for the result to hold. In Fig. 6b, the entries are not considered Gaussian but the distribution of the surviving species still matches the truncated Gaussian.*

Given the random equilibrium \mathbf{x}^* , recall the definition of \mathcal{S} in (3). We introduce the following quantities:

$$\hat{p} = \frac{|\mathcal{S}|}{n}, \quad \hat{m} = \frac{1}{|\mathcal{S}|} \sum_{i \in [n]} x_i^*, \quad \hat{\sigma}^2 = \frac{1}{|\mathcal{S}|} \sum_{i \in [n]} (x_i^*)^2.$$

Notice that in the definitions of \hat{m} and $\hat{\sigma}^*$ we can replace $\sum_{i \in [n]}$ by $\sum_{i \in \mathcal{S}}$.

Denote by $Z \sim \mathcal{N}(0, 1)$ a standard Gaussian random variable and by Φ the cumulative Gaussian distribution function:

$$\Phi(x) = \int_{-\infty}^x \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} du.$$

Recall the definition of the set \mathcal{A} in (10).

Heuristics 1. Let $(\alpha, \mu) \in \mathcal{A}$. The following system of three equations and three unknowns (p, m, σ)

$$\sigma\sqrt{p}\Phi^{-1}(1-p) + \alpha(1 + \mu pm) = 0, \quad (14)$$

$$1 + \mu pm + \frac{\sigma\sqrt{p}}{\alpha}\mathbb{E}(Z | Z > -\delta) = m, \quad (15)$$

$$(1 + \mu pm)^2 + (1 + \mu pm)\frac{2\sigma\sqrt{p}}{\alpha}\mathbb{E}(Z | Z > -\delta) + \frac{\sigma^2 p}{\alpha^2}\mathbb{E}(Z^2 | Z > -\delta) = \sigma^2 \quad (16)$$

where

$$\delta = \delta(p, m, \sigma) = \frac{\alpha}{\sigma\sqrt{p}}(1 + \mu pm), \quad (17)$$

admits a unique solution (p^*, m^*, σ^*) and

$$\hat{p} \xrightarrow[n \rightarrow \infty]{a.s.} p^*, \quad \hat{m} \xrightarrow[n \rightarrow \infty]{a.s.} m^* \quad \text{and} \quad \hat{\sigma} \xrightarrow[n \rightarrow \infty]{a.s.} \sigma^*.$$

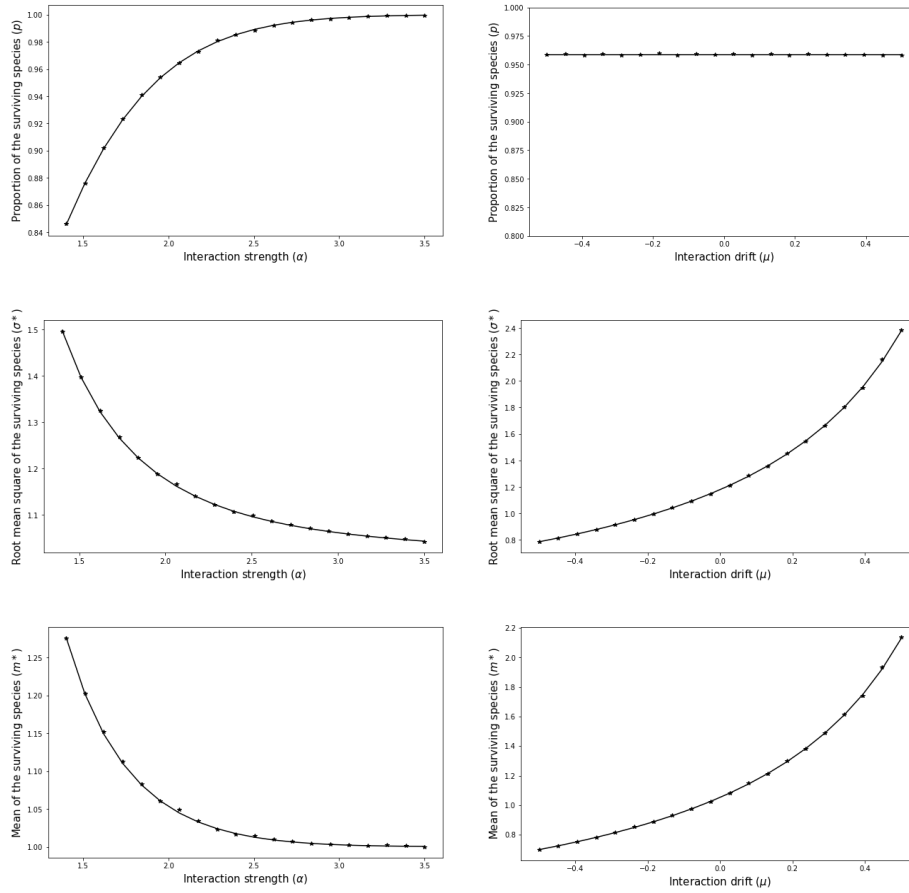
Associated to this solution (p^*, m^*, σ^*) is $\delta^* = \delta(p^*, m^*, \sigma^*)$.

There is a strong matching between the parameters obtained by solving (14)-(16) and their empirical counterparts obtained by Monte-Carlo simulations. This is illustrated in Fig. 4. In Fig. 5, we illustrate the sensitivity of σ^* to the parameters (α, μ) .

Remark 4. The heuristics above substantially simplifies in the centered model case, where $\mu = 0$ and $B = \frac{A}{\alpha\sqrt{n}}$. Following (12), assume that $\alpha > \sqrt{2}$. Then the system with two unknowns (p, σ)

$$\begin{cases} \sigma\sqrt{p}\Phi^{-1}(1-p) + \alpha & = 0 \\ 1 + \frac{2\sigma\sqrt{p}}{\alpha}\mathbb{E}(Z | Z > -\delta) + \frac{\sigma^2 p}{\alpha^2}\mathbb{E}(Z^2 | Z > -\delta) & = \sigma^2 \end{cases} \quad \text{where } \delta = \frac{\alpha}{\sigma\sqrt{p}}$$

admits a unique solution (p^*, σ^*) . Moreover, $\hat{p} \xrightarrow[n \rightarrow \infty]{a.s.} p^*$ and $\hat{\sigma} \xrightarrow[n \rightarrow \infty]{a.s.} \sigma^*$.



(a) Parameters (p^*, σ^*, m^*) versus α .

(b) Parameters (p^*, σ^*, m^*) versus μ .

Figure 4: The plots represent a comparison between the theoretical solutions (p^*, σ^*, m^*) of (14)-(16) and their empirical Monte Carlo counterpart (the star marker) as functions of the interaction strength α (left) and the interaction drift μ (right). Matrix B has size $n = 500$ and the number of Monte Carlo experiments is 200. In Column (4a), $\mu = 0$ and $\alpha > \sqrt{2}$ on the x -axis (which guarantees a unique and stable equilibrium \mathbf{x}^*). When interaction α^{-1} increases, the number of surviving species p^* decrease but their variance σ^* and mean m^* increase. In Column (4b), $\alpha = 2$ and $\mu \in (-0.5, 0.5)$ on the x -axis. The interaction drift appears to have no impact on the proportion p^* of surviving species, whereas it influences their variances and means.

