

# Out-of-equilibrium inference of feeding rates through population data from generic consumer-resource stochastic dynamics

José A. Capitán

*Complex Systems Group, Department of Applied Mathematics, Universidad Politécnica de Madrid, Spain*

David Alonso

*Computational and Theoretical Ecology. CEAB-CSIC, Blanes, Catalonia, Spain*

---

## Abstract

Statistical models are often structurally unidentifiable, because different sets of parameters can lead to equal model outcomes. To be useful for prediction and parameter inference from data, stochastic population models need to be identifiable, this meaning that model parameters can be uniquely inferred from a large number of model observations. In particular, precise estimation of feeding rates in consumer-resource dynamics is crucial, because consumer-resource processes are central in determining biomass transport across ecosystems. Model parameters are usually estimated at stationarity, because in that case model analyses are often easier. In this contribution we analyze the problem of parameter redundancy in a multi-resource consumer-resource model, showing that model indentifiability depends on whether the dynamics have reached stationarity or not. To be precise, we: (i) Calculate the steady-state and out-of-equilibrium probability distributions of predator's abundances analytically using generating functions, which allow us to unveil parameter redundancy and carry out proper maximum likelihood estimation. (ii) Conduct *in silico* experiments by tracking the abundance of consumers that are either searching for or handling prey, data then used for maximum likelihood parameter estimation. (iii) Show that, when model observations are recorded out of equilibrium, feeding parameters are truly identifiable, whereas if sampling is done solely at stationarity, only ratios of rates can be inferred from data (i.e., parameters are redundant). We discuss the implications of our results when inferring parameters of general dynamical models.

**Keywords:** Stochastic consumer-resource models, Master equation, Model identifiability, Parameter inference, Generating function, Multi-resource Holling type II and III feeding dynamics

---

## 1. Introduction

The consumer-resource interaction is a key driver of the transport of biomass across ecological communities (Belgrano, 2005; Dunne, 2006; Getz, 2011; Perkins et al., 2022; Valdovinos et al., 2022). In classical approaches, biomass flow is quantified by inter-dependent changes in species abundances, which are usually measured by the density of individuals of each species (or their biomass) in a given region (Bersier et al., 2002). This quantification has relied mainly on differential equations describing the spatial and temporal variability of species abundances. Consumer-resource interactions are included in these equations via mathematical functions called *functional responses*, which encapsulate and simplify the complexities of feeding interactions, providing a coarse-grained representation of the underlying mechanisms (Lafferty et al., 2015; Abrams, 2022). Functional responses are defined as the per capita, average feeding rate of consumers on their resources (Berryman, 1992). Such functions have been derived from conceptual considerations about feeding dynamics (ranging from simple predator-prey equations to more complex food-web

---

*Email addresses:* ja.capitan@upm.es (corresponding author) (José A. Capitán), dalonso@ceab.csic.es (David Alonso)

models (McPeck, 2019; Abrams, 2022)), but also from feeding experiments (Rosenbaum and Rall, 2018; Barbier et al., 2021; Abrams, 2022), where a consumer is observed feeding on a resource population. Holling’s seminal papers (Holling, 1959a,b) show how the behavior of individual consumers gives rise to an average feeding rate over the whole consumer population, which then can be used in population-level models. As shown in Palamara et al. (2021, 2022), Holling’s type II functional response can be rigorously derived from processes at the individual level when mass action (individuals meet at random), chemostatic conditions (the total number of consumers and the total number of resources are kept constant) and stationarity are assumed. Functional responses have been generalized to multi-species systems (Koen-Alonso, 2007; Robertson et al., 2022) and to include a variety of foraging mechanisms.

Elemental processes at the lower level of aggregation (i.e., individual level) in consumer-resource models are driven by feeding parameters. In the simplest scenario one can imagine, the feeding process involves two predator behavioral stages: the search for and the handling of preys. The search or hunting behavior of the available prey ends with a predator attack, moment at which the predator starts handling the prey and stops searching for other preys. This predator-prey complex is usually irreversible (unless the prey escapes) and relaxes into a free predator again ready to search for more prey. This relaxation process takes some (handling) time, during which the individual predator kills, handles, and diggest the prey. Both attack and relaxation processes occur at certain rates that are not known a priori. These rates could be inferred from population abundances of the different types involved in the feeding process, given some hypotheses about how these abundances vary according to the two elemental processes. In particular, the distributions of the population numbers of each behavioral type (free and handling consumer numbers, in the example above) can be used to infer feeding parameters, provided that the statistical model is fitted to empirical abundance data. In this contribution, we focus on the problem of parameter inference using stochastic population models and *in silico* (simulated) abundance data.

Stochastic models are to be designed so that model parameters can be *uniquely* determined from a large number of model realizations. In that case, the model is known as *structurally identifiable*, because model parameters can be inferred without redundancy from data, regardless the quantity or quality of the data (Bellman and Åström, 1970). In particular, dynamic consumer-resource models should allow per-capita feeding rates (associated to elemental feeding processes) to be estimated unambiguously from experimental population abundance data. The problem of structural identifiability is general and of multidisciplinary character. For example, it has been thoroughly analyzed for deterministic dynamics in the context of biological modeling approaches (see Villaverde et al. (2016) and references therein), but also in social sciences (Helbing et al., 2015; Jusup et al., 2022). In this contribution, we focus on the problem of structural identifiability in stochastic dynamical models. Our aim is to unveil scenarios for feeding rate redundancy using a multi-resource, Holling type II feeding dynamics (Holling, 1959a,b). For that purpose, we consider a single predator population, in which individuals are able to search for and handle  $S$  different resource species, leading to  $S + 1$  different behavioral types, namely:  $S$  types associated to predators handling each resource species, and free predators that are not handling any resource. We assume mass action and chemostatic conditions to derive the joint probability distribution of the different behavioral types, at two regimes: in the transient dynamics (out of equilibrium) and in the long run (once the steady-state is reached). We show analytically that, at equilibrium, the model is not identifiable because only the ratio between relaxation and attack rates can be extracted from fitting model to data. In this sense, the model at equilibrium is *set identifiable*, this meaning that parameters can be precisely determined only on a subset of the parameter space. We also show that true feeding rates cannot be properly estimated via a standard fitting procedure such as maximum likelihood (ML) parameter estimation using synthetic (model-simulated) data at stationarity. This *in silico* experiment, which mimics an empirical setup to estimate feeding parameters that assumes the system has reached stationarity, only allows for precise estimation of quotients of rates at equilibrium.

However, if model observations are taken out of equilibrium at different sampling times, feeding parameters can be uniquely estimated via ML. In this case, the likelihood function associated to independent observations of the system’s state exhibits a maximum at precise parameter values, making the model com-

pletely identifiable. This allows for proper parameter inference from data, providing a practical solution for parameter redundancy at equilibrium. In this line, to avoid potential ambiguities in model calibration from time series data, we stress that parameter inference should be done out of equilibrium. We also show that, for the multi-resource feeding dynamics studied, parameter redundancy arises if sampling is done for independent dynamical observations taken at the exact same time (either close to equilibrium or not). In actual experiments, this is not a shortcoming for inference.

In order to carry out the analysis, we solve the stochastic model at equilibrium by computing the joint probability distribution of behavioral types, which turns out to be a multinomial distribution. In addition, we solve the master equation to obtain the temporal dependence of the joint probability distribution in the case of equal relaxation rates, which leads also to a multinomial with probability parameters depending on time. Thanks to these calculations, we can conduct parameter inference, both at equilibrium and out-of-equilibrium, using synthetic datasets based on model observations. The paper is organized as follows. Section 2 summarizes the background on parameter redundancy and shows that the explicit computation of the generating function can be used to assess parameter redundancy. The analytical solution of the stochastic model is explained in detail in Section 3. Section 4 is devoted to analyze several *in silico* experiments designed to illustrate parameter identification failure, showing that feeding parameters can be only fully identifiable if the model is sampled out of equilibrium at different sampling times. At equilibrium, the model allows for the precise determination of dimensionless parameter ratios. Finally, we conclude discussing the implications and potential extensions of our work.

## 2. Stochastic model identifiability via generating functions

Precise inference of model parameters cannot be done for every statistical model: a model is defined as *identifiable* if it is theoretically possible to infer true parameter values after retrieving an infinite number of observations from the model. Mathematically, this condition amounts to saying that different parameter sets lead to strictly different distributions of observable quantities pertaining to the sample space. More precisely, let  $\mathcal{G} = (\mathcal{S}, \mathcal{P})$  be a statistical model with sample space  $\mathcal{S}$ , onto which a vector of observable variables  $\mathbf{x}$  takes values, and a set  $\mathcal{P}$  of probability distributions  $P_{\boldsymbol{\theta}}(\mathbf{x})$  whose parameter vector  $\boldsymbol{\theta}$  is a member of the parameter space  $\Theta$ :

$$\mathcal{P} = \{P_{\boldsymbol{\theta}} | \boldsymbol{\theta} \in \Theta\}.$$

Model  $\mathcal{G}$  is *identifiable* if the mapping  $\boldsymbol{\theta} \mapsto P_{\boldsymbol{\theta}}$ , which assigns a probability distribution to a set of parameters, is *one-to-one* (Lehmann and Casella, 2006). As said, this means that distinct parameter sets should correspond to distinct probability distributions, i.e.:  $P_{\boldsymbol{\theta}_1}(\mathbf{x}) = P_{\boldsymbol{\theta}_2}(\mathbf{x}) \Leftrightarrow \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$  almost surely for all  $\mathbf{x}$ .

For an identifiable model, the map  $\boldsymbol{\theta} \mapsto P_{\boldsymbol{\theta}}$  can be inverted to learn the true parameter set if the model can be observed an infinite number of times. This is true because of the strong law of large numbers, by which probabilities can be estimated by averaging frequencies over a large number of observations (Loève, 1977). Thus, with an infinite number of observations, finding the true probability distribution  $P_{\boldsymbol{\theta}_0}$  is feasible. Because the map  $\boldsymbol{\theta} \mapsto P_{\boldsymbol{\theta}}$  is invertible, the true parameter set  $\boldsymbol{\theta}_0$  which generated the distribution  $P_{\boldsymbol{\theta}_0}$  can be unambiguously extracted from the distribution.

The easiest way to check for parameter redundancy is by finding an exhaustive summary (Cole, 2020), which is a vector  $\boldsymbol{\kappa}(\boldsymbol{\theta})$  of model parameter combinations that uniquely determine the statistical model  $\mathcal{G}$ , so that  $\boldsymbol{\kappa}(\boldsymbol{\theta}_1) = \boldsymbol{\kappa}(\boldsymbol{\theta}_2) \Leftrightarrow P_{\boldsymbol{\theta}_1}(\mathbf{x}) = P_{\boldsymbol{\theta}_2}(\mathbf{x})$  for all  $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$ . In general, any model will have many suitable exhaustive summaries. Knowledge of values of the exhaustive summary fully determines model outcomes. If the parameter vector  $\boldsymbol{\theta}$  can be determined uniquely from an exhaustive summary  $\boldsymbol{\kappa}(\boldsymbol{\theta})$ , then the model lacks of parameter redundancy. See Castro and De Boer (2020) for more advanced ways of testing structural identifiability.

Stochastic model identifiability can be assessed if one can compute the distribution  $P_{\mathbf{n}}(t)$  of observing state  $\mathbf{n}$  at time  $t$ , which will depend on model parameters. For that purpose, the master (forward Kolmogorov) equation has to be solved (Gillespie, 1991), and from  $P_{\mathbf{n}}(t)$  the exhaustive summary could

be eventually extracted. Equivalently, one can obtain the multivariate  $z$ -transform of  $P_{\mathbf{n}}(t)$ , known as generating function, defined as

$$G(\mathbf{z}, t) = \sum_{n_1, \dots, n_S \geq 0} P_{\mathbf{n}}(t) \prod_{i=1}^S z_i^{n_i}, \quad (1)$$

where the vector  $\mathbf{z} = (z_1, \dots, z_S)$  is conjugate to the state vector  $\mathbf{n} = (n_1, \dots, n_S)$ . This vector  $\mathbf{z}$  can take values over  $\mathbb{R}^S$ . A series expansion will eventually lead to the probabilities  $P_{\mathbf{n}}(t)$ . Having  $G(\mathbf{z}, t)$  is equivalent to knowing  $P_{\mathbf{n}}(t)$ , because the generating function allows to compute every moment of the probability distribution, thus uniquely determining it. Therefore, a set of parameter combinations that uniquely determine the generating function can be used as an exhaustive summary to assess parameter redundancy.

We illustrate this idea using a well-known birth-death-immigration stochastic dynamics for a single population (Kendall, 1949). The elemental processes are defined by the following reaction scheme:



The first process stands for the arrival of a new individual  $X$  by immigration at rate  $\mu$ , irrespectively of the number of individuals present in the system. The second one refers to reproduction at a per-capita rate  $\beta$ , this meaning that if there are  $n$  individuals, they will reproduce with probability  $\beta n$  per unit time. Finally, individual death occurs at a per-capita rate  $\delta$ . Since birth and death rates are linear on population numbers, the master equation can be solved analytically in terms of the generating function

$$G(z, t) = \sum_{n \geq 0} P_n(t) z^n,$$

which results in (Kendall, 1949)

$$G(z, t) = \frac{\left(\frac{\beta}{\delta} - 1\right)^{\frac{\mu}{\beta}} \left[ e^{(\beta - \delta)t} (1 - z) + \frac{\beta}{\delta} z - 1 \right]^{n_0}}{\left[ \frac{\beta}{\delta} e^{(\beta - \delta)t} (1 - z) + \frac{\beta}{\delta} z - 1 \right]^{n_0 + \frac{\mu}{\beta}}}$$

for an initial condition assuming  $n(0) = n_0$  individuals at time  $t = 0$ . We focus on the problem of estimating elemental process rates ( $\beta$ ,  $\delta$ , and  $\mu$ ) and assume that the initial condition  $n_0$  is known in advance, so we do not need to infer it from data. The generating function —and, henceforth, the probability  $P_n(t)$ — is uniquely determined by specifying the parameter combinations  $\mu/\beta$ ,  $\beta - \delta$ , and  $\beta/\delta$ . Thus, an exhaustive summary is  $\boldsymbol{\kappa}(\boldsymbol{\theta}) = (\mu/\beta, \beta - \delta, \beta/\delta)$ . If vector  $\boldsymbol{\kappa}(\boldsymbol{\theta})$  is known, then the parameter vector  $\boldsymbol{\theta} = (\beta, \delta, \mu)$  can be uniquely determined and the model is identifiable. Therefore, the whole parameter set can be inferred from model observations via maximum likelihood parameter estimation (MLE) techniques, for example.

On the other hand, let  $\beta < \delta$  and assume that model observations are conducted at stationarity ( $t \rightarrow \infty$ ). The generating function reduces to

$$G(z, \infty) = \left(1 - \frac{\beta}{\delta}\right)^{\frac{\mu}{\beta}} \left(1 - \frac{\beta}{\delta} z\right)^{-\frac{\mu}{\beta}}. \quad (5)$$

Besides the trivial issue that  $n_0$  cannot be inferred from steady-state observations in models with a single attractor like this one, there is a major issue in determining elemental process rates. According to (5), an exhaustive summary that fully determines steady-state model realizations is  $\boldsymbol{\kappa}(\boldsymbol{\theta}) = (\mu/\beta, \beta/\delta)$  for

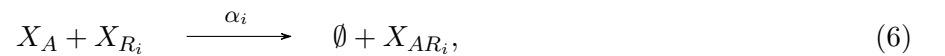
$\theta = (\beta, \delta, \mu)$ . Therefore, only the ratios  $\mu/\beta$  and  $\beta/\delta$  are truly identifiable, and the model remains parameter redundant (set identifiable) in this limit.

This simple example illustrates a problem that we expect to occur frequently, in stochastic population models, when it comes to estimating individual process rates. First, the master equation cannot be solved for general, multi-species birth-death processes when transition probability rates depend non-linearly on population numbers. Second, among the solvable models, it is much more frequent to find steady-state solutions of the master equation (thanks to detailed-balance conditions, for example) rather than the general temporal dependence of  $P_{\mathbf{n}}(t)$ . Therefore, having computed the steady-state solution  $P_{\mathbf{n}}$  analytically, one could naively formulate a MLE method to estimate model rates by sampling the model at stationarity. As the example shows, this can lead to infinite possible estimates. Although model samples are generated at precise values of elemental process rates, they cannot be inferred from stationary distributions. The issue precisely stems by sampling the model after the steady state has been reached.

We conjecture this issue to be general (at least for models defined by elemental process rates with a polynomial dependence on population abundances), because at equilibrium there is no well-defined temporal scale, and the generating function could depend only on ratios of rates. Therefore, parameters should be estimated out of equilibrium. This has important implications for population models that are not analytically tractable, for which numerical log-likelihood profiles or simulation are to be applied. Moreover, if the MLE procedure is based on abundance data, one should be aware of whether data were collected at stationarity or not to be truly confident of ML estimates. In the rest of the paper, we demonstrate this pathology of MLE at equilibrium using a solvable multi-species consumer-resource model and show that the issue is resolved when data for MLE is collected out of equilibrium. In the discussion section, we highlight that natural extensions of this model would also be parameter redundant at equilibrium.