3.2 Distribution of surviving species

In the previous section, the proportion \hat{p} of the surviving species, their mean \hat{m} and second moment $\hat{\sigma}^2$ have been described as empirical counterparts of

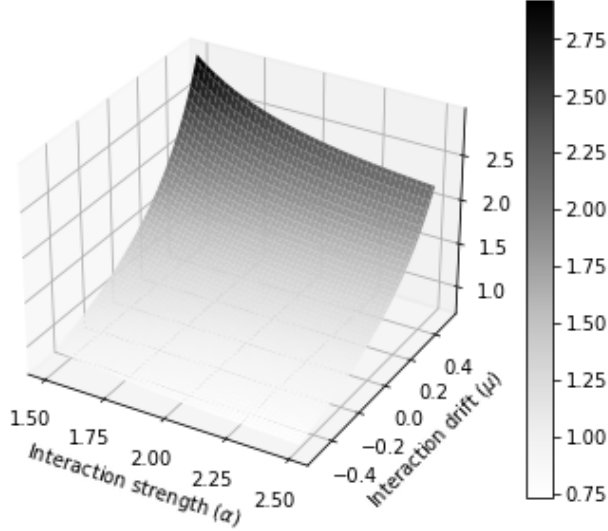


Figure 5: The 3D plot represents $\sigma^* = \sigma^*(\alpha, \mu)$, solution of the system (14)-(16). In contrast to the proportion of surviving species p^* , we observe that μ has a major influence over σ^* . The graph for the theoretical value of m^* has approximately the same behavior with respect to μ and α .

the solutions $p^*, m^*, (\sigma^*)^2$ of a system of equations. While establishing this system of equations, we will provide the following representation (see (19)) of the abundance x_k^* of a surviving species:

$$x_k^* = 1 + \mu p^* m^* + \frac{\sigma^* \sqrt{p^*}}{\alpha} Z_k,$$

where $Z_k \sim \mathcal{N}(0, 1)$ and $Z_k > -\delta^* = -\delta(p^*, m^*, \sigma^*)$, δ being defined in (17). We take here advantage of this representation to characterize x_k 's distribution, which turns out to be a truncated Gaussian.

Heuristics 2. Let $(\alpha, \mu) \in \mathcal{A}$, \mathbf{x}^* the solution of (7) and let (p^*, m^*, σ^*) the solution of the system (14)-(16). Recall the definition (17) of δ and denote by $\delta^* = \delta(m^*, p^*, \sigma^*)$. Let $x_k^* > 0$ a positive component of \mathbf{x}^* , then:

$$\mathcal{L}(x_k^*) \xrightarrow{n \rightarrow \infty} \mathcal{L} \left(1 + \mu p^* m^* + \frac{\sigma^* \sqrt{p^*}}{\alpha} Z \mid Z > -\delta^* \right),$$

where $Z \sim \mathcal{N}(0, 1)$. Otherwise stated, asymptotically x_k^* admits the following density

$$f(y) = \frac{\mathbf{1}_{\{y > 0\}}}{\Phi(\delta^*)} \frac{\alpha}{\sigma^* \sqrt{2\pi p^*}} \exp \left\{ -\frac{1}{2} \left(\frac{\alpha}{\sigma^* \sqrt{p^*}} y - \delta^* \right)^2 \right\}.$$

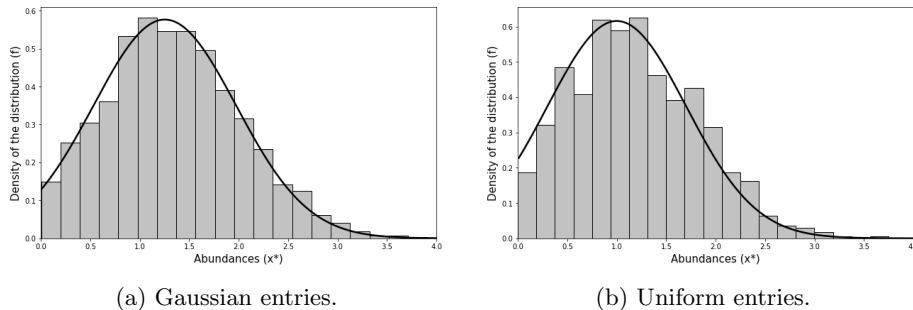


Figure 6: Distribution of surviving species. The x -axis represents the value of the abundances and the histogram is built upon the positive components of equilibrium \mathbf{x}^* . The solid line represents the theoretical distribution for parameters (α, μ) as given by Heuristics 2. In Fig. (6a), the entries are Gaussian $\mathcal{N}(0, 1)$ and the parameters are set to $(n = 2000, \alpha = 2, \mu = 0.2)$. In Fig. (6b), the entries are uniform $\mathcal{U}(-\sqrt{3}, \sqrt{3})$ with variance 1 and the parameters are set to $(n = 2000, \alpha = \sqrt{3}, \mu = 0)$. Notice in particular that the theoretical distribution matches with non-Gaussian entries.

The heuristics simply follows from the fact that if x_k^* is a surviving species then

$$x_k^* = 1 + \mu p^* m^* + \frac{\sigma^* \sqrt{p^*}}{\alpha} Z_k$$

conditionally on the fact that the right hand side of the equation is positive, that is $Z_k > -\delta^*$. A simple change of variable yields the density - details are provided in Appendix B.

Fig. 6 illustrates the matching between the theoretical distribution obtained in Heuristics 2 and a histogram obtained by Monte-Carlo simulations. It also illustrates the validity of the heuristics beyond the gaussianness assumption of the entries.

3.3 Construction of the heuristics

We first discuss Heuristics 1 and establish Equations (14), (15) and (16).

Equation (14). We first recall a result on order statistics of a Gaussian sample. Consider a family $(Z_k)_{k \in [n]}$ of i.i.d. random variables $\mathcal{N}(0, 1)$ and the associated order statistics

$$Z_1^* \leq Z_2^* \leq \dots \leq Z_n^*.$$

Consider an index $[n\alpha] \in [n]$ where $\alpha \in (0, 1)$ is fixed, then the typical location of $Z_{[n\alpha]}^*$ is $\Phi^{-1}(\alpha)$:

$$Z_{[n\alpha]}^* \simeq \Phi^{-1}(\alpha) \quad \text{as } n \rightarrow \infty, \quad (18)$$

see for instance [Smi49, BDH78].

Let \mathbf{x}^* be the equilibrium of (1) and consider the random variable

$$\check{Z}_k = \sum_{i \in \mathcal{S}} B_{ki} x_i^* = (B\mathbf{x}^*)_k.$$

We assume that asymptotically the x_i^* 's are independent from the B_{ki} 's, an assumption supported by the chaos hypothesis, see for instance Geman and Hwang [GH82]. Denote by $\mathbb{E}_{\mathbf{x}^*} = \mathbb{E}(\cdot | \mathbf{x}^*)$ the conditional expectation with respect to \mathbf{x}^* . Notice that conditionally to \mathbf{x}^* , the \check{Z}_k 's are independent Gaussian random variables, whose two first moments can easily be computed, see Appendix B, Section B.1 for the details:

$$\mathbb{E}_{\mathbf{x}^*} \check{Z}_k = \mu \hat{p} \hat{m} \quad \text{and} \quad \text{var}_{\mathbf{x}^*}(\check{Z}_k) = \frac{\hat{p} \hat{\sigma}^2}{\alpha^2}.$$

Notice that the fact that $\mathbb{E}_{\mathbf{x}^*}$ and $\text{var}_{\mathbf{x}^*}(\check{Z}_k)$ only depend on \hat{p} , $\hat{\sigma}$ and \hat{m} which are (supposedly) converging quantities supports the idea that \check{Z}_k is unconditionally a Gaussian random variable with moments:

$$\mathbb{E} \check{Z}_k = \mu p^* m^* \quad \text{and} \quad \text{var}(\check{Z}_k) = \frac{p^* (\sigma^*)^2}{\alpha^2},$$

where p^* , m^* , σ^* are resp. the limits of \hat{p} , \hat{m} , $\hat{\sigma}$. We now introduce the standard Gaussian random variables $(Z_k)_{k \in [n]}$ where

$$Z_k = \frac{\check{Z}_k - \mathbb{E} \check{Z}_k}{\sqrt{\text{var}(\check{Z}_k)}} = \alpha \frac{\check{Z}_k - \mu p^* m^*}{\sigma^* \sqrt{p^*}}.$$

Consider the equilibrium $\mathbf{x}^* = (x_k^*)_{k \in [n]}$. If $k \in \mathcal{S}$, that is $x_k^* > 0$, we have

$$1 - x_k^* + (B\mathbf{x}^*)_k = 0 \quad \Rightarrow \quad 1 + (B\mathbf{x}^*)_k > 0.$$

This identity has two implications:

$$x_k^* = 1 + (B\mathbf{x}^*)_k \quad \text{and} \quad 1 + (B\mathbf{x}^*)_k > 0 \quad \text{if } k \in \mathcal{S}.$$

Relying on the representation $(B\mathbf{x}^*)_k = \check{Z}_k$, we obtain the representation

$$x_k = 1 + (B\mathbf{x}^*)_k = 1 + \mu p^* m^* + \frac{\sigma^* \sqrt{p^*}}{\alpha} Z_k \quad \text{if } k \in \mathcal{S}. \quad (19)$$

and the condition:

$$1 + (B\mathbf{x}^*)_k = 1 + \mu p^* m^* + \frac{\sigma^* \sqrt{p^*}}{\alpha} Z_k > 0.$$

If $k \notin \mathcal{S}$ then

$$1 + (B\mathbf{x}^*)_k = 1 + \mu p^* m^* + \frac{\sigma^* \sqrt{p^*}}{\alpha} Z_k \leq 0$$

by the non invadability condition. Otherwise stated,

$$\begin{cases} Z_k \leq -\frac{\alpha(1+\mu p^* m^*)}{\sigma^* \sqrt{p^*}} & \text{if } k \notin \mathcal{S}, \\ Z_k > -\frac{\alpha(1+\mu p^* m^*)}{\sigma^* \sqrt{p^*}} & \text{if } k \in \mathcal{S}. \end{cases}$$

Considering the order statistics of the Z_k 's we obtain:

$$Z_1^* \leq \dots \leq Z_i^* \leq -\frac{\alpha(1+\mu p^* m^*)}{\sigma^* \sqrt{p^*}} \leq Z_{i+1}^* \leq \dots \leq Z_n^*.$$

Now, there are exactly $n - |\mathcal{S}| = n(1 - \hat{p})$ indices before the threshold corresponding to the components of \mathbf{x}^* equal to zero. In particular, index $i = n(1 - \hat{p})$ corresponds to the value

$$Z_i^* \simeq -\frac{\alpha(1+\mu p^* m^*)}{\sigma^* \sqrt{p^*}}$$

Relying on (18), we finally obtain

$$\Phi^{-1}(1 - \hat{p}) = -\frac{\alpha(1+\mu p^* m^*)}{\sigma^* \sqrt{p^*}}.$$

It remains to replace \hat{p} by its limit p^* to obtain (14).

Equation (15). Our starting point is the following generic representation of an abundance at equilibrium (either of a surviving or vanishing species):

$$\begin{aligned} x_k^* &= \left(1 + \mu p^* m^* + \frac{\sigma^* \sqrt{p^*}}{\alpha} Z_k\right) \mathbf{1}_{\{Z_k > -\delta^*\}} \\ &= (1 + \mu p^* m^*) \mathbf{1}_{\{Z_k > -\delta^*\}} + \left(\frac{\sigma^* \sqrt{p^*}}{\alpha} Z_k\right) \mathbf{1}_{\{Z_k > -\delta^*\}} \end{aligned}$$

Summing over \mathcal{S} and normalizing,

$$\begin{aligned} \frac{1}{|\mathcal{S}|} \sum_{k \in \mathcal{S}} x_k^* &= (1 + \mu p^* m^*) \frac{1}{|\mathcal{S}|} \sum_{k \in \mathcal{S}} \mathbf{1}_{\{Z_k > -\delta^*\}} + \frac{\sigma \sqrt{p^*}}{\alpha} \frac{1}{|\mathcal{S}|} \sum_{k \in \mathcal{S}} Z_k \mathbf{1}_{\{Z_k > -\delta^*\}}, \\ \hat{m} &\stackrel{(a)}{=} (1 + \mu p^* m^*) + \frac{\sigma \sqrt{p^*}}{\alpha} \frac{n}{|\mathcal{S}|} \frac{1}{n} \sum_{k \in [n]} Z_k \mathbf{1}_{\{Z_k > -\delta^*\}}, \\ \hat{m} &\stackrel{(b)}{\simeq} (1 + \mu p^* m^*) + \frac{\sigma \sqrt{p^*}}{\alpha} \frac{1}{\mathbb{P}(Z > -\delta^*)} \mathbb{E}(Z \mathbf{1}_{\{Z > -\delta^*\}}), \\ \hat{m} &\simeq (1 + \mu p^* m^*) + \frac{\sigma \sqrt{p^*}}{\alpha} \mathbb{E}(Z | Z > -\delta^*). \end{aligned}$$

where (a) follows from the fact that $|\mathcal{S}| = \sum_{k \in \mathcal{S}} \mathbf{1}_{\{Z_k > -\delta^*\}}$ (by definition of \mathcal{S}), (b) from the law of large numbers $\frac{1}{n} \sum_{k \in [n]} Z_k \mathbf{1}_{\{Z_k > -\delta^*\}} \xrightarrow{n \rightarrow \infty} \mathbb{E} Z \mathbf{1}_{\{Z > -\delta^*\}}$ and $\frac{|\mathcal{S}|}{n} \xrightarrow{n \rightarrow \infty} \mathbb{P}(Z > -\delta^*)$ with $Z \sim \mathcal{N}(0, 1)$. It remains to replace \hat{m} by its limit m^* to obtain (15).

Eq. (16) can be obtained similarly. Details are provided in Appendix B, see Section B.2.

4 Switching between equilibria: changing interaction strength

In the previous sections, the strength α of the interactions was fixed, cf. equation (2). However, in nature interactions between species are constantly changing due to e.g. abiotic factors such as temperature, which affect the rate at which individuals forage for prey, etc. Our contribution is rooted in the framework of asymptotic dynamics, but many recent ecological studies highlight the importance of taking into account both transient dynamics (out-of-equilibrium abundance values due to frequent perturbations) and shifts between equilibria due to changing environmental conditions [Has01, FN11, NA16]. In the sequel, we discuss a more general framework.

The model and intuition. If we restrict ourselves to a specific environment, a possible ecological interpretation of the fluctuation of interaction strength corresponds to the relationship between the size of the habitat and the probability of contact between individuals from two interacting species (e.g. think of a pool of freshwater containing piscivorous fishes and their prey species – interactions, be them competition or predation, would be potentially more frequent if the volume of water was reduced). In physics, think of particles in motion in a given volume: if the number of particles and the temperature stay constant, reducing the volume should increase the number of interactions between particles.