### 3. Multi-resource Holling Type II feeding dynamics

We start by defining a model for consumer-resource dynamics. Assume that an individual consumer  $X_A$  can feed on a range of  $S$  alternative resource species,  $X_{R_i}$ , with  $i = 1, \dots, S$ . We consider two processes: first, the attack of resource  $X_{R_i}$  by the consumer  $X_A$ , which results in the formation of a compound  $X_{AR_i}$  formed by the consumer handling the resource. Second, handling consumers do not search for prey until the compound  $X_{AR_i}$  “relaxes” because the predator stops handling the prey, then becoming a consumer searching for prey. The stochastic dynamics can be summarized by  $2S$  individual feeding reactions:



The first channel represents feeding interactions, by taking into account different resource preferences determined by distinct attack rates ( $\alpha_i$ ) for the resources. The second reaction models consumer-resource compound relaxation, which occurs at rates  $\nu_i$  (depending on resource type). Mean handling times, given by  $\nu_i^{-1}$ , are also characteristic for individual consumers feeding on each resource type. Our aim in this section is to solve analytically the stochastic model, by obtaining the joint probability distribution of the number of searching predators, as well as the number of consumers handling on each resource, both out-of-equilibrium and at stationarity.

The total consumer population,  $n_A^0$ , is partitioned into  $S + 1$  classes,

$$n_A^0 = n_A + \sum_{i=1}^S n_i, \quad i = 1, \dots, S, \quad (8)$$

where  $n_A$  refers to the number of (free) consumers searching for resources, and  $n_i$  is the number of predators handling on resource  $i$ . In this section, we first write the reaction scheme (6)–(7) as transition probability

rates between discrete states, then solve the master equation at stationarity, and finally obtain the temporal dependence of the probability distribution in the particular case in which these two conditions hold: (i) relaxation rates are equal,  $\nu_1 = \dots = \nu_S$ , and (ii) initially there are only  $n_A = n_A^0$  free consumers. Either for finite times or as  $t \rightarrow \infty$ , we obtain multinomial probability distributions.

The master equation determines the temporal evolution of the probability of finding a given discrete configuration of consumer types,  $(n_A, n_1, \dots, n_S)$ , at a given time  $t$ . As in Palamara et al. (2021), we assume chemostatic conditions, so the overall number of consumers,  $n_A^0$ , is kept constant. Therefore, the number of predators  $n_A$  is immediately determined by the number of consumers handling on each resource,  $\{n_i\}$ , via Eq. (8). Accordingly, we define the state vector of the stochastic process as  $\mathbf{n} = (n_1, \dots, n_S)$ , considering only variations in handling predator numbers. Given a configuration vector  $\mathbf{n}$ , the number of free predators is determined as  $n_A = n_A^0 - \sum_{i=1}^S n_i$ .

Let  $\mathbf{e}_i = (0, \dots, 1, \dots, 0)$ ,  $i = 1, \dots, S$ , be the vectors of the canonical basis of  $\mathbb{R}^S$ . Transition probability rates are defined as

$$T(\mathbf{n} + \mathbf{e}_i | \mathbf{n}) = \alpha_i \frac{n_A n_{R_i}}{N}, \quad T(\mathbf{n} - \mathbf{e}_i | \mathbf{n}) = \nu_i n_i,$$

for  $i = 1, \dots, S$ , each one corresponding to chemical reactions (6) and (7), respectively. They account for the conditional probability for transition  $\mathbf{n} \rightarrow \mathbf{n} + \mathbf{e}_i$ , i.e, the event of a consumer that starts handling on resource  $i$ , as well as the conditional probability for transition  $\mathbf{n} \rightarrow \mathbf{n} - \mathbf{e}_i$ , which represents the relaxation of the compound formed by the consumer and the  $i$ -th resource. For handling predators formation, we assume also chemostatic conditions for resources levels, meaning that resource densities  $x_{R_i} := n_{R_i}/N$  are kept constant —here  $N$  stands for the number of sites of the system (system volume). In addition, a law of mass action is assumed because the probability of consumer-resource encounter has to be proportional to resource availability. The effective per capita attack rate of a free consumer at a given density level of the  $i$ -th resource type is a resource-specific constant parameter defined by  $\theta_i := \alpha_i x_{R_i}$ . The second transition probability represents  $AR_i$  compounds relaxation, leading to free predators after a mean handling time  $\nu_i^{-1}$ . This probability is proportional to the number of handling predators  $n_i$  if one of the  $AR_i$  compounds stops handling the resource.

Using Eq. (8), transition probability rates become

$$T(\mathbf{n} + \mathbf{e}_i | \mathbf{n}) = \theta_i \left( n_A^0 - \sum_{j=1}^S n_j \right), \quad T(\mathbf{n} - \mathbf{e}_i | \mathbf{n}) = \nu_i n_i, \quad (9)$$

where  $n_A^0$  and  $\nu_i, \theta_i$  for  $i = 1, \dots, S$ , are constant parameters. Let  $P_{\mathbf{n}}(t)$  be the probability of observing a state vector  $\mathbf{n}$  at time  $t$ . The master equation expresses the balance of probabilities as an infinite system of coupled differential equations,

$$\frac{dP_{\mathbf{n}}}{dt} = \sum_{i=1}^S \{ T(\mathbf{n} | \mathbf{n} + \mathbf{e}_i) P_{\mathbf{n} + \mathbf{e}_i} + T(\mathbf{n} | \mathbf{n} - \mathbf{e}_i) P_{\mathbf{n} - \mathbf{e}_i} - [T(\mathbf{n} + \mathbf{e}_i | \mathbf{n}) + T(\mathbf{n} - \mathbf{e}_i | \mathbf{n})] P_{\mathbf{n}} \}. \quad (10)$$

The probability balance can be interpreted as follows. The probability of visiting state  $\mathbf{n}$  at time  $t$  increases at rate  $T(\mathbf{n} | \mathbf{n} - \mathbf{e}_i) = \theta_i \left( n_A^0 - (n_i - 1) - \sum_{j \neq i}^S n_j \right)$  for the transition  $\mathbf{n} - \mathbf{e}_i \rightarrow \mathbf{n}$ , where a free consumer starts handling a resource of type  $i$ . The probability  $P_{\mathbf{n}}(t)$  also increases through the transition  $\mathbf{n} + \mathbf{e}_i \rightarrow \mathbf{n}$  at rate  $T(\mathbf{n} | \mathbf{n} + \mathbf{e}_i) = \nu_i (n_i + 1)$ , when a handling consumer leaves the resource and starts searching for other resources. The negative terms in the right-hand side of (10) simply balance the transitions from  $\mathbf{n}$  to any other state, which force the probability  $P_{\mathbf{n}}(t)$  to decrease.

Our aim is to accurately estimate feeding rates  $\{\alpha_1, \dots, \alpha_S, \nu_1, \dots, \nu_S\}$ . We first compute the steady-state distribution for arbitrary parameter values, and show that the distribution of feeding types is a multinomial  $\mathcal{M}(n_A^0, \{p_A, p_1, \dots, p_S\})$  with index  $n_A^0$  and probability parameters  $\{p_A, p_1, \dots, p_S\}$ . Labeling

$p_0 := p_A$ , the multinomial density function is defined as

$$P_{\mathbf{n}} = \begin{cases} \frac{n_A^0!}{\prod_{k=0}^S n_k!} \prod_{k=0}^S p_k^{n_k}, & \text{when } n_A + \sum_{k=1}^S n_k = n_A^0, \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

Here, vector  $\mathbf{n} = (n_A, n_1, \dots, n_S)$  gathers the number of searching predators,  $n_A$ , together with the numbers  $n_i$  of predators handling on each resource ( $i = 1, \dots, S$ ). Probability parameters are

$$p_i = \frac{\theta_i}{\nu_i} \frac{1}{1 + \sum_{k=1}^S \frac{\theta_k}{\nu_k}} \quad (12)$$

for  $i = 1, \dots, S$ , and

$$p_A = p_0 = \frac{1}{1 + \sum_{k=1}^S \frac{\theta_k}{\nu_k}} = 1 - \sum_{k=1}^S p_k \quad (13)$$

is the probability associated to the number of free predators,  $n_A = n_A^0 - \sum_{i=1}^S n_i$ . To carry out the derivation, we use a detailed-balance condition satisfied by transition probabilities. The details can be found in Appendix A.

Second, using the generating function formalism, we find an analytical solution for out-of-equilibrium probabilities  $P_{\mathbf{n}}(t)$  for the particular case of equal relaxation rates,  $\nu_1 = \dots = \nu_S$ , and for an initial state formed only by searching predators. In the second case, we also show that the distribution is multinomial. From the form of the distributions (and the generating function), it is apparent that the model is perfectly identifiable out of equilibrium.

In the general case of different relaxation rates, we cannot analytically solve the master equation. However, it is a linear system of differential equations, so it could be easily integrated numerically. Besides numerical integration, there are other exact methods—based on matrix exponentiation (Keeling and Ross, 2009) or Laplace transforms (Etienne and Alonso, 2007)—or approximate—dimensionality reduction (Keeling and Ross, 2009)—that allow the computation of probabilities over time. Although the parameter estimation analysis presented below would be numerically feasible in the case of distinct relaxation rates, we do not discuss it in this contribution.

### 3.1. Out-of-equilibrium probability distribution

Calculating the out-of-equilibrium probability distribution involves integrating the master equation (10) as a function of time. Usually this equation has to be integrated numerically. In this case, however, an analytical solution is feasible because transition probability rates depend linearly on population numbers (Kendall, 1949; Nisbet and Gurney, 2004). We solve the master equation by introducing the generating function defined in Eq. (1). Transforming  $P_{\mathbf{n}}(t)$  into the generating function  $G(\mathbf{z}, t)$  converts the infinite system of coupled differential equations into a partial differential equation.

We start by obtaining the PDE satisfied by the generating function  $G(\mathbf{z}, t)$  defined in Eq. (1). By multiplying the master equation, Eq. (10), by the product  $z_1^{n_1} \dots z_S^{n_S}$ , and summing over the population variables  $n_i$  ( $i = 1, \dots, S$ ) over non-negative integers, we obtain the following PDE for the generating function  $G$ ,

$$\frac{\partial G}{\partial t} = \sum_{i=1}^S (1 - z_i) \left( \nu_i \frac{\partial G}{\partial z_i} + \theta_i \sum_{j=1}^S z_j \frac{\partial G}{\partial z_j} - \theta_i n_A^0 G \right).$$

Equivalently, it can be rewritten as

$$\frac{\partial G}{\partial t} = -n_A^0 m(\mathbf{z}) G + \sum_{i=1}^S [(1 - z_i) \nu_i + z_i m(\mathbf{z})] \frac{\partial G}{\partial z_i}, \quad (14)$$

where we have introduced the notation

$$m(\mathbf{z}) := \sum_{j=1}^S \theta_j (1 - z_j). \quad (15)$$

We leave the details of the derivation of Eq. (14) from (10) to Appendix B.

We assume that the initial condition for the probability distribution is the following, without loss of generality:

$$P_{\mathbf{n}}(0) = \prod_{i=1}^S \delta_{n_i, \hat{n}_i},$$

i.e., initially the state of the system is defined by the population vector  $\hat{\mathbf{n}} = (\hat{n}_1, \dots, \hat{n}_S)$ , where  $\hat{n}_i$  stands for the initial number of predators handling resource  $i$ , and there are  $\hat{n}_A = n_A^0 - \sum_{j=1}^S \hat{n}_j$  free predators—we assume that  $n_A^0 - \sum_{j=1}^S \hat{n}_j \geq 0$  holds. In the initial condition above,  $\delta_{n,m}$  stands for the Kronecker symbol, i.e.,  $\delta_{n,m} = 1$  if  $n = m$  and 0 otherwise. In terms of  $G(\mathbf{z}, t)$ , the initial condition reduces to

$$G(\mathbf{z}, 0) = \prod_{i=1}^S z_i^{\hat{n}_i}. \quad (16)$$

It is important to remark that Eq. (14) immediately implies that the probability distribution  $P_{\mathbf{n}}(t)$  is normalized for all  $t$ : indeed, if  $z_i = 1$  for all  $i$ , then it follows from (14) that  $\frac{\partial}{\partial t} G(\mathbf{1}, t) = 0$ , i.e.,  $G(\mathbf{1}, t)$  must be a constant—here we use the notation  $\mathbf{1} = (1, \dots, 1)$  for a vector of ones. Thus, by definition of  $G(\mathbf{z}, t)$ ,

$$G(\mathbf{1}, t) = \sum_{n_1, \dots, n_S \geq 0} P_{\mathbf{n}}(t) = 1,$$

which follows from the normalization of the probability distribution. Therefore, the condition  $G(\mathbf{1}, t) = 1$  is satisfied by the PDE and automatically implies the normalization of the distribution.

Now we solve the first-order PDE (14) using the method of characteristics curves. An equivalent way of writing the initial PDE in a very compact form is as the following dot product,  $\mathbf{v} \cdot \nabla H = 0$ , where the solution  $g = G(\mathbf{z}, t)$  is viewed in the  $(S + 2)$ -dimensional space  $(\mathbf{z}, t, g)$  as the manifold  $H(\mathbf{z}, t, g) = G(\mathbf{z}, t) - g = 0$ , and the vector field  $\mathbf{v}$  is defined as

$$\mathbf{v} = (\mathbf{w}, -1, n_A^0 m(\mathbf{z}) G(\mathbf{z}, t)).$$

Here the row vector  $\mathbf{w} = (w_i)$  has  $S$  components defined as  $w_i = \nu_i(1 - z_i) + m(\mathbf{z})z_i$ , for  $i = 1, \dots, S$ . Finding the integral surface  $H$  amounts to solving the PDE.

Since the gradient of  $H$  is normal to the integral surface at each and every point,  $(\mathbf{z}, t, g)$ , then any curve on the surface should have a tangential direction, defined by the vector  $\mathbf{v}$ , that should be perpendicular to  $\nabla H$  at every point. Now, we write an arbitrary curve over the manifold  $H = 0$  in parametric form as  $(\mathbf{z}(s), t(s), g(s))$ , and the tangent vector  $(\mathbf{z}'(s), t'(s), g'(s))$  has to be parallel to vector  $\mathbf{v}$ : this immediately yields the following non-linear system of  $S + 1$  coupled ODEs,

$$\begin{aligned} \frac{dz_i}{ds} &= -\nu_i(1 - z_i) - m(\mathbf{z})z_i, \quad \text{for } i = 1, \dots, S, \\ \frac{dG}{ds} &= -n_A^0 m(\mathbf{z})G. \end{aligned} \quad (17)$$

where we have chosen  $s = -t$  and used that  $\frac{dg}{ds} = \frac{dG}{dt}$  because  $g$  equals  $G(\mathbf{z}, t)$  over any characteristic curve within the manifold  $H = 0$ .

This system could be easily solved if we knew the implicit time dependence of function  $m(t) := m(\mathbf{z}(t))$ , see Eq. (15), because  $z_i$  are indeed functions of time. Thus, the ODE system can be expressed by adding



an extra equation for  $m(t)$  as an additional variable, which is obtained by multiplying the  $i$ -th equation in (17) by  $\theta_i$  and summing over all the equations for  $z_i$ . This process yields

$$\frac{dm}{dt} = -m^2 + Qm + \sum_{j=1}^S \nu_j \theta_j (1 - z_j), \quad (18)$$

for  $Q := \sum_{j=1}^S \theta_j$ . The coupled ODE system to solve for the time dependence of the  $S + 2$  unknowns  $(\mathbf{z}, m, G)$  is

$$\begin{aligned} \frac{dz_i}{dt} &= -\nu_i(1 - z_i) - mz_i, \quad \text{for } i = 1, \dots, S, \\ \frac{dm}{dt} &= -m^2 + Qm + \sum_{j=1}^S \nu_j \theta_j (1 - z_j), \\ \frac{dG}{dt} &= -n_A^0 m G. \end{aligned} \quad (19)$$

In the particular case in which all the relaxation rates are equal,  $\nu_1 = \dots = \nu_S =: \nu$ , the system decouples because the second equation is closed for  $m$ , according to Eq. (15), so we can find the time dependence of function  $m(t) := m(\mathbf{z}(t))$ . The condition of equal relaxation rates amounts to assuming that the predator spends the same average time  $\nu^{-1}$  handling any type of resource. Therefore, the equation for  $m(t)$  reduces to a logistic growth equation if the assumption of equal handling times (equal relaxation rates) holds:

$$\frac{dm}{dt} = -m^2 + (\nu + Q)m. \quad (20)$$

From now on in this section we will assume that relaxation rates  $\nu_j$  are equal, in order to find explicit, closed-form solutions for the PDE satisfied by the generating function.