From a model standpoint, let μ be fixed, $\alpha = \alpha(t) : \mathbb{R}^+ \rightarrow (\sqrt{2}, \infty)$. Consider

$$\frac{dx_k}{dt} = x_k (1 - x_k + (B_t \mathbf{x})_k), \quad k \in [n], \quad (20)$$

where matrix B_t admits the following representation

$$B_t = \frac{A}{\alpha(t)\sqrt{n}} + \frac{\mu}{n} \mathbf{1}\mathbf{1}^T \quad \text{and} \quad (\alpha(t), \mu) \in \mathcal{A}.$$

Remark 5. *Following Theorem 2, condition $(\alpha(t), \mu) \in \mathcal{A}$ guarantees that there exists a unique equilibrium $\mathbf{x}^*(t)$ for every $t \in \mathbb{R}^+$.*

We focus on the case of a system that fluctuates between two equilibrium points (Figure 7a). We consider a sudden incident, most often irreversible in the short term, which reduces a portion of habitat, e.g. forest fires. The system transits from a feasible state to a state with vanishing species due to the change of the strength of interactions, modelled by the following *step function*:

$$\alpha(t) = \alpha_1 \mathbf{1}_{[0, t_0)} + \alpha_2 \mathbf{1}_{[t_0, +\infty)}, \quad (\alpha_1, \alpha_2, t_0) \in (\sqrt{2}, +\infty)^2 \times \mathbb{R}_+ \quad (21)$$

The change of model parameter occurs at t_0 which causes a change in the strength of the interactions going from a value α_1 to α_2 . One may choose $\alpha_2 < \alpha_1$ and the difference (or ratio) between the two values represents the intensity of the incident.

In large dimension, it is possible to characterize this change by its impact on the number of surviving species in the system (20). At a given time t , the proportion of surviving species $p = p(t) \in [0, 1]$ can be computed by resolving the system in Heuristics 1. This function, associated to the step function α given in (21), is represented in Figure 7b.

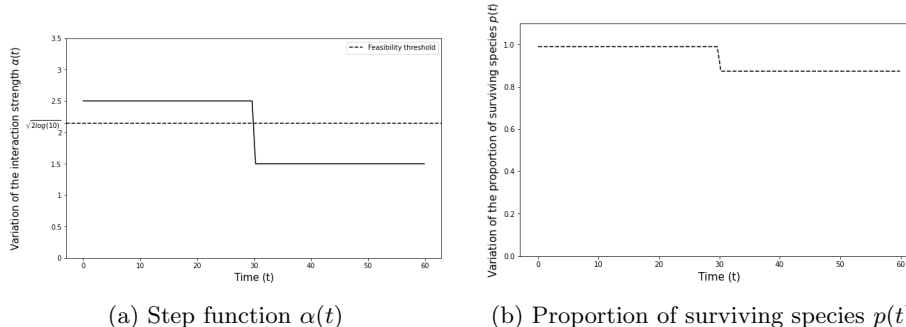


Figure 7: (a) Variation of the interaction strength through time, used in the dynamics of a ten-species system depicted in Figure 8 ($\alpha_1 = 2.5$, $\alpha_2 = 1.5$). The dotted line represents the feasibility threshold associated to the system. (b) Variation of the proportion of surviving species depending of the variation of $\alpha(t)$ in (a).

In the case of a sudden incident, the proportion of surviving species predicted by the heuristics has a form similar to $\alpha(t)$ i.e. a step response. In the feasible state, $p(t)$ is close to 1 (i.e. all species coexist); after the transition occurs, some species vanish and here $p(t) \approx 0.87$. Beware that the heuristics results follow instantaneously the change of α ; however, there is a smoother transition in the dynamics between the two equilibria (respectively corresponding to α_1 and α_2) due to the return rate to equilibrium, see for instance [NC97], [ABLH18]). This transition is illustrated in Figure 8.

Simulations. We provide hereafter a simulation-based analysis of the impact of the sudden incident on a given ecosystem: Define a ten-species system (20) with a fixed matrix of interactions A with Gaussian entries $\mathcal{N}(0, 1)$ and consider $\alpha = \alpha(t)$ as in Fig. 7a.

This scenario has a mixed impact on the community, see Fig. 8. While some species benefit from this change through an increase in their abundances, others are severely affected by this shift, some of which become extinct. This phenomenon can be understood as follows: at first ($t \leq t_0$), the system admits a feasible equilibrium state with $\alpha = \alpha_1 = 2.5 > \sqrt{2 \log(10)} \simeq 2.14$ and the abundances converge to this equilibrium (see Figure 7a). When the transition occurs at $t = t_0$, Theorem 2 ensures the convergence to a new equilibrium defined by parameter α_2 . Since $\alpha_2 = 1.5$ is below the feasibility threshold $\sqrt{2 \log(10)} \simeq 2.14$, some species vanish. In other words, this sudden change

of model parameter causes an increase of interaction strengths, which has a negative impact on species diversity.

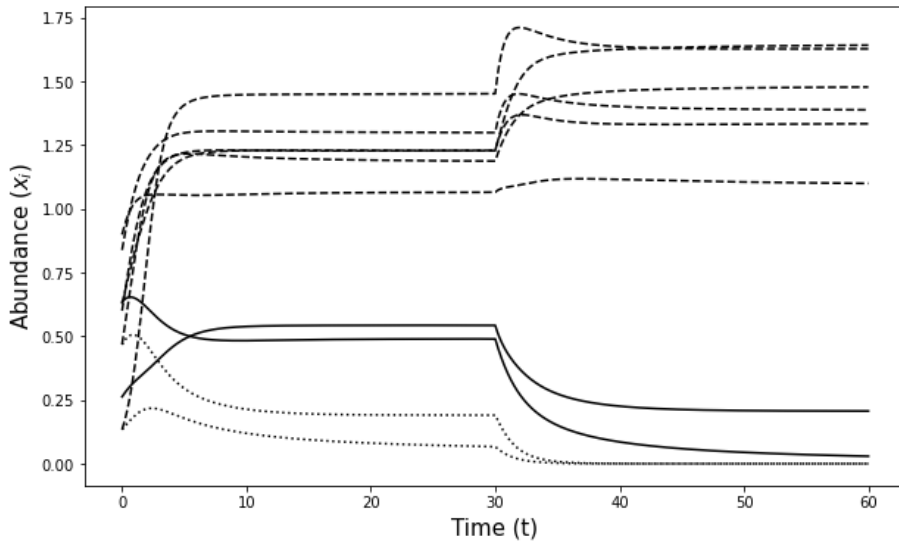


Figure 8: Abundance dynamics in the case of a community of ten species. The matrix of interactions A and the initial conditions are common and we apply the function of variation $\alpha(t)$ given in Figure 7a. The dashed lines represents species which benefit from habitat variation; solid lines represent species suffering from the change. Dotted lines represent species undergoing extinction.

Evolution of diversity Finally, we illustrate the evolution of diversity using diversity indicators more suited to the description of changes such as the one represented in Fig. 8 [Jos06]. We introduce here Shannon diversity H' , a standard measure of biodiversity in ecology, which is given by

$$H' = - \sum_i \frac{x_i}{\sum_j x_j} \log \left(\frac{x_i}{\sum_j x_j} \right) \quad (22)$$

and ranges from 0 (one species completely dominates the community) to $\log(n)$, when each species is equally abundant. When many species become rare while others become more abundant, H' drops. Because H' varies before species actually vanish, it is a more sensitive index of community diversity than species richness.

The Hill number of order 1, defined as $e^{H'}$, is a diversity measure that takes into account species abundances and varies between 1 and n , i.e. it behaves like an “effective species richness”, see e.g. [Jos07].

In Fig. 9a, we represent the mean of this diversity measure over time for a hundred-species system and observe a negative impact of the variation of the

strength of the interactions on diversity. Parameter μ has no impact on diversity at equilibrium (similarly, μ has no impact on the proportion of persistent species), but the lower the value of μ , the slower the transition to a new equilibrium. In other words, the more generally “competitive” the ecosystem is (i.e. very negative values of μ), the longer it takes for transient dynamics to settle near equilibria.

The evolution of the Hill number of order 1 complements the evolution of species richness: when α decreases, the expected number of surviving species decreases (Fig. 7); at the same time, $e^{H'}$ decreases even more drastically as the abundance distribution of surviving species gets more heterogeneous. Figure 9b also shows that the variability of the Hill number of order 1 among simulations increases drastically when α decreases. The conclusion is that the more species are lost, the more difficult it is to predict the diversity index as σ^* depends on α and strongly influences $e^{H'}$.

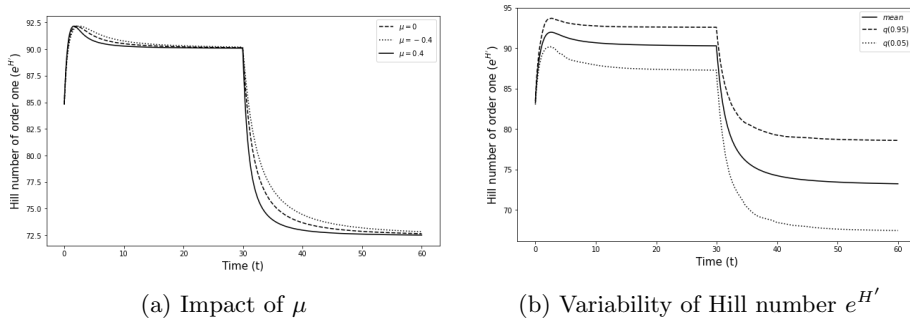


Figure 9: Dynamics of the Hill number of order 1 in the case of an ecosystem of a hundred species. The initial conditions are similar for each species. We define an interaction matrix A and let the dynamics of Lotka-Volterra evolve according to model (20) and we apply the function of variation $\alpha(t)$ of Figure 7a. For each time step, we compute $e^{H'}$. We repeat this scheme a large number of times (here 500), and we average the time series. In (a), we apply this procedure for different values of μ . In (b), we apply this procedure for a fixed $\mu = 0$ and compute the quantiles of the 500 trajectories.

Theoretical estimation of diversity Standard mathematical methods (Taylor’s theorem) can be used to obtain a theoretical approximation of the Hill number of order 1 (details are provided in Appendix B.4):

$$e^{H'} \approx np^* \left(\frac{3}{2} - \frac{1}{2} \frac{(\sigma^*)^2}{(m^*)^2} \right). \quad (23)$$

This estimator is based on the properties of the persistent species (p^*, m^*, σ^*) calculated by solving the fixed point equation of the heuristics (1). These three properties depend on the type of the interactions, as indicated by parameters

(α, μ) (Figure 4). We compare the accuracy of this estimator with two examples in which the strength of the interactions (α) and the interaction drift (μ) vary (Figure 10).

On the one hand, a shift of the interaction drift μ does not affect the proportion of surviving species. Furthermore, the impact of μ on σ^* and m^* is proportional i.e. $\frac{\sigma^*}{m^*}$ is equal to a constant. For these reasons, μ does not affect the Hill number (Figure 10b). On the other hand, if α increases, then $p^*, m^*, \sigma^* \xrightarrow{\alpha \rightarrow +\infty} 1$ and $e^{H'} \rightarrow n$ which makes sense because when α becomes very large, all abundances converge to 1. If α decreases: p^* decreases, and σ^* increases faster than m^* . This confirms that $e^{H'}$ decreases even more drastically as the abundance distribution of surviving species gets more heterogeneous (Figure 10a).

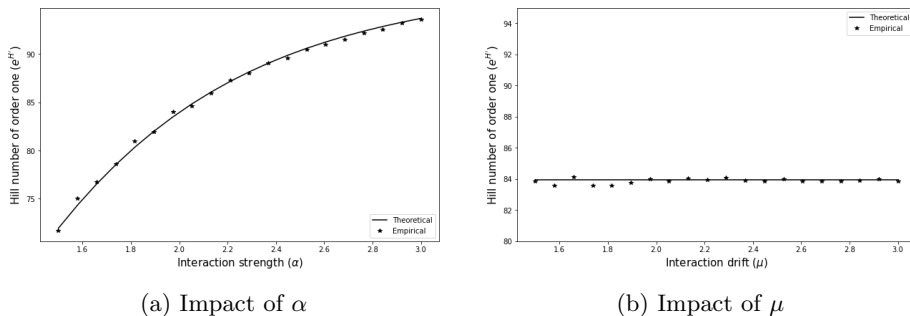


Figure 10: Evolution of the Hill number of order 1 as a function of α (a) and μ (b). The theoretical solutions (solid line) are computed by resolving (1) and integrating the parameter (p^*, σ^*, m^*) in equation (23). The empirical solutions (star marker) are computed by a Monte-Carlo experiment (100 experiments): we define a matrix B of size 100×100 , solve the LCP problem and calculate the associated Hill number $e^{H'}$ using (22).

5 Discussion

In this paper, our main interest was to describe the impact of the strength α and mean μ of interactions in large LV models on the conditions of coexistence of interacting species. Combining results from Takeuchi [Tak96] with standard RMT results, we have provided insights into the study of stability of large random ecosystems - see [Sto18, GGRA18].

We have characterized the unique equilibrium properties of the surviving species by resolving a system of equations. From a physicist point of view, similar equations were obtained by Oppen and Diederich [OD92] and studied in a more general framework by Bunin using the dynamical cavity method [Bun17] and Galla [Gal18] using generating functional techniques.

The coexistence of many species in random ecosystems was also studied

by Servan et al. [SCG⁺18] and Pettersson et al. [PSNJ20], where a more generic case was analyzed with different growth rates. Grilli et al. [GAS⁺17] identified the key quantities regulating the parameter space leading to feasible communities. In contrast to previous approaches, an important feature of our model is the monitoring of the impact of interactions by a normalization factor ($\alpha\sqrt{n}$). From an ecological point of view, one might expect that the larger the number of species, the weaker the interactions will be due to some dilution of interactions among potential interaction partners, which would justify the use of such normalization factors. From a mathematical viewpoint, the normalizing parameter α captures the range of a unique equilibrium and the threshold for feasibility.

In nature, interactions between species are constantly changing and affected by the environment. Under the assumptions that environmental conditions influence interaction strengths, we have endeavoured in Section 4 to study the consequences of a sudden change of environmental conditions, expressed through a decrease in parameter α . Solving numerically the Lotka-Volterra system confirms the predictions given by heuristics, i.e. that a decrease in α negatively affects equilibrium species richness. A more precise representation of biodiversity dynamics can be obtained through Hill numbers of order 1 which also decreases after the sudden change in α . The dynamics of this diversity measure suggests that the mean of interaction coefficients, μ , affects the duration of transient dynamics, with shorter transient dynamics being associated with more mutualistic interactions (i.e. higher positive values of μ).

Many questions naturally arise as a follow-up. First, a mathematical proof of the heuristics presented here is a challenging prospect because the LCP procedure induces a statistical dependence that is a priori difficult to handle. However, looking for this proof will certainly help extend the results to other underlying assumptions on the parameters of the LV system.

Regarding the extension of the heuristics to other assumptions, two situations could be of particular interest: non-centered elliptical matrix models as in Bunin [Bun17] and LV models in which species growth rates are also controlled as in [SCG⁺18]. We are confident that such extensions are possible, but might hinge on more sophisticated developments, in particular to include growth rates in the calculus of order statistics.

In this paper we have only considered the case of a full interaction matrix with parameters (α, μ) . However, food webs are often structured in compartments [BDB⁺11] and/or obey hierarchies (e.g. larger species eat smaller ones) [BAB⁺19]. By a numerical analysis of LV systems, one could use the same tools to study more patterned matrices [AT12]. Recent studies emphasize the sparsity of real food webs [BSHM17]. Beyond the feasibility studied by Akjouj and Najim [AN21], one could also study the existence and stability of a unique equilibrium in a sparse context.