Integrating Eq. (20) yields

$$m(t) = \frac{\nu + Q}{1 + Ke^{-(\nu+Q)t}}, \quad (21)$$

for  $K$  an arbitrary constant to be calculated as function of the initial conditions,  $\hat{z}_j := z_j(0)$ . Then solving for  $K$  in Eq. (21) at  $t = 0$  gives

$$K = \frac{\nu + R}{Q - R} = \frac{\nu + \sum_{j=1}^S \theta_j \hat{z}_j}{\sum_{j=1}^S \theta_j (1 - \hat{z}_j)}, \quad (22)$$

where we have defined  $R := \sum_{j=1}^S \theta_j \hat{z}_j$ . Observe that the initial condition vector  $\hat{\mathbf{z}}$  depends on the vector of  $z$ -variables,  $\mathbf{z}$ , via the explicit solution of the differential equations for  $z_i$  in Eq. (17). Because the solution of the system of ODEs satisfied by the characteristic curves will depend on the vector  $\hat{\mathbf{z}}$  of initial conditions, we have to substitute  $\hat{z}_j$  as function of  $\mathbf{z}$  at the end of the calculation in order to get the desired expression for the generating function  $G(\mathbf{z}, t)$ .

Therefore, in the particular case of equal handling times, the system of ODEs (19) turns out to be uncoupled. Upon substitution of Eq. (21) into (19), it is immediate to solve the equation for  $G$ , which gives

$$G(\hat{\mathbf{z}}, t) = G_0 (K + e^{(\nu+Q)t})^{-n_A^0},$$

where  $K$  and  $G_0$  depend on the initial conditions  $\hat{\mathbf{z}}$ . Imposing the initial condition  $G(\hat{\mathbf{z}}, 0)$  given by Eq. (16), we deduce that  $G_0 = (1 + K)^{n_A^0} \prod_{j=1}^S \hat{z}_j^{\hat{n}_j}$ , and

$$G(\hat{\mathbf{z}}, t) = \left( \frac{1 + K}{K + e^{(\nu+Q)t}} \right)^{n_A^0} \prod_{j=1}^S \hat{z}_j^{\hat{n}_j}. \quad (23)$$

Recall that the desired dependence on  $\mathbf{z}$  is obtained once we make the substitution  $\hat{z}_j = \hat{z}_j(\mathbf{z}, t)$ . So it only remains to find the dependence of the initial conditions  $\hat{z}_j$  in terms of  $\mathbf{z}$ .

From Eq. (21) we observe that the ODEs for  $z_i(t)$  are equivalent, so we have to solve just one of the equations, for example, the one for  $z_j$ :

$$\frac{dz_j}{dt} = -(1 - z_j)\nu - \frac{(\nu + Q)z_j}{1 + Ke^{-(\nu+Q)t}},$$

subject to the initial condition  $z_j(0) = \hat{z}_j$ . Then we immediately find that

$$z_j(t) = \frac{QK - \nu e^{(\nu+Q)t} + e^{\nu t}[\nu + Q\hat{z}_j - QK(1 - \hat{z}_j)]}{Q(K + e^{(\nu+Q)t})}, \quad (24)$$

Multiplying this expression by  $\theta_j$  and summing over  $j = 1, \dots, S$ , we get

$$Q - m(\mathbf{z}) = \frac{KQ - \nu e^{(\nu+Q)t} + e^{\nu t}[(1 + K)R + \nu - KQ]}{K + e^{(\nu+Q)t}},$$

because  $\sum_{j=1}^S \theta_j z_j = Q - m(\mathbf{z})$  by (15). Given that  $\hat{z}_j$  are functions of  $\mathbf{z}$ ,  $K$  becomes itself a function of  $\mathbf{z}$ , which can be determined as follows. From Eq. (22) we can solve for  $R = \sum_{j=1}^S \theta_j \hat{z}_j$  in terms of  $K$  as  $R = (KQ - \nu)/(1 + K)$  which, upon substitution into the last expression, yields

$$Q - m(\mathbf{z}) = \frac{KQ - \nu e^{(\nu+Q)t}}{K + e^{(\nu+Q)t}}.$$

In this equality, any dependence on the initial condition  $\hat{\mathbf{z}}$  has been removed. We can solve for  $K$  in terms of  $m$ , giving

$$K = K(\mathbf{z}, t) = \left( \frac{\nu + Q}{m(\mathbf{z})} - 1 \right) e^{(\nu+Q)t}. \quad (25)$$

Finally, an explicit expression of initial conditions  $\hat{z}_j$  as function of  $\mathbf{z}$  is determined by substituting  $K(\mathbf{z}, t)$ , given by Eq. (25), into (24), which results in an equation from which we can eliminate  $\hat{z}_j$  in terms of  $\mathbf{z}$ . This manipulation gives

$$\hat{z}_j = 1 - \left( 1 + \frac{\nu}{Q} \right) \frac{m(\mathbf{z}) + [(1 - z_j)Q - m(\mathbf{z})]e^{Qt}}{m(\mathbf{z}) + [\nu + Q - m(\mathbf{z})]e^{(\nu+Q)t}}. \quad (26)$$

Substituting Eqs. (25) and (26) into Eq. (23) yields the final expression for the generating function, expressed in closed form as function of  $\mathbf{z}$ :

$$G(\mathbf{z}, t) = \left[ 1 + \frac{m(\mathbf{z})(e^{-(\nu+Q)t} - 1)}{\nu + Q} \right]^{n_A^0} \prod_{j=1}^S \left[ 1 - \left( 1 + \frac{\nu}{Q} \right) \frac{m(\mathbf{z}) + [Q(1 - z_j) - m(\mathbf{z})]e^{Qt}}{m(\mathbf{z}) + [\nu + Q - m(\mathbf{z})]e^{(\nu+Q)t}} \right]^{\hat{n}_j}. \quad (27)$$

Recall the definitions  $m(\mathbf{z}) := \sum_{j=1}^S \theta_j(1 - z_j)$  and  $Q := \sum_{j=1}^S \theta_j$ , and that the condition  $n_A^0 - \sum_{j=0}^S \hat{n}_j \geq 0$  must hold for the last expression to make sense. This function satisfies the initial condition (16) as well as the normalization condition  $G(\mathbf{1}, t) = 1$ . The generating function has to be expanded in power series of  $z_i$  in order to obtain the explicit expression of the probability distribution  $P_{\mathbf{n}}(t)$ . Such task is hard in general. As shown by the derivation itself, it is very difficult to find the out-of-equilibrium distribution from the master equation, in contrast with the solution at stationarity: for the former, we cannot find it unless assuming equal handling rates, and we cannot carry out the series expansion to obtain the probability distribution  $P_{\mathbf{n}}(t)$  for arbitrary initial conditions.

However, the series expansion can be carried out for the particular case  $\hat{n}_1 = \dots = \hat{n}_S = 0$ , which means that, initially, the population is formed solely by  $n_A = n_A^0$  free predators and no individual consumer is initially handling any prey. In this case, Eq. (27) reduces to

$$G(\mathbf{z}, t) = \left( p_0(t) + \sum_{j=1}^S p_j(t)z_j \right)^{n_A^0}, \quad (28)$$

with

$$p_0(t) = \frac{1 + e^{-(\nu + \sum_{k=1}^S \theta_k)t} \sum_{i=1}^S \frac{\theta_i}{\nu}}{1 + \sum_{k=1}^S \frac{\theta_k}{\nu}}, \quad p_j(t) = \frac{\theta_j}{\nu} \frac{1 - e^{-(\nu + \sum_{k=1}^S \theta_k)t}}{1 + \sum_{k=1}^S \frac{\theta_k}{\nu}}, \quad (29)$$

for  $j = 1, \dots, S$ . It is easy to check that  $0 \leq p_j(t) \leq 1$  for  $j = 0, 1, \dots, S$  and arbitrary time  $t$ . In addition,  $\sum_{j=0}^S p_j(t) = 1$  holds for any time. Therefore, the parameter combinations  $p_j(t)$  can be interpreted as probabilities. As we immediately show below, they are precisely the probabilities associated to the  $j$ -th population in a *multinomial* distribution.

First, using the binomial theorem, we can expand  $G(\mathbf{z}, t)$  as

$$G(\mathbf{z}, t) = \left( p_0(t) + \sum_{j=1}^S p_j(t) z_j \right)^{n_A^0} = \sum_{n_A=0}^{n_A^0} \binom{n_A^0}{n_A} p_0^{n_A} \left( \sum_{j=1}^S p_j z_j \right)^{n_A^0 - n_A}.$$

Then, by the multinomial theorem (Knuth, 1997), it holds

$$\left( \sum_{j=1}^S p_j z_j \right)^{n_A^0 - n_A} = \sum_{\substack{n_1, \dots, n_S \geq 0 \\ \sum_i n_i = n_A^0 - n_A}} \frac{(n_A^0 - n_A)!}{n_1! \dots n_S!} \prod_{k=1}^S (p_k z_k)^{n_k},$$

which substituted into the identity above yields

$$P_{\mathbf{n}}(t) = \sum_{\substack{n_A=0 \\ n_A=n_A^0 - \sum_i n_i}}^{n_A^0} \frac{n_A^0!}{n_A! n_1! \dots n_S!} p_0^{n_A} p_1^{n_1} \dots p_S^{n_S}$$

after identifying the sought probability  $P_{\mathbf{n}}(t)$  as the coefficient of the series expansion associated with the power  $z_1^{n_1} \dots z_S^{n_S}$ . Given the restriction  $n_A = n_A^0 - \sum_i n_i$  that affects the sum above, it is only one term that contributes to the sum, i.e., the one that verifies the restriction. Hence we obtain the definition (11) of the multinomial distribution. In addition, from the expression above it is apparent that if  $\sum_{j=1}^S n_j > n_A^0$ , then  $P_{\mathbf{n}}(t) = 0$ . So we recover the multinomial distribution  $\mathcal{M}(n_A^0, \{p_j(t)\})$  with  $n_A^0$  trials and probabilities  $p_j(t)$  given by Eq. (29).

This result agrees with those obtained at equilibrium, in the limit  $t \rightarrow \infty$ . As explained in Appendix A, we also found a multinomial distribution in the stationary regime. However, the out-of-equilibrium distribution for  $\nu_1 = \nu_2 = \dots = \nu_S$  is not multinomial in general: note that the generating function for non-zero initial populations, Eq. (27), does not match the one associated to a multinomial distribution, which is given by Eq. (28), unless there are no handling predators initially.  $P_{\mathbf{n}}(t)$  follows a multinomial distribution only if  $\hat{\mathbf{n}} = \mathbf{0}$ . Irrespective of the set of initial conditions, however, the generating function (27) has the correct limit as  $t \rightarrow \infty$ . Indeed, if we take this limit on both sides of Eq. (27) we get

$$\lim_{t \rightarrow \infty} G(\mathbf{z}, t) = \left( 1 - \frac{m(\mathbf{z})}{\nu + Q} \right)^{n_A^0} = \left( p_A + \sum_{j=1}^S p_j z_j \right)^{n_A^0},$$

where  $p_A = \nu/(\nu + Q)$  and  $p_j = \theta_j/(\nu + Q)$ , which are precisely the probabilities given by Eqs. (12) and (13) in the particular case of equal relaxation rates. Therefore, in this limit we obtain the generating function (28) of the multinomial distribution  $\mathcal{M}(n_A^0, \{p_A, p_1, \dots, p_S\})$  with  $n_A^0$  trials. In sum, what we have shown here is that, in the case of equal handling times, and starting by an initial population formed solely by  $n_A^0$  searching consumers, the joint probability distribution of all behavioral types is multinomial in the transient as well as at stationarity. Consistently, the probabilities  $p_j(t)$  derived in the out-of-equilibrium regime, given by Eq. (29), satisfy  $p_j(t) \rightarrow p_j$  ( $j = 1, \dots, S$ ) and  $p_0(t) \rightarrow p_A$  in the limit  $t \rightarrow \infty$ , for  $p_A$  and  $p_j$  given by Eqs. (12) and (13), respectively.

Out of equilibrium, feeding rates are perfectly identifiable, as can be seen from the generating function by computing an exhaustive summary. We assume that resource densities are known in advance and not subject to MLE, as well as the total number of consumers  $n_A^0$ . The generating function (28) is completely determined once the  $S + 1$  parameter combinations gathered in vector  $\kappa(\theta) = \left(\nu + \sum_{k=1}^S \theta_k, \frac{\theta_1}{\nu}, \dots, \frac{\theta_S}{\nu}\right)$  are known, as can be seen immediately from Eq. (29). Therefore, the exhaustive summary determines uniquely the  $S$  attack rates as  $\alpha_i = \theta_i/x_{R_i}$  ( $i = 1, \dots, S$ ) and the reciprocal mean handling time,  $\nu$ . Thus, the model is structurally identifiable out of equilibrium.

#### 4. Effect of steady-state and out-of-equilibrium sampling

The two solvable models discussed above (Sections 2 and 3) illustrate the issue of parameter redundancy when steady-state probability distributions are used: in that regime, any likelihood function based on steady-state distributions will allow only for precise estimation of ratios of rates, not the rates themselves. Therefore, we have to use time-dependent probability distributions to ensure structural identifiability. However, as opposed to structural issues in model design, limitations in the dataset often preclude the whole parameter set to be inferred from data —this is known as *practical identifiability* (Wieland et al., 2021). In this section we focus on the effect of the quality of the data: is the problem of identification failure fully solved if we retain the temporal dependence in the probability  $P_{\mathbf{n}}(t)$ , regardless of the times at which model samples are taken?

To estimate parameters through maximum likelihood, we think of a procedure based on tracking population abundances in *in silico* (simulated) model samples (we name this methodology as “feeding experiment”). We want to estimate unique values for feeding rates based only on feeding experiments, in which one may have access to the population abundances of each behavioral type (in our case, the number of free consumers,  $n_A$ , and the number of consumers handling resource  $i$ ,  $n_i$ , for  $i = 1, \dots, S$ ). We distinguish between two situations: (i) abundance data is available for different independent replicas of the system at different times at the steady state, and (ii) abundances for different behavioral types are available for out-of-equilibrium independent replicas of the system, taken at different finite times along the transient regime before reaching stationarity. We use the distributions computed in Section 3 to infer, if possible, unique values for attack rates and reciprocal mean handling times.

In both cases, our *in silico* feeding experiments comprise  $n_A^0$  individual consumers feeding on  $S$  different resources at controlled resource densities ( $x_{R_i}$  constant,  $i = 1, \dots, S$ ), under Holling type II stochastic dynamics —cf. reaction schemes in Eqs. (6)–(7). The parameter values that we would like to estimate from abundance data are: attack rate  $\alpha_i$  for each resource, as well as the reciprocal mean handling time over each resource,  $\nu_i$  (relaxation rate). Assume a fixed system volume  $N$ . Suppose we have a detection device that can record the number of free consumers,  $n_A(t_j)$ , as well as the number of consumers handling on each resource  $i$ ,  $n_i(t_j)$  ( $i = 1, \dots, S$ ), at time  $t_j$ . In order to use the multinomial distribution (11) with probabilities (29) as the probability of any configuration of the system, every replica of the experiment starts initially with  $n_A(0) = n_A^0$  free consumers, and no consumer is handling any prey. From now on, for the sake of simplicity, we consider that all predator individuals have equal attack and relaxation rates across resources ( $\alpha := \alpha_1 = \dots = \alpha_S$ ,  $\nu := \nu_1 = \dots = \nu_S$ ). Therefore, only two parameters  $\alpha$  and  $\nu$  are to be estimated.

We replicate the experiment a number of times  $T$ , each replica simulating the stochastic dynamics (using Gillespie (1977) method) with  $n_A^0$  free predators initially, and record abundances at time  $t_j$  for each realization  $j = 1, \dots, T$ . As shown in Appendix C, it is important that at least two sampling times are different (i.e.,  $t_j \neq t_k$  for some  $j \neq k$ ). Otherwise the model is not identifiable. In practice, from an empirical point of view, it is very unlikely that independent system samples are obtained at precisely the same exact time.