Finally, variations of the interaction strength highlight the impact of habitat destruction. Many theoretical studies provide mathematical formulas for the return rate to equilibrium [NC97, ABLH18]. A further theoretical study of

model (20) could provide a more quantitative answer. In this article, we have limited the analysis to the case of a single sudden incident, but other types of fluctuations for the interaction strength could be considered for a better understanding of habitat conservation phenomena. For example, a seasonal model could be appropriate to describe the evolution of the dynamics over the seasons.

References

- [ABLH18] J.-F. Arnoldi, A. Bideault, M. Loreau, and B. Haegeman. How ecosystems recover from pulse perturbations: A theory of short-to long-term responses. *Journal of Theoretical Biology*, 436:79–92, January 2018.
- [AN21] I. Akjouj and J. Najim. Feasibility of sparse large lotka-volterra ecosystems. *arXiv preprint arXiv:2111.11247*, 2021.
- [AT12] S. Allesina and S. Tang. Stability criteria for complex ecosystems. *Nature*, 483(7388):205, 2012.
- [BAB⁺19] U. Brose, P. Archambault, A. Barnes, L-F. Bersier, T. Boy, J. Canning-Clode, E. Conti, M. Dias, C. Digel, A. Dissanayake, A. Flores, K. Fussmann, B. Gauzens, C. Gray, J. Häussler, M. Hirt, U. Jacob, M. Jochum, S. Kéfi, O. McLaughlin, M. MacPherson, E. Latz, K. Layer-Dobra, P. Legagneux, Y. Li, C. Madeira, N. Martinez, V. Mendonça, C. Mulder, S. Navarrete, E. O’Gorman, D. Ott, J. Paula, D. Perkins, De. Piechnik, I. Pokrovsky, D. Raffaelli, B. Rall, B. Rosenbaum, R. Ryser, A. Silva, Es. Sohlström, N. Sokolova, M. Thompson, R. Thompson, F. Vermandele, C. Vinagre, S. Wang, J. Wefer, R. Williams, E. Wieters, G. Woodward, and A. Iles. Predator traits determine food-web architecture across ecosystems. *Nature Ecology and Evolution*, 3(6):919–927, 2019.
- [BC12] C. Bordenave and D. Chafaï. Around the circular law. *Probab. Surv.*, 9:1–89, 2012.
- [BDB⁺11] E. B. Baskerville, A. P. Dobson, T. Bedford, S. Allesina, T. M. Anderson, and M. Pascual. Spatial guilds in the serengeti food web revealed by a bayesian group model. *PloS Computational Biology*, 7(12):e1002321, 2011.
- [BDH78] A. Balkema and L. De Haan. Limit distributions for order statistics. i. *Theory of Probability & Its Applications*, 23(1):77–92, 1978.
- [BN21] P. Bizeul and J. Najim. Positive solutions for large random linear systems. *arXiv:1904.04559*, to be published in *Proceedings of the A.M.S.*, 2021.
- [BS10] Z. D. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. Springer, New York, second edition, 2010.
- [BSHM17] D. M. Busiello, S. Suweis, J. Hidalgo, and A. Maritan. Explorability and the origin of network sparsity in living systems. *Scientific Reports*, 7(1):12323, December 2017.

- [Bun17] G. Bunin. Ecological communities with lotka-volterra dynamics. *Physical Review E*, 95(4):042414, 2017.
- [CDMF09] M. Capitaine, C. Donati-Martin, and D. Féral. The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *The Annals of Probability*, 37(1), January 2009.
- [Cle22] M. Clenet. Equilibrium and surviving species in a large lotka-volterra system of differential equations. <https://github.com/maxime-clenet/>, 2022.
- [CPS09] R. W. Cottle, J-S. Pang, and R.E. Stone. *The linear complementarity problem*. SIAM, 2009.
- [DVR⁺18] M. Dougoud, L. Vinckenbosch, R. P Rohr, L-F. Bersier, and C. Mazza. The feasibility of equilibria in large ecosystems: A primary but neglected concept in the complexity-stability debate. *PLoS computational biology*, 14(2):e1005988, 2018.
- [FN11] T. Fukami and M. Nakajima. Community assembly: alternative stable states or alternative transient states? *Ecology Letters*, 14(10):973–984, 2011.
- [Gal18] T. Galla. Dynamically evolved community size and stability of random Lotka-Volterra ecosystems. *EPL (Europhysics Letters)*, 123(4):48004, September 2018. arXiv: 1808.06660.
- [GAS⁺17] J. Grilli, M. Adorisio, S. Suweis, G. Barabás, J. R. Banavar, S. Allesina, and A. Maritan. Feasibility and coexistence of large ecological communities. *Nature Communications*, 8(1):14389, April 2017.
- [GGRA18] T. Gibbs, J. Grilli, T. Rogers, and S. Allesina. Effect of population abundances on the stability of large random ecosystems. *Physical Review E*, 98(2):022410, 2018.
- [GH82] S. Geman and C.-R. Hwang. A chaos hypothesis for some large systems of random equations. *Z. Wahrsch. Verw. Gebiete*, 60(3):291–314, 1982.
- [Has01] A. Hastings. Transient dynamics and persistence of ecological systems. *Ecology Letters*, 4(3):215–220, 2001.
- [HS98] J. Hofbauer and K. Sigmund. *Evolutionary games and population dynamics*. Cambridge University Press, 1998.
- [Jos06] L. Jost. Entropy and diversity. *Oikos*, 113(2):363–375, 2006.
- [Jos07] L. Jost. Partitioning diversity into independent alpha and beta components. *Ecology*, 88(10):2427–2439, 2007.

- [Lam19] A. Lamperski. Lemke’s algorithm for linear complementarity problems. <https://github.com/AndyLamperski/lemkelcp>, 2019.
- [LM96] R. Law and R. D. Morton. Permanence and the Assembly of Ecological Communities. *Ecology*, 77(3):762–775, April 1996.
- [Lot25] A. J. Lotka. *Elements of physical biology*. Williams & Wilkins, 1925.
- [May72] R.M. May. Will a large complex system be stable? *Nature*, 238(5364):413, 1972.
- [Mur72] K. G. Murty. On the number of solutions to the complementarity problem and spanning properties of complementary cones. *Linear Algebra and its Applications*, 5(1):65–108, 1972.
- [MY88] K. G. Murty and F-T. Yu. *Linear complementarity, linear and non-linear programming*, volume 3. Citeseer, 1988.
- [NA16] B. C. Nolting and K. C. Abbott. Balls, cups, and quasi-potentials: quantifying stability in stochastic systems. *Ecology*, 97(4):850–864, 2016.
- [NC97] M. G. Neubert and H. Caswell. Alternatives to resilience for measuring the responses of ecological systems to perturbations. *Ecology*, 78(3):653–665, April 1997.
- [OD92] M. Opper and S. Diederich. Phase transition and 1/f noise in a game dynamical model. *Physical Review Letters*, 69(10):1616–1619, September 1992. Publisher: American Physical Society.
- [PSNJ20] S. Pettersson, V. M. Savage, and M. Nilsson Jacobi. Predicting collapse of complex ecological systems: quantifying the stability–complexity continuum. *Journal of The Royal Society Interface*, 17(166):20190391, May 2020. Publisher: Royal Society.
- [SCG⁺18] C. A. Serván, J. A. Capitán, J. Grilli, K. E. Morrison, and S. Allesina. Coexistence of many species in random ecosystems. *Nature Ecology & Evolution*, 2(8):1237–1242, August 2018.
- [Smi49] N. V. Smirnov. Limit distributions for the terms of a variational series. *Trudy Matematicheskogo Instituta imeni VA Steklova*, 25:3–60, 1949.
- [Sto18] L. Stone. The feasibility and stability of large complex biological networks: a random matrix approach. *Scientific reports*, 8(1):8246, 2018.
- [TA80] Y. Takeuchi and N. Adachi. The existence of globally stable equilibria of ecosystems of the generalized volterra type. *Journal of Mathematical Biology*, 10(4):401–415, 1980.

- [Tak96] Y. Takeuchi. *Global dynamical properties of Lotka-Volterra systems*. World Scientific, 1996.
- [Tao13] T. Tao. Outliers in the spectrum of iid matrices with bounded rank perturbations. *Probability Theory and Related Fields*, 155(1):231–263, February 2013.
- [Vol26] V. Volterra. Fluctuations in the abundance of a species considered mathematically. *Nature*, 118(2972):558–560, 1926.

A Simulation details

Simulations were performed in Python. All the figures and the code are available on Github [Cle22].

Simulations on the properties of surviving species are performed in two different ways. The theoretical solutions are obtained resolving numerically the system of equations of heuristics 1. We use a solver (cf. `scipy.optimize`) to find a local minimum of the function defined by the system of equations (a modification of the Powell hybrid method). The empirical solutions are computed using a Monte Carlo experiment. We simulate a large number of matrix matrix B , we resolve the associated LCP problem using the Lemke’s algorithm. Then, we use the LCP solution to calculate the properties of the surviving species: proportion of survivors, etc. Finally, we make an average on the ensemble of experiments. The Lemke algorithm is implemented in the `lemkelcp` package and can be found on Github [Lam19]. The dynamics of the Lotka-Volterra are achieved by a Runge-Kutta of order 4 (RK4) implemented in the code.

B Remaining computations

B.1 Details on Equation (14): Moments of \check{Z}_k .

We compute hereafter the conditional mean and variance of $\check{Z}_k = (B\mathbf{x}^*)_k$ with respect to \mathbf{x}^* . We rely on the following identities:

$$\mathbb{E}B_{ki} = \frac{\mu}{n}, \quad \mathbb{E}(B_{ki})^2 = \frac{1}{\alpha^2 n} + \frac{\mu^2}{n^2} \simeq \frac{1}{\alpha^2 n}, \quad \mathbb{E}B_{ki}B_{kj} = \frac{\mu^2}{n^2} \quad (i \neq j).$$

We first compute the conditional mean:

$$\mathbb{E}_{\mathbf{x}^*}(\check{Z}_k) = \sum_{i \in [n]} \mathbb{E}(B_{ki})x_i^* = \sum_{i \in \mathcal{S}} \mathbb{E}(B_{ki})x_i^* = \frac{\mu}{n} \sum_{i \in \mathcal{S}} x_i^* = \mu \frac{|\mathcal{S}|}{n} \frac{1}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} x_i^* = \mu \hat{p} \hat{m}.$$

We now compute the second moment:

$$\begin{aligned}
\mathbb{E}_{\mathbf{x}^*}(\check{Z}_k^2) &= \mathbb{E}_{\mathbf{x}^*} \left(\sum_{i \in \mathcal{S}} B_{ki} x_i^* \right)^2 = \mathbb{E}_{\mathbf{x}^*} \sum_{i,j \in \mathcal{S}} B_{ki} B_{kj} x_i^* x_j^*, \\
&= \sum_{i \in \mathcal{S}} \mathbb{E}(B_{ki}^2) x_i^{*2} + \sum_{i \neq j} \mathbb{E}(B_{ki} B_{kj}) x_i^* x_j^*, \\
&= \frac{1}{\alpha^2 n} \sum_{i \in \mathcal{S}} x_i^{*2} + \frac{\mu^2}{n^2} \sum_{i \neq j} x_i^* x_j^*, \\
&\stackrel{(a)}{\simeq} \frac{\hat{p} \hat{\sigma}^2}{\alpha^2} + \frac{\mu^2 \hat{p}^2}{|\mathcal{S}|^2} \sum_{i,j \in \mathcal{S}} x_i^* x_j^* = \frac{\hat{p} \hat{\sigma}^2}{\alpha^2} + \frac{\mu^2 \hat{p}^2}{|\mathcal{S}|^2} \left(\sum_{i \in \mathcal{S}} x_i^* \right)^2 = \frac{\hat{p} \hat{\sigma}^2}{\alpha^2} + \mu^2 \hat{p}^2 \hat{m}^2,
\end{aligned}$$

where the approximation in (a) follows from the fact that

$$\frac{1}{|\mathcal{S}|^2} \sum_{i,j \in \mathcal{S}} x_i^* x_j^* = \frac{1}{|\mathcal{S}|^2} \sum_{i \neq j} x_i^* x_j^* + \mathcal{O} \left(\frac{1}{|\mathcal{S}|} \right).$$

We can now compute the variance:

$$\text{var}_{\mathbf{x}^*}(\check{Z}_k) = \mathbb{E}_{\mathbf{x}^*}(\check{Z}_k^2) - (\mathbb{E}_{\mathbf{x}^*} \check{Z}_k)^2 = \frac{\hat{p} \hat{\sigma}^2}{\alpha^2}.$$

B.2 Details on Equation (16).

As for the proof of (15), we start from the generic representation of x_k^* :

$$\begin{aligned}
x_k^* &= \left(1 + \mu p^* m^* + \frac{\sigma^* \sqrt{p^*}}{\alpha} Z_k \right) \mathbf{1}_{\{Z_k > -\delta^*\}} \\
&= (1 + \mu p^* m^*) \mathbf{1}_{\{Z_k > -\delta^*\}} + \frac{\sigma \sqrt{p^*}}{\alpha} Z_k \mathbf{1}_{\{Z_k > -\delta^*\}}.
\end{aligned}$$

Taking the square, we get

$$\begin{aligned}
x_k^{*2} &= (1 + \mu p^* m^*)^2 \mathbf{1}_{\{Z_k > -\delta\}} \\
&\quad + 2(1 + \mu p^* m^*) \frac{\sigma^* \sqrt{p^*}}{\alpha} Z_k \mathbf{1}_{\{Z_k > -\delta\}} + \frac{(\sigma^*)^2 p^*}{\alpha^2} Z_k^2 \mathbf{1}_{\{Z_k > -\delta^*\}}.
\end{aligned}$$

Summing over \mathcal{S} and normalizing, we get

$$\begin{aligned}
\frac{1}{|\mathcal{S}|} \sum_{k \in \mathcal{S}} (x_k^*)^2 &= (1 + \mu p^* m^*)^2 \frac{1}{|\mathcal{S}|} \sum_{k \in \mathcal{S}} \mathbf{1}_{\{Z_k > -\delta^*\}} \\
&\quad + 2(1 + \mu p^* m^*) \frac{\sigma^* \sqrt{p^*}}{\alpha} \frac{1}{|\mathcal{S}|} \sum_{k \in \mathcal{S}} Z_k \mathbf{1}_{\{Z_k > -\delta^*\}} \\
&\quad \quad \quad + \frac{(\sigma^*)^2 p^*}{\alpha^2} \frac{1}{|\mathcal{S}|} \sum_{k \in \mathcal{S}} Z_k^2 \mathbf{1}_{\{Z_k > -\delta^*\}}.
\end{aligned}$$

Finally, we conclude by replacing the empirical means by their limits

$$\frac{1}{|\mathcal{S}|} \sum_{k \in \mathcal{S}} Z_k^i \mathbf{1}_{\{Z_k > -\delta^*\}} = \mathbb{E}(Z^i | Z > -\delta^*), \quad i = 1, 2.$$

and get

$$\begin{aligned} \hat{\sigma}^2 &= (1 + \mu p^* m^*)^2 + 2(1 + \mu p^* m^*) \frac{\sigma^* \sqrt{p^*}}{\alpha} \mathbb{E}(Z | Z > -\delta^*) \\ &\quad + \frac{(\sigma^*)^2 p^*}{\alpha^2} \mathbb{E}(Z^2 | Z > -\delta^*). \end{aligned}$$

It remains to replace $\hat{\sigma}$ by its limit σ^* to obtain (16).