Thus, we have  $T$  independent stochastic realizations of the system, each replica is an observation at a

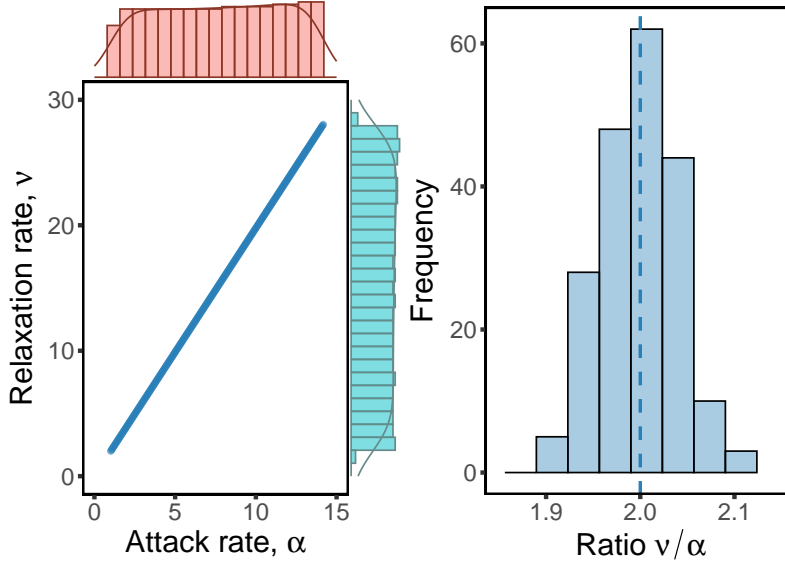


Figure 1: **Estimation using equilibrium samples reveals parameter redundancy.** We generated a single realization of the data matrix  $M(\{t_j\})$ , including  $T = 100$  independent model realizations starting from  $n_A^0 = 100$  searching predators at  $t = 0$ , using Gillespie’s stochastic simulation algorithm up to  $T = 100$  different ending times uniformly drawn in the interval  $[50\tau, 100\tau]$  to ensure stationarity, with  $\tau = (\nu + \alpha \sum_{i=1}^S x_{R_i})^{-1}$  the characteristic time of the dynamics. True rates are  $\alpha = 5$ ,  $\nu = 10$ , and individual consumers feed on  $S = 5$  resources with densities  $\mathbf{x}_R = (0.9, 0.7, 0.5, 0.3, 0.1)$ . Left: for that synthetic dataset, we run a L-BFGS-B optimization algorithm for the log-likelihood function (30) starting from different initial conditions, and we recorded the values of  $(\alpha^*, \nu^*)$  at which the maximum was reached. Optimal parameter values follow a perfect straight line. The distributions of  $\alpha^*$  and  $\nu^*$  are compatible with uniform distributions, and the optimal values of  $\log \mathcal{L}$  keep constant up to round-off errors. Overall, this indicates a line of infinite optima. Right: we replicated the previous experiment 200 times and, for each replica, we estimated the ratio  $\nu/\alpha$  by fitting the slope of the straight line. We obtained the distribution in the right panel, which is compatible with a Gaussian with mean 2 and variance extracted from data. The vertical, dashed line indicates the true ratio.

different time. The observational data is the following matrix

$$M(\{t_j\}) = \begin{pmatrix} n_{1,1}(t_1) & n_{1,2}(t_2) & \cdots & n_{1,T}(t_T) \\ \vdots & \vdots & \ddots & \vdots \\ n_{S,1}(t_1) & n_{S,2}(t_2) & \cdots & n_{S,T}(t_T) \end{pmatrix},$$

where columns comprise vectors  $\mathbf{n}_j = (n_{1,j}, \dots, n_{S,j})$  reporting observed abundances of handling predators at sampling time  $t_j$ . Free predator abundance at time  $t_j$  is computed with  $n_A(t_j) = n_A^0 - \sum_{i=1}^S n_{i,j}(t_j)$ . If each column of  $M(\{t_j\})$  is a sample of the Holling type II multi-resource dynamics, then the probability  $P_{\mathbf{n}_j}(t_j|\boldsymbol{\theta})$ , computed using Eqs. (11) and (29), is the probability of observing a population vector  $(n_A(t_j), \mathbf{n}_j)$  given a model parameter vector,  $\boldsymbol{\theta} := (\alpha, \nu)$  —notice that resource densities,  $x_{R_i} = n_{R_i}/N$ , the system volume,  $N$ , and the total number of consumers,  $n_A^0$ , are kept constant by assumption and therefore are not regarded as parameters in  $\boldsymbol{\theta}$ . Matrix  $M(\{t_j\})$  gathers  $T$  independent observations, so the function that assigns a likelihood to parameter vector  $\boldsymbol{\theta}$  given the data matrix  $M(\{t_j\})$  is multiplicative,  $\mathcal{L}(\boldsymbol{\theta}|M(\{t_j\})) = \prod_{j=1}^T P_{\mathbf{n}_j}(t_j|\boldsymbol{\theta})$ . The log-likelihood function,

$$\log \mathcal{L}(\boldsymbol{\theta}|M(\{t_j\})) = \sum_{j=1}^T \log P_{\mathbf{n}_j}(t_j|\boldsymbol{\theta}), \quad (30)$$

is more convenient for numerical purposes.

Now, consider the effect of sampling times and assume that matrix  $M(\{t_j\})$  is obtained when  $t_j$  are “large enough” compared to the characteristic time of the dynamics, which is given by  $\tau = (\nu +$

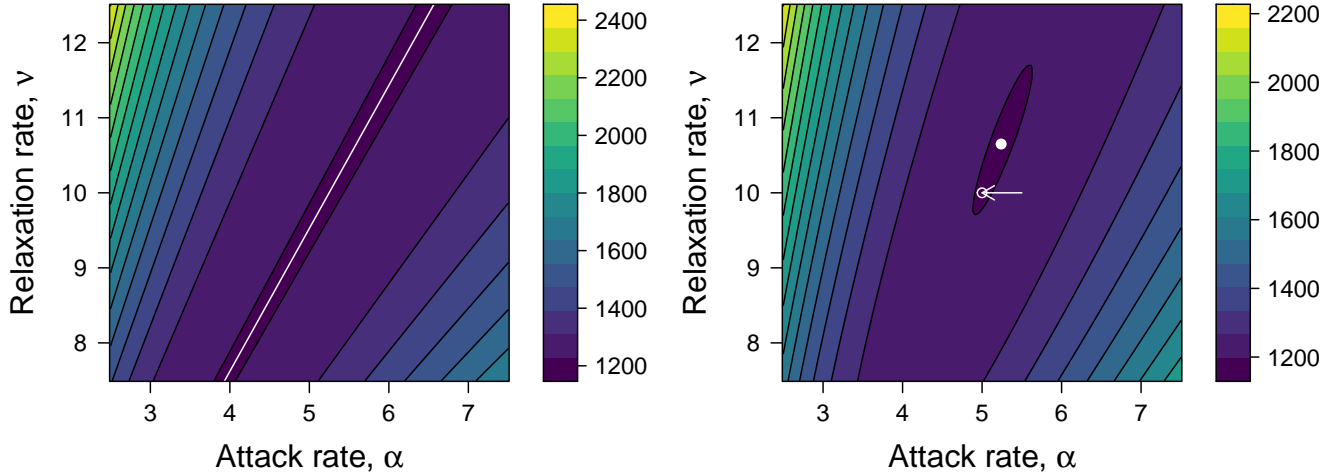


Figure 2: **Heatmaps of negative log-likelihood for stationary and out-of-equilibrium samples unveils redundancy** (for steady-state sampling). Representation of  $-\log \mathcal{L}$ , cf. Eq. (30), in the  $(\alpha, \nu)$  space, obtained for two sets of  $T = 100$  model samples generated for  $(\alpha, \nu) = (5, 10)$ : (i) taken at stationarity (sampling times  $\{t_j\}$  drawn uniformly in  $[10\tau, 20\tau]$ , left panel), and (ii) taken out of equilibrium (sampling times  $\{t_j\}$  drawn uniformly in  $[0, 10\tau]$ , right panel),  $\tau$  being the characteristic time scale. Left: a line of infinite minima is obtained. Right: The arrow marks the true pair  $(\alpha, \nu) = (5, 10)$  used for data generation, and the white point marks the minimum of  $-\log \mathcal{L}$ . In both panels, remaining parameter values are the same as in Fig. 1.

$\alpha \sum_{i=1}^S x_{R_i})^{-1}$ —see Eq. (29). In Fig. 1, *in silico* data (i.e., abundances of each behavioral type) were generated by stochastic simulation of a system with  $S$  resources (with fixed densities  $x_{R_i}$ ) consumed by  $n_A^0$  predator individuals. A number  $T$  of independent stochastic simulations started with initial conditions  $n_A = n_A^0$  and  $n_i = 0$  ( $i = 1, \dots, S$ ) at  $t = 0$ , each of which were stopped up to a time  $t_j$  drawn uniformly in  $[50\tau, 100\tau]$ , which is enough to reach stationarity. We maximized the log-likelihood function (30) using the L-BFGS-B optimization scheme (Zhu et al., 1997), which was used for numerical optimization throughout the contribution. Each realization of the data matrix  $M(\{t_j\})$  leads to a line of maximum likelihood in the  $(\alpha, \nu)$  space (see also Fig. 2, left panel), whose slope can vary over different replicas but can be determined precisely, providing to a consistent, unbiased estimation of the ratio  $\nu/\alpha$  (Fig. 1, right panel). Using data sampled at stationarity, even if ML is based on a time-dependent log-likelihood, only parameter ratios are identifiable.

From Eqs. (A.10)–(A.11) in Appendix A, it is immediate to check that the dimensionless ratio  $\nu/\alpha$  satisfies

$$\frac{\nu}{\alpha} = \frac{\langle n_A \rangle}{\langle n_{AR} \rangle} \sum_{i=1}^S x_{R_i}, \quad (31)$$

with  $\langle n_{AR} \rangle := \left\langle \sum_{i=1}^S n_i \right\rangle = \sum_{i=1}^S \langle n_i \rangle$  the average total number of handling consumers. In Appendix C we show that, when sampling times coincide ( $t_1 = t_2 = \dots = t_T = t$ ) and  $t \rightarrow \infty$ , the likelihood function does exhibit a line of infinite critical points with slope given by (31), in which quantities  $\langle n_A \rangle$  and  $\langle n_{AR} \rangle$  are substituted by averages across samples,  $\langle n_A \rangle_T := \frac{1}{T} \sum_{k=1}^T n_{A,k}$  and  $\langle n_{AR} \rangle_T := \frac{1}{T} \sum_{k=1}^T \sum_{i=1}^S n_{i,k}$ , respectively. Therefore, our numerical experiment (Fig. 1 and Fig. (2), left panels) is consistent with a submanifold of degenerated log-likelihood optimum values in the parameter space.

ML estimation techniques produce proper estimates for individual rates  $\alpha$  and  $\nu$  when data sample is taken out of equilibrium (and, importantly, at least two sampling times are different; see Appendix C). During the transient regime, samples were taken at times  $\{t_j\}$  drawn from a uniform distribution over the interval  $[0, 10\tau]$ ,  $\tau = (\nu + \alpha \sum_{i=1}^S x_{R_i})^{-1}$  being the characteristic dynamical time scale. Figure 2 compares

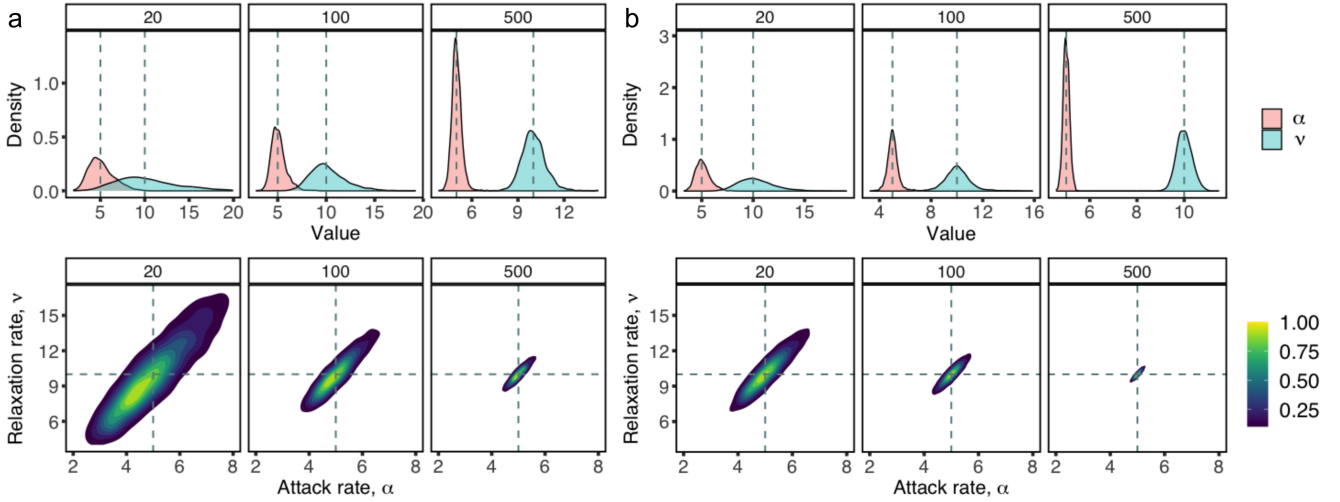


Figure 3: **Distributions of ML parameter estimates are unbiased for different number of initial consumers and increased number of temporal samples.** (a) The estimation procedure was repeated three times, for  $n_A^0 \in \{20, 100, 500\}$ , using  $T = 20$  samples in simulated out-of-equilibrium data. Upper panels show the distributions of  $\alpha^*$  and  $\nu^*$ , which tend to concentrate around true values (dashed lines) as  $n_A^0$  increases. Parameter estimates show bias for low  $n_A^0$  values, and the bias reduces as  $n_A^0$  increase. This phenomenon is more clearly observed in the density plots represented in lower panels for  $(\alpha^*, \nu^*)$ , the lighter the color indicating the higher density of estimated pairs. (b) Same as (a) for  $T = 100$  temporal samples. Compared to panel (a), we observe that variability is decreased as  $T$  increases. In both panels, remaining parameters are the same as in Fig. 1. 3000 estimation experiments were conducted to obtain distributions.

the log-likelihood landscapes obtained using steady-state (left panel) and out-of-equilibrium model samples (taken at times uniformly drawn over the interval  $[10\tau, 20\tau]$ ). A parameter screening in the  $(\alpha, \nu)$  plane shows a line of infinite maxima  $\nu^* \propto \alpha^*$  at equilibrium, confirming the results reported in Fig. 1. This evidences parameter redundancy when out-of-equilibrium samples are missing. On the contrary, a single log-likelihood maximum value arises and parameters can be truly inferred from the data when out-of-equilibrium samples are included (we have checked that, if samples are taken at times within the interval  $[0, 20\tau]$ , we also find a single maximum in log-likelihood, which implies that redundancy is avoided even when the data contain steady-state and out-of-equilibrium observations together). The accuracy of the estimates  $\alpha^*$  and  $\nu^*$  depends on data quality. For example, if the number of samples  $T$  taken is large, then it is expected that estimates will be close to actual values. Similarly, if the overall number of individuals  $n_A^0$  monitored in the experiment is large, then better estimates for model parameters are expected.

This phenomenon is illustrated in Figure 3, which shows the distributions of ML estimates  $\alpha^*$  and  $\nu^*$  for  $T = 20$  and  $T = 100$  samples, and for three values of  $n_A^0$ . We observe that distributions become closer to true values as  $T$  and  $n_A^0$  increase. The joint distributions of parameter estimates (bottom panels) show also that estimates are consistent: although for small samples and low values of  $n_A^0$  the joint distributions are asymmetric and not centered in true values and show certain bias, as  $T$  and  $n_A^0$  increase the distributions tend to center around actual parameter values with small variability.

These distributions allow to calculate 95% confidence intervals for parameter estimates. We calculated the 2.5% and 97.5% percentiles of the empirical distributions obtained in Fig. 3 (upper panels), these values corresponding to the limits of the sought confidence interval. Intervals are listed in Table 1 for the set of  $T$  and  $n_A^0$  values reported in Fig. 3. As described, intervals are not centered in actual values for small samples, and tend to be centered (with decreasing spread) as  $T$  and  $n_A^0$  increase.

Usually, one can only have access to one single realization of empirical observations. In this case, we still can obtain ML parameter estimates as well as confidence intervals following the log-likelihood ratio test (Cox and Snell, 1989). We illustrate this procedure using the same data matrix as in Fig. 2 (right panel), for which the ML parameter estimates are  $(\alpha^*, \nu^*) = (5.24, 10.65)$ . We obtained log-likelihood profiles shown

95% Confidence Intervals

$n_A^0$	True value	$T = 20$ times	$T = 100$ times
20	$\alpha = 5$	[2.86, 8.25]	[3.95, 6.81]
100	$\alpha = 5$	[3.86, 6.89]	[4.22, 6.04]
500	$\alpha = 5$	[4.47, 5.69]	[4.75, 5.28]
20	$\nu = 10$	[4.86, 17.76]	[7.42, 14.32]
100	$\nu = 10$	[7.25, 14.55]	[8.08, 12.54]
500	$\nu = 10$	[8.67, 11.66]	[9.40, 10.67]

Table 1: **Confidence intervals calculated using parameter distributions.** Here we present true values and 95% confidence intervals for the *in silico* experiments shown in Fig. 3. The empirical distributions of parameter estimates, shown in that figure, are used to compute the corresponding percentiles that form the confidence interval. The sample of estimated parameters used to compute intervals has a size of 3000 estimates for each  $(T, n_A^0)$  combination.

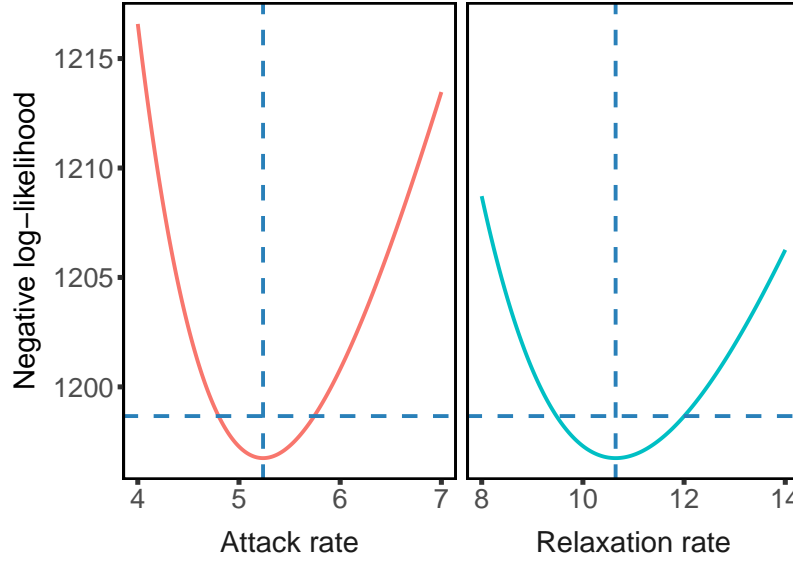


Figure 4: **Confidence interval estimation using negative log-likelihood profiles** for the same synthetic matrix  $M(\{t_j\})$  used for the heatmap in Fig. 2 (right panel). Negative log-likelihood profiles  $\ell_1(\alpha)$  and  $\ell_2(\nu)$  (left and right panels, respectively) yield confidence intervals as the abscissae of the intercept of the profiles  $\ell_1(\alpha)$  and  $\ell_2(\nu)$  with the horizontal dotted lines whose ordinates are equal to  $\ell_1(\alpha^*) + \chi_{1,0.95}^2/2$  and  $\ell_2(\nu^*) + \chi_{1,0.95}^2/2$ , respectively.  $(\alpha^*, \nu^*)$  refers to the point (marked with vertical, dotted lines) at which the maximum of negative log-likelihood is reached.

in Fig. 4 by minimizing negative log-likelihood over  $\nu$  for fixed  $\alpha$ ,  $\ell_1(\alpha) := -\min_{\nu} \log \mathcal{L}(\alpha, \nu | M(\{t_i\}))$  (left panel), and minimizing over  $\alpha$  for fixed  $\nu$ ,  $\ell_2(\nu) := -\min_{\alpha} \log \mathcal{L}(\alpha, \nu | M(\{t_i\}))$  (right panel). Then, the 95% confidence interval for  $\alpha$  can be defined as the set of values satisfying  $|\ell_1(\alpha) - \ell_1(\alpha^*)| \leq \chi_{1,0.95}^2/2$ ,  $\chi_{1,0.95}^2$  being the 95% percentile of the chi-squared distribution with one degree of freedom. Similarly, the 95% confidence interval for  $\nu$  is defined as the set of values satisfying  $|\ell_2(\nu) - \ell_2(\nu^*)| \leq \chi_{1,0.95}^2/2$ . Note that  $\ell_1(\alpha^*) = \ell_2(\nu^*)$  equals the minima obtained for the negative log-likelihood at the point  $(\alpha^*, \nu^*)$  as well. This yields the intervals  $\alpha \in [4.80, 5.73]$  and  $\nu \in [9.51, 11.98]$  at a 95% confidence level, which are consistent with their counterpart in Table 1 for  $T = 100$  and  $n_A^0 = 100$ .

In case the experimental detection device could only count the number of free consumers at each sampling time, then we should resort on the one-dimensional marginal probability distribution of free consumers to build the correct likelihood function for our estimation procedure. Because any marginal of a multinomial is again a multinomial distribution (in particular, a binomial distribution if all variables but one are marginalized), the marginal distribution of the number of free predators is given by the binomial (A.1)



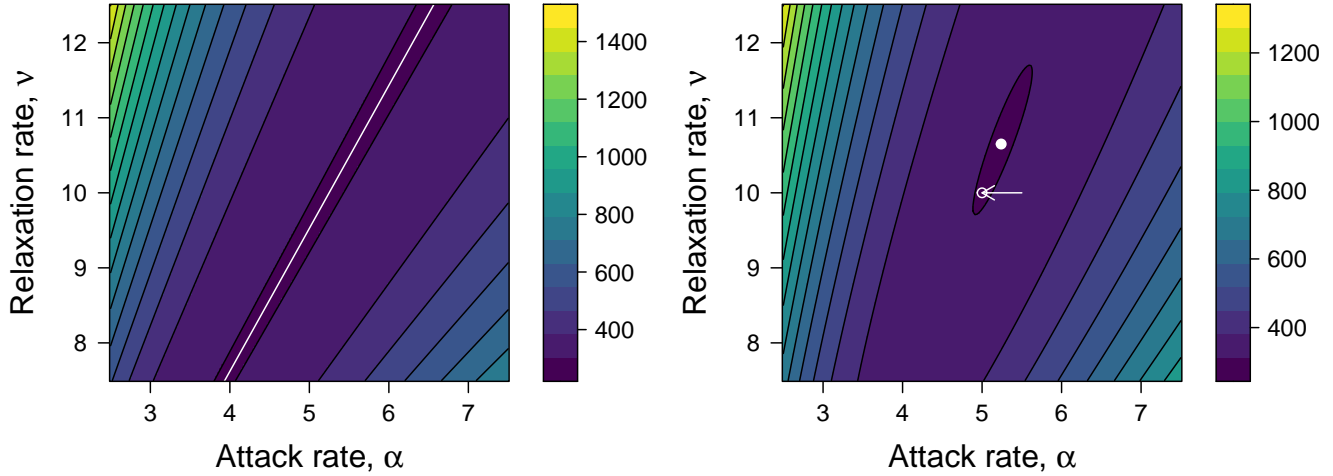


Figure 5: **Heatmaps of negative log-likelihood using only free predator abundances for stationary and out-of-equilibrium samples unveils redundancy** (for steady-state sampling). Representation of  $-\log \mathcal{L}$ , cf. Eq. (32) (i.e., using only free predators abundance information), in the  $(\alpha, \nu)$  space, obtained for the same two sets of model samples generated for Fig. 2: true parameter values are  $(\alpha, \nu) = (5, 10)$  (arrow in the right panel), and stationary (out-of-equilibrium) samples yield the log-likelihood represented in the left (right) panel. Observe that the heatmaps only differ from those in Fig. 2 in the colorbar values, because here we are using only partial information to evaluate  $-\log \mathcal{L}$  for the same model observations as in that figure. Again, we find parameter redundancy for stationary sampling, whereas a single minimum is obtained for out-of-equilibrium sampling (the white point marks the minimum of  $-\log \mathcal{L}$ ). Remaining parameter values were chosen as in Fig. 1.

with

$$p_A(t) = \frac{\nu + \alpha e^{-(\nu + \alpha \sum_{k=1}^S x_{R_k})t} \sum_{i=1}^S x_{R_i}}{\nu + \alpha \sum_{k=1}^S x_{R_k}},$$

as obtained in Eq. (29). Then we can maximize the log-likelihood function

$$\mathcal{L}(\boldsymbol{\theta} | M(\{t_j\})) = \prod_{j=1}^T P_{n_{A,j}}(t_j | \boldsymbol{\theta}) \quad (32)$$

where, again, independence has been assumed in stochastic realizations, and  $\boldsymbol{\theta} = (\alpha, \nu)$  represents the vector of varying model parameters. Here

$$M(\{t_j\}) = (n_{A,1}(t_1), n_{A,2}(t_2), \dots, n_{A,T}(t_T))$$

stands for an empirical data vector with  $T$  observational abundances  $n_{A,j}$  of the number of free predators. We use the time-dependent log-likelihood to estimate  $\alpha$  and  $\nu$  depending on the sampling scheme used to generate the vector  $M(\{t_j\})$ . The ratio  $\nu/\alpha$  can be inferred if the experiment only reports steady-state abundance data, but not rates themselves. Model identifiability is illustrated with the plots of log-likelihood out-of and at equilibrium in Fig. 5, where only searching predators abundance data is used for ML parameter estimation. As in the first experiment, in which the abundances of each behavioral type were available, unique parameter estimates can be inferred from out-of-equilibrium model observations.

Parameter estimation is illustrated with a second *in silico* experiment based only on free predators abundance vectors —cf. Eq. (4). Figure 6 shows the distributions obtained in this case, gathering 3000 ML parameter estimates for varying  $n_A^0$ , the overall number of consumers, and  $T$ , the number of observations (columns) used to construct vector  $M(\{t_j\})$ . Distributions for each parameter (upper panels) are very similar to those reported in Fig. 3, which took information of the full vector of abundances. Only for

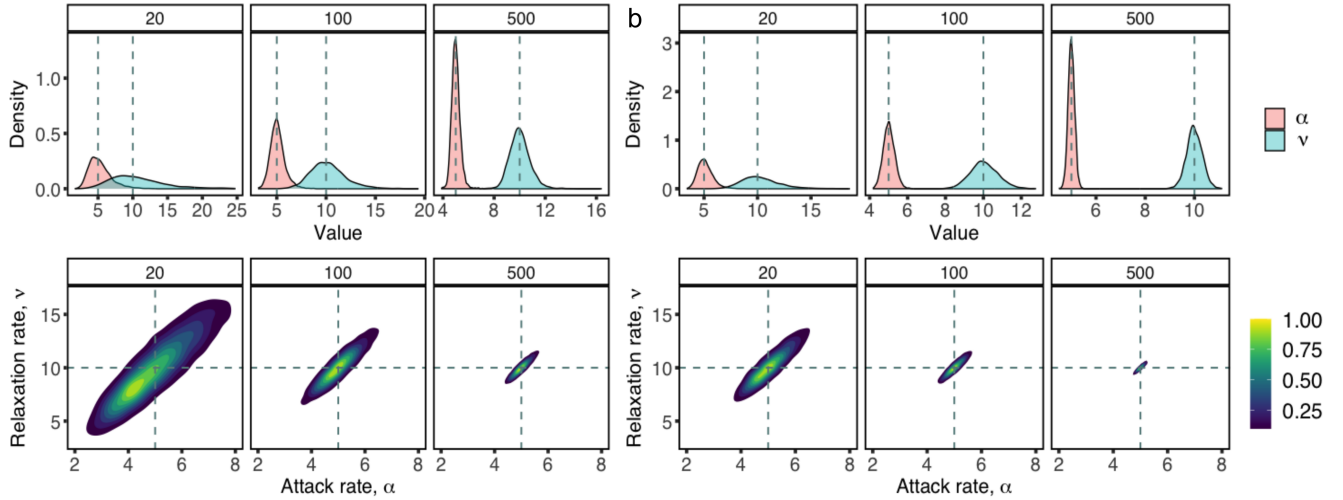


Figure 6: **Distributions of ML parameter estimates using only free predator abundances are unbiased for increasing number of initial consumers and number of temporal samples.** Same as Fig. 3 for a ML estimation procedure based on searching predator numbers. In this case, we have only used the number of free predators recorded at each sampling time, see Eq. (4). Distributions are quite similar to those obtained by using the vector of population handling predator abundances,  $\mathbf{n} = (n_1, \dots, n_S)$ , which are shown in Fig. 3. In both panels, parameters are the same as in Fig. 1.

95% Confidence Intervals			
$n_A^0$	True value	$T = 20$ times	$T = 100$ times
20	$\alpha = 5$	[2.93, 9.34]	[3.95, 6.80]
100	$\alpha = 5$	[3.82, 6.94]	[4.47, 5.64]
500	$\alpha = 5$	[4.44, 5.71]	[4.75, 5.25]
20	$\nu = 10$	[4.81, 20.27]	[7.36, 14.40]
100	$\nu = 10$	[7.12, 14.62]	[8.69, 11.54]
500	$\nu = 10$	[8.64, 11.75]	[9.37, 10.63]

Table 2: **Confidence intervals using only free predators abundances** calculated using parameter distributions (Fig. 6). The table lists actual values and 95% confidence intervals for parameters of the *in silico* experiments shown in Fig. 6. Again, 3000 samples of estimated parameters were used to compute intervals for each  $(T, n_A^0)$  combination.

small-size samples (i.e., for  $n_A^0$  and  $T$  small), distributions and density plots (lower panels) appear to be more spread in this case. This effect can be seen more clearly in confidence intervals, which are calculated using the empirical distributions in Fig. 6: reported intervals (Table 2), based only on free predator abundances, are wider than those listed in Table 1, for  $n_A^0 = 20$  and  $T = 20$ . Once the ML estimation procedure contains more information, as  $n_A^0$  and  $T$  increase, the intervals estimated from both synthetic experiments are consistently comparable.

## 5. Discussion

In this contribution we have studied model identifiability in two stochastic population models, and we have shown that, when independent model observations include samples taken at different sampling times in the transient regime, the model is truly identifiable in the sense that model parameters can be uniquely estimated from observations. On the contrary, if population numbers are only observed at stationarity, then the model is set-identifiable because only a combination of parameters can be inferred from data. We derived model redundancy conditions using the generation function formalism in both a single-population birth-death-immigration model, and a multi-resource consumer-resource stochastic population model. For the latter, synthetic data were used to calibrate the model in two *in silico* experiments: one in which

only free predator population numbers are considered, and another in which the whole matrix of searching and handling predator numbers is available. If this model was to be calibrated using actual, experimental data, and independent observations were taken at stationarity (or at equal sampling times), the problem of parameter redundancy would persist: the log-likelihood function would produce an infinite set of maxima, because the critical points conditions would reduce to one single equation —see the derivation of Eq. (C.1) in Appendix C.

Importantly, if temporal samples contain both transient and stationary observations, model parameters remain perfectly identifiable. Therefore, an important implication of our work is that time series must include out-of-equilibrium observations to infer parameters without redundancy. Several statistical methods allow one to separate transient and steady-state regimes in a time series: for example, the Chow test (Chow, 1960) allows to separate a positive slope in abundance time series from a slope compatible with zero, proper of a steady-state regime. For empirical data, the experiment has to be designed to collect out-of-equilibrium samples, by monitoring the experiment from an initial condition far from equilibrium.

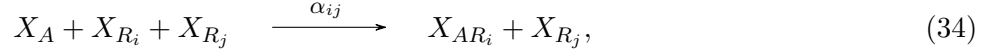
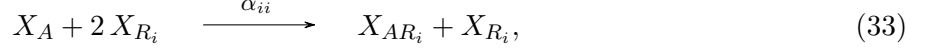
For stochastic processes other than the multi-resource consumer-resource model analyzed here, we can state that, as long as the distributions of observational variables depend on parameter combinations whose intersection is a manifold that does not reduce to a unique point, then identification failure will emerge. This is particularly important at equilibrium, because many models are calibrated using steady-state distributions (which are the ones commonly available, unlike the out-of-equilibrium distributions, which are difficult to derive analytically). If those steady-state distributions depend on parameter ratios, as in the examples discussed here, then the issue of parameter inference will arise. Because many quantities at equilibrium depend on dimensionless parameter combinations, we anticipate that the problem of parameter redundancy can be very frequent in stochastic models commonly used. This has important implications on ML parameter estimation using empirical data, because if the system under study is sampled once the steady-state has been reached, the problem of parameter redundancy is likely to arise.

Additional information should be available to conduct precise parameter inference at stationarity. The heatmaps in Fig. 2 represent the log-likelihood function values, Eq. (30), as the two parameters vary between a minimum and a maximum value along the axes, and show that when samples are taken close enough to stationarity, then a line of optimal values arise. This agrees with the expression found in Appendix A for the joint stationary distribution (as well as its marginals), which depends on the compound parameter controlled by the dimensionless ratio  $\nu/\alpha$  (the relaxation rate in units of the attack rate). This nondimensional quotient is the one that can be statistically inferred from data at equilibrium. If we were able to sort out, experimentally, individuals in the different behavioral modes at stationarity (attacking *vs* handling), we could identify, as shown in Figure 2, the true attack rate  $\alpha$  on resources, respectively, provided that we had external information to previously determine the relaxation rate  $\nu$ , and confidence intervals could be assigned (see Table 1). Regarding to the rest of parameters, that is, the total number of consumers,  $n_A^0$ , resource densities  $x_{R_i}$ , and the volume,  $N$ , of the system, we believe that they will be fairly easy to initially fix and control in the setting of any experiment.

We stress that any additional information used to solve parameter redundancy at equilibrium should come from monitoring the system out of equilibrium. In particular, precise estimates of  $\tau$  would break redundancy at stationarity. Observe that, if the characteristic time  $\tau$  and the equilibrium distribution are known, one can obtain non-redundant parameter estimates in the case of equal handling times. However, estimating the characteristic time requires knowledge of the transient dynamics before reaching the steady state. One could perturb an experimental system by varying populations along a proper direction away from the equilibrium point (along an eigenvector of the Jacobian matrix evaluated at equilibrium). Then the eigenvalue would help measure the relaxation time. Methods alternative to our ML approach, such as the one proposed in Appendix D (based on the initial decay of the abundance of free predators), imply breaking stationarity or having precise knowledge of the transient state. These methods are, therefore, equivalent to sampling the system during the transient dynamics before stationarity.