B.3 Density of the distribution of the persistent species.

Assume that $x^* > 0$, and let $f = \mathbb{R} \rightarrow \mathbb{R}$ be a bounded continuous test function. We have

$$\begin{aligned} \mathbb{E}f(x_k^*) &= \mathbf{E} \left[f \left(1 + \frac{\sigma^* \sqrt{p^*}}{\alpha} Z_k + \mu p^* m^* \right) \mid Z_k > -\delta^* \right], \\ &= \int_{-\infty}^{\infty} f \left(1 + \mu p^* m^* + \frac{\sigma^* \sqrt{p^*}}{\alpha} u \right) \frac{\mathbf{1}_{\{u > -\delta^*\}}}{1 - \Phi(-\delta^*)} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2p^*}} du, \\ &= \int_0^{\infty} f(y) e^{-\frac{1}{2} \left(\frac{\alpha}{\sigma^* \sqrt{p^*}} y - \delta^* \right)^2} \frac{\alpha}{\sqrt{2\Phi(\delta^*)} p^* \sigma^*} dy, \end{aligned}$$

hence the density of x_k^* .

B.4 Theoretical estimation of the diversity index

Recall that $|\mathcal{S}| = n\hat{p}$ is the number of surviving species and that

$$p_i = \frac{x_i}{\sum_{j \in \mathcal{S}} x_j}$$

is the frequency of (surviving) species i .

To find a theoretical estimate of Hill number of order 1, we proceed by expansion and set

$$p_i = \frac{1}{|\mathcal{S}|} + \delta_i, \quad |\delta_i| \ll \frac{1}{|\mathcal{S}|}$$

where δ_i represents the deviation of species i from the standard frequency if all surviving species have the same frequency. Notice that $\sum_{i \in \mathcal{S}} \delta_i = 0$.

$$H' = - \sum_{i \in \mathcal{S}} p_i \log(p_i) = - \sum_{i \in \mathcal{S}} \left(\frac{1}{|\mathcal{S}|} + \delta_i \right) \log \left(\frac{1}{|\mathcal{S}|} + \delta_i \right).$$

We use the Taylor-Young formula of order 2 to decompose the log:

$$\begin{aligned}\log\left(\frac{1}{|\mathcal{S}|} + \delta_i\right) &= \log\left(\frac{1}{|\mathcal{S}|}\right) + |\mathcal{S}|\delta_i - \frac{\delta_i^2|\mathcal{S}|^2}{2} + \delta_i^3\varepsilon(\delta_i), \\ &\approx \log\left(\frac{1}{|\mathcal{S}|}\right) + |\mathcal{S}|\delta_i - \frac{\delta_i^2|\mathcal{S}|^2}{2}.\end{aligned}$$

$$\begin{aligned}H' &\approx -\sum_{i \in \mathcal{S}} \left(\frac{1}{|\mathcal{S}|} + \delta_i\right) \left(\log\left(\frac{1}{|\mathcal{S}|}\right) + |\mathcal{S}|\delta_i - \frac{\delta_i^2|\mathcal{S}|^2}{2}\right), \\ &= -\sum_{i \in \mathcal{S}} \left[\frac{1}{|\mathcal{S}|} \log\left(\frac{1}{|\mathcal{S}|}\right) + \delta_i - \frac{\delta_i^2|\mathcal{S}|}{2} + \delta_i \log\left(\frac{1}{|\mathcal{S}|}\right) + |\mathcal{S}|\delta_i^2 - \frac{\delta_i^3|\mathcal{S}|^2}{2}\right], \\ &= \log(|\mathcal{S}|) - \sum_{i \in \mathcal{S}} \frac{\delta_i^2|\mathcal{S}|}{2} + \sum_{i \in \mathcal{S}} \frac{\delta_i^3|\mathcal{S}|^2}{2}.\end{aligned}$$

Notice that $\sum_{i=1}^{|\mathcal{S}|} \frac{\delta_i^3|\mathcal{S}|^2}{2}$ is negligible since $|\delta_i| \ll |\mathcal{S}|^{-1}$. The term 1 corresponds to the maximum value that the Shannon diversity index can take if $|\mathcal{S}|$ are present in the system. It remains to develop the second term of the r.h.s.

$$\begin{aligned}-\frac{1}{2} \sum_{i \in \mathcal{S}} \delta_i^2|\mathcal{S}| &= -\frac{|\mathcal{S}|}{2} \sum_{i \in \mathcal{S}} \left(\frac{x_i}{\sum_{j \in \mathcal{S}} x_j} - \frac{1}{|\mathcal{S}|}\right)^2, \\ &= -\frac{|\mathcal{S}|}{2} \sum_{i \in \mathcal{S}} \left(\frac{x_i^2}{(\sum_{j \in \mathcal{S}} x_j)^2} - \frac{2}{|\mathcal{S}|} \frac{x_i}{\sum_{j \in \mathcal{S}} x_j} + \frac{1}{|\mathcal{S}|^2}\right), \\ &= -\frac{|\mathcal{S}|}{2} \sum_{i \in \mathcal{S}} \left(\frac{x_i^2}{(\sum_{j \in \mathcal{S}} x_j)^2}\right) + \frac{1}{2}, \\ &= -\frac{|\mathcal{S}|}{2} \frac{\sum_{i \in \mathcal{S}} x_i^2}{|\mathcal{S}|^2 \left(\frac{1}{|\mathcal{S}|} \sum_{j \in \mathcal{S}} x_j\right)^2} + \frac{1}{2}, \\ &= -\frac{1}{2} \frac{\frac{1}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} x_i^2}{\left(\frac{1}{|\mathcal{S}|} \sum_{j \in \mathcal{S}} x_j\right)^2} + \frac{1}{2}, \\ &= -\frac{1}{2} \frac{\hat{\sigma}^2}{(\hat{m})^2} + \frac{1}{2}, \\ &= -\frac{1}{2} \left(\frac{\hat{\sigma}^2}{\hat{m}^2} - 1\right).\end{aligned}$$

Finally the Hill number of order 1 can be computed as:

$$\begin{aligned}e^{H'} &\approx e^{\log(|\mathcal{S}|) - \frac{|\mathcal{S}|}{2} \sum_{i=1}^{|\mathcal{S}|} \delta_i^2}, \\ &\approx |\mathcal{S}| \left(1 - \frac{|\mathcal{S}|}{2} \sum_{i=1}^{|\mathcal{S}|} \delta_i^2\right) = |\mathcal{S}| \left(1 - \frac{1}{2} \frac{\hat{\sigma}^2}{(\hat{m})^2} + \frac{1}{2}\right) = \frac{|\mathcal{S}|}{2} \left(3 - \frac{\hat{\sigma}^2}{(\hat{m})^2}\right).\end{aligned}$$

Replacing $|\mathcal{S}|$ by np^* and $\hat{\sigma}$ and \hat{m} by their limits, we get the desired result:

$$e^{H'} \approx \frac{np^*}{2} \left(3 - \frac{(\sigma^*)^2}{(m^*)^2} \right) .$$