As for the generality of our results, the multinomial distribution derived in section 3, either at stationarity or out of equilibrium, can be extended to a situation in which predator’s feeding efficiency varies

non-linearly in response to both the same or different resources. Let  $\alpha_{ij}$  ( $i, j = 1, \dots, S$ ) measure the effect of resource densities  $x_{R_i}$  and  $x_{R_j}$  on the formation of consumers handling on resource  $i$ . The effect of the density of the same resource or alternative resources on the formation of handling consumers of the focal resource  $i$  can be taken into account, for example, if we assume the following  $3 \times S$  chemical reaction scheme:



where the second reaction holds for  $i \neq j$ . In this way, depending on the balance between  $\alpha_{ii}$  and  $\alpha_{ij}$ , consumer's feeding rate on resource  $R_i$  can be enhanced or damped by the presence of an alternative resource  $R_j$  more than it is by the presence of the same type of resource  $R_i$ . Therefore, consumer efficiency on a focal resource is affected not only by the density of the focal resource but also by the presence of alternative resources.

Proceeding in the same way as in Palamara et al. (2022), the functional response that arises from those fundamental processes corresponds to a multispecies version of the Holling type III functional response. It is easy to see that, according to the reaction scheme (33)–(35) and because of mass action, the per capita consumption rate of a free/searching individual  $X_A$  feeding on a focal resource  $i$  becomes

$$\theta_i := \alpha_{ii}(x_{R_i}^0)^2 + \sum_{j \neq i} \alpha_{ij} x_{R_i}^0 x_{R_j}^0. \quad (36)$$

If only a single resource were present, the first term would correspond to a Holling type III functional response, with Hill exponent (i.e., the exponent to which resource densities are raised)  $r = 2$  (Real, 1977). The definition of per-capita consumption rates  $\theta_i$  would be the only change in the derivation of the distributions, at stationarity or out of equilibrium, for this scheme of elemental processes. All the results derived for the multi-resource Holling type II case treated in section 3 hold in this more general case, because transition probability rates are exactly the same except for the replacement of parameter  $\theta_i$  by the one in Eq. (36), which is still a constant parameter that depends on resource densities under the hypothesis of chemostatic conditions. Indeed, note that the two channels specified by Eqs. (33) and (34) can be subsumed into the transition probability rate for increasing the number of handling predators by one unit —first equation of (9), with  $\theta_i$  given by (36). Reaction (35) leads to the same transition rate for decreasing the number of handling predators by one unit —second equation of (9). As a consequence, the previous derivation of the multi-resource Holling Type II stationary and out-of-equilibrium distributions applies here replacing parameter  $\theta_i$  by the new definition. In both type II and III cases, the formulae that define probability parameters of the multinomial distributions remain valid with the replacement of  $\theta_i$  by its expression in (36). Consequently, the experiments designed to carry out ML parameter estimation can be generalized to the Holling type III dynamics as well, together with the issues of parameter identifiability. Our results can be extended to more general consumer-resource contexts, unveiling the issue of parameter redundancy to be very common in stochastic population models.

We used the assumption of equal handling times to obtain closed-form solutions for the out-of-equilibrium joint distribution of population abundances. When handling times for different resources are not equal, our methodology applies if the master equation is integrated numerically to obtain probability distributions as functions of time. Given these probabilities, one can compute numerically log-likelihoods and estimate the  $2S$  parameters  $\{\alpha_i, \nu_i\}$  ( $i = 1, \dots, S$ ) without redundancy. It is important to note that, in this case, the degeneracy of the approach when samples are taken at equilibrium increases because, in this case, only  $S$  ratios ( $\alpha_i/\nu_i$ ) are identifiable —see Eqs. (12)–(13). If only abundance data at stationarity is available, to solve redundancy one would need prior knowledge of  $S$  additional quantities relating the unknown rates.

On a side note, it is important to observe that the assumption of equal handling times does not impose any restriction in the case of one single resource: for the standard Holling type II consumer-resource feeding dynamics, the calculations that follow are fully general for the case  $S = 1$ , and the multinomial out-of-equilibrium solution is exact in that case —provided that the dynamics starts with  $n_A^0$  searching consumers initially.

Traditional methods for estimating attack and handling rates in consumer-resource interactions rely on depletion experiments, where a fixed number of consumers are placed with a known initial resource density and the abundance of resources decreases over time due to consumption (Rogers, 1972). Parameters like attack rate and handling time are inferred by fitting the depletion curve to a model of resource abundance. These experiments run out-of-equilibrium, without resource replenishment. In contrast, our approach assumes chemostatic conditions, where consumed resources are continuously replenished, and the total number of consumers remains constant. By tracking consumer behavior —distinguishing between searching and handling states— through animal trajectories, one could apply our maximum likelihood approach to estimate process rates. Monitoring the early stages of the experiment is crucial to avoid parameter redundancy. Experimental setups capable of classifying individual behaviors, such as searching for or digesting resources, can support this approach and allow parameter estimation under controlled conditions (Bartumeus et al., 2002, 2005, 2003). Because maximum likelihood estimators are asymptotically those with minimal variance, we expect that our approach is better than simply fitting a curve to data, as in depletion experiments.

Another shortcoming of the inference approach followed in this contribution comes from the independence of measurements at different times, needed to construct the likelihood function to uniquely estimate feeding rates. Such assumptions would require the use of different replicates for every measurement, which would importantly increase the cost of hypothetical field or laboratory experiments due to the repetition of realizations. In principle, one could not realistically use the same replica to gather different measurement at different times, because the independence hypothesis would be broken, together with sampling the system close to equilibrium as time elapses. Other inference approaches, not based on multiplicative likelihood functions, should be devised to extract meaningful parameter estimates from controlled experiments.

Classical feeding experiments have been performed for a variety of systems and taxa, where feeding rates have been measured mainly using the averaged expressions at fixed resource levels (Novak and Stouffer, 2021; Stouffer and Novak, 2021; DeLong, 2021). Here, we propose to use the distribution of behavioral types, thus allowing for refined estimates of parameters associated to elemental process. For the multi-resource Holling type II stochastic dynamics (as well as the extensions discussed here), we managed to give a complete analysis of the temporal dynamics, as well as its steady state. We described the potential inference methods to use to estimate, when identifiable, population parameters or characterized the subsets of the parameter space in which the model is set-identifiable. For models that are not fully solvable like those studied here, one can always resort to numerical integration of the master equation in order to compute numerically the likelihood function and infer model parameters. We envision the application of the methods presented in this contribution in the design and realization of novel tests of predator-prey theory for controlled laboratory experiments (Rosenbaum and Rall, 2018) or field experiments (Barbier et al., 2021; Abrams, 2022).

## Acknowledgements

We acknowledge advice and fruitful discussions with Gian Marco Palamara and the rest of members of our Unique-Priority project team. Funding was provided by grants CRISIS (PGC2018-096577-B-I00) to D.A. and J.A.C., UNIQUE (PID2021-127202NB-C21) to D.A., and PRIORITY (PID2021-127202NB-C22) to J.A.C., all funded by MCIN/AEI/10.13039/501100011033 and “ERDF. A way of making Europe”.

## Code availability

Code to reproduce all findings is available at [https://github.com/vankampen92/PROJECT\\_HOLLING\\_II\\_nD](https://github.com/vankampen92/PROJECT_HOLLING_II_nD)

## Appendix A. Steady-state distribution derivation

It can be shown that, for single resource case ( $S = 1$ ), the steady-state probability distribution of the number of free predators is a binomial  $B(n_A^0, p)$  with probability parameter  $p = (1 + \frac{\theta}{\nu})^{-1}$ ,

$$P_{n_A} = \binom{n_A^0}{n_A} p^{n_A} (1 - p)^{n_A^0 - n_A}, \quad (\text{A.1})$$

$n_A$  being the number of searching consumers (Palamara et al., 2022). One can take advantage of knowing that distribution to rigorously derive the functional response associated to the single resource case ( $S = 1$ ), which turns out to be the Holling type II functional response (Palamara et al., 2022). We here extend the binomial distribution result to a multi-resource case by solving the master equation at stationarity, which yields an analytical expression for the joint steady-state distribution  $P_{\mathbf{n}}$ . We can do this because transition probability rates (9) are linear in population numbers (Kendall, 1949; Nisbet and Gurney, 2004). As a result, the single-resource binomial distribution is extended to a multi-resource case, leading to a multinomial distribution for the joint probability distribution of observing a vector  $(n_A, \mathbf{n})$  of the populations of behavioral types, as we show next.

The stationary distribution  $P_{\mathbf{n}}$  satisfies the (infinite) system of equations defined by setting the right-hand side of Eq. (10) equal to zero. This distribution can be calculated recursively because, if the following *detailed-balance conditions* hold,

$$T(\mathbf{n}|\mathbf{n} + \mathbf{e}_i)P_{\mathbf{n}+\mathbf{e}_i} = T(\mathbf{n} + \mathbf{e}_i|\mathbf{n})P_{\mathbf{n}}, \quad (\text{A.2})$$

for  $i = 1, \dots, S$ , then all the terms in the right-hand side of Eq. (10) cancel out. This can be seen as follows. According to Eq. (9), detailed-balance conditions (A.2) reduce to

$$\nu_i(n_i + 1)P_{\mathbf{n}+\mathbf{e}_i} = \theta_i \left( n_A^0 - \sum_{j=1}^S n_j \right) P_{\mathbf{n}}, \quad (\text{A.3})$$

for  $i = 1, \dots, S$ . Making the change  $n_i \rightarrow n_i - 1$  in the last equation one gets

$$T(\mathbf{n} - \mathbf{e}_i|\mathbf{n})P_{\mathbf{n}} = T(\mathbf{n}|\mathbf{n} - \mathbf{e}_i)P_{\mathbf{n}-\mathbf{e}_i},$$

which together with (A.2) cancel out the  $i$ -th term in the right-hand side of the master equation. Therefore, if detailed-balance conditions (A.3) hold, then the probability distribution that satisfies them is the stationary distribution.

Let us apply Eq. (A.3) to the  $i$ -th and  $j$ -th handling predator's populations: this yields

$$P_{\mathbf{n}} = \frac{\nu_i}{\theta_i} \frac{n_i + 1}{n_A^0 - \sum_{k=1}^S n_k} P_{\mathbf{n}+\mathbf{e}_i}, \quad P_{\mathbf{n}} = \frac{\nu_j}{\theta_j} \frac{n_j + 1}{n_A^0 - \sum_{k=1}^S n_k} P_{\mathbf{n}+\mathbf{e}_j}, \quad (\text{A.4})$$

and, subsequently,

$$P_{\mathbf{n}+\mathbf{e}_i} = \frac{\nu_j}{\theta_j} \frac{\theta_i}{\nu_i} \frac{n_j + 1}{n_i + 1} P_{\mathbf{n}+\mathbf{e}_j}.$$

Making the change  $n_i \rightarrow n_i - 1$  in the last equation one gets

$$P_{\mathbf{n}} = \frac{\nu_j}{\theta_j} \frac{\theta_i}{\nu_i} \frac{n_j + 1}{n_i} P_{\mathbf{n}-\mathbf{e}_i+\mathbf{e}_j}. \quad (\text{A.5})$$

This equation is the fundamental identity we are going to use to calculate recursively the distribution  $P_{\mathbf{n}}$ , because it has the property of keeping constant the total number of handling predators  $\sum_{j=1}^S n_j$  after each recursion step.

Now, apply Eq. (A.5) to resources  $i = 1$  and  $j = 2$ : this gives

$$P_{\mathbf{n}} = \frac{\theta_1}{\nu_1} \frac{\nu_2}{\theta_2} \frac{n_2 + 1}{n_1} P_{(n_1-1, n_2+1, \dots, n_S)}.$$

Apply again the recursion (A.5) to resources  $i = 1$  and  $j = 2$  in  $P_{(n_1-1, n_2+1, \dots, n_S)}$ :

$$P_{\mathbf{n}} = \left( \frac{\theta_1}{\nu_1} \frac{\nu_2}{\theta_2} \right)^2 \frac{(n_2 + 1)(n_2 + 2)}{n_1(n_1 - 1)} P_{(n_1-2, n_2+2, \dots, n_S)}.$$

By repeating the iteration  $n_1 - 2$  additional times one gets

$$P_{\mathbf{n}} = \left( \frac{\theta_1}{\nu_1} \frac{\nu_2}{\theta_2} \right)^{n_1} \binom{n_1 + n_2}{n_1} P_{(0, n_1+n_2, \dots, n_S)},$$

where  $\binom{m}{p} = \frac{m!}{p!(m-p)!}$  stands for the binomial coefficient  $m$  choose  $p$ . We observe that we have written an arbitrary probability  $P_{\mathbf{n}}$  in terms of the probability of a configuration that has no predators handling on the first resource.

The recursion can be applied to  $P_{(0, n_1+n_2, \dots, n_S)}$  using the positions  $i = 2$  and  $j = 3$ . After  $n_1 + n_2$  recursion steps, the number of individuals handling on the second resource will be exactly zero:

$$P_{\mathbf{n}} = \left( \frac{\theta_1}{\nu_1} \frac{\nu_2}{\theta_2} \right)^{n_1} \left( \frac{\theta_2}{\nu_2} \frac{\nu_3}{\theta_3} \right)^{n_1+n_2} \binom{n_1 + n_2}{n_1} \binom{n_1 + n_2 + n_3}{n_1 + n_2} P_{(0, 0, n_1+n_2+n_3, \dots, n_S)},$$

which simplifies to

$$P_{\mathbf{n}} = \left( \frac{\theta_1}{\nu_1} \right)^{n_1} \left( \frac{\theta_2}{\nu_2} \right)^{n_2} \left( \frac{\nu_3}{\theta_3} \right)^{n_1+n_2} \binom{n_1 + n_2 + n_3}{n_1, n_2, n_3} P_{(0, 0, n_1+n_2+n_3, \dots, n_S)},$$

where we have used known relations between the binomial and the multinomial coefficients, the latter being defined as  $\binom{m}{s_1, \dots, s_q} = \frac{m!}{\prod_{k=1}^q s_k!}$  (with  $s_1 + \dots + s_q = m$ ). Here, probabilities are expressed in terms of configurations in which no individual consumers feed on resources 1 and 2.

Iterating the process until all the populations in every class have moved to the  $S$ -th one, we obtain

$$P_{\mathbf{n}} = \binom{\sum_{k=1}^S n_k}{n_1, \dots, n_S} \left( \frac{\nu_S}{\theta_S} \right)^{\sum_{k=1}^{S-1} n_k} \left[ \prod_{k=1}^{S-1} \left( \frac{\theta_k}{\nu_k} \right)^{n_k} \right] P_{(0, \dots, 0, \sum_{k=1}^S n_k)}. \quad (\text{A.6})$$

To finish our derivation we just need to evaluate  $P_{(0, \dots, 0, q)}$  for an arbitrary overall number  $q = \sum_{k=1}^S n_k$  of predators handling on resource  $S$ . This can be achieved using the first identity in Eq. (A.4), which for  $i = S$  can be written, after changing  $n_S \rightarrow n_S - 1$ , as

$$P_{\mathbf{n}-\mathbf{e}_S} = \frac{\nu_S n_S}{\theta_1(n_A^0 + 1 - \sum_{k=1}^S n_k)} P_{\mathbf{n}}.$$

We choose  $\mathbf{n} = (0, \dots, 0, q) = q\mathbf{e}_S$  in the last equation. Upon  $q$  iterations, we obtain

$$\begin{aligned} P_{(q-1)\mathbf{e}_S} &= \frac{\nu_S}{\theta_S} \frac{q}{n_A^0 + 1 - q} P_{q\mathbf{e}_S}, \\ P_{(q-2)\mathbf{e}_S} &= \left( \frac{\nu_S}{\theta_S} \right)^2 \frac{q(q-1)}{(n_A^0 + 1 - q)(n_A^0 + 2 - q)} P_{q\mathbf{e}_S}, \\ &\vdots \\ P_0 &= \left( \frac{\nu_S}{\theta_S} \right)^q \frac{q!}{(n_A^0 + 1 - q) \cdots (n_A^0 - 1) n_A^0} P_{q\mathbf{e}_S}, \end{aligned}$$

where  $\mathbf{0} = (0, \dots, 0)$ . Equivalently,

$$P_{q\mathbf{e}_S} = \binom{n_A^0}{q} \left( \frac{\theta_S}{\nu_S} \right)^q P_{\mathbf{0}}.$$

Setting  $q = \sum_{k=1}^S n_k$  in the last equation and substituting into (A.6) we obtain the joint stationary distribution  $P_{\mathbf{n}}$  up to a normalization constant  $P_{\mathbf{0}}$ ,

$$P_{\mathbf{n}} = \frac{n_A^0!}{n_1! \cdots n_S! n_A!} \left[ \prod_{k=1}^S \left( \frac{\theta_k}{\nu_k} \right)^{n_k} \right] P_{\mathbf{0}}, \quad (\text{A.7})$$

where we have used Eq. (8) and

$$\binom{n_A^0}{\sum_{k=1}^S n_k} \binom{\sum_{k=1}^S n_k}{n_1, \dots, n_S} = \frac{n_A^0!}{n_1! \cdots n_S! n_A!}.$$

Observe that the distribution calculated in (A.7) up to normalization is non-zero over the hyperplane defined by  $n_A + \sum_{k=1}^S n_k = n_A^0$ . State vectors outside that plane have probability zero of being realized.

In the form of Eq. (A.7), the joint stationary distribution remarkably resembles a multinomial distribution, see Eq. (11). To write Eq. (A.7) as a multinomial distribution, first we have to calculate the normalization constant  $P_{\mathbf{0}}$ . It follows immediately after evaluating the probability of observing a total number  $q$  of handling predators,  $q = \sum_{k=1}^S n_k$ , which can be done as follows.

Let  $P_q$  be the probability that the overall number of handling predators,  $\sum_{k=1}^S n_k$ , is equal to  $q$ . It can be calculated (up to normalization) by summing Eq. (A.7) over the hyperplane  $\sum_{k=1}^S n_k = q$ :

$$P_q = \sum_{\sum_i n_i = q} \frac{n_A^0!}{n_1! \cdots n_S! n_A!} \left[ \prod_{k=1}^S \left( \frac{\theta_k}{\nu_k} \right)^{n_k} \right] P_{\mathbf{0}},$$

which can be written alternatively as

$$P_q = \binom{n_A^0}{q} \sum_{\sum_i n_i = q} \frac{q!}{n_1! \cdots n_S!} \left[ \prod_{k=1}^S \left( \frac{\theta_k}{\nu_k} \right)^{n_k} \right] P_{\mathbf{0}}.$$

According to the multinomial theorem (Knuth, 1997), the sum above can be explicitly expressed as

$$\sum_{\sum_i n_i = q} \frac{q!}{n_1! \cdots n_S!} \prod_{k=1}^S \left( \frac{\theta_k}{\nu_k} \right)^{n_k} = \left( \sum_{k=1}^S \frac{\theta_k}{\nu_k} \right)^q,$$

which implies that

$$P_q = \binom{n_A^0}{q} \left( \sum_{k=1}^S \frac{\theta_k}{\nu_k} \right)^q P_{\mathbf{0}}. \quad (\text{A.8})$$

Therefore, if we impose normalization of the distribution  $P_q$  we get

$$P_{\mathbf{0}}^{-1} = \sum_{q=0}^{n_A^0} \binom{n_A^0}{q} \left( \sum_{k=1}^S \frac{\theta_k}{\nu_k} \right)^q = \left( 1 + \sum_{k=1}^S \frac{\theta_k}{\nu_k} \right)^{n_A^0}, \quad (\text{A.9})$$

where in the last equality we have used the binomial theorem.

The joint distribution of the vector of population abundances of different behavioral types,  $(n_A, n_1, \dots, n_S)$ , is given by Eqs. (A.7) and (A.9). It is a multinomial distribution  $\mathcal{M}(n_A^0, \{p_A, p_1, \dots, p_S\})$ , as defined in Eq. (11), with  $n_A^0$  trials and probabilities given by Eqs. (12) and (13), which we reproduce here:

$$p_i = \frac{\theta_i}{\nu_i} \frac{1}{1 + \sum_{k=1}^S \frac{\theta_k}{\nu_k}}$$



for predators handling on resources  $i = 1, \dots, S$ , and

$$p_A = \frac{1}{1 + \sum_{k=1}^S \frac{\theta_k}{\nu_k}}$$

for free predators. Observe that  $p_A + \sum_{k=1}^S p_k = 1$  is satisfied. In addition, Eq. (13) extends naturally, for more than one resource, the probability parameter  $p = (1 + \frac{\theta}{\nu})^{-1}$  of the binomial distribution (A.1).

Observe also that the average population number of each behavioral type is  $\langle n_i \rangle = n_A^0 p_i$ . This immediately leads to the average number of predators feeding on a focal resource of class  $i$ ,

$$\langle n_i \rangle = \frac{n_A^0}{1 + \sum_{k=1}^S \frac{\theta_k}{\nu_k}} \frac{\theta_i}{\nu_i}, \quad (\text{A.10})$$

and a mean number of free predators equal to

$$\langle n_A \rangle = \frac{n_A^0}{1 + \sum_{k=1}^S \frac{\theta_k}{\nu_k}}. \quad (\text{A.11})$$

This automatically yields the multi-resource Holling type II functional response extension (see Palamara et al. (2022) for the derivation using a different methodology).

In addition to the joint, steady-state probability distribution of different behavioral types, by summing over the corresponding variables we can obtain the marginal distributions of each single type, which are binomials  $B(n_A^0, p_i)$  (Loève, 1977). In particular, from our derivation above we find that the distribution of the total number of handling consumers is a binomial  $B(n_A^0, 1 - p_A)$ : using Eqs. (A.9) and (13), the probability that  $\sum_{k=1}^S n_k = q$ , which is given by Eq. (A.8), can be rewritten as

$$P_q = \binom{n_A^0}{q} (1 - p_A)^q p_A^{n_A^0 - q}.$$

This immediately implies that the distribution of the number of free predators,  $n_A = n_A^0 - \sum_{k=1}^S n_k = n_A^0 - q$ , follows a binomial distribution  $B(n_A^0, p_A)$ ,

$$P_{n_A} = \binom{n_A^0}{n_A} p_A^{n_A} (1 - p_A)^{n_A^0 - n_A},$$

as expected (Palamara et al., 2022).

If we were to use the multinomial steady-state distribution (11) with probabilities given by Eqs. (12) and (13), to estimate feeding rates via ML (as above, we assume that  $n_A^0$  is fixed and not subject to MLE), we could write the following exhaustive summary:  $\boldsymbol{\kappa} = (\theta_1/\nu_1, \dots, \theta_S/\nu_S)$ , which is tantamount to  $\boldsymbol{\kappa} = (\alpha_1/\nu_1, \dots, \alpha_S/\nu_S)$  because resource densities  $x_{R_i}$  are kept constant (chemostatic conditions). Since the log-likelihood is determined by  $S$  ratios and there are  $2S$  individual rates  $(\alpha_1, \dots, \alpha_S, \nu_1, \dots, \nu_S)$  to be estimated, the model is structurally unidentifiable (parameter redundant) if MLE is conducted using the steady-state distribution.

## Appendix B. Derivation of the PDE satisfied by the generating function

This Appendix is devoted to derive the PDE (14) satisfied by the generating function (1) in the out-of-equilibrium multi-resource Holling type II case. First we multiply Eq. (10) by  $z_1^{n_1} \dots z_S^{n_S}$  and sum over all non-negative integer vectors  $\mathbf{n}$ . The sum can be split as

$$\frac{\partial G}{\partial t} = A + B,$$

with terms  $A$  and  $B$  defined as

$$A := \sum_{n_1, \dots, n_S \geq 0} \sum_{i=1}^S z_1^{n_1} \cdots z_S^{n_S} [T(\mathbf{n}|\mathbf{n} + \mathbf{e}_i)P_{\mathbf{n}+\mathbf{e}_i} - T(\mathbf{n} - \mathbf{e}_i|\mathbf{n})P_{\mathbf{n}}]$$

and

$$B := \sum_{n_1, \dots, n_S \geq 0} \sum_{i=1}^S z_1^{n_1} \cdots z_S^{n_S} [T(\mathbf{n}|\mathbf{n} - \mathbf{e}_i)P_{\mathbf{n}-\mathbf{e}_i} - T(\mathbf{n} + \mathbf{e}_i|\mathbf{n})P_{\mathbf{n}}],$$

respectively.

First, focus on the  $A$  term and the transition probabilities given by Eq. (9), which are linear functions of the population numbers. In order to evaluate the sums in terms of  $G$  and its partial derivatives, we can use the identity

$$\frac{\partial G}{\partial z_i} = \sum_{n_1, \dots, n_S \geq 0} (n_i + 1)P_{\mathbf{n}+\mathbf{e}_i} z_1^{n_1} \cdots z_S^{n_S},$$

which follows from the definition (1) and the fact that  $P_{-\mathbf{e}_i}(t) = 0$  for every index  $i$ , which has to be imposed for population numbers to remain positive. We also take into account the identity

$$z_i \frac{\partial G}{\partial z_i} = \sum_{n_1, \dots, n_S \geq 0} n_i P_{\mathbf{n}} z_1^{n_1} \cdots z_S^{n_S}, \quad (\text{B.1})$$

which directly follows from the definition of the generating function. Using both identities, we can immediately write the term  $A$  as function of  $z_i$  and the first order partial derivative of the generating function  $G$  with respect to  $z$ , as follows:

$$A = \sum_{n_1, \dots, n_S \geq 0} \sum_{i=1}^S [\nu_i(n_i + 1)P_{\mathbf{n}+\mathbf{e}_i} - \nu_i n_i P_{\mathbf{n}}] z_1^{n_1} \cdots z_S^{n_S} = \sum_{i=1}^S \nu_i (1 - z_i) \frac{\partial G}{\partial z_i}.$$

Second, and taking into account the expressions of transition probabilities given by Eqs. (9), we observe that the  $B$  term can be written as the following sum:

$$B = \sum_{n_1, \dots, n_S \geq 0} \sum_{i=1}^S z_1^{n_1} \cdots z_S^{n_S} \left\{ n_A^0 \theta_i (P_{\mathbf{n}-\mathbf{e}_i} - P_{\mathbf{n}}) + \theta_i \left[ P_{\mathbf{n}} \sum_j n_j - \left( n_i - 1 + \sum_{j \neq i} n_j \right) P_{\mathbf{n}-\mathbf{e}_i} \right] \right\}.$$

The following identities also follow from the definition of the generating function:

$$\begin{aligned} z_i^2 \frac{\partial G}{\partial z_i} &= \sum_{n_1, \dots, n_S \geq 0} (n_i - 1) P_{\mathbf{n}-\mathbf{e}_i} z_1^{n_1} \cdots z_S^{n_S}, \\ z_i z_j \frac{\partial G}{\partial z_j} &= \sum_{n_1, \dots, n_S \geq 0} n_j P_{\mathbf{n}-\mathbf{e}_i} z_1^{n_1} \cdots z_S^{n_S} \quad (i \neq j). \end{aligned}$$

They are based also in the property  $P_{-\mathbf{e}_i}(t) = 0$  for all  $i$ , which must hold for all  $i$  so that configurations with negative populations are assigned probability zero. Using these two identities and (B.1), together with

$$z_i G = \sum_{n_1, \dots, n_S \geq 0} P_{\mathbf{n}-\mathbf{e}_i} z_1^{n_1} \cdots z_S^{n_S},$$

which follows also from  $P_{-\mathbf{e}_i}(t) = 0$ , we find an expression for the  $B$  term in terms of  $G$  and its derivatives,

$$B = \sum_{i=1}^S \theta_i (1 - z_i) \left( -n_A^0 G + \sum_{j=1}^S z_j \frac{\partial G}{\partial z_j} \right).$$

Gathering the contributions of  $A$  and  $B$  terms, we obtain the PDE (14) satisfied by the generating function  $G(\mathbf{z}, t)$ .

## Appendix C. ML estimation with equal sampling times

In this section we show that sampling data at equal times,  $t_1 = \dots = t_T = t$  ( $t > 0$ ), makes the model unidentifiable. Here we write the multinomial distribution as

$$P_{\mathbf{n}}(n_A^0, \{p_i\}) = C(\mathbf{n}) \left(1 - \sum_{i=1}^S p_i\right)^{n_A} \prod_{i=1}^S p_i^{n_i},$$

where  $\mathbf{n} = (n_1, \dots, n_S)$  is the vector of handling predators abundances,  $C(\mathbf{n})$  is the multinomial coefficient,  $n_A = n_A^0 - \sum_{j=1}^S n_j$  is the number of free predators, and probability parameters  $p_j$  are given, for  $j = 1, \dots, S$ , by Eq. (29). Assume equal rates,  $\alpha_j = \alpha$  and  $\nu_j = \nu$  for  $j = 1, \dots, S$ . Then the multinomial probability parameters are expressed as

$$p_j(t) = \frac{x_{R_j} \alpha}{\nu} \cdot \frac{1 - e^{-(\nu + \alpha \Lambda_R)t}}{1 + \alpha \Lambda_R / \nu}.$$

Here we have defined the total resource density  $\Lambda_R := \sum_{i=1}^S x_{R_i}$ . Then we can write  $p_i = x_{R_i} h(t; \alpha, \tau)$ , where

$$h(t; \alpha, \tau) = \alpha \tau (1 - e^{-t/\tau})$$

and  $\tau = (\nu + \alpha \Lambda_R)^{-1}$  is the characteristic time scale.

The log-likelihood function is expressed as

$$\log \mathcal{L}(\alpha, \nu | M(t)) = T \log C(\mathbf{n}) + \sum_{k=1}^T \left\{ n_{AR,k} \log h(t; \alpha, \tau) + n_{A,k} \log(1 - \Lambda_R h(t; \alpha, \tau)) + \sum_{i=1}^S n_{i,k} \log x_{R_i} \right\},$$

where we have used that  $p_i = x_{R_i} h(t; \alpha, \tau)$  and defined the  $k$ -th sampled total handling predators abundance as the sum  $n_{AR,k} = \sum_{i=1}^S n_{i,k}$ , and the  $k$ -th sampled number of free predators as  $n_{A,k} = n_A^0 - \sum_{i=1}^S n_{i,k} = n_A^0 - n_{AR,k}$  —the last equality holds for every sample,  $k = 1, \dots, T$ . Observe that the dependence on model parameters  $\alpha$  and  $\nu$  has been isolated in function  $h$ . Therefore, the first and last terms in the log-likelihood function do not contribute to derivatives with respect to those parameters.

Thus, the critical points of the log-likelihood satisfy the equations

$$\begin{aligned} \frac{\partial}{\partial \alpha} \log \mathcal{L} &= \frac{\partial h}{\partial \alpha} \sum_{k=1}^T \left( \frac{n_{AR,k}}{h} - \frac{n_{A,k} \Lambda_R}{1 - \Lambda_R h} \right) = 0, \\ \frac{\partial}{\partial \nu} \log \mathcal{L} &= \frac{\partial h}{\partial \nu} \sum_{k=1}^T \left( \frac{n_{AR,k}}{h} - \frac{n_{A,k} \Lambda_R}{1 - \Lambda_R h} \right) = 0. \end{aligned}$$

Then one can check that the system formed by  $\frac{\partial h}{\partial \alpha} = 0$  and  $\frac{\partial h}{\partial \nu} = 0$  has no solution satisfying at the same time the obvious restrictions  $\alpha, \nu, t > 0$ . Therefore the two conditions for critical points reduces to one single equation,

$$\sum_{k=1}^T \left( \frac{n_{AR,k}}{h} - \frac{n_{A,k} \Lambda_R}{1 - \Lambda_R h} \right) = 0,$$

from which we can eliminate the function  $h(t; \alpha, \tau)$  to obtain the condition to be satisfied by model parameters at the critical points. This defines a manifold in the parameter space, given by

$$\frac{\alpha [1 - e^{-(\nu + \alpha \Lambda_R)t}]}{\nu + \alpha \Lambda_R} = \frac{\langle n_{AR} \rangle_T}{\Lambda_R n_A^0}, \quad (\text{C.1})$$

where we have defined  $\langle n_{AR} \rangle_T := \frac{1}{T} \sum_{k=1}^T n_{AR,k}$  as the average across samples of handling consumers' total abundance. The quantity  $\langle n_{AR} \rangle_T$ , relative to the product  $\Lambda_R n_A^0$ , determines only one single restriction to

be satisfied by model parameters, in the  $(\alpha, \nu)$  space. This immediately implies that the two parameters cannot be uniquely determined, yielding (for every  $t > 0$ ) an infinite number of critical points satisfying Eq. (C.1), as we wanted to prove. In the case of exactly equal sampling times (either out of equilibrium or at stationarity), the model remains set-identifiable. At least two sampling times must be distinct in order for  $\alpha$  and  $\nu$  being non-redundant.

Finally, observe that, in the limit  $t \rightarrow \infty$ , the manifold defined by Eq. (C.1) reduces to the straight line whose intercept is equal to zero and whose slope is

$$\frac{\nu}{\alpha} = \frac{\langle n_A \rangle_T}{\langle n_{AR} \rangle_T} \Lambda_R,$$

as confirmed by the numerical optimization of the equilibrium log-likelihood function —here  $\langle n_A \rangle_T := \frac{1}{T} \sum_{k=1}^T n_{A,k} = n_A^0 - \langle n_{AR} \rangle_T$ . The slope of the line of infinite maxima is determined by the average number of free predators across samples, relative to the average total number of handling predators, times the overall density of resources. This is consistent with Eq. (31) if, at equilibrium,  $\langle n_A \rangle$  and  $\langle n_{AR} \rangle$  are estimated from data by their corresponding averages across samples, i.e., by  $\langle n_A \rangle_T$  and  $\langle n_{AR} \rangle_T$ , respectively.

## Appendix D. Macroscopic model dynamics

In this section we integrate the macroscopic dynamics

$$\frac{dx_i}{dt} = -\nu x_i + \theta_i \left( n_A^0 - \sum_{j=1}^S x_j \right),$$

for  $x_i = \langle n_i \rangle$  the mean-field abundance of consumers feeding on resource  $i = 1, \dots, S$ . Once the abundances are solved as function of  $t$ , one can obtain the average number of free consumers as  $x_A = \langle n_A \rangle = n_A^0 - \sum_{i=1}^S x_i$ . Initially, all handling predator abundances are set equal to zero ( $x_i(0) = 0$ ), which means that all consumers are initially free,  $x_A(0) = n_A^0$ .

This model is linear and can be written in a compact form as

$$\frac{d\mathbf{x}}{dt} = n_A^0 \boldsymbol{\theta} - J\mathbf{x},$$

for vector  $\boldsymbol{\theta} = (\theta_i) = \alpha(x_{R_i})$ , matrix  $J = \nu I + \boldsymbol{\theta} \mathbf{1}^T$  (which, as a rank-one perturbation of a diagonal matrix, is invertible by the Sherman-Morrison's formula) and initial condition  $\mathbf{x}(0) = \mathbf{0}$ . The equilibrium abundances are given by  $\mathbf{x}^* = n_A^0 J^{-1} \boldsymbol{\theta}$ . If we write  $\mathbf{y} = \mathbf{x} - \mathbf{x}^*$  then the dynamics reduces to  $\frac{d\mathbf{y}}{dt} = -J\mathbf{y}$  with initial condition  $\mathbf{y}(0) = -\mathbf{x}^*$ . In that form, it can be solved by matrix exponentiation,

$$\mathbf{y}(t) = e^{-Jt} \mathbf{y}(0) = -n_A^0 e^{-Jt} J^{-1} \boldsymbol{\theta}.$$

Thus, in terms of the original variables  $\mathbf{x}$ , we obtain

$$\mathbf{x}(t) = \mathbf{y}(t) + \mathbf{x}^* = n_A^0 (I - e^{-Jt}) J^{-1} \boldsymbol{\theta}. \quad (\text{D.1})$$

The problem is reduced to calculate the exponential  $e^{-Jt}$  and the inverse  $J^{-1}$ .

In order to sum the exponential series, we write  $J = \nu H$  and compute the powers of  $H = I + \mathbf{h} \mathbf{1}^T$  with  $\mathbf{h} = \nu^{-1} \boldsymbol{\theta}$ . Given that  $H$  is a rank-one perturbation of the identity, we expect that

$$H^n = I + c_n \mathbf{h} \mathbf{1}^T,$$

for a sequence  $c_n$  to be determined, which obviously satisfies  $c_1 = 1$ . Now compute  $H^{n+1}$  as

$$H^{n+1} = (I + c_n \mathbf{h} \mathbf{1}^T)(I + \mathbf{h} \mathbf{1}^T) = I + ((\mathbf{1}^T \mathbf{h} + 1)c_n + 1) \mathbf{h} \mathbf{1}^T,$$

from which we obtain the recurrence  $c_{n+1} = (\mathbf{1}^T \mathbf{h} + 1)c_n + 1$  to be solved with initial condition  $c_1 = 1$ . Using that  $\mathbf{1}^T \mathbf{h} = \nu^{-1} \mathbf{1}^T \boldsymbol{\theta} = Q/\nu$ , and solving the recurrence for coefficients  $c_n$ , we obtain

$$J^n = \nu^n I + \frac{(\nu + Q)^n - \nu^n}{Q} \boldsymbol{\theta} \mathbf{1}^T.$$

Then, by summing the series, it is straightforward to write the exponential of  $-tJ$  as, again, a rank one perturbation of a diagonal matrix:

$$e^{-Jt} = e^{-\nu t} I + \frac{1}{Q} \left( e^{-(\nu+Q)t} - e^{-\nu t} \right) \boldsymbol{\theta} \mathbf{1}^T. \quad (\text{D.2})$$

On the other hand, by the Sherman-Morrison's formula,

$$J^{-1} = \nu^{-1} (I - (\nu + Q)^{-1} \boldsymbol{\theta} \mathbf{1}^T),$$

and therefore  $J^{-1} \boldsymbol{\theta} = \frac{1}{\nu+Q} \boldsymbol{\theta}$ . Substituting this result and Eq. (D.2) into Eq. (D.1) we finally obtain

$$\mathbf{x}(t) = \frac{n_A^0}{\nu + Q} (1 - e^{(\nu+Q)t}) \boldsymbol{\theta}. \quad (\text{D.3})$$

The mean abundance of free predators is  $x_A(t) = n_A^0 - \mathbf{1}^T \mathbf{x}(t)$ :

$$x_A(t) = \frac{n_A^0}{\nu + Q} (\nu + Q e^{(\nu+Q)t}) = n_A^0 (\tau \nu + (1 - \tau \nu) e^{-t/\tau}), \quad (\text{D.4})$$

for  $\tau = (\nu + Q)^{-1}$ . Observe that, asymptotically,  $\lim_{t \rightarrow \infty} \mathbf{x}(t) = \frac{n_A^0}{\nu+Q} \boldsymbol{\theta}$  and  $\langle n_A \rangle_\infty = \lim_{t \rightarrow \infty} x_A(t) = \frac{n_A^0 \nu}{\nu+Q}$ . The total abundance of handling predators at equilibrium,  $\langle n_{AR} \rangle_\infty = \lim_{t \rightarrow \infty} \mathbf{1}^T \mathbf{x}(t) = \frac{n_A^0 Q}{\nu+Q}$ . Naturally, because  $Q = \alpha \sum_{i=1}^S x_{R_i}$ , the expressions obtained for the averages  $\langle n_A \rangle_\infty$  and  $\langle n_{AR} \rangle_\infty$  satisfy Eq. (31). Thus, the ratio  $\nu/\alpha$  can be estimated using steady-state averages. Additional information is necessary to fully determine individual rates  $\alpha$  and  $\nu$ .

A possible solution would be having external information about the characteristic time of the dynamics,  $\tau$ . Here we propose using the initial decay of the number of free predators as function of  $t$ . For  $t \ll \tau$ , we can approximate  $x_A(t) = n_A^0 (1 - Qt)$ , so we can use the slope of  $x_A(t)$  at  $t = 0$  to determine  $\alpha$  via  $Q = \alpha \sum_{i=1}^S x_{R_i}$ . We have done so using stochastic realizations. and we estimated  $\alpha$  using the slope at  $t = 0$  of each realization. We generated long-time realizations to estimate the ratio  $\nu/\alpha$  from Eq. (31), i.e.,  $\nu/\alpha = (\langle n_A \rangle_\infty / \langle n_{AR} \rangle_\infty) \sum_{i=1}^R x_{R_i}$ . Results are reported in Fig. D.1. Observe that estimates obtained from this approach are more biased than the estimates obtained by maximum likelihood (compare with Fig. 3), yielding very poor estimates for  $n_A^0$  small. In addition, the width of the parameter distributions does not reduce as  $n_A^0$  increases. This is not surprising because ML estimators are, under natural regularity conditions, those with minimal variance. More importantly, the estimation procedure cannot be based solely on steady-state abundances: additional information is needed from an out-of-equilibrium regime (in this case, the rate at which the abundance of free predators decays for shorter times).

## References

- Peter A Abrams. Food web functional responses. *Frontiers in Ecology and Evolution*, 10:984384, 2022.
- Matthieu Barbier, Laurie Wojcik, and Michel Loreau. A macro-ecological approach to predation density-dependence. *Oikos*, 130(4):553–570, 2021.
- F. Bartumeus, J. Catalan, U. L. Fulco, M. L. Lyra, and G. M. Viswanathan. Optimizing the encounter rate in biological interactions: Lévy versus brownian strategies. *Phys. Rev. Lett.*, 88:097901, Feb 2002.

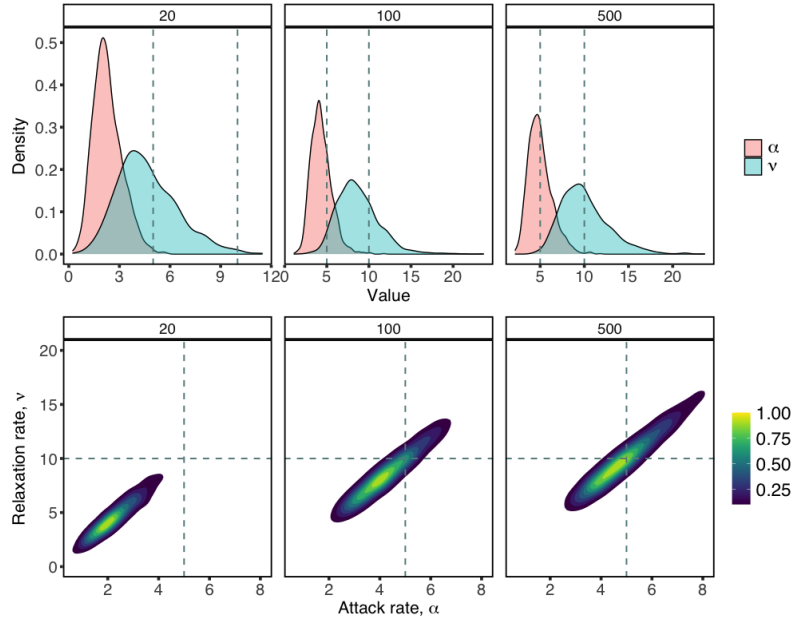


Figure D.1: **Distributions of parameter estimates based on the deterministic dynamics, for different number of initial consumers.** Same as Fig. 3 for a estimation procedure based on the deterministic model. We generated 1500 stochastic realizations from  $t = 0$  to  $t = 100\tau$ . Then the slope at  $t = 0$  in each stochastic realization, which equals  $Q = \alpha \sum_{i=1}^S x_{R_i}$ , was used to estimate  $\alpha$ . For that purpose, we fit a linear model to the function  $n_A(t)/n_A^0$  using the first 20 times in each realization. Second, we used the asymptotic values  $\langle n_A \rangle_\infty$  and  $\langle n_{AR} \rangle_\infty$  to estimate the ratio  $\nu/\alpha$ : from each realization we average  $n_A(t)$  and  $n_{AR}(t)$  neglecting the first 500 temporal points, enough to reach stationarity. Compared to distributions in Fig. 3, we obtain with this method wider distributions and more biased estimates, especially for small  $n_A^0$ . Parameters are the same as in Fig. 1.

Frederic Bartumeus, Francesc Peters, Salvador Pueyo, Cèlia Marrasé, and Jordi Catalan. Helical Lévy walks: adjusting searching statistics to resource availability in microzooplankton. *Proceedings of the National Academy of Sciences of the United States of America*, 100(22):12771–12775, 2003.

Frederic Bartumeus, M. G E. da Luz, G. M. Viswanathan, and J. Catalan. Animal search strategies: A quantitative random-walk analysis. *Ecology*, 86(11):3078–3087, 2005.

Andrea Belgrano. *Aquatic food webs: an ecosystem approach*. Oxford University Press, 2005.

Ror Bellman and Karl Johan Åström. On structural identifiability. *Mathematical biosciences*, 7(3-4): 329–339, 1970.

Alan A Berryman. The origins and evolution of predator-prey theory. *Ecology*, 73(5):1530–1535, 1992.

Louis-Félix Bersier, Carolin Banašek-Richter, and Marie-France Cattin. Quantitative descriptors of food-web matrices. *Ecology*, 83(9):2394–2407, 2002.

Mario Castro and Rob J De Boer. Testing structural identifiability by a simple scaling method. *PLOS Computational Biology*, 16(11):e1008248, 2020.

Gregory C Chow. Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society*, pages 591–605, 1960.

Diana Cole. *Parameter redundancy and identifiability*. Chapman and Hall/CRC, 2020.

David R. Cox and E. Joyce Snell. *Analysis of binary data*, volume 32. CRC press, 1989.

- John P DeLong. *Predator ecology: Evolutionary ecology of the functional response*. Oxford University Press, 2021.
- Jennifer A Dunne. The network structure of food webs. *Ecological networks: linking structure to dynamics in food webs*, pages 27–86, 2006.
- Rampal S Etienne and David Alonso. Neutral community theory: how stochasticity and dispersal-limitation can explain species coexistence. *Journal of Statistical Physics*, 128(1):485–510, 2007.
- Wayne M Getz. Biomass transformation webs provide a unified approach to consumer–resource modelling. *Ecology letters*, 14(2):113–124, 2011.
- Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361, 1977.
- Daniel T. Gillespie. *Markov processes: an introduction for physical scientists*. Elsevier, 1991.
- Dirk Helbing, Dirk Brockmann, Thomas Chadeaux, Karsten Donnay, Ulf Blanke, Olivia Woolley-Meza, Mehdi Moussaid, Anders Johansson, Jens Krause, Sebastian Schutte, et al. Saving human lives: What complexity science and information systems can contribute. *Journal of Statistical Physics*, 158:735–781, 2015.
- C.S. Holling. The components of predation as revealed by a study of small-mammal predation of the european pine sawfly. *The Canadian Entomologist*, 91(5):293–320, 1959a.
- C.S. Holling. Some characteristics of simple types of predation and parasitism. *The Canadian Entomologist*, 91(7):385–398, 1959b.
- Marko Jusup, Petter Holme, Kiyoshi Kanazawa, Misako Takayasu, Ivan Romić, Zhen Wang, Sunčana Geček, Tomislav Lipić, Boris Podobnik, Lin Wang, et al. Social physics. *Physics Reports*, 948:1–148, 2022.
- Matthew J. Keeling and Joshua V. Ross. Efficient methods for studying stochastic disease and population dynamics. *Theoretical population biology*, 75(2-3):133–141, 2009.
- David G. Kendall. Stochastic processes and population growth. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(2):230–282, 1949.
- Donald E. Knuth. *The art of computer programming. Volume 1: Fundamental Algorithms*, volume 1. Pearson Education, 1997.
- Mariano Koen-Alonso. A process-oriented approach to the multispecies functional response. In *From energetics to ecosystems: the dynamics and structure of ecological systems*, pages 1–36. Springer, 2007.
- Kevin D. Lafferty, Giulio DeLeo, Cheryl J. Briggs, Andrew P. Dobson, Thilo Gross, and Armand M. Kuris. A general consumer-resource population model. *Science*, 349(6250):854–857, 2015.
- Erich L. Lehmann and George Casella. *Theory of point estimation*. Springer Science & Business Media, 2006.
- Michel Loève. *Elementary probability theory*. Springer, 1977.
- Mark A. McPeck. Mechanisms influencing the coexistence of multiple consumers and multiple resources: resource and apparent competition. *Ecological Monographs*, 89(1):e01328, 2019.
- Roger M. Nisbet and William S.C. Gurney. *Modelling Fluctuating Populations*. Blackburn Press, 2004. ISBN 9781930665903.

- Mark Novak and Daniel B. Stouffer. Systematic bias in studies of consumer functional responses. *Ecology Letters*, 24:580–593, 2021.
- Gian Marco Palamara, José A. Capitán, and David Alonso. The stochastic nature of functional responses. *Entropy*, 23(5), 2021.
- Gian Marco Palamara, José A. Capitan, and David Alonso. The implicit assumptions of classic functional responses and their multi-species extensions. *bioRxiv*, pages 2022–07, 2022.
- Daniel M. Perkins, Ian A. Hatton, Benoit Gauzens, Andrew D. Barnes, David Ott, Benjamin Rosenbaum, Catarina Vinagre, and Ulrich Brose. Consistent predator-prey biomass scaling in complex food webs. *Nature Communications*, 13(1):4990, 2022.
- Leslie A. Real. The kinetics of functional response. *The American Naturalist*, 111(978):289–300, 1977.
- Matthew D. Robertson, Mariano Koen-Alonso, Paul M. Regular, Noel Cadigan, and Fan Zhang. Accounting for a nonlinear functional response when estimating prey dynamics using predator diet data. *Methods in Ecology and Evolution*, 13(4):880–893, 2022.
- David Rogers. Random search and insect population models. *The Journal of Animal Ecology*, pages 369–383, 1972.
- Benjamin Rosenbaum and Björn C. Rall. Fitting functional responses: Direct parameter estimation by simulating differential equations. *Methods in Ecology and Evolution*, 9(10):2076–2090, 2018.
- Daniel B. Stouffer and Mark Novak. Hidden layers of density dependence in consumer feeding rates. *Ecology Letters*, 24(3):520–532, 2021.
- Fernanda S. Valdovinos, Kayla R.S. Hale, Sabine Dritz, Paul R. Glaum, Kevin S. McCann, Sophia M. Simon, Elisa Thébault, William C. Wetzal, Kate L. Wootton, and Justin D. Yeakel. A bioenergetic framework for aboveground terrestrial food webs. *Trends in ecology & evolution*, 2022.
- Alejandro F. Villaverde, Antonio Barreiro, and Antonis Papachristodoulou. Structural identifiability of dynamic systems biology models. *PLoS computational biology*, 12(10):e1005153, 2016.
- Franz-Georg Wieland, Adrian L. Hauber, Marcus Rosenblatt, Christian Tönsing, and Jens Timmer. On structural and practical identifiability. *Current Opinion in Systems Biology*, 25:60–69, 2021.
- Ciyu Zhu, Richard H. Byrd, Peihuang Lu, and Jorge Nocedal. Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on mathematical software (TOMS)*, 23(4):550–560, 1997